Reinforcement Learning Based Text Style Transfer without Parallel Training Corpus

Hongyu Gong* **Suma Bhat*** **Lingfei Wu†** **Jinjun Xiong †** **Wen-mei Hwu***
*University of Illinois at Urbana-Champaign, USA
†T. J. Watson Research Center, IBM
*{hgong6, spbhat2, w-hwu}@illinois.edu †{wuli, jinjun}@us.ibm.com

Abstract

Text style transfer rephrases a text from a source style (e.g., informal) to a target style (e.g., formal) while keeping its original meaning. Despite the success existing works have achieved using a parallel corpus for the two styles, transferring text style has proven significantly more challenging when there is no parallel training corpus. In this paper, we address this challenge by using a reinforcement-learning-based generator-evaluator architecture. Our generator employs an attention-based encoder-decoder to transfer a sentence from the source style to the target style. Our evaluator is an adversarially trained style discriminator with semantic and syntactic constraints that score the generated sentence for style, meaning preservation, and fluency. Experimental results on two different style transfer tasks (sentiment transfer and formality transfer) show that our model outperforms state-of-the-art approaches. Furthermore, we perform a manual evaluation that demonstrates the effectiveness of the proposed method using subjective metrics of generated text quality.

1 Introduction

Text style transfer is the task of rewriting a piece of text to a particular style while retaining the meaning of the original text. It is a challenging task of natural language generation and is at the heart of many recent NLP applications, such as personalized responses in dialogue system (Zhou et al., 2017), formalized texts (Rao and Tetreault, 2018), cyberspace purification by rewriting offensive texts (Niu and Bansal, 2018; Santos et al., 2018), and poetry generation (Yang et al., 2018).

Recent works on supervised style transfer with a parallel corpus have demonstrated considerable success (Jhamtani et al., 2017b; Rao and Tetreault, 2018). However, a parallel corpus may not always be available for a transfer task. This has prompted studies on style transfer without parallel corpora. These hinge on the common idea of separating the content from the style of the text (Shen et al., 2017; Fu et al., 2018; Santos et al., 2018). This line of research first encodes the context via a style-independent representation, and then transfers sentences by combining the encoded content with style information. In addition, an appropriate training loss is chosen to change the style while preserving the content. However, these approaches are limited by their use of loss functions that must be differentiable with respect to the model parameters, since they rely on gradient descent to update the parameters. Furthermore, since focusing only on semantic and style metrics in style transfer, they ignore other important aspects of quality in text generation, such as language fluency.

In this paper, we propose a system trained using reinforcement-learning (RL) that performs text style transfer without accessing to a parallel corpus. Our model has a generator-evaluator structure with one generator and one evaluator with multiple modules. The generator takes a sentence in a source style as input and transfers it to the target style. It is an attention-based sequence-to-sequence model, which is widely used in generation tasks such as machine translation (Luong et al., 2015). More advanced model such as graph-to-sequence model can also exploited for this generation task (Xu et al., 2018b). The evaluator consists of a style module, a semantic module and a language model for evaluating the transferred sentences in terms of style, semantic content, and fluency, respectively. Feedback from each evaluator is sent to the generator so it can be updated to improve the transfer quality.

Our style module is a style discriminator built using a recurrent neural network, predicting the likelihood that the given input is in the target style.

Proceedings of NAACL-HLT 2019, pages 3168–3180
Minneapolis, Minnesota, June 2 - June 7, 2019.

We train the style module adversarially to be a target style classifier while regarding the transferred sentences as adversarial samples. An adversarial training renders style classification more robust and accurate. As for the semantic module, we used word movers' distance (WMD), a state-of-the-art unsupervised algorithm for comparing semantic similarity between two sentences (Kusner et al., 2015; Wu et al., 2018b), to evaluate the semantic similarity between input sentences in the source style and the transferred sentences in the target style. We also engaged a language model to evaluate the fluency of the transferred sentences.

Unlike prior studies that separated content from style to guarantee content preservation and transfer strength, we impose explicit semantic, style and fluency constraints on our transfer model. Moreover, employing RL allows us to use other evaluation metrics accounting for the quality of the transferred sentences, including non-differentiable ones.

We summarize our contributions below:
(1) We propose an RL framework for text style transfer. It is versatile to include a diverse set of evaluation metrics as the training objective in our model.
(2) Our model does not rely on the availability of a parallel training corpus, thus addressing the important challenge of lacking parallel data in many transfer tasks.
(3) The proposed model achieves state-of-the-art performance in terms of content preservation and transfer strength in text style transfer.

The rest of our paper is organized as follows: we discuss related works on style transfer in Section 2. The proposed text style transfer model and the reinforcement learning framework is introduced in Section 3. Our system is empirically evaluated on sentiment and formality transfer tasks in Section 4. We report and discuss the results in Section 5 and Section 6. The paper is concluded in Section 7.

2 Related Works

Text style transfer has been explored in the context of a variety of natural language applications, including sentiment modification (Zhang et al., 2018b), text simplification (Zhang and Lapata, 2017), and personalized dialogue (Zhou et al., 2017). Depending on whether the parallel corpus is used for training, two broad classes of style transfer methods have been proposed to transfer the text from the source style to the target style. We will introduce each line of research in the following subsections.

Style transfer with parallel corpus. Style transfer with the help of a style parallel corpus can be cast as a monolingual machine translation task. For this, a sequence-to-sequence (seq2seq) neural network has been successfully applied in a supervised setting. Jhamtani et al. transfer modern English to Shakespearean English by enriching a seq2seq model with a copy mechanism to replicate the source segments in target sentences (Jhamtani et al., 2017a).

Style transfer without parallel corpus. Scarce parallel data in many transfer tasks has prompted a recent interest in studying style transfer without a parallel corpus (e.g., (Zhang et al., 2018a)). Li et al. propose to delete words associated with the source style and replace them with similar phrases associated with the target style. Clearly, this approach is limited to transfers at the lexical level and may not handle structural transfer. Most existing unsupervised approaches share a core idea of disentangling content and style of texts. For a given source sentence, a style-independent content representation is firstly derived. Then, in combination with the target style, the content representation is used to generate the sentence following the target style.

Approaches to extract the content include variational auto-encoders (VAE) and cycle consistency. VAEs are commonly used to learn the hidden representation of inputs for dimensionality reduction, and have been found to be useful for representing the content of the source (Hu et al., 2017; Mueller et al., 2017; Shen et al., 2017; Fu et al., 2018). Cycle consistency is an idea borrowed from image style transfer for content preservation (Zhu et al., 2017). It proposes to reconstruct the input sentence from the content representation, by forcing the model to keep the information of the source sentence (Santos et al., 2018).

The transferred sentences are generated based on the content representation and the target style. One way to achieve this is with the use of a pre-trained style classifier. The classifier scores the transfer strength of the generated sentences and guides the model to learn the target text style (Santos et al., 2018; Prabhumoye et al., 2018). Another way is to learn the style embedding, which can

be concatenated with the content embedding as the representation of the target sentence (Fu et al., 2018). The decoder then constructs the sentences from their hidden representations.

We note that previous works rely on gradient descent in their model training, and therefore their training losses (e.g., content and style loss) were limited to functions differentiable with respect to model parameters. Also, very few works consider other aspects of transfer quality beyond the content and the style of the generated sentences. This is in part due to their reliance on a differentiable training objective. We propose an RL-based style transfer system so that we can incorporate more general evaluation metrics in addition to preserving the semantic meaning of content and style transfer strength.

Reinforcement learning. RL has recently been applied to challenging NLP tasks (Yu et al., 2017). RL has advantages over supervised learning in that it supports non-differentiable training objectives and does not need annotated training samples. Benefits of using RL have been demonstrated in image captioning (Guo et al., 2018), sentence simplification (Zhang and Lapata, 2017), machine translation (Wu et al., 2018a) and essay scoring (Wang et al., 2018). A recent work on the task of sentiment transfer applied reinforcement learning to handle its BLEU score-based training loss (a non-differentiable function) (Xu et al., 2018a). Similar to the style transfer works discussed above, it also disentangled the semantics and the sentiment of sentences using a neutralization module and an emotionalization module respectively. Our work is different from these related works in that the semantic preservation and transfer strength are taken care of by the use of discriminators without explicitly separating content and style. An additional aspect that we focus here is the notion of fluency of the transferred sentences, which has not been explored before.

3 Model

Our style transfer system consists of the following modules: a generator, a style discriminator, a semantic module and a language model as shown in Fig. 1. We next describe the structure and function of each component. A closer view of our system is presented in Fig. 2.

Generator. The generator in our system takes a sentence in the source style as input and transfers it to the target style. For this, we use a recurrent encoder-decoder model combined with attention mechanism, which can handle variable-length input and output sequences (Sutskever et al., 2014; Cho et al., 2014). We could also leverage recently proposed more advanced encoder-decoder models (Xu et al., 2018b,c) to exploit rich syntactic information for this task, which we leave it as future work. Both the encoder and the decoder are recurrent neural layers with gated recurrent units (GRU). The encoder takes one word from the input at each time step, and outputs a hidden state vector $\bar{h}_s$ at time s. Similarly, the decoder outputs a hidden representation h_t at time t.

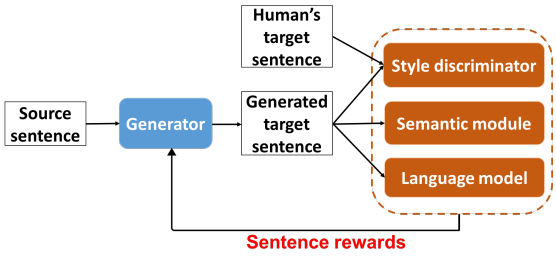

Figure 1: Model overview: the generator transfers the input source sentence to the generated target sentence. The generated sentences are collectively evaluated by the style discriminator, the semantic module and language module respectively. The style discriminator is adversarially trained with both human- and model-generated sentences. These three modules evaluate the generated sentences in terms of transfer strength, content preservation and fluency, and the rewards are sent to train the generator.

Suppose that the input sequence consists of T words $x = \{x_1, \ldots, x_T\}$, and the generated target sentence y is also a sequence of words $\{y_1, \ldots, y_{T'}\}$. We use vec$(\cdot)$ to denote the embedding of a word.

The gated recurrent unit dynamically updates its state h_t based on its previous state h_{t-1} and current input i_t. Its computation can be abstracted as $h_t = \text{GRU}(h_{t-1}, i_t)$. For the encoder, the input i_t is the embedding of the t-th input source word,

$$\bar{h}_t = \text{GRU}(\bar{h}_{t-1}, \text{vec}(x_t)). \quad (1)$$

For the decoder, the input to the recurrent unit is the embedding of the t-th generated target word.

$$h_t = \text{GRU}(h_{t-1}, \text{vec}(y_t)). \quad (2)$$

An attention mechanism is commonly adopted in text generation, such as machine translation (Bahdanau et al., 2015; Luong et al., 2015). We

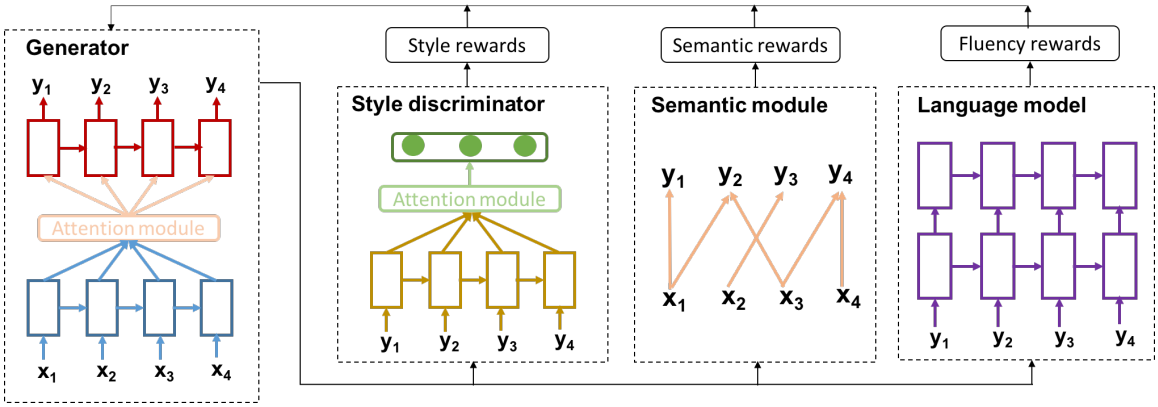

Figure 2: A detailed view of each component in the text style transfer system.

apply the attention mechanism to the decoding step so that the decoder learns to attend to source words and generates words. In this work, we use the attention mechanism similar to that used in (Luong et al., 2015). At the t-th decoding step, the attention $\alpha_t(s)$ is the weight of the s-th encoder state $\bar{h}(s)$.

The encoder hidden states are linearly weighted by the attention as the context vector at time t.

$$c_t = \sum \alpha_t(s)\bar{h}_s. \quad (3)$$

Combining the attention over the source sentence, the decoder produces a new hidden state $\tilde{h}_t$,

$$\tilde{h}_t = \tanh(W_c[c_t; h_t]). \quad (4)$$

The hidden vector $\tilde{h}_t$ is then used to predict the likelihood of the next word in the target sentence over the target vocabulary.

$$\mathbb{P}(y_t|y_{<t}, x) = \text{softmax}(W_s\tilde{h}_t). \quad (5)$$

where W_c and W_s are decoder parameters.

Style discriminator. The style discriminator evaluates how well the generated sentences are transferred to the target style. It is a classifier built on a bidirectional recurrent neural network with attention mechanism. The style discriminator is pre-trained to minimize the cross-entropy loss in the style classification task. This style classifier predicts the likelihood that an input sentence is in the target style, and the likelihood is taken as the style score of a sentence.

The pre-training does not guarantee that the neural network model will learn robust style patterns. So we resort to adversarial training as done in generative adversarial networks (GAN) (Yu et al., 2017; Wang and Lee, 2018). Accordingly, the style discriminator is later adversarially trained to distinguish the original (human-written) sentences from the model-generated ones so that the classifier learns to classify the text style well.

Semantic module. This evaluates how well the content from the input is preserved in the generated sentences. We use word mover's distance (WMD), which is the state-of-the-art approach (known for its robustness and efficiency) to measure the dissimilarity between the input and output sentences based on word embeddings (Kusner et al., 2015; Wu et al., 2018b). We take the negative of the WMD distance and divide it by the sequence length to yield the *semantic score* of a generated sentence. Previous works have also used cycle reconstruction loss to measure content preservation by reconstructing input sentences from generated sentences (Santos et al., 2018).

Language model. The style and the semantic modules do not guarantee the fluency of the transferred sentences. This fluency is achieved using a language model. The language model we use is a two-layer recurrent neural network pre-trained on the corpus in the target style so as to maximize the likelihood of the target sentences (Mikolov et al., 2010; Jozefowicz et al., 2016). The language model estimates the probability of input sentences. We take the logarithm of the probability and divide it by the sequence length as the *fluency score*.

3.1 Reinforcement Learning

The output sentences from the generator are sent to the semantic, style and language model modules for evaluation. These modules give feedback to the generator for the purpose of tuning it and to improve the quality of the generated sentences. We emphasize that despite the fact that our chosen evaluation metrics are not differentiable with respect to the generator parameters, they are still us-

able here. This is made possible by our use of the RL framework (the REINFORCE algorithm) to update the parameters of the generator (Williams, 1992).

In the RL framework, we define the state and the action for our style transfer task as follows. The state s_t at time t is the input source sequence and the first $t-1$ words that are already generated in the target sequence, i.e., $s_t = (X, Y_{1:t-1})$. The action a_t at time t is the t-th word to be generated in the output sequence, i.e., $a_t = y_t$. Suppose that the target vocabulary is V, and the maximum length of the decoder is T'. The generator G is parameterized with a parameter set θ, and we define the expected reward of the current generator as $J(G_\theta)$. The total expected reward is

$$J(G_\theta) = \sum_{t=1}^{T'} \mathbb{E}_{Y_{1:t-1} \sim G_\theta}[\sum_{y_t \in V} \mathbb{P}_\theta(y_t|s_t) Q(s_t, y_t)], \tag{6}$$

where $\mathbb{P}_\theta(y_t|s_t)$ is the likelihood of word y_t given the current state, and $Q(s_t, y_t)$ is the cumulative rewards that evaluate the quality of the sentences extended from $Y_{1:t}$. Suppose that $r(s_t, y_t)$ is the reward of word y_t at state s_t. The total reward, Q, is defined as the sum of the word rewards.

$$Q(s_t, y_t) = \sum_{\tau=t}^{T'} \gamma^{\tau-t} r(s_\tau, y_\tau), \tag{7}$$

where γ $(0 < \gamma < 1)$, is a discounting factor so that the future rewards have decreasing weights, since their estimates are less accurate.

If we only consider one episode, i.e., $Y_{1:t-1}$ has been given for every y_t, the reward $J(G_\theta)$ can be written as

$$J(G_\theta) = \sum_{t=1}^{T'} \sum_{y_t \in V} \mathbb{P}_\theta(y_t|s_t) Q(s_t, y_t). \tag{8}$$

Sequence sampling. By design, the three evaluation modules in Fig. 1 only evaluate complete sentences instead of single words or partial sentences. This means that we cannot obtain $r(s_t, y_t)$ directly from the evaluation modules at any time instance before the end of the sentence. One way around this problem is rolling out (Yu et al., 2017), where the generator 'rolls out' the given sub-sentence $Y_{1:t}$ at time step t to generate complete sentences by sampling the remaining part of the sentence $\{Y^n_{t+1:T'}\}$.

Previous works have adopted different sampling strategies, including Monte Carlo search, multinomial sampling and beam search. Starting from the given segment $Y_{1:t}$, Monte Carlo search explores the sub-sequence which leads to the best complete sentence (Yu et al., 2017). This leads to a good estimate of the sentence rewards but comes at significant computational cost. In many applications, the other two sampling strategies have been adopted for their efficiency. In multinomial sampling, each word y_τ $(t < \tau \leq T')$ is sampled from the vocabulary according to the likelihood $\mathbb{P}(y_\tau|s_\tau)$ predicted by the generator (ODonoghue et al., 2016; Chatterjee and Cancedda, 2010). The beam search process, on the other hand, keeps track of the k (a user-specified parameter) most likely words at each decoding step rather than just one word (Wu et al., 2018a). While this yields an accurate estimate of the reward for each action, multinomial sampling allows us to explore the diversity of generated texts with a potentially higher reward later on. This is the trade-off between exploitation and exploration in RL.

To balance the estimation accuracy and the generation diversity, we combine the ideas of beam search and multinomial sampling. Given a source sentence, we first generate a reference target sentence $Y^{\text{ref}}_{1:T}$ using beam search. To estimate the reward at each time step t, we draw samples of complete sentences $\{Y^l_{1:T'}\}$ by rolling out the subsequence $Y^{\text{ref}}_{1:t}$ using multinomial sampling. The evaluation scores of the sampled sentences are used as reward $r(y_t, s_t)$. More details about the sampling process are in Appendix.

Reward estimation. We estimate the reward as follows. We draw N samples of complete sentences starting from $Y_{1:t}$: $\{Y^{(n)}_{1:T'}\}^N_{n=1}$. The complete sentences are then fed into the three evaluation modules. Let f_{style} be the style score given by the style module, f_{semantic} be semantic score by the semantic module, and f_{lm} be the fluency score given by the language model.

We score the action y_t at state s_t by the average score of the complete sentences rolled out from $Y_{1:t}$. This action score is defined as the weighted sum of the scores given by the three modules.

$$f(s_t, y_t) = \frac{1}{N} \sum_{n=1}^{N} (\alpha \cdot f_{\text{style}}(Y^{(n)}_{1:T'}) + \beta \cdot f_{\text{semantic}}(Y^{(n)}_{1:T'}, Y^{\text{real}}_{1:T'}) + \eta \cdot f_{\text{lm}}(Y^{(n)}_{1:T'})), \tag{9}$$

where the hyperparameters α, β and η are positive.

In our experiments, we set $\alpha = 1.0$, $\beta = 0.5$ and $\eta = 0.5$ heuristically.

Given the scores from the evaluation modules, we define the reward $r(s_\tau, y_\tau)$ of word y_τ at state s_τ as

$$r(s_\tau, y_\tau) = \begin{cases} f(s_\tau, y_\tau) - f(s_{\tau-1}, y_{\tau-1}), & \tau > 1, \\ f(s_1, y_1), & \tau = 1. \end{cases} \quad (10)$$

We then obtain the discounted cumulative reward $Q(s_t, y_t)$ from the rewards $\{r(s_\tau, y_\tau)\}_{\tau>t}$ at each time step using Eq. 7.

The total reward of $J(G_\theta)$ can be derived from the cumulative rewards $\{Q(s_t, y_t)\}$ using Eq. 8. We define the generator loss L_θ as the negative of reward $J(G_\theta)$, $L_G(\theta) = -J(G_\theta)$.

According to Eq. 8, we can find the gradient $\nabla_\theta L_\theta$ of the generator loss as,

$$\nabla_\theta L_G(\theta) = -\sum_{t=1}^{T'} \nabla_\theta \mathbb{P}_\theta(y_t|s_t) Q(s_t, y_t). \quad (11)$$

3.2 Adversarial Training

The style discriminator is pre-trained on corpora in the source and target styles, and is used to evaluate the strength of style transfer. We note that this pre-training may not be sufficient for the style classifier to learn robust patterns and to provide accurate style evaluation. Indeed, in our experiments we found that even though the generator was trained to generate target sentences by maximizing the style rewards, the one-shot pre-training was insufficient to render the sentences in the target style.

Borrowing the idea of adversarial training proposed in GANs, we continuously trained the style discriminator using the generated target sentences. Toward this, we used a combination of a randomly sampled set of human-written target sentences $\{Y^{(k)}_{\text{human}}\}$ and model-generated sentences $\{Y^{(k)}_{\text{model}}\}$. Here the model-generated instances act as adversarial training samples, using which, the style discriminator was trained to distinguish the model outputs from human-written sentences. Let the discriminator D be parameterized by a parameter set ϕ. We define the prediction of the style discriminator, $D(Y)$, as the likelihood that the sentence Y is in the target style. The objective of this adversarial training amounts to minimizing the discriminator loss L_D:

$$L_D(\phi) = \frac{1}{K}\Big(-\sum_{k=1}^{K} \log(1 - D_\phi(Y^{(k)}_{\text{model}})) - \sum_{k=1}^{K} \log D_\phi(Y^{(k)}_{\text{human}}) \Big) . \quad (12)$$

4 Experiments

In this work, we considered two textual style transfer tasks, that of sentiment transfer (**ST**, involving negative and positive sentiments) and formality transfer (**FT**, involving informal and formal styles) using two curated datasets. We experimented with both transfer directions: positive-to-negative, negative-and-positive, informal-to-formal and formal-to-informal.

Dataset. For our experiments with style transfer we used a sentiment corpus and a formality corpus described below.

	Vocabulary	Type	Train	Dev	Test
Sentiment	9,640	Negative	176,878	25,278	50,278
		Positive	267,314	38,205	76,392
Formality	21,129	Informal	50,711	1,019	1,327
		Formal	50,711	1,019	1,019

Table 1: Data sizes of sentiment and formality transfer.

(1) Sentiment corpus. The sentiment corpus consists of restaurant reviews collected from the Yelp website (Shen et al., 2017). The reviews are classified as either negative or positive.

(2) Formality corpus. We use the Grammarly's Yahoo Answers Formality Corpus (GYAFC) (Rao and Tetreault, 2018), which is a collection of sentences posted in a question-answer forum (Yahoo Answers) and written in an informal style. In addition, these sentences have been manually rewritten in a formal style. We used the data from the section *family and relationships*. Note that even though the corpus is parallel, we did not use the parallel information.

Table 1 shows the train, dev and test data sizes as well as the vocabulary sizes of the corpora used in this work.

Model settings. The word embeddings used in this work were of dimension 50. They were first trained on the English WikiCorpus and then tuned on the training dataset. The width of the beam search (parameter k) was 8 during the RL and the inference stage.

Type	Source sentence	Transferred sentence
Negative-to-Positive	Crap fries , hard hamburger buns , burger tasted like crap !	Love you fries, burgers , always fun burger , authentic !
Positive-to-Negative	I was very impressed with this location .	I was very disappointed with this location .
Informal-to-Formal	It defenitely looks like he has feelings for u do u show how u feel u should ! !	It is like he is interested in you you should show how you feel .
Formal-to-Informal	I believe you 're a good man most likely she loves you quite a bit.	I think you 're a good man she kinda loves you .

Table 2: Example transferred sentences.

Pre-training. We pre-trained the generator, the style discriminator and the language model before the reinforcement learning stage. We discuss each of these steps below.

Generator pre-training. We pre-trained the generator to capture the target style from the respective target corpus. This pre-training occurred before setting up the reward from the evaluator to update its parameters in reinforcement learning. During pre-training, we used a set of target instances with a given instance serving as the input as well as the expected output. Using this set we trained the generator in a supervised manner with the cross-entropy loss as the training objective. Pre-training offered two immediate benefits for the generator: (1) the encoder and decoder learned to capture the semantics and the target style from the target corpus; (2) the generator had a good set of initial parameters that led to faster model training. This second aspect is a significant gain, considering that reinforcement learning is more time consuming than supervised learning.

Style discriminator pre-training. The style discriminator in our work was built using a bidirectional recurrent neural network. It was pre-trained using training corpora consisting of sentences in both the source and the target styles. We trained it to classify the style of the input sentences with the cross-entropy classification loss.

Language model pre-training. The language model was a two-layer recurrent neural network. Taking a target sentence $y = \{y_1, \ldots, y_{T'}\}$ as the input, the language model predicted the probability of the t-th word y_t given the previous subsequence $y_{1:t-1}$. The language model was pre-trained on the training corpus in target style to maximize the probability of y_t $(1 \leq t \leq T')$.

Baselines. We considered two state-of-the-art methods of unsupervised text style transfer that use non-parallel training corpus.

(1) Cross alignment model (CA). The CA model assumes that the text in the source and target style share the same latent content space (Shen et al., 2017). The style-independent content representation generated by its encoder is combined with available style information to transfer the sentences to the target style. We used their publicly available model for ST, and trained the model for FT separately with its default parameters.

(2) Multi-decoder seq2seq model (MDS). MDS consists of one encoder and multiple decoders (Fu et al., 2018). Similar to the cross alignment transfer, its encoder learns style-independent representations of the source, and the style specific decoder will rewrite sentences in the target style based on the content representation. We trained the model with its default parameters for both the tasks.

4.1 Evaluation

We used both automatic and human evaluation to validate our system in terms of content preservation, transfer strength and fluency.

4.1.1 Automatic evaluation

Aligning with prior work, we used the automatic metrics of content preservation, transfer and fluency that have been found to be well correlated with human judgments (Fu et al., 2018). For comparison, in Appendix, we also report our style and semantic metrics as provided by the evaluator.

Content preservation. A key requirement of the transfer process is that the original meaning be retained. Here we measure this by an embedding based sentence similarity metric s_{sem} proposed by (Fu et al., 2018). The embedding we used was based on the word2vec (CBOW) model (Mikolov et al., 2013). It was first trained on the English WikiCorpus and then tuned on the training dataset. Previous works used pre-trained GloVe embedding (Pennington et al., 2014), but we note that it does not have embeddings for Internet slang commonly seen in sentiment and formality datasets.

Transfer strength. The transfer strength s_{style} captures the degree to which the style transfer was

Sentiment	Negative-to-Positive				Positive-to-Negative			
Metric	Content	Style	Overall	Perplexity	Content	Style	Overall	Perplexity
CA	**0.894**	0.836	0.432	103.11	0.905	0.836	0.435	185.35
MDS	0.783	0.988	0.437	**98.89**	0.756	0.860	0.402	**156.98**
RLS	0.868	0.98	**0.460**	119.24	0.856	0.992	**0.459**	174.02
Formality	Informal-to-Formal				Formal-to-Informal			
Metric	Content	Style	Overall	Perplexity	Content	Style	Overall	Perplexity
CA	0.865	0.558	0.339	238.05	0.789	0.956	0.432	317.40
MDS	0.519	0.435	0.237	278.65	0.546	0.998	0.353	352.86
RLS	0.885	0.601	**0.358**	**208.33**	0.873	0.982	**0.462**	**267.78**

Table 3: Automatic evaluation of text style transfer systems on sentiment and formality transfer.

carried out and was quantified using a classifier. An LSTM-based classifier was trained for style classification on a training corpus (Fu et al., 2018). The classifier predicts the style of the generated sentences with a threshold of 0.5. The prediction accuracy is defined as the percentage of generated sentences that were classified to be in the target style. The accuracy was used to evaluate transfer strength, and the higher the accuracy is, the better the generated sentences fit in target style.

Overall score. We would like to point out that there is a trade-off between content preservation and transfer strength. This is because the outputs resulting from unchanged input sentences show the best content preservation while having poor transfer strength. Likewise, for given inputs, sentences sampled from the target corpora have the strongest transfer strength while barely preserving any content if at all. To combine the evaluation of semantics and style, we use the overall score s_{overall}, which is defined as a function of s_{sem} and s_{style}: $s_{\text{overall}} = \frac{s_{\text{sem}} * s_{\text{style}}}{s_{\text{sem}} + s_{\text{style}}}$ (Fu et al., 2018).

Fluency. This is usually evaluated with a language model in many NLP applications (Peris and Casacuberta, 2015; Tüske et al., 2018). We used a two-layer recurrent neural network with gated recurrent units as a language model, and trained it on the target style part of the corpus. The language model gives an estimation of perplexity (PPL) over each generated sentence. Given a word sequence of M words $\{w_1, \ldots, w_M\}$ and the sequence probability $p(w_1, \ldots, w_M)$ estimated by the language model, the perplexity is defined as:

$$\text{PPL} = p(w_1, \ldots, w_M)^{-\frac{1}{M}}. \quad (13)$$

The lower the perplexity on a sentence, the more fluent the sentence is.

4.1.2 Human annotation

Noting the best overall score of our system in both directions of the tasks considered (to be discussed in the section that follows), we performed human annotations for content, style and fluency to validate the automatic scores. We chose a sample of 100 sentences generated by our system for each transfer task and collected three human judgments per sentence in each evaluation aspect. The annotation guidelines were:

Content preservation. Following the annotation scheme adopted by (Rao and Tetreault, 2018), we asked annotators to rate the semantic similarity between the original and transferred sentence on a scale from 1 to 6. Here "1" means completely dissimilar, "2" means dissimilar but on the same topic, "3" means dissimilar while sharing some content, "4" means roughly similar, "5" means almost similar, and "6" means completely similar.

Transfer strength. Annotators were given pairs of original and transferred sentences and were asked to decide which one was more likely to be in the target style. We define transfer strength to be the percentage of transferred sentences that were classified to be in the target style.

Fluency. Similar to the annotation of content, annotators scored sentences for fluency on a scale of 1(*not fluent*) to 6 (*perfectly fluent*).

5 Results

Some example sentences transferred by our system are shown in Table 2. More transferred sentences generated by our system and those by the baseline methods can be found in the Appendix. We first report the results of the automatic evaluation of our proposed system (denoted as "RLS") and the two baselines–the cross alignment model (CA) (Shen et al., 2017) and the multi-decoder seq2seq model (MDS) (Fu et al., 2018)–in Table 3.

Sentiment transfer. We notice that CA was the best in preserving content, MDS generated the most fluent target sentences and our model achieved the best trade-off between meaning and

Metric	Negative-to-positive	Positive-to-negative	Informal-to-formal	Formal-to-informal
Content (1-6)	5.19	5.20	4.96	5.33
Style accuracy	0.90	0.91	0.83	0.86
Fluency (1-6)	5.51	5.61	5.33	5.21

Table 4: Human judgments of transferred sentences

style with the highest overall score. Looking at the Overall score, it is notable that despite the differences in performance between the models studied here, each one performs similarly in both directions. This could be interpreted to mean that with respect to difficulty of transfer, style transfer is equivalent in both the directions for this task.

Formality transfer. For this task, we notice that our model outperforms the baselines in terms of content preservation, transfer strength and fluency with the best Overall score and perplexity. This suggests that our model is better at capturing formality characteristics compared to the baselines. We also note that the style strength of all models for informal-to-formal transfer is significantly lower than that for formal-to-informal transfer. This suggests that the informal-to-formal transfer is harder than the reverse. A plausible explanation is that informal sentences are more diverse and thus easier to generate than formal sentences. For example, informality can be achieved by multiple ways, such as by using an abbreviation (e.g., "u" used as "you") and adding speech markers (e.g., "hey" and "ummm"), while formality is achieved in a more restricted manner.

Another challenge for informal-to-formal transfer is that informal data collected from online users usually contain non-negligible spelling errors such as "defenetely", "htink" and "realy". Words being the smallest semantic units in all the models considered here, these spelling errors could affect the transfer performance.

For each direction of transfer, we average the scores by annotators for each evaluation item, and report the results in Table 4. Our transferred sentences are shown to have good quality in content, style and fluency in subjective evaluations.

6 Discussion

To gain insights into the ways in which our approach performs the intended style transfer, we randomly sampled the generated sentences in the informal-to-formal transfer task. We found that the forms of rewriting can be broadly classified as: lexical substitution, word removal, word insertion and structural change. We show the following examples to these forms of re-writing, where the changed parts are highlighted.

(1) Lexical substitution. The informal sentence "I do **n't** know what **u** mean" was transferred to "I do **not** know what **you** mean";

(2) Word removal. The informal sentence "**And** I dont know what I should do" was rewritten as "I do not know what I should do";

(3) Word insertion. In the example instance "depends on the woman" that was changed to "**It** depends on the woman", we see that a subject was added to generate a complete formal sentence.

(4) Structural change. A small number of instances were also rewritten by making structural changes. For example, the informal sentence "**Just** tell them , **what are they gonna do** , slap you ??" was transferred to a formal version as "**You should** tell them , **they can not** slap you". Other ways of style transfer by incorporating evaluation metrics of structural diversity are left for future work.

7 Conclusion

We proposed a reinforcement-learning-based text style transfer system that can incorporate any evaluation metric to enforce semantic, stylistic and fluency constraints on transferred sentences. We demonstrated its efficacy via automatic and human evaluations using curated datasets on two different style transfer tasks. We will explore and incorporate other metrics to improve other aspects of generated texts such as the structural diversity in the future work.

Acknowledgments

This work is supported by IBM-ILLINOIS Center for Cognitive Computing Systems Research (C3SR) - a research collaboration as part of the IBM AI Horizons Network. We thank the NAACL anonymous reviewers for their constructive suggestions.

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *ICLR*.

Samidh Chatterjee and Nicola Cancedda. 2010. Minimum error rate training by sampling the translation lattice. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 606–615. Association for Computational Linguistics.

Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734.

Zhenxin Fu, Xiaoye Tan, Nanyun Peng, Dongyan Zhao, and Rui Yan. 2018. Style transfer in text: Exploration and evaluation. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 663–670.

Tszhang Guo, Shiyu Chang, Mo Yu, and Kun Bai. 2018. Improving reinforcement learning based image captioning with natural language prior. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 751–756.

Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P. Xing. 2017. Controllable text generation. *CoRR*, abs/1703.00955.

Harsh Jhamtani, Varun Gangal, Eduard Hovy, and Eric Nyberg. 2017a. Shakespearizing modern language using copy-enriched sequence-to-sequence models. *EMNLP 2017*, 6:10.

Harsh Jhamtani, Varun Gangal, Eduard H. Hovy, and Eric Nyberg. 2017b. Shakespearizing modern language using copy-enriched sequence-to-sequence models. *CoRR*, abs/1707.01161.

Rafal Jozefowicz, Oriol Vinyals, Mike Schuster, Noam Shazeer, and Yonghui Wu. 2016. Exploring the limits of language modeling. *arXiv preprint arXiv:1602.02410*.

Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger. 2015. From word embeddings to document distances. In *International Conference on Machine Learning*, pages 957–966.

Juncen Li, Robin Jia, He He, and Percy Liang. 2018. Delete, retrieve, generate: a simple approach to sentiment and 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)*, volume 1, pages 1865–1874.

Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421.

Tomáš Mikolov, Martin Karafiát, Lukáš Burget, Jan Černockỳ, and Sanjeev Khudanpur. 2010. Recurrent neural network based language model. In *Eleventh Annual Conference of the International Speech Communication Association*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 3111–3119.

Jonas Mueller, David Gifford, and Tommi Jaakkola. 2017. Sequence to better sequence: continuous revision of combinatorial structures. In *International Conference on Machine Learning*, pages 2536–2544.

Tong Niu and Mohit Bansal. 2018. Polite dialogue generation without parallel data. *Transactions of the Association for Computational Linguistics*, 6:273–389.

Brendan ODonoghue, Rémi Munos, Koray Kavukcuoglu, and Volodymyr Mnih. 2016. Pgq: Combining policy gradient and q. *arXiv preprint arXiv:1611.01626*.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1532–1543.

Álvaro Peris and Francisco Casacuberta. 2015. A bidirectional recurrent neural language model for machine translation. *Procesamiento del Lenguaje Natural*, 55:109–116.

Shrimai Prabhumoye, Yulia Tsvetkov, Ruslan Salakhutdinov, and Alan W Black. 2018. Style transfer through back-translation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 866–876. Association for Computational Linguistics.

Sudha Rao and Joel Tetreault. 2018. Dear sir or madam, may i introduce the yafc corpus: Corpus, benchmarks and metrics for formality style transfer. *arXiv preprint arXiv:1803.06535*.

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*.

Tianxiao Shen, Tao Lei, Regina Barzilay, and Tommi Jaakkola. 2017. Style transfer from non-parallel text by cross-alignment. In *Advances in Neural Information Processing Systems*, pages 6830–6841.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

Zoltán Tüske, Ralf Schlüter, and Hermann Ney. 2018. Investigation on LSTM recurrent n-gram language models for speech recognition. In *Interspeech 2018, 19th Annual Conference of the International Speech Communication Association, Hyderabad, India, 2-6 September 2018.*, pages 3358–3362.

Yau-Shian Wang and Hung-yi Lee. 2018. Learning to encode text as human-readable summaries using generative adversarial networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 4187–4195.

Yucheng Wang, Zhongyu Wei, Yaqian Zhou, and Xuanjing Huang. 2018. Automatic essay scoring incorporating rating schema via reinforcement learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 791–797.

Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256.

Lijun Wu, Fei Tian, Tao Qin, Jianhuang Lai, and Tie-Yan Liu. 2018a. A study of reinforcement learning for neural machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3612–3621.

Lingfei Wu, Ian EH Yen, Kun Xu, Fangli Xu, Avinash Balakrishnan, Pin-Yu Chen, Pradeep Ravikumar, and Michael J Witbrock. 2018b. Word mover's embedding: From word2vec to document embedding. *arXiv preprint arXiv:1811.01713*.

Jingjing Xu, SUN Xu, Qi Zeng, Xiaodong Zhang, Xuancheng Ren, Houfeng Wang, and Wenjie Li. 2018a. Unpaired sentiment-to-sentiment translation: A cycled reinforcement learning approach. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 979–988.

Kun Xu, Lingfei Wu, Zhiguo Wang, and Vadim Sheinin. 2018b. Graph2seq: Graph to sequence learning with attention-based neural networks. *arXiv preprint arXiv:1804.00823*.

Kun Xu, Lingfei Wu, Zhiguo Wang, Mo Yu, Liwei Chen, and Vadim Sheinin. 2018c. Exploiting rich syntactic information for semantic parsing with graph-to-sequence model. *arXiv preprint arXiv:1808.07624*.

Cheng Yang, Maosong Sun, Xiaoyuan Yi, and Wenhao Li. 2018. Stylistic chinese poetry generation via unsupervised style disentanglement. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 3960–3969.

Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. 2017. Seqgan: Sequence generative adversarial nets with policy gradient. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 2852–2858.

Xingxing Zhang and Mirella Lapata. 2017. Sentence simplification with deep reinforcement learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 584–594.

Ye Zhang, Nan Ding, and Radu Soricut. 2018a. Shaped: Shared-private encoder-decoder for text style adaptation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, volume 1, pages 1528–1538.

Yi Zhang, Jingjing Xu, Pengcheng Yang, and Xu Sun. 2018b. Learning sentiment memories for sentiment modification without parallel data. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1103–1108.

Ganbin Zhou, Ping Luo, Rongyu Cao, Fen Lin, Bo Chen, and Qing He. 2017. Mechanism-aware neural machine for dialogue response generation. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 3400–3407.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 2242–2251.

A Appendices

A.1 Sequence Sampling in Reinforcement Learning

The generator G transfers a source sentence X into a sentence in target style. In this work, we use beam search of width k to find a reference target sentence $Y^{\text{ref}}_{1:T'}$. In RL, we need to estimate the reward of each action y_t in the reference sentence

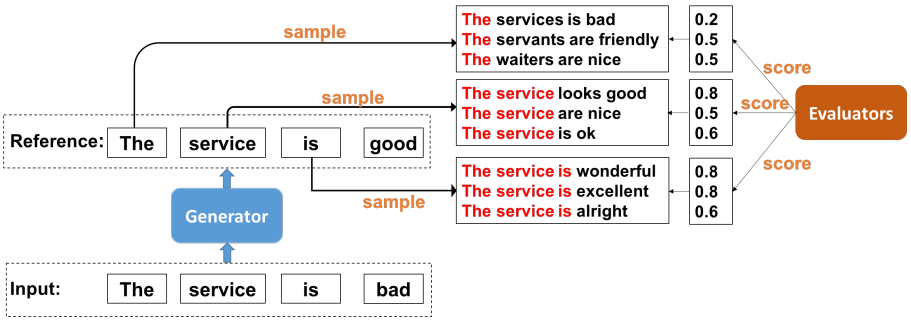

Figure 3: Sequence sampling: red words are sub-sequences in reference target sentence based on which the remaining sub-sequences are sampled. The sampled complete sentences are sent to the evaluator for scoring.

$Y^{\text{ref}}_{1:T'}$. Fig. 3 shows the sampling and scoring process.

Suppose that the reference target sentence $Y^{\text{ref}}_{1:T'}$ is "The service is good". At the first time step, i.e., when $t = 1$, we start from the sub-sequence $Y^{\text{ref}}_{1:1}$, i.e., the sub-sentence "The". We use multinomial sampling to roll out "The" to complete sentences, which are "The service is bad", "The servants are friendly" and "The waiters are nice" in Fig. 3. These sampled sentences are then sent to the evaluator for scoring in terms of style, content and fluency. Their scores are 0.2, 0.5 and 0.5 respectively, and we average them as the action score $f(y_1, s_1)$ of the first word y_1 at its state s_1. The score $f(y_1, s_1) = 0.4$ is sent back to the generator, which will be used to obtain the reward $r(y_1, s_1)$ as described in Eq. 10. Similarly when $t = 2$, we sample three complete sentences based on the sub-sentence "The service": "The service looks good", "The service are nice" and "The service is ok".

A.2 Experiments

Automatic evaluation metrics. We reported the automatic evaluation results of all text style transfer systems in Table 3, where we used the evaluation metrics adopted by previous works (Fu et al., 2018; Santos et al., 2018). Here we report the style and semantic scores given by the evaluator in our system in Table 5. Recall that semantic score given by our evaluator was the negative of word movers' distance between the generated sentence and the source sentence divided by the sentence length. The larger the semantic score was, the better the content was preserved in the generated sentence. As for the style evaluation, we used a bidirectional recurrent neural network as style classifier. It predicted the likelihood that an input sentence was in target style, which was taken as the *style score* of the generated sentences. Again, the larger the style score was, the better the generated sentence fitted in target style.

As shown in Table 5, the results given by the semantic and style modules of our evaluator are very similar to those given by Fu et al.. In sentiment transfer task, CA model does best in content preservation and MDS does best in transfer strength. As for FT, our model outperforms the two baselines in terms of semantic and style scores.

Examples and Analysis. We list some example transferred sentences given by our model and two baseline systems in Table 6. In the first example of negative-to-positive transfer, our model adheres to the topic of food service while baselines change to topic of food. Similarly in the first example of positive-to-negative transfer, our model preserves the topic of chicken while CA model talks about pizza and MDS model talks about customer service. Semantic similarity as explicit semantic constraints in our model is shown to be better at preserving the topic of source sentences.

There is still space to improve content preservation in all models. In the second example of informal-to-formal transfer, all transferred sentences miss the segment of "take a deep breathe" in the source sentence. In the second example of formal-to-informal transfer, the three transferred sentences miss part of source information. The source sentence is a rhetorical question, which truly means "people hardly understand the meaning behind their behavior". This is a hard example, and all models do not capture its semantic meaning accurately.

FT task is more challenging compared with ST given that the sentence structure is more complicated with a larger vocabulary in the formality dataset.

Table 5: Semantic and style scores given by our evaluator on all systems.

	Sentiment				Formality			
	Negative-to-Positive		Positive-to-Negative		Informal-to-Formal		Formal-to-Informal	
Metric	Semantic	Style	Semantic	Style	Semantic	Style	Semantic	Style
CA	**-1.293**	0.806	**-1.346**	0.818	-1.212	0.646	-1.281	0.851
MDS	-1.412	**0.855**	-1.662	0.822	-1.508	0.568	-1.445	**0.878**
RLS	-1.315	0.846	-1.458	**0.847**	**-0.935**	**0.782**	**-0.903**	0.872

Table 6: Example transferred sentences of all systems.

Type	Transferred sentence
Negative-to-positive	Source: I 've noticed the food service sliding down hill quickly this year . CA: I have enjoyed the food here and this place is perfect . MDS: Food is the best staff . RLS: I 've noticed the food service was perfect this time .
	Source: The chicken tenders did n't taste like chicken , wtf ? CA: The food tastes good , just like spicy ! MDS: And the food is the food in the food in well . RLS: . The chicken were like chicken, you can find what you want .
Positive-to-negative	Source: I recommend ordering the " special chicken " really good ! CA: I would give the pizza ... how they really really good ? MDS: They are the worst customer service . RLS: I would say chicken were very bad .
	Source: My experience was brief , but very good . CA: My experience was ok , but , very good . MDS: Worst , i would never go to going back . RLS: My experience was bad .
Informal-to-formal	Source: Well that is just the way it is i guess . CA: It is the best thing i think that is not . MDS: That is for the way . RLS: It is the way I think .
	Source: Like i said he already knows that you like him , so just take a deep breathe and ask him . CA: I think that she likes you , but perhaps you will get a relationship and and ask her . MDS: If you find him and i think that you have been in a relationship . RLS: I believe he knows that you like him, so go to ask him .
Formal-to-informal	Source: Well, if you are really attracted to this guy, then smile and speak nicely to him . CA: If you to tell her the way that is you and get married . MDS: The way of guys are not if you are not . RLS: Well , if you really like this guy , then smile to him .
	Source: Men are unintelligent! What person understands the meaning behind their behavior ? CA: Men are not of his meaning . MDS: Men are understands all men are not ? RLS: Men are stupid ! Why girl loves the mind ?