INTELIGENCIA ARTIFICIAL EN LA TRANSCRIPCIÓN DE ENTREVISTAS[1]

DRA. VERÓNICA YÉPEZ-REYES
https://orcid.org/0000-0003-3617-0418
Pontificia Universidad Católica del Ecuador
vyepezr@puce.edu.ec

DR. JORGE CRUZ-SILVA
https://orcid.org/0000-0002-5327-2152
Pontificia Universidad Católica del Ecuador
jacruz@puce.edu.ec

Recibido: 17 de noviembre del 2023 / Aceptado: 21 de febrero del 2024
doi: https://doi.org/10.26439/contratexto2024.n41.6750

RESUMEN. Las entrevistas, fundamentales para el ejercicio periodístico y la investigación cualitativa, capturan el significado profundo del pensamiento de la humanidad. En 2023, las herramientas de inteligencia artificial (IA) se popularizaron, y vimos que se las empezó a usar en la grabación, transcripción y subtitulado de discursos. El objetivo del estudio es identificar la IA más adecuada para transcribir grabaciones en español y priorizar la completitud de tareas, la eficacia y la eficiencia. La IA seleccionada se aplicará a un corpus de 450 entrevistas cortas, para luego codificarlas y analizar su contenido. El artículo se centra en cuatro herramientas de transcripción con IA en español: Office 365 (Word) Transcribe, Amazon Transcribe, Notta y Whisper. La tecnología permite aprovechar la riqueza de la grabación original sin que medie la intervención, y posible modificación, de quien la transcribe, sea persona o asistente virtual. Los resultados señalan la rapidez de la transcripción y la posibilidad de las IA de procesar y alojar en línea documentos escritos. En cuanto a las posibilidades de interacción con el texto, se observa el rol fundamental de los equipos de investigación en la comprensión profunda y análisis del contenido con el apoyo proporcionado por las IA en las tareas de transcripción.

PALABRAS CLAVE: transcripción / inteligencia artificial / entrevistas / automatización / interacción humano-computadora

1 Esta investigación es parte del Proyecto Fact Checking e Inteligencia Artificial, código QINV0438-IINV502000000, financiado por la Dirección de Investigación de la Pontificia Universidad Católica del Ecuador.

ARTIFICIAL INTELLIGENCE IN INTERVIEW TRANSCRIPTION

ABSTRACT. Interviews, crucial for journalistic practice and qualitative research, capture the profound meaning of human thought. In 2023, artificial intelligence (AI) tools became widespread, including their use in recording, transcribing, and subtitling speeches. The study aims to identify the most suitable AI for transcribing recordings in Spanish, prioritizing task completeness, efficiency, and effectiveness. The selected AI will be applied to a corpus of 450 short interviews, which will then be coded and analyzed for content. The article focuses on four Spanish-language AI transcription tools: Office 365 (Word) Transcribe, Amazon Transcribe, Notta, and Whisper. The technology allows harnessing the richness of the original recording without the intervention, and potential modification, of the person or virtual assistant transcribing it. The results highlight the speed of transcription and the ability of AIs to process and host written documents online. Regarding possibilities for interacting with the text, the fundamental role of research teams in the deep understanding and analysis of content is observed, with support provided by AIs in transcription tasks.

KEYWORDS: transcription / artificial intelligence / interviews / automatization / human-computer interaction

INTELIGÊNCIA ARTIFICIAL NA TRANSCRIÇÃO DE ENTREVISTA

RESUMO. Entrevistas, fundamentais para o exercício jornalístico e pesquisa qualitativa, capturam o significado profundo do pensamento humano. Em 2023, ferramentas de inteligência artificial (IA) se popularizaram, incluindo seu uso na gravação, transcrição e legendagem de discursos. O objetivo do estudo é identificar a IA mais adequada para transcrever gravações em espanhol, priorizando a completude das tarefas, eficácia e eficiência. A IA selecionada será aplicada a um corpus de 450 entrevistas curtas, que serão posteriormente codificadas e analisadas quanto ao conteúdo. O artigo foca em quatro ferramentas de transcrição com IA em espanhol: Office 365 (Word) Transcribe, Amazon Transcribe, Notta e Whisper. A tecnologia permite aproveitar a riqueza da gravação original sem a intervenção, e possível modificação, da pessoa ou assistente virtual que a transcreve. Os resultados destacam a rapidez da transcrição e a capacidade das IAs em processar e hospedar documentos escritos online. Em relação às possibilidades de interação com o texto, observa-se o papel fundamental das equipes de pesquisa na compreensão profunda e análise de conteúdo, com o suporte proporcionado pelas IAs nas tarefas de transcrição.

PALAVRAS-CHAVE: transcrição / inteligência artificial / entrevistas / automatização / interação humano-computador

INTRODUCCIÓN

"Quizás todos vivimos en lo que podría denominarse una sociedad de la entrevista, en donde las entrevistas son fundamentales para dar sentido a nuestras vidas" (Silverman, 2013, p. 124). Esta cita, pese a haberse escrito hace más de una década, sigue siendo de actualidad. Una búsqueda rápida de la palabra entrevista (*interview*) en la base de datos Ebsco arroja más de dos millones de entradas y Scopus registra casi un millón. Hay documentos escritos de entrevistas, como método de obtención de información, que datan del año 400 a. C. cuando Tucídides lo empleaba para obtener testimonios de los participantes de las guerras del Peloponeso (Kvale, 2011). El término entrevista consiste en una visión compartida por una díada sobre un punto de vista (entre-vista), un mutuo intercambio al mirar y organizar el mundo a través del diálogo (Castells, 2002).

La entrevista es la manera privilegiada para ampliar información respecto de la temática de una investigación (Lopezosa et al., 2023), pero son pocos los autores que, al abordar la metodología cualitativa, se detienen a analizar el proceso de transcripción. El texto seminal de Blecua (1983), sobre crítica textual, sostiene que "en cuanto el mensaje oral se fija en la escritura se convierte en un texto. La crítica textual, en efecto, puede trabajar sobre tradiciones orales, pero solo cuando quedan fijadas en forma de texto" (p. 17). En este estudio, los investigadores y periodistas consultados sostuvieron, enfáticamente, lo mismo: toda entrevista debe transcribirse, pues la entrevista deja de ser un registro oral para convertirse en un documento para el análisis. Así, la transcripción "transforma una grabación de audio o de video (información primaria) en representaciones espaciales, icónicas y textuales (información secundaria) del habla y la conducta corporal en la interacción lingüística" (Greco et al., 2019).

Hoy en día la tecnología brinda diversas posibilidades para el uso directo de las grabaciones originales, sin pasar por la intermediación que implica la transcripción. Estas tecnologías son las responsables de que el análisis de contenido ya no se remita únicamente al análisis del texto escrito, sino que se sumerja dentro de la amplia gama de estudios del discurso (Angermuller et al., 2014), en donde la multimodalidad, su eje principal, incorpora textos, imágenes, sonidos, movimiento, artefactos y múltiples modos y formas de comunicación.

Adicionalmente, las implicaciones que tiene la inteligencia artificial (IA) en todos los campos, pero particularmente en las ciencias sociales, en la educación y en la comunicación, han generado la necesidad de la producción de investigaciones interdisciplinarias dentro del campo de las ciencias sociales computacionales con un interés creciente en las aplicaciones e implicaciones sociales de la IA (Ligo et al., 2021).

En el ámbito de la IA aplicada a la comunicación, pocas investigaciones se centran en el análisis detallado de las metodologías cualitativas relativas al proceso de transcripción. Frecuentemente, las discusiones omiten este paso intermedio entre la conducción

de entrevistas y el posterior análisis de contenido, que se limita a proporcionar explicaciones extensas sobre la entrevista y la interpretación de la información recopilada. Sin embargo, la transcripción no debe considerarse simplemente como una tarea trivial, sino como un proceso que enfrenta una serie de desafíos prácticos y técnicos específicos dentro del contexto de la investigación en inteligencia artificial relacionada con la comunicación. Como señala Kvale (2011), es un malentendido considerar la transcripción como una "tarea menor" (p. 124). Hay consenso entre los autores de que la transcripción no se lleva a cabo como un fin en sí misma, sino siempre como un medio para lograr un objetivo mayor (Kreuz & Riordan, 2018, p. 96).

Kvale (2011) pone en cuestión tres aspectos de la transcripción: ¿cuál es el procedimiento de la transcripción?, ¿cuál es el uso que se dará al documento escrito? y ¿quién transcribe? El procedimiento para la transcripción se relaciona con el sistema que ha de aplicarse a los registros orales para mantener la unidad y la consistencia del discurso, así todos serán equivalentes para su análisis.

En el campo de la lingüística, la transcripción es un tema relevante: la transformación de las grabaciones del lenguaje oral en texto escrito es central (McMullin, 2023). Ahondando en el campo de la etnometodología y el análisis de la conversación (EMCA, por sus siglas en inglés), la transcripción es fundamental, como destaca el proyecto colaborativo de la emcawiki.net, que provee una base de datos especializada de publicaciones que abordan la EMCA.

De manera general se identifican dos grandes sistemas de transcripción: el ortográfico y el fonético (Nagy, 2014). Estos sistemas, *grosso modo*, diferencian el eje en donde se centra la atención: el primero se enfoca en lo que se dijo, en la intención comunicativa, en contraposición con el segundo, que se centra en cómo esto se enunció (Kreuz & Riordan, 2018).

Por un lado, el sistema ortográfico busca ser fiel a las reglas de la escritura para generar un texto bien documentado y transparente. En ello encuentra, entre otros, un problema en el uso de los signos de puntuación. Mientras el habla transcurre de corrido, con inflexiones, silencios y frases que modulan el relato, la puntuación deviene de signos arbitrarios provistos por quien hace la transcripción. La puntuación y sus signos no son parte sustancial del corpus original, sino son libertades que se permite quien transcribe.

Por otro lado, la transcripción fonética procura incluir todos los detalles del habla que contiene la grabación. Esto es particularmente relevante para el campo del análisis de la conversación (CA, por sus siglas en inglés), el cual se basa en detalladas transcripciones de conversaciones. Para ello, el CA ha desarrollado sofisticados sistemas, que pueden haber iniciado con las convenciones clásicas de transcripción de Gail Jefferson y que se han refinado durante más de cuarenta años (Wagner, 2022). Estas convenciones, a partir de símbolos empleados, proveen información importante acerca de la entonación,

el alargue de segmentos y la unión entre palabras y momentos. Este tipo de transcripción procura plasmar el habla y su interacción en un documento.

Este estudio tiene como objetivo identificar la IA más adecuada para transcribir grabaciones en español en el ámbito de la investigación, que priorice la completitud de tareas, la eficacia y la eficiencia. Por esta razón, la transcripción ortográfica es el ámbito que se aborda con mayor profundidad. Para ello, Point y Baruch (2023) examinan cuatro estrategias para la transcripción en la investigación: (a) hacerla por cuenta propia, (b) contratar el servicio de transcripción, (c) no transcribir o (d) utilizar inteligencia artificial. Estas estrategias se enmarcan dentro de la interacción humano-computador sin cuestionamientos, como si se tratara de una parte inherente al proceso de transcripción y que se da de forma similar a lo que O'Brien (2020) analiza respecto del proceso de traducción.

En el contexto de la comunicación, hacer la entrevista por cuenta propia propone que el comunicador se sumerja en la entrevista, revise y transcriba el contenido completo para revivir el momento, ya que adopta el papel de oyente e interpelador para introducir cambios de significado en el material traducido (Creswell, 2013, p. 74). La transcripción, por sí sola, proporciona al comunicador la capacidad de recordar la conversación y descubrir elementos que podrían haber pasado desapercibidos durante la interacción de la entrevista. Volver a escuchar la entrevista y transcribirla permite regresar al registro sonoro y revivir el momento del diálogo junto con las memorias que evoca. Esta estrategia de transcripción es reconocida como la más apropiada y altamente valorada en la investigación horizontal, que busca generar preguntas y respuestas diversas con el otro, en contraposición a reproducir los mismos discursos hegemónicos frente a las nuevas realidades (Cornejo & Rufer, 2020, p. 8). Es evidente que la transcripción es una tarea tediosa y prolongada; el tiempo requerido puede oscilar entre tres y ocho horas para transcribir una hora de grabación, pues depende de la velocidad de tipeo (McMullin, 2023, p. 141).

La segunda estrategia es la de contratar o contar con un equipo de asistentes de investigación, personas encargadas de llevar a cabo el proceso de transcripción. La transcripción por encargo es un oficio que se remonta a los primeros papiros escritos en el 1100 a. C. y que luego se convertirán en el símbolo de la cultura griega: "La publicación de libros la hacían los copistas, el autor reunía varios escribas y les iba dictando el libro. Estos utilizaban letras mayúsculas sin separar las palabras … Los escribas cobraban por cada línea escrita" (Ossa, 1993, p. 63). La transcripción por encargo implica que alguien por fuera de la díada entrevistador-entrevistado genere el documento escrito. Este contrato supone una confianza del investigador en la fidelidad de la transcripción producida. Así mismo, el investigador está consciente de que toda transcripción implica una reducción de la riqueza de detalles de la propia interacción lingüística al plasmarse en un texto escrito (Wagner, 2022).

La tercera estrategia, mencionada por los autores, es la de no transcribir y, en su lugar, realizar el análisis y la codificación de la entrevista directamente a partir del audio

original. Hoy en día las grabaciones se pueden reproducir en el mismo dispositivo utilizado para la transcripción (computador, laptop, tableta). Hasta hace pocos años, este uso de archivos digitales de audio y video era impensable, dado que se requerían múltiples dispositivos para grabar, para reproducir, para tipear e imprimir las transcripciones. Hoy en día, distintos *softwares* de análisis cualitativo de datos (CAQDAS, por sus siglas en inglés) permiten la codificación y análisis de entrevistas a partir de la reproducción directa de la grabación original, sin necesidad de intermediar su transcripción (ejs. NVivo y ATLAS.ti). En esta práctica de uso del material original (sin alteraciones) para el proceso de codificación, la transcripción no llega a omitirse en su totalidad. Existen partes de la grabación (denominadas unidades) que requieren ser transcritas para la presentación de un informe final de investigación (Point & Baruch, 2023). Estas unidades son frases literales que se transcriben tal y como fueron enunciadas y que, en muchos casos, en un primer ciclo de análisis se utilizan para la codificación "en vivo" mediante el uso literal de las expresiones de la locución, es decir, palabras o frases propias del entrevistado. La codificación en vivo es distinta de la codificación descriptiva, la cual es provista de términos propuestos por el investigador (Saldaña, 2016).

Finalmente, la cuarta estrategia constituye el eje de este estudio: el empleo de inteligencia artificial (IA) para la transcripción. Los asistentes de voz personales, como Google Home Mini, Alexa de Amazon, Siri de Apple y Cortana de Microsoft, pueden realizar el proceso de reconocimiento de voz y su transcripción a texto (Nagaraj et al., 2023). No obstante, las grabaciones de entrevistas a terceros no siempre pueden ser comprendidas a cabalidad por estos dispositivos, por lo que existe un amplio número de plataformas en línea, provistas con IA, capaces de convertir una grabación a un texto, con alto grado de fidelidad.

Las plataformas en línea reemplazan y dejan de lado a programas de transcripción ampliamente utilizados hasta hace diez años, como es el caso del programa Transcriber. Este *software* gratuito permite ver la gráfica del sonido grabado para, manualmente, posibilitar el que se introduzcan secciones y se tipee la transcripción. Otra herramienta popular, oTranscribe, tiene incorporado el sistema de reproducción de texto y posibilita tipearlo en línea para su descarga y uso en otros documentos digitales.

Un informe realizado por *Consumer Reports* (Waddell, 2022) examinó siete programas que generan subtitulado automático y los resultados reportaron que todos cometieron algún tipo de error. Ese estudio resalta el desafío que tienen los programas para subtitular mientras ocurre la locución de manera totalmente correcta. Esto se atribuye a la dificultad de encontrar una lengua, acento o dialecto estandarizada y sin altibajos, así como a la calidad del audio y la transmisión. El estudio de García-Prieto y Figuereo-Benítez (2022) destaca dos proyectos de generación automatizada de subtitulado con el uso de inteligencia artificial: ELITR y Deep Sync.

En cuanto a la calidad de la grabación para una transcripción fidedigna, si bien

los teléfonos celulares inteligentes hoy en día sustituyen a las grabadoras de audio al tener esta capacidad incorporada, aún se comercializan grabadoras digitales y dictáfonos con kits completos de grabación y reconocimiento de voz que incluyen micrófonos, audífonos, así como una transcripción automática, a la par que cuentan con aplicaciones propias para dispositivos móviles que cierran el círculo de la entrevista y transcripción.

No obstante, lo que atañe a esta investigación son las plataformas de transcripción que tienen incorporadas IA. Más allá de permitir el acceso desde cualquier dispositivo con conexión a internet, generan transcripciones con alta fidelidad, para uso inmediato y sin intervención del investigador, como se explica en los siguientes apartados.

La inteligencia artificial y el proceso de transcripción

La inteligencia artificial (IA) es un concepto que involucra a máquinas inteligentes y algoritmos que ocupan cada vez más áreas de la vida contemporánea, y en cada una presentan resultados que, en determinados aspectos, superan las capacidades humanas.

Existe una amplia discusión filosófica sobre si se debe denominar "inteligente" a una máquina. Dos experimentos clásicos son ampliamente utilizados para apoyar y refutar la denominación de IA. El científico británico Alan Turing fue el primero en emplear el término inteligencia artificial con la prueba de Turing, que determina que una máquina es inteligente si puede mantener una conversación y ser capaz de imitar las respuestas que daría una persona, hasta el punto de engañar al ser humano. El semiólogo John Searle, en 1980, refutó a Turing con el experimento de la habitación china, el cual consiste en aislar a una persona en una habitación y proveerle de instrucciones para poder leer un texto escrito en mandarín. Quien está afuera de la habitación, un hablante de mandarín, envía información y recibe una respuesta, por lo que asume que la persona dentro de la habitación habla su idioma. Searle aclara que ni la persona ni la computadora en tal escenario están realmente pensando, al contrario, siguen instrucciones de entrada y salida. Tanto la prueba de Turing como la habitación china aparecen limitadas, porque hoy conocemos que las personas reaccionan diferente a un mismo estímulo y también pueden responder igual a un estímulo distinto, y lo que hace la IA es capturar una imitación de ese comportamiento (Berkemer & Grottke, 2023).

Seifert et al. (2018) clasifican tres tipos de inteligencia artificial: tecnologías de IA orientadas al comportamiento, sistemas que piensan y actúan racionalmente, y sistemas de *hardware* de inspiración biológica. Para fines de esta investigación se aborda solamente la primera categoría de tecnologías de IA orientadas al comportamiento, la cual involucra tres grandes ámbitos: tecnologías semánticas, procesamiento del lenguaje natural (NLP, por sus siglas en inglés) y modelado cognitivo.

En las tecnologías semánticas, el procesamiento digital se basa en la ejecución de reglas sintácticas que, para considerar el contenido, emplean conexiones entre hechos,

eventos, conceptos y categorías que explican las relaciones y funciones de una gran cantidad de datos con el empleo de métodos estadísticos. NLP se refiere a cómo las máquinas entienden, interpretan y procesan el lenguaje humano que constituye un área fundamental dentro de la investigación de IA, dado que el uso de ironía, metáforas y comparaciones difiere diametralmente del lenguaje formal de comandos que utilizan las computadoras. En ese contexto es donde también cobra importancia la pragmática del uso de la lengua, que permite a las personas agregar palabras que sirven para adornar la conversación y que las personas reconocen como tal y omiten en el análisis de contenido, pero las máquinas son incapaces (aún) de reconocerlas (Claeser et al., 2023). El NLP es tan complejo que todavía ningún sistema ha pasado con éxito la prueba de Turing (Seifert et al., 2018, p. 58).

El modelado cognitivo se refiere a entender el funcionamiento de los procesos cognitivos de las personas como la memoria a largo plazo, el pensamiento lógico y el razonamiento, y se lo analiza a partir de errores frecuentes en el razonamiento de las personas para comprender la percepción humana y los procesos de procesamiento de información. Esta es también una tarea compleja, puesto que incluso las metáforas en relación a la digitalidad que construyen las personas evocan conocimientos culturales previos, los cuales están ya arraigados en su sistema conceptual (Girón-García & Esbrí-Blasco, 2019).

Dentro de esta categoría de las tecnologías de IA orientadas al comportamiento se inserta la transcripción. El proceso de reconocimiento de voz implica múltiples etapas. Comienza con la captura de señales de audio y muestras de habla mediante micrófonos incorporados en dispositivos. Luego, se realiza una descomposición de estas señales, que eliminan el ruido de fondo y ajustan el tono, volumen y tempo del habla para su análisis. Posteriormente, la información digital se convierte en frecuencias, que preparan el terreno para la interpretación del habla humana. Un componente crítico en este proceso es la modelización acústica, que genera representaciones matemáticas de fonemas, las unidades fundamentales del sonido que distinguen unas palabras de otras. Además, se formulan hipótesis contextualmente informadas para comprender el significado del discurso.

Las redes neuronales son el enfoque más utilizado, principalmente debido a su notable capacidad para procesar información (Benkerzaz et al., 2019). Esta característica las ha posicionado como uno de los modelos más importantes en el ámbito de la inteligencia artificial. Además, los avances recientes en IA han impulsado la amplia utilización de las redes neuronales. Estas redes se estructuran en tres capas fundamentales: la capa de entrada, la capa oculta y la capa de salida.

La precisión en el reconocimiento de voz se apoya en algoritmos específicos. Tres de estos algoritmos destacan por su relevancia (O'Shaughnessy, 2024):

1. Modelo oculto de Markov (HMM). Este algoritmo aborda la diversidad del habla, incluyendo variaciones en pronunciación, velocidad y acento. Proporciona un marco eficaz para modelar la estructura temporal de las señales de audio y voz, así como la secuencia de fonemas que componen una palabra.
2. Alineación temporal dinámica (DTW). El DTW compara secuencias de habla con diferencias en velocidad y sincroniza grabaciones de audio con distintas velocidades y longitudes.
3. Redes neuronales artificiales (ANN). Las ANN, inspiradas en las redes neuronales humanas, han impulsado avances significativos en el reconocimiento de voz, ya que utilizan técnicas de aprendizaje profundo para mejorar la precisión.

Es esencial diferenciar entre el reconocimiento del habla y el reconocimiento de la voz, dos tecnologías relacionadas pero distintas. El reconocimiento del habla se enfoca en su conversión en texto legible, que mejora la eficiencia y automatiza procesos. Por otro lado, el reconocimiento de la voz se centra en la autenticación y seguridad, que convierte a la voz en datos digitales basados en características únicas del hablante, como tono, tonalidad y ritmo, con aplicaciones clave en la autenticación de identidad, como el desbloqueo de dispositivos.

La velocidad de procesamiento de los datos es el factor primordial que se reconoce en la incorporación de la tecnología de transcripción. Para poder realizar la transcripción es necesario, a veces, bajar la velocidad de la locución. Acelerar o ralentizar la reproducción de videos es una tecnología que entró en funcionamiento en 2017 para la plataforma de alojamiento de videos YouTube, y en 2021 para la aplicación de mensajería instantánea WhatsApp, lo que posibilita un menor o mayor tiempo de duración de una locución.

El presente estudio plantea explorar el uso de la IA en el proceso de transcripción de entrevistas grabadas en español. Para ello, se seleccionaron plataformas de IA con la aplicación de criterios específicos de exclusión que se explican en la siguiente sección.

METODOLOGÍA

En este estudio se eligieron 10 plataformas que utilizan IA para la transcripción de entrevistas grabadas en español. La elección de plataformas fue intencional (Reales et al., 2022), pues se consideraron aquellas que tienen una mayor probabilidad de producir una transcripción eficaz, relativa al tiempo empleado, y eficiente, en cuanto a la calidad del texto escrito. Es importante contemplar que la elección de estas IA no es exhaustiva, dada la vertiginosidad de las tecnologías pronto aparecerán otras alternativas con capacidades superiores.

Luego de una primera revisión, se excluyeron las IA de pago, puesto que la investigación plantea la sustentabilidad de su uso en un entorno académico. La única herramienta

de pago que se utilizó fue Office 365, porque tanto docentes como estudiantes pueden contar con licencia provista por el centro de estudios o de investigación.

Se excluyeron además aquellas que presentaban complicaciones para la carga de archivos de audio (Media.io, Veed.io, Bear File Converter), las de licencias gratuitas de corta duración (Transkriptor, Cockatoo) y aquellas que no permitían elegir el idioma (Otter.ai). Finalmente, se eligieron cuatro plataformas de IA: Office 365, con su complemento Transcribe en la aplicación de procesador de palabras Word, Amazon Transcribe, Notta y Whisper. Estas permitían su libre utilización durante el periodo de prueba y entendían audio en español (véase la Tabla 1) (Covella, 2005).

Tabla 1

Plataformas de IA para transcripción

Programa	Licencia	Formatos de audio	Voces	Exportar
Office 365 (Word) Transcribe	profesional ($)	mp3, mp4, wav, mp4a	sí	.docx
Amazon Transcribe	1 año gratis	mp3, mp4, wav, flac, amr, ogg, webm	sí	.json
Notta	Gratis 120 min/mes	mp3, m4a, caf, aiff, avi, rmvb, flv, mp4, mov, wmv, wma	no	--
Whisper	Cuenta gratuita Google Drive | Colaboratory	mp3, m4a, caf, aiff, avi, rmvb, flv, mp4, mov, wmv, wma	no	.txt, .tsv, .srt, .json, .vtt

Nota. Datos corresponden a septiembre del 2023.

Una vez seleccionada la herramienta de IA, se utilizó el modelo propuesto por Covella (2005) que revisa la completitud de tareas, eficiencia, eficacia y satisfacción. Esta herramienta será utilizada para la transcripción de 450 entrevistas cortas para su codificación y análisis de contenido en una investigación posterior.

Para efectos de esta investigación se eligieron entrevistas cortas, sin voces solapadas, con turnos bien marcados entre entrevistador y entrevistado. Su finalidad fue establecer la herramienta de IA que realice la transcripción más fidedigna para su posterior análisis y que no implique un tiempo excesivo de revisión y corrección del texto final.

El análisis de la completitud de tareas, eficiencia en cuanto a la velocidad y eficacia respecto de la fidelidad se realizó sobre tres entrevistas cortas con dos voces: (a) dos voces de mujer (33 s), (b) dos voces de hombre (34 s) y (c) mixta, mujer y hombre con una mayor duración (56 s). Esto no garantiza que en entrevistas más largas el comportamiento sea igual, pero da luces sobre las posibilidades (y limitaciones) de las plataformas.

Se realizaron entrevistas a un panel de cinco expertos: investigadores en áreas de la salud, psicología, periodismo, ciencias sociales y humanidades, con el fin de comprender su percepción respecto del ejercicio de la transcripción y la incorporación de asistentes virtuales a la tarea. Cada entrevista fue semiestructurada, "planificada y flexible con el propósito de obtener descripciones del mundo de la vida del entrevistado con respecto a la interpretación del significado de los fenómenos descritos" (Kvale, 2011, p. 163). Las entrevistas tuvieron una duración promedio de 30 minutos.

A partir de estas entrevistas, una plataforma que no fue incluida en el análisis pero que luego fue recomendada por uno de los investigadores entrevistados es Riverside. Sobre su utilización, fue interesante conocer que, debido al número de errores de concordancia en las transcripciones, una vez que se cuenta con la transcripción en texto, el investigador lo envía a ChatGPT para que corrija errores de concordancia en el texto que faciliten su lectura.

RESULTADOS

Este estudio buscó encontrar la herramienta de IA más adecuada para la transcripción, con la más alta fidelidad y el menor tiempo posible, de un corpus de 450 entrevistas cortas que puedan luego ser empleadas para su codificación y análisis del contenido. Los documentos de texto resultantes serán importados y analizados con un *software* especializado para el manejo de datos cualitativos.

1. Se analizaron las siguientes IA:
2. Microsoft Word de Office 365 con su complemento Transcribir
3. Amazon Transcribe, que forma parte de Amazon Web Services (AWS), una colección de distintos servicios de datos de la multinacional Amazon
4. Notta, una herramienta de transcripción de pago que cuenta con una licencia gratuita pero limitada en tiempo de uso
5. Whisper, la herramienta de IA de la empresa OpenAI, de código abierto que se instala en Colaboratory de Google Research, el equipo de investigación de Google que desarrolla herramientas de IA

Estas cuatro IA se probaron en reiteradas ocasiones y luego se emplearon para la transcripción de las 3 entrevistas elegidas con las que se obtuvieron los resultados que se resumen en la Tabla 2. Se identificaron dos variables que impactan en eficiencia, eficacia y completitud de tareas: error de voz se refiere a la identificación del locutor y

error en el texto alude a las transcripciones incorrectas del audio locutado. En la discusión se aborda con más detalle cada una de las variables.

Tabla 2

Datos del uso de las IA para transcripción

	IA	# error voz	# error texto	Tiempo*
Voces masculinas	Office 365	0	3	00:33.85
	AWS	0	1	01:01.94
	Notta	No hay voces	2	00:06.96
	Whisper	No hay voces	0	05:07.77
Voces femeninas	Office 365	4	3	00:29.97
	AWS	2	1	01:02.47
	Notta	No hay voces	4	00:07.29
	Whisper	No hay voces	0	07:44.43
Voces mixtas	Office 365	5	4	00:25.32
	AWS	2	0	01:02.24
	Notta	No hay voces	1	00:46.10
	Whisper	No hay voces	1	08:57.10

Nota. (*) Tiempo empleado para transcribir el texto.

Las grabaciones de solo voces masculinas y solo voces femeninas tienen una duración aproximada de 30 segundos. Este dato es importante, en tanto Office demora un poco menos del tiempo de reproducción a velocidad normal en transcribir el texto (33" y 29"), pero la plataforma Notta rompe el récord al transcribirlas en apenas segundos, tiempo mucho menor de lo que toma la reproducción a velocidad normal de la entrevista.

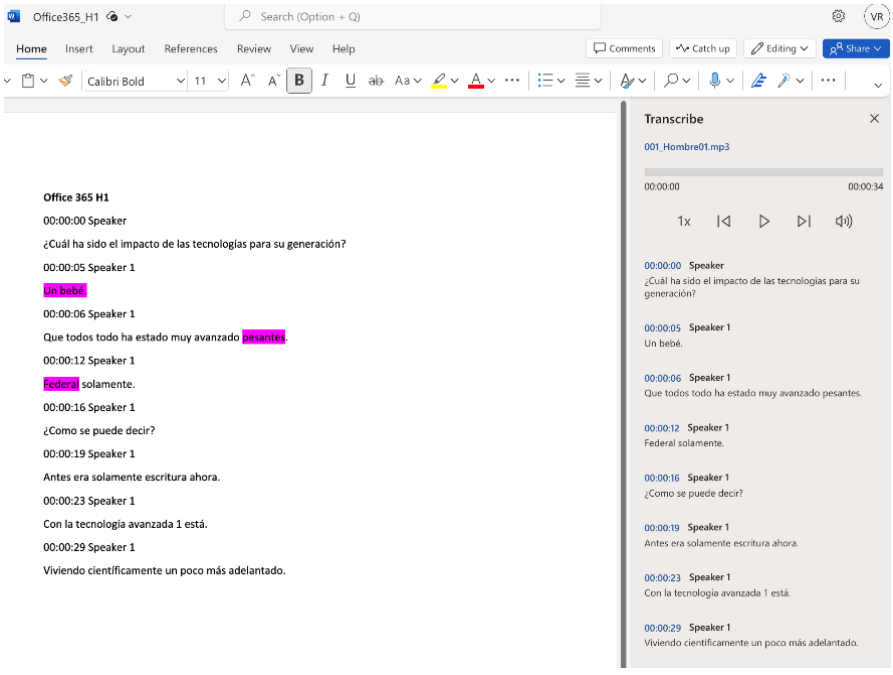
Eficiencia y eficacia en la transcripción

Las plataformas provistas con IA requieren algún tipo de instalación que, si bien debe realizarse en la mayoría de los casos por una única ocasión, es importante tomar en cuenta su facilidad de instalación y uso. En las plataformas analizadas, Microsoft lleva la delantera. Su utilización es tan simple como escribir en el procesador de palabras: pulsar en el ícono de micrófono, elegir "Transcribir", cargar el archivo y seleccionar el idioma (véase la Figura 1). Esta misma tecnología la incorpora el programa de reuniones virtuales Teams, que permite la transcripción y subtitulado durante el encuentro virtual, es empleada por investigadores y periodistas para entrevistas a través de videoconferencias (Teams, Zoom, Webex, etcétera) que permiten contar con la grabación y su transcripción.

En cuanto a formatos aceptados, Microsoft admite únicamente cuatro formatos de grabación para la transcripción (.wav, .mp4, .m4a, .mp3), lo que excluye otros formatos como los de código abierto, por ejemplo, la extensión de archivo .ogg.

Figura 1

Transcripción en Microsoft Word 365



AWS requiere de la carga de archivos en su propia nube dentro del repositorio Amazon S3 Bucket (activo durante 90 días) para poder luego transcribirlo mediante la aplicación Amazon Transcribe, que permite identificar voces y canales, pues especifica el número de voces a particionar. La transcripción separa voces, pero no tiene buena puntuación. Cada pausa marca el inicio, con mayúscula, de una nueva frase.

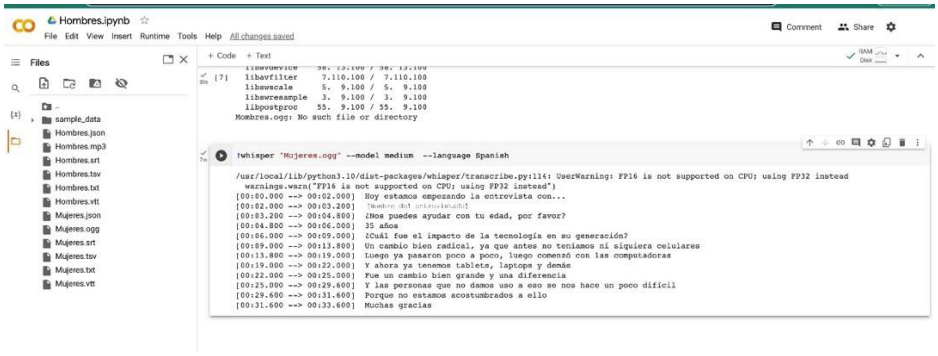
Notta no separa voces, pero crea diferentes bloques de acuerdo con las pausas del locutor o cambio de voz. Para los fines de esta investigación, la restricción de tiempo en versiones gratuitas no fue un problema, pero la versión gratuita solo admite grabaciones de hasta 3 minutos. El tiempo de subida de transcripción es el más corto de todos con una transcripción limpia y de alta calidad, con buena puntuación y ortografía.

Whisper es el más complicado de instalar y utilizar, puesto que requiere codificación de cada paso. No cuenta con una versión de aplicación propia, sino que está integrado a Colaboratory de Google Research. Además, es un programa con acceso abierto de la empresa OpenAI, que aún no ha desarrollado una interfaz pensada en la experiencia del usuario (UX). El tiempo de procesamiento es más largo que el de las otras IA, con un nivel medio de fidelidad. El nivel alto de fidelidad duplica la prolongada espera. Sin embargo,

la calidad de la transcripción es óptima, gramatical y ortográficamente, con la descarga de archivos con múltiples formatos (véase la Figura 2).

Figura 2

Transcripción de Whisper



La ausencia de UX de Whisper, unida a los tiempos largos de procesamiento, hace que, pese a la fidelidad del texto en relación con la grabación, sea difícil inclinarse por esta herramienta.

Además de la rapidez, otra utilidad de las plataformas con IA para la transcripción es la posibilidad de trabajar en línea, lo que permite la accesibilidad a los archivos indistintamente del lugar físico del investigador o comunicador. Nuevamente, el uso que se haga de la transcripción requerirá archivos de distinta índole. Para programación y *big data*, la descarga de la transcripción en formatos de datos semiestructurados, como .json (JavaScript Object Notation) o .xlm (Extensible Markup Language), permite manejar el flujo de datos con mayor facilidad (Lv et al., 2019). Para el trabajo periodístico y de investigación los documentos de texto con formato o sin él serían suficientes.

Como se observa en la Tabla 2, no existe una diferencia relevante en el timbre de voz para la transcripción cuando hay voces femeninas o mixtas; no obstante, los errores de identificación de voces se eliminan con el timbre de voz masculino. Ahora bien, en el análisis emprendido, las voces de los entrevistados presentaron errores en la transcripción, primordialmente por la falta de vocalización de las respuestas y no tanto por el empleo de modismos o regionalismos. Esto ocasionó un número de errores particularmente al inicio de la grabación en donde, sobre todo las voces femeninas, se mezclaron en todas las IA, salvo en Whisper.

Whisper, desde un inicio, genera locuciones diferenciadas para las dos voces como locuciones separadas, aunque el programa en sí no realiza identificación de voces como sí lo hacen Office y AWS. La transcripción de nombres propios y de apellidos generó el mayor número de errores en todas las transcripciones. Cabe notar que Whisper logró descifrar correctamente el apellido de una entrevistada, que incluso los investigadores

no habían conseguido transcribir puesto que era casi imposible escuchar la voz atenuada en las últimas letras del nombre. En cualquier caso, si bien existen errores en todas las IA, el número es siempre muy bajo.

La decisión de si la grabación debe transcribirse o no, dado que puede codificarse directamente y extraer solo el texto requerido desde el audio original, es un cuestionamiento recurrente en la literatura (McMullin, 2023; Nagy, 2014; Point & Baruch, 2023; Saldaña, 2016). A esto se une la futilidad de extensos documentos de transcripción a los que rara vez regresa el investigador, cuando lo esencial que responde a las preguntas de la investigación es marcado, subrayado, anotado y extraído en un primer paso de codificación. El segundo ciclo de codificación implica la generación de categorías, las cuales devienen de los códigos iniciales, por lo que regresar al texto original en la búsqueda de nuevas unidades raramente sucede. El análisis de contenido es una suerte de embudo en donde la riqueza del texto original se filtra durante el proceso que no retorna al punto inicial. Esto lo manifiesta el panel de investigadores expertos consultados durante la investigación.

La transcripción, como se ha recalcado, incorpora la impronta de quien realiza este proceso, sea un investigador, periodista, asistente o inteligencia artificial; integra puntuación, ritmo y pausas; inserta letras y completa palabras que la grabación no logró captar a cabalidad. En este sentido, si bien se busca al máximo la fidelidad de la transcripción con respecto a la entrevista, es también cierto que la díada de la entrevista se rompe cuando involucra la participación de un tercero, el cual, más que comprometer el sentido original del contenido, incorpora su presencia en el documento.

Al ser consultada sobre esta participación externa, una investigadora de la salud mencionó que, por el volumen de las entrevistas recabadas en su estudio, su equipo de investigación —constituido fundamentalmente por estudiantes universitarios— es el encargado de la transcripción. Por ello, la investigadora veía la necesidad de revisar, antes del análisis, la fidelidad del texto y, en algunos casos, corroborar con la grabación original aquellas partes que no presentaran un sentido claro. Tras esa revisión, el trabajo se desarrolla únicamente con las unidades elegidas y no con todo el documento transcrito.

El desempeño de una inteligencia artificial en el ámbito del periodismo y la comunicación no difiere significativamente en este contexto. Si el comunicador reconoce desde el principio la posibilidad de errores por parte de la IA, realizará una revisión exhaustiva antes del análisis, la codificación y la categorización. No obstante, la sustitución de un asistente de investigación por uno virtual implica, en parte, liberar a esa persona de la ejecución de tareas más bien mecánicas que aportan poco en términos de aprendizaje significativo. La inclusión de una IA en este proceso, por otro lado, acelera considerablemente el tiempo de generación de documentos escritos. Esta aceleración puede posibilitar

un enfoque más detallado en la extracción de unidades de análisis y codificación, lo que permite que en el ejercicio periodístico y de investigación, los equipos se sumerjan en análisis profundos y discusiones profesionales y académicas más enriquecedoras.

Es importante notar que la adopción tecnológica presenta limitaciones en cuanto a elementos de *hardware* y *software*: obsolescencia programada de dispositivos, actualizaciones de programas y permisos, cambios en las condiciones de uso, ajustes en el modelo de negocio *freemium*[2]. Las competencias digitales del equipo periodístico y de investigación pueden también presentar limitaciones respecto al uso, no solo de la aplicación, sino de actividades en los procesos de activación, carga y descarga de archivos, etcétera.

CONCLUSIONES

Los resultados de este estudio brindan una perspectiva para identificar la IA más adecuada en la búsqueda de transcripciones óptimas de entrevistas en español, en el ámbito periodístico y de la comunicación. Los datos sugieren que, aunque algunas IA pueden ofrecer una velocidad de procesamiento más rápida (eficiencia), esto no debe comprometer la fidelidad de la transcripción (eficacia). La elección de la plataforma debe equilibrar eficiencia y eficacia, así como considerar que debe completar las tareas como una acción mínima requerida, que satisfaga las necesidades del equipo periodístico o de investigación.

Se argumenta que un grado de transcripción de las entrevistas en investigación es siempre necesario. Contar con las letras plasmadas en el documento ofrece posibilidades de análisis inconcebibles solo con la grabación de audio o video. En el primer caso es posible poner en paralelo dos momentos de la locución y analizarlos simultáneamente. En audio o video, esta misma acción produciría imágenes y sonidos sobrepuestos, ininteligibles. La espacialidad, entonces juega a favor de la transcripción.

Pero, como se ha visto, optar por la transcripción completa del documento mediante el uso de IA puede ser de gran utilidad en la investigación. Esto no significa apuntalar el discurso apocalíptico de la pérdida de empleos por causa de la tecnología, al contrario, permite utilizar al máximo el potencial de las máquinas de análisis de datos. De este modo, se puede delegar las actividades propias de la inteligencia humana a los investigadores que son capaces de desmenuzar, dar sentido y abstraer importantes categorías de análisis, a partir de la riqueza de la experiencia humana que es compartida a través de las entrevistas.

2 La cuenta *freemium* se refiere a un uso gratuito de una herramienta hasta cierto nivel de capacidad o tiempo. Esto determina que si el usuario requiere mayor capacidad o tiempo de uso deberá pagar una suscripción.

Quizá uno de los elementos más significativos de este estudio radica en la capacidad de utilizar las IA en línea, lo que proporciona a los investigadores la flexibilidad de acceder a archivos y realizar transcripciones desde cualquier ubicación, lo que puede mejorar la eficiencia y la adaptabilidad en proyectos de investigación. Estos resultados contribuyen a la comprensión de cómo la IA puede ser aprovechada eficazmente en la investigación cualitativa, al tiempo que subraya la importancia de una elección informada del tipo de plataforma para la transcripción en función de las necesidades de velocidad y precisión en cada proyecto.

La evaluación de herramientas útiles para la investigación y el periodismo es un tema de permanente interés para el ejercicio profesional. Por ese motivo, se considera importante abordar estudios similares en un futuro aplicados a otros campos. Por ejemplo, en el doblaje se incorpora un nivel más de complejidad en el uso de IA, puesto que compromete la utilización de al menos dos lenguas con las particularidades que representa la identificación de la voz y la fidelidad en la traducción.

Para concluir, este estudio da luces sobre velocidad de procesamiento y posibilidades de uso de apenas cuatro de las muchas IA para transcripción que al momento tienen la capacidad de hacerlo en español. Sin lugar a duda, estas herramientas se seguirán perfeccionando y podrán sustituir a las tradicionales formas de transcripción y, entonces, tendrá más sentido el que la transcripción no sea un tema central en los manuales de metodologías cualitativas, porque lo principal no es el proceso mecánico de llevar el contenido de la entrevista al texto escrito, sino su análisis y el aporte del investigador.

CONFLICTOS DE INTERÉS

Los autores declaran no tener conflictos de interés.

CONTRIBUCIÓN DE AUTORES

Conceptualización: V. YR.; Curación de datos: V. YR.; Investigación, V. YR., J. CS.; Metodología, V. YR. y J. CS.; Escritura: V. YR. y J. CS.; Visualización: J. CS.; Revisión y Edición: V. YR.

REFERENCIAS

Angermuller, J., Maingueneau, D., & Wodak, R. (2014). *The Discourse Studies Reader: Main currents in theory and analysis*. John Benjamins Publishing Company. http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=800889&lang=es&site=ehost-live

Benkerzaz, S., Elmir, Y., & Dennai, A. (2019). A Study on Automatic Speech Recognition. *Journal of Information Technology Review*, *10*(3), 77-85. https://doi.org/10.6025/jitr/2019/10/3/77-85

Berkemer, R., & Grottke, M. (2023). Learning Algorithms. What is Artificial Intelligence Really Capable of? En P. Klimczak & C. Petersen (Eds.), *AI – Limits and Prospects of Artificial Intelligence* (pp. 9-42). Transcript Verlag. https://doi.org/doi:10.1515/9783839457320-003

Blecua, A. (1983). *Manual de crítica textual*. Castalia.

Castells, M. (2002). *The rise of the network society* (vol. 1). Blackwell.

Claeser, D., Pritzkau, A., Schade, U., & Winandy, S. (2023). Let's Fool That Stupid AI: Adversarial Attacks against Text Processing AI. En P. Klimczak & C. Petersen (Eds.), *AI – Limits and Prospects of Artificial Intelligence* (pp. 267-284). Transcript Verlag. https://doi.org/doi:10.1515/9783839457320-012

Cornejo, I., & Rufer, M. (2020). *Horizontalidad : hacia una crítica de la metodología.* CLACSO / CALAS. https://biblioteca.clacso.edu.ar/clacso/se/20201023034518/Horizontalidad.pdf

Covella, G. J. (2005). *Medición y evaluación de calidad en uso de aplicaciones web* [Tesis de maestría, Universidad Nacional de La Plata]. Repositorio Institucional de la UNLP. https://sedici.unlp.edu.ar/handle/10915/4082

Creswell, J.W. (2013). *Qualitative inquiry and research design: choosing among five approaches.* SAGE.

García-Prieto, V., & Figuereo-Benítez, J. C. (2022). Accesibilidad de los contenidos televisivos para personas con discapacidad: limitaciones y propuestas de mejora. *Contratexto*, (38), 289-311. https://doi.org/10.26439/contratexto2022.n038.5779

Girón-García, C., & Esbrí-Blasco, M. (2019). Analysing the Digital World and its Metaphoricity: Cybergenres and Cybermetaphors in the 21st Century. *Cultura, Lenguaje y Representación*, *22*, 21-35. https://doi.org/10.6035/CLR.2019.22.2

Greco, L., Galatolo, R., Horlacher, A. S., Piccoli, V., Ticca, A. C., & Ursi, B. (2019). Some theoretical and methodological challenges of transcribing touch in talk-in-interaction. *Social Interaction. Video-Based Studies of Human Sociality*, *2*(1). https://doi.org/10.7146/si.v2i1.113957

Kreuz, R. J., & Riordan, M. A. (2018). The art of transcription: Systems and methodological issues. In A. H. Jucker, K. P. Schneider & W. Bublitz (Eds.), *Methods in Pragmatics* (pp. 95-120). De Gruyter Mouton. https://doi.org/doi:10.1515/9783110424928-003

Kvale, S. (2011). *Las entrevistas en investigación cualitativa*. Morata.

Ligo, A. K., Rand, K., Bassett, J., Galaitsi, S. E., Trump, B. D., Jayabalasingham, B., Collins, T., & Linkov, I. (2021). Comparing the Emergence of Technical and Social Sciences Research in Artificial Intelligence. *Front. Comput. Sci.*, *3*, 1-13. https://doi.org/10.3389/fcomp.2021.653235

Lopezosa, C., Codina, L., & Boté-Vericad, J.-J. (2023). *Testeando ATLAS.ti con OpenAI: hacia un nuevo paradigma para el análisis cualitativo de entrevistas con inteligencia artificial*. Universitat Pompeu Fabra, Departamento de Comunicación. https://repositori.upf.edu/handle/10230/56449

Lv, T., Yan, P., & He, W. (2019). On Massive JSON Data Model and Schema. *Journal of Physics: ConferenceSeries*,*1302*(2),1-4.https://doi.org/10.1088/1742-6596/1302/2/022031

McMullin, C. (2023). Transcription and Qualitative Methods: Implications for Third Sector Research. *Voluntas*, *34*, 140-153. https://doi.org/10.1007/s11266-021-00400-3

Nagaraj, P., Muneeswaran, V., Rohith, B., Sai Vasanth, B., Veda Varshith Reddy, G., & Koushik Teja, A. (23-25 de enero de 2023). *Automated YouTube Video Transcription to Summarized Text Using Natural Language Processing*. 2023 International Conference on Computer Communication and Informatics (ICCCI). Institute of Electrical and Electronics Engineers, Coimbatore, India. https://doi.org/10.1109/ICCCI56745.2023.10128375

Nagy, N. (2014). Chapter 12: Transcription. En R. J. Podesva & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 235-256). Cambridge University Press. https://doi.org/10.1017/CBO9781139013734

O'Brien, S. (2020). *Translation, human-computer interaction and cognition*. Routledge.

O'Shaughnessy, D. (2024). Trends and developments in automatic speech recognition research. *Computer Speech & Language*, *83*, 1-33. https://doi.org/10.1016/j.csl.2023.101538

Ossa, F. (1993). *Historia de la escritura*. Planeta.

Point, S., & Baruch, Y. (2023). (Re)thinking transcription strategies: Current challenges and future research directions. *Scandinavian Journal of Management*, *39*(2), 1-10. https://doi.org/10.1016/j.scaman.2023.101272

Reales, L., Robalino, G., Peñafiel, A., Cárdenas, J., & Cantuña-Vallejo, P. (2022). El muestreo intencional no probabilístico: herramienta de investigación científica en carreras de Ciencias de la Salud. *Universidad & Sociedad*, *14*(S5), 681-691. https://rus.ucf.edu.cu/index.php/rus/article/view/3338

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3.ª ed.). SAGE.

Seifert, I., Bürger, M., Wangler, L., Christmann-Budian, S., Rohde, M., Gabriel, P., & Zinke, G. (2018). *Potential of Artificial Intelligence in Germany's Producing Sector*. PaiCE

Scientific Assistance. https://www.digitale-technologien.de/DT/Redaktion/EN/Downloads/Publikation/PAiCE_AI_Study.pdf?__blob=publicationFile&v=1

Silverman, D. (2013). *Doing qualitative research*. SAGE.

Waddell, K. (24 de agosto de 2022). *Lost in Transcription: Auto-Captions Often Fall Short on Zoom, Facebook, Google Meet, and YouTube*. Consumer Reports. https://www.consumerreports.org/disability-rights/auto-captions-often-fall-short-on-zoom-facebook-and-others-a9742392879/

Wagner, J. (2022). Conversation Analysis: Transcriptions and Data. En C. A. Chapelle (Ed.), *The Concise Encyclopedia of Applied Linguistics* (pp. 296-303). Wiley.