

How to add a new language to EasyAlign ?

Read the 2 communications available at <http://latlntic.unige.ch/phonetique/> (EasyAlign at Interspeech 2011 and EasyAlignSpanish at 5thCFE 2011)

Here are the usual steps to adapt a new language to EasyAlign. Some are easy, other may be insurmountable. Read it through first. Basically, 2 things are needed:

1. Few hours of clean speech data
2. A grapheme-to-phoneme conversion tool

Here are the detailed steps:

1. Get the ISO-639-3 code of your language on <http://www.sil.org/iso639-3/codes.asp> . If it does not exist (e.g. dialect), suggest one !
2. Make a phoneme inventory with SAMPA alphabet format (<http://www.phon.ucl.ac.uk/home/sampa>). Make a text file with a phoneme per line and a word as an example as follows (in French)

sampa	example
p	pas
t	tas
k	cas
b	bas
...	

Name it with the code and .Table extension (e.g. fra.Table)

3. Find or record 3 hours of reading aloud with a minimum of 6 different speakers (both genders). Use a good quality recording. It may be useful later for another purpose. The text must be varied. Beware that non-frequent phonemes appear many times. (>50-100 times) Try to make files of 10 minutes or less. Use WAV format. Name them with .wav extension.
4. Make a macro-segmentation of it at utterance level.i.e make a TextGrid for each sound file with an interval tier named 'ortho' and the orthographic transcription. You can use the step#1 Macro-segmentation of EasyAlign. Place a single boundary in the middle of the pause. Make sure you have a pause at the beginning and the end of each utterance and to name the TextGrid with the same name as the sound file but with the .TextGrid extension
5. Find a grapheme-to-phoneme conversion tool working in command-line mode. Preferably it takes two arguments : the input text file (with orthographic transcription) and the output text file (with the phonetic transcription). It must preserve spaces and newlines. i.e. each line must have the same number of words (and thus space) in the input and the output. The number of lines in the output file must be the same as in the input file.
6. Find a nice 20 second oral example (for the demo mode) with its transcription

7. Send it to me (phoneme inventory + g2p tool + example + TextGrids) so I can develop the first step. I'll return to you a new version of EasyAlign with your language partially integrated.
8. Now go through step#2 Phonetization with all your files to make the 'phono' tier. Don't forget to save. Verify meticulously the phonetic transcription from the sound. In practical, listen to the sentence, one by one, and check the phono tier.
9. Send to me all the new verified TextGrid files.
10. Hopefully I'll train the acoustic models of the phones and finish up the integration
11. Not done ! Evaluation: record another few minutes of few new speakers. 3 minutes x 5 speakers would be nice. Transcribe, phonetize, check the phonetization. Then I'll send to you TextGrid and Sound to align manually. Then I'll do some statistics to evaluate the precision.