

AI 문학번역에서 프롬프트 엔지니어링이 번역 오류와 창의성에 미치는 영향

김현웅(한국외국어대학교)

이상빈(한국외국어대학교)

1. 서론

대형언어모델(이하 “LLM”)¹⁾의 번역물이 신경망기계번역(이하 “NMT”)의 품질을 능가하는 것으로 밝혀지면서 LLM 번역 연구도 다양한 분야에서 이루어지기 시작했다(이창수, 2024; 홍서연, 2025; Jiao et al., 2023; Wang et al., 2023). 문학번역 분야도 예외는 아니다. NMT 번역기로는 한계가 명확했던 문학번역을 LLM이라면 극복이 가능할지 모른다는 기대 때문이었다(마승혜, 2018; 이준호, 2019; Almahasees, 2017; Voigt & Jurafsky, 2012).

하지만 LLM도 이내 한계를 드러냈다. 문학번역에서는 오역이나 누락이 없어야 함은 물론이고, 자연스러운 목표어로 원문의 다층적 의미를 전달할 수 있어야 하는데, LLM 역시 NMT 번역기처럼 출발어의 형태적 특성을 그대로 옮기거나 누락과 과잉번역(over-translation) 같은 오류를 범했다(이창수,

1) 생성형 AI는 사전학습 데이터를 이용해 텍스트, 이미지, 비디오, 음성 등 ‘어떠한 형태로든’ 콘텐츠를 생성하는 모델을 아우르는 개념이다. 따라서 텍스트만을 다루는 본 연구에서는 LLM이라는 표현이 좀 더 적절하다. 다만, 맥락상 기술적 구분이 필요한 경우에는 LLM 대신 생성형 AI라는 용어를 사용하였다.

2024; 홍서연, 2025; Karpinska & Iyyer, 2023; Pourkamali & Sharifi, 2024). 이 탓에 LLM도 유독 문학번역에 있어서는 품질이 기대에 못 미친다는 평가를 받았다(Khoshafah, 2023; Karabayeva & Kalizhanova, 2024; Manapbayeva et al., 2024).²⁾

이에 본 연구는 LLM 문학번역의 한계를 극복하기 위한 시도로 프롬프트 엔지니어링(prompt engineering)에 주목했다. 프롬프트 엔지니어링이란 생성형 AI가 사용자의 의도에 부합하는 결과물을 생성할 수 있도록 프롬프트를 체계적이고 전략적으로 설계하는 작업을 뜻한다. 이는 본래 자연어 처리와 생성형 AI 분야에서 출발한 개념이지만, 최근에는 AI 번역을 연구하는 번역학계로 관련 논의가 확대되고 있다(곽은주 등, 2025; 이상빈, 2025, pp. 129-156; 전현주, 2025; Moreno García & Mangiron, 2024).

본 연구의 목적은 프롬프트 엔지니어링 전략 가운데 하나인 ‘RGC 프롬프트’가 문학번역에서의 오류를 줄이고 창의성을 높이는 데 기여할 수 있는지를 확인하는 것이다. RGC 프롬프트란 AI가 수행할 ‘역할’(Role), AI가 달성할 ‘목표’(Goal), 작업의 기본 ‘맥락’(Context) 등을 담고 있는 전략적 프롬프트를 뜻한다. 본고에서 다룬 연구질문은 다음과 같다.

- (1) RGC 프롬프트는 LLM 문학번역의 번역 오류 감소에 기여하는가?
- (2) RGC 프롬프트는 LLM 문학번역의 번역 창의성 제고에 기여하는가?

이 질문에 답하기 위해, 프롬프트 엔지니어링을 거의 거치지 않은 기본 프롬프트와 세 가지 세부 정보를 담은 RGC 프롬프트를 사용해 소설 「술 권하는 사회」(현진건, 1921)를 각기 영역한 후, 두 번역물의 차이를 ‘번역 오류’와 ‘번역 창의성’의 측면에서 정량·정성적으로 분석한다. 미리 밝히는데, 본 연구의 핵심 용어인 번역 오류와 번역 창의성은 매우 제한적인 의미로 사용된다. 전자는 ‘과잉번역’, ‘누락’, ‘오역’이라는 부정적 결과를 통칭하고, 후자는 긍정적 측면을 내포하는 ‘다시쓰기’와 ‘첨가’를 가리킨다.

2) 반대로, LLM의 문학번역 활용 가능성을 긍정적으로 평가한 연구도 있다(우취영과 박지숙, 2024; Farghal & Haider, 2024; Guan, 2024).

2. 선행연구 검토

2.1 LLM 번역과 한계

LLM의 번역 성능이 주목할 만하다는 연구 결과가 적지 않다. 예컨대 LLM은 유창성이나 정확도 측면에서 기존 NMT 번역기를 능가하는 것으로 평가되기도 한다(이창수, 2024; 홍서연, 2025; Jiao et al., 2023; Wang et al., 2023). 하지만 LLM이 기계번역의 고질적 한계까지 모두 극복한 것은 아니다. NMT 번역에서 자주 확인됐던 과잉번역, 누락, 오역 등이 LLM 번역에서도 여전히 나타나기 때문이다(이창수, 2024; 홍서연, 2025; Karpinska & Iyyer, 2023; Pourkamali & Sharifi, 2024).

직역이라는 용어에서 연상되는 번역상의 문제도 LLM에 와서는 상당 부분 해소된 것으로 간주한다(Raunak et al., 2023; Zhao et al., 2023). 하지만 LLM도 ‘직역’(부자연스러운 축자역) 문제에서 자유롭지 않다는 점이 다양한 연구를 통해 실증적으로 확인되고 있다(이아형, 2023; Kong & Macken, 2025; Manapbayeva et al., 2024; Zhang et al., 2025). 가령, 리 등(Li et al., 2025)은 정보 중심 텍스트(informative text)를 대상으로 다양한 LLM에서 번역체(translationese)를 찾아 분석했고,³⁾ 그 결과 높은 비율의 직역 사례를 확인할 수 있었다. 대표 LLM인 GPT-4의 경우, 영중 번역에서는 전체의 43%, 독영 번역에서는 전체의 50%가량이 번역체였다.

직역이 특히 문제가 되는 경우는 표현 중심 텍스트(expressive text)를 다룰 때다. 예컨대, 단순한 의미 전달을 넘어 읽기 경험까지도 유지해야 하는 문학번역 분야에서는 과도한 직역이 가독성을 떨어뜨려 몰입을 방해하기 쉽다(신상범, 2016, p. 96). 따라서 문학번역에서는 원문의 문장 구성 및 형태를 그대로 유지하기보다는 독자의 수용성을 높일 수 있도록 어휘·문법(lexicogrammar)을 창의적으로 재구성하거나 필요한 내용을 덧붙이는 전략이 필요하다. 조원석 등(2022)에 따르면, 전문 번역가는 원문 소설의 문화적 메시지를 전달하기 위해 다시쓰기와 첨가 전략 등을 다채롭게 구사한다. 또한

3) ALMA-7B, ALMA-13B, Mistral-7B, GPT-3.5, GPT-4를 활용해 신문 기사, 과학 논문, 위키피디아 페이지, 소셜미디어의 번역을 분석하였다.

박옥수(2011)의 논문에서도 의미를 풀어쓰거나 내용을 추가하는 전략으로 번역 소설의 가독성을 제고한 사례들이 소개된다.

2.2 RGC 프롬프트

프롬프트 엔지니어링이란 생성형 AI가 사용자의 의도에 부합하는 응답을 생성할 수 있도록 프롬프트를 체계적으로 설계하는 과정이나 방법론을 의미한다. 프롬프트 엔지니어링은 다양한 실천적 방법을 통해 구현되는데, 이를 프롬프트 엔지니어링 기법이라 한다. 예를 들어, 수행시키고자 하는 작업과 유사한 예제 몇 가지를 제시해 패턴을 학습시키고 이를 따라 하도록 유도하는 퓨샷(few-shot) 방식은 가장 잘 알려진 기법이다. 이상빈(2025)이 다양한 사례를 통해 보여주었듯, 프롬프트 엔지니어링은 난해한 번역 과업에서 돌파구를 마련할 수 있다.

프롬프트가 번역에 미치는 영향은 최근 들어 조금씩 연구되기 시작했다. 대표적인 사례는 곽은주 등(2025), 류친(2024), 박정희와 강병규(2024)가 있는데, 이러한 연구는 프롬프트의 설계 방식과 세부 내용 등에 있어 큰 차이를 보인다. 예컨대, 박정희와 강병규(2024)는 번역 브리프, 줄거리, 인물 정보, 병렬코퍼스(퓨샷 효과 연출)와 같은 변인을 프롬프트 설계(GPT Builder)에 반영한 후, 중한 문학번역의 품질을 조사하였다. 그들에 따르면 다양한 조건을 고려한 프롬프트는 파파고(Papago)나 딥엘(DeepL) 같은 NMT 번역기보다 높은 수준의 번역물을 생성한다.

선행연구와 달리 본 연구는 한영 문학번역이라는 틀을 기반으로 ‘역할(Role) 부여’, ‘목표(Goal) 제시’, ‘맥락(Context) 제공’이라는 새로운 조건에 주목한다. 생성형 AI에게 가상의 전문인 역할을 부여하는 역할 부여는 다양한 분야에서 응답 품질을 개선한 것으로 알려져 있다(Chen et al., 2025; Kong et al., 2024). 또한 원하는 출력 형식을 명확히 지정하는 목표 제시 기법과, 생성형 AI가 맥락을 명확히 파악할 수 있도록 추가 정보를 제공하는 맥락 제공 기법도 프롬프팅의 기본 요소로 정의될 만큼 널리 받아들여지고 있다(전현주, 2025, p. 278). 이 세 가지 기법을 번역의 관점에서 간단히 소개하면 다음과 같다.

- 역할 부여4): 생성형 AI에게 독창적인 지식이나 전문성을 갖춘 특정 역할을 부여해, 주어진 맥락에 어울리는 응답을 생성하도록 유도하는 기법이다(Kong et al., 2024, pp. 1-2). 가령 “너는 수십 년간 한국문학을 영어로 번역해 온 전문 번역가야.”라는 문장을 입력해 ‘한국문학 번역가’라는 가상의 역할을 지정하는 것이다. LLM의 응답 범위를 해당 역할이나 관점으로 제한함으로써 맥락에 알맞거나 맞춤형된 결과를 생성하는 것이 핵심 원리다(Phoenix & Taylor, 2024, p. 106).
- 목표 제시: 사용자가 원하는 응답 형식을 구체적으로 설정하는 것은 프롬프트 설계의 기본 원칙과도 같다(Vairamani & Nayyar, 2025, pp. 5-6). LLM에게 기계번역을 요청하는 때에도 번역의 목표를 명확하게 제시해 주는 것이 좋다. 예컨대 “번역에서 원문의 톤과 스타일을 보존하라”는 “원문을 번역하라”보다 효과적인 지시문이 될 수 있다.
- 맥락 제공: 생성형 AI가 맥락에 알맞은 응답을 생성할 수 있도록 업무에 대한 맥락 정보를 제공하는 방법이다(Auffarth, 2023, p. 42; Tabatabaian, 2024, p. 2). LLM으로 번역물을 생성하는 상황이라면, 모델이 원문의 맥락을 고려하여 번역을 수행할 수 있도록 작품의 배경이나 주제, 주인공의 성격 묘사 등을 사전에 알려줄 수 있다. “이 원작 소설은 20세기 초 한국을 배경으로 한 사실주의 작품으로, 지식인 ‘나’의 내면 심리와 사회적 억압을 섬세하게 그려낸다...” 등이 맥락 제공에 해당한다.

프롬프트에 사용자의 의도를 충분히 반영하기 위해, 여러 기법을 동시에 활용할 수 있다. 앞서 소개한 역할·목표·맥락을 모두 고려한 소위 “RGC 프롬프트”(Role-Goal-Context prompt)도 점차 다양한 분야에서 그 효과가 논의되기 시작했다(Hattenhauer, 2024; Tabatabaian, 2024, p. 5). 가령, RGC 프롬프트는 요약의 정확성과 번역물의 BLEU 점수, 혹은 번역물의 문법적 정확성을 높이는 데 효과적인 것으로 보고된 바 있다(Higginbotham & Matthews, 2024; Yaqin et al., 2025). 본 연구에서도 제한적이거나 이 같은 결과에 주목해, 문학번역을 위한 RGC 프롬프트를 설계한 후 실험을 진행하였다.

4) 페르소나(persona)라는 명칭으로도 불린다.

3. 연구 방법

본 연구에서는 두 종류의 프롬프트, 즉 프롬프트 엔지니어링을 거의 거치지 않은 ‘기본 프롬프트’(baseline control prompt)와 역할·목표·맥락을 고려한 ‘RGC 프롬프트’를 사용해 한국어 단편소설을 영어로 번역한 후, RGC 프롬프트가 번역 오류를 줄이고 번역 창의성을 제고할 수 있는지 정량·정성적으로 분석한다.

실험에 사용할 원문으로 현진건(1921)의 「술 권하는 사회」를 선정했다. 이 소설을 택한 이유는 다음과 같이 크게 세 가지다. 첫째, 이 소설에는 오늘날 통용되지 않는 예스러운 표현과 문법이 다수 등장한다. 따라서 이 소설을 LLM으로 번역하면 오류가 발생할 가능성이 크고 창의적 접근이 요구되는 때도 많을 것으로 추정했다(텍스트의 질적 측면). 둘째, 단편소설 중에서도 분량(토큰 수: 7,038)⁵⁾이 적은 편에 속해, 텍스트 전체를 ChatGPT에 업·출력(번역)하기에 용이해 보였다(실험의 실용적 측면). 셋째, 이 작품은 ‘퍼블릭 도메인’에 속하므로 생성형 AI 도구에서 자유롭게 활용할 수 있었다(연구의 윤리적 측면). 원문 텍스트 파일은 《위키문헌》⁶⁾에서 추출했으며, 실험 결과의 타당성을 높이기 위해 연구자가 직접 발췌문의 오자 및 띄어쓰기를 일부 수정하였다.

3.1 프롬프트 설계

대조군에 해당하는 기본 프롬프트의 지시문은 쿼리(query), 즉 역할·목표·맥락을 포함하지 않은 기본 입력문으로 구성하였다. 그러나 쿼리만으로 구성된 단순 프롬프트(“Please translate the Korean text into English.”)로 예비 실험을 수행한 결과, 번역물에 중대한 문제가 발생했다.⁷⁾ 이를테면 아래 예

-
- 5) LLM에서 토큰은 모델이 텍스트를 쪼개어 처리하는 기본 단위이다. 예를 들어 “ChatGPT”는 “Chat”, “G”, “PT”라는 단위로 나눌 수 있다. 영어에서 1 토큰은 약 ¼ 단어 혹은 4자(문자) 정도에 해당한다.
 - 6) 위키문헌은 저작권이 자유로운 원문 텍스트와 그 번역문을 수록하고 있는 위키피디아 형식의 웹사이트다(<https://ko.wikisource.org/>).
 - 7) 분석 대상으로 삼은 구간의 길이는 기본 프롬프트의 경우 평균 3,154 토큰, RGC 프롬프트의 경우 평균 3,336 토큰이었다. 한편 예비 실험에서 확인한 번역물의 길

시에서 볼 수 있듯이, 원문의 문장 요소가 과도하게 삭제되거나 단순화되는 경향이 나타났다.

Source Text (ST): 나리는 혀를 억지로 돌리어 코와 입으로 대답을 하였다. 그래도 몸은 꿈적도 않는다. 도리어 그 개개 풀린 눈을 자려는 것처럼 스스로 감는다. 아내는 눈만 비비고 서 있다.

Target Text (TT): [...] he muttered, barely moving. His eyes slowly closed.

이런 번역물을 분석 대상으로 삼는 것은 연구의 목적에 부합하지 않으므로, 실제 실험에서는 ‘원문의 구조와 디테일을 최대한 보존하라’(이 지시문은 RGC 프롬프트에도 포함됨)는 요청을 쿼리에 추가해 출력물의 품질을 통제하였다. 실제 실험에서 기본 프롬프트로 사용한 입력문은 아래와 같다.

Please translate the Korean text into English, preserving its structure and details as much as possible.⁸⁾

한편, 실험군인 RGC 프롬프트의 지시문은 역할·목표·맥락을 모두 고려해 RGC 프롬프트의 형태로 설계하였다. 이는 실제 번역 상황에서 비교적 손쉽게 적용할 만한, 실현 가능성이 높은 프롬프트 설계 전략이기도 했다. 다음은 RGC 프롬프트의 입력문이다.

You are a professional translator who translates Korean novels into English. Now you will translate a Korean short story into English, preserving its structure and details as much as possible.

Follow the conditions⁹⁾ below:

- Ensure the translation reads smoothly and naturally, avoiding any awkward phrasing.

이는 1,979 토큰에 불과했다.

- 8) 내용을 명확히 전달하기 위해 모든 지시문은 초고자원 언어인 영어로 작성했다.
- 9) 이러한 내용을 당연한 것으로 해석해서는 안 된다. 인간과 달리 AI(기본 상태)는 기본 번역 원칙들을 내재화한 채 번역을 출력하지 않는다. AI가 특정 내용에 집중하도록 하려면, 해당 내용을 프롬프트에 적시해야 한다.

- *Reflect the tone and style of the original text.*
- *The text may contain old expressions or grammar, so be sure to fully understand its meaning.*

The text you will translate is part of a short story published in Korea in 1921. The story follows a woman who has waited alone for years for her husband, who went to study abroad in Tokyo. She never knew what his studies were about, but she believed that they were something that could grant any desires. When her husband finally returns, however, he appears anxious and depressed, and begins coming home drunk. One day, he arrives home intoxicated and blames 'society' for his behavior. However, the wife, not knowing what the word 'society' means, assumes it is the name of a restaurant. Unable to find comfort in her, the husband leaves—and the wife bitterly blames 'society' for driving him to drink.

Now, please translate the following Korean text into English like a professional translator, taking into account your role, as well as the conditions and the context I provided above.

위 영어 입력문의 내용을 기법별로 요약하면 다음과 같다.

- 역할 부여: ChatGPT에게 ‘한국어 소설을 영어로 번역하는 전문 번역사’의 역할을 부여한 후, 기본 프롬프트에서와 같은 취지의 쿼리를 덧붙여 구체적인 번역 업무를 요청하였다.
- 목표 제시: 번역 과정에서 ChatGPT가 따라야 할 목표(과업 세부 내용)를 다음과 같이 제시하였다. 첫째, 어색한 표현을 피하고, 매끄럽고 자연스럽게 읽히도록 번역하라. 둘째, 원문의 톤과 스타일을 보존하라. 셋째, 오래된 표현이나 문법이 포함되어 있을 수 있으니, 그 뜻을 정확히 파악하라.
- 맥락 제공: 인풋(input) 텍스트의 종류, 출판 연도, 출판 국가 등의 언어 외적 정보와 소설의 줄거리 정보를 간략하게 제시하였다.

세 가지 내용을 제시한 후에는 ‘리포커스’(refocus)—영어 프롬프트의 마지막 단락—를 추가하여 지시 사항을 강조하였다. 긴 프롬프트¹⁰⁾를 입력할 때 유용한 리포커스는 생성형 AI가 지시문의 핵심 내용을 잊지 않도록 유도

하여 출력물 내용의 타당성을 높인다(Berryman & Ziegler, 2025, p. 125).

3.2 번역물 생성 및 데이터 코딩

이렇게 설계한 기본 프롬프트와 RGC 프롬프트를 원문과 함께 입력해 번역물을 출력했다. 실험 결과의 신뢰도를 높이기 위해 프롬프트 유형별로 총 다섯 개의 번역문을 생성했다. 이때 각 번역물은 모두 “새 채팅”에서 생성하여, 이전 대화 기록의 간섭을 최소화하였다. ChatGPT가 한 번에 출력할 수 있는 최대 출력 토큰 수에 도달해 출력이 중단되면,¹¹⁾ “Please continue.”와 같은 프롬프트를 입력하여 정상적으로 출력을 재개했다. 모든 과정은 2025년 5월 8일 ChatGPT-4o Plus 환경에서 이루어졌다.

출력된 번역물에 대해서는 정량분석과 정성분석을 진행하였다. 정량분석의 경우, 번역물 10개를 모두 분석 대상으로 삼아 그 결과를 표로 정리한 뒤, ‘집단 간 평균 비교’ 등의 통계치를 산출하여 경향성도 파악했다. 이와 달리, 정성분석에서는 각 프롬프트 유형의 첫 번째 번역물(‘출력 1’)만을 골라 분석하였다. 지면 제약상 모든 번역물을 제시하기는 어려우며, 그렇다고 매번 다른 번역물에서 사례를 고른다면 일관성을 확보할 수 없기 때문이다.

다만, 번역 중 일부 구간에서 원문의 특정 구간이 번역물에 전해지지 않고 통제로 빠지는 일이 발생했다. 이처럼 문제가 되는 구간은 전체 분석의 타당성을 확보하기 위해 분석 대상에서 제외할 수밖에 없었다. 이에 더해, 누락 이후의 구간도 분석 범위에 포함하지 않았다. ChatGPT가 누락 이후의 구간을 온전히 ‘이해했다’고 보기가 어렵기 때문이다. 결과적으로 모든 번역물 10개의 후반부 30%가량이 일괄적으로 분석 대상에서 제외되었고, 나머지 70%에 해당하는 부분만을 분석 대상으로 삼았다.

10) 프롬프트가 지나치게 길면 출력물의 품질에도 영향을 줄 수 있다. 다만, 본 실험의 목표가 역할·목표·맥락의 통합 효과를 확인하는 것이기에 프롬프트의 길이를 만족스럽게 통제하기란 쉽지 않았다. 역할, 목표, 맥락을 분리하여 프롬프트를 만들 경우, 모델의 학습 효과로 인한 문제가 더욱더 커진다.

11) GPT 모델은 한 번에 출력할 수 있는 텍스트의 최대 길이(토큰 수)가 제한되어 있는데, GPT-4o의 경우 4,096 토큰을 초과하는 텍스트를 한 번에 출력할 수 없다. 본 실험에서 생성한 각 번역물의 길이는 4,000 토큰 안팎이었다.

구체적인 내용 분석은 ‘번역 오류’와 ‘번역 창의성’을 중심으로 이루어졌다. 본 연구에서 창의성은 바이에르-호엔바르터(Bayer-Hohenwarter, 2010, pp. 85-97)가 제시한 정의를 조작적으로 적용한 것이다. 바이에르-호엔바르터는 창의성의 구성 요소로 “수용성”, “유연성”, “참신성”, “유창성”을 제시한 바 있다. 수용성은 내용상 오류가 없고 스킴포스에 부합하는 번역, 나아가 독자 수용에 문제가 없는 번역을 뜻한다. 유연성은 단순한 축자역이나 직역에서 벗어난 번역을 가리킨다. 참신성은 비교 가능한 번역과 차이를 보이는 번역을, 유창성은 상대적으로 짧은 시간에 만든 번역을 의미한다. 바이에르-호엔바르터는 이 네 가지 지표를 종합하여 창의성을 분석했지만, 본 연구에서는 LLM 번역의 기본 특성을 고려해 수용성과 유연성만을 취하기로 했다. LLM은 출력 때마다 조금이라도 다른 번역을 거의 즉각적으로 생성하므로, 참신성이나 유창성을 측정하는 것은 무의미하거나 어려워 보였기 때문이다. 따라서 본 연구에서는 ‘창의적인 번역’을 매우 협소하게나마 ‘원문의 메시지 전달에 문제가 없으면서도 직역의 틀에서 벗어난 결과물, 그리고 독자 수용에 문제가 없는 결과물’로 정의한다.¹²⁾

번역 오류와 창의성은 각기 두 가지 이상의 하위 형태로 코딩하였다. <표 1>에서 확인할 수 있듯이, 번역 오류는 ‘과잉(번역)’, ‘누락’, ‘오역’으로, 창의성은 ‘다시쓰기’와 ‘첨가’라는 형태로 분류했다.

표 1
데이터 코딩 기준

번역 유형	번역 현상	설명
번역 오류	과잉	원문의 의미를 과도하게 확대해 번역물에 전함.
	누락	원문이 제시하는 정보를 번역물에 전하지 않음.
	오역	원문이 제시하는 정보를 왜곡해 번역물에 전함.
번역 창의성	다시쓰기	문장 구성, 어휘, 표현 방식을 변형하는 등, 원문을 재구성하여 독자 수용을 도울 수 있음.
	첨가	원문이 제시하지 않은 정보를 번역물에 새로 추가하여 독자 수용을 도울 수 있음.

12) 이는 쿠스마울(Kußmaul, 2000)의 관점과도 일치한다. 쿠스마울은 창의성의 핵심을 ‘원문과 달라짐’(ibid., p. 121) 및 ‘적절함’(ibid., p. 124)으로 정의한다.

연구자는 <표 1>의 코딩 기준에 따라 원문과 번역물을 대조하며 번역 현상을 직접 분류하였다. 이때 코딩 단위는 단어, 구, 문장 수준 가운데 각 사례를 아우를 수 있는 가장 작은 단위를 골라 적용하였다. 한국어 원문에 없던 주어를 영어 번역물에 표시하는 등의 의무적 변이는 고려하지 않았다.

코딩은 두 저자가 단계적으로 수행하였다. 1차 코딩은 제1저자가 주도적으로 했고, 그 결과를 교신저자가 검토하여 최종 데이터로 확정했다. 검토 과정에서 확인한 다섯 건의 불일치에 대해서는 논의를 통해 결정하였다.

4. 분석 결과

4.1 정량분석

먼저, 기본 프롬프트와 RGC 프롬프트로 출력한 모든 번역물을 정량적으로 분석하였다. <표 2>는 번역물에 나타난 과잉, 누락, 오역의 발생 횟수, 그리고 다시쓰기와 첨가의 발생 횟수¹³⁾를 정리한 것이다.

표 2
번역물에 나타난 번역 현상의 발생 횟수

번역물	프롬프트 유형	번역 오류				번역 창의성		
		과잉	누락	오역	합계	다시쓰기	첨가	합계
출력 1	기본	7	25	23	55	34	4	38
	RGC	13	17	22	52	62	11	73
출력 2	기본	3	9	17	29	31	6	37
	RGC	15	20	19	54	64	20	84
출력 3	기본	7	22	30	59	27	10	37
	RGC	15	24	20	59	57	20	77

13) 기본 프롬프트로 얻은 번역물과 RGC 프롬프트로 얻은 번역물이 길이 면에서 크게 다르지 않았다(평균 차이: 약 190 토큰). 따라서 일정 토큰당 오류 수가 아닌, 좀 더 직관적인 오류 발생 횟수로 소개한다.

출력 4	기본	5	8	21	34	41	6	47
	RGC	11	23	24	58	56	20	76
출력 5	기본	9	31	22	62	33	2	35
	RGC	12	15	22	49	49	13	62
평균	기본	6.2	19.0	22.6	47.8	33.2	5.6	38.8
	RGC	13.2	19.8	21.4	54.4	57.6	16.8	74.4

<표 2> 및 관련 데이터를 바탕으로 다음과 같은 세 가지 결론을 도출할 수 있다. 첫째, 전체적으로 번역 오류는 기본 프롬프트와 RGC 프롬프트 간에 뚜렷한 차이가 없었다. 각 출력에서 번역 오류의 총합을 따져보면, 두 프롬프트가 비슷하거나 같은 경우(출력 1과 출력 3), 기본 프롬프트가 더 큰 경우(출력 5), RGC 프롬프트가 더 큰 경우(출력 2와 출력 4) 등으로 다양하게 나타났다. 하지만 이들 출력의 합계 평균은 기본 프롬프트의 경우 47.8회, RGC 프롬프트의 경우 54.4회로 나타나, 두 프롬프트 간에 유의미한 차이가 없었다($t(8) = -0.94, p = .376, \text{Cohen's } d = 0.59$).

둘째, 번역 오류의 하위 유형에서는 프롬프트 간 차이가 있었다. 과잉 번역은 기본 프롬프트의 경우 평균 6.2회, RGC 프롬프트의 경우 평균 13.2회 발생하여, 프롬프트 간 유의미한 차이를 보였다($t(8) = -5.40, p < .001, \text{Cohen's } d = 3.42$). 이와 달리, 누락과 오역의 발생 수치는 프롬프트 간 큰 차이가 없었다. 누락은 기본 프롬프트의 경우 평균 19.0회, RGC 프롬프트의 경우 평균 19.8회 나타났고, 오역은 기본 프롬프트의 경우 평균 22.6회, RGC 프롬프트의 경우 평균 21.4회 나타났다(누락: $t(8) = -0.17, p = .873, \text{Cohen's } d = 0.10$; 오역: $t(8) = 0.53, p = .614, \text{Cohen's } d = 0.33$).

셋째, 번역 창의성은 RGC 프롬프트에서 더 많이 발생했다($t(8) = -8.56, p < .001, \text{Cohen's } d = 5.41$). 다시쓰기는 기본 프롬프트의 경우 평균 33.2회, RGC 프롬프트의 경우 평균 57.6회 발생했고, 첨가는 기본 프롬프트에서 평균 5.6회, RGC 프롬프트에서 평균 16.8회 나타났다. 두 하위 영역에서 프롬프트 간 차이는 통계적으로 유의미했다(다시쓰기: $t(8) = -7.01, p < .001, \text{Cohen's } d = 4.43$; 첨가: $t(8) = -4.69, p < .01, \text{Cohen's } d = 2.97$).

4.2 정성분석

지금부터는 앞서 논의한 번역 변이들이 어떤 양상으로 나타났는지를 영역별로 확인한다.

4.2.1 오류: 과잉번역

맥락상 오역은 아니지만 원문의 뜻을 지나치게 확대한 경우는 RGC 프롬프트를 사용할 때 훨씬 더 자주 발생했다. <예문 1>과 <예문 2>가 대표적인 사례다.

<예문 1>

ST: [아내는] 두어 번 얼굴을 쓰다듬자 불현듯 밖으로 나왔다.

TT-기본: After rubbing her face once or twice, she rushed outside.

TT-RGC: She rubbed her face a couple of times and, with a sudden burst of energy, stepped outside.

<예문 1>에서 아내는 남편의 귀가를 기다리다 깜빡 잠이 들고, 마침내 남편이 돌아오자 선잠에서 깨어 그를 맞이하러 나간다. TT-RGC에 덧붙여진 “with a sudden burst of energy”는 “불현듯”의 뜻을 과하게 표현한 것이다.

<예문 2>

ST: [남편은] 도리어 그 개개 풀린 눈을 자려는 것처럼 스프르르 감는다.

TT-기본: Instead, his bleary eyes slowly closed as if preparing for sleep.

TT-RGC: Instead, his bleary eyes fluttered shut again as if to sleep, ignoring everything around him.

<예문 2>에서 남편은 술에 취해 몸을 꿈적하지도 못하고 잠이 든다. TT-RGC에는 “ignoring everything around him”이 추가되었는데, 이러한 확장은 문장의 기본 의미를 전달하는 데 있어 문제가 되지 않으나 원문의 의미론적 범위를 넘어선 것이다.

4.2.2 오류: 누락

번역에서 일부 정보가 완전히 소실되는 경우는 두 프롬프트에서 비슷하

게 나타났다. 이와 같은 사례는 <예문 3>과 <예문 4>에서 잘 드러난다.

<예문 3>

ST: 할멈은 중간턱을 넘어설 제, 몇 번 혀를 차고는, 갈 데로 가버렸다.

TT-기본: [할멈은 중간턱을 넘어설 제] The old woman clucked her tongue a few times and departed.

TT-RGC: As the servant stepped down from the threshold, she clicked her tongue several times and returned to her room.

<예문 3>에서 할멈은 술에 취한 나리(남편)가 가까스로 일어나자, 혀를 차며 다른 곳으로 가버린다. TT-RGC와 달리 TT-기본에서는 “할멈은 중간턱을 넘어설 제”라는 정보가 완전히 사라졌다.

<예문 4>

ST: 막 그의 남편이 서울서 중학을 마쳤을 제 그와 결혼하였고, 그러자마자 고만 동경(東京)에 부급한 까닭이다.

TT-기본: They married just as her husband finished middle school in Seoul, and he had immediately gone off to study in Tokyo.

TT-RGC: They had married just after he'd finished middle school in Seoul, and no sooner had they wed than he left for [부급] Tokyo.

<예문 4>에서는 TT-RGC에서 누락이 발생했다. 부급은 ‘타향으로 공부하러 가다’라는 뜻인데, TT-RGC에는 이 의미가 반영되지 않았다.

4.2.3 오류: 오역

원문의 일부 의미를 왜곡한 경우도 두 프롬프트에서 비슷하게 발생했다. 오역 사례는 <예문 5>와 <예문 6>에서 확인할 수 있다.

<예문 5>

ST: 그 상처를 누른 채 그는 바느질고리에 눈을 주었다.

TT-기본: Still pressing the wound, she glanced at the sewing ring.

TT-RGC: Still pressing the wound, she glanced toward her sewing basket.

<예문 5>에 등장하는 옛말 “고리”는 상자 같은 물건을 일컫는다. 따라서

“바느질고리”는 바느질 도구를 담은 용기를 뜻한다. TT-RGC에서는 “sewing basket”처럼 올바르게 번역됐지만, TT-기본에서는 고리가 일차적 의미(둥글게 만든 물건)인 “ring”으로 번역되어 의미가 이상하게 변했다.

<예문 6>

ST: 검푸른 밤 빛이 허연 길 위에 그물그물 깃들었을 뿐이었다.

TT-기본: The bluish night shadows wavered over the pale path.

TT-RGC: Only the bluish hue of the night filtered through the pale street
like a net.

<예문 6>에서도 어휘의 외형적 형태에서 비롯된 황당한 오역이 확인된다. 부사 “그물그물”은 불빛 따위가 비치지 않고 몹시 침침해진 모양을 가리킨다. 이 어휘가 TT-RGC에서는 “like a net”(그물처럼)으로 번역되어, 원문을 이해하기가 어렵게 되었다.

4.2.4 창의성: 다시쓰기

앞서 확인했듯이, 번역 창의성의 하위 영역인 다시쓰기와 첨가는 TT-RGC에서 훨씬 더 자주 발생했다. 다시쓰기 사례는 <예문 7>과 <예문 8>에서 잘 드러난다.

<예문 7>

ST: 걸핏하면 성을 낸다. 입버릇 모양으로 화난다, 화난다 하였다.

TT-기본: At the slightest thing, he would flare up in anger, often repeating, “I’m furious, furious,” as if it were a habit.

TT-RGC: [...] snapping at her over the smallest things. It became a habit—
he’d lose his temper at the slightest provocation.

<예문 7>에는 걸핏하면 화를 내는 남편이 언급된다. TT-기본에서 “화난다, 화난다”는 원문의 화법적 특징이 그대로 반영된 “I’m furious, furious”로 번역되었다. 이에 반해, TT-RGC에서는 원문의 표층 구조를 크게 벗어난 번역이 이루어졌다.

<예문 8>

ST: “무슨 말이야, 무슨 말이야. 왜 오늘 일을 내일로 미루어. 할 말이 있거든 지금 해!”

TT-기본: “What do you mean, in the morning? Why postpone today’s matters? If you have something to say, say it now.”

TT-RGC: “What are you saying? What’s with this ‘talk later’ nonsense? If there’s something to say, say it now!”

<예문 8>에서는 아내가 남편(화자)의 하소연 섞인 술주정을 내일로 미루려 하자, 남편은 지금 당장 이야기를 하자며 아내에게 소리친다. ST의 밑줄 친 부분은 TT-기본에서 다소 어색한 “Why postpone today’s matters?”로 직역되었다. 이에 반해, TT-RGC에서는 어휘문법적으로 크게 다른 “What’s with this ‘talk later’ nonsense?”(‘나중에 얘기하자’니, 무슨 말 같지도 않은 소리야?)로 번역되어, 남편의 감정 상태와 문맥적 의미를 적절히 전달한다.

4.2.5 창의성: 첨가

원문에 없던 정보가 추가된 사례는 다음과 같다.

<예문 9>

ST: “옳지, 누가 나에게 술을 권했던 말이요? 내가 술이 먹고 싶어서 먹었던 말이요?”

TT-기본: “Right. Who told me to drink? Did I want to drink?”

TT-RGC: “Right. Who told me to drink? Did I drink because I wanted to? Because I enjoyed it?”

<예문 9>에서는 술에 취해 귀가한 남편이 술을 마시게 된 피치 못할 이유가 있었음을 아내에게 항변하고 있다. 남편은 억울하던 말투로, 아내에게 비꼬듯 반문한다. 발췌문에서 볼 수 있듯이, TT-기본은 남편의 말을 거의 그대로 옮긴 것이다. 이와 달리 TT-RGC는 “Because I enjoyed it?”(내가 술을 즐겨 마시니까?)을 덧붙인 것으로, 남편의 비꼼과 감정 상태를 강조한다.

<예문 10>

ST: 저…… 첫째는 환증이 술을 권하고 둘째는 ‘하이칼라’가 약주를 권

하지요.

TT-기본: Well... first, frustration did. Second, the high-collar folks offered you drinks.

TT-RGC: Hmm... First, it's your temper that made you drink, and second, those 'high collars'—those fancy types.

<예문 10>에는 “하이칼라”라는 표현이 등장한다. 하이칼라는 ‘깃을 높이 세운 옷차림’에서 유래한 일본식 조어로, ‘유행을 좇아 서양풍을 뽐내며 거들먹거리는 사람’을 가리킨다(Yumoto, 1996/2004, p. 228). 영어를 차용한 표현이지만 실제 영어권에서는 위와 같은 의미로 사용되지 않기에, 번역물 내에서 그 의미를 명확히 할 필요가 있다. 두 번역물에서 “하이칼라”는 비슷한 방식으로 음차되었지만, TT-RGC에서는 “those fancy types”가 추가되어 의미가 상대적으로 명확해졌다.

5. 결론

5.1 분석 결과 요약

본 연구에서는 역할·목표·맥락을 기반으로 설계한 RGC 프롬프트와 프롬프트 엔지니어링을 거의 거치지 않은 기본 프롬프트를 통해 「술 권하는 사회」의 영어 번역물을 생성한 다음, RGC 프롬프트가 번역 오류 감소 및 창의성 제고에 기여하는지를 살펴보았다.

분석 결과, 오류의 발생 횟수는 기본 프롬프트와 RGC 프롬프트 간에 유의미한 차이가 없었으나, 특정 오류 유형에서는 프롬프트 간에 뚜렷한 차이가 드러났다. 특히 과잉번역은 기본 프롬프트에서 평균 6.2회, RGC 프롬프트에서 평균 13.2회 발생했는데, 이는 RGC 프롬프트가 과잉번역의 증가와 관련되어 있음을 의미한다. 반면 누락과 오역의 경우에는 프롬프트 유형 간 큰 차이가 관찰되지 않았다. 누락은 기본 프롬프트와 RGC 프롬프트에서 각각 평균 19.0회와 19.8회 발생했고, 오역은 각각 평균 22.6회와 21.4회 발생했다. 오류 사례들을 자세히 확인한 결과, 누락이나 과잉번역은 작품 해석에

중대한 영향을 미치지 않았으나, 오역은 작품 이해를 방해할 수 있는 문제적 표현을 포함하고 있었다.

한편, 창의성 측면에서는 RGC 프롬프트가 긍정적인 효과를 가져온 것으로 나타났다. 다시쓰기는 기본 프롬프트와 RGC 프롬프트에서 각각 33.2회와 57.6회 발생했고, 첨가는 각각 5.6회와 16.8회 발생하면서 프롬프트 간에 유의미한 차이를 드러냈다. 정성분석에 따르면, 다시쓰기를 통해 번역이 영어권 화법에 가까워지고 첨가를 통해 의미가 명확해지는 등, 일부 사례에서는 독자 수용성이 높아진 듯 보였다.

따라서 본 연구의 틀 속에서, 앞서 제시한 두 가지 연구질문에 대해 다음과 같은 결론을 내릴 수 있다. 첫째, RGC 프롬프트는 번역 오류 감소에 뚜렷한 효과가 없다. 둘째, RGC 프롬프트는 번역 창의성 개선에 뚜렷한 효과가 있다.

5.2 연구의 한계와 의의

본 연구에는 다음과 같은 한계가 존재한다. 첫째, 본 연구는 제한된 실험 환경과 대상을 통해 이루어졌기에 그 결과를 일반화할 수 없다. 다른 AI 모델, 다른 언어 쌍, 다른 원문을 대상으로 실험을 진행하면 세부 결과가 달라질 수도 있으며, 같은 취지의 프롬프트라도 문구를 달리하면 출력 결과에 차이가 나타날 수 있다(Jiao et al., 2023). 둘째, 본 연구에서는 역할, 목표, 맥락 제시 기법을 하나의 입력문으로 한꺼번에 입력하였기에, 결과에 대한 해석이 제한적일 수밖에 없었다. 후속 연구에서는 각 기법이 번역에 미치는 영향을 개별적으로 분석함으로써 프롬프트 엔지니어링이 번역 품질에 미치는 영향을 좀 더 면밀히 파악해야 한다. 셋째, 코딩의 신뢰도를 제고할 필요가 있었다. 연구자가 외부 전문가와 함께 코딩을 수행했거나 복수의 코딩 기법을 적용했다면, 좀 더 객관적인 결과 도출도 가능했을 것이다.

위와 같은 한계에도 불구하고, 본 연구는 국내에서 LLM 번역에 프롬프트 엔지니어링을 접목한 초기 연구라는 점에서 큰 의미가 있다. 특히, 그간 잘 알려지지 않은 RGC 프롬프트를 구상해 번역 현상을 다섯 가지 영역으로 세분화해 살펴봤다는 점은 선행연구와 차별화되는 대목이다. 본 연구가 소개한 실험 설계는 비록 간단한 구상을 통해 이루어졌지만, 프롬프트 엔지니

어링의 효과를 탐구하려는 연구자들에게 유용한 참고 사례가 될 수 있다.

참고문헌

<1차 자료>

현진건. (1921). 술 권하는 사회. 위키문헌. Retrieved May 2, 2025, from https://ko.wikisource.org/wiki/술_권하는_사회

<2차 자료>

곽은주, 탁진영, 전현주. (2025). 생성형 AI 프롬프팅과 번역 효과에 관한 연구. 통역과 번역, 27(1), 1-39.

류친. (2024). GPT를 활용한 중-한 문학번역에 대한 고찰 — 장편 소설 “인생(活著)”의 관용구 번역을 중심으로. 문화와융합, 46(1), 273-287.

마승혜. (2018). 문학작품 기계번역의 한계에 대한 상세 고찰. 통번역학연구, 22(3), 65-88.

박옥수. (2011). 한영 문학번역에서 충실성과 가독성을 반영하는 방식: 「무진기행」의 세 편의 번역물 분석을 중심으로. 비교문학, 55, 53-79.

박정희, 강병규. (2024). 챗GPT에 기반한 중국문학작품 번역 방법론 탐구. 중국문학, 119, 169-200.

신상범. (2016). 한영 문학번역의 역동적 등가와 가독성 연구: 『부랑일기』를 중심으로. 통번역교육연구, 14(1), 95-117.

우취영, 박지숙. (2024). 챗GPT 등장 이후의 인공지능(AI)과 문학 번역 — 채만식 『민족의 죄인』 속의 대화문을 중심으로. 중국학보, 110, 117-143.

이상빈. (2025). AI와 통번역의 미래. 한국문화사.

이아형. (2023). 챗GPT의 번역 수업 활용 방안 고찰 — 한중과학기술번역을 중심으로. 중국어문학지, 85, 153-176.

이준호. (2019). 문학번역 적용을 위한 기계번역의 현주소. 통번역학연구, 23(1), 143-167.

이창수. (2024). 챗GPT, 파파고, 인간 번역가 간의 한영 문학번역 차이점 연

- 구. 번역학연구, 25(2), 11-37.
- 전현주. (2025). 디지털 시대의 창의적인 글쓰기 도구, 생성형 AI의 프롬프팅에 관한 연구. *디지털콘텐츠학회논문지*, 26(2), 275-285.
- 조원석, 이화연, 김슬기, 이상빈. (2022). 한국단편소설 번역의 자기교정 사례 분석 — 브루스/주찬 풀턴의 번역 *A Dream of Good Fortune*을 기반으로. *통번역학연구*, 26(4), 139-160.
- 홍서연. (2025). AI 번역기의 한러 번역성능 비교 — 파파고, 구글, 챗GPT를 중심으로. *통번역학연구*, 29(1), 207-233.
- Almahasees, Z. (2017). Machine translation quality of Khalil Gibran's *The Prophet*. *Arab World English Journal for Translation and Literary Studies*, 1(4), 151-159.
- Auffarth, B. (2023). *Generative AI with LangChain*. Packt Publishing.
- Bayer-Hohenwarter, G. (2010). Comparing translational creativity scores of students and professionals: Flexible problem-solving and/or fluent routine behaviour? In S. Göpferich, F. Alves & I. M. Mees (Eds.), *New approaches in translation process research* (pp. 83-111). Samfundslitteratur.
- Berryman, J., & Ziegler, A. (2025). *Prompt engineering for LLMs*. O'Reilly Media.
- Chen, Y.-C., Lee, S.-H., Sheu, H., Lin, S.-H., Hu, C.-C., Fu, S.-C., Yang, C.-P., & Lin, Y.-C. (2025). Enhancing responses from large language models with role-playing prompts: A comparative study on answering frequently asked questions about total knee arthroplasty. *BMC Medical Informatics and Decision Making*, 25(196), 1-8.
- Farghal, M., & Haider, A. S. (2024). Translating classical Arabic verse: Human translation vs. AI large language models (Gemini and ChatGPT). *Cogent Social Sciences*, 10(1), 1-15.
- Guan, X. (2024). A Bard is born: A new era of poetry translation by ChatGPT-4. *Translation Review*, 120(1), 23-39.
- Hattenhauer, R. (2024). *ChatGPT & Co*. Rheinwerk Computing.
- Higginbotham, G. Z., & Matthews, N. S. (2024). *Prompting and in-context*

- learning: Optimizing prompts for Mistral Large*. Optimus Research. <https://doi.org/10.21203/rs.3.rs-4430993/v1>
- Jiao, W., Wang, W., Huang, J., Wang, X., Shi, S., & Tu, Z. (2023). *Is ChatGPT a good translator? Yes with GPT-4 as the engine*. <https://doi.org/10.48550/arXiv.2301.08745>
- Karabayeva, I., & Kalizhanova, A. (2024). Evaluating machine translation of literature through rhetorical analysis. *Journal of Translation and Language Studies*, 5(1), 1-9.
- Karpinska, M., & Iyyer, M. (2023). Large language models effectively leverage document-level context for literary translation, but critical errors persist. In P. Koehn, B. Haddow, T. Kocmi & C. Monz (Eds.), *Proceedings of the Eighth Conference on Machine Translation* (pp. 419-451). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.wmt-1.41>
- Khoshafah, F. (2023). *ChatGPT for Arabic-English translation: Evaluating the accuracy*. Research Square. <https://doi.org/10.21203/rs.3.rs-2814154/v2>
- Kong, A., Zhao, S., Chen, H., Li, Q., Qin, Y., Sun, R., Zhou, X., Wang, E., & Dong, X. (2024). Better zero-shot reasoning with role-play prompting. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4099-4113). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.naacl-long.228>
- Kong, D., & Macken, L. (2025). *Decoding machine translationese in English-Chinese news: LLMs vs. NMTs*. <https://doi.org/10.48550/arXiv.2506.22050>
- Kußmaul, P. (2000). Types of creative translating. In A. Chesterman, N. Gallardo San Salvador & Y. Gambier (Eds.), *Translation in context* (pp. 117-126). John Benjamins.
- Li, Y., Zhang, R., Wang, Z., Zhang, H., Cui, L., Yin, Y., Xiao, T., & Zhang, Y. (2025). Lost in literalism: How supervised training shapes translationese in LLMs. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics* (pp. 12875-12894). Association

- for Computational Linguistics. <https://doi.org/10.48550/arXiv.2503.04369>
- Manapbayeva, Z., Zaurbekova, G., Ayazbekova, K., Kazezova, A., & Pirmanova, K. (2024). AI in literary translation: ChatGPT-4 vs. professional human translation of Abai's poem 'Spring.' *Procedia Computer Science*, 251, 526-531.
- Moreno García, L. D., & Mangiron, C. (2024). Exploring the potential of GPT-4 as an interactive transcreation assistant in game localisation: A case study on the translation of Pokémon names. *Perspectives*, 1-18.
- Phoenix, J., & Taylor, M. (2024). *Prompt engineering for generative AI*. O'Reilly Media.
- Pourkamali, N., & Sharifi, S. E. (2024). *Machine translation with large language models: Prompt engineering for Persian, English, and Russian directions*. <https://doi.org/10.48550/arXiv.2401.08429>
- Raunak, V., Menezes, A., Post, M., & Hassan, H. (2023). Do GPTs produce less literal translations? In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (pp. 1041-1050). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.acl-short.90>
- Tabatabaian, M. (2024). *Prompt engineering using ChatGPT*. Mercury Learning and Information.
- Vairamani, A. D., & Nayyar, A. (2025). *Prompt engineering*. CRC Press.
- Voigt, R., & Jurafsky, D. (2012). Towards a literary machine translation: The role of referential cohesion. In D. Elson, A. Kazantseva, R. Mihalcea & S. Szpakowicz (Eds.), *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature* (pp. 18-25). Association for Computational Linguistics.
- Wang, L., Lyu, C., Ji, T., Zhang, Z., Yu, D., Shi, S., & Tu, Z. (2023). Document-level machine translation with large language models. In H. Bouamor, J. Pino & K. Bali (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 16646-16661). Association for Computational Linguistics. <https://doi.org/>

10.18653/v1/2023.emnlp-main.1036

- Yaqin, L. N., Hassan, H., & Yusof, B. (2025). Performance and accuracy of ChatGPT in generating Malay academic texts: A comparative study with expert corrections. *LLT Journal: A Journal on Language and Language Teaching*, 28(1), 495-517.
- Yumoto, K. (2004). 일본 근대의 풍경 (정선태, 최정옥, 손애리, 이차원, 이경근, 손윤민, Trans.). (Original work published 1996)
- Zhang, R., Zhao, W., & Eger, S. (2025). How good are LLMs for literary translation, really? Literary translation evaluation with humans and LLMs. In *Proceedings of the 2025 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 9815-9836). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.naacl-long.548>
- Zhao, Y., Zhang, M., Chen, X., Deng, Y., Geng, A., Liu, L., Liu, X., Li, W., Jiang, Y., Yang, H., Han, Y., Tao, S., Xie, N., Li, X., Ma, M., & Zhang, Z. (2023). *Human evaluation for translation quality of ChatGPT: A preliminary study* (pp. 282-287). https://doi.org/10.26615/issn.2683-0078.2023_023

Effects of prompt engineering on error reduction and creativity in AI literary translation

Hyeon-woong Kim

Department of English Translation, Graduate School, Hankuk University of Foreign Studies, Seoul, South Korea

Sang-Bin Lee (corresponding author, sblee0110@naver.com)

Department of English for International Conferences and Communication, Hankuk University of Foreign Studies, Seoul, South Korea

Abstract

This study explores how highly calibrated prompts influence the translation of Hyŏn Chin'gŏn's 1921 short story "Sul kwŏnhanŭn sahoe" (*A Society That Drives You to Drink*). It draws comparisons between English translations generated with a Role-Goal-Context (RGC) prompt and those produced with a baseline control prompt ("Please translate the Korean text into English, preserving its structure and details as much as possible."). The results show no significant difference in the overall frequency of translation errors between the two prompt types. However, translations created with the RGC prompt exhibit a higher rate of over-translation, whereas rates of omission and mistranslation remain relatively stable across both conditions. Moreover, RGC-prompted translations demonstrate a greater use of creative strategies, such as paraphrasing and elaboration. These findings suggest that while RGC prompts may not reduce the overall incidence of translation errors, they enhance the contextual sensitivity and readability of the output.

Keywords: Prompt engineering; Role-Goal-Context (RGC) prompt; AI-assisted translation; ChatGPT; Large Language Model (LLM); translation of Korean literature

키워드: 프롬프트 엔지니어링, RGC 프롬프트, AI 번역, 챗GPT, 대형언어모델, 한국문학 번역

김현웅(제1저자, <https://orcid.org/0009-0002-7293-8092>)
한국외국어대학교 일반대학원 영어번역학과 석·박사통합과정(4학기)
khwoong24@gmail.com

이상빈(교신저자, <https://orcid.org/0000-0003-2140-3266>)
한국외국어대학교 영어대학 영어통번역학과 교수
sblee0110@naver.com

논문 투고일: 2025년 8월 4일
1차 심사 완료일: 2025년 9월 1일
2차 심사 완료일: 2025년 9월 7일
게재 확정일: 2025년 9월 15일