**Pre-training on High-Resource Speech Recognition for**

**Low Research Speech-to-Text Translation**

Speech Translation (ST) has many applications, such as documenting languages or speeches to support operators that need to understand foreign-spoken customers, with very effective results in end-to-end ST tasks [1]. Typically, it can be designed as a pipeline composed of an automatic speech recognition (ASR) component, followed by a machine translation (MT). ST requires several hours of audio with its corresponding translation texts, which are likely to obtain in high-resource language contexts. However, this kind of data is scarce for low-resource languages, where audio data or text transcription is missing to implement ST [2]. To this end, authors have proposed several approaches by taking advantage of transfer learning in the NLP area or implementing novel approaches to obtain the corresponding transcription [1, 2, 3]. For example, Bansal et al [1] use the parameters of pre-trained encoder-decoders in high-resource languages to initialize encoder-decoders in low-resource languages. Their results show outstanding results even though the languages are not the same (e.g., using the parameters of an encoder learned in French ASR data, authors obtain an improvement of around 2.0 points regarding BLEU of the Spanish-English translation task). Also, Anastasopoulos & Chiang [2] proposed an ensemble architecture where two similar data with different goal tasks are merged to obtain better performances for another task, which in the author’s case is speech transcription (related to language documentation). Finally, Conneau et al [3] propose an unsupervised framework to create cross-lingual speech recognition. They show the virtues in several tasks for low-source language: (1) multilingual pretraining outperforms monolingual training (especially English-only), (2) the learned representation transfers well to unseen languages, and (3) the cross-lingual representation improves the low-source language understanding. However, although the learning transfer can provide effective solutions, certain limitations should be considered. In this regard, Lauscher et al [4] propose four key questions to acknowledge learning transfer and provide answers by performing several experiment setups and the corresponding analysis of the results. Some results highlight the dependency on the linguistic (dis)similarity between the source and target language, the pretraining corpus size, and the task type.

**References**

[1]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*.

[2]Anastasopoulos, A., & Chiang, D. (2018). Leveraging translations for speech transcription in low-resource settings. *arXiv preprint arXiv:1803.08991*.

[3] Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., ... & Stoyanov, V. (2019). Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

[4] 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*.