Lyrics Generation Applying Metaphor Generation

Kazuki Murakami and **Asuka Terai**
Future University / 116-2 Kamedanakano, Hakodate Hokkaido, Japan
g2122068@fun.ac.jp and aterai@fun.ac.jp

Abstract

The development of deep-learning technology has improved the quality of several tasks in the field of natural language processing. Examples of these tasks include response, dialogue, and text-generation systems. However, some contributions can be made to the literature. In particular, there are challenges in text generation in the music genre. In this study, we propose a lyrics generation system that considers topics and includes metaphors. First, we analyzed the features of a corpus of Japanese lyrics based on the number of moras, Latent Dirichlet Allocation topic estimation, and semantic textual similarity using sentence-BERT(S-BERT). We thereafter propose a system based on the results of the analysis. In this system, succeeding lyrics are generated as input to the preceding lyrics. A figurative verb was generated as a noun in the extracted template lyrics. The template verb was replaced with a figurative verb and the replaced template was complemented by a pretrained BERT mask language model fine-tuned with the lyrics corpus to generate lyrics. The system outputs lyrics with metaphors that reflect the characteristics of the actual lyrics and their coherence with topics in the input lyrics.

1 Introduction

Natural language processing (NLP) techniques have evolved rapidly over the last decade. NLP involves various tasks such as machine translation, summerization, dialogue systems, question answering systems, and text generation. Presently, general-purpose language modeling tools such as GPT and BERT are available to everyone over the web to tackle these tasks. Among these tasks, text generation has garnered the attention of several researchers. Generating texts of comparable quality to that of humans is also an important independent task.

Several studies have been conducted on lyric-generating systems for text generation in music. Zhang et al.(Zhang et al., 2020) proposed a user-interactive lyrics generation system. Qian et al. (Qian et al., 2023) proposed a lyrics generation system that incorporates text and music information. Some lyrics generation systems limited to raps and based on transformers were also proposed by Xue et al.(Xue et al., 2021) and Nikolov et al.(Nikolov et al., 2020). However, Rodrigues et al. (Rodrigues et al., 2022) demonstrated the feasibility of using generative models for English and Portuguese lyrics generation by fine-tuning the GPT-2. Conversely, the problem has been reported that some phrases generated by the system have lost their meaning. This contextual collapse has been frequently reported in GPT-2.

Watanabe et al.(Watanabe et al., 2014) suggested the importance of maintaining topic coherence in lyrics generation. For instance, lyrics often include the lyricist's assertion, and sometimes include a topic that represents the entire work such as "love" or "dreams." Thus, the topic is an important factor in lyrics generation. Wang and Zhao (Wang and Zhao, 2019) improved the connectivity and coherence of lyrical topics using generative models. A multichannel seq2seq model with attention addresses this problem. To create an appropriate corpus, the lyrics' topics were extracted using topic classification. We studied Chinese song lyrics and made significant improvements to them.

In contrast, the importance of metaphorical expressions in lyrics has also been highlighted. Kagita et al.(Kagita et al., 2013) analyzed real lyrics and found that several types of lyrics exist. Among them, figurative terms such as simile and "uniqueness of co-occurred terms" were found to be more memorable. The term "uniqueness of co-occurred terms" refers to the use of a combination of nouns and verbs that do not normally co-occur. These can be regarded as "verb-metaphor." Verb-metaphors are the targets of metaphor generation. Additionally, few stud-

ies have proposed a computational approach for generating metaphors in Japanese, and their targets are metonymy and similes. For instance, Takahashi and Ohshima(Takahashi and Ohshima, 2019) demonstrated the effectiveness of expressing metonymy (personification metaphors) using a Kronecker product. Mitsuishi and Shimada(Mitsuishi and Shimada, 2019) created a dataset of similes in Japanese, and proposed a neural network system to automatically generate novel metaphorical expressions. However, few studies have focused on metaphors in lyrics generation, although it is not a study aimed at lyrics generation, as one of these types of metaphor generation methods, Umemura and Kano (Umemura and Kano, 2021) presented a model for a metaphor generator based on the concept metaphor theory (Lawler, 1983) using templates and Japanese case frames (Daisuke and Sadao, 2005). For machine learning, the authors considered that there were insufficient learning data for catchphrases and adopted a rule-based method. The results exhibit superior performance compared to the rule-based method based on parsing (Iwama and Kano, 2018).

In this study, we propose a lyrics generation system that maintains topic coherence and applies metaphorical generation. We also use one of these models based on a transformer, more specifically BERT, to generate natural language for song lyrics to address the contextual collapse problem.

2 Lyrics Analysis

To propose a system that reflects existing lyrics' characteristics, lyrics analysis was performed.

2.1 Methods

2.1.1 Lyrics Corpus Collecting

We extracted Japanese lyrics from the Uta-net website[1] as a Japanese lyrics corpus. The Japanese songs were recorded by omitting duplicates. In Table 1, each lyric sample is preprocessed into a pair of preceding and succeeding sentences, and thereafter subjected to data cleaning. Lyric pairs were collected using the following procedure:

(I) Lyrics comprising only stop words must be omitted. For instance, the symbols (, !, etc...) and lyrics like "La La La" have been removed.

(II) Saving lyrics with three or more characters.

(III) Duplicate sentences were removed.

(IV) In the case of songs that contained languages other than Japanese, only parts in which the Japanese sentences were continuous were extracted.

Ultimately, 5.12 million Japanese lyrics pairs were collected. After applying this rule, depending on the results of the topic classification described later, the lyrics were divided into a corpus of lyrics for each topic. The divided lyrics corpus was used for fine-tuning BERT.

2.1.2 Mora and sementic similarity

The collected lyrics were analyzed from the perspectives of mora and semantic textual similarity. A mora is a segmental unit of sound with temporal length. Unlike syllables, which are defined by phonological structures, they depend on the length of the sounds within each language. Therefore, the moras differ among languages. In Japanese, one kana character is essentially one mora. Vowels were counted in the analysis, because 1 Japanese kana character contains one vowel. The distribution of moras in all lyrics was investigated.

Semantic textual similarity was determined by the relationship between the preceding and succeeding lyrics. In this study, the cosine distance is adopted to calculate the semantic textual similarity. Sentence-BERT(Reimers and Gurevych, 2019) was used to calculate the semantic textual similarity.

2.1.3 Topic Classification

Topic classification is a technique for predicting and classifying document topics based on words that appear in a document. Latent Dirichlet Allocation (LDA) has been used for topic classification (Blei et al., 2003). LDA classifies topics based on the distinctiveness of words. There are various methods for calculating the distinctiveness of a word. Among them, term saliency (Chuang et al., 2012) has been applied. The distinctiveness of word w in topic T is defined as follows:

$$distinctiveness(w) = \sum_{T} P(T|w) \log \frac{P(T|w)}{P(T)}$$

Stop-word lists were used to create document feature vectors. For the stop-word list, Slothlib(Ohshima et al., 2007) stop-words were adopted. Slothlib stop-words can be used to remove symbols and stop-words such as 'a,' 'an,' and 'the.'

[1] https://www.uta-net.com/

Preceding Lyric	Succeeding Lyric
流れるまんま流されたら (If we are swept away by the flow)	抗おうか美しい鰭で （Shall we resist with our beautiful fins?）

Table 1: Lyrics pair example from “美しい鰭” by スピッツ

2.2 Results

2.2.1 Mora and semantic textual similarity

The analysis result of the mora of the lyrics corpus is shown in Fig.1,2,3.

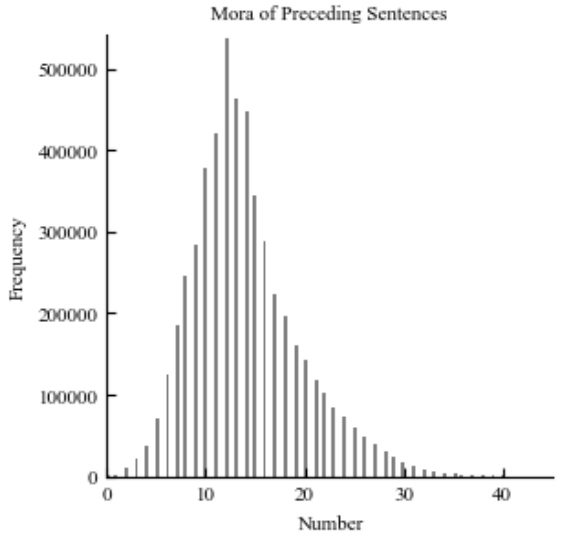

Figure 1: Distribution of mora of the preceding sentence of the lyric pair.

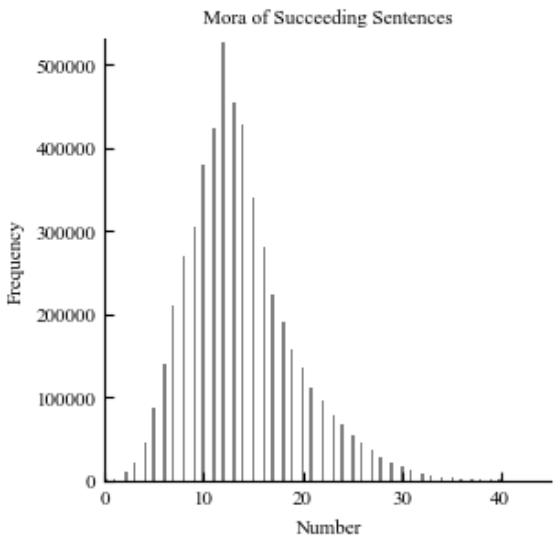

Figure 2: Distribution of mora of the succeeding sentence of the lyric pair.

Fig.1 and 2 show that the lyrics mora were concentrated at approximately 13 with a high frequency, regardless of whether the lyrics preceded or succeeded. Successful sentences tended to occur slightly more frequently. Fig.3 shows that the difference in moras between the preceding and succeeding lyrics was concentrated at zero, and the difference tended to be small. Based on these facts, it can be said that the smaller the difference, the closer the mora is to the real lyrics. It can also be said that the more succeeding moras, the closer they are to the real lyrics.

Fig.4 shows that the cosine distance between all lyrics data pairs was approximately 0.706, on average, at 95 % confidence. The range of values was adjusted such that the closer it was to 0, the closer the semantic distance, and the closer it was to 2, the farther away.

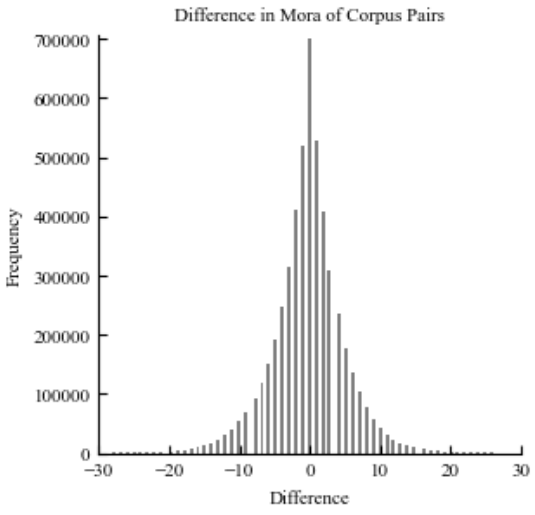

Figure 3: Difference of mora between the lyric pair.

2.2.2 Topic Classification

Fig.5 shows that fixing the number of topics to six was the best because the perplexity was low

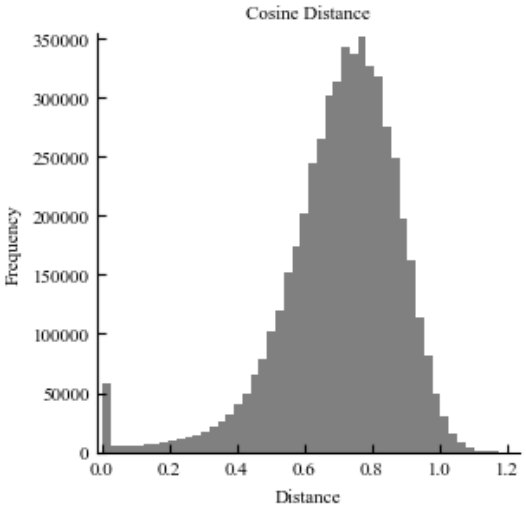

Figure 4: Cosine distance between the lyric pair.

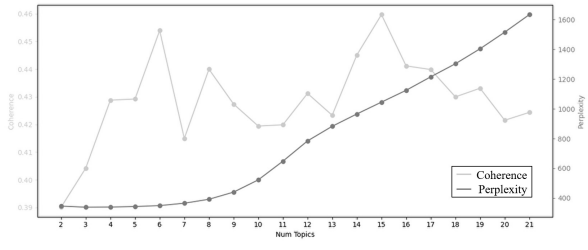

Figure 5: Evaluation of LDA topic classification by perplexity and coherence metrics

and the coherence was high. Then, the number of topics was fixed at six, and all collected Japanese lyrics were classified. The classification results were used as a criterion for template selection. The lyrics corpus was then divided into six topics based on LDA results.

Fig.6 shows the frequency distribution by topic for the entire lyrics corpus. Fig.7 shows the mean value of the cosine distance for each topic. Table 2 summarizes the results of the word-by-topic classification.

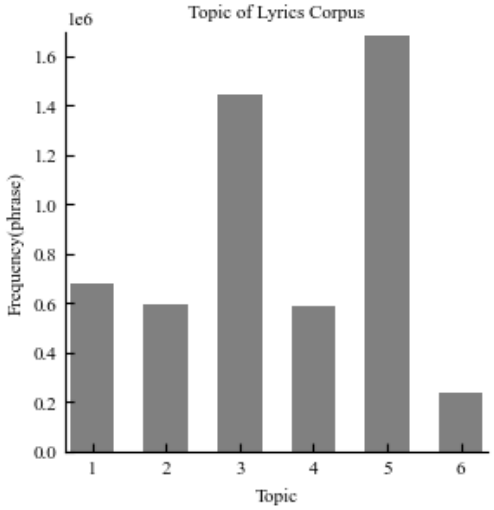

Figure 6: Topic classification results for lyrics corpus by LDA. The vertical axis shows the number of words belonging to each topic.

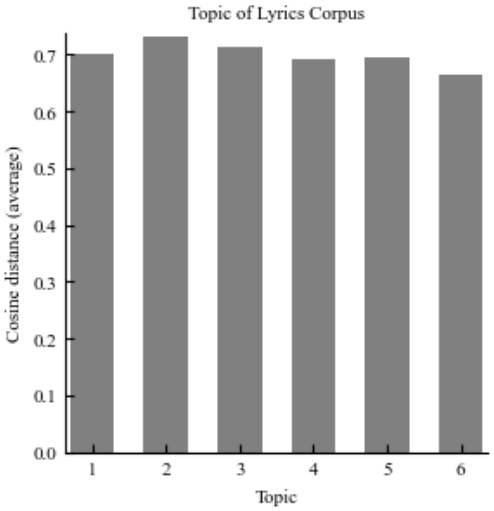

Figure 7: Mean value of the cosine distance for each topic.

Rank	Topic1	Topic2	Topic3	Topic4	Topic5	Topic6
1	君(You)	夢(Dream)	世界(World)	恋(Love)	お前(You)	キミ(You)
2	僕(I)	空(Sky)	生き(Live)	夜(Night)	今日(Today)	恋(Love)
3	好き(Like)	心(Heart)	人間(Human)	花(Flower)	人生(Life)	好き(Like)
4	心(Heart)	風(Wind)	全て(All)	夢(Dream)	明日(Tomorrow)	魔法(Magic)
5	言葉(Words)	明日(Tomorrow)	無い(Nothing)	海(Sea)	早く(Quickly)	ハト(Pigeon)

Table 2: Top five keywords in each topic

3 Lyrics Generation

We proposed a lyrics generation system based on the results of lyrics analysis. Fig.8 shows the configuration of the system. The system generates succeeding lyrics as input to the preceding lyrics.

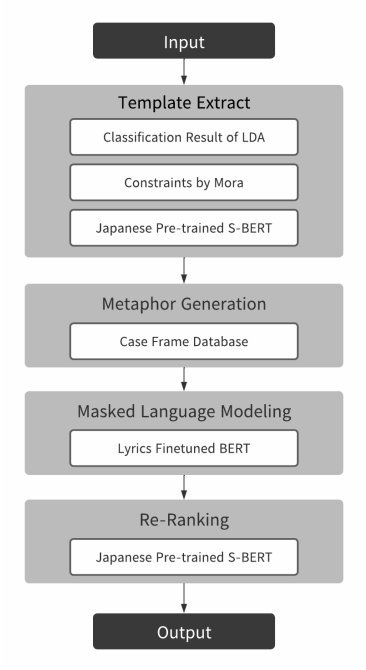

Figure 8: Configuration of the system.

3.1 Methods

3.1.1 Template Extract

The template was extracted from a Japanese lyric corpus. Template extraction was performed using the semantic textual similarity of Sentence-BERT (S-BERT)(Reimers and Gurevych, 2019). S-BERT is a modification of the pretrained BERT that uses new learning methods to calculate semantic sentence embedding, which can be compared using cosine distance. S-BERT reduces the effort required to determine the pair that is most similar to BERT. The Japanese pretrained S-BERT was used.

Based on the results of the mora in the lyrics analysis, the constraint is that the mora difference from the input sentence is up to one. At 95 % confidence intervals, the mean difference is approximately from -0.25 to -0.24. A mora is an integer value set to 1 by rounding.

According to the results of the LDA topic classification, the succeeding sentences of the lyrics data for the same topic in the Japanese lyrics corpus as the input sentence were taken as correct data. In the correct data, the top five lyrics with cosine distances close to this value were extracted as templates.

3.1.2 Metaphor Generation

When the template lyrics contained a noun and verb, a verb metaphor was generated based on the noun in the template. A figurative verb was generated for the noun in the template, and replacing the verb in the template with the figurative verb with the template created lyrics containing metaphors. Referring to previous studies(Umemura and Kano, 2021), metaphors were generated using the following procedure.

(I) Obtain a set of particle verbs that co-occur with nouns. The Kyoto University case frame(Daisuke and Sadao, 2005) was used to obtain co-occurrence.

(II) Obtain nouns that co-occur with the verbs(1).

(III) Calculate the degree of abstraction for nouns(2) and select nouns with an absolute value of 0.5 or more from the input noun as a candidate for the starting region.

(IV) Obtain verbs co-occurring with nouns(3). Verbs that match the verbs obtained in verbs(1) were excluded.

(V) Select verbs with low tf-idf values from verbs(4).

(VI) Combining nouns and verbs(4) into metaphorical expressions.

The case frame was an arrangement of usage and its related nouns organized by usage. Using a case

frame allows verb metaphors (the acquisition of rare co-occurring pairs of nouns and verbs). When a figurative verb was selected, the target verb in the template was removed.

Previous studies (Umemura and Kano, 2021) used the degree of polysemy for scoring; however, its effectiveness has not yet been confirmed. In this study, tf-idf was used for scoring to generate novel metaphors. Tf-idf was originally a value measuring the importance of a word in a document. The formula for the tf-idf of verb *v* for noun *n* is defined as follows. In the following, co-occurrence frequency of verb *v* with noun *n* is referred to as f_{vn}, total number of nouns in the case frame as N, and number of nouns co-occurring with verb v as N_v, term frequency of verb v in noun n as tf_{vn}, Inverse Document Frequency of verb v in nouns as idf_v.

$$tf_{vn} = \frac{f_{vn}}{\sum_v f_{vn}}, \quad idf_v = \log \frac{N}{N_v}$$

$$tf\text{-}idf_{vn} = tf_{vn} \cdot idf_v$$

3.1.3 Masked Language Modeling

The nouns obtained in generating metaphors are replaced with nouns in the template, and the verb is replaced with verbs in the template. GINZA(Matsuda, 2020) was used to determine the main noun and main verb in the sentence. Afterwards, insert [MASK] after the metaphorical verb. If no figurative verb was generated, insert [MASK] before the metaphorical noun. This insertion procedure solves grammatical problems caused by verb and noun substitutions in templates.

The BERT model was compatible with the template used. This was because of the masked language modeling of the BERT training method, which hides arbitrary words in a given string of training data and learns to predict those hidden words. Combining BERT and templates also makes it possible to maintain the context of lyrics, which is an issue in GPT. In addition, BERT was a state-of-the-art masked language model at the time of this study. Using fine-tuned BERT not only corrects grammatical errors but also leads to topic coherence. Owing to the nature of fine-tuning, development is also possible using a limited amount of data.

For masked language modeling, a Japanese pretrained BERT[2] fine-tuned on a lyrics corpus was used. The architecture of the BERT model is similar to the original BERT model Devlin et al.'s (2019). Fine-tuning was performed with the following configurations 3.

3.1.4 Re-ranking

The same S-BERT model was used for template extraction. The template based on the generated metaphor and the input sentence was re-ranked and presented in order from those whose cosine distance was close to 0.706, based on the lyrics analysis results of semantic textual similarity between the preceding and succeeding lyrics.

3.2 Results

Examples of each sequence are listed in Table4,5,6,7,8. A toatl of five examples of the output are listed in Table9 when the preceding lyric "流れるまんま流されたら（If we are swept away by the flow）", whose the actual succeeding lyric is "抗おうか美しい鰭で（Shall we resist with our beautiful fins?）". There were three topics when the input lyric was 3. The numbers of moras and topics of these outputs are also shown.

3.3 Discussions

First, we discuss the output lyrics from the perspective of moras. As summarized in Table9, sentence 1 is a complete match between the input sentence and the mora, and the rank is 1. Sentence 2 has 16 moras, which is 3 more than the input sentence but more than sentences 3, 4, and 5, which have fewer moras than the input sentence. From the moras analysis, it can be said that the output result is appropriate.

Next, we discuss the output lyrics from the perspective of grammar. As summarized in table9, there were no grammatical errors in Japanese. For sentences 1 and 5, rare metaphorical expressions were generated. This is because 'feelings' never 'scatter' and 'shadows' never 'feel.' Although sentences 2, 3, and 4 are originally metaphorical expressions, they contain idioms that are used in everyday life such as "change the world." In other words, these sentences cannot be considered as rare metaphorical expressions.

Finally, we discussed the output lyrics from the perspective of topic coherence. As summarized in table9, sentences 1, 2, and 5 maintained topic coherence, based on the LDA topic classification results. However, Sentences 3 and 4 did not maintain

[2] https://github.com/cl-tohoku/bert-japanese/

Model architecture	BERT base model
Optimizer	AdamW
Epochs	5
Adam's epsilon	1e-12

Table 3: Configurations of fine-tuning

Template
溢れる想いが四方に散った
(Overflowing feelings scattered in all directions)
挫けそうになった思い出
(Memories that almost crushed me)
ふりむけばこの世界の
(If I turn around, this world's)
くだらないぜその人生
(That life is silly)
不安の影奪い去っていく
(The shadow of anxiety is taken away)

Table 4: Examples of template.

Metaphor Pair
想い, -(Feeling,-)
思い出，語る(Memories,)
世界，変える(World,)
人生，送る(Life,)
不安，感じる(Anxiety,)

Table 5: Examples of metaphor pair(noun and verb).

Masked Sentence
[MASK]想いが四方に散った
([MASK] feelings scattered in all directions)
挫けそうに語る[MASK]思い出
(Desperately talking [MASK] Memories)
変える[MASK]この世界の
(Change [MASK] of this world's)
送る[MASK]その人生
(Send [MASK] that life)
不安の影感じる[MASK]
(Feel the shadow of anxiety [MASK])

Table 6: Examples of metaphor pair(noun and verb).

Results of Masked Language Modeling
その想いが四方に散った
(This feelings scattered in all directions)
挫けそうに語る遠い思い出
(Desperately talking far memories)
変えるよこの世界の
(I'll change this world's)
送るよその人生
(I'll send that life)
不安の影感じるまま
(Feeling the shadow of anxiety)

Table 7: Examples of metaphor pair(noun and verb).

topic coherence. In summary, three of the five output lyrics maintained topical coherence. The topic number of the input lyrics was three, which was the second-largest topic, as shown in Fig.6. When the number of lyrics included in a topic is small, the accuracy of the fine-tuned BERT model may be affected by the training data. In addition, there was no significant difference in the mean value of the cosine distance for each topic, as shown in Fig.7. Therefore, the cosine distance is not considered to affect topic classification.

4 Conclusions

In this study, we analyzed the Japanese lyrics corpus from the perspective of LDA topic classification and moras. We also propose the application of metaphor generation to lyrics generation and a system for thematically aware lyrics generation. Consequently, lyrics that were grammatically correct and included metaphorical expressions were generated. In addition, more than half of the outputs maintained topic coherence. However, it is necessary to evaluate the output lyrics of this system in humans.

References

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Rank	Final Output	Score
1	その想いが四方に散った(This feelings scattered in all directions)	0.0084
2	挫けそうに語る遠い思い出(Desperately talking far memories)	0.0298
3	変えるよこの世界の(I'll change this world's)	0.0457
4	送るよその人生(I'll send that life)	0.0612
5	不安の影感じるまま(Feeling the shadow of anxiety)	0.1336

Table 8: Output results and distance. Score shows the distance from the value 0.706 of cosine distance. Lower scores represent greater similarity.

Number	Output Sentences	Mora	Topic
1	その想いが四方に散った(This feelings scattered in all directions)	13	3
2	挫けそうに語る遠い思い出(Desperately talking far memories)	16	3
3	変えるよこの世界の(I'll change this world's)	10	1
4	送るよその人生(I'll send that life)	10	4
5	不安の影感じるまま(Feeling the shadow of anxiety)	12	3

Table 9: Examples of outputs for the input as "流れるまんま流されたら（If we are swept away by the flow）." Mora shows the mora of output sentence. Topic shows the results of topic classification of output sentences by LDA. The topic of the input sentence was 3.

Jason Chuang, Christopher D. Manning, and Jeffrey Heer. 2012. Termite: Visualization techniques for assessing textual topic models. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, page 74–77, New York, NY, USA. Association for Computing Machinery.

Kawahara Daisuke and Kurohashi Sadao. 2005. kaku-framejisyo no zenjitekijidoukoutiku. *Journal of natural language processing*, 12(2):109–131.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Kango Iwama and Yoshinobu Kano. 2018. Japanese advertising slogan generator using case frame and word vector. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 197–198, Tilburg University, The Netherlands. Association for Computational Linguistics.

Risako Kagita, Ryosuke Yamanishi, Yoko Nishihara, and Jun'ichi Fukumoto. 2013. Extraction of impressive phrase from lyric focusing on characteristic expression. *IPSJ Technical Report (Web)*, 2013-MUS-101(6):1–6.

John M. Lawler. 1983. Metaphors we live by. *Language*, 59(1):201–207.

Hiroshi Matsuda. 2020. Ginza - universal dependencies niyoru jitsuyoutekinihongokaiseki. *Natural Language Processing*, 27(3):695–701.

Yuto Mitsuishi and Kazutaka Shimada. 2019. Selection of figurative expression using automatically generated data set. *IFAT Technical Report*, 133(8):1–6.

Nikola I. Nikolov, Eric Malmi, Curtis Northcutt, and Loreto Parisi. 2020. Rapformer: Conditional rap lyrics generation with denoising autoencoders. In *Proceedings of the 13th International Conference on Natural Language Generation*, pages 360–373, Dublin, Ireland. Association for Computational Linguistics.

Hiroaki Ohshima, Satoshi Nakamura, and Katsumi Tanaka. 2007. Slothlib: A programming library for research on web search. *The Database Society of Japan Letters*, 6(1):113–116.

Tao Qian, Fan Lou, Jiatong Shi, Yuning Wu, Shuai Guo, Xiang Yin, and Qin Jin. 2023. UniLG: A unified structure-aware framework for lyrics generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 983–1001, Toronto, Canada. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. 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 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Matheus Augusto Rodrigues, Alcione Oliveira, Alexandra Moreira, and Maurilio Possi. 2022. Lyrics generation supported by pre-trained models. *The International FLAIRS Conference Proceedings*, 35.

Katsurou Takahashi and Hiroaki Ohshima. 2019. Genereting personification metaphors by transformation matrix. *DBS Technical Report*, 169(15):1–4.

Kanako Umemura and Yoshinobu Kano. 2021. A catchphrase generator which superposes meanings by automatic metaphor generation. *Japanese Society for Artificial Intelligence Study Group Materials Language/Speech Understanding and Dialogue Processing Workshop*, 91:05.

Jie Wang and Xinyan Zhao. 2019. Theme-aware generation model for chinese lyrics.

Kento Watanabe, Yuichiroh Matsubayashi, Kentaro Inui, and Masataka Goto. 2014. Modeling structural topic transitions for automatic lyrics generation. In *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing*, pages 422–431, Phuket,Thailand. Department of Linguistics, Chulalongkorn University.

Lanqing Xue, Kaitao Song, Duocai Wu, Xu Tan, Nevin L. Zhang, Tao Qin, Wei-Qiang Zhang, and Tie-Yan Liu. 2021. DeepRapper: Neural rap generation with rhyme and rhythm modeling. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 69–81, Online. Association for Computational Linguistics.

Rongsheng Zhang, Xiaoxi Mao, Le Li, Lin Jiang, Lin Chen, Zhiwei Hu, Yadong Xi, Changjie Fan, and Minlie Huang. 2020. Youling: an AI-assisted lyrics creation system. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 85–91, Online. Association for Computational Linguistics.