VOICE2TEXT AI

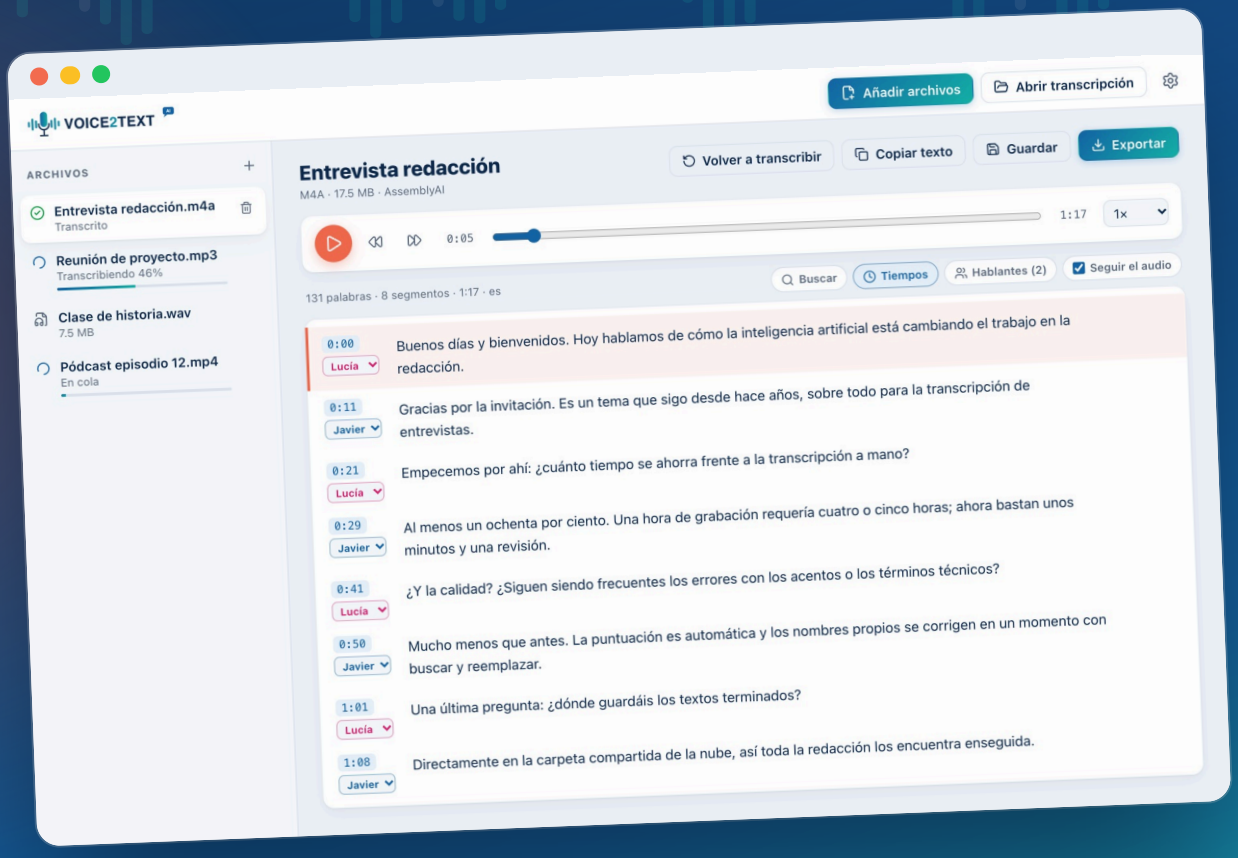
MANUAL DE USUARIO

Voice2Text

De la voz al texto, en cualquier idioma.

Transcribe archivos de audio y vídeo con inteligencia artificial, corrige el texto y guárdalo donde quieras.

Whisper sin conexión · 12 formatos · Nube · 9 idiomas



Versión 1.1 · macOS · Windows · Linux · Español

Índice

01	Bienvenido a Voice2Text	3
02	Instalación	4
03	La ventana principal	5
04	Elegir motor e idioma	6
05	Servicios en línea: tu audio sale del ordenador	8
06	Durante la transcripción	10
07	Corregir la transcripción	11
08	Guardar y exportar	13
09	Guardar en la nube	15
10	Ajustes e idiomas	17
11	Atajos de teclado	19
12	Preguntas frecuentes	20

01 Bienvenido a Voice2Text

> Voice2Text convierte grabaciones de audio y vídeo en texto escrito. Entrevistas, reuniones, clases, pódcasts, mensajes de voz: arrastra el archivo a la ventana, elige el motor de inteligencia artificial y en pocos minutos tendrás el texto, listo para corregir y guardar.

Qué puedes hacer

- **Cualquier formato.** MP3, M4A, WAV, FLAC, OGG, AAC, AIFF, vídeos MP4, MOV, MKV, WebM y muchos más; con ffmpeg instalado, prácticamente todos.
- **Incluso sin Internet.** Whisper funciona directamente en el ordenador: el audio nunca sale de tu escritorio.
- **Quién habla.** Con AssemblyAI y Deepgram las voces se separan y puedes poner nombre a cada persona.
- **Corrección cómoda.** El texto está dividido en frases vinculadas al audio: un clic y vuelves a escuchar el punto exacto.
- **Cualquier formato de salida.** Word, PDF, ODT, RTF, TXT, Markdown, HTML, subtítulos SRT y VTT, CSV, JSON y el proyecto que se puede reabrir.
- **Directamente en la nube.** Google Drive, OneDrive, Dropbox, Box, Proton Drive, iCloud y otros, mediante sus carpetas en el ordenador.
- **Nueve idiomas.** La interfaz habla italiano, inglés, francés, español, alemán, chino, japonés, ruso y árabe; las traducciones son archivos XML que puedes corregir.

Requisitos

macOS 11 o posterior (Mac con Apple o Intel), Windows 10/11 de 64 bits o Linux de 64 bits (Ubuntu, Debian, Fedora y derivadas). Para Whisper sin conexión conviene tener al menos 8 GB de memoria; los servicios en línea solo necesitan conexión.

02 Instalación

> Descarga el archivo para tu sistema e instálalo como cualquier otro programa.

macOS

Abre `Voice2Text_1.1.1_universal.dmg` y arrastra el icono de Voice2Text a la carpeta **Aplicaciones**. Al abrirlo por primera vez, macOS puede indicar que el desarrollador no está verificado: abre **Ajustes del Sistema** → **Privacidad y seguridad** y elige **Abrir igualmente**, o haz clic derecho en la app y elige **Abrir**. Solo hace falta una vez.

Windows

Ejecuta `Voice2Text_1.1.1_x64-setup.exe`. Si aparece la ventana azul de Windows SmartScreen, haz clic en **Más información** y después en **Ejecutar de todas formas**. La instalación no requiere permisos de administrador.

Linux

En Ubuntu y Debian instala el paquete `.deb` (doble clic o `sudo apt install ./Voice2Text_1.1.1_amd64.deb`); en Fedora, el paquete `.rpm`. También puedes hacer ejecutable el `.AppImage` (`chmod +x`) e iniciarlo.

> **Opcional: ffmpeg.** Voice2Text lee por sí mismo los formatos más comunes. Para los poco habituales (OPUS, WMA, AMR, AVI…) instala ffmpeg: `brew install ffmpeg` en Mac, `winget install ffmpeg` en Windows, `sudo apt install ffmpeg` en Linux. Se detecta automáticamente.

03 La ventana principal

> Todo el trabajo se hace en una sola ventana: a la izquierda la lista de archivos y en el centro el archivo seleccionado.

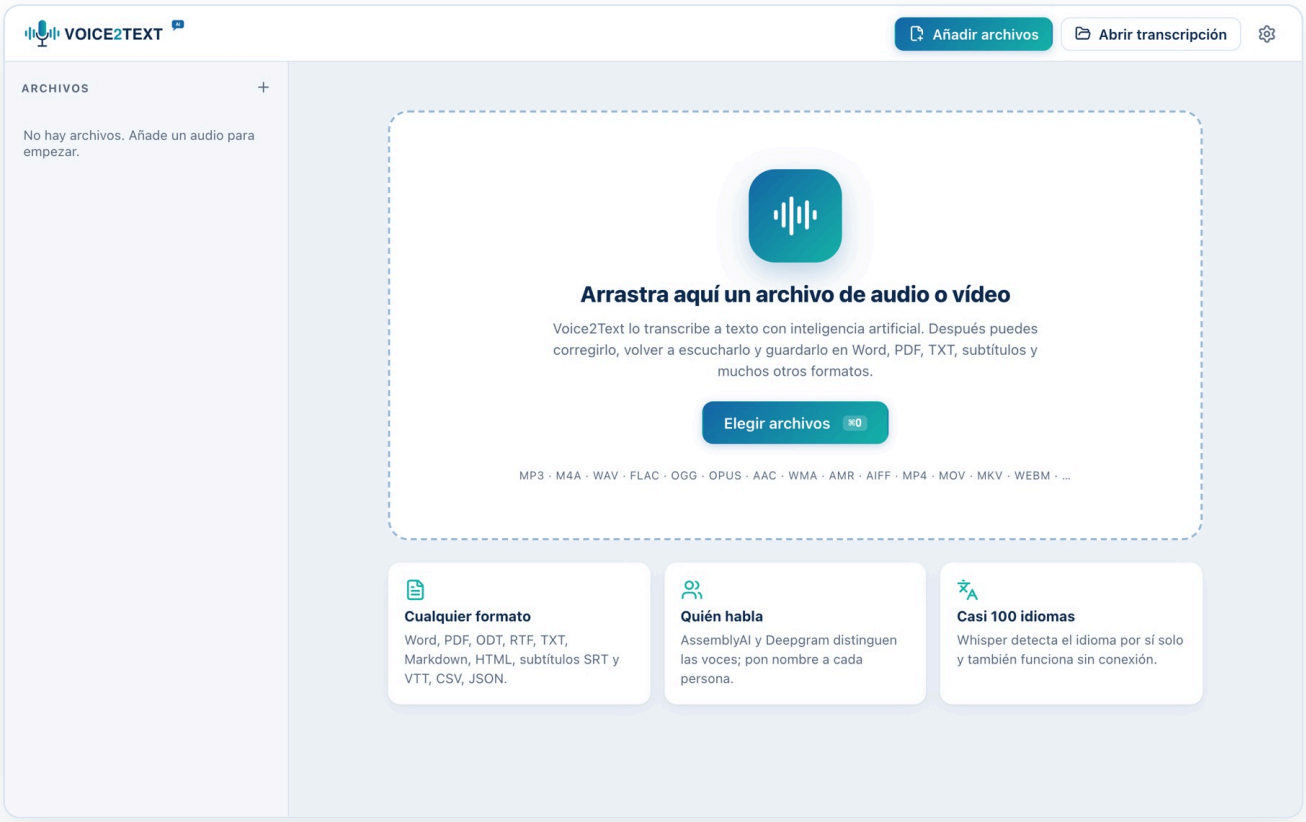


Figura 1. La pantalla de inicio: arrastra un archivo al recuadro o usa «Elegir archivos».

- **Barra superior**: **Añadir archivos**, **Abrir transcripción** (un proyecto `.v2t` guardado) y el engranaje de **Ajustes**.
- **Lista de archivos** a la izquierda: cada archivo muestra su estado (listo, en cola, transcribiendo, transcrito, error) y su avance. El punto tras el nombre indica cambios sin guardar.
- **Zona central**: los ajustes de la transcripción, el avance o el texto que hay que corregir.

Añadir archivos

Arrastra uno o varios archivos de audio o vídeo a la ventana o usa **Añadir archivos** (⌘O / Ctrl+O). Puedes añadir todos los que quieras: se transcriben uno tras otro. Arrastrando un archivo `.v2t` vuelves a abrir una transcripción guardada.

04 Elegir motor e idioma

> Antes de transcribir, elige quién hará el trabajo: Whisper en tu ordenador o un servicio en línea.

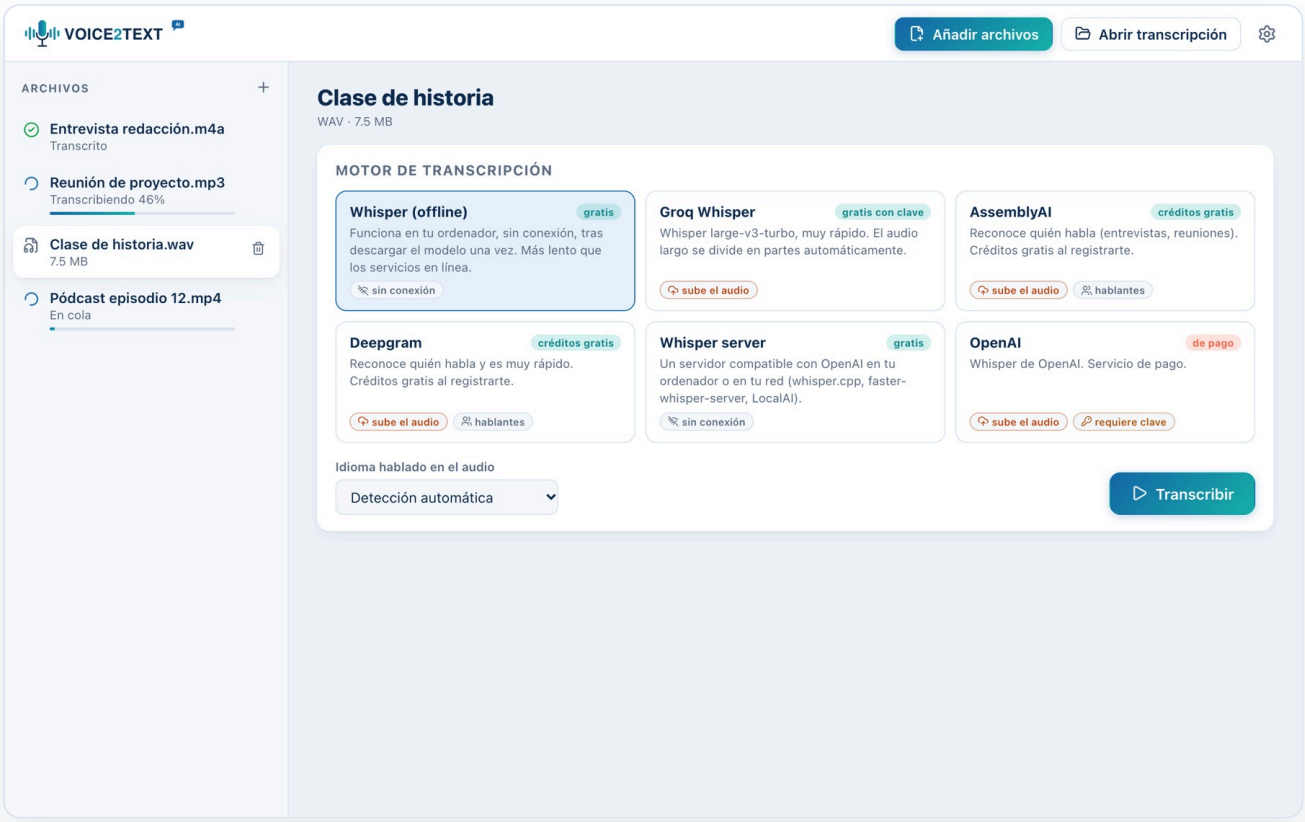


Figura 2. Elección del motor y del idioma para el archivo seleccionado.

Motor	Coste	Características
Whisper (sin conexión)	gratis	Funciona sin Internet; el modelo se descarga una vez (de 45 a 600 MB). Más lento que los servicios en línea.
Groq Whisper	gratis con clave	Muy rápido; el audio largo se divide en partes automáticamente.
AssemblyAI	créditos gratis	Reconoce quién habla. Ideal para entrevistas y reuniones.
Deepgram	créditos gratis	Reconoce quién habla y es muy rápido.
Whisper server	gratis	Un servidor Whisper en tu ordenador o en tu red (whisper.cpp, faster-whisper-server, LocalAI).
OpenAI	de pago	Whisper de OpenAI.

El idioma del audio

Con **Detección automática**, Voice2Text reconoce por sí solo el idioma hablado. Si lo conoces, elígelo: el resultado es más preciso, sobre todo en los primeros segundos.

Las claves de los servicios en línea

Groq, AssemblyAI, Deepgram y OpenAI piden una clave personal: en **Ajustes** → **Motores** encontrarás un campo para cada uno y el enlace «obtener clave». El registro en Groq es gratuito; AssemblyAI y Deepgram regalan créditos al registrarte.

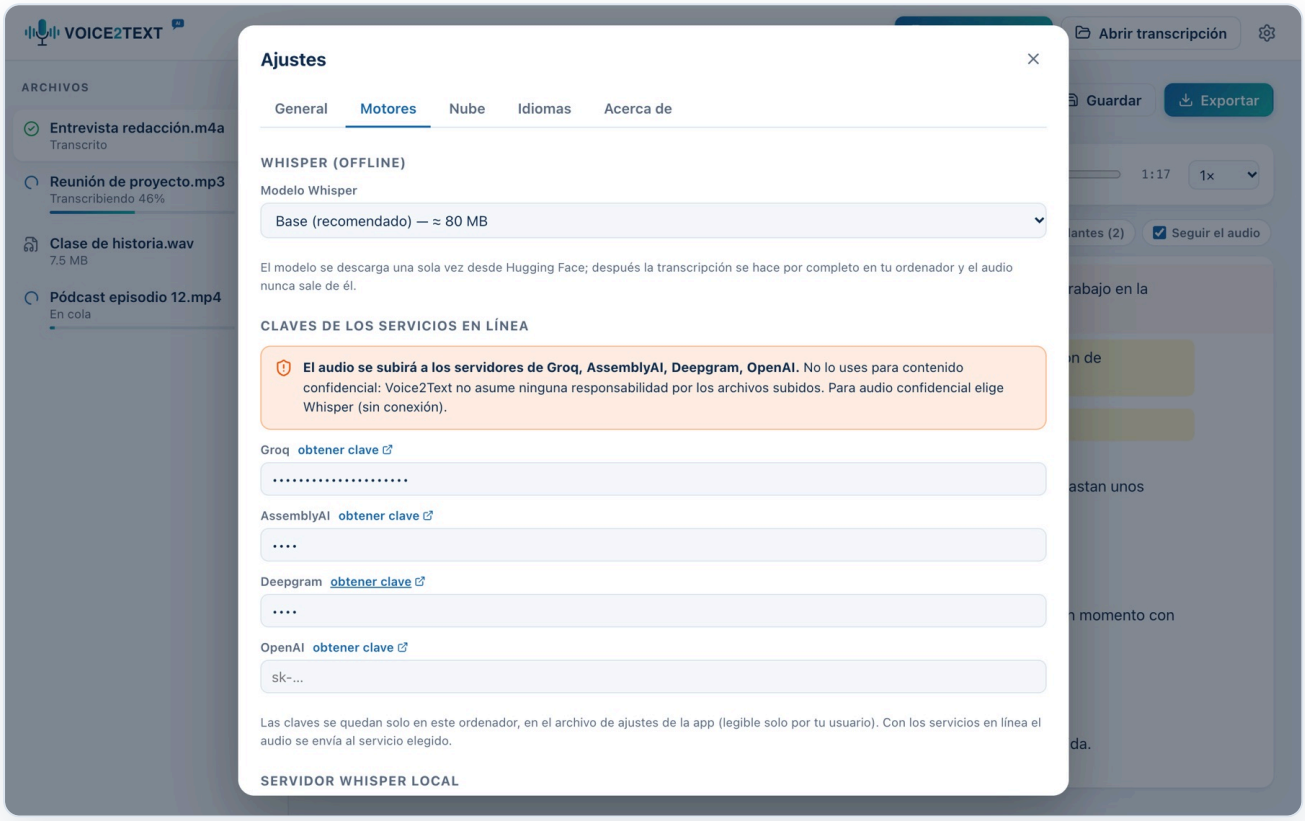


Figura 3. Ajustes → Motores: modelo Whisper y claves de los servicios.

Pulsa **Transcribir** para empezar. Con varios archivos en espera aparece también **Transcribir todos** en la barra lateral.

05 Servicios en línea: tu audio sale del ordenador

> Cuando usas un servicio de inteligencia artificial en línea, el archivo de audio se sube a Internet. Voice2Text te lo dice con claridad y pide tu confirmación.

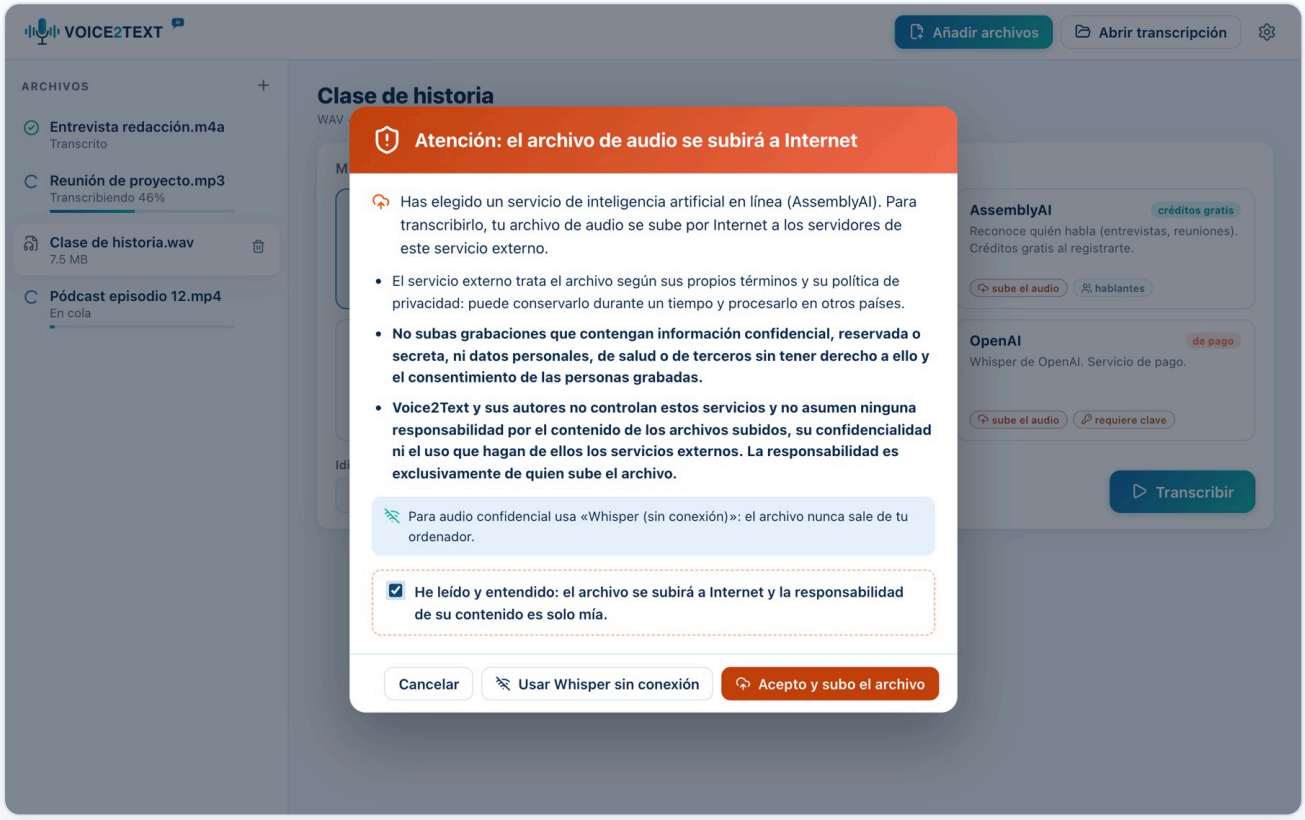


Figura 4. El aviso obligatorio antes del primer envío a un servicio en línea.

La primera vez que usas Groq, AssemblyAI, Deepgram u OpenAI aparece este aviso. Para continuar debes marcar la casilla **«He leído y entendido»**; también puedes pasar directamente a **Whisper sin conexión**.

> **El archivo de audio se sube a los servidores del servicio externo**, que lo trata según sus propios términos y su política de privacidad: puede conservarlo durante un tiempo y procesarlo en otros países.
>
> **No subas grabaciones con información confidencial**, reservada o secreta, ni datos personales, de salud o de terceros sin tener derecho a ello y el consentimiento de las personas grabadas.
>
> **Voice2Text y sus autores no asumen ninguna responsabilidad** por el contenido de los archivos subidos, su confidencialidad ni el uso que hagan de ellos los servicios externos. La responsabilidad es exclusivamente de quien sube el archivo.

La aceptación vale para cada servicio por separado y se registra con la fecha. Mientras haya un servicio en línea seleccionado, un recuadro naranja te lo recuerda. En **Ajustes → Acerca de** puedes volver a leer el aviso y hacer que aparezca de nuevo.

> **¿Audio confidencial?** Usa **Whisper (sin conexión)**: la transcripción se hace por completo en tu ordenador.

06 Durante la transcripción

> Voice2Text muestra en todo momento cómo va el trabajo.

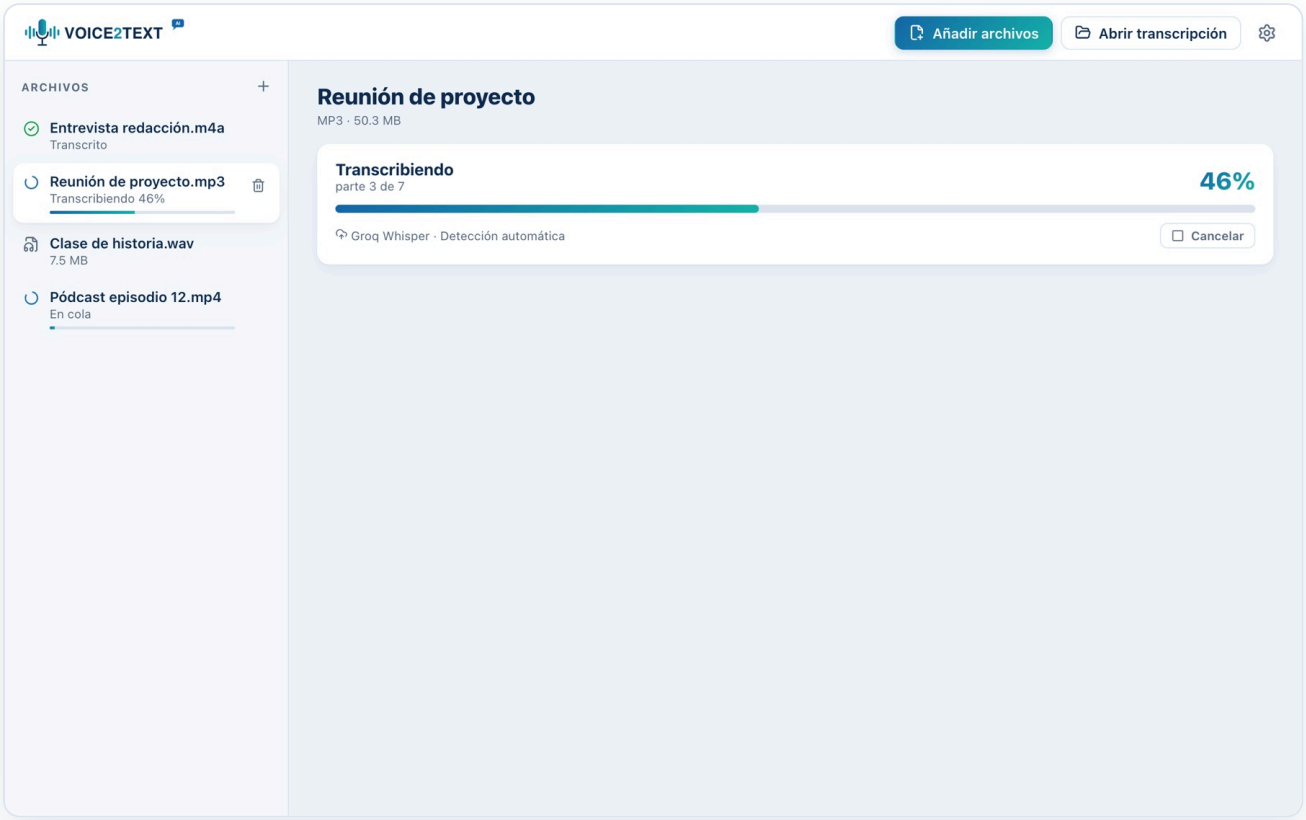


Figura 5. Una transcripción en curso con Groq: avance y parte en proceso.

- **Leyendo el audio**: el archivo se decodifica en el ordenador.
- **Preparando el modelo** (solo Whisper sin conexión, la primera vez): el modelo elegido se descarga y se conserva.
- **Enviando el audio** y **Transcribiendo**: con archivos largos ves la parte en proceso («parte 3 de 7»).
- **Esperando al servicio**: con los planes gratuitos, si se supera el límite de solicitudes, Voice2Text espera y vuelve a intentarlo solo.

Puedes cancelar en cualquier momento con **Cancelar**, seguir trabajando con otros archivos y dejar la transcripción en marcha: cuando el texto esté listo recibirás una notificación.

07 Corregir la transcripción

> El texto está dividido en frases, cada una con su tiempo y, si se reconoce, la persona que habla.



Figura 6. El editor: reproductor de audio arriba y frases editables; la frase que se está escuchando aparece resaltada.

Escuchar

El reproductor de arriba reproduce el audio original: `Espacio` reproduce y pausa (cuando no estás escribiendo), los botones saltan 5 segundos y puedes elegir la velocidad de 0,75× a 2×. Haz clic en el tiempo de una frase para escucharla; con **Seguir el audio** la frase en reproducción siempre queda a la vista.

Editar

- Haz clic en una frase y escribe: los cambios son inmediatos.
- `Intro` divide la frase en la posición del cursor; `Mayús+Intro` añade un salto de línea en la misma frase.
- `Retroceso` al principio de una frase la une con la anterior; el botón **Unir** la junta con la siguiente.
- La papelera elimina una frase (por ejemplo, un ruido reconocido como palabra).

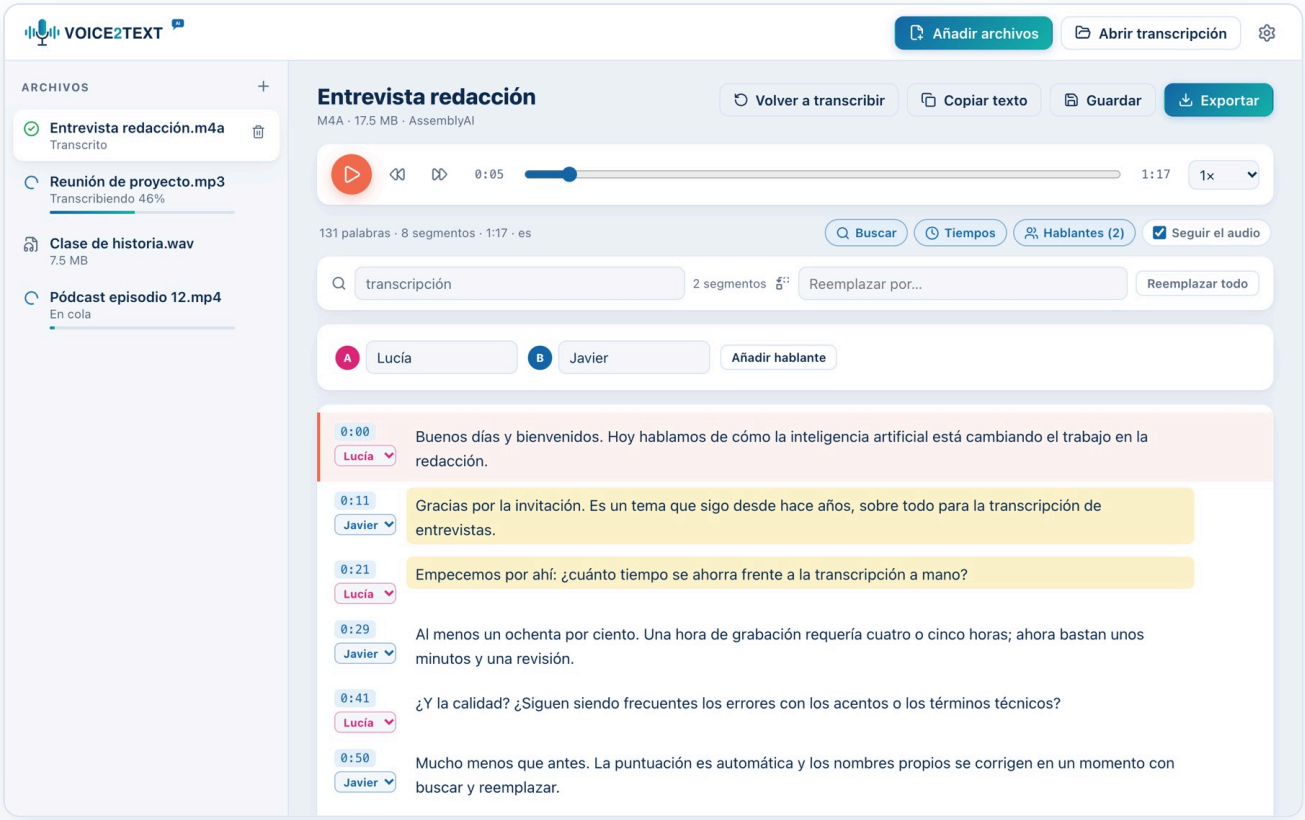


Figura 7. Buscar y reemplazar, y los nombres de los hablantes.

Buscar y reemplazar

Buscar (⌘F / Ctrl+F) resalta las frases que contienen la palabra; **Reemplazar todo** corrige de una vez un nombre mal escrito en toda la transcripción.

Hablantes

Con AssemblyAI y Deepgram cada frase tiene una etiqueta (Hablante A, Hablante B…). En el panel **Hablantes** escribe el nombre real: aparece en todas partes, también en los archivos exportados. Puedes cambiar el hablante de una frase desde el menú junto al tiempo, añadir más o asignarlos a mano aunque el motor no los reconozca.

08 Guardar y exportar

> Guarda el proyecto para seguir corrigiendo más tarde y exporta el texto en el formato que necesites.

El proyecto .v2t

Guardar (`⌘S` / `Ctrl+S`) crea un archivo `.v2t` con la transcripción, las correcciones y los nombres de los hablantes. Vuelve a abrirlo con **Abrir transcripción** (`⇧⌘O` / `Ctrl+Mayús+O`) o arrastrándolo a la ventana. `⇧⌘S` guarda con otro nombre.

Exportar

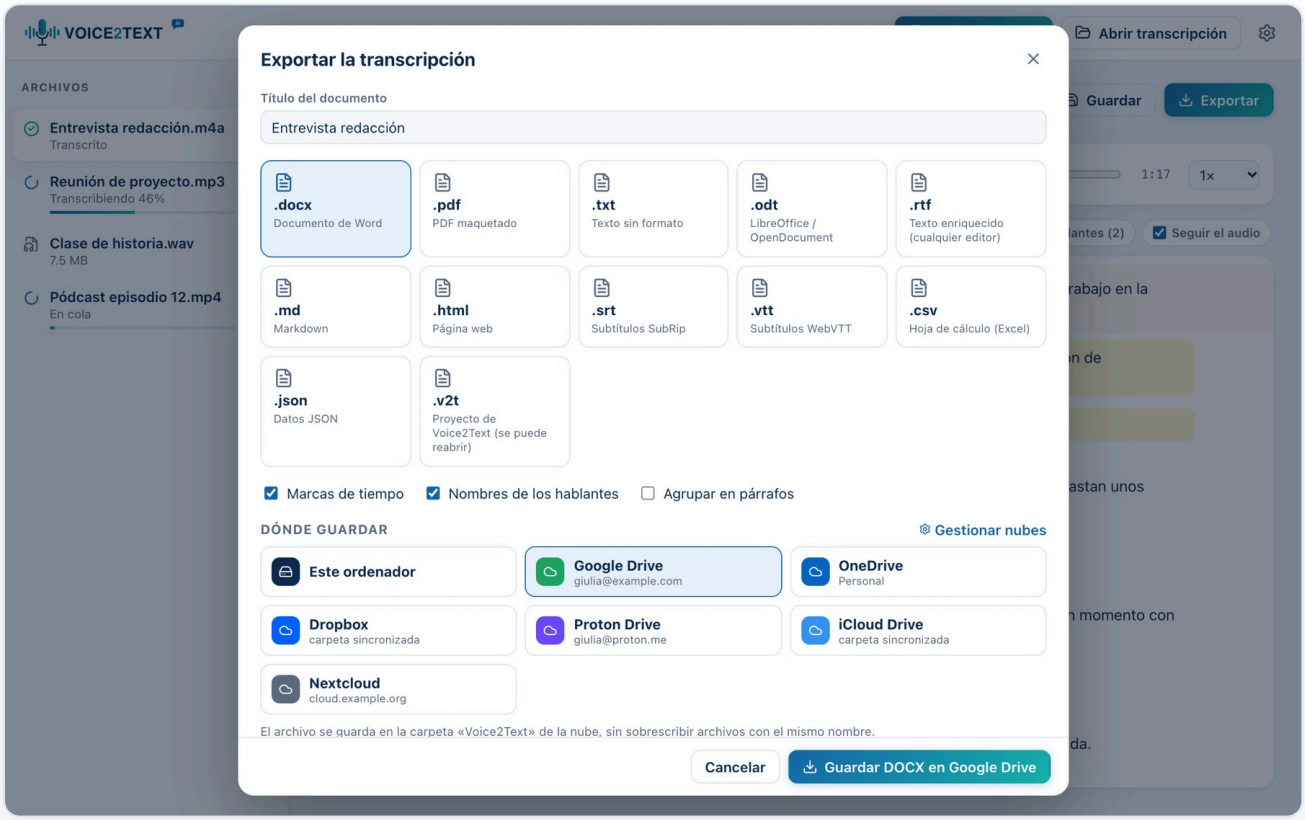


Figura 8. La ventana Exportar: formato, opciones y destino.

Pulsa **Exportar** (`⌘E` / `Ctrl+E`), escribe el título del documento y elige el formato:

Formato	Ideal para
.docx Word · **.odt** LibreOffice · **.rtf**	documentos para editar, maquetados con título y datos de la grabación
.pdf	documentos para imprimir o enviar; se crea con el cuadro de impresión del sistema («Guardar como PDF»)
.txt · **.md** · **.html**	texto sin formato, Markdown, página web
.srt · **.vtt**	subtítulos para vídeo (YouTube, Premiere, DaVinci, VLC…)
.csv · **.json**	hojas de cálculo (Excel, Numbers) y datos para otros programas
.v2t	el proyecto que se puede reabrir, también para guardarlo en la nube

Las opciones permiten incluir las **marcas de tiempo**, los **nombres de los hablantes** y **agrupar las frases en párrafos**. El árabe y los demás idiomas de derecha a izquierda se maquetan correctamente.

09 Guardar en la nube

> Voice2Text guarda los archivos directamente en tu nube usando las carpetas que las apps de los servicios crean en el ordenador. No hace falta conectar cuentas ni escribir contraseñas.

Cómo funciona

Al instalar la app de Google Drive, OneDrive, Dropbox, Box, **Proton Drive**, iCloud Drive, pCloud, MEGA, Nextcloud o Sync.com, aparece una carpeta sincronizada en el ordenador. Voice2Text la encuentra por sí solo: en la ventana **Exportar**, en **Dónde guardar**, elige la nube y pulsa el botón. El archivo se escribe en la subcarpeta `Voice2Text` y la app del servicio lo sube enseguida.

- Los archivos nunca se sobrescriben: si ya existe uno con el mismo nombre, el nuevo pasa a llamarse «nombre (2)».
- Puedes guardar en la nube cualquier formato, incluido el proyecto `.v2t`.
- El PDF pasa por el cuadro de impresión: allí elige «Guardar como PDF» y la carpeta de la nube.

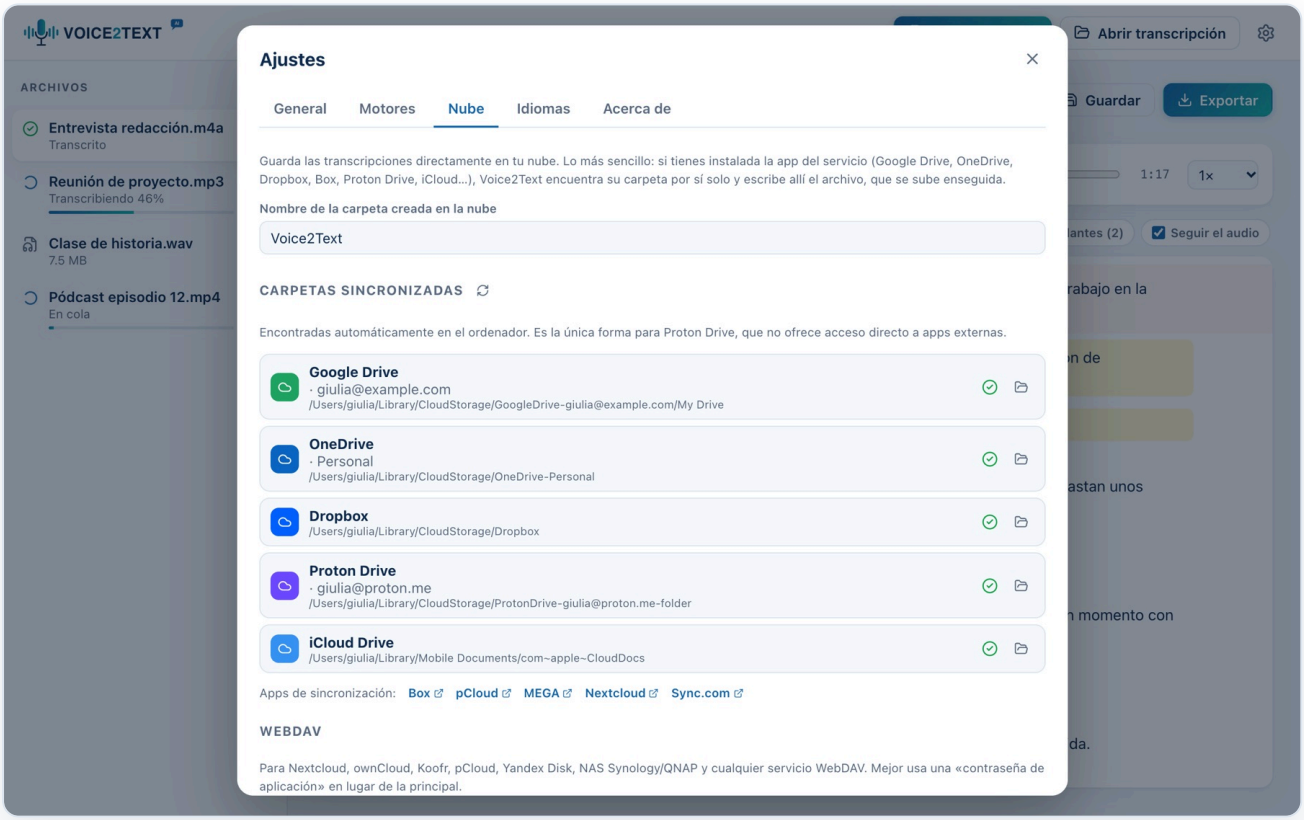


Figura 9. Ajustes → Nube: las carpetas encontradas en el ordenador y la conexión WebDAV.

Ajustes → Nube

- **Nombre de la carpeta**: la subcarpeta creada en la nube (por defecto `Voice2Text`).
- **Carpetas sincronizadas**: la lista de nubes encontradas; el icono de carpeta la abre. Si falta algún servicio, los enlaces de abajo llevan a las apps que hay que instalar.
- **Añadir otra carpeta**: para otras nubes o un disco de red (NAS).

- **WebDAV**: para Nextcloud, ownCloud, Koofr, pCloud, Yandex Disk y NAS Synology/QNAP sin app instalada. Elige el servicio, escribe la dirección, el usuario y una contraseña de aplicación y pulsa **Probar la conexión**.

> **Proton Drive** no permite que otras apps accedan directamente a la cuenta: por eso Voice2Text usa su carpeta sincronizada, que mantiene el mismo cifrado de extremo a extremo.

10 Ajustes e idiomas

> El engranaje de arriba a la derecha (`⌘,` / `Ctrl+,`) abre los ajustes.

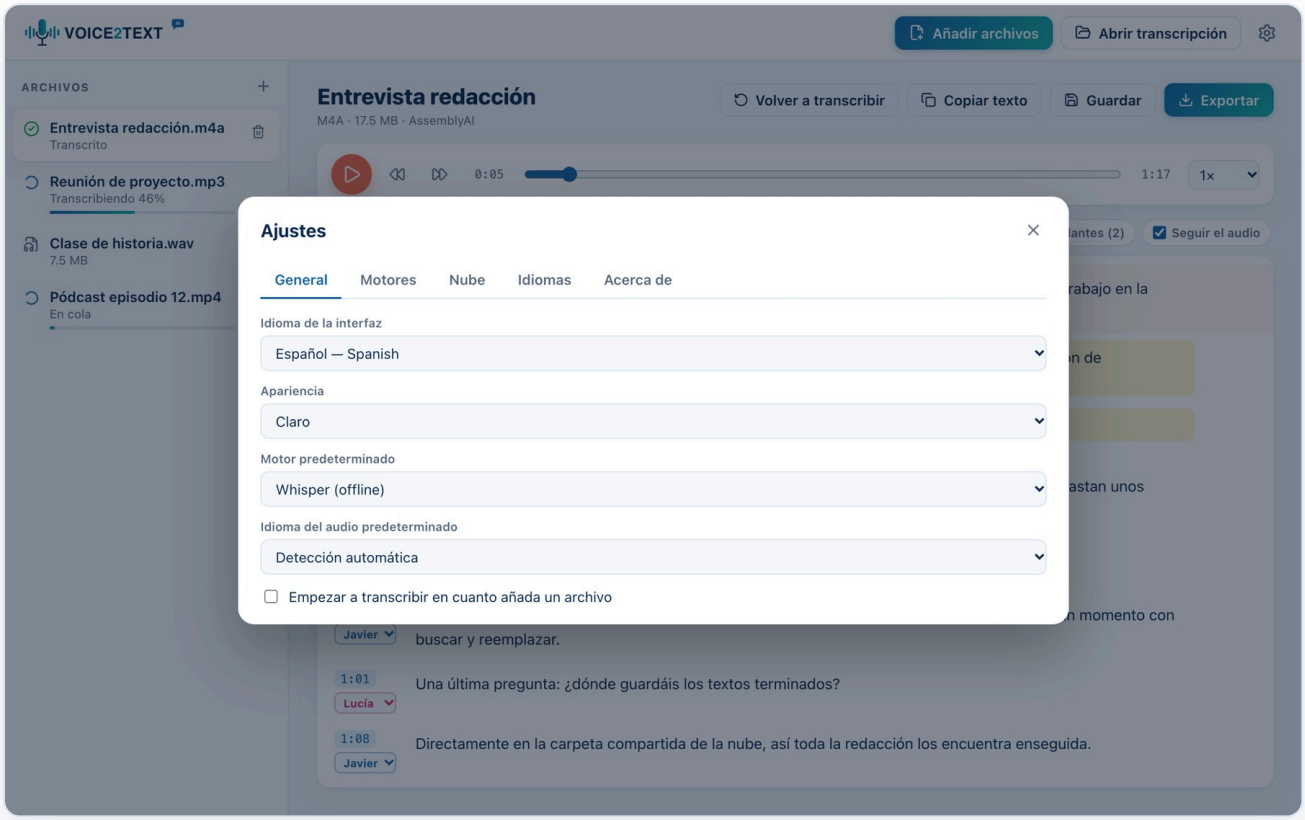


Figura 10. Ajustes → General.

- **General**: idioma de la interfaz, apariencia clara, oscura o automática, motor e idioma del audio predeterminados, inicio automático de la transcripción.
- **Motores**: modelo de Whisper sin conexión (mínimo, base, pequeño, grande turbo), espacio ocupado por los modelos, claves de los servicios, servidor Whisper local, presencia de ffmpeg.
- **Nube**: carpetas sincronizadas y WebDAV (capítulo anterior).
- **Idiomas**: los archivos de traducción de la interfaz.
- **Acerca de**: versión, licencia, aviso sobre los servicios en línea.

Los idiomas son archivos XML

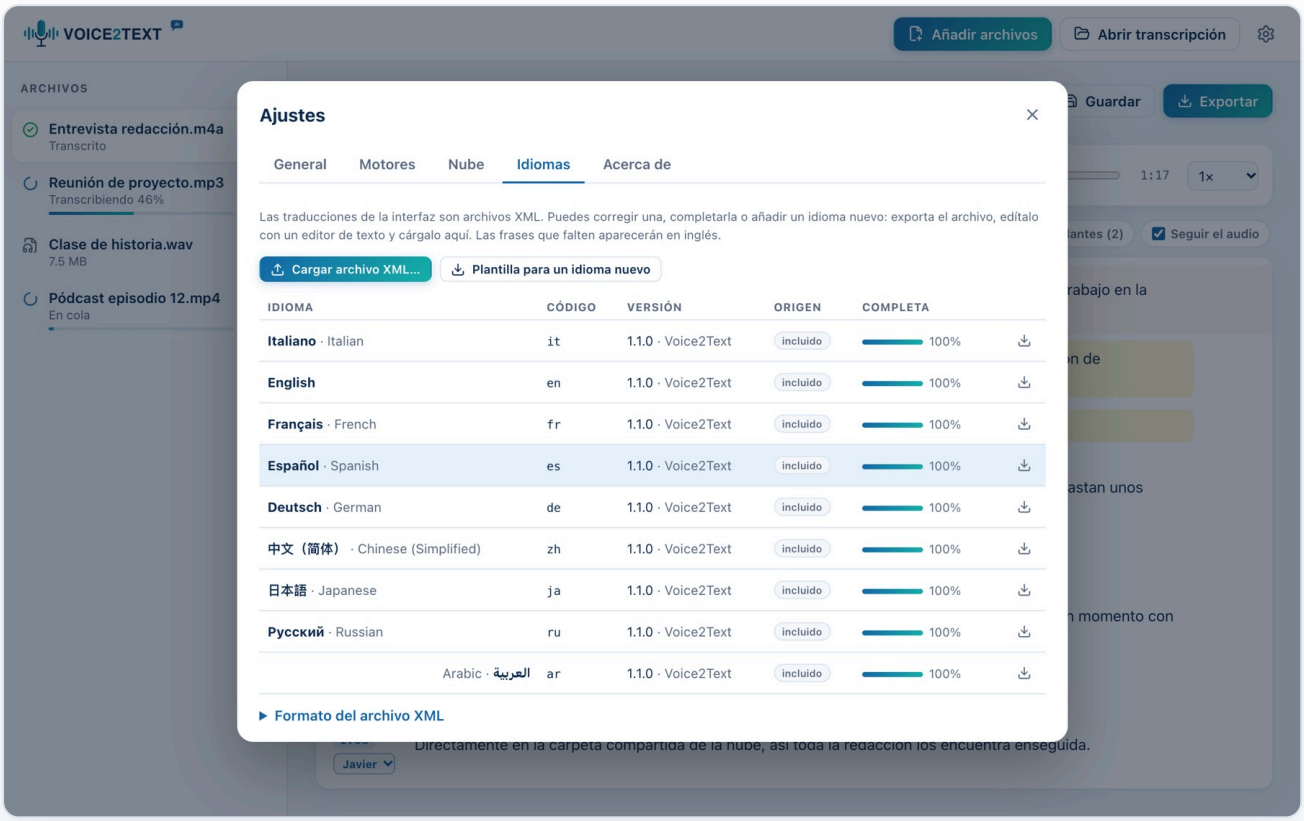


Figura 11. Ajustes → Idiomas: idiomas disponibles, grado de traducción y herramientas para los archivos XML.

Cada idioma de la interfaz es un archivo XML. Para corregir una traducción o añadir una nueva:

1. Exporta el archivo del idioma (flecha en su fila) o descarga la **Plantilla para un idioma nuevo**.
2. Ábrelo con un editor de texto y traduce solo el texto entre las etiquetas, sin cambiar los atributos `key` ni los marcadores entre llaves como `{n}`.
3. Pulsa **Cargar archivo XML...**: el idioma aparece enseguida en la lista y en el menú de idiomas.

```
<language code="pt" name="Português" english="Portuguese" dir="ltr" version="1.0">
  <string key="app.addFiles">Adicionar ficheiros</string>
  <string key="progress.piece">parte {i} de {n}</string>
</language>
```

Un archivo con el mismo código que un idioma incluido lo sustituye (para las correcciones); siempre puedes volver al original. Las frases que falten aparecen en inglés; la columna **Completa** muestra cuántas están traducidas. Para los idiomas de derecha a izquierda usa `dir="rtl"`.

11 Atajos de teclado

> En Windows y Linux usa **Ctrl** en lugar de ⌘.

Teclas	Acción
⌘O	Añadir archivos de audio o vídeo
⇧⌘O	Abrir una transcripción guardada (.v2t)
⌘S · ⇧⌘S	Guardar · Guardar como
⌘E	Exportar
⌘F	Buscar y reemplazar
⌘,	Ajustes
Espacio	Reproducir / pausa (fuera del texto)
Intro · **Mayús+Intro**	Dividir la frase · Salto de línea en la frase
Retroceso al inicio de la frase	Unir con la frase anterior
Esc	Salir del texto · cerrar la ventana abierta

12 Preguntas frecuentes

> Respuestas a las dudas más comunes.

El formato de mi archivo no se lee.

Instala ffmpeg (capítulo Instalación) o usa AssemblyAI o Deepgram, que aceptan cualquier formato porque reciben el archivo original.

Whisper sin conexión es lento.

Elige el modelo **Base** o **Mínimo** en Ajustes → Motores. «Grande turbo» es el más preciso, pero sin tarjeta gráfica tarda. Para archivos largos, Groq es mucho más rápido.

El texto está en otro idioma o traducido.

Indica el idioma del audio en lugar de la detección automática y usa **Volver a transcribir**.

«Clave del servicio no válida».

Revisa la clave en Ajustes → Motores: cópiala de nuevo desde la web del servicio, sin espacios.

Mi nube no aparece.

Comprueba que la app del servicio esté instalada y sincronizada y pulsa **Buscar de nuevo** en Ajustes → Nube. Si no, añade la carpeta a mano.

El reproductor dice «audio no disponible».

El archivo de audio original se movió o se cambió de nombre después de guardar el proyecto. El texto sigue disponible; para volver a escucharlo, devuelve el archivo a su sitio.

¿Dónde están mis claves y mis ajustes?

Solo en tu ordenador, en el archivo de ajustes de la app, legible únicamente por tu usuario. Nunca se envían a Voice2Text.



VOICE2TEXT

De la voz al texto, en cualquier idioma.

Voice2Text es un programa de overminder. Licencia PolyForm Strict 1.0.0: uso no comercial permitido; redistribución, modificaciones y obras derivadas no permitidas.

© 2026 overminder · Versión 1.1