This is Hossep’s notebook to share on how he did alignments

# Log

I generally followed the strategy in <https://github.com/jhdeov/interlingual-MFA>

* I didn’t use any beams, but just speaker adaptation.
* The original dictionary from <https://github.com/jhdeov/armenian-intonation> was turned to lowercase
* I ran the Armenian data against english\_mfa, spanish\_mfa, french\_mfa, and german\_mfa. These models were trained over 1000s of hours of data because they’re rich people.
* I kept a copy of the phone\_mapping that I used. For non-English, I suspect that there are probably more accurate possible mappings. For example, Armenian /ʁ, ɾ, r/ are all mapped to French /ʁ/, because I don’t know French – you can see French’s phones and other lang’s phones here [[Fr](https://mfa-models.readthedocs.io/en/latest/dictionary/French/French%20MFA%20dictionary%20v2_0_0a.html#french-mfa-dictionary-v2-0-0a), [Gr](https://mfa-models.readthedocs.io/en/latest/dictionary/German/German%20MFA%20dictionary%20v2_0_0a.html#german-mfa-dictionary-v2-0-0a), [Sp](https://mfa-models.readthedocs.io/en/latest/dictionary/Spanish/Spanish%20MFA%20dictionary%20v2_0_0a.html#spanish-mfa-dictionary-v2-0-0a), [En](https://mfa-models.readthedocs.io/en/latest/dictionary/English/English%20MFA%20dictionary%20v2_0_0a.html#english-mfa-dictionary-v2-0-0a)]
* Below are the snippets that I used.
* I also ran an Armenian-based aligner [[Hy](https://github.com/jhdeov/armenianMFA/)] just for fun, no beam changes

# CODE

# convert dictionary

python convertPronDict.py /Users/Hovsep/Desktop/word-pronunciations-EA.tsv /Users/Hovsep/Desktop/phoneMappingFr.txt /Users/Hovsep/Desktop/pronDictIntermediate.txt

# validate dictionary

mfa validate /Users/Hovsep/Desktop/EA /Users/Hovsep/Desktop/pronDictIntermediate.txt french\_mfa --ignore\_acoustics

# align

mfa align /Users/Hovsep/Desktop/EA /Users/Hovsep/Desktop/pronDictIntermediate.txt french\_mfa /Users/Hovsep/Desktop/EA\_Fr --clean --overwrite --Speaker\_characters=4

# convert alignments

python convertAlignments.py wordTranscriptions.pkl /Users/Hovsep/Desktop/EA\_Fr

# TODOs:

* It would be nice to play with inter-annotator agreement via these aligners
* Perhaps the non-English models can be improved by one of you refining the phone-mapping files, by using knowledge of European phonetics wrt to English/Armenian
* The pauses may be problematic: <https://montreal-forced-aligner.readthedocs.io/en/latest/user_guide/corpus_creation/training_dictionary.html?highlight=pause#silence-probability-and-correction-factors> [[1](https://github.com/MontrealCorpusTools/Montreal-Forced-Aligner/issues/386),[2](https://github.com/MontrealCorpusTools/Montreal-Forced-Aligner/issues/222)];