




```
8""""
8      eeee eeee eeee eeee eeee
8eee 8 8 8 " 8 8 8 8      8 8 8 e e e eeee eeee eeee eeee      8 8 8 8 8
8eee 8 8 8 " 8 8 8 8      8e 8 8 8 8 8 8 " 8 8 8 8 8      eee eee 8e
88    8eee8 8eee 8e 8eee 8eee8e eeee 88 8 8 8eee8 8e 8eee 8eee8 8eee 8eee8e eeee 88 8 88 8 88
88    88 8 88 88 88 88 8      88 8 8 88 8 88 88 88 88 88 8      88 8 88 8 88
88    88 8 8ee88 88 88ee 88 8      88ee8ee8 88 8 88 8ee88 88 88ee 88 8      88 8 88 8 88eee
```

TRANSCRIPTION AUDIO EN TEXTE À L'AIDE DE WHISPER STANDALONE POUR PC

YOLAINE CHÉNARD, SIMON CÔTÉ-LAPOINTE ET JONATHAN LAPORTE
BIBLIOTHÉCAIRES

TABLE DES MATIERES

PROCÉDURE POUR TRANSCRIRE DE L'AUDIO EN FORMAT TEXTE AVEC WHISPER STANDALONE POUR PC	2
À propos	2
Préparation et formats des fichiers supportés	2
Version simplifiée et version complète	2
 Exercice 1 : installation du logiciel.....	2
Téléchargement / Installation	2
 Exercice 2 : expérimenter la transcription automatique - utiliser le fichier audio de la formation	5
Fonctionnement simplifié “One Click Transcribe”	5
 Exercice 3 : Expérimenter la transcription par l'invite de commande - utiliser le fichier audio de la formation	10
Version complète en invite de commandes.....	10
Créer une commande pour la transcription audio à l'aide du générateur de commande.....	10
Ouvrir l'Invite de commandes	11
Comprendre une ligne de commande	14
Informations complémentaires sur quelques commandes	14
Diarisation des locuteurs	14
Formats de fichiers de sortie et horodatage	14
Pour en apprendre davantage	15

PROCÉDURE POUR TRANSCRIRE DE L'AUDIO EN FORMAT TEXTE AVEC WHISPER STANDALONE POUR PC

À propos

Whisper est un script développé par OpenAI qui est disponible gratuitement en téléchargement. La version Whisper Standalone permet d'utiliser le script directement sur votre ordinateur PC (la formation n'abordera pas *Whisper Transcription* pour MAC¹). Étant donné qu'il est installé en local sur votre ordinateur, ceci constitue une solution fiable pour le traitement de données sensibles, car il y a moins d'enjeux de fuite de données contrairement aux solutions infonuagiques ou en ligne.

Préparation et formats des fichiers supportés

La plupart des formats de fichiers de type audio et vidéo sont supportés. Whisper utilise le décodeur ffmpeg. Principaux formats de fichiers pris en charge : mp3, wav, ogg, flac, m4a, mp4, etc.

Si vous travaillez à partir de fichiers dans un cloud, il est préférable de les télécharger sur votre ordinateur, car le temps d'accès au cloud peut empêcher ou ralentir le traitement.

Version simplifiée et version complète

Whisper Standalone fonctionne de deux façons possibles :

1. l'option simplifiée "One Click Transcribe" qui ne demande pas d'intervention de codage
2. l'option via l'invite de commandes, plus complexe, mais qui permet d'ajouter des options sur mesure telles que la traduction, l'identification automatique des intervenants, etc.



EXERCICE 1 : INSTALLATION DU LOGICIEL

Téléchargement / Installation

1. Visitez le site : <https://github.com/Purfview/whisper-standalone-win> . Vous y trouvez plusieurs informations sur le fonctionnement de Whisper Standalone, voir notamment la section *Discussions* pour poser des questions.
2. Cliquez sur "Standalone executables"

```
Administrator: 41% | 124/306 | 03:48 < 05:35 | 64 audio seconds/s
d:\whisper-fast\dist\whisper-faster.exe "D:\Temp\Infinity.Pool.test.mkv" --language=English --model=medium
Standalone Faster-Whisper v125 running on: CPU
[00:00.760 -> 00:02.760] Peking inspired yet?
[00:07.890 -> 00:08.890] Thank you.
[00:08.890 -> 00:10.890] I thought you said you were hungry.
[00:12.890 -> 00:17.890] There's a boat tour of the Teakla Island formations this afternoon.
[00:17.890 -> 00:22.890] I was thinking we could go on that and make reservations in town for dinner.
[00:22.890 -> 00:25.890] We could try the Chinese place.
[00:25.890 -> 00:28.890] I don't think I'd survive another dinner in town.
[00:29.890 -> 00:31.890] Even the idea that...
[00:31.890 -> 00:33.890] Does anyone think it's a real town?
[00:36.890 -> 00:38.890] Why would they have a Chinese place?
[00:43.280 -> 00:46.280] Is it okay if I go? I'll meet you on the beach.
[00:47.280 -> 00:49.280] No, sure.
[01:23.270 -> 01:24.270] Someone's making a statement.
[01:24.270 -> 01:26.270] One of the locals, I guess.
[01:29.270 -> 01:31.270] What do you think he's trying to say?
[01:31.270 -> 01:35.270] He's saying that he wants to put a long knife right through her.
[01:35.270 -> 01:40.270] And after you die, he'll hang your body at the airport to scare off the other tourists.
[01:42.270 -> 01:44.270] Seems a bit extreme.
[01:46.270 -> 01:48.270] The Lakotons are a melodramatic people.
[01:54.740 -> 01:56.740] I loved your book.
[01:58.740 -> 01:59.740] Sorry?
[02:00.740 -> 02:03.740] You're James Foster. I loved your book.
```

[Standalone executables](#) of OpenAI's Whisper & Faster-Whisper for those who don't want to bother with Python.

¹ Pour les MAC, il est possible de télécharger une version nommée *Whisper Translate* via App Store : <https://apps.apple.com/us/app/whisper-transcription/id1668083311>. L'option gratuite est très limitée.

Ce qui vous mène à cette page : <https://github.com/Purfview/whisper-standalone-win/releases>

3. Identifiez la section titrée **Faster-Whisper-XXL r245.4**. Ouvrir la section "Assets".

➔ Faster-Whisper-XXL r245.4

Standalone Faster-Whisper implementation using optimized CTranslate2 models.

Includes all Standalone Faster-Whisper features + some additional ones. Read [here](#).

Last included commit: [#245](#)

Faster-Whisper-XXL includes all needed libs.

Some new stuff in r245.4:

Fixed: 'unload_model' error when using batched inference

Some new stuff in r245.3:

Fixed: Model didn't free VRAM when --model_preload False

Fixed: "cuda:1" didn't work with MDX

Fixed: Some issues in "One Click Transcribe.bat"

Fixed: Matplotlib backend error [Colab]

Fixed: cuDNN libs not found error [Linux]

Changed: --ff_mdx_kim2 to --ff_vocal_extract mdx_kim2

Changed: --mdx_device to --voc_device

Added: distil-large-v3.5 model

Updated: pyinstaller to 6.12.0




Link to [the changelog](#).

▶ Assets 7

👍 73 🗨️ 6 ❤️ 30 🚀 4 🏠 13 101 people reacted

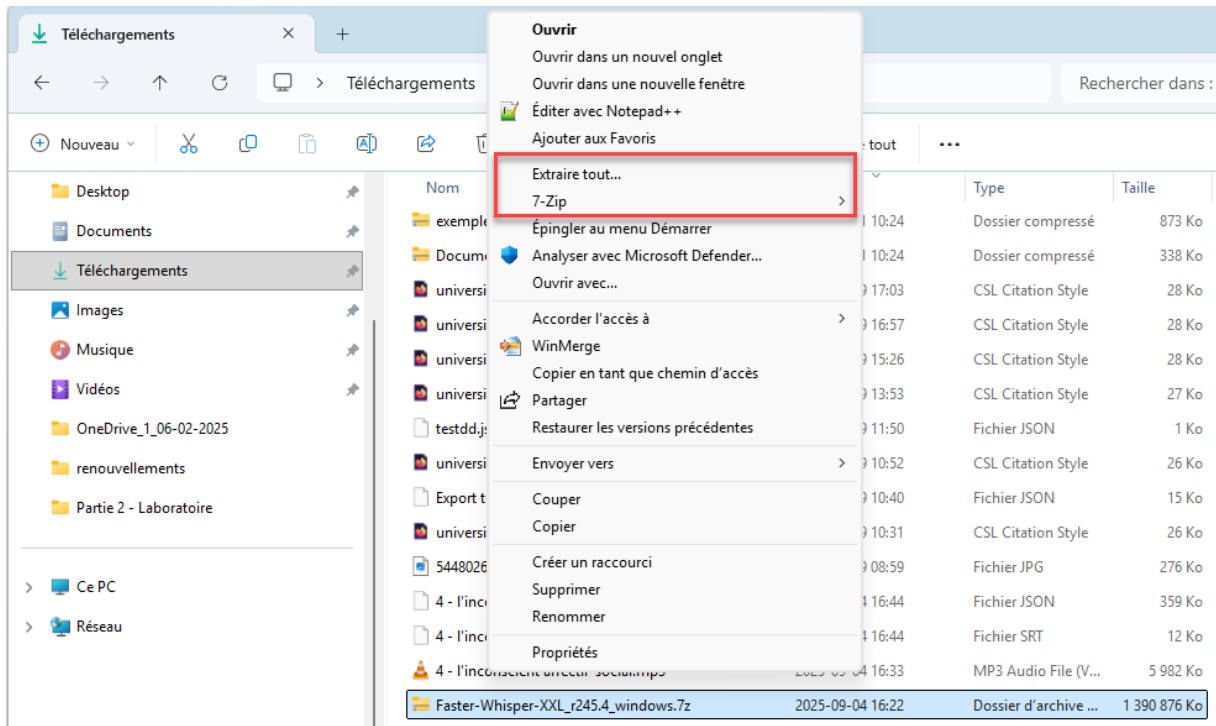
4. Cliquez sur le lien de la version pour Windows disponible, par exemple **Faster-Whisper-XXL r245.4**, pour télécharger le logiciel

▼ Assets 7





 Faster-Whisper-XXL_r192.3.1_linux.7z	856 MB	Apr 8, 2024
 Faster-Whisper-XXL_r192.3.4_windows.7z	1.06 GB	Apr 22, 2024
 Faster-Whisper-XXL_r245.1_windows.7z	1.35 GB	Jan 13
 Faster-Whisper-XXL_r245.4_linux.7z	1.54 GB	Apr 13
 Faster-Whisper-XXL_r245.4_windows.7z	1.33 GB	Apr 13
 Source code (zip)		Mar 25, 2024
 Source code (tar.gz)		Mar 25, 2024

5. Télécharger puis extraire le contenu du fichier .7z téléchargé. À l'aide du menu contextuel (clic droit sur le fichier), sélectionner "Extraire tout" (ou utiliser 7-Zip → si vous n'avez pas d'application pour extraire les fichiers, veuillez [télécharger 7-Zip](#) ou [NanaZip](#)). Cette opération peut prendre plusieurs minutes.

Attention : ne pas télécharger ou extraire le dossier dans le dossier "Windows" car cela pourrait causer des problèmes.



6. Ce qui vous donnera un dossier nommé "Faster-Whisper-XXL_r245.4_windows". Double-cliquez pour accéder au sous-dossier "Faster-Whisper-XXL" contenant les éléments suivants :

 faster-whisper-xxl.exe	2025-04-13 09:03	Application	40 362 Ko
 One Click Transcribe.bat	2025-04-09 18:35	Fichier de comma...	2 Ko
 _xxl_data	2025-04-13 09:03	Dossier de fichiers	
/ Il y a longtemps			
 ffmpeg.exe	2023-12-31 02:44	Application	80 857 Ko

→Vous devez disposer de suffisamment d'espace mémoire pour faire fonctionner ce logiciel.

7. Choisir la version **Whisper Standalone** la plus appropriée selon vos besoins :

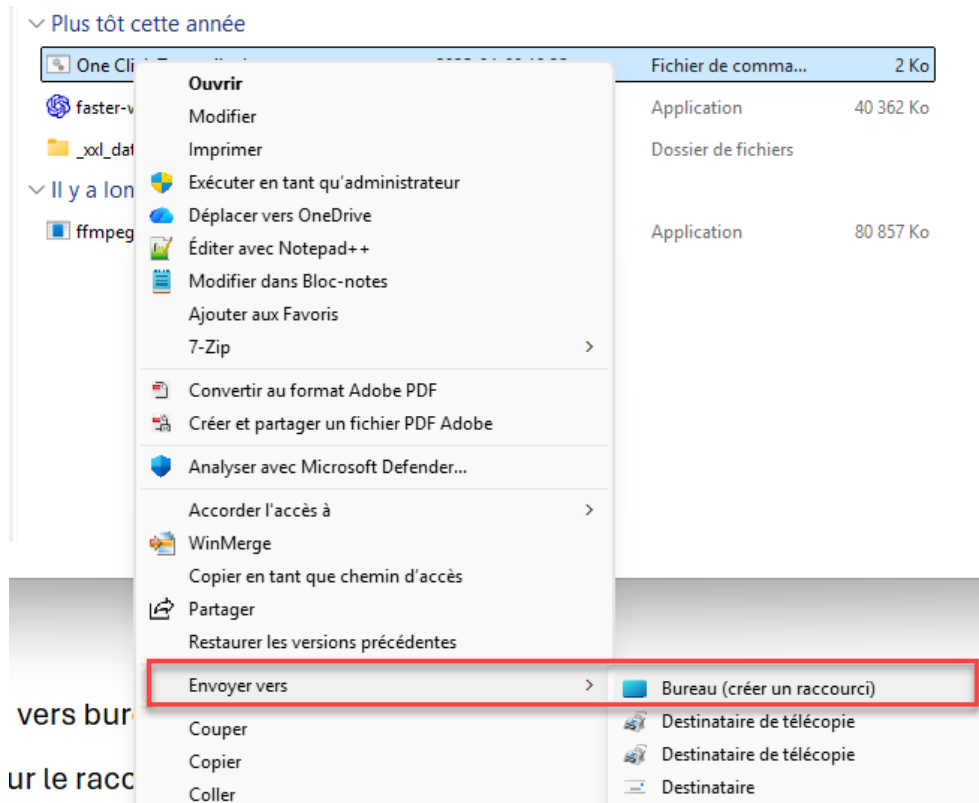
- [Fonctionnement simplifié "One Click Transcribe"](#) qui ne demande pas d'intervention de codage
- [Fonctionnement via l'invite de commande](#), plus complexe, mais qui permet d'ajouter des options sur mesure telles que la traduction, l'identification automatique des intervenants, etc.



EXERCICE 2 : EXPÉRIMENTER LA TRANSCRIPTION AUTOMATIQUE - utiliser le fichier audio de la formation

Fonctionnement simplifié "One Click Transcribe"

1. Faites un clic droit sur "One Click Transcribe.bat" et Cliquez sur "Envoyer vers > raccourci vers bureau"

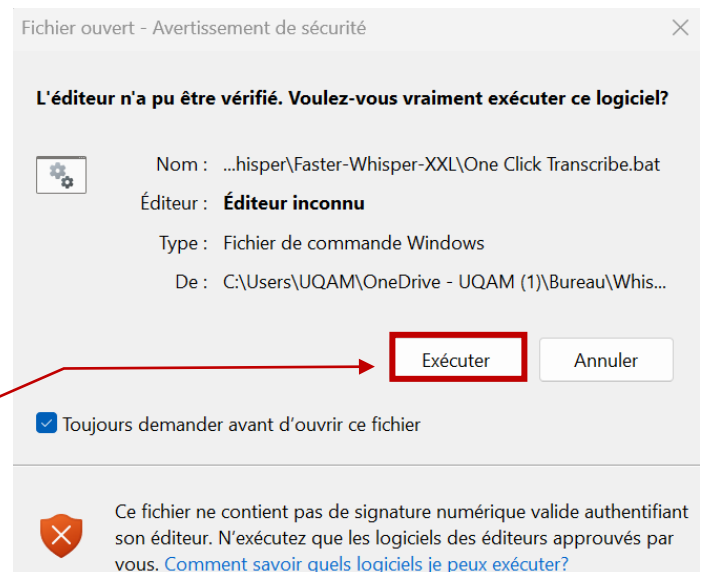


2. Copier votre fichier audio localement. Vous pouvez vous créer un dossier "Entrevues" et y déposer vos fichiers audios. Si votre fichier audio est très long, il est recommandé de le scinder en plusieurs parties au préalable afin que le traitement soit plus rapide.

3. Accédez à votre Bureau

Glisser-déposer le fichier audio sur le raccourci créé sur le bureau puis enfoncer la touche Entrée/Enter de votre clavier.

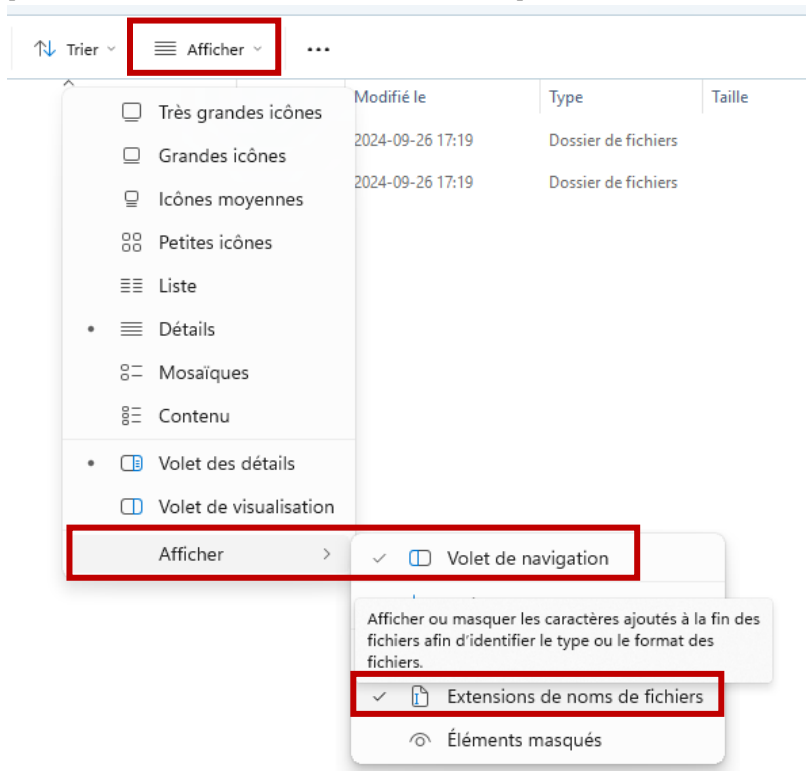
Si ce message apparaît, autoriser l'opération -> Exécuter



Appuyez sur une touche pour continuer. Le logiciel se fermera automatiquement.

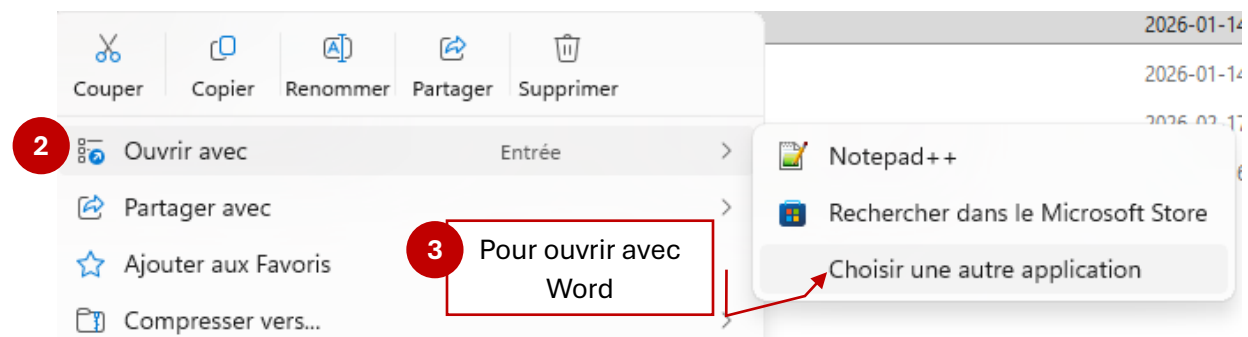
5. Une fois le traitement complété, 2 fichiers seront produits. Avec cette méthode, les fichiers se trouveront au même emplacement que le fichier d'origine.
 - **.json** : fichier comprenant tous les détails de l'extraction (probabilités pour chaque mot, horodatage pour chaque mot, etc.)
 - **.str** : fichier de type sous-titre qui contient les plages temporelles segmentées par phrases.

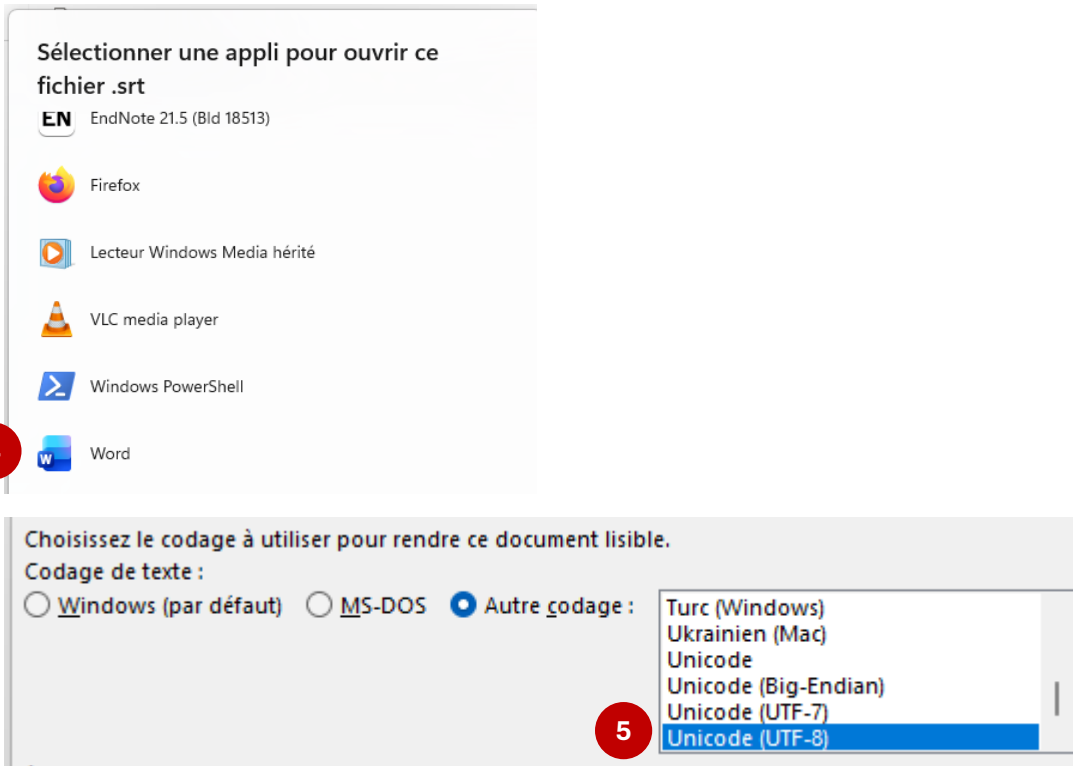
Note : si vous ne voyez pas les extensions des fichiers veuillez changer l'affichage [Afficher/Extensions de noms de fichiers] :



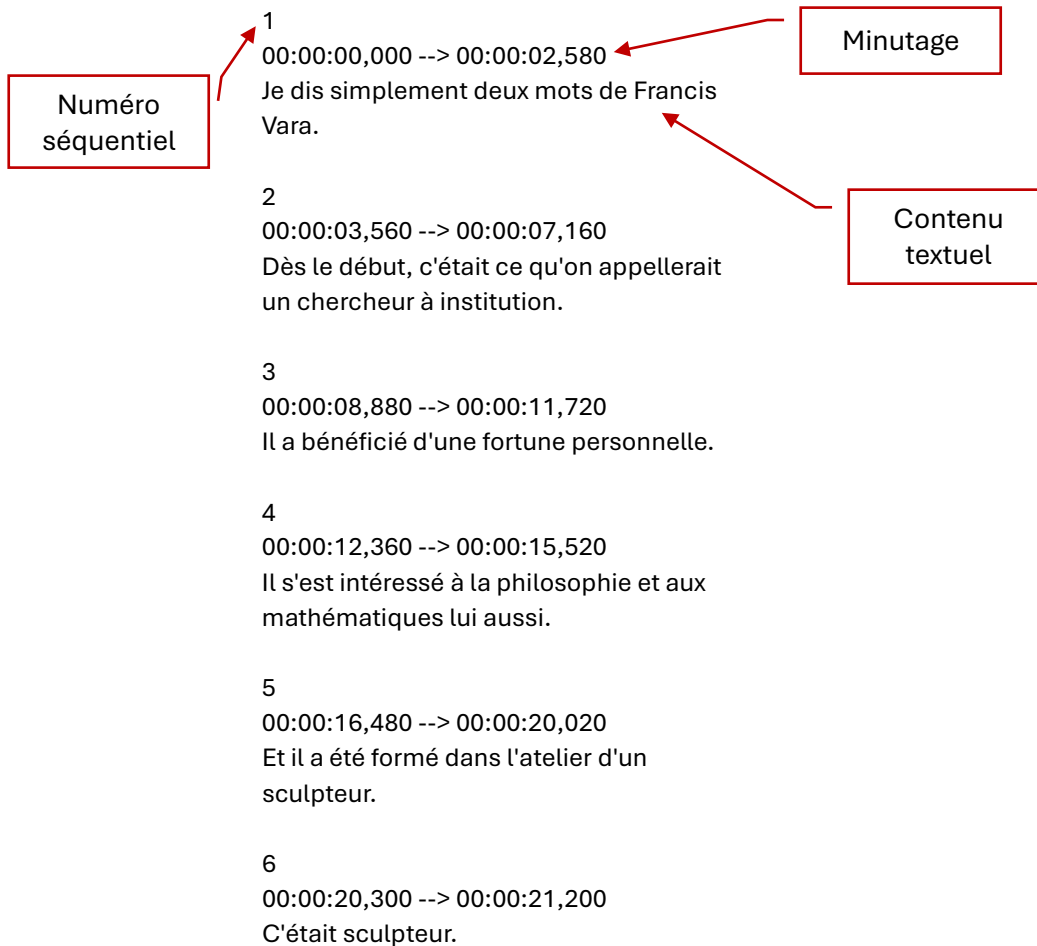
6. On recommande l'utilisation du fichier .str pour les utilisations courantes.
7. On peut ouvrir le fichier .str avec le [logiciel gratuit Notepad++](#) (ou BlocNote) ou encore, avec Word :

1 Clic droit sur le fichier .str

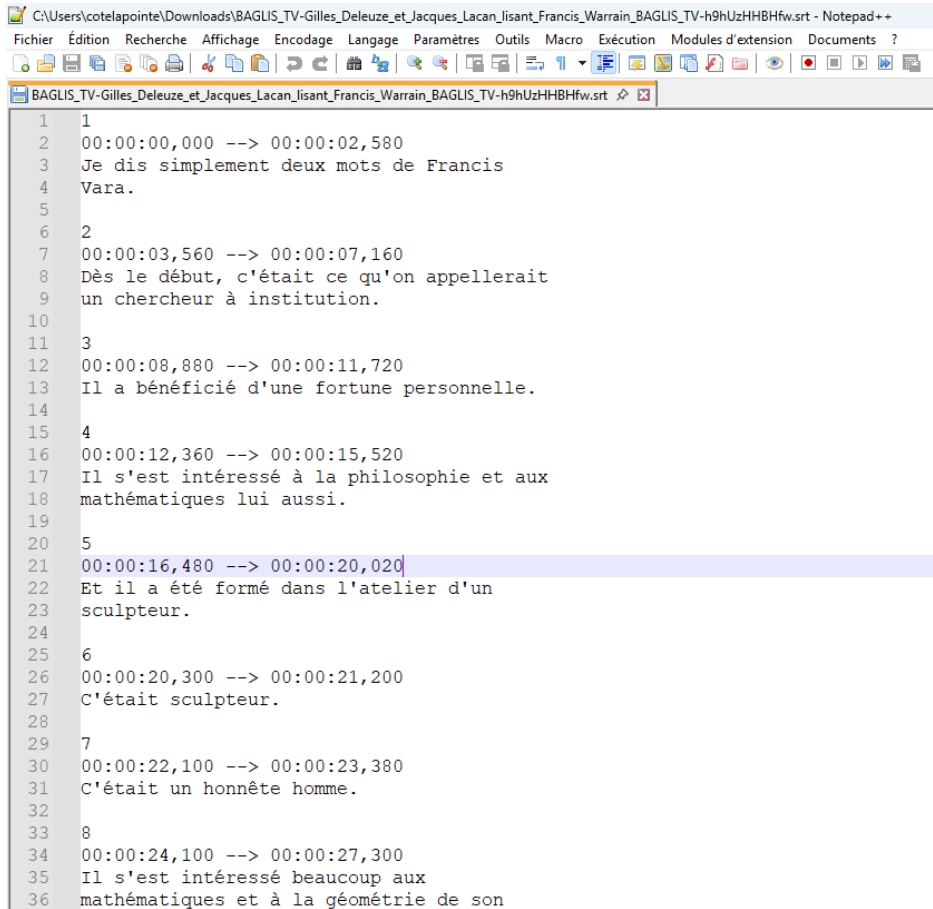




8. Voici à quoi ressemble le contenu d'un fichier :



Chaque segment est identifié par un numéro, sur la ligne suivante se trouve le minutage et ensuite le contenu textuel.



```
C:\Users\cotelapointe\Downloads\BAGLIS_TV-Gilles_Deleuze_et_Jacques_Lacan_lisant_Francis_Warrain_BAGLIS_TV-h9hUzHHBHfw.srt - Notepad++
Fichier  Édition  Recherche  Affichage  Encodage  Langage  Paramètres  Outils  Macro  Exécution  Modules d'extension  Documents  ?

BAGLIS_TV-Gilles_Deleuze_et_Jacques_Lacan_lisant_Francis_Warrain_BAGLIS_TV-h9hUzHHBHfw.srt

1 1
2 00:00:00,000 --> 00:00:02,580
3 Je dis simplement deux mots de Francis
4 Vara.
5
6 2
7 00:00:03,560 --> 00:00:07,160
8 Dès le début, c'était ce qu'on appellerait
9 un chercheur à institution.
10
11 3
12 00:00:08,880 --> 00:00:11,720
13 Il a bénéficié d'une fortune personnelle.
14
15 4
16 00:00:12,360 --> 00:00:15,520
17 Il s'est intéressé à la philosophie et aux
18 mathématiques lui aussi.
19
20 5
21 00:00:16,480 --> 00:00:20,020
22 Et il a été formé dans l'atelier d'un
23 sculpteur.
24
25 6
26 00:00:20,300 --> 00:00:21,200
27 C'était sculpteur.
28
29 7
30 00:00:22,100 --> 00:00:23,380
31 C'était un honnête homme.
32
33 8
34 00:00:24,100 --> 00:00:27,300
35 Il s'est intéressé beaucoup aux
36 mathématiques et à la géométrie de son
```



EXERCICE 3 : EXPÉRIMENTER LA TRANSCRIPTION PAR L'INVITE DE COMMANDE - utiliser le fichier audio de la formation

VERSION COMPLÈTE EN INVITE DE COMMANDES

Créer une commande pour la transcription audio à l'aide du générateur de commande

Pour simplifier la création de votre commande de transcription, utilisez notre générateur automatique:
<https://web-bibliotheques.uqam.ca/site/whisper>

Plusieurs paramètres s'offrent à vous.

Générateur de commande – Faster-Whisper-XXL

Choisir les options → copier la commande. Dans les chemins Windows : \ simple (ex. C:\Tools\video.mp4).

The screenshot shows the 'Générateur de commande – Faster-Whisper-XXL' interface. It features several sections for configuring transcription parameters:

- Presets:** Transcription simple, Réunion (2-6 locuteurs), Qualité max + diarize.
- Chemin vers l'exécutable:** .\faster-whisper-xxl.exe (marked with a red circle '1').
- Fichier audio/vidéo:** Ex. C:\Tools\video.mp4.
- Dossier de sortie:** output.
- Langue:** Français.
- Tâche:** transcribe.
- Modèle:** medium (recommandé).
- Formats de sortie:** txt, text (texte brut sans horodatage), srt, vtt, lrc, json, tsv, all.
- Compute type:** auto.
- Options utiles:** --sentence, --vad_filter.
- Diarisation des locuteurs:** off.
- Nombre de locuteurs (si connu):** Active diarize pour utiliser ce champ.
- Actions:** Copier la commande, Réinitialiser.
- Commande prête à copier/coller:**

```

".\faster-whisper-xxl.exe" --model medium --language fr --task transcribe --output_format txt srt --output_dir "output"
--compute_type auto --sentence --vad_filter True
                    
```

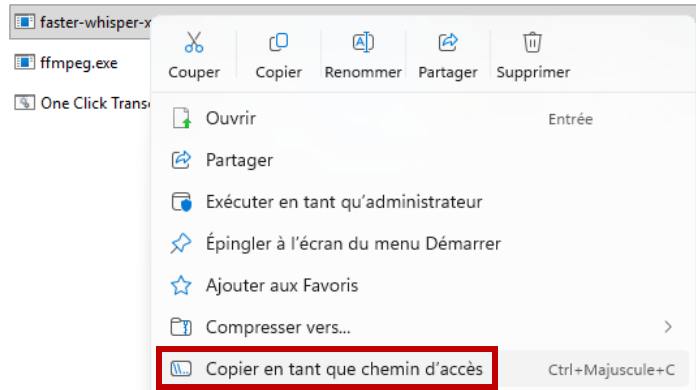
Annotations:

- A red box labeled 'Paramètres rapides' points to the preset buttons.
- A red box labeled 'La diarisation permet d'identifier les différents locuteurs' points to the 'Diarisation des locuteurs' section.
- A red box labeled 'Lorsque les différents paramètres seront choisis, copier la commande et la coller dans l'invite de commande' points to the 'Copier la commande' button and the command output area.
- A red circle '1' highlights the executable path field.
- A red circle '2' highlights the 'Copier la commande' button.

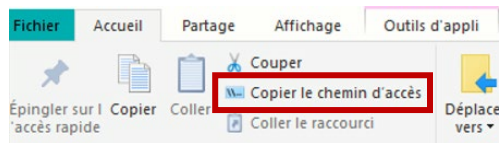
- 1 Vous devez copier le chemin vers l'exécutable et le coller dans **Chemin vers l'exécutable**. Ensuite, copier le chemin de l'emplacement de votre fichier audio et le coller dans le champ **Fichier audio/vidéo**.

Comment faire?

Avec Windows 11 : À l'aide du clic droit de la souris sur l'exécutable se trouvant dans l'emplacement où vous avez installé Whisper : **faster-whisper-xdl.exe**, sélectionner « Copier en tant que chemin d'accès ». Ensuite, coller le lien dans du formulaire. Même processus pour le chemin où se trouve le fichier audio. Il est recommandé d'enregistrer localement votre fichier audio sur votre ordinateur.

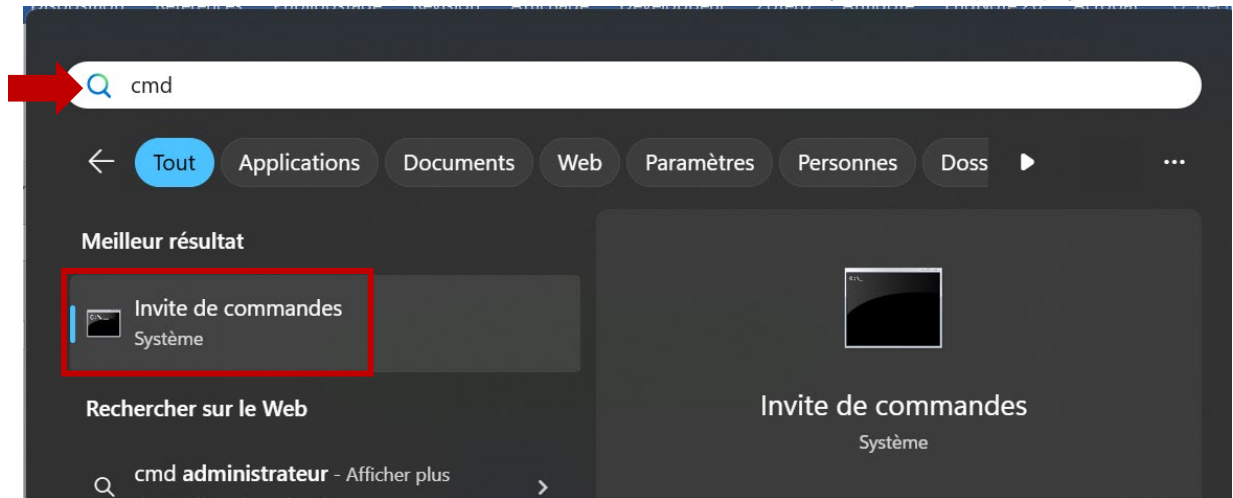
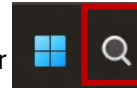


Version antérieure de Windows : Vous trouverez l'option **Copier en tant que chemin d'accès** dans le ruban de la barre d'outils :



- 2 **Ouvrir l'Invite de commandes**

1. Cliquer sur la loupe dans la barre des tâches de l'ordinateur
2. Écrire CMD dans rechercher pour ouvrir l'Invite de commandes (Command prompt)



Double cliquer pour ouvrir l'Invite de commandes (Command prompt).

3. Une fois le générateur de commande complété, copier la ligne de commande que vous collerez dans l'Invite de commande

```

Microsoft Windows [version 10.0.26200.7623]
(c) Microsoft Corporation. Tous droits réservés.

C:\Users\chenard_Y>"C:\Users\chenard_Y\OneDrive - UQAM\Bureau\Whisper\Faster-Whisper-XXL\faster-whisper-xxl.exe" "C:\Users\chenard_Y\OneDrive - UQAM\Bureau\Whisper\Faster-Whisper-XXL\BAGLIS_TV-Gilles_Deleuze_et_Jacques_Lacan_lisant_Francis_Warrain_BAGLIS_TV-h9hUzHHBfw.mp4" --model medium --language fr --task transcribe --output_format=text --output_dir "output" --compute_type auto --vad_filter True --diarize reverb_v1 --num_speakers 2

```

4. Faire « Entrée/Enter » pour démarrer le traitement.

Si aucun message d'erreur n'apparaît, il ne reste plus qu'à attendre. Le logiciel effectuera les étapes suivantes :

- Téléchargement du modèle pour la transcription sur l'ordinateur. La ligne model.bin indique l'état du téléchargement. Le temps varie selon votre connexion internet.
- Diarisation de l'audio : détection des intervenants (cela peut prendre un certain temps), selon la solution choisie, on verra apparaître la transcription au fur et à mesure qu'elle se fera.
- Le temps de traitement varie selon la puissance de votre ordinateur et la durée de votre fichier audio. Même si ça ne semble pas travailler pendant de longues minutes, laissez rouler le script.

```

Microsoft Windows [version 10.0.26200.7628]
(c) Microsoft Corporation. Tous droits réservés.

C:\Users\UQAM>"C:\Users\UQAM\OneDrive - UQAM (1)\Bureau\Whisper\Faster-Whisper-XXL\faster-whisper-xxl.exe" "C:\Users\UQAM\OneDrive - UQAM (1)\Bureau\Whisper\Faster-Whisper-XXL\BAGLIS_TV-Gilles_Deleuze_et_Jacques_Lacan_lisant_Francis_Warrain_BAGLIS_TV-h9hUzHHBfw.mp4" --model medium --language fr --task transcribe --output_format txt srt --output_dir "output" --compute_type auto --sentence --vad_filter True --without_timestamps True --diarize reverb_v1 --num_speakers 2

Model not found at: C:\Users\UQAM\OneDrive - UQAM (1)\Bureau\Whisper\Faster-Whisper-XXL\models\faster-whisper-medium
Attempting to download:

config.json: 2.26kB [00:00, ?B/s]
vocabulary.txt: 460kB [00:00, 462kB/s]
tokenizer.json: 2.20MB [00:01, 1.33MB/s]
tokenizer.json: 2.20MB [00:01, 1.57MB/s] | 0.00/1.53G [00:00<?, ?B/s]
model.bin: 1% | 10.5M/1.53G [00:12<29:11, 867kB/s]

```

- Si vous n'avez pas changé l'emplacement par défaut, on retrouvera le fichier généré dans le dossier output. Vérifier son emplacement lorsque le traitement sera terminé :

```

[06:41.200 --> 06:43.380] de même dans la bibliothèque de Gilles Deleuze,
[06:43.520 --> 06:45.560] qui en fait mention quelque part.
[06:46.100 --> 06:48.180] Voilà quelques éléments
[06:49.980 --> 06:51.020] de postérité,
[06:51.300 --> 06:54.600] comme quoi, même si tout cela est très limité,
[06:56.280 --> 06:57.800] la voix de Vronsky ne s'est pas
[06:57.800 --> 06:58.860] perdue dans le désert.

Transcription speed: 0.54 audio seconds/s
Subtitles are written to 'C:\Users\chenard_Y\output' directory.

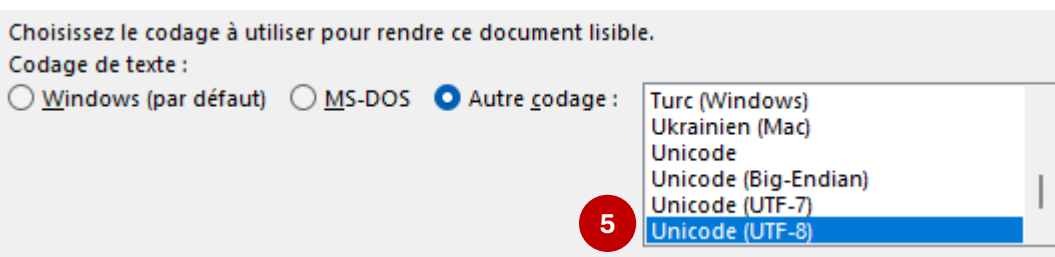
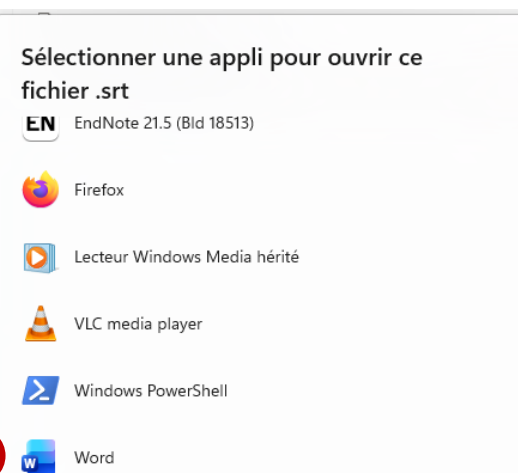
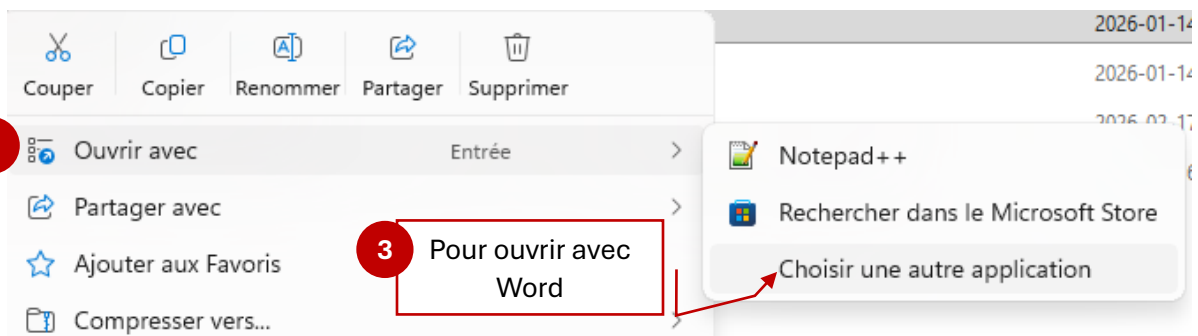
Operation finished in: 0:20:37.866

C:\Users\chenard_Y>

```

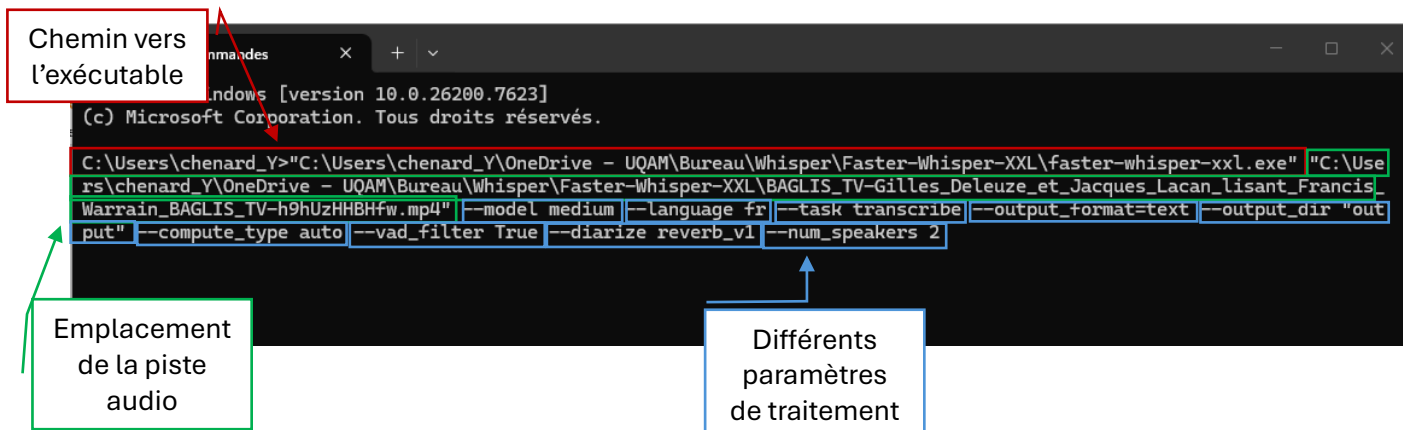
5. On recommande l'utilisation du fichier .str pour les utilisations courantes.
6. On peut ouvrir le fichier .str avec le [logiciel gratuit Notepad++](#) (ou BlocNote) ou encore, avec Word :

1 Clic droit sur le fichier .str



Comprendre une ligne de commande

Chaque partie donne des indications sur le traitement du fichier. Il est possible d'entrer une ligne manuellement dans l'invite de commande (pour utilisatrices et utilisateurs avancés).



Informations complémentaires sur quelques commandes

Diarisation des locuteurs

La diarisation est l'identification automatique des intervenants

--diarize

pyannote_v3.0 - script le plus rapide pour le processeur

pyannote_v3.1 - Identique à la v3.0, mais plus rapide avec CUDA (GPU)

reverb_v1 - Réputé meilleur que pyannote v3

reverb_v2 - La plus lente, mais réputée la meilleure

Autres options

--num_speakers - Nombre d'intervenants, si connu.

--min_speakers - Nombre minimal d'intervenants. Sans effet si num_speakers est fourni.

--max_speakers - Nombre maximal d'intervenants. Sans effet si num_speakers est fourni.

--speaker - Remplace la chaîne « SPEAKER » par votre propre mot.

--diarize_device - « CUDA » ou « CPU ». Automatique, inutile de modifier ce paramètre.

--diarize_threads - Nombre de threads. Automatique, inutile de modifier ce paramètre.

--diarize_dump - Enregistre le résultat de la diarisation dans un fichier.

--diarize_only : Diariser sans transcription.

-embeddings : Inclure les embeddings des locuteurs.

--diarize_ff : Activer la diarisation après les filtres --ff... Automatique.

À noter : Mentions légales : Les modèles de réverbération sont réservés à un usage personnel et non commercial.

Pour en savoir plus : [Speaker Diarization · Purview/whisper-standalone-win · Discussion #322 · GitHub](#)

Formats de fichiers de sortie et horodatage

--output_formats txt srt vtt

ou

-f txt srt

Vous pouvez choisir un ou plusieurs formats de sortie séparés par un espace :

Variables possibles : 'json', 'lrc', 'txt', 'text', 'vtt', 'srt', 'tsv', 'all'

À noter : Si vous utilisez uniquement le format de sortie **txt**, vous obtiendrez un long fichier texte sans aucun horodatage. L'horodatage est une fonctionnalité inhérente aux formats de sous-titres comme **srt** et **vtt**.

Pour en apprendre davantage

- [GitHub - openai/whisper: Robust Speech Recognition via Large-Scale Weak Supervision](#)
- <https://github.com/Purfview/whisper-standalone-win/discussions/231>
- ["One Click Transcribe" tool · Purfview/whisper-standalone-win · Discussion #337 · GitHub](#)