

# ChatGPT를 활용한 문화소 번역연구

-소설 『대장금』의 한중 번역을 중심으로-

방영풍\* 서연주\*\*

## -차 례-

1. 들어가는 말
2. 연구대상과 연구방법
3. ChatGPT를 활용한 문화소 번역
  - 3.1 문화 관련어
  - 3.2 관용어
4. 번역 품질 평가에 대한 결과
5. 맺는 말

---

\* 제1저자. 충남대학교 중어중문학과 한/중어 통·번역 석사학위과정생

\*\* 교신저자. 충남대학교 중어중문학과 조교수

## [국문초록]

본 연구는 조선 시대를 배경으로 한 역사 소설 『대장금』에 등장하는 문화소를 ChatGPT를 활용하여 번역, 그 결과를 인간 번역과 신경망 기계 번역인 파파고 및 바이두와 비교하고, ChatGPT의 번역 성능을 분석하는 것을 목적으로 한다. 연구 대상은 총 184개 문장으로, 인명 및 호칭어, 지명, 관직명, 음식명, 의학 용어 등 다양한 문화 관련 어휘와 관용어를 포함하였다. ChatGPT 4.0과 3.5는 특히 의학 용어와 같은 문화 관련 어휘 번역에서 인간 번역보다 우수한 성과를 보였으며, 다른 영역에서도 인간 번역에 근접한 결과를 나타냈다. 특히 ChatGPT 4.0은 3.5보다 더 높은 정확성과 완결성을 보였다. 반면, 파파고와 바이두는 직역 경향이 강하여 일부 번역의 정확성과 가독성이 낮게 평가되었다. 이는 ChatGPT가 방대한 학습 데이터와 지식을 바탕으로 번역 성능을 강화한 결과로 해석할 수 있다. 본 연구는 ChatGPT가 문화소 번역에서 신경망 기계 번역의 한계를 극복할 가능성을 제시하며, 향후 번역 품질 향상을 위한 방향성을 시사한다.

**주제어** : #신경망 기계번역 #생성형 인공지능 #한중 번역 #문학번역  
#문화소 번역 #문화소 #ChatGPT

## 1. 들어가는 말

본 연구는 조선 시대 배경의 역사 소설『대장금(大長今)』에 등장한 문화소를 ChatGPT를 활용하여 번역한 결과에 관한 것이다. 현대 사회에서 인공지능이 점점 발전함에 따라 기계 번역(Machine Translation)은 번역 학계에서 가장 뜨겁고 논쟁적인 주제 중 하나로 자리 잡았으며, 그 영향력도 점차 커지고 있다.

기계 번역이라는 용어는 1949년 워렌 위버(Warren Weaver)가 처음으로 언급했으며, 1988년에 통계 기반 기계 번역(Statistical Machine Translation, SMT)이 도입되면서 대규모 번역 데이터를 분석하여 문맥에 맞는 번역을 시도하는 방식이 우선 자리 잡았다. 통계 기반 기계 번역(SMT) 기술은 1990년대 초 IBM에서 제안한 방법으로, 수많은 병렬 말뭉치와 코퍼스에서 통계 분석을 통해 모델 매개변수를 학습하고, 이러한 모델을 기반으로 입력 문장을 번역하는 방식이었다.(황영숙, 2010: 91, 임순정, 2012: 185)

2016년 9월, 구글에서 최초로 딥러닝 기술의 발전과 함께 통계 기반 기계 번역 발전에 일조한 신경망 기계 번역(Neural Machine Translation, NMT) 서비스를 개시했다. 신경망 기계 번역은 문장 전체의 구조와 맥락을 더 잘 이해하고 자연스러운 번역을 제공하게 되었으며(서재형외, 2021: 70), 2022년부터는 대규모 언어 모델(Large Language Model, LLM)을 기반으로 한 생성형 기계번역, 대표적으로 ChatGPT 번역이 합류하게 되면서 번역계에 새로운 바람을 일으키고 있게 되었다.(이창수, 2023: 539)

신경망 기계 번역(NMT)은 원문(ST)을 전체적으로 분석한 후, 문장의 전체 맥락을 고려하여 문장을 최적 단위의 형태소로 분해한다. 그런 다음 각 형태소에 대한 최적의 번역어를 선택하고, 언어 데이터에 포함된 유사한 번역 쌍을 활용하여 최종 번역문(TT)을 생성한다.(남철진, 2018: 3) 이를 통해 번역가는 어려운 번역 과정에서 힌트를 얻고, 더 정확하고 풍부한 표현을 만들어낼 수 있으며, 시간과 비용 면에서도 효율성을 높일 수 있다.(문려화, 2023: 20)

신경망 기계 번역은 통계 기반 기계 번역을 넘어 경제성과 번역 품질

모두에서 우수한 성능을 보여준다.(강병규, 이지은, 2018: 264) 현재 한국에서는 구글 번역, 파파고 번역, 중국에서는 바이두 번역, 유다오 번역 등과 같은 플랫폼이 번역가에게 필수적인 도구로 자리 잡고 있다.

대규모 언어 모델(LLM)인 ChatGPT는 기계 번역 작업을 수행하는데 있어 원문에 대한 뜻을 파악하고 적합한 답을 형성하는 과정에서 번역 기능을 수행한다. ChatGPT는 훈련 자료로부터 언어 사이의 문법 구조와 법칙을 ‘습득’하여 이를 기반으로 번역 작업을 진행한다.(신세연 외, 2023: 158) 따라서 사용자는 ChatGPT와의 대화 과정에서 자신의 필요에 따른 프롬프트를 수시로 입력해 가며 자신이 구상했던 최적의 번역문을 기대할 수 있게 되었다.

대규모 언어 모델(LLM)의 출시 이후, 신경망 기계 번역과 생성형 기계 번역 간의 구분은 점점 뚜렷해지고 있다. 신경망 기계 번역은 특히 문학 번역에 취약한데, 문장 하나하나를 직역하여 의미 전달이 완전하게 이루어지지 않으며, 문학적 가치를 살리지 못하는 경우가 많기 때문이다.(김동미 2018, 홍은영, 탁진영 2020) 또한 문맥을 적절히 파악하지 못해 오류가 빈번하게 발생하며, 특히 문화적 요소를 번역하는 데에 한계가 드러났다.(김지은 2023, 문려화 2023, 노금송, 왕원 2024) 신경망 기계 번역 과정이 불투명하고 닫힌 알고리즘을 사용하기 때문에 번역가가 시스템 안에 간섭할 수 없는 점도 신경망 기계 번역의 한계로 지적된다.(지인영, 김희동 2020)

한편, 생성형 기계 번역은 출시된 지 오래되지 않아 신경망 기계 번역만큼 관련 연구가 활성화되지는 않았지만, 적어도 문학 분야에서 만큼은 기존 어떠한 번역 방법보다 더 뛰어난 성능을 보여주고 있으며, 더 많은 가치를 창출할 수 있다고 평가된다.(박정희, 강병규 2024) 이는 높은 활용 가능성과 효용성을 바탕으로 신경망 기계 번역보다 우수한 번역 완성도를 보이며, 인간 번역에 더 가깝고 인정을 받는다. 문학의 미적 전달 측면에서는 여전히 인간 번역에 미치지 못한 수준이지만(윤여법 2023, 이유정 2023; 이창수 2024; 신천 외 2024, 황지연 외 2024), 인간에게 인지시키며, 창의적인 보조 도구가 될 수 있기에(마승혜 2024), 인간 번역가와 협력할 때 시너지 효과를 기대할 수 있는 것이다.

이러한 배경아래 본 연구는 다음과 같은 세 가지 필요성을 가진다. 첫째, 본 연구는 기계 번역 분야에서 연구 성과가 상대적으로 적은 한중 번역을 대상으로 선정했다. 이때 중국에서도 번역 출판된 유민주 역사 소설 『대장금』을 바탕으로,<sup>1)</sup> 기계 번역이 한중 번역 시 어떤 성능을 보일지 비교 탐구하고자 한다. 이를 통해 문학 번역 중 한중 기계 번역 분야의 연구를 보완할 수 있을 것이라 기대한다.

둘째, 본 연구는 문화소 구성 요인 전반으로 번역 대상을 확장했다. 현재까지 기계 번역을 통한 문화소 번역에 대한 연구는 기본적으로 관용 표현을 위주로 진행되었다. 본고에서는 조선 시대에 사용되었던 특정 문화 관련 어휘(인명 및 호칭어, 지명 및 장소명, 관직명, 음식명, 의학 용어), 관용어(고사성어, 속담 및 관용구)가 포함된 문장을 뽑아 총 184개 문장을 연구 대상으로 삼았는데, 한 문장 안에 복수의 문화소가 등장하기도 하여 도합 257개의 문화소가 분석대상이 되었다.

셋째, 본고에서는 네 가지 주요 기계 번역 도구—ChatGPT 3.5, ChatGPT 4.0, 파파고, 바이두 번역을 사용하여 인간 번역을 기준으로 번역 결과를 비교 분석한다. 신경망 기계 번역의 예시는 한중 번역에서 오류가 빈번하게 발생하는 구글 번역기는 제외하고, 국내에서 활용도가 높고 한국어와 관련된 번역에서 강점을 보이는 파파고 번역과(김혜림, 2022: 289) 중국에서 광범위하게 사용되는 바이두 번역을 따랐다. 생성형 기계 번역은 근래 가장 대중적인 ChatGPT를 활용하되, 무료 버전인 ChatGPT 3.5와 유료 버전인 ChatGPT 4.0을 구분지었다.<sup>2)</sup>

본 연구를 통해 이들 한중 번역 결과를 비교 분석하고 품질 평가를

1) 본 논문에서 참조한 번역본 『大長今』은 2005년 薛舟와 徐丽红이 번역하고 北京人民文学出版社에서 인증 및 출판한 것으로 공신력과 정확성을 갖춘 도서이다. 薛舟, 徐丽红 역시 『신경숙의 엄마를 부탁해(번역판 '请照顾好我妈妈')』(2021), 김은희, 이인희의 『Signal(번역판 '信号')』(2023) 등 한국의 유명 작가 작품들을 중국어로 번역해 꾸준히 문화 가교 역할을 하고있는 베테랑 작가들이다.

2) 본 연구에서는 ChatGPT 3.5와 ChatGPT 4.0에 동일한 프롬프트를 설정했다. 특정 문화 관련 어휘를 번역할 때는 “이제부터 조선시대를 배경으로 한 한국 역사 소설의 문장들을 보낼 것입니다. 이 문장들을 시대적 배경을 고려하여 가독성 있게 자연스럽게 중국어로 번역해 주세요.”라고 했고, 관용어를 번역할 때는 “이제부터 보내는 문장에 관용어가 포함되어 있습니다. 가독성을 고려하여 충실하게 중국어로 번역해 주세요.”라고 설정했습니다.

진행하는 과정에서 생성형 기계 번역이 신경망 기계 번역, 인간 번역에 비해 어떤 강점과 약점을 가지고 있는지, 또 문학 번역에서의 활용 가치가 얼마나 있는지를 확인할 수 있을 것이다. 특히 ChatGPT의 유, 무료 버전의 번역 결과를 비교하는 과정에서는 생성형 기계 번역 기술의 진화 과정까지 엿볼 수 있을 것이다.

## 2. 연구대상과 연구방법

본고에서 다루는 『대장금』은 조선시대 궁중을 배경으로, 한국의 전통적 가치관, 문화, 사회적 구조, 그리고 역사적 사건들이 복합적으로 얽힌 작품이다. 이러한 문화적 요소들은 단순한 언어 변환으로는 온전히 전달되기 어렵다. 본 연구를 통해 기존 연구에서 신경망 기계 번역의 한계로 지적되어 온 문학적, 문화적 맥락에 대한 번역을 ChatGPT의 번역에서 어느 정도나 잘 수행하는지 알 수 있을 것이다.

본 연구에서 집중적으로 다루는 ‘문화소’에 대해서는 학자마다 정의와 분류를 달리 하고 있다.<sup>3)</sup> 한국에서는 김효중(2004: 72)이 ‘문화소’라는 개념을 처음으로 도입하여 특정 상황에서 문화적 요소가 유발하는 행위와 이해를 설명했으며, 이근희(2003: 6)가 ‘문화와 밀접한 상관이’라는 용어를 사용했다. 이후 이근희(2015)는 이를 ‘문화 관련 어휘’로 변경하며, ‘원천 언어를 사용하는 사회의 역사적, 사회적, 경제적, 정치적 배경과 언어적 관습 등에서 비롯된 독특한 문화적 어휘’라고 정의했다.

특히 이근희(2003)는 문화소를 여섯 개의 범주로 나누어 다음과 같

3) Newmark(1988)는 ‘Cultural words’를 특정 문화와 밀접하게 연관된 단어로 정의하며, 번역 시 문화적 차이를 고려해야 한다고 강조했다. Vermeer(1990)는 문화소를 특정 문화의 고유 개념으로 간주하고, 사회적 현상으로 정의하였다. Stolze(1992)는 도착어 문화권에서 특정한 연상을 유도하는 텍스트적 요소로 문화소를 설명하며, 문화적 차이에 따른 부등성을 여러 측면에서 분류했다. 중국에서는 1980년대 말부터 문화 간 소통에서 출발어와 목표어의 고유 특성을 강조하는 ‘文化因素(문화인소)’라는 개념이 등장했다. 이후 여러 학자들이 ‘文化缺省’과 ‘文化负载词’ 등의 용어를 사용했고, 张南峰(2004)이 ‘文化专有项’이라는 용어를 제시하면서 중국에서 인정받아 널리 사용되고 있다.

이 “(1) 고유명사: 사람 이름, 건물 이름, 단체명, 책 제목 등 (2) 특정 문화와 관련된 어휘: 의복, 음식, 주거, 지역, 사회적 관습 등 (3) 특정 사건과 관련된 단어 (4) 관용적 표현 (5) 단위 체계: 화폐, 거리, 높이, 무게 등 (6) 언어적 관습에서 유래한 표현: 지위나 직위와 이름을 함께 나타내는 표현, 날짜 표기 방식 등.”의 세부 항목으로 분류하였으며, 기존 연구들에서 대표적인 분류의 기준으로 활용되고 있다.(최미란 2021, 최연연 2023, 박금은 2024 재인용) 이후 이근희(2015)는 ‘관용구’를 ‘관용어’로 변경하며, 이를 ‘관용구, 사자성어, 은어 및 신조어’로 구성되는 것이라 정의했다.

본 논문에서는 이 두 가지를 종합하여 문화소 번역의 분류 기준을 만들었다. 우선 문화 관련 어휘에는 인명 및 호칭어, 지명 및 장소명, 관직명, 음식명, 의학 용어 등이 포함된다. 『대장금』은 조선시대 문화의 각양각색을 잘 드러내고 있는 만큼, 현재에는 더 이상 쓰이지 않는 궁중, 관식, 음식, 의학 용어들이 사건 전개에 따라 대거 등장한다. 따라서 관용어를 제외한 나머지 항목들을 세세히 구분하는 작업이 오히려 기준을 더욱 모호하게 만드는 면이 있었다. 본 연구에서는 이를 통틀어 ‘문화 관련어’라 명했다. 고사성어와 속담 및 관용구는 관용어로 분류하였다.

번역 결과를 비교 평가하는 기준은 노금송과 왕원(2024)이 제시한 번역 품질 평가 모델을 바탕으로 삼았으며, 문항 점수를 연구 목적에 맞춰 일부 수정하였다. 노금송과 왕원은 한국 소설에 나타난 한국어 문화소(특정 문화 관련 어휘, 관용어) 관련 문장을 번역 텍스트로 삼고, 기유미(2018), 정해연(2018), 권도경 외(2019), 황지연·양슬아(2020) 등의 선행 연구에서 제시된 인간 평가 모델을 종합하여 점수제 평가 기준을 만들었다. 이 평가 기준은 10점 만점을 기준으로 하며, 각 항목에 따라 오류가 발생한 부분에 대해서 감점하는 방식을 취한다.

<표 1> 번역 품질평가 모델

상위 항목	하위 항목	비중
정확성	문화소(고유명사, 특정 문화 관련어, 관용 표현 의미)의 정확성	3
	일반 단어의 정확성	1

	문법 표현의 정확성	1
완결성	원문의 의도와 내용을 누락 없이 반영한 번역문의 완성도	2
가독성	도착 언어에서 문장이 자연스럽고, 모국어 독자의 수용도	1
일관성	절과 절의 논리성	1
형식성	오타, 인용구 및 문장부호 등	1
합계		10

본 연구에서는 문화소 번역이 주인공름, 문화 관련어와 관용어 번역을 중심으로 ‘정확성’ 항목을 더 세분화하였다. 번역문에서 특정 문화 관련 어휘, 관용어 중 하나라도 오역이 발생할 경우 3점을 감점하였으며, 그 외 다른 단어에서 오류가 발생하면 1점을 감점하여 차등을 두었다. 또한 번역문에 문법적 오류가 있을 경우에는 1점을 감점했다. 아울러 문화소와 문법 번역에서 오역이 발생하면 문장 전체의 가독성도 자연스럽게 저하된다는 점을 감안하여, 해당 경우가 발생하면 가독성과 정확성면에서 중복 감점이 되지 않게끔 기존 평가 모델에서 2점으로 설정했던 것을 1점으로 하향 조정하였다.

### 3. ChatGPT를 활용한 문화소 번역

아래는 위의 번역 품질평가 모델을 사용해 평가 과정을 구체적으로 설명한 예시이다.

<예1>

ST: “그다음엔 곰탕이 저리 되었으니 무 조치를 끊어야겠다. 무를 다듬어 손가락으로 떠냐. 너무 두껍지 않게 뜨는 거 잊지 말고.....”

TT1(인간 번역): “牛肉汤已经煮好, 该把萝卜放进去了。你把萝卜收拾好, 然后用勺刮, 注意不要刮得太厚.....”

TT2(ChatGPT3.5): “接下来是炖好的牛骨汤, 要开始煮萝卜酱了。把萝卜切好, 尽量切薄些, 别忘了.....”

TT3(ChatGPT4.0): “接下来炖了牛骨汤, 还得煮萝卜酱汤。把萝卜切好, 用勺子挖起来, 不要挖得太厚, 别忘了……”

TT4(파파고 번역): “接下来牛骨汤已经做好了, 要煮萝卜条了。把萝卜收拾好, 用勺子舀起来。不要忘记浮起来不要太厚……”

TT5(바이두 번역): “然后牛骨汤变凉了, 要煮萝卜丝了。把萝卜捣碎后用勺子舀起来。别忘了舀得太厚……”

<예1>에 등장하는 “조치”는 바득하게 만든 찌개나 찜을 가리키는 말이다.<sup>4)</sup> “무 조치”는 현대어로는 “무 찌개”를 의미한다. 인간 번역은 “조치”의 의미를 잘못 이해하여 ‘곰탕에 무를 넣어야 한다’고 번역, 원문의 의미를 크게 왜곡했다. 이로 인해 문화소 번역 오류로 3점 감점하여 최종 7점이라 평가하였다. ChatGPT 3.5 번역 역시 ‘무 조치’를 “萝卜酱”으로 잘못 번역해, 정확성 부분의 오류로 3점 감점했다. 또한, “그다음엔 곰탕이 저리 되었으니”를 是자문으로 오역했으므로 문법 정확성에서 1점 감점, 최종 6점이라 평가하였다. ChatGPT 4.0 번역은 “萝卜酱汤”으로 정확한 등가 번역을 구현했다. 그러나 “그다음엔 곰탕이 저리 되었으니” 구문에서 문법 오류가 발생하여 ‘이어서 곰탕을 끓였다’로 오역했기에 문법 정확성 항목에서 1점 감점되어 최종 9점으로 평가했다.

파파고 번역은 “무 조치”를 “萝卜条”로 잘못 번역하여 문화소 오류로 3점 감점, “그다음엔”이라는 접속사를 오역하여 문법 어순 오류가 발생, 문법 정확성 항목에서 1점 감점하여 최종 6점이라 평가하였다. 바이두 번역은 “무 조치”를 “萝卜丝”로 오역해 문화소 오류로 3점 감점, “저리 되었으니”를 “变凉”으로, “무를 다듬어”를 “把萝卜捣碎”로 오역하여 일반 단어 정확성 항목에서 1점 감점, “그다음엔”이라는 접속사 어순 오류로 인해 문법 정확성 항목에서 1점 감점하여 최종 5점이라 평가하였다. 본 연구는 이상과 같은 번역 품질평가 방식을 사용하여 진행될 것이다.

『대장금』은 조선 시대를 배경으로 음식 문화, 의료 문화, 관직 문화,

4) 본 연구에서는 문화소에 대한 설명을 위해 『국립국어원 표준국어대사전』, 『고려대 한국어사전』, 『단국대 한국한자어사전』, 『新华字典』등을 참고 자료로 삼았다.

궁중 문화를 다양하게 다룬 소설이다. 상술했듯 이 소설에 등장하는 관직명, 음식명, 약물명, 병명 등은 고유명사이면서 특정 문화와 관련된 어휘이자 특정 사건과 관련된 어휘로, 중복되는 영역이 크다. 따라서 본 연구에서는 이를 통틀어 ‘문화 관련어’로 분류한다. 관용어로는 사자성어, 속담 및 관용구를 포함했다. 이 절에서는 이상 두 방면에 대한 구체적인 연구 결과를 다룬다.

### 3.1 문화 관련어

문화 관련어는 의(衣), 식(食), 주(住), 지역, 사회, 관습 등으로 구분할 수 있는 매우 광범위한 범주를 포괄한다.(이근희, 2003: 6) 『대장금』에 등장하는 문화 관련어는 절대 다수가 조선시대에 사용된 경우로 한정되기 때문에, 현재 한국어를 모국어로 사용하는 사람들에게도 낯설게 느껴질 수 있다. 『대장금』에 등장하는 인명 및 호칭어는 61개, 지명 및 장소명은 132개, 관직명은 122개, 음식명은 105개, 의학 용어는 286개이다. 해당 절에서는 이 중 각 문화소가 포함된 문장 20개씩을 연구 대상으로 선정해 총 100개의 문장을 분석 대상으로 삼은 셈이 되었다. 이 가운데 등장하는 인명 및 호칭어는 33개, 지명 및 장소명은 26개, 관직명은 35개, 음식명은 43개, 의학 용어는 36개이다. 우선 호칭어의 예시를 살펴보겠다.

<예2>

ST: “똑똑히 듣거라! 같은 정3품이라도 당상관과 당하관으로 구분한다. 또한 당상관은 영감, 당하관은 나으리라 호칭한다. 알고 있으렸다!”

TT1(인간 번역): “你听清楚了! 同为正三品, 可以分为堂上官和堂下官两种。堂上官称为令监, 堂下官称为大人。这个你知道吗?”

TT2(ChatGPT3.5): “听好了! 即使是正三品官员, 也要分为堂上官和堂下官。堂上官称为‘老爷’, 堂下官则称为‘大人’, 记住了吗!”

TT3(ChatGPT4.0): “听好了! 虽然同为正三品官员, 但会区分为堂上官和堂下官。另外, 堂上官称为‘令监’, 堂下官称为‘大人’。记住了吗?”

TT4(과과과 번역): “听清楚了! 即使是同样的正三品, 也分为党上官和党下

官。另外，还称呼堂上官为老头，堂下官为好。应该知道吧！”

TT5(바이두 번역)：“听清楚了！即使是同样的正三品，也分为堂上官和堂下官。而且堂上官称呼老头子，堂下官称呼大人。我知道！”

<예2>에 등장하는 “영감”은 정삼품과 종이품의 벼슬아치를 이르던 말, “나으리”는 당하관을 높여 부르던 말로 사용되었다. 중국어에서는 이와 대응하는 용어로 ‘令監’과 ‘大人’이 이미 존재한다.(百度百科) 인간 번역과 ChatGPT 4.0의 번역은 이국화 번역 전략을 적절히 사용하여 정확하게 번역되었다. 조선과 대응했던 명·청(明清) 시대의 중국에서는 관료들을 ‘老爺’, 혹은 ‘大人’이라 불렀으며, 이 두 호칭 간에 등급 차이는 없었다. ChatGPT 3.5의 번역은 “영감”을 “老爺”로 번역하여 가독성이 높아지긴 했으나, 정확성은 떨어졌다. 파과고 번역은 문맥을 완전히 잘못 짚어 “令監”과 “大人”을 각각 “老头(늙은이)”와 “好(좋다)”로 번역하여 고유명사 번역 오류가 발생했다. 또한, “당상관”과 “당하관”을 각각 “党上官”과 “党下官”으로 번역했는데, 이 둘은 모두 중국어에 아예 없는 표현으로, 독자들에게 혼란을 줄 가능성이 크다. 바이두 번역도 문맥을 제대로 파악하지 못해 “영감”을 “老头(늙은이)”로 오역했다.

### <예3>

ST: 한쪽에서는 생각시들이 야채를 비롯한 각종 재료들을 다듬었다.

TT1(인간 번역): 丫头们在旁边择菜, 或者准备其他材料.

TT2(ChatGPT3.5): 一旁, 思娘们正忙着处理各种食材, 包括蔬菜.

TT3(ChatGPT4.0): 一旁, 宫女们正在清理各种蔬菜和食材.

TT4(파과고 번역): 一边是思绪们整理蔬菜等各种材料.

TT5(바이두 번역): 一边是思想枯萎, 修剪了包括蔬菜在内的各种材料.

<예3>에서 “생각시”라는 표현은 조선 시대 지밀(至密), 침방(針房), 수방(繡房)에 소속된 궁녀 중, 관례(冠禮)를 치르지 않아 새앙머리를 땀은 어린 궁녀를 가리키는 말이다. “丫头”는 어린 여자아이 또는 심부름을 하는 어린 여자아이를 의미한다. 인간 번역은 “생각시”를 “丫头”라

고 의역하여 등가 번역을 구현했다. 반면, ChatGPT 4.0은 “생각시”를 “宮女”라고 번역했다. 이는 상위 개념으로서의 의미가 전달되긴 했지만 ‘어린이’라는 의미가 부족하다. 이를 ‘小宮女’라 번역했으면 보다 정확했을 것이다. ChatGPT 3.5는 “생각시”를 “思娘们(중국어에 없는 표현)”로 번역했고, 파파고 번역은 “思绪(생각)”로, 바이두 번역은 “思想枯萎(의기소침함)”로 잘못 번역하여 문화적 맥락 파악에 실패했으며, 이로 인해 문장의 가독성이 크게 떨어졌다. 또한 ‘생각시들이’가 주어였는데, 파파고와 바이두 번역은 조사 “는” 앞의 “한쪽에서”를 주어로 오해하고 이를 “是” 자문으로 번역하여 주술 관계를 혼동한 문법적 오류가 추가로 발생했다.

<예4>

ST: 조선시대에는 타락색(駝酪色)이라 하여, 우유를 취급하는 기관이 따로 있었다.

TT1(인간 번역): 朝鮮时代把牛奶叫做駝酪, 并有专门负责供应牛奶的部门。

TT2(ChatGPT3.5): 朝鮮时代有一个专门管理牛奶的机构, 称为駝乳色(駝酪色)。

TT3(ChatGPT4.0): 朝鮮时代设有专门管理奶制品的机构, 称为“駝酪色”(駝酪色)。

TT4(파파고 번역): 朝鮮时代名为墮酪色(駝酪色), 另有经营牛奶的机构。

TT5(바이두 번역): 朝鮮时代, 因为是墮酪色, 所以有专门经营牛奶的机构。

<예4>에서 “타락색”은 조선시대에 임금이 음용할 우유에 관한 일을 맡아보았던 사복시(司僕寺)에 딸린 부서(部署)를 의미한다. 타락(駝酪)<sup>5)</sup>은 우유 또는 양유를 끓여 만든 음료이다. 인간 번역에서는 “타락색”을 생략하고, 조선 시대에 우유를 “駝酪”이라고 불렀다는 부분에만 초점을 맞추어 번역했다. 그 결과 원문에서 우유를 공급하는 기관이 ‘타락색’이라는 의미를 제대로 전달하지 못했다. ChatGPT 4.0 번역은 원문을 완전히 이해하여 “타락색”을 ‘기관’이라는 의미로 정확하게 번역했다. 반면 ChatGPT 3.5 번역에서는 “락(酪)”을 “乳”로 오역하여 어휘 대체 오류가 나타났다. 파파고 번역은 “타락색”을 축자 번역하여

5) 타락은 몽골제국 시기에 몽골 유목민들의 발효 유제품 ‘타락(tarak)’이 고려로 전해지며 유래된 말로, 한자로는 ‘駝酪’이라 표기되었다.

문장의 가독성이 떨어졌다. 바이두 번역 또한 같은 오류를 범하여 문맥에 맞지 않는 오역을 보였고, “하여”를 인과관계로 인식하여 “因为”와 “所以”로 번역하면서 논리적 관계를 왜곡하는 문법적 오류가 발생했다.

<예5>

ST:의녀를 내의(內醫), 간병의(看病醫), 초학의(初學醫) 세 등급으로 대별하여 각각 하는 일과 녹봉을 구분했다.

TT1(인간 번역): 此时医女被区分为内医、看病医、初学医三个等级, 各司其责, 各领其俸.

TT2(ChatGPT3.5): 医女分为内医、看病医和初学医三个等级, 分别从事不同的工作, 领取不同的俸禄.

TT3(ChatGPT4.0): 把医女分为内医、看病医和初学医三类, 分别划分她们的职责和俸禄.

TT4(과파고 번역): 将医女分为内衣(醫)、看护(看病醫)、草医学(初醫)三个等级, 分别区分了工作和俸禄.

TT5(바이두 번역): 将医女分为内医、看护医、初学医三个等级, 分别区分了所做的工作和俸禄.

<예5>에서 인간 번역, ChatGPT 3.5 번역과 ChatGPT 4.0 번역은 등가 번역을 실현하여 정확하게 “内医、看病医、初学医”로 번역하였다. 그러나 ChatGPT 4.0 번역에서는 “의녀를”에서 ‘를’을 把자문으로 번역하여 주어와 생략된 비문이 되었다. 과파고 번역은 “内医、看病医、初学医”가 조선 시대에 사용된 관직임을 인식하지 못해, 문맥에 맞지 않는 동음이의어인 “内衣、看护、草医学(속옷, 간병, 초의학)”로 번역해 가독성이 떨어졌다. 또한 번역문에서 “医女”가 서술어로 사용되어 주어와 누락되었으며, 피동문의 오류가 보였다. 바이두 번역은 “간병의”를 “看护医(간호 의사)”로 번역하는 문화소 오역이 발생했지만, 과파고 번역보다는 정확도가 높았다. 그러나 동일한 문법적 오역으로 인해 문장이 자연스럽게 읽혔다.

<예6>

ST: ST: “우선 식어도 별 문제 없는 어선과 오이선부터 준비할 것이니, 그동안 너는 생선을 씻어. 포는 내가 뜯 거야.”

TT1(인간 번역): “首先准备凉了也能吃的鱼膳和黄瓜膳, 你洗鱼, 鱼脯我来切。”

TT2(ChatGPT3.5): “先准备那些即使凉了也没关系的鱼膳和黄瓜膳, 你先去洗鱼, 鱼片我来切。”

TT3(ChatGPT4.0): “我会先准备那些即使凉了也无碍的鱼膳和黄瓜膳, 你这期间先把鱼洗干净, 鱼片我来切。”

TT4(과과고 번역): “首先会准备凉了也没问题的渔船和黄瓜船, 这段时间你去洗鱼吧。我要升炮了。”

TT5(바이두 번역): 先准备凉了也没什么问题的渔船和黄瓜船, 这段时间你洗鱼吧, 我来开炮。

한국 요리 선(膳)은 채소, 두부, 쇠고기 따위를 잘게 썰거나 다져서 만든 음식을 통틀어 이르는 말이다. 인간 번역, ChatGPT 3.5 번역, ChatGPT 4.0 번역은 “어선과 오이선”을 적절하게 “鱼膳和黄瓜膳”으로 번역했다. 그러나 인간 번역에서는 “포는 내가 뜯 거야”에서 “포”를 ‘어포’로 착각하여 “鱼脯(어포, 쥐포)”로 번역, 일반 단어 번역 오류가 발생했다. 과과고 번역과 바이두 번역은 “어선과 오이선”이 음식을 가리킨다는 사실을 인식하지 못하고 각각 “渔船和黄瓜船(어선과 오이배)”으로 직역하여 의미 전달에 실패했다. 또한 “포는 내가 뜯 거야”를 각각 “我要升炮了(나는 포를 쏘아 올리겠다)”와 “我来开炮(나는 포를 썰게)”로 잘못 번역하여 번역문의 일관성과 가독성이 크게 떨어졌다.

<예7>

ST: 용종, 즉 종기(腫氣) 제주도 사람들에게 기생충 질환과 함께 가장 흔한 병 중의 하나였다. 기생충 질환이 많은 것은 따뜻하고 습한 기후 때문이었다.

TT1(인간 번역): 臃肿, 即脓肿, 对于济州岛上的人来说, 这就跟寄生虫疾病一样, 都是最常见的疾病, 寄生虫疾病起因于天气的温暖和潮湿。

TT2(ChatGPT3.5): 痈肿 (即肿气) 是济州岛居民常见的疾病之一, 与寄生虫病同样普遍。寄生虫病多发的原因是温暖潮湿的气候。

TT3(ChatGPT4.0): 痈肿, 即疮疖, 是济州岛居民中最常见的疾病之一, 与寄

生虫病一起流行。寄生虫病多发的原因是温暖潮湿的气候。

TT4(과과고 번역) : 对于济州岛人来说,雍钟,即肿块(腫)与寄生虫疾病一起是最常见的疾病之一。寄生虫病之所以多,是因为温暖潮湿的气候。

TT5(바이두 번역) : 肿块, 即肿块, 对济州岛人来说, 与寄生虫疾病一起是最常见的疾病之一。寄生虫疾病多是因为温暖潮湿的气候。

<예7>에서 나오는 “옹종”은 작은 종기이다. 중국에서 이대 대응한 병명은 “痈肿”이다.(百度百科) ChatGPT3.5 번역과 ChatGPT4.0 번역은 “옹종”을 “痈肿”으로 적절하게 번역했지만 ChatGPT3.5 번역은 “종기(腫氣)”를 “肿气”로 오역해 일반 단어 번역 오류가 보인다. ChatGPT4.0 번역은 “기생충 질환과 함께”를 “与寄生虫病一起”로, 즉, 옹종과 기생충 질환 모두가 동시 유행하는 것으로 오역했다. 인간번역의 경우, “옹종”을 “臃肿”으로 완전히 잘못 번역했다. “臃肿”은 ‘붓는다’는 뜻이기에 “옹종”과는 완전히 다른 의미를 지닌다. 과과고 번역은 “옹종”을 중국어에는 없는 표현인 “雍钟”으로 음역하여 의미 전달이 제대로 되지 못했다. 바이두 번역은 “옹종, 즉 종기(腫氣)”를 “肿块, 即肿块(종양, 즉 종양)”으로 오역하여 번역문의 가독성과 정확성이 떨어졌다.

<예8>

ST: “제음단은 음한 기운을 보하는 환약으로, 유산한 여인에게 좋습니다.”

TT1(인간 번역) : “济饮丹能补气血阴虚, 对流产后的女人很有好处。”

TT2(ChatGPT3.5) : “济阴丹是一种补充阴寒之气的药丸, 适合小产的女性服用。”

TT3(ChatGPT4.0) : “济阴丹是一种补充阴寒气息的丸药, 特别适合流产后的女性。”

TT4(과과고 번역) : “我的音坛是滋补阴寒气息的药丸, 对流产的女人很好。”

TT5(바이두 번역) : “我的阴丹妙药是补阴的丸药, 对流产的女人有好处。”

<예8>에 나오는 “제음단”은 중국에도 존재하는 알약이다. 이에 대응하는 번역은 “济阴丹”이다.(百度百科) 따라서 이국화 전략을 활용해 ChatGPT 3.5 번역과 ChatGPT 4.0 번역처럼 “济阴丹”으로 번역하는 것이 가장 정확하다. 인간번역은 “济饮丹”으로 음역하여 정확하지 않

았다. 또한 “丹”은 알약을 의미하므로 “환약”과 중복되어 생략해도 무방하다. 따라서 인간번역, ChatGPT 3.5 번역, ChatGPT 4.0 번역을 종합해 “济阴丹能补气血阴虚”가 더 적절한 번역이다. 과과고 번역과 바이두 번역은 문맥을 이해하지 못하고 각각 “我的音坛(중국어에 없는 표현)”과 “我的阴丹妙药(나의 음단 묘약)”로 직역하여 원문과 전혀 관련 없는 표현으로 오류를 범했다.

### 3.2 관용어

관용어에는 관용구, 사자성어, 은어, 신조어 등이 포함된다.(이근희, 2003: 6) 본 논문에서는 『대장금』 전체에 등장하는 고사성어 57개, 속담 및 관용구 27개가 포함된 84개의 문장을 대상으로 연구를 진행하였다. 이근희(2015)에 따르면, 관용어 번역 전략에는 “1) 직역, 2) 도착 언어권의 속담으로 번역, 3) 도착 언어권의 유행어로 번역, 4) 도착 언어권의 관용표현으로 번역, 5) 의역(의미를 풀이하여 번역)”이라는 다섯 가지 방법이 존재한다. 본고에서는 각 기계 번역이 기본적으로 어떠한 전략을 활용하여 번역을 수행하는지도 함께 살펴보았다.

<예9>

ST: “언감생심 꺼내달라는 청을 드리겠습니까? 그저 한 번만 얼굴이라도 볼 수 있게 해주십시오.”

TT1(인간 번역): “奴婢哪敢求您放人? 只想请您让我们见上一面。”

TT2(ChatGPT3.5): “我怎么敢请求您把他释放出来呢? 只求能让我见他一面就足够了。”

TT3(ChatGPT4.0): “岂敢妄想提出这样的请求? 只求您让我见他一面。”

TT4(과과고 번역): “我请求你拿出你的想法吗? 就让我见一面吧。”

TT5(바이두 번역): “能拜托您拿出什么感想吗? 就让我见一面吧。”

<예9>에서, 인간 번역, ChatGPT 3.5 번역, 그리고 ChatGPT 4.0 번역은 등가 번역과 의역을 적절히 사용하여 원문의 의미를 정확하게 전달했다. 특히 “언감생심”이라는 표현을 각각 “哪敢”, “怎么敢”, “岂敢

妄想”으로 번역해, 원문의 뉘앙스를 살리면서도 자연스러운 문맥을 유지했다. 이때 ChatGPT 3.5 번역은 외연화 전략을 사용해 원문에 없던 ‘그를 석방하다’라는 내용을 추가하여 번역문의 논리적 흐름과 가독성을 높였다. 반면, 파파고와 바이두 번역은 “언감생심”을 각각 “拿出你的想法吗(네 생각을 내볼래)?”와 “拿出什么感想吗(어떤 감상을 내볼래)?”로 잘못 해석하여 직역하였고, 이로 인해 원문의 의미와 맞지 않는 번역이 되었다. 이러한 오역은 등가 번역이 실패한 사례로, 문맥적으로 부자연스럽고 원문의 의도를 제대로 전달하지 못했다.

<예10>

ST: 누군가 장금을 가둔 것이라면, 모래밭에서 바늘 찾거나 한 가지였다.

TT1(인간 번역): 如果有人把长今藏起来了, 那么这样找下去无异于海底捞针。

TT2(ChatGPT3.5): 如果有人把长今关起来, 那就像大海捞针一样难找。

TT3(ChatGPT4.0): 如果有人把长今关起来了, 那就像是海底捞针一样。

TT4(파파고 번역): 要是有人关了长今, 就等于沙地里捞针。

TT5(바이두 번역): 如果说是有人关了长今, 那就是在沙地里找针。

<예10>에서 인간 번역, ChatGPT 3.5 번역, 그리고 ChatGPT 4.0 번역은 “모래밭에서 바늘 찾기”를 각각 “海底捞针”과 “大海捞针”으로 번역, 비유적으로 잘 표현하여 가독성을 높이고 원문의 의미를 충실히 전달했다. 반면, 파파고 번역과 바이두 번역은 각각 “沙地里捞针(모래밭에서 바늘 건지기)”과 “在沙地里找针(모래밭에서 바늘 찾기)”으로 잘못된 축자역을 도출하여 원문의 비유적 의미를 제대로 전달하지 못했다.

<예11>

ST: “춤은 내가 출 터이니, 너는 굶이나 보고 떡이나 먹거라.”

TT1(인간 번역): “舞还是由我来跳, 你就只管看热闹、吃点心就行了。”

TT2(ChatGPT3.5): “跳舞的事我来, 你就少管闲事, 坐着吃糕吧。”

TT3(ChatGPT4.0): “舞我来跳, 你就旁观吃饼吧。”

TT4(파파고 번역): “跳舞由我来跳, 你就看跳神吃打糕吧。”

TT5(바이두 번역): “跳舞我来跳, 你去跳大神吃年糕吧。”

<예11>에서, 인간 번역과 ChatGPT 3.5 번역은 원문의 의미를 잘 살리면서 “긋이나 보고 떡이나 먹거라”를 각각 “你就只管看热闹、吃点心就行了”와 “少管闲事”라는 표현으로 의역하는 등가 번역을 구현해 원문의 뉘앙스를 잘 유지했다. 반면, ChatGPT 4.0은 “你就旁观吃饼吧”로 의역해 관용어 번역의 등가성은 떨어졌지만, 가독성은 높았다. 한편, 파파고 번역과 바이두 번역은 각각 “看跳神吃打糕”와 “跳大神吃年糕”라는 직역을 도출해 원문의 의미를 왜곡하고 가독성을 저하시켰다.

관용어에 대한 번역 결과를 분석해 보면 인간 번역과 생성형 기계 번역 모두 중국어에 대응하는 관용어를 사용하거나 의역하는 경향이 보인다. 이로써 문장의 수사적 표현을 살리면서 중국 독자에게 더 자연스럽게 가독성이 높은 번역을 구현할 수 있었다. 고사성어의 경우 신경망 기계 번역의 결과도 우수한 편이었는데, 속담 및 관용구 번역 면에서는 생성형 기계 번역의 수준이 이보다 훨씬 앞섰다.

『대장금』의 다른 예시를 좀 더 들여보자면, 인간 번역은 “백지장도 맞들면 낫다”를 중국어의 완전 대응 표현인 “三个臭皮匠，顶个诸葛亮”으로 번역했으며, ChatGPT 3.5는 이를 “连一张薄纸两个人抬都更轻松”으로 의역했다. ChatGPT 4.0 역시 이를 중국의 관용구인 “众人拾柴火焰高”로 등가 번역했다. 또한 인간 번역은 “될성부른 나무는 떡잎부터 알아본다”를 “人看从小，马看蹄爪”라는 중국 관용구로 등가 번역했으며, ChatGPT 3.5는 “成大树的苗子，从嫩芽就能看出来”로 의역했다. ChatGPT 4.0도 이를 “苗不秀，不成竹，有潜质的东西从一开始就能看出来”로 등가 번역과 의역을 모두 구현하여 매끄럽고 가독성이 좋은 번역을 제시했다.

반면에 신경망 기계번역은 직역하는 경향이 강해 이로 인한 오역이 자주 발생하는 것을 볼 수 있었다. 예를 들어 파파고 번역은 “참 귓구멍에 말뚝을 박다”를 “把桩子扎在耳朵眼儿上”으로 직역했고, 바이두 번역 역시 “在耳孔钉了木桩”으로 축자 번역하여 가독성이 떨어졌다. “십년 체증이 다 내려간 기분이다”의 경우 파파고 번역은 “感觉10年的积食都没了”로, 바이두 번역은 “感觉十年的积食都过去了”로 직역하여 원

문의 의미 전달이 제대로 이루어지지 않았다.

#### 4. 번역 품질 평가 결과

본 절에서는 위의 평가 기준을 사용하여 5가지 번역 모델을 대상으로, 『대장금』에서 등장하는 인명 및 호칭어, 지명 및 장소명, 관직명, 음식명, 의학 용어, 고사성어 등 총 257개의 문화소를 포함한 184개의 문장에 대한 번역 결과물을 평가하고 비교 분석하였다. 해당 품질 평가의 결과는 다음과 같다.

<표 2> 문화소 세부 항목별 번역 평가 결과

	인간 번역	ChatGPT 3.5 번역	ChatGPT 4.0 번역	파파고 번역	바이두 번역
인명 및 호칭어	9.7	7.5	7.85	5.7	6.45
지명 및 장소명	9.75	9.1	9.55	7.65	7.3
관직명	9.85	9.6	9.2	6.2	7.55
음식명	9.65	7.05	7.65	6.1	5.55
의학용어	9.2	9.35	9.6	7	7.15
고사성어	10	9.3	9.35	8.25	8.15
속담 및 관용구	9.6	9.4	9.45	7.7	6.95

<표 2>에 따르면, 특히 의학 용어 번역에서 ChatGPT 4.0(9.6)과 ChatGPT 3.5(9.35)가 인간 번역(9.2)보다 우수한 성능을 보였다. 조선시대 배경의 의학 용어를 번역할 때는 사전 지식이 필수적이지만, 번역가가 해당 분야에 익숙하지 않거나 지식이 부족할 경우 번역 품질에 편차가 발생할 수 있다. 특히 “웅종”, “신부전증”, “허증”, “오령산”, “제음단” 등의 번역에서 인간 번역은 번역가의 주관적인 판단으로 인해 오역이 발생하는 한계를 보였다. 또한 신경망 기계번역은 코퍼스에 조선시대 배경의 인명, 호칭어, 지명, 관직명, 음식명, 의학 용어 등에

대한 자료가 부족한 까닭에 오역이 가장 빈번하게 나타났다. 반면, 생성형 인공지능은 방대한 학습량을 기반으로 하여 조선시대 배경의 의학 용어 번역에서 더 우수한 결과를 나타냈다.

아래는 문화소 유형별 점수 결과이다.

<표 3> 문화 관련어 및 관용어의 번역 결과

	인간 번역	ChatGPT 3.5 번역	ChatGPT 4.0 번역	파파고 번역	바이두 번역
문화 관련어	9.65	8.49	8.76	6.56	6.82
관용어	9.8	9.35	9.4	7.98	7.55

<표 3>에 따르면, 특정 문화 관련 어휘, 관용어 번역에서 다섯 가지 번역 시스템 모두 문화 관련어보다 관용어 번역에서 더 높은 점수를 기록했다. 이는 조선시대 배경의 문화소보다 현대에서 더 빈번하게 사용되는 관용어가 기계 번역 코퍼스에 포함될 가능성이 높고, 번역가에게도 관용어 번역이 상대적으로 더 용이하기 때문일 것이다. 또한 관용어 번역에서 생성형 인공지능인 ChatGPT 4.0(9.4)과 ChatGPT 3.5(9.35)는 신경망 기반 기계 번역인 파파고(7.98)와 바이두 번역(7.55)보다 훨씬 뛰어난 성능을 보였으며, 인간 번역에 가까운 번역 품질을 제공했다.

<표 4> 번역 품질 평가 결과

	인간 번역	ChatGPT 3.5 번역	ChatGPT 4.0 번역	파파고 번역	바이두 번역
정확성(5)	4.71	3.9	4.08	2.84	2.76
완결성(2)	1.97	1.96	1.99	1.96	1.94
가독성(1)	0.99	0.94	0.91	0.51	0.56
일관성(1)	1	0.99	0.98	0.69	0.76
형식성(1)	1	1	1	0.95	0.99
합계	9.73	8.92	9.08	7.27	7.19

인간 번역은 정확성(4.71), 가독성(0.99)에서 가장 높은 점수를 얻

어, 모든 번역 시스템 중에서 최고로 완전한 결과를 보였다. 그다음으로 우수한 결과를 기록한 것은 ChatGPT 4.0 번역으로, 정확성(4.08), 완결성(1.99), 가독성(0.91)에서 고르게 높은 점수를 받아 생성형 인공지능의 발전 양상 및 가능성을 보여주었다. ChatGPT 4.0 번역은 정확성에서 인간 번역보다 약간 떨어졌으나, 완결성(1.99)에서는 인간 번역(1.97)을 거의 능가한 성과를 보였다. 종합적으로 ChatGPT 4.0은 9.08점으로 다른 기계 번역보다 월등히 높은 점수를 기록하며 만족스러운 결과를 나타냈다.

ChatGPT 3.5 번역은 정확성(3.9)에서 3위를 차지했지만, 완결성(1.96)과 가독성(0.94)에서는 인간 번역과 ChatGPT 4.0에 근접한 성과를 기록하며 총점 8.92점을 받았다. 반면, 신경망 기반 기계 번역인 파파고와 바이두 번역은 정확성과 가독성에서 인간 번역, ChatGPT 3.5, ChatGPT 4.0에 비해 낮은 점수를 기록했다. 파파고는 정확성 2.84점, 가독성 0.51점을, 바이두는 정확성 2.76점, 가독성 0.56점을 받았으며, 일관성 점수는 각각 0.69점과 0.76점으로 신경망 기계 번역의 한계를 드러냈다.

종합적으로 다섯 가지 번역 방법 중 인간 번역이 여전히 가장 높은 점수를 기록했으며, 그 뒤를 이어 유료 버전인 ChatGPT 4.0과 무료 버전인 ChatGPT 3.5가 차례로 우수한 성과를 보임을 확인할 수 있었다. 신경망 기계 번역 시스템인 파파고 번역과 바이두 번역은 직역을 하는 경향이 강해, 문장 맥락에 맞지 않는 오역이 많아 상대적으로 낮은 성능을 보였다. 이때, 의료 용어와 같은 전문 분야 번역에 대해서는 생성형 기계 번역이 인간 번역보다 더 높은 정확도를 보이는바, 이는 생성형 기계 번역이 인간 번역을 보완하여 더욱 정확한 번역을 할 수 있는 길잡이가 될 수 있는 가능성을 시사한다.

신경망 기계 번역과 생성형 기계 번역은 모두 코퍼스를 지속적으로 업데이트하고 학습하고 있어, 코퍼스의 크기가 두 시스템 간의 차이를 결정짓는 이유는 아니다. 원문 문맥에 대한 이해도가 두 시스템 간의 주요 차이를 만드는 것이다. 신경망 기계 번역의 원문 파악 문제는 그 작업 원리에서 기인한다. 신경망 기계 번역은 원문을 최적 단위인 형태소로 분할한 후, 그에 맞는 번역어를 찾아 문장을 재구성하여 번역문을

생성하는 방식으로 작동하는데, 이 과정에서 원문의 문맥을 충분히 이해하지 못하고 직역을 하게 되는 경우가 많다. 따라서 신경망 기계 번역에서는 원문 전체의 문맥을 더욱 깊이 있게 이해할 수 있도록 개선이 필요하다.

본 연구에 따르면, 인간 번역은 의학 용어와 같은 기술 번역에서 가장 취약하다. 기계 번역이 빠르게 발전하는 현재 상황에서 번역가는 전문 용어 번역에 더욱 집중해야 하며, 지속적인 연구와 학습을 통해 기술 용어 번역의 부족함을 보완해야 한다. 현재, 신경망 기계 번역과 생성형 기계 번역은 다의어 번역에서 자주 오류가 발생한다는 점이 지적된다. 한국어는 다의어가 매우 많은 언어이기 때문에, 한중 번역에서는 다의어 번역에 특히 주의를 기울여야 한다. 본 연구에서도 이러한 오류가 확인되는데, 문학번역에서 직역을 최소화할 수 있는 기술이 추가된다면, 다의어 번역의 오류 또한 감소 할 것으로 판단된다.

## 5. 나가는 말

본 연구에서는 ChatGPT 유, 무료 버전을 중심으로 조선시대 배경의 문화소 번역 성능을 분석하고, 다양한 번역 시스템 간의 비교를 통해 다음과 같은 결론을 도출하였다.

우선 ChatGPT 4.0은 ChatGPT 3.5에 비해 가독성과 일관성 측면에서는 다소 부족한 모습을 보이나, 정확성과 완결성 측면에서는 더욱 우수한 성과를 나타냈다. 종합적으로 평가하였을 때, ChatGPT 4.0이 ChatGPT 3.5보다 기술적으로 더 발전된 언어 모델을 기반으로 하여 번역의 정교함과 정확성을 높였으며, 더 자연스러운 번역 품질을 제공하는 것으로 판단된다.

ChatGPT 3.5와 4.0은 신경망 기반 기계 번역 시스템인 과파고와 바이두보다 훨씬 더 정확하고 일관된 번역을 제공했다. 문화 관련 어휘의 번역에서 생성형 기계 번역은 신경망 기계 번역보다 훨씬 우수한 결과를 냈으며, 인간 번역에 근접하거나, 이를 뛰어넘는 결과를 가져오기도 했다. 즉, 생성형 기계 번역이 기존의 신경망 기계번역과 인간 번역의

한계를 어느 정도 초월하거나 극복하고 있는 것이다.

특히 의학 용어와 같은 전문적인 문화소 번역에서 ChatGPT 3.5와 4.0은 방대한 학습 데이터와 사전 지식을 기반으로 인간 번역보다 더 정확한 번역 결과를 제공했다. 이는 생성형 인공지능이 인간 번역가보다 넓고 깊은 지식 기반을 가지고 있음에 기인할 것이다.

본 논문에서는 ChatGPT 3.5와 ChatGPT 4.0에 동일하게 간단한 프롬프트를 사용하였다. 프롬프트 설정을 더욱 구체화하면 번역의 정확성이 크게 향상될 수 있을 것이다. 생성형 기계 번역에서는 프롬프트를 통해 번역의 방향과 맥락을 명확하게 학습시킴으로써 다양한 창의적인 번역 결과를 생성할 수 있기 때문이다. 비록 생성형 기계 번역의 알고리즘 특성상 동일한 프롬프트를 다른 시점에 입력할 경우 번역의 결과에 차이가 발생할 수 있지만, 이는 생성형 기계번역의 한계이자 성과라 할 수 있을 것이다. 다면적인 프롬프트 설정과 생성형 기계 번역의 품질에 대한 연구는 차후 과제로 남긴다.

[Abstract]

## A Study on the Translation of Cultural Elements Using ChatGPT: Focusing on the Korean-Chinese Translation of the Novel 『Dae Jang Geum』

Pang, YONGFENG

Seo, Yeonju(Chungnam National University)

This study aims to analyze the translation performance of ChatGPT by comparing its results with human translations and neural machine translations (Papago and Baidu) for cultural elements found in the historical novel Dae Jang Geum, set in the Joseon Dynasty. The research includes a total of 184 sentences encompassing various cultural terms such as names and titles, place names, official titles, food, and medical terminology, as well as idiomatic expressions. ChatGPT 4.0 and 3.5 showed superior performance, especially in translating culturally specific terms like medical terminology, often surpassing human translations, and yielded results close to human quality in other areas. Notably, ChatGPT 4.0 demonstrated higher accuracy and completeness than 3.5. In contrast, Papago and Baidu displayed a tendency toward literal translation, resulting in lower accuracy and readability in some translations. These results suggest that ChatGPT's translation performance is enhanced by its extensive training data and knowledge. This study indicates that ChatGPT has the potential to overcome some limitations of neural machine translation for cultural terms and suggests directions for

improving translation quality in the future.

**Key words** : #NeuralMachineTranslation #Generative AI #KoreanChinese Translation #LiteraryTranslation #Cultural wordsTranslation #Cultural words #ChatGPT

## [참고문헌]

### □ 단행본

- 김효중(2004), 「새로운 번역을 위한 패러다임」, 푸른사상사, 37-79면.  
이근희, 『번역의 이론과 실제』, 한국문화사.  
유민주, 『대장금1』, 『대장금2』, 『대장금3』, 은행나무, 2003.  
薛舟, 徐丽红, 『大长今』, 人民文学出版社, 2005.  
张南峰(2004), 「艾克西拉的文化专有项翻译策略评介」, 『中国翻译』第25卷 第1期, 20-25면.  
Newmark, Peter (1998), 『A textbook of translation, Great Britain: Redwood Books』  
Stolze (1992), 『Übersetzungstheorien: Eine Einführung』  
Vermeer, H. (1990), 『Skopos und Translationsauftra - Aufsätze. Frankfurt: IKO』  
『국립국어원 표준국어대사전』  
『고려대 한국어사전』, 고려대학교 민족문화연구원.  
『단국대 한국한자어사전』  
『新华字典』(最新版12版), 新华辞书社

### □ 논문 및 기타 자료

- 감동미, 「인공지능과 인간의 문학(詩)번역 작품 비교 연구: 한용운의 『님의 침묵』을 중심으로」, 『인문사회 21』 제9권, 2018, 67-82면.  
강병규, 이지은, 「신경망 기계번역의 작동 원리와 번역의 정확률- 중한 번역을 실례로」, 『중어중문학』 한국중어중문학회, 2018 253-295면.  
김지은, 「희곡의 기계번역에 대한 고찰-희곡 『버스 정류장』의 등장인물을 중심으로」, 『한중언어문화연구』 제68권, 한국중국어언어문화연구회, 2023, 3-32면.  
김혜림, 「중한 기계번역 포스트에디팅 가이드라인에 관한 예비연구 -포스트에디팅 교육을 위한 가이드라인」, 『중국어언어연구』 제99권, 한국중국어언어학회, 2022, 277-311면.  
노금송, 왕원, 「한국어 문화소에 관한 기계번역과 인간번역의 비교 분석 연구」, 『한국학연구』 제73권, 인하대학교 한국학연구소, 2024, 299-318면.  
남철진, 「中韓(韓中) 번역을 통해 본 구글번역(GNMT)의 문제」, 『중국어 중국학』 영남대학교 중국연구센터, 2018, 1-27면.  
마승혜, 「AI 문학번역, 어디까지 가능한가—챗GPT가 번역한 한국 문학작품 제목의 기능 분석을 중심으로」, 『번역학연구』 제25권, 한국번역학회, 2024, 57-85면.  
문려화, 「기계번역을 통한 한중 소설 번역의 양상과 전망 - 윤대녕의 소설 ‘누가 고양이

- 이를 죽였나'를 중심으로, 『이화어문논집』 제60집, 이화어문학회, 2023, 5-32면.
- 박금은, 「일한 자막번역의 문화소 번역전략 연구 - 드라마 『심야식당』을 중심으로, 『문화와융합』 제46권, 한국문화융합학회, 2024, 55-68면.
- 박정희, 강병규, 「챗GPT에 기반한 중국문학작품 번역 방법론 탐구, 『중국문학』 제119권, 한국중국어문학회, 2024, 169-200면.
- 서재형, 박찬준, 어수경, 문현석, 임희석, 「인공신경망 기계번역에서 디코딩 전략에 대한 연구, 『한국융합학회논문지』 제12권, 한국융합학회, 2021, 69-80면.
- 신세연, 오용토야, 투굴두르, 송보나, 「생성형 인공지능 번역기와 번역가 번역을 이용한 다국어 번역 비교 연구 : 엘런 튜링의 '계산 기계와 지능'을 중심으로, 『에듀테크 인턴트연구』, 제5권, 한국에듀테크인턴트학회, 2023, 155-170면.
- 신천, 조혜진, 「위화 소설 『허삼관 매월기』 중한번역을 통해 본 문학번역에서의 챗GPT의 활용가능성, 『언어학연구』 제29권, 한국언어연구학회, 2024, 51-65면.
- 이근희, 「문화와 밀접한 상관어(相關語)의 번역 전략, 『번역학연구』 제4권, 한국번역학회, 2003, 5-27면.
- 이창수, 「챗GPT 출현 이후 기계 번역과 인간 번역 간의 번역 문체 차이 변화 연구, 『번역학연구』 제24권, 한국번역학회, 2023, 539-561면.
- 이창수, 「챗GPT, 파과고, 인간 번역가 간의 한영 문학번역 차이점 연구, 『번역학연구』 제25권, 한국번역학회, 2024, 11-37면.
- 임순정, 「구글 번역기를 통해서 본 통계기반방식 번역기의 특징, 『통번역교육연구』 제10권, 한국통번역교육학회, 2012, 183-210면.
- 이유정, 「현대시 인공지능(AI) 번역의 오류 양상 연구 - ChatGPT- 3.5를 활용한 김소월 시 번역결과물을 중심으로-, 『문화와융합』 제45권, 2023, 97-110면.
- 윤여법, 「ChatGPT의 한영 번역 도구로서의 정확성 분석, 『한국초등교육』 제34권 4호, 서울교육대학교 초등교육연구원, 2023, 215-231면.
- 지인영, 김희동, 「신경망기계번역 기술 진화와 번역품질 분석, 『통번역학연구』 제24권, 한국외국어대학교 통번역연구소, 2020, 199-223면.
- 최미란, 「馮驩才의 장편소설 『三寸金蓮』에 나타난 문화소 번역 전략 연구, 『한중인문학연구』, 통권제73집, 한중인문학회, 2021, 127-152면.
- 최연연, 「중국 드라마 「보보경심(步步惊心)」 한국어 자막에 반영된 문화소 번역의 실제, 『문화와융합』 제 45권, 한국문화융합학회, 2023, 261-274면.
- 홍은영, 탁진영, 「인공지능시대의 번역: 스티브 잡스의 연설문을 중심으로, 『인문사회 21』 제11권, 2020, 1531-1544면.
- 황영숙, 「병렬말뭉치의 구축과 활용 ; 통계기반 기계번역 기술의 소개 및 최신 연구동향 분석, 『언어사실과 관점』 제25권, 연세대학교 언어정보연구원, 2010,

89-114면.

황지연, 이미령, 원다인, 「표현적 텍스트의 기계 번역 활용 가능성 고찰 - K-pop그룹 뉴진스 노래 가사 번역을 중심으로」, 『통번역학연구』 제28권, 한국외국어대학교 통번역연구소, 2024, 177-207면.