

UNIVERSIDAD DE SALAMANCA
FACULTAD DE TRADUCCIÓN Y DOCUMENTACIÓN
GRADO EN TRADUCCIÓN E INTERPRETACIÓN
Trabajo de Fin de Grado

LOST IN TRANSCRIPTION: LA TRADUCCIÓN DEL HUMOR BRITÁNICO EN LOS SUBTÍTULOS AUTOMÁTICOS DE YOUTUBE

Análisis de la calidad del reconocimiento
automático del habla y la traducción automática
neuronal en la subtitulación de vídeos con
elevada carga cultural

Autora:

Lorena Sánchez Díaz

Tutor:

Jesús Torres del Rey

Salamanca, 2025

*Humor is the first of the gifts to
perish in a foreign tongue*

Virginia Woolf

RESUMEN

El subtítulo automático se ha convertido en una herramienta habitual para acceder a contenidos en otros idiomas, pero su eficacia en situaciones comunicativas complejas sigue siendo motivo de debate. Este estudio se centra en analizar la calidad de los subtítulos en español generados automáticamente por YouTube en vídeos humorísticos británicos, un tipo de contenido especialmente exigente por su uso del lenguaje figurado, referencias culturales y acentos regionales. Para ello, se han seleccionado tres vídeos cortos y se han analizado dos fases clave del sistema: la transcripción automática del audio (ASR) y su posterior traducción al español mediante redes neuronales (NMT). La metodología combina un modelo de evaluación basado en MQM con una codificación específica para los errores de transcripción. Los resultados muestran una notable acumulación de errores que afectan directamente a la comprensión del mensaje y a la transmisión del humor. La traducción automática, lejos de corregirlos, tiende a mantener los fallos previos. Aunque los subtítulos permiten hacerse una idea general del contenido, no son suficientes para captar los matices del original ni para garantizar una experiencia comunicativa completa. Todo ello plantea dudas sobre la fiabilidad de estos sistemas en contextos que van más allá de lo puramente informativo y sugiere la necesidad de intervención humana en materiales que exigen sensibilidad lingüística y cultural.

Palabras clave: subtítulo automático, reconocimiento del habla, traducción automática, humor, YouTube, MQM, errores de traducción.

ABSTRACT

Automatic subtitling has become a common tool for accessing content in other languages, its effectiveness in complex communicative contexts remains controversial. This study examines the quality of Spanish subtitles automatically generated by YouTube for British humorous videos—a particularly challenging type of content due to its figurative language, cultural references, and regional accents. Three clips were selected and analyzed through two key stages: automatic speech recognition (ASR) and subsequent neural machine translation (NMT) into Spanish. The methodology combines an evaluation model based on MQM with a specific coding system for transcription errors.

The results reveal a significant number of inaccuracies that directly affect both message comprehension and the transmission of humor. Instead of correcting these issues, the translation stage tends to reproduce them. While the subtitles may offer a general idea of the content, they fall short of conveying the original nuance and do not ensure a complete communicative experience. These findings raise concerns about the reliability of such systems in contexts that go beyond purely informative content and highlight the ongoing need for human involvement in materials requiring linguistic and cultural sensitivity.

Keywords: automatic subtitling, speech recognition, machine translation, humor, YouTube, MQM, translation errors.

ÍNDICE

LISTA DE SIGLAS EMPLEADAS	3
INTRODUCCIÓN	4
1. Motivación para el tema	4
2. Criterios de selección inicial y proceso de búsqueda.....	4
3. Observaciones iniciales.....	5
4. Objetivos	5
5. Hipótesis	6
6. Estructura del trabajo	6
MARCO TEÓRICO	7
1. El subtitulado	7
2. La traducción del humor en el subtitulado.....	8
3. Tecnologías del procesamiento automático del lenguaje.....	9
3.1. La traducción automática	9
3.2. El reconocimiento automático del habla	11
4. El subtitulado automático de YouTube	12
5. La evaluación de la calidad de los subtítulos.....	13
MARCO APLICADO	14
1. Plataforma	14
2. Vídeos seleccionados	14
3. Modelo de análisis	16
3.1. Modelo MQM	16
3.2. Códigos específicos para ASR	21
4. Metodología	22
4.1. Extracción de los datos del corpus	22
4.2. Procedimiento de análisis.....	23
5. Análisis y resultados	27

5.1. Vídeo nº1	30
5.2. Vídeo nº2	39
5.3. Vídeo nº3	45
6. Observaciones adicionales	52
7. Análisis comparativo final	53
CONCLUSIONES.....	55
BIBLIOGRAFÍA.....	56
ANEXOS.....	60
Anexo I. Hoja de análisis vídeo 1 (Excel)	60
Anexo II. Hoja de análisis vídeo 2 (Excel)	61
Anexo III. Hoja de análisis vídeo 3 (Excel).....	62
Anexo IV. Acceso a los documentos completos del análisis	63

LISTA DE SIGLAS EMPLEADAS

ADC	<i>Analog-to-Digital Converter</i> (convertidor analógico a digital)
API	<i>Application Programming Interface</i> (interfaz de programación de aplicaciones)
ASR	<i>Automatic Speech Recognition</i> (reconocimiento automático del habla)
BBC	British Broadcasting Corporation
LLM	<i>Large Language Models</i> (modelos de lenguaje masivos)
LM	Lengua Meta
LO	Lengua Origen
MQM	<i>Multidimensional Quality Metrics</i> (métricas de calidad multidimensional)
MT	<i>Machine Translation</i> (traducción automática)
NMT	<i>Neural Machine Translation</i> (traducción automática neuronal)
OBS	<i>Open Broadcaster Software</i> (software de grabación y transmisión en directo)
RBMT	<i>Rule-based Machine Translation</i> (traducción automática basada en reglas)
RNN	<i>Recurrent Neural Networks</i> (redes neuronales recurrentes)
SMT	<i>Statistical Machine Translation</i> (traducción automática estadística)
TA	Traducción Automática
TAN	Traducción Automática Neuronal
TAV	Traducción Audiovisual
TM	Texto Meta
WER	<i>Word Error Rate</i> (tasa de error de palabra)

INTRODUCCIÓN

1. Motivación para el tema

En los últimos años, el subtitulado automático ha adquirido una creciente relevancia en el ámbito de la traducción audiovisual, especialmente con la expansión de plataformas como YouTube. La posibilidad de generar subtítulos al momento y en varios idiomas lo convierte en una herramienta con un gran potencial, tanto para mejorar la accesibilidad como para facilitar que los contenidos lleguen a audiencias internacionales. Sin embargo, este avance tecnológico plantea dudas sobre la calidad real de los subtítulos producidos de forma automática, sobre todo en contextos lingüísticos complejos.

El presente trabajo surge del interés por comprobar hasta qué punto esta tecnología puede facilitar la comprensión de vídeos en lenguas extranjeras, en este caso, del inglés al español, sin necesidad de escuchar el audio original. Dada la complejidad que conlleva el humor, con su carga cultural, referencias implícitas y juegos de palabras, se considera un terreno especialmente exigente y, por tanto, idóneo para poner a prueba la fiabilidad del sistema de subtitulado automático de YouTube.

2. Criterios de selección inicial y proceso de búsqueda

Para abordar esta cuestión, se definieron previamente una serie de criterios fijos de búsqueda orientados a localizar vídeos humorísticos en inglés británico con una fuerte carga cultural y lingüística. Se estableció que los vídeos debían suponer un reto para los sistemas automáticos de reconocimiento del habla (ASR) y traducción automática neuronal (NMT). En particular en lo relativo a acentos marcados, ritmo acelerado, expresiones idiomáticas, diferencias sociolingüísticas y referencias culturales específicas.

El proceso de búsqueda se realizó dentro de la propia plataforma de YouTube, por ser ampliamente utilizada, de fácil acceso y que permite consultar sus propios subtítulos automáticos. Inicialmente, se preseleccionaron varios vídeos que parecían adecuados según los criterios mencionados anteriormente. No obstante, tras una revisión más detallada, solo tres de ellos, emitidos originalmente por la BBC, cumplían con todos los requisitos establecidos, por lo que fueron finalmente seleccionados para el análisis.

3. Observaciones iniciales

Una vez seleccionados los vídeos, se llevó a cabo una revisión preliminar del material escogido. Durante esta primera visualización, se detectaron ciertas deficiencias que despertaron el interés por profundizar en el análisis, ya que los subtítulos automáticos presentaban incoherencias, omisiones y expresiones poco naturales, tanto en inglés como en español. Estas primeras impresiones permitieron intuir que los sistemas ASR y NMT de YouTube aún tienen dificultades para captar elementos esenciales del mensaje original, especialmente cuando el contenido se apoya en recursos como la ironía, el doble sentido o los contrastes dialectales.

Otro aspecto que llamó la atención fue cómo algunos errores detectados en la transcripción se trasladaban directamente a la traducción, lo que hizo pensar que la calidad de la transcripción podía estar influyendo directamente en el resultado final. Esta posible relación entre ambas fases planteó una línea de análisis que se exploraría con más detalle a lo largo del trabajo.

4. Objetivos

El objetivo principal de este trabajo es evaluar la calidad de los subtítulos automáticos generados por YouTube en vídeos originalmente en inglés. El análisis se centra tanto en la transcripción en la lengua origen como en su posterior traducción automática al español, en casos donde el contenido presenta un alto grado de dificultad por motivos lingüísticos y culturales. Para ello, se analizan tanto el reconocimiento automático del habla (ASR) como la traducción automática neuronal (NMT)

De forma más específica, se plantean los siguientes objetivos:

1. **Analizar la precisión del sistema de reconocimiento automático del habla (ASR)**, identificando los errores más comunes, sus posibles causas y su gravedad dentro del discurso.
2. **Evaluar la calidad de la traducción automática (NMT)** del inglés al español, prestando especial atención a la fidelidad del mensaje, la naturalidad del lenguaje, la correcta transmisión de referencias culturales y humorísticas, así como al impacto que tienen los errores según su nivel de gravedad.

3. **Determinar si los subtítulos automáticos traducidos permiten una comprensión autónoma del contenido**, sin necesidad de recurrir al audio original.
4. **Establecer una tipología de errores** basada en modelos de calidad de la traducción, adaptados a los requisitos específicos.
5. **Reflexionar sobre las limitaciones y el potencial de mejora** de los sistemas de subtitulado automático, valorando su utilidad real para el público.

5. Hipótesis

Este estudio parte de la idea de que los subtítulos generados automáticamente por YouTube, mediante el ASR y la NMT, probablemente no ofrecerán una traducción lo bastante precisa ni contextualizada cuando se aplican a vídeos con carga humorística y cultural elevada. Se prevé que los errores no solo afecten a la fidelidad del contenido, sino que comprometan la inteligibilidad del mensaje, especialmente en aquellos casos donde la interpretación del chiste o la referencia cultural dependa del tono, la pronunciación o la estructura pragmática del discurso. En consecuencia, se espera que, si bien estos subtítulos pueden ofrecer una comprensión general del contenido, resulten insuficientes para captar por completo el sentido del mensaje sin escuchar el audio.

6. Estructura del trabajo

A partir de estos planteamientos iniciales, el presente trabajo se organiza en varios apartados. En primer lugar, se presenta el marco teórico que sustenta el análisis, seguido de una descripción detallada de la metodología utilizada. A continuación, se expone el análisis de cada uno de los vídeos seleccionados, así como un apartado comparativo que sintetiza los hallazgos más significativos. Por último, se discuten las principales conclusiones del estudio y se plantean posibles líneas de investigación futuras.

MARCO TEÓRICO

1. El subtítulo

El subtítulo es una modalidad de la traducción audiovisual (TAV) definida por Díaz Cintas (2003, 32) como:

una práctica lingüística que consiste en ofrecer, generalmente en la parte inferior de la pantalla, un texto escrito que pretende dar cuenta de los diálogos de los actores, así como de aquellos elementos discursivos que forman parte de la fotografía (cartas, pintadas, leyendas, pancartas, etc.) o de la pista sonora (canciones, voces en off, etc.) [...]. Los subtítulos han de estar sincronizados con la imagen y los diálogos, deben ofrecer un recuento semántico adecuado de los mismos y permanecer en pantalla el tiempo suficiente para que los espectadores puedan leerlos.

Talaván, Ávila-Cabrera y Costal (2016) destacan que el subtítulo no solo cumple una función traductora, sino también de accesibilidad, permitiendo que personas con discapacidad auditiva accedan a los contenidos.

El subtítulo puede clasificarse en función de la relación entre la lengua original del contenido y la lengua de los subtítulos. Según Díaz Cintas y Remael (2021, 9-12), las dos modalidades principales son la subtitulación intralingüística, en la que se transfiere el contenido oral al modo escrito dentro de la misma lengua, y la subtitulación interlingüística, en la que los subtítulos se traducen a otro idioma.

En los últimos años, se ha extendido el uso de la subtitulación automática, generada por sistemas de inteligencia artificial sin intervención humana. Esta modalidad combina tecnologías como el reconocimiento automático del habla (ASR) y la traducción automática neuronal (NMT). Según Georgakopoulou (2020), plataformas como YouTube aplican estos sistemas para ofrecer subtítulos generados automáticamente a gran escala, aunque su fiabilidad varía en función de diferentes factores.

El subtítulo profesional se rige por normas técnicas estandarizadas que garantizan calidad y accesibilidad. Tal y como indican Díaz Cintas y Remael (2021, 98), los subtítulos deben presentarse en un máximo de dos líneas con 35-42 caracteres por línea, manteniendo una sincronización precisa con el diálogo (1-6 segundos en pantalla) y una velocidad de lectura de 12-17 caracteres por segundo (Talaván *et al.*, 2016, 79). La puntuación incluye guiones para cambios de hablante y cursiva para elementos no

verbales. La identificación de hablantes se marca cuando no es evidente (ibid., 86). Estos estándares, combinados con técnicas de condensación lingüística y adaptación cultural (Díaz Cintas y Remael 2021), aseguran que los subtítulos cumplan su función comunicativa sin saturar al espectador.

2. La traducción del humor en el subtitulado

Cuando el contenido que se subtitula incluye elementos humorísticos, el proceso traductológico se complica aún más, ya que el humor introduce retos adicionales vinculados tanto a la comprensión como a la adaptación cultural y textual. En primer lugar, es necesario aclarar qué se entiende por humor, ya que este fenómeno social y universal está presente en todas las culturas y épocas, y juega un papel fundamental en la creación de lazos y en la facilitación de la comunicación (Vaz de Almeida 2020, citado en Davila *et al.* 2022, 63). El humor puede entenderse como cualquier elemento textual orientado a provocar la risa o la diversión en los receptores, dentro de un proceso comunicativo (Zabalbeascoa 2001, 251). Sin embargo, su naturaleza es compleja, ya que, como señala Zabalbeascoa (ibid.), intervienen de forma simultánea factores lingüísticos, culturales, sociales y psicológicos.

El proceso de traducir el humor en productos audiovisuales suele implicar distintas fases: la detección del intertexto humorístico, la documentación para entender la referencia y, finalmente, la traducción, donde se debe decidir cómo transferir la comicidad y la intertextualidad (Botella Tejera 2017, 91). En el caso del subtitulado, el traductor se enfrenta, además, a limitaciones de espacio, tiempo y sincronización, así como a la necesidad de condensar el mensaje sin perder la función humorística (Vélez 2020, 5).

Para afrontar estos problemas, los traductores pueden elegir entre diversas estrategias: mantener la referencia original cuando sea reconocible, sustituirla por otra más cercana al público meta, o neutralizarla si su comprensión resulta improbable (Lorenzo 2005, 142 citado en Botella Tejera 2017, 94). Tal y como subraya Zabalbeascoa (2001, 257), la prioridad en la traducción humorística debe ser siempre provocar la risa y el entretenimiento, aunque ello suponga sacrificar la equivalencia léxica o sintáctica del original: «¿Si los chistes no funcionan como chistes buenos en el TM, qué importa que se hayan traducido fielmente cada una de las palabras que los componían o su contenido informativo?» (ibid.).

A nivel práctico, Gottlieb (Davila *et al.* 2023, 64) señala estrategias específicas como la expansión, condensación, paráfrasis o imitación, mientras que Mariani (2022, citado en Davila *et al.* 2023,64) remarca que los juegos de palabras son especialmente complejos y exigen creatividad y flexibilidad. Además, Botella Tejera (2017, 96) insiste en que el traductor audiovisual debe atender siempre a la relevancia cultural y a los «principios cooperativos» del texto, para que la referencia humorística tenga un impacto similar en la nueva audiencia.

En definitiva, subtítular humor requiere una combinación de pericia lingüística, sensibilidad cultural y manejo técnico, donde el objetivo principal es que la audiencia meta experimente una comicidad equivalente a la de la audiencia original.

3.Tecnologías del procesamiento automático del lenguaje

En los últimos años, el subtitulado ha experimentado una transformación significativa gracias a los avances en las tecnologías del procesamiento automático del lenguaje. Actualmente, herramientas como la traducción automática (TA) y el reconocimiento automático del habla (ASR) permiten generar subtítulos de manera más rápida y eficiente.

3.1. La traducción automática

El término traducción automática (TA o MT, esta última del inglés *machine translation*) es, como señalan Forcada *et al.* (2016, 95), ambiguo, ya que puede hacer referencia tanto al proceso de traducir mediante una máquina, como al producto final resultante de dicho proceso, e incluso al propio software utilizado. Por su parte, François Yvon (2023) lo define como un campo científico interdisciplinar que reúne a lingüistas, lexicólogos, informáticos y profesionales de la traducción con un objetivo común: diseñar y desarrollar recursos electrónicos y programas informáticos capaces de traducir automáticamente un documento en una lengua origen (LO) a un texto equivalente en una lengua meta (LM).

A lo largo de los años, esta tecnología ha ido evolucionando para adaptarse a las necesidades de la sociedad, pudiendo diferenciar tres paradigmas principales: la traducción automática basada en reglas (*rule-based* MT) la traducción automática estadística (*statistical* MT o SMT) y la traducción automática neuronal (*neural* MT o NMT).

El primer enfoque de la traducción automática fue la *rule-based* MT o traducción basada en reglas. Surgió en los años 50 con el objetivo de crear un sistema que fuera capaz de traducir textos sin intervención humana. Este método se fundamentaba en normas lingüísticas predefinidas por expertos de manera deductiva y diccionarios que establecían equivalencias entre pares de idiomas. No obstante, su desarrollo era muy costoso y resultaba imposible ofrecer toda la información necesaria (Forcada *et al.* 2016, 151).

De ahí que, a finales del siglo XX, con el aumento de grandes corpus bilingües, apareciera la traducción automática estadística. Las reglas fueron remplazadas por modelos probabilísticos, ahora, el texto origen se segmentaba en frases y se buscaba la traducción más probable en tablas de frases. Aunque la SMT supuso un gran avance en comparación con la RBMT, seguía teniendo ciertas limitaciones para generar traducciones contextualmente precisas (Kenny 2022, 36; Forcada *et al.* 2016, 177).

Posteriormente, la llegada de la TAN (traducción automática neuronal) en 2014 revolucionó la industria. A diferencia de los enfoques anteriores, este nuevo método utiliza redes neuronales artificiales para procesar enormes volúmenes de datos y producir así traducciones más precisas y fluidas (Kenny 2022, 39). En 2015, un sistema desarrollado en la Universidad de Standford superó con creces a los sistemas SMT en una prueba comparativa, especialmente en pares de idiomas complejos como inglés-alemán (Bentivogli *et al.*, 2016 citado en Kenny 2022, 39). Fue entonces cuando, según Bentivogli *et al.* (ibid.), comenzó lo que se conoce como «la nueva era NMT», en la que ellos mismos llegaron a comparar a los sistemas NMT con la calidad humana.

La traducción automática neuronal funciona a través de un modelo llamado *encoder-decoder* (codificador-decodificar). Primero, el texto origen se introduce en el codificador, que convierte cada palabra en una representación numérica (*word embeddings*). Estas representaciones no son simples equivalencias de palabras, sino que reflejan el significado en función del contexto en el que aparecen. De esta forma, las palabras con significados similares tienen representaciones matemáticas más cercanas entre sí. Después, esta información pasa al decodificador, que se encarga de generar la traducción en la lengua meta, eligiendo las palabras que más probabilidad tienen de ser correctas según lo que el sistema ha aprendido durante su entrenamiento (Bahdanau, Cho y Bengio 2015, 1).

Las redes neuronales artificiales, que imitan el funcionamiento del cerebro humano, son la base de estos sistemas. Están formadas por capas de nodos interconectados que procesan y aprenden patrones.

Un elemento clave de las arquitecturas neuronales de la última década es el mecanismo de atención (*attention mechanism*). En los primeros modelos neuronales, en especial los basados en redes recurrentes (RNN), la información de la oración de origen se comprimía en un solo vector, lo que podía hacer que se perdieran detalles importantes, sobre todo en oraciones largas. Este problema se resolvió con la introducción del mecanismo de atención, que evalúa constantemente todas las palabras del texto origen y asigna números a cada una de ellas en función de su relevancia en la traducción (Ashraf 2024, 657; Vaswani *et al.* 2017).

Actualmente, los sistemas más avanzados utilizan arquitecturas Transformer, que ha reemplazado en gran medida a las RNN. A diferencia de estas, se basan exclusivamente en mecanismos de atención para establecer dependencias globales entre todos los elementos de la oración, sin procesarlas secuencialmente. Esto ha permitido que la traducción automática sea más rápida y fluida, mejorando también la precisión en estructuras gramaticales complejas. Herramientas como DeepL y Google Translate han incorporado esta tecnología, lo que ha resultado en una mejora significativa en la calidad de sus traducciones (Ashraf 2024, 657; Vaswani *et al.* 2017).

3.2. El reconocimiento automático del habla

Otro de los sistemas empleados habitualmente en los procesos de subtítulo automático es el reconocimiento automático del habla (ASR). Esta tecnología convierte el lenguaje hablado en texto para que pueda ser leído y procesado por sistemas informáticos a tiempo real. Este proceso comienza con la captura de las vibraciones producidas por la voz humana, las cuales son transformadas en señales digitales mediante un convertidor analógico a digital (ADC, por sus siglas en inglés), que mide las ondas sonoras y filtra la información relevante. La señal digital resultante se segmenta en fragmentos muy breves y a cada segmento se le asigna un fonema (Amazon Web Services 2021; Ahmad 2022).

Estos fonemas se procesan posteriormente con modelos matemáticos que los comparan con patrones de palabras y frases conocidas para determinar la secuencia de texto más probable que corresponde al audio original (Ahmad 2022). Finalmente, el software presenta la transcripción en formato de texto. En este sentido, el ASR utiliza tres tipos de

modelos fundamentales: el modelo acústico, que relaciona las señales de audio con unidades lingüísticas como los fonemas; el modelo lingüístico, que identifica patrones y secuencias plausibles de palabras para mejorar la precisión; y el modelo de pronunciación, que vincula la representación ortográfica de las palabras con sus representaciones fonéticas (ibid.).

La evolución del ASR ha llevado a la incorporación de técnicas de aprendizaje profundo, como las redes neuronales recurrentes, que permiten al sistema aprender directamente de grandes volúmenes de datos y mejorar su robustez frente a variaciones del hablante y condiciones acústicas adversas (Ahmad 2022; Hannun *et al.* 2014, 1). Sin embargo, el rendimiento del ASR puede verse afectado por varios factores, entre ellos el ruido ambiental, la calidad del audio, la variedad de acentos y pronunciaciones, así como la complejidad del vocabulario utilizado (Jurafsky y Martin 2024, 333; Amazon Web Services 2021).

4. El subtítulo automático de YouTube

YouTube, que forma parte de Google, utiliza tecnologías avanzadas de esta empresa para generar subtítulos automáticos en los vídeos. La base de este sistema es el software que se acaba de mencionar, el reconocimiento automático del habla (ASR).

Google ofrece a los desarrolladores la **API de Cloud Speech-to-Text**, un servicio en la nube que emplea modelos de aprendizaje profundo y redes neuronales para analizar y transcribir con alta precisión audios en múltiples idiomas y acentos. Esta API puede procesar tanto audios en tiempo real como grabados, además de distinguir entre varios hablantes y detectar automáticamente el idioma, lo que mejora la calidad de los subtítulos.

Un ejemplo práctico de esta tecnología es **Live Captions**, una herramienta de Google disponible tanto en su navegador Chrome como en dispositivos Android, que permite generar subtítulos automáticos en tiempo real para cualquier tipo de audio reproducido en el dispositivo. Live Captions utiliza algoritmos de reconocimiento de voz y aprendizaje automático similares a los de la API de Cloud Speech-to-Text, pero está diseñada para el usuario final (García 2021).

Para facilitar la accesibilidad global, YouTube utiliza la **API de Traducción de Google Cloud**, que se basa en la tecnología de traducción automática neuronal de Google. Esta

API traduce textos de forma dinámica mediante modelos personalizados previamente entrenados o mediante modelos de lenguaje masivos (LLM, Large Language Models) especializados en traducción. Estos modelos avanzados, basados en redes neuronales profundas, mejoran la calidad y coherencia de las traducciones al captar mejor el contexto y las sutilezas del lenguaje (Google Cloud Translation API 2025).

Como señalan los propios desarrolladores, estas tecnologías son muy avanzadas, pero la precisión puede verse afectada por factores como la calidad del audio, ruido de fondo, rapidez del habla o terminología específica. Por ello, YouTube permite que los creadores editen manualmente los subtítulos automáticos para corregir errores y mejorar su accesibilidad (YouTube Help s.f.).

5. La evaluación de la calidad de los subtítulos

A pesar de los avances tecnológicos en la generación automática de subtítulos, es fundamental evaluar la calidad de estos para garantizar su utilidad y accesibilidad real, ya que estos sistemas pueden cometer errores en múltiples puntos de su proceso, y cuanto antes ocurran, más perjudiciales pueden ser para el resultado final (Di Gangi *et al.*, 2022, 352, citado en Baños 2023, 66). La precisión de los subtítulos no solo depende de la tecnología empleada, sino también del contexto, la claridad del audio y las necesidades específicas de los usuarios.

La evaluación de la calidad del subtitulado puede realizarse mediante métodos automáticos y manuales. Los modelos automáticos, como la tasa de error de palabra (WER) o la puntuación BLEU, usan métricas cuantitativas que comparan los subtítulos generados con transcripciones de referencia, mientras que la evaluación manual implica la revisión por expertos humanos que valoran aspectos como la precisión, fluidez, coherencia y adecuación cultural (Pangeanic 2022)

Entre los principales modelos de evaluación humana destaca el MQM (*Multidimensional Quality Metrics*), que permite clasificar y jerarquizar distintos tipos de errores en la traducción y subtitulado, proporcionando un análisis detallado de la calidad lingüística y funcional (Multidimensional Quality Metrics, 2025). Como se observará más adelante, este enfoque resulta especialmente útil para evaluar subtítulos automáticos en contextos complejos, como el humor o la cultura, donde la simple coincidencia textual no basta para garantizar la calidad.

MARCO APLICADO

1. Plataforma

YouTube ha sido elegida para este análisis por ser la plataforma de video más utilizada a nivel mundial. Fundada en 2005 y adquirida por Google en 2006, cuenta con más de 2700 millones de usuarios activos mensuales en 2025, lo que representa aproximadamente el 47% de los usuarios de internet a nivel mundial (Affmaven 2024; Formación Avanzada Europea 2024).

A lo largo de los años, YouTube ha pasado de ser una simple plataforma de videos a convertirse en un referente en la distribución de contenido audiovisual. Su accesibilidad desde diversos dispositivos y su interfaz intuitiva han sido elementos clave que han favorecido su consolidación como la principal fuente de consumo de videos en línea. Además, su modelo gratuito ha facilitado el acceso a una amplia gama de contenidos, tanto educativos como de entretenimiento (Formación Avanzada Europea 2024).

Esta accesibilidad y diversidad de material han permitido localizar fácilmente videos con una fuerte carga cultural y humorística, cumpliendo con los criterios específicos del análisis.

2. Vídeos seleccionados

Se han seleccionado tres vídeos con alta carga cultural y humorística para evaluar cómo responde el sistema de subtulado y traducción automáticos de YouTube ante estas características. A continuación, se describen los vídeos seleccionados:

El primero de los vídeos seleccionados es [The Problem with Christmas TV Adverts](#), publicado el 23 de noviembre de 2018. Pertenece al programa *The Mash Report*¹ de la BBC (British Broadcasting Corporation). El *sketch* critica los anuncios navideños británicos, que suelen caracterizarse por un tono excesivamente emotivo y una representación idealizada de la Navidad, alejada en muchos casos de la realidad cotidiana. Se pone de relieve cómo estos anuncios apelan a los sentimientos con fines claramente comerciales, buscando generar una respuesta emocional que incentive el consumo. En él

¹ Programa de comedia británico emitido por la BBC entre 2017 y 2020. Parodia el formato de los noticieros y, mediante sketches y comentarios irónicos, ofrece una crítica satírica de la actualidad política, social y cultural del Reino Unido ((British Comedy Guide, s.f.)

aparecen varios personajes con funciones bien diferenciadas. El hilo conductor lo lleva un presentador que, con un tono serio e irónico, introduce el tema y da paso a las distintas intervenciones. A continuación, se presentan dos testimonios ficticios que relatan cómo los anuncios navideños les afectan emocionalmente. Estos personajes parodian casos reales y refuerzan el mensaje crítico del *sketch*. Uno de ellos llama especialmente la atención por su marcado acento del norte de Inglaterra y la velocidad con la que habla, lo que añade complejidad tanto a la comprensión como, seguramente, a la traducción automática de su discurso. Finalmente, interviene un supuesto profesor o experto que explica, en clave de humor, los mecanismos psicológicos que utilizan los anuncios para manipular las emociones del público.

El segundo vídeo seleccionado es [*Vanity Overshadows the Gasforth Police Department*](#), un episodio de la serie *The Thin Blue Line*, que originalmente fue emitida por la BBC entre 1995 y 1996. Este vídeo fue subido a la página oficial de YouTube de *BritBox*, un servicio de *streaming* de la BBC, el 6 de agosto de 2019. La serie es una comedia ambientada en una comisaría policial ficticia llamada Gasforth, que retrata con humor e ironía las dificultades y absurdos de la vida en el cuerpo policial. En este capítulo, la comisaría se prepara para la visita de un equipo de televisión que les realizará una entrevista, lo que lleva a los policías a esforzarse por mejorar su apariencia y proyectar una imagen impecable. El inspector Fowler (Rowan Atkinson) critica esta preocupación excesiva por la imagen, aunque finalmente él mismo también cae en esa vanidad.

El tercer vídeo seleccionado es [*Posh Family Reacts to Northern Nanny*](#), un extracto subido el 27 de agosto de 2021 del popular programa británico *The Catherine Tate Show*², que originalmente se emitió entre 2004 y 2007 en la BBC Two. En este *sketch*, una familia de clase alta espera a su niñera habitual, pero la agencia les envía a otra en su lugar: una mujer del norte de Inglaterra con un acento muy cerrado y una forma de hablar extremadamente rápida. La familia mantiene en todo momento un acento *posh*³, caracterizado por una pronunciación refinada y asociada a la élite británica, lo que acentúa aún más el contraste con la forma de hablar de la nueva niñera. Esta diferencia genera una situación cómica en la que los miembros de la familia insisten en que no entienden nada de lo que ella dice. El *sketch* parodia los acentos regionales y actúa como una crítica

²Programa de comedia británico de *sketches* protagonizado por Catherine Tate. Parodia estereotipos sociales y acentos mediante personajes humorísticos (Valleriani 2024).

³Adjetivo informal del inglés británico que describe a una persona, acento o estilo asociado a la clase alta; en español puede traducirse como «pijo», «refinado» o «estirado», según el contexto (WordReference, s.f.).

humorística a las diferencias de clase en el Reino Unido, mostrando cómo el lenguaje y la pronunciación funcionan como marcadores sociales.

3. Modelo de análisis

3.1. Modelo MQM

Para elaborar la parrilla de evaluación de la [Tabla 1](#) se ha tomado como referencia el modelo MQM (*Multidimensional Quality Metrics*), en su versión más reciente (MQM 2025) como marco general. No obstante, la clasificación específica de errores se basa principalmente en la adaptación propuesta por Ortiz Boix (2016) en su tesis doctoral, donde aplica la primera versión del modelo (*Multidimensional Quality Metrics, Draft Version* 2012). Resulta especialmente pertinente, ya que, al igual que en el presente estudio, Ortiz Boix se centra en la identificación y clasificación de errores en productos audiovisuales, en su caso aplicados al doblaje de documentales de vida silvestre.

MQM no ofrece una métrica de evaluación de traducción fija, sino que se trata de un marco flexible para definir modelos específicos para una tarea de traducción empleando una terminología y tipos de errores comunes (Campo Bayón 2019, 41). Esta flexibilidad ha permitido diseñar una parrilla propia adaptada a las necesidades del estudio, que se centra en tres categorías principales: *Precisión* (P), *Convenciones Lingüísticas* (F) y *Estilo* (E)⁴. Las dos primeras también están presentes en el modelo empleado por Ortiz Boix, si bien con denominaciones ligeramente distintas, pues el término «*Fluency*» que utiliza Ortiz Boix, se corresponde, en esta propuesta, con «*Convenciones Lingüísticas*» en línea con la última versión de MQM. A estas dos categorías, en este contexto, se suma la de *Estilo*, incluida por su relevancia en el análisis de productos audiovisuales con contenido cultural y humorístico, donde el modo en que se transmite el mensaje en la lengua meta resulta crucial para preservar tanto la fidelidad al original como el tono y las sutilezas del humor.

El hecho de poder adaptar el modelo MQM a distintos contextos ha permitido priorizar las categorías más relevantes para este estudio. Por ejemplo, aunque el modelo actual de MQM sugiere tratar la terminología como un nivel independiente, en este caso se ha

⁴ Traducciones propias de los términos originales en inglés *Accuracy*, *Linguistic Conventions* (en la versión anterior del modelo denominada *Fluency*, de ahí la abreviatura [F] para *Convenciones Lingüísticas*) y *Style*, respectivamente.

optado por mantener la propuesta del modelo anterior e incluirla dentro de la categoría *Precisión*, ya que los textos por analizar no contienen una gran cantidad de terminología especializada y, por tanto, no constituyen un aspecto prioritario a la hora de evaluar la calidad de la traducción. Asimismo, se han incorporado nuevas subcategorías, que, aunque no aparecen en el modelo actual de MQM ni en la propuesta de Ortiz Boix, han resultado útiles para el análisis, como es el caso de *Sin sentido* (P2.8) o *Pérdida de significado cultural* (P2.9).

Además, uno de los rasgos distintivos de estos modelos es la capacidad para jerarquizar los errores en distintos niveles de gravedad. Gracias a ello, y a una parrilla organizada en diferentes niveles con sus correspondientes subgrupos, se puede identificar con mayor exactitud el tipo de fallo cometido, lo que mejora la precisión del análisis y facilita su corrección.

A continuación, se describen las tres categorías principales y los subgrupos correspondientes de este modelo de análisis:

3.1.1. Precisión (P)

Los errores que surgen cuando el contenido del texto meta no refleja con exactitud el significado esencial del texto original, debido a alteraciones, omisiones o añadidos en el mensaje⁵.

En este grupo se incluyen los siguientes subgrupos:

- **Terminología (P1):** uso incorrecto de términos normativos del campo temático o elección inadecuada del equivalente en la lengua meta. Ejemplo: traducir *allergy* como «intolerancia» en lugar de «alergia».
- **Mala interpretación (P2):** errores debidos a una transferencia errónea o ambigua del significado original, subdividido en:
 - a) P2.1: traducción demasiado literal. Ejemplo: traducir *confidence issues* como «problemas de confianza» en vez de «problemas de autoestima».

⁵ «Errors occurring when the target content does not accurately correspond to the propositional content of the source text because of distortion, omission, or addition to the message» (MQM 2025).

- b) P2.2: términos no traducibles. Ejemplo: traducir *bluetooth* como «diente azul» en vez de mantenerse igual.
 - c) P2.3: invención/sin sentido. Ejemplo: traducir *cheers!* por «¡quesos!» en vez de «¡salud!».
 - d) P2.4: fecha u hora incorrectas. Ejemplo: interpretar mal una fecha anglosajona: 03/05/2024 como 5 de mayo en lugar de 3 de marzo.
 - e) P2.5: errores en números o unidades. Traducir *the room is 10 by 12 feet* por «la habitación mide 10 por 12 metros» en vez de «la habitación mide aproximadamente 3 por 3,6 metros».
 - f) P2.6: entidades o nombres propios mal traducidos. Ejemplo: traducir William Shakespeare por Guillermo Shakespeare en lugar de mantener William Shakespeare, pues los nombres de figuras históricas no deben ser traducidos a menos que exista una forma asentada.
- **Omisión (P3):** eliminación no justificada de información del original.
 - **Adición (P4):** inclusión de información no presente en el original.
 - **Sin traducir (P5):** segmentos o palabras no traducida de forma injustificada.

3.1.2. Convenciones lingüísticas (F)

Errores relacionados con la corrección lingüística del texto, que incluyen problemas de gramática, expresiones idiomáticas y aspectos mecánicos⁶.

En este grupo se incluyen los siguientes subgrupos:

- **Ortografía (F1)**
 - a) F1.1: uso incorrecto de mayúsculas y/o minúsculas.
 - b) F1.2: diacríticos mal empleados o ausentes.
- **Tipografía y puntuación (F2)**

⁶ «Errors related to the linguistic well-formedness of the text, including problems with grammaticality, idiomaticity, and mechanical correctness» (MQM 2025).

- a) F2.1: puntuación incorrecta.
- b) F2.2: uso inadecuado de comillas, corchetes u otros signos.
- **Gramática (F3)**
 - a) F3.1: morfología incorrecta (sufijos, prefijos, verbos...).
 - b) F3.2: uso erróneo de palabras funcionales (preposiciones, pronombres...).
 - c) F3.3: concordancia incorrecta.
 - d) F3.4: orden de palabras inadecuado.
- **Ininteligible (F4):** fragmentos imposibles de comprender por problemas estructurales o léxicos.

3.1.3. Estilo (E)

Errores que, aunque gramaticalmente aceptables, resultan inapropiados por desviarse de las guías de estilo organizacional o por emplear un registro lingüístico inadecuado⁷.

En este grupo se incluyen los siguientes subgrupos:

- **Redacción poco natural (E1):** expresiones forzadas o antinaturales.
- **Estilo inapropiado (E2):** registro o tono inadecuado.
- **Uso repetitivo de términos (E3):** repetición excesiva que afecta a la calidad estilística del texto.

La tabla resultante que se muestra a continuación ofrece, por tanto, una clasificación detallada que permite detectar, analizar y corregir errores de forma precisa y eficaz, adaptándose a las necesidades específicas del análisis de productos audiovisuales con contenido cultural y humorístico, y aprovechando la flexibilidad que ofrece el modelo MQM.

⁷ «Errors occurring in a text that are grammatically acceptable but are inappropriate because they deviate from organizational style guides or exhibit inappropriate language style» (MQM 2025).

Tabla 1. Parrilla de evaluación basada en MQM para traducción audiovisual.

PRECISIÓN (P)		
	Terminología (P1)	
	Mala interpretación (P2)	Demasiado literal (P2.1)
		No debería haberse traducido (P2.2)
		Invención/ sin sentido (P2.3)
		Fecha/hora (P2.4)
		Números y unidades (P2.5)
		Entidades/ nombres propios (P2.6)
	Omisión (P3)	
	Adición (P4)	
	Sin traducir (P5)	
CONVENCIONES LINGÜÍSTICAS (F)		
	Ortografía (F1)	Mayúsculas/minúsculas (F1.1)
		Diacríticos (F1.2)
	Tipografía (F2)	Puntuación (F2.1)
		Comillas, corchetes u otros signos (F2.2)
	Gramática (F3)	Morfología (F3.1)
		Palabras funcionales (F3.2)
		Concordancia (F3.3)
		Orden de las palabras (F3.4)
	Ininteligible (F4)	
ESTILO (E)		
	Redacción poco natural (E1)	
	Estilo inapropiado (E2)	
	Uso excesivo de términos repetitivos (E3)	

3.2. Códigos específicos para ASR

Además del modelo de análisis anterior, se ha implementado un sistema de anotación con corchetes para clasificar el tipo de error cometido exclusivamente por el ASR. Los códigos empleados son:

- [T-]: omisión de palabras o fragmentos presentes en la transcripción original.
- [T+]: adición de elementos no presentes en la transcripción original.
- [Tx]: error o falta de transferencia del contenido, que da lugar a una palabra o frase incorrecta en el subtítulo automático.

La letra T se utiliza en estos códigos para hacer referencia a la *transcripción*, ya que los errores señalados se producen al comparar el subtítulo automático en inglés (B) con la transcripción original (manual) del inglés (A). Es decir, se trata de un proceso intralingüístico e intersemiótico, no de una traducción entre lenguas. Por este motivo, no es apropiado aplicar aquí la categoría de *Precisión* del modelo MQM, diseñada específicamente para evaluar traducciones interlingüísticas. En su lugar, se ha optado por un sistema de codificación específico propio que permita identificar y clasificar los errores característicos del proceso de reconocimiento automático del habla (ASR), así como por reutilizar las dos categorías del modelo MQM (*Convenciones lingüísticas* y *Estilo*) que se pueden aplicar desde una perspectiva monolingüe o intralingüística.

Es importante destacar que este sistema de codificación parte de la idea de que el ASR funciona como un proceso de entrada y salida: recibe una señal de audio (entrada) y genera una transcripción escrita (salida). Los errores marcados con códigos T reflejan fallos en la entrada, es decir, en cómo el sistema interpreta el audio. Por su parte, los errores formales etiquetados con el modelo MQM (categoría F o *Convenciones lingüísticas* y categoría E o *Estilo*) afectan a la salida textual, evaluando aspectos como puntuación, ortografía o uso de mayúsculas. Así, ambos sistemas se complementan y permiten analizar tanto el reconocimiento del contenido como la calidad del subtítulo resultante.

4. Metodología

4.1. Extracción de los datos del corpus

Con el objetivo de obtener resultados precisos y fiables, se llevó a cabo un proceso de análisis estructurado y cuidadosamente supervisado.

Como ya se comentó previamente, primero se seleccionaron los vídeos objeto de estudio atendiendo a un criterio fundamental: que presentaran una notable carga cultural y un componente humorístico relevante.

Posteriormente, se utilizó OBS Studio⁸ para grabar el audio directamente desde YouTube de cada uno de los tres vídeos seleccionados. A continuación, los archivos de audio se introdujeron en la plataforma Sonix.ai⁹, que generó automáticamente la transcripción del audio original en formato Word, junto con la segmentación temporal y la identificación de los distintos hablantes. Esta transcripción fue revisada manualmente para asegurar su exactitud, comparando cada fragmento con el contenido real del audio. Asimismo, se verificó que los intervalos temporales asignados por la herramienta coincidieran con las intervenciones reales de cada hablante, realizando los ajustes necesarios cuando fue preciso para garantizar una correspondencia con los subtítulos.

En cuanto a los subtítulos generados automáticamente por YouTube en inglés, se accedió a la transcripción completa a través del menú de opciones del vídeo (tres puntos verticales > *Mostrar transcripción*). Esta función permite visualizar todos los fragmentos de subtítulos en una columna lateral, desde donde pueden copiarse manualmente. Desde ahí, los subtítulos fueron pegados en un documento de Word, y se comprobó que coincidieran exactamente con lo mostrado en pantalla, tanto en contenido como en número de caracteres visibles por fragmento.

Finalmente, la transcripción de los subtítulos en español se llevó a cabo manualmente, viendo el vídeo en YouTube y copiando palabra por palabra cada línea tal como aparecía en pantalla, respetando su segmentación original. Este procedimiento fue necesario porque, a diferencia de los subtítulos automáticos en inglés, YouTube no ofrece la

⁸ *Software* libre y de código abierto para grabación de video y transmisión en vivo (OBS Project 2025)

⁹ Herramienta de transcripción avanzada que convierte archivos de audio y vídeo en texto escrito con gran precisión (Sonix.ai 2025)

posibilidad de acceder a una transcripción completa cuando los subtítulos han sido generados automáticamente en otro idioma

4.2. Procedimiento de análisis

A continuación, se trasladaron las transcripciones de cada vídeo, tanto en inglés como en español, a una hoja de cálculo de Excel titulada *Análisis*, con el fin de facilitar un análisis más preciso y sistemático. Para ello, se estructuraron los datos en distintas columnas, cada una con una función específica:

- Columna A: transcripción del audio original en inglés. Los sonidos no textuales relevantes, como marcas suprasegmentales o prosódicas y otros sonidos del ambiente que influyen en la comprensión, se indican entre corchetes.
- Columna B: transcripción de los subtítulos automáticos generados por YouTube en inglés.
- Columna C: anotación de los errores detectados en la comparación entre el audio original (columna A) y los subtítulos en inglés (columna B).
- Columna D: número de caracteres presentes en cada fragmento de subtítulo en inglés (columna B).
- Columna E: transcripción de los subtítulos automáticos generados por YouTube en español.
- Columna F: anotación de los errores detectados en la traducción de los subtítulos del inglés al español (comparación entre columna B y E).
- Columna G: número de caracteres por fragmento en los subtítulos en español (columna E).
- Columna H: propuestas de mejora para la traducción o segmentación de los subtítulos.
- Columna I: aspectos que han sido corregidos adecuadamente en la traducción del inglés al español.

Una vez estructurada la hoja de cálculo, se procedió al análisis de errores. El primer paso consistió en comparar la transcripción del audio original (**columna A**) con los subtítulos automáticos generados en inglés (**columna B**), siguiendo el modelo de categorías de la tabla de análisis (ver en la [Tabla 1](#)) y los códigos específicos para ASR, estos últimos utilizados solo en procesos de traducción intralingüística. Esta comparación permitió

identificar divergencias y errores relacionados con el reconocimiento automático del habla (ASR).

Para facilitar la interpretación visual de los datos, se aplicó un sistema de codificación con colores, negritas y signos gráficos. La convención fue la siguiente:

- Las oraciones de la columna A aparecen **en negrita**.
- Las palabras **en verde** de la columna A permiten identificar visualmente qué palabra en B es incorrecta y cuál sería su forma correcta. Es decir, al ver una palabra mal transcrita en B, se puede localizar de inmediato su versión correcta en A gracias al marcado en verde.
- Los errores detectados en B se marcan en **rojo y en negrita**.
- Las omisiones (ya sea de sonidos, palabras o signos de puntuación) se indican entre **llaves ({}).**
- Cuando el sistema ASR no distingue un cambio de hablante la frase correspondiente se **subraya** para señalar que pertenece al siguiente interlocutor. Esto sucede porque el sistema no detecta cuándo habla una persona diferente y continúa el subtítulo en la misma línea, como si fuera el mismo hablante, sin separar el texto entre interlocutores.

Cada error se codificó mediante dos sistemas complementarios, como se adelantó en el apartado 3:

3. **Códigos del modelo de análisis** (ver en la [Tabla 1](#)): combinaciones de letras y números que identifican el tipo de error (por ejemplo, F2.1 = *Convenciones lingüísticas > Puntuación*). Se asocian a la salida textual.
4. **Códigos específicos para ASR**: la letra T en combinación con los signos «-», «+» y «x» (ver la [explicación](#)), indicando respectivamente omisión, adición, y error o falta de transferencia del contenido. Se asocian a la entrada textual.
5. Además, se añadió la etiqueta «**No distinción hablante**» para aquellos casos en los que el ASR no detectó el cambio de hablante.

En las **columnas D y G**, correspondientes al número de caracteres por línea de subtítulo, se marcaron en rojo aquellas que superaban el límite recomendado de 42 caracteres por línea, considerado un umbral óptimo para garantizar una lectura cómoda y fluida en subtítulo, como se señala en el marco teórico según Díaz Cintas y Remael (2021, 98).

	A	B	C	D
1	AUDIO ORIGINAL INGLÉS	TRANSCRIPCIÓN SUBTITULADO INGLÉS	ANÁLISIS	Nº CARACTERES
2		[00:00:00 – 00:00:19]		
3		Man 1 (Presenter):		
4		Britain is in the grip of a vomiting		36
5		epidemic caused by nauseating TV adverts		42
6	for big shops [laughter].	for big shots [(laughter)]{.}	[Tx] + [T-] + F2.1	14
7	The lavish Christmas commercials,	the lavish Christmas commercials{.}	F2.1	32
8		typically featuring some kind of bear or		41
9	child with confidence issues, are so syrupy	child with confidence issues{.}(are) so syrupy	F2.1 + [T-]	39
10		and cloying that they're causing viewers		41
11		to be physically sick and sometimes even		41
12	shit their pants [laughter].	{shit} their pants [(laughter)]{.}	[T-] + [T-] + F2.1	11
13		[00:00:20 – 00:00:48]		
14		Woman 1 (Donna Sheridan):		
15	I was having my [me] tea and this advert came on	I was a tamati and the sandwich came on	[Tx] + [Tx]	40
16		about a mum watching a nativity play and		41
17	she's really nervous and a kid singing	she's really nervous and kids singing	[Tx]	38
18	and. But then it's brilliant and and then	and{.} but then it's brilliant and (and) then	F2.1 + [T-]	37
19	the whole room just fills with love [laughter] [vomiting sound].	the whole room just fills with love [(laughter)]	[T-] + [T-]	36
20	I'm I'm just sick immediately in my lap [soft laughter].	I'm I'm just sick immediately in my lap [(soft laughter)]{.}	[T-] + F2.1	40
21	some came out of my nose. Then the Waitrose	some came out of my nose{.} then the way it	F2.1 + [Tx]	41
22	comes on. I nuked again all	was outward comes on{.} I nuked again all	[Tx] + F2.1	39

Ilustración 1. Captura de pantalla de la hoja de cálculo de Excel con las columnas A, B, C y D.

En una segunda fase, se compararon los subtítulos en inglés (**columna B**) con los subtítulos automáticos en español (**columna E**). Este análisis se realizó siguiendo un procedimiento similar al anterior, pero centrándose ahora en la calidad de la traducción automática. A diferencia de los errores propios del ASR (T-, T+, Tx), en esta etapa los errores se relacionaron con aspectos de precisión, convenciones lingüísticas y estilo.

Para distinguir los errores originados por la traducción automática (de B a E) de aquellos que ya estaban presentes en la transcripción original (de A a B), se adoptó un sistema de doble etiquetado. Por ejemplo, si un subtítulo en español presentaba un *Sin sentido* (P2.8) debido a un error de transcripción previa ([Tx]), se anotaba como P2.8[Tx].

Los errores de traducción en esta fase se marcaron en **color naranja**, diferenciándose así visualmente de los errores en inglés. Para reforzar la comprensión visual, se utilizaron también otros colores complementarios:

- **Azul y rosa** para distinguir tipos de errores diferentes contiguos.

	E	F	G
1	TRANSCRIPCIÓN SUBTITULADO ESPAÑOL	ANÁLISIS	Nº CARACTERES
2	[00:00:00 – 00:00:19]		
3	Hombre 1 (Presentador):		
4	Gran Bretaña está en las garras de una	P2.1	38
5	epidemia de vómitos causada por los		
6	nauseabundos anuncios de televisión		71
7	de los peces gordos	F3.2 + P2.1[Tx]	19
8	los lujosos anuncios navideños que		34
9	suelen presentar algún tipo de oso o		36
10	niño con problemas de confianza tan	P2.1	
11	empalagosos		47
12	y almibarados que están haciendo que los		
13	espectadores se sientan		64
14	físicamente enfermos y a veces [incluso]	P2.1 + P3	
15	se		33
16	caguen en los pantalones	P4[T-]	24
17	[00:00:20 – 00:00:48]		
18	Mujer 1 (Donna Sheridan):		
19	Yo era un tamati y apreció el sándwich	P2.3[Tx] + P2.3	38
20	sobre una madre viendo una obra de teatro de	P2.1/ E1	
21	la natividad y está		64
22	muy nerviosa y los niños cantando		33
23	y pero luego es brillante y y luego	F2.1 + P4[T-]	35
24	toda la habitación se llena de amor	F2.1	35
25	Estoy [estoy] simplemente enfermo	E1 (toda la oración) +	
26	inmediatamente en mi regazo	P3 + F3.3 + F3.4	53
27	se me salió algo de la nariz y luego aparece la	F3.1 + P.4	
28	forma en que		60

Ilustración 2. Captura de pantalla de la hoja de cálculo de Excel con las columnas E, F y G

En la **columna H**, se ofrecieron propuestas de mejora para los subtítulos en español, orientadas a lograr una traducción más idiomática, fluida y adaptada al contexto cultural y humorístico del vídeo. Finalmente, la **columna I** recogió aquellas situaciones en las que el motor de traducción automática logró corregir de manera adecuada errores presentes en el subtítulo original en inglés. Estas correcciones se marcaron en **verde**, para facilitar su identificación.

	H	I
2		
3		
4	en medio de	
5		
6	que emiten las grandes empresas	
7		
8		
9	de autoestima	
10		
11	lleguen a vomitar	
12		caguen
13		
14		
15	estaba tomando té y apareció este anuncio	
16	obra de teatro navideña	
17		
18		y
19		
20	Me, me siento mal al instante, como si todo cayera en mi regazo	
21		
22		Mayúscula (reconoce que hay cambio de frase)

Ilustración 3. Captura de pantalla de la hoja de cálculo de Excel con las columnas H e I.

5. Análisis y resultados

En esta sección se presenta un análisis conjunto, de carácter cuantitativo y cualitativo, de los errores detectados en el sistema de subtítulo automático de YouTube. El objetivo es evaluar la calidad y fiabilidad de sus dos componentes principales: el reconocimiento automático del habla (ASR) y la traducción automática neuronal (NMT).

El análisis está dividido en dos apartados por cada vídeo analizado:

1. Análisis de la transcripción automática en inglés (ASR)
2. Análisis de la traducción automática al español (NMT)

Cada apartado incluye un análisis **cuantitativo**, que incluye:

- El número total de errores detectados
- El número de líneas de subtítulo con errores

- La media de errores por línea
- La clasificación por tipo y subtipo, de acuerdo con los modelos de evaluación establecidos
- La gravedad de los errores, dividida en tres niveles:
 - **Leve** (1 punto): no impide la comprensión global. Ejemplo: escribir «christmas» en lugar de «Christmas». La falta de mayúscula solo vulnera una convención ortográfica; el lector sigue entendiendo la palabra sin dificultad.
 - **Grave** (5 puntos): entorpece la comprensión o distorsiona el contenido. Ejemplo: el subtítulo omite la acotación «[vomiting sound]». Al suprimirse la indicación sonora, el espectador pierde un elemento clave (el ruido de arcada) y se debilita el efecto cómico.
 - **Crítico** (10 puntos): imposibilita la comprensión del mensaje o anula el efecto humorístico. Ejemplo: traducir «I was having my tea and this advert came on» como «I was a tamati and the sandwich came on». Surgen palabras inventadas («tamati»), y cambios de sustantivo («advert» → «sandwich»). Esto rompe la lógica de la frase; el resultado carece de sentido.

Además, en el caso específico de los errores atribuidos al ASR (códigos T, véase en el [apartado ya mencionado](#)), se añade un análisis **cualitativo** que examina las posibles **causas** mediante una clasificación en cinco categorías:

- A) Habla demasiado rápida y/o acento cerrado o difícil de entender.
- B) Referencia cultural complicada
- C) Interferencia por sonidos externos al diálogo (risas, ruidos, etc.)
- D) Contenido reconocido como inapropiado por el sistema.
- E) Contenido no suficientemente audible o considerado irrelevante por el sistema.
- F) Otro motivo no determinado

Para facilitar esta evaluación, se ha elaborado una nueva hoja de cálculo de Excel (titulada *Resultados*) para cada vídeo, en la que se ha registrado todo lo anterior y, además, se han

cuantificado todos los errores detectados por línea de subtítulo. Asimismo, cada error ha sido codificado por colores según su gravedad:

- **Azul claro** para errores leves
- **Azul medio** para errores graves
- **Azul oscuro** para errores críticos

En el caso concreto de los errores del ASR (códigos T), cuando se ha podido determinar una causa específica, se ha añadido además la letra (A-F) indicativa de la causa (de color morado) junto al código. Por ejemplo, un error clasificado como [Tx] podría aparecer resaltado en azul medio, acompañado de la letra C, indicando que es un error de tipo grave cuya causa son interferencias por sonidos externos al diálogo (risas, ruidos, etc.).

	TRANSCRIPCIÓN SUBTITULADO INGLÉS	ANÁLISIS	Nº ERRORES
1			
2	[00:00:00 – 00:00:19]		
3	Man 1 (Presenter):		
4	Britain is in the grip of a vomiting		
5	epidemic caused by nauseating TV adverts		
6	for big shots [[laughter]] {,}	[Tx] C+ [T-] E + F2.1	3
7	the lavish Christmas commercials{,}	F2.1	1
8	typically featuring some kind of bear or		
9	child with confidence issues{,}{are} so syrupy	F2.1 + [T-] C	2
10	and cloying that they're causing viewers		
11	to be physically sick and sometimes even		
12	{shit} their pants [[laughter]] {,}	[T-] D+ [T-] E+ F2.1	3
13	[00:00:20 – 00:00:48]		
14	Woman 1 (Donna Sheridan):		
15	I was a tamati and the sandwich came on	[Tx] A+ [Tx] A	2
16	about a mum watching a nativity play and		
17	she's really nervous and kids singing		
18	and{,} but then it's brilliant and {and} then	F2.1 + [T-] A	2
19	the whole room just fills with love [[laughter]] [[vomiting sound]] .	[T-] E + [T-] E	2
20	I'm I'm just sick immediately in my lap [[soft laughter]] {,}	[T-] E+ F2.1	2
21	some came out of my nose{,} then the way it	F2.1+ [Tx] A	2
22	was outward comes on{,} I puked again all	[Tx] A+ F2.1	2
23	over the table [[soft laughter]] {,} I'm the cat{,} and watch the	[T-] E+ F2.1 + [Tx] A + F2.1 + [Tx] A	5

Ilustración 4. Captura de pantalla de la hoja de cálculo de Excel (titulada “Resultados”) con la codificación por colores según la gravedad del error, la letra correspondiente al análisis cualitativo y el número de errores por línea de subtítulo.

Este procedimiento se repetirá para los tres vídeos seleccionados, lo que permitirá comprobar el rendimiento de estas tecnologías en contextos lingüísticos exigentes. A su vez, este enfoque permitirá identificar patrones recurrentes de error y comparar el funcionamiento de las herramientas en distintos escenarios.

Además, en relación con la subcategoría *Puntuación* (F2.1), es importante matizar que en el análisis de la traducción automática de los tres vídeos no se han tenido en cuenta aquellos errores de puntuación que no han sido detectados en la fase previa (ASR). Se parte de la base de que, si el ASR no introduce signos como puntos o comas en su salida, el subtítulo resultante carecerá de esa información. No obstante, cabe señalar que la traducción automática trata los signos de puntuación como *tokens* (Pan *et al.* 2023, 31), al igual que las palabras, por lo que en algunos casos sí podría llegar a corregir o añadir puntuación ausente. Aun así, en este análisis no se ha considerado como error la falta de signos de puntuación cuando esta se debe únicamente a que el sistema de reconocimiento automático del habla (ASR) no los introdujo.

5.1. Vídeo nº1

5.1.1 Análisis de la transcripción automática en inglés (ASR)

En el primer vídeo analizado, compuesto por **41 líneas de subtítulo**, se identificaron **78 errores** originados por el sistema de subtulado automático de YouTube. De las 41 líneas, **30** presentaban al menos un error, lo que representa el 73,2 % del total. El promedio de errores por línea fue de **2,6**, lo que evidencia una densidad elevada de fallos por unidad de subtítulo. Además, considerando que el subtítulo contiene un total de 1.462 caracteres, el número medio de errores por carácter es de 0,053, lo que equivale aproximadamente a **un error cada 19 caracteres**.

Este análisis se estructura en dos partes:

a) Clasificación mediante los códigos T relacionados con el sistema ASR:

- [T-] Omisión de palabras o fragmentos: 26 errores.
- [Tx] Error de léxico: 12 errores.
- [T+] Adición de contenido inexistente en el audio: 4 errores.

Total de errores ASR (códigos T): 42

b) Clasificación mediante la *Tabla 1* (modelo MQM):

- F2.1 Puntuación: 32 errores.
- F1.1 Uso incorrecto de mayúsculas: 1 error.

Además, se detectaron 4 errores de «No distinción de hablante».

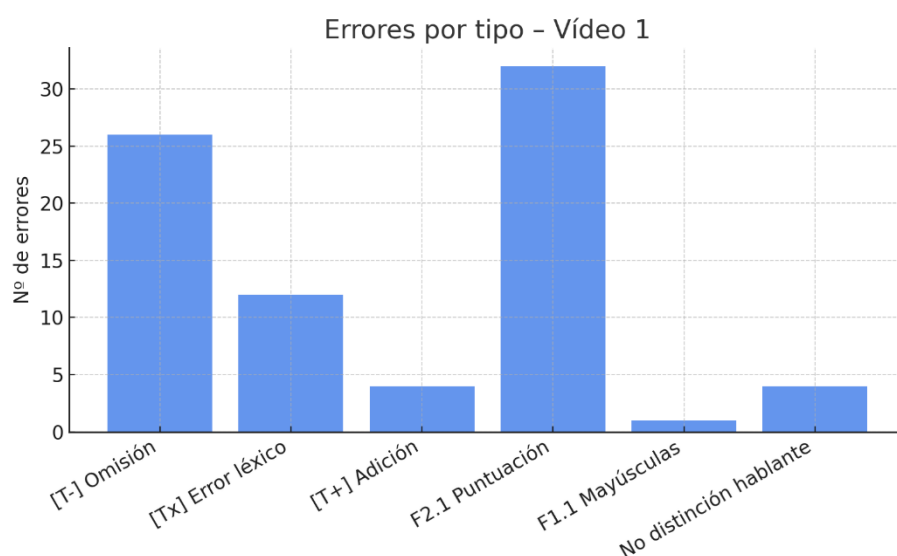


Ilustración 5. Captura de pantalla de un gráfico de barras generado con Matplotlib, que muestra el número de errores por tipo en el vídeo 1.

Seguidamente, se ha aplicado una escala basada en el impacto comunicativo del error según su impacto. En la **Tabla 2** se puede ver cuántos errores son leves, graves o críticos, y además a qué tipo o categoría pertenecen.

Tabla 2. Distribución de errores por gravedad y tipo.

Tipo de error	Leve	Grave	Crítico	Total
No distinción de hablante	—	4	—	4
T+	2	1	1	4
T-	20	5	1	26
Tx	1	2	9	12
F2.1	32	—	—	32
F1.1	1	—	—	1
Total y % sobre el total	56 (71,79%)	12 (15,38%)	10 (12,82%)	78 (100%)

Entre los errores leves más frecuentes se encuentran la ausencia de signos de puntuación y la omisión del indicador de risa:

penguins lost in a forest **[[laughter]]** Oh{,} God{,} [T-] E + F2.1 + F2.1 + F2.1

Ilustración 6. Captura de pantalla de la hoja de cálculo de Excel que muestra ejemplos de errores de puntuación y [T-], clasificados como leves.

Por el contrario, no identificar correctamente un cambio de hablante se considera un error grave, ya que, aunque no invalida el discurso, dificulta su correcta interpretación.

La presencia de 11 errores críticos refleja las claras dificultades del sistema ASR para gestionar situaciones en las que intervienen lenguaje acelerado, acentos regionales o una elevada carga humorística. Aplicando la escala de puntuación, el total obtenido asciende a 316 puntos.

A continuación, se detallan algunos de los errores marcados como críticos:

- «Those animals **only** a carpet» (*own me*)
- «And other **[]** really cause...» (omisión de *bullshit*)
- «**I'm** the cat **and watch** the cat in a **ball**» (*and, I didn't wash, bowl*)
- «**Not** a lot of people» (adición de *not*, lo correcto sería: *a lot of people*)

Por último, se ha llevado a cabo el análisis cualitativo de los errores ASR con el fin de identificar las causas que los originan. Este análisis se ha aplicado a los 42 errores marcados con códigos T, y se han dado las siguientes combinaciones:

Tabla 3. Causas de los errores ASR códigos T.

Tipo error	A	C	D	E	F	Total
T-	1	2	2	21	—	26
Tx	9	1	—	—	1	11
T+	1	—	—	1	2	4
Total	11	3	2	22	3	42

Los resultados muestran que los errores más frecuentes están relacionados con la omisión debido a la falta de audibilidad del contenido (E). A continuación, se muestra dos oraciones en la que el sistema ha omitido los ruidos de risas y arcadas, ya sea porque los considera secundarios a la conversación o porque no lo haya oído lo suficiente.

the whole room just fills with love **[[laughter]]** **[[vomiting sound]]** | **[T-] E + [T-] E**

Ilustración 7. Captura de pantalla de la hoja de cálculo de Excel que muestra ejemplos de errores clasificados como [T-] E.

[[Music]] | **[T-] E**

Ilustración 8. Captura de pantalla de la hoja de cálculo de Excel que muestra ejemplos de errores clasificados como [T-] E.

A estos le siguen los errores de léxico debidos a la velocidad y/o al acento del hablante (A). Por ejemplo:

- «then the **way it was outward** comes on (Then the **Waitrose** comes on)
- «I was **a tamati** and **the sandwich** came on» (I was **having my te** and **this advert** came on)

Tanto los errores de [T-] E como [Tx] A parecen representar desafíos especialmente relevantes para el sistema.

Además, En varios casos, el sistema omite palabras malsonantes o vulgares (D), incluso cuando se pronuncian con claridad. Una posible explicación es que el ASR esté programado para filtrar este tipo de lenguaje con el fin de cumplir las políticas de contenido de plataformas como YouTube. Sin embargo, esta censura automática puede afectar al tono original del mensaje.

moon and other **[[bullshit]]** **[[laughter]]** really cause [T-] D+ [T-] E

Ilustración 9. Captura de pantalla de la hoja de cálculo de Excel que muestra un error clasificado como [T-] D.

5.1.2. Análisis del sistema de traducción automática (NMT)

A continuación, se evalúa la calidad de la traducción automática al español proporcionada por YouTube, partiendo de los subtítulos generados automáticamente en inglés. El análisis se ha realizado sobre las 40 líneas de subtítulo, de las cuales **26** presentaban **al menos un error**, lo que representa el 65 % del total. Se han identificado **38 errores**, de los cuales 14 están directamente relacionados con fallos previos en la transcripción automática (ASR). Esto supone una media de **1,54 errores por línea**. Además, teniendo en cuenta un total de **1.657 caracteres**, se obtiene una media de 0,023 errores por carácter, lo que equivale aproximadamente a **un error cada 43 caracteres**.

Tabla 4. Tipos de errores en la traducción automática al español.

Tipo de error	Nº de errores	Porcentaje
Precisión	21	55,26 %
Convenciones lingüísticas	13	34,21 %
Estilo	3	7,89 %
No distinción de hablante	1	2,63 %
Total	38	100 %

La mayoría de los errores se corresponden con problemas de precisión, es decir, fallos que afectan al contenido del mensaje (traducciones literales, invenciones, omisiones...).

niño con problemas de confianza tan empalagosos	P2.1
y almibarados que están haciendo que los espectadores se sientan	
físicamente enfermos y a veces {incluso} se caguen en los pantalones	P2.1 + P3 P4[T-]

Ilustración 10. Captura de pantalla de la hoja de cálculo de Excel que muestra ejemplos de errores de Precisión (P).

Le siguen los errores lingüísticos formales, y en menor medida, los de estilo. Además, se ha identificado un caso adicional en el que no distingue el cambio de hablante, distinto de los ya contabilizados en la fase de transcripción automática al inglés.

5.1.2.1. Errores de Precisión (21 errores)

Tabla 5. Errores de Precisión: subtipos y frecuencia.

Subtipo	Descripción	Nº de errores
P2.1	Traducción demasiado literal	6 (2 de ellos derivados de error previo)
P2.3	Invención / sin sentido	5 (4 de ellos derivados de error previo)
P3	Omisión de contenido	3
P4	Adición de contenido	5 (4 de ellos derivados de error previo)
P5	Fragmento sin traducir	2

Dentro de la categoría de *Precisión*, una parte significativa de los errores está directamente condicionada por fallos previos en la transcripción automática. Concretamente, 10 de los 21 errores de precisión (47,61%). Esta dependencia se observa especialmente en los subtipos de invención o sin sentido (P2.3) y adición de contenido (P4), donde el 80 % de los casos están provocados por una mala transcripción del original.

A continuación, se muestran algunos ejemplos:

afectados tienen el estómago revuelto por un
manipulador adverso que

P2.3[Tx]

Ilustración 11. Captura de pantalla de la hoja de cálculo de Excel que muestra un ejemplo de P2.3[Tx] (causado por A: habla demasiado rápida y/o acento cerrado o difícil de entender).

[Música]

P4[T-]

Ilustración 12. Captura de pantalla de la hoja de cálculo de Excel que muestra un ejemplo de P4[T-] (causado por E: contenido no suficientemente audible o considerado irrelevante por el sistema.).

5.1.2.2. Errores de Convenciones Lingüísticas (13 errores)

Tabla 6. Errores de Convenciones Lingüísticas: subtipos y frecuencia

Subtipo	Descripción	Nº de errores
F1.1	Uso incorrecto de mayúsculas/minúsculas	4 (todos derivados de un error previo)
F2.1	Puntuación	1
F3.1	Morfología	1
F3.2	Palabras funcionales mal empleadas	2
F3.3	Errores de concordancia	4
F3.4	Orden de las palabras	1

En esta categoría, los errores más frecuentes están relacionados con el uso inadecuado de mayúsculas y minúsculas (F1.1) y con problemas de concordancia (F3.3). Cabe señalar que todos los errores F1.1 tienen su origen en una transcripción incorrecta del texto original realizada por el sistema de reconocimiento automático del habla (ASR). El ASR

no logra identificar correctamente factores como la entonación o el orden de los elementos dentro de una frase. En cambio, el sistema de traducción automática sí es capaz de detectar que se trata de frases distintas y aplica las mayúsculas correspondientes. No obstante, no introduce los signos de puntuación que deberían acompañar a dichas mayúsculas, lo que genera incoherencias.

estaba hacia afuera V omitó de nuevo	F1.1[F2.1]
por toda la mesa S oy el gato y miren al	F1.1[F2.1]
gato en una bola M e pasó algo similar el	F1.1[F2.1]

Ilustración 13. Captura de pantalla de la hoja de cálculo de Excel que muestra errores de tipo F1.1 derivados de un F2.1 del ASR.

5.1.2.3. Errores de Estilo (3 errores)

Tabla 7. Errores de Estilo: subtipos y frecuencia

Subtipo	Descripción	Nº de errores
E1	Redacción poco natural	3

Los tres errores de esta categoría corresponden al subtipo E1, relacionado con expresiones que presentan una redacción poco natural en español. En ocasiones, el resultado es artificial y poco idiomático. Un ejemplo de E1 puede identificarse en la siguiente oración:

enfermedades físicas Y a sabes n orovirus?	E1+ F1.1 [F2.1] + F3.2 + F2.1
--	-------------------------------

Ilustración 14. Captura de pantalla de la hoja de cálculo de Excel que muestra, entre otros, un error de tipo E1.

En este caso, «enfermedades físicas» no sería del todo incorrecto, pero sí resulta poco natural en español y afecta a la fluidez que se espera. Por tanto, una alternativa más adecuada en este contexto podría ser «malestares», ya que se ajusta al uso habitual del idioma en situaciones comunicativas similares.

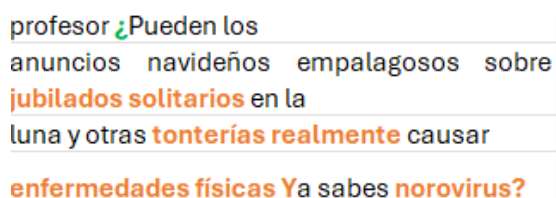
Por otro lado, los errores también han sido evaluados según su impacto sobre la comprensión, el tono y la funcionalidad comunicativa del mensaje:

Tabla 8. Errores según su gravedad.

Gravedad N° de errores Porcentaje		
Leve	28	73,68%
Grave	6	15,79%
Crítico	4	10,53%
Total	38	100%

De nuevo, la mayoría de los errores identificados son de carácter leve. No obstante, los 9 errores clasificados como graves o críticos (26,32 %) sí afectan de manera significativa a la calidad de la traducción. Al aplicar la escala de puntuación por gravedad, el total acumulado asciende a 98 puntos.

Por último, se observa una diferencia significativa en la puntuación total asignada a cada sistema en función del número y la gravedad de los errores identificados: el reconocimiento automático del habla (ASR) acumula 316 puntos, frente a los 98 de la traducción automática. Además, se pueden observar algunas mejoras en la fase de traducción con la incorporación de algunos signos de puntuación que el ASR no identificaba. Un ejemplo de ello puede verse en la Ilustración 15, donde se añade el signo de interrogación de apertura en la pregunta.



profesor ¿Pueden los
anuncios navideños empalagosos sobre
jubilados solitarios en la
luna y otras tonterías realmente causar
enfermedades físicas Ya sabes norovirus?

Ilustración 15. Captura de pantalla de la hoja de cálculo de Excel que muestra en color verde la incorporación del signo de interrogación.

5.2. Vídeo nº2

5.2.1. Análisis de la transcripción automática en inglés (ASR)

En este segundo vídeo, compuesto por 55 líneas de subtítulo, se han detectado errores en 44 de ellas, lo que supone un 80 % de líneas afectadas. El número total de errores asciende a **109**, con una media de **2 fallos por línea**. demás, teniendo en cuenta que el total de caracteres es de **1.694**, se obtiene una media de **0,064 errores por carácter**, lo que equivale aproximadamente a **un error cada 16 caracteres**.

Los errores se distribuyen del siguiente modo:

a) Según los códigos de error propios del sistema ASR:

- Omisiones ([T-]): 30
- Reconocimientos incorrectos ([Tx]): 16

Total errores: 46

b) Clasificación mediante la *Tabla 1* (modelo MQM):

- Puntuación incorrecta (F2.1): 52
- División incorrecta del contenido (F4): 1
- Error en el etiquetado de eventos sonoros (E2): 1

Además, se identificaron 9 errores por falta de distinción de hablantes, un problema recurrente que interfiere en la correcta interpretación del discurso.



Ilustración 16. Captura de pantalla de un gráfico de barras generado con Matplotlib, que muestra el número de errores por tipo en el vídeo 2.

Respecto a la gravedad de los errores detectados, los resultados se reparten de la siguiente forma:

Tabla 9. Distribución de errores por gravedad y tipo

Tipo de error	Leve	Grave	Crítico	Total
No distinción de hablante	–	9	–	9
Tx	4	8	4	16
E2	–	1	–	1
F4	–	1	–	1
F2.1	52	–	–	52
Total y % sobre el total	56 (70,9 %)	19 (24,1 %)	4 (5,1 %)	79 (100 %)

En este vídeo, la mayoría de los errores se consideran leves, con una fuerte presencia de fallos en el uso de signos de puntuación. Aun así, los errores más graves siguen teniendo un peso considerable en la comprensión, especialmente en aquellas líneas afectadas por cambios de tono o interrupciones. Algunos ejemplos ilustrativos son:

- «you are talking complete donkey **twiddle**» (*twaddle*)
- «my father's **turnout** police dress uniform» (*Trinidad*)

En cuanto al análisis cualitativo, se observan los siguientes patrones entre tipo de error y su causa probable:

Tabla 10. Causas de los errores ASR códigos T.

Tipo error	A	C	E	F	Total
T-	–	8	20	3	31
Tx	11	2	–	2	15
Total	11	10	20	5	46

Como en el resto de los fragmentos analizados, las omisiones debidas a baja audibilidad (E) vuelven a ser predominantes. También destacan los errores de reconocimiento vinculados al acento o al ritmo del habla (A). En la mayoría de los casos en los que se ha registrado un error tipo [Tx] A, el fallo coincide con las intervenciones del personaje principal de la serie, Rowan Atkinson. Su forma de hablar, entrecortada, exagerada y cargada de entonaciones, supone un reto considerable para el sistema, que no logra identificar con precisión lo que dice.

Man 1 (Rowan Atkinson): but Ritter the whole point of this	[Tx] A
---	--------

Ilustración 17. Captura de pantalla de la hoja de cálculo de Excel que muestra un error de tipo [Tx] A en la intervención de Rowan Atkinson.

Man 1 (Rowan Atkinson): mistaken{.} comfortable you're talking	F2.1 + [Tx] A
---	---------------

Ilustración 18. Captura de pantalla de la hoja de cálculo de Excel que muestra un error de tipo [Tx] A en la intervención de Rowan Atkinson.

También se observan errores provocados por sonidos que dificultan entender bien lo que se dice (C). En este vídeo hay muchas risas de fondo, y en la mayoría de los casos, un personaje empieza a hablar antes de que terminen. Como los sonidos se solapan, el sistema tiene muchos problemas para reconocer correctamente lo que se está diciendo.

evening{.} all{.} [[laughter]] {Detective Constable Poole,} TV's sex cop always ready to get on the	F2.1 + F2.1 + [T-]E + [T-]C
---	-----------------------------

Ilustración 19. Captura de pantalla de la hoja de cálculo de Excel que muestra, entre otros, un error de tipo [Tx] C.

5.2.2. Análisis del sistema de traducción automática (NMT)

La evaluación de la traducción al español revela errores en **27 de las 55 líneas analizadas**, lo que supone un **49,1 % del total**. Se han identificado **30 errores**, de los cuales **11 derivan de fallos** presentes ya en la fase de transcripción. Considerando un total de

1.735 caracteres, se obtiene una media de 0,017 errores por carácter, lo que equivale aproximadamente a **un error cada 58 caracteres**.

Tabla 11. Tipos de errores en la traducción automática al español.

Categoría principal	Nº de errores	Porcentaje
Precisión	22	75,86 %
Convenciones lingüísticas	4	13,79 %
Estilo	1	3,45 %
No distinción de hablante	1	3,45 %
Total	29	100 %

5.2.2.1. Errores de Precisión (22 errores)

Los errores más frecuentes están relacionados con la precisión del mensaje. A continuación se muestran los subtipos detectados:

Tabla 12. Errores de Precisión: subtipos y frecuencia.

Subtipo	Descripción	Nº de errores
P2.3	Invención / sin sentido	11 (10 de ellos derivados de ASR)
P2.1	Traducción literal excesiva	7
P3	Omisión de contenido	2 (1 de ellos derivado de ASR)
P5	Fragmento sin traducir	1
P1	Traducción incorrecta	1

En total, 11 de los 22 errores de precisión están directamente condicionados por una transcripción previa errónea, lo que refleja una vez más la repercusión de los fallos del sistema ASR sobre la calidad final de la traducción. En casi todos estos casos, con la excepción de uno, se trata de errores clasificados como P2.3 (invención o sin sentido), en los que el sistema de traducción ha mantenido en español el mismo contenido erróneo generado en inglés. Es razonable pensar que, de no haberse producido esos errores en la fase de transcripción, la traducción automática habría ofrecido un resultado mucho más

preciso. A continuación, se presenta un ejemplo en el que puede observarse cómo un error de transcripción en inglés afecta directamente a la traducción. El sistema interpreta la frase original «mistaken. Constable, you're talking» como «mistaken comfortable you're talking», lo que da lugar a una traducción completamente incoherente en español: «equivocado cómodo estás hablando...».

equivocado cómoda estás hablando tonterías

F3.3 + P2.3[Tx]

Ilustración 20. Captura de pantalla de la hoja de cálculo de Excel que muestra, entre otros, un error de tipo P2.3[Tx] (causado por A: habla demasiado rápida y/o acento cerrado o difícil de entender).

5.2.2.2. Errores de Convenciones Lingüísticas (4 errores)

Tabla 13. Errores de Convenciones Lingüísticas: subtipos y frecuencia.

Subtipo	Descripción	Nº de errores
F2.1	Puntuación	1
F3.2	Uso incorrecto de palabras funcionales	1
F3.3	Concordancia gramatical	2

Dentro de esta categoría, se ha identificado un caso en el que no se utilizan correctamente las palabras funcionales (F3.2). En la frase «luciendo un poco gris sintiéndome un poco viejo nuevo gris lejos te haremos parecer 10 años más», el adjetivo «nuevo» aparece sin el artículo «un» que debería precederlo según las normas gramaticales del español. Esta omisión da lugar a una construcción anómala que rompe la fluidez del mensaje y contribuye a una expresión poco natural en la lengua meta.

5.2.2.3. Errores de Estilo (1 error)

Tabla 14. Errores de Estilo: subtipos y frecuencia.

Subtipo	Descripción	Nº de errores
E1	Redacción poco natural	1

A pesar de que los errores lingüísticos y de estilo tienen una presencia más limitada, siguen contribuyendo a una sensación general de traducción poco cuidada en algunos tramos.

Respecto al impacto comunicativo de los errores, los resultados son los siguientes:

Tabla 15. Errores según su gravedad.

Grado de gravedad	Nº de errores	Porcentaje
Leve	15	51,72 %
Grave	9	31,03 %
Crítico	5	17,24 %
Total	29	100 %

Aunque más de la mitad de los errores se consideran leves, los casos graves y críticos representan casi el 50 %, y afectan de forma clara a la calidad y comprensión del subtítulo.

Algunos ejemplos son:

venido a trabajar **liciendo** respetable **porque** P2.3 + P2.1

Ilustración 21. Captura de pantalla de la hoja de cálculo de Excel que muestra errores graves y críticos.

uniforme de policía de mi padre **{de Trinidad}** P3[Tx]

Ilustración 22. Captura de pantalla de la hoja de cálculo de Excel que muestra un error grave.

La puntuación total alcanzada en esta fase asciende a 110 puntos, que, sumados a los 255 puntos del sistema ASR, da un total de 365 puntos para este vídeo. Esta cifra representa el valor más bajo entre los tres fragmentos analizados. Esto puede deberse a que los personajes elaboran frases cortas y estructuras simples. Además, la mayoría de los personajes hablan con un acento muy cercano al estándar, similar al que probablemente se emplea en el entrenamiento del sistema. También cabe destacar que, a medida que avanza el vídeo, pueden observarse algunas mejoras en la traducción. Comienzan a

aparecer comas y signos de interrogación de forma más consistente, e incluso el sistema logra reconocer correctamente ciertos fragmentos más complejos. Un ejemplo significativo es que, en esta fase, la traducción automática identifica correctamente «Boots» como un nombre propio, algo que el sistema de transcripción (ASR) no había conseguido reconocer previamente, pues lo transcribió como «boots».

bolso de otra persona en **Boots** Dios mío, Dios mío, Dios mío,

Ilustración 23. Captura de pantalla de la hoja de cálculo de Excel que muestra una mejora en el sistema de traducción automática.

5.3. Vídeo nº3

5.3.1. Análisis de la transcripción automática en inglés (ASR)

El tercer fragmento analizado consta de 39 líneas de subtítulos, de las cuales **36** contienen al menos un error, lo que equivale a un **92,3 %** de líneas afectadas. En total, se identificaron 125 errores, alcanzando una media de 3,2 fallos por línea, la cifra más alta de los tres vídeos. Una posible explicación de este aumento en la tasa de errores podría ser la mayor carga cultural y la complejidad lingüística del contenido, tanto en el léxico empleado como en la pronunciación de los hablantes. Si se considera la extensión total del texto (1.210 caracteres), la densidad de errores se sitúa en torno a **0,103 por carácter**, lo que equivale a **un error cada 10 caracteres**, aproximadamente.

A continuación, se detallan las dos partes:

a) Clasificación mediante los códigos T relacionados con el sistema ASR:

- Omisiones ([T-]): 22
- Inserciones no justificadas ([T+]): 1
- Reconocimientos léxicos incorrectos ([Tx]): 21

Total de errores: 44

b) Clasificación mediante la *Tabla 1* (modelo MQM):

- Problemas de puntuación (F2.1): 65
- Uso inadecuado de mayúsculas/minúsculas (F1.1): 3
- Morfología incorrecta (F3.1): 3
- Uso erróneo de palabras funcionales (F3.2): 1
- Concordancia gramatical deficiente (F3.3): 1

Total de errores: 73

Además, se identificaron 9 casos en los que el sistema no logra distinguir adecuadamente entre diferentes hablantes.



Ilustración 24. Captura de pantalla de un gráfico de barras generado con Matplotlib, que muestra el número de errores por tipo en el vídeo 3.

En cuanto a la gravedad de los errores detectados, los resultados obtenidos en este tercer vídeo son los siguientes:

Tabla 16. Distribución de errores por gravedad y tipo.

Tipo de error	Leve	Grave	Crítico	Total
F2.1	65	—	—	65
F1.1	3	—	—	3
F3.1	2	1	—	3
F3.2	1	—	—	1
F3.3	—	1	—	1
No distinción de hablante	—	9	—	9
T-	14	8	—	22
T+	1	—	—	1
Tx	1	11	9	21
Total y % sobre el total	88 (69,84 %)	29 (23,02 %)	9 (7,14 %)	126 (100 %)

La mayoría de los errores detectados son de carácter leve, sobre todo aquellos relacionados con la falta de signos de puntuación.

be strong{.} my darlings{.} because there's	F2.1 + F2.1
something else you should know{.} something	F2.1
much worse{.} this person{.} the agency are	F2.1 + F2.1 + F3.1

Ilustración 25. Captura de pantalla de la hoja de cálculo de Excel que muestra errores de puntuación.

No obstante, los fallos más graves, aunque menos numerosos, tienen un impacto considerable en la comprensión, especialmente cuando coinciden con acentos marcados o ritmo acelerado. En total, la puntuación total alcanzó los 323 puntos.

Entre los errores críticos se incluyen omisiones de fragmentos importantes, como sonidos dramáticos que forman parte del efecto cómico, así como fallos léxicos derivados del desconocimiento de ciertos dialectos del inglés. Este tipo de errores se hace especialmente

evidente en expresiones propias del norte de Inglaterra, como *pair of tykes*, que ha sido transcrito incorrectamente como *pair of takes*, o *an'er* (contracción coloquial de *aren't I*), que ha sido interpretado por el sistema como «AR». La velocidad del habla, unida a la pronunciación cerrada y a la complejidad léxica, contribuye a que el sistema proporcione resultados muy alejados del contenido original. También se han identificado errores relacionados con el reconocimiento incorrecto de vocabulario, como ocurre en la frase «We were beginning to think you didn't want your bage», donde el sistema transcribe *bage* en lugar de *brioche*. Este tipo de fallos puede deberse a la presencia de préstamos lingüísticos, en este caso, un galicismo, que no forman parte del léxico propio del inglés y que, por consiguiente, sistema no es capaz de procesar adecuadamente. Ante esta dificultad, tiende a sustituir la palabra desconocida por otra fonéticamente similar dentro del idioma objetivo, generando resultados incorrectos pero plausibles desde un punto de vista acústico.

Por otro lado, el análisis cualitativo permite profundizar en las causas de los errores detectados. A continuación se detallan las combinaciones observadas entre tipo de error y su origen estimado:

Tabla 17. Causas de los errores ASR códigos T.

Tipo error	A	B	C	D	E	F	Total
T-	–	–	1	–	21	–	22
Tx	4	11	4	–	–	1	20
T+	1	–	–	–	–	–	1
Total	5	11	5	–	21	1	43

Las omisiones relacionadas con falta de audibilidad (E) encabezan la clasificación, seguidas por errores de reconocimiento vinculados a las referencias culturales (B), al habla (A) o a la interferencia por otros sonidos. En el caso de las causas clasificadas como B y A, una parte importante de los errores están relacionados con las intervenciones de la niñera, quien combina un acento cerrado del norte de Inglaterra con un estilo de habla coloquial. Otro error recurrente a lo largo del vídeo está relacionado con la incapacidad

del sistema para distinguir entre variantes del inglés, concretamente entre el británico y el estadounidense, a pesar de ser las dos variedades más comunes. Un ejemplo significativo se da cuando los niños se refieren a su madre con el término «mummy», propio del inglés británico. Sin embargo, el sistema lo transcribe erróneamente como «mommy», forma característica del inglés americano. Aunque pueda parecer un detalle menor, este tipo de errores refleja la falta de adaptación del sistema al contexto geográfico y cultural del contenido, lo que compromete la adecuación del subtítulo al público objetivo.

there you are{,} Mommy{,} we were beginning to F1.1 + F2.1 + [Tx]B + F2.1

Ilustración 26. Captura de pantalla de la hoja de cálculo de Excel que muestra, entre otros, un error de [Tx]B.

5.3.2. Análisis del sistema de traducción automática (NMT)

En cuanto a la traducción automática al español, se detectaron errores en 27 de las 39 líneas, lo que representa un **69,2 % del total**. En total, se identificaron **35 errores**, de los cuales 11 están directamente relacionados con fallos previos del sistema de transcripción (ASR). En relación con la extensión del subtítulo (1.226 caracteres), la densidad de errores se sitúa en torno a **0,029 por carácter**, lo que equivale aproximadamente a **un error cada 34 caracteres**.

Tabla 18. Tipos de errores en la traducción automática al español.

Categoría principal	Nº de errores	Porcentaje
Precisión	22	59,46 %
Convenciones lingüísticas	11	29,73 %
Estilo	4	10,81 %
Total	37	100 %

5.3.2.1. Errores de Precisión (24 errores)

Como ya se observó en los vídeos anteriores, los errores de precisión vuelven a ser los más frecuentes en este caso. A continuación, se detallan los distintos subtipos identificados:

Tabla 19. Errores de Precisión: subtipos y frecuencia.

Subtipo	Descripción	Nº de errores
P2.3	Invención / sin sentido	12 (10 de ellos derivados de ASR)
P2.1	Traducción excesivamente literal	7 (1 de ellos derivado de ASR)
P4	Adición de contenido	1
P5	Fragmento sin traducir	2 (1 de ellos derivado de ASR)

Los errores del subtipo P2.3 son especialmente problemáticos, pues generan mensajes sin sentido. En este grupo, la influencia de una transcripción deficiente es muy significativa. Por ejemplo, el sistema traduce «CL» en lugar de «Chloe», o «en T» en lugar de «a tiempo», manteniendo las distorsiones introducidas previamente en la transcripción al inglés.

5.3.2.2. Errores de Convenciones Lingüísticas (11 errores)

Tabla 20. Errores de Convenciones Lingüísticas: subtipos y frecuencia.

Subtipo	Descripción	Nº de errores
F3.1	Morfología	3 (1 de ellos derivado de ASR)
F3.3	Concordancia	5
F3.4	Orden de las palabras	2
F2.2	Interrogación/exclamación incorrecta	1

La mayoría de estos errores no dependen directamente de fallos en la transcripción, sino que se deben a limitaciones propias del sistema de traducción automática a la hora de generar estructuras sintácticas naturales y correctas en español. En este caso concreto, se observa una tendencia constante a traducir el pronombre «you» como «tú», sin tener en cuenta que, en muchos contextos del vídeo, el interlocutor es plural (*vosotros*).

Un ejemplo representativo es la frase original «Thomas, Chloe, Mummy has something to tell you», en la que el sistema produce la traducción «mami tiene algo que decirte», ignorando que se trata de un mensaje dirigido a más de una persona. Es probable que, ante la duda, el sistema opte por la opción más frecuente o segura: el singular.

5.3.2.3. Errores de Estilo (2 errores)

Dos de los errores detectados corresponden a expresiones poco naturales en español (E1), que no comprometen directamente el significado del mensaje, pero sí afectan a su fluidez y a la naturalidad del discurso. A esto se suman otros dos errores de tipo E2, relacionados con un estilo inadecuado que no respeta el tono ni el registro del original.

Un ejemplo claro se da cuando la niñera utiliza la expresión «hiya», un saludo informal y afectuoso característico del norte de Inglaterra. En la traducción, este matiz se pierde por completo al ser sustituido por un simple «hola», lo que elimina la referencia sociolingüística implícita. Esta omisión resulta especialmente relevante en este vídeo, cuyo humor se apoya en gran parte en la representación irónica de las diferencias sociales británicas a través de las formas del habla.

Para concluir esta sección, se presenta la valoración de los errores en función de su gravedad:

Tabla 21. Errores según su gravedad.

Grado de gravedad N° de errores Porcentaje		
Leve	19	54,3 %
Grave	15	42,9 %
Crítico	3	8,6 %
Total	35	100 %

Nuevamente, aunque la mayoría de los errores se consideran leves, los clasificados como graves o críticos, que representan casi la mitad del total, afectan directamente a la comprensión del mensaje en partes clave. En conjunto, la traducción automática alcanza una puntuación de 124 puntos. Sumando ambas fases, el total acumulado asciende a 447 puntos, la cifra más alta registrada entre los tres vídeos. Además, a diferencia de los otros vídeos, en los que la fase de traducción automática logró corregir parcialmente algunos errores procedentes del ASR, en este caso la capacidad de recuperación es mucho más limitada. Únicamente se ha detectado la incorporación de algunos signos de interrogación, y en una proporción menor que en los otros fragmentos analizados.

6. Observaciones adicionales

Durante el desarrollo de este trabajo y al margen del análisis realizado, se ha podido comprobar que algunos de los vídeos analizados presentaban cambios en los subtítulos automáticos ofrecidos por YouTube. En ciertos casos, estos ajustes introducían mejoras en aspectos como la puntuación y, ocasionalmente, en la selección léxica, lo que daba lugar a versiones algo más precisas y acordes con el tono original del mensaje, incluido el humor. Aunque estas variaciones no se han evaluado de forma cuantitativa ni forman parte del análisis, su aparición sugiere que el sistema de subtitulado automático de la plataforma podría estar sujeto a procesos de actualización o ajuste tras la publicación del vídeo, ya sea por mejoras en los modelos de reconocimiento y traducción o por retroalimentación del propio sistema. Este hecho invita a considerar la dimensión evolutiva de este tipo de herramientas.

7. Análisis comparativo final

El análisis conjunto de los tres vídeos permite extraer algunas conclusiones transversales sobre el funcionamiento de los sistemas de subtítulo automático, tanto en la fase de transcripción como en la de traducción. Aunque cada fragmento presenta sus propias particularidades, se repiten ciertos problemas estructurales que evidencian las limitaciones del sistema.

Uno de los elementos más destacables es que el sistema parece tener una capacidad muy reducida para adaptarse al contexto sociolingüístico y cultural de los vídeos. Esto se aprecia, por ejemplo, en su tendencia a ignorar los dialectos regionales del inglés (como el del norte de Inglaterra en el vídeo 3) o en su incapacidad para distinguir entre formas británicas y americanas, como «mummy» y «mommy».

También resulta evidente la falta de flexibilidad para interpretar el número del interlocutor en la traducción, con errores repetidos como la traducción sistemática de «you» por «tú», incluso cuando claramente se alude a más de una persona. Este fallo, más allá de lo gramatical, revela una limitación en la capacidad del sistema para inferir información pragmática del discurso. Además, la falta de detección de cambios de hablante refuerza este problema, ya que impide al sistema interpretar correctamente quién habla y a quién se dirige el mensaje.

Otro aspecto especialmente problemático es la gestión de los ruidos. En los tres vídeos, el sistema tiende a ignorar risas, sonidos ambientales o efectos dramáticos, como si no formaran parte del discurso, sin marcar su presencia ni indicar pausas. Esta omisión no solo afecta al tono del mensaje, sino que también dificulta la interpretación del contenido, especialmente en contextos humorísticos donde el sonido cumple una función expresiva.

Además, el sistema de ASR todavía no es capaz de filtrar eficazmente los ruidos externos, lo que interfiere directamente en su capacidad para reconocer el habla. Como resultado, se producen omisiones de fragmentos completos o errores de reconocimiento léxico graves, no tanto porque no sepa lo que se dice, sino porque no puede oírlo bien. Por lo tanto, todavía hay limitaciones técnicas en la detección y el procesamiento del audio.

Asimismo, a lo largo de los tres vídeos se mantiene una constante: el sistema arrastra los errores de una fase a otra, lo que refuerza la idea de que la calidad de la traducción automática depende en gran medida del rendimiento del ASR. Cuando este falla, la NMT

no es capaz de reinterpretar o corregir el mensaje, sino que mantiene o incluso amplifica los errores. Es decir, hay una transmisión directa de fallos entre fases, sin mecanismos efectivos de recuperación.

Por otro lado, aunque se observan pequeñas mejoras puntuales, como la inclusión de signos de interrogación o el reconocimiento correcto de algunos nombres propios, estas correcciones no son sistemáticas y parecen más casuales que resultado de una evolución en el tratamiento del contenido.

De esta forma, el sistema automatizado analizado muestra serios problemas para gestionar situaciones reales de comunicación oral, especialmente cuando estas incluyen acentos no estándar, referencias culturales, humor o habla acelerada. No solo se ve superado por las condiciones acústicas y lingüísticas, sino que carece de mecanismos de compensación que le permitan mejorar o ajustar la salida en función del contexto.

CONCLUSIONES

El presente trabajo ha permitido evaluar de forma rigurosa la calidad de los subtítulos automáticos en español generados por YouTube en vídeos de humor británico, a través del análisis de dos fases clave del proceso: el reconocimiento automático del habla (ASR) y la traducción automática neuronal (NMT). La elección de contenidos con alta carga cultural y lingüística ha demostrado ser un criterio eficaz para poner a prueba los límites actuales de estas tecnologías.

Los resultados confirman la hipótesis inicial: los subtítulos automáticos acumulan numerosos errores que afectan tanto a la comprensión del mensaje como a la transmisión efectiva del humor. En la fase del ASR, destacan las omisiones, así como los errores léxicos y de puntuación, que se ven agravados por factores como los acentos regionales, la rapidez del habla y el ruido ambiental. A su vez, la traducción automática tiende a reproducir estos errores, dando lugar a formulaciones inexactas y dificultando la adaptación de referencias culturales y los matices humorísticos. Aunque se observan algunas mejoras puntuales, son insuficientes para garantizar una experiencia comunicativa satisfactoria.

En conjunto, puede decirse que los subtítulos automáticos de YouTube pueden ofrecer una comprensión general del contenido, pero no alcanzan los estándares mínimos de calidad necesarios para textos con alta complejidad comunicativa. Este análisis refuerza, por tanto, la necesidad de la intervención humana en entornos que requieren sensibilidad lingüística y cultural. Si bien el subtitulado automático puede servir como apoyo inicial o pretraducción, no puede sustituir al criterio profesional en textos donde el contexto, la intención y el tono son elementos clave.

Más allá de los resultados técnicos, este trabajo también ha permitido poner en práctica las competencias adquiridas a lo largo del Grado, aplicándolas a un caso real y actual. Además, el proceso ha servido para reflexionar de forma crítica sobre los límites de la automatización en tareas tradicionalmente humanas.

Finalmente, los hallazgos obtenidos abren la puerta a nuevas líneas de investigación. Entre ellas, cabe plantearse si los subtítulos automáticos de YouTube evolucionan con el tiempo tras la publicación del vídeo, si existen mecanismos internos de revisión automática que los mejoren, o si el rendimiento del sistema de reconocimiento del habla podría optimizarse mediante el entrenamiento con diferentes dialectos y acentos. También podría analizarse el impacto de modelos híbridos que combinen automatización con revisión humana. En definitiva, aunque el subtitulado automático ha avanzado mucho en los últimos años, aún queda un largo camino por recorrer para que sea realmente eficaz en contextos de alta exigencia comunicativa.

BIBLIOGRAFÍA

- Ahmad, Zahra. 2022. «Introduction to Automatic Speech Recognition». *Medium*. Acceso el 19 de abril de 2025. <https://syrian.medium.com/introduction-to-automatic-speech-recognition-14f1a10704e6>.
- AffMaven. 2024. «YouTube Statistics 2024». *AffMaven*. Acceso el 20 de abril de 2025. <https://affmaven.com/es/youtube-statistics/>.
- Amazon Web Services. 2021. «¿Qué es el reconocimiento automático del habla (ASR)?». Acceso el 1 de junio de 2025. <https://aws.amazon.com/what-is/speech-to-text/>.
- Ashraf, Mudasir. 2024. «Innovations and Challenges in Neural Machine Translation: A Review». *International Journal of Science and Research (IJSR)* 13, n.º 10: 656–662. <https://dx.doi.org/10.21275/SR241008113336>.
- Bahdanau, Dzmitry, Kyunghyun Cho y Yoshua Bengio. 2015. «Neural Machine Translation by Jointly Learning to Align and Translate». En *Actas de la conferencia ICLR 2015*. <https://doi.org/10.48550/arXiv.1409.0473>.
- Baños, Rocío. 2023. «Key challenges in using automatic dubbing to translate educational YouTube videos». *Linguistica Antverpiensia, New Series: Themes in Translation Studies* 22: 30–54. <https://doi.org/10.52034/lans-tts.v22i.763>.
- Botella Tejera, Carla. 2017. «La traducción del humor intertextual audiovisual. Que la fuerza os acompañe». *MonTI. Monografías de Traducción e Interpretación*, n.º 9: 77–100. Acceso 12 de mayo de 2025. <https://raco.cat/index.php/MonTI/article/view/338762>.
- British Comedy Guide. s. f. «The Mash Report». Acceso el 27 de mayo de 2025. https://www.comedy.co.uk/tv/the_mash_report/.
- Davila Silva, Melany, Melany Rodríguez, María Béjar y Dayana Montes. 2023. «Problemas y estrategias en la traducción humorística de sitcoms». *MANTSIARI* 1, n.º 1: 62–75. <https://doi.org/10.33539/mant.v1i1.3119>.
- Díaz Cintas, Jorge. 2003. *Teoría y práctica de la subtitulación. Inglés-Español*. Barcelona: Ariel.

Díaz Cintas, Jorge y Aline Remael. 2021. *Audiovisual Translation: Subtitling*. 3.^a ed. Londres: Routledge

Do Campo Bayón, María. 2019. «La traducción automática dentro del contexto de una lengua minorizada: ¿Qué tipo de motor se adapta mejor al caso especial del gallego?». Trabajo de Fin de Máster. Universitat Autònoma de Barcelona. Acceso el 15 de junio de 2025.

https://ddd.uab.cat/pub/trerecpro/2019/hdl_2072_359319/Mariado_Campo_TFM.pdf.

Forcada, Mikel L., Felipe Sánchez Martínez y Juan Antonio Pérez Ortiz. 2016. *Manual de informática y de tecnologías para la traducción*. Alicante: Universidad de Alicante. Acceso el 12 de mayo de 2025. <http://hdl.handle.net/10045/53085>.

Formación Avanzada Europea. 2024. «La historia de YouTube: desde sus humildes comienzos hasta la gigantesca plataforma de video». *Formación Avanzada Europea*. Acceso 20 de abril de 2025. <https://www.formacionavanzadaeuropea.com/post/la-historia-de-youtube-desde-sus-humildes-comienzos-hasta-la-gigantesca-plataforma-de-video>.

García, José. 2021. «Los subtítulos automáticos en tiempo real de Google Chrome ya están disponibles: así funcionan y se activan». *Xataka*. Acceso el 19 de marzo de 2025. <https://www.xataka.com/aplicaciones/subtitulos-automaticos-tiempo-real-google-chrome-estan-disponibles-asi-funcionan-se-activan>.

Georgakopoulou, Alexandra. 2020. «Automatic Subtitling in the Era of AI: Challenges and Perspectives.» *Journal of Specialised Translation*. Acceso el 20 mayo de 2025. <https://www.jostrans.org/article/download/7487/6989?inline=1>.

Google Cloud. s.f. «Speech-to-Text Documentation». Acceso el 1 de junio de 2025. <https://cloud.google.com/speech-to-text/docs>.

Google Cloud Translation API. s.f. «Cloud Translation Documentation». Acceso el 1 de junio de 2025. <https://cloud.google.com/translate/docs>.

Hannun, Awni, Carl Case, Jared Casper, Bryan Catanzaro, *et al.* 2014. «Deep Speech: Scaling up End-to-End Speech Recognition». arXiv (1412.5567). <https://arxiv.org/abs/1412.5567>.

Jurafsky, Daniel, y James H. Martin. 2024. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*. Third Edition Draft. January 12, 2025. [S. l.]: s. n. Acceso el 21 de mayo de 2025. <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>.

Kenny, Dorothy. 2022. «Machine Translation for Subtitling: Challenges and Opportunities». *The Journal of Specialised Translation* 37: 2-19. Acceso el 10 de febrero de 2025. <https://zenodo.org/records/6653406>.

Multidimensional Quality Metrics. 2025. «The MQM Framework». Acceso el 3 de abril de 2025. <https://themqm.org/>.

OBS Studio. 2025. «OBS – Open Broadcaster Software». Acceso el 2 de marzo de 2025. <https://obsproject.com/es>.

Ortiz Boix, Carla. 2015. «Evaluación de la calidad de la traducción audiovisual: estudio empírico sobre subtitulación para sordos en televisión». Tesis doctoral. Universitat Autònoma de Barcelona. Acceso el 17 de enero de 2025. <https://www.tdx.cat/bitstream/handle/10803/400020/cob1de1.pdf?sequence=1&isAllowed=y>.

Pangeanic. 2022. «Estimación de calidad en traducción automática: Un análisis exhaustivo». *Pangeanic*. Acceso el 15 de mayo de 2025. https://blog.pangeanic.com/es/tecnicas-medir-calidad-traduccion-automatica?utm_source.

Sonix.ai. 2025. «Transcripción automática de audio y vídeo en línea». Acceso el 15 de mayo de 2025. <https://sonix.ai/es>.

Talaván, Nuria, Ignacio Ávila-Cabrera y Diego Costal. 2016. *El subtitulado y la accesibilidad: función traductora y social*. Barcelona: Editorial UOC.

Valleriani, Luca. 2024. «Accent enregisterment through humour in The Catherine Tate Show». *English Language and Linguistics* 28, n.º 4: 730–750. <https://doi.org/10.1017/S1360674324000650>.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, *et al.* 2017. «Attention is All You Need». En *Advances in Neural Information Processing Systems*, vol. 30, 5998–6008. Acceso el 17 de abril de 2025.

https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Vélez Guasca, Yury Daniela. 2020. «Análisis descriptivo de la traducción subtitulada y doblaje de culturemas en la película Forrest Gump». Tesis de Grado. Universidad del Valle. Acceso el 17 de mayo de 2025. <https://bibliotecadigital.univalle.edu.co/server/api/core/bitstreams/1444b4d5-b6b3-4580-b906-b2be6f5dd44c/content>.

WordReference. 2025. «Traducción de ‘posh’». Acceso el 17 de mayo de 2025. <https://www.wordreference.com/es/translation.asp?tranword=posh>.

YouTube Help. s.f. «Get automatic captions for your videos». Accedido el 1 de junio de 2025. <https://support.google.com/youtube/answer/6373554?hl=en>.

Yvon, François. 2023. «Machine Translation». *Oxford Bibliographies Online*. <https://doi.org/10.1093/obo/9780199772810-0170>.

Zabalbeascoa Terran, Patrick. 2001. «La traducción del humor en textos audiovisuales». En *La traducción para el doblaje y la subtitulación*, coordinado por Miguel Duro, 251–263. Madrid: Cátedra, Signo e Imagen. Acceso el 29 de mayo de 2025. https://www.academia.edu/3239437/La_traducci%C3%B3n_del_humor_en_textos_audiovisuales.

ANEXOS

Anexo I. Hoja de análisis vídeo 1 (Excel)

AUDIO ORIGINAL INGLÉS	TRANSCRIPCIÓN SUBTITULADO INGLÉS	ANÁLISIS	Nº ERRORES	Nº CARACTERES	TRANSCRIPCIÓN SUBTITULADO ESPAÑOL	ANÁLISIS	Nº ERRORES Y CAUSA	Nº CARACTERES	PROPUESTAS DE MEJORA	COSAS QUE HA RECUPERADO
	[00:00:00 – 00:00:19] Man 1 (Presenter): Britain is in the grip of a vomiting epidemic caused by nauseating TV adverts				[00:00:00 – 00:00:19] Hombre 1 (Presentador): 36 Gran Bretaña está en las garras de una epidemia de vómitos causada por los 42 nauseabundos anuncios de televisión	P2.1	1	38	en medio de	
for big shops [laughter]. The lavish Christmas commercials,	for big shots [laughter]] (.) the lavish Christmas commercials(.) typically featuring some kind of bear or	[Tx] C+ [T-] E + F2.1 F2.1	3 1		14 de los peces gordos 32 los lujosos anuncios navideños que 41 suelen presentar algún tipo de oso o niño con problemas de confianza tan 39 empalagosos y alimbarados que están haciendo que los 41 espectadores se sientan	F3.2 + P2.1[Tx] P2.1	2	71 19 34 36	que emiten las grandes empresas	
child with confidence issues, are so syrupy	child with confidence issues(.] (are) so syrupy	F2.1 + [T-] C	2		41 típicamente enfermos y a veces (inclusive) se 11 caguen en los pantalones	P2.1 + P3 P4[T-]	1	47	de autoestima	
	and cloying that they're causing viewers				[00:00:20 – 00:00:48] Woman 1 (Donna Sheridan): I was a tamati and the sandwich came on	P2.1 + P3 P4[T-]	2	64 33 24		
shit their pants [laughter].	to be physically sick and sometimes even (shit) their pants [laughter]](.) [00:00:20 – 00:00:48] Woman 1 (Donna Sheridan): I was a tamati and the sandwich came on	[T-] D + [T-] E+ F2.1	3		40 Yo era un tamati y aprecié el sándwich 41 la navidad y está 38 muy nerviosa y los niños cantando 37 y pero luego es brillante y y luego toda la habitación se llena de amor	P2.1 + P3 P4[T-]	1	38	estaba tomando té y apareció este anuncio	
I was having my [me] tea and this advert came on	about a mum watching a nativity play and she's really nervous and kids singing (and) but then it's brilliant and (and) then the whole room just fills with love [laughter] [vomiting sound]].	[Tx] A+ [Tx] A F2.1 + [T-] A [T-] E + [T-] E	2 2 2		40 Yo era un tamati y aprecié el sándwich 41 la navidad y está 38 muy nerviosa y los niños cantando 37 y pero luego es brillante y y luego toda la habitación se llena de amor	P2.3[Tx]+P2.3 P2.1 P4[T-]	2 1 1	35	obra de teatro navideña	Y
and. But then it's brilliant and and then	the whole room just fills with love [laughter] [vomiting sound].	F2.1 + [T-] A [T-] E + [T-] E	2 2		Estoy (estoy) simplemente enfermo 40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
I'm I'm just sick immediately in my lap [soft laughter].	I'm I'm just sick immediately in my lap [soft laughter]](.)	[T-] E + F2.1	2		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
some came out of my nose. Then the Waitrose comes on. I puked again all	some came out of my nose(.)] then the way it was outward comes on(.)] I puked again all over the table [soft laughter]](.)] I'm the cat(.)] and watch the cat in a ball [soft laughter]](.)] I had a similar thing last [00:00:49 – 00:01:18] Man 2 (Wayne Hayes): when I watched one about a baby penguins lost in a forest [laughter]. Oh, God. and he makes friends with an owl [laughter] [vomiting sound]] (Oh, God) and then the word love appears in the sky [laughter] [vomiting sound]]. Oh, God. God. Those animals owe me a carpet [laughter].	F2.1 + [Tx] A [Tx] A+ F2.1 [T-] E+ F2.1 + [Tx] A F2.1 + [Tx] A [Tx] A+ [T-] E+ F2.1 + No distinción hablante	2 2 5 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
when I watched one about baby	penguins lost in a forest [laughter]. Oh, God. and he makes friends with an owl [laughter] [vomiting sound]] (Oh, God) and then the word love appears in the sky [laughter] [vomiting sound]]. Oh, God. God. Those animals owe me a carpet [laughter].	[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	1 4 3 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
And he makes friends with an owl [laughter] [vomiting sound] . Oh, God. And	God. Those animals owe me a carpet [laughter].	[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	1 4 3 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
God. Those animals owe me a carpet [laughter].		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	1 4 3 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
Professor Henry Brubaker joins us. Professor, can sickly sweet Christmas	Professor Henry Brubaker joins his(.) professor(.]) can sickly-sweet christmas adverts about lonely pensioners on the	[Tx] A + F2.1 F2.1 + F1.1	2 2		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
moon and other bulshit [laughter] really cause	moon and other [bulshit] [laughter]] really cause	[T-] D+ [T-] E	2		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
physical illness? Mm, you know norovirus, the	physical illness(?) (Mm) you know norovirus(.)] the	F2.1 + [T-] E+ F2.1 + No detección hablante	4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
so-called winter vomiting bug, is a	so-called winter vomiting bug(.)] it's a	F2.1 + [T-] F	2		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
myth spread by John Lewis [laughter]. Actually,	myth spread by John Lewis [laughter]](.)] actually(.]	[T-] E + F2.1 + F2.1	3		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
turned by manipulative adverts that	turned by manipulative adverse that you used weaponized emotion to make us buy pants and chocolate [laughter]](.)	[Tx] A + [T-] F [T-] E+ F2.1	2 2		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
pants and chocolate [laughter].		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo por toda la mesa Soy el gato y miren al gato en una bola Me pasó algo similar el	E1 (toda la oración) + P3 + F3.3 + F3.4/ E1 (toda la oración) + P3 + 4 F3.3 + F3.4 F3.1 + P.4 F1.1[F2.1] F1.1[F2.1] F1.1[F2.1]	2 1 1 1	53 60 35 39	Me, me siento mal al instante, como si todo cayera en mí regazo	
		[T-] E + F2.1 + F2.1 + F2.1 [T-] E+ [T-] E+ [T-] C F2.1 + [T-] E+ [T-] E F2.1 F2.1 + [Tx] A + [T-] E F2.1	4 2 4 4 4		40 inmediatamente en mi regazo se me salió algo de la nariz y luego aparece la forma en que estaba hacia afuera Vomité de nuevo 					

Anexo II. Hoja de análisis vídeo 2 (Excel)

AUDIO ORIGINAL INGLÉS	TRANSCRIPCIÓN SUBTITULADO INGLÉS	ANÁLISIS	Nº ERRORES	Nº CARACTERES	TRANSCRIPCIÓN SUBTITULADO ESPAÑOL	ANÁLISIS	Nº ERRORES	Nº CARACTERES	PROPUESTAS DE MEJORA	COISAS QUE HA RECUPERADO
	[Music] 00:00:18				6 [Música] 00:00:18					
The BBC not here? Good. There was a terrible queue at the chemist's, but I thought the occasion warranted a new comb [Laughter] Patricia, have you got nits? [Laughter]	Man 1 (Rowan Atkinson): (The BBC not here?) good. There was a terrible queue at the chemist's, but I thought the occasion warranted a Newcomb. (Laughter) Patricia have you got nits? (Laughter)	[T] E + F2.1	2		Hombre 1 (Rowan Atkinson): 38 bueno había una cola terrible en la 35 farmacia pero pensé que la ocasión 37 ameritaba un Newcomb (buenos linderos	P2.3 [T]	1	35	Bien.	
No, I haven't got nits. I've had my hair done. I'm trying to look nice for the telly.	Woman 1 (Blonde woman): Now I haven't got nits. I've had my hair done. I'm trying to look nice for the telly.	[T4] C + F2.1	2		Mujer 1 (Mujer rubia): 39 ahora no tengo linderos me he arreglado el pelo 39 estoy tratando de verme bien para la 4 telly.	P5 [T4]	1	37	signo interrogación	
Patricia, the whole point of this exercise is to show us as we are.	Man 1 (Rowan Atkinson): But fitter the whole point of this exercise is to show us as we are. what sort of example are you setting to the constables? (Laughter) Gotcha [Incomprehensible]. What are you thinking of? (F)	[T4] A	1		Hombre 1 (Rowan Atkinson): 34 pero fitter el objetivo de esta 33 ejercicio es mostrarnos tal como somos ¿qué 44 tipo de ejemplo estás dando a los alguaciles? 37 gases gástricos felix ¿en qué estás 16 pensando tus piernas tus piernas	P2.1 + P1	2	46	qué ejemplo/pelicos	signo interrogación
What? Your legs, your legs.	Woman 2 (Young woman): what? your legs. your legs.	F2.1 + [T] E + [T] C + F4 + F2.1	5		Mujer 2 (mujer joven): 34 pero fitter el objetivo de esta 33 ejercicio es mostrarnos tal como somos ¿qué 44 tipo de ejemplo estás dando a los alguaciles? 37 gases gástricos felix ¿en qué estás 16 pensando tus piernas tus piernas	P2.3 [T4]	1	35	signo interrogación	
you naughty girl. They have no business [Laughter] being here. Get rid of them. [Laughter] What image thing, sir? It's about personal	Man 1 (Rowan Atkinson): you naughty girl. They have no business. (Laughter) being here. get rid of them. (Laughter) What image thing, sir? It's about personal	F2.1	1		Hombre 1 (Rowan Atkinson): 34 pero fitter el objetivo de esta 33 ejercicio es mostrarnos tal como somos ¿qué 44 tipo de ejemplo estás dando a los alguaciles? 37 gases gástricos felix ¿en qué estás 16 pensando tus piernas tus piernas	P3	1	32		
empowerment. It's about me being the me I want to be. Unless I'm very much	Woman 2 (Young woman): empowerment. It's about me being the me I want to be. Unless I'm very much	F2.1	1		Mujer 2 (mujer joven): 40 empoderamiento personal se trata de que 35 sea la yo que quiero ser			39		
mistaken. Constable, you're talking complete donkey twaddle. [Laughter] I'm not talking donkey	Man 1 (Rowan Atkinson): mistaken. comfortable you're talking complete donkey twaddle. (Laughter) I'm not talking donkey	F2.1 + [T4] A	2		Hombre 1 (Rowan Atkinson): 35 y esa no está muy 38 equivocada cómoda estás hablando tonterías	P2.3 [T4]	1	18		
donkey twaddle, sir. I have an obligation to be a	Woman 2 (Young woman): twaddle. (sir.) I have an obligation to be a role model to show young girls that being a strong and in control woman does not mean denying my essential femininity.	F2.1 + F2.1	2		Mujer 2 (mujer joven): 41 hablando 35 torerías señor tengo la obligación de ser un 40 modelo a seguir para mostrarle a las jóvenes que 40 ser una mujer fuerte y en control no significa negar mi feminidad esencial	P2.1	1	48	ejemplo a seguir	
Or to put it another way, you want to look nice on the television. [Laughter] I think it's my duty, sir. As a feminist [Laughter]	Man 1 (Rowan Atkinson): or what another way. (sir.) you want to look nice on the television. (Laughter) I think it's my duty. (sir.) as a feminist. (Laughter)	F2.1	1		Hombre 1 (Rowan Atkinson): 36 o de qué otra manera quieras verte 36 en la televisión como feminista	P2.1	1	34	dicho de otra forma	
Really? What has got into you girls today? Vanity, thy name is woman. You will spend your next break taking your skirt down. [Laughter]	Man 1 (Rowan Atkinson): Really? What has got into you girls today? vanity. (sir.) thy name is woman. (sir.) you will spend your next break taking your skirt down. (Laughter)	[T] C + F2.1	2		Hombre 1 (Rowan Atkinson): 37 qué os pasa hoy en día chicas? 39 vanidad tu nombre es mujer pasarás 38 tu próximo descanso bajándote la falda	P2.1	1	34	Como diría Shakespeare: la vanidad tiene nombre de mujer	signo interrogación
[Laughter] [Inaudible] Good morning to you, sir. At least you've	Man 2 (Man with the police helmet): [Laughter] [Inaudible] good morning to you. (sir.) at least you've	[T] E + [T] F + F2.1 + F2.1 + No distinción hablante	5		Hombre 2 (Hombre con casco de policía): 40 buenos días a usted señor al menos ha			39		
come to work looking respectable. Why	Man 1 (Rowan Atkinson): come to work looking respectable. (sir.)	F2.1 + No distinción del hablante	2		Hombre 1 (Rowan Atkinson): 36 venido a trabajar siendo respetable porque	P2.3 + P2.1	1	44	luciendo. / Oh,	
thank you. [Laughter]	Man 2 (Man with the police helmet): thank you. (Laughter)	F2.1 + [T] E	2		9 gracias 11 [Risal] Buenos			7		
Evening, all. [Laughter] Detective Constable Poole, TV's sex cop always ready to get on the job.	Man 3 (Man with glasses): evening. (all.) (Laughter) Detective Constable Poole. TV's sex cop always ready to get on the job. [Laughter]	F2.1 + F2.1 + [T] E + [T] C	4		Hombre 3 (Hombre con gafas): 51 noches, toda la 4 televisión es un policía sexual listo para ponerse a 10 trabajar	P2.3	1	15	a todos	
So, you think it's all right if I wore my father's Trinidad police dress uniform? [Laughter]	Man 4 (Man in white uniform): no. (sir.) you think it's all right if I wore my father's Trinidad police dress uniform? (Laughter)	F2.1	1		Hombre 4 (Hombre con uniforme blanco): 40 así que creen que está bien si llevo el 37 uniforme de policía de mi padre (de Trinidad)	P2 [T4]	1	31		
Now listen to me, all of you. This is a serious documentary about serious police work. We will	Man 1 (Rowan Atkinson): (Now listen to me.) all of you. This is a serious documentary about serious police work. (sir.) we will present ourselves as God made us and as the Queen expects. (sir.) what's grey away? (F)	[T] C + F2.1	2		Hombre 1 (Rowan Atkinson): 40 todos ustedes este es un documental serio 39 sobre el trabajo policial serio 39 nos presentaremos como Dios nos hizo y como 34 la reina espera que es gris lejos	P3.3	1	41	sobre el trabajo serio en la policía	
Raymond. Grey away. [Laughter] I have no idea.	Woman 1 (Blonde woman): Raymond. I saw away. (Laughter) I have no idea.	F2.1 + No distinción del hablante + [T4] F + F2.1	4		Mujer 1 (Mujer rubia): 33 Raymond gris lejos no tengo idea	P2.3 [T4]	1	32	Gris Fuera	
Looking a bit grey. Feeling a bit old. [Laughter] New grey away we'll make you look 10 years younger. [Laughter]	Woman 1 (Blonde woman): looking bit grey. (sir.) feeling a bit old. (Laughter) new grey away we'll make you look 10 years younger. (Laughter)	F2 + F2.1 + F2.1 + [T] E	4		Mujer 1 (Mujer rubia): 38 luciendo un poco gris sintiéndome un poco viejo nuevo 39 gris lejos lo haremos parecer 10 años más 7 joven	P2.3 [T4] + P2.3 [T4]	1	41	Gris Fuera te hará	
Good Lord, I must have picked up someone else's bag at Boots. [Laughter] Good lordy Lord. Good lordy, lordy. Do do ding dum de. [Laughter] [T4] I'll take it back at lunchtime.	Man 1 (Rowan Atkinson): good lord. (sir.) I must have picked up someone else's bag boots. (Laughter) good lordy Lord. (sir.) good lordy, lordy. do do ding dum de. (Laughter) (Inaudible) I'm I'll take it back at lunchtime. (F)	[T4] B + F2.1 + [T] E + F2.1	1		Hombre 1 (Rowan Atkinson): 37 bolso de otra persona en Boots Dios mío, Dios mío, Dios mío, 38 accidentalmente do do ding dum de. (in) Lo 30 devolvérete a la hora del almuerzo	P1	1	32	coger	reconoce que "Boots" es nombre propio
Messages [Inaudible]. But it was just a thought. [Laughter]	Man 5 (Man with a toupee): Messages [Inaudible] but it was just a thought. (Laughter)	[T4] C + [T] F + F2.1 + [T] E	4		Hombre 5 (Hombre con peluca): 23 pero fue solo una idea 7 [Música]			23		
	[Music]									

Anexo II. Documento original extraído de Excel que recoge el análisis completo de errores detectados en los subtítulos automáticos del vídeo 2.

Anexo III. Hoja de análisis vídeo 3 (Excel)

AUDIO ORIGINAL INGLÉS	TRANSCRIPCIÓN SUBTITULADO INGLÉS	ANÁLISIS	Nº ERRORES	Nº CARACTERES	TRANSCRIPCIÓN SUBTITULADO ESPAÑOL	ANÁLISIS	Nº ERRORES	Nº CARACTERES	PROPUESTAS DE MEJORA	COSAS QUE HA RECUPERADO
	[00:00:08 – 00:00:08] Boy 1 (Son):				[00:00:04 – 00:00:08] Chico 1 (Hijo):					
There you are, mummy. We were beginning to think you didn't want your brinche. [Soft laughter] Thomas.	there you are, Mommy, we were beginning to think you didn't want your babe. [Soft laughter] Thomas	F2.1 + F2.1 + [Tx]B + F2.1	4	4	40 ahí estás mami estábamos empezando a pensar que no querías tu babe Thomas	P5[Tx]	1	36	36 brinche	
	[00:00:08 – 00:00:12] Woman 1 (Mother):	[Tx]A + F2.1 + [Tx]B + F2.1		4	[00:00:08 – 00:00:12] Mujer 1 (Madre):		1	34	34 deciros	detecta signo interrogación
Chloe. Mummy has something to tell you.	Clay, mommy has something to tell you.				[00:00:12 – 00:00:16] Chico 1 (Hijo):	F3.3	1	34	34 deciros	
	[00:00:12 – 00:00:16] Boy 1 (Son):				[00:00:16 – 00:00:24] Mujer 1 (Madre):	E1 + P2.3	2	30	30 ¿Qué ocurre, mi dulce corazón? Cuéntalo todo	
What is it, dear heart? Tell all. [Laughter] Alice can't come to work today, so	what is it, dear heart? tell all. [Laughter] Alice can't come to work today, so	F2.1 + F2.1 F2.1 + [T]-[E] + No distinción hablante + F2.1	2 4	1	26 ¿qué es querido corazón? élle a 37 todos Alice no puede venir a trabajar hoy así que			14	14 mami tuvo que	
	[00:00:16 – 00:00:24] Woman 1 (Mother):	[Tx]B	1	1	[00:00:16 – 00:00:24] Mujer 1 (Madre):	P2.1	1	45	45 contratar a una niñera	
mummy's had to arrange for an agency nanny to come round. [Dramatic music] Mummy? No. We must all be very	mommy's had to arrange for an agency Nanny to come around. [dramatic music] mommy? No. We must all be very	F2.1 F2.1 + [T]-[E] + [Tx]B + F2.1	1 5	35	14 mami tuvo que 36 arreglar que una niñera de la agencia viniera					
	[00:00:24 – 00:00:27] Girl 1 (Daughter):				[00:00:24 – 00:00:27] Chica 1 (Hija):			30	30	detecta coma
	[00:00:27 – 00:00:29] Woman 1 (Mother):				[00:00:27 – 00:00:29] Mujer 1 (Madre):			31	31	
brave. I don't feel very well.	brave. I don't feel very well.	F2.1 + F2.1 + No distinción hablante	3	28	28 valientes no me siento muy bien					
	[00:00:29 – 00:00:36] Boy 1 (Son):				[00:00:29 – 00:00:36] Chico 1 (Hijo):					
	[00:00:36 – 00:00:39] Woman 1 (Mother):				[00:00:36 – 00:00:39] Mujer 1 (Madre):					
Be strong, my darlings, because there's something else you should know. Something much worse. This person, the agency is	be strong, my darlings, because there's something else you should know. Something much worse. This person, the agency are	F2.1 + F2.1 F2.1 F2.1 + F2.1 + F3.1	2 1 3	41	37 sean fuertes mis queridos porque hay algo más que deberían saber algo 37 mucho peor esta persona la agencia nos está	F3.4 P2.3 [F2.1] + F3.1	1 2	36 43	36 queridos míos 43 la que la agencia nos envía	
us. She's from the north. [Dramatic music][Laughter][Child coughs] Newcastle, I'm told, but could be as far as Sunderland. [Laughter]	us. She's from the north. [Dramatic music][Laughter][Child coughs] Newcastle, I'm told, but could be as far as Sunderland. [Laughter]	F2.1 F2.1 + [T]-[E] + [T]-[E] + F2.1 + F2.1 [Tx] + [T]-[E]	1 6 2	17	17 es del norte de Newcastle me han dicho pero podría estar tan lejos 40 como allí Lo siento	P2.3 [F2.1] + F3.1 + P2.1 P2.3[Tx]	3 1	66 21	66 es del norte, de Newcastle, me han dicho. Pero podría ser de más lejos. 21 de Sutherland, lo siento	
I'm sorry. [Doorbell rings] [Sound of door opening]	I. I'm sorry. [Doorbell rings][Sound of door opening]	[Tx]B + F2.1 + F2.1 + [T]-[E] + [T]-[E] [Tx]B + F2.1 + F2.1 + [T]-[E] + [T]-[E]	5 5	12	12 X					
	[00:01:02 – 00:01:05] Woman 2 (Northern nanny):				[00:01:02 – 00:01:05] Mujer 2 (Niñera del norte):					
Hiya, I'm Gina from the agency. [Laughter] Are you Mrs. Montgomery? [Squeaking sound][Laughter][Nanny sighs] Hey, I tell you what, I thought	hi yeah, I'm Gina from the agency. [Laughter] are you Mrs. Montgomery? [Squeaking sound][Laughter][Nanny sighs] hey, I tell you what, I thought	[Tx]B + F2.1 + [T]-[E] + F2.1 F2.1 + [T]-[E] + [T]-[E] + F2.1 + F2.1 + No distinción hablante	4 7	41	41 hola si soy Gina de la agencia ¿eres la Sr. Montgomery hey te digo lo que pensé que	E2 P5 + P2.1	1 2	39 40	39 ey 40 oye, ¿sabes qué?	detecta signo interrogación
	[00:01:13 – 00:01:29] Woman 2 (Northern nanny):				[00:01:13 – 00:01:29] Mujer 2 (Niñera del norte):					
I was going to be late. The traffic was that bad. But never mind. Here I am on time [Laughter]	I was going to be late. The traffic was that bad. But never mind. Here I am on time [Laughter]	F2.1 F2.1 + F2.1 + [Tx]A + [T]-[E]	1 4	38	38 iba a llegar tarde el tráfico estaba muy mal pero no importa aquí estoy en T	E1 P2.3[Tx]	1 1	36 39	36 había mucho tráfico	
no reason. And you've got a lovely place, Mrs. M. You don't mind if I call you Mrs. M? Do you [Laughter] only Mrs.	no reason and you got a lovely Place Mrs. M. you don't mind if I call you Mrs. M? do you [Laughter] only Mrs.	[Tx]C + F3.1 + F3.1 + F2.1 + F2.1 [Tx]E + F2.1 [T]-[E] + F2.1 + [Tx]B + F2.1 F2.1 + [T]-[E] + F2.1 + F2.1 + F3.1	5 2 2 5	37	37 medifolia y tienes un lugar encantador Sr. M. no te importa si te llamo Sr. M. ¿solo te llamas Sr. M? 38 Mong hola si cómo la Ramas	P2.3[Tx] + F2.1 P2.3[T]	2 1	44 20	44 una casa 20 detecta signo interrogación	
Mong [Inaudible] Hiya. What do you call yours, then? [Laughter] See, I'm gonna have my hands full with you. Pair of tykes, an'er.	Mong [Inaudible] hi yeah. what what you call use. then? [Laughter] see, I'm going to have me hand full with you. pair of takes AR.	F2.1 + [Tx]B + [Tx]A + F2.1 [Tx]A + [Tx]B + F2.1 + [Tx]B + F2.1 + F2.1 + [T]-[E] + No distinción hablante	4 8	12	12 solo te llamas Sr. M. 39 emm como viéis voy a tener mis manos ocupadas 25 contigo un par de tomas AR 40 yo mami mami qué está diciendo no lo	P2.3 [F2.1] + F2.1 P2.3[Tx] P2.3[T]	1 1 1	41 26 36	41 pero suena muy enfadada 24 no la provocáis 36 evitad	
Mummy! Mummy! What's she saying? [Laughter] I don't	I mommy! mommy! what is she saying? [Laughter] I don't				[00:01:39 – 00:01:39] Chico 1 (Hija):					
	[00:01:39 – 00:01:47] Woman 1 (Mother):				[00:01:39 – 00:01:47] Mujer 1 (Madre):					
know, Chloe, but she sounds awfully angry, so don't provoke her in any way. [Laughter] Avoid looking her in the eyes. [Laughter] She'd	know, Chloe, but she sounds awfully angry. so I don't provoke her in any way. [Laughter] avoid looking her in the eyes. [Laughter] she	F2.1 + [Tx]C + F2.1 + F2.1 [Tx]C F2.1 + [T]-[E] + F2.1 + [T]-[E] + F3.1 + No distinción hablante	4 1 6	26	26 no la provoco de ninguna manera evita mirarla a los ojos ella	P2.3[Tx] + F2.1 F3.3	2 1	41 24	41 pero suena muy enfadada 24 no la provocáis	
	[00:01:47 – 00:01:49] Boy 1 (Son):				[00:01:47 – 00:01:49] Chico 1 (Hijo):					
kill us. Shh, shh. Thomas. Mummy won't let	kill us. Shh. I shh. I shh. Thomas. I mommy won't let	F2.1 + F2.1 + F2.1 + F2.1 + [Tx]A + No distinción hablante	6	38	38 nos mata shh shh Thomas mami no dejará que	F3.1[Tx]	1	42	42 nos va a matar	
	[00:01:49 – 00:01:52] Woman 1 (Mother):				[00:01:49 – 00:01:52] Mujer 1 (Madre):					
anything happen to you. Right, I'll start	anything happen to you. right. I'll start	F2.1 + F2.1 + No distinción hablante	3	39	39 nada te pase ¿Verdad? Comenzaré	F3.3 + F2.2	2	31	31 os	detecta signos interrogación
	[00:01:49 – 00:01:52] Woman 2 (Northern nanny):				[00:01:49 – 00:01:52] Mujer 2 (Niñera del norte):					
with the breakfast things. But first, why don't I make us all a nice milky brew? And who fancies a couple of bikkies as well, eh? Run for your lives, children.	with a breakfast things. but first why don't I make us all a nice milky brew? and who fancies a couple of bikkies as well, eh? Run for your lives. I children.	F2.1 F2.1 [Tx]B F2.1 + F2.1 + No distinción hablante + F2.1 + F2.1	1 1 1 5	37	37 con cosas para despertar pero primero ¿por qué no nos preparo a todos una rica infusión con leche 38 y a quién le apetece un par de cerezas grandes 36 también? corran por sus vidas niños	P2.1[Tx] + P4 F3.4	2 1	46 48	46 detecta signo interrogación 48 galletitas	detecta signo interrogación
	[00:01:52 – 00:01:56] Woman 1 (Mother):				[00:01:52 – 00:01:56] Mujer 1 (Madre):					
We're all gonna die. [Laughter] [screaming]	we're all gonna die. [Laughter] [Screaming]	F2.1 + [T]-[E] + [T]-[E]	3	20	20 todos vamos a morir			19	19	
			126	1210			37	1226		

Anexo III. Documento original extraído de Excel que recoge el análisis completo de errores detectados en los subtítulos automáticos del vídeo 3.

Anexo IV. Acceso a los documentos completos del análisis

A continuación se incluyen los enlaces a las hojas de análisis de los subtítulos automáticos de cada vídeo:

- **VÍDEO 1:**
<https://docs.google.com/spreadsheets/d/1T8gtOkSNJ8-DQ3oSxHpPsQSLZHckTZD8/edit?usp=sharing&ouid=100660638544021331676&rtpof=true&sd=true>

- **VÍDEO 2:**
<https://docs.google.com/spreadsheets/d/1Mux6nvTBFtQ8tDQD2HWt5rcnsEXwqLAT/edit?usp=sharing&ouid=100660638544021331676&rtpof=true&sd=true>

- **VÍDEO 3:**
<https://docs.google.com/spreadsheets/d/1rR9DFdfqKXWyF8u3bwKP0MG4o6gfPciD/edit?usp=sharing&ouid=100660638544021331676&rtpof=true&sd=true>