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

Acedo Ortega, Anna; Sánchez-Gijón, Pilar, dir. Evaluación de la transcripción generada automáticamente con Whisper aplicada a la traducción audiovisual y la accesibilidad a contenido audiovisual. Bellaterra: Universitat Autònoma de Barcelona, 2023. (Tecnologies de la Traducció)

This version is available at <https://ddd.uab.cat/record/283618>

under the terms of the  license

TRABAJO DE FIN DE MÁSTER

2022-2023



**Evaluación de la transcripción generada automáticamente
con Whisper aplicada a la traducción audiovisual y la
accesibilidad a contenido audiovisual**

MÁSTER EN TRADUMÁTICA: TECNOLOGÍAS DE LA TRADUCCIÓN
FACULTAD DE TRADUCCIÓN E INTERPRETACIÓN

AUTORÍA

Anna Acedo Ortega

TUTORÍA

Pilar Sánchez-Gijón

Barcelona, 26 de julio de 2023

Datos del TFM / Dades del TFM / Dissertation data

Título: Evaluación de la transcripción generada automáticamente con Whisper aplicada a la traducción audiovisual y la accesibilidad a contenido audiovisual

Títol: Avaluació de la transcripció generada automàticament amb Whisper aplicada a la traducció audiovisual i la accessibilitat a contingut audiovisual

Title: Evaluation of Whisper's automatically generated transcription applied to audiovisual translation and accessibility to audiovisual content

Autor/a: Anna Acedo Ortega

Autor/a: Anna Acedo Ortega

Author: Anna Acedo Ortega

Tutor/a: Pilar Sánchez Gijón

Tutor/a: Pilar Sánchez Gijón

Tutor: Pilar Sánchez Gijón

Centro: Universidad Autónoma de Barcelona (UAB)

Centre: Universitat Autònoma de Barcelona (UAB)

Centre: Autonomous University of Barcelona (UAB)

Estudios: Máster oficial en Tradumática: Tecnologías de la Traducción

Estudis: Màster oficial en Tradumàtica: Tecnologies de la Traducció

Studies: Official master's degree in Tradumatics: Translation Technologies

Palabras clave / Paraules clau / Keywords

Tradumática, Reconocimiento del habla, Traducción Audiovisual, Whisper, Open AI

Traducció, Reconeixement de la parla, Traducció Audiovisual, Whisper, Open AI

Translation, Speech Recognition, Audiovisual Translation, Whisper, Open AI

Resumen del TFM / Resum del TFM / Abstract

El presente estudio busca analizar la calidad de la transcripción de la herramienta Whisper desde un punto de vista general de la corrección de la lengua. Para ello, se hará una comparación de dicha transcripción con los subtítulos generados automáticamente de la herramienta YouTube Studio y, finalmente, con la transcripción humana creada por la autora del presente trabajo. El objetivo final de este estudio es observar la corrección lingüística o nivel de acierto de la transcripción de Whisper, la viabilidad del uso de esta herramienta que usa el reconocimiento del habla con transcripción automática, aplicado al campo de la traducción audiovisual. Para llevar a cabo estas tareas, se ha realizado un estudio con 25 muestras en 5 idiomas y 1 dialecto: español, inglés, francés, catalán, valenciano y japonés. Se espera obtener resultados que reflejen si el uso de la transcripción generada a partir del reconocimiento del habla puede ayudar en la tarea de la traducción audiovisual y la accesibilidad a este tipo de contenido.

El present estudi busca analitzar la qualitat de la transcripció de l'eina Whisper des d'un punt de vista general de la correcció de la llengua. Per a això, es farà una comparació d'aquesta transcripció amb els subtítols generats automàticament de l'eina YouTube Studio i, finalment, amb la transcripció humana creada per l'autora del present treball. L'objectiu final d'aquest estudi és observar la correcció lingüística o nivell d'encert de la transcripció de Whisper, la viabilitat de l'ús d'aquesta eina que usa el reconeixement de la parla amb transcripció automàtica, aplicat al camp de la traducció audiovisual. Per a dur a terme aquestes tasques, s'ha realitzat un estudi amb 25 mostres en 5 idiomes i 1 dialecte: espanyol, anglès, francès, català, valencià i japonès. S'espera obtenir resultats que reflecteixin si l'ús de la transcripció generada

a partir del reconeixement de la parla pot ajudar en la tasca de la traducció audiovisual i l'accessibilitat a aquesta mena de contingut.

The present study wants to analyze the quality of the transcription automatically generated by Whisper from a general point of view of language correctness. For this purpose, a comparison will be made between the transcription and the subtitles automatically generated by YouTube Studio and, finally, with the human transcription created by the author of this paper. The final purpose of this study is to observe the linguistic correctness or level of accuracy of the Whisper transcription, and the feasibility of using this tool that uses speech recognition with automatic transcription, applied to the field of audiovisual translation. To determine this, a study has been carried out with 25 samples in 5 languages and 1 dialect: Spanish, English, French, Catalan, Valencian and Japanese. It is expected to obtain results that reflect whether the use of transcription generated from speech recognition can help in the task of audiovisual translation and accessibility to this type of content.

Aviso legal / **Avís legal** / *Legal notice*

© AUTOR, Barcelona, 2023. Todos los derechos reservados.

Ningún contenido de este trabajo puede ser objeto de reproducción, comunicación pública, difusión y/o transformación, de forma parcial o total, sin el permiso o la autorización de su autora.

© AUTOR, Barcelona, 2023. Tots els drets reservats.

Cap contingut d'aquest treball pot ser objecte de reproducció, comunicació pública, difusió i/o transformació, de forma parcial o total, sense el permís o l'autorització de la seva autora.

© Autor, Barcelona, 2023. All rights reserved.

None of the content of this academic work may be reproduced, distributed, broadcasted and/or transformed, either in whole or in part, without the express permission or authorization of the author.

TABLA DE CONTENIDOS

I. Introducción	6
1. Breve descripción de la propuesta	6
2. Motivación y justificación del trabajo.....	7
3. Objetivos e hipótesis (o preguntas de investigación)	8
II. Marco teórico.....	8
1. Inteligencia artificial (IA): desarrollo y logros más destacados	8
1.1 Breve explicación del aprendizaje automático (machine learning) y del aprendizaje profundo (deep learning).....	10
2. Introducción al del reconocimiento del habla.....	12
3. Bases de la subtítulos aplicada a la traducción audiovisual.....	13
3.1 Enfoque prescriptivo.....	13
3.2 Enfoque descriptivo	15
3.3 Características generales de la subtítulos	15
3.3.1 Aspectos técnicos de la subtítulos	16
3.3.2 Aspectos ortotipográficos de los subtítulos	17
3.3.3 Aspectos lingüísticos.....	17
4. Comunicación audiovisual y accesibilidad a los medios de comunicación	18
4.1 Contenido audiovisual y accesibilidad para personas con discapacidad auditiva.....	22
III. Marco práctico.....	25
1. Procedimiento utilizado en la preparación del módulo de trabajo de Whisper	26
1.1 Código de Python inicial – diciembre 2022-marzo 2023.....	26
1.2 Metodología de la parte práctica	33
♦ Español.....	37
♦ Inglés.....	39
♦ Francés.....	42
♦ Catalán y Valenciano.....	44
♦ Japonés.....	47
1.3 Resultados de las transcripciones de Whisper y YouTube: análisis y comparación.....	50
1.3.1 Análisis y comparación de la puntuación de las transcripciones de Whisper y YouTube	50
1.3.2 Análisis y comparación de la corrección gramatical y terminológica de las transcripciones de Whisper y YouTube	57
IV. Conclusión.....	58
V. Referencias bibliográficas.....	60
VI. Anexos.....	65
1. Pasos a seguir para instalar Whisper en el ordenador.....	65
2. Resultados obtenidos de la ejecución del código que aparecen en el terminal de VS Code	68

TABLA DE CONTENIDOS: Figuras, tablas y gráficos

Figura 1: Artificial Intelligence Timeline (Anyoha, 2017)	10
Figura 2. Código de Python inicial.....	29
Figura 3: Muestra de Gradio original (Gradio, 2022).....	29
Figura 4: Muestra de Gradio personalizada. Código de Python inicial	30
Figura 5: Muestra de Gradio personalizada-Error. Código de Python inicial	30
Figura 6: Código de Python definitivo	32
Figura 7: Exportación de audio con QuickTime Player en MacBook Air.....	34
Figura 8: Gráfica de WER. GitHub de Whisper (Radford et al., 2022).....	36
Tabla 1: ES_coloquial_vulgar.m4a_result.....	37
Tabla 2. ES_formal_cientifico_2.m4a_result.....	38
Tabla 3. ES_formal_varios_interlocutores.m4a_result.....	39
Tabla 4. ES_formal.m4a_result	39
Tabla 5. EN_coloquial_varios_interlocutores.m4a_result.....	40
Tabla 6. EN_formal_cientifico_2.m4a_result.....	40
Tabla 7. EN_formal_cientifico.m4a_result.....	41
Tabla 8. EN_informal_vulgar_2.m4a_result.....	41
Tabla 9. EN_informal_vulgar.m4a_result.....	42
Tabla 10. EN_informal.m4a_result.....	42
Tabla 11. FR_formal_cientifico.m4a_result.....	43
Tabla 12. FR_informal_coloquial.m4a_result.....	43
Tabla 13. FR_informal_vulgar.m4a_result.....	44
Tabla 14. CA_coloquial.m4a_result	44
Tabla 15. CA_formal_2.m4a_result.....	45
Tabla 16. CA_formal.m4a_result	45
Tabla 17. CA_VA_varios_interlocutores_coloquial.m4a_result.....	46
Tabla 18. CA_varios_interlocutores_coloquial.m4a_result.....	46
Tabla 19. CA_vulgar_varios_interlocutores.m4a_result.....	47
Tabla 20. JA_informal_vlog.m4a_result.....	47
Tabla 21. JA_informal_coloquial_musica.m4a_result.....	48
Tabla 22. JA_informal_2_varios_interlocutores.m4a_result.....	48
Tabla 23. JA_formal_interferencias.m4a_result.....	49
Tabla 24. JA_formal_2.m4a_result	49
Tabla 25. JA_formal_1.m4a_result	50
Figura 9. Análisis de la puntuación en catalán.....	51
Figura 10. Análisis de la puntuación en inglés.....	52
Figura 11. Análisis de la puntuación en español.....	53
Figura 12. Análisis de la puntuación en francés.....	53
Figura 13. Análisis de la puntuación en japonés.....	54
Gráfico 1. Análisis de puntos en todas las lenguas	55
Tabla 26. Análisis de puntos en todas las lenguas.....	56
Gráfico 2. Análisis de comas en todas las lenguas.....	56
Tabla 27. Análisis de comas en todas las lenguas.....	56

I. Introducción

1. Breve descripción de la propuesta

En el presente trabajo de fin de máster se trabajará con el sistema de reconocimiento de voz Whisper, de OpenAI. Según los creadores (Radford et al., 2022):

“Whisper is an automatic speech recognition (ASR) system trained on 680,000 hours of multilingual and multitask supervised data collected from the web. We show that the use of such a large and diverse dataset leads to improved robustness to accents, background noise and technical language. Moreover, it enables transcription in multiple languages, as well as translation from those languages into English. We are open-sourcing models and inference code to serve as a foundation for building useful applications and for further research on robust speech processing.”

Respecto al marco teórico y para poder darle sentido a la parte práctica de este trabajo, se tiene por objetivo hacer una búsqueda general de los avances de inteligencia artificial y una breve explicación del reconocimiento del habla.

Por otro lado, se abordará la comunicación multimedia audiovisual, enfocándose en cómo es esta a través de las redes sociales, ya que los materiales a usar en la parte práctica son todos archivos multimedia extraídos de las redes sociales. Se valorará si estas herramientas pueden serles útiles tanto a las personas oyentes como no oyentes, para crear con mayor rapidez o disfrutar del contenido audiovisual.

Respecto al marco práctico, se realizará una serie de pruebas divididas en dos grandes bloques. El material utilizado es variado, tanto en registro de formalidad como de ámbito temático. Se analizarán archivos en formato audiovisual enfocado en dos perspectivas: el discurso formal e informal. Dentro de estos dos registros de formalidad de la lengua, el formal se ha dividido en lenguaje científico-técnico: los archivos analizados son del campo de la medicina, jurídico, político y periodístico.

Por otro lado, lenguaje informal se ha dividido en estándar, coloquial y vulgar: los archivos analizados son videoblogs, recetas de cocina, entrevistas y fragmentos de películas donde hay un uso de lenguaje vulgar, peyorativo u ofensivo.

Se evaluará si hay censura o no de terminología específica. Además, se quiere demostrar la capacidad de transcribir correctamente de Whisper para poder utilizar los resultados con fines profesionales en la traducción general y, en específico, de la traducción audiovisual.

Es interesante también, a su vez, observar si los sonidos de fondo de un archivo multimedia pueden llegar a influir desfavorablemente en el reconocimiento de voz y posterior transcripción utilizando Whisper. Este análisis se comparará con la transcripción proporcionada por YouTube y se cotejarán los resultados con la transcripción humana realizada por la autora del presente trabajo. Las lenguas de trabajo en la parte práctica son español, catalán, inglés, francés y japonés.

El objetivo último de la comparación de estos dos tipos de transcripciones generadas automáticamente es llegar a justificar o refutar si la transcripción automática generada con inteligencia artificial facilita la tarea de la traducción audiovisual y la accesibilidad al contenido audiovisual, además de valorar si se puede utilizar la TA (traducción automática) o la TAPE (Traducción automática con posesición) con los resultados obtenidos.

2. Motivación y justificación del trabajo

Los avances en tecnología han resultado en un aumento notorio en la cantidad de datos audiovisuales disponibles, lo que los convierte en un recurso muy valioso para la investigación y el análisis. En este trabajo, mi objetivo es explorar las posibilidades del reconocimiento de voz en el análisis de datos audiovisuales.

El reconocimiento de voz desempeña un papel importante en el campo de la traducción porque puede transcribir rápidamente el lenguaje oral en texto escrito. Esto, a su vez, facilita mucho la traducción de archivos audiovisuales, como conferencias, podcasts, etc. También puede facilitar la comunicación en tiempo real entre personas que no hablan el mismo idioma, ya que se puede traducir automáticamente el contenido de una conversación casi en tiempo real. Esto ayuda a superar las barreras lingüísticas y facilita la comunicación entre personas.

El reconocimiento de voz es importante en el campo de la traducción aplicada a los vídeos de las redes sociales porque permite subtítular y traducir automáticamente los vídeos casi en tiempo real. Esto facilita que los usuarios entiendan el contenido de los vídeos, aunque hablen idiomas diferentes.

Teniendo en cuenta que hoy en día vivimos rodeados de tecnología, estamos constantemente interconectados, con acceso a cualquier información del planeta, es interesante estudiar los puntos antes mencionados en este trabajo. Sería especialmente útil para aquellas plataformas de redes sociales con una audiencia global, ya que aumenta la accesibilidad y la participación de los usuarios que, de otro modo, podrían tener dificultades para entender el contenido. Además, la traducción

mediante el reconocimiento de voz puede ayudar a aumentar el alcance de los vídeos al hacerlos más accesibles a un público más amplio, lo que conduce a una mayor participación y a una posible propagación del contenido.

Este estudio tiene por objetivo, pues, ver cuáles son los desafíos y las limitaciones del uso del reconocimiento de voz aplicado a la traducción mediante Whisper, con el análisis de datos audiovisuales y evaluar su eficacia comparado con la traducción humana y con otros métodos de transcripción ya existentes o más arcaicos.

3. Objetivos e hipótesis (o preguntas de investigación)

- ♦ Comprobar la corrección lingüística y eficacia de la transcripción del reconocimiento de voz de Whisper respecto al sistema de reconocimiento de voz de YouTube Studio.
- ♦ Demostrar si la transcripción generada por inteligencia artificial es lo suficientemente buena y útil para poder compararse con la transcripción humana o si, por el contrario, aún no llega a un nivel de corrección suficiente, desde el punto de vista de la corrección lingüística, para poder implementar los resultados en la traducción audiovisual y la accesibilidad a contenido audiovisual.

II. Marco teórico

1. Inteligencia artificial (IA): desarrollo y logros más destacados

La primera vez que se habló de inteligencia artificial (IA) como tal (*AI* en inglés), fue en 1956, cuando John McCarthy acuñó el término como tema de la Conferencia de Dartmouth (en inglés, *Dartmouth Summer Research Project on Artificial Intelligence* (DSRPAI)).

Pero no fue realmente 4 años antes, tal y como se explica en el artículo *The History of Artificial Intelligence* (Anyoha, 2017), que se empezó a plantear el tema de la inteligencia artificial. El pionero en este campo de estudio fue Alan Turing. Por primera vez se empezó a plantear y documentar sobre la inteligencia artificial, concretamente en 1950, cuando Turing escribió su artículo académico *Computing Machinery and Intelligence* (Turing, 1950) en el que analizaba cómo construir máquinas inteligentes y cómo poner a prueba su inteligencia mediante la resolución de problemas y toma de decisiones. No obstante, hubo dos grandes obstáculos en el proceso para llevar a cabo esta idea.

John McCarthy en 2004 dio su propia definición de inteligencia en su artículo *What is artificial intelligence?* Tal como citan los expertos de IBM, para McCarthy la IA “[...] es la ciencia y la ingeniería de la fabricación de máquinas inteligentes, especialmente programas informáticos inteligentes. Está relacionada con la tarea similar de usar computadoras para entender la inteligencia humana, pero la IA no tiene que limitarse a métodos que son biológicamente observables”. Una de las muchas aplicaciones de la inteligencia artificial es el reconocimiento del habla, en la que nos centraremos en el presente trabajo.

Tal expone Anhoya (2017), en primer lugar, en aquella época —antes de 1949— los ordenadores solo podían ejecutar órdenes, pero no almacenarlas. En segundo lugar, era sumamente caro utilizar un ordenador ya que era un recurso que casi únicamente las grandes empresas tecnológicas y las universidades prestigiosas podían permitirse, por lo que para obtener fuentes de financiación y demostrar que la inteligencia artificial merecía la pena, se necesitaba una prueba de concepto.

Pero ¿qué es una prueba de concepto? Este término proviene del inglés *proof of concept*. Jobin et al. (2022), en su artículo *What does the proof-of-concept (POC) really prove? A historical perspective and a cross-domain analytical study*, lo definen como la implementación de un método o de una idea que tiene que demostrarse ser cierta, aunque puede ser que dicha idea esté inacabada. Esta prueba se lleva a cabo con el objetivo de demostrar que dicha idea, concepto o teoría es viable y digna de ser explorada y/o patentada, por lo que para ello se necesitaría de la financiación que únicamente grandes instituciones podrían otorgar, como bien se ha dicho antes, debido a su alto coste en un campo aún inmaduro y poco conocido, a riesgo de perder grandiosas sumas de dinero, así como de tiempo, en un experimento que puede que no dé frutos.

A pesar de las dificultades, en 1955, Allen Newell, Cliff Shaw y Herbert Simon iniciaron la prueba de concepto con *Logic Theorist*. El *Logic Theorist* era un programa diseñado para imitar las habilidades de resolución de problemas de un ser humano y fue financiado por la Corporación de Investigación y Desarrollo (RAND). Muchos lo consideran el primer programa de inteligencia artificial y se presentó en la antes mencionada Conferencia de Dartmouth. Aunque dicha conferencia no dio los resultados que McCarthy esperaba, se consolidó la idea de que la IA era útil y factible, por lo que se podría obtener la financiación esperada. En los siguientes 17 años la IA creció y se consolidó con ordenadores más rápidos, accesibles a un gran público, con mayor capacidad para almacenar datos.

Anyoha (2017) documentó que de esta manera destacados investigadores de la Conferencia de Dartmouth llevaron sus proyectos para que fueran aprobados por el gobierno y así obtener financiación en investigación sobre la IA, en la cual contribuyó la Agencia de Proyectos de Investigación Avanzada de Defensa (del inglés *Defense Advanced Research Projects Agency*, abreviado DARPA). Al gobierno estadounidense le interesaba especialmente una máquina que fuera capaz de transcribir y traducir el lenguaje hablado, así como en el procesamiento de datos de alto rendimiento.

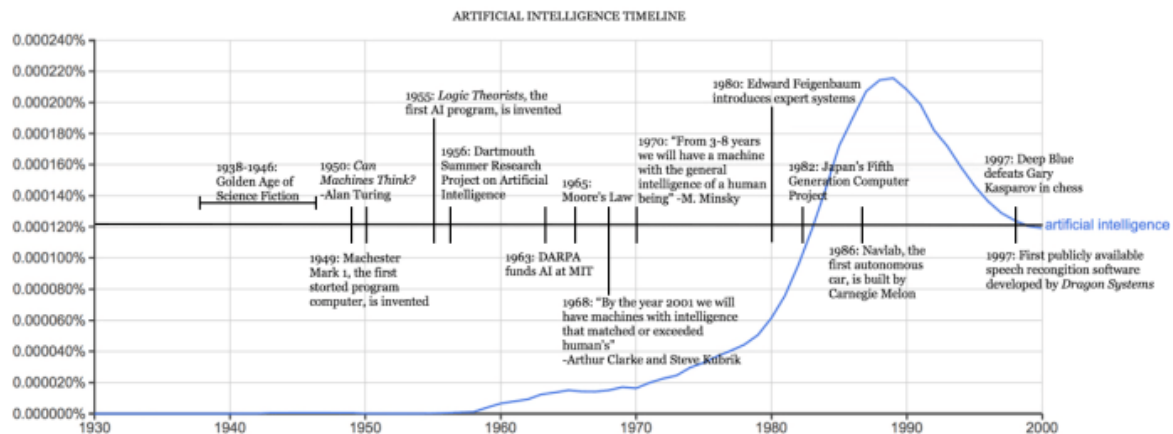


Figura 1: Artificial Intelligence Timeline (Anyoha, 2017)

Tal como explica Anyoha (2017) en su artículo, en 1980, tras una década en la cual la investigación y la financiación se estancaron, la IA se reavivó gracias la ampliación de las herramientas algorítmicas y el aumento de los fondos. John Hopfield y David Rumelhart popularizaron las técnicas de «*deep learning*», también conocido como aprendizaje profundo en español.

Para poder entender estos términos, previamente hay que conocer los conceptos básicos del aprendizaje automático y, derivado de este, el aprendizaje profundo.

1.1 Breve explicación del aprendizaje automático (*machine learning*) y del aprendizaje profundo (*deep learning*)

Según el artículo de Alameda (2019), donde citan a José Luis Espinoza:

“[...] El aprendizaje automático es un maestro del reconocimiento de patrones, y es capaz de convertir una muestra de datos en un programa informático capaz de extraer inferencias de nuevos conjuntos de datos para los que no ha sido entrenado previamente”.

El aprendizaje automático utiliza modelos para resolver tareas y se dividen en tres tipos de modelos: geométricos, probabilísticos y lógicos. Estos, a su vez, se dividen en subgrupos, ya que están conformados por distintos algoritmos con características específicas para resolver las tareas. (Moreno, A. et al., 1994).

El aprendizaje profundo es una subrama del aprendizaje automático. "En realidad, el aprendizaje profundo se compone de redes neuronales. "Profundo" en el aprendizaje profundo se refiere a una red neuronal compuesta por más de tres capas (que incluirían las entradas y la salida), que se puede considerar un algoritmo de aprendizaje profundo.

La diferencia entre el aprendizaje profundo y el aprendizaje automático es cómo aprende cada algoritmo. El aprendizaje profundo automatiza gran parte de la fase de extracción de características del proceso, lo que elimina parte de la intervención humana manual necesaria y permite el uso de conjuntos de datos más grandes. El aprendizaje profundo se podría considerar como "*machine learning scalable*", tal como Lex Fridman señaló en la conferencia del MIT de 2019 *Deep Learning Basics: Introduction and Overview*. El *machine learning* tradicional, o "*non-deep*", depende más de la intervención humana para aprender. Los expertos humanos determinan la jerarquía de características para comprender las diferencias entre las entradas de datos, lo que por lo general requiere más datos estructurados para aprender. (IBM, n./a.)

Anyoha (2017) expone que Edward Feigenbaum introdujo sistemas expertos que imitaban el proceso de toma de decisiones de un experto humano. El programa preguntaba a un experto en un campo cómo responder en una situación determinada y, una vez aprendido esto para prácticamente todas las situaciones, los no expertos podían recibir consejos de ese programa.

El gobierno japonés financió en gran medida los sistemas expertos y otras iniciativas relacionadas con la IA como parte de su Proyecto Informático de Quinta Generación (FGCP, del inglés *Fifth Generation Computer Project*). Entre 1982 y 1990 invirtieron 400 millones de dólares con el objetivo de revolucionar el procesamiento informático. Por desgracia, la mayoría de los objetivos no llegaron a dar frutos. El FGCP dejó de financiarse y la IA perdió protagonismo.

Durante las décadas de 1990 y 2000, se alcanzaron muchos de los objetivos más importantes de la inteligencia artificial. En 1997, el software de reconocimiento de voz, desarrollado por Dragon Systems, se implantó en Windows.

Anyoha (2017) Un hecho que marcó la historia de IA fue la Ley de Moore. Esta estima que la memoria y la velocidad de los ordenadores se duplica cada año. Como caso anecdótico pero que muestra los grandes avances que se habían conseguido en esa década, Deep Blue (superordenador de la empresa tecnológica IBM) fue capaz de derrotar en un torneo de ajedrez a Gary Kasparov (campeón

del mundo de ajedrez) en 1997 y como Alpha Go de Google fue capaz de derrotar al campeón chino de Go, Ke Jie en 2017.

Aunque los algoritmos no mejoren mucho, los macrodatos y la computación masiva hacen que la inteligencia artificial mejore a marchas forzadas, aunque es probable que ley de Moore se esté ralentizando un poco.

2. Introducción al del reconocimiento del habla

El reconocimiento del habla es la capacidad de los ordenadores y computadoras para entender y procesar lo que decimos en lenguaje humano. Tal como se explica en el artículo *A historical perspective of speech recognition* (Huang et al., 2014), en 1971 se lanzó en EE. UU. el mayor proyecto conocido en la historia del reconocimiento del habla. Se trata del proyecto *Search Understanding Research* de ARPA (*Advanced Research Projects Agency*), con un presupuesto de quince millones de dólares y una duración de cinco años.

En ese mismo año, un grupo de estudio sobre el reconocimiento del habla presidido por Allen Newell. Este recomendó que se utilizaran muchas más fuentes de conocimiento para abordar el desarrollo del reconocimiento del habla. El informe creado analizaba seis niveles de conocimiento: acústico, paramétrico, fonémico, léxico, oracional y semántico. De entre los avances logrados a lo largo de cuatro décadas, destacan los siguientes: Hearsay-I y Hearsay-II , Dragon NaturallySpeaking, más conocido como Dragon for PC, Harpy y Sphinx I y II.

- ♦ Hearsay-I fue uno de los primeros sistemas capaces de reconocer el habla de forma continua.
- ♦ Dragon NaturallySpeaking, creado en 1997, es un software de reconocimiento el habla para para Windows. Este permite a los usuarios interactuar con muchas aplicaciones diferentes del ordenador su PC. Por ejemplo, permite la entrada de datos en Excel por voz, el sistema de dictado en Word que muchos conocemos, etc.
- ♦ El sistema Harpy introdujo el concepto de búsqueda en haz (*beam search*).
- ♦ Sphinx-I, desarrollado en 1987, fue el primer sistema en demostrar el reconocimiento del habla independiente del hablante. Sphinx-II, desarrollado en 1992, mejoró con parámetros para equilibrar la entrenabilidad y eficiencia del reconocimiento del habla.

Según las evaluaciones del habla financiadas por DARPA, la tasa de error de palabras en el reconocimiento del habla se ha utilizado como métrica principal para evaluar el progreso.

Huang et al. (2014) exponen que, en la actualidad, hay herramientas de investigación de código abierto que permiten construir sistemas de reconocimiento del habla, como Whisper. El uso de una gran cantidad de datos en la nube ha permitido mejorar continuamente los modelos acústicos

y lingüísticos. Esto ha dado lugar a la búsqueda por voz en teléfonos móviles de empresas como Apple, Google y Microsoft.

El aprendizaje automático y la disponibilidad de grandes cantidades de datos y tecnología han impulsado el avance en el reconocimiento de voz. Esto ha permitido que los ordenadores comprendan de manera más unificada diferentes aspectos del habla, como el sonido, el significado y la gramática.

Debido a que ahora disponemos de una gran cantidad de datos, esto ha ayudado infinitamente en el entrenamiento de los sistemas de reconocimiento de voz.

Huang et al. (2014) concluyen que aún nos queda mucho trabajo por delante: es necesario trabajar en etiquetar más datos para que los sistemas puedan seguir mejorando y adaptándose a diferentes situaciones y usuarios. Por ejemplo, se han hecho avances para adaptar los sistemas a diferentes hablantes y entornos, pero todavía hay dificultades en el rendimiento del reconocimiento del habla en muestras donde haya presencia de ruido y variabilidad del habla. Esto lo podremos ver en los resultados obtenidos con Whisper en el presente trabajo.

En resumen, el reconocimiento de voz ha avanzado significativamente gracias al aprendizaje automático y al acceso a grandes cantidades de datos. Aunque todavía hay desafíos por superar, se espera que los sistemas sigan mejorando y adaptándose a diferentes situaciones y necesidades en el futuro.

3. Bases de la subtitulación aplicada a la traducción audiovisual

3.1 Enfoque prescriptivo

La traducción audiovisual (TAV) es una variante de la traducción bastante reciente y no ha sido considerada como traducción hasta hace unas dos décadas, tal como explica Carla Botella (2007) en su artículo *Aproximación al estudio del doblaje y la subtitulación desde la perspectiva prescriptivista y la descriptivista*: la traducción audiovisual. Según los estudios prescriptivos, se consideraba la subtitulación como una adaptación, más que una traducción. El motivo, desde el enfoque prescriptivo, es que solo se puede dar por correcta una única traducción con validez universal e inalterable a posibles reediciones del producto traducido (por ejemplo, una película en DVD), mientras que el resto de las traducciones se consideran erróneas. Es por este motivo que tomando como válida una única traducción en la que no se cuestiona el porqué de dicha traducción, no se tienen en cuenta condicionamientos sociales, tipo de lector, solo busca la fidelidad que lleve a la equivalencia. Además, se debe respetar fielmente el texto original y no alterar el significado, por lo que se aprecia un claro enfrentamiento entre los objetivos e intereses de los estudios prescriptivos y la TAV.

Chaume (2004) señala que “la traducción y ajuste de un guion de un texto audiovisual y la posterior interpretación de esta traducción por parte de los actores, bajo la dirección del director de doblaje y los consejos del asesor lingüístico, cuando esta figura existe”. (Chaume, 2004)

Por otro lado, hablando de la subtitulación, Díaz Cintas (2003) apunta que “se puede definir 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 o de la pista sonora”.

La traducción audiovisual se divide, a su vez, en otras dos ramas: la subtitulación y el doblaje. Es importante remarcar que en este campo de estudio hay que hacer necesariamente cambios del texto original, tal como expone Botella (2007), debido a los siguientes motivos:

- ♦ En la subtitulación, como es habitual en el mundo de la traducción, para poder adaptarlo a la cultura de llegada hay que recurrir a técnicas como la compensación, la omisión, la neutralización, etc. y no será posible una traducción literal como así requieren los estudios prescriptivos. Si hay más de una persona en la escena, comúnmente solo se podrán traducir los segmentos de un personaje y, a lo sumo, dos, ya que no sería posible debido a los tiempos requeridos de la subtitulación, de los cuales se hablará más adelante.
- ♦ En el doblaje no tiene la problemática de segmentación y síntesis, pero debe tener en cuenta que cada lengua tiene una forma de expresarse y, por consecuencia, la extensión será diferente. Los cambios originales en el texto se aprecian en el doblaje debido a otra problemática como es la sincronización de la articulación visual y los sonidos. Por lo tanto, en la traducción del doblaje se pueden llegar a hacer cambios importantes respecto al original e incluso respecto a la subtitulación, porque tiene que coincidir y quedar natural lo que se oye con la forma en la que se mueven los labios de la persona a la que se está doblando.

Como apunta Botella (2007): «Si el traductor debe respetar estas imágenes, tendrá que sacrificar palabras, referencias e incluso fragmentos del texto original. Obviamente, la vulneración del texto original, la falta de equivalencia entre el original y su traducción impide a todas luces que la escuela prescriptivista pueda considerar el doblaje o el subtitulado como traducción.»

3.2 *Enfoque descriptivo*

Debido a las diferencias entre los estudiosos del enfoque prescriptivo, surge el enfoque descriptivo, en el que el texto traducido tiene una importancia similar o igual al texto original. Desde este punto de vista, se parte de la base en que la sociedad y la cultura tienen una importancia y un peso en la traducción, no solo en la audiovisual, sino también en el resto de las tipologías de traducción. Debido a esta contemplación de la cultura de llegada, las variedades del habla de distintos estratos sociales, etc. no se pueden omitir o pasar por alto esos componentes a la hora de traducir. Es importante remarcar que con este enfoque se establecen las normas de traducción de cada época y de cada sociedad, normal que perduran hoy en día pero que están sujetas a cambios socioculturales.

A partir de ese momento, el doblaje y la subtitulación se empiezan a incluir en corpus de estudio, así como explicar los procesos interlingüísticos que se han empleado en la traducción de estos.

Desde el estudio descriptivo, el traductor tendrá que adaptarse a las características propias de la subtitulación y el doblaje. Se aceptan, pues, distintas traducciones de una misma obra, debido a que están sujetas a las normas y a la recepción social que hayan tenido, teniendo en cuenta que debe adaptarse a los distintos medios.

Según Botella (2007), “el hecho de elegir la subtitulación supondrá una preferencia por la extranjerización, debido, entre otros factores, al anteriormente mencionado *gossiping effect*, mientras que optar por el doblaje responderá a una preferencia por la familiarización o, en palabras de Venutti (1995), *domestication*.”

La TAV se originó a principios del siglo XX con el objetivo de hacer accesible los filmes sonoros a otras lenguas y culturas (Chaume, 2013).

3.3 *Características generales de la subtitulación*

La traducción de subtítulos presenta unas particularidades que la hacen única dentro del ámbito de la TAV. Al contrario que en el resto de las modalidades de traducción, aquí el texto meta no sustituye al texto original, sino que ambos aparecen de manera sincrónica en la pantalla (Díaz Cintas y Remael, 2007:8; Pedersen, 2011:9). Se dice que los mejores subtítulos son aquellos que pasan desapercibidos (Díaz Cintas y Remael, 2007:185) y que no suponen un esfuerzo añadido al espectador (Bermejo, 2019). No existe una regla de oro que establezca cómo han de ser los subtítulos convencionales, aunque la mayoría de las empresas tienen unas directrices similares en cuanto a duración, tamaño y posición en pantalla. A continuación, se presenta un

resumen de las consideraciones técnicas y lingüísticas que se han seguido para la realización de los subtítulos de este proyecto, basadas en las recomendaciones de Díaz Cintas y Remael (2007:81-96) para la creación de subtítulos convencionales para un público oyente. (Suárez, 2022).

3.3.1 Aspectos técnicos de la subtitulación

Como explica Suárez (2022) en su estudio, unos de los límites que caracterizan la subtitulación son los límites de espacio y tiempo. En cuanto al formato de los subtítulos, hay que tener en cuenta las siguientes consideraciones:

- ◆ Restricciones espaciales: un subtítulo no debe tener más de dos líneas y tampoco se debería exceder el máximo de 37 caracteres por línea.
- ◆ Posición de los subtítulos: Los subtítulos deben posicionarse en el centro de la pantalla, en la parte inferior. Esto es un requisito indiscutible que únicamente se alterará cuando interfieran con otros subtítulos; dado el caso, estos se situarán en la parte superior central de la imagen.
- ◆ Restricciones temporales: es sumamente importante que los subtítulos sean legibles y que respeten unos parámetros de tiempo establecidos. Para ellos, se establece una velocidad máxima de lectura que se mide en caracteres por segundo (cps), que suelen ser 17-20 cps. Además, el tiempo que los subtítulos permanecen en pantalla también determina su calidad: si aparece muy poco tiempo, podría darse un efecto *flash* en el que el espectador podría no ser capaz de leerlos a tiempo; por el contrario, los que tienen una duración demasiado larga pueden provocar el efecto opuesto, ser leídos más de una vez (Cerezo Merchán, 2019:31).

Como indica Suárez (2002), es fundamental que los subtítulos estén sincronizados; el texto ha de aparecer y desaparecer a la vez que la pista de audio. Además, se recomienda dejar una pausa de 2 o 3 fotogramas entre subtítulos para que el cerebro asimile que aparece un nuevo subtítulo. Esta pausa ha de ajustarse con cada vídeo. “En caso de que haya más de una persona hablando a la vez y no sea posible crear subtítulos diferentes debido a la velocidad del discurso, será necesario señalar de manera clara que hay más de un hablante en un mismo subtítulo”. Por último, en cada cambio de plano debe haber ausencia de subtítulos.

3.3.2 Aspectos ortotipográficos de los subtítulos

Suárez (2002) en su trabajo expone que la subtitulación sigue unas convenciones concretas a la hora de enfrentarse a las reglas tipográficas y de puntuación, que pueden variar dependiendo del país o, incluso, del cliente (Díaz Cintas y Remael, 2007:103). En el capítulo 5 de su monografía, Díaz Cintas y Remael (2007:102-143) sugieren diversos consejos referentes a la puntuación; en esta sección se han incluido los más relevantes para este proyecto, teniendo en cuenta los casos que han aparecido en el material de trabajo:

- ◆ Comas: a pesar de que la coma es el signo ortográfico más versátil, no se debe abusar de su uso para indicar que la frase continúa en el siguiente subtítulo puesto que puede confundirse con un punto.
- ◆ Puntos suspensivos: se utilizarán para indicar vacilaciones, ideas inacabadas, pausas exageradamente largas o interrupciones, entre otros casos. Por cuestiones de espacio, no se recomienda indicar por medio de puntos suspensivos que una oración continúa en el siguiente subtítulo; la falta de una coma o un punto, junto con la ausencia de mayúscula inicial, son indicadores suficientes de esa continuidad.
- ◆ Paréntesis: se desaconseja su uso.
- ◆ Comillas inglesas: usadas para introducir estilo indirecto o citas textuales. Solo se ha añadido este signo al inicio y al final de la cita, independientemente del número de subtítulos que ocupe el mensaje.

3.3.3 Aspectos lingüísticos

Debido a las convenciones espaciotemporales expuestas anteriormente, en muy pocas ocasiones será posible producir una transcripción literal o verbatim en la que los subtítulos se correspondan totalmente con la pista de audio, con una sincronización absoluta y sin problemas de legibilidad. Para poder transmitir el mismo mensaje que la pista de audio ciñéndose a las restricciones de tiempo y espacio, Díaz Cintas y Remael (2007:144-179) nombran diversas estrategias que se pueden seguir (Suárez 2002):

- ◆ Reducción: es uno de los rasgos lingüísticos más característicos de la subtitulación. Puesto que el discurso oral es, normalmente, mucho más rápido que el escrito, es necesario buscar la manera de transmitir el mismo mensaje con un menor número de palabras para poder encajar el discurso dentro de los límites de espacio y tiempo. Según

el grado de reducción, se puede hablar de condensación y reformulación (reducción parcial), o de omisión (reducción total).

- ◆ Condensación y reformulación: estas estrategias se utilizan para mantener la integridad del mensaje. Consisten, entre otros ejemplos, en simplificar la información por medio de tiempos o perífrasis verbales más simples, usar sinónimos más cortos, remplazar sustantivos con pronombres, unir dos frases en una o cambiar de estilo indirecto a directo.
- ◆ Omisión: siempre que se pueda, se priorizará una reducción parcial frente a una total. Sin embargo, a veces se presentan situaciones en las que la única solución posible es la eliminación de una parte del texto que no sea relevante. Este es el caso, por ejemplo, de las repeticiones, los vocativos, construcciones agramaticales, interjecciones y expresiones carentes de significado («Oh», «ah», «ok», «ya sabes», «bueno», etc.) (Georgakopoulou, 2009:27).
- ◆ Segmentación: se trata del otro rasgo distintivo de este ámbito de la TAV. Es primordial ofrecer una segmentación adecuada; una incorrecta segmentación dificulta la comprensión de los subtítulos, los hace pesados y el exceso de esfuerzo por parte del espectador puede traducirse en una pérdida de interés en el producto. Díaz Cintas y Remael (2007:172-178) sugieren lo siguiente:
 - colocar una frase por línea;
 - mantener las unidades de significado juntas (adjetivo + sustantivo/adverbio, verbo + adverbio, artículo + sustantivo, preposición + sintagma preposicional, etc.), las preguntas con sus respuestas, y las formas verbales compuestas; o no separar verbos de objetos directos o indirectos;
 - dividir las líneas después de un signo de puntuación y antes de una conjunción o preposición.

4. Comunicación audiovisual y accesibilidad a los medios de comunicación

La comunicación audiovisual es un proceso mediante el cual se transmiten mensajes utilizando medios que involucran tanto elementos visuales como sonoros, como la televisión, el cine, la publicidad, el videojuego y los contenidos en línea, entre otros. (Grau, 2002)

La obra de McLuhan *Understanding Media: The Extensions of Man* explora cómo los medios de comunicación han transformado la forma en que las personas perciben el mundo y se relacionan entre sí.

Tal como explica Gabay (2018) en su estudio *Evolución de las redes sociales, de la comunicación y del marketing: análisis de su interrelación*, que a su vez cita a Moragas (2012) donde indica que la comunicación y la cultura son consideradas “dos caras de una misma moneda”.

Según Gabay (2018), la cultura se puede definir de dos formas: la cultura entendida en sentido aristocrático, una vertiente con enfoque colonialista que no da cabida a la diversidad cultural y, por otra parte, la cultura entendida desde el punto más amplio y diverso: “[...] referidas al conjunto de formas de vida y de organización del hombre inserto en una sociedad”.

La cultura de masas se originó tras la aparición de los medios de comunicación de masas, como lo son la prensa en el siglo XV, el cine y la radio en el siglo XIX y, por último, la televisión e Internet en el siglo XX. “La cultura se concibe como una construcción histórica de un sistema de valores, indisociables de las construcciones discursivas de los medios de comunicación. La visión intercultural toma relevancia en el análisis de las redes sociales y su alcance local y global. Los medios de comunicación y las tecnologías tienen un papel crucial en el proceso de la cultura”. (Gabay, 2018).

En la década de 1970 la televisión se hizo más accesible para todo el mundo y la publicidad se hizo más diversa. Las tecnologías de la información hacen posible la rápida unión de la cultura y la comunicación a principios de nuestro siglo.

En 1982 la UNESCO definió que: «[...] en su sentido más amplio, la cultura puede considerarse actualmente como el conjunto de los rasgos distintivos, espirituales y materiales, intelectuales y afectivos que caracterizan a una sociedad o un grupo social.» Esta visión intercultural toma relevancia si se analizan luego las redes sociales y los aspectos locales y globales que abarcan. Los medios de comunicación y los nuevos usos de las tecnologías ocupan espacios centrales, cada vez más grandes en la cultura moderna. Es gracias a la comunicación audiovisual, en la que se incluyen los medios de comunicación y el avance de las tecnologías, que la cultura da forma a las redes sociales y, al mismo tiempo, estas se ven influenciadas por el avance de la tecnología. (Gabay, 2018).

Gabay (2018) apunta que los medios dejan de ser continuos o discontinuos para estar siempre a disposición del usuario siempre conectado. Esto también lo defienden Valdés et al. (2017) en su artículo, que se expondrá más adelante y en el que podremos contrastar la información sobre la comunicación audiovisual y los medios de comunicación.

Gabay (2018) cita a Moragas (2012) y este, a su vez, a Castells (2009) para explicar que la distribución de la comunicación ya no está solo en manos de grandes organizaciones, sino que también pequeñas instituciones pueden convertirse en medios de comunicación gracias a la “autocomunicación de masas”.

La comunicación audiovisual ahora se centra en la producción de contenido, en vez de centrarse en la posibilidad de difundir dicho contenido. Debido a que vivimos rodeados de información

que se va renovando y actualizando constantemente, existe como consecuencia una sobreproducción de contenido, por lo que se prima la calidad a la cantidad como parámetro y se deben escoger con cautela las fuentes para que sean fiables. Gabay (2018)

Actualmente la comunicación audiovisual, que a su vez intenta difundir la cultura, extraen la información para los usuarios tanto de noticias, información o hechos tanto a nivel internacional como local, así pues, se entiende que la información internacional y la local son complementarias y no excluyentes. De este modo, la tecnología pasó de considerarse un instrumento a formar parte la cultura y de la cotidianeidad.

Para continuar, tal como exponen Valdés et al. (2017) en su estudio *Comunicación audiovisual, nuevos paradigmas y tendencias*, la comunicación audiovisual ha sufrido grandes cambios, mejoras, redefiniciones, etc. Existen unos nuevos paradigmas de la comunicación, los medios intentan generar servicios multiplataforma a los cuales el usuario accede desde múltiples terminales. Esas transformaciones llevan a cambios en los sistemas productivos y en el consumo (Lavín y Silva, 2015). Como consecuencia, se ha desarrollado más la televisión temática, se consolida la televisión interactiva con programas en los que el espectador puede participar.

Valdés et al. (2017) nos presentan una serie de datos muy valiosos en cuanto al avance de la comunicación audiovisual y la accesibilidad: En 2015 la AIMC (Asociación para la Investigación de Medios de Comunicaciones) realizó una encuesta donde se recogieron casi 21.000 respuestas y se llevó a cabo un estudio que reveló que hubo un cambio en las tendencias: se daba paso al fenómeno de la multitarea al ver la televisión junto al acceso simultáneo a Internet.

Valdés et al. (2017) apunta que los contenidos más consumidos son noticias (42,3 %), películas (40,3 %), entrevistas (35,6 %), series de ficción extranjeras (34,5 %) y retransmisiones deportivas (30,1 %). Por otro lado, la televisión se usa un 21,2 % de los encuestados para conectarse a Internet, duplicaba el uso de 2013 a 2015, y ahora será un porcentaje mucho mayor, ya que, con el nacimiento de las televisiones inteligentes, tenemos acceso a cualquier aplicación y web desde la televisión. Debido a ese uso del Internet a través de la televisión, un 84,9 % del contenido que se visualiza son contenidos audiovisuales online.

Según el estudio de AIMC, (Valdés et al., 2017), un 75,2 % de los internautas declaraban usar las redes sociales a diario, entre ellas estaban Facebook, Twitter, Google+, LinkedIn, Instagram. Los principales usos de las redes sociales son las relaciones de amistad, familiares, estar informado de la actualidad, compartir pasatiempos y las relaciones laborales.

La transmisión de vídeos en YouTube a través de dispositivos móviles ha crecido dos veces más rápido que la transmisión en ordenadores de escritorio. Desde los aproximadamente la comercialización de las televisiones inteligentes (2010 en adelante), la televisión a la carta o *streaming*

se ha incrementó. Hoy en día podría asegurar que es un porcentaje altísimo, puesto que casi todo el mundo goza de tener a su disposición servicios de *streaming* como Netflix, HBO, Prime Video, Disney+, etc. Esto ha producido que los contenidos de la televisión corriente se fuesen especializando y la oferta se tematizara. (Valdés et al, 2017). En Estados Unidos los consumidores con acceso a medios como Netflix consumía en 2015 un 20 % más de contenido que los consumidores con únicamente televisión en vivo, ya que gracias a los servicios de *streaming* se puede ver contenido sin anuncios (López, 2015). Netflix se ha convertido en el servicio más popular de *streaming* del mundo (Barredo, 2016).

La industria de la comunicación ofrece cadenas públicas y privadas, emisiones por cable, canales premium y plataformas digitales. Una de las novedades es la fascinación del espectador por la imagen audiovisual como exponente de un espectáculo donde lo real se convierte en sí mismo en valor noticia. Esto provoca que las noticias sean cada vez más cortas, el entretenimiento se vuelva noticia, se eleve a personas sin mérito periodístico a rango de celebridad, se coloquen anuncios entre noticias reales y que la red posibilite el acceso directo del público a las fuentes informativas sin la mediación de los comunicadores.

Esto significa que los dispositivos móviles irrumpen cada vez más en la comunicación audiovisual. Hoy en día consumismo mucho más contenido en redes sociales antes mencionadas, pero más aún con el auge de TikTok en 2020, que ofrece una gran cantidad de contenido y además muy variado con un alto porcentaje de *engagement*. (Peña et al., 2022)

Gracias a esta diversificación de la comunicación audiovisual, se crean nuevos puestos de trabajo según Vargas (2016) citado en el artículo de Valdés et al. (2017), como, por ejemplo: editor multiplataforma, editor de audiencias, editor *SEO* y analítica, editor multimedia, jefe de *lab media*, diseño web, *community manager*, productor periodístico y editor de visualizaciones y datos.

Valdés et al. (2017) razonan que la presencia de teléfonos móviles en el entorno tecnológico y el uso de *smartphones* ha desembocado en un cambio en la producción de contenidos audiovisuales. También explican que las redes sociales son formas de interacción sociales social por medio de Internet, un intercambio de información dinámico entre personas, grupos e instituciones que llegan a identificarse por necesidades, gustos o similitudes en temas generales.

Por consiguiente, en la web hay que combinar texto, fotos, audio y vídeo. Internet convierte a la comunicación en bidireccional y lo masivo cede a espacios ante lo personal. Además, el servicio es más personalizado para los usuarios y con Internet existe la posibilidad de ver, escuchar, leer y recibir solo los mensajes que desee.

Para concluir, Valdés et al. (2017) constatan que el servicio de contenido audiovisual en las redes sociales debe ser un medio al alcance de todos y que se pueda personalizar. Por otro lado, en las redes sociales existe un lenguaje diferente y tiene que seguir patrones específicos, como por

ejemplo una redacción de la narrativa sencilla, dinámica, con soportes gráficos y fácil de leer. También predomina la idea de que la información se actualiza constantemente en Internet. Finalmente, el usuario es el protagonista y se vuelve emisor. Para que pueda existir un buen intercambio de mensajes a través de las redes sociales, el abordaje debe ser horizontal, lo cual significa que el medio debe entenderla demanda de los usuarios. De ahí que haya surgido la presencia del gestor de comunidad digital, que no deja de ser alguien especializado en la cultura social media y el comportamiento de los usuarios en las redes sociales.

4.1 Contenido audiovisual y accesibilidad para personas con discapacidad auditiva

Retomando el tema tratado en el apartado 3.3. *Características generales de la subtitulación*, en este apartado se desarrollará en mayor profundidad cómo debe ser la subtitulación para sordos (SpS) o con discapacidad auditiva y la accesibilidad para este tipo de público, así como qué normas se han ido estableciendo con el paso del tiempo para hacer más estándar la traducción y la subtitulación.

Según el artículo de Díaz Cintas: *La accesibilidad a los medios de comunicación audiovisual a través del subtitulado y de la audiodescripción* publicado en 2010, el concepto de accesibilidad ha evolucionado con el tiempo. Es importante remarcar que la accesibilidad es un concepto «que implica la integración social de personas con discapacidades no solo físicas sino también sensoriales y cognitivas.» Díaz Cintas afirma que este colectivo ha sufrido una gran marginación por parte de la sociedad. Esto ha afectado a lo largo de los años el mundo de la comunicación audiovisual. Es por eso por lo que se está llevando a cabo desde la última década una gran labor desde no solo la traducción general, sino también la traducción audiovisual. Es este el punto en el que se quiere enfocar este estudio con el desarrollo de la accesibilidad a vídeos con Whisper. La transcripción y la traducción audiovisual juegan un papel fundamental en la accesibilidad de contenidos de todo tipo, desde películas disponibles para un público general, series de televisión, hasta vídeos que podemos encontrar en las redes sociales que se ven a diario, bajo los intereses individuales de los usuarios de estas.

Tal y como expone Díaz Cintas (2010), una de las dificultades ha sido hacer accesibles los medios de comunicación audiovisual, ya que tiene una posición dominante a la hora de educar, informar, vender, etc. La comunicación audiovisual es uno de los pilares dentro de la sociedad.

A pesar de que los medios de comunicación audiovisual sea un tema sobre el que se ha escrito, el material destinado a la accesibilidad existente es muy escaso. Tal como apuntaba Gabay (2018), se puede contrastar y refutar con Díaz Cintas (2010), que el acceso a los medios de comunicación

audiovisual y a la información tiene un peso enorme en la formación individual y sociocultural de todo ser humano.

Díaz Cintas (2010), expone que antiguamente lo esencial para poder expandir el conocimiento individual era tener acceso a libros. En comparación, hoy en día lo primordial para llevar a cabo esa tarea individual sería tener acceso a Internet y la comunicación y, por ende, a la comunicación audiovisual. Es por esto por lo que se deben hacer accesibles cuantos materiales se pueda. No obstante, es una labor en la que deberían verse implicadas más instituciones relevantes. Tal y como apuntaban Valdés et al. (2017), la cultura podía entenderse antiguamente desde un punto de vista colonial y no diverso.

Si, históricamente, el acceso a los libros era la clave para ampliar el conocimiento y el saber, hoy en día se puede extrapolar a tener acceso a la televisión y a internet. En referencia a lo que cita el autor respecto a la accesibilidad para las personas con discapacidad auditiva, no se ha desarrollado tanto como se debería. El mayor número de medios que se adaptan y se deben hacer accesibles son materiales como el cine, vídeos o DVD, televisión, Internet, obras teatrales, ópera, congresos, exposiciones en los museos, etc. El objetivo de hacer accesibles todo tipo de material sería llegar a una gran o completa integración de los individuos con discapacidad auditiva o personas sordas.

La Asamblea General de las Naciones Unidas impulsa enormemente hacer accesibles programas de televisión, películas, teatro y otros materiales culturales. Por otra parte, según la Convención de las Naciones Unidas no es que haya definidas unas normas, pero sí que hay un acuerdo aproximado de los procedimientos a llevar a cabo para lograr que todos estos formatos sean accesibles:

1. El subtitulado para sordos (SpS) y personas con discapacidad auditiva
2. La audiodescripción (AD) para personas ciegas y con discapacidad visual
3. La interpretación de lengua de signos (ILS)

Estos tres procedimientos son los más estudiados recientemente. A pesar de que el tanto el SPS como la AD últimamente son cada vez más objeto de estudio, hasta hace relativamente poco no encajaban con las disciplinas académicas tradicionales de manera de manera similar a lo que ocurría con la traducción audiovisual, como se vio anteriormente en el artículo de Botella (2007), entre las cuales podemos encontrar el subtitulado, el doblaje, el *voice-over*, etc.

Citando a Díaz Cintas (2010), el SPS y AD “los convertían en objetos de estudio poco deseables”, es decir, lo que el autor intenta transmitir es que hay un sector de personas privilegiadas en la sociedad que se dedicaban a estudiar, por ejemplo, la traducción generalizada pero que a su vez no les interesaba estudiar a fondo y hacer los materiales accesibles para el colectivo con discapacidades auditivas y, por ende, hacerlos accesible.

No obstante, el presente estudio no solo busca trabajar con la accesibilidad con Whisper para las personas sordas o con discapacidades auditivas, sino también para personas oyentes.

Díaz Cintas (2010) en su estudio cita a Delabastita (1989) y Mayoral (2001) y destacan la gran recepción y acogida que ha tenido la labor de hacer accesibles a los medios de comunicación audiovisual en la traducción general y la TAV. El aumento en el interés por la accesibilidad va a la misma marcha que los siguientes desarrollos técnicos:

- ◆ El cambio de la tecnología analógica a la digital
- ◆ La obligatoriedad por ley de las cadenas públicas de ofrecer unos porcentajes mínimos de programas accesibles con SpS, ADE y ILS
- ◆ Concesión de licencias como la prestación de servicios mínimos de accesibilidad

Los progresos que ha habido en cuanto a la legislación española para hacer accesibles los medios de comunicación y la comunicación audiovisual en concreto figuran las siguientes leyes:

Según Díaz Cintas (2010) en 2003 AENOR publicó en España la norma UNE 153010 titulada *Subtitulado para personas sordas y personas con discapacidad auditiva. Subtitulado a través de teletexto*. Esta norma consiste en ofrecer en la parte inferior de la pantalla un texto escrito de lo que se dice, quién lo dice, cómo se dice, lo que se oye y lo que se ve. Estas normas las podemos ver analizadas anteriormente en el estudio de Suárez (2022). Todo el programa audiovisual con SPS debe tener tres componentes principales: la palabra oral, la imagen y los subtítulos. Estos han de estar sincronizados con la imagen.

La información subtitulada se ha de presentar en pantalla el tiempo suficiente para que los espectadores puedan leerla con un ritmo que permita una lectura cómoda del texto, así como el disfrute de las imágenes y se atenga a una velocidad de lectura adecuada para el público al que va dirigido.

El subtitulado para personas sordas se originó en el Reino Unido en 1979. La cadena televisiva *BBC* retransmitió un documental sobre niños sordos titulado *Quietly in Switzerland*. En comparación, no fue hasta septiembre 1990, once años más tarde que nuestros vecinos anglosajones, que en España también se transmitiría por primera vez un programa accesible. La cadena pionera fue TV3. Más tarde se le sumarían a nivel nacional TVE.

Díaz Cintas (2010) afirma que actualmente tenemos en nuestro poder una solución a medias, ya que se intenta satisfacer tanto a personas oyente como personas con discapacidad auditiva una misma pista de subtítulos, aunque esta situación se ha visto mejorada en los últimos años, con las normas UNE, ya que ha habido un aumento en la producción de contenido o material para personas con discapacidad auditiva.

No obstante, al no haber una norma que todas las empresas sigan, esto conduce a una heterogeneidad en las prácticas aplicadas en cuanto a la TAV y los SPS.

III. Marco práctico

Es indiscutible el hecho de que vivimos en una época completamente tecnológica, mucho más que antes, aunque seguramente menos que en los años venideros. Se producen constantemente cambios y evoluciones de cualquier tipo de tecnología, desde la que menos conocemos hasta la que tenemos al alcance de nuestras manos en nuestras vidas diarias.

Puesto que vivimos en una era completamente digital en la que la mayoría de la población tiene acceso a Internet, podemos comunicarnos sin depender directamente de los grandes medios y podemos crear y compartir contenido en redes sociales. Es importante, pues, que el contenido sea accesible y se pueda crear de manera eficiente. Dentro de esta creación de contenido, entra en juego la traducción audiovisual. A diario se traducen miles de contenidos audiovisuales, por lo que llegar a poder usar herramientas de reconocimiento del habla que transcriban automáticamente, puede llegar a ser un gran avance para la traducción automática, la traducción audiovisual y la accesibilidad a este tipo de contenido.

A continuación, se realizará un análisis de archivos en formato audiovisual enfocado en dos perspectivas: el discurso formal e informal. Se evaluarán aspectos de corrección lingüística para demostrar la capacidad de transcribir correctamente de Whisper para poder utilizar los resultados con fines profesionales en la traducción general y, en específico, de la traducción audiovisual. El análisis de los resultados de Whisper se comparará con la transcripción proporcionada por YouTube y se cotejarán los resultados con la transcripción humana.

El objetivo último de la comparación de estos dos tipos de transcripciones generadas automáticamente es llegar a justificar o refutar si la transcripción automática generada con inteligencia artificial facilita la tarea de la traducción audiovisual y la accesibilidad al contenido audiovisual, además de valorar si se puede utilizar la TA (traducción automática) o la TAPE (Traducción automática con posesición) con los resultados obtenidos.

1. Procedimiento utilizado en la preparación del módulo de trabajo de Whisper

Este es el procedimiento que se ha llevado a cabo en el presente estudio para la instalación de Whisper.

1.1 Código de Python inicial – diciembre 2022-marzo 2023

El código que se muestra a continuación tiene elementos que no se han utilizado para los resultados finales del presente trabajo puesto que la metodología usada en esta parte inicial ha sido descartada. Dicha metodología consiste en que Whisper transcribe en tiempo real una grabación de voz, se valora su corrección lingüística y se sopesa si esta transcripción es lo suficientemente bueno sin posesición humana para usar la traducción automática.

Este código únicamente ha dado resultados prematuros en Google Colab, entorno en el que se decide no seguir la investigación por dificultades en la ejecución del código de Python.

El aspecto positivo del uso del entorno de Google Colab es que permite crear y ejecutar el código con un entorno GPU, que es mucho más rápida que una CPU, además de no tener que cargar con otras dificultades o importunidades que le puedan ocasionar al usuario, como pueden ser la instalación de programas o software para poder ejecutar el código, el uso de la memoria del ordenador personal, etcétera.

El aspecto negativo de esta interfaz es que solamente se puede usar y mantener el entorno GPU durante 24 horas, después se borran los datos y hay que volver a iniciar el proceso desde cero.

```
pip install git+https://github.com/openai/whisper.git
pip install --upgrade --no-deps --force-reinstall
git+https://github.com/openai/whisper.git
import whisper
model = whisper.load_model("base")

def transcribe(audio):

    #time.sleep(3)
    # load audio and pad/trim it to fit 30 seconds
    audio = whisper.load_audio("prueba1ES.m4a")
    audio = whisper.pad_or_trim(audio)

    # make log-Mel spectrogram and move to the same device as the model
    mel = whisper.log_mel_spectrogram(audio).to(model.device)

    # detect the spoken language
    _, probs = model.detect_language(mel)
    print(f"Detected language: {max(probs, key=probs.get)}")
```

```

    # decode the audio
    options = whisper.DecodingOptions()
    result = whisper.decode(model, mel, options)
    return(result.text)
! pip install gradio
import gradio as gr
import time
gr.Interface(
    title = 'OpenAI Whisper - Prueba TFM Anna Acedo',
    fn=transcribe,
    inputs=[
        gr.inputs.Audio(source="microphone", type="filepath")
    ],
    outputs=[
        "textbox"
    ],
    live=True).launch()
model = whisper.load_model("base")
# load audio and pad/trim it to fit 30 seconds
audio = whisper.load_audio("prueba1ES.m4a")
audio = whisper.pad_or_trim(audio)

# make log-Mel spectrogram and move to the same device as the model
mel = whisper.log_mel_spectrogram(audio).to(model.device)

# detect the spoken language
_, probs = model.detect_language(mel)
print(f"Detected language: {max(probs, key=probs.get)}")

# decode the audio
options = whisper.DecodingOptions()
result = whisper.decode(model, mel, options)

# print the recognized text
print(result.text)

```

Detected language: es
 Hola, soy Ana Cedo, hoy es 6 de enero del 2023 y esta es una grabación de prueba para la teciis de mi master en tecnologías de la traducción.

```

model = whisper.load_model("base")
# load audio and pad/trim it to fit 30 seconds
audio = whisper.load_audio("prueba2CA.m4a")
audio = whisper.pad_or_trim(audio)

# make log-Mel spectrogram and move to the same device as the model
mel = whisper.log_mel_spectrogram(audio).to(model.device)

```

```
# detect the spoken language
_, probs = model.detect_language(mel)
print(f"Detected language: {max(probs, key=probs.get)}")

# decode the audio
options = whisper.DecodingOptions()
result = whisper.decode(model, mel, options)

# print the recognized text
print(result.text)
```

Detected language: es
 Hola, bon dia a su clana, a que se se la proba número 5, no al whisper, a que está audio en concreto, esticraban en catalá, para ver ahora como es el rendiment de whisper en catalá. Y, ve, a hoy fondiamos la suelallad, al selvent blau, y que tenga un buen día

```
model = whisper.load_model("base")
# load audio and pad/trim it to fit 30 seconds
audio = whisper.load_audio("prueba3JA.m4a")
audio = whisper.pad_or_trim(audio)

# make log-Mel spectrogram and move to the same device as the model
mel = whisper.log_mel_spectrogram(audio).to(model.device)

# detect the spoken language
_, probs = model.detect_language(mel)
print(f"Detected language: {max(probs, key=probs.get)}")

# decode the audio
options = whisper.DecodingOptions()
result = whisper.decode(model, mel, options)

# print the recognized text
print(result.text)
```

Detected language: ja
 はいこんにちは私はアナセドですよろしくお願ひしますあの今日はとても晴れているんですけどとても寒いです最近命は面白くないです

```
model = whisper.load_model("base")
# load audio and pad/trim it to fit 30 seconds
audio = whisper.load_audio("prueba4CA.m4a")
audio = whisper.pad_or_trim(audio)

# make log-Mel spectrogram and move to the same device as the model
mel = whisper.log_mel_spectrogram(audio).to(model.device)

# detect the spoken language
```

```
_, probs = model.detect_language(mel)
print(f"Detected language: {max(probs, key=probs.get)}")

# decode the audio
options = whisper.DecodingOptions()
result = whisper.decode(model, mel, options)

# print the recognized text
print(result.text)
```

Detected language: ca
 Les classes de lletí eren aborrides, però també ho eren les matemàtiques, les ciències naturals, i en manú mesura les classes d'història. Malgrat tot apreníem, encara que no sabia dir si apreníem per que els professors ens aduïen amb la seva capacitat de evitar que ens distraiguessin.

Figura 2. Código de Python inicial.

En esta primera versión se incluye Gradio, un entorno o interfaz que dispone de 2 apartados. En un primer apartado el usuario se encuentra con un espacio en el que poder grabar en tiempo real la voz del interlocutor. En un segundo apartado el usuario encuentra la transcripción que proporciona Whisper. Los resultados de esa transcripción en tiempo real permiten a Whisper poder traducir al cabo de unos segundos. Cabe resaltar que no hace falta definir el idioma en el que el interlocutor habla, sino que, igual que en el código de Python que veremos a continuación, lo identifica por sí solo.

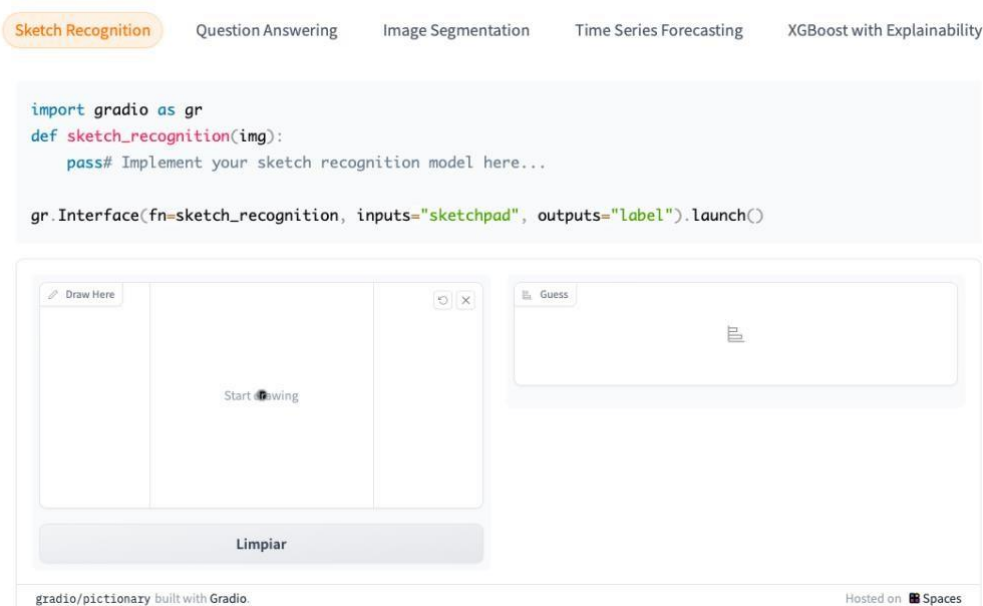


Figura 3: Muestra de Gradio original (Gradio, 2022)

```

Note: opening Chrome Inspector may crash demo inside Colab notebooks.

> create a public link, set "share=True" in "launch()".
Running on https://localhost:7860/

```

OpenAI Whisper - Prueba TFM Anna Acedo

audio

Record from microphone

Limpiar

output

Avisar

Figura 4: Muestra de Gradio personalizada. Código de Python inicial.

OpenAI Whisper - Prueba TFM Anna Acedo

audio

Record from microphone

Error

output

Avisar

Figura 5: Muestra de Gradio personalizada-Error. Código de Python inicial.

El presente trabajo ha derivado en optar por el análisis de archivos de audio en formato .m4a de vídeos extraídos y recopilados en su totalidad de la plataforma TikTok. Se ha utilizado un terminal MacBook Air del 2015 con el sistema operativo macOS Monterey, versión 12.2.1, con un procesador 1,6GHz Intel Core i5 de doble núcleo y una memoria de 6 GB 1600 MHz DDR3.

Se han analizado un total de 25 archivos de audio en inglés (EE. UU., Reino Unido), francés (Francia), catalán, valenciano, español (España) y japonés (Japón).

```

import whisper
import os
from datetime import datetime

#import gradio as gr
#import time

def run_whisper(filepath):
    model = whisper.load_model("base")

    # load audio and pad/trim it to fit 30 seconds
    audio = whisper.load_audio("prueba1ES.m4a")
    audio = whisper.pad_or_trim(audio)

```



```

# make log-Mel spectrogram and move to the same device as the model
mel = whisper.log_mel_spectrogram(audio).to(model.device)

# detect the spoken language
_, probs = model.detect_language(mel)
print(f"Detected language: {max(probs, key=probs.get)}")

# decode the audio
options = whisper.DecodingOptions(fp16 = False)
result = whisper.decode(model, mel, options)

# print the recognized text
print(result.text)

def transcribe(audiopath, resultpath="result.txt"):

    model = whisper.load_model("base")
    #time.sleep(3)
    # load audio and pad/trim it to fit 30 seconds
    audio = whisper.load_audio(audiopath)
    audio = whisper.pad_or_trim(audio)

    # make log-Mel spectrogram and move to the same device as the model
    mel = whisper.log_mel_spectrogram(audio).to(model.device)

    # detect the spoken language
    _, probs = model.detect_language(mel)
    print(f"Detected language: {max(probs, key=probs.get)}")
    result = model.transcribe(audio)
    print(result["text"])

    #decode the audio
    #options = whisper.DecodingOptions()
    #result = whisper.decode(model, mel, options)
    #return result.text

    with open(resultpath, "w") as file:

```

```

file.write(result["text"])

#def run_gradio():
#gr.Interface(
#title = 'OpenAI Whisper - Prueba TFM Anna Acedo',
#fn=transcribe,
#inputs=[
#gr.inputs.Audio(source="microphone", type="filepath")
#],
#outputs=[
#"textbox"
#],
#live=True).launch(share=True)

if __name__ == "__main__":
#run_whisper("prueba1ES.m4a")
#run_whisper("prueba2CA.m4a")
#run_whisper("prueba3JA.m4a")
#run_whisper("prueba4CA.m4a")
#run_whisper()
#run_gradio()
#run_gradio()
for elem in os.listdir("./audios"):
if elem.endswith(".m4a"):
transcribe("./audios/"+elem, os.path.join("RESULTADOS", elem+"_result.txt"))
print(datetime.now().strftime("%Y-%m-%d %H:%M:%S") + " - Starting...");

```

Figura 6: Código de Python definitivo.

Este es el script de Python creado para el presente trabajo. Utiliza la herramienta Whisper de OpenAI para transcribir archivos de audio y evaluar el reconocimiento de voz. A continuación, explicaré el código línea por línea:

- ◆ Líneas 1-3. En esta sección, se importan los módulos necesarios. **‘whisper’** es el módulo que proporciona la funcionalidad de la herramienta Whisper de OpenAI. **‘os’** se utiliza para manipular rutas de archivo y directorios, y **‘datetime’** se utiliza para obtener la fecha y hora actuales.

- ♦ Línea 8. La función '**run_whisper()**' carga el modelo de Whisper utilizando la función '**load_model()**' y especifica "base" como el nombre del modelo.
- ♦ Líneas 12-13. Se carga el archivo de audio utilizando la función '**load_audio()**' y se realiza un ajuste de duración para que coincida con 30 segundos utilizando la función '**pad_or_trim()**'.
- ♦ Línea 16. Se calcula el espectrograma logarítmico de Mel del audio utilizando la función '**log_mel_spectrogram()**'. Luego, se mueve al mismo dispositivo que el modelo utilizando el atributo '**device**' del modelo.
- ♦ Líneas 19-20. Se utiliza la función '**detect_language()**' para detectar el idioma hablado en el audio. La función devuelve una tupla con información sobre el idioma detectado y las probabilidades asociadas. Luego, se imprime el idioma detectado en la consola.
- ♦ Líneas 23-24. Se configuran las opciones de decodificación mediante la creación de una instancia de '**DecodingOptions**' y se utiliza la función '**decode()**' para transcribir el audio utilizando el modelo Whisper. El resultado de la transcripción se almacena en la variable '**result**', y se imprime el texto reconocido en la consola.
- ♦ Líneas 29-44. La función '**transcribe()**' sigue un enfoque similar a '**run_whisper()**'. Carga el modelo, el audio, realiza el ajuste de duración, calcula el espectrograma de Mel, detecta el idioma y transcribe el audio utilizando la función '**transcribe()**' del modelo. Luego, guarda el resultado en un archivo de texto especificado por '**resultpath**'.
- ♦ Líneas 67-78. Esta sección se ejecutará solo si el archivo se ejecuta directamente, no si se importa como módulo. Aquí, itera a través de los archivos en el directorio "./audios" y selecciona los archivos que tienen la extensión ".m4a". Luego, llama a la función '**transcribe()**' para cada archivo, especificando la ruta del archivo de audio y la ruta de salida para el resultado transcripción. Finalmente, imprime la fecha y hora actuales. De esta manera, podemos calcular cuánto tiempo ha tardado en ejecutarse el código y proporcionar los resultados.

1.2 Metodología de la parte práctica

Para esta parte del trabajo se han analizado 25 archivos de audio en formato .m4a que, a su vez, están extraídos de archivos de vídeo .mp4 de la plataforma TikTok. En una fase experimental previa del trabajo se hizo uso de archivos de audio procedentes de WhatsApp o grabaciones simples hechas con un dispositivo móvil, pero se descartó dicha idea puesto que, para evaluar la calidad de transcripción de YouTube hace falta archivos de vídeo. Es por este motivo por el que la dirección del trabajo se cambió y ha resultado en el siguiente análisis.

La base del trabajo es evaluar la transcripción de ambas herramientas para después valorar si se puede hacer uso total o parcial para la traducción audiovisual utilizando la traducción automática. Cabe destacar que se ha elegido utilizar como objeto de estudio Whisper y compararlo con YouTube porque ambas son herramientas gratuitas y accesibles para todo el mundo. La desventaja de Whisper es que, al ser *open source*, se requieren conocimientos más avanzados para poder utilizarlo.

A continuación, veremos analizados y corregidos las transcripciones de Whisper: 25 muestras en 5 idiomas y de YouTube: 18 muestras en 4 idiomas. El motivo por el cual las muestras no coinciden en número se verá más adelante, en el apartado Conclusiones. Los materiales de referencia son parte tanto de la parte práctica como del apartado Anexos. Estos están en sus respectivas carpetas que siguen una jerarquía específica. En primer lugar, hay la carpeta “CODIGO_AUDIOS_RESULTADOS_WHISPER”, que contiene el código de Python previamente visto en este trabajo, otra carpeta con los archivos de audio extraídos de los vídeos de TikTok, llamada “AUDIOS”. Los archivos de audio se han obtenido de los vídeos con la herramienta QuickTime Player que proporcionaba una manera fácil de exportar los audios de cualquier vídeo utilizado para la parte práctica de este trabajo.

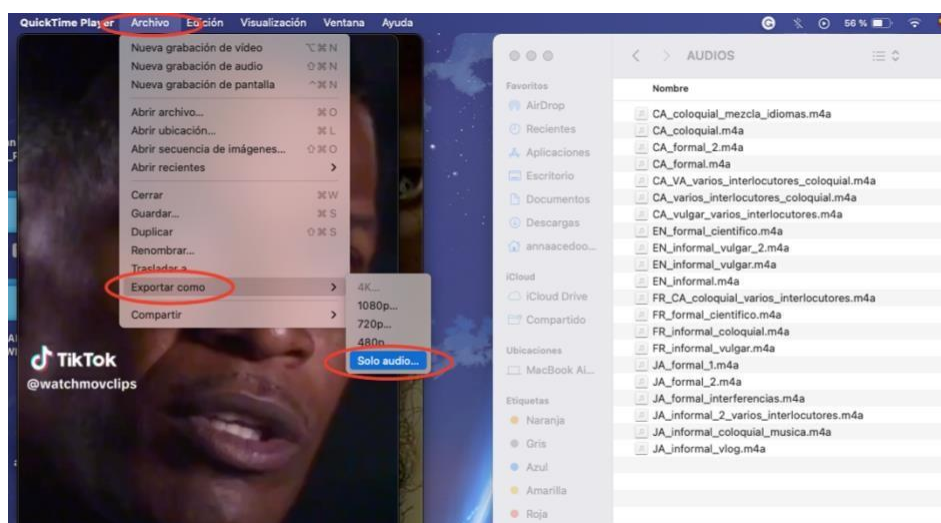


Figura 7: Exportación de audio con QuickTime Player en MacBook Air.

Por último, hay otra carpeta con los resultados de Whisper en formato .txt, llamada “RESULTADOS”. Estos archivos se han generado y almacenado en dicha carpeta automáticamente con el código de Python.

En segundo lugar, hay la carpeta “Documentos”, que contiene el presente trabajo, un archivo Excel en el que hay una fórmula que hace un recuento de caracteres por línea, para así no superar el máximo que exige la norma, comentada previamente en el apartado 1.2. Aspectos técnicos de la

subtitulación. También hay otro documento Excel con una fórmula que hace recuento de la puntuación de cada muestra transcrita. Dado que los archivos .srt contienen marcas de tiempo y están divididos por líneas de subtitulación, se ha utilizado el recuento de caracteres por línea con el fin de analizar la calidad de las marcas de tiempo y la longitud de los subtítulos generados automáticamente de YouTube. En esta segunda carpeta también hay un archivo de vídeo con una de las pruebas iniciales que se hicieron en Google Colab con el antiguo código, comentada por la autora del presente trabajo.

En tercer lugar, hay la carpeta “SUBTITULOS_YOUTUBE” donde se han almacenado manualmente los archivos .srt exportados desde YouTube. Es importante mencionar que el método que se ha de seguir en YouTube para poder obtener las transcripciones de los subtítulos generados automáticamente es realmente tedioso si se cuenta con una gran cantidad de archivos de vídeo. Al contrario que con Whisper, que gracias al código de Python se almacenan manualmente sin esfuerzo humano, en YouTube Studio se tiene que, primero de todo, disponer de una cuenta de YouTube, subir los archivos de vídeo, añadir parámetros e información sobre el vídeo que el usuario se dispone a subir, como por ejemplo si el contenido es apto para niños o no, el idioma del vídeo, si se permiten remixes del vídeo, etc.

Además, algo que dificulta también el trabajo es que con una cuenta de usuario normal hay un límite de vídeos que YouTube deja subir. Para poder proceder, fue necesario verificar la identidad de la propietaria de la cuenta proporcionando una fotografía del DNI a Google. Pasado un periodo de 1 día máximo, se recibe una notificación conforme se ha aprobado o denegado el acceso a subir más vídeos. Otra limitación de las transcripciones generadas automáticamente por YouTube es que solamente dispone, por el momento, de 14 idiomas: alemán, coreano, español, francés, indonesio, inglés, italiano, japonés, neerlandés, portugués, ruso, turco, ucraniano y vietnamita. Si hay varias pistas de audio en un vídeo, los subtítulos automáticos estarán en el idioma predeterminado. Por contra, el modelo de Whisper dispone de 96 idiomas:

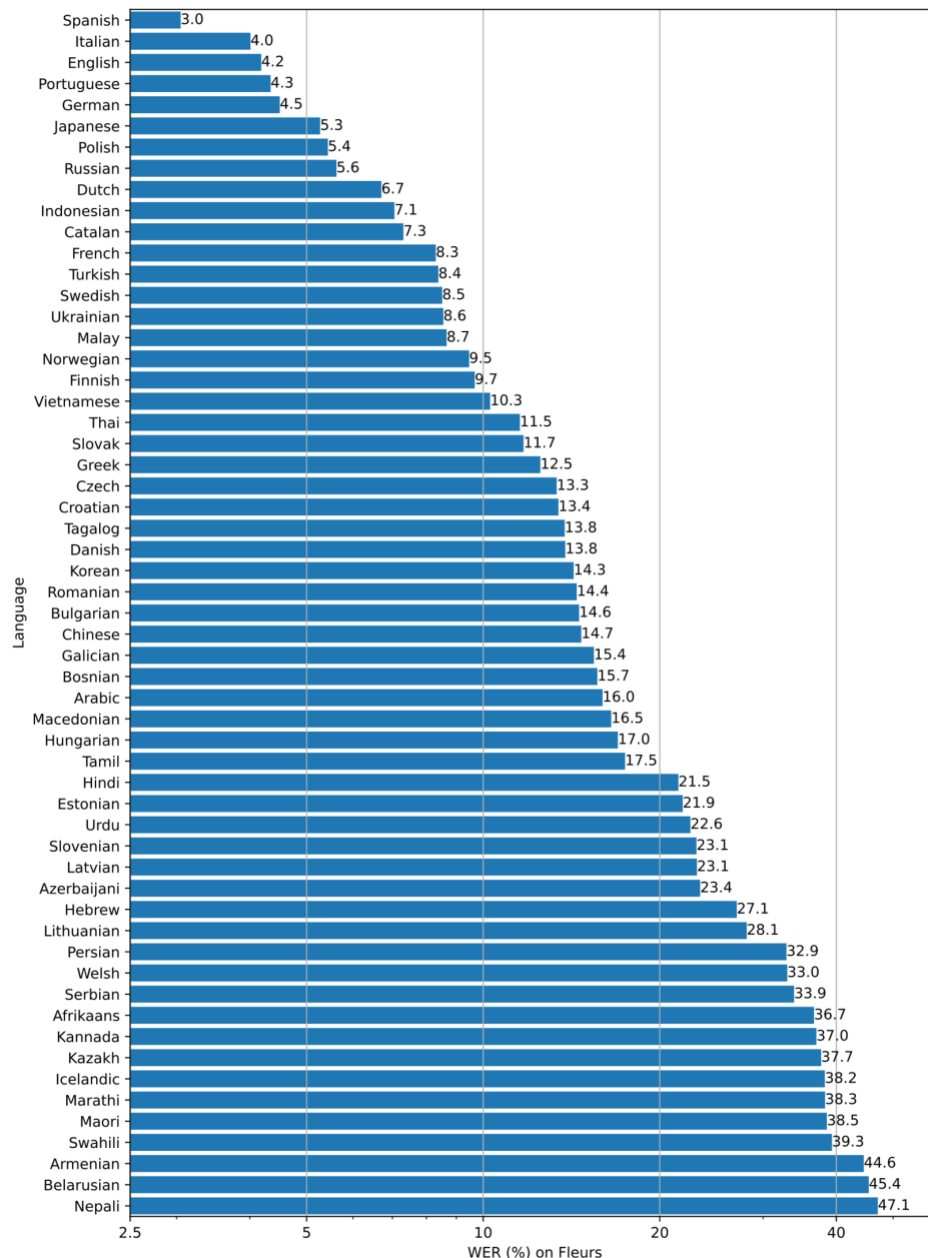


Figura 8: Gráfica de WER. GitHub de Whisper (Radford et al., 2022)

En el repositorio de GitHub los creadores de Whisper explican que esta gráfica muestra el margen de error por palabra (WER):

“Whisper’s performance varies widely depending on the language. The figure below shows a WER (Word Error Rate) breakdown by languages of the Fleurs dataset using the large-v2 model (The smaller the numbers, the better the performance). Additional WER scores corresponding to the other models and datasets can be found in Appendix D.1, D.2, and D.4. Meanwhile, more BLEU (Bilingual Evaluation Understudy) scores can be found in Appendix D.3.”

Finalmente, hay la carpeta “VIDEOS”, en la que hay almacenados los materiales de referencia en formato .mp4 extraídos de TikTok.

A continuación, se exponen los resultados de la transcripción con Whisper con guardado automático con formato .txt, así como los resultados de la transcripción con YouTube con guardado manual en formato .srt. Se han elegido estos formatos por la versatilidad que proporcionan con otras herramientas TAO, como pudieran ser SDL Trados, MemoQ, Phrase, Subtitle Edit, Aegisub, SDL Passolo, OmegaT, etc. La parte positiva de los archivos .txt es que es un tipo de archivo sin formato y fácilmente se puede trasladar la información a dichas herramientas. Por otro lado, la parte positiva de los archivos .srt es que suelen ser de los más usados en el mundo de la traducción audiovisual y subtitulación, por lo que son altamente compatibles con las herramientas TAO y muchos reproductores multimedia.

◆ *Español*

Transcripción de Whisper	(—)Vale, voy a deciros de qué va la hay que ver . Trata sobre una putita que es una máquina de polla , no sea(,) por lo menada , por lo tarde, por la noche, polla, polla, polla, polla, polla, polla. (—)¿Cuántas polla son? (—)Un montón. (—)Entonces la tía un día se encuentra con un hijo puta que tiene un perazo de pollo . El eios como Charles Bronson en la gran evasión , Cava Tunes . La tía ha encontrado la polla de su vino , ¿verdad? Y siente algo que no había sentido nunca. Dola . Dola . (—)Chute, vich .
Transcripción de YouTube	vale(,) Voy a deciros de que va like a verte (.) trata sobre una putita que es una máquina de [] o sea(,) por la mañana(,) por la tarde(,) por la noche(,) [] [] [] [] [] (¿)Cuántos pollos (?) Son un montón(.) entonces la tía un día se encuentra con un hijo [] que tiene un pedazo de pollo necios como Charles Bronson en la gran evasión (,) cava túneles (.) ha encontrado la [] de su vida y siente algo que no había sentido (nunca)(...)dolor qué dolor(.)
Corrección	—Vale, voy a deciros de qué va <i>Like a Virgin</i> . Trata sobre una putita que es una máquina de follar; o sea, por lo mañana, por lo tarde, por la noche, polla, polla, polla, polla, polla, polla. —¿Cuántas pollas son? —Un montón. —Entonces, la tía un día se encuentra con un hijo puta que tiene un pedazo de pollón. El tio es como Charles Bronson en <i>La gran evasión</i> , cava túneles. La tía ha encontrado la polla de su vida y siente algo que no había sentido nunca...dolor, dolor. —¿Chew...Toby Chew?

Tabla 1: *ES_coloquial_vulgar.m4a_result*

Transcripción de Whisper	¿Tienes caspador Matiti seborreca ? Entonces este vídeo es para ti. La dermatiti(s) seborr(e)ica es un trastorno de la piel que (a) afecta varias partes del cuerpo, especialmente al cuero cabelludo. También se le conoce como caspa, excemas seborreico, soriasis seborr(e)ica y(,) en los bebés(,) se le llama costra láctea. Los síntomas son descamaciones de la piel, manzas cubiertas con escamas blancas, amarillas o costras, en-rojamiento de la piel y picazón. Aún se (d)esconoce las
--------------------------	--

	causas de la dermatiti(s) seborr(e)ica, pero se relaciona con un hongo llamado humana-secia o por una respuesta irregular del sistema inmunitario.
Transcripción de YouTube	(¿) tienes caspa (o) dermatitis seborreica(?) entonces este vídeo es para ti(.) la dermatitis seborr(e)ica es un trastorno de la piel que afecta (a) varias partes del cuerpo(.) especialmente al cuero cabelludo(.) también se le conoce como caspa(.) eccemas de borreico (.) psoriasis seborr(e)ica y(.) en los bebés(.) se le llama costra láctea(.) los síntomas son descamaciones de la piel(.) manchas cubiertas con escamas blancas(.) amarillas o costras(.) enrojecimiento de la piel y picazón(.) Aún se desconocen las causas de la dermatitis seborreica pero se relaciona con un hongo llamado manaseia o por una respuesta irregular del sistema inmunitario(.)
Corrección	¿Tienes caspa o dermatitis seborreica? Entonces este vídeo es para ti. La dermatitis seborreica es un trastorno de la piel que afecta a varias partes del cuerpo, especialmente al cuero cabelludo. También se le conoce como caspa, eccema seborreico, psoriasis seborreica y, en los bebés, se le llama costra láctea. Los síntomas son descamaciones de la piel, manchas cubiertas con escamas blancas, amarillas o costras, enrojecimiento de la piel y picazón. Aún se desconocen las causas de la dermatitis seborreica, pero se relaciona con un hongo llamado “Malassezia” o por una respuesta irregular del sistema inmunitario.

Tabla 2. ES_formal_cientifico_2.m4a_result

Transcripción de Whisper	(—)¿Eres la volición insta de la exportación reproductiva? (—)Sí, sí. (—)¿Sí? Y de hecho, esta legislatura hemos conseguido que, que , claramente, reconocida como una forma de violencia contra las mujeres, que hasta ahora no ocurría. Lo que queríamos haber introducido, además de esto, es que, en España , esos contratos son nulos de pleno derecho, porque no es una práctica legal. (—)Y(.) de hecho, está penado. (—)Pero si tú haces el contrato en otro país, entonces no es perseguible. Entonces, para nosotres, podría haber sido una buena forma de contribuir a eliminar esta práctica(.) para perseguir los contratos también a través del mundo .
Transcripción de YouTube	(—)¿ esa abolicionista de la explotación reproductiva?(—)Sí(.) sí.(—)(¿)sí?(—)Y(.) de hecho(.) esta legislatura hemos conseguido que quede claramente reconocida como una forma de violencia contra las mujeres(.) que hasta ahora no ocurría(.) lo que queríamos haber introducido(.) además de esto(.) es que en España esos contratos son nulos de pleno derecho(.) porque no es una práctica legal.(—)Y(.) de hecho(.) está penado.(—)Pero si tú haces el contrato en otro país(.) entonces no es perseguible(.) entonces(.) para nosotras(.) podría haber sido una buena forma de contribuir a eliminar esta práctica(.) perseguir los contratos también a través del derecho internacional(.)
Corrección	—¿Eres abolicionista de la explotación reproductiva? —Sí, sí. —¿Sí? —Y, de hecho, esta legislatura hemos conseguido que quede claramente reconocida como una forma de violencia contra las mujeres, que hasta ahora no ocurría. Lo que queríamos haber introducido, además de esto, es que en España esos contratos son nulos de pleno derecho, porque no es una práctica legal. —Bueno y, de hecho, está penado. —Pero si tú haces el contrato en otro país, entonces no es perseguible. Entonces, para nosotras podría haber sido una buena forma de contribuir a eliminar esta práctica, perseguir los contratos también a través del derecho internacional.

Tabla 3. ES_formal_varios_interlocutores.m4a_result

Transcripción de Whisper	Bueno, como saben, la gestación subrogada es una práctica que no es legal en España, la nueva ley de derechos sexuales y derechos reproductivos, lo reconoce como una forma de violencia contra las mujeres. Y les pido(.) por favor(.) que nunca se olviden ustedes(.) los medios de comunicación, que son quienes tienen la oportunidad de mostrarles esta realidad al conjunto de la ciudadanía(.) de las mujeres que están detrás, que no se olviden de las mujeres que están detrás. Hay un sesgo de discriminación por pobreza, claro , hace poco veíamos un reportaje en el que un señor(.) que llevaba una agencia de gestación sobre la gente.
Transcripción de YouTube	bueno como saben la gestación subrogada es una práctica que no Es legal en España(.) la nueva ley de derechos sexuales y derechos reproductivos lo reconoce como una forma de violencia contra las mujeres y les pido(.) por favor(.) que nunca se olviden ustedes(.) los medios de comunicación que son quienes tienen la oportunidad de mostrarles esta realidad Al conjunto de la ciudadanía(.) de las mujeres que están detrás(.) que no se olviden de las mujeres que están detrás(.) Hay un sesgo de discriminación por pobreza claro(.) hace poco veíamos un reportaje en el que un señor que llevaba una agencia de gestación subrogada en
Corrección	Bueno, como saben, la gestación subrogada es una práctica que no es legal en España, la nueva ley de derechos sexuales y derechos reproductivos lo reconoce como una forma de violencia contra las mujeres y les pido, por favor, que nunca se olviden ustedes, los medios de comunicación, que son quienes tienen la oportunidad de mostrarles esta realidad al conjunto de la ciudadanía, de las mujeres que están detrás, que no se olviden de las mujeres que están detrás. Hay un sesgo de discriminación por pobreza claro, hace poco veíamos un reportaje en el que un señor, que llevaba una agencia de gestación subrogada

Tabla 4. ES_formal.m4a_result

◆ Inglés

Transcripción de Whisper	(-)(")What are you convinced people are pretending to enjoy("?) sparkling water? (-)Facts(!) (-)Delete it(!) (-)that tastes like stout it if I had it tastes disgusting.(.)(")Pretend(ing) to enjoy their friend(s') success(.)(-)wow.(.) (-"Wow, fam) I'm gas(sed) you're doing so well for yourself(...")(-)I'm wrong . (-)This one is for me facts- Aki(.), bro. (-)Facts(.) (-)Aki socks (!) I'm sorry guys. I'm carrying you to(...)-aki-socks , bro(!). I hate it(!)(")boil(ed) potatoes There's a U
Transcripción de YouTube	(-)(")what are you convinced people are pretending to enjoy("?) sparkling water.(.) (-)fat(!)(-)that tastes like stuff(.) in fact (Fam,) it tastes (disgusting.)(")pretending to enjoy their friend's success.(.)(-)wow.) (-)wow(.), Fam(.), I'm gassed you're doing so well for yourself(...)(-)um bro.(.)(-)this one is for me facts(:) Aki(.), bro(.) box (-)Akee) sucks(!) (I'm) sorry guys(.) I'm Caribbean too(...) Aki sucks(.), bro(!) I hate it(!) (")boiled potatoes(") four potatoes (tastes like) pennies
Corrección	—"What are you convinced people are pretending to enjoy?" Sparkling water. —Facts! —Delete it! —That tastes like static. —Fam, it tastes disgusting. [...] "Pretending to enjoy their friends' success."

	<p>—Wow.</p> <p>—"Wow, fam, I'm gassed. You're doing so well for yourself..."</p> <p>—Damn bro.</p> <p>—This one is for me facts: Akee, bro.</p> <p>—Facts. [laughter]</p> <p>—Akee sucks! I'm sorry guys. I'm Caribbean too... Akee sucks, bro! I hate it! [...] "Boiled potatoes". Boiled potatoes taste like pennies.</p>
--	--

Tabla 5. EN_coloquial_varios_interlocutores.m4a_result

Transcripción de Whisper	Here are five signs you might have small intestine bacteria low growth . So(.) number one, and this is probably one of the most common symptoms of SIBO, but it's that chronic bloating and stomach distension. Now the reason why this happens is because that bacteria in your gut loves carbohydrates. When you have an overgrowth in the small intestine, anytime you consume a meal, they're essentially creating fermentation from those carbohydrates, producing gas in the stomach, which is causing a lot of that discomfort and distension. So(.) number two, and this is also a really common symptom associated with SIBO, but it's chronic diarrhea.
Transcripción de YouTube	here are five signs you might have small intestine bacterial overgrowth(.) so(.) number one(.) and this is probably one of the most common symptoms of sibo (.) but it's that chronic bloating and stomach distension(.) now the reason why this happens is because that bacteria in your gut loves carbohydrates(.) when you have an overgrowth in the small intestine(.) anytime you consume a meal they're essentially creating fermentation from those carbohydrates producing gas in the stomach(.) which is causing a lot of that discomfort and distension(.) so(.) number two(, and) this is also a really common symptom associated with sibo (,) but it's chronic diarrhea
Corrección	Here are five signs you might have Small Intestine Bacterial Overgrowth. So, number one, and this is probably one of the most common symptoms of SIBO, but it's that chronic bloating and stomach distension. Now the reason why this happens is because that bacteria in your gut loves carbohydrates. When you have an overgrowth in the small intestine, anytime you consume a meal, they're essentially creating fermentation from those carbohydrates producing gas in the stomach, which is causing a lot of that discomfort and distension. So, number two, and this is also a really common symptom associated with SIBO, but it's chronic diarrhea.

Tabla 6. EN_formal_cientifico_2.m4a_result

Transcripción de Whisper	Seen the Neural Anatomy and Function Lecture, we covered how the nervous system is composed of two major cell types. We have the neurons and glial cells, or glia, or neuroglia. The difference between the two is that glia don't have a direct role in information processing, but they do have several functions that are important for the healthy functioning of the nervous system. Glial cells are located within the central nervous system, which includes the brain and spinal cord, and the peripheral nervous system, which includes the nerves that are in the brain.
Transcripción de YouTube	you've seen the neuron anatomy and function lecture we call it how the nervous system is composed of two major cell types(:) we have the neurons and glial cells(.) or glia(.) or neuroglia(.) the difference between the two is that glia don't have a direct role in information processing(.) but they do have several functions that are important for the healthy functioning of the nervous system(.) real cells are located

	within the central nervous system(,) which includes the brain and spinal cord(,) and the peripheral nervous system(,) which includes the nerves that connect the brain
Corrección	Seen the Neural Anatomy and Function lecture, we covered how the nervous system is composed of two major cell types: we have the neurons and glial cells, or glia, or neuroglia. The difference between the two is that glia don't have a direct role in information processing, but they do have several functions that are important for the healthy functioning of the nervous system. Glial cells are located within the central nervous system, which includes the brain and spinal cord, and the peripheral nervous system, which includes the nerves that connect the brain.

Tabla 7. EN_formal_cientifico.m4a_result

Transcripción de Whisper	(-)What does Marcellus Wallace look like? (-)What? (-)What? (-)What ain't no country I ever heard of? (.)They speak English and what? (-)What? (-)English motherfucker, do you speak it? (-)Yes, (-)and you know what I'm saying. (-Yes.) (-)Describe what Marcellus Wallace looks like. (-)What? (-)Say what again. Say what again. I dare you. I double dare you(,) motherfucker. Say what? One more goddamn times.
Transcripción de YouTube	(-)what does Marellus Wallace look like(.) (-What?) (-)what country you broke?)(-)what(?) what English [_] do you speak it?)(-)yes(.)(-)then you know what I'm saying(.) describe what Marcellus Wallace looks like (-)what it's like (-Say) what again(.) say what again(.) I dare you(.) I double dare you(.) [_] (.) say what one more goddamn time(.)
Corrección	—What does Marsellus Wallace look like? —What? —What country you from? —What? —What ain't no country I ever heard of. They speak English in what? —What? —English, motherfucker, do you speak it? —Yes. —Then you know what I'm saying. —Yes. —Describe what Marsellus Wallace looks like. —What? I- —Say what again. Say what again. I dare you. I double dare you, motherfucker. Say what One more goddamn time.

Tabla 8. EN_informal_vulgar_2.m4a_result

Transcripción de Whisper	Leonardo DiCaprio had a problem saying the word nigga. (")Pow, pal, it's just, it's tough for me to say this. (") And(, uh,) I remember saying, you know, Jackson, (going:) I'm gonna get over him, motherfucker. This is another Tuesday, motherfucker. I'm gonna talk about these motherfuckers. Look at what you, look at who I am, motherfucker. (Fuck.) And it was like, (...) you know,(?) and I said, (:") yeah,(") I said, (:")Leo, we're not friends.(") I said, (:") this is just another Thursday. This is your property. These are humans. These are your property. He looks, excuse me, more.
Transcripción de YouTube	Leonardo DiCaprio had a problem saying the word [_] (") pow(,) pow(,) is just(,) it's tough for me to say this(.) and(,) uh I remember Samuel Jackson going(:) get over it(,) I it's just another Tuesday(,) [_] (.) I don't give a [_] about these

	[____] (.) look what you (,) look at who I am (,) [____] (.) [____] (.) and it was like(...) you know(?) and I said(:") yeah(".) I said(:") Leo we're not friends(.) I said(:") this is just another Thursday(.) you this is your property(.) these are humans(.) these are (your) property(.) ("—")you excuse me(.)
Corrección	Leonardo DiCaprio had a problem saying the word nigga. "Pal, pal, it's just, it's tough for me to say this." And, uh, I remember Samuel Jackson going: "Get over it, motherfucker. This is just another Tuesday, motherfucker. I don't give a fuck about these motherfuckers. Look at- Look at who I am, motherfucker. Fuck." And it was like... you know? And I said: "Yeah." I said, "Leo, Leo, we're not friends." I said: "This is just another Thursday. This is your property. These aren't humans. These are your property." "—Will you excuse me one moment?"

Tabla 9. EN_informal_vulgar.m4a_result

Transcripción de Whisper	This is huge. Moderna has begun its first clinical trials of an HIV vaccine. It actually uses the same mRNA technology as Moderna's COVID-19 vaccine. Our researchers have been trying to make an HIV vaccine for decades. And if this works, it could help up to 1.2 million Americans who currently have the virus.
Transcripción de YouTube	this is huge(.) moderna has begun its first clinical trials of an HIV vaccine(.) it actually uses the same mRNA technology as modernist covid-19 vaccine(.) our researchers have been trying to make an HIV vaccine for decades(.) and if this works(.) it could help up to 1.2 million Americans who currently have the virus(.)
Corrección	This is huge. Moderna has begun its first clinical trials of an HIV vaccine. It actually uses the same mRNA technology as Moderna's COVID-19 vaccine. Our researchers have been trying to make an HIV vaccine for decades. And if this works, it could help up to 1.2 million Americans who currently have the virus.

Tabla 10. EN_informal.m4a_result

◆ *Francés*

Transcripción de Whisper	à qui on a donné le nom de cellules gliales. Ce qui est très intéressant à ce sujet, c'est que dans le cadre du principe d'une neurone , la théorie d'une neurone , on croyez que les neurones étaient les cellules fonctionnelles du cerveau. Et les cellules gliales étaient là(.) plus ou moins(.) juste en soutien. Et bien(.) par définitions(.) cela veut juste dire que vous (n') utilisez que 10-%. de votre cerveau(.) parce que c'est tout ce que les neurones représentent dans le cerveau-10-%. Pendant longtemps, les cellules gliales(.) encore une fois(.) n'étaient considérées que comme un tissu conjonctif qui apportait un soutien physique et nuage-
Transcripción de YouTube	à qui on a donné le nom de cellules gliales(.) ce qui est très intéressant à ce sujet(.) c'est que dans le cadre du principe du neurone(.) la théorie du neurone(.) on croyait que les neurones étaient les cellules fonctionnelles du cerveau et les cellules gliales étaient là(.) plus ou moins(.) juste en soutien(.) et bien(.) par définition(.) cela veut juste dire que vous n'utilisez que 10 % de votre cerveau(.) parce que(.) c'est tout ce que les neurones représentent dans le cerveau (:) 10-% pendant longtemps les cellules gliales(.) encore une fois(.) n'étaient considérées que comme un tissu conjonctif qui apportait un soutien physique et nutritionnel

Corrección	[...] à qui on a donné le nom de cellules gliales. Ce qui est très intéressant à ce sujet, c'est que dans le cadre du principe du neurone, la théorie du neurone, on croyait que les neurones étaient les cellules fonctionnelles du cerveau et les cellules gliales étaient là, plus ou moins juste en soutien. Et bien, par définition, cela veut juste dire que vous n'utilisez que 10 % de votre cerveau, parce que c'est tout ce que les neurones représentent dans le cerveau : 10 %. Pendant longtemps, les cellules gliales, encore une fois, n'étaient considérées que comme un tissu conjonctif qui apportait un soutien physique et nutritionnel [...]
------------	---

Tabla 11. FR_formal_cientifico.m4a_result

Transcripción de Whisper	2 types de gens, (t'as) les gens qui te disent « eh », (et) les gens qui disent « mais ». (T'as un mec) Tu vas aller (lui) dire (:) « voilà, on va faire (... On va faire) un vélo de dingue ». (« Et ouais !) on va faire ça ».) alors tu as deux cas, tu as le mec qui dit « (Ah) ouais mais (...Mais) si on fait ça, ça raconte des chers, mais si on met ça, on va le voler ». Ça c'est les « mais ». Et puis, (t'as un mec), tu le racontes la même phrase, il dit « ouais, on va mettre ça et ça, il va pouvoir faire ça et on va rajouter ça ». Tarnat (en as) 2. Alors ceux qui font des « eh », c'est les entrepreneurs. Et ceux qui font des « mais », c'est (les) gestionnaire(s).
Transcripción de YouTube	deux types de gens(.) tu as les gens qui te disent (« hey ») (et) les gens qui disent (« mais »)(.) t'as un mec(.) tu vas lui dire (:) « voilà on va faire (...) on va faire un vélo de dingue (», «) et ouais (!) on va faire ça(.) alors tu as deux cas(.) tu as le mec qui dit (:) (« ah ouais mais si on fait ça(.) ça va (me) coûter cher(...) mais si on met ça(.) on va le voler (».) ça c'est les mets(.) et puis(.) tu as le mec(.) tu lui racontes la même phrase(.) il dit (:) « ouais mais on va mettre ça et ça(.) il va pouvoir faire ça et on va rajouter ça (».) tu en as deux (.) alors ceux qui font des haies (,) ees (les) entrepreneurs mais ceux qui font des messes-et les gestionnaires(.)
Corrección	[...] 2 types de gens, t'as les gens qui te disent « et » et les gens qui disent « mais ». T'as un mec, tu vas lui dire : « Voilà, on va faire... On va faire un vélo de dingue », « Et ouais ! On va faire ça ». Alors t'as 2 cas, t'as le mec qui dit : « Ah ouais mais... Mais si on fait ça, ça va me coûter cher... Mais si on met ça, on va le voler ». Ça c'est les « mais ». Et puis, t'as un mec, tu lui racontes la même phrase, il dit « Ouais, on va mettre ça et ça, il va pouvoir faire ça et on va rajouter ça ». Tu en as 2. Alors ceux qui feront des « et », c'est les entrepreneurs et ceux qui font des « mais », c'est les gestionnaires.

Tabla 12. FR_informal_coloquial.m4a_result

Transcripción de Whisper	Hé ! Tu veux que je suis-face-à un café ? Je t'en-mètre, Pierre, tu veux un café, tu vas le faire toi-même, hein ? Je ne-veux pas le temps de passer l'aspect aujourd'hui, hein ? Je vais déjà-n'aider-que Mayenne. Qu'est-ce que tu as dit là ? Non, non, mais rappelle-ce que tu as dit là ? Est-ce que moi, je peux arranger le salon ? Ah bah(.) écoute Pierre, si toi tu veux arranger le salon, si tu veux passer l'aspect, tu peux le faire, hein ? Tu le fais contre-chante du bureau. Voilà, comme ça, j'ai-y-à-régler. Oui Pierre, je sais que tu t'affes et que moi, je ne t'affes pas, oui, je comprends ça. («) Et un-juste-«), mais n'importe quoi, mec !
Transcripción de YouTube	Sin resultados

Corrección	Hé ! Tu veux que je fasse un café ? Je t'emmérde, Pierre, tu veux un café, tu vas le faire toi-même, hein ? J'ai pas le temps de passer l'aspi aujourd'hui, hein ? J'vais déjeuner avec Marianne. Qu'est-ce que t'as dit là ? Non, non, mais répète ce que t'as dit là ! « Est-ce que moi, je peux ranger le salon ? » Ah bah, écoute Pierre, si toi tu veux ranger le salon, si tu veux passer l'aspi, tu peux le faire, hein ? Tu le fais quand tu rentres du bureau. Voilà, comme ça, c'est réglé. Oui, Pierre, je sais que tu taffes et que moi, je ne taffe pas, oui, je comprends ça. « C'est injuste » Mais n'importe quoi, mec !
------------	--

Tabla 13. FR_informal_vulgar.m4a_result

◆ *Catalán y Valenciano*

Transcripción de Whisper	Ben dia no hi ha gent, ja sóc a Ginebra, que és la meva habitació de hotel. Jo he decidit que Ginebra se l'anà bastida molt clàssic(a), perquè és una ciutat així com (molt) a l'Europa i a(,) molt dalt . Llavors jo he postat per(:) a Samarretta, a base i qui, si m'ha de col·laborar del l'única, els meus pantalons negres, preciosos, xulíssims, cap tan un col·lazo, amb els meus moquesins vegades que m'encanten, i porto aquest mocador per donar-li una mica d'aquell , perquè sigui un punt més emocional , perquè si no sigues veritat que hi ha gent .
Transcripción de YouTube	Sin resultados
Corrección	Bon dia, nois i noies, soc a Ginebra, aquesta és la meva habitació d'hotel. Jo he decidit que a Ginebra s'hi ha d'anar vestida molt clàssica, perquè és una ciutat així com molt europea, molt tal. Llavors jo he apostat per: samarreta basiquíssima de color blanc, de l'Uniqlo, els meus pantalons negres, preciosos, <i>xulíssims</i> , que em fan un <i>culazo</i> amb els meus mocassins vegades que m'encanten, i, em, porto aquest mocador per donar-li una mica de què, perquè sigui un punt més emocionant, perquè si no si que és veritat que anar només

Tabla 14. CA_coloquial.m4a_result

Transcripción de Whisper	(—) i els prepara a l'Uni. Bonat, de nou. En ningú s'havia escapa(, a) aquestes alçades, la gran polèmica generada a Espanya, sobre el tema del delictes de sedició, (i) la reforma del codi penal. Aquest passat dijós , el Congrés va votar a favor de la presa en consideració de la reforma del codi penal proposada per peixó i podem d'acord amb esquerra publicana de Catalunya. Però atenció. - La presa en consideració.
Transcripción de YouTube	Sin resultados
Corrección	[...] —Partit Junts per Catalunya. —Bona tarda, de nou. A ningú (no) se li escapa, a aquestes alçades, la gran polèmica generada a Espanya, sobre el tema del delictes de sedició i la reforma del codi penal. Aquest passat dijous, el Congrés va votar a favor de la presa en consideració de la

	reforma del codi penal, proposada per PSOE i Podemos d'acord amb Esquerra Republicana de Catalunya. Però atenció, la presa en consideració.
--	---

Tabla 15. CA_formal_2.m4a_result

Transcripció de Whisper	no han comut , ha vingut aquí, en aquest pla , sense fer cap pacte amb cap formació política, sense tenir certes escoles com acabaria aquesta investidura, però amb una única certesa, que és que nosaltres anem directes a l'oposició. Per primer cop, en aquest saludable ens hi han aquest ajuntament, hi ha representats de box . Tenen dret a estar aquí, això no ho pot questionar ningú, perquè els han votat una part de la ciutadania. Però jo sí que els vull deixar molt clar que nosaltres no pensem normalitzar la seva presència i encara menys el seu discurs.
Transcripció de YouTube	Sin resultados
Corrección	Barcelona en Comú ha vingut aquí, en aquest ple, sense fer cap pacte amb cap formació política, sense tenir certeses (de) com acabaria aquesta investidura, però amb una única certesa, que és que nosaltres anem directes a l'oposició. Per primer cop, en aquest Saló de Plens, i en aquest Ajuntament, hi ha representats de Vox. Tenen dret a estar aquí, això no ho pot qüestionar ningú, perquè els han votat una part de la ciutadania. Però jo sí que els hi vull deixar molt clar que nosaltres no pensem normalitzar la seva presència i encara menys el seu discurs.

Tabla 16. CA_formal.m4a_result

Transcripció de Whisper	(—)Et pots parlar de la pèl·la? (—)Ja ets que vulguis. (—)No m'he espas sentit a parlar amb l'Ancia . (—)Però no hi ha pèl·la, (de “No,) el catalán no és (lo mismo que) el que valencià. (—)Ja haurà pèl·les. Estan existint i a l'espera. (—)Els valencià, (i) el catalán, són cosin(s) germans. Estan destinats a d'or saber i a portar saber com ho va dir. (—)Què tal (va), per aquí(,) per Barcelona, t'agrada la ciutat? (—)M'encanta. (—)La gent que em permudar-te algun dia? (?) (—)Jo soc valenciana i de la terretre ara mateix a ningú. (—Ets molts guapa, eh!) (—)Tu, m'heu sentit com a emocionar-me(, no em diguis això... Avui...). (—Tens cara,) No sé(,) com de modelos.
Transcripció de YouTube	Sin resultados
Corrección	— Hola. — Et puc fer una petita pregunta? — Sí (és) clar, totes les que vulgues. — Jo és que només per sentir-te parlar en valencià ja m'emociono. — ¡Ah, claro! Pero por favor que no hayan peleas, eh, de “No, el catalán no es lo mismo que el valenciano. — Hi haurà peleas. Estan existint, ja, les peleas. — El valenciano y el catalán, són cosins germans i estan destinats a d'ur-se bé i a portar-se bé, como se diga. — Què tal va, per aquí, per Barcelona, t'agrada la ciutat? — M'encanta. — Lo suficient com per mudar-te algun dia? — Jo soc valenciana i de la terreta no me meneja ningú.

	<p>—Ets molt guapa, eh!</p> <p>—Tu més! Em, puc emocionar-me, no em diguis això...Avui...</p> <p>—Tens cara, no sé, com de <i>modelo</i>, saps?</p>
--	---

Tabla 17. CA_VA_varios_interlocutores_coloquial.m4a_result

Transcripción de Whisper	<p>(—Jo a la Juliana Canet) No li puc dir que no. (—)</p> <p>Què tal, Carlos, com estàs?</p> <p>(—)Molt bé, Julia, engane. (—)</p> <p>!) és la tiva cita ideal. (?)</p> <p>(—)La tiva cita ideal és aquella que vols tornar a tenir una segona molt ràpid, (no?) que acabes i dius: (: "Ens veiem demà?) Com ho tens?(-)No, és que no podia, però...</p> <p>(—)Molt bona resposta. La mi, que tinc d'atant múnica, (en) tant(s) horaris diferents i tal de diferents, hi ha cites meravellos(es) per esmorzar, que se m'han de les mans i acabes de no ser on(.) i ha cites destupentes al vespre. Jo sóc bastant urbà. No m'imagino massa, n'ha citat el camp(,) la primera. Tu hi ha saps que jo soc una gran organitzadora de cites, eh?</p>
Transcripción de YouTube	Sin resultados
Corrección	<p>—Jo a la Juliana Canet no li puc dir que no.</p> <p>—Què tal, Carlos, com estàs?</p> <p>—Molt bé, Juliana Canet, i tu?</p> <p>—Quina és la teva cita ideal?</p> <p>—La meva cita ideal és aquella que vols tornar a tenir una segona molt ràpid, no? Que acabes i dius: "Ens veiem demà? Com ho tens? No, és que no podia, però..."</p> <p>—Molt bona resposta.</p> <p>—N'he i amics n'han tingut, de tan boniques, en tants horaris diferents i tants llocs diferents! Hi ha cites meravelloses per esmorzar, que se te'n van de les mans i acabes a no ser on... Hi ha cites estupendes al vespre. Jo soc bastant urbà, no m'imagino massa una citat el camp, la primera.</p> <p>—Tu ja saps que jo soc una gran organitzadora de cites, eh!</p>

Tabla 18. CA_varios_interlocutores_coloquial.m4a_result

Transcripción de Whisper	<p>(—)Home, es que fan pintada no menjar conya(, vale?). Home, enjar que el només si demana si us plau de ja no és que dalt menjar. Primer de tot(-)ja no et pique. Ja no et pique és una persona (que) gràmosa, et menja el conya si t'estima. I... (mmm... mmm...) Oh... ("?) (—)És que no ho sé, perquè fa pintada guarríssim, però no menja(, no menja). (—)No, no menja. Tot el resto(ho) fa perfecte, però menja(r) (, no menja). El nou dia de la l'olita, el mat i ni a més... No, és que mai de la vida menjar(à) el conya, igual que el Pablo Castellano, no ve (de) la Maria Pomo, és un tipus de la luna.</p>
Transcripción de YouTube	Sin resultados

Corrección	<p>—Homes que fan pinta de no menjar cony, vale? O menjar-te'l només si demanes si us plau, de genolls, que te'l mengin. Primer de tot: Gerard Piqué. Gerard Piqué és una persona que només et menja el cony si t'estima. I... mmm... mmm... O?</p> <p>—És que no ho sé, perquè fa pintada <i>guarrísim</i>, però no menja, no menja.</p> <p>—No, no menja. Tota la resta ho fa perfecte, però menjar, no menja. El <i>nòvio</i> de la Lola Lolita: no menja ni que el matin. No, és que mai de la vida menjarà cony, igual que el Pablo Castellano, <i>nòvio</i> de la Maria Pombo, és un tipus de gent.</p>
------------	---

Tabla 19. CA_vulgar_varios_interlocutores.m4a_result

◆ Japonés

Transcripción de Whisper	<p>こんにちは、今日のデイリ(一)ブロックのボイソーダーは日本語であげてみたいと思います(。思)ってよりたくさん日本語でアップロードしてくださいって(いう)リクエストがたくさんあったんですけど(、)ちょっと時間くかつちゃってすみませんが(、)今日は日本語で頑張っていきます(。)私の生まれ育ちがオ(一)ストリアで小学校から大学生まず全部オ(一)ストリアで帰っているんでやっぱり英語がメインとしてずっとしゃべってきたんですけど家ではやっぱり親が両方日本人なので日本語ずっとしゃべってきたし(、)小学校から中学校</p>
Transcripción de YouTube	<p>こんにちは(、)今日のデイリーブロックのボイスオーバーは日本語であげてみたいと思います(。)思ったよりたくさん日本語でアップロードしてくださいというリクエストがたくさんあったんですけど(、)ちょっと時間かかつちゃってすいませんが(、)今日は日本語で頑張っていきます(。)私の生まれ育ちがオーストラリアで小学校から大学生まで全部オーストラリアで通っているんでやっぱり英語がメインとしてずっと喋ってきたんですけど家ではやっぱり親が両方日本人なので日本語ずっと喋ってきたし(、)小学校から中学校</p>
Corrección	<p>こんにちは、今日のデイリーブログのボイスオーバーは日本語であげてみたいと思います。思ったよりたくさん日本語でアップロードしてくださいというリクエストがたくさんあったんですけど、ちょっと時間かかつちゃって、すみませんが、今日は日本語で頑張っていきます。私の生まれ育ちがオーストラリアで小学校から大学生まで全部オーストラリアで通っているんでやっぱり英語がメインとしてずっとしゃべってきたんですけど家ではやっぱり親が両方日本人なので日本語ずっとしゃべってきたし、小学校から中学校</p>

Tabla 20. JA_informal_vlog.m4a_result

Transcripción de Whisper	<p>ご飯が止まらない(！) お家で簡単、トロトロ、ツヤツヤ、ご飯が進む(、) おー、ノコバタ(一) チュキンカレイ(。) ごちそうさまでした、モモニタ、ヨーグルと入れて、みんな大好き、ハチミツ、カレーコ、おさじ 1、おしよ入れまして、まずまずでございます、下味付けましたそしたら、ニンニクとしょうがを入れて、フライパンを固むけて</p>
--------------------------	---

Transcripción de YouTube	合板が止まるな (一) い (!) おうちで簡単 (、) とろとろ (、) つやつや (、) ご飯が進む、 あの子 バターチキンカレー (。) ごちそうさまでした (、) もも肉 (、) ヨーグルト入れて (、) 皆大好き (、) ハチミツ (、) カレー粉大さじ 1 (、) お塩入れまして (、) まずまず でございます (、) 下味つけました (。) そしたらにんにくと生姜 (が) を入れて (、) フライパンを傾け (て、)
Corrección	ご飯が止まらない！お家で簡単、トロトロ、ツヤツヤ、ご飯が進む、おー、濃厚バターチキンカレー。ごちそうさまでした、もも肉、ヨーグルト入れて、みんな大好き、ハチミツ、カレー粉大さじ 1、お塩入れまして、混ぜ混ぜでございます、下味付けました。そしたら、ニンニクとしょうがを入れて、フライパンを傾けて、

Tabla 21. JA_informal_coloquial_musica.m4a_result

Transcripción Whisper	なぜセンチュリーなのでしょうか (。)(あの、まあ、) 事故が少ないということですね (。)(例えば衝突時の安全性とか (。)(誰が何とも) どういう風に安全性があったりとか (、) いろいろと限られた上でこれが (。)(あの、) 例えば大統領者のような ボ=ダン ガラスというものはございません (。)(テジトシヤー とセンチュリーが来た (ら) 当然センチュリーの (方が) 安全というのは誰が見てもわかることですので (。)
Transcripción de YouTube	なぜセンチュリーなのでしょうか (。)(あの、まあ、) 事故が少ないということとですね (。)(例えば衝突時の安全性とか (。)(具体的には) どういう風 (に) 安全性 だ たりとか 色々 と加 工 された上でこれが (。)(あの、) 例えば大統領者のような防弾ガラスというものはございません (。)(軽自動車とセンチュリー (が) 比較しては) 当然 セン手 の (方が) 安全 って というのは誰が見て (も) わかることですので (。)
Corrección	なぜセンチュリーなのでしょうか。あの、まあ、事故が少ないということですね。例えば衝突時の安全性とか。具体的にはどういう風に安全性があったりとか、いろいろと加味された上でこれが。あの、例えば大統領車のような防弾ガラスというものはございません。軽自動車とセンチュリーが来たら当然センチュリーの方が安全というのは誰が見てもわかることです。

Tabla 22. JA_informal_2_varios_interlocutores.m4a_result

Transcripción de Whisper	やっぱり使いともの方の 2階の人(がない) 、みだいろことを呼んでいるん ですけれども 、そうじゃなくて、皆さん言っているのは記憶ということ でそ うです。そう、使えてくれてますかって言ったら、今、 やったらお小伝伝 という人に伝 い ます。みたいなことを言ったんですけれども、
Transcripción de YouTube	やっぱり近いところの方の理解はいない (、) とみたいなこと (を) 言うんですけ (れ) ども (、) いや そうじゃなくて (、) 皆さん言って (い) るのは記録ということ でしょ って (。)(そう (、) 伝えてくれてますかって言ったら (、) 今 (、) 言ったお言葉通りに使えます (。)(みたいなことを言ったんですけれども (、)

Corrección	やっぱり近いところの方の理解がない、みたいなことを言うんですけども、そうじゃなくて、皆さん言っているのは記録ということですよって。そう、伝えてくれてますかって言ったら、今、言ったお言葉通りに伝えます。みたいなことを言ったんですけども、
------------	---

Tabla 23. JA_formal_interferencias.m4a_result

Transcripción de Whisper	今日 5号 1 時すぎ 東京 8歳を自治 の中学校で 理科の実験中に 複数の生徒が体調不良を訴えているということです。(。 テース 調によりますと 今日 5号 1 時すぎ 8歳を自治の 8歳を自学院 8歳を自治学校 で 理科の実験中に気分が悪くなった生徒が複数人いると 199 番通報がありました。(。 。) 当時中学 2 年の生徒 36 人が 理科の 事業 を受けていて
Transcripción de YouTube	(今日午後 1 時すぎ) 東京八王子市の中学校で理科の実験中に複数の生徒が体調不良を訴えているということです。(。 。) 警視庁によりますと 今日午後 1 時過ぎ八王子市の八王子学園八王子中学校で理科の実験中に気分が悪くなった生徒が複数人いると 119 番通報がありました。(。 。) 当時中学 2 年の生徒 36 人が理科の授業を受けていて
Corrección	今日午後 1 時すぎ 東京八王子市の中学校で 理科の実験中に 複数の生徒が体調不良を訴えているということです。警視庁によりますと 今日午後 1 時すぎ 八王子市の八王子学園八王子中学校で 理科の実験中に気分が悪くなった生徒が複数人いると 199 番通報がありました。当時中学 2 年の生徒 36 人が 理科の授業を受けていて

Tabla 24. JA_formal_2.m4a_result

Transcripción de Whisper	警察と 証報 によりますと 25 日午後 3 時半ごろ、 4回地市区場長の改造側 で近くに住む 13 歳の男子中学生が、 動給生 と小学 6 年の妹を (。 。) 合わせて 3 人で、川に入って遊んでいたところ、妹が 覚 れそうになり、助けに入るため川に入りましたが、姿が見えなくなったということです。19 番通報を受け、発生からおよそ 15 分後にかけつけた 急車隊 が、水心 1.5m の川底に沈んでいた
Transcripción de YouTube	(警察と) 消防によりますと 25 日午後 3 時半頃 (。 。) 四日市市生桑町の 改造 川で近くに住む 13 歳の男子中学生が (。 。) 同級生と小学 6 年の妹 (。 。) 合わせて 3 人で (。 。) 川に入って遊んでいたところ (。 。) 妹が溺れそうになり (。 。) 助けに入るため川に入りましたが (。 。) 姿が見えなくなったということです (。 。) 119 番通報を受け (。 。) 発生からおよそ 15 分後に駆けつけた救助隊が (。 。) 推進 1.5m の川底に沈んでいた
Corrección	警察と消防によりますと 25 日午後 3 時半ごろ、四日市市生桑町の手蔵川で近くに住む 13 歳の男子中学生が、同級生と小学 6 年の妹、合わせて 3 人で、川に入って遊んでいたところ、妹が溺れそうになり、助けに入るため川に入りましたが、姿が見えなくなったということです。19 番通報を受け、発生からおよそ 15 分後にかけつけた救助隊が、水心 1.5m の川底に沈んでいた

1.3 Resultados de las transcripciones de Whisper y YouTube: análisis y comparación

El esfuerzo empleado en la generación de subtítulos generados automáticamente por reconocimiento de voz de Whisper en el terminal especificado anteriormente es de 10 minutos. Esto es teniendo en modelo de lengua y el código de Python ya creados y listos para ser utilizados. De no ser así, conllevaría horas de investigación y pruebas para un usuario con ningún conocimiento o, en su defecto, con conocimientos muy básicos de Python. Muy posiblemente el rendimiento de esta herramienta sería mucho más rápido y eficiente en un terminal con mayor capacidad de procesamiento.

Por otro lado, el esfuerzo empleado en la generación de subtítulos automáticos de YouTube con la cantidad de vídeos analizados es de 1 hora y media / 2 horas, por los motivos expuestos anteriormente en el apartado 1.2 *Metodología del marco práctico*. Es importante mencionar que con los subtítulos generados automáticamente no se ha podido obtener las transcripciones de la lengua catalana puesto que no está disponible. El vídeo de referencia “FR_informal_vulgar.mp4”, YouTube no ha proporcionado la transcripción generada automáticamente.

1.3.1 Análisis y comparación de la puntuación de las transcripciones de Whisper y YouTube

En primer lugar, veamos el análisis detallado sobre la puntuación en cada lengua, comparando los tres tipos de transcripción. Las siguientes figuras están extraídas del documento Excel “Formula_recuento_puntuacion”, ubicado en la carpeta “DOCUMENTOS”. Las últimas tres columnas, en color verde, solo aplican al japonés, ya que tienen grafías diferentes para los signos de puntuación.

Idioma	Muestra	Transcripción	Texto	,	.	:	;	¿	?	¡	—	-	,	o	!
CA	CA_coloquia	Whisper	Ben dia no hi ha gent, ja se	13	3	0	0	0	0	0	0	0	1		
CA		YouTube		0	0	0	0	0	0	0	0	0	0		
CA	Corrección	Humana	Bon dia, nois i noies, soc a	15	2	1	0	0	0	0	0	0	1		
CA	CA_formal_2	Whisper	i els prepara a l'Uni. Bonat	5	6	0	0	0	0	0	0	0	0		
CA		YouTube		0	0	0	0	0	0	0	0	0	0		
CA	Corrección	Humana	— Partit Junts per Catalunya	7	5	0	0	0	0	0	0	2	0		
CA	CA_formal.n	Whisper	no han comut, ha vingut a	10	4	0	0	0	0	0	0	0	0		
CA		YouTube		0	0	0	0	0	0	0	0	0	0		
CA	Corrección	Humana	Barcelona en Comú ha ving	10	4	0	0	0	0	0	0	0	0		
CA	CA_VA_vari	Whisper	Et pots parlar de la pèl·la?	6	12	0	0	0	2	0	0	0	2		
CA		YouTube		0	0	0	0	0	0	0	0	0	0		
CA	Corrección	Humana	— Hola. — Et puc fer una p	18	15	0	0	0	4	1	3	14	5		
CA	CA_varios_in	Whisper	No li puc dir que no. Què t	13	11	0	0	0	3	0	0	0	0		
CA		YouTube		0	0	0	0	0	0	0	0	0	0		
CA	Corrección	Humana	— Jo a la Juliana Canet no l	13	10	1	0	0	6	0	2	8	0		
CA	CA_vulgar_v	Whisper	Home, es que fan pintada	12	17	0	0	0	0	0	0	0	0		
CA		YouTube		0	0	0	0	0	0	0	0	0	0		
CA	Corrección	Humana	— Homes que fan pinta de	13	17	2	0	0	2	0	0	3	1		

Figura 9. Análisis de la puntuación en catalán

En primer lugar, se puede observar que YouTube no admite transcripción automática en catalán, por lo que queda descartado del análisis de puntuación. La distancia de edición entre la transcripción de Whisper y la humana, en términos de puntuación, no es demasiada. Vemos que en los discursos formales la transcripción de Whisper ha acertado más en los signos de puntuación.

Para continuar, cuando hay varios interlocutores, Whisper no utiliza la raya (guion largo de diálogo), por lo que en ese aspecto hay mayor diferencia con la transcripción humana. Además, Whisper también es capaz de utilizar signos de interrogación y exclamación.

A pesar de que los resultados del recuento no son los mismos, esto se debe o bien por la rapidez del habla del interlocutor o bien porque Whisper interpreta las pausas en el discurso de forma errónea. Aun y así, utiliza puntos, comas, signos de exclamación y de interrogación. En ninguna de las muestras en este idioma ha conseguido utilizar ni dos puntos, ni puntos suspensivos ni punto y coma.

Idioma	Muestra	Transcripción	Texto	,	.	:	;	¿	?	¡	—	-	,	°	!
EN	EN_coloquial	Whisper	What are you convinced pe	1	6	0	0	0	1	0	0	0	1		
EN	EN_coloquial	YouTube	what are you convinced pec	0	0	0	0	0	0	0	0	0	0		
EN	Corrección	Humana	—"What are you convinced	5	18	1	0	0	1	0	5	11	0		
EN	EN_formal_d	Whisper	Here are five signs you migh	8	5	0	0	0	0	0	0	0	0		
EN	EN_formal_d	YouTube	here are five signs you migh	0	0	0	0	0	0	0	0	0	0		
EN	Corrección	Humana	Here are five signs you migh	9	5	0	0	0	0	0	0	0	0		
EN	EN_formal_d	Whisper	Seen the Neural Anatomy at	7	4	0	0	0	0	0	0	0	0		
EN	EN_formal_d	YouTube	you've seen the neuron anat	0	0	0	0	0	0	0	0	0	0		
EN	Corrección	Humana	Seen the Neural Anatomy at	7	3	1	0	0	0	0	0	0	0		
EN	EN_informal	Whisper	What does Marcellus Walla	2	7	0	0	0	10	0	0	0	0		
EN	EN_informal	YouTube	what does Marcellus Wallac	0	0	0	0	0	0	0	0	0	0		
EN	Corrección	Humana	—What does Marsellus Wal	3	10	0	0	0	8	0	0	13	1		
EN	EN_informal	Whisper	Leonardo DiCaprio had a pr	19	12	0	0	0	0	0	0	0	0		
EN	EN_informal	YouTube	Leonardo DiCaprio had a pr	0	0	0	0	0	0	0	0	0	0		
EN	Corrección	Humana	Leonardo DiCaprio had a pr	11	13	3	0	0	2	0	0	1	1		
EN	EN_informal	Whisper	This is huge. Moderna has b	1	6	0	0	0	0	0	0	0	1		
EN	EN_informal	YouTube	this is huge moderna has be	0	1	0	0	0	0	0	0	0	1		
EN	Corrección	Humana	This is huge. Moderna has b	1	6	0	0	0	0	0	0	0	1		

Figura 10. Análisis de la puntuación en inglés

En primer lugar, se puede observar que YouTube no utiliza ningún signo de puntuación a excepción de un caso: cuando se utilizan cifras que van separadas por puntos.

La distancia de edición entre la transcripción de Whisper y la humana, en términos de puntuación, es mayor cuanto más coloquial es un discurso. No obstante, en la última muestra informal, vemos que es igual a la humana, puesto que la dicción del interlocutor era buena y la claridad del sonido también.

De nuevo, vemos que en los discursos formales la transcripción de Whisper ha acertado más en los signos de puntuación, a pesar de que en algunos casos haya utilizado signos de puntuación de más. Además, cuando hay varios interlocutores, Whisper no utiliza la raya (guion largo de diálogo), por lo que en ese aspecto hay mayor diferencia con la transcripción humana.

A pesar de que los resultados del recuento no son los mismos, esto se debe o bien por la rapidez del habla del interlocutor o bien porque Whisper interpreta las pausas en el discurso de forma errónea. Aun y así, utiliza puntos, comas, signos de exclamación y de interrogación. En ninguna de las muestras en este idioma ha conseguido utilizar ni dos puntos, ni puntos suspensivos ni punto y coma.

Idioma	Muestra	Transcripción	Texto	,	.	:	;	¿	?	¡	—	-	,	.	!
ES	ES_coloquia	Whisper	Vale, voy a deciros de qué	13	9	0	0	2	2	0	0	0	0		
ES	ES_coloquia	YouTube	vale Voy a deciros de que v	0	0	0	0	0	0	0	0	0	0		
ES	Corrección	Humana	—Vale, voy a deciros de qu	13	6	0	1	2	2	0	0	5	0		
ES	ES_formal_c	Whisper	¿Tienes caspador Matiti se	7	5	0	0	1	1	0	0	0	0		
ES	ES_formal_c	YouTube	tienes caspa dermatitis se	0	0	0	0	0	0	0	0	0	0		
ES	Corrección	Humana	¿Tienes caspa o dermatitis	9	5	0	0	1	1	0	0	0	0		
ES	ES_formal_v	Whisper	¿Eres la volición insta de la	15	6	0	0	1	1	0	0	0	0		
ES	ES_formal_v	YouTube	esa abolicionista de la expl	0	0	0	0	0	0	0	0	0	0		
ES	Corrección	Humana	—¿Eres abolicionista de la	6	2	0	0	2	2	0	0	4	0		
ES	ES_formal.m	Whisper	Bueno, como saben, la ges	8	3	0	0	0	0	0	0	0	0		
ES	ES_formal.m	YouTube	bueno como saben la gest	0	0	0	0	0	0	0	0	0	0		
ES	Corrección	Humana	Bueno, como saben, la ges	11	1	0	0	0	0	0	0	0	0		

Figura 11. Análisis de la puntuación en español

De nuevo se puede observar que YouTube no utiliza ningún signo de puntuación. La distancia de edición entre la transcripción de Whisper y la humana, en términos de puntuación, es mayor cuanto más coloquial es un discurso. En la tercera muestra se observa una sobrepuntuación tanto en comas como puntos.

La distancia de edición –en cuanto a la puntuación– sería menor en la mayoría de los casos, puesto que tres de las cuatro muestras son discursos formales. Se puede ver de nuevo que cuando hay varios interlocutores, Whisper no utiliza la raya (guion largo de diálogo), por lo que en ese aspecto hay mayor diferencia con la transcripción humana.

A pesar de que los resultados del recuento no son los mismos, esto se debe o bien por la rapidez del habla del interlocutor o bien porque Whisper interpreta las pausas en el discurso de forma errónea. Aun y así, utiliza puntos, comas, signos de exclamación y de interrogación. En ninguna de las muestras en este idioma ha conseguido utilizar ni dos puntos, ni puntos suspensivos ni punto y coma.

Idioma	Muestra	Transcripción	Texto	,	.	:	;	¿	?	¡	—	-	,	.	!
FR	FR_formal_c	Whisper	à qui on a donné le nom d	5	5	0	0	0	0	0	0	0	0		
FR	FR_formal_c	YouTube	à qui on a donné le nom de	0	0	0	0	0	0	0	0	0	0		
FR	Corrección	Humana	[...] à qui on a donné le nom	10	3	1	0	0	0	0	0	0	0		
FR	FR_informal	Whisper	2 types de gens, les gens d	12	8	0	0	0	0	0	0	0	0		
FR	FR_informal	YouTube	deux types de gens tu as le	0	0	0	0	0	0	0	0	0	0		
FR	Corrección	Humana	[...] 2 types de gens, t'as le	14	7	2	0	0	0	0	1	0	0		
FR	FR_informal	Whisper	Hé ! Tu veux que je suis fa	20	4	0	0	0	7	0	2	0	4		
FR		YouTube		0	0	0	0	0	0	0	0	0	0		
FR	Corrección	Humana	Hé ! Tu veux que je fasse u	21	4	0	0	0	6	0	3	0	3		

Figura 12. Análisis de la puntuación en francés

Al igual que en el resto de las lenguas, YouTube no utiliza ningún signo de puntuación. Sorprendentemente, YouTube no ha sido capaz de transcribir la última muestra en francés.

La distancia de edición entre la transcripción de Whisper y la humana en esta lengua ha sido menor. En términos de puntuación, Whisper ha puntuado muy bien a pesar de que dos de las tres muestras fueran discursos informales.

Se puede ver de nuevo que cuando hay varios interlocutores, Whisper no utiliza la raya (guion largo de diálogo), por lo que en ese aspecto hay mayor diferencia con la transcripción humana.

A pesar de que los resultados del recuento no son los mismos, esto se debe o bien por la rapidez del habla del interlocutor o bien porque Whisper interpreta las pausas en el discurso de forma errónea. Aun y así, utiliza puntos, comas, signos de exclamación y de interrogación. En ninguna de las muestras en este idioma ha conseguido utilizar ni dos puntos, ni puntos suspensivos ni punto y coma.

Idioma	Muestra	Transcripción	Texto	,	.	:	;	¿	?	¡	—	-	,	。	！	
JA	JA_formal_1	Whisper	警察と証報によりますと	0	1	0	0	0	0	0	0	0	0	8	1	0
JA	JA_formal_1	YouTube	消防によりますと25日午	0	1	0	0	0	0	0	0	0	0	0	0	0
JA	Corrección	Humana	警察と消防によりますと	0	1	0	0	0	0	0	0	0	0	9	1	0
JA	JA_formal_2	Whisper	今日5号1時すぎ 東京8歳	0	0	0	0	0	0	0	0	0	0	0	0	0
JA	JA_formal_2	YouTube	東京八王子市の中学校で	0	0	0	0	0	0	0	0	0	0	0	0	0
JA	Corrección	Humana	今日午後1時すぎ 東京八	0	0	0	0	0	0	0	0	0	0	0	2	0
JA	JA_formal_in	Whisper	やっぱり使いともの方	0	0	0	0	0	1	0	0	0	0	7	2	0
JA	JA_formal_in	YouTube	やっぱり近いところの方	0	0	0	0	0	0	0	0	0	0	0	0	0
JA	Corrección	Humana	やっぱり近いところの方	0	0	0	0	0	0	0	0	0	0	7	2	0
JA	JA_informal	Whisper	なぜセンチュリーなの	0	0	0	0	0	0	0	0	0	0	0	0	0
JA	JA_informal	YouTube	なぜセンチュリーなの	0	0	0	0	0	0	0	0	0	0	0	0	0
JA	Corrección	Humana	なぜセンチュリーなの	0	0	0	0	0	0	0	0	0	0	4	6	0
JA	JA_informal	Whisper	ご飯が止まらないお家	0	0	0	0	0	0	0	0	0	0	15	0	0
JA	JA_informal	YouTube	合板が止まるないおうち	0	0	0	0	0	0	0	0	0	0	0	0	0
JA	Corrección	Humana	ご飯が止まらない！お家	0	0	0	0	0	0	0	0	0	0	16	2	1
JA	JA_informal	Whisper	こんにちは、今日のデイ	0	0	0	0	0	0	0	0	0	0	1	0	0
JA	JA_informal	YouTube	こんにちは 今日のデイ	0	0	0	0	0	0	0	0	0	0	0	0	0
JA	Corrección	Humana	こんにちは、今日のデイ	0	0	0	0	0	0	0	0	0	0	5	2	0

Figura 13. Análisis de la puntuación en japonés

YouTube en estas muestras ha utilizado el punto en el único caso en que había una cifra separada por este signo ortográfico. Este tipo de punto: “.” es el que se usa en japonés para las cifras. Por el contrario, en las pausas en un discurso se usa este otro tipo de punto propio del japonés: “。”. Lo mismo sucede con las comas, en japonés se utiliza este tipo de coma: “、” . Asimismo, otros signos de puntuación también son diferentes en este idioma. Por ejemplo: signo de interrogación “?” vs “?”, signo de exclamación “!” vs “!”, paréntesis “()” vs “()”, comillas simples “ ’ ’ ” vs “ 「 」 ” y,

por último, comillas dobles “« »” vs “『 』 ”. Puede parecer que tanto el signo de interrogación como el de exclamación sean el mismo, pero si se observa bien, hay un espacio antes o después que está por defecto. Visualmente, parece que sobren espacios, pero es una separación que no se puede eliminar, sino que la grafía está diseñada así. Esto se debe a que el japonés es un idioma cuya escritura no usa espacios entre palabras. Por otro lado, es importante remarcar que en la escritura formal no se utilizan ni signos de exclamación ni de interrogación. En la escritura formal, la interrogación se marca con la partícula *Ka* a final de la oración interrogativa y se le añade un punto final: “か。”. Vemos pues, que en la tercera muestra Whisper ha añadido erróneamente un interrogante de estilo occidental.

La distancia de edición entre la transcripción de Whisper y la humana en esta lengua ha sido mayor que en otras lenguas.

A pesar de que los resultados del recuento no son los mismos, esto se debe o bien por la rapidez del habla del interlocutor o bien porque Whisper interpreta las pausas en el discurso de forma errónea. Aun y así, utiliza puntos, comas, signos de exclamación y de interrogación. En ninguna de las muestras en este idioma ha conseguido utilizar ni dos puntos, ni puntos suspensivos ni punto y coma.

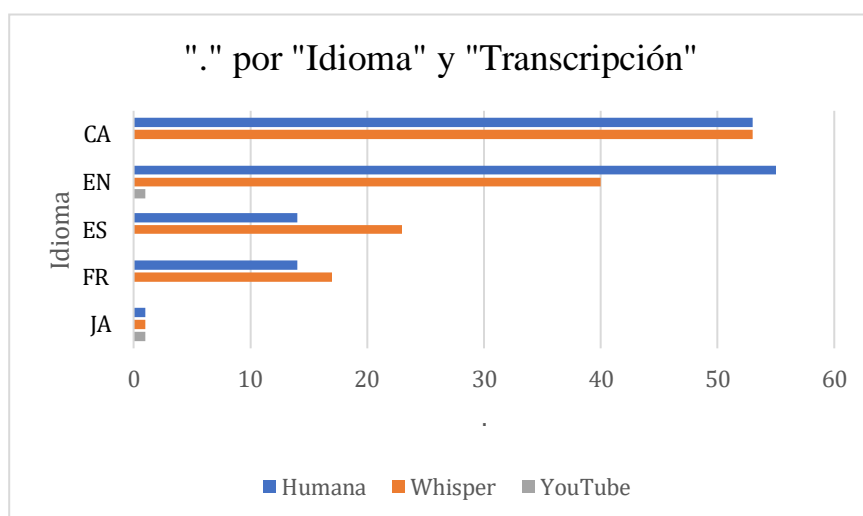


Gráfico 1. Análisis de puntos en todas las lenguas

Suma de .	Transcripción			Total general
Idioma	Humana	Whisper	YouTube	
CA	53	53	0	106
EN	55	40	1	96
ES	14	23	0	37
FR	14	17	0	31
JA	1	1	1	3
Total general	137	134	2	273

Tabla 26. Análisis de puntos en todas las lenguas

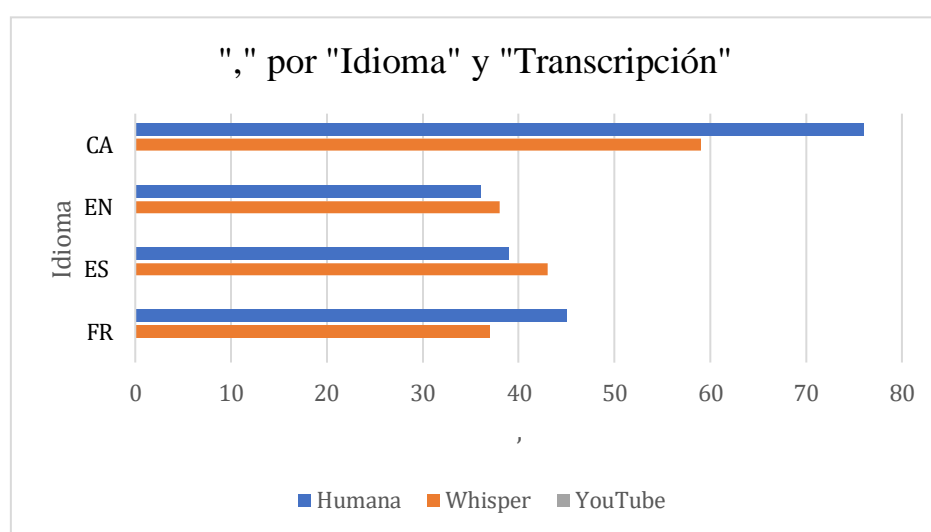


Gráfico 2. Análisis de comas en todas las lenguas

Suma de ,	Transcripción				
Idioma	Humana	Whisper	YouTube	Total general	
CA		76	59	0	135
EN		36	38	0	74
ES		39	43	0	82
FR		45	37	0	82
Total general		196	177	0	373

Tabla 27. Análisis de comas en todas las lenguas

1.3.2 Análisis y comparación de la corrección gramatical y terminológica de las transcripciones de Whisper y YouTube

Como se ha expuesto en el punto anterior, en discursos de registro formal hay menos errores que en los discursos de registro informal en todos los idiomas. Por lo general, el reconocimiento del habla y transcripción ha sido mejor con Whisper. Este no solo proporciona puntuación (comas y puntos, signos de interrogación) sino también acentuación. Por el contrario, en YouTube no había ninguna muestra con puntuación, a excepción de las cifras separadas por puntos. Esto ha generado más errores y más esfuerzo humano en el momento de corregir cada texto. Ya que con YouTube no había puntos finales de frase, ha sido necesario poner mayúsculas de inicio de frase allí donde fuera necesario, además de todos los otros signos de puntuación.

Se aprecia un mejor desempeño de ambas herramientas en todas las lenguas en discurso formal, mientras que, en discurso informal, como por ejemplo en el caso de diálogos, interrupciones, muestras con ruido de fondo o distorsiones y mezcla de idiomas; había un peor desempeño tanto en Whisper como en YouTube.

Es de suma importancia añadir que Whisper no censura el lenguaje verbal ofensivo o vulgar, como se ha podido ver en las muestras donde había insultos o palabras malsonantes. Por contra, YouTube censura ese tipo de terminología con el uso de este símbolo: [__].

Los mejores resultados de Whisper se han visto sobre todo en español e inglés, mientras que, en catalán, francés y japonés no lo ha conseguido lograr satisfactoriamente. Esto puede deberse a queo bien el modelo no haya sido entrenado con ese tipo de términos en esos idiomas, o bien sea porque, por lo general, como veíamos en la gráfica que analiza el WER, esos idiomas tienen una tasa de error más elevada, independientemente de si se utiliza un registro formal o informal.

En el caso del catalán y valenciano, el desempeño también ha sido mejor en ambas herramientas en lenguaje formal, mientras que en informal ha sido mejor Whisper.

Por lo general, viendo los resultados obtenidos de las correcciones, se puede afirmar que Whisper tiene un mejor desempeño en todos los ámbitos, además de tener una distancia de edición mucho menor a la de YouTube. Whisper utiliza mejor la puntuación, la acentuación, la gramática y el uso de nombres propios. Esto último se puede ver en casos como “Moderna”, la compañía farmacéutica; COVID-19, la enfermedad. Otros nombres propios lo han transcrito correctamente ambas herramientas: “mRNA”, “HIV”, “Wallace”, “La gran evasión”, “Leonardo DiCpario”.

Ha habido algunos errores de Whisper que han sido cometidos a causa del aglutinamiento de sonidos o de la cadencia del interlocutor, como, por ejemplo: *¿Tienes ~~caspador~~ ~~Matiti~~ ~~seborreica~~? > ¿Tienes caspa o dermatitis seborreica?

Esto se debe a la rapidez o el grado de vocalización del hablante. En estos casos en que se aglutinan sonidos, la herramienta no es capaz de discernir un sonido entre otro, cosa que una persona nativa de cada lengua lo podría hacer a la perfección y sin dificultades en la mayor parte de los casos.

Siempre influye negativamente, tanto para ambas herramientas como para un humano el hecho de que la calidad de sonido sea mala, haya interrupciones o sonidos de fondo que dificulten la audición.

También ha habido malas transcripciones en ambas herramientas en palabras demasiado coloquiales o vulgares, así como interferencias de idioma, como es el caso de las muestras en catalán y valenciano, donde a veces hay palabras en español.

En japonés el acierto de los *kanjis*, palabras, cifras y la gramática ha sido muy bueno en Whisper en la mayoría de las muestras, así como la puntuación. El único ejemplo en el que se ha visto superado en corrección por YouTube ha sido en las muestras formales de este idioma, especialmente en los *kanjis*.

IV. Conclusión

Recordando los objetivos de este trabajo, por un lado: comprobar la corrección lingüística y eficacia de la transcripción del reconocimiento de voz de Whisper respecto al sistema de reconocimiento de voz de YouTube Studio. Y, por otro lado: demostrar si la transcripción generada por inteligencia artificial es lo suficientemente buena y útil para poder compararse con la transcripción humana o si, por el contrario, aún no llega a un nivel de corrección suficiente, desde el punto de vista de la corrección lingüística, para poder implementar los resultados en la traducción audiovisual y la accesibilidad a contenido audiovisual; se puede determinar con certeza que Whisper tiene un mejor desempeño en todos los niveles tras analizar, corregir y comparar los resultados de las transcripciones en 5 lenguas, con distintos registros y niveles de formalidad.

Empezando por la rapidez en la que se hacen las transcripciones, pasando por el correcto uso de la gramática, la puntuación y la cantidad de terminología acertada y acabando por la censura de la lengua.

Whisper es una herramienta que podría ayudar mucho en el mundo de la traducción audiovisual, además de en otros campos y ámbitos de la vida. No obstante, si pensamos en la aplicación a la traducción con herramientas TAO y de subtitulación, vemos que el margen de error aún es demasiado alto como para poder utilizar los resultados obtenidos en su totalidad. Se podría incluso llegar a decir que parcialmente quizás tampoco ayudase mucho en la tarea de la traducción. A modo de reflexión final, puede llegar a ser más rápido una transcripción humana, ya que somos capaces de

diferencias sonidos aglutinados, podemos entender mejor a un interlocutor por mucho que haya interferencias en la calidad del audio, al menos mejor que cualquier herramienta.

No obstante, y como punto favorable, Whisper es una herramienta increíble y sumamente potente que puede llegar a ahorrar mucho esfuerzo humano, ya sea mental, de tiempo y físico. Dependiendo de para qué se requiera el uso de esta herramienta, es mejor utilizar las transcripciones automáticas, aún y teniendo que corregir pequeños errores.

Hoy en día ya hay herramientas que tienen Whisper integrado, como por ejemplo Subtitle Edit. Otras herramientas, como por ejemplo SDL Trados, antes utilizaba Microsoft Word para transcribir y ahora utiliza su propio motor.

Respecto al uso de Whisper aplicado al contenido audiovisual, pude llegar a ser muy útil porque conlleva un menor esfuerzo humano y rapidez a la hora de realizar esta tarea. Una desventaja, comparando con YouTube es que, de momento, no proporciona marcas de tiempo. Al menos no el modelo base que se ha utilizado en el presente trabajo.

Comparando las transcripciones de Whisper con las transcripciones de YouTube, consideraría el uso de las transcripciones de YouTube solamente por las marcas de tiempo. Estas transcripciones son más fáciles de aplicar al contenido audiovisual. Para otras cosas, lo desaconsejo, puesto que las correcciones que se tengan que realizar requieren más tiempo con YouTube que directamente realizar la transcripción manualmente.

Finalmente, tras la comparación de la puntuación y la corrección gramatical, el acierto en las palabras transcritas, nombres propios, terminología específica, tipos de formalidad del lenguaje y otros valores analizados en la parte práctica, concluyo que todavía hay demasiados errores como para utilizar las transcripciones con Whisper aplicando la TA (traducción automática), ya que los errores de transcripción pueden llegar a ser tantos que generaría todavía más errores de traducción automática. Solo consideraría el uso de las transcripciones de Whisper con TA tras haber hecho una posesición exhausta y minuciosa de los resultados de dichas transcripciones.

V. Referencias bibliográficas

Ada Colau [@adacolau]. (2023). TikTok.

<https://www.tiktok.com/@adacolau/video/7246409257929739546?lang=es&q=ada%20colau%20&t=1687629472386>

Alameda, T. (2019). *Machine learning”: ¿qué es y cómo funciona?* BBVA NOTICIAS. <https://www.bbva.com/es/machine-learning-que-es-y-como-funciona/>. Fecha de consulta: 8 de febrero de 2023.

Anyoha, R. (2017). *The History of Artificial Intelligence*. Harvard University, Estados Unidos. <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>. Fecha de consulta: 17 de enero de 2023)

Botella, C. (2007). *Aproximación al estudio del doblaje y la subtitulación desde la perspectiva prescriptivista y la descriptivista: la traducción audiovisual*. Universidad de Alicante. Tonos digital: Revista de estudios filológicos, ISSN 1577-6921, Nº. 13. (Botella, 2007). https://www.um.es/tonosdigital/znum13/secciones/tritonos_A_doblaje.htm. Fecha de consulta: 1 de marzo de 2023.

Chaume, F. [Frederic]. (2013). *Panorámica de la investigación en traducción para el doblaje*. TRANS, (17), 13-34 <https://doi.org/10.24310/TRANS.2013.v0i17.3225>

Defense Advanced Research Projects Agency. (n.d.). <https://www.darpa.mil>. Fecha de consulta: 4 de febrero de 2023.

De Mitjans Audiovisuals SA, C. C. (n.d.). *ésAdir > xulo | xula*. <https://esadir.cat/entrades/fitxa/node/xulo>. Fecha de consulta: 25 de mayo de 2023.

De Mitjans Audiovisuals SA, C. C. (n.d.-a). *ésAdir > er. . .* <https://esadir.cat/entrades/fitxa/id/5318>. . Fecha de consulta: 28 de mayo de 2023.

De Moragas, M. de (2012). *Comunicación, medios y cultura* (pp. 123-155). En De Moragas, M. (Ed.), *La comunicación: de los orígenes a Internet*. Barcelona: Gedisa. <http://www.fondationeuropea.com/wp-content/uploads/2015/05/LA-COMUNICACIÓN-De-los-or%C3%ADgenes-a-Internet.pdf>. Fecha de consulta: 4 de abril de 2023.

Díaz Cintas, J. (2010). *La accesibilidad a los medios de comunicación audiovisual a través del subtitulado y de la audiodescripción*, pág 157-180. ESLETRA. https://cvc.cervantes.es/lengua/esletra/pdf/04/020_diaz.pdf. Fecha de consulta: 23 de marzo de 2023.

Diccionari català-valencià-balear. (n.d.). *Vulgues*. <https://dcvb.iec.cat>. Fecha de consulta: 28 de mayo de 2023.

Diccionari català-valencià-balear. (n.d.). *Manejar*. <https://dcvb.iec.cat>. Fecha de consulta: 28 de mayo de 2023.

Fridman. L. (2019) [@LexFridman]. *Deep Learning Basics: Introduction and Overview*.

<https://www.youtube.com/watch?v=O5xeyoRL95U>. Fecha de consulta: 15 de enero de 2022.

Gabay, M. M. (2018). *Evolución de las redes sociales, de la comunicación y del marketing: análisis de su interrelación*. Universidad de San Andrés, Comunidad Autónoma de Buenos Aires, Argentina.

<https://repositorio.udesar.edu.ar/jspui/bitstream/10908/15927/1/%5BP%5D%5BW%5D%20M.%20Mar.%20Gabay,%20Mar%C3%ADa%20Mercedes.pdf> Fecha de consulta: 4 de abril de 2023.

Grau, J. (2002). *Antropología Audiovisual*. Bellaterra Edicions. Bellaterra, Cerdanyola del Vallès, Barcelona.

https://www.academia.edu/767287/Antropolog%C3%ADa_audiovisual. Fecha de consulta: 28 de marzo de 2023.

IBM. (n.d.). *¿Qué es la inteligencia artificial (IA)?*

https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwimz7HG14GAAXVoi_OHHTAKDZcQFnoECA8QAQ&url=https%3A%2F%2Fwww.ibm.com%2Fes-es%2Ftopics%2Fartificial-intelligence&usq=AOvVaw2M0zHgB8We0vhc1vZJN8ss&opi=89978449.

Fecha de consulta: 24 de mayo de 2023.

Jobin, C., Hooge, S., & Le Masson, P. (2020). *What does the proof-of-concept (POC) really prove? A historical perspective and a cross-domain analytical study*. XXIXème conférence de l'Association Internationale de Management Stratégique (AIMS). <https://hal.science/hal-02570321/document>. Fecha de consulta: 24 de mayo de 2023.

Juliana Canet [@julianacenet] Modelet suïs. TikTok.

<https://www.tiktok.com/@julianacenet/video/7237892928251661595?q=juliana%20canet%20ginebra&t=1688920911658>

Juliana Canet [@julianacenet]. Homes que no mengen cony. TikTok.

<https://www.tiktok.com/@julianacenet/video/7238682658698448155?lang=es>

Juliana Canet [@julianacenet]. Entrevista amb Oscar Cuevas. TikTok.

<https://www.tiktok.com/@julianacenet/video/7243473849973591323?lang=es&q=juliana%20canet%20carlos%20cuevas&t=1688921354543>

Juliana Canet . [@julianacenet]. Entrevista en català i valencià. TikTok.

<https://www.tiktok.com/@julianacenet/video/7233526279457344795?lang=es&q=juliana%20canet%20valencia&t=1688921396825>

Junts per Catalunya. [@junts_x_castellbisbal]. TikTok.

https://www.tiktok.com/@junts_x_castellbisbal/video/7171900597963132165?lang=es&q=delicte%20de%20sedicio&t=1687629756957

McQuail, D. (1994). *Teoría de la comunicación de masas*. Paidós.

<https://locucionucvcohortelvii.files.wordpress.com/2012/06/introduccion-a-la-teoria-de-la-comunicacion-de-masas-mcquail-denis.pdf>. Fecha de consulta: 4 de abril de 2023.

McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. McGraw-Hill Education.

<https://designopendata.files.wordpress.com/2014/05/understanding-media-mcluhan.pdf>. Fecha de consulta: 4 de abril de 2023.

Mitchell, T. (1997). *Machine Learning*, McGraw Hill. ISBN 0-07-042807-7.

<https://www.cin.ufpe.br/~cavmj/Machine%20-%20Learning%20-%20Tom%20Mitchell.pdf>. Fecha de consulta: 7 de febrero de 2023.

Moreno, A. et al. (1994). *Aprendizaje automático*. Edicions UPC. Universitat Politècnica de Catalunya. ISBN 84-7653-460-4.

<https://upcommons.upc.edu/bitstream/handle/2099.3/36157/9788483019962.pdf?sequence=1&isAllowed=y> Fecha de consulta: 8 de febrero de 2023.

Openai. (n.d.). *GitHub - openai/Whisper: Robust Speech Recognition via Large-Scale Weak Supervision*.

GitHub. <https://github.com/openai/whisper/tree/main>. Fecha de consulta: 18 de diciembre de 2022.

Peña, S.; Larrondo, A.; Morales, J. *Current affairs on TikTok*. (2022). *Virality and entertainment for digital natives*.

Profesional de la Información, 2022, vol. 31, no 1. <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/86803/63085>

Pérez, F. (2007). *Introducción a las series de tiempo: métodos paramétricos*. Universidad de Medellín, pág. 13-

14. [ISBN 9789589801079](https://www.amazon.es/dp/B000000000). Fecha de consulta: 2 de marzo de 2023.

Suárez, R. (2002). *La subtitulación del discurso oral: análisis comparativo entre una charla espontánea y una conferencia semiespontánea*. pág. 11-14. Dipòsit digital de documents de la UAB. Universitat autònoma de Barcelona.

<https://ddd.uab.cat/record/266438>. Fecha de consulta: 1 de marzo de 2023.

Turing, A.M. (1950). *Computing Machinery and Intelligence*. Mind, Volumen LIX, Edición 236, pág. 433–460.

<https://academic.oup.com/mind/article/LIX/236/433/986238>. Fecha de consulta: 1 de febrero de 2023.

Valdés, A. T. B., Pérez, Z. C., & Condel, A. A. D. V. (2017). *Comunicación audiovisual, nuevos paradigmas y tendencias*. Alcance. Revista Cubana de Información y Comunicación, vol. 6 núm. 12, pág. 26-55. ISSN 2411-9970

<http://scielo.sld.cu/pdf/ralc/v6n12/ralc03117.pdf> Fecha de consulta: 23 de marzo de 2023.

Huang, X., Baker, J. & Reddy, R. (2014). *A historical perspective of speech recognition*. Commun. ACM 57, 1 (2014), 94–103.

https://www.researchgate.net/publication/262157198_A_Historical_Perspective_of_Speech_Recognition Fecha de consulta: 27 de febrero de 2023.

Hui, J. (2019). *Speech Recognition — Weighted Finite-State Transducers (WFST)*. Medium. <https://jonathan-hui.medium.com/speech-recognition-weighted-finite-state-transducers-wfst-a4ece08a89b7>. Fecha de consulta: 28 de febrero de 2023.

Team, G. (n.d.). *Gradio*. <https://gradio.app/> <https://gradio.app>. Fecha de consulta: 24 de diciembre de 2022. <https://www.termcat.cat/es/diccionaris-en-linia/289/ca/S>

Pulp Fiction. [@_the_extra]. TikTok.

https://www.tiktok.com/@_the_extra/video/7179682947107212590?lang=es&q=pulp%20fiction%20motherfucker%20i%20dare%20you&t=1688406206574

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Robust speech recognition via large-scale weak supervision*. In International Conference on Machine Learning. PMLR. Universidad de Cornwell, Estados Unidos. <https://cdn.openai.com/papers/whisper.pdf>. Fecha de consulta: 18 de diciembre de 2022.

[@Shxtsngigs].

TikTok <https://www.tiktok.com/@shxtsngigs/video/7250491679197695259?lang=es&q=%40shxtsngigs%20akee&t=1688921603647>

[@jakekocherhans]. SIBO symptoms. TikTok.

<https://www.tiktok.com/@jakekocherhans/video/7227640217878990122?lang=es&q=sibo%20symptoms&t=1688407656825>

[@ekgscience]. Glial cells.

<https://www.tiktok.com/@ekgscience/video/7169481396413664513?q=glia%20cells&t=1683651097750>

[@watchmovclips]. Jamie Foxx interview about Leonardo DiCaprio in Django Unchained.

TikTok <https://www.tiktok.com/@watchmovclips/video/7227900128953224474?lang=es&q=leonardo%20di%20caprio%20n%20word&t=1688405758494>

[@yahooneews]. HIV Vaccine. TikTok.

<https://www.tiktok.com/@yahooneews/video/7059426451774475567?q=aids%20news&t=1683655405348>

[@s1gma_guffy]. Reservoir Dogs clip. TikTok.

https://www.tiktok.com/@s1gma_guffy/video/7216267105178619141?lang=es&q=reservoir%20dog%20n%20word&t=1688406546151

Platon, A. [@alvaroplato]. Dermatitis seborreica. TikTok.

<https://www.tiktok.com/@alvaroplato/video/6985216047364295941?lang=es&q=dermatitis%20seborreica&t=1688408036507>

[@ahorapodemos]. Irene Montero sobre la explotación reproductiva. TikTok.

<https://www.tiktok.com/@ahorapodemos/video/7221525494561279237?lang=es&q=discurso%20viente%20alquiler&t=1688407822383>

[@aljiuss_malaga]. Irene Montero sobre la explotación reproductiva. TikTok.

https://www.tiktok.com/@aljiuss_malaga/video/7216085915721043206?lang=es&q=discurso%20viente%20alquiler&t=1688407822383

[@anti_survie2]. Cellules gliales, partie 2. TikTok.

https://www.tiktok.com/@anti_survie2/video/7245683647435574554?lang=es&q=cellules%20gliales%20&t=1688364141816

[@businessfrancaisoff]. 2 types de gens. TikTok.

<https://www.tiktok.com/@businessfrancaisoff/video/7214613082042125595?lang=es&q=contenu%20en%20français&t=1688364220310>

[@tatty.macleod]. stay at home girlfriend... the French edition. TikTok.

<https://www.tiktok.com/@tatty.macleod/video/7179219753305771269>

[@ayumisawada] TikTok.

<https://www.tiktok.com/@ayumi.sawada/video/7207779417152572673?lang=es&q=理科議会&t=1688324428092>

[@arutiyuji]. センチュリー. TikTok.

<https://www.tiktok.com/@arutiyuji/video/7072674185985756417?lang=es&q=議会&t=1688324557427>

[@onakasuitaichi]. 濃厚バターチキンカレーレシピ. TikTok.

<https://www.tiktok.com/@onakasuitaichi/video/7192904586103983361?lang=es&q=レシピ&t=1688364000531>

[@saki.pan]. デイリーブロック. TikTok.

<https://www.tiktok.com/@saki.pan/video/7137210641323953409?lang=es&q=日本語を話すブログ&t=1688324882773>

VI. Anexos

1. Pasos a seguir para instalar Whisper en el ordenador:

Abrir el terminal y darle el siguiente comando: ‘pip install whisper’

```
Last login: Fri Mar 24 11:58:59 on console
/Users/annaacedoortega/.zprofile:7: no such file or directory: /opt/homebrew/bin/brew
/Users/annaacedoortega/.zprofile:9: no such file or directory: /opt/homebrew/bin/brew
/Users/annaacedoortega/.zprofile:10: no such file or directory: /opt/homebrew/bin/brew
annaacedoortega@MacBook-Air-de-Anna ~ % pip install whisper
Collecting whisper
  Downloading whisper-1.1.10.tar.gz (42 kB)
    Preparing metadata (setup.py) ... done
Collecting six
  Installing collected packages: six, whisper
  DEPRECATION: whisper is being installed using the legacy 'setup.py install' method, because it does not have a 'pyproject.toml' and the 'wheel' package is not installed. pip 23.1 will enforce this behavior change. A possible replacement is to enable the '--use-pep517' option. Discussion can be found at https://github.com/pypa/pip/issues/8559
  Running setup.py install for whisper ... done
Successfully installed six-1.16.0 whisper-1.1.10

[notice] A new release of pip available: 22.3.1 -> 23.0.1
[notice] To update, run: pip install --upgrade pip
annaacedoortega@MacBook-Air-de-Anna ~ % pip install --upgrade pip
Requirement already satisfied: pip in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (22.3.1)
Collecting pip
  Downloading pip-23.0.1-py3-none-any.whl (2.1 MB)
    2.1/2.1 MB 6.4 MB/s eta 0:00:00
Installing collected packages: pip
  Attempting uninstall: pip
    Found existing installation: pip 22.3.1
    Uninstalling pip-22.3.1:
      Successfully uninstalled pip-22.3.1
  Successfully installed pip-23.0.1
annaacedoortega@MacBook-Air-de-Anna ~ % pip install whisper
Requirement already satisfied: whisper in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (1.1.10)
Requirement already satisfied: six in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from whisper) (1.16.0)
annaacedoortega@MacBook-Air-de-Anna ~ %
```

A continuación, darle el siguiente comando: ‘pip install gradio’, que es el entorno web en el que se puede grabar en vivo la voz del locutor y, a continuación, transcribe el audio y lo pasa a texto.

```
annaacedoortega@MacBook-Air-de-Anna ~ % pip install whisper
Requirement already satisfied: whisper in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (1.1.10)
Requirement already satisfied: six in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from whisper) (1.16.0)
annaacedoortega@MacBook-Air-de-Anna ~ % pip install gradio
Collecting gradio
  Downloading gradio-3.23.0-py3-none-any.whl (15.8 MB)
    15.8/15.8 MB 7.8 MB/s eta 0:00:00
Collecting huggingface-hub
  Downloading huggingface_hub-0.13.3-py3-none-any.whl (199 kB)
    199.8/199.8 kB 5.6 MB/s eta 0:00:00
Requirement already satisfied: Jinja2 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from gradio) (3.1.2)
Collecting markdown-it-py[linkify]>=2.0.0
  Downloading markdown_it_py-2.2.0-py3-none-any.whl (84 kB)
    84.5/84.5 kB 3.3 MB/s eta 0:00:00
Collecting pyyaml
  Downloading PyYAML-6.0-cp310-cp310-macosx_10_9_x86_64.whl (197 kB)
    197.6/197.6 kB 4.7 MB/s eta 0:00:00
Collecting aiofiles
  Downloading aiofiles-23.1.0-py3-none-any.whl (14 kB)
Collecting fastapi
  Downloading fastapi-0.95.0-py3-none-any.whl (57 kB)
    57.1/57.1 kB 2.3 MB/s eta 0:00:00
Collecting pandas
  Downloading pandas-1.5.3-cp310-cp310-macosx_10_9_x86_64.whl (12.0 MB)
    12.0/12.0 MB 8.2 MB/s eta 0:00:00
Requirement already satisfied: numpy in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from gradio) (1.23.5)
Collecting websockets>=10.0
  Downloading websockets-10.4-cp310-cp310-macosx_10_9_x86_64.whl (97 kB)
    97.9/97.9 kB 3.6 MB/s eta 0:00:00
Requirement already satisfied: typing-extensions in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from gradio) (4.5.0)
Collecting mdit-py-plugins<0.3.3
  Downloading mdit_py_plugins-0.3.3-py3-none-any.whl (50 kB)
    50.5/50.5 kB 1.4 MB/s eta 0:00:00
Collecting httpx
  Downloading httpx-0.23.3-py3-none-any.whl (71 kB)
    71.5/71.5 kB 2.9 MB/s eta 0:00:00
Collecting aiohttp
  Downloading aiohttp-3.8.4-cp310-cp310-macosx_10_9_x86_64.whl (358 kB)
    358.1/358.1 kB 7.0 MB/s eta 0:00:00
Collecting semantic-version
  Downloading semantic_version-2.10.0-py2.py3-none-any.whl (15 kB)
Collecting pydub
  Downloading pydub-0.25.1-py2.py3-none-any.whl (32 kB)
Collecting uvicorn
  Downloading uvicorn-0.21.1-py3-none-any.whl (57 kB)
    57.8/57.8 kB 2.3 MB/s eta 0:00:00
Requirement already satisfied: MarkupSafe in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from gradio) (2.1.2)
Collecting python-multipart
  Downloading python_multipart-0.0.6-py3-none-any.whl (45 kB)
    45.7/45.7 kB 1.8 MB/s eta 0:00:00
Collecting altair>=4.2.0
  Downloading altair-4.2.2-py3-none-any.whl (813 kB)
    813.6/813.6 kB 4.5 MB/s eta 0:00:00
Collecting orjson
  Downloading orjson-3.8.9-cp310-cp310-macosx_10_9_x86_64.macosx_11_0_arm64.macosx_10_9_universal2.whl (488 kB)
    488.6/488.6 kB 8.0 MB/s eta 0:00:00
Requirement already satisfied: requests in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from gradio) (2.28.1)
Collecting ffpay
  Downloading ffpay-0.3.0.tar.gz (4.8 kB)
  Preparing metadata (setup.py) ... done
Collecting matplotlib
  Downloading matplotlib-3.7.1-cp310-cp310-macosx_10_12_x86_64.whl (7.4 MB)
    7.4/7.4 MB 8.2 MB/s eta 0:00:00
```

```

Collecting matplotlib
  Downloading matplotlib-3.7.1-cp310-cp310-macosx_10_12_x86_64.whl (7.4 MB)
    7.4/7.4 MB 8.2 MB/s eta 0:00:00
Collecting fsspec
  Downloading fsspec-2023.3.0-py3-none-any.whl (145 kB)
    145.4/145.4 kB 5.8 MB/s eta 0:00:00
Collecting pydantic
  Downloading pydantic-1.10.7-cp310-cp310-macosx_10_9_x86_64.whl (2.9 MB)
    2.9/2.9 MB 7.0 MB/s eta 0:00:00
Collecting pillow
  Downloading Pillow-9.4.0-2-cp310-cp310-macosx_10_10_x86_64.whl (3.3 MB)
    3.3/3.3 MB 7.2 MB/s eta 0:00:00
Collecting toolz
  Downloading toolz-0.12.0-py3-none-any.whl (55 kB)
    55.0/55.0 kB 1.7 MB/s eta 0:00:00
Collecting entrypoints
  Downloading entrypoints-0.4-py3-none-any.whl (5.3 kB)
Collecting jsonschema>=3.0
  Downloading jsonschema-4.17.3-py3-none-any.whl (90 kB)
    90.4/90.4 kB 3.1 MB/s eta 0:00:00
Collecting packaging>=20.9
  Downloading packaging-23.0-py3-none-any.whl (42 kB)
    42.7/42.7 kB 1.6 MB/s eta 0:00:00
Requirement already satisfied: tqdm>=4.42.1 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from huggingface-hub->gradio) (4.65.0)
Requirement already satisfied: filelock in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from huggingface-hub->gradio) (3.10.2)
Collecting mdurl==0.1
  Downloading mdurl-0.1.2-py3-none-any.whl (10.0 kB)
Collecting linkify-it-py<3,>=1
  Downloading linkify-it-py-2.0.0-py3-none-any.whl (19 kB)
Collecting pytz>=2020.1
  Downloading pytz-2023.3-py2.py3-none-any.whl (502 kB)
    502.3/502.3 kB 8.2 MB/s eta 0:00:00
Collecting python-dateutil>=2.8.1
  Downloading python_dateutil-2.8.2-py2.py3-none-any.whl (247 kB)
    247.7/247.7 kB 4.7 MB/s eta 0:00:00
Collecting async-timeout<5.0,>=4.0.0a3
  Downloading async_timeout-4.0.2-py3-none-any.whl (5.8 kB)
Collecting frozenlist==1.1.1
  Downloading frozenlist-1.3.3-cp310-cp310-macosx_10_9_x86_64.whl (35 kB)
Collecting attrs>=17.3.0
  Downloading attrs-22.2.0-py3-none-any.whl (60 kB)
    60.0/60.0 kB 2.4 MB/s eta 0:00:00
Collecting multidict<7.0,>=4.5
  Downloading multidict-6.0.4-cp310-cp310-macosx_10_9_x86_64.whl (29 kB)
Collecting aiohttp>=1.1.2
  Downloading aiohttp-1.3.1-py3-none-any.whl (7.6 kB)
Requirement already satisfied: charset-normalizer<4.0,>=2.0 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from aiohttp->gradio) (2.1.1)
Collecting yarl<2.0,>=1.0
  Downloading yarl-1.8.2-cp310-cp310-macosx_10_9_x86_64.whl (61 kB)
    61.0/61.0 kB 2.5 MB/s eta 0:00:00
Collecting starlette<0.27.0,>=0.26.1
  Downloading starlette-0.26.1-py3-none-any.whl (66 kB)
    66.9/66.9 kB 1.8 MB/s eta 0:00:00
Collecting httpcore<0.17.0,>=0.15.0
  Downloading httpcore-0.16.3-py3-none-any.whl (69 kB)
    69.6/69.6 kB 3.2 MB/s eta 0:00:00
Collecting rfc3986[idna2008]<2,>=1.3
  Downloading rfc3986-1.5.0-py2.py3-none-any.whl (31 kB)
Requirement already satisfied: certifi in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from httpx->gradio) (2022.9.24)
Collecting sniffio
  Downloading sniffio-1.3.0-py3-none-any.whl (10 kB)
Collecting kiwisolver>=1.0.1
  Downloading kiwisolver-1.4.4-cp310-cp310-macosx_10_9_x86_64.whl (65 kB)
    65.5/65.5 kB 2.8 MB/s eta 0:00:00
  Downloading attrs-22.2.0-py3-none-any.whl (60 kB)
    60.0/60.0 kB 2.4 MB/s eta 0:00:00
Collecting multidict<7.0,>=4.5
  Downloading multidict-6.0.4-cp310-cp310-macosx_10_9_x86_64.whl (29 kB)
Collecting aiohttp>=1.1.2
  Downloading aiohttp-1.3.1-py3-none-any.whl (7.6 kB)
Requirement already satisfied: charset-normalizer<4.0,>=2.0 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from aiohttp->gradio) (2.1.1)
Collecting yarl<2.0,>=1.0
  Downloading yarl-1.8.2-cp310-cp310-macosx_10_9_x86_64.whl (61 kB)
    61.0/61.0 kB 2.5 MB/s eta 0:00:00
Collecting starlette<0.27.0,>=0.26.1
  Downloading starlette-0.26.1-py3-none-any.whl (66 kB)
    66.9/66.9 kB 1.8 MB/s eta 0:00:00
Collecting httpcore<0.17.0,>=0.15.0
  Downloading httpcore-0.16.3-py3-none-any.whl (69 kB)
    69.6/69.6 kB 3.2 MB/s eta 0:00:00
Collecting rfc3986[idna2008]<2,>=1.3
  Downloading rfc3986-1.5.0-py2.py3-none-any.whl (31 kB)
Requirement already satisfied: certifi in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from httpx->gradio) (2022.9.24)
Collecting sniffio
  Downloading sniffio-1.3.0-py3-none-any.whl (10 kB)
Collecting kiwisolver>=1.0.1
  Downloading kiwisolver-1.4.4-cp310-cp310-macosx_10_9_x86_64.whl (65 kB)
    65.5/65.5 kB 2.8 MB/s eta 0:00:00
Collecting pyparsing>=2.3.1
  Downloading pyparsing-3.0.9-py3-none-any.whl (98 kB)
    98.3/98.3 kB 4.2 MB/s eta 0:00:00
Collecting fonttools>=4.22.0
  Downloading fonttools-4.39.3-py3-none-any.whl (1.0 MB)
    1.0/1.0 MB 8.6 MB/s eta 0:00:00
Collecting cyclex==0.10
  Downloading cyclex-0.11.0-py3-none-any.whl (6.4 kB)
Collecting contourpy>=1.0.1
  Downloading contourpy-1.0.7-cp310-cp310-macosx_10_9_x86_64.whl (244 kB)
    244.2/244.2 kB 5.5 MB/s eta 0:00:00
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from requests->gradio) (1.26.12)
Requirement already satisfied: idna<4,>=2.5 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from requests->gradio) (3.4)
Collecting h11==0.8
  Downloading h11-0.14.0-py3-none-any.whl (58 kB)
    58.3/58.3 kB 2.3 MB/s eta 0:00:00
Collecting click>=7.0
  Downloading click-8.1.3-py3-none-any.whl (96 kB)
    96.6/96.6 kB 3.8 MB/s eta 0:00:00
Collecting anyio<5.0,>=3.0
  Downloading anyio-3.6.2-py3-none-any.whl (80 kB)
    80.6/80.6 kB 2.9 MB/s eta 0:00:00
Collecting pyrsistent==0.17.0,!=0.17.1,!=0.17.2,>=0.14.0
  Downloading pyrsistent-0.19.3-cp310-cp310-macosx_10_9_universal2.whl (82 kB)
    82.5/82.5 kB 3.7 MB/s eta 0:00:00
Collecting uc-micro-py
  Downloading uc_micro_py-1.0.1-py3-none-any.whl (6.2 kB)
Requirement already satisfied: six>=1.5 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from python-dateutil>=2.8.1->pandas->gradio) (1.16.0)
Installing collected packages: uc-micro-py, toolz, sniffio, semantic-version, pypaml, python-multipart, python-dateutil, pyrsistent, pyparsing, pydantic, pillow, packaging, orjson, multidict, mdurl, kiwisolver, h11, fsspec, frozenlist, fonttools, entrypoints, cyclex, contourpy, click, attrs, async-timeout, aiofiles, yarl, uvicorn, pandas, matplotlib, markdown-it-py, linkify-it-py, jsonschema, huggingface-hub, anyio, aiohttp, starlette, mdit-py-plugins, httpcore, altair, aiohttp, httpx, fastapi, gradio
DEPRECATION: ffpay is being installed using the legacy 'setup.py install' method, because it does not have a 'pyproject.toml' and the 'wheel' package is not installed. pip 23.1 will enforce this behavior change. A possible replacement is to enable the '--use-pep517' option. Discussion can be found at https://github.com/pypa/pip/issues/8559
Running setup.py install for ffpay ... done
Successfully installed aiofiles-23.1.0 aiohttp-3.8.4 aiohttp-1.3.1 altair-4.2.2 anyio-3.6.2 async-timeout-4.0.2 attrs-22.2.0 click-8.1.3 contourpy-1.0.7 cyclex-0.11.0 entrypoints-0.4 fastapi-0.95.0 ffpay-0.3.0 fonttools-4.39.3 frozenlist-1.3.3 fsspec-2023.3.0 gradio-3.23.0 h11-0.14.0 httpcore-0.16.3 httpx-0.23.3 huggingface-hub-0.13.3 jsonschema-4.17.3 kiwisolver-1.4.4 linkify-it-py-2.0.0 markdown-it-py-2.2.0 matplotlib-3.7.1 mdit-py-plugins-0.3.0 mdurl-0.1.2 multidict-6.0.4 orjson-3.8.9 packaging-23.0 pandas-1.5.3 pillow-9.4.0 pydantic-1.10.7 pydub-0.25.1 pyparsing-3.0.9 pyrsistent-0.19.3 python-dateutil-11.2.8.2 python-multipart-0.0.6 pytz-2023.3 pypaml-6.0 rfc3986-1.5.0 semantic-version-2.10.0 sniffio-1.3.0 starlette-0.26.1 toolz-0.12.0 uc-micro-py-1.0.1 uvicorn-0.21.1 websockets-10.4 yarl-1.8.2
annaesdoortegaaMacBook-Air-de-Anna ~ %

```



```

Downloading starlette-0.26.1-py3-none-any.whl (66 kB)
Collecting httpcore<0.17.0,>=0.15.0
  Downloading httpcore-0.16.3-py3-none-any.whl (69 kB)
Collecting rfc3986(idna2008)<2,>=1.3
  Downloading rfc3986-1.5.0-py2.py3-none-any.whl (31 kB)
Requirement already satisfied: certifi in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from httpx>gradio) (2022.9.24)
Collecting sniffio
  Downloading sniffio-1.3.0-py3-none-any.whl (10 kB)
Collecting kiwisolver<=1.0.1
  Downloading kiwisolver-1.4.4-cp310-cp310-macosx_10_9_x86_64.whl (65 kB)
Collecting pydantic<=2.3.1
  Downloading pydantic-2.3.1-py3-none-any.whl (98 kB)
Collecting fonttools<=4.22.0
  Downloading fonttools-4.22.0-py3-none-any.whl (1.0 MB)
Collecting clicker<=7.0
  Downloading clicker-8.1.3-py3-none-any.whl (96 kB)
Collecting anyio<5.0,>=3.0
  Downloading anyio-3.6.2-py3-none-any.whl (80 kB)
Collecting pyrsistent<=0.17.0,!=0.17.1,!=0.17.2,>=0.14.4
  Downloading pyrsistent-0.17.3-cp310-cp310-macosx_10_9_universal2.whl (82 kB)
Collecting uc-micro-py
  Downloading uc_micro_py-1.0.1-py3-none-any.whl (6.2 kB)
Requirement already satisfied: six>=1.5 in /Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages (from python-dateutil>=2.8.1->pandas>gradio) (1.16.0)
Installing collected packages: rfc3986, pytz, pydub, ffmpeg, websockets, uc-micro-py, toolz, sniffio, semantic-version, pyyaml, python-multipart, python-dateutil, pyrsistent, pydantic, pillow, p
ackaging, orjson, multidict, mdurl, kiwisolver, h11, fsspec, frozenlist, fonttools, entrypoints, cyclar, contourpy, click, attrs, async-timeout, aiofiles, yarl, uvicorn, pandas, matplotlib, markdown-it-py
, linkify-it-py, jsonschema, huggingface-hub, anyio, aiohttp, starlette, mdit-py-plugins, httpcore, altair, aiohttp, httpx, fastapi, gradio
DEPRECATION: ffray is being installed using the legacy 'setup.py install' method, because it does not have a 'pyproject.toml' and the 'wheel' package is not installed. pip 23.1 will enforce this behavio
ur change. A possible replacement is to enable the '--use-pep517' option. Discussion can be found at https://github.com/pypa/pip/issues/8559
Running setup.py install for ffray ... done
Successfully installed aiofiles-23.1.0 aiohttp-3.8.4 aiohttp-1.3.1 altair-4.2.2 anyio-3.6.2 async-timeout-4.0.2 attrs-22.2.0 click-8.1.3 contourpy-1.0.7 cyclar-0.11.0 entrypoints-0.4 fastapi-0.95.0 ffray
-y-3.0 fonttools-4.22.0 frozenlist-1.3.3 fsspec-2023.3.0 gradio-3.23.0 h11-0.14.0 httpcore-0.16.3 httpx-0.23.3 huggingface-hub-0.13.3 jsonschema-4.17.3 kiwisolver-1.4.4 linkify-it-py-2.0.0 markdown-it-py
-2.2.0 matplotlib-3.7.1 mdit-py-plugins-0.3.3 mdurl-0.1.2 multidict-6.0.4 orjson-3.8.9 packaging-23.0 pandas-1.5.3 pillow-9.4.0 pydantic-1.10.7 pydub-0.25.1 pydantic-3.0.9 pyrsistent-0.17.3 python-dateut
il-2.8.2 python-multipart-0.6.0 pytz-2023.3 pyyaml-6.0 rfc3986-1.5.0 semantic-version-2.10.0 sniffio-1.3.0 starlette-0.26.1 toolz-0.12.0 uc-micro-py-1.0.1 uvicorn-0.21.1 websockets-10.4 yarl-1.8.2
annaedoortega@MacBook-Air-de-Anna ~ % pip install ffray
Collecting ffray
  Downloading ffray-1.4.tar.gz (5.1 kB)
  Preparing metadata (setup.py) ... done
Installing collected packages: ffray
DEPRECATION: ffray is being installed using the legacy 'setup.py install' method, because it does not have a 'pyproject.toml' and the 'wheel' package is not installed. pip 23.1 will enforce this behavi
our change. A possible replacement is to enable the '--use-pep517' option. Discussion can be found at https://github.com/pypa/pip/issues/8559
Running setup.py install for ffray ... done
Successfully installed ffray-1.4
annaedoortega@MacBook-Air-de-Anna ~ % brew install ffray
Running 'brew update --auto-update' ...
#

```

Por último, ejectuar el siguiente comando: ‘pip install ffmpeg’ desde Homebrew, que es para hacerlo accesible a un ordenador MacOS. Sin esto, no se podría completar la instalación de Whisper en este tipo de terminal.

```

ackaging, orjson, multidict, mdurl, kiwisolver, h11, fsspec, frozenlist, fonttools, entrypoints, cyclar, contourpy, click, attrs, async-timeout, aiofiles, yarl, uvicorn, pandas, matplotlib, markdown-it-py
, linkify-it-py, jsonschema, huggingface-hub, anyio, aiohttp, starlette, mdit-py-plugins, httpcore, altair, aiohttp, httpx, fastapi, gradio
DEPRECATION: ffray is being installed using the legacy 'setup.py install' method, because it does not have a 'pyproject.toml' and the 'wheel' package is not installed. pip 23.1 will enforce this behavio
ur change. A possible replacement is to enable the '--use-pep517' option. Discussion can be found at https://github.com/pypa/pip/issues/8559
Running setup.py install for ffray ... done
Successfully installed ffray-1.4
annaedoortega@MacBook-Air-de-Anna ~ % brew install ffray
Running 'brew update --auto-update' ...
#> Homebrew is run entirely by unpaid volunteers. Please consider donating:
#> https://github.com/Homebrew/brew#donations
#> Downloading https://formulae.brew.sh/api/formula.jws.json
#> Downloading https://formulae.brew.sh/api/cask.jws.json
#> Fetching dependencies for ffray: highway, libmath, jpeg-turbo, xz, zstd, openssl, webp, jpeg-xl, aom, xriib24, david, freetype, fontconfig, ca-certificates, openssl@1.1, libnghttp2, unbound, gnutils, gl
ib, libwebp, libxml2, icu4c, harfbuzz, libunibreak, libass, mbedtls, librist, libvidstab, libvpx, rav1e, libsmptexture, mpg123, libandfile, rubberband, sdl2, snappy, avt-av1, libarchive, pango and tesseract
#> Fetching highway
#> Downloading https://ghcr.io/v2/homebrew/core/highway/manifests/1.0.4
#> Downloading https://ghcr.io/v2/homebrew/core/highway/blobs/sha256:d6d822c1989e1a382593ee64a7d59146c63ce153008f56740648f98ecbab2e3c
#> Downloading from https://pkg-containers.githubusercontent.com/ghcr1/blobs/sha256:d6d822c1989e1a382593ee64a7d59146c63ce153008f56740648f98ecbab2e3c?tee=2023-03-30T18:03:45.033Z
#> Fetching libmath
#> Downloading https://ghcr.io/v2/homebrew/core/libmath/manifests/3.1.7
#> Downloading https://ghcr.io/v2/homebrew/core/libmath/blobs/sha256:5e98e640b156cf2cbab7ee8fca4e2d464c192b522b34955843a1ab8f4c7130eb
#> Downloading from https://pkg-containers.githubusercontent.com/ghcr1/blobs/sha256:5e98e640b156cf2cbab7ee8fca4e2d464c192b522b34955843a1ab8f4c7130eb?tee=2023-03-30T18:03:45.033Z
#> Fetching jpeg-turbo
#> Downloading https://ghcr.io/v2/homebrew/core/jpeg-turbo/manifests/2.1.5.1
#> Downloading https://ghcr.io/v2/homebrew/core/jpeg-turbo/blobs/sha256:2836e3a4f9f53a16f532d901c1c8d8dc411a6cc29a984e993f6a377e0f
#> Downloading from https://pkg-containers.githubusercontent.com/ghcr1/blobs/sha256:2836e3a4f9f53a16f532d901c1c8d8dc411a6cc29a984e993f6a377e0f?tee=2023-03-30T18:03:45.033Z
#> Fetching xz
#> Downloading https://ghcr.io/v2/homebrew/core/xz/manifests/5.4.2
#> Downloading https://ghcr.io/v2/homebrew/core/xz/blobs/sha256:e37a99ec525b581132f833cd99f6cd29f3299e1b9972d6d3cac4938176d1
#> Downloading from https://pkg-containers.githubusercontent.com/ghcr1/blobs/sha256:e37a99ec525b581132f833cd99f6cd29f3299e1b9972d6d3cac4938176d1?tee=2023-03-30T18:03:45.033Z
#> Fetching zstd
#> Downloading https://ghcr.io/v2/homebrew/core/zstd/manifests/1.5.4
#> Downloading https://ghcr.io/v2/homebrew/core/zstd/blobs/sha256:833ef757e78b4c27e29f7abaa45ee962b622ffa4c1de5e02ef4fff13394a3e
#> Downloading from https://pkg-containers.githubusercontent.com/ghcr1/blobs/sha256:833ef757e78b4c27e29f7abaa45ee962b622ffa4c1de5e02ef4fff13394a3e?tee=2023-03-30T18:03:45.033Z
#> Fetching openssl
#> Downloading https://ghcr.io/v2/homebrew/core/openssl/manifests/3.1.6
#>

```


olviden ustedes los medios de comunicación, que son quienes tienen la oportunidad de mostrarles esta realidad al conjunto de la ciudadanía de las mujeres que están detrás, que no se olviden de las mujeres que están detrás. Hay un sesgo de discriminación por pobreza, claro, hace poco veíamos un reportaje en el que un señor que llevaba una agencia de gestación sobre la gente.

Detected language: fr

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

à qui on a donné le nom de cellules gliales. Ce qui est très intéressant à ce sujet, c'est que dans le cadre du principe d'une eurone, la théorie d'une eurone, on croit que les neurones étaient les cellules fonctionnelles du cerveau. Et les cellules gliales étaient là plus ou moins juste en soutien. Et bien par définitions cela veut juste dire que vous utilisez que 10% de votre cerveau parce que c'est tout ce que les neurones représentent dans le cerveau, 10%. Pendant longtemps, les cellules gliales encore une fois n'étaient considérées que comme un tissu conjonctif qui apportait un soutien physique et nuage.

Detected language: en

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

What does Marcellus Wallace look like? What? What country you from? What? What ain't no country I ever heard of? They speak English and what? What? English motherfucker, do you speak it? Yes, and you know what I'm saying. Describe what Marcellus Wallace looks like. What? Say what again. Say what again. I dare you. I double dare you motherfucker. Say what? One more goddamn times.

Detected language: en

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

This is huge. Moderna has begun its first clinical trials of an HIV vaccine. It actually uses the same mRNA technology as Moderna's COVID-19 vaccine. Our researchers have been trying to make an HIV vaccine for decades. And if this works, it could help up to 1.2 million Americans who currently have the virus.

Detected language: ja

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

ご飯が止まらないお家で簡単、トロトロ、ツヤツヤ、ご飯が進むおー、ノコバタチューキンカレイごちそうさまでした、モモニク、ヨーグルと入れて、みんな大好き、ハチミツ、カレーコ、おさじ 1、おしよ入れまして、まずまずでございます、下味付けましたそしたら、ニンニクとしょうがを入れて、フライパンを固むけて

Detected language: ca

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

Home, es que fan pintada no menjar conya. Home, enjar que el només si demana si us plau de ja no és que dalt menjar. Primer de tot ja no et pique. Ja no et pique és una persona gràmosa, et menja el conya si t'estima. I... Oh... És que no ho sé, perquè fa pintada guarríssim, però no menja. No, no menja. Tot el resto fa perfecte, però menja. El nou dia de la l'olita, el mat i ni a més... No, és que mai de la vida menjar el conya, igual que el Pablo Castellano, no ve la Maria Pomo, és un tipus de la luna.

Detected language: en

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

Leonardo DiCaprio had a problem saying the word nigga. Pow, pal, it's just, it's tough for me to say this. And I remember saying, you know, Jackson, I'm gonna get over him, motherfucker. This is another Tuesday, motherfucker. I'm gonna talk about these motherfuckers. Look at what you, look at who I am, motherfucker. And it was like, you know, and I said, yeah, I said, Leo, we're not friends. I said, this is just another Thursday. This is your property. These are humans. These are your property. He looks, excuse me, more.

Detected language: ja

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

なぜセンチュリーなのでしょうか事故が少ないということですね例えば衝突時の安全性とか誰が何ともどういう風に安全性があったりとかいろいろと限られた上でこれが例えば大統領者のようなボーダンガラスというのはございませんケジトンシャーとセンチュリーが来た当然センチュリーの安全というのは誰が見てもわかることです

Detected language: ca

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")
i els prepara a l'Uni. Bonat, de nou. En ningú s'havia escapa aquestes alçades, la gran polèmica generada a Espanya, sobre el tema del delict de sedició, la reforma del codi penal. Aquest passat dií, el Congrés va votar a favor de la presa en consideració de la reforma del codi penal proposada per peixó i podem d'acord amb esquerra publicana de Catalunya. Però atenció. La presa en consideració.
Detected language: es
/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:
UserWarning: FP16 is not supported on CPU; using FP32 instead
warnings.warn("FP16 is not supported on CPU; using FP32 instead")
¿Tienes caspador Matiti seborreca? Entonces este vídeo es para ti. La dermatiti seborrica es un trastorno de la piel que afecta varias partes del cuerpo, especialmente al cuero cabelludo. También se le conoce como caspa, excemas seborreico, soriasis seborreca y en los bebés se le llama costra láctea. Los síntomas son descamaciones de la piel, manchas cubiertas con escamas blancas, amarillas o costras, en rojamiento de la piel y picazón. Aún se desconoce las causas de la dermatiti seborreca, pero se relaciona con un hongo llamado humana seca o por una respuesta irregular del sistema.
Detected language: en
/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:
UserWarning: FP16 is not supported on CPU; using FP32 instead
warnings.warn("FP16 is not supported on CPU; using FP32 instead")
Seen the Neural Anatomy and Function Lecture, we covered how the nervous system is composed of two major cell types. We have the neurons and glial cells, or glia, or neuroglia. The difference between the two is that glia don't have a direct role in information processing, but they do have several functions that are important for the healthy functioning of the nervous system. Glial cells are located within the central nervous system, which includes the brain and spinal cord, and the peripheral nervous system, which includes the nerves that are in the brain.
Detected language: en
/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:
UserWarning: FP16 is not supported on CPU; using FP32 instead
warnings.warn("FP16 is not supported on CPU; using FP32 instead")
What are you convinced people are pretending to enjoy sparkling water? Facts Delete it that tastes like stout if I had it tastes disgusting Pretend to enjoy their friend success wow I'm gas you're doing so well for yourself. I'm wrong. This one is for me facts. Aki bro. Facts Aki socks I'm sorry guys. I'm carrying you to aki socks, bro. I hate it boil potatoes There's a U-
Detected language: ja
/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:
UserWarning: FP16 is not supported on CPU; using FP32 instead
warnings.warn("FP16 is not supported on CPU; using FP32 instead")
やっぱり使いともの方の 2 階の人、みだいろことを呼んでいるんですけども、そうじゃなくて、皆さん言っているのは記憶ということですので。そう、使えてくれますか?って言ったら、今、やったらお小伝伝という人に伝えます。みたいなことを言ったんですけども、
Detected language: ja
/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:
UserWarning: FP16 is not supported on CPU; using FP32 instead
warnings.warn("FP16 is not supported on CPU; using FP32 instead")
今日 5 号 1 時すぎ 東京 8 歳を自治の中学校で 理科の実験中に 複数の生徒が体調不良を訴えているということです ケース調によりますと 今日 5 号 1 時すぎ 8 歳を自治の 8 歳を自学院 8 歳を自治学校で 理科の実験中に気分が悪くなった生徒が複数人いると 199 番通報がありました当時中学 2 年の生徒 36 人が 理科の事業を受けていて
Detected language: ja
/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:
UserWarning: FP16 is not supported on CPU; using FP32 instead
warnings.warn("FP16 is not supported on CPU; using FP32 instead")
こんにちは、今日のデイリブブロックのボイソーダーは日本語であげてみたいと思いますってよりたくさん日本語でアップロードしてくださいってリクエストがたくさんあったんですけどちょっと時間くかつちゃってすみません今日は日本語で頑張っていきます私の生まれ育ちがオーストラリアで小学校から大学生まで全部オーストラリアで帰っているんでやっぱり英語がメインとしてずっとしゃべってきたんですけど家ではやっぱり親が両方日本人なので日本語ずっとしゃべってきたし小学校から中学校
Detected language: ca
/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:
UserWarning: FP16 is not supported on CPU; using FP32 instead
warnings.warn("FP16 is not supported on CPU; using FP32 instead")
no han comut, ha vingut aquí, en aquest pla, sense fer cap pacte amb cap formació política, sense tenir certes escoles com acabaria aquesta inversió, però amb una única certesa, que és que nosaltres anem directes a l'oposició. Per primer cop, en aquest saludable ens hi han aquest ajuntament, hi ha representats de box. Tenen dret a estar aquí, això no ho pot qüestionar ningú, perquè els han votat una part de la ciutadania. Però jo sí que els vull deixar molt clar que nosaltres no pensem normalitzar la seva presència i encara menys el seu discurs.
Detected language: ja

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

警察と証報によりますと 25 日午後 3 時半ごろ、4 回地市区場長の改造側で近くに住む 13 歳の男子中学生が、動給生と小学 6 年の妹を合わせて 3 人で、川に入って遊んでいたところ、妹が覚れそうになり、助けに入るため川に入りましたが、姿が見えなくなったということです。19 番通報を受け、発生からおおよそ 15 分後にかけつけた急車隊が、水心 1.5m の川底に沈んでいた

Detected language: ca

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

Ben dia no hi ha gent, ja sóc a Ginebra, que és la meua habitació de hotel. Jo he decidit que Ginebra se l'anà bastida molt clàssic, perquè és una ciutat així com a l'Europa i a molt dalt. Llavors jo he postat per a Samarretta, a base i qui, si m'ha de col·laborar del l'única, els meus pantalons negres, preciosos, xulíssims, cap tan un col·lazo, amb els meus moquesins vegades que m'encanten, i porto aquest mocador per donar-li una mica d'aquell, perquè sigui un punt més emocional, perquè si no sigues veritat que hi ha gent.

Detected language: es

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

¿Eres la volición insta de la exportación reproductiva? Sí, sí. Y de hecho, esta legislatura hemos conseguido que, que, claramente, reconocida como una forma de violencia contra las mujeres, que hasta ahora no ocurría. Lo que queríamos haber introducido, además de esto, es que, en España, esos contratos son nulos de pleno derecho, porque no es una práctica legal. Y de hecho, está penado. Pero si tú haces el contrato en otro país, entonces no es perseguible. Entonces, para nosotros, podría haber sido una buena forma de contribuir a eliminar esta práctica para perseguir los contratos también a través del mundo.

Detected language: ca

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

Et pots parlar de la pèl·la? Ja ets que vulguis. No m'he espas sentit a parlar amb l'Ancia. Però no hi ha pèl·la, el catalán no és el que valencià. Ja haurà pèl·les. Estan existint i a l'espera. Els valencià, el catalán, són cosin germans. Estan destinats a d'or saber i a portar saber com ho va dir. Què tal, per aquí per Barcelona, t'agrada la ciutat? M'encanta. La gent que em permudar-te algun dia. Jo soc valenciana i de la terretre ara mateix a ningú. Tu, m'heu sentit com a emocionar-me. No sé com de modelos.

Detected language: es

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

Vale, voy a deciros de qué va la hay que ver. Trata sobre una putita que es una máquina de polla, no sea por lo menada, por lo tarde, por la noche, polla, polla, polla, polla, polla, polla. ¿Cuántas polla son? Un montón. Entonces la tía un día se encuentra con un hijo puta que tiene un perazo de pollo. El cíos como Charles Bronson en la gran evasión, Cava Tunes. La tía ha encontrado la polla de su vino, ¿verdad? Y siente algo que no había sentido nunca. Dola. Dola. Chuto, vich.

Detected language: fr

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

C'est tout ce que je fais pour mes enfants aujourd'hui, comme je vous disais à mon père.

Detected language: fr

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

2 types de gens, les gens qui te disent « eh », les gens qui disent « mais ». Tu vas aller dire « voilà, on va faire un vélo de dingue ». Et on va faire ça alors tu as deux cas, tu as le mec qui dit « ouais mais si on fait ça, ça raconte des chers, mais si on met ça, on va le voler ». Ça c'est les mais. Et puis tu le racontes la même phrase, il dit « ouais, on va mettre ça et ça, il va pouvoir faire ça et on va rajouter ça ». Tarnat 2. Alors ceux qui font des « eh », c'est les entrepreneurs. Et ceux qui font des mais, c'est gestionnaire.

Detected language: en

/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/whisper/transcribe.py:114:

UserWarning: FP16 is not supported on CPU; using FP32 instead

warnings.warn("FP16 is not supported on CPU; using FP32 instead")

Here are five signs you might have small intestine bacteria low growth. So number one, and this is probably one of the most common symptoms of SIBO, but it's that chronic bloating and stomach distension. Now the reason why this happens is because that bacteria in your gut loves carbohydrates. When you have an overgrowth in the small intestine, anytime you consume a meal, they're essentially creating fermentation from those carbohydrates, producing gas in the stomach, which is

causing a lot of that discomfort and distension. So number two, and this is also a really common symptom associated with SIBO, but it's chronic diarrhea.

annaacedoortega@MacBook-Air-de-Anna CODIGO_AUDIOS_RESULTADOS %