VOICE2TEXT AI

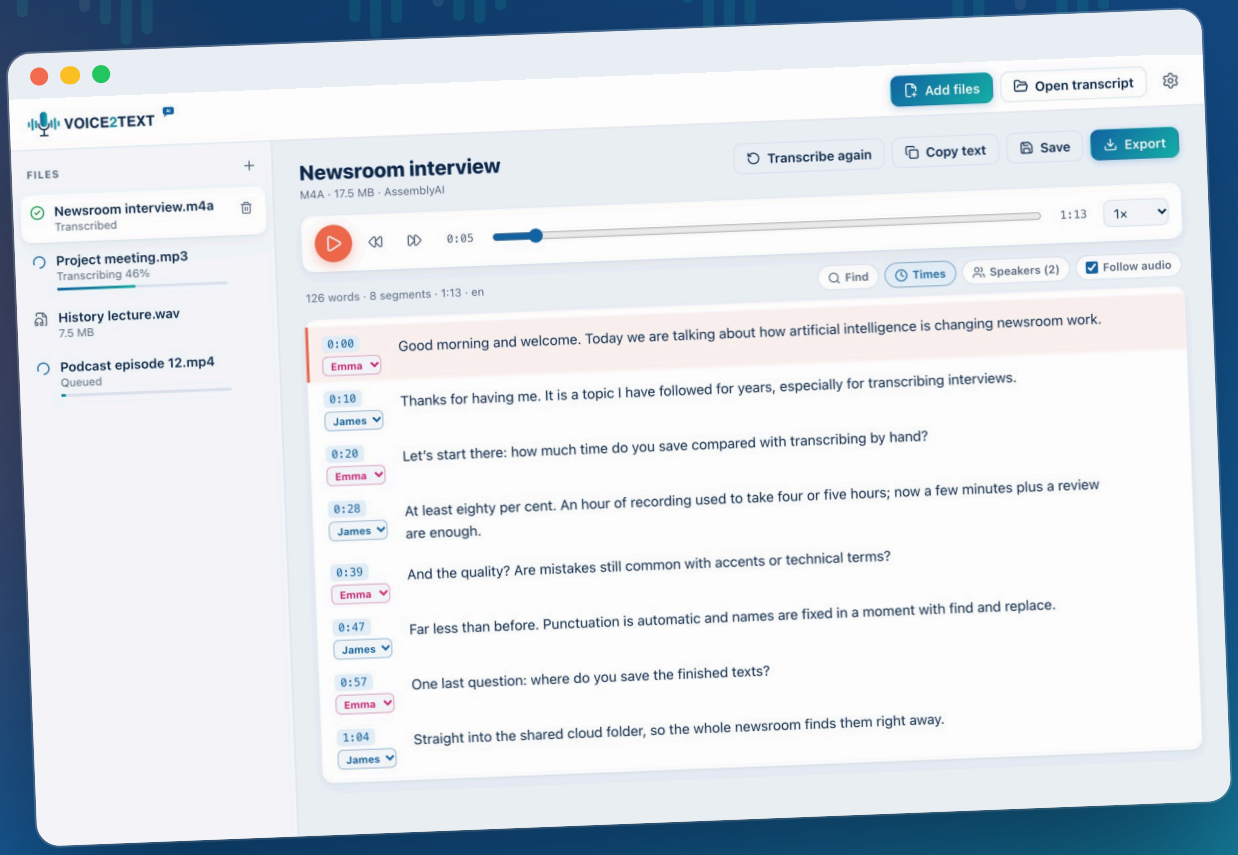
USER GUIDE

Voice2Text

From voice to text, in any language.

Transcribe audio and video files with artificial intelligence, edit the text and save it wherever you like.

Offline Whisper · 12 formats · Cloud · 9 languages



Version 1.1 · macOS · Windows · Linux · English

Contents

01	Welcome to Voice2Text	3
02	Installation	4
03	The main window	5
04	Choosing engine and language	6
05	Online services: your audio leaves the computer	8
06	During transcription	10
07	Editing the transcript	11
08	Saving and exporting	13
09	Saving to the cloud	15
10	Settings and languages	17
11	Keyboard shortcuts	19
12	Frequently asked questions	20

01 Welcome to Voice2Text

> Voice2Text turns audio and video recordings into written text. Interviews, meetings, lectures, podcasts, voice messages: drop the file into the window, choose the AI engine and within minutes you have the text, ready to edit and save.

What you can do

- **Any format.** MP3, M4A, WAV, FLAC, OGG, AAC, AIFF, MP4, MOV, MKV and WebM video and many more; with ffmpeg installed, practically all of them.
- **Even without Internet.** Whisper runs right on your computer: the audio never leaves your desk.
- **Who is speaking.** With AssemblyAI and Deepgram the voices are told apart and you can give each person a name.
- **Easy editing.** The text is split into sentences linked to the audio: one click and you hear the exact spot again.
- **Any output format.** Word, PDF, ODT, RTF, TXT, Markdown, HTML, SRT and VTT subtitles, CSV, JSON and the re-openable project.
- **Straight to the cloud.** Google Drive, OneDrive, Dropbox, Box, Proton Drive, iCloud and others, through their folders on your computer.
- **Nine languages.** The interface speaks Italian, English, French, Spanish, German, Chinese, Japanese, Russian and Arabic; translations are XML files you can correct.

Requirements

macOS 11 or later (Apple or Intel Mac), 64-bit Windows 10/11, or 64-bit Linux (Ubuntu, Debian, Fedora and derivatives). For offline Whisper at least 8 GB of memory is recommended; online services only need a connection.

02 Installation

> Download the file for your system and install it like any other program.

macOS

Open `Voice2Text_1.1.1_universal.dmg` and drag the Voice2Text icon into the **Applications** folder. On first launch macOS may say the developer cannot be verified: open **System Settings** → **Privacy & Security** and choose **Open Anyway**, or right-click the app and choose **Open**. You only need to do this once.

Windows

Run `Voice2Text_1.1.1_x64-setup.exe`. If the blue Windows SmartScreen window appears, click **More info** and then **Run anyway**. Installation does not require administrator rights.

Linux

On Ubuntu and Debian install the `.deb` package (double-click or `sudo apt install ./Voice2Text_1.1.1_amd64.deb`); on Fedora the `.rpm` package. Alternatively make the `.AppImage` executable (`chmod +x`) and run it.

> **Optional: ffmpeg.** Voice2Text reads the most common formats by itself. For rare ones (OPUS, WMA, AMR, AVI...) install ffmpeg: `brew install ffmpeg` on Mac, `winget install ffmpeg` on Windows, `sudo apt install ffmpeg` on Linux. It is found automatically.

03 The main window

> All the work happens in a single window: the list of files on the left, the selected file in the middle.

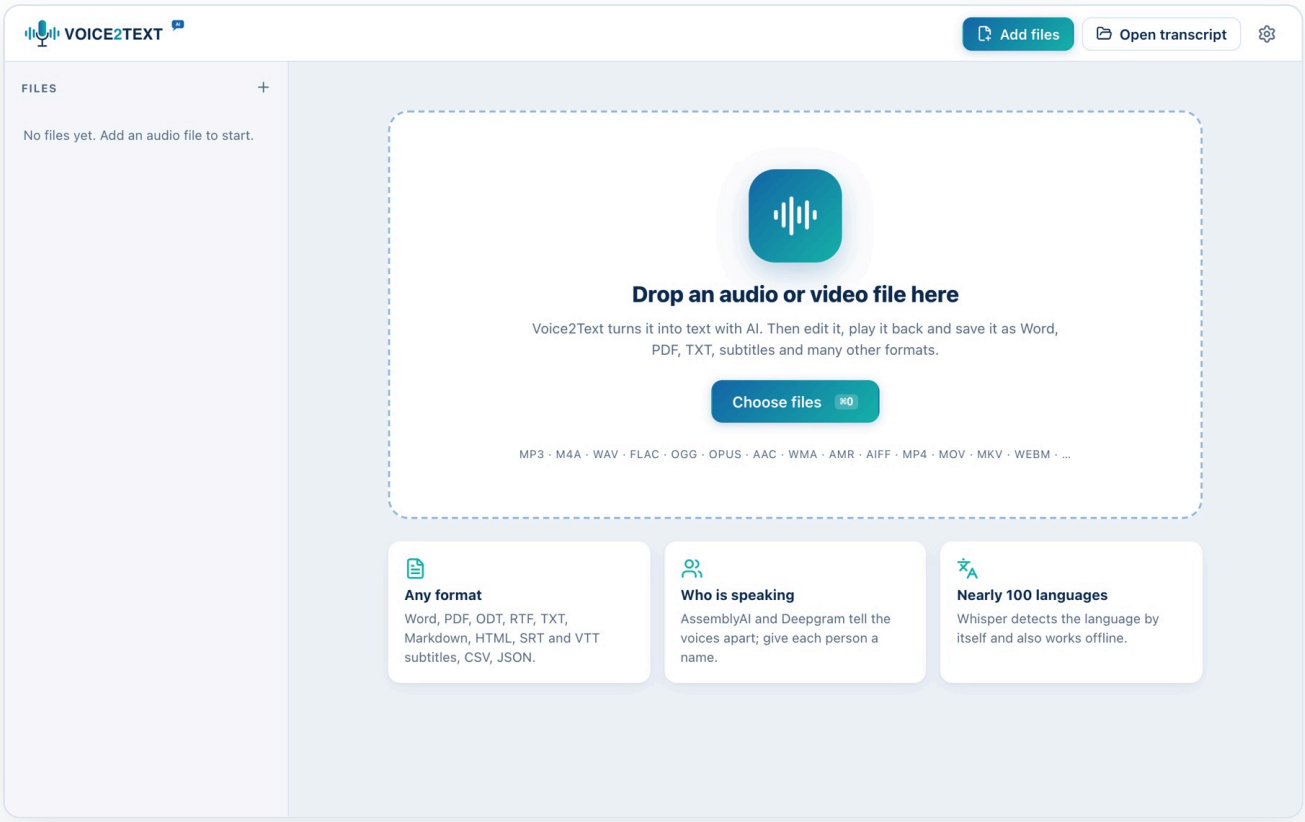


Figure 1. The start screen: drop a file onto the box or use “Choose files”.

- **Top bar**: **Add files**, **Open transcript** (a saved `.v2t` project) and the **Settings** gear.
- **File list** on the left: each file shows its status (ready, queued, transcribing, transcribed, error) and progress. A dot after the name means unsaved changes.
- **Central area**: the transcription settings, the progress or the text to edit.

Adding files

Drag one or more audio or video files onto the window, or use **Add files** (`⌘O` / `Ctrl+O`). Add as many as you like: they are transcribed one after another. Dropping a `.v2t` file reopens a saved transcript.

04 Choosing engine and language

> Before transcribing, choose who does the work: Whisper on your computer or an online service.

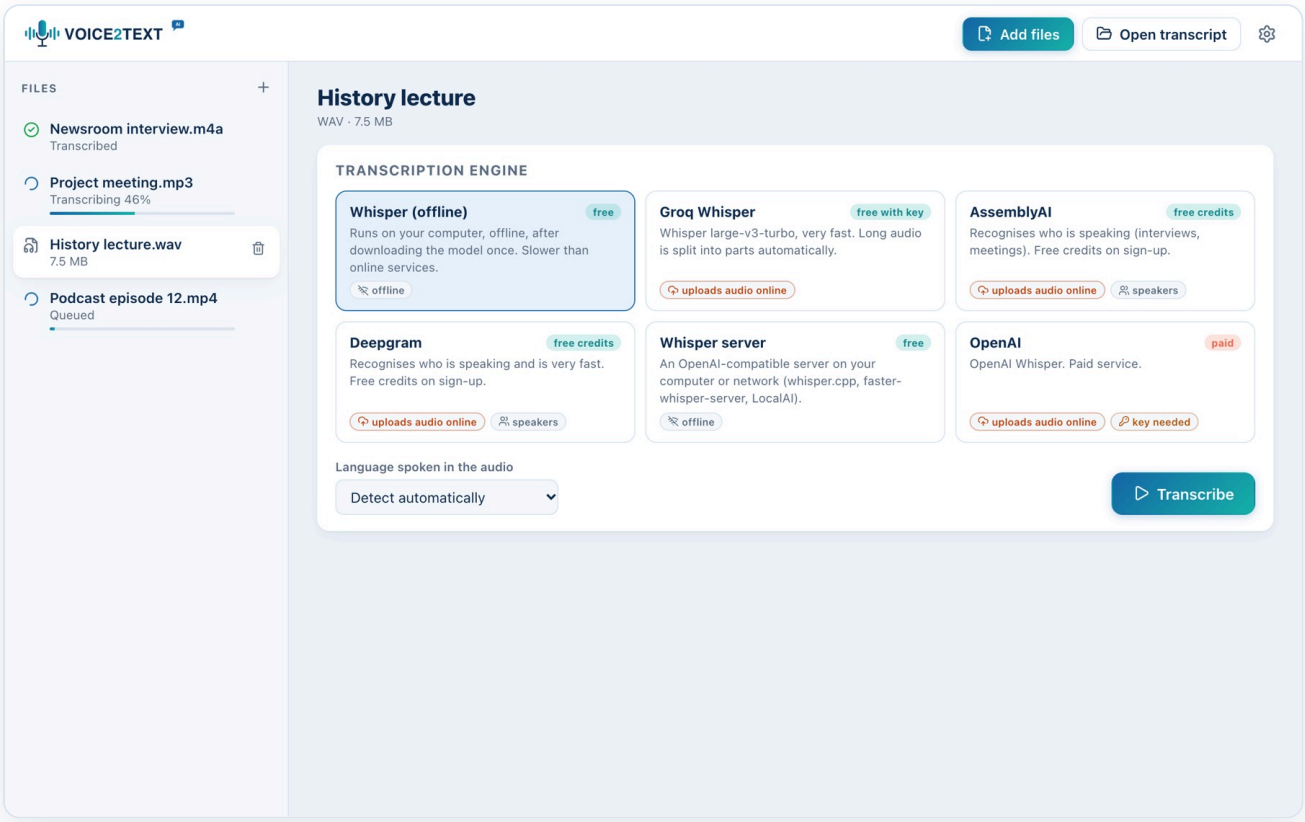


Figure 2. Choosing the engine and the language for the selected file.

Engine	Cost	Features
Whisper (offline)	free	Works without Internet; the model is downloaded once (45 to 600 MB). Slower than online services.
Groq Whisper	free with key	Very fast; long audio is split into parts automatically.
AssemblyAI	free credits	Recognises who is speaking. Ideal for interviews and meetings.
Deepgram	free credits	Recognises who is speaking and is very fast.
Whisper server	free	A Whisper server on your computer or network (whisper.cpp, faster-whisper-server, LocalAI).
OpenAI	paid	OpenAI Whisper.

The audio language

With **Detect automatically** Voice2Text works out the spoken language by itself. If you know it, choose it: the result is more accurate, especially in the first seconds.

Keys for online services

Groq, AssemblyAI, Deepgram and OpenAI need a personal key: in **Settings** → **Engines** you find a field for each and a “get a key” link. Signing up to Groq is free; AssemblyAI and Deepgram give credits on sign-up.

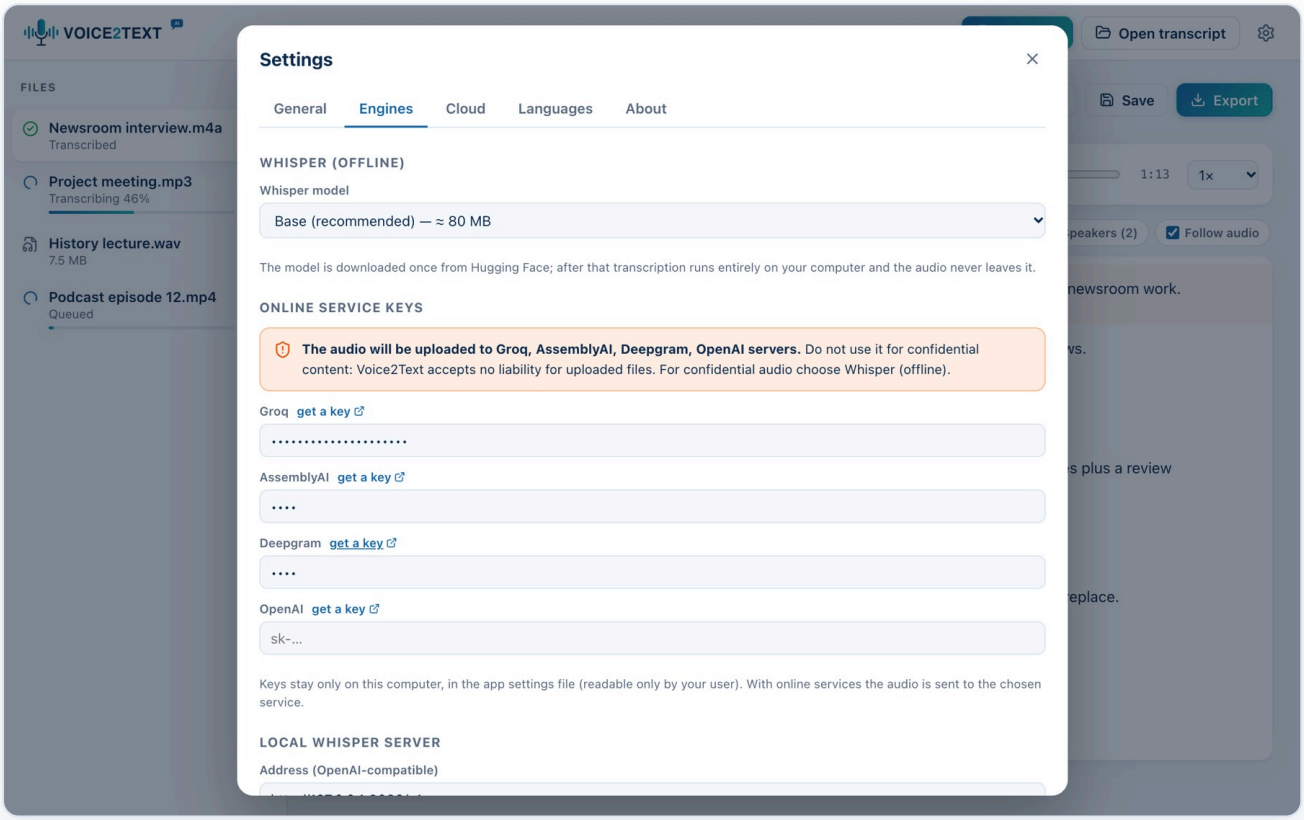


Figure 3. Settings → Engines: Whisper model and service keys.

Press **Transcribe** to start. With several files waiting, **Transcribe all** also appears in the sidebar.

05 Online services: your audio leaves the computer

> When you use an online AI service, the audio file is uploaded over the Internet. Voice2Text tells you so clearly and asks for your confirmation.

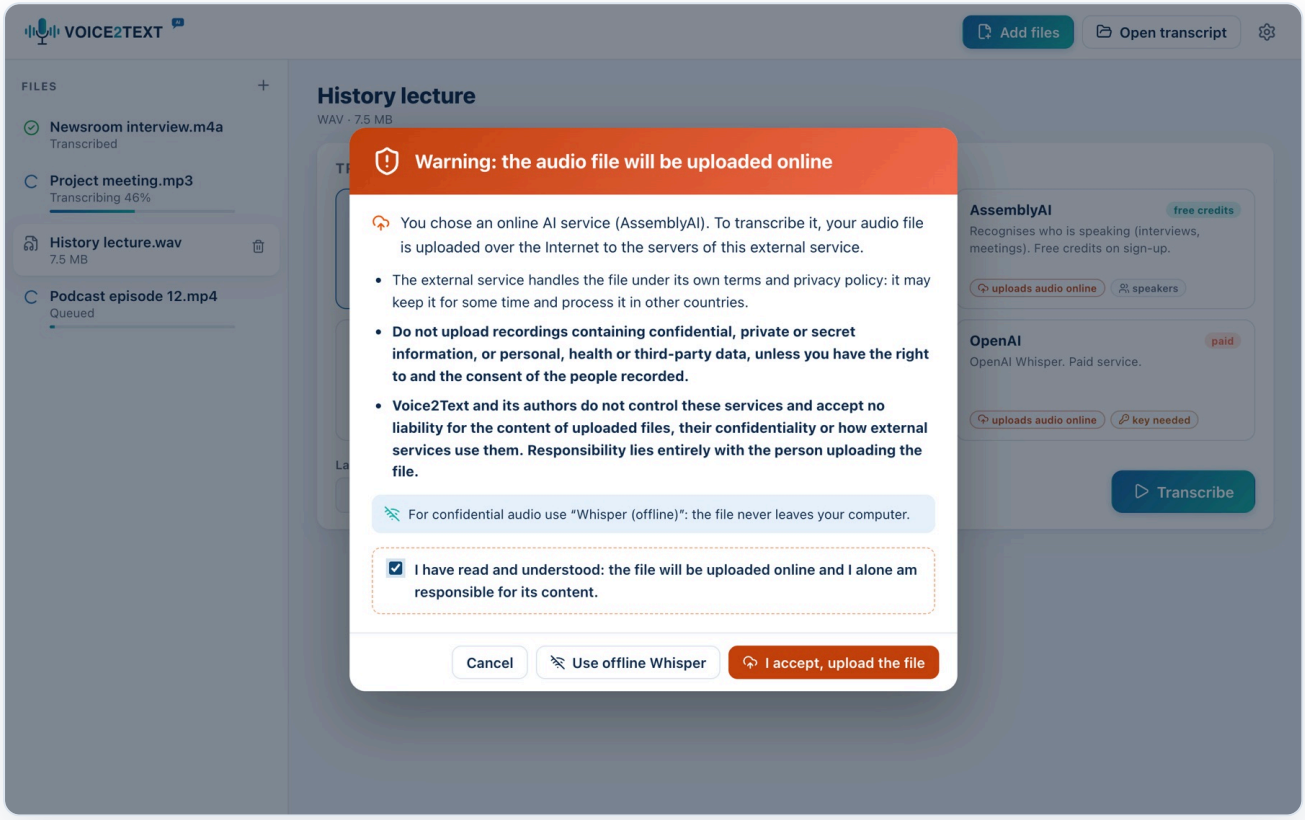


Figure 4. The mandatory warning before the first upload to an online service.

The first time you use Groq, AssemblyAI, Deepgram or OpenAI this warning appears. To continue you must tick **"I have read and understood"**; alternatively you can switch straight to **offline Whisper**.

> **The audio file is uploaded to the external service's servers**, which handle it under their own terms and privacy policy: they may keep it for some time and process it in other countries.
>
> **Do not upload recordings with confidential information**, private or secret information, or personal, health or third-party data, unless you have the right to and the consent of the people recorded.
>
> **Voice2Text and its authors accept no liability** for the content of uploaded files, their confidentiality or how external services use them. Responsibility lies entirely with the person uploading the file.

Acceptance applies to that single service and is recorded with the date. While an online service is selected, an orange box reminds you of it. In **Settings** → **About** you can reread the warning and make it appear again.

Confidential audio? Use **Whisper (offline)**: transcription happens entirely on your computer.

06 During transcription

> Voice2Text shows at every moment how far the work has got.

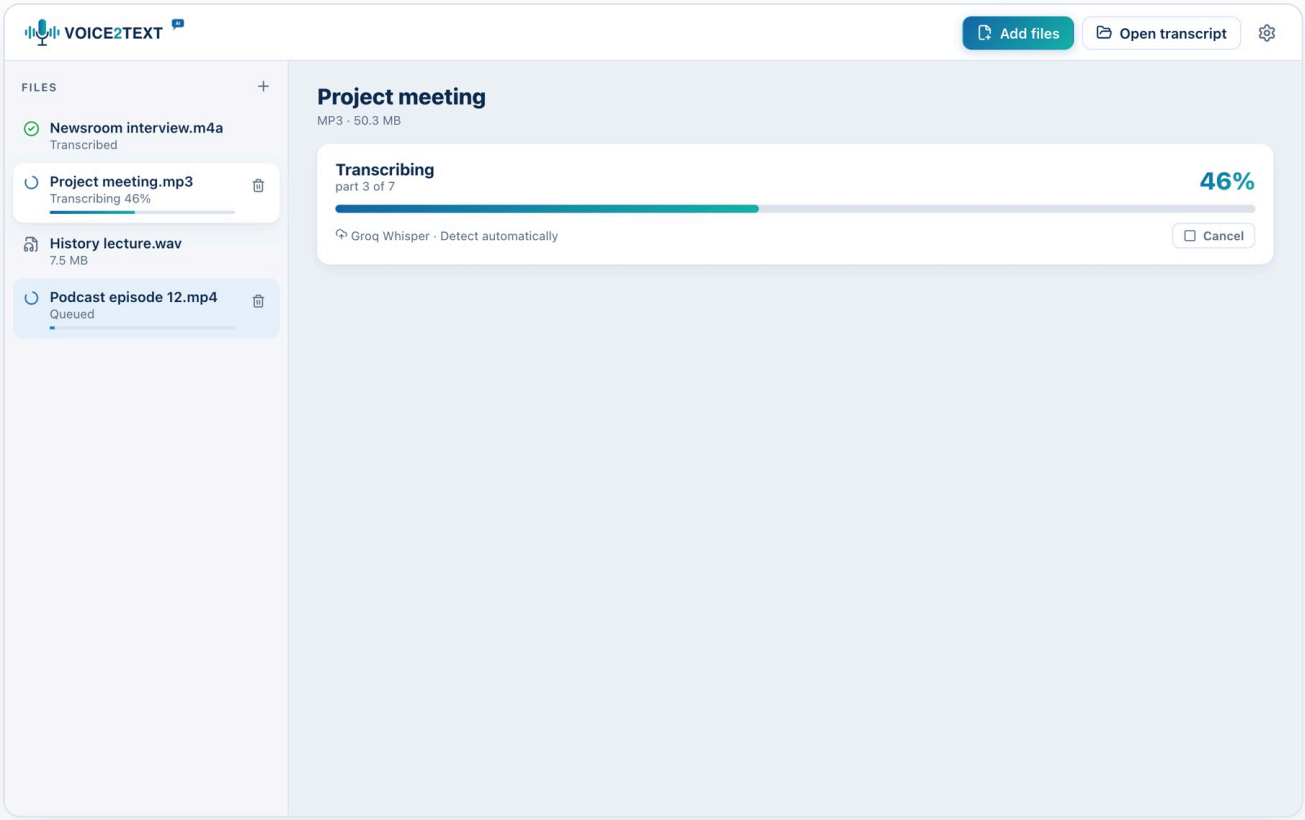


Figure 5. A transcription in progress with Groq: progress and part being processed.

- **Reading audio**: the file is decoded on your computer.
- **Preparing the model** (offline Whisper only, the first time): the chosen model is downloaded and kept.
- **Uploading audio** and **Transcribing**: with long files you see the part being processed (“part 3 of 7”).
- **Waiting for the service**: on free plans, if the request limit is exceeded, Voice2Text waits and retries by itself.

You can cancel at any time with **Cancel**, keep working on other files and leave the transcription running: when the text is ready you get a notification.

07 Editing the transcript

> The text is split into sentences, each with its time and, when recognised, the person speaking.

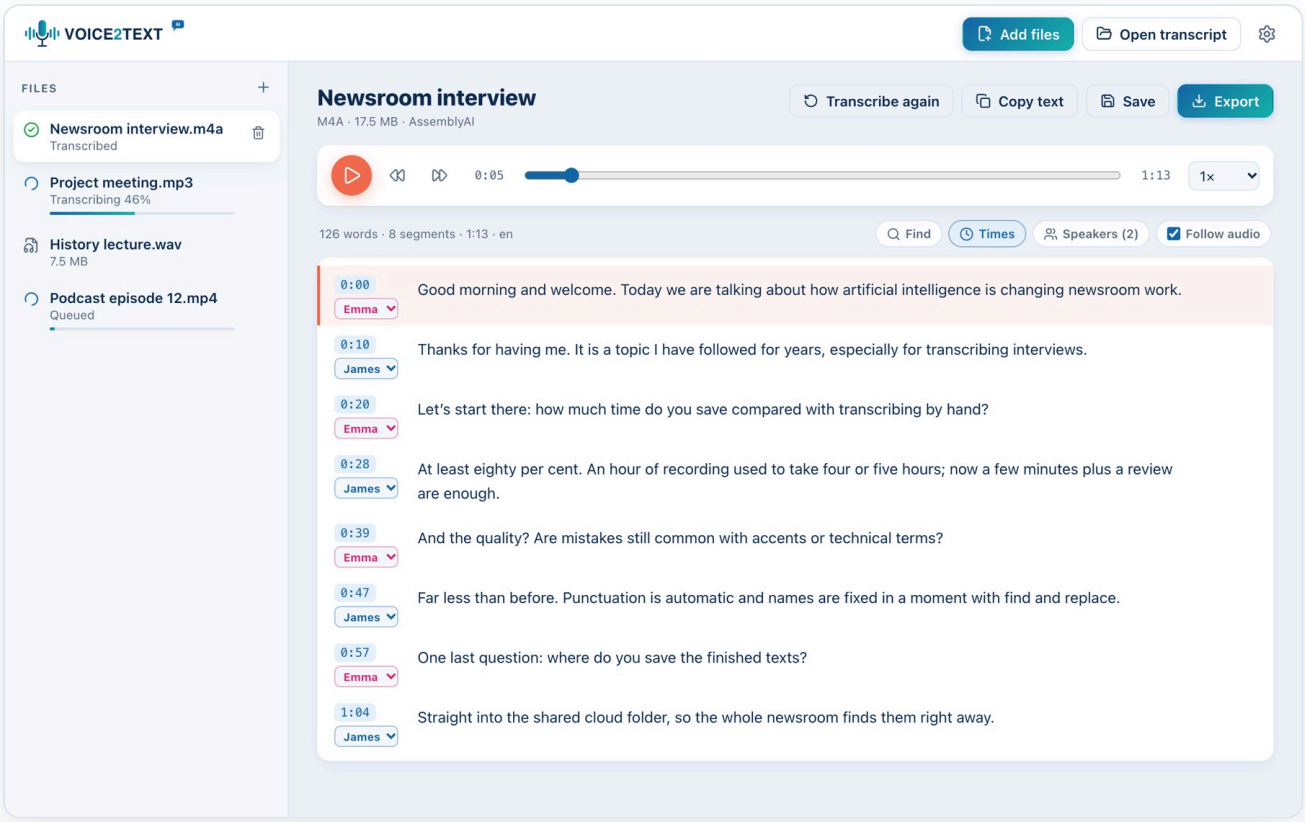


Figure 6. The editor: audio player at the top and editable sentences; the sentence being played is highlighted.

Listening

The player at the top plays the original audio: **Space** plays and pauses (when you are not typing), the buttons jump 5 seconds and you can pick a speed from 0.75× to 2×. Click the time of a sentence to hear it; with **Follow audio** the sentence being played always stays in view.

Editing

- Click a sentence and type: changes are immediate.
- **Enter** splits the sentence at the cursor; **Shift+Enter** adds a line break inside the sentence.
- **Backspace** at the start of a sentence merges it with the previous one; the **Merge** button joins it with the next one.
- The bin deletes a sentence (for example a noise recognised as a word).

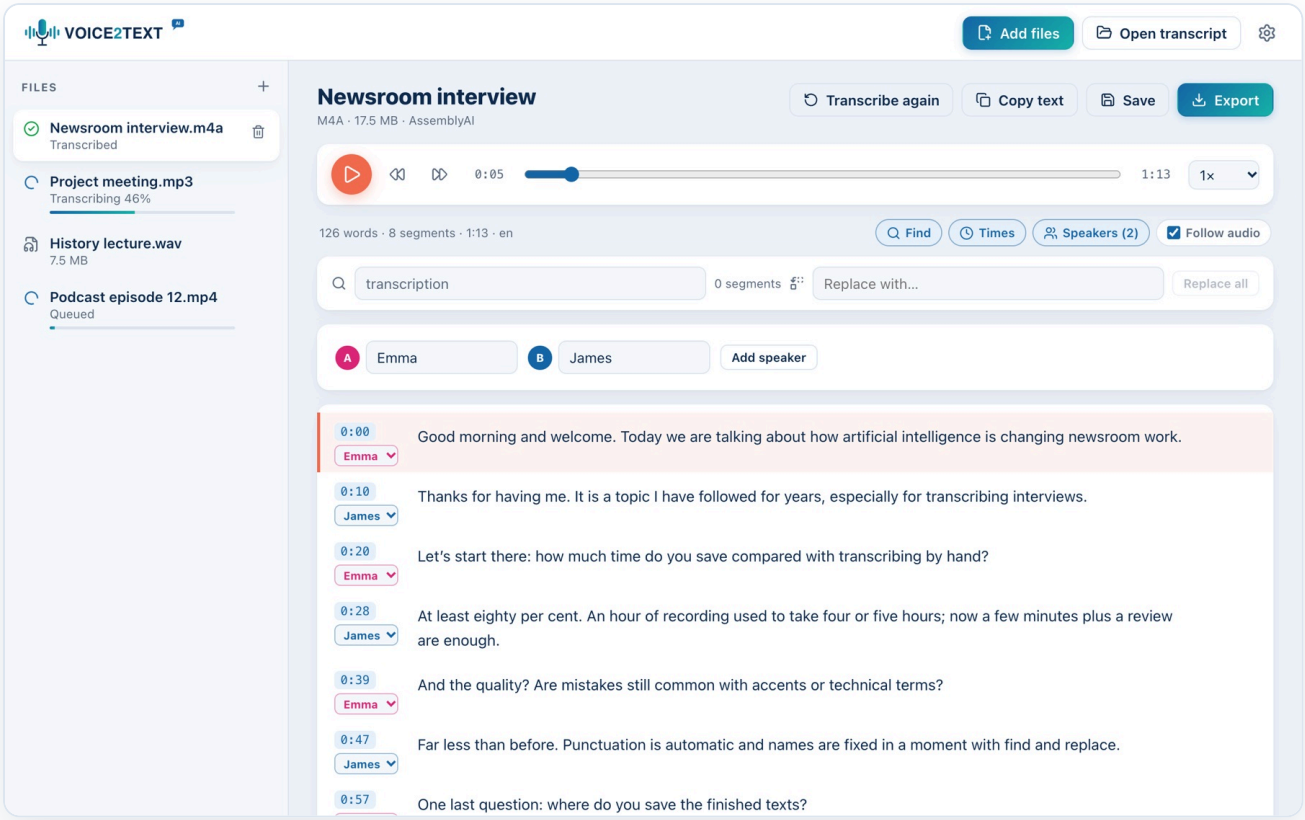


Figure 7. Find and replace, and the names of the people speaking.

Find and replace

Find (⌘F / Ctrl+F) highlights the sentences containing the word; **Replace all** fixes a misspelled name throughout the transcript in one go.

Speakers

With AssemblyAI and Deepgram each sentence has a label (Speaker A, Speaker B…). In the **Speakers** panel type the real name: it appears everywhere, including exported files. You can change a sentence's speaker from the menu next to its time, add more speakers or assign them by hand even when the engine does not recognise them.

08 Saving and exporting

> Save the project to continue editing later and export the text in the format you need.

The .v2t project

Save (⌘S / Ctrl+S) creates a `.v2t` file with the transcript, your edits and the speakers' names. Reopen it with **Open transcript** (⇧⌘O / Ctrl+Shift+O) or by dropping it onto the window. ⇧⌘S saves under another name.

Exporting

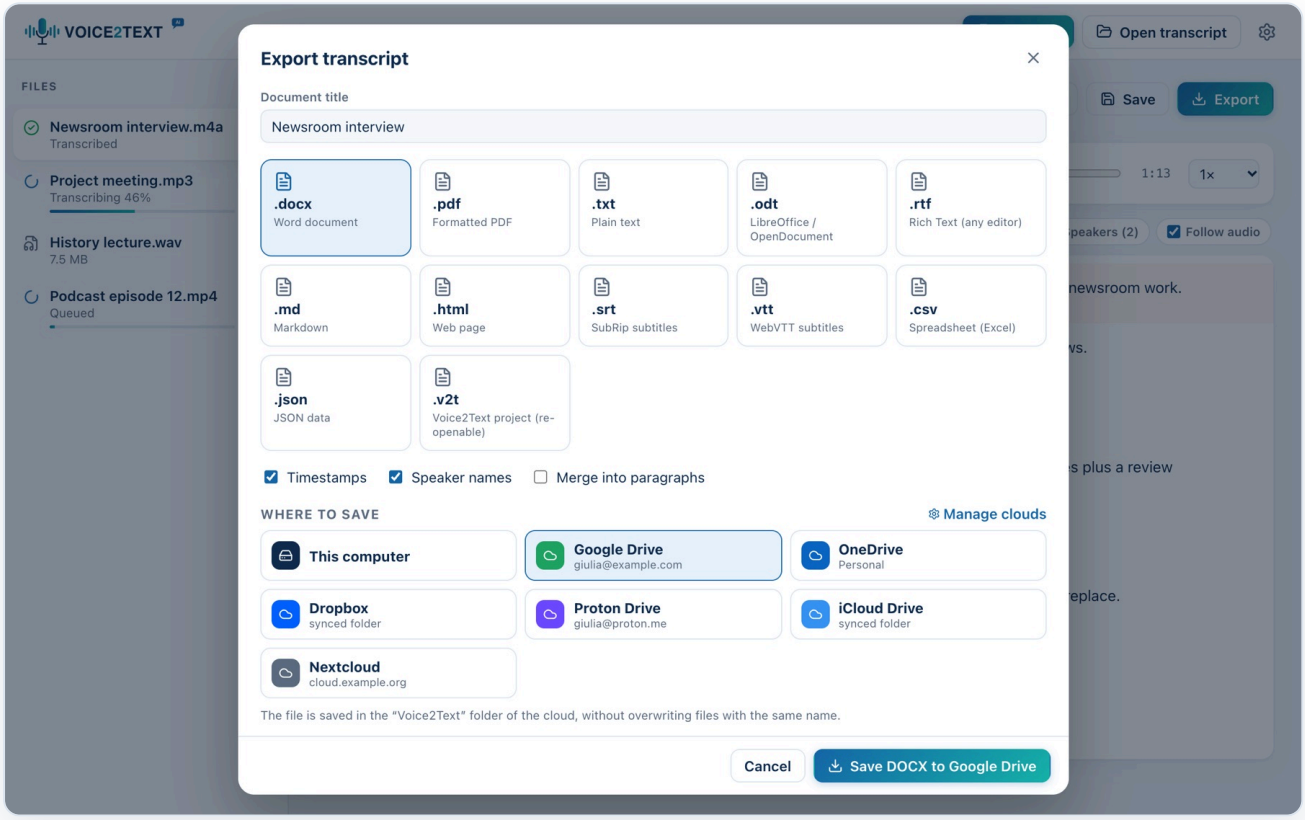


Figure 8. The Export window: format, options and destination.

Press **Export** (⌘E / Ctrl+E), type the document title and choose the format:

Format	Best for
.docx Word · **.odt** LibreOffice · **.rtf**	documents to edit, laid out with title and recording details
.pdf	documents to print or send; created through the system print dialog (“Save as PDF”)
.txt · **.md** · **.html**	plain text, Markdown, web page
.srt · **.vtt**	video subtitles (YouTube, Premiere, DaVinci, VLC…)
.csv · **.json**	spreadsheets (Excel, Numbers) and data for other programs
.v2t	the re-openable project, which you can also save to the cloud

The options choose whether to include **timestamps**, **speaker names** and whether to **merge sentences into paragraphs**. Arabic and other right-to-left text is laid out correctly.

09 Saving to the cloud

> Voice2Text saves files straight into your cloud using the folders that the services' apps create on your computer. You do not need to connect accounts or enter passwords.

How it works

When you install the app of Google Drive, OneDrive, Dropbox, Box, **Proton Drive**, iCloud Drive, pCloud, MEGA, Nextcloud or Sync.com, a synced folder appears on your computer. Voice2Text finds it by itself: in the **Export** window, under **Where to save**, choose the cloud and press the button. The file is written into the `Voice2Text` subfolder and the service's app uploads it right away.

- Files are never overwritten: if a file with the same name already exists, the new one becomes "name (2)".
- You can save any format to the cloud, including the `.v2t` project.
- The PDF goes through the print dialog: there choose "Save as PDF" and the cloud folder.

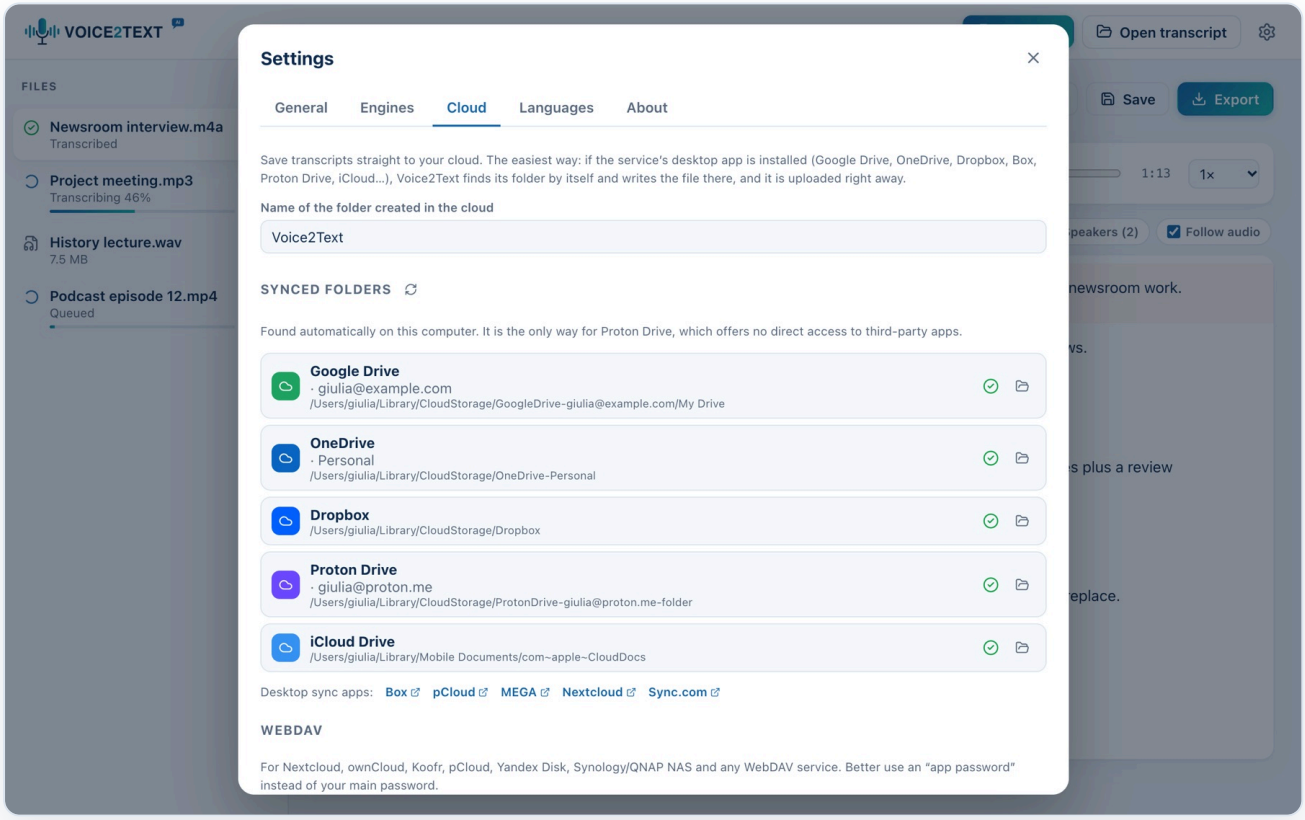


Figure 9. Settings → Cloud: the folders found on the computer and the WebDAV connection.

Settings → Cloud

- **Folder name**: the subfolder created in the cloud (default `Voice2Text`).
- **Synced folders**: the clouds found; the folder icon opens it. If a service is missing, the links below lead to the apps to install.
- **Add another folder**: for other clouds or a network drive (NAS).

- **WebDAV**: for Nextcloud, ownCloud, Koofr, pCloud, Yandex Disk and Synology/QNAP NAS without an installed app. Choose the service, enter address, user and an app password, then press **Test connection**.

> **Proton Drive** does not let other apps access the account directly: that is why Voice2Text uses its synced folder, which keeps the same end-to-end encryption.

10 Settings and languages

> The gear at the top right (`⌘,` / `Ctrl+,`) opens the settings.

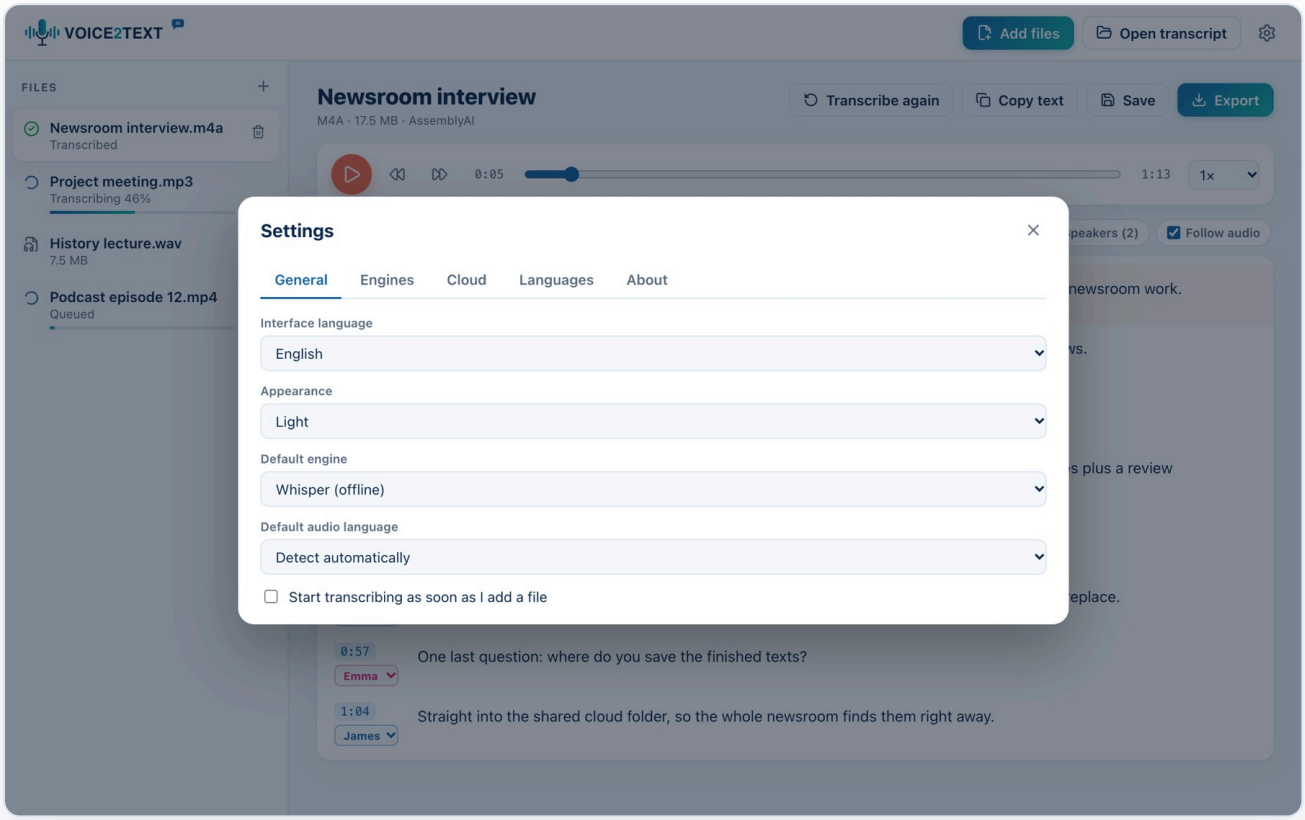


Figure 10. Settings → General.

- **General**: interface language, light, dark or automatic appearance, default engine and audio language, automatic start of transcription.
- **Engines**: offline Whisper model (tiny, base, small, large turbo), space used by models, service keys, local Whisper server, ffmpeg detection.
- **Cloud**: synced folders and WebDAV (previous chapter).
- **Languages**: the interface translation files.
- **About**: version, licence, warning about online services.

Languages are XML files

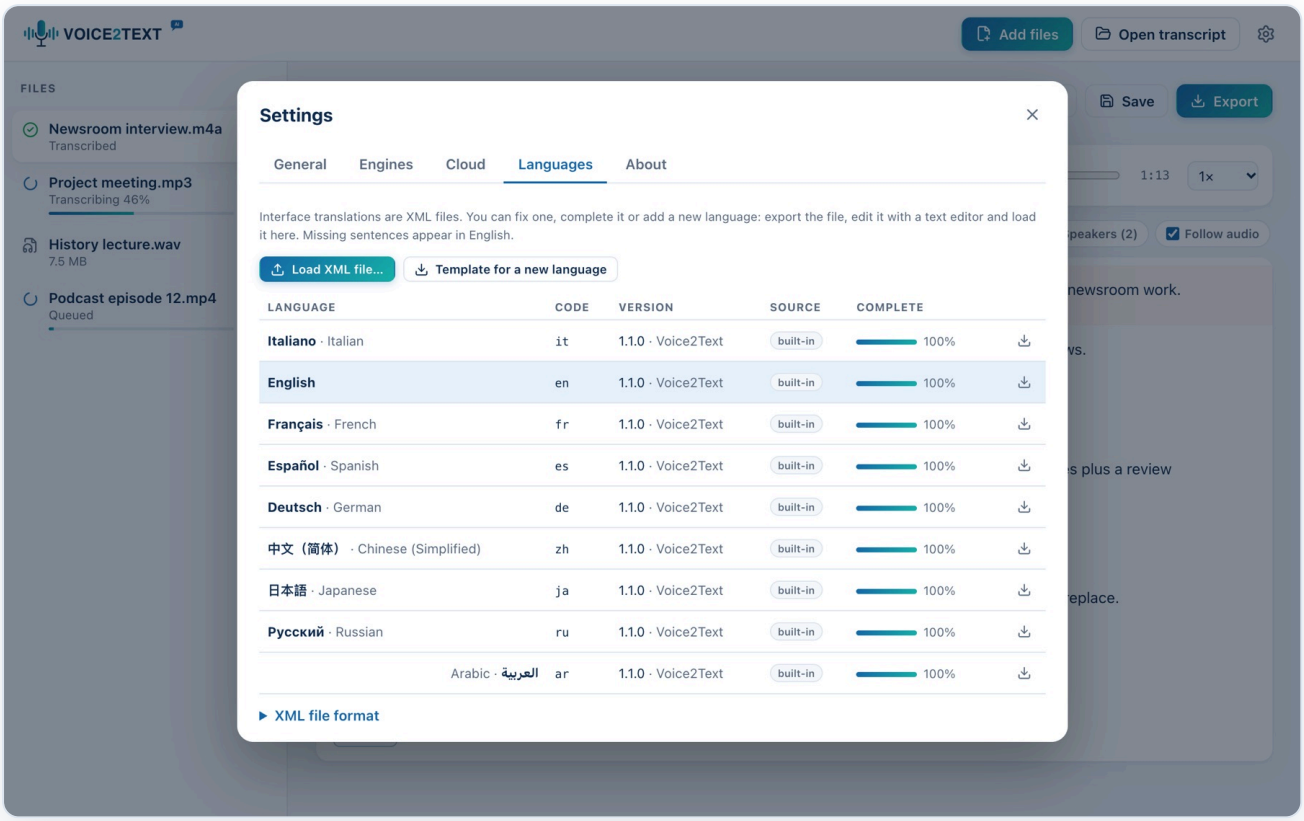


Figure 11. Settings → Languages: available languages, completeness and tools for XML files.

Each interface language is an XML file. To fix a translation or add a new one:

1. Export the language file (arrow in its row) or download the **Template for a new language**.
2. Open it in a text editor and translate only the text between the tags, keeping the `key` attributes and curly-brace placeholders such as `{n}` unchanged.
3. Press **Load XML file...**: the language immediately appears in the list and in the language menu.

```
<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>
```

A file with the same code as a built-in language replaces it (for corrections); you can always go back to the original. Missing sentences appear in English; the **Complete** column shows how many are translated. For right-to-left languages use `dir="rtl"`.

11 Keyboard shortcuts

> On Windows and Linux use **Ctrl** instead of ⌘.

Keys	Action
⌘O	Add audio or video files
⇧⌘O	Open a saved transcript (.v2t)
⌘S · ⇧⌘S	Save · Save as
⌘E	Export
⌘F	Find and replace
⌘,	Settings
Space	Play / pause (outside the text)
Enter · Shift+Enter	Split sentence · Line break in sentence
Backspace at sentence start	Merge with previous sentence
Esc	Leave the text · close the open window

12 Frequently asked questions

> Answers to the most common questions.

My file's format cannot be read.

Install ffmpeg (Installation chapter) or use AssemblyAI or Deepgram, which accept any format because they receive the original file.

Offline Whisper is slow.

Choose the **Base** or **Tiny** model in Settings → Engines. "Large turbo" is the most accurate but takes time without a graphics card. For long files Groq is much faster.

The text is in another language or translated.

Set the audio language instead of automatic detection and use **Transcribe again**.

"Invalid service key".

Check the key in Settings → Engines: copy it again from the service's website, without spaces.

My cloud does not appear.

Make sure the service's app is installed and syncing, then press **Search again** in Settings → Cloud. Alternatively add the folder by hand.

The player says "audio unavailable".

The original audio file was moved or renamed after the project was saved. The text is still available; to listen again, put the file back where it was.

Where are my keys and settings?

Only on your computer, in the app's settings file, readable only by your user. They are never sent to Voice2Text.



VOICE2TEXT

From voice to text, in any language.

Voice2Text is a program by overminder. PolyForm Strict 1.0.0 licence: non-commercial use allowed; redistribution, modifications and derivative works not allowed.

© 2026 overminder · Version 1.1