VOICE2TEXT AI

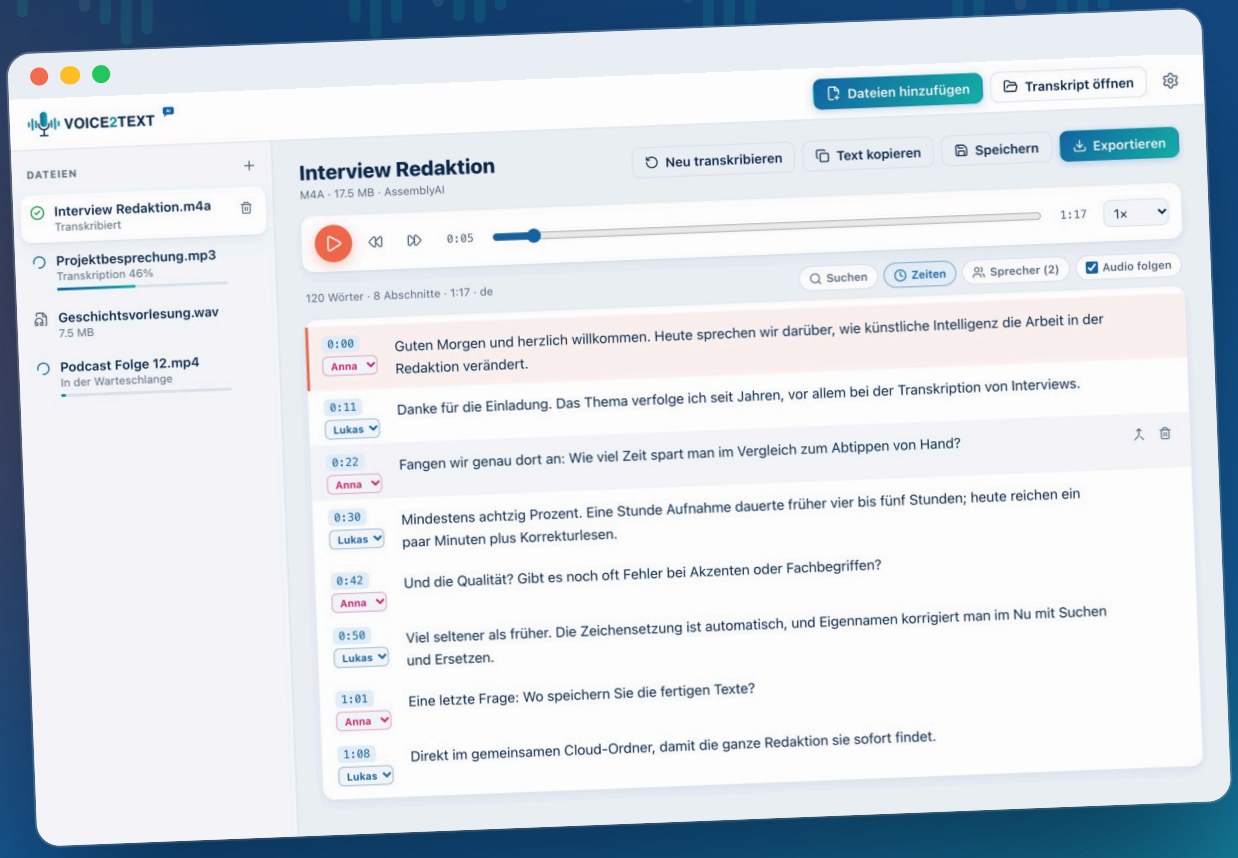
BENUTZERHANDBUCH

Voice2Text

Von der Stimme zum Text, in jeder Sprache.

Audio- und Videodateien mit künstlicher Intelligenz transkribieren, den Text korrigieren und speichern, wo Sie möchten.

Offline-Whisper | 12 Formate | Cloud | 9 Sprachen



Version 1.1 | macOS · Windows · Linux | Deutsch

Inhalt

01	Willkommen bei Voice2Text	3
02	Installation	4
03	Das Hauptfenster	5
04	Dienst und Sprache wählen	6
05	Online-Dienste: Ihr Audio verlässt den Computer	8
06	Während der Transkription	10
07	Das Transkript korrigieren	11
08	Speichern und exportieren	13
09	In der Cloud speichern	15
10	Einstellungen und Sprachen	17
11	Tastenkürzel	19
12	Häufige Fragen	20

01 Willkommen bei Voice2Text

> Voice2Text verwandelt Audio- und Videoaufnahmen in geschriebenen Text. Interviews, Besprechungen, Vorlesungen, Podcasts, Sprachnachrichten: Datei ins Fenster ziehen, KI-Dienst wählen, und nach wenigen Minuten liegt der Text vor – bereit zum Korrigieren und Speichern.

Was Sie tun können

- **Jedes Format.** MP3, M4A, WAV, FLAC, OGG, AAC, AIFF, Videos in MP4, MOV, MKV, WebM und viele mehr; mit installiertem ffmpeg praktisch alle.
- **Auch ohne Internet.** Whisper läuft direkt auf dem Computer: Das Audio verlässt nie Ihren Schreibtisch.
- **Wer spricht.** Mit AssemblyAI und Deepgram werden die Stimmen getrennt, und Sie können jeder Person einen Namen geben.
- **Bequemes Korrigieren.** Der Text ist in Sätze unterteilt, die mit dem Audio verknüpft sind: ein Klick, und Sie hören genau die Stelle noch einmal.
- **Jedes Ausgabeformat.** Word, PDF, ODT, RTF, TXT, Markdown, HTML, SRT- und VTT-Untertitel, CSV, JSON und das wieder öffnbare Projekt.
- **Direkt in die Cloud.** Google Drive, OneDrive, Dropbox, Box, Proton Drive, iCloud und weitere, über ihre Ordner auf dem Computer.
- **Neun Sprachen.** Die Oberfläche spricht Italienisch, Englisch, Französisch, Spanisch, Deutsch, Chinesisch, Japanisch, Russisch und Arabisch; die Übersetzungen sind XML-Dateien, die Sie korrigieren können.

Systemvoraussetzungen

macOS 11 oder neuer (Mac mit Apple- oder Intel-Chip), Windows 10/11 64 Bit oder Linux 64 Bit (Ubuntu, Debian, Fedora und Derivate). Für Offline-Whisper sind mindestens 8 GB Arbeitsspeicher empfehlenswert; Online-Dienste brauchen nur eine Verbindung.

02 Installation

> Laden Sie die Datei für Ihr System herunter und installieren Sie sie wie jedes andere Programm.

macOS

Öffnen Sie `Voice2Text_1.1.1_universal.dmg` und ziehen Sie das Voice2Text-Symbol in den Ordner **Programme**. Beim ersten Start meldet macOS eventuell, dass der Entwickler nicht verifiziert ist: Öffnen Sie **Systemeinstellungen** → **Datenschutz & Sicherheit** und wählen Sie **Dennoch öffnen**, oder klicken Sie mit der rechten Maustaste auf die App und wählen Sie **Öffnen**. Das ist nur einmal nötig.

Windows

Starten Sie `Voice2Text_1.1.1_x64-setup.exe`. Erscheint das blaue Fenster von Windows SmartScreen, klicken Sie auf **Weitere Informationen** und dann auf **Trotzdem ausführen**. Für die Installation sind keine Administratorrechte nötig.

Linux

Unter Ubuntu und Debian installieren Sie das Paket `.deb` (Doppelklick oder `sudo apt install ./Voice2Text_1.1.1_amd64.deb`), unter Fedora das Paket `.rpm`. Alternativ machen Sie die `.AppImage` ausführbar (`chmod +x`) und starten sie.

> **Optional: ffmpeg.** Voice2Text liest die gängigen Formate selbst. Für seltene Formate (OPUS, WMA, AMR, AVI…) installieren Sie ffmpeg: `brew install ffmpeg` auf dem Mac, `winget install ffmpeg` unter Windows, `sudo apt install ffmpeg` unter Linux. Es wird automatisch gefunden.

03 Das Hauptfenster

> Die ganze Arbeit findet in einem einzigen Fenster statt: links die Dateiliste, in der Mitte die ausgewählte Datei.

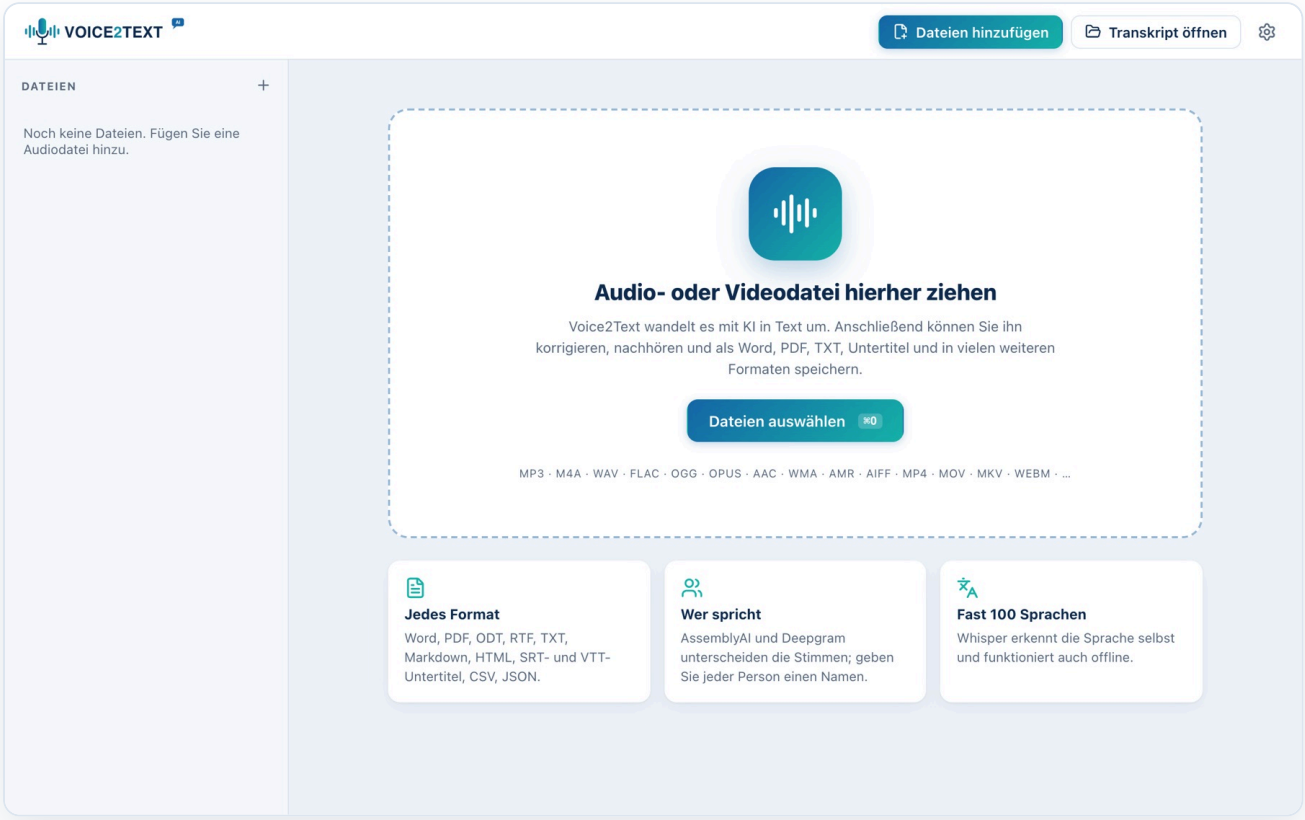


Abbildung 1. Der Startbildschirm: Datei in den Rahmen ziehen oder „Dateien auswählen" verwenden.

- **Obere Leiste**: **Dateien hinzufügen**, **Transkript öffnen** (ein gespeichertes `.v2t`-Projekt) und das Zahnrad der **Einstellungen**.
- **Dateiliste** links: Jede Datei zeigt ihren Status (bereit, in der Warteschlange, Transkription, transkribiert, Fehler) und den Fortschritt. Ein Punkt hinter dem Namen bedeutet ungespeicherte Änderungen.
- **Mittlerer Bereich**: die Einstellungen der Transkription, der Fortschritt oder der zu korrigierende Text.

Dateien hinzufügen

Ziehen Sie eine oder mehrere Audio- oder Videodateien auf das Fenster oder verwenden Sie **Dateien hinzufügen** (⌘O / Strg+O). Sie können beliebig viele hinzufügen: Sie werden nacheinander transkribiert. Eine hineingezogene `.v2t`-Datei öffnet ein gespeichertes Transkript.

04 Dienst und Sprache wählen

> Wählen Sie vor dem Transkribieren, wer die Arbeit erledigt: Whisper auf Ihrem Computer oder ein Online-Dienst.

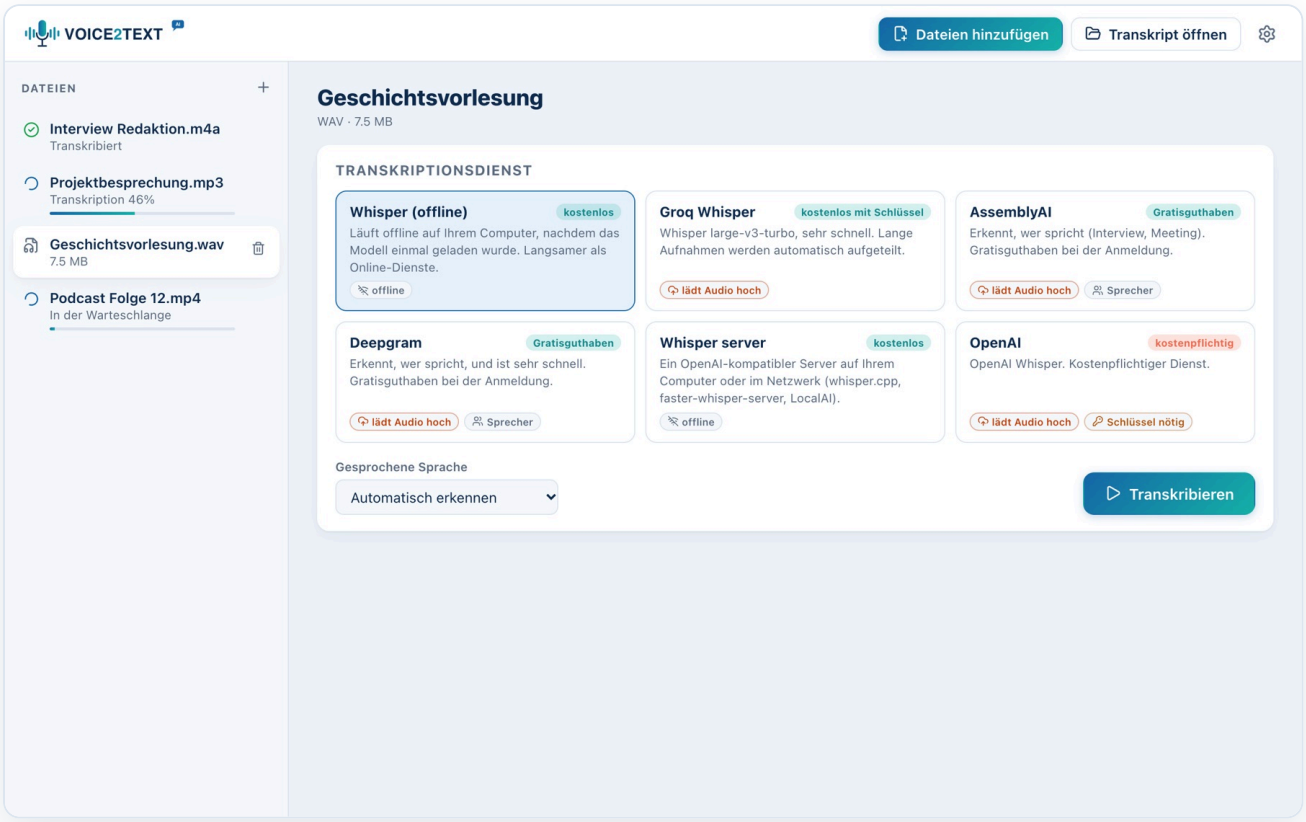


Abbildung 2. Auswahl von Dienst und Sprache für die gewählte Datei.

Dienst	Kosten	Eigenschaften
Whisper (offline)	kostenlos	Funktioniert ohne Internet; das Modell wird einmal geladen (45 bis 600 MB). Langsamer als Online-Dienste.
Groq Whisper	kostenlos mit Schlüssel	Sehr schnell; lange Aufnahmen werden automatisch aufgeteilt.
AssemblyAI	Gratisguthaben	Erkennt, wer spricht. Ideal für Interviews und Besprechungen.
Deepgram	Gratisguthaben	Erkennt, wer spricht, und ist sehr schnell.
Whisper server	kostenlos	Ein Whisper-Server auf Ihrem Computer oder im Netzwerk (whisper.cpp, faster-whisper-server, LocalAI).
OpenAI	kostenpflichtig	OpenAI Whisper.

Die Sprache der Aufnahme

Mit **Automatisch erkennen** ermittelt Voice2Text die gesprochene Sprache selbst. Wenn Sie sie kennen, wählen Sie sie aus: Das Ergebnis wird genauer, besonders in den ersten Sekunden.

Schlüssel für Online-Dienste

Groq, AssemblyAI, Deepgram und OpenAI benötigen einen persönlichen Schlüssel: Unter **Einstellungen → Dienste** finden Sie für jeden ein Feld und den Link „Schlüssel holen“. Die Anmeldung bei Groq ist kostenlos; AssemblyAI und Deepgram schenken bei der Anmeldung Guthaben.

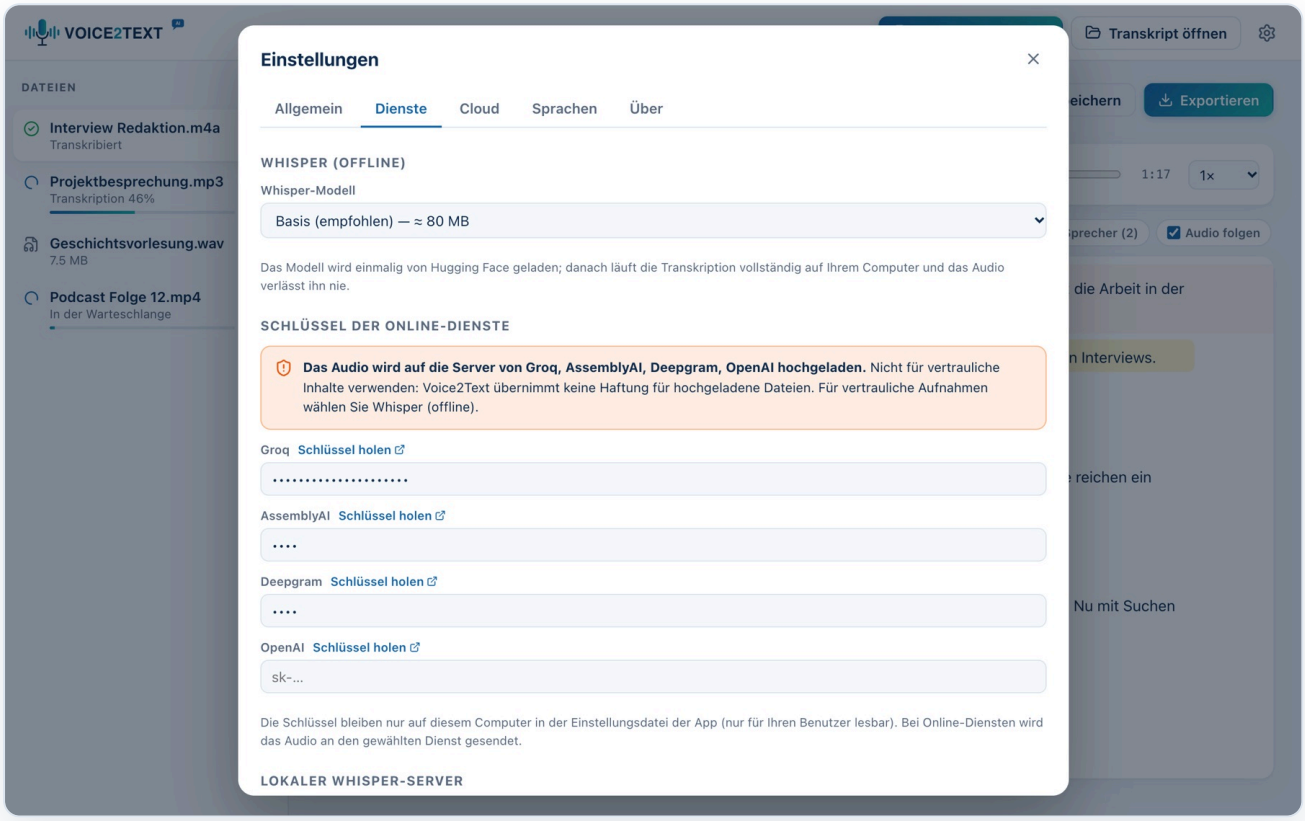


Abbildung 3. Einstellungen → Dienste: Whisper-Modell und Schlüssel der Dienste.

Drücken Sie **Transkribieren**, um zu beginnen. Warten mehrere Dateien, erscheint in der Seitenleiste auch **Alle transkribieren**.

05 Online-Dienste: Ihr Audio verlässt den Computer

> Wenn Sie einen Online-KI-Dienst nutzen, wird die Audiodatei über das Internet hochgeladen. Voice2Text sagt Ihnen das deutlich und bittet um Ihre Bestätigung.

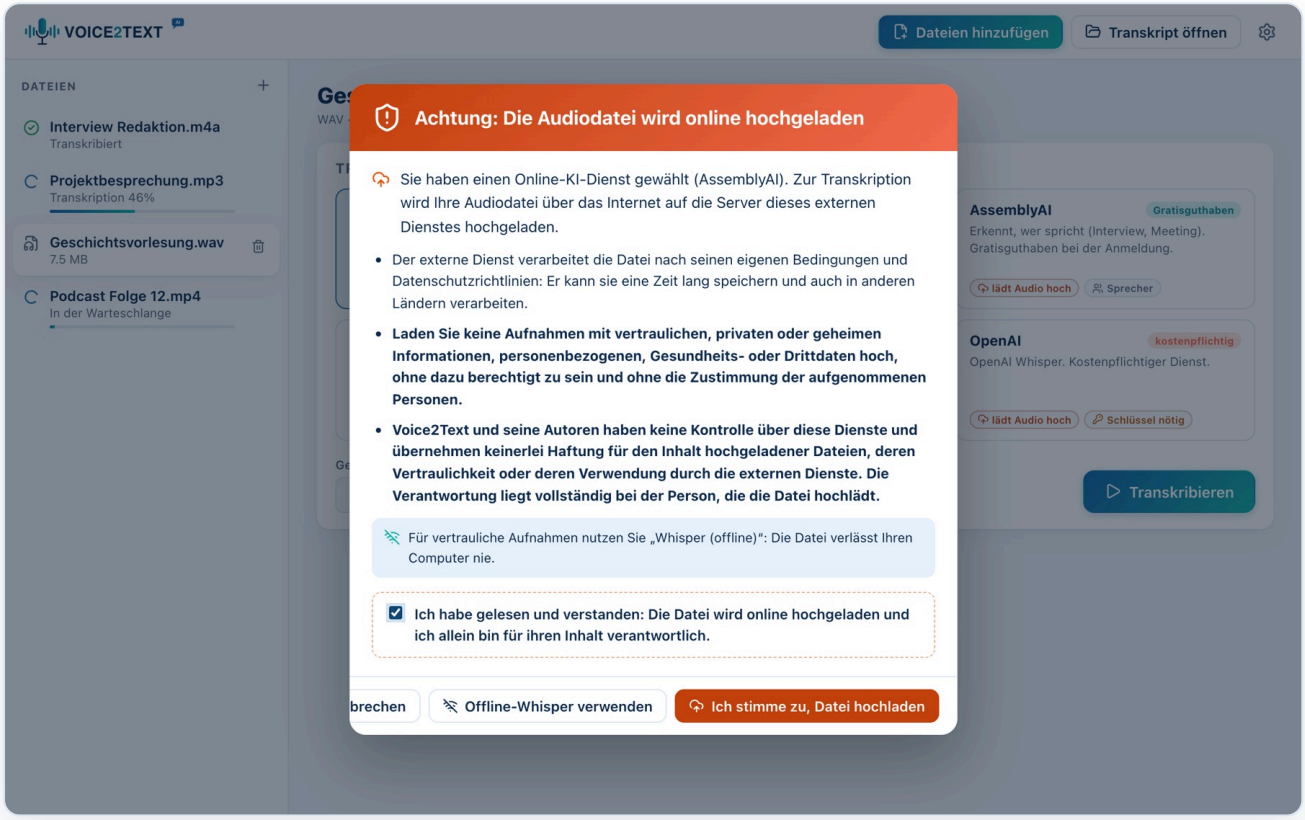


Abbildung 4. Der verpflichtende Hinweis vor dem ersten Hochladen zu einem Online-Dienst.

Wenn Sie Groq, AssemblyAI, Deepgram oder OpenAI zum ersten Mal verwenden, erscheint dieser Hinweis. Um fortzufahren, müssen Sie **„Ich habe gelesen und verstanden“** ankreuzen; alternativ können Sie direkt zu **Offline-Whisper** wechseln.

Die Audiodatei wird auf die Server des externen Dienstes hochgeladen, der sie nach seinen eigenen Bedingungen und Datenschutzrichtlinien verarbeitet: Er kann sie eine Zeit lang speichern und auch in anderen Ländern verarbeiten.

Laden Sie keine Aufnahmen mit vertraulichen Informationen hoch, keine privaten oder geheimen Inhalte und keine personenbezogenen, Gesundheits- oder Drittdaten, ohne dazu berechtigt zu sein und ohne die Zustimmung der aufgenommenen Personen.

Voice2Text und seine Autoren übernehmen keinerlei Haftung für den Inhalt hochgeladener Dateien, deren Vertraulichkeit oder deren Verwendung durch die externen Dienste. Die Verantwortung liegt vollständig bei der Person, die die Datei hochlädt.

Die Zustimmung gilt jeweils für einen Dienst und wird mit Datum gespeichert. Solange ein Online-Dienst gewählt ist, erinnert ein orangefarbener Kasten daran. Unter **Einstellungen** → **Über** können Sie den Hinweis erneut lesen und wieder anzeigen lassen.

> **Vertrauliche Aufnahme?** Verwenden Sie **Whisper (offline)**: Die Transkription erfolgt vollständig auf Ihrem Computer.

06 Während der Transkription

> Voice2Text zeigt jederzeit, wie weit die Arbeit ist.

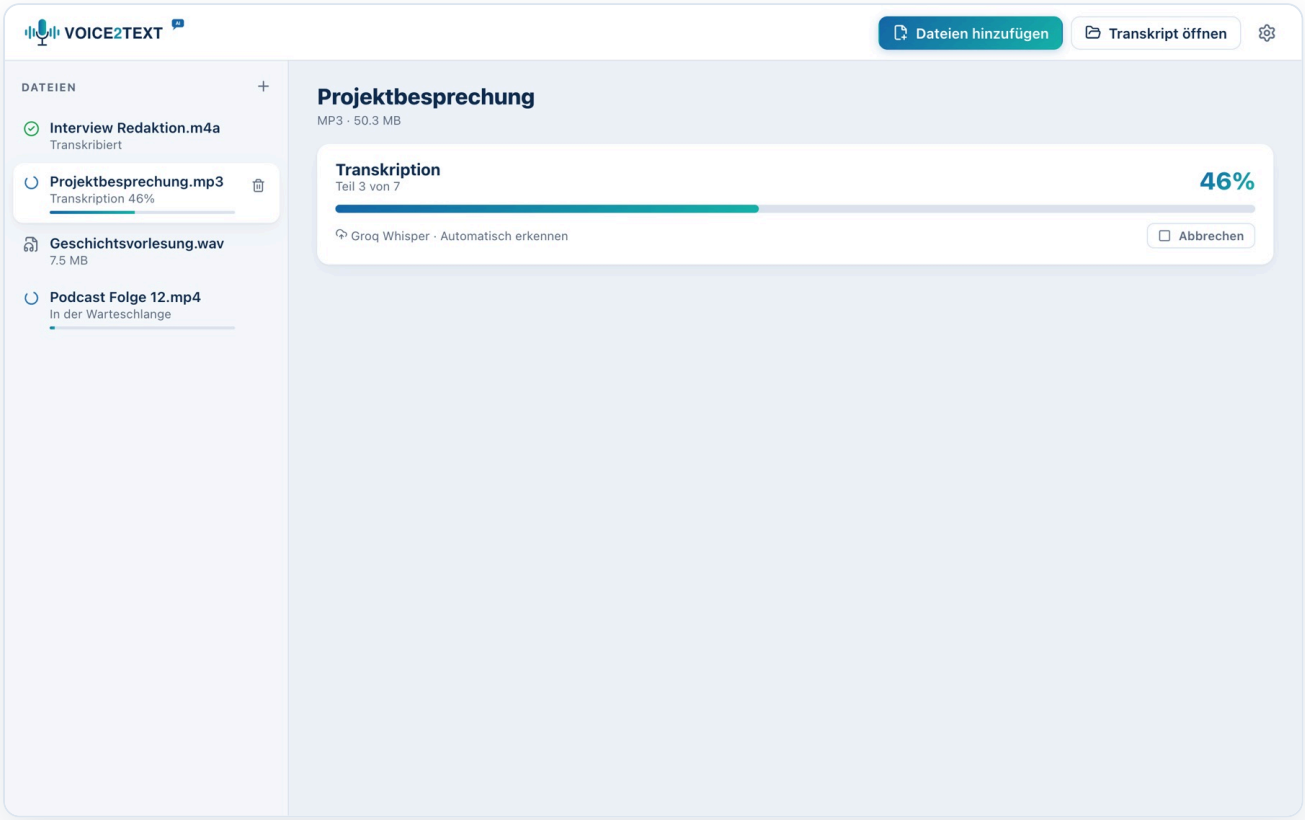


Abbildung 5. Eine laufende Transkription mit Groq: Fortschritt und bearbeiteter Teil.

- **Audio wird gelesen**: Die Datei wird auf dem Computer dekodiert.
- **Modell wird vorbereitet** (nur Offline-Whisper, beim ersten Mal): Das gewählte Modell wird geladen und gespeichert.
- **Audio wird hochgeladen** und **Transkription**: Bei langen Dateien sehen Sie den aktuellen Teil („Teil 3 von 7“).
- **Warten auf den Dienst**: Bei Gratisplänen wartet Voice2Text bei Überschreitung des Anfragelimits und versucht es selbstständig erneut.

Sie können jederzeit mit **Abbrechen** abbrechen, an anderen Dateien weiterarbeiten und die Transkription laufen lassen: Ist der Text fertig, erhalten Sie eine Benachrichtigung.

07 Das Transkript korrigieren

> Der Text ist in Sätze unterteilt, jeder mit seiner Zeit und – falls erkannt – der sprechenden Person.

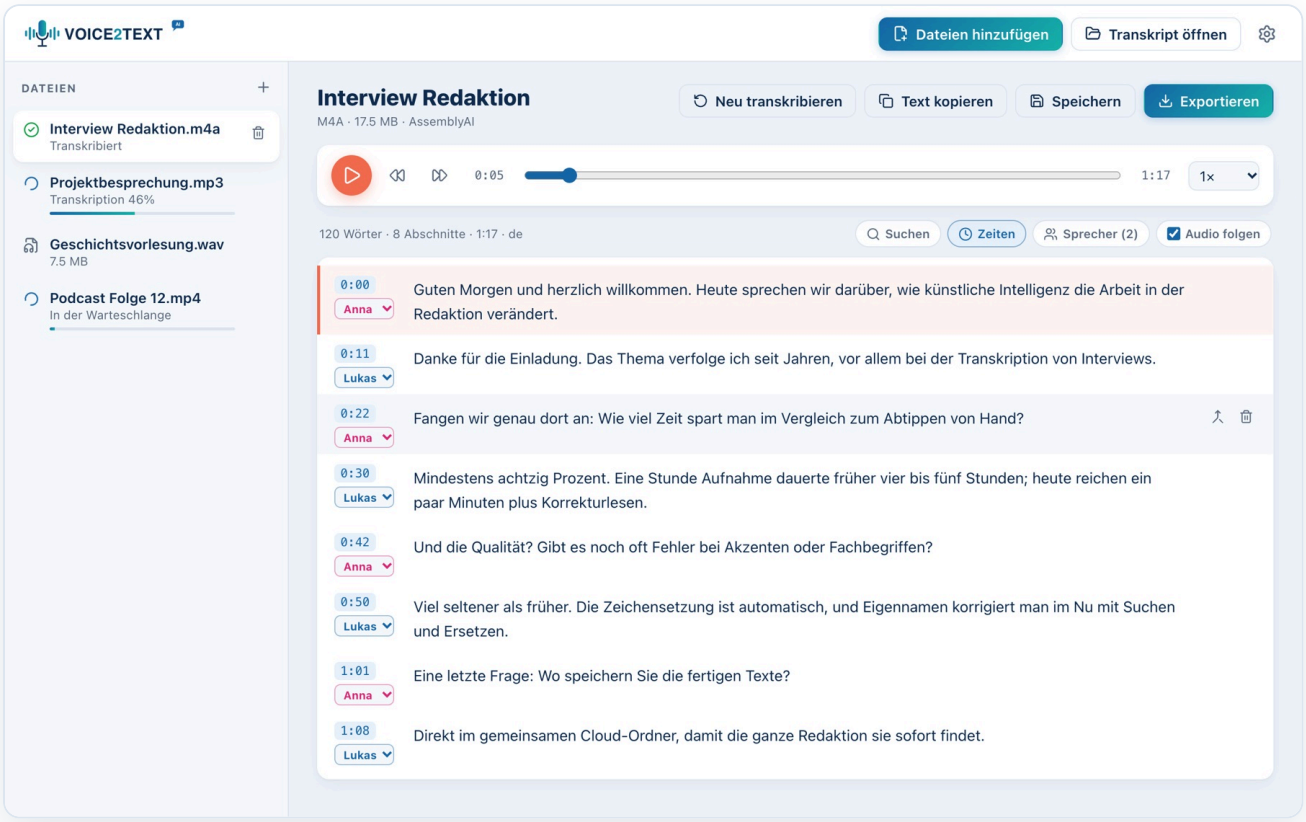


Abbildung 6. Der Editor: oben der Audioplayer, darunter bearbeitbare Sätze; der gerade abgespielte Satz ist hervorgehoben.

Anhören

Der Player oben spielt das Originalaudio ab: `Leertaste` startet und pausiert (wenn Sie nicht schreiben), die Tasten springen 5 Sekunden, und die Geschwindigkeit ist von 0,75× bis 2× wählbar. Klicken Sie auf die Zeit eines Satzes, um ihn zu hören; mit **Audio folgen** bleibt der abgespielte Satz immer sichtbar.

Bearbeiten

- Klicken Sie auf einen Satz und schreiben Sie: Änderungen gelten sofort.
- `Eingabe` teilt den Satz an der Cursorposition; `Umschalt+Eingabe` fügt im selben Satz einen Zeilenumbruch ein.
- `Rücktaste` am Satzanfang verbindet ihn mit dem vorherigen; die Schaltfläche **Verbinden** fügt ihn mit dem nächsten zusammen.
- Der Papierkorb löscht einen Satz (etwa ein Geräusch, das als Wort erkannt wurde).

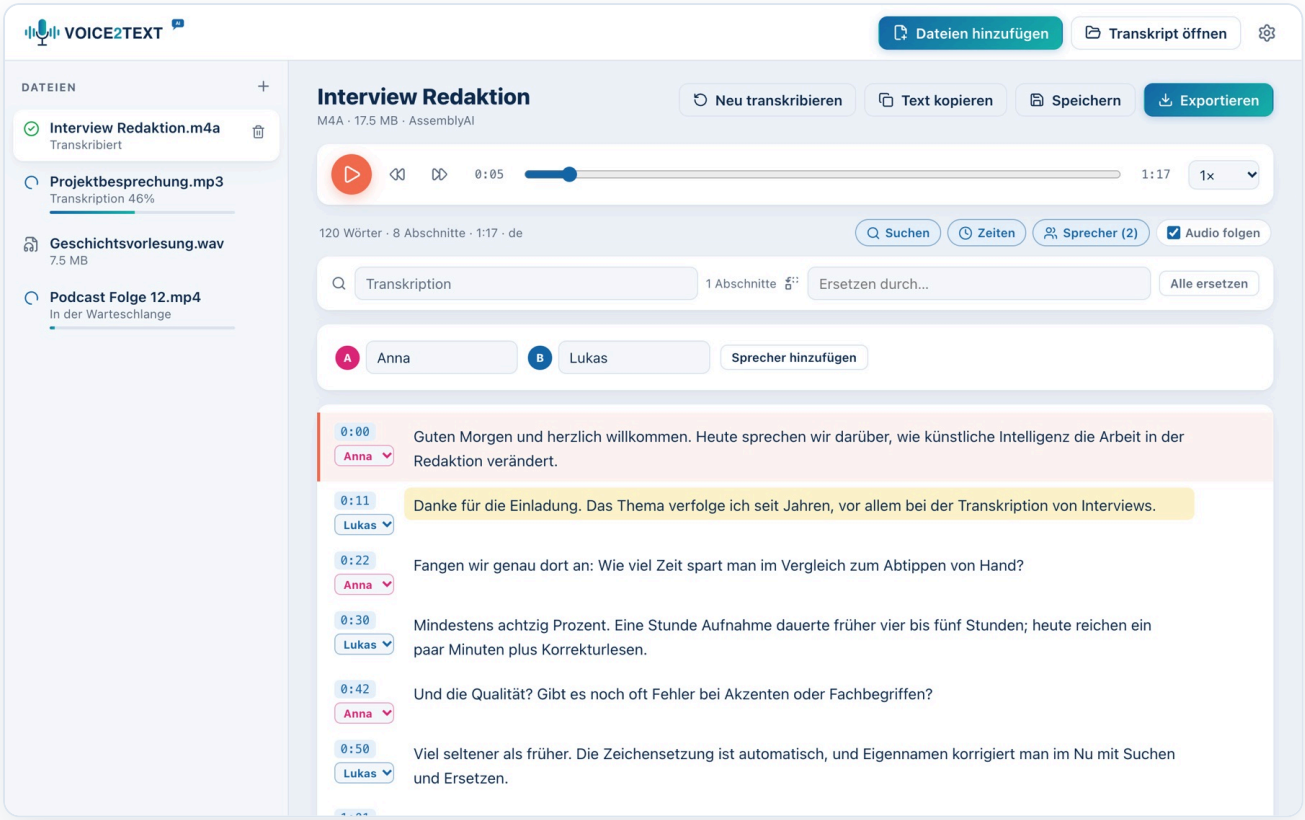


Abbildung 7. Suchen und Ersetzen sowie die Namen der Sprecher.

Suchen und Ersetzen

Suchen (⌘F / Strg+F) hebt die Sätze hervor, die das Wort enthalten; **Alle ersetzen** korrigiert einen falsch geschriebenen Namen im ganzen Transkript auf einmal.

Sprecher

Mit AssemblyAI und Deepgram hat jeder Satz eine Kennzeichnung (Sprecher A, Sprecher B…). Tragen Sie im Bereich **Sprecher** den echten Namen ein: Er erscheint überall, auch in exportierten Dateien. Den Sprecher eines Satzes ändern Sie im Menü neben der Zeit; Sie können weitere hinzufügen oder sie manuell zuordnen, auch wenn der Dienst sie nicht erkennt.

08 Speichern und exportieren

> Speichern Sie das Projekt, um später weiterzukorrigieren, und exportieren Sie den Text im benötigten Format.

Das .v2t-Projekt

Speichern (⌘S / Strg+S) erstellt eine `.v2t`-Datei mit dem Transkript, den Korrekturen und den Sprechernamen. Mit **Transkript öffnen** (⇧⌘O / Strg+Umschalt+O) oder durch Hineinziehen öffnen Sie es wieder. ⇧⌘S speichert unter anderem Namen.

Exportieren

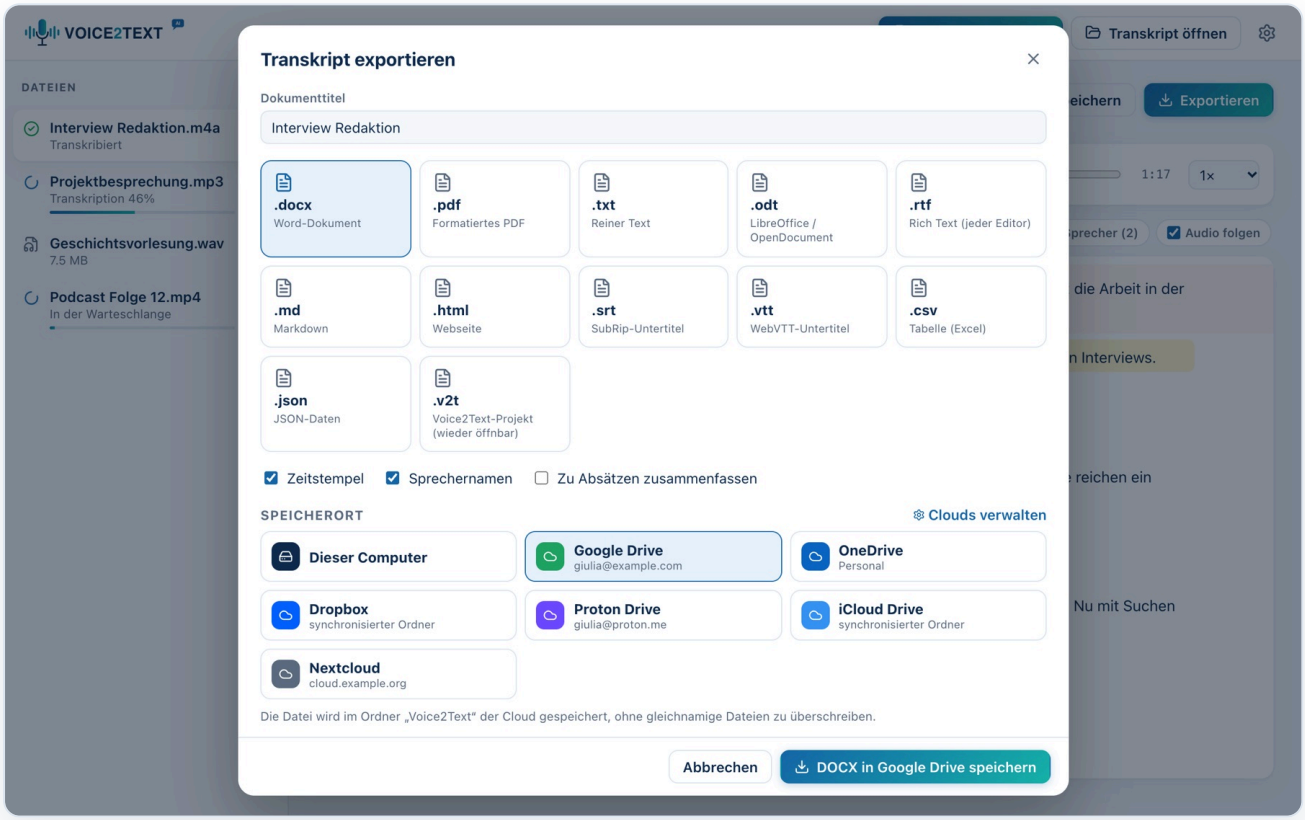


Abbildung 8. Das Fenster Exportieren: Format, Optionen und Speicherort.

Drücken Sie **Exportieren** (⌘E / Strg+E), geben Sie den Dokumenttitel ein und wählen Sie das Format:

Format	Ideal für
.docx Word · **.odt** LibreOffice · **.rtf**	bearbeitbare Dokumente mit Titel und Angaben zur Aufnahme
.pdf	Dokumente zum Drucken oder Versenden; entsteht über den Druckdialog des Systems („Als PDF sichern“)
.txt · **.md** · **.html**	reiner Text, Markdown, Webseite
.srt · **.vtt**	Untertitel für Videos (YouTube, Premiere, DaVinci, VLC…)
.csv · **.json**	Tabellen (Excel, Numbers) und Daten für andere Programme
.v2t	das wieder öffnbare Projekt, auch zum Speichern in der Cloud

Die Optionen legen fest, ob **Zeitstempel** und **Sprechernamen** enthalten sind und ob die Sätze **zu Absätzen zusammengefasst** werden. Arabisch und andere rechts-nach-links-Sprachen werden korrekt gesetzt.

09 In der Cloud speichern

> Voice2Text speichert Dateien direkt in Ihrer Cloud – über die Ordner, die die Apps der Dienste auf dem Computer anlegen. Sie müssen keine Konten verbinden und keine Passwörter eingeben.

So funktioniert es

Wenn Sie die App von Google Drive, OneDrive, Dropbox, Box, **Proton Drive**, iCloud Drive, pCloud, MEGA, Nextcloud oder Sync.com installieren, entsteht auf dem Computer ein synchronisierter Ordner. Voice2Text findet ihn selbst: Wählen Sie im Fenster **Exportieren** unter **Speicherort** die Cloud und drücken Sie die Schaltfläche. Die Datei wird im Unterordner `Voice2Text` abgelegt, und die App des Dienstes lädt sie sofort hoch.

- Dateien werden nie überschrieben: Gibt es bereits eine gleichnamige Datei, heißt die neue „Name (2)“.
- Sie können jedes Format in der Cloud speichern, auch das `.v2t` -Projekt.
- Das PDF entsteht über den Druckdialog: Wählen Sie dort „Als PDF sichern“ und den Cloud-Ordner.

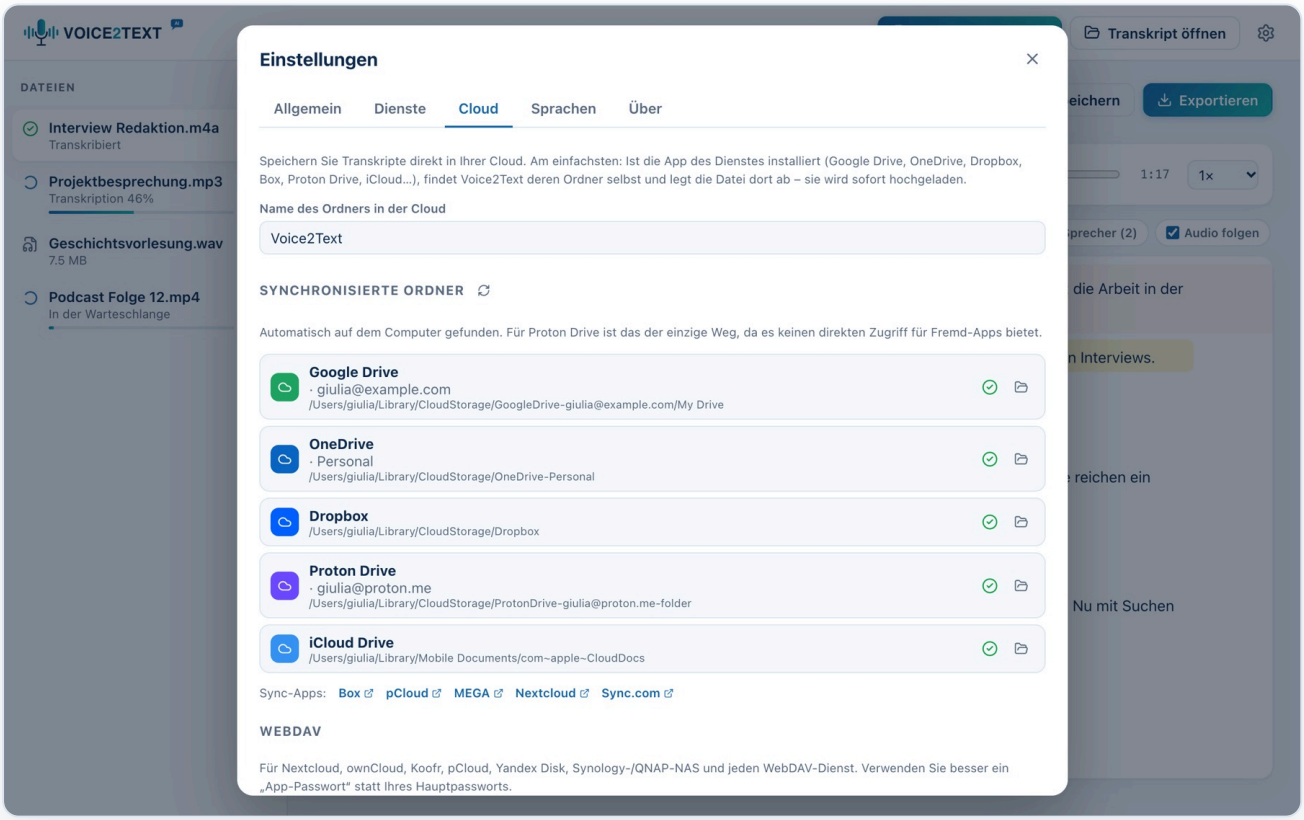


Abbildung 9. Einstellungen → Cloud: die auf dem Computer gefundenen Ordner und die WebDAV-Verbindung.

Einstellungen → Cloud

- **Ordnername**: der in der Cloud angelegte Unterordner (Standard `Voice2Text`).
- **Synchronisierte Ordner**: die gefundenen Clouds; das Ordnersymbol öffnet sie. Fehlt ein Dienst, führen die Links darunter zu den passenden Apps.
- **Weiteren Ordner hinzufügen**: für andere Clouds oder ein Netzlaufwerk (NAS).

- **WebDAV**: für Nextcloud, ownCloud, Koofr, pCloud, Yandex Disk und Synology-/QNAP-NAS ohne installierte App. Dienst wählen, Adresse, Benutzer und ein App-Passwort eingeben, dann **Verbindung testen** drücken.

> **Proton Drive** erlaubt anderen Apps keinen direkten Zugriff auf das Konto: Deshalb nutzt Voice2Text den synchronisierten Ordner, der dieselbe Ende-zu-Ende-Verschlüsselung bietet.

10 Einstellungen und Sprachen

> Das Zahnrad oben rechts (`⌘,` / `Strg+,`) öffnet die Einstellungen.

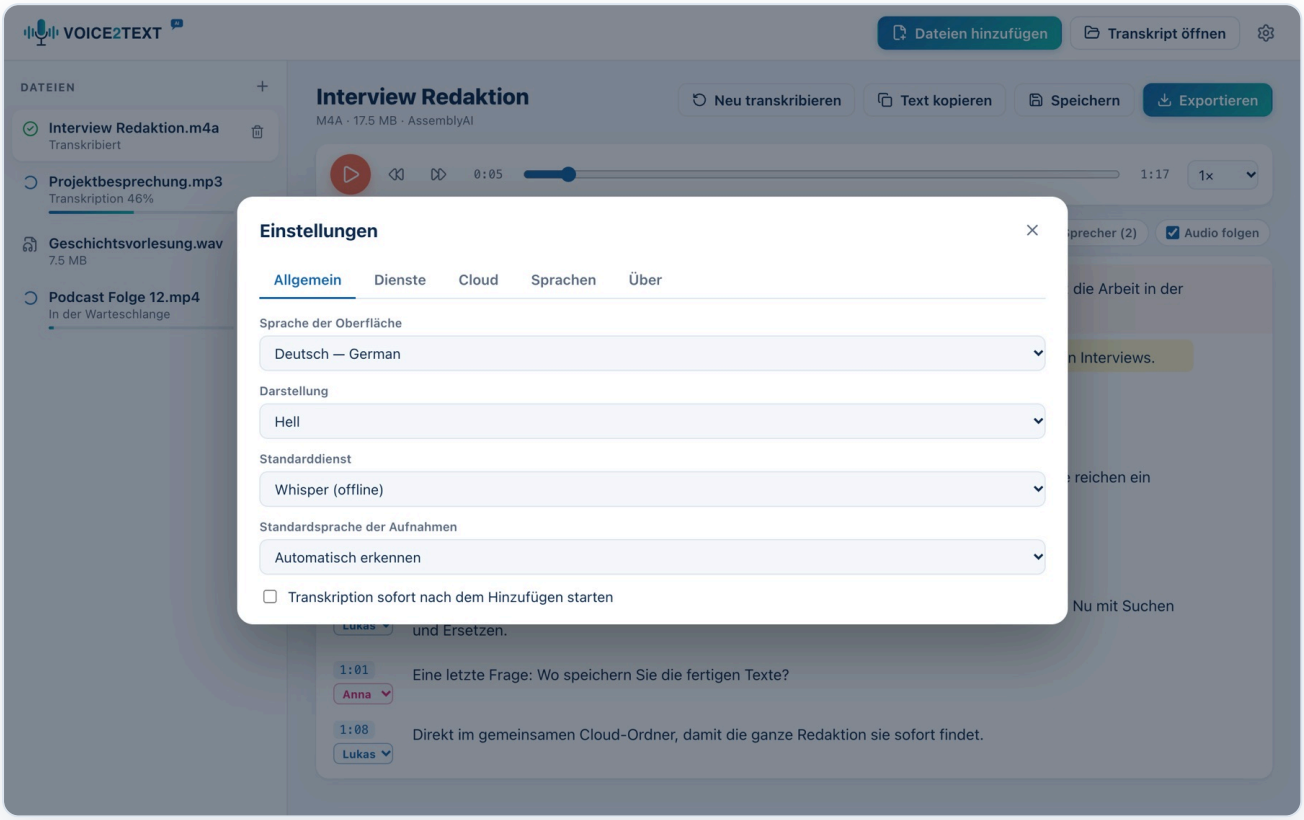


Abbildung 10. Einstellungen → Allgemein.

- **Allgemein**: Sprache der Oberfläche, helle, dunkle oder automatische Darstellung, Standarddienst und Standardsprache der Aufnahmen, automatischer Start der Transkription.
- **Dienste**: Modell für Offline-Whisper (winzig, Basis, klein, groß Turbo), Speicherplatz der Modelle, Schlüssel der Dienste, lokaler Whisper-Server, Erkennung von ffmpeg.
- **Cloud**: synchronisierte Ordner und WebDAV (vorheriges Kapitel).
- **Sprachen**: die Übersetzungsdateien der Oberfläche.
- **Über**: Version, Lizenz, Hinweis zu Online-Diensten.

Sprachen sind XML-Dateien

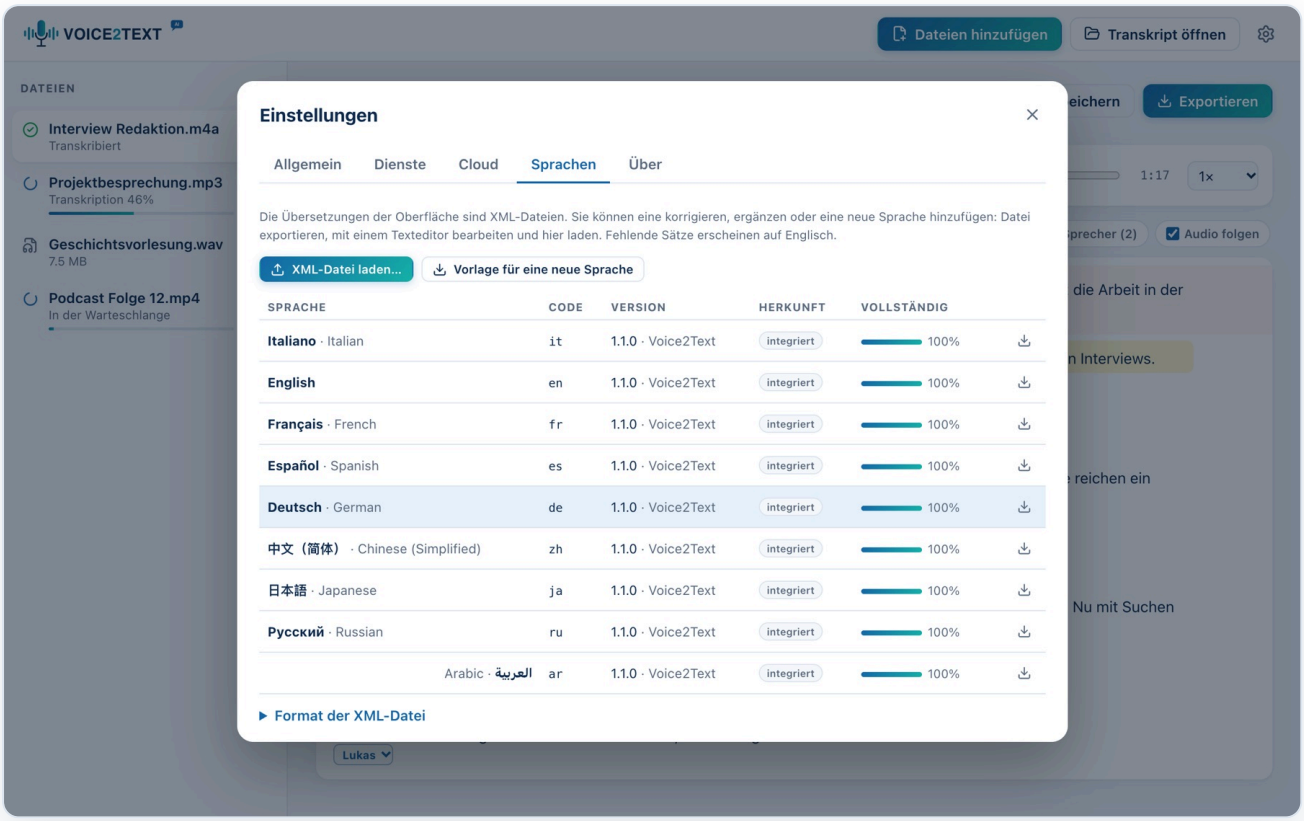


Abbildung 11. Einstellungen → Sprachen: verfügbare Sprachen, Vollständigkeit und Werkzeuge für XML-Dateien.

Jede Sprache der Oberfläche ist eine XML-Datei. So korrigieren Sie eine Übersetzung oder fügen eine neue hinzu:

1. Exportieren Sie die Sprachdatei (Pfeil in der Zeile) oder laden Sie die **Vorlage für eine neue Sprache** herunter.
2. Öffnen Sie sie in einem Texteditor und übersetzen Sie nur den Text zwischen den Tags; lassen Sie die `key`-Attribute und Platzhalter in geschweiften Klammern wie `{n}` unverändert.
3. Drücken Sie **XML-Datei laden...**: Die Sprache erscheint sofort in der Liste und im Sprachmenü.

```
<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>
```

Eine Datei mit dem Code einer integrierten Sprache ersetzt diese (für Korrekturen); Sie können jederzeit zum Original zurückkehren. Fehlende Sätze erscheinen auf Englisch; die Spalte **Vollständig** zeigt, wie viele übersetzt sind. Für rechts-nach-links-Sprachen verwenden Sie `dir="rtl"`.

11 Tastenkürzel

> Unter Windows und Linux verwenden Sie **Strg** statt ⌘.

Tasten	Aktion
⌘O	Audio- oder Videodateien hinzufügen
⇧⌘O	Gespeichertes Transkript öffnen (.v2t)
⌘S · ⇧⌘S	Speichern · Speichern unter
⌘E	Exportieren
⌘F	Suchen und Ersetzen
⌘,	Einstellungen
Leertaste	Abspielen / Pause (außerhalb des Textes)
Eingabe · **Umschalt+Eingabe**	Satz teilen · Zeilenumbruch im Satz
Rücktaste am Satzanfang	Mit dem vorherigen Satz verbinden
Esc	Text verlassen · offenes Fenster schließen

12 Häufige Fragen

> Antworten auf die häufigsten Fragen.

Das Format meiner Datei wird nicht gelesen.

Installieren Sie ffmpeg (Kapitel Installation) oder verwenden Sie AssemblyAI bzw. Deepgram, die jedes Format akzeptieren, weil sie die Originaldatei erhalten.

Offline-Whisper ist langsam.

Wählen Sie unter Einstellungen → Dienste das Modell **Basis** oder **Winzig**. „Groß Turbo“ ist am genauesten, braucht aber ohne Grafikkarte Zeit. Für lange Dateien ist Groq viel schneller.

Der Text ist in einer anderen Sprache oder übersetzt.

Geben Sie statt der automatischen Erkennung die Sprache der Aufnahme an und verwenden Sie **Neu transkribieren**.

„Ungültiger Dienstschlüssel“.

Prüfen Sie den Schlüssel unter Einstellungen → Dienste: Kopieren Sie ihn erneut von der Website des Dienstes, ohne Leerzeichen.

Meine Cloud erscheint nicht.

Stellen Sie sicher, dass die App des Dienstes installiert ist und synchronisiert, und drücken Sie unter Einstellungen → Cloud auf **Erneut suchen**. Alternativ fügen Sie den Ordner manuell hinzu.

Der Player meldet „Audio nicht verfügbar“.

Die Originaldatei wurde nach dem Speichern des Projekts verschoben oder umbenannt. Der Text bleibt erhalten; zum Anhören legen Sie die Datei wieder an ihren Platz.

Wo sind meine Schlüssel und Einstellungen?

Nur auf Ihrem Computer, in der Einstellungsdatei der App, die nur Ihr Benutzer lesen kann. Sie werden nie an Voice2Text gesendet.



VOICE2TEXT

Von der Stimme zum Text, in jeder Sprache.

Voice2Text ist ein Programm von overminder. Lizenz PolyForm Strict 1.0.0: nichtkommerzielle Nutzung erlaubt; Weitergabe, Änderungen und abgeleitete Werke nicht erlaubt.

© 2026 overminder · Version 1.1