Lecture 16: Forced Alignment, ASR

LING 1340/2340: Data Science for Linguists

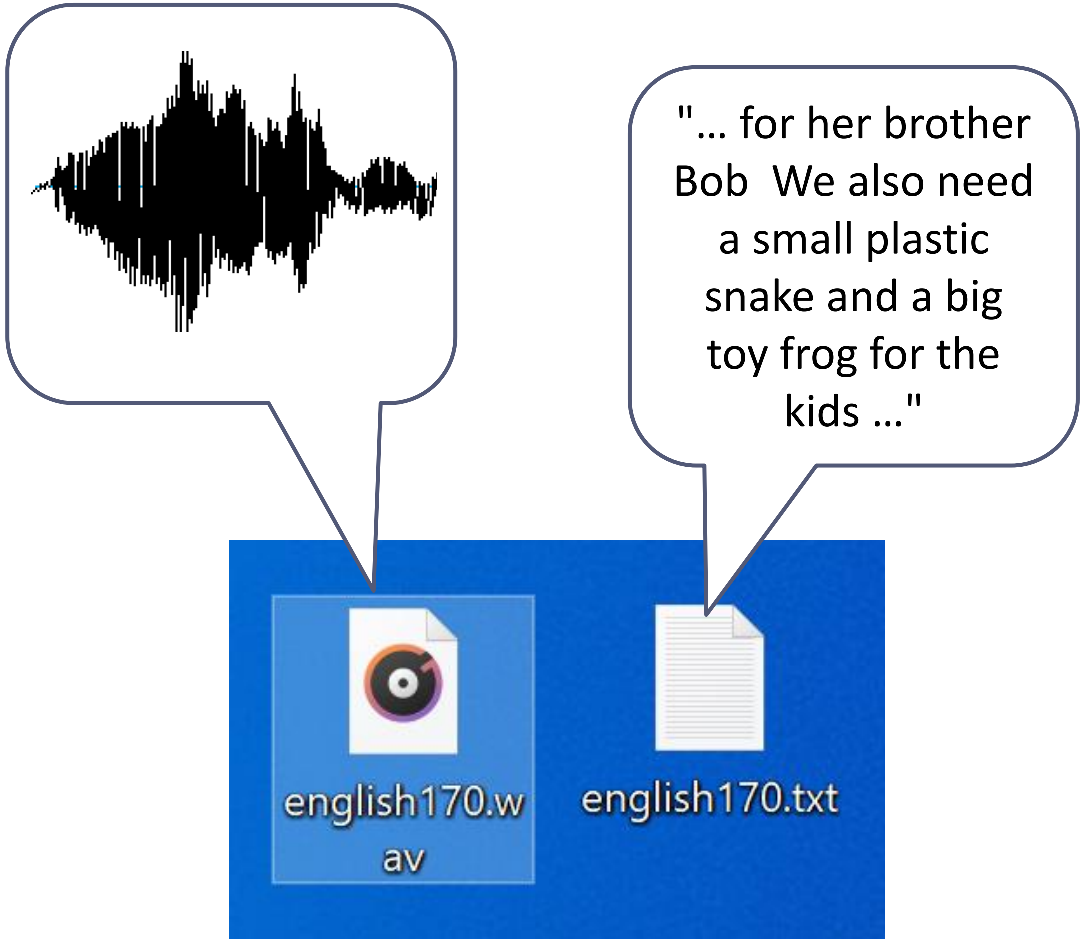
Na-Rae Han

Objectives

- Forced alignment demo: Montreal Forced Aligner
- ASR demo: SpeechRecognition library
- ASR theory → Nope, next class

Forced alignment

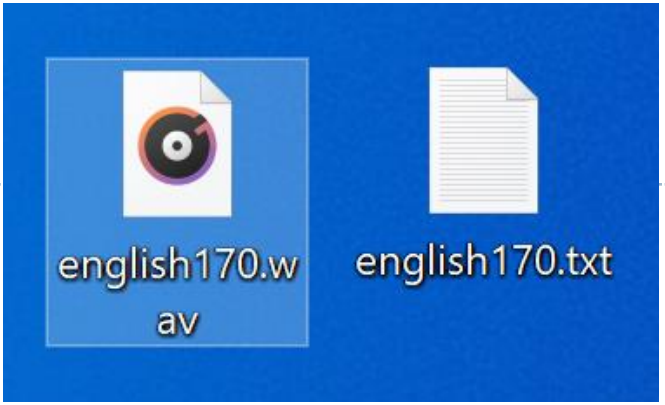
- "**Forced alignment**": automatic synchronization of a sequence of phones with an audio file.
- Purpose: speed up manual segmentation and annotation
  - Rather than manually creating phonetic transcription from scratch, correct output from forced aligner
  - Makes life easier for linguists doing speech-focused research!

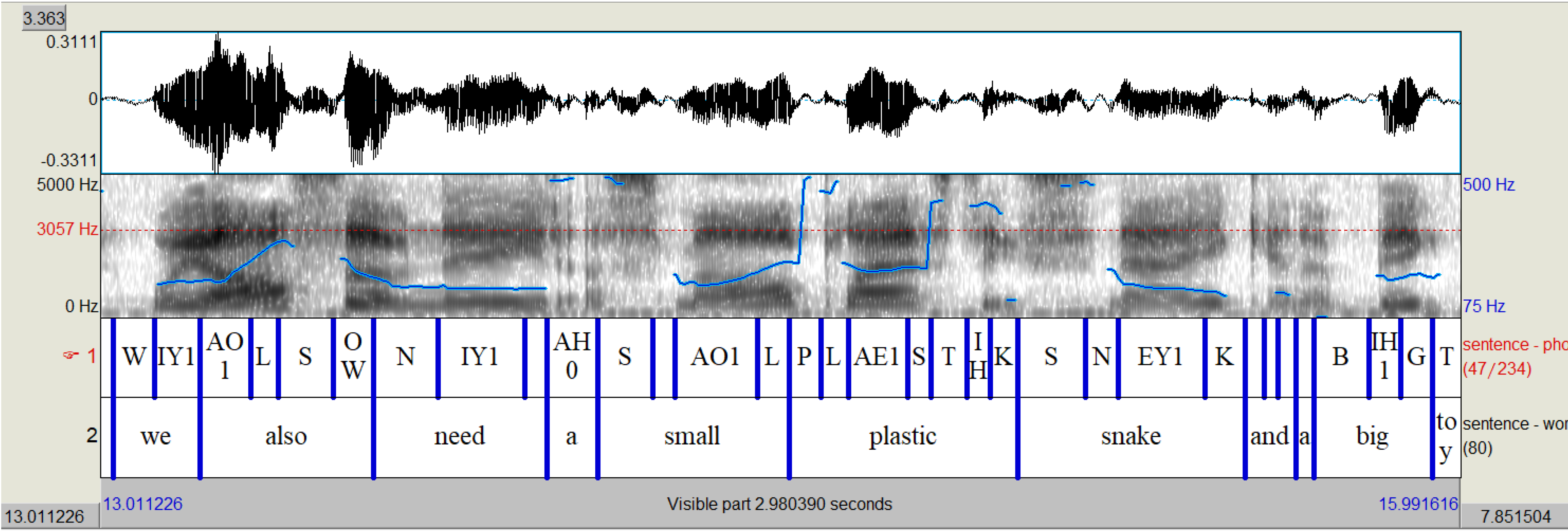


"… for her brother Bob We also need a small plastic snake and a big toy frog for the kids …"

Forced alignment

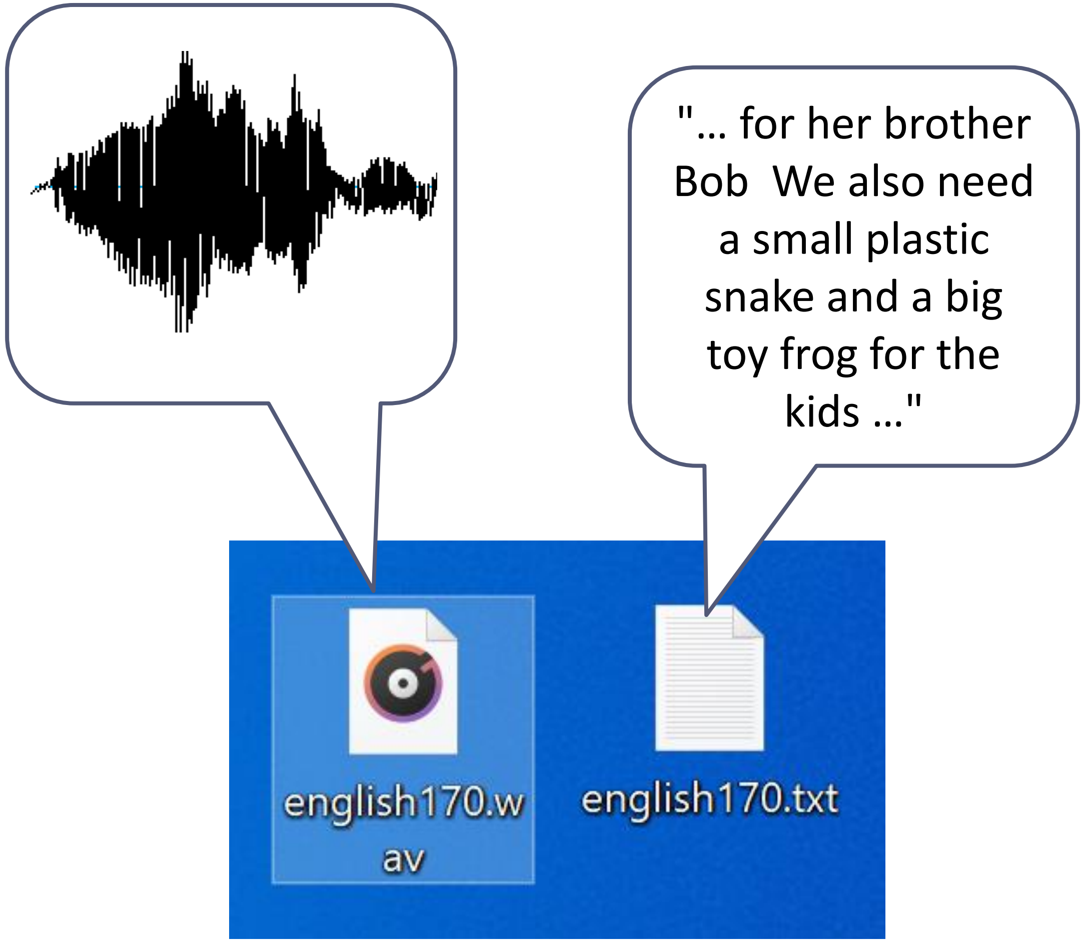
- You have: a speech file (.wav), a transcript file (.txt) →
- You want:




3.363
0.3111
0
-0.3311
5000 Hz
3057 Hz
0 Hz
500 Hz
75 Hz
1
W IY1 AO1 L S OW N IY1 AH0 S AO1 L P L AE1 S T IH K S N EY1 K B IH1 G T
sentence - pho
(47/234)
2
we also need a small plastic snake and a big toy
sentence - wor
(80)
13.011226
13.011226
Visible part 2.980390 seconds
15.991616
7.851504

Sound wave, words, phones

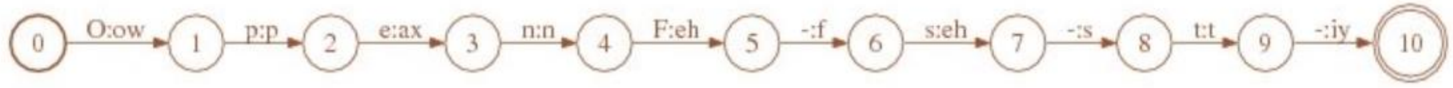
- What additional linguistic information is needed?
  - Pronunciation dictionary
    - Phonemic representations for "brother", "we", "also"…
    - More broadly: **orthography → phone mapping** (G2P, "grapheme-to-phoneme")
  - Acoustic model
    - How phonemic representation relates to sound wave



"… for her brother Bob We also need a small plastic snake and a big toy frog for the kids …"

Demo: Montreal Forced Aligner

- Home page:
  - https://montreal-forced-aligner.readthedocs.io/en/latest/
- GitHub project page:
  - https://github.com/MontrealCorpusTools/Montreal-Forced-Aligner
- Builds on popular/standard libraries:
  - **Kaldi** ASR toolkit
    - [home] [GitHub repo]
  - which builds on **OpenFST**
    - [home]



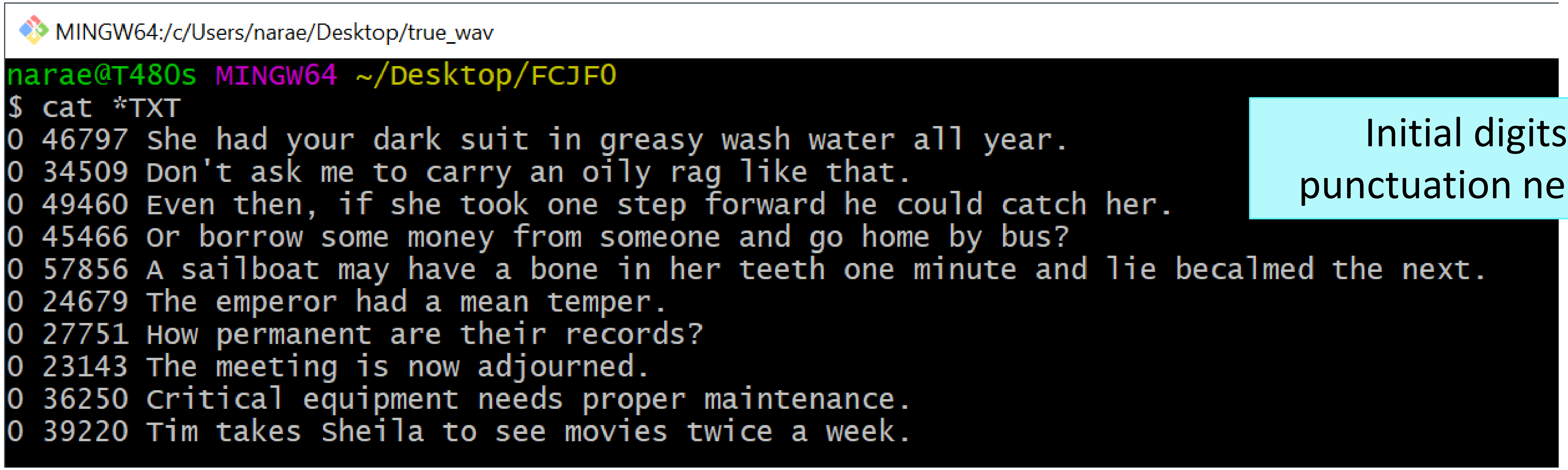
Steps (latest MFA version 2.0)

- Install Kaldi, MFA
  - Windows users: For ver 2.0, you need WSL (**W**indows **S**ubsystem for **L**inux, essentially Linux on Windows!) to use full G2P functionality. Alternatively: install older ver 1.0.1 available here, which is Windows-native.
- Prepare data to align
  - Speech files (WAV format, single-channel)
  - Transcript files (.lab or .txt format; no punctuation)

We'll use TIMIT data for demo
(pretend it came with audio files
and .TXT transcripts only)

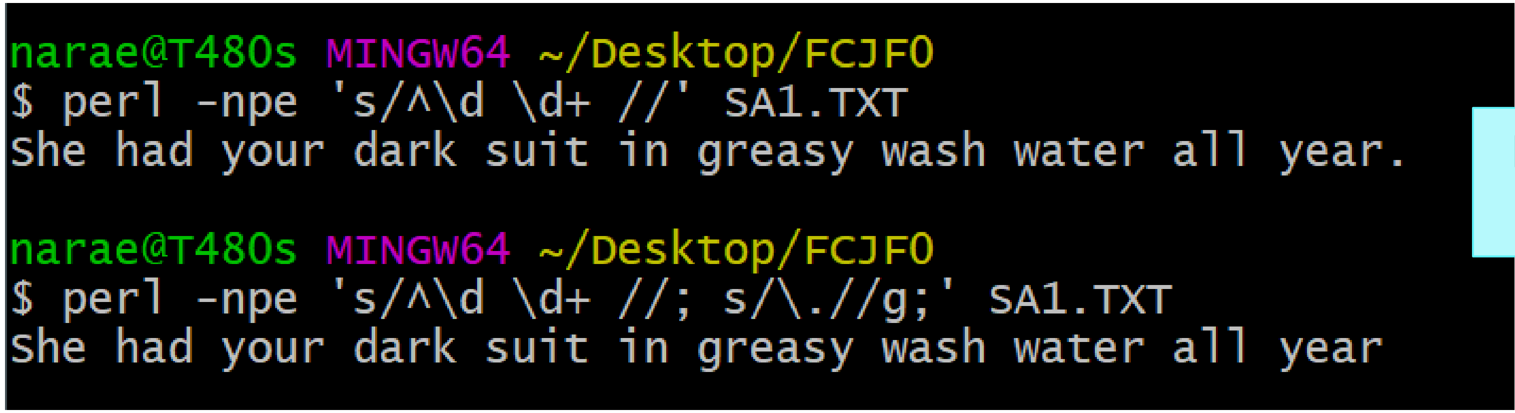
- Download language models (pre-trained, MFA offers many)
  - A pronunciation dictionary for the language
    - If not available: produce one by running language-specific G2P (grapheme-to-phoneme) on your transcript files
  - An acoustic model for the language
- Run:
  - `mfa align <input-dir> <pron-dict> <acoustic-model> <output-dir>`
- New TextGrid files in the output dir! Examine.

Cleaning transcript files



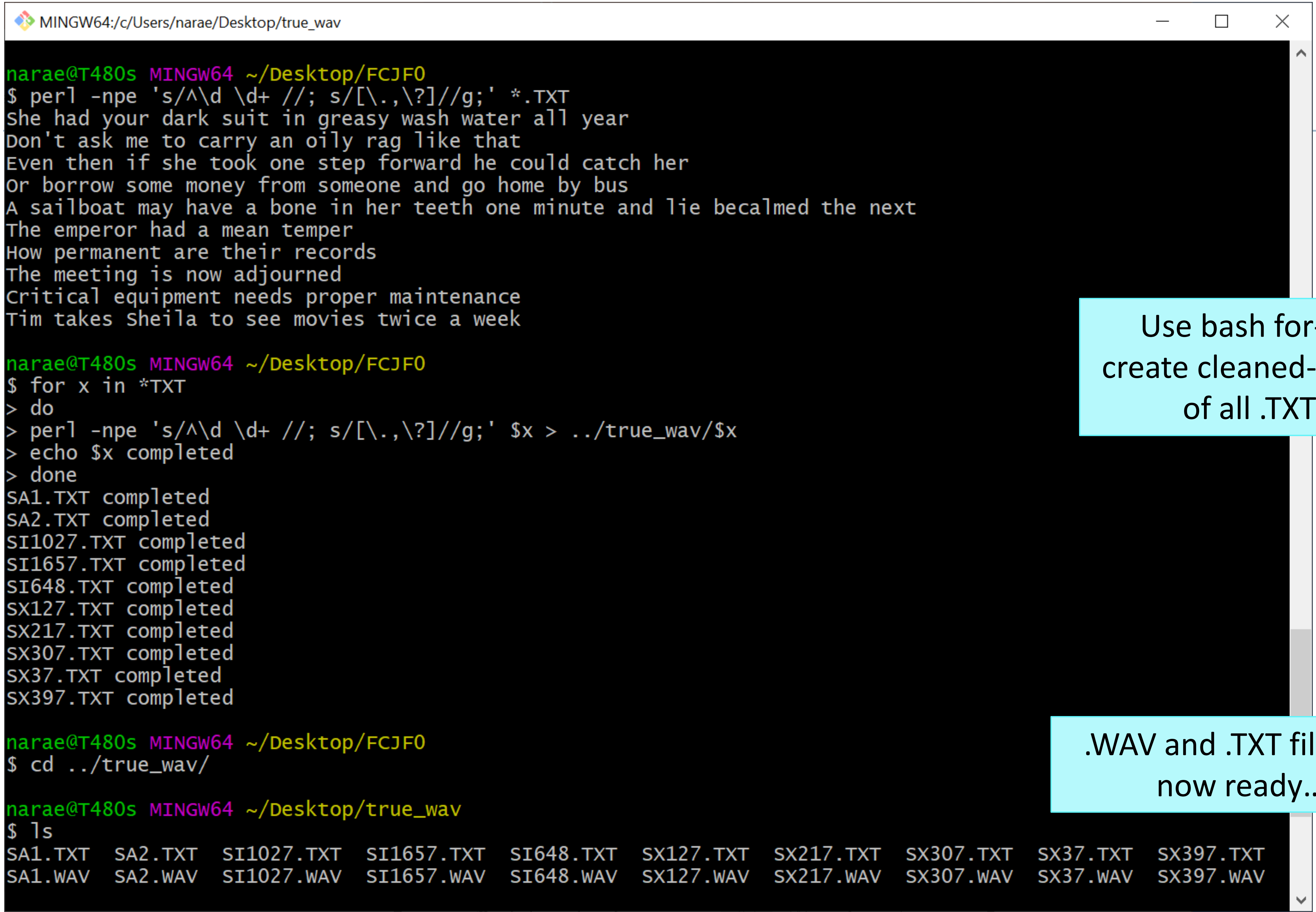
```
narae@T480s MINGW64 ~/Desktop/FCJF0
$ cat *TXT
0 46797 She had your dark suit in greasy wash water all year.
0 34509 Don't ask me to carry an oily rag like that.
0 49460 Even then, if she took one step forward he could catch her.
0 45466 Or borrow some money from someone and go home by bus?
0 57856 A sailboat may have a bone in her teeth one minute and lie becalmed the next.
0 24679 The emperor had a mean temper.
0 27751 How permanent are their records?
0 23143 The meeting is now adjourned.
0 36250 Critical equipment needs proper maintenance.
0 39220 Tim takes Sheila to see movies twice a week.
```

Initial digits and punctuation need to go



```
narae@T480s MINGW64 ~/Desktop/FCJF0
$ perl -npe 's/^\d \d+ //' SA1.TXT
She had your dark suit in greasy wash water all year.

narae@T480s MINGW64 ~/Desktop/FCJF0
$ perl -npe 's/^\d \d+ //; s/\.//g;' SA1.TXT
She had your dark suit in greasy wash water all year
```

Perl + regular expressions to clean up



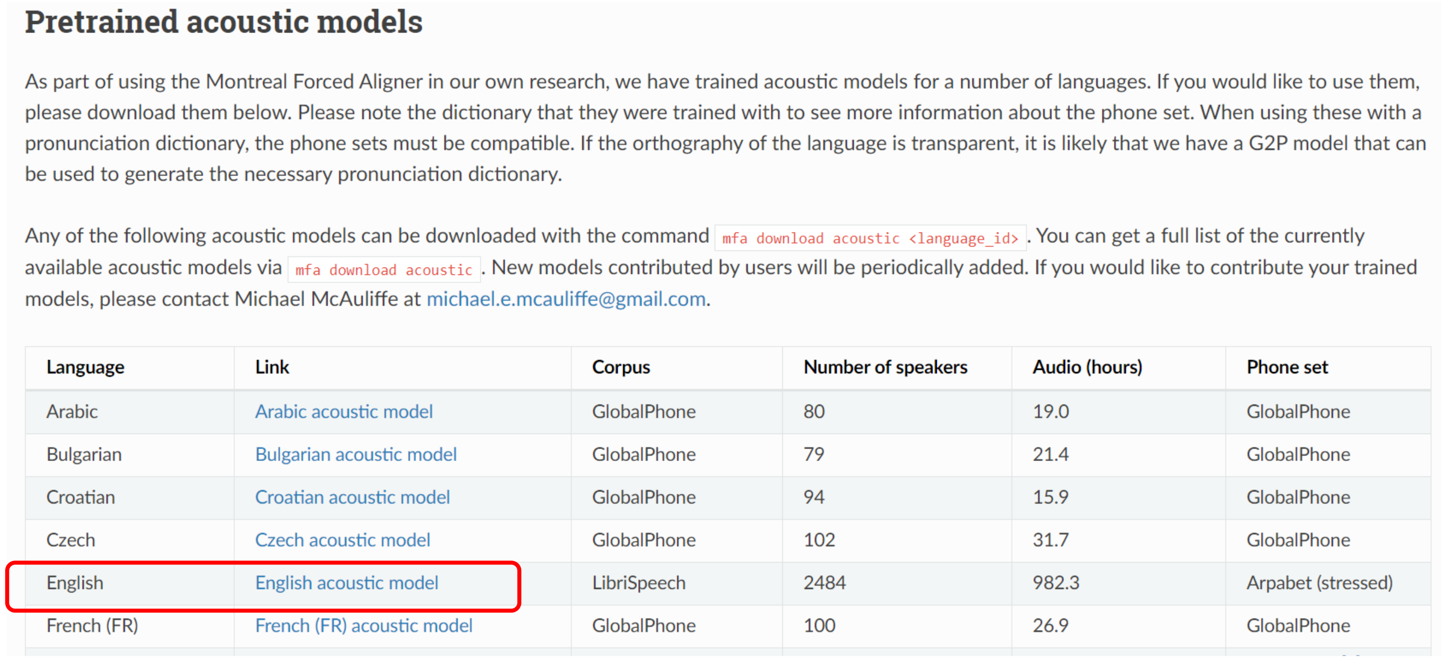
```
narae@T480s MINGW64 ~/Desktop/FCJF0
$ perl -npe 's/^\d \d+ //; s/[\.,\?]//g;' *.TXT
She had your dark suit in greasy wash water all year
Don't ask me to carry an oily rag like that
Even then if she took one step forward he could catch her
Or borrow some money from someone and go home by bus
A sailboat may have a bone in her teeth one minute and lie becalmed the next
The emperor had a mean temper
How permanent are their records
The meeting is now adjourned
Critical equipment needs proper maintenance
Tim takes Sheila to see movies twice a week

narae@T480s MINGW64 ~/Desktop/FCJF0
$ for x in *TXT
> do
> perl -npe 's/^\d \d+ //; s/[\.,\?]//g;' $x > ../true_wav/$x
> echo $x completed
> done
SA1.TXT completed
SA2.TXT completed
SI1027.TXT completed
SI1657.TXT completed
SI648.TXT completed
SX127.TXT completed
SX217.TXT completed
SX307.TXT completed
SX37.TXT completed
SX397.TXT completed

narae@T480s MINGW64 ~/Desktop/FCJF0
$ cd ../true_wav/

narae@T480s MINGW64 ~/Desktop/true_wav
$ ls
SA1.TXT  SA2.TXT  SI1027.TXT  SI1657.TXT  SI648.TXT  SX127.TXT  SX217.TXT  SX307.TXT  SX37.TXT  SX397.TXT
SA1.WAV  SA2.WAV  SI1027.WAV  SI1657.WAV  SI648.WAV  SX127.WAV  SX217.WAV  SX307.WAV  SX37.WAV  SX397.WAV
```

Use bash for-loop to create cleaned-up version of all .TXT files

.WAV and .TXT files are now ready…

Download language models

- MFA's pre-trained models:
  - https://montreal-forced-aligner.readthedocs.io/en/latest/pretrained_models.html

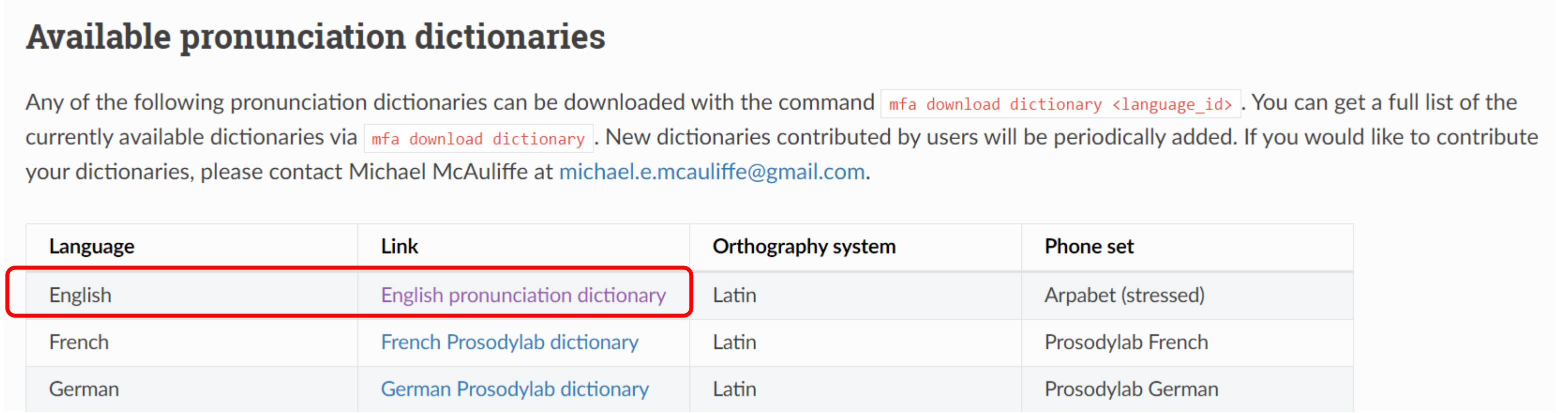


Pretrained acoustic models

As part of using the Montreal Forced Aligner in our own research, we have trained acoustic models for a number of languages. If you would like to use them, please download them below. Please note the dictionary that they were trained with to see more information about the phone set. When using these with a pronunciation dictionary, the phone sets must be compatible. If the orthography of the language is transparent, it is likely that we have a G2P model that can be used to generate the necessary pronunciation dictionary.

Any of the following acoustic models can be downloaded with the command `mfa download acoustic <language_id>`. You can get a full list of the currently available acoustic models via `mfa download acoustic`. New models contributed by users will be periodically added. If you would like to contribute your trained models, please contact Michael McAuliffe at michael.e.mcauliffe@gmail.com.

Language	Link	Corpus	Number of speakers	Audio (hours)	Phone set
Arabic	Arabic acoustic model	GlobalPhone	80	19.0	GlobalPhone
Bulgarian	Bulgarian acoustic model	GlobalPhone	79	21.4	GlobalPhone
Croatian	Croatian acoustic model	GlobalPhone	94	15.9	GlobalPhone
Czech	Czech acoustic model	GlobalPhone	102	31.7	GlobalPhone
English	English acoustic model	LibriSpeech	2484	982.3	Arpabet (stressed)
French (FR)	French (FR) acoustic model	GlobalPhone	100	26.9	GlobalPhone



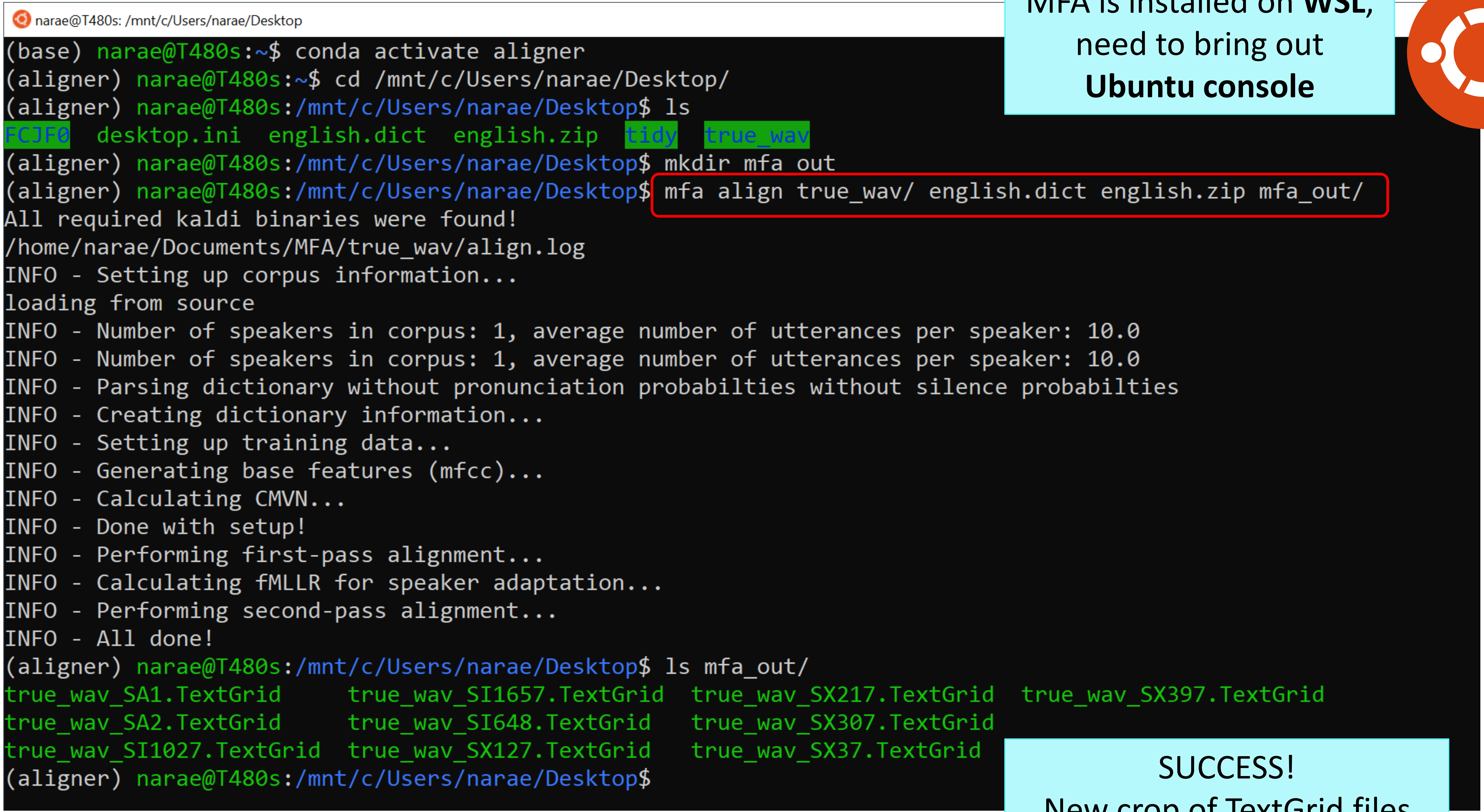
Available pronunciation dictionaries

Any of the following pronunciation dictionaries can be downloaded with the command `mfa download dictionary <language_id>`. You can get a full list of the currently available dictionaries via `mfa download dictionary`. New dictionaries contributed by users will be periodically added. If you would like to contribute your dictionaries, please contact Michael McAuliffe at michael.e.mcauliffe@gmail.com.

Language	Link	Orthography system	Phone set
English	English pronunciation dictionary	Latin	Arpabet (stressed)
French	French Prosodylab dictionary	Latin	Prosodylab French
German	German Prosodylab dictionary	Latin	Prosodylab German

CMU pronouncing dictionary

MFA is installed on **WSL**, need to bring out **Ubuntu console**

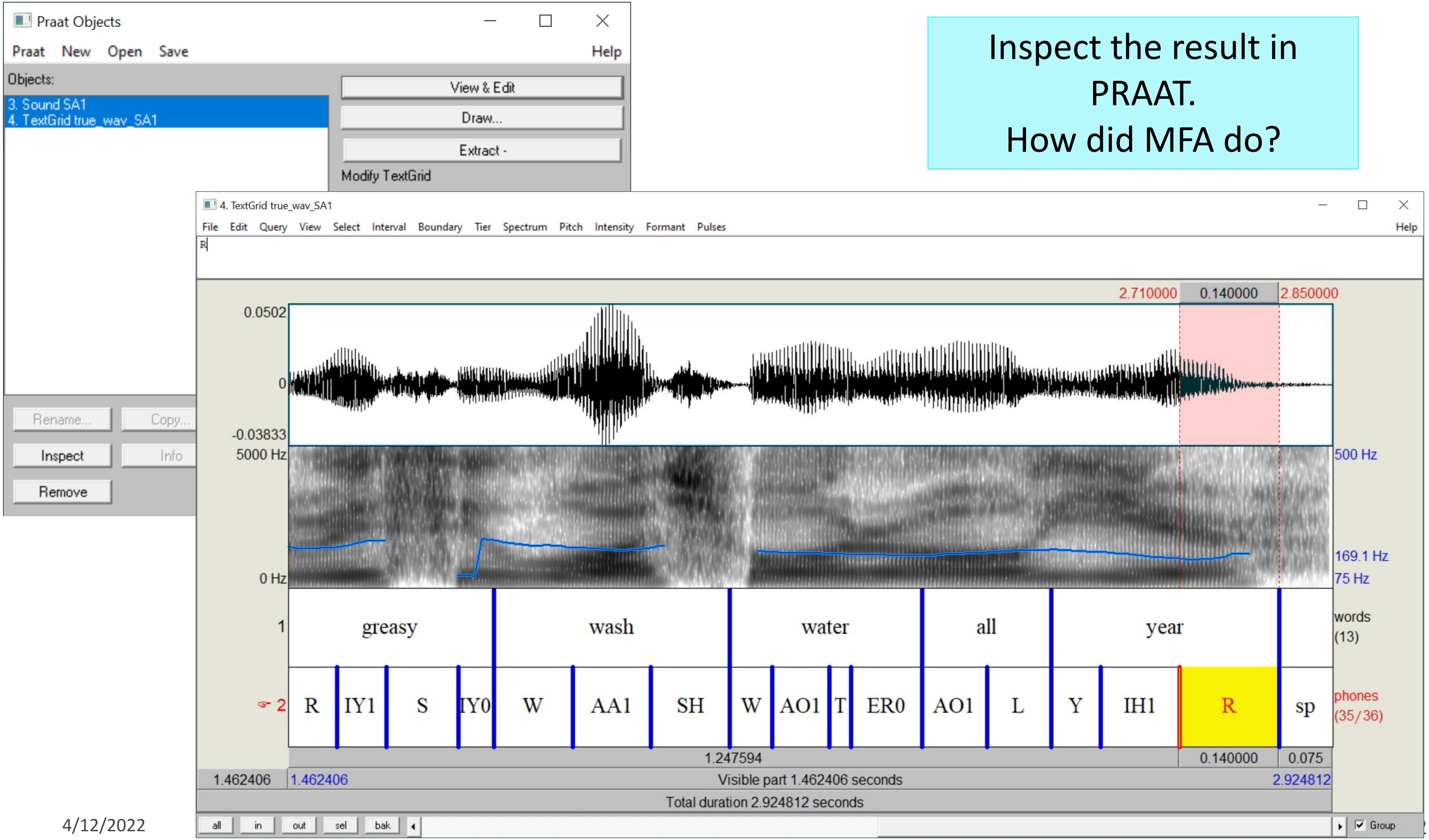


narae@T480s: /mnt/c/Users/narae/Desktop

```
(base) narae@T480s:~$ conda activate aligner
(aligner) narae@T480s:~$ cd /mnt/c/Users/narae/Desktop/
(aligner) narae@T480s:/mnt/c/Users/narae/Desktop$ ls
FCJF0  desktop.ini  english.dict  english.zip  tidy  true_wav
(aligner) narae@T480s:/mnt/c/Users/narae/Desktop$ mkdir mfa_out
(aligner) narae@T480s:/mnt/c/Users/narae/Desktop$ mfa align true_wav/ english.dict english.zip mfa_out/
All required kaldi binaries were found!
/home/narae/Documents/MFA/true_wav/align.log
INFO - Setting up corpus information...
loading from source
INFO - Number of speakers in corpus: 1, average number of utterances per speaker: 10.0
INFO - Number of speakers in corpus: 1, average number of utterances per speaker: 10.0
INFO - Parsing dictionary without pronunciation probabilties without silence probabilties
INFO - Creating dictionary information...
INFO - Setting up training data...
INFO - Generating base features (mfcc)...
INFO - Calculating CMVN...
INFO - Done with setup!
INFO - Performing first-pass alignment...
INFO - Calculating fMLLR for speaker adaptation...
INFO - Performing second-pass alignment...
INFO - All done!
(aligner) narae@T480s:/mnt/c/Users/narae/Desktop$ ls mfa_out/
true_wav_SA1.TextGrid     true_wav_SI1657.TextGrid  true_wav_SX217.TextGrid  true_wav_SX397.TextGrid
true_wav_SA2.TextGrid     true_wav_SI648.TextGrid   true_wav_SX307.TextGrid
true_wav_SI1027.TextGrid  true_wav_SX127.TextGrid   true_wav_SX37.TextGrid
(aligner) narae@T480s:/mnt/c/Users/narae/Desktop$
```

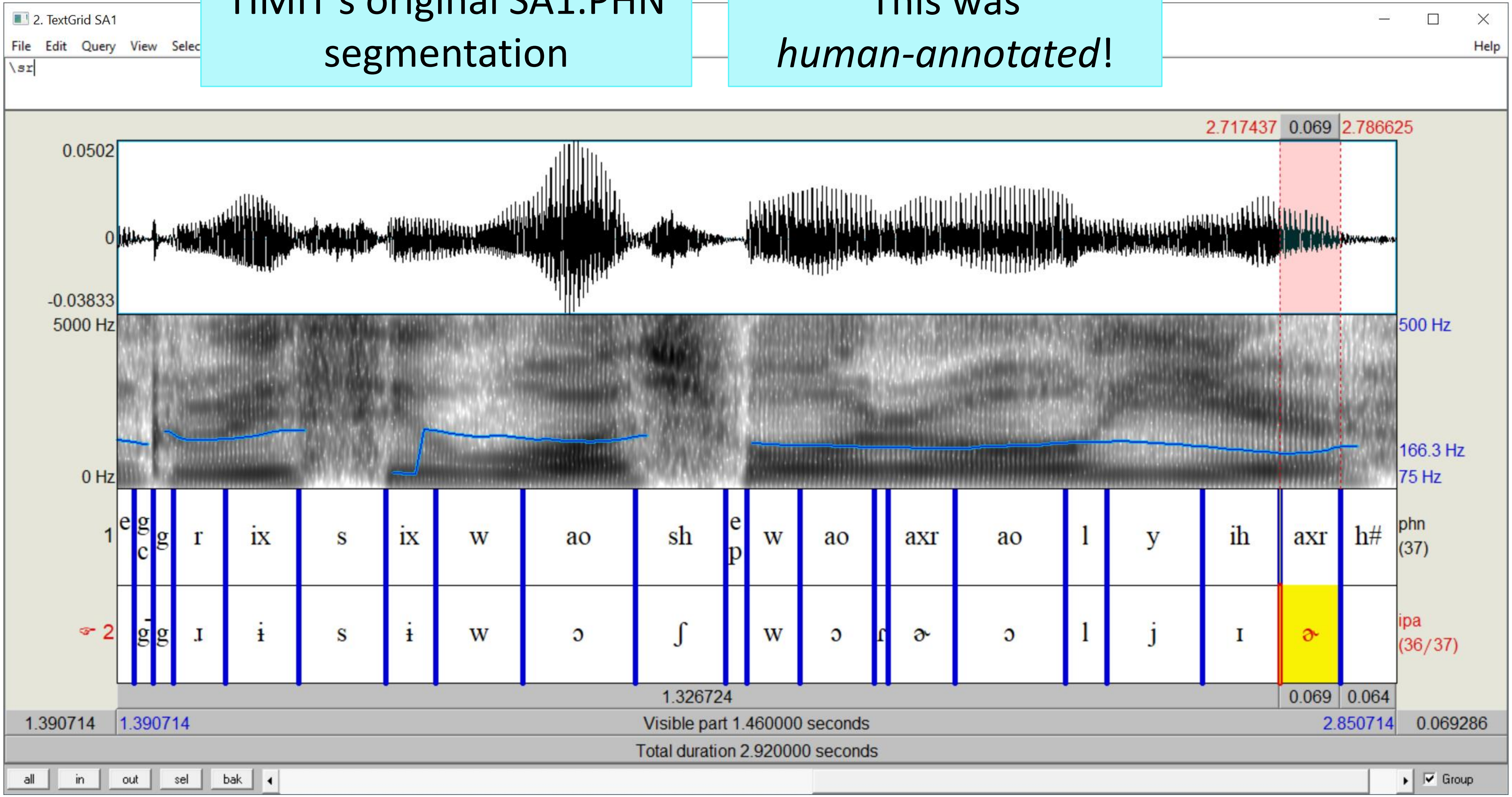
SUCCESS!
New crop of TextGrid files

Inspect the result in PRAAT.
How did MFA do?



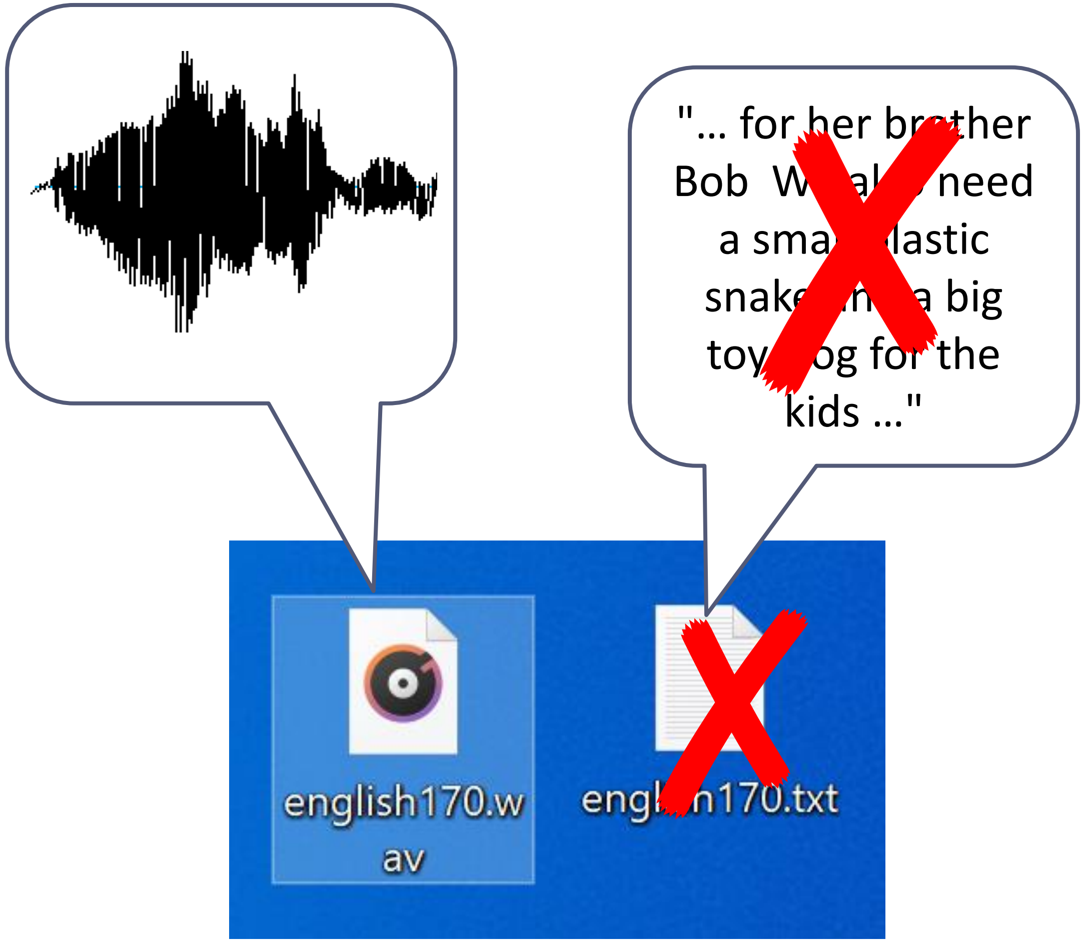
Compare with TIMIT's original SA1.PHN segmentation

This was *human-annotated*!



But, what if we don't have a transcript?

- Suppose all we have is audio files…
- Hire people to manually transcribe
- … Or, go for ASR (automatic speech recognition)
  - Forced alignment is based on ASR.



"… for her br ther Bob W al need a sma lastic snake n a big toy og for the kids …"

ARS demo

- With SpeechRecognition library
- In Jupyter Notebook!

Spoken language + Python

- Praat in Python
  - Libraries: `Praat-textgrids`, `Parselmouth`
- `SpeechRecognition` library
  - https://github.com/Uberi/speech_recognition
  - Speech recognition module for Python, supporting several engines and APIs, online and offline.
- DataCamp course: Spoken Language Processing in Python
  - https://learn.datacamp.com/courses/spoken-language-processing-in-python
  - Libraries covered: `wave`, `SpeechRecognition`, `PyDub`

DataCamp subscription until Jul 10!

Wrapping up

- Next class:
  - ELAN demo by Lindsey
  - Project presentation: Misha
  - Intro to ASR

- Final project submission:
  - May 1 (Sun) 6pm
  - If using 2-day late pass, email and LET ME KNOW before SUNDAY! (Final grade is due on Wed)