Towards Zero-Shot Multilingual Poetry Translation

Wai Lei Song nlp2ct.jacky@gmail.com
Haoyun Xu nlp2ct.haoyun@gmail.com
Derek F. Wong derekfw@um.edu.mo
Runzhe Zhan nlp2ct.runzhe@gmail.com
Lidia S. Chao lidiasc@um.edu.mo
Shanshan Wang* nlp2ct.shanshan@um.edu.mo
NLP^{2}CT Lab, Department of Computer and Information Science, University of Macau

Abstract

The application of machine translation in the field of poetry has always presented significant challenges. Conventional machine translation techniques are inadequate for capturing and translating the unique style of poetry. The absence of a parallel poetry corpus and the distinctive structure of poetry further restrict the effectiveness of traditional methods. This paper introduces a zero-shot method that is capable of translating poetry style without the need for a large-scale training corpus. Specifically, we treat poetry translation as a standard machine translation problem and subsequently inject the poetry style upon completion of the translation process. Our injection model only requires back-translation and easily obtainable monolingual data, making it a low-cost solution. We conducted experiments on three translation directions and presented automatic and human evaluations, demonstrating that our proposed method outperforms existing online systems and other competitive baselines. These results validate the feasibility and potential of our proposed approach and provide new prospects for poetry translation.

1 Introduction

The process of translating poetry presents an intricate challenge within the field of machine translation. The prevalence of the neural machine translation (NMT) paradigm (Vaswani et al., 2017), which necessitates copious amounts of data to effectively train a model capable of producing accurate translations (Koehn and Knowles, 2017). Unfortunately, the availability of parallel corpora that can be leveraged towards the training of a robust poetry translation system is currently inadequate. On the other hand, poetry is a manifestation of the unique imagination and creativity of the poet, as well as their distinctive writing style. As Chakrabarty et al. (2021) has pointed out, although NMT systems may succeed in translating the essence meaning of the poetry, the translation process inevitably disregards the writing style.

As a result, two distinct research paths have emerged within the field of poetry translation. In an effort to preserve the poet's distinctive writing style, Genzel et al. (2010) initially employed statistical machine translation, with a focus on maintaining the rhythm of the original

*Corresponding Author.

Proceedings of Machine Translation Summit XIX, Vol. 1: Research Track, pages 324–335
September 4–8, 2023, Macau SAR, China.
©2023 The authors. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

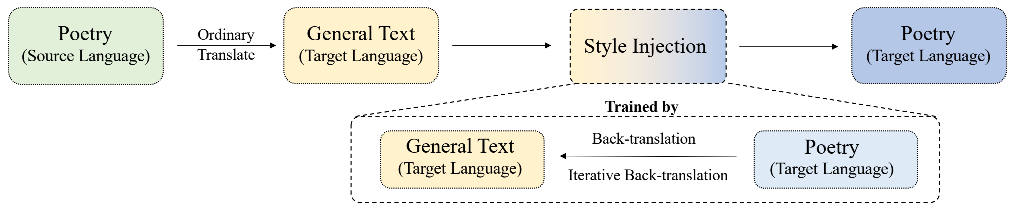

Figure 1: An illustration of the proposed zero-shot poetry translation method.

work. This approach was subsequently refined by Ghazvininejad et al. (2018), who introduced the additional constraints of rhythm and rhyme into neural poetry translation. This yielded an acceptability rate of 78.2% in terms of translation quality. Additionally, Yang et al. (2019) explored the use of unique tokens to govern the structure of the translated poetry. Meanwhile, From the perspective of overcoming the low-resource challenge, researchers have been proved to be an effective method by utilizing out-of-domain data, such as song lyrics (Shen et al., 2020; Liu et al., 2019). However, the existence of gaps in genre and writing style between poetry and other forms of text, improvements made to current poetry do not necessarily result in significant advancements. This is because these methods remain limited to enhancing the representation of general domain text, rather than enabling the model to comprehend the intricate patterns present in poetry. Encouragingly, as Chakrabarty et al. (2021) has provided poetic parallel corpora and demonstrated that fine-tuning with such data can greatly enhance the model's ability to adapt in poetry translation. Regrettably, the collection of parallel text and the construction of corpora are prohibitively expensive for research purposes. To address the aforementioned concerns in a cost-effective manner, we proposed a novel zero-shot method for poetry translation that consists of two stages. During the first stage, the model concentrates on learning to translate general domain text, thereby guaranteeing the preservation of the meaning of the source sentence and the fluency of the generated translation. In the second stage, the model acquires the ability to inject poetic style into the sentences generated in the first stage, resembling the process of style transfer (Huang et al., 2020; Li et al., 2020; Malmi et al., 2020). To achieve this goal, we collected over 2 million poetry texts in the target languages, including English, Portuguese, and Chinese. We utilized a back-translation approach to separate the poetic style, resulting in a pseudo-parallel corpus that goes from general text to poetry text. Our proposed method outperforms several competitive baselines, as evidenced by both automatic evaluation metrics and human evaluation results.

The contributions of our work are as follows:

- We are the first to propose a new approach for poetry translation without requiring a parallel poetry corpus. And the proposed method can be extended to other language pairs with monolingual data.
- We proposed a new human evaluation framework for poetry translation (Section 3.3) and invited several professional poets to evaluate the translation results.
- We will release the collected monolingual data and the created pseudo-parallel corpus for the purpose of research.

2 Methodology

To overcome the current absence of parallel poetry corpus, a zero-shot poetry translation method was proposed. As displayed in Figure 1, the proposed approach comprises of two distinct stages:

ordinary translation and style injection.

2.1 Related work

Poetry translations continue to predominantly rely on parallel corpora. To address this, one potential approach is to explore alternative datasets that share similar attributes to train the model effectively (Shen et al., 2020; Liu et al., 2019). Utilizing a multilingual parallel poetry corpus for fine-tuning pre-trained models has demonstrated promising results, indicating that poetry within a language family can be more effectively modeled through this approach (Chakrabarty et al., 2021). Furthermore, by leveraging the available dataset, modifying the model's mechanism or incorporating specific notations can enhance its ability to grasp the underlying structure of the poem more effectively (Ghazvininejad et al., 2018; Yang et al., 2019).

Text style transfer is the process of modifying the style of a sentence by rephrasing the original sentence in a different style while retaining its original meaning (Toshevska and Gievska, 2021). Until now, recent research has focused on a certain area of style transfer, such as Personal style (Pennebaker et al., 2003; Argamon et al., 2003; Peersman et al., 2016), Formality (Sheikha and Inkpen, 2010; Heylighen and Dewaele, 1999), Politeness (Brown et al., 1987; Andersson and Pearson, 1999; Chhaya et al., 2018), Offensiveness Pavlopoulos et al. (2019); Zampieri et al. (2019), Genre (Dewdney et al., 2001) and Sentiment (Russell, 1980; Susanto et al., 2020). Indeed, there are various methods utilized in different areas for achieving text style transfer. These methods may vary depending on the specific domain or application. Establishing a pseudo-parallel corpus by using an augmentation method such as back-translation is one direction of style transfer (Zhang et al., 2020b). Representation learning involves feeding sentences with a particular input style into the encoder while embedding the target style into the decoder. This enables the generation of desired outputs with different styles (Zhang et al., 2018; Liu et al., 2020). One approach to text style transfer involves removing certain words or adjusting the latent representation of a sentence, followed by regenerating the sentence to alter its style (Li et al., 2018; Sudhakar et al., 2019). The advantage of this method lies in its effectiveness in effectively changing the style compared to representation learning methods. However, one common challenge is that the model may struggle to maintain the same meaning of the sentences and may introduce grammar errors during the style transfer process.

2.2 Formulation

Given a poetry text $\mathbf{x}$ in the source language, the neural poetry translation model parameterized with θ generates the translation $\hat{\mathbf{y}}$ in the target language based on the conditional probability:

$$\hat{\mathbf{y}} = \arg\max \prod_{i=1}^{I} \log P(y_i | y_{<i}, \mathbf{x}; \theta) \tag{1}$$

Through the comparison of the golden truth $\mathbf{x}$ and the model hypothesis $\mathbf{y}$, the optimization of the model parameters θ occurs during the training process. We argue that the acquisition of robust parameters θ, necessary for the production of high-quality poetry translation $\mathbf{y}$, is a formidable task due to the aforementioned challenges. The primary obstacle in poetry translation lies in the model's requirement to not only translate the intended meaning accurately but also to adhere to the structural and stylistic conventions of the target language's poetry. This poses a significant translated challenge in terms of wording and sentence structure. Furthermore, directly translating the source language poetry text may result in the loss of the poem's intended meaning, due to the agency of the poetry text. This risk of meaning loss is further compounded by the inherent difficulty in producing high-quality translations when working with a single line of a poem at a time, as this approach lacks the necessary contextual and structural information to accurately convey the intended meaning, regardless of whether the source

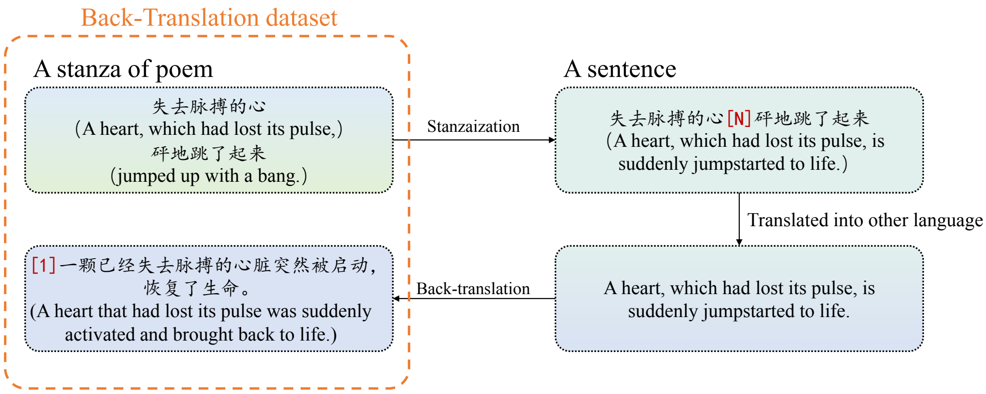

Figure 2: An illustration of poetry stanzaization process and back-translation process. The token known as "[number] serves the purpose of indicating the number of line breaks present in the original poems.

language or target language text is read.

Consequently, to address these challenges, as displayed in Figure 1, a two-stage translation approach is proposed, utilizing two distinct models, namely θ_{trans} and θ_{style}. In the first stage, model θ_{trans} is employed to translate the source sentence $\mathbf{x}$ into ordinary text $\hat{y}_{\text{general}}$. In the subsequent stage, model θ_{style} is utilized to inject the ordinary text with poetic style, producing the poetic text $\hat{y}$.

2.3 Poetry Stanzaization

A stanza is a fundamental poetry unit comprised of poetic lines that follow a specific principle or set of principles, such as syntax, meter, alliteration, lineation, or arc of thought. In some poetry styles, a stanza is created through end rhymes. The identification of stanzas is based on their intervals and other units of lines before and after them, often mirroring the first stanza. Stanzas possess a periodic nature, directing readers through a poem's text with their organized lines and transitions between stanzaic intervals. The line serves as both a rhythmical and structural unit, perceptible to both readers and listeners of poetry. Traditional mechanisms of closure are employed to delimitate stanzas from each other. As a unit of measure, a stanza is connected to adjacent units to form higher-level structures, while functioning as a structure built of lower-level poetic lines.

In order to guarantee the preservation of the meaning of poetry during the two-stage translation process, it is suggested that the stanza be encoded with the assistance of an additional signal. More specifically, as shown in Figure 2, the verse (or stanza) contains two lines. Therefore, a stanzaization process was utilized to properly handle the poem. the two sentences are concatenated into "A heart, which had lost its pulse, is suddenly jumpstarted to life." The token "[N]" is primarily utilized to represent the place after the stanzaization process, where the line break occurs. Furthermore, it is utilized to guarantee the model's consistent translation of an equivalent number of lines of poetry, thereby ensuring the consistency of evaluation conditions. In addition, when applying the stanzaization process to poetic works that are often written in a prose-like style, the resulting outcomes will typically consist of long sentences. To prevent the issue of excessively long spliced sentences, any sentence exceeding 100 words will be divided into shorter sentences based on punctuation or the natural ending point of the original sentence. The objective is to maximize the preservation of the sentence's meaning while minimizing the frequency of sentence cuts.

2.4 Stage I: Ordinary Translation

To acquire an ordinary translation model θ_{trans}, there exist three candidate systems, namely online system[1], in-house trained system, and pre-trained system[2][3]. In order to ascertain the preferred choice of an ordinary translation model, we employed English-to-Chinese translation tasks as our experimental framework. Our team opted to utilize the WMT'17 English⇒Chinese shared task data[4] to train our in-house systems, with the intention of securing a competitive translation model. In order to evaluate the performance of these systems, we compared them using the WMT'17 English⇒Chinese test set and determined their F-score via BERTScore Zhang et al. (2020a). Our results indicate that the online system achieved the highest performance on the test set[5], and as such, we have selected it to serve as the translation model for the first stage of our method for all languages.

2.5 Stage II: Style Injection

The utilization of Back-Translation (BT) is a prevalent technique in establishing a pseudo-parallel corpus that serves as an input for training the style injection model. One advantage of the BT method is that it only requires easily obtainable monolingual data, thereby reducing the need for costly parallel corpus construction. It is widely recognized that the era, experience, and emotional state of poets play a profound role in the style of their poemsYu and Liu (2021). Nonetheless, Rabinovich et al. (2017) research has revealed that sentences generated by machine translation tend to adopt the stylistic features of the machine translation process itself, rather than the specific style of the original author. Therefore, another benefit is that sentences generated through BT tend to adopt the stylistic features of machine translation, which can facilitate the injection of poetic style into the translated sentences. To illustrate, Figure 2 displayed the use of an online system to apply the BT method and establish a pseudo-parallel poetry corpus.

The monolingual poetry text $\mathbf{y}_{\text{poe}}$ in the target language is subjected to the BT method through an online system, resulting in the generation of an ordinary text $\mathbf{y}^*_{\text{general}}$, which can be formulated as: $\mathbf{y}_{\text{poe}} \rightarrow \mathbf{x}_{\text{general}} \rightarrow \mathbf{y}^*_{\text{general}}$. The sentence pair $\{\mathbf{y}^*_{\text{general}}, \mathbf{y}_{\text{poe}}\}$, comprising the original poetry text $\mathbf{y}_{\text{poe}}$ and its back-translated counterpart $\mathbf{y}^*_{\text{general}}$, is employed as the source and target data for training the style injection model θ_{style}. Furthermore, we incorporate an additional measure to our methodology by placing a distinct token labeled "[number]" at the beginning of each sentence. This token is used to indicate the number of lines present in each stanza, thereby providing an added layer of control over the overall structure of the poetic style. For example, as shown in Figure 2,"[1]" denotes that there are two lines in the current stanza. If only one line in the stanza the token will display "[0]".

One potential weakness of poetry stanzaization is that it will decrease the amount of training data because it merges multiple lines into one training instance. For example, the 1.6M modern Chinese poetry lines become 30K long sentences after applying the stanzaization process.

[1]https://fanyi.baidu.com
[2]https://huggingface.co/Helsinki-NLP/opus-mt-en-zh
[3]https://huggingface.co/facebook/m2m100_418M
[4]http://www.statmt.org/wmt17/translation-task.html
[5]Pre-trained: 0.742; In-house:0.809; Baidu Online System: 0.832

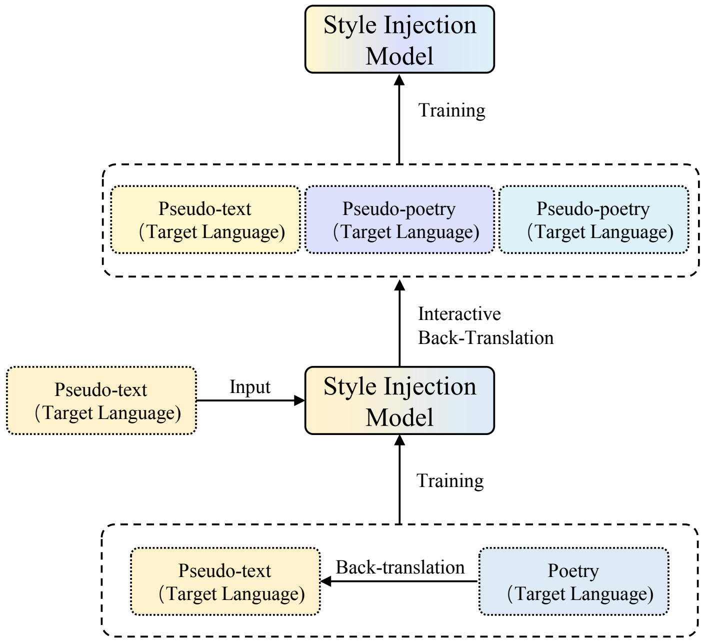

Figure 3: An illustration of Iterative Back-translation (IBT) process.

We have noted that certain words found within poems possess additional nuances of meaning that cannot be accurately conveyed through the back-translation process. Additionally, the act of performing back-translations may also result in alterations to the original intended meaning. Consequently, it is not uncommon for some of the generated sentences to contain erroneous meanings. Hence, we further use Iterative Back-translation Hoang et al. (2018) to augment the data for training θ_{style}. Hoang et al. (2018) introduced the Iterative Back-translation (IBT) methodology, and according to his approach, duplicating the bilingual corpus and appending it to the corpus dataset can be advantageous to the training model without any detrimental impact. Intuitively, the back-translation method is utilized to generate the pseudo-parallel dataset. The act of implementing the iterative back-translation method to enrich the dataset holds the potential to improve the model's capabilities and enhance its overall performance. As shown in Figure 3, IBT repeats the BT process to establish an extensive pseudo-parallel corpus for building a better translation system, and augmenting the pseudo-data by IBT is also a direct and inexpensive method. It is also helpful to improve the robustness of the style injection system by providing diverse pseudo-data.

3 Experimental Settings

We conduct experiments on English, Portuguese, and Chinese poetry translation task to verify the effectiveness of the proposed method.

3.1 Dataset

In this study, we have compiled a comprehensive collection of poetry in different languages. Specifically, a significant corpus of 1.6M sentences of modern Chinese poetry was collected, which was subsequently subjected to the process of stanzaization, resulting in a final compilation of 300k lines. Regarding the English and Portuguese poetry, we have refrained from

Lanugage	Split	Train	Dev	Test
Chinese	BT	300K	3K	876
	ConPP	300K		
	UnPP	300K		
	PSen	275K		
English	BT	290K	1K	100
	ConPP	270K		
	UnPP	270K		
	PSen	270K		
Portuguese	BT	380K	1K	100
	ConPP	350K		
	UnPP	350K		
	PSen	350K		

Table 1: Statistics of the pseudo-parallel corpora used to train style injection model. The BT corpus $\{\mathbf{y}^*_{\mathrm{general}}, \mathbf{y}_{\mathrm{poe}}\}$ is built by back-translation. The ConPP $\{\mathbf{y}^*_{\mathrm{general}}, \mathbf{y}^{'}_{\mathrm{poe}}\}$, UnPP $\{\mathbf{y}^*_{\mathrm{general}}, \hat{\mathbf{y}}_{\mathrm{poe}}\}$, and PSen $\{\mathbf{y}^-_{\mathrm{general}}, \mathbf{y}_{\mathrm{poe}}\}$ refer to Controllable Pseudo-Poems corpus, Uncontrollable Pseudo-Poems corpus, and Pseudo-Sentences corpus respectively. These corpora are built by iterative back-translation.

categorizing them based on style and instead collected a total of 290K and 380K stanzas, respectively. The primary sources of data for this compilation were online platforms, including forums[6][7], website[8][9], and other online resources[10][11].

Table 1 shows the statics of pseudo-parallel data used to train style injection model θ_{style}, including BT data and several IBT data variants.

- **Controllable Pseudo-Poems (ConPP)**: The synthetic parallel corpus $\{\mathbf{y}^*_{\mathrm{general}}, \mathbf{y}^{'}_{\mathrm{poe}}\}$ is built by applying iterative translation to translate the $\mathbf{y}^*_{\mathrm{general}}$ by prepending a "[number]" token before the $\mathbf{y}_{\mathrm{poe}}$, which eventually produces a new synthetic parallel corpus as: $\mathbf{y}^*_{\mathrm{general}} \rightarrow \mathbf{y}^{'}_{\mathrm{poe}}$.
- **Uncontrollable Pseudo-Poems (UnPP)**: The synthetic parallel corpus $\{\mathbf{y}^*_{\mathrm{general}}, \hat{\mathbf{y}}_{\mathrm{poe}}\}$ is built by applying iterative translation to translate the $\mathbf{y}^*_{\mathrm{general}}$ without prepending the "[number] token before the $\hat{\mathbf{y}}_{\mathrm{poe}}$.
- **Pseudo-Sentences (PSen)**: The synthetic parallel corpus $\{\mathbf{y}^-_{\mathrm{general}}, \mathbf{y}_{\mathrm{poe}}\}$ is built by applying iterative translation to translate the $\mathbf{y}_{\mathrm{poe}}$ into the $\mathbf{y}^-_{\mathrm{general}}$ in the IBT process, which can be formulated as: $\mathbf{y}_{\mathrm{poe}} \rightarrow \mathbf{y}^-_{\mathrm{general}}$.

To create the test sets for both the one-stage and two-stage models, a random selection of 200 English poems (comprising 876 stanzas in total) was made from the English-Chinese Poetry Translation Website[12]. The test set of Portuguese poems and English poems translated

[6] http://www.chinawriter.com.cn/
[7] https://poets.org/poems
[8] http://www.zgshige.com/
[9] https://allpoetry.com/
[10] http://vvchem.com/
[11] https://www.luso-poemas.net/
[12] www.poetrysky.com

from English and Chinese respectively, are randomly selected 100 stanzas in total. As for the training of the various style injection models, the aforementioned test set was employed, with a random selection of 3,000 stanzas of modern Chinese poetry, 1,000 stanzas of English and Portuguese poetry serving as the validation set.

3.2 Model Training

Transformer architecture Vaswani et al. (2017) implemented by *fairseq* toolkit[13] is used to train the in-house NMT system, style injection and iterative back-translation models with shared vocabulary of 30K BPE Sennrich et al. (2015) tokens. Both the encoder and decoder block consist of 6 layers with 8 attention heads. The embedding size and hidden size are set to 512. We train all the models with a learning rate of 3e-5 and use 16K tokens per batch.

3.3 Evaluation

As for the automatic evaluation, we use BLEU and BERTScore to evaluate the performance of different MT systems. Furthermore, the evaluation method BERTScore calculates a sentence's score depending on the contextualized embedding similarity of references and system results, which could better understand each token's meaning during the sentence than the BLEU method, and more similar to the human evaluation score Zhang et al. (2020a). We report the recall score (BS.R), precision score (BS.P), and F1 (BS.F) score of the BERTScore respectively. For human evaluation, given the absence of any preceding research on evaluation metrics for poetry translation, our study relied upon human evaluation as a way to acquire more objective and authentic evaluations. We invited experts in the field of modern Chinese poetry to modify the evaluation framework proposed by Yi et al. (2018), and proposed a new evaluation framework that comprises five distinct perspectives:

- **Poeticity**: The translated poem exhibits a structure and poetic quality that is consistent with the poetry style of the target language.
- **Fluency**: The translated poem employs the diction and grammar that are characteristic of the poetry in the target language.
- **Coherence**: The meaning translated in the content of the translated poem is equivalent to that of the original poem in the source language.
- **Meaningfulness**: Translating poetry entails conveying a significant meaning and message.
- **Overall Impression**: For the overall impression score of the translated poem.

We randomly selected 30 Chinese poems and then asked four experts in the field of poetry to evaluate four types of translation generated by humans, the online system, "BT" model, and "IBT:ALL" model, respectively. The evaluation was conducted in a blind way, i.e., the experts did not know the type of translation during the evaluation process.

4 Main Results

Table 2 illustrates that the injection of poetic style using the model trained with all the IBT variants, following the utilization of the online translation system, outperforms the one-stage and other two-stage baselines. This superiority is further confirmed by the results of the human evaluation presented in Table 3, despite all the automatic translation methods being unable to surpass human translation. The automatic evaluation and human evaluation results serve as evidence for the efficacy of the two-stage translation approach, with particular emphasis on the effectiveness of the style injection model compare to the online system.

[13]https://github.com/pytorch/fairseq/

	English ⇒ Modern Chinese				Chinese ⇒ English				English ⇒ Portuguese			
	BLEU	**BS.P**	**BS.R**	**BS.F**	**BLEU**	**BS.P**	**BS.R**	**BS.F**	**BLEU**	**BS.P**	**BS.R**	**BS.F**
One-stage Neural Poetry Translation												
Pre-trained Transformer	7.62	71.4	70.1	70.6	11.46	81.6	81.2	81.3	12.82	74.8	75.3	75.0
Baidu Online system	14.23	74.6	74.7	74.5	18.87	83.5	83.5	83.4	18.62	77.0	77.8	77.4
Two-stage Neural Poetry Translation												
BT	11.69	76.4	76.1	76.2	19.05	85.5	85.0	85.2	18.64	78.0	78.6	78.2
IBT:All	14.04	**78.3**	**77.7**	**77.9**	19.34	**86.5**	**85.8**	**86.1**	19.04	78.7	78.7	78.5
IBT:PSen	12.08	76.7	76.4	76.5	**19.54**	86.2	85.7	86.0	17.70	**79.0**	**79.1**	**78.9**
IBT:ConPP	12.29	76.8	76.6	76.6	18.28	85.4	85.1	85.2	**19.72**	78.9	79.0	78.8
IBT:UnPP	**14.35**	78.2	77.6	**77.9**	17.93	86.2	85.7	86.0	18.81	78.3	78.6	78.3

Table 2: Statistics of the pseudo-parallel corpora used to train style injection model. The BT corpus $\{\mathbf{y}^*_{\text{general}}, \mathbf{y}_{\text{poe}}\}$ is built by back-translation. The ConPP $\{\mathbf{y}^*_{\text{general}}, \mathbf{y}'_{\text{poe}}\}$, UnPP $\{\mathbf{y}^*_{\text{general}}, \hat{\mathbf{y}}_{\text{poe}}\}$, and PSen $\{\mathbf{y}^-_{\text{general}}, \mathbf{y}_{\text{poe}}\}$ refer to Controllable Pseudo-Poems corpus, Uncontrollable Pseudo-Poems corpus, and Pseudo-Sentences corpus respectively. These corpora are built by iterative back-translation. The IBT:ALL refers train style injection model by using all corpus.

5 Analysis

As shown in Table 2, regarding the second stage of neural poetry translation, it was observed that the scores achieved for the English and Chinese poems were higher than those attained for the Portuguese poems. This may be attributed to the fact that the stylistic obviously differences between the Chinese and English poems were more pronounced. Consequently, the stylistic injected in these poems became more evident. Nonetheless, it is important to take into consideration that the golden poems were translated by a translator who may not possess a thorough understanding of the poetic style in the target language. As such, we speculate that better results can be achieved in terms of transferring poetic style by leveraging the additional Chinese poetry that has been translated from other languages during the training process.

	Poeticity	**Fluency**	**Coherence**	**Meaningfulness**	**Impression**
Human Translation	**3.54**	**3.76**	**3.79**	**3.65**	**3.69**
Baidu Online System	2.95	3.06	3.05	2.91	3.00
BT	3.15	3.17	3.23	<u>3.21</u>	3.21
IBT:ALL	<u>3.17</u>	<u>3.18</u>	<u>3.25</u>	3.16	<u>3.24</u>

Table 3: Human evaluation of the selected 30 poems (English⇒Modern Chinese). **Bold** values denote the highest scores of each evaluation perspective, and <u>underlined</u> values denote the highest scores of each evaluation perspective among the neural poetry translation systems.

Based on the results of the human evaluation, Table 3 demonstrates that the models' capacity to translate modern Chinese poetry exceeds that of the most advanced online translation systems. The "BT" model stands out for its superior capacity to conserve the meaning of poetry, while the "IBT:ALL" model particularly excels in preserving the poetic nature of the text. This may be attributed to the continuous superimposition of poetic imagery during the process of data enhancement, which enables the "IBT:ALL" model to effectively capture the poeticity of the text. Conversely, the process of data enhancement may introduce incomplete or even incorrect meanings, which can impact the ability of the "IBT:ALL" model to preserve the meaning of poetry during translation, thus rendering it inferior to the "BT" model in this regard.

6 Conclusions and Future Work

This paper introduces a zero-shot poetry translation method, which reduces the difficulty of the whole poetry translation by splitting the process of translating the meaning of the poem and translating the poetic meaning in poetry translation. The proposed method also avoids the disadvantage of lack of parallel poetry corpus and reduces the cost of training. The experimental results, derived from both automatic and human evaluation, provide evidence of the efficacy of our proposed method. In the poetry translation, the model achieved superior outcomes in comparison to contemporary online systems. Moreover, the proposed technique was evaluated by humans and found to inject poetic meaning into the translated poems, thereby bringing them closer to the standards of human translation and aligning more closely with human expectations for poetry. Moving forward, our future work will focus on exploring the potential of our proposed method on more languages.

Acknowledgement

This work was supported in part by the Science and Technology Development Fund, Macau SAR (Grant Nos. FDCT/0070/2022/AMJ, FDCT/060/2022/AFJ), the Multi-year Research Grant from the University of Macau (Grant No. MYRG2020-00054-FST), and the Research Program of Guangdong Province (Grant No. 2220004002576). This work was performed in part at SICC which is supported by SKL-IOTSC, and HPCC supported by ICTO of the University of Macau.

References

Andersson, L. M. and Pearson, C. M. (1999). Tit for tat? the spiraling effect of incivility in the workplace. *Academy of management review*, 24(3):452–471.

Argamon, S., Koppel, M., Fine, J., and Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text & Talk*, 23(3):321–346.

Brown, P., Levinson, S. C., and Levinson, S. C. (1987). *Politeness: Some universals in language usage*, volume 4. Cambridge university press.

Chakrabarty, T., Saakyan, A., and Muresan, S. (2021). Don't go far off: An empirical study on neural poetry translation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7253–7265, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Chhaya, N., Chawla, K., Goyal, T., Chanda, P., and Singh, J. (2018). Frustrated, polite, or formal: Quantifying feelings and tone in email. In *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pages 76–86.

Dewdney, N., Van Ess-Dykema, C., and MacMillan, R. (2001). The form is the substance: Classification of genres in text. In *Proceedings of the ACL 2001 Workshop on Human Language Technology and Knowledge Management*.

Genzel, D., Uszkoreit, J., and Och, F. (2010). "poetic" statistical machine translation: Rhyme and meter. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 158–166, Cambridge, MA. Association for Computational Linguistics.

Ghazvininejad, M., Choi, Y., and Knight, K. (2018). Neural poetry translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 67–71, New Orleans, Louisiana. Association for Computational Linguistics.

Heylighen, F. and Dewaele, J.-M. (1999). Formality of language: definition, measurement and behavioral determinants. *Interner Bericht, Center "Leo Apostel", Vrije Universiteit Brüssel*, 4.

Hoang, V. C. D., Koehn, P., Haffari, G., and Cohn, T. (2018). Iterative back-translation for neural machine translation. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, pages 18–24, Melbourne, Australia. Association for Computational Linguistics.

Huang, Y., Zhu, W., Xiong, D., Zhang, Y., Hu, C., and Xu, F. (2020). Cycle-consistent adversarial autoencoders for unsupervised text style transfer. *arXiv preprint arXiv:2010.00735*.

Koehn, P. and Knowles, R. (2017). Six challenges for neural machine translation. In *Proceedings of the First Workshop on Neural Machine Translation*, pages 28–39, Vancouver. Association for Computational Linguistics.

Li, J., Jia, R., He, H., and Liang, P. (2018). Delete, retrieve, generate: A simple approach to sentiment and style transfer. *arXiv preprint arXiv:1804.06437*.

Li, J., Li, Z., Mou, L., Jiang, X., Lyu, M. R., and King, I. (2020). Unsupervised text generation by learning from search. *arXiv preprint arXiv:2007.08557*.

Liu, D., Fu, J., Zhang, Y., Pal, C., and Lv, J. (2020). Revision in continuous space: Unsupervised text style transfer without adversarial learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8376–8383.

Liu, Z., Fu, Z., Cao, J., de Melo, G., Tam, Y.-C., Niu, C., and Zhou, J. (2019). Rhetorically controlled encoder-decoder for modern chinese poetry generation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1992–2001.

Malmi, E., Severyn, A., and Rothe, S. (2020). Unsupervised text style transfer with padded masked language models. *arXiv preprint arXiv:2010.01054*.

Pavlopoulos, J., Thain, N., Dixon, L., and Androutsopoulos, I. (2019). Convai at semeval-2019 task 6: Offensive language identification and categorization with perspective and bert. In *Proceedings of the 13th international Workshop on Semantic Evaluation*, pages 571–576.

Peersman, C., Daelemans, W., Vandekerckhove, R., Vandekerckhove, B., and Van Vaerenbergh, L. (2016). The effects of age, gender and region on non-standard linguistic variation in online social networks. *arXiv preprint arXiv:1601.02431*.

Pennebaker, J. W., Mehl, M. R., and Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, 54(1):547–577.

Rabinovich, E., Patel, R. N., Mirkin, S., Specia, L., and Wintner, S. (2017). Personalized machine translation: Preserving original author traits. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1074–1084, Valencia, Spain. Association for Computational Linguistics.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.

Sennrich, R., Haddow, B., and Birch, A. (2015). Neural machine translation of rare words with subword units. *arXiv preprint arXiv:1508.07909*.

Sheikha, F. A. and Inkpen, D. (2010). Automatic classification of documents by formality. In *Proceedings of the 6th international conference on natural language processing and knowledge engineering (nlpke-2010)*, pages 1–5. IEEE.

Shen, L., Guo, X., and Chen, M. (2020). Compose like humans: Jointly improving the coherence and novelty for modern chinese poetry generation. In *2020 International Joint Conference on Neural Networks, IJCNN 2020, Glasgow, United Kingdom, July 19-24, 2020*, pages 1–8. IEEE.

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

Susanto, Y., Livingstone, A. G., Ng, B. C., and Cambria, E. (2020). The hourglass model revisited. *IEEE Intelligent Systems*, 35(5):96–102.

Toshevska, M. and Gievska, S. (2021). A review of text style transfer using deep learning. *IEEE Transactions on Artificial Intelligence*, page 1–1.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., von Luxburg, U., Bengio, S., Wallach, H. M., Fergus, R., Vishwanathan, S. V. N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008.

Yang, Z., Cai, P., Feng, Y., Li, F., Feng, W., Chiu, E. S.-Y., and Yu, H. (2019). Generating classical Chinese poems from vernacular Chinese. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6155–6164, Hong Kong, China. Association for Computational Linguistics.

Yi, X., Sun, M., Li, R., and Yang, Z. (2018). Chinese poetry generation with a working memory model. *arXiv preprint arXiv:1809.04306*.

Yu, M. and Liu, C. (2021). "creation" and "production": Reflections on microsoft xiaobing's poetry writing software. *Modern Literary Magazine (05) p.158-165*.

Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N., and Kumar, R. (2019). Predicting the type and target of offensive posts in social media. *arXiv preprint arXiv:1902.09666*.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., and Artzi, Y. (2020a). Bertscore: Evaluating text generation with BERT. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

Zhang, Y., Ge, T., and Sun, X. (2020b). Parallel data augmentation for formality style transfer. *arXiv preprint arXiv:2005.07522*.

Zhang, Y., Xu, J., Yang, P., and Sun, X. (2018). Learning sentiment memories for sentiment modification without parallel data. *arXiv preprint arXiv:1808.07311*.