Pen to Code: Emulating Shakespeare's Prose through Style Transfer

Easha Meher Koppisetty, Zhifei Yu, Jiayue Zhang

1 Introduction

Translating modern English to Shakespearean English requires balancing semantic accuracy with stylistic fidelity, making this project uniquely fascinating for its blend of linguistic artistry and computational challenges. This study explores NLP models, including T5, Seq2Seq with attention, and Style Transformer, using a curated dataset of paired sentences. Fine-tuning T5 leveraged contextual strength, while Seq2Seq offered a traditional approach. Evaluation with BLEU scores highlighted challenges in preserving meaning while achieving stylistic transformation, providing insights into refining stylistic translation with advanced techniques.

2 Background and Related Work

The task of style transfer has gained significant traction in NLP research, with applications ranging from formal-to-informal language transformation to domain-specific text adaptation. Style transfer methods aim to modify stylistic features of a text while retaining its core meaning. Unlike sentiment or domain adaptation tasks, Shakespearean style transfer involves nuanced lexical, syntactical, and poetic transformations, making it a challenging NLP problem.

2.1 Transformer-based Models

he T5 model, introduced in "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer" (Raffel et al., 2020), provides a robust framework for modern-to-Shakespearean English transformation. Its text-to-text approach enables effective style transfer by fine-tuning on paired datasets, leveraging pre-trained knowledge to reduce data requirements. Related studies, such as Riley et al. (2020), demonstrated T5's adaptability for style transfer using style vectors and few-shot learning, while Xu et al. (2023) employed contrastive learning for state-of-the-art performance in complex style tasks. These insights emphasize T5's potential when combined with strategies like external dictionaries, style conditioning, and advanced pre-training techniques.

2.2 Sequence-to-Sequence with Copy Mechanism

Sequence-to-Sequence models, introduced by Sutskever et al. (2014) and enhanced with attention by Bahdanau et al. (2015), have achieved state-of-the-art results in tasks like translation and summarization. Incorporating a Copy Mechanism (Jhamtani et al., 2021) enhances these models for modern-to-Shakespearean English transformation by allowing direct copying of input words while adjusting stylistic elements. The use of external dictionaries to map modern words to Shakespearean equivalents further refines stylistic accuracy, enabling a robust and authentic style transformation system.

2.3 Style Transformer

Unpaired style transfer techniques, such as the Style Transformer, leverage adversarial learning and Transformer architectures to modify text without requiring parallel corpora. Dai et al. (Dai et al., 2019) introduced a novel approach that eschews explicit disentanglement of content and style, enabling robust stylistic transformations. This method aligns well with Shakespearean style transfer, where obtaining paired training data is often infeasible.

By synthesizing insights from these methods, this project seeks to evaluate their applicability and effectiveness in Shakespearean language transformation.

3 Data

The dataset (Garnavaurha, 2022), sourced from Kaggle and supplemented with modern paraphrases of 16 Shakespearean plays from SparkNotes, forms the basis for this style transfer project. It consists of paired modern English and Shakespearean sentences, divided into training (10,000 pairs), validation (1,000 pairs), and test (1,000 pairs) subsets, with an average sentence length of 15 words and a preprocessing limit of 20 words to ensure stability. A vocabulary of approximately 7,200 tokens was created by filtering words with a frequency threshold of 5 and enriched with pre-trained GloVe embeddings for contextual representation. The dataset's balance of modern and archaic styles, along with its syntactic diversity, makes it ideal for style transfer while presenting unique challenges in handling poetic constructs and syntactic inversions.

4 Methods

4.1 T5

The T5 (Text-to-Text Transfer Transformer) model was implemented for Shakespeare's translation with key features:

4.1.1 Tokenization and Preprocessing

The T5 tokenizer uses subword tokenization for robust sequence handling. Maximum sequence length filtering ensures efficiency during training and inference. Input and target texts are preprocessed to conform to length constraints, optimizing consistency and performance.

4.1.2 Training Configuration

The model was fine-tuned with a learning rate of 2×10^{-5}, batch size of 8, and weight decay of 0.01 for regularization. Training was conducted over one epoch to balance performance and prevent overfitting.

4.1.3 Unknown Word Handling

Subword tokenization effectively handles out-of-vocabulary words by breaking them into smaller, known components, maintaining semantic integrity and robust processing of rare or unseen words.

4.1.4 Performance Evaluation

Performance was evaluated using BLEU scores, measuring n-gram overlap to assess the stylistic and semantic accuracy of the translations.

4.2 Seq2Seq

4.2.1 Problem Setup

The Seq2Seq model translates Shakespearean English to modern English using an encoder-decoder architecture. Input sentences are tokenized with a maximum sequence length of 25 tokens, and the vocabulary size is capped at 12,000 after pruning low-frequency words.

4.2.2 Model Features

Attention Mechanism. The attention mechanism dynamically focuses on relevant input tokens, improving the handling of longer and more complex sentences.

Copy Mechanism. This enables direct copying of input words or phrases, preserving rare or domain-specific terms critical for semantic fidelity.

4.2.3 Unknown Words

Unknown words are replaced with a special `<UNK>` token. The copy mechanism addresses this limitation by directly copying unknown words from the input, enhancing translation quality.

4.2.4 Training Details

The model uses the Adam optimizer with a learning rate of 0.001 and a batch size of 32. Sequence lengths are capped at 25 tokens. Pre-trained embeddings of size 192 are fine-tuned to adapt to the domain. For models with Sentinel Loss, a weight (λ) of 2.0 balances copying and generation.

4.2.5 Training Process

The encoder uses a unidirectional LSTM to encode input into a context vector. Attention-enabled decoders dynamically weigh encoder outputs, and the copy mechanism integrates seamlessly for token copying. Training is performed over 10 epochs, with checkpoints to monitor progress.

4.2.6 Decoding Strategy

Greedy decoding generates outputs one word at a time based on the highest probability, ensuring efficiency. Future improvements could include a beam search for better output diversity.

4.2.7 Integration with Experiments

Experiments (Section 5.1) validate the model configurations, highlighting the benefits of attention and copy mechanisms, and provide insights into performance variations with different LSTM sizes and Sentinel Loss.

4.3 Style Transformer

4.3.1 Problem Setup

Given a modern English sentence x, the task is to generate a Shakespearean sentence y, preserving semantic content while altering the style. The Style Transformer employs a GAN-like framework with an attention-based sequence-to-sequence architecture:

$$y = f_{\text{StyleTransformer}}(x)$$

4.3.2 Training Configuration

Our experiment focused on training and evaluating a Style Transformer model to translate modern English sentences into Shakespearean English using a dataset of paired sentences split into training, validation, and test sets. Preprocessing involved removing sentences exceeding 20 words to meet sequence length constraints, constructing a combined vocabulary to handle linguistic variations, and replacing rare words with <UNK> tokens to manage vocabulary size effectively. The model employed the Adam optimizer with learning rates of 0.001 and 0.0005 for the generator (Model F) and discriminator (Model D), respectively, and a weight decay of 0.01 for regularization. The training incorporated temperature annealing, dropout decay, and loss terms—cycle-consistency, adversarial, and self-reconstruction—balanced at 1.0, 0.1, and 1.0 to ensure effective style transfer and output coherence.

4.3.3 Training Process and Evaluation

The training process begins with pre-training the encoder-decoder model for reconstruction, ensuring the preservation of semantic meaning. Following this, the Style Transformer is fine-tuned using adversarial loss to achieve stylistic consistency. The training involves iterative updates of the discriminator D and the generator, progressively enhancing the quality of the style transfer through adversarial dynamics. The generated output is then evaluated using the BLEU score.

5 Experiments

5.1 T5

5.1.1 Implementation Overview

We implemented the T5 model using the Hugging Face Transformers library, fine-tuning it on a parallel dataset for modern English to Shakespearean English translation. The configurations used were a learning rate of 2×10^{-5}, batch size of 8, and 1 training epoch.

5.1.2 Results Analysis

Fine-tuning T5 for Shakespearean style transfer revealed challenges. The model often retained modern English structures, made superficial modifications, or produced incorrect transformations, failing to capture the Shakespearean style. Examples are shown in Appendix A.1.

5.1.3 Problem Verification

To identify the root cause, we conducted validation experiments:

1. **Dataset Validation:** Issues persisted across different Shakespearean parallel datasets, indicating the problem was not dataset-specific.

2. **Implementation Verification:** Successful performance on a summarization task confirmed our pipeline's correctness.

3. **Model Validation:** Similar translation patterns with GPT-2 suggested a broader challenge with pretrained models for this task.

5.1.4 Analysis and Potential Solutions

The primary issues included data imbalance (T5's pretraining corpus of 750GB versus a 1.81MB fine-tuning dataset) and misalignment between T5's pretraining objectives and our task, which requires substantial surface-level modifications. Potential solutions involve freezing encoder parameters, expanding and augmenting the dataset, and incorporating style-sensitive loss functions with regularization terms or a style classifier. These solutions remain future work due to time and resource constraints. Alternative approaches, such as Seq2Seq models with copy mechanisms and Style Transformer, showed more promise.

5.2 Seq2Seq with Copy Mechanism

5.2.1 Experimental Configurations

We tested several configurations to evaluate the impact of attention, copy mechanisms, and hyperparameter variations. The baseline configuration, referred to as "Simple Seq2Seq," excluded attention and copy mechanisms. An enhanced version incorporated an attention mechanism to enable dynamic focus on input tokens. The "Copy Mechanism" configuration included both attention and the ability to replicate tokens directly from the input sequence, providing a robust solution for handling rare or domain-specific terms. Additionally, we experimented with a variation that added Sentinel Loss (2.0) to balance copying and generation. To assess the influence of model complexity, we tested configurations with different LSTM sizes, specifically 128 and 256 units, to evaluate the trade-offs between model capacity and overfitting.

5.2.2 Results and Analysis

The copy mechanism significantly improved performance by retaining rare terms and structures. Smaller LSTM sizes reduced overfitting, while larger sizes improved BLEU scores. Sentinel Loss negatively impacted results, likely due to over-regularization.

5.2.3 Discussion

The copy mechanism enhances translation quality, especially for domain-specific terms and unknown words. Larger LSTM sizes further improve structure capture. Future work could explore fine-tuning embeddings or alternative loss functions for better handling of rare patterns.

5.3 Style Transformer

Fine-tuning the T5 model for Shakespearean style transfer revealed notable challenges, as shown below::

> **No Semantic Sense:**
> *Input:* “It is the lark that sings so out of tune, harsh and.”
> *Output:* “I love the size of this dance floor!”

> **Partial Style Transfer:**
> *Input:* “Oh, then I see that madmen have no ears.”
> *Output:* “I seeth now that people nay listen to reason.”

These examples reveal that the model frequently failed to fully adapt to the Shakespearean style. It either preserved the modern English structure, applied only superficial changes, or produced outputs that were entirely irrelevant to the intended stylistic transformation.

6 Conclusions

The Seq2Seq model demonstrated the most reliable and coherent results in translating modern English to Shakespearean English, successfully capturing the linguistic and stylistic nuances of Shakespearean text. Despite its theoretical advantages, the Style Transformer struggled to produce outputs that met expectations. Challenges such as unstable adversarial training, limited BLEU score improvements, and inconsistent style transfer were observed.

The T5 (Text-to-Text Transfer Transformer) model, while known for its state-of-the-art capabilities in text generation tasks, failed to achieve significant improvements in this specific task. This could be attributed to insufficient domain-specific fine-tuning or a lack of robust Shakespearean-specific pretraining.

Metrics such as BLEU scores are essential but not always sufficient for evaluating generated outputs' stylistic and semantic fidelity.

7 Future Work

Future work could explore domain-specific pretraining for models like T5 and the Style Transformer using an expanded Shakespearean corpus to better capture stylistic nuances. Expanding the dataset with more Shakespearean works and modern-English equivalents would enhance linguistic representation. Optimizing hyperparameters and stabilization techniques, particularly for the Style Transformer, could improve adversarial training and convergence. Incorporating new evaluation metrics, such as human judgments or style-focused measures, would better assess subjective aspects of style transfer. Hybrid models combining Seq2Seq and transformer architectures, along with advanced inference strategies like constrained beam search, could enhance output quality. Interactive feedback mechanisms for real-time input and extending the approach to other languages or literary styles would demonstrate its versatility.

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations (ICLR)*.

Zihan Dai, Chunting Zhou, Guoyin Wang, Zhe Gan, Ruoming Pang, and Lawrence Carin. 2019. Style transformer: Unpaired text style transfer without disentangled latent representation. Association for Computational Linguistics.

Garnavaurha. 2022. Shakespearify dataset. `https://www.kaggle.com/datasets/garnavaurha/shakespearify`.

Harsh Jhamtani, Varun Gangal, Eduard Hovy, and Eric Nyberg. 2021. Shakespearizing modern language using copy-enriched sequence-to-sequence models. 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.

Parker Riley, Debanjan Ghosh, and Kathleen McKeown. 2020. Textsettr: Few-shot text style extraction and tunable targeted restyling. *arXiv preprint arXiv:2010.03802*.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*, volume 27, pages 3104–3112. Curran Associates, Inc.

Ruiqi Xu, Xindi Peng, Wanlu Zhao, and Xubo Liu. 2023. Specializing small language models towards complex style transfer via latent attribute pre-training. *arXiv preprint arXiv:2309.10929*.

A Additional Results

A.1 Style Transfer Examples

No Style Transfer:
Input: “Yes your highness, I will leave today.”
Output: “Yes your highness, I will leave today.”

Partial Style Transfer:
Input: “All right, my lord.”
Output: “All the right, mine lord.”

Failed Style Transfer:
Input: “Madam, this glove.”
Output: “Madam, this tee.”

A.2 BLEU Scores for Different Configurations

Table 1: BLEU Scores for Different Configurations

Model Configuration	BLEU Score
Simple Seq2Seq	13.23
Attention	13.91
Copy Mechanism (default)	31.62
Copy Mechanism (+SL)	13.16
Copy Mechanism (LSTM size=128)	33.72
Copy Mechanism (LSTM size=256)	34.60

A.3 Analysis of Copy Mechanism (Default) Outputs

Key observations include:

- For the sentence *"I have half a mind to hit you before you speak again."*, the model generated *"I have a unk to hit you before you speak."* While the intent ("to hit you") was preserved, less frequent phrases such as "half a mind" were not accurately generated.

- The input *"He's married to Octavia, madam."* resulted in *"He's married, Octavia."* The omission of "madam" suggests difficulty in capturing peripheral contextual information.

- For the emotionally charged phrase *"May you die of the worst disease!"*, the output was *"The worst lepidus are you!"*, introducing unrelated terms ("lepidus") and highlighting the model's struggles with uncommon input patterns.