# SINGABLE AND CONTROLLABLE NEURAL LYRIC TRANSLATION: A LATE-BREAKING SHOWCASE

**Longshen Ou   Xichu Ma   Ye Wang**
School of Computing, National University of Singapore

## ABSTRACT

The development of general-domain neural machine translation (NMT) methods has advanced significantly in recent years, but the lack of naturalness and musical constraints in the outputs makes them unable to produce singable lyric translations. This paper bridges the singability quality gap by formalizing lyric translation into a constrained translation problem, converting theoretical guidance and practical techniques from translatology literature to prompt-driven NMT approaches, exploring better adaptation methods. We instantiate our approach to an English-Chinese lyric translation system, that achieves high control effectiveness without sacrificing text quality. In our subjective evaluation, our model shows a 75% relative enhancement in overall quality compared to naive fine-tuning.

## 1. INTRODUCTION

In the age of entertainment globalization, the appreciation of foreign language songs and artists' global reach is on the rise. Yet, the challenge remains: most commercial songs lack multilingual versions, and existing translations often neglect musical constraints, making them unsingable. This language barrier hinders the full potential of the music industry. Song translation, with its unique requirements, poses challenges even for skilled human translators due to music constraints and style requirements. Suppose we can construct lyric-specific machine translation (MT) systems to produce drafts that satisfy these constraints and requirements. In this case, the difficulty and cost of lyric translation will be vastly reduced, as lyricists and translators can start with such automatic drafts and focus on post-processing for creativity.

## 2. CHALLENGES

However, obtaining singable lyrics from MT systems is challenging. Figure 1 shows two sentences of lyrics from the song *Let It Go*, together with an MT output and a singable translation. We observe a notable quality gap between them. While the MT output correctly translates the source, it ignores all the criteria that matter to make the output singable: (1) The second sentence of the MT outputs is unnatural because of incoherent vocabulary selection and lack of aesthetics. (2) Overcrowded syllables in the first sentence of the MT outputs force performers to break music notes into pieces, diverging the intended rhythm pattern. (3) The two-syllable word in the red box is situated across a musical pause (blue box), causing an unnatural pronunciation. (4) The end-syllables (purple text) are not of the same rhyme pattern, making the output miss a key chance for being poetic.

## 3. APPROACH

### 3.1 Problem Formulation

To achieve a comprehensive and language-independent method, we define "singable translation" as following the "Pentathlon Principle" from [1]: that quality, singable translations are obtained by balancing five aspects—singability, rhythm, rhyme, naturalness, and sense. Table 1 lists these aspects, corresponding requirements, and how we actualize them in our model. Particularly, we identify (1)–(3) as the controlling aspects of our model and realize them with prompt-based control, while (4) and (5) are achieved from the perspectives of adaptation and pretraining.

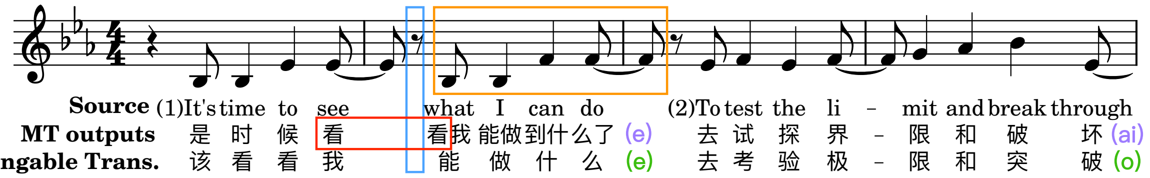
Source (1)It's time to see what I can do (2)To test the li - mit and break through
MT outputs 是 时 候 看 看我能做到什么了 (e) 去 试 探 界 - 限 和 破 坏 (ai)
Singable Trans. 该 看 看 我 能 做 什 么 (e) 去 考 验 极 - 限 和 突 破 (o)

**Figure 1**: Translation comparison of a general-domain NMT system (2nd row), already been adapted with parallel lyric data, versus a singable translation (3rd row).

| Aspects | Requirements | Our Actualization |
|---|---|---|
| (1) Singability | Outputs are suitable for singing with the given melodies. | Enhance music-lyric compatibility by prompt-based necessary word boundary control. |
| (2) Rhythm | Outputs follow rhythm patterns in the music. | Prompt-based length (number of syllables) control. |
| (3) Rhyme | Outputs fulfil certain rhyme patterns. | Prompt-based end-rhyme control and paragraph-level rhyme ranking. |
| (4) Naturalness | Outputs read like lyrics originally composed in the target language. | Adapting with back-translation of in-domain target-side monolingual data. |
| (5) Sense | Outputs are fidelity to the meaning of source sentences. | Large-scale general-domain pretraining. |

**Table 1**: The "Pentathlon Principle" and the actualizations in our model.

### 3.2 Prompt-Based Control

To implement prompt-based control for aspects (1)–(3), we introduce special tokens for output property control: $l_{\text{tgt}}$, $r_{\text{tgt}}$, and $\mathbf{b}_{\text{tgt}}$, for desired syllable count, end-rhyme type, and word boundaries. During training, these prompts, derived from target-side sentences, serve as additional inputs to guide the generation. For inference, prompts originate from music or source sentences. For system workflow, please refer to Figures 2b and 2c.

### 3.3 Word Boundary Control

We enhance music–lyric compatibility to refine the singability (1). We noticed that for multi-syllable words, intra-word musical rests tend to reduce pronunciation acceptability [2, 3]. Abruptly highlighted syllables are also observed to have similar adverse effects due to abrupt pitch and tension shifts. We address this by ensuring *word boundaries* (points between syllables from different words) align with *melody boundaries*, positions *at* musical pauses and *before* highlighted notes (e.g., downbeats). In our model, such alignment is achieved by a list of special tokens $\mathbf{b}_{\text{tgt}}$, indicating whether a word boundary is necessary or optional after each syllable.

### 3.4 Reverse-Order Decoding

To achieve better rhyme control (3), we train our model to generate the translation from the last word to the first word for each sentence, inspired by human translators' rhyming technique [1]. This method fixes important words, the word that should be in rhyme first, and hence secures the rhyme control. This reverse-order paradigm also aids in rhyme ranking because we can treat the end-word probability as a greedy prediction of sentence quality. By aggregating the probability according to rhyme category and observing the rhyme distribution, we can pick the most suitable rhyme for each sentence and, further, for a paragraph containing multiple sentences.

### 3.5 Back Translation

Naturalness (4) and sense (5) can be guarenteed when large-scale datasets available, but community-contributed parallel lyric data is not only limited in scale but also suffers from quality issues [4–6]. To overcome these problems, we seek help from target-side monolingual lyric data. We incorporate back-translation (BT) [7], i.e., adopt existing translation systems in the backward direction to translate the target-side lyric to the source language, to obtain synthesized parallel data that augments the limited parallel data.

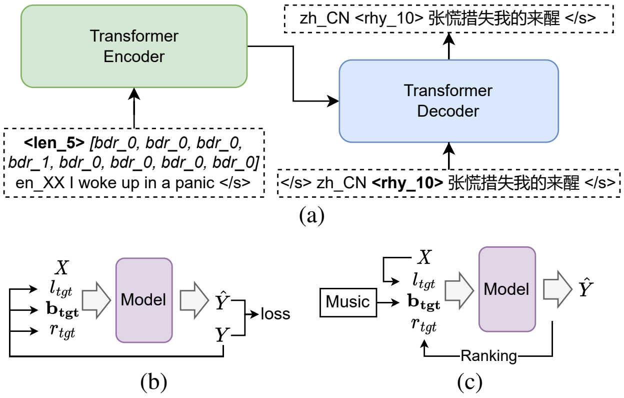
Transformer Encoder
Transformer Decoder
zh_CN <rhy_10> 张慌措失我的来醒 </s>
<len_5> [bdr_0, bdr_0, bdr_0, bdr_1, bdr_0, bdr_0, bdr_0, bdr_0] en_XX I woke up in a panic </s>
</s> zh_CN <rhy_10> 张慌措失我的来醒 </s>
(a)
X
$l_{tgt}$
$\mathbf{b_{tgt}}$
$r_{tgt}$
Model
$\hat{Y}$
Y
loss
(b)
Music
X
$l_{tgt}$
$\mathbf{b_{tgt}}$
$r_{tgt}$
Model
$\hat{Y}$
Ranking
(c)

**Figure 2**: (a): Structure of our English-to-Chinese lyric translation system. (b): Workflow of the fine-tuning stage. (c) Workflow of the inference stage.

## 4. RESULTS

Taken together, these innovations form our final lyric translation method [8]. In the evaluation, we instantiate our techniques with Multilingual BART (refer to Figure 2 for structure and workflow), producing the Singable Translation (Row 3) in Figure 1. [1]

In the objective evaluation, our model consistently surpasses the baseline without BT and control across all metrics, achieving superior BLEU and TER scores, and near-perfect control accuracies for length, rhyme, and word boundaries. [2] This proficiency indicates our model's capability to generate singable lyric translations without compromising text quality. Furthermore, in human assessments, our model excels over both the baseline and GagaST [3], leading by 75.0% and 20.2%, respectively, in the overall singable translation quality metric. Importantly, our model demonstrates a marked improvement in music-lyric compatibility (+74.7% and +10.2%).

## 5. CONCLUSION

In this paper, we presented novel methods to improve the singability and fluency of translated lyrics. We efficiently managed word boundaries using prompt-based techniques and strengthened rhyme control with reverse-order decoding. We further optimized stanza rhyming through our rhyme ranking mechanism and enhanced sense and naturalness with back-translation. These methods blend linguistic accuracy with musical harmony for lyric translation.

[1] Additional case studies and demos are featured in `https://www.oulongshen.xyz/lyric_translation`.

[2] Please refer to the paper [8] for detailed objective results.

## 6. REFERENCES

[1] P. Low, “Singable translations of songs,” *Perspectives: Studies in Translatology*, vol. 11, no. 2, pp. 87–103, 2003.

[2] H. Franco, L. Neumeyer, and H. Bratt, “Modeling intra-word pauses in pronunciation scoring,” in *STiLL-Speech Technology in Language Learning*, 1998.

[3] F. Guo, C. Zhang, Z. Zhang, Q. He, K. Zhang, J. Xie, and J. Boyd-Graber, “Automatic song translation for tonal languages,” in *Findings of the Association for Computational Linguistics: ACL 2022*. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 729–743. [Online]. Available: https://aclanthology.org/2022.findings-acl.60

[4] Y. Li, “英文歌词翻译存在的问题及应遵循原则,” 山西青年, 2020.

[5] Y. Zhang, “英文歌词文言文翻译中的创造性叛逆问题分析,” 英语广场, 2022.

[6] H. Xie and Q. Lei, “归化异化视角下线上音乐平台歌词翻译分析,” 海外英语, 2022.

[7] R. Sennrich, B. Haddow, and A. Birch, “Improving neural machine translation models with monolingual data,” in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 86–96. [Online]. Available: https://aclanthology.org/P16-1009

[8] L. Ou, X. Ma, M.-Y. Kan, and Y. Wang, “Songs across borders: Singable and controllable neural lyric translation,” in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 447–467. [Online]. Available: https://aclanthology.org/2023.acl-long.27