# Evaluating the Effectiveness of Large Language Models in Translating Korean Poetry into English: Leveraging Document-Level Context to Preserve Literary Quality

Jiyeon Lee

Universität Mannheim, 68161, Mannheim, Germany

**Abstract.** This paper investigates the efficacy of large language models (LLMs) in translating Korean poetry into English while preserving its literary quality and cultural nuances. By leveraging advanced AI tools, this study aims to explore whether LLMs can maintain the stylistic and contextual integrity that is important in literary translation. A dataset comprising 50 Korean poems was translated using LLMs, and the outputs were evaluated through human assessments, focusing on error types and fidelity to the original texts. The findings highlight the critical role of human oversight in translation processes, despite significant advances in AI capabilities.

**Keywords:** Large Language Model · Literary translation · Prompt Engineering · line, stanza, and Whole poem-level evaluation

## 1 First Section

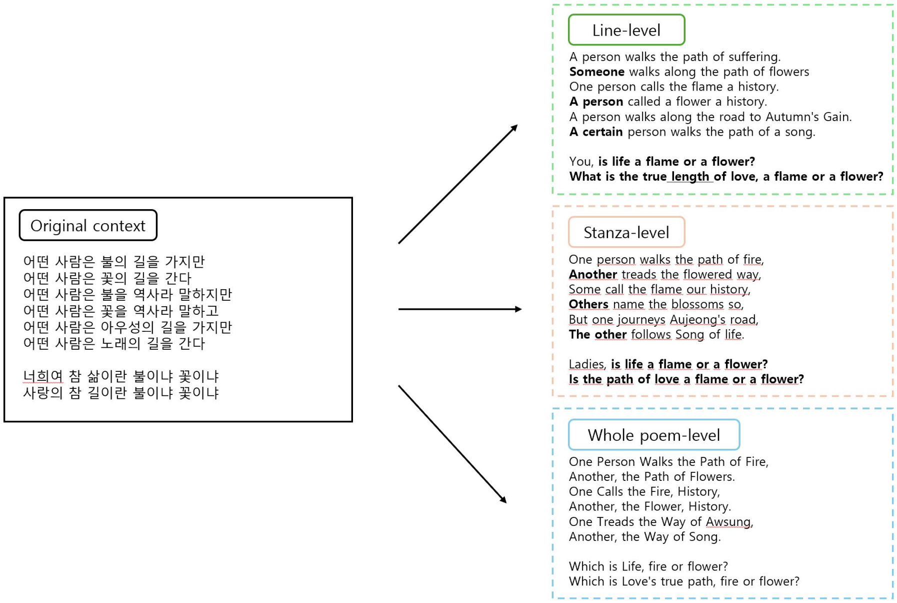
### 1.1 Introduction

The advent of artificial intelligence models has revolutionized many aspects of technology, with literary translation. The task of translating Korean poetry into English presents unique challenges due to the dense, culturally and poetically rich nature of the source material and the structural and grammatical complexities of the Korean language. This paper evaluates the performance of large language models in this domain, aiming to assess their ability to handle the nuances and subtleties inherent in literary texts.

Historically, translation has been predominantly a human effort, where the translator's interpretive skills play a crucial role in conveying the author's original intent and emotional tone. However, with the rise of neural networks and machine learning, AI models like LLMs have shown promising results in translating texts from one language to another. Despite their proficiency in handling large datasets and complex patterns, questions remain about their ability to fully capture the depth of literary works, particularly in poetry where every word and phrase may carry significant symbolic weight.

This study, therefore, not only tests the technical capabilities of LLMs in translating Korean poems but also examines how well these translations retain the poetic style, cultural context, subtle linguistic cues, and cultural nuances often lost in translation. Through a combination of AI-driven translation and meticulous human evaluation, this research seeks to determine the practical limits of current AI technologies in literary translation and the ongoing necessity for human expertise in achieving translations that are true to their source.

**why Korean Poetry** Despite the advantages of technology in translating huge texts quickly and providing unique interpretations, poetry translation demands a nuanced understanding often beyond the capabilities of machines alone. Korean poetry, rich in cultural subtleties and deep meanings, requires the sensitivity and contextual awareness that typically come from human translators. Therefore, the most effective translations often result from a collaborative effort where humans and large language models work together. This synergy allows the preservation of emotional depth and cultural nuances in translation, while also introducing innovative perspectives that can enhance our appreciation and understanding of the original works.[8, 6]



어떤 사람은 불의 길을 가지만
어떤 사람은 꽃의 길을 간다
어떤 사람은 불을 역사라 말하지만
어떤 사람은 꽃을 역사라 말하고
어떤 사람은 아우성의 길을 가지만
어떤 사람은 노래의 길을 간다

너희여 참 삶이란 불이냐 꽃이냐
사랑의 참 길이란 불이냐 꽃이냐

A person walks the path of suffering.
**Someone** walks along the path of flowers
One person calls the flame a history.
**A person** called a flower a history.
A person walks along the road to Autumn's Gain.
**A certain** person walks the path of a song.

You, **is life a flame or a flower?**
**What is the true length of love, a flame or a flower?**

One person walks the path of fire,
**Another** treads the flowered way,
Some call the flame our history,
**Others** name the blossoms so,
But one journeys Aujeong's road,
**The other** follows Song of life.

Ladies, **is life a flame or a flower?**
**Is the path of love a flame or a flower?**

One Person Walks the Path of Fire,
Another, the Path of Flowers.
One Calls the Fire, History,
Another, the Flower, History.
One Treads the Way of Awsung,
Another, the Way of Song.

Which is Life, fire or flower?
Which is Love's true path, fire or flower?

**Fig. 1.** An example of line, stanza, and whole poem level translation

**why Human evaluation** Similar to BLEU and BLEURT, many MT (Machine Translation) evaluation metrics fail to effectively assess translation quality. Translators have noted that overly literal translations are a major flaw in current MT systems. Moreover, monolingual human evaluations comparing MT outputs with human-generated reference translations have revealed additional issues concerning readability and fluency.[11]
Furthermore, poetry is a literary form that deeply relies on subtle expressions, employing stylistic features like meter, rhyme, and imagery to deliver complex emotional and intellectual themes. When it comes to translating poetry, it is crucial to preserve these elements, as they are essential to the poem's overall effect and significance. In contrast to prose, each word and structure in poetry carries an array of potential interpretations, which makes translating it a particularly delicate endeavor.

## 2 Second Section

### 2.1 Background

Large language models (LLMs) have made notable strides in machine translation, especially at the sentence level. However, their effectiveness in translating longer texts such as paragraphs and documents, particularly within literary contexts, has not been thoroughly examined. This is largely due to the complexities and costs associated with such comprehensive evaluations. This research focuses on assessing how well LLMs can handle document-level context in literary translations, underscoring both their capabilities and ongoing challenges in this area.

The results demonstrate that LLMs when translating at the document level, tend to produce translations of higher quality with fewer errors in mistranslation, grammar, and style compared to sentence-level translations.[10] Nonetheless, they are still susceptible to significant mistakes, including occasional omissions of content. Therefore, human oversight is essential to ensure the translation faithfully represents the original author's voice and maintains the overall translation quality.
Moreover, the study highlights the distinct difficulties in translating literary works, such as poetry, which demand a deep understanding of nuanced language and cultural nuances. For instance, translating a Korean poem into English requires not only linguistic accuracy but also sensitivity to convey the original emotion, intent context, and stylistic elements like meter, rhyme, and imagery. While LLMs can provide a foundational translation, human translators are crucial for refining these efforts to capture the poetic and artistic qualities of the source material.

In collaboration with human translators, LLMs help to bridge the gap between cultural contexts and emotional expressions from Korean to English. This research aims to advance translation studies by examining how literary traditions interact and by developing better methods for translating literary content.

# 3 Third Section

This paper focuses on translating modern Korean poetry at each line, stanza, and whole poem level using a Large language model. In this section, we not only treat 50 modern Korean poetry but also is adjusted to different levels so that the translation can be performed at each level. Also, this paper outlines how to leverage Mistral AI using LM studio to translate this modern Korean poetry dataset at both the line, stanza, and whole poem levels.

## 3.1 Dataset collection

**Source and target language** As the source language, this paper selects Korean which belongs to different language families that have unique structural features and different writing systems. As the target language, this paper selected English[15]

**Selecting works** For a source poem to be included, this paper collects 50 modern Korean poems that were written from 1910, and must have been translated poetry from the source poem. to compare the results translated into three levels.

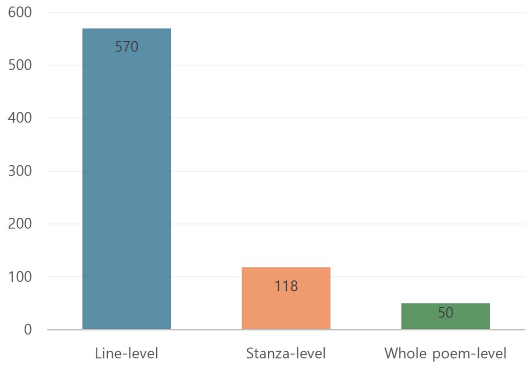


**Fig. 2.** The number of line, stanza, and poem

Fig2 presents the quantity of lines, stanzas, and whole poems analyzed in the study. The bar graph consists of a total of 570 lines at the line-level, 118 stanzas at the stanza-level, and 50 poems at the whole-poem level.

**Characteristics in modern Korean poetry** modern Korean poetry encapsulates a deep tension between traditional Korean values and the pressures of modernity, exploring complex themes such as national identity, individual freedom, and the stark realities of political conflict. This period marked a significant departure from classical poetic forms, embracing free verse and experimental structures that profoundly shaped contemporary Korean literature. Additionally, modern Korean poetry captures the nuances of contemporary life, touching on rapid modernization, cultural identity, and personal introspection. The unique structural features of the Korean language, including its agglutinative syntax and extensive system of honorifics, enhance the depth and subtlety of the poetry. These linguistic characteristics present significant translation challenges, as they enrich the aesthetic experience but also demand precision in conveying nuanced emotions and cultural contexts. [8, 6]

\

**Table 1.** Examples of modern Korean poetry[1]

| | |
|---|---|
| 1 | 단 한 사람의 가슴도 제대로 지피지 못했으면서 무성한 연기만 내고 있는 내 마음의 군불이여 꺼지려면 아직 멀었느냐 |
| 2 | 죽는 날까지 하늘을 우러러 한 점 부끄럼이 없기를, 잎새에 이는 바람에도 나는 괴로워했다. 별을 노래하는 마음으로 모든 죽어 가는 것을 사랑해야지. 그리고 나한테 주어진 길을 걸어가야겠다.<br>오늘 밤에도 별이 바람에 스치운다. |

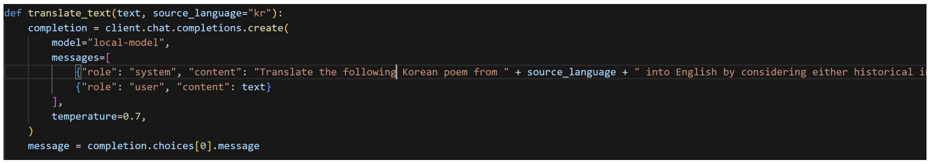
### 3.2 Translate with large language models

In this section, this paper focuses on translating a dataset containing 50 modern Korean poetry[1] at each level using a large language model. More specifically, This paper employs the Mistral 7B Instruct v0.2[4] with LM studio[2] to consider the following translating three levels. this paper gives them the same strategy and then allows this paper to compare their results.

**Prompting for translation:**

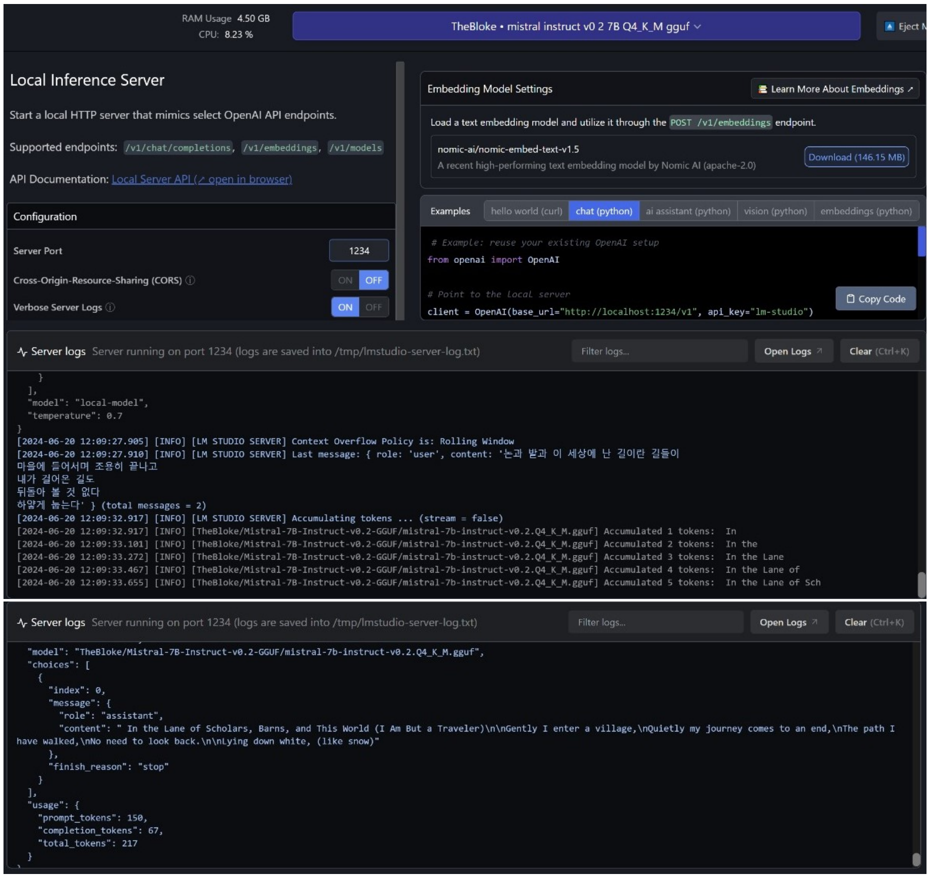
- Line-level translation: In this approach, each line of the poem is translated independently from the others, with the translated lines produced in the target language.
- Stanza-level translation: This method involves translating each stanza of the poem separately from the other stanzas, resulting in the translated stanzas being generated in the target language.
- Whole poem-level translation: The entire poem is translated as a single unit, distinct from other poems, and the complete translated poem is generated in the target language.

This methodology is used to create the results of translating and prompting.[7]



```python
def translate_text(text, source_language="kr"):
    completion = client.chat.completions.create(
        model="local-model",
        messages=[
            {"role": "system", "content": "Translate the following Korean poem from " + source_language + " into English by considering either historical i
            {"role": "user", "content": text}
        ],
        temperature=0.7,
    )
    message = completion.choices[0].message
```

**Fig. 3.** The message for prompting: Translate the following Korean poem from source_language into English by considering either historical information, the Author's Life, the poetic style, or elements such as the order of lines, rhythm, rhyme, Unity, Consistency, Metaphors, Figurative Language and so on.[14]



Local Inference Server

Start a local HTTP server that mimics select OpenAI API endpoints.

Supported endpoints: `/v1/chat/completions`, `/v1/embeddings`, `/v1/models`

API Documentation: Local Server API (↗ open in browser)

Configuration

Server Port: 1234

Cross-Origin-Resource-Sharing (CORS): OFF

Verbose Server Logs: ON

Embedding Model Settings

Load a text embedding model and utilize it through the `POST /v1/embeddings` endpoint.

nomic-ai/nomic-embed-text-v1.5
A recent high-performing text embedding model by Nomic AI (apache-2.0)

```
# Example: reuse your existing OpenAI setup
from openai import OpenAI

# Point to the local server
client = OpenAI(base_url="http://localhost:1234/v1", api_key="lm-studio")
```

Server logs Server running on port 1234 (logs are saved into /tmp/lmstudio-server-log.txt)

```
    }
  ],
  "model": "local-model",
  "temperature": 0.7
}
[2024-06-20 12:09:27.905] [INFO] [LM STUDIO SERVER] Context Overflow Policy is: Rolling Window
[2024-06-20 12:09:27.910] [INFO] [LM STUDIO SERVER] Last message: { role: 'user', content: '논과 밭과 이 세상에 난 길이란 길들이
마음에 들어서며 조용히 끝나고
내가 걸어온 길도
뒤돌아 볼 것 없다
하얗게 눕는다' } (total messages = 2)
[2024-06-20 12:09:32.917] [INFO] [LM STUDIO SERVER] Accumulating tokens ... (stream = false)
[2024-06-20 12:09:32.917] [INFO] [TheBloke/Mistral-7B-Instruct-v0.2-GGUF/mistral-7b-instruct-v0.2.Q4_K_M.gguf] Accumulated 1 tokens:  In
[2024-06-20 12:09:33.101] [INFO] [TheBloke/Mistral-7B-Instruct-v0.2-GGUF/mistral-7b-instruct-v0.2.Q4_K_M.gguf] Accumulated 2 tokens:  In the
[2024-06-20 12:09:33.272] [INFO] [TheBloke/Mistral-7B-Instruct-v0.2-GGUF/mistral-7b-instruct-v0.2.Q4_K_M.gguf] Accumulated 3 tokens:  In the Lane
[2024-06-20 12:09:33.467] [INFO] [TheBloke/Mistral-7B-Instruct-v0.2-GGUF/mistral-7b-instruct-v0.2.Q4_K_M.gguf] Accumulated 4 tokens:  In the Lane of
[2024-06-20 12:09:33.655] [INFO] [TheBloke/Mistral-7B-Instruct-v0.2-GGUF/mistral-7b-instruct-v0.2.Q4_K_M.gguf] Accumulated 5 tokens:  In the Lane of Sch
```

Server logs Server running on port 1234 (logs are saved into /tmp/lmstudio-server-log.txt)

```
  "model": "TheBloke/Mistral-7B-Instruct-v0.2-GGUF/mistral-7b-instruct-v0.2.Q4_K_M.gguf",
  "choices": [
    {
      "index": 0,
      "message": {
        "role": "assistant",
        "content": " In the Lane of Scholars, Barns, and This World (I Am But a Traveler)\n\nGently I enter a village,\nQuietly my journey comes to an end,\nThe path I
have walked,\nNo need to look back.\n\nLying down white, (like snow)"
      },
      "finish_reason": "stop"
    }
  ],
  "usage": {
    "prompt_tokens": 150,
    "completion_tokens": 67,
    "total_tokens": 217
  }
```

**Fig. 4.** The prompting using LM Studio

**Prompting** This experiment in Fig4 employs a Python code for prompting and translating. It connects with a Local Model Inference Server and then sends the text of the poem to the server, which then processes the translations. The dataset in which included 50 modern Korean poems translates Korean poems into English at the line, stanza, and whole-poem levels, following the specified prompt strategy. [9]

## 4 Fourth Section

This paper compares the representation of translation quality. Unlike MT evaluation metrics such as BLEU and BLEURT, this study does not utilize these automatic metrics to assess the results. Instead, it considers alternative methods to evaluate the reliability and quality of the translations.

### 4.1 Evaluating Three levels poetry translation

**Recruiting annotator** For annotation purposes, this paper requires proficiency in both Korean, the source language, and English, the target language. Consequently, the author of this paper, who is a native speaker, has been hired to perform the annotation tasks.

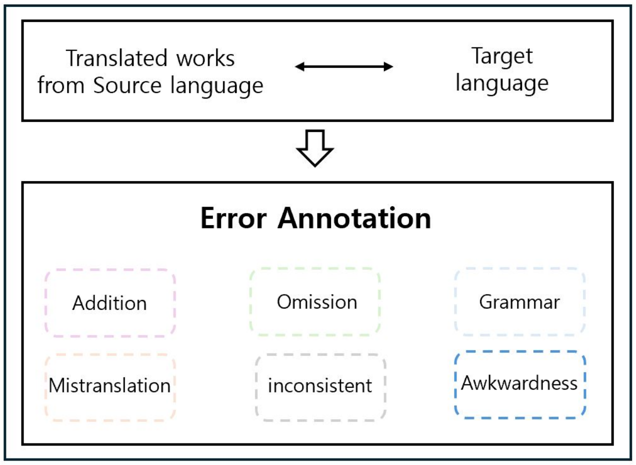
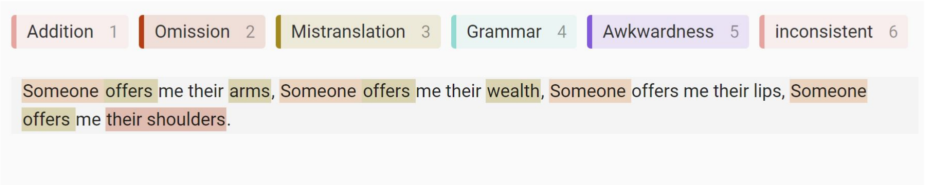


**Fig. 5.** The process of annotation task

**Annotation task** This paper utilizes Multidimensional Quality Metrics (MQM) to conduct an annotation task, employing an annotation tool specifically for error annotation. The Fig5 illustrates the methodology used to identify and categorize errors in translations from a source language to a target language. Following the translation of poetry from Korean to English, this study assesses the preservation of the original meaning, rhyme, and other elements by comparing the translated poem with both the source and target languages.[7, 3]



Someone offers me their arms, Someone offers me their wealth, Someone offers me their lips, Someone offers me their shoulders.

**Fig. 6.** The annotation tool for error annotation.[12, **?**]

Specifically, this paper highlights these certain error types in MQM when this paper evaluates and annotates.[4]

\

**Table 2.** Annotation Guideline based on MQM. [13, 5]

| Error category | Description |
|---|---|
| Mistranslation | Translation does not accurately represent the source. |
| Addition | Translation includes information not present in the source. |
| Omission | Translation is missing content from the source |
| Grammar | Issues related to the grammar or syntax of the text. other than spelling and orthography (especially inconsistency of the tenses and conditionals) |
| Inconsistent | Terminology is used in a consistent manner within the text. |
| Awkwardness | Translation has stylistic problems. |

## 5 Fifth Section

### 5.1 Results

In this section, the paper examines the results derived from translations at three levels, based on data from human evaluations. It is generally noted that translations at the stanza and whole poem levels outperform those at the line level according to human evaluation. These findings indicate that the model effectively leverages stanza-level context to produce superior translations compared to line-level translations.

Additionally, the paper explores poetic elements like rhymes and rhythms. It is observed that stanza-level translation more effectively maintains the poem's structural and thematic integrity, though it may introduce additional errors due to the complexity of translating larger text blocks.[7]

\

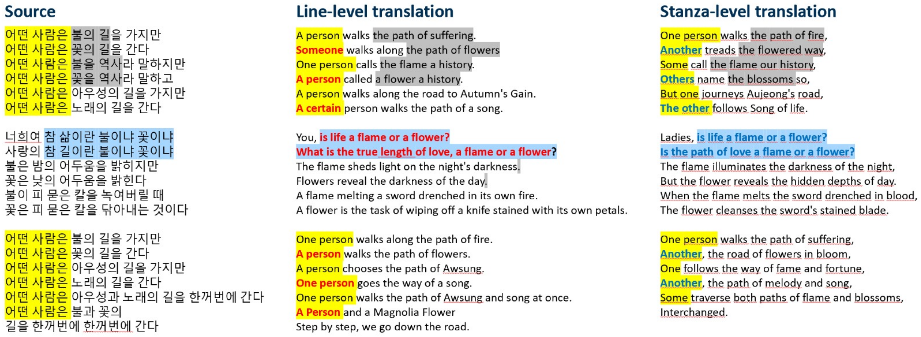
**Table 3.** Total counts of all error type from this paper annotation

| Type | Line | Stanza | Whole poem |
|---|---|---|---|
| Mistranslation | 943 | 625 | 710 |
| Omission | 12 | 13 | 6 |
| Addition | 47 | 61 | 54 |
| Grammar | 26 | 21 | 13 |
| Inconsistent | 38 | 27 | 15 |
| Awkwardness | 113 | 43 | 78 |

1. **Mistranslation**:
   - Mistranslation is the most frequent error across all levels, with the highest incidence at the line-level. This suggests that individual lines may be more susceptible to errors in meaning when translated independently. Also this can be due to linguistic differences, cultural contexts, or specific poetic components that is hard to translate

2. **Omission**:
   - Omission errors are relatively low across all levels, indicating a general accuracy in content inclusion during translation.
3. **Addition**:
   - Addition errors, where extraneous content is added, are more frequent at the stanza and whole poem- level, possibly due to the complexity of translating larger text blocks.
4. **Grammar**:
   - Grammar errors are evenly distributed in line and stanza translations but significantly drop at the whole poem level, suggesting better grammatical coherence when the entire poem is considered as a unit.
5. **Inconsistent**:
   - Inconsistency in translation seems to increase with the complexity of the text (stanza level), but markedly decreases when translating entire poems, suggesting that broader context helps maintain consistency.
6. **Awkwardness**:
   - Awkward phrasings often occur when translators attempt to adhere too closely to the source language structure, resulting in unnatural expressions in the target language. And awkward translations are most common at the line level, indicating that shorter text segments might be prone to less natural translation outputs without the broader contextual cues available in longer sections.

**Human evaluation prefers translation at Stanza-level** The human evaluation shows a preference for stanza and whole poem level translation over other granularities, such as line-level. This preference is largely because stanza-level translation allows for a better balance between maintaining the poem's original structure and capturing its emotional and thematic depth. This can consider the context provided by the entire stanza, leading to translations that are more coherent and faithful to the original intent and tone of the poem. These level translations preserve the integrity of poetic devices like metaphors and rhyme schemes that might span multiple lines, thus preserving the aesthetic and rhythmic qualities of the poem.



| Source | Line-level translation | Stanza-level translation |
|---|---|---|
| 어떤 사람은 불의 길을 가지만 | A person walks the path of suffering. | One person walks the path of fire, |
| 어떤 사람은 꽃의 길을 간다 | Someone walks along the path of flowers | Another treads the flowered way, |
| 어떤 사람은 불을 역사라 말하지만 | One person calls the flame a history. | Some call the flame our history, |
| 어떤 사람은 꽃을 역사라 말하고 | A person called a flower a history. | Others name the blossoms so, |
| 어떤 사람은 아우성의 길을 가지만 | A person walks along the road to Autumn's Gain. | But one journeys Aujeong's road, |
| 어떤 사람은 노래의 길을 간다 | A certain person walks the path of a song. | The other follows Song of life. |
| | | |
| 너희여 참 삶이란 불이냐 꽃이냐 | You, is life a flame or a flower? | Ladies, is life a flame or a flower? |
| 사랑의 참 길이란 불이냐 꽃이냐 | What is the true length of love, a flame or a flower? | Is the path of love a flame or a flower? |
| 불은 밤의 어두움을 밝히지만 | The flame sheds light on the night's darkness. | The flame illuminates the darkness of the night, |
| 꽃은 낮의 어두움을 밝힌다 | Flowers reveal the darkness of the day. | But the flower reveals the hidden depths of day. |
| 불이 피 묻은 칼을 녹여버릴 때 | A flame melting a sword drenched in its own fire. | When the flame melts the sword drenched in blood, |
| 꽃은 피 묻은 칼을 닦아내는 것이다 | A flower is the task of wiping off a knife stained with its own petals. | The flower cleanses the sword's stained blade. |
| | | |
| 어떤 사람은 불의 길을 가지만 | One person walks along the path of fire. | One person walks the path of suffering, |
| 어떤 사람은 꽃의 길을 간다 | A person walks the path of flowers. | Another, the road of flowers in bloom, |
| 어떤 사람은 아우성의 길을 가지만 | A person chooses the path of Awsung. | One follows the way of fame and fortune, |
| 어떤 사람은 노래의 길을 간다 | One person goes the way of a song. | Another, the path of melody and song, |
| 어떤 사람은 아우성과 노래의 길을 한꺼번에 간다 | One person walks the path of Awsung and song at once. | Some traverse both paths of flame and blossoms, |
| 어떤 사람은 불과 꽃의 | A Person and a Magnolia Flower | Interchanged. |
| 길을 한꺼번에 한꺼번에 간다 | Step by step, we go down the road. | |

**Fig. 7.** An example of comparing between line and stanza level translation

**Stanza-level is better than line-level** Stanza-level translation is generally considered better than line-level translation for several reasons. First, line-level translation can lead to a fragmented understanding of the poem, as each line is often translated in isolation, potentially missing the broader thematic elements developed across multiple lines. This isolation can disrupt the natural flow and coherence found in the original language. In contrast, stanza-level translation takes into account the interplay of lines within a stanza, enabling a more holistic approach that captures subtleties and nuances that line-level translations often miss.

For example, Looking at the Fig 7, this is the result of translation to the line and stanza level. In the yellow highlight, the same words are used but refer to different people. Since 'fire(불)' and 'flower(꽃)' are contrasting, they can't describe just one person. In the line-level translation, both words were simply translated as "a person." But in the stanza-level translation, they were correctly linked to different individuals, showing a better understanding of the poem's context.

And next the blue highlights, These parts of the poem need extra attention to rhyme, as they end with the same sound, " 이냐." The stanza translation does this well, capturing the rhyme at the beginning of the sentences with the word "is." However, the line-level translation missed the rhyme completely.

**Whole poem sentence Analysis** This paper identify several significant errors in Addition and Mistranslation. Whole poem level translation usually appear to fill in cultural, historical poetic gaps that might not be clear from just the lines or stanzas helping readers get the picture. it also help connect different parts of the poem together, making sure everything flows well as a whole, particularly in longer poems where themes might require more extensive explain.

**Between Stanza-level and Whole poem-level translation** Comparing stanza-level and whole poem-level translations presents a nuanced discussion. Whole poem-level translation considers the entire work at once, potentially capturing overarching themes and stylistic consistencies better than stanza-level translation. However, this approach can sometimes overlook the intricacies and local context of individual stanzas, which are crucial for the poem's impact and meaning. Stanza-level translation, on the other hand, strikes a balance by focusing on the smaller units of meaning (stanzas) without losing sight of the poem's overall thematic and structural integrity.

**Preserving poetic component**

1. **Line-level translation**:
   - The granularity tends to focus narrowly on the literal meaning of individual lines without considering the poem's broader context or its poetic form. As a result, important aspects like rhyme schemes, rhythmic patterns, and thematic connections between lines lost, leading to a translation that might adhere to the text's surface meaning but lacks its poetic essence and cohesion.
2. **Stanza-level translation**:
   - Stanza-level translation, however, offered a more effective approach to preserving poetic components. By treating the stanza as the unit of translation, the model was able to maintain the structural and thematic integrity of the poem.
3. **Whole poem-level translation**:
   - Addition errors, where extraneous content is added, are more frequent at the stanza level, possibly due to the complexity of translating larger text blocks

## 6 Sixth Section

### 6.1 Conclusion

In this paper, human evaluation on a Korean-English dataset identifies contextual sensitivity, artistic integrity, and lexical cohesion as three main sources of inconsistency. Also, this paper demonstrates that LLMs, by utilizing paragraph and whole poem-level context, produce translations that are more consistent and make sense than line-level translations, while also containing fewer mistranslations and grammatical issues.

The choice of granularity in poetry translation—line-level, stanza-level, or whole poem-level—has significant implications for how effectively the poetic components of the original work are preserved and conveyed. Stanza-level translation, in particular, is considered because it balances the need to maintain the structural and thematic integrity of the poem while capturing the aesthetic and emotional nuances.

### 6.2 Limitations

We have demonstrated that the LLMs uses stanza-level context to generate translations that are superior to those made by line-level equivalents. Nonetheless, there are still potential confounding factors.

Firstly, Poetry is subjective, and different readers may derive different meanings from the same piece. Human interpretation and evaluators can offer diverse perspectives on whether the translation resonates in a way that is true to the original's intent, reflecting a range of emotional and intellectual responses. Moreover, Human evaluation is insightful, However, this work is time-consuming

Secondly, From the annotations, this paper observes that all of them suffer from occasional omissions and addition of content from the source. awkwardness as well. Moreover, Human evaluation is insightful, However, this work is time-consuming

Finally, Korean has a low source regarding translation. In training GPT models, there is more high-quality training data available for English than Korean.

### 6.3 Discussion

This study highlights the sophisticated capabilities and limitations of large language models (LLMs) in translating Korean poetry into English. The findings reveal significant variations in translation accuracy between different textual levels—line, stanza, and whole poem. These discrepancies underscore the importance of developing translation processes that are sensitive to the structural and thematic integrity of the original works.

Looking to the future, this research identifies several avenues for improving the effectiveness of LLMs in literary translation tasks:

1. **Training Customization**:
   - Enhancing training protocols to focus more on literary texts may increase the models' adjustment to poetic forms and cultural subtleties.
2. **Contextual Deepening**:
   - Creating mechanisms that enable models to incorporate broader literary contexts, such as entire poems or a poet's complete works, could improve the quality of translations.

## References

1. Korean poetry in translation.
2. Label studio.
3. Lm studio.
4. Mistral ai.
5. Multidimensional quality metrics.
6. Kim Byong-sun. The present conditions and tasks in constructing the database of korean literary materials centering on the korean poetry corpus. *The Review of Korean Studies*, 8(4):105–139, 2005.
7. Marzena Karpinska and Mohit Iyyer. Large language models effectively leverage document-level context for literary translation, but critical errors persist, 2023.
8. Aziza Talat kizi Khidoyatova. The peculiarities of the development of korean literature in the 1970s–2000s. *International Journal of Multicultural and Multireligious Understanding*, 10(11):241–246, 2023.
9. Chenyang Lyu, Zefeng Du, Jitao Xu, Yitao Duan, Minghao Wu, Teresa Lynn, Alham Fikri Aji, Derek F. Wong, Siyou Liu, and Longyue Wang. A paradigm shift: The future of machine translation lies with large language models, 2024.
10. Matt Post and Marcin Junczys-Dowmunt. Escaping the sentence-level paradigm in machine translation, 2024.
11. Katherine Thai, Marzena Karpinska, Kalpesh Krishna, Bill Ray, Moira Inghilleri, John Wieting, and Mohit Iyyer. Exploring document-level literary machine translation with parallel paragraphs from world literature, 2022.
12. Longyue Wang, Chenyang Lyu, Tianbo Ji, Zhirui Zhang, Dian Yu, Shuming Shi, and Zhaopeng Tu. Document-level machine translation with large language models, 2023.
13. Longyue Wang, Zhaopeng Tu, Yan Gu, Siyou Liu, Dian Yu, Qingsong Ma, Chenyang Lyu, Liting Zhou, Chao-Hong Liu, Yufeng Ma, Weiyu Chen, Yvette Graham, Bonnie Webber, Philipp Koehn, Andy Way, Yulin Yuan, and Shuming Shi. Findings of the wmt 2023 shared task on discourse-level literary translation: A fresh orb in the cosmos of llms, 2023.
14. Biao Zhang, Ankur Bapna, Melvin Johnson, Ali Dabirmoghaddam, Naveen Arivazhagan, and Orhan Firat. Multilingual document-level translation enables zero-shot transfer from sentences to documents. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4176–4192, Dublin, Ireland, May 2022. Association for Computational Linguistics.
15. Jiacheng Zhang, Huanbo Luan, Maosong Sun, Feifei Zhai, Jingfang Xu, Min Zhang, and Yang Liu. Improving the transformer translation model with document-level context. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 533–542, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

\

**Table 4.** Generative AI Tools Declaration

| Tool | Purpose | Where | Useful |
|---|---|---|---|
| Mistral AI | Prompting | Fig3,4 | + |
| LM studio | Prompting | Fig3,4 | + |
| Label studio | error annotation | Fig 6 | + |
| GPT4 | checking Code | translated.py | + |
| GPT4 | Reference and Create latex elements | Reference | - |
| GPT4 | getting insight for prompting | prompting | + |