

La transcripción automática de archivos de audio y vídeo con Inteligencia Artificial (IA) para la mejora de la interacción oral

Automatic transcription of audio and video files with Artificial Intelligence (AI) to improve oral interaction

JORGE MARTÍ CONTRERAS¹
UNIVERSITAT JAUME I
jmarti@uji.es

Resumen:

La lingüística de corpus es empleada para conocer con más exactitud el uso real de la lengua, pero para ello se han de transcribir los documentos orales. Hoy en día existen corpus orales accesibles en línea, pero cada uno de ellos ha sido creado con una finalidad específica. Actualmente, disponemos de la Inteligencia Artificial generativa para ayudarnos a crear estos corpus orales. En este artículo se analizan 13 programas gratuitos que se pueden emplear para hacer transcripciones orales y se estudia su precisión.

Palabras clave: transcripción automática, inteligencia artificial, interacción oral, improvisación teatral

Abstract:

Corpus linguistics is used to gain a more accurate understanding of the actual use of language, but this requires the transcription of oral documents. Nowadays, there are oral corpora accessible online, but each of them has been created for a specific purpose. Currently, we have generative Artificial Intelligence to assist in the creation of these oral corpora. This article analyzes 13 free programs that can be used for oral transcriptions and examines their accuracy.

Keywords: automatic transcription, artificial intelligence, oral interaction, improv theater

¹ Este artículo ha sido elaborado dentro del marco del Grupo de Innovación Educativa en Español como Lengua Extranjera y del Grupo de Investigación de Español Como Lengua Extranjera de la Universitat Jaume I.

1. Marco teórico

La lingüística de corpus se emplea para análisis de elementos y recursos lingüísticos tanto de forma diacrónica como sincrónica y cada vez son más los fragmentos de corpus que se emplean en la enseñanza de lenguas extranjeras. Actualmente existen corpus de transcripción de textos reales orales tanto de Español como L1 (lengua nativa) como de LE/L2/LH (Lengua Extranjera, Segunda Lengua y Lengua de Herencia).

Algunos de los corpus orales en español (publicados en abierto y de acceso gratuito) más destacados de L1 son el corpus Val.Es.Co 3.0 (Valencia Español Coloquial) de la Universitat de València liderado por el Dr. Pons, ESLORA de la Universidad de Santiago de Compostela coordinado por la Dra. Vázquez, el CORLEC (Corpus Oral de Referencia del Español Contemporáneo) de la Universidad Autónoma de Madrid y el COREC (Corpus Oral de Referencia de Español en Contacto) capitaneado por la Dra. Palacios; estos cuatro corpus pueden tener algún rasgo diatópico. Por otro lado, el corpus COSER (Corpus Oral y Sonoro del Español Rural) dirigido por la Dra. Fernández-Ordóñez registra a hablantes de más de 73 años en ámbito rural. Existen varios proyectos panhispánicos como el PRESEEA (Proyecto para el Estudio Sociolingüístico del español de España y América) coordinado por el Dr. Moreno o AMERESCO (América y España español coloquial) también de la Universitat de València y dirigido por el Dr. Briz. Otros corpus que contienen tanto textos escritos como orales son CREA (Corpus de Referencia del Español Actual) y CORPES (Corpus del Español del siglo XXI), destacaremos el Corpus del español de Mark Davies con más de 5 millones de palabras de registro oral. En cuanto a corpus de LE/L2/LH existen corpus que solo contienen producciones escritas de aprendices de LE como CAES (Corpus de Aprendices de Español), a nosotros, por la naturaleza de esta investigación en la que se analiza la transcripción de textos orales, nos interesan más las producciones e interacciones orales, entre los que sobresale el *Columbia Corpus de Conversaciones para E/LE* (Columbia) dirigido por la Dra. Ruiz, diseñado para la práctica de E/LE (Español como Lengua Extranjera) los hablantes que aparecen en los vídeos son de L1. En cuanto a LH (Lengua de Herencia) encontramos el corpus *Spanish in Texas* de la Dra. Bullock y la Dra. Toribio. En el caso de querer tener en cuenta el sesgo de edad, se debe consultar el COLA (Corpus Oral de Lenguaje Adolescente) coordinado por la Dra. Jørgensen de la Universidad de Bergen (Noruega). Para finalizar esta introducción a los corpus, señalaremos el CORELE (Corpus Oral de Español como Lengua Extranjera) que es un proyecto del Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid en el que se recogen las hablas de estudiantes de niveles A1 y A2 y B1 y SPLLOC II (Spanish Learner Language Oral Corpora) coordinado por la Dra. Domínguez y en la que se incluyen investigadores de las universidades de Southampton, Newcastle y York.

Existen programas para realizar la transcripción, segmentación y añadir etiquetas, como por ejemplo *TranscriberAG*. El uso de estas herramientas es un primer paso antes de que los datos sean accesibles a través de búsquedas en KWIC o WORKBENCH.

El grupo GIELE (Grupo de Investigación de Español como Lengua Extranjera) de la Universitat Jaume I, está creando un corpus oral, COCREA-Game, basado en conversaciones orales espontáneas a partir de improvisaciones teatrales realizadas por hablantes de Español como L1. Para ello se han empleado herramientas de Inteligencia Artificial (IA) para transcribir el texto de forma automática.

2. Objetivo

El objetivo principal de este artículo es analizar las herramientas gratuitas que ofrece la Inteligencia Artificial (IA) para facilitar el proceso de transcripción a quienes investigan.

Los objetivos secundarios de esta transcripción son realizar un análisis lingüístico de corpus teniendo en cuenta que en trabajos posteriores se establecerá el nivel de lengua (Consejo de Europa, 2002 y 2020; Instituto Cervantes, 2006; ACTFL, 2015 y 2024) de los textos transcritos y se etiquetará el corpus obtenido siguiendo los estándares de otros corpus descritos en el apartado anterior.

3. Metodología

Para realizar el proceso de transcripción es necesario disponer de un archivo de audio y/o vídeo que haya sido seleccionado por ser de especial interés de estudio. En nuestro caso, como se quiere analizar el habla en interacciones orales espontáneas se ha seleccionado un fragmento de una obra de teatro improvisado, en el que el público propone un tema y los actores con apenas 20 segundos para crear una dramaturgia mínima salen al escenario para crear la historia de forma espontánea.

El fragmento seleccionado tiene por título *Foto, foto* y fue registrado con público y con permiso de la organización y de los actores. Tiene una duración total de 3 minutos y en él aparecen dos turistas en el Coliseo romano haciendo fotos y viendo apariciones de personajes no reales en las fotos (se puede consultar la transcripción completa en *Anexo I*). Para la grabación se empleó un teléfono móvil *Oneplus 2T 5G* con una cámara principal Sony IMX766 de 50 MP con la que se puede realizar grabaciones de vídeo y voz a 10 metros de distancia con calidad adecuada para ser transcritas.

Para la realización de la investigación que se presenta se ha usado en primer lugar un programa para eliminar ruidos de fondo en la grabación y acto seguido, se han empleado programas de IA de transcripción automática. Hemos transcrito siempre el mismo fragmento para poder realizar comparaciones de los resultados obtenidos. Hemos adaptado la

transcripción al sistema Val.Es.Co (Briz y Grupo Val.Es.Co, 2002) (*Anexo I*). Finalmente, ha sido calculado el porcentaje de precisión entre la transcripción de cada uno de los programas y la transcripción según el sistema Val.Es.Co (Briz y Val.Es.Co, 2002) a través de la página Prepostseo², esta página realiza una comparativa entre dos textos, verifica cuantitativamente la coincidencia entre ambos textos y resalta el texto que es exacto; por otro lado, se puede destacar que admite no solo texto, sino también diferentes tipos de archivo (*.doc, *.txt, *.pdf) o incluso URL de páginas web.

4. Programas de IA para la transcripción

A continuación, se analizan trece programas que se pueden emplear para la transcripción automática de textos.

4.1. Auphonic

Es una herramienta diseñada por Georg Holzmann para la edición, producción y mejora de pódcast. En la versión gratuita (hasta dos horas al mes) limpia el audio (elimina música y ruidos de fondo). Por otro lado, también tiene la opción de transcripción y editar resúmenes, pero es una versión de pago.

Es muy recomendable emplear este programa antes de iniciar la transcripción gratuita automática en otros programas, ya que el sonido es mucho más nítido, lo que permite obtener transcripciones más precisas.

Se puede encontrar en <https://auphonic.com/>

4.2. Otter.ai

Permite realizar hasta 300 min/mes de transcripciones y transcribe en una tercera parte de la duración del vídeo. Alinea el texto con el audio y vídeo. Actualmente solo realiza transcripción en inglés (de Reino Unido o de Estados Unidos), por lo que la transcripción que ofrece es excesivamente sesgada:

Speaker 1 (0:08). Yeah. Romano, Oh, wow. Woman,

Speaker 2 (0:22). Yeah, and who

(0:42). Look At

(0:48).Yes, Mas, MAs. Because

Speaker 2 (1:40). siempre. Senores, el coliseo, minutos cuenta, la

(1:44). quien OTOs estan previras

Speaker 2 (1:49). o esta Sora ahora, esta previred hace

(1:54). mucho tiempo que no se pueden, no se permitas

de

(1:57). todos a Que paso en este lugar.

(2:28). Mirabi.

² Se puede acceder a través de <https://www.prepostseo.com/es/text-compare>

El porcentaje de precisión en la transcripción realizada por *Otter.ai* con respecto al *Anexo I* es del 0 %.

Su página web es <https://otter.ai/>

4.3. Google Docs (Escribir por voz / *Voice Typing*)

Es un transcriptor de voz en tiempo real, no está diseñado para la transcripción de archivos de audio o de vídeo, pero, aun así, se pueden realizar transcripciones situando un altavoz cerca de un micrófono (esto hace que se pierda calidad en la transcripción). Como es evidente, no reconoce a los hablantes. En cambio, la calidad de la transcripción, a pesar de que se haga en un único párrafo es de suficiente calidad y permite ser editada a posteriori.

Este deseo romano wow aquí colgada porque soy un Boomer y vamos a hacer unas fotos no venía a Italia de vacaciones vale vale salir con el Coliseo del uy y eso qué es seguro que no había nadie son dos romanos ampliada por favor madre mía está mucho miedo levanta la mano derecha más más el pie izquierdo para atrás está frío señores el Coliseo va a cerrar en dos minutos no estarán haciendo fotos están prohibidas a estas horas hace mucho tiempo que no se pueden y no se permita hacer fotos aquí qué pasó en este lugar estás hablando no puedes hacer fotos si no hay nadie los macarrones al pesto dios mío su familia parece pero mira mira cómo le mandan los pechitos parece que estoy haciendo sin perder selfie yo saco.

Para activarlo primero se debe configurar el idioma en el que se va a registrar el audio (Archivo> Idioma) y después se debe activar el micrófono (Herramientas> Escribir por voz).

El porcentaje de precisión en la transcripción realizada por *Google Docs* con respecto al *Anexo I* es del 14,2 %.

El programa está disponible en <https://docs.google.com/>

4.4. Dictation (Voice Notepad)

Es un programa excesivamente limitado ya que tiene problemas de reconocimiento. Al igual que *Google Docs* no se puede abrir un archivo de audio o vídeo y necesita un micrófono externo, por lo que tiene problemas de reconocimiento de audio. El texto que es capaz de reconocer es el siguiente:

Ya entonces cada hora canción romano señores el Coliseo va a cerrar en 2 minutos.

El porcentaje de precisión en la transcripción realizada por *Dictation* con respecto al *Anexo I* es del 1,0 %.

Está disponible en línea en <https://dictation.io/speech>

4.5. Microsoft OneNote

Al igual que los dos programas anteriores (*Google Docs* y *Dictation.io*) necesita de un micrófono externo. Aunque es muy limitado, el resultado es mejor que el programa anterior, aunque omite fragmentos cuando no es capaz de procesarlos:

Guau Uy qué te juro que no había nadie cuando tú con todo vamos ampliada ampliada madre mía a ti no no no otra fíjate fíjate lo que hay en el bolsillo vamos tú y yo mira de bolsillo mira de botillo soy yo señores el corifeo va a cerrar en 2 minutos estarán haciendo fotos están prohibidas a estas horas ahora hace mucho tiempo que no se pueden no se permite hacer cosas aquí qué pasó en este lugar hablando no puedes hacer fotos no puedes hacer fotos si baila bien no no los macarrones al pesto lo entonces que yo lo vi mira aquí en tu familia pareja cómo levantar los pechitos guapa guapa

El porcentaje de precisión en la transcripción realizada por *OneNote* con respecto al *Anexo I* es del 6,0 %.

Accesible desde Windows 10 y 11, se debe entrar en *Microsoft OneNote* en una área para escribir (puede ser en *OneNote*, en *Google Docs*, en el *Bloc de notas*, en un correo electrónico...), a continuación se configura el idioma en que se va hablar (pulsar de forma simultánea las teclas de *Windows* y la *vocal I*) y posteriormente, cuando esté en el área de escritura se ha de activar el micrófono (pulsar de forma simultánea las teclas de *Windows* y la *consonante H*).

4.6. Microsoft 365

Es muy similar a las tres herramientas anteriores (*Google Docs*, *Dictation.io* y *OneNote*) la gran diferencia es que además de disponer el micrófono —con idénticas características—, se le puede cargar un archivo de vídeo o de audio sin necesidad de emplear el micrófono, por lo que la calidad de la transcripción aumenta considerablemente. Transforma el archivo de vídeo directamente en un archivo de audio en MP3 (esto dificulta la investigación si se centrara en la gestualidad y necesita una alineación).

Cuando finaliza la transcripción aparece en la parte derecha el texto transcrito con oradores y marcas de tiempo; se puede seleccionar que se inserte solo el texto en un nuevo documento, o también con los oradores y/o con oradores y marcas de tiempo. El programa no reconoce siempre a cada una de las personas que intervienen (de hecho, parece que asigna principalmente a un hablante las producciones de los tres participantes), pero es un buen punto de partida ya que se puede editar el texto obtenido muy fácilmente, además el texto está alineado con el audio:

00:00:00 **Orador** Dile.

00:00:00 **Orador** O K espera, espera, espera.

00:00:03 **Orador** Ya mi cuenta es para algo.
00:00:08 **Orador** Mira qué chulo.
00:00:10 **Orador** Me convenció. Romano, Oh, Mira, llegó la Cámara aquí colgada.
00:00:15 **Orador** Porque soy un bumen.
00:00:19 **Orador 1** ¿Vamos a tener unas botas, no?
00:00:20 **Orador 1** Sí que parece que venía Tania de vacaciones.
00:00:25 **Orador 1** Dani, Dani, a ver a romano.
00:00:27 **Orador 1** Pero Ponte tú, Ponte tú para salir con el Coliseo.
00:00:38 **Orador 1** ¿Uy, Ah, y esto qué es?
00:00:41 **Orador 1** Esto es, te juro que no había nadie.
00:00:43 **Orador 1** He hecho un fútbol.
00:00:46 **Orador 1** Son dos romanos, ampliala, ampliala.
00:00:53 **Orador 1** ¿Sí que la *****, madre mía, da mucha miedo a ti, no mames, porque Ponte un gol mío, Eh? Levanta la mano derecha más más el pie izquierdo para atrás.
00:01:07 **Orador** Guau.
00:01:09 **Orador 1** Ostras, he dicho, osea.
00:01:15 **Orador 1** Han frío, han frío, dale, han frío.
00:01:21 **Orador 1** Espera, Fíjate, Fíjate lo que hay en el bolsillo. Hago zoom, solo el bolsillo, mira el bolsillo, mira en el bolsillo si soy yo.
00:01:33 **Orador 1** Madre mía, de mí me vas en el bolsillo, tío, es que yo soy feo de nacimiento.
00:01:38 **Orador 1** Yo yo un mensaje amplio, no me está, señores.
00:01:43 **Orador 1** ¿El Coliseo va a cerrar en 2 minutos, lo estarán haciendo?
00:01:49 **Orador 1** Otros están prohibidas a estas horas, ahora están prohibidas.
00:01:54 **Orador 1** Hace mucho tiempo que no se pueden y no se permite hacer fotos aquí.
00:01:58 **Orador 1** ¿Qué pasó en este lugar? ¿De qué estás hablando?
00:02:00 **Orador** No.
00:02:02 **Orador** Hay.
00:02:05 **Orador 1** Ninguna voz tú, no me sé, señor, aquí.
00:02:10 **Orador 1** No me estaba aquí.
00:02:12 **Orador 1** Acuérdate, no puedes hacer fotos, no puedes hacer fotos.
00:02:16 **Orador 1** Si no hay nadie.
00:02:20 **Orador 1** Mira, voy a hacer una foto.
00:02:21 **Orador 1** Aquí vas en tomar los macarrones al pesto.
00:02:26 **Orador 1** Dios mío es ese que yo lo vi, mira esta vez su familia, parece.
00:02:32 **Orador 1** Pero mira, mira tú Cómo levanta los pechitos, parece que estoy haciendo un gato caca.
00:02:39 **Orador 1** ¿Y cuál es tu que sabe que vamos a celebrar? Echa aquí con ellos sefi, yo salgo de odio, de la cuesta.

El porcentaje de precisión en la transcripción realizada por *Microsoft 365* con respecto al *Anexo I* es del 18,0 %.

Si se tiene usuario, se puede encontrar en <https://www.office.com/y> a continuación acceder a Dictar> Transcribir > Cargar audio.

4.7. Descript

Es muy similar a las tres herramientas anteriores (*Google Docs*, *Dictation.io* y *OneNote*) la gran diferencia es que además de disponer el micrófono —con idénticas características—, se le puede cargar un archivo de vídeo o de audio sin necesidad de emplear el micrófono, por lo que la calidad de la transcripción aumenta considerablemente. Transforma el archivo de vídeo directamente en un archivo de audio en MP3 (esto dificulta la investigación si se centrara en la gestualidad y necesita una alineación).

Cuando finaliza la transcripción aparece en la parte derecha el texto transcrito con oradores y marcas de tiempo; se puede seleccionar que se inserte solo el texto en un nuevo documento, o también con los oradores y/o con oradores y marcas de tiempo. El programa no reconoce siempre a cada una de las personas que intervienen (de hecho, parece que asigna principalmente a un hablante las producciones de los tres participantes), pero es un buen punto de partida ya que se puede editar el texto obtenido muy fácilmente, además el texto está alineado con el audio:

00:00:00 **Orador** Dile.
00:00:00 **Orador** O K espera, espera, espera.
00:00:03 **Orador** Ya mi cuenta es para algo.
00:00:08 **Orador** Mira qué chulo.
00:00:10 **Orador** Me convenció. Romano, Oh, Mira, llegó la Cámara aquí colgada.
00:00:15 **Orador** Porque soy un bumén.
00:00:19 **Orador 1** ¿Vamos a tener unas botas, no?
00:00:20 **Orador 1** Sí que parece que venía Tania de vacaciones.
00:00:25 **Orador 1** Dani, Dani, a ver a romano.
00:00:27 **Orador 1** Pero Ponte tú, Ponte tú para salir con el Coliseo.
00:00:38 **Orador 1** ¿Uy, Ah, y esto qué es?
00:00:41 **Orador 1** Esto es, te juro que no había nadie.
00:00:43 **Orador 1** He hecho un fútbol.
00:00:46 **Orador 1** Son dos romanos, ampliala, ampliala.
00:00:53 **Orador 1** ¿Sí que la *****, madre mía, da mucha miedo a ti, no mames, porque Ponte un gol mío, Eh? Levanta la mano derecha más más el pie izquierdo para atrás.
00:01:07 **Orador** Guau.
00:01:09 **Orador 1** Ostras, he dicho, osea.
00:01:15 **Orador 1** Han frío, han frío, dale, han frío.
00:01:21 **Orador 1** Espera, Fíjate, Fíjate lo que hay en el bolsillo. Hago zoom, solo el bolsillo, mira el bolsillo, mira en el bolsillo si soy yo.
00:01:33 **Orador 1** Madre mía, de mí me vas en el bolsillo, tío, es que yo soy feo de nacimiento.
00:01:38 **Orador 1** Yo yo un mensaje amplio, no me está, señores.

00:01:43 **Orador 1** ¿El Coliseo va a cerrar en 2 minutos, lo estarán haciendo?
00:01:49 **Orador 1** Otros están prohibidas a estas horas, ahora están prohibidas.
00:01:54 **Orador 1** Hace mucho tiempo que no se pueden y no se permite hacer fotos aquí.
00:01:58 **Orador 1** ¿Qué pasó en este lugar? ¿De qué estás hablando?
00:02:00 **Orador** No.
00:02:02 **Orador** Hay.
00:02:05 **Orador 1** Ninguna voz tú, no me sé, señor, aquí.
00:02:10 **Orador 1** No me estaba aquí.
00:02:12 **Orador 1** Acuérdate, no puedes hacer fotos, no puedes hacer fotos.
00:02:16 **Orador 1** Si no hay nadie.
00:02:20 **Orador 1** Mira, voy a hacer una foto.
00:02:21 **Orador 1** Aquí vas en tomar los macarrones al pesto.
00:02:26 **Orador 1** Dios mío es ese que yo lo vi, mira esta vez su familia, parece.
00:02:32 **Orador 1** Pero mira, mira tú Cómo levanta los pechitos, parece que estoy haciendo un gato caca.
00:02:39 **Orador 1** ¿Y cuál es tu que sabe que vamos a celebrar? Echa aquí con ellos sefi, yo salgo de odio, de la cuate.

El porcentaje de precisión en la transcripción realizada por *Microsoft 365* con respecto al *Anexo I* es del 18,0 %.

Si se tiene usuario, se puede encontrar en <https://www.office.com/y> a continuación acceder a Dictar> Transcribir > Cargar audio.

4.8. Google Pinpoint

Esta herramienta fue desarrollada para facilitar el trabajo de periodistas, pues después de subir el audio y procesarlo, ofrece la transcripción e indica la referencia temporal del inicio de cada fragmento. Para transcribir fragmentos de obras de teatro no es el mejor programa, pero sí que es recomendable para entrevistas de larga duración en el que se quiera tener localizado el fragmento de vídeo alineado al texto.

► 00:00

La cámara aquí colgada porque soy un Boomer y nos vamos a hacer unas fotos, no sí que parece que venía a Italia de vacaciones.

► 00:34

Son dos romanos, levanta la mano derecha más más el pie izquierdo amplio.

► 01:20

Fíjate, fíjate lo que hay en el bolsillo. Hago bolsillo. Mira en el bolsillo. Mira en el bolsillo, si soy yo, madre mía, qué miedo das un bolsillo, tío. Es que yo soy feo de nacimiento, señores. El Coliseo va a cerrar en dos minutos no estarán haciendo fotos, claro, otros están prohibidas a estas horas. Ahora está prohibidas. Hace mucho tiempo que no se pueden y no se permite hacer fotos aquí, qué pasó en este lugar y quién estás hablando?

► 02:04

Fíjate, fíjate lo que hay en el bolsillo. Hago bolsillo. Mira en el bolsillo. Mira en el bolsillo, si soy yo, madre mía, qué miedo das un bolsillo, tío. Es que yo soy feo de nacimiento, señores. El Coliseo va a cerrar en dos minutos no estarán haciendo fotos, claro, otros están prohibidas a estas horas. Ahora está prohibidas. Hace mucho tiempo que no se pueden y no se permite hacer fotos aquí, qué pasó en este lugar y quién estás hablando?

► 02:04

No puedes hacer fotos, no puedes hacer fotos si no hay nadie los macarrones, al pesto dios mío, es ese que yo lo he visto su familia los pechitos, parece que estoy haciendo un Waka Waka Leticia Sabater con ellos un selfie yo saco los adicionales.

El porcentaje de precisión en la transcripción realizada por *Google Pinpoint* con respecto al *Anexo I* es del 19,0 %.

Accesible en <https://journaliststudio.google.com/pinpoint/>

4.9. Youtube (Automatic Captions)

Cuando se sube un vídeo a la plataforma de *Youtube*, el sistema te ofrece la posibilidad de añadir subtítulos —pueden ser subtítulos ya creados en un archivo que se debe subir, subtítulos creados con la opción de *Auto-sync* de forma automática, escritos de forma manual o creados en español por Youtube—. El programa no identifica a los participantes cuando se trata de vídeos con varias personas. Existe mucha precisión en la transcripción (con un único hablante la precisión en la transcripción puede llegar a ser hasta el 100%).

Un, Mira qué chulo el Coliseo
romano Wow Mira llevo la cámara aquí
colgada porque soy un
Boomer vamos a hacer unas fotos no que
para venido a Italia de vacaciones Dale
dale a Pero ponte tú ponte tú para salir
con el
Uy Ah Y eso qué
es está Te juro que no había nadie
cuando hecho una
foto son dos romanos amplía

amplí s que da miedo madre mía da mucho
miedo te voy a hacer yo un a ti ponte
ponte ponte
Levanta la mano derecha más más el pie
izquierdo tíralo para atrás eh eh saca
la
lengua
ostras he dicho ostras
solo amplio amplio
amplio espera fíjate fíjate lo que hay
en el bolsillo hago zoom solo al
bolsillo mira en el bolsillo mira en el
bolsillo si soy yo
y madre mía Qué miedo ser un bolsillo
tío Es que yo soy feo de
nacimiento con un mensaje
ampl señores el coliseo va a cerrar en
dos minutos No estarán haciendo fotos
las fotos están prohibidas a estas horas
ahora están
prohibidas Hace mucho tiempo que no se
pueden y no se permite hacer fotos aquí
Qué pasó en este lugar Con quién estás
hablando una voz tú no ves un señor
aquí no ves Nada aquí acuérdate no
puedes hacer fotos no puedes hacer fotos
si no hay
nadie Mira voy hacer una foto hacia aquí
me sent mal los macarrones al
pesto Dios mío es ese que yo lo he mirao
su familia
parece Pero mira mira cómo levanda los
pechitos parece que estoy haciendo guaca
guaca y pare leticias a
bater yfi con ellos sefi yo saco los
lados de la muerte muerte

Las transcripciones generadas (se pueden editar en la plataforma) pueden ser descargadas desde el apartado Studio>Subtitles, también se permite editar la sincronización del texto y el vídeo.

El porcentaje de precisión en la transcripción realizada por *Youtube* con respecto al *Anexo 1* es del 3,0 %. En muchas ocasiones con un único hablante la fiabilidad de precisión ronda el 100 %.

Está disponible en <https://studio.youtube.com/>

4.10. Veed.io

Es una herramienta de inteligencia artificial que necesita un tercio de la duración del vídeo para crear los subtítulos. Se puede elegir el idioma y

además la variedad (español de España, Honduras o Cuba). También se pueden insertar de forma automática la traducción de las transcripciones en más de 100 idiomas.

Es una herramienta de transcripción automática que es gratuita, pero con algunas limitaciones (sí que están disponibles en la versión de pago). Por ejemplo, se obtiene la transcripción generada automáticamente, pero no se puede descargar el texto completo —aunque como cada frase se puede seleccionar, cortar y pegar, si el investigador lo desea, se genera poco a poco el texto en otro documento—. Tampoco es posible descargar el vídeo de forma gratuita (ni con los subtítulos en el idioma original ni con las traducciones automáticas). Por otro lado, una mejora respecto a otros programas es que aparecen en color naranja aquellas palabras susceptibles de no estar bien transcritas.

Mira qué chulo.
El coliseo romano.
Mira, llevo la cámara aquí
colgada porque soy un boomer.
Dale, dale. Pero ponte tú.
Ponte tú para salir con el Coliseo.
¿Y eso qué es?
Te juro que eno había nadie cuando he hecho una foto
Son dos romanos
Amplíala, amplíala.
Madre mía, da mucho miedo. Te voy a hacer.
Levanta la mano derecha. Más, más.
El pie izquierdo tirado para atrás.
Osstrás.
he dicho ostras.
Solo ha peleado amplio.
Espera, fíjate.
Fíjate lo que hay en el bolsillo.
Hago zoom solo al bolsillo.
Mira en el bolsillo.
Mira en el bolsillo
Si soy yo.
Madre mía, qué miedo.
Va a ser un bolsillo, tío.
Es que yo soy feo de nacimiento.
Lleva con un mensaje. Amplía.
Lleva un estadio.
Señores, el coliseo va a cerrar en dos minutos.
¿No estarán haciendo fotos?
Las fotos están prohibidas a estas horas
Ahora están prohibidas.
Hace mucho tiempo que no se pueden
y no se permite hacer fotos aquí.
¿Qué pasó en este lugar?

¿Con quién estás hablando?
Tú no ves el señor que no ves **nada**.
Aquí no puedes hacer fotos si no hay nadie.
Mira, voy a hacer una foto hacia aquí. **Más**.
Entonces los macarrones al pesto.
Dios **mío**.
Es ese que yo lo he visto. Mira.
Está **viendo** su familia. Parece.
Pero **mira**, mira cómo levanta los pechitos.
Parece que estoy haciendo guaca.
Guaca sin parecer. **He** dicho Sabater.

El porcentaje de precisión en la transcripción realizada por *Veed.io* con respecto al *Anexo I* es del 3,0%.

Se puede acceder a través de <https://www.veed.io/>

4.11. Happyscribe

Es un programa que ofrece, según indican en su página, una transcripción con un 85% de precisión (ponen a disposición de los usuarios una transcripción realizada por una persona, en lugar de por IA, con un pago de 1,75 €/min y con la precisión del 99%)

Si se quiere bajar la transcripción se debe realizar un pago, pero, al igual que en *Veed.io* se pueden cortar y pegar fragmentos de transcripción en un nuevo documento. Diferencia entre hablantes —aunque no siempre coincide el hablante con el fragmento asignado— y además realiza alineación del texto con el vídeo.

Hablante 1 (00:08) Mira qué chulo. El coliseo romano. Wow. Mira, llevo la cámara aquí colgada porque soy un boomer. ¿Vamos a hacer unas fotos, no? Sí que parece por venir a Italia de vacaciones.

Hablante 2 (00:24) Dale, dale.

Hablante 1 (00:26) Pero ponte tú. Ponte tú para salir con el coliseo detrás.

Hablante 3 (00:39) ¿Y eso qué es?

Hablante 4 (00:41) ¿Esto?

Hablante 3 (00:42) Te juro que no había nadie cuando.

Hablante 4 (00:43) He hecho una foto.

Hablante 5 (00:46) Son dos romanos. Amplíala, amplíala. Madre mía, da mucho miedo.

Hablante 4 (00:57) Te voy a hacer un grati.

Hablante 5 (01:00) Levanta la mano derecha. Más, más.

Hablante 4 (01:03) El pie izquierdo tirado para atrás. Ostras.

Hablante 5 (01:11) He dicho ostras.

Hablante 1 (01:12) Solo.

Hablante 4 (01:16) Ha peleado. Amplio, amplio. Espera, fíjate. Fíjate lo que hay en el bolsillo.

Hablante 5 (01:24) Hago zoom solo al bolsillo. Mira el de bolsillo. Mira en el bolsillo. Si soy yo.

Hablante 4 (01:33) Madre mía, qué miedo.
Hablante 3 (01:34) Va a ser un bolsillo, tío.
Hablante 4 (01:35) Es que yo soy feo de nacimiento.
Hablante 3 (01:39) Aquí va con un mensaje. Amplía.
Hablante 2 (01:42) Señores, el coliseo va a cerrar en 2 min. ¿No estarán haciendo fotos? Las fotos están prohibidas a estas horas. Ahora están prohibidas. Hace mucho tiempo que no se pueden y no se permite hacer fotos aquí.
Hablante 3 (01:58) ¿Qué pasó en este lugar?
Hablante 5 (02:00) ¿De quién estás hablando? Tú no ves el señor aquí.
Hablante 3 (02:10) No ves nada.
Hablante 2 (02:11) Aquí no puedes hacer fotos. No puedes hacer fotos si no hay nadie.
Hablante 5 (02:19) Mira, voy a hacer una foto hacia aquí.
Hablante 4 (02:22) Más.
Hablante 3 (02:22) Entonces los macarrones al pesto.
Hablante 5 (02:26) Dios mío.
Hablante 1 (02:27) Es ese que yo lo he visto.
Hablante 3 (02:29) Mira. Está viendo su familia.
Hablante 4 (02:30) Parece.
Hablante 5 (02:32) Pero mira, mira cómo levanta los pechitos.
Hablante 4 (02:36) Parece que estoy haciendo guaca guaca y parece Letizia se va a material y.
Hablante 3 (02:42) Se fi con ellos.
Hablante 5 (02:44) Sefi, yo saco los ojos de la hueste.

El porcentaje de precisión en la transcripción realizada por *Happyscribe* con respecto al *Anexo I* es del 18,0 %.

El programa está disponible en <https://www.happyscribe.com>

4.12. Speechnotes

Este programa permite transcribir hasta un máximo de 50 min en la versión gratuita. Puede realizar la transcripción en múltiples idiomas. La transcripción se realiza en la mitad de tiempo que dura la grabación. Detecta automáticamente a los hablantes (después de decirle previamente cuánto son) y transcribe a cada hablante con un color.

Mira.

Qué chulo.

El Colegio romano, Mira, llevo la Cámara aquí colgada porque soy un. A Italia de vacaciones.

Dale, dale a ver.

Pero Ponte tú, Ponte tú para salir con el Coliseo.

¿Y eso qué es? Te juro que no había nadie cuando eché una foto.

Son dos romanos ampliada, ampliada, madre mía.

El Coliseo va a cerrar en 2 minutos. ¿No estarán haciendo fotos? Las fotos están prohibidas a estas horas. Ahora están prohibidas. Hace mucho tiempo que no se pueden y no se permite hacer fotos aquí.

¿Qué pasó en este lugar?

De.

¿Quién estás hablando, tu no ves el señor, aquí no ves nada aquí?

Acuérdate, no puedes hacer fotos.

No puedes.

Hacer.

Sí pues, dice Sabater, todo lo hace con ellos, sefi, yo saco el rato de la muerte.

El porcentaje de precisión en la transcripción realizada por Speechnotes con respecto al *Anexo I* es del 16,0%.

Se puede acceder a esta herramienta en <https://speechnotes.co/>

4.13. Klang.ai

Es la herramienta más avanzada de las que se han presentado hasta el momento, además de asignar el turno de palabra a cada hablante con mayor exactitud y tener mayor precisión en la transcripción, añade signos de puntuación, de interrogación y exclamación de manera adecuada. El mayor defecto es el número limitado de idiomas que se puede elegir para realizar la transcripción (entre los que se encuentra el español).

Por otro lado, realiza un resumen en inglés del vídeo a partir de lo que interpreta del texto, tal y como lo haría el público, por ejemplo, en la transcripción no aparece la palabra “fantasma” pero sí, y de forma adecuada, en el resumen:

Speaker 1 is taking a photo at the Colosseum and notices something unusual in the picture. They exclaim that there are two “romanos” (Romans) in the picture, even though they didn't see anyone when they took the photo. They ask others to zoom in on the picture to see what it is. When they do, they are startled and make jokes about it, saying they will create a “gratis” (free) ghost photo. They continue to make jokes and tell each other to pose in certain ways for more ghost photos.

At one point, Speaker 1 realizes that the “ghost” appears to have something in its pocket. They tell the others to zoom in on the pocket and are surprised to see what appears to be a message. Speaker 1 then reminds the group that photography is not allowed at the Colosseum at this time and that they could get in trouble for taking pictures.

Despite this warning, Speaker 1 continues to take pictures and point out strange things in them, such as a person who looks like a celebrity and a creature that resembles Sefi, a character from a Spanish TV show. The others in the group continue to express surprise and excitement as Speaker 1 points out these strange sights in the photos.

Otra mejora que ofrece es la *sugerencia cambios* o bien en la transcripción o bien en la asignación del hablante, y el porcentaje de fiabilidad de la transcripción en cada una de las intervenciones de los hablantes, aquí hay una muestra de las 29 sugerencias que ofrecía:

1. Speaker 1 0:30	22%
¡Ah!	
2. Speaker 1 0:39	86%
¿Y eso qué es? ¿Esto es...? Te juro que no había nadie cuando eché la foto.	
3. Speaker 1 0:45	85%
Son dos romanos. ¡Amplíala! ¡Amplíala!	
4. Speaker 1 0:52	70%
¡Sí que da miedo!	
5. Unknown 0:55	79%
¡Sí! ¡Madre mía!	
6. Speaker 1 0:56	77%
Da mucho miedo. ¡Te voy a hacer yo un gratis! ¡Ponte! ¡Ponte!	
7. Unknown 0:59	100%
¡Ponte!	
8. Speaker 1 1:00	84%
¡Levanta la mano derecha!	

Se puede descargar el texto completo de forma gratuita, pero en formato TXT y sin que se muestren los hablantes ni la temporización (sí que están disponibles en la versión de pago, así como la descarga del archivo en otros formatos como en formato de tabla CSV, de texto DOC o de subtítulos SRT).

¡Ah! ¿Y eso qué es? ¿Esto es...? Te juro que no había nadie cuando eché la foto. Son dos romanos. ¡Amplíala! ¡Amplíala! ¡Sí que da miedo!
¡Sí! ¡Madre mía!
Da mucho miedo. ¡Te voy a hacer yo un gratis! ¡Ponte! ¡Ponte!
¡Ponte!
¡Levanta la mano derecha!
¡Más! ¡El pie izquierdo!
¡Tira lo para atrás! ¡Sacad la lengua! ¡Ostras! ¡He dicho ostras solo!
¡Amplío! ¡Amplío! ¡Amplío!
¡Espera! ¡Fíjate! ¡Fíjate lo que hay en el bolsillo! ¡Hago zoom solo al bolsillo! ¡Mirad en el bolsillo!

¡Mirad en el bolsillo!
¡Si soy yo! ¡Madre mía!
¡Qué miedo!
¡Va a ser un bolsillo, tío! ¡Es que yo soy feo de nacimiento! Aquí va con un mensaje, Amplía. Lleva un mensaje. Señores. El Coliseo va a cerrar en dos minutos. No estarán haciendo fotos. Las fotos están prohibidas a estas horas. Ahora están prohibidas. Hace mucho tiempo que no se pueden y no se permite hacer fotos aquí. ¿Qué pasó en este lugar? ¿Con quién estás hablando? ¿Tú no ves el señor aquí? ¿No ves nada aquí? Acuérdate. No puedes hacer fotos. No puedes hacer fotos. ¡Si no hay nadie! Mira. Voy a hacer una foto. Hacia aquí. Más el tamal, los macarrones al pesto. Dios mío. ¡Es ese! ¡Que yo lo he visto! ¡Mira! ¡Su familia aparece! Mira cómo levanta los

pechitos. Parece que esté haciendo un guaca guaca. ¡Y parece Leticia Sabater!
¡Dios mío! Y Sefi con ellos.
¡Sefi! ¡Yo salgo!
¡Sefi de la muerte! ¡Sorte!

El porcentaje de precisión en la transcripción realizada por *Klang.ai* con respecto al *Anexo I* es del 29,0%.

Accesible en <https://klang.ai/>

5. Conclusiones

Es necesario que se sigan desarrollando nuevas herramientas de IA, pues todavía existen muchos errores en la transcripción automática. Investigadores como Bazán Gil *et al.* (2021: 6) hablan de la tasa de error por palabra y consideran aceptable una transcripción con un máximo del 20 % de errores (que está muy alejado del casi 30 % de acierto que hemos observado con *Klang.ai*, aunque eso sí, se debe tener en cuenta que la comparación que hemos realizado ha sido del texto transcrito original y el editado según el sistema de transcripción en el que incluyen símbolos).

En el caso de los medios de comunicación, pueden emplear la lingüística de corpus para realizar la transcripción, cortado, categorización (o etiquetado) de textos y la distribución de información según la necesidad individual de cada usuario; en el campo de la lingüística aplicada a la enseñanza de idiomas se necesita la lingüística de corpus para la transcripción, cortado, nivelación (A1, A2, B1, B2, C1 o C2) de cada fragmento y etiquetado del texto para estudiar elementos pragmáticos o lingüísticos, y mejorar, por ejemplo, la interacción oral.

Entre los programas gratuitos analizados se puede realizar una clasificación en cinco grupos: (1) programas para eliminar ruidos de fondo y que sea más nítido el audio para poderlo emplear en el resto de programas (*Auphonic*), (2) programas con uso de micrófono como elemento auxiliar (*Google docs*, *Dictation*, *OneNote* y *Microsoft 365*) todos ellos con un muy bajo índice de fiabilidad en la transcripción, a excepción de *Microsoft 365* ya que permite también subir directamente el archivo en vídeo o audio para que sea transcrito, (3) programas sin detección de turnos de habla (*Descript*, *Google Pinpoint*) con una media del 16,5 % de correspondencia, frente al 3 % de otros programas (*Youtube Automatic Captions* y *Veed.io*), (4) programas con identificación de los hablantes y turnos de habla (*Happyscribe* y *Speechnotes*) con el 17 % de similitud, y, finalmente, (5) programas de transcripción sin identificación de hablantes, pero con la inclusión de signos de puntuación (*Klang.ai*), por lo que se alcanza el 29 % de similitud.

Por otro lado, y de cara a futuros trabajos, sería conveniente que se creen nuevas etiquetas de textos más simples que las que proporciona ELAN (Carcelén y Uclés, 2019: 27) para que se pueda alinear el texto con los ítems del MCER (Consejo de Europa, 2002) y ACTFL (2015 y 2024), para

así abordar de una manera más efectiva el proceso de enseñanza-aprendizaje de la lengua.

6. BIBLIOGRAFÍA

- AMERICAN COUNCIL ON THE TEACHING OF FOREIGN LANGUAGES (ACTFL) (2015 [2012]): Performance Descriptors for Language Learners. https://www.actfl.org/uploads/files/general/ACTFLPerformance_Descriptors.pdf
- AMERICAN COUNCIL ON THE TEACHING OF FOREIGN LANGUAGES (ACTFL) (2024): ACTFL Proficiency Guidelines. https://www.actfl.org/uploads/files/general/Resources-Publications/ACTFL_Proficiency_Guidelines_2024.pdf
- BAZÁN-GIL, V.; PÉREZ-CERNUDA, C.; MARROYO-NÚÑEZ, N.; SAMPEDRO-CANET, P.; DE-IGNACIO-LEDESMA, D. (2021): "Inteligencia artificial aplicada a programas informativos de radio. Estudio de caso de segmentación automática de noticias en RNE". *Profesional de la información*, núm. 30 (33), pp. 1-12. Recuperado de <https://doi.org/10.3145/epi.2021.may.20>
- BRIZ GÓMEZ, Antonio y Grupo Val. Es.Co (2002): "Sistema de transcripción". *Corpus de conversaciones coloquiales*. Madrid: Arco Libros. <https://www.uv.es/valesco/sistema.pdf>
- BRIZ GÓMEZ, Antonio y ALBELDA MARCO, Marta (2009): "Estado actual de los corpus de la lengua española hablada y escrita: I+D". *Anuario del Instituto Cervantes*. Madrid: Instituto Cervantes, pp. 165-226.
- CARCELÉN GUERRERO, Andrea y UCLÉS RAMADA, Gloria (2019): "Corpus y construcción de un corpus oral multidialectal. El corpus Ameresco". *Normas*, 9, 17-36. <https://doi.org/10.7203/Normas.v9i1.16007>
- CONSEJO DE EUROPA [2002 (2001)]: *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: MECD- Anaya. Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/
- CONSEJO DE EUROPA [2020 (2018)]: *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación. Volumen complementario*. Estrasburgo: Servicio de publicaciones del Consejo de Europa. Descargado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco_complementario/mcer_volumen-complementario.pdf
- GARCÍA RUIZ-CASTILLO, Carlos (2024): "Corpus orales de aprendientes de ELE para el estudio de la conversación en la L2 disponibles en Internet". *Hiroshima Studies in Language and Language Education*, 37, 117-132.
- INSTITUTO CERVANTES (2006): *Plan curricular del Instituto Cervantes. Niveles de referencia para el español*. Madrid: Instituto Cervantes- Biblioteca nueva. En https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/

CORPUS CITADOS

- BULLOCK, Barbara E. y TORIBIO, Almeida Jackqueline (dirs.) (2024): *Spanish in Texas Corpus*, V1. <https://doi.org/10.18738/T8/GBYLBX>
- DAVIES, Mark (2024): *Corpus del Español de Mark Davies*. Brigham Young University. <https://www.corpusdelespanol.org/>
- DOMÍNGUEZ, Laura (2010): *Spanish Learner Language Oral Corpora (SPLLOC 2)*. University of Southampton, University of Newcastle y University of York. <https://web-archive.southampton.ac.uk/www.splloc.soton.ac.uk/>

- ESLORA (2023): *Corpus para el estudio del español oral (ESLORA)*. Universidad de Santiago de Compostela. <http://eslora.usc.es>, versión 2.2 de noviembre de 2023, ISSN. 2444-1430.
- FERNÁNDEZ ORDÓÑEZ, Inés (2024): *Corpus Oral y Sonoro del Español Rural (COSER)*. Universidad Autónoma de Madrid. <http://www.corpusrural.es/>
- JØRGENSEN, Annette Myre (2023): *Corpus Oral de Lenguaje Adolescente (COLA)*. <https://blogg.hiof.no/colam-esp/>
- MORENO FERNÁNDEZ, Francisco (dir.) (2024): *Proyecto para el estudio Sociolingüístico del Español de España y de América (PRESEEA)*. Universidad de Alcalá <https://preseea.uah.es/>
- PALACIOS ALCAINE, Azucena (dir.) (2017-): *Corpus COREC (Corpus Oral de Referencia del Español en Contacto)*. Universidad Autónoma de Madrid. <https://espanolcontacto.fe.uam.es/>
- PONS BORDERÍA, Salvador (dir.) (2024): *Corpus Val.Es.Co. 3.0*. Universitat de València <http://valesco.es>
- REAL ACADEMIA ESPAÑOLA (2024a): *Corpus de Referencia del Español Actual (CREA)*, <https://corpus.rae.es/creanet.html>
- REAL ACADEMIA ESPAÑOLA (2024b): *Corpus Diacrónico del Español (CORDE)*, <https://corpus.rae.es/cordenet.html>
- REAL ACADEMIA ESPAÑOLA y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (ASALE) (2024a): *Corpus del Español del Siglo XXI, versión 1.1 (CORPES XXI)*. Real Academia Española. <https://www.rae.es/corpes/>
- ROJO, Guillermo y PALACIOS, Ignacio (dirs.) (2022): *Corpus de aprendices de español (CAES)*. Universidad de Santiago de Compostela <https://galvan.usc.es/caes>
- RUIZ FAJARDO, Guadalupe (dir.) (2017): *Columbia Corpus de Conversaciones para E/LE* (Columbia). <https://edblogs.columbia.edu/corpusdeconversaciones/>

PROGRAMAS INFORMÁTICOS CITADOS

- PACI, Giulio y ROUXI, Elie (2014): TranscriberAG [programa informático] <https://transag.sourceforge.net/>

Anexo I

Transcripción completa del vídeo *Foto, foto* (Arteatre y Aspaiet) según el sistema de transcripción de Val.Es.Co (Briz y Grupo Val. Es.Co, 2002):

V: °(Mira qué chulo)° ↑ °(el Coliseo romano↑)° guaaaaau↓
J: Oooooi
V: Guaaaay. Mira, llevo la cámara aquí colgada porque soy un boomer↑ ¿vamos a hacer unas fotos, no?
J: Sí ↑
V: Que para eso hemos venido a Italia de vacaciones.
J: ¡Dale, dale! ¡Al romano!↑
V.: Sí, pero ponte ahí. Ponte tú, ponte tú, para salir con el Coliseo romano. A ver ↓
(3'')
V: ¡Mira, mira!↑
J: ¡Ah!↑
V: ¿Y eso qué es?

J: ¿Es-to es-...? ↓
V: Te juro que no había nadie cuando eché la foto.
D: Ohhhhhhh, ohhhhhhh ↑
V: Son dos romanos. →
J: ¡¡Amplíala! ¡Amplíala! ↑
V: ¡Sí que da miedo! ¡Sí! ↑
J: ¡Madre mía! ↑
V: Da mucho miedo. ↓
J: ¡Te voy a hacer yo una a ti! ↑
V: § [No, no ,no]
J: § [¡Ponte!]
V: § [Da mucho miedo]
J: § [¡Ponte!]
J: ¡Levanta la mano derecha! ¡Más! ¡Más! ¡El pie izquierdo! ¡Tira lo pa' trás! ¡Sacad la lengua! ///
(3'')
J: ¡Ostras! ¡He dicho ostras solo! ///
V: § [Hay un... hay un]
J: § [¡Amplío!] ¡Amplío! ¡Amplío!
J: ¡Espera! ¡Fíjate! ¡Fíjate lo que hay en el bolsillo! ¡Hago zoom solo al bolsillo! ¡Mira en el bolsillo! ¡Mirad en el bolsillo! (2'') ¡Si soy yo!
J: ¡Madre mía!/
V: ¡Qué miedo das en un bolsillo tío! ///
J: ¡Es que yo soy feo de nacimiento! /
V: Aquí va con un mensaje, Amplía. Lleva un mensaje.
S: Señores. El Coliseo va a cerrar en dos minutos. Ehhh... No estarán haciendo fotos. Las fotos están prohibidas a estas horas. Ahora están prohibidas. Hace mucho tiempo que no se pueden y no se permite hacer fotos aquí.
V: ¿Qué pasó en este lugar?
J: ¿Con quién estás hablando?
V: ¿Tú no ves al señor aquí? ¿No ves nada aquí?
S: § [Acuérdate.] No puedes hacer fotos.
J: § [Qué... qué...]
S: No puedes hacer fotos.
J: ¡Si no hay nadie! Mira. Voy a hacer una foto. [Hacia aquí.]
V: § [Me han sentado mal los macarrones al pesto]
J: DIOS MÍO.
V: ¡Es ese! ¡Que yo lo he vis-! ¡Mira! ¡Su familia aparece!
J: Pero... Mira cómo levanta los pechitos. Parece que esté haciendo un guaca guaca. ¡Y parece Leticia Sabater!
V: ¡Dios mío!
J: ¡Selfie! ¡Y yo salgo con (()