# Shakespearean to Formal Modern English: A Stylized Neural Machine Translation Approach

**Dipti Lohia, Sharanya Kamath, Yamini Kashyap, Avantika Borikar**
College of Information and Computer Sciences
University of Massachusetts Amherst

## Abstract

Archaic texts are hard to understand for modern readers, and direct translation into modern language often results in grammatical mistakes or unprofessional sentence structures. In this project, we are interested in neural machine translation of text followed by style transfer to convert archaic Shakespearean texts into formally correct modern English. We will be implementing two models - an NMT and a text style transfer model, and perform iterative bidirectional knowledge transfer between these two models so that they learn from each other.

## 1 Introduction

Linguistic style is an integral aspect of natural language communication. It conveys the social context in which communication occurs and defines particular ways of using language to engage with the audiences to which the text is accessible. Text style transfer is the process of changing the style of a sentence by rewriting it in a new style while maintaining its semantic content. Stylized neural machine translation seeks to convert sentences of one style into sentences of a different style, which is important for a variety of Natural Language Processing (NLP) applications, such as non-native speaker assistants and for education purposes.

In this study, we aim to perform stylized neural machine translation. Our stylized Neural Machine Translation (NMT) model aims to translate sentences into target language while changing the style of the source sentences and preserving their style-independent content. By the traditional sequential approach (Cohn and Lapata, 2007), we aim to conduct this in two steps: First step, is to translate the sentence from source language to target language and second step is to apply style transfer on the target translated language sentence. But as this process involves decoding in two steps, this approach is computationally expensive as well as has potential error propagation problem.

The performance of Neural machine translation depends on the availability of parallel data. We also aim to analyze the performance by implementing an NMT model (Shakespeare to Modern English student model) that leverages the existing style transfer model (Informal English to Formal English teacher model) to guide the learning process of the student model on the parallel data corpus (Chen et al., 2017). This will facilitate direct optimization of the intended NMT model without any requirement of separate two decoder computations thus will avoid any error propagation. Further, We would also study the model performance by extending the previous student teacher model to have a bidirectional knowledge transfer between them (Wu et al., 2021a).

## 2 Related work

RNN based encoder-decoder methods, first introduced by (Cho et al., 2014b), have been the most common way to approach the machine translation problem in the past. (Cho et al., 2014a) uses this method alongside a gated recursive convolutional network (grCNN) for translating French to English. Even though the grCNN learns the grammatical sentence structure automatically, it faces a bottleneck as the input sentence size increases. (Kabbara and Cheung, 2016) uses an LSTM based encoder-decoder architecture for the style transfer task. This method is capable of automatically capturing stylistic nuances instead of hard-coding the feature vectors. The three criteria they use for evaluation is soundness, coherence, and effectiveness of the output text. This is mainly done by human-evaluation based on Likert ratings. In our approach we plan to use automatic evaluation metrics such as BLEU along with human-evaluation.

A much newer study proposed by (Hu et al., 2020) gives an exhaustive insight into controllable text generation, neural machine translation, and text style transfer. They also explain the major challenge faced in these problem statements - the unavailability of parallel corpus i.e. parallel sentences with same semantic meaning but different styles. (Shen et al., 2017) and (Fu et al., 2018) approach the challenge of non-availability of parallel data by using adversarial networks to separate content and style representations. They also introduce two new metrics for evaluation - content preservation and transfer strength. (Sudhakar et al., 2019) solves the same problem by using Generative Style Transformer (GST) which is part of a larger 'Delete Retrieve Generate' framework.

(Wu et al., 2021b) deals with first translating a source language to a target language and then applying certain style transfers to the output. They perform this stylized neural machine translation (NMT) by implementing bidirectional knowledge transfer and distillation between an NMT model and a informal-to-formal style transfer model. We will be following this paper as a reference for our Shakespearean to modern English translation with a formal language style transfer.

# 3 Your approach

The framework we propose consists of two models: (1) A Text Style Transfer Model and (2) A stylized Neural Machine Translation (NMT) Model. Our stylized NMT model aims to convert sentences from Shakespearean English (source language) to Modern English (target language) while changing the style of the source sentences (informal to formal).

Figure 1 illustrates the architecture of the solution and the iterative bidirectional knowledge transfer procedure between the above mentioned models. Initially, the text style transfer model will be trained on a text style transfer corpus $D_{style}$ and the NMT model will be trained on a machine translation corpus $D_{mt}$. This training will be done to minimize the negative likelihood on their individual training corpora. The NMT model is trained to translate a source language sentence into corresponding target language sentence, and the style transfer model is trained to convert an informal sentence to a formal sentence. Hereafter, we will perform iterative dual knowledge transfer between these models using sequence-level knowledge distillation as proposed in (Kim and Rush, 2016). Through dual knowledge transfer, we aim to improve the performance of both the models by making them act as teacher and student in an iterative manner.

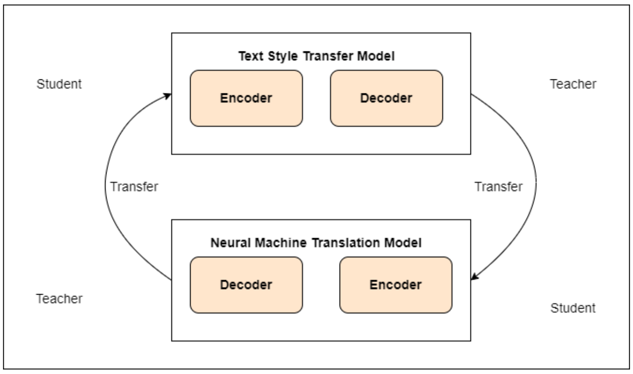


Figure 1: Proposed framework

## 3.1 Baselines

To validate the performance of our model, we would be comparing our proposed architecture with two baseline models for our approach.

### 3.1.1 Dictionary

We aim on using the dictionary generated by (Xu et al., 2012) which maps commonly used Shakespearean style words to Modern English words. For instance: 'eyne' meaning eyes and 'facinorous' meaning wicked. We can perform word by word replacement of the input i.e the Shakespearean text to produce the target output i.e the modern English text. This method, even though has limitations on the quality of the output they produce, they can serve as a good baseline model.

### 3.1.2 Statistical MT model

(Xu et al., 2012) introduced a statistical machine translation model. The model performs word alignment using the GIZA++ (Och and Ney, 2003), extracts phrase pairs and decodes using Moses (Koehn et al., 2007).

The above two baseline models are available at : https://github.com/cocoxu/Shakespeare

## 3.2 Evaluation

### 3.2.1 BLEU score

Bi-Lingual Evaluation Understudy (BLEU) was proposed by (Papineni et al., 2002) which calculates the similarity between the reference translations - the ground truth and reference translations

- the machine translated text. It is a position independent metric which calculates the number of token matches that both the ground truth and target text have. We will use this metric to evaluate the first half of our model, that performs machine translation from Shakespearean English to modern English.

#### 3.2.2 Human Evaluation

The most commonly used metric is human evaluation. The second part of our model that performs style transfer and gives a formal touch to the intermediate Modern English text can be evaluated with the help of our peers.

### 3.3 Schedule

- Week 1: We will spend some time gaining knowledge on Seq to Seq models, LSTMs with attention mechanism.
- Week 2: Every team member will source datasets and setup baseline code. Implement and evaluate baselines, capture the results.
- Week 3: Dipti and Yamini will work on building and training BERT-based Neural machine translation model and capturing the results. Sharanya and Avantika will work on setting up BERT based style transfer model.
- Week 4: Once the results are ready from NMT model, we will work on running the Style transfer model and capture the results.
- Week 5: After the initial phase of our approach, we will learn about bidirectional knowledge transfer using distillation models.
- Week 6: Everyone will work on setting up the code and structure for bidirectional knowledge transfer.
- Week 7: Everyone will analyze the output and debug and fix any errors.
- Week 8: Everyone will work on final report.

## 4 Data

### 4.1 Shakespearean texts scraped data

We will use the dataset curated by (Xu et al., 2012) for training our NMT model. This dataset is created by crawling the nfs.sparknotes.com website and downloading the full texts of Shakespeare's sonnets and plays with parallel modern English translations. More text is craped from www.enotes.com and combined with the sparknotes data after performing appropriate sentence alignment. This preprocessed dataset is available openly at `https://github.com/cocoxu/Shakespeare`.

### 4.2 GYAFC (Grammarly's Yahoo Answers Formality Corpus)

GYAFC (Rao and Tetreault, 2018) is a dataset containing 110K pairs of informal/formal sentence pairs. This dataset has already undergone a few preprocessing steps which include removal of questions, URLs, and too short or too long sentences. We will be using this dataset to train our text style transfer model.

## 5 Tools

We will make use of the PyTorch framework for most of our implementations. The PyTorch-Transformers library consists of pre-trained BERT implementation which we will use for our NMT and style transfer models. We will use the BertTokenizer library for converting our text corpus into tokens suitable for training our models.

We will use the GPU resources available on `https://colab.research.google.com` and `https://www.kaggle.com` for training and evalutaion of our models.

## References


Chen, Y., Liu, Y., Cheng, Y., and Li, V. O. K. (2017). A teacher-student framework for zero-resource neural machine translation.

Cho, K., Van Merriënboer, B., Bahdanau, D., and Bengio, Y. (2014a). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014b). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Cohn, T. and Lapata, M. (2007). Machine translation by triangulation: Making effective use of multi-parallel corpora. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 728–735, Prague, Czech Republic. Association for Computational Linguistics.

Fu, Z., Tan, X., Peng, N., Zhao, D., and Yan, R. (2018). Style transfer in text: Exploration and evaluation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Hu, Z., Lee, R. K.-W., Aggarwal, C. C., and Zhang, A. (2020). Text style transfer: A review and experimental evaluation. *arXiv preprint arXiv:2010.12742*.

Kabbara, J. and Cheung, J. C. K. (2016). Stylistic transfer in natural language generation systems using recurrent neural networks. In *Proceedings of the Workshop on Uphill Battles in Language Processing: Scaling Early Achievements to Robust Methods*, pages 43–47.

Kim, Y. and Rush, A. M. (2016). Sequence-level knowledge distillation. *arXiv preprint arXiv:1606.07947*.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., et al. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions*, pages 177–180.

Och, F. J. and Ney, H. (2003). A systematic comparison of various statistical alignment models. *Computational linguistics*, 29(1):19–51.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Rao, S. and Tetreault, J. (2018). Dear sir or madam, may i introduce the gyafc dataset: Corpus, benchmarks and metrics for formality style transfer. *arXiv preprint arXiv:1803.06535*.

Shen, T., Lei, T., Barzilay, R., and Jaakkola, T. (2017). Style transfer from non-parallel text by cross-alignment. *Advances in neural information processing systems*, 30.

Sudhakar, A., Upadhyay, B., and Maheswaran, A. (2019). Transforming delete, retrieve, generate approach for controlled text style transfer. *arXiv preprint arXiv:1908.09368*.

Wu, X., Liu, J., Li, X., Xu, J., Chen, Y., Zhang, Y., and Huang, H. (2021a). Improving stylized neural machine translation with iterative dual knowledge transfer. In *IJCAI*.

Wu, X., Liu, J., Li, X., Xu, J., Chen, Y., Zhang, Y., and Huang, H. (2021b). Improving stylized neural machine translation with iterative dual knowledge transfer. In *Proceedings of the 30th international joint conference on artificial intelligence (IJCAI), Montreal, QC, Canada*, pages 19–27.

Xu, W., Ritter, A., Dolan, W. B., Grishman, R., and Cherry, C. (2012). Paraphrasing for style. In *Proceedings of COLING 2012*, pages 2899–2914.