



HAL
open science

Whisper : Retranscription automatique rapide et sans difficulté

Alioscha Massein, Sylvain Besson, Orline Poulat

► **To cite this version:**

Alioscha Massein, Sylvain Besson, Orline Poulat. Whisper : Retranscription automatique rapide et sans difficulté. Café des données, Maison des Sciences de l'Homme Lyon Saint-Etienne; PUD Panels, Sep 2024, Lyon (Maison des Sciences de l'Homme), France. hal-04707311

HAL Id: hal-04707311

<https://hal.science/hal-04707311v1>

Submitted on 24 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



Café des Données

Whisper : Retranscription automatique rapide et sans difficulté



Le café des données c'est quoi ?

- Un mardi par mois, de 9h à 10h.
- Un aspect de l'utilisation et du traitement de données par café
- Orienté pour les SHS.
- Pensé pour la discussion avec les intervenant-es et les personnes présentes.

PROGRAMME

PUD-panels : Accompagner les chercheur-ses avec leurs données

22 octobre 2024

PUD-Panels & Hélène Kieffer

Choisir un entrepôt pour ses données : quels critères ?

26 novembre 2024

Marie-Emiia Herbet

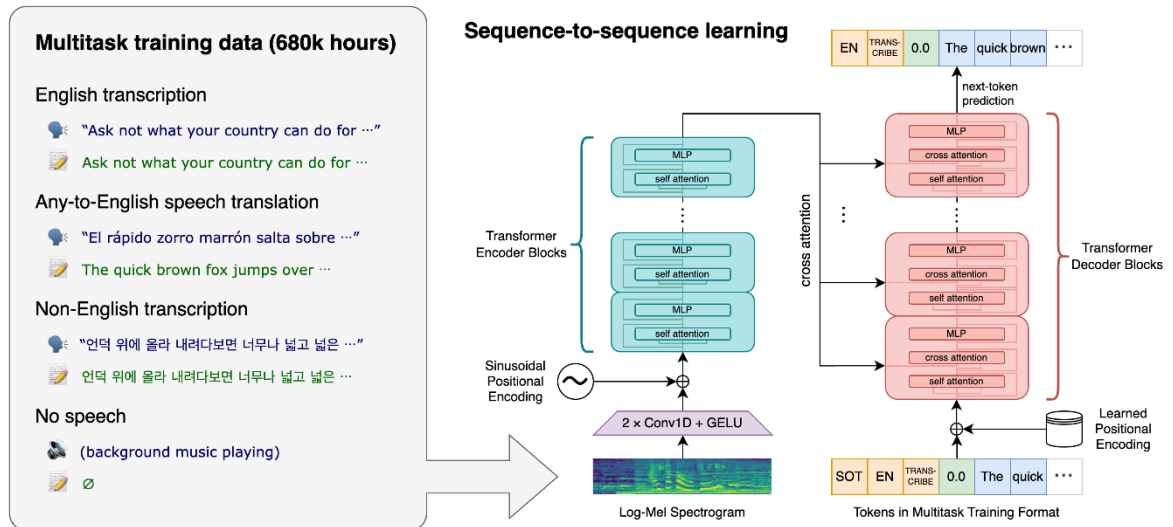
GeoVistory : Une introduction au web sémantique

21 janvier 2025

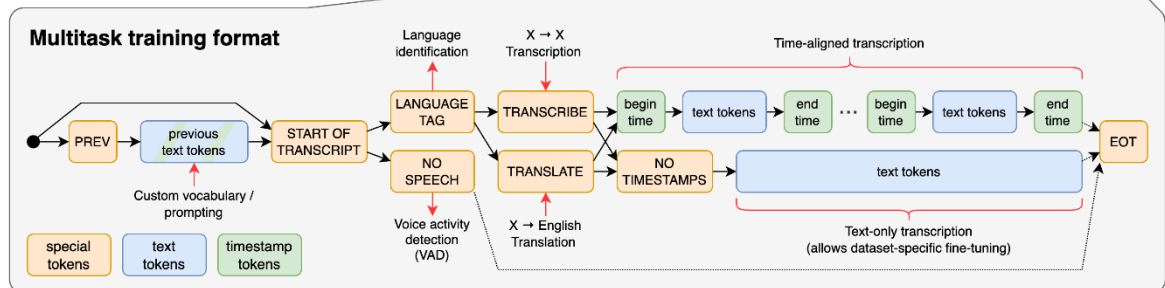
Morgane Pica, Vincent Alamericy

Whisper, c'est quoi ?

- Un modèle d'IA
- Qui fonctionne à partir d'audio encodé en « Transformers »

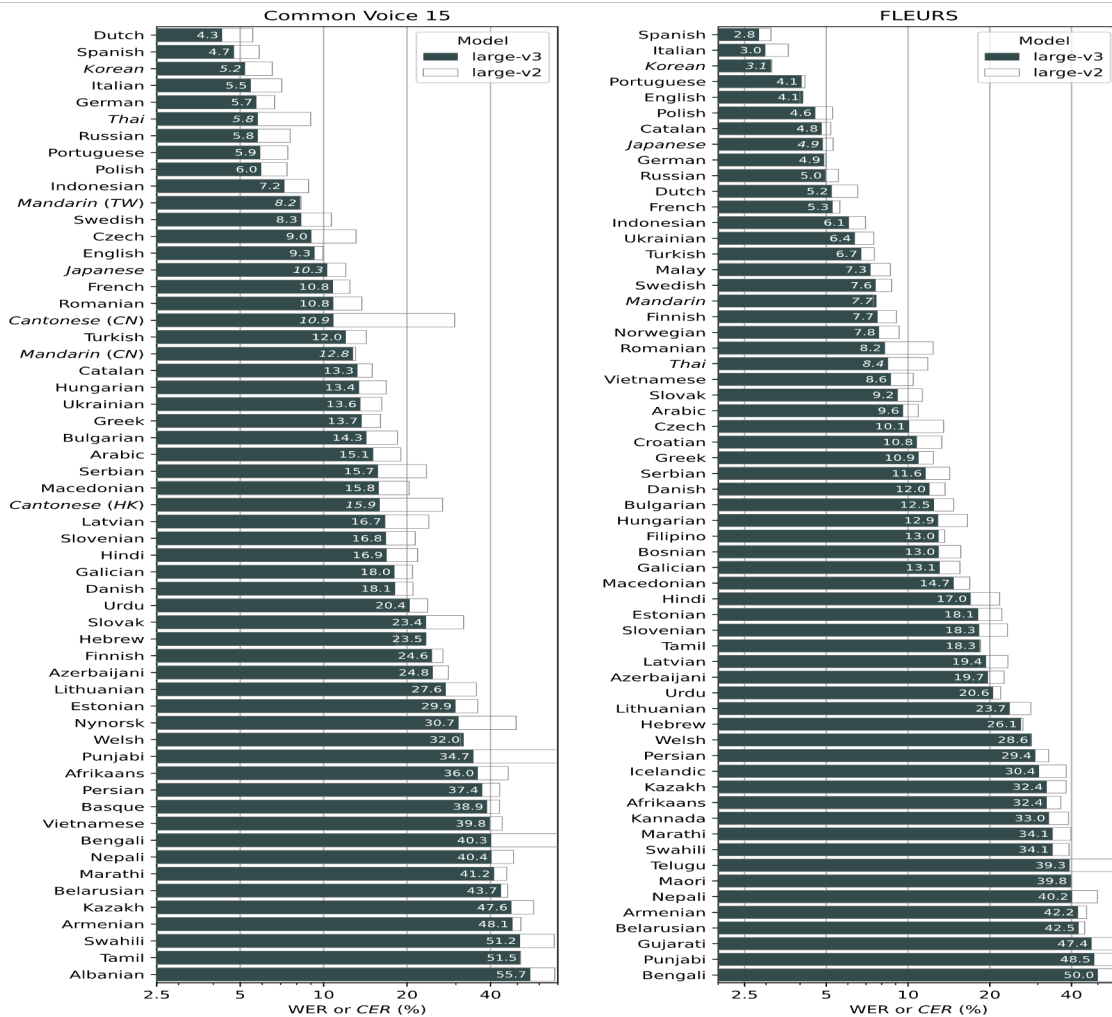


Enfin la structure est très proche des modèles de langues comme Llama, GPT,...



<https://github.com/openai/whisper>

Et ça peut retranscrire quelles langues ?



Presque une soixantaine de langues et de dialectes différents. Les taux d'erreur peuvent fortement varier en fonction de la langue de transcription ou de traduction.



Et comment ça fonctionne ?

1. En ligne de commande

The following command will transcribe speech in audio files, using the `medium` model:

```
whisper audio.flac audio.mp3 audio.wav --model medium
```



The default setting (which selects the `small` model) works well for transcribing English. To transcribe an audio file containing non-English speech, you can specify the language using the `--language` option:

```
whisper japanese.wav --language Japanese
```



Adding `--task translate` will translate the speech into English:

```
whisper japanese.wav --language Japanese --task translate
```



2. Dans un script python

```
import whisper

model = whisper.load_model("base")
result = model.transcribe("audio.mp3")
print(result["text"])
```





Avantages :

- S'adapte à vos besoins les plus précis
- Peut permettre la diarisation
- Non dépendant d'une connexion à un serveur
- On peut utiliser d'autres modèles que Whisper (insanely-fast-whisper,...)

Inconvénients :

- Non trivial, demande des connaissances en informatique (Shell, python)
- Dépend fortement des capacités de l'ordinateur
- Occupe de la place de stockage (taille du modèle).



3. En local avec des applications



Buzz

<https://github.com/chidiwilliams/buzz>



Vibe

<https://github.com/theWH1teagle/vibe>

SocioTranscribe

<https://framagit.org/justinmponcet/whisper>

Avantages :

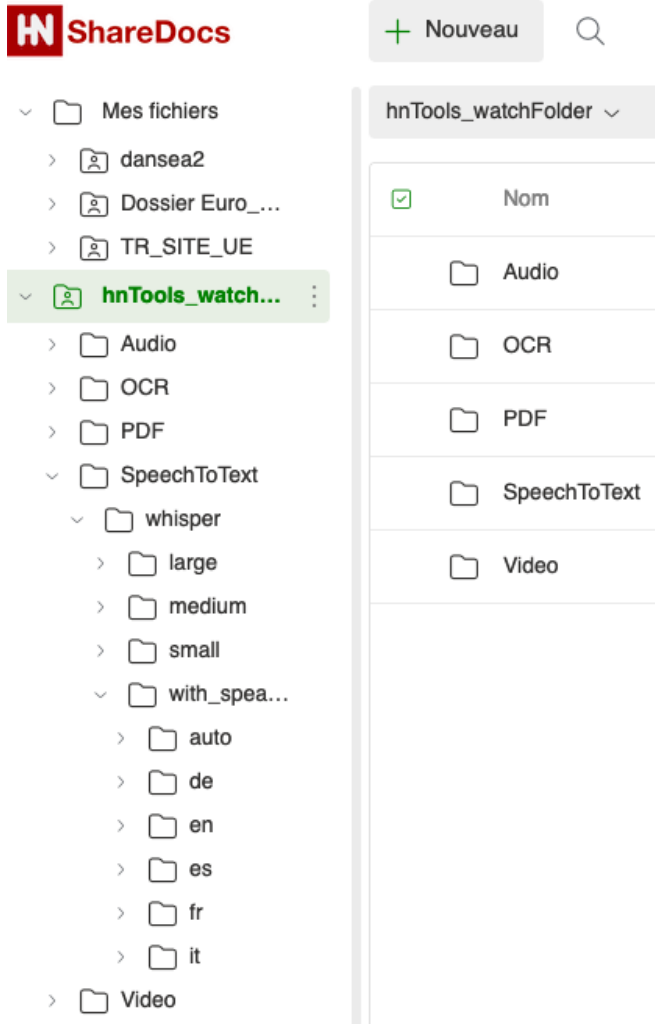
- Simple d'utilisation (Buzz & Vibe)
- Avec ou sans diarisation (Vibe)
- Produit de la recherche CNRS (SocioTranscribe)
- En local (pour les données)

Inconvénients :

- Dépend toujours des performances de l'ordinateur
- SocioTranscribe demande quelques compétences en informatique pour l'installation
- Limité dans leur usage aux fonctions qui y ont été prévu



4. En ligne avec ShareDocs de HumaNum depuis juin 2023



Dans le répertoire hnTools_watchFolder :

- Attention ce répertoire ne s'affiche qu'à la **2nde connexion avec au moins 30min depuis la première.**
- Répertoire de type **First In – First Out.**
- Pour Huma-Num : quantification / test pour voir l'évolution d'usage de l'outil.
- Extraire l'audio de la vidéo pour faciliter le traitement
- Fichiers **supprimés automatiquement** au bout de 21j et **non sauvegardés.**



4. (Bis) Créer un compte ShareDocs personnel en 2 mots

- Créer un **compte HumanId** à l'adresse : <https://humanid.huma-num.fr/>



Réinitialiser mon mot de passe

HN se connecter avec HumanID

première visite sur HumanID ?

HN Créer un compte HumanID

connectez-vous avec un compte externe ⓘ

 se connecter avec eduGAIN

 se connecter avec ORCID

HAL se connecter avec HAL

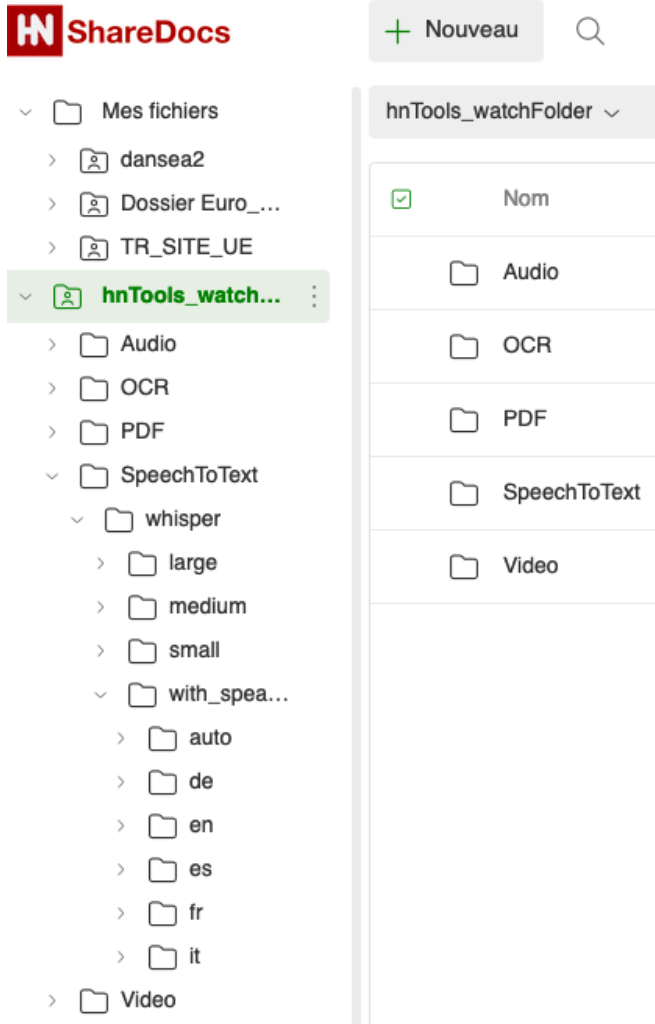
 se connecter avec LinkedIn

 se connecter avec Twitter

 se connecter avec Google



4. En ligne avec ShareDocs de HumaNum depuis juin 2023



Avantages :

- Ne nécessite pas de laisser l'ordinateur allumé
- Avec diarisation
- Permet de faire tourner le modèle « large »

Inconvénients :

- Nécessite un compte HumanId avec un accès à ShareDocs
- Temps d'attente variable : principe de file d'attente.



4. (Bis) Créer un compte ShareDocs personnel en 2 mots
- Créer un **compte HumanId** à l'adresse : <https://humanid.huma-num.fr/>
- Demander l'accès au service ShareDocs : renseigner le **titre** de projet, une courte **description** de celui-ci, et une **volumétrie**.
- Il vous faudra ensuite patienter quelques jours afin qu'Huma-Num vous donne vos accès personnels.



Plateforme de stockage
et de partage de fichiers
(Web et clients WebDAV)

[Demander l'accès](#)

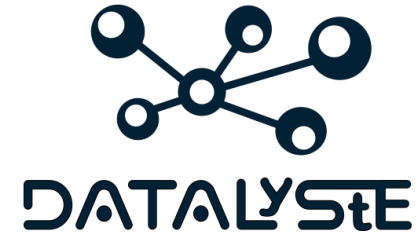
Attention à distinguer le compte personnel du compte projet :

- Compte personnel : 90 Go max
- Compte projet : 300 Go max
- Chaque membre d'un compte projet devra avoir au préalable un compte personnel créé (qui sert notamment d'accès au compte projet).
- Pour toute question concernant Huma-Num : online.poulat@msh-lse.fr



5. En ligne via DataLyste

Datalyste est le nom de [l'Atelier de la données](#) sur le site de Lyon-Saint-Etienne. Il propose des services, des outils, des recommandations et un accompagnement sur les questions relatives aux données pour le personnel de la recherche.



Dans ce cadre, DataLyste s'est doté d'un service de transcription avec Whisper sur des serveurs de l'Ecole Centrale de Lyon. Une interface est en cours de programmation, mais vous pouvez d'ores et déjà utiliser ce service via la liste de diffusion prévue à cet effet : retranscription@groupes.renater.fr

Ce service est pensé dans le cadre de la retranscription d'un nombre de fichier important et sera à terme disponible via vos identifiants institutionnels.



5. Retour d'expérience

Conditions:

- réalisation de deux entretiens en groupe de deux heures chacun en français
- Un groupe avec 3 locuteurs et un groupe avec 6 locuteurs
- Enregistrement dans une pièce fermée (sous-sol de la MSH) avec un micro HF et un téléphone portable



	Datalyste	ShareDocs	NoScribe
Vitesse d'exécution	7 à 10 minutes pour 2h de transcription	Inconnu; mis sur une liste d'attente (en moyenne une nuit)	Variable selon l'ordinateur; Avec mon PC sans carte graphique, 1h pour 1h de retranscription (beaucoup plus rapide avec carte graphique)
Ergonomie	Actuellement, la demande doit être faite auprès d'Orline ou Aljoscha	Service à l'intérieur de sous-dossier dans ShareDocs; pas très intuitif	Interface très simple et intuitive
Qualité de la retranscription	99% d'exactitude en moyenne	99% d'exactitude en moyenne	/
Qualité de la diarisation	95% d'exactitude en moyenne	37% d'exactitude en moyenne	/

Remarques:

- Les résultats sont plus ou moins les mêmes entre l'enregistrement avec l'enregistrement HF et l'enregistrement avec un portable
- Plus le nombre de locuteurs augmentent, plus le taux d'erreurs augmente que ce soit pour la transcription ou pour la diarisation
- Il est également possible d'utiliser Whisper sur des langues moins usitées (ex: Thai)
- Les conditions de prise de son étaient optimales (très peu de bruits parasites)



Pour approfondir

- <https://www.css.cnrs.fr/fr/whisper-pour-retranscrire-des-entretiens2/>
- <https://mate-shs.cnrs.fr/actions/tutomate/tuto24-retranscription-elise-tancoigne/>
- <https://mate-shs.cnrs.fr/wp-content/uploads/2020/06/tuto24-slides-tancoigne-retranscription.pdf>

- <https://agepouvoir.hypotheses.org/494>
- <https://shs.hal.science/halshs-01339474/>

- <https://documentation.huma-num.fr/sharedocs-stockage/>