# To be a Knight-errant Novel Master: Knight-errant Style Transfer via Contrastive Learning

Anonymous ACL submission

## Abstract

Knight-errant style writing is a challenging task for novice writers due to the highly condensed terminology and highly literary language culture of the knight-errant works. To tackle this problem, in this paper, we propose a new large-scale parallel knight-errant dataset and model the knight-errant writing as a text style transfer (TST) task between modern style and knight-errant style. We establish the benchmark performance of six current SOTA models for knight-errant style transfer. Empirical results demonstrate that the existing SOTA TST models are unable to accurately identify and generate knight-errant style sentences. Therefore, we propose Knight, a TST framework based on contrastive learning. Knight uses multiple strategies to construct positive and negative samples, making it significantly better than existing SOTA models in terms of content fluency, style transfer accuracy, and factuality. The data and code are publicly available [1].

## 1 Introduction

The lack of literary sophistication is a frequently-appeared phenomenon (Bereiter and Scardamalia, 1987; Bryson et al., 1991) in novice writers, leading to their inability to write subtle literary works such as knight-errant novels. Therefore, for many years researchers have been dedicated to building intelligent writing systems (Levinson, 1989; Heidorn, 2000; Jhamtani et al., 2017; Carlson et al., 2018) to assist novice writers in their writing. In recent years, due to the progress in text style transfer (TST) techniques (Hu et al., 2017; Prabhumoye et al., 2018; Li et al., 2022b), some researchers have been addressing the problem via text style transfer (TST) approach (Carlson et al., 2018; Jhamtani et al., 2017; Wang et al., 2020). TST aims to change the style of the input text and keep its content unchanged. However, due to the lack of parallel datasets, most studies (Taele et al., 2020; Chakrabarty et al., 2020) focus on unsupervised TST approaches, which can achieve some results in some simple TST tasks, though the sentences generated fail to reach a satisfactory quality in knight-errant transfer (Section 6.2).

[1]https://anonymous.4open.science/r/knight-errant-style-transfer-C2E1/

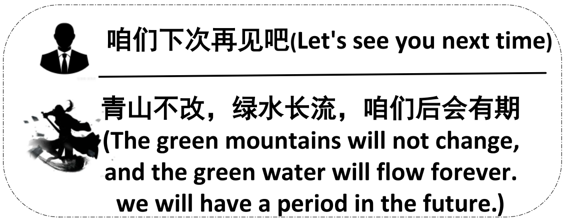



Figure 1: An example of knight-errant style transfer.

In this paper, we choose knight-errant style for which is comparatively difficult for novice writers to follow and model the knight-errant writing as a text style transfer task between modern style and knight-errant style. Specifically, we propose a new subtask of text style transfer named knight-errant style transfer, and supply a large-scale fine-grained parallel knight-errant dataset **KE**. KE is derived from human-written knight-errant novels, and we construct parallel data via back translation and manual annotation. We show an example in Figure 1, where the source sentence is in modern style and the target sentence is in knight-errant style. From Figure 1, we can observe that the knight-errant style utilizes extensive rhetorical techniques such as simile, metaphor, and metonymy to enhance the literary character of the work. This is a very challenging task as it requires the model to be capable of capturing the concept of knight-errant style and generate sentences in the corresponding style without changing the main content.

To establish a comprehensive and reliable benchmark for researchers to evaluate, we employ six state-of-the-art approaches encompassing unsupervised and supervised TST methods as baselines. Empirical results demonstrate that unsupervised methods perform badly on multiple metrics on this task. Some methods with supervision achieve some results. However, since the models trained with the mle (maximum likelihood estimate) method have difficulties in distinguishing different styles at the sentence level (Paulus et al., 2017), the generated results are still unsatisfactory.

To be able to identify different styles at the sentence level, we propose a new TST model named **Knight** based on contrastive learning. Knight requires only simple methods to construct positive and negative samples to improve the performance significantly compared to current SOTA models. In addition, we train the knight model with the prompt method, so that given different prompt prefixs, only one model is enough to generate different knight-errant style texts. This is very cost-effective and prevents us from consuming a lot of resources to train multiple TST models.

Our main contributions can be summarized as:

- We propose a practical task of knight-errant style transfer and a new knight-errant dataset **KE**, which has many potential applications in knight-errant style writing.
- We establish the baseline performance of this task and discuss the key challenges of the task, models.
- We propose a contrastive learning model **Knight** trained with the prompt method, which achieve state-of-the-art performance against multiple strong baselines.

## 2 Related Work

### 2.1 Text Style Transfer

Text style transfer based on deep learning has been extensively studied in recent years, which has achieved encouraging results on styles of expertise (Cao et al., 2020), offensiveness (Santos et al., 2018), sentiment (Fu et al., 2017; Li et al., 2022b), formality (Jain et al., 2019; Liu et al., 2020b), poetry (Shang et al., 2019) and other stylized text generation tasks (Gao et al., 2019; Cao et al., 2020; Syed et al., 2020). However, due to the lack of parallel data, only a few researchers focus on supervised TST methods. Jhamtani et al. (2017) explore neural machine translation (NMT) method to transform text from modern style to Shakespearean style, while a statistical machine learning approach (Carlson et al., 2018) is employed for style transfer using different versions of the Bible as parallel datasets. Rao and Tetreault (2018) use a crowd-sourcing technique to rewrite Yahoo answers to create the GYAFC dataset for TST evaluation.

| Dataset | Number | Task | Number of styles |
|---|---|---|---|
| Yelp | 1000 | Sentiment | 2 |
| Amazon | 1000 | Sentiment | 2 |
| GYAFC | 112594 | Formality | 2 |
| TCFC | 2000 | Formality | 2 |
| MTFC | 4277 | Formality | 2 |
| Bible | 32320918 | Bible | 2 |
| KE(Ours) | 1224065 | Knight-errant | 6 |

Table 1: Comparison between different parallel datasets.

Due to the difficulty of collecting parallel data, most of the existing studies have studied text style transfer with unsupervised methods. A common pattern is to first separate the latent space as content and style representation, then adjust the style-related representation and generate stylistic sentences through the decoder. Hu et al. (2017); Fu et al. (2017); Li et al. (2019) assume that appropriate style regularization can achieve the separation. Style regularization may be implemented as an adversarial discriminator or style classifier in an automatic encoding process. Additionally, another line of work argues that it is unnecessary to disentangle style and content from latent space. Their main approach is to use unsupervised machine translation to construct stylized text based on cyclic reconstruction (Dai et al., 2019; Liu et al., 2020b) and back-translation(Jin et al., 2020).

### 2.2 Dataset

Most of the existing parallel style transfer datasets focus on a coarse-grained style transfer, which generally consist of only two styles. Popular datasets include sentiment modification datasets Yelp, Amazon (He and McAuley, 2016) and IMDB (Li et al., 2019). TCFC (Wu et al., 2020a) and GYAFC (Rao and Tetreault, 2018) focus on formality transfer. In contrast, our dataset contains parallel datasets of six different styles, including four Chinese knight-errant styles and two English literary styles. Moreover, the size of the dataset reaches the level of millions. A comparison with other parallel datasets is shown in Table 1.

### 2.3 Contrastive Learning

Contrastive learning is a popular representation learning method that has been first applied in visual understanding (He et al., 2020; Chen et al., 2020; Hjelm et al., 2018). The core idea is to minimize the distance between the feature representations of different views of the same image (positive example), while maximizing the distance between the feature representations of views of different images (negative example).

In NLP, contrastive learning methods are mainly used in natural language understanding tasks and pre-training tasks. For example, Fang et al. (2020) uses contrastive learning to train self-supervised language models, (Gao et al., 2021; Yan et al., 2021) uses contrastive learning to learn sentence representations, Zhang et al. (2021) improve document clustering performance via contrastive learning. Recently, several studies have applied contrastive learning to text generation tasks (Liu and Liu, 2021; Cao and Wang, 2021), and they all require sophisticated methods for constructing positive and negative samples. Our contrastive learning approach is designed for text style transfer tasks and requires only simple methods for constructing positive and negative samples to significantly improve model performance.

## 3 Dataset Construction Process

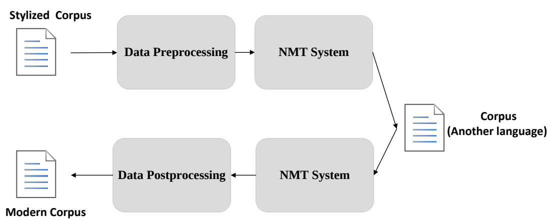



Figure 2: The processes of corpus construction.

| Author | Train | Valid | Test |
|---|---|---|---|
| Gu Long(zh) | 325226 | 92922 | 46460 |
| Jin Yong(zh) | 149390 | 42682 | 21341 |
| Liang Yusheng(zh) | 264655 | 75616 | 37807 |
| Wen Ruian(zh) | 99781 | 28508 | 14254 |
| Shakespeare(en) | 18395 | 1218 | 1462 |
| Le Morte d'Arthur(en) | 3065 | 876 | 437 |

Table 2: Statistics of dataset **KE** (parallel data). The source corpus is modern style, and the target corpus is knight-errant style.

In this section, we describe how to construct our TST dataset **KE** in detail, which has parallel text in modern style and knight-errant style. The goal of the dataset **KE** is to transfer modern style Chinese or English sentences into knight-errant style sentences, and finally can be used to promote the development of knight-errant writing as well as style transfer community.

### 3.1 Chinese Knight-errant Corpus

For the style transfer of Chinese knight-errant text, we select four well-known Chinese knight-errant novel masters[2] who have a high reputation in China. These masters are all famous for their knight-errant novels and have distinctive styles, collectively known in China as the "Four Great Masters of Knight-errant Fiction". We collect their works to build knight-errant style corpus.

First, we collect novels from a knight-errant novel website[3] and cut the text into sentences. To minimize noise, in the preprocessing process, we removed sentences with less than 3 words or more than 128 words, chapter headings and other irrelevant symbols. Finally, we get about one million knight-errant style sentences.

Ideally, the dataset should be constructed by collecting human labeled modern style and knight-errant style parallel data. However, annotating millions of parallel data for training is economically unacceptable. Therefore, we propose a back-translation based approach to get modern style sentences shown in Figure 2. Specifically, applying the NMT (Neural Machine Translation) system, we translate knight-errat style sentences from Chinese to English and then translate them from English back to Chinese. However, the NMT system cannot well translate some knight-errant domain specific vocabulary such as "Five Poisonous Sects(五毒教)" and "Eighteen Ways of Beating the Dragon"(降龙十八掌)", so we manually constructed a domain-specific vocabulary in Chinese, ensuring that they do not change before and after translation.

### 3.2 English Knight-errant Corpus

For English knight-errant style data, we choose the famous knight-errant novel "Le Morte d'Arthur"[4] as the English dataset, and translate English to Chinese and back to English with the help of the NMT

[2] 金庸(Jin Yong), 古龙(Gu Long), 温瑞安(Wen Ruian), 梁羽生(Liang Yusheng)

[3] http://www.wuxia.net.cn/

[4] Le Morte d'Arthur is a 15th-century Middle English knight-errant prose reworking by Sir Thomas Malory of tales about the legendary King Arthur.

| Dataset | Source Sentence | Target Sentence |
|---|---|---|
| Jin Yong | 他们都惊呆了。<br>(They were stunned.) | 这一下变起俄顷，众人都吓得呆了。<br>(Everyone was shocked by the sudden change.) |
| Gu Long | 突然，他说：“是谁？”<br>(Suddenly, he said, " who? ") | 语声突顿，大喝一声：“是谁？”<br>(He suddenly stopped and shouted, "who is it?") |
| Liang Yusheng | 突然，剑从鞘里出来，变成了一道银色的彩虹。<br>(Suddenly, the sword was pulled out,<br>and its appearance was like a rainbow.) | 倏地宝剑出鞘，化作一道银虹。<br>(Suddenly the sword was unsheathed<br>and turned into a silver rainbow.) |
| Wen Ruian | 这是一个遗憾。<br>(This is a pity.) | 这就令人惋惜莫已了。<br>(This is lamentable already.) |
| Le Morte d'Arthur | I will put up with you | I shall abide you, |
| Shakespeare | You are an honest man . | Th' art an honest man . |

Table 3: Examples in the **KE** dataset. Our dataset provides both modern style and knight-errant style sentences. The sentences in the brackets are the translation of the corresponding sentence.

system, through which we get the parallel data pairs of English modern style and English knight-errant style. Moreover, to expand the number of English datasets, we additionally add the Shakespeare dataset (Jhamtani et al., 2017) due to the knight-errant style that pervades much of Shakespeare's work (Rose, 1985).

### 3.3 Quality of Corpus

To ensure data quality, following previous works (Li et al., 2022a; Maynez et al., 2020), we use the NLI (Nature Language Inference) score to detect the modern style sentences, and only sentences with content relevance above 90% will be retained. Furthermore, following (Wu et al., 2020b), the modern style classifier is employed to select the modern style sentences with a high confidence. All modern style sentences in the dataset KE have more than 95% probability of being predicted as modern style by the classifier. Finally, for each writer's corpus, we divided the training, validation, and test sets according to a ratio of 7:2:1. The statistical information of all datasets is shown in Table 2, some examples are shown in Table 3.

## 4 Contrastive Learning Methodology

In this section, we describe our contrastive learning model **Knight** for text style transfer. We first describe our problem definition, then we introduce contrastive learning framework in detail.

### 4.1 Task Definition

Existing supervised TST models (Jhamtani et al., 2017; Carlson et al., 2018) mostly follow the sequence-to- sequence (seq2seq) framework. Given a set of style-labelled sentences $\mathcal{D} = \{(X_i, S_i)\}_{i=1}^{M}$, where $M$ is the total number of sentences. $X_i$ denotes the $i^{th}$ source sentence, and $S_i$ denotes the corresponding style label, which belongs to a source style label set: $S_i \in S_M$ (e.g., modern/kinght-errant). The goal of TST is to transfer sentence $X_i$ with style $S_i$ to a sentence $Y_i$ sharing the same content while having a different style $\tilde{S}_i$. We employ transformer architecture (Vaswani et al., 2017), which is composed of an encoder Transformer $Enc(X; \theta_E)$ and a decoder Transformer $Dec(H; \theta_D)$. Specifically, the encoder Transformer maps sentence $X$ into a sequence of hidden states $\mathbf{E} = (e_0, e_1, ..., e_{|X|})$.

$$\mathbf{E} = Trans^{Enc}(X), \tag{1}$$

The decoder Transformer computes the current hidden state $o_t$ by self-attention to the encoder hidden states $E$ and proceeding tokens $y_{0:t-1}$ .

$$o_t = Trans^{Dec}(y_{0:t-1}, \mathbf{E}), \tag{2}$$

Note that during training, we can obtain $\mathbf{O} = (o_1, ..., o_{|Y|})$ in parallel.

$$\mathbf{O} = Trans^{Dec}(Y, \mathbf{E}), \tag{3}$$

The probability of $y_t$ can be estimated using a linear projection and a softmax function:

$$p(y_t|y_{0:t-1}, X) = softmax(W^o o_t), \tag{4}$$

The loss function of the sequence-to-sequence model minimizes the negative log-likelihood of the training data:

$$\mathcal{L}_{NLL} = -\frac{1}{|Y|}\sum_{t=1}^{|Y|} logP(y_t|y_{0:t-1}, X). \tag{5}$$

### 4.2 Knight: Knight-errant style transfer with Contrastive Learning

Previous work on text style transfer has mostly focused on coarse-grained style transfer, such as sentiment polarity conversion (Hu et al., 2017; Dai et al., 2019; Li et al., 2022b; Rao and Tetreault, 2018) and text formality conversion (Wu et al., 2020a). In this task, we propose a fine-grained dataset, for example, the works of both *Jin Yong* and *Gu Long* belong to the knight-errant style, but *Jin Yong*'s works are more mature and stable, while Gu Long's works are more indolent and unrestrained. It is difficult for mle-trained models to distinguish between these two different knight-errant styles and generate corresponding fine-grained style sentences. However, using contrastive learning, we can pull the distance of these two different styles in the semantic space, which assists the model to discriminate different styles precisely.

Therefore, we designed a contrastive learning based training target, which drives the TST model to learn preferences for fine-grained knight-errant style sentences. Specifically, let a modern style text $X$ have a set of positive knight-errant samples $P$ and another set of negative knight-errant negative samples $N$. To get the sentence representation for similarity computation, we add a multi-layer perceptron (MLP) to the decoder's last layer. The sentences representation and contrastive learning objective is:

$$h = MLP(Trans^{Dec}(Y, \mathbf{E})) \tag{6}$$

$$\mathcal{L}_{CL} = -\frac{1}{|P|} \sum_{\substack{y_i, y_j \in P \\ y_i \neq y_j}} log \frac{exp(sim(h_i, h_j)/\tau)}{\sum_{y_k \in P \cup N} exp(sim(h_i, h_k)/\tau)} \tag{7}$$

where $h_i$, $h_j$ are the representations of generated sentences, positive samples $P$. $h_k$ are the representations of union set of $P$ and $N$. $sim(\cdot, \cdot)$ calculates the cosine similarity between sentence representations. $\tau$ is a temperature and is set to 1.0. Moreover, positive samples $P$ and negative samples $N$ are included in the same batch of training, so the model obtains a better representation of distinguishing correct reference from error by comparing the two types of samples, thus maximizing the probability of positive samples and minimizing the likelihood of corresponding negative samples.

#### 4.2.1 Negative Sample Construction

Here we describe three strategies for constructing negative samples $N$ that modify the references.

**Other Authors' Works (OAW)** For Jin Yong's works, Gu Long's works are naturally a kind of negative sample. Therefore, during the style transfer of Jin Yong's works, we treat other authors' works as negative samples as a contrastive example, so that the model can identify what kind of sentences conform to Jin Yong's style during the training process, thus generating sentences that conform to Jin Yong's style.

To improve the ability of the model to retain the correct textual content while generating the corresponding styles, we next propose two methods for constructing negative samples of content.

**Ramdom Mask and Fill (RMF)** Content consistency is one of the main challenges (Dai et al., 2019; Li et al., 2022b; Cao et al., 2020) of text style transfer task. We use the ability of the language model to insert erroneous information in the correct human reference. Specifically, we use the [MASK] token to randomly replace one or several word tokens in the sentence, and language model Bert (Devlin et al., 2018) is used to predict the [MASK] token. Notably, we choose the set of tokens with the lowest prediction probability for filling to simulate extrinsic content errors. Note that bert model is not fine-tuned on the knight-errant dataset, and thus tokens it predicts will also introduce style errors .

**Low Confidence Generation (LCG)** While the previous approach constructs negative samples at the token level, and here we propose a way to construct negative samples at the sentence level. We fine-tune the Bart model (Lewis et al., 2019) on knight-errant style transfer task so that the fine-tuned Bart can generate knight-errant style sentences. For each generated sentence, we check the model confidence on the tokens of each proper noun by considering all beams at the last decoding step as candidates with beam sizes of 5. If the probability is below the threshold we set, we keep it as a negative sample for the sentence with low confidence do not align with the target style.

#### 4.2.2 Positive Sample Construction

Following (Cao and Wang, 2021; Xu et al., 2021), we use human reference as a natural positive example. In order to create multiple positive samples, we use sentences generated by fine-tune 10000 steps of Bart on the training set as our positive samples.

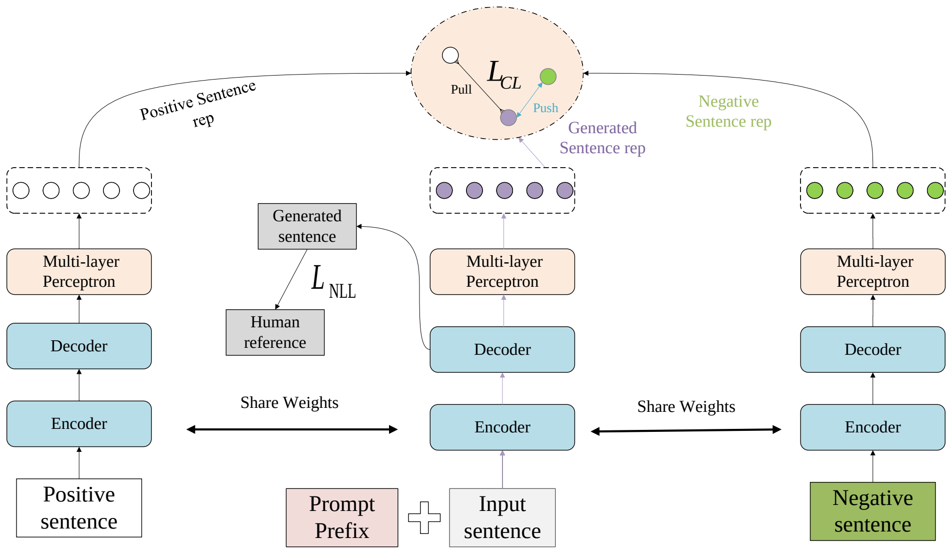



Figure 3: General overview of our contrastive text style transfer model **Knight**.

#### 4.2.3 Training objective

Combining the negative log-likelihood loss $\mathcal{L}_{NLL}$ and our contrastive learning loss $\mathcal{L}_{CL}$, the the final loss function is formulated as: $\mathcal{L} = \mathcal{L}_{NLL} + \lambda\ \mathcal{L}_{CL}$, where $\lambda$ is a hyper-parameter. Moreover, inspired by the application of the prompt method (Liu et al., 2021), for each dataset we add a different prompt prefix, so that a single model can generate a different knight-errant style.

## 5 Experiments

We re-implemented the six SOTA models from previous TST studies on the KE dataset. Further ablation study is conducted to give a detailed analysis of the knowledge and structure implications.

### 5.1 Baselines

We choose the following SOTA method to compare with our model and establish the benchmark performance of knight-errant style transfer on the dataset. For fairness, we classify the compared models into two classes. (A) Supervised Models. (B) Unsupervised Models.

The unsupervised models selected are: (1) **ControlGen** (Hu et al., 2017) utilizes VAE model to learn content representations and reconstructs style vectors by adversarial training. (2) **FGIM** (Wang et al., 2019) uses the method of editing latent representations to control the direction of style generation. (3) **Style Transformer** (Dai et al., 2019) that uses cyclic reconstruction to learn content and style vectors without parallel data.

The supervised models selected are: (1) **Moses** (Koehn et al., 2007) is a statistical machine translation system. (2) **OpenNMT** (Klein et al., 2017) is an open-source neural machine translation framework, which is widely used in text generation tasks (Jhamtani et al., 2017). (3) **Bart** (Lewis et al., 2019) is a SOTA pre-trained generative language model proposed by FaceBook. We choose multilingual bart (Liu et al., 2020a) for training.

### 5.2 Implementation Details

Our contrastive learning model is initialized from BART (Liu et al., 2020a) privided by Huggingface (Wolf et al., 2020). Specifically, the encoder and decoder are all 12-layer transformers with 16 attention heads, hidden size is 1,024 and feed-forward dim is 4,096, which amounts to 406M trainable parameters. We train our framework using the Adam optimizer (Kingma and Ba, 2017) with the initial learning rate 1e-5, and we employ a linear schedule for the learning rate. Drop is set to 0.1. All models are trained on 8 RTX 3090 GPUs, the number of training steps is 50,000 for Chinese and 10,000 for English. We run each model five times to average the scores.

### 5.3 Evaluation Metrics

Following (Li et al., 2022b, 2019; Fu et al., 2017), we make an automatic evaluation on five aspects:

**Content Retention (BLEU (Average BLEU)** and **Rouge(Rouge-L)**) verifies whether the generated sentences retain the original content (Papineni et al., 2002; Lin, 2004).

| | Jin Yong | | | | | | Gu Long | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | S-Acc | BLEU | Rouge | PPL↓ | NLI | Human | S-Acc | BLEU | Rouge | PPL↓ | NLI | Human |
| ControlGen | 61.38 | 2.42 | 17.32 | 96.49 | 15.73 | 5.31 | 76.91 | 2.21 | 14.22 | 97.55 | 19.21 | 6.80 |
| FGIM | 63.18 | 3.72 | 31.86 | 88.88 | 22.25 | 10.28 | 64.25 | 5.52 | 34.40 | 87.33 | 21.59 | 11.54 |
| Style Transformer | 61.47 | 4.32 | 45.27 | 73.95 | 71.54 | 20.45 | 75.12 | 3.13 | 41.09 | 88.06 | 30.75 | 20.08 |
| Moses | 78.23 | 20.85 | 52.35 | 49.23 | 84.04 | 58.45 | 85.40 | 27.75 | 62.35 | 26.56 | 86.01 | 60.42 |
| OpenNMT | 80.02 | 21.85 | 53.85 | 46.21 | 86.05 | 55.94 | 88.32 | 30.50 | 65.52 | 23.95 | 89.43 | 62.78 |
| Bart | 89.73 | 21.86 | 58.85 | 26.21 | 90.23 | 74.32 | 92.25 | 32.86 | 68.45 | 20.45 | 91.25 | 73.94 |
| Knight(ours) | **94.74** | **23.36** | **61.42** | **19.23** | **93.70** | **79.64** | **94.54** | **35.45** | **69.75** | **18.45** | **93.87** | **79.61** |
| | Wen Ruian | | | | | | Liang Yusheng | | | | | |
| Model | S-Acc | BLEU | Rouge | PPL↓ | NLI | Human | S-Acc | BLEU | Rouge | PPL↓ | NLI | Human |
| ControlGen | 10.92 | 1.24 | 4.23 | 98.55 | 36.74 | 6.34 | 82.45 | 5.65 | 12.51 | 91.55 | 16.40 | 8.42 |
| FGIM | 87.26 | 2.65 | 10.84 | 84.11 | 19.03 | 15.45 | 78.74 | 5.38 | 35.69 | 84.11 | 31.95 | 12.32 |
| Style Transformer | 58.33 | 7.25 | 17.32 | 76.42 | 59.68 | 18.62 | 76.38 | 8.33 | 39.89 | 72.48 | 65.15 | 16.55 |
| Moses | 85.25 | 26.44 | 58.16 | 20.45 | 85.57 | 55.64 | 81.24 | 17.65 | 55.32 | 39.88 | 84.44 | 60.54 |
| OpenNMT | 85.56 | 27.86 | 60.18 | 20.92 | 87.70 | 57.68 | 84.71 | 19.20 | 57.45 | 36.84 | 87.43 | 61.44 |
| Bart | 91.45 | 29.53 | 64.73 | 15.03 | 91.47 | 70.40 | 89.23 | 20.49 | 61.44 | 34.98 | 90.85 | 72.24 |
| Knight(ours) | **94.05** | **32.09** | **66.30** | **11.80** | **93.66** | **78.60** | **92.15** | **22.74** | **63.04** | **27.87** | **91.33** | **80.23** |
| | Le Morte d'Arthur | | | | | | Shakespeare | | | | | |
| Model | S-Acc | BLEU | Rouge | PPL↓ | NLI | Human | S-Acc | BLEU | Rouge | PPL↓ | NLI | Human |
| ControlGen | 83.86 | 11.63 | 19.25 | 94.35 | 23.04 | 4.64 | 66.56 | 3.37 | 4.69 | 98.46 | 31.47 | 5.76 |
| FGIM | 11.41 | 23.61 | 23.34 | 87.09 | 6.24 | 14.34 | 9.65 | 10.36 | 9.39 | 99.17 | 3.53 | 16.58 |
| Style Transformer | 57.76 | 18.11 | 28.24 | 62.33 | 49.95 | 38.46 | 52.62 | 28.98 | 47.16 | 80.36 | 60.35 | 48.59 |
| Moses | 88.45 | 50.50 | 50.01 | 20.76 | 63.67 | 55.40 | 82.30 | 41.94 | 37.88 | 31.32 | 67.44 | 62.48 |
| OpenNMT | 86.44 | 51.37 | 52.05 | 30.28 | 62.14 | 55.64 | 81.29 | 42.74 | 40.81 | 30.50 | 66.15 | 58.54 |
| Bart | 90.77 | 55.95 | 56.10 | 25.56 | 67.57 | 72.54 | 85.22 | 44.22 | 42.02 | 27.45 | 69.45 | 62.45 |
| Knight(ours) | **93.44** | **57.56** | **58.49** | **23.66** | **72.44** | **78.24** | **90.90** | **46.75** | **45.30** | **24.30** | **73.11** | **76.44** |

Table 4: Benchmark and evaluation results for dataset **KE**. ↓ means the smaller the better. We bold the best results.

**Style Control** (S-Acc) measures the style accuracy of the transferred sentences. We train a classifier on the training set of each dataset using XLM-Roberta (Conneau et al., 2019).

**Fluency** (PPL) is usually measured by the perplexity of the transferred sentence. To get the ppl score, we fine-tune GPT-2 (Radford et al., 2019) on the training set for each style.

**Factuality** (NLI Score) is applied to determine the factual consistency of two sentences and is widely employed in text generation tasks (Li et al., 2022a; Maynez et al., 2020).

**Human Evaluation** Following (Madotto et al., 2019; Li et al., 2022b), We randomly sampled 50 sentences generated on the target style and distributed a questionnaire at Amazon Mechanical Turk asking each worker to rank the content retention (0 to 5), style transfer(0 to 5 ) and fluency(0 to 5): human score = Average($\sum score_{sty}$+$\sum score_{con}$+$\sum score_{flu}$), human score $\in$ [0,100]. Three workers are recruited for human evaluation.

# 6 Results and Analysis

## 6.1 Result of Model Performance

Table 4 shows the performance of the different models on our proposed dataset. From this table, we obtain the following observations: (1) The unsupervised methods perform pretty badly on our dataset, yet which achieve good performance on tasks such as sentiment polarity conversion, formality conversion, etc in unsupervised setting (Dai et al., 2019; Li et al., 2019, 2022b). This indicates that our proposed task is so challenging that good performance is not achievable using unsupervised methods. (2) Supervised models outperform unsupervised methods in terms of content retention (BLEU, Rouge), style transfer strength (S-Acc), faithfulness (NLI-S), and fluency (PPL) due to the additional supervision information. The above phenomenon shows that in the application of TST in industry, a supervised method should be preferred. (3) Our proposed contrastive learning model **Knight** significantly outperforms all SOTA models in several automatic and manual metrics, and especially in both faithfulness and style accuracy , demonstrating the remarkable effect of our proposed contrastive learning strategy.

## 6.2 Case Study

Two examples of transferred sentences in Chinese and English are given in Table 5 . From which, it is intuitively clear that ControlGen and FGIM almost destroy the semantic content of the sentence , introducing grammar and factual errors. Although

Style Transformer preserves part of the semantics of the sentence, it still does not generate sentences with correct style, which is the reason that it has a higher BLEU and Rouge Score but fails in NLI-S. In contrast, the supervised approach performs well in terms of sentence content retention, and factual correctness, which confirms that the introduction of supervised signals is significantly effective for complicated TST tasks. However, while supervised models such as OpenNMT can generate sentences that are verbally fluent and free of factual errors, the fact that mle training is based on individual words makes it impossible to distinguish between different styles at the sentence level.

As a contrast, due to the application of contrastive learning, Knight model can distinguish between different styles of representations in the latent space. Therefore, Knight model is far more stylistically accurate than other models, which makes it generate knight-errant styles precisely.

In addition, as seen in the Table 5, using different prompt prefixes, Knight model can generate different fine-grained styles of text, which indicates that Knight is capable of clearly identifying each different style of text by contrastive learning. And from the generated results, we can see the subtle differences between the different styles. For example, Shakespeare likes to employ ***thee*** instead of ***you***, while the Arthur style prefers ***thou***.

## 6.3 Ablation Study

To investigate the effect of different components on the overall performance, we further perform an ablation study on our model and the results are shown in Table 6. From which, we obtain the following observations: (1) Each positive and negative example plays a facilitating role in the model. (2) Using the OAW method maximizes style accuracy, indicating differences between different author styles. LCG and RMF improve BLEU and Rouge score, suggesting that introducing content negative samples improves the model's content retention ability. (3)Negative examples bring about a significant improvement over positive examples. We speculate that this is due to the positive sample is more similar to the human reference and the model can easily distinguish them.

# 7 Conclusion

In this paper, we propose a new challenging parallel knight-errant dataset. Moreover, we establish the benchmark performance of six current SOTA models, and we build a TST model based on contrastive learning for distinguishing knight-errant styles precisely. We believe this work has many promising applications for the knight-errant writing industry. In the future, we are interested in applying contrastive learning to unsupervised models to solve similar TST problems and we will try to apply our model to practical industry.

| | | Kneght-errant(zh) |
|---|---|---|
| | Source | 所有人都开心的欢呼<br>(Everyone cheered happily.) |
| U | ControlGen | 所有人都跑了<br>(Everyone ran away.) |
| | FGIM | 都到这些，所有的人都很着急<br>(All to these, all the people are very anxious.) |
| | Style Trans | 所有人听了，都说：“大叫起来”<br>(Hearing this, all the heroes said, "shout.") 、 |
| S | OpenNMT | 所有人都在开心的欢呼<br>(Everyone cheered happily.) |
| | Knight(Jin ) | 群雄一听，尽皆喝彩<br>(Hearing this,all the heroes applauded.) |
| | Knight(Gu ) | 听到这些，群豪都欢呼<br>(Hearing this, the group of heroes all cheered.) |
| | Knight(Liang ) | 众英雄听了，齐声喝彩<br>(Hearing this, the heroes applauded in unison) |
| | Knight(Wen) | 群雄听了，都是欢呼<br>(Hearing this, all the heroes cheered.) |
| | Human | 听到这里，群豪齐声喝彩<br>(Hearing this,<br>the group of heroes applauded in unison.) |
| | | **Knight-errant(en)** |
| | Source | I can tell that you don't know who I am. |
| U | ControlGen | you my swear please please am my you. |
| | FGIM | i see how long you two sons are. |
| | Style Trans | I can tell that you don't know who I am . |
| S | OpenNMT | I know you know me not. |
| | Knight(Shakes) | I can tell you that I am unknown unto thee. |
| | Knight(Arthur) | I can tell you that thou know'st me not. |
| | Human | I see thou know'st me not . |

Table 5: Examples of model outputs, where red denotes successful style transfers, blue denotes content errors, and green denotes grammar errors, better looked in color. For the Knight model, we show the results generated by different prompt prefixes. U, S refer to unsupervised and supervised models. More examples are in the appendix.

| Stragey | Jin Yong | | | Shakespear | | |
|---|---|---|---|---|---|---|
| | S-ACC | BLEU | NLI | S-Acc | BLEU | NLI |
| Bart | 89.73 | 21.86 | 90.23 | 85.22 | 44.22 | 69.45 |
| +OAW | **92.96** | 22.06 | 90.24 | **89.45** | 44.32 | 69.55 |
| +RMF | 89.86 | 24.56 | **93.81** | 85.59 | **46.64** | **72.70** |
| +LCG | 90.23 | **24.73** | 92.11 | 85.50 | 46.43 | 71.81 |
| +Pos | 89.69 | 21.81 | 90.62 | 85.32 | 44.20 | 69.81 |

Table 6: Model ablation study results on Jin Yong and Shakespear dataset. We bold the best results.

## References


Carl Bereiter and Marlene Scardamalia. 1987. An attainable version of high literacy: Approaches to teaching higher-order skills in reading and writing. Curriculum inquiry, 17(1):9–30.

Mary Bryson, Carl Bereiter, Marlene Scardamalia, and Elana Joram. 1991. Going beyond the problem as given: Problem solving in expert and novice writers. Complex problem solving: Principles and mechanisms, 61:84.

Shuyang Cao and Lu Wang. 2021. Cliff: Contrastive learning for improving faithfulness and factuality in abstractive summarization. arXiv preprint arXiv:2109.09209.

Yixin Cao, Ruihao Shui, Liangming Pan, Min-Yen Kan, Zhiyuan Liu, and Tat-Seng Chua. 2020. Expertise style transfer: A new task towards better communication between experts and laymen. arXiv preprint arXiv:2005.00701.

Keith Carlson, Allen Riddell, and Daniel Rockmore. 2018. Evaluating prose style transfer with the bible. Royal Society open science, 5(10):171920.

Tuhin Chakrabarty, Smaranda Muresan, and Nanyun Peng. 2020. Generating similes effortlessly like a pro: A style transfer approach for simile generation. arXiv preprint arXiv:2009.08942.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In International conference on machine learning, pages 1597–1607. PMLR.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116.

Ning Dai, Jianze Liang, Xipeng Qiu, and Xuanjing Huang. 2019. Style transformer: Unpaired text style transfer without disentangled latent representation. arXiv preprint arXiv:1905.05621.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Hongchao Fang, Sicheng Wang, Meng Zhou, Jiayuan Ding, and Pengtao Xie. 2020. Cert: Contrastive self-supervised learning for language understanding. arXiv preprint arXiv:2005.12766.

Zhenxin Fu, Xiaoye Tan, Nanyun Peng, Dongyan Zhao, and Rui Yan. 2017. Style transfer in text: Exploration and evaluation.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. Simcse: Simple contrastive learning of sentence embeddings. arXiv preprint arXiv:2104.08821.

Xiang Gao, Yizhe Zhang, Sungjin Lee, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2019. Structuring latent spaces for stylized response generation. arXiv preprint arXiv:1909.05361.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 9729–9738.

Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In proceedings of the 25th international conference on world wide web, pages 507–517.

George Heidorn. 2000. Intelligent writing assistance. Handbook of natural language processing, pages 181–207.

R Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio. 2018. Learning deep representations by mutual information estimation and maximization. arXiv preprint arXiv:1808.06670.

Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P. Xing. 2017. Toward controlled generation of text.

Parag Jain, Abhijit Mishra, Amar Prakash Azad, and Karthik Sankaranarayanan. 2019. Unsupervised controllable text formalization. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 6554–6561.

Harsh Jhamtani, Varun Gangal, Eduard Hovy, and Eric Nyberg. 2017. Shakespearizing modern language using copy-enriched sequence-to-sequence models. arXiv preprint arXiv:1707.01161.

Di Jin, Zhijing Jin, Joey Tianyi Zhou, Lisa Orii, and Peter Szolovits. 2020. Hooks in the headline: Learning to generate headlines with controlled styles. arXiv preprint arXiv:2004.01980.

Diederik P. Kingma and Jimmy Ba. 2017. Adam: A method for stochastic optimization.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M Rush. 2017. Opennmt: Open-source toolkit for neural machine translation. arXiv preprint arXiv:1701.02810.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, 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.

Paul Levinson. 1989. Intelligent writing: The electronic liberation of text. Technology in Society, 11(4):387–400.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461.

Dianqi Li, Yizhe Zhang, Zhe Gan, Yu Cheng, Chris Brockett, Ming-Ting Sun, and Bill Dolan. 2019. Domain adaptive text style transfer. arXiv preprint arXiv:1908.09395.

Wei Li, Wenhao Wu, Moye Chen, Jiachen Liu, Xinyan Xiao, and Hua Wu. 2022a. Faithfulness in natural language generation: A systematic survey of analysis, evaluation and optimization methods. arXiv preprint arXiv:2203.05227.

Xiangyang Li, Xiang Long, Yu Xia, and Sujian Li. 2022b. Low resource style transfer via domain adaptive meta learning.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In Text summarization branches out, pages 74–81.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. arXiv preprint arXiv:2107.13586.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020a. Multilingual denoising pre-training for neural machine translation. Transactions of the Association for Computational Linguistics, 8:726–742.

Yixin Liu and Pengfei Liu. 2021. Simcls: A simple framework for contrastive learning of abstractive summarization. arXiv preprint arXiv:2106.01890.

Yixin Liu, Graham Neubig, and John Wieting. 2020b. On learning text style transfer with direct rewards. arXiv preprint arXiv:2010.12771.

Andrea Madotto, Zhaojiang Lin, Chien-Sheng Wu, and Pascale Fung. 2019. Personalizing dialogue agents via meta-learning. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 5454–5459.

Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On faithfulness and factuality in abstractive summarization. arXiv preprint arXiv:2005.00661.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 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.

Romain Paulus, Caiming Xiong, and Richard Socher. 2017. A deep reinforced model for abstractive summarization. arXiv preprint arXiv:1705.04304.

Shrimai Prabhumoye, Yulia Tsvetkov, Ruslan Salakhutdinov, and Alan W Black. 2018. Style transfer through back-translation. arXiv preprint arXiv:1804.09000.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. OpenAI blog, 1(8):9.

Sudha Rao and Joel Tetreault. 2018. Dear sir or madam, may I introduce the GYAFC dataset: Corpus, benchmarks and metrics for formality style transfer. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pages 129–140, New Orleans, Louisiana. Association for Computational Linguistics.

Mark Rose. 1985. Othello's occupation: Shakespeare and the romance of chivalry. English Literary Renaissance, 15(3):293–311.

Cicero Nogueira dos Santos, Igor Melnyk, and Inkit Padhi. 2018. Fighting offensive language on social media with unsupervised text style transfer. arXiv preprint arXiv:1805.07685.

Mingyue Shang, Piji Li, Zhenxin Fu, Lidong Bing, Dongyan Zhao, Shuming Shi, and Rui Yan. 2019. Semi-supervised text style transfer: Cross projection in latent space. arXiv preprint arXiv:1909.11493.

Bakhtiyar Syed, Gaurav Verma, Balaji Vasan Srinivasan, Anandhavelu Natarajan, and Vasudeva Varma. 2020. Adapting language models for non-parallel author-stylized rewriting. In AAAI, pages 9008–9015.

Paul Taele, Jung In Koh, and Tracy Hammond. 2020. Kanji workbook: A writing-based intelligent tutoring system for learning proper japanese kanji writing technique with instructor-emulated assessment. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 13382–13389.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Advances in neural information processing systems, pages 5998–6008.

Ke Wang, Hang Hua, and Xiaojun Wan. 2019. Controllable unsupervised text attribute transfer via editing entangled latent representation.

Yunli Wang, Yu Wu, Lili Mou, Zhoujun Li, and Wenhan Chao. 2020. Formality style transfer with shared latent space. In Proceedings of the 28th International Conference on Computational Linguistics, pages 2236–2249.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pages 38–45, Online. Association for Computational Linguistics.

Yu Wu, Yunli Wang, and Shujie Liu. 2020a. A dataset for low-resource stylized sequence-to-sequence generation. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 9290–9297.

Yu Wu, Yunli Wang, and Shujie Liu. 2020b. A dataset for low-resource stylized sequence-to-sequence generation. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 9290–9297.

Shusheng Xu, Xingxing Zhang, Yi Wu, and Furu Wei. 2021. Sequence level contrastive learning for text summarization. arXiv preprint arXiv:2109.03481.

Yuanmeng Yan, Rumei Li, Sirui Wang, Fuzheng Zhang, Wei Wu, and Weiran Xu. 2021. Consert: A contrastive framework for self-supervised sentence representation transfer. arXiv preprint arXiv:2105.11741.

Dejiao Zhang, Feng Nan, Xiaokai Wei, Shangwen Li, Henghui Zhu, Kathleen McKeown, Ramesh Nallapati, Andrew Arnold, and Bing Xiang. 2021. Supporting clustering with contrastive learning. arXiv preprint arXiv:2103.12953.

# A Appendix

## A.1 More Implementation Experiment Details

For ControlGen, we use a reference implementation in Texar-tf v0.2.4[5], which uses an undirectional GRU encoder and an attention GRU decoder. The train setting of ControlGen is 10 reconstruction epochs and 2 transfer epochs. For FGIM, we use the author's published repo[6], which uses a 2-layer transformer encoder and a 2-layer transformer decoder. The train setting of FGIM is 200 train epochs (reconstruction and transfer in the same step). For style-transformer, we use the author's published repo[7] in fastnlp, which is combined with a 4-layer transformer encoder and a 4-layer transformer decoder. The train setting of the style transformer is 500 pretrain steps, and 400 train steps (5 style transformer updates and 10 discriminator updates in the same step). For Moses, we use a phrase-based statistical translation model. For OpenNMT, we use a three-layer bidirectional LSTM structure, and the number of training steps is set to 6000. For Bart, we use the same model structure as Knight, and the rest of experimental setup is the same as Knight.

## A.2 More Generation Examples

To demonstrate more examples of generation to verify the effectiveness of the model, we selected 5 generated sentences from KE dataset, as shown in Table 7 and Table 8.

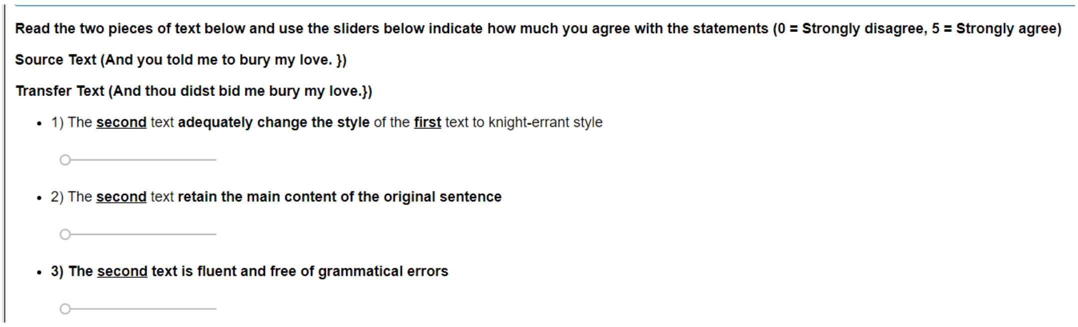
| | Knight-errant(English) |
|---|---|
| Input | And now you've changed ? |
| Knight-errant | And art thou changed? |
| Human Ref. | And art thou changed ? |
| Input | And you told me to bury my love . |
| Knight-errant | And thou didst bid me bury my love. |
| Human Ref. | And badest me bury love. |
| Input | I'm in a rush. |
| Knight-errant | I am in haste. |
| Human Ref. | I stand on sudden haste. |
| Input | Please , I beg you , don't scold me . |
| Knight-errant | I prithee, do not scold me. |
| Human Ref. | I pray thee , chide not . |
| Input | I'm a man, my lady,<br>that God has made for himself to ruin. |
| Knight-errant | I am a man, lady,<br>that God hath made himself to fall. |
| Human Ref. | One, gentlewoman,<br>that God hath made, himself to mar. |

Table 7: English knight-errant style generation results on KE dataset.

## A.3 Details on Human Evaluation

For the results generated by each method, following (Li et al., 2022b), we randomly selected 50 sentences to be submitted in the Amazon Mechanical Turk[8] questionnaire. We pay our workers 5 cents per sentence. As shown in Figure 4, the questionnaire asked to judge the generated sentences on three dimensions: strength of style transfer, degree of content retention, and text fluency. To minimize the impact of spamming, we require each worker to be a native English speaker with a 95% or higher approval rate and a minimum of 1,000 hits.

[5]https://github.com/asyml/texar/tree/v0.2.4/examples/textstyletransfer
[6]https://github.com/Nrgeup/controllable−text−attribute−transfer
[7]https://github.com/fastnlp/style-transformer
[8]https://www.mturk.com/


Read the two pieces of text below and use the sliders below indicate how much you agree with the statements (0 = Strongly disagree, 5 = Strongly agree)

Source Text (And you told me to bury my love. })

Transfer Text (And thou didst bid me bury my love.})

- 1) The second text adequately change the style of the first text to knight-errant style
- 2) The second text retain the main content of the original sentence
- 3) The second text is fluent and free of grammatical errors

Figure 4: Human evaluation questionnaire. We randomly sampled 50 sentences generated on the knight-errant style and distributed a questionnaire at Amazon Mechanical Turk asking each worker to rank the content retention (0 to 5), style transfer(0 to 5 ), and fluency(0 to 5).

| | Knight-errant(Chinese) |
|---|---|
| Input | 她感到一股淡淡的香味<br>(She felt a faint scent.) |
| Knight-errant | 一股淡淡的香气扑鼻而来<br>(A faint fragrance hit her nose.) |
| Human Ref. | 在鼻边一嗅，觉有一股淡淡的香气<br>(A sniff at the nose, feel a faint aroma.) |
| Input | 这些话是真诚的<br>(These words are sincere) |
| Knight-errant | 这几句话说得甚是诚恳<br>(These words were said with great sincerity) |
| Human Ref. | 这几句话情辞真挚，十分恳切。<br>(These words are sincere and earnest.) |
| Input | 他心里悲凉，心绪凄凉。<br>(His heart was sad and his mind was bleak.) |
| Knight-errant | 他心中一凉,思潮起伏,甚是凄凉<br>(His heart in a desolate, thoughts ups and downs, very desolate.) |
| Human Ref. | 他心中悲痛，意兴萧索<br>(His heart was saddened and his mood was depressed.) |
| Input | 他们都惊讶地看着对方<br>(They both looked at each other in amazement) |
| Knight-errant | 众人面面相觑,都吃了一惊<br>(Everyone looked at each other in dismay, all taken aback.) |
| Human Ref. | 众人愕然相顾。<br>(The crowd looked at each other in consternation.) |
| Input | 没有灯光，没有声音<br>(There are no lights, no sound.) |
| Knight-errant | 四下无灯,更无声息<br>(All around no lights,and even less noise.) |
| Human Ref. | 四下里黑沉沉地，既无灯火，又无人声<br>(All around, it was dark and dreary, with no lights and no sound.) |

Table 8: Chinese knight-errant style generation results on KE dataset.