

# Automated Korean Poetry Generation Using LSTM Autoencoder<sup>\*</sup>

Eun-Soon You<sup>1,\*\*</sup>, Soohwan Kang<sup>2</sup>, and Su-Yeon O<sup>3</sup>

<sup>1</sup> Department of French language and culture, Inha University  
100 Inha-ro, Michuhol-gu, Incheon, 22212 Republic of Korea  
tesniere@naver.com

<sup>2</sup> Department of Korean Studies, Inha University  
100 Inha-ro, Michuhol-gu, Incheon, 22212 Republic of Korea  
22151147@inha.edu

<sup>3</sup> Department of Cultural Contents and Management, Inha University  
100 Inha-ro, Michuhol-gu, Incheon, 22212 Republic of Korea  
djs02061@naver.com

**Abstract.** Automatically composing poem is considered a highly challenging task, and it has received increasing attention across various fields. The computer has to ensure readability, meaningfulness, and vocabulary adequacy as well as the semantics of the poet's production, which are otherwise realized by their imagination and inspiration. In this study, a model for Korean poem generation based on long short-term memory (LSTM) is proposed with the aim of creating poems that imitate the writing style of four poets who represent Korea's modern poetry. To do this, 1000 poems by the target poets were collected, and their styles were defined using natural language processing (NLP). Following this, each sentence of the poem was preprocessed, and training was performed using LSTM. When a user selects the desired poet and enters a keyword, the model automatically generates a poem based on that poet's style. The poems produced showed some errors in syntactic structure and semantic delivery, but they successfully reproduced the characteristic vocabulary and emotions of the poet.

**Keywords:** Poem Generation, Natural Language Generation, Long Short Term Memory (LSTM), Writing Style.

## 1 Introduction

The question of whether a computer is capable of writing text with creative features such as poetry (and if so, how it will differ from human creation) has become a popular one among researchers in the fields of natural language generation (NLG), computational

---

<sup>\*</sup> Copyright © by the paper's authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). In: J.-T. Kim, J. J. Jung, E. You, O.-J. Lee (eds.): Proceedings of the 1st International Workshop on Computational Humanities and Social Sciences (Computing4Human 2020), Pohang, Republic of Korea, 15-February-2020, published at <http://ceur-ws.org>

<sup>\*\*</sup> Corresponding author.

creativity, and, broadly, artificial intelligence (AI). While a significant number of studies have long attempted to achieve automatic responses to a series of questions, automatic poetry generation has been receiving increasing attention in recent times. This is a challenging research area, because it requires a very high level of skill to satisfy the formal conditions and content of the poem.

Automated poetry creation not only indicates technological progress, but it represents a new creative approach altogether that is entirely different from the existing concept and principles of poetry creation. The shift in the perception of poetry creation methods dates back to 1920, even before computers were accessible. It can be observed in a poem by Tristan Tzara, a poet who participated in a new European art movement called Dadaism in the early 20th century. His poem [5] is as follows:

To make a Dadaist poem:  
 Take a newspaper.  
 Take a pair of scissors.  
 Choose an article as long as you are planning to make your poem.  
 Cut out the article.  
 Then cut out each of the words that make up this article and put them in a bag.  
 Shake it gently.  
 Then take out the scraps one after the other in the order in which they left the bag.  
 Copy conscientiously.  
 The poem will be like you.  
 And here are you a writer, infinitely original and endowed with a sensibility that is charming though beyond the understanding of the vulgar.

Zara's poem, published in 1920, proposed a complete departure from traditional poetry. In it, the poet finds adequate poetic words and combines them according to the rules; their intentions, feelings, or causality cannot be found. This recalls the definition of an algorithm, meaning a procedure or method for solving a problem or a step for performing a task.

With the advent of computers, experimental attempts were made to generate poems. Bailey [1] suggested semi-automatic poem generation, emphasizing the potential of computer use in poem creation. The French Atelier of Literature Assisted by Maths and Computers (ALAMO) group [2] proposed 'rimbaudelaire,' a method to combine existing poems in order to create new ones, in which the structure of a poem by Rimbaud was filled with the vocabulary of Baudelaire's poems.

The use of deep learning algorithms such as recurrent neural networks (RNNs) and LSTM in computational creativity has evolved the concept of automatic poetry generation. In this context, the present study proposes a Korean poetry generation model based on deep learning which aims to imitate the writing style of a particular poet.

First, four poets, Kim So Wol, Yoon Dong-Joo, Baek Seok, and Jeong Ji-yong, who represent Korea's modern poetry and remain popular amongst Koreans, were selected. These poets wrote noteworthy poems within the forms of free poetry and lyric poetry during the Japanese occupation.

The rests of the paper are organized as follows. We start by reviewing previous works in the Section 2. Section 3 describes the approach adapted in our experiments. And we

illustrate the evaluation in Section 4. The conclusion of this paper and future work are demonstrated in Section 5.

## 2 Related Work

If Bailey [1] demonstrated the possibility of computer-generated poetry, Gervás [4] marked the beginning of the automatic poetry generation, and various studies have been since. Wu and Tosa [10] proposed a poem generation system based on Haiku phrase corpus. When the user enters a word or phrase, the system finds expressions containing it in the corpus and creates a poem by combining them. Manurung and Thompson [6] developed a system using genetic algorithms. The poetry generation system, called McGonagall, finds one of several candidate poems with no grammatical error and clear meaning transmission according to stochastic search. Das and Gambäck [3] presented a syllable-based poetry generator. When a user enters a sentence, the syllabification engine understands its rhythm and generates the appropriate sentence that follows it.

Recently, along with the advance of machine learning, poetry generation using deep learning have emerged. Wang et al. [9] propose the machine poetry generator based on LSTM for imitating Chinese poet Du Fu’s writing styles. Given the first character, this model produces a poem reflecting the tone and rhythm of Du Fu’s poem. Zugarini and Maggini [11] also suggested a system using LSTM to generate terrestres, a feature of Dante Alighieri.

Poetic generation researches in various languages such as English, Japanese, Chinese, and Italian are being actively conducted. Regarding Korean poem generation, Park et al. [7] introduced a model for generating poems using Sequence Generative Adversarial Networks (SeqGAN). Korean poem generation is just beginning. In this context, we present a Korean poetry generation model based on deep learning aiming to imitate the style of a particular poet.

## 3 Approaches

### 3.1 Experimental Workflow

The whole experiment workflow is shown in Figure 1. At the beginning of the experiment, we first collected 400 poems written by four poets. The poetic work of the target authors is usually not enough to successfully train deep neural networks, so we collected a total of 1000 pieces by adding other poems written during the Japanese colonial period. Before training the data using LSTM, we removed old language, Chinese character, etc. from the poem in the pre-processing process and then numbered each line of the poem. When a user enters a keyword and clicks a poet, a poem is generated.

### 3.2 Stylistic Analysis

We attempted a quantitative analysis of poetry text to define the style of each poet. To do this, we extracted high frequency vocabularies using part-of-speech (POS) tagging

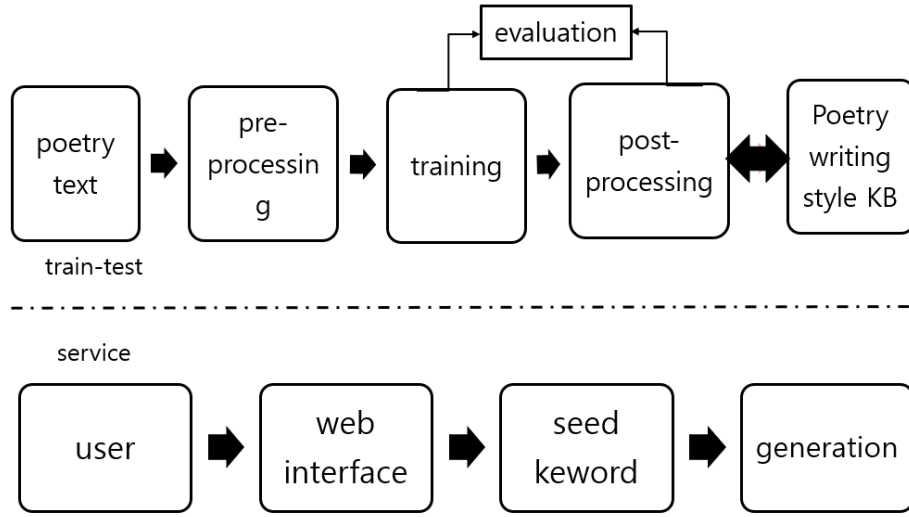


Fig. 1: The workflow of the whole experiment.

and made a cloud composed of top frequent word as shown in Figure 2. Besides, we analyzed the co-occurrence patterns of words through bi-gram.

From the results of the stylistic analysis, certain characteristics could be determined. First, vocabulary related to nature appeared frequently in the poems by the specified poets; second, the primary emotion influencing their poetry is sorrow; and third, sensory expressions representing nature are often used. Finally, the use of first-person pronouns ‘I’ and determiner such as ‘this’ was high. Table 1 shows the results of the stylistic analysis.

Categories	Authors	Examples
Nature	윤동주 (Yun Dong-Joo)	밤 (night), 하늘 (sky), 별 (star), etc.
	정지용 (Jeong Ji-yong)	바다 (sea), 물 (water), etc.
	김소월 (Kim So Wol)	산 (mountain), 나무 (tree), etc.
	백석 (Baek Suk)	새 (bird), 개구리 (frog), etc.
Emotion	슬프 (sad), 외롭 (alone), 괴롭 (painful), 서럽 (sorrow), etc.	
Sense	붉은 (red), 푸른 (blue), 밝은 (bright), 검은 (black), 높은 (high), 뜨거운 (hot), etc.	
Pronoun	나 (I)	
Determiner	이 (this), 그 (that), etc.	

Table 1: The results of the stylistic analysis.



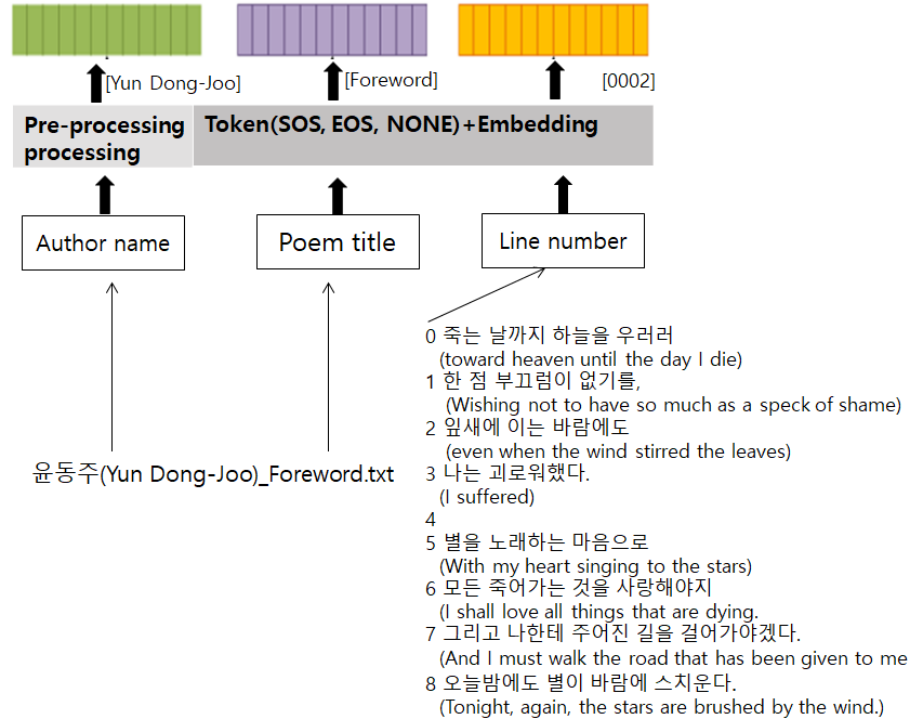


Fig. 3: The training process for Korean poetry generation.

sentence meanings could not be gathered despite the absence of syntactic errors. In order to improve on these aspects, a significant amount of further study is required.

#### 4.2 Imitating the style of a poet

When a user clicks on one of the four poets and enters the desired keyword, a new poem reflecting the poet's style is created. Even if the same keyword is entered repeatedly, new results will be generated each time. The reproductions of the poets' vocabulary and emotions were analyzed in 60 poems. Words related to nature, such as rivers, skies, mountains, and the sea appeared in the title and content of the poem, and emotional words representing sorrow and loneliness were used. However, some meaninglessly repeated words affected the readability of the poems.

### 5 Conclusion

Automatic poetry composition is a highly challenging problem because poetry is a genre of literature that expresses human imagination and creativity. Several studies have been conducted to generate poems in various languages such as English, Chinese, and

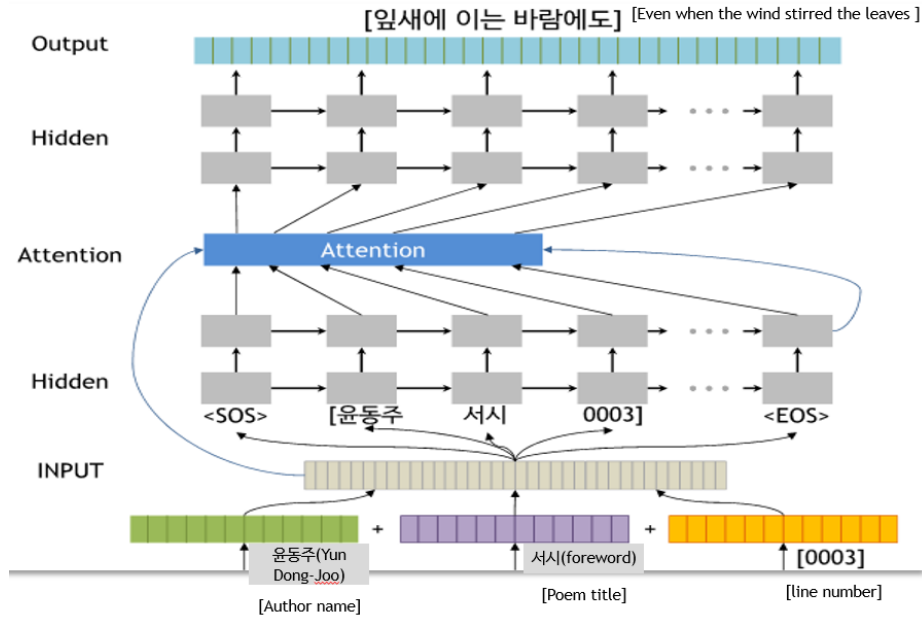


Fig. 4: Korean Poetry Generation Model based on LSTM.

Japanese. In this paper, an LSTM-based approach to produce a Korean poem with a specific style is presented. When a user enters a keyword and clicks on a poet, the model creates a poem that reflects the poet's writing style. Korean poem generation research is just beginning; this work is expected to contribute to Korean text generation research.

## Acknowledgment

This research was supported by the Korea Creative Content Agency, under the Ministry of Culture, Sports and Tourism.

## References

1. Bailey, R.W.: Computer-assisted poetry: the writing machine is for everybody. *Computers in the Humanities* pp. 283–295 (1974)
2. Collectifs, G.: *Atlas de Littérature Potentiel. Folio essais*, Gallimard, Paris, France (Jan 1988), [https://www.ebook.de/de/product/10458470/gall\\_collectifs\\_atlas\\_de\\_litt\\_potentiel.html](https://www.ebook.de/de/product/10458470/gall_collectifs_atlas_de_litt_potentiel.html)
3. Das, A., Gambäck, B.: Poetic machine: Computational creativity for automatic poetry generation in bengali. In: Colton, S., Ventura, D., Lavarac, N., Cook, M. (eds.) *Proceedings of the 5th International Conference on Computational Creativity (ICCC 2014)*. pp. 230–238. [computationalcreativity.net](http://computationalcreativity.net), Ljubljana, Slovenia (Jun 2014)

4. Gervás, P.: Wasp: Evaluation of different strategies for the automatic generation of spanish verse. In: Time for AI and Society - Proceedings of the AISB Symposium on Creative & Cultural Aspects and Applications of AI & Cognitive Science. pp. 93–100. Birmingham, UK (Apr 2000)
5. Lewis, P.: The Cambridge Introduction to Modernism. Cambridge University Press, Cambridge, UK (2007). <https://doi.org/10.1017/cbo9780511803055>
6. Manurung, R., Ritchie, G., Thompson, H.: Using genetic algorithms to create meaningful poetic text. *Journal of Experimental & Theoretical Artificial Intelligence* **24**(1), 43–64 (Mar 2012). <https://doi.org/10.1080/0952813x.2010.539029>
7. Park, Y.H., Jeong, H.J., Kang, I.M., Park, C.Y., Choi, Y.S., Lee, K.J.: Automatic generation of korean poetry using sequence generative adversarial networks. In: Proceedings of the 2018 Annual Conference on Human and Language Technology, Human and Language Technology. pp. 580–583 (Oct 2018)
8. Stent, A., Marge, M., Singhai, M.: Evaluating evaluation methods for generation in the presence of variation. In: Gelbukh, A.F. (ed.) Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2005). Lecture Notes in Computer Science, vol. 3406, pp. 341–351. Springer Berlin Heidelberg, Mexico City, Mexico (Feb 2005). [https://doi.org/10.1007/978-3-540-30586-6\\_38](https://doi.org/10.1007/978-3-540-30586-6_38)
9. Wang, K., Tian, J., Gao, R., Yao, C.: The machine poetry generator imitating du fu's styles. In: Proceedings of the 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD 2018). pp. 261–265. IEEE, Chengdu, China (May 2018). <https://doi.org/10.1109/icaibd.2018.8396206>
10. Wu, X., Tosa, N., Nakatsu, R.: New hitch haiku: An interactive renku poem composition supporting tool applied for sightseeing navigation system. In: Natkin, S., Dupire, J. (eds.) Proceedings of the 8th International Conference on Entertainment Computing (ICEC 2009). Lecture Notes in Computer Science, vol. 5709, pp. 191–196. Springer Berlin Heidelberg, Paris, France (Sep 2009). [https://doi.org/10.1007/978-3-642-04052-8\\_19](https://doi.org/10.1007/978-3-642-04052-8_19)
11. Zugarini, A., Melacci, S., Maggini, M.: Neural poetry: Learning to generate poems using syllables. In: Tetko, I.V., Kurková, V., Karpov, P., Theis, F.J. (eds.) Artificial Neural Networks and Machine Learning – ICANN 2019: Text and Time Series – Proceedings of the 28th International Conference on Artificial Neural Networks. Lecture Notes in Computer Science, vol. 11730, pp. 313–325. Springer International Publishing, Munich, Germany (Sep 2019). [https://doi.org/10.1007/978-3-030-30490-4\\_26](https://doi.org/10.1007/978-3-030-30490-4_26)