

Fall 2018 Progress Update - Yda

Kaldi and Pytorch - Going through tutorials

Pytorch - Decided on due to comfortability with python and (hopefully) ease of use

Expect to try first tests with german data in next week or two

Expected issues:

- Data for certain phonemes not present in standard german
- Differences in lexicon for certain words

Fall 2018 Progress Update - Yda

Data:

- University of Hamburg Language Technology group has open resources <https://www.inf.uni-hamburg.de/en/inst/ab/lt/resources/data.html>
 - Use for tutorials at the moment, can use for training in future full tests
- Deutsche Welle
 - Newscasts in standard German, also spoken slowly with transcripts available
 - Downloadable as MP3, can be divided into segments as well
 - Can use as standard training data or testing
- Dialects
 - Have not found available dataset to test on yet
 - Possible to record ~2 or 3 different dialects from 1 or 2 native speakers each in the coming months