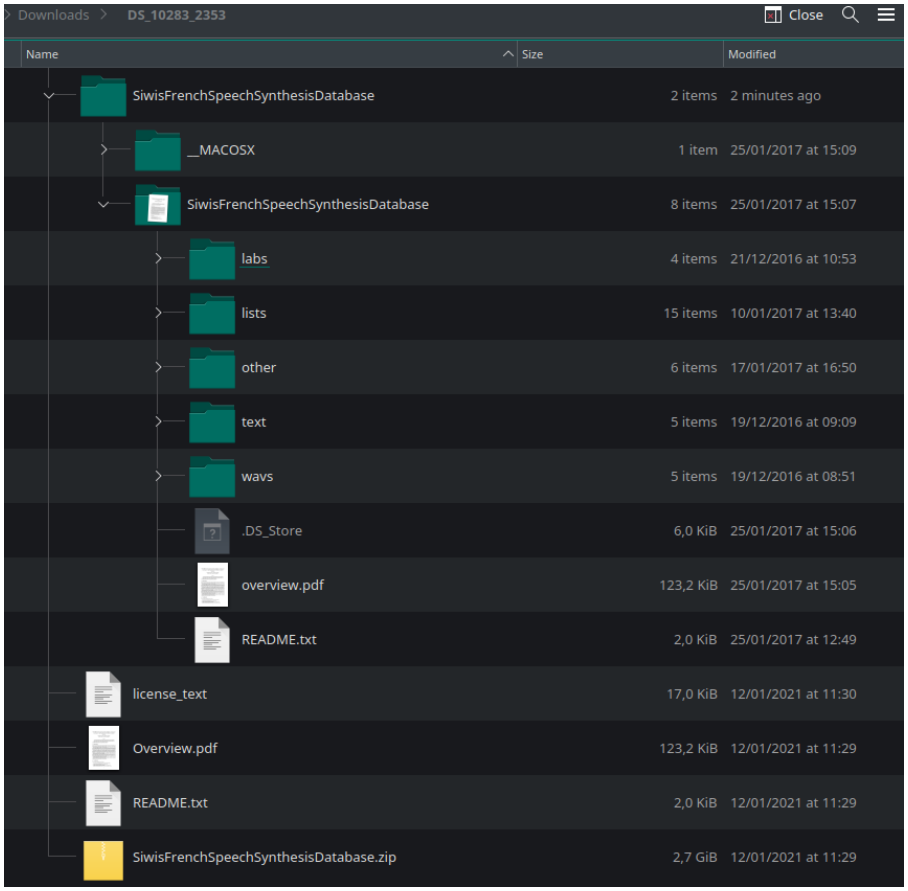
Use Case 1a: aligning French speech data

In this notebook we will do a simple use case by aligning French speech data. We will use the SIWIS French Speech Synthesis Database (Yamagishi et al., 2016) that contains utterances from political debates, book (chapters) and more spoken by one speaker.

1) Download resources

First, download the data from the link: https://datashare.ed.ac.uk/handle/10283/2353. This zip contains another zip (SiwisFrenchSpeechSynthesisDatabase.zip) with the data. You should see the following structure:



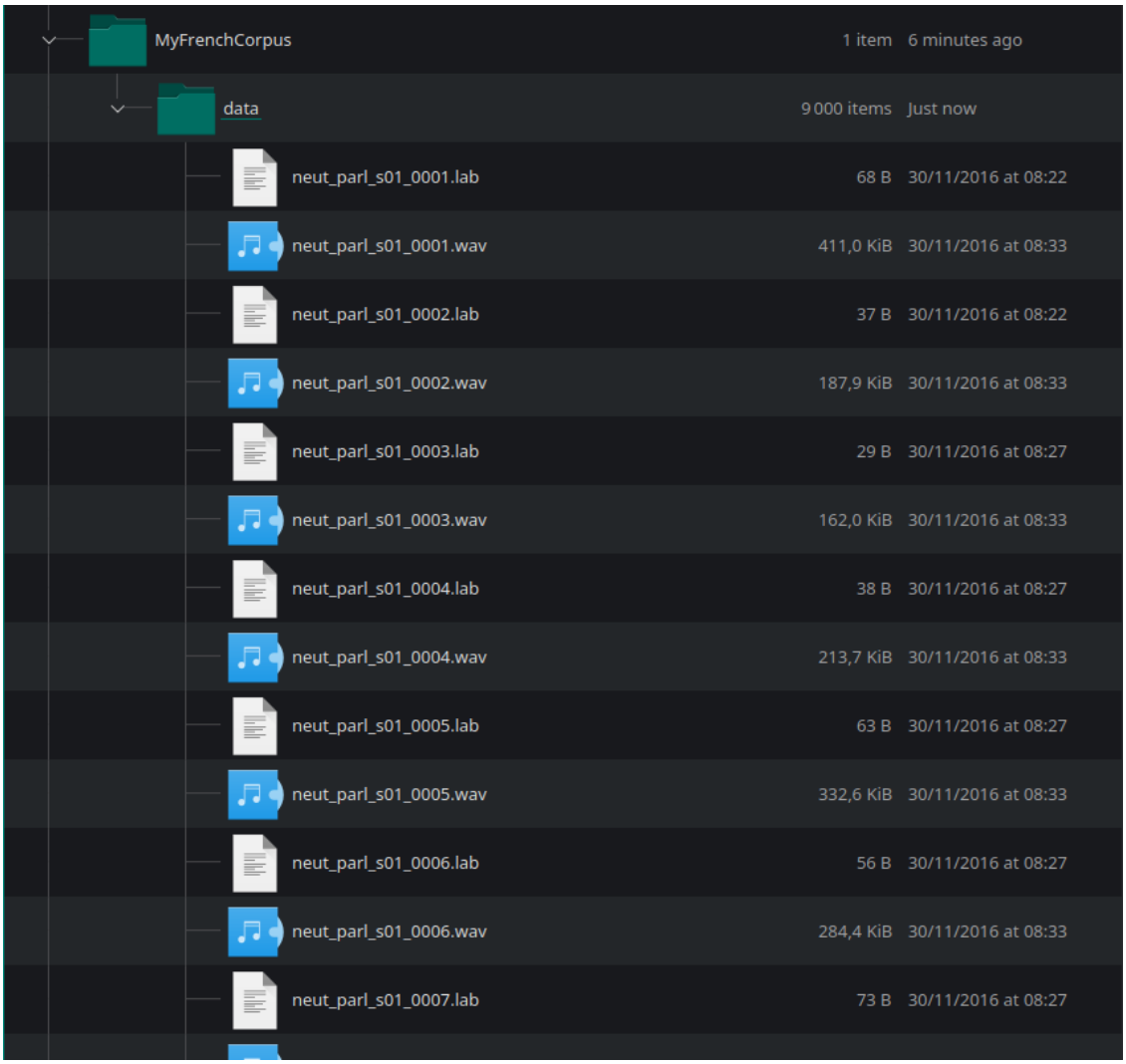
We need the wav files from dataset. Download also the UseCase1_transcriptions.zip.

Note: These are not the .lab files from the SIWIS dataset!

2) Create a corpus

The next step is the basis for the alignments. Just create a folder that will contain the data of your corpus. In this case, I created a folder in the documents directory called "MyFrenchCorpus" and another folder that will contain the data.

You can create your corpus folder anywhere you want on your computer, but be sure to note path. Now we can copy & paste the wav files we want to align. In this case, I used the subset "part1" of the SIWIS corpus, but you can use any or all of the subsets. Copy & paste also the respective lab files from the UseCase1_transcriptions.zip. The folder structure should look like this:



In general, the MFA needs both wav files and annotation files for the automatic alignment. The annotation files contain the utterances of the wav files either in the TextGrid formant or as .lab files. The .lab files are simple text files with .lab suffix instead of .txt. Regardless of the format, the annotation files contain only the orthographic transcriptions of the utterances. Furthermore, wav files and corresponding annotation files need ot have the same file names.

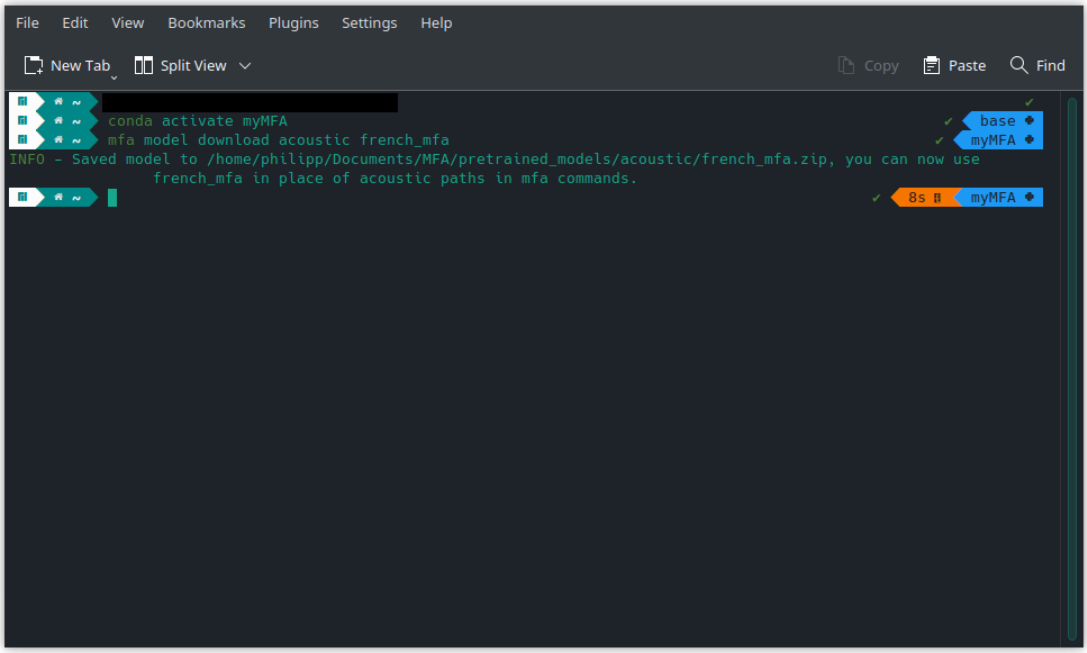
3) open Anaconda and activate your MFA environment

In case Anaconda is not opened, open your terminal/console/command line and activate the environment in which you installed the MFA by (replace "myMFA" by the name you use for the environment):

[ ]:
```
conda activate myMFA
```

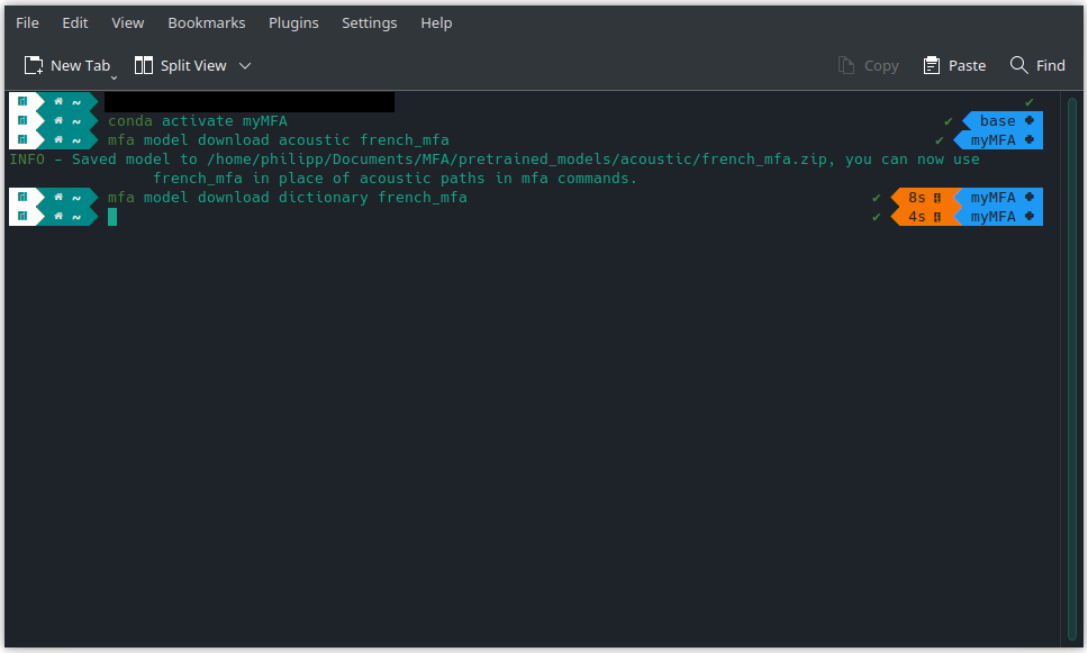
Before we can align the french data, we have to install an acoustic model and dictionary. Pre-trained acoustic models and dictionary can be found here: https://mfa-models.readthedocs.io/en/latest/ . In this case, we install the "french_mfa" acoustic model:

[ ]:
```
mfa model download acoustic french_mfa
```

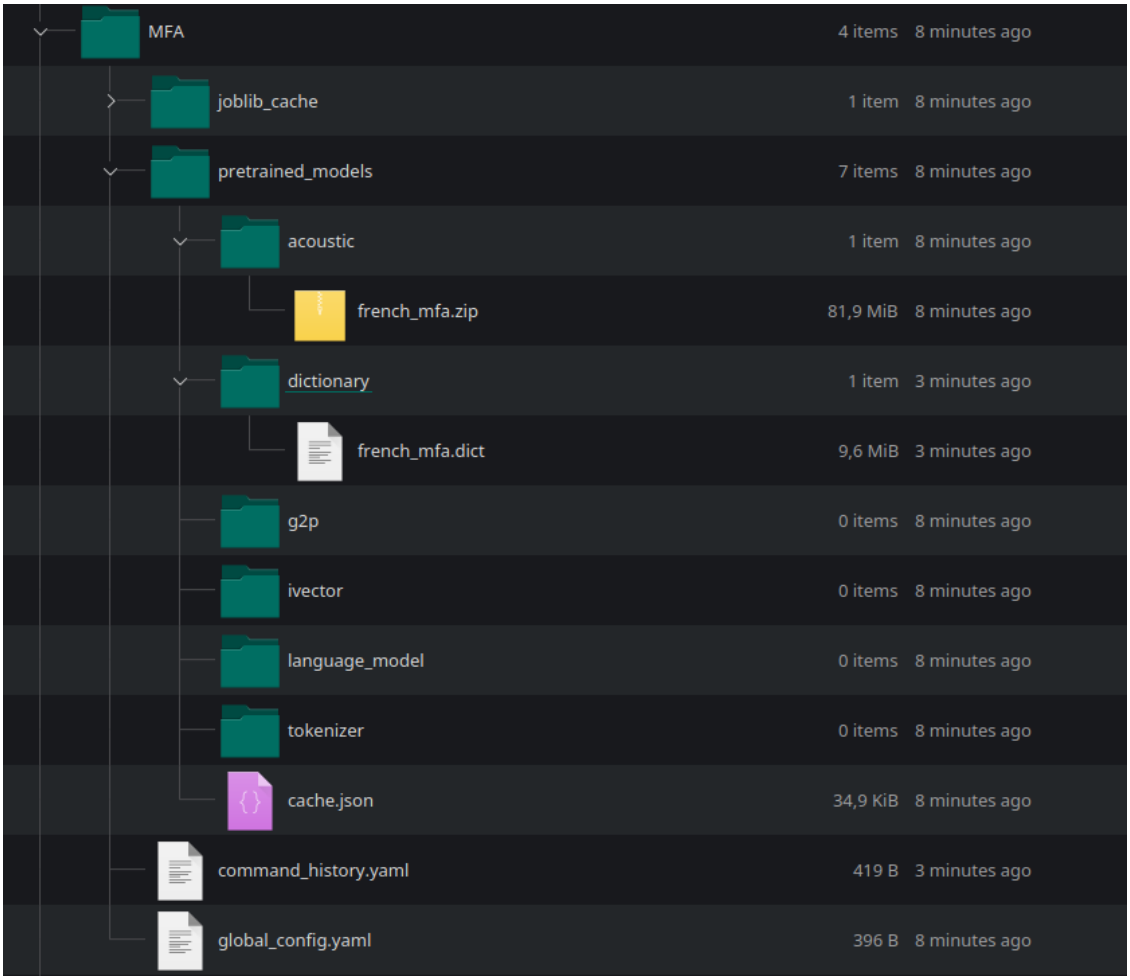


In the next step, we download the corresponding french dictionary:

[ ]:
```
mfa model download dictionary french_mfa
```



After you first downloaded an acoustic model, dictionary, g2p model etc., you will see that a folder "MFA" was created in your documents directory. This folder contains all models you downloaded:



If you want to delete a specific acoustic model, dictionary, etc., simply delete the respective file. If you want to delete all models, just delete the entire "MFA" folder.

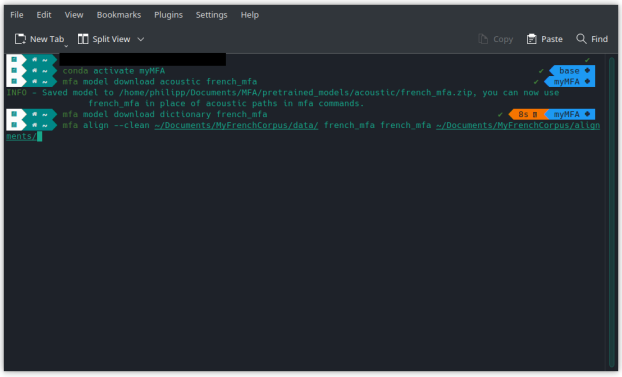
4) Generate alignments for the French data

Now that we have a corpus, an acoustic model and a dictionary, we can create the alignments for the data. We have to type in the following command to the command line:

```
[ ]: mfa align --clean /path/to/your/data /path/to/your/dictionary /path/to/your/
     ↪acoustic_model /path/to/your/output/
```

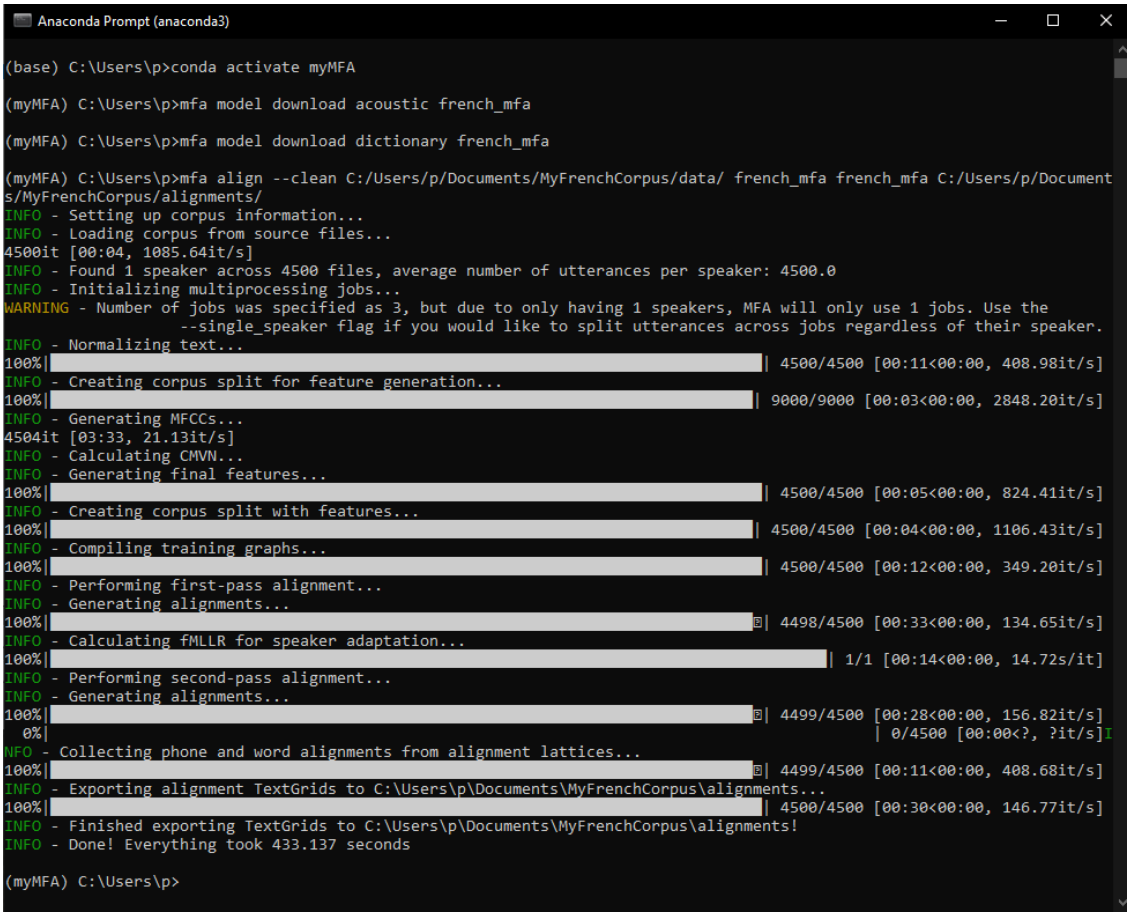
Since this will be the first alignment, the –clean flag can be omitted. You can use the –clean flag if you run the aligner mutliple times, since the files from previous runs will be removed. On Linux and Mac, you can abbreviate your home directory with ~/. Note: use the full path on Windows: C:/Users/...

Since we downloaded the french_mfa acoustic model and dictionary, we do not have to enter the full paths, we can just enter french_mfa:

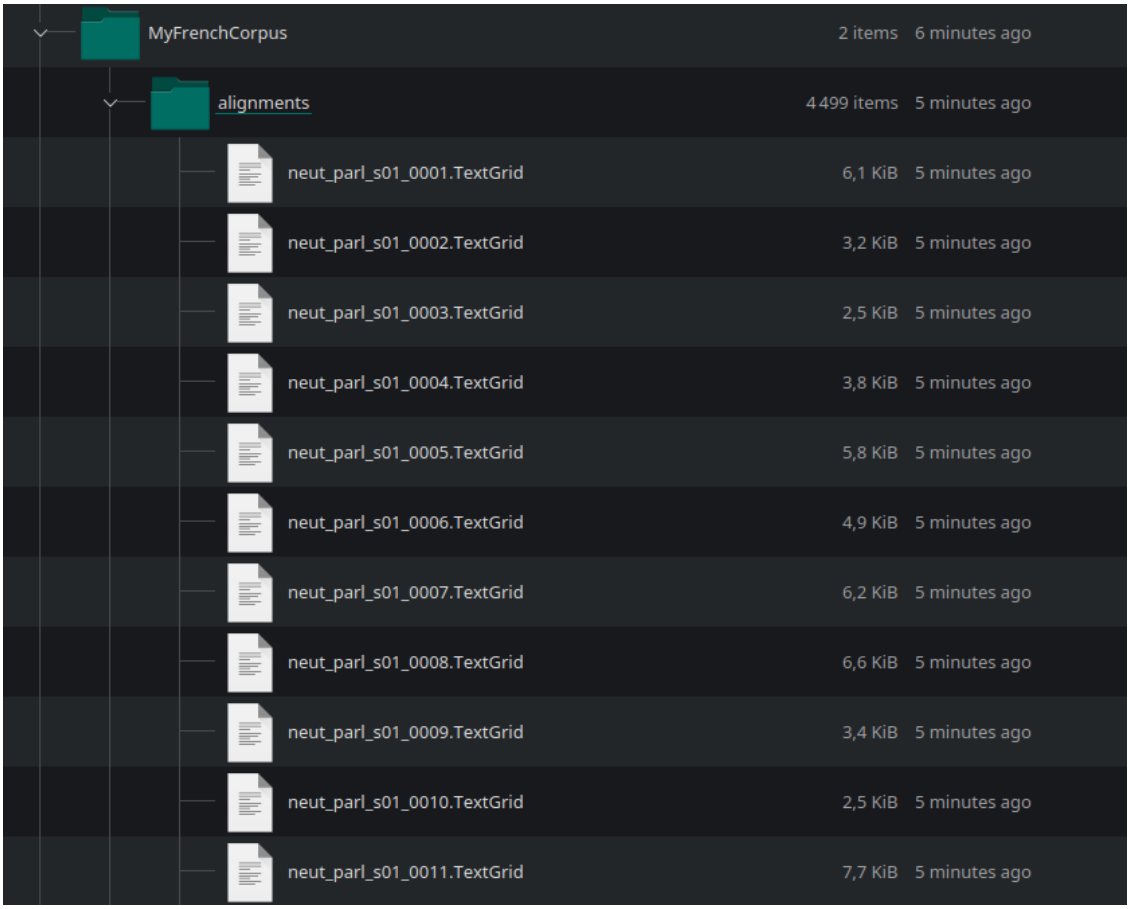


This may take some minutes, because we have a lot of data. You should see this after a successful run:

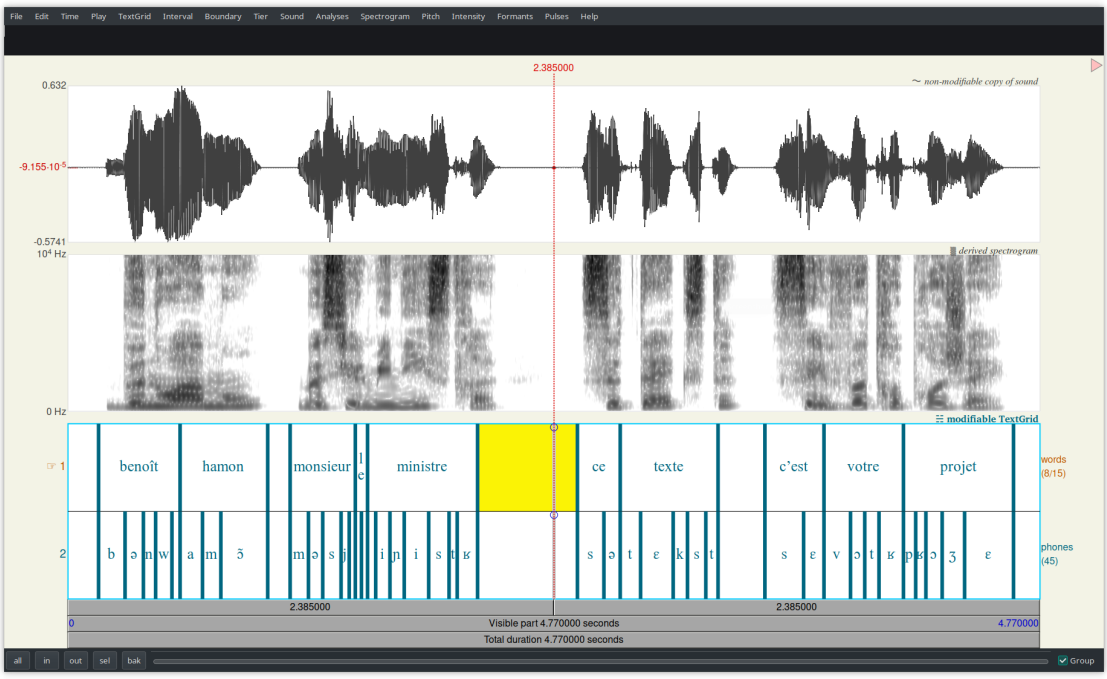
And here is an example for Windows:



You should find a folder called “alignments” in your corpus directory containing the TextGrids:



When you open them in Praat, you will see that word and phone alignments were generated:



References

Yamagishi, Junichi; Honnet, Pierre-Edouard; Garner, Philip; Lazaridis, Alexandros. (2017). The SIWIS French Speech Synthesis Database, 2016 [dataset]. University of Edinburgh. School of Informatics. The Centre for Speech Technology Research. https://doi.org/10.7488/ds/1705.