# Pre-training on High-Resource Speech Recognition for Low Research Speech-to-Text Translation

In our research we will recreate the model results of the paper, “Insert Paper Title Here” and improve upon this line of research by adding the following capabilities: “FIRST”, “SECOND”, “THIRD”.

The state-of-the-art (SOFT) technologies for creating speech to text translation include a technique call pre-training and fine-tuning. In this technique a model is trained on a high resource language and fine-tined on low-resource language. **High-resource languages** are defined as [definition of high-resource language]. **Low-resource languages** are defined as [insert definition of low-resource language]. Pre-training and fine-tuning is a technique which consists of the following steps: [step 1, step 2,…,step 3]. Research has shown that this method can increase performance metrics in [insert name of metric scores][insert citation]. In Ref. [1], the authors show that adding the transcriptions of audio in a high resource language can add to the performance of a multi-source shared attention model. They compared their performance to several baseline ensemble models evaluated using 2 parallel datasets: *Mboshi-French* and *Ainu-English*. The results show that the authors were able to reduce the Character-Error-Rate (CER) by [error rate1, error rate2] for Mboshi-French and Ainue-English.

The datasets are part of the open-source data set CALLHOME and are freely available for download. For development, our team plans on using the HOOPER GPU cluster available to all students at GMU. For guidance we plan on leveraging Dr. Antonis Anastasopoulos as a resource.

The current code for the model is available on GitHub and is written in DYNET. Our first contribution to the field will be to recode the model in PyTorch. PyTorch is the most popular framework for developing Neural Network models and converting the code will make this research more freely available to the broader research community. We will then use this recoded model as our baseline for measuring the change in performance of our work. The next phase will be to improve the model.

We plan on contributing to the field of NLP Endangered language research by building on the previous work completed my Dr. Antonis Anastasopoulos in Ref [1]. The model described is a multi-source shared attention model. Many authors have shown that the performance of a model can be improved by pre-training on a high-resource language and fine-tunning on a down stream task [1,2]. In addition, researcher have shown that multi-lingual models, especially those languages who are in the same language family, can improve performance in few-shot setting (i.e. those scenarios where only a few translations or transcriptions are available for training) [3]. Out team will extend the model developed in [1] to include several additional high-resource translations and transcriptions as training data. Performance of the model will be compared to the original paper and to the baseline model which is defined as the PyTorch version of the shared-attention network.

# Bibliography

1. Anastasopoulos, A., & Chiang, D. (2018). Leveraging translations for speech transcription in low-resource settings. arXiv preprint arXiv:1803.08991.
2. Bansal, S., Kamper, H., Livescu, K., Lopez, A., & Goldwater, S. (2018). Pre-training on high-resource speech recognition improves low-resource speech-to-text translation. arXiv preprint arXiv:1809.01431.
3. Lauscher, A., Ravishankar, V., Vulić, I., & Glavaš, G. (2020). From zero to hero: On the limitations of zero-shot cross-lingual transfer with multilingual transformers. arXiv preprint arXiv:2005.00633.