# Examining the Effects of Cross-Lingual Pre-Training on Downstream Translation: A Case Study in Middle English

**Alex Lyman**
University of Pennsylvania
alyman@seas.upenn.edu

## 1 Introduction

One of the problems inherent to Low-Resource Machine Translation is that of data sparsity. Researchers working in low-resource contexts must make sensible decisions about model architecture to maximize the effectiveness of limited data and often, limited compute. Today, state of the art machine translation systems leverage large transformer-based language models which are massively pretrained on text from one or more languages, usually including the source or target language. In this study, we explore the effect of cross-lingual pretraining on models finetuned on the downstream task of translating Modern English to Middle English, a previously unseen language. We introduce a dataset of several thousand aligned Middle-English/Modern English sentence pairs, and fine-tune several checkpoints of T5 (Raffel et al., 2020) and MT5 (Xue et al., 2021). We then compare the performance of cross-lingual pretrained models to those without cross-lingual pretraining.

## 2 Methods

We introduce a dataset of roughly 58000 aligned Middle-English/Modern English sentence pairs comprising John Wycliffe's Bible, the complete works of Geoffrey Chaucer, and other contemporary poetry and prose.[1] Using publicly-available checkpoints from the huggingface hub, we finetune T5-small, T5-base, and T5-Large as well as mT5-small, mT5-base, and mT5-Large until convergence using a single A100 gpu. We evaluate our models using the sacreBLEU metric(Post, 2018).

## 3 Results

We find that mT5 performs marginally better (2-3 BLEU points) than T5 at each size checkpoint, suggesting that cross-lingual pretraining provides benefits on downstream translation tasks. This increase in performance is commensurate with scaling up model size. (mT5-small performs similarly to T5-base, mT5-base performs similarly to T5-large.) Our experiments suggest that cross-lingual pretraining can be a useful technique for improving the performance of machine translation systems on low-resource language pairs, including previously-unseen languages like Middle English. We suspect that although mT5 is never specifically trained on Middle English, features from other languages in the training set allow for a more robust 'understanding' of Middle English.

| **Checkpoint Size** | **T5** | **mT5** |
|---|---|---|
| Small | 7.1 | 8.9 |
| Base | 8.8 | 12.6 |
| Large | 12.4 | 15.7 |

Table 1: Test accuracy (BLEU) on English/Middle English translation.

## 4 Conclusion

In this study, we investigated the impact of cross-lingual pretraining on the task of translating Modern English to Middle English. Our results show that models pretrained on multiple languages, such as mT5, outperformed their monolingual counterparts, indicating that cross-lingual pretraining can be an effective technique for improving machine translation performance on low-resource language pairs. Our experiments also demonstrated (unsurprisingly) that larger models perform better, with mT5-large achieving the highest accuracy. Our findings suggest that cross-lingual pretraining could be a useful approach for improving machine translation systems for historical languages and other low-resource language pairs.

[1] Data available at: https://huggingface.co/datasets/Qilex/EN-ME

## References

Matt Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.