

T&I REVIEW

December 2023. Vol. 13. No. 2.

다국어 『조선왕조실록』번역 프로젝트 수행 가능성 모색*

- 인간과 기계번역의 협업 패러다임을 기반으로 -

전현주**

Chun Hyunju (2023). Exploring the possibility of a multilingual translation project for “Annals of the Joseon Dynasty” : Based on the paradigm of collaboration between human and machine translation. *Driven by the possibility of using artificial intelligence (AI) translation technology for Korean classics and the need for further research on the subject, this paper aims to specify and discuss the possibility of carrying out a multilingual translation project for Annals of the Joseon Dynasty (hereinafter referred to as ‘Annals’). This study includes prior considerations that take into account the unique characteristics of translation for Korean classics, construction of a customized AI translation corpus and categories of related metatexts, cases of AI translation integration, and methods of collaboration between human and machine translation. In particular, there is a need for further discussion on the construction of a customized AI learning corpus for Annals and the selection and categorization of metatexts to be applied, given the recognition of the construction of a high-quality learning corpus as the most important step in measuring the success or failure of project implementation. (Shinhan University, Korea)*

Keywords: Annals of Joseon Dynasty (Joseon Wangjo Sillok), meta-text, translation corpus for AI learning, *The Annals of King T’aejo, The Veritable Records of King Sejong*

주제어: 조선왕조실록, 메타텍스트, AI 학습용 번역 코퍼스, 태조실록, 세종실록

* 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019S1A5 A2A01048542)

** 신한대학교, 부교수

1. 서론

1.1. 연구의 배경

최근 인공지능(AI) 기술, 특히 기계번역의 발전은 번역 작업의 패러다임을 획기적으로 전환시키고 있다. 파파고, 구글 번역, Bing 번역, 딥엘 등의 서비스는 이제 일상 대화의 수준을 넘어서 전문적인 문서까지도 상당한 수준으로 번역할 능력을 갖추게 되었다. 특히, 우리는 2022년 11월에 공개된 ChatGPT¹⁾를 포함한 생성형 AI가 번역 서비스에 특화된 플랫폼이 아님에도 불구하고 고급 문맥 인식과 자연스러운 번역 결과를 생성하는 GPT(Generative Pretrained Transformer) 세상을 경험하고 있다. 그리고 이보다 앞선 2016년에 공개된 신경망 기계번역(Neural Machine Translation, NMT)(Wu, Y., et. al. 2016; 전현주 2020: 62; 2022: 172)은 딥러닝 기반 모델을 통하여 문맥을 보다 잘 이해하고 통합된 방식으로 전체 문장을 번역할 뿐 아니라 이전의 통계적 기계번역(Statistical Machine Translation, SMT) 방식보다 급격히 향상된 번역품질 서비스를 직접적으로 경험하고 있다.

AI는 대량의 언어 데이터를 학습하고, 복잡한 언어 패턴을 인식하며, 문맥을 파악하는 능력을 발전시키면서, 번역가들이 고전 작품의 번역에 있어서 오랫동안 직면해온 시대적, 언어적, 문화적 장벽을 뛰어넘는 데 큰 도움을 주고 있다. 가령, “ETRI, 인공지능으로 한문고서 번역해 문화유산 접근성 높인다²⁾”, “승정원일기 등 한문고전 AI가 번역한다³⁾”, 그리고 “읽기 어려운 한문고전, AI가 번역한다⁴⁾” 등의 사례에서 보다시피 AI 기술의 눈부신 발전은 고전 작품의 번역에 있어서도 새로운 지평을 열고 있음을 알 수 있다. 이는 고전 작품들이 더욱 폭넓은 독자층에게 소개되는 기회를 마련하고, 작품의 본질과 다층적인 의미를 보다 풍부하게 전달할 수 있는 가능성을 확장시킬 수 있음을 의미한다. 본 연구는 이러한 가능성에 주목하며 인간과 AI 기계번역의 협업을 기반으로 『조선왕조실록』의 영역(英譯) 및 다국어 번역 프로젝트 수행의 당위성을 논의하고 이를 수행하기 위한 실질

1) 3.5 version 기준(2023년 3월 14일 3.5버전에 비하여 신뢰성, 독창성, 그리고 문제해결능력을 향상시킨 4.0 version을 공개하였다.) <https://openai.com/blog/chatgpt>

2) <https://www.aitimes.kr/news/articleView.html?idxno=28968>

3) <https://www.yna.co.kr/view/AKR20210114072200005>

4) <https://www.helloddd.com/news/articleView.html?idxno=91377>

적인 고려사항 및 실행 가능성을 모색한다. 이와 더불어 AI 번역 기술을 고전 번역에서 구체적으로 적용하기 위하여 『조선왕조실록』에 특화된 맞춤형 AI 학습용 코퍼스 구축의 필요성을 역설하고자 한다.

1.2. 연구의 필요성 및 목적

우리 민족 기록문화의 대표적인 상징이자, 국보 제 151호 및 유네스코 세계기록문화 유산으로도 등재된 『조선왕조실록』은 원문이 한문으로 기술되어 있으므로 한글을 사용하는 로컬 독자의 가독성을 확보하기 위하여 번역은 필수적으로 수반되어야 할 중요한 과정이다. 번역은 비단 『조선왕조실록』에만 국한되지 않고 20세기 이전에 저술된 대부분의 고전 작품 및 기록물에도 모두 적용되어야 하는 필연성을 지니고 있다. 『조선왕조실록』의 번역은 1968년부터 세종대왕기념사업회가 주관하다가 1972년부터 민족문화추진회(현 한국고전번역원) 주관으로 28년에 걸친 한역 작업을 거쳐 1993년에 한글본이 413책으로 간행된 바 있다. 오랜 기간 동안 각고의 노력 끝에 완역된 『조선왕조실록』임에도 불구하고 체계적인 사전준비 미흡, 주관 기관 간의 분할 번역작업, 원고 위촉제 번역 등으로 인한 번역수준의 균질화 및 번역 어휘의 통일성 등의 문제(서정문 2011: 요약보고문)가 끊임없이 제기되었다. 이 때문에 다시 번역을 수행하는 신국역 사업이 2012년부터 추진되기 시작하였으며, 현재도 진행 중이며 완역 시기는 아직 미정인 상태이다. 그럼에도 2006년부터 국사편찬위원회⁵⁾ 및 한국고전번역원 홈페이지⁶⁾에서 『조선왕조실록』의 국역본과 현재까지 진행된 신역본을 대상으로 원본과 한글번역본 열람서비스를 제공하기 시작하였다⁷⁾.

한편, 한국어로 완역된 『조선왕조실록』의 편찬에 힘입어 국사편찬위원회는 2012년 1월 초 “조선왕조실록 번역”이라는 제하의 기사⁸⁾에서 400억 원의 예산을 투입하여 2033년 완역을 목표로 영어 번역 프로젝트를 시작한다는 야심찬 기획 소식을 전하였다. 이 대장정의 번역 프로젝트는 경복궁 1차 복원 기간과 맞먹는

5) <http://sillok.history.go.kr>

6) db.itkc.or.kr

7) 『조선왕조실록』의 인터넷 서비스는 세 기관의 협력으로 이뤄지고 있다. 원문은 서울대 규장각이 제공하고, 번역은 한국고전번역원에서 담당하며, 국사편찬위원회는 그 자료들을 모아서 데이터베이스로 구축하여 서비스하고 있다.

8) <https://news.kmib.co.kr/article/view.asp?arcid=0005725734&code=11171211>

장장 20년의 기간이 소요될 것이라 예상했다. 그리고 2012년부터 3년간의 ‘터 닷기’ 기간 동안 용어와 표기의 기준을 마련한 뒤 샘플을 추출하여 시험번역을 거친 후 본격적인 번역 작업은 2014년쯤 시작할 전망이라고 하였다.

영역(英譯) 프로젝트가 시작된 지 10여 년이 지난 현재, 국사편찬위원회에서 주관한 『조선왕조실록』의 영역 현황을 살펴보면, 2017년 3월 『세종실록』의 번역본 *The Veritable Records of King Sejong I*이 발간되었다. 1차 발간된 영역본은 세종의 재위기간 중 즉위년인 1418년 8월부터 세종 17년인 1435년 4월까지의 내용을 담고 있으며, 이후부터 세종 32년(1450년)까지의 나머지 내용은 현재 영역 작업이 진행 중이다. 기 번역된 내용은 국사편찬위원회에서 원문(한자)과 한글 및 영어로 번역문에 대한 데이터 서비스를 제공하고 있다⁹⁾.

이러한 괄목할만한 성과는 그 자체로서 탁월한 인정을 받아야 하지만 한편으로는 완역의 실현가능성에 대한 일말의 불안감 또한 떨쳐 버릴 수 없다. 현재 수준의 번역 투입 예산과 인력, 그리고 번역방식을 적용하면 완역까지는 앞으로도 족히 수 십 년 이상 더 소요될 것이 자명하기 때문이다. 2012년에 시작한 『조선왕조실록』의 한역 현대화 작업 또한 2018년 4월 현재 아직 10%가량밖에 달성하지 못하였다. 총 574권 가운데 6년간 62권을 수행하는데 그쳤다. 이러한 속도라면 100% 완역까지는 40년 이상 걸리는데, 영역 작업은 이 보다 더 많은 시간을 요할 것이다¹⁰⁾.

이러한 상황에서 필자는 번역에 소요되는 시간과 예산, 그리고 번역을 수행하는 주체와 번역 수행방식의 한계점을 극복하고 『조선왕조실록』의 한역과 영역을 넘어 다국어 번역의 실현가능성에 주목하고 이에 논의의 초점을 옮겨가고자 한다. 그리고 딥러닝과 강화학습, 그리고 생성형AI 등 초고도화 된 AI 기술의 접목이 일상화된 이 시대에 한국의 기록 문화유산의 백미인 『조선왕조실록』의 영어/다국어 버전의 번역 프로젝트 수행을 위하여 기계번역의 활용을 적극적으로 제안한다.

이를 위하여 인간과 기계번역의 협업 메커니즘에 기반한 『조선왕조실록』의 번역 프로젝트 수행의 당위성을 제안하고 실행 가능성을 모색하고자 한다. 이러한 논의는 『조선왕조실록』의 역사적인 가치와 한국의 지식 문화유산의 세계적인 확산 및 보급을 위한 기초 인프라 구축의 필요성에 대한 인식 확산에 기여할 것이

9) <http://esillok.history.go.kr/about/translationProjectInfo.do>

10) 승정원일기도 1994년부터 번역을 시작했지만 진척률이 20%에 불과하다. 총 2386권 가운데 해마다 40-50권 가량 번역한다. 완역까지 41년 걸린다. <https://www.etnews.com/20180416000283>

다. 그리고 무엇보다도 『조선왕조실록』의 영역 완간을 넘어 다국어 번역 프로젝트 수행은 한국 고전의 세계사적인 의미 확대는 물론 관련 한국문화 콘텐츠의 글로벌 위상 제고에도 크게 기여할 것이다. 뿐만 아니라 해당 프로젝트 수행 과정에서 양산되는 중간결과물 및 최종결과물의 다각적인 활용방식을 확장하고 관련 분야의 파급효과를 확장하는데 직간접적으로 기여할 것이다.

2. AI 기반의 고전 번역 프로젝트 수행의 타당성

2.1. 선행연구

우선 『조선왕조실록』의 번역학적인 접근과 관련된 연구 결과를 살펴보면, 『조선왕조실록』번역 성과물의 연계성 및 활용 방안 고찰(박현주 2018), 『조선왕조실록』을 통해서 한국에서의 번역 개념의 역사 조망(이영훈 2012), 『조선왕조실록』번역사업과 남북한 체제경쟁(권오현 2021), 『조선왕조실록』의 번역에 나타난 오류(김주원 2006) 등이 있다. 그리고 최근에는 기계번역을 활용한 우리나라 고전 작품의 번역 필요성을 요청하는 연구들이 눈에 띄는데, 가령 고전문언문(한문) 기계번역용 말뭉치 구축 방안 제안(허철 2022), 고전문언문(古典文言文) 기계번역의 현황과 과제(김우정 2021), AI 시대 중국 고전문학 연구의 새로운 방법론 탐색(김준연 2021) 등을 꼽을 수 있다. 이러한 성과를 통해서 고전 기록물의 중요성과 대표성의 측면에서 한국학술지인용색인(KCI)에 등재된 『조선왕조실록』관련 연구실적은 사실 매우 척박한 수준에 머물러 있다고 할 수 있다.

대신 한국고전번역원에서 지원한 2009년부터 2023년 현재까지 수행된 연구프로젝트에 대한 보고서 형태의 다양한 결과물이 있다. 가령, 고전번역사업 추진의 효율화 및 성과활용 극대화 방안 기획연구(2009)와 『국역조선왕조실록』현대화사업 수립계획을 위한 기초조사 연구(2009), 『승정원일기』번역종합계획 수립을 위한 기초조사연구(2010), 『승정원일기』자동번역시스템의 활용방안 연구(2021) 등 기계번역을 적용한 프로젝트 수행의 가능성을 진단하기 위한 연구도 활발하게 진행되고 있다. 이러한 연구경향은 기계번역과 인간 번역의 공존 패러다임을 기반으로 고전 번역 프로젝트의 수행 가능성을 모색하는 본 연구의 논지와 상당부분

부합하는 측면이 있어서 매우 고무적이다. 따라서 이와 같은 기계번역 적용 가능성을 목표로 수행한 연구 결과물을 적극적으로 검토하여 논의의 주제와 접목키로 한다.

2.2. 프로젝트 수행의 당위성 및 타당성

본 연구와 관련하여 『2022 고전번역연감』¹¹⁾에 따르면 최근 고전번역의 연구 추세가 ‘고전번역의 새로운 방법론 모색 및 DB와 기계번역의 가능성 탐색’으로 ‘기계번역’의 특징 및 과제에 대한 연구의 필요성을 제기한 점은 매우 시의적절하다 할 수 있다. 이는 고전번역과 AI 기계번역의 접목 가능성 및 필요성에 대한 사업 주관기관의 수용 전략의 변화를 가능할 수 있는 중요한 측면을 함의하기 때문이다. 이러한 전략 수정의 변화 이면에는 국사편찬위원회가 주관하는 『조선왕조실록』의 영역 프로젝트는 『세종실록』의 본문은 모두 번역하지만 나머지 왕대의 실록은 주요 사실만 선별하여 번역할 계획(박현주 2018: 67-68)이며, 이마저도 사업종료 시점은 미확정이라는 현실적인 한계 등이 주요 원인으로 작용했을 것으로 추측된다. 『조선왕조실록』의 영역 프로젝트를 시작한 지 10년(1단계: 2012년 - 2022년)이 지났으며, 새로이 시작된 2단계(2023년 -)에서 다시 10년의 시간을 투입한다 해도 현재의 방식을 적용한다면 『조선왕조실록』의 영역본 완간은 여전히 요원할 것이기 때문이다.

이러한 상황에서 고전 번역에 박차를 가할 필요성과 그 과정에서 AI 기술의 적용 가능성을 탐구하는 것은 현재와 미래의 학문적, 문화적 가치 보존에 있어 매우 중요한 과제이다. 특히, 『조선왕조실록』과 같은 방대한 역사적 자료의 영역 프로젝트에 AI 번역 기술을 도입하는 것은 시간적, 경제적, 그리고 학문적인 측면에서 충분히 당위성과 타당성을 갖는다.

AI 기반의 고전 번역을 수행해야 할 당위성을 요약하면 다음과 같다: 1) 자원의 효율적 활용: 인간 번역가만을 활용한 기존의 번역 방식은 대규모의 비용과 시간이 투입되어야 하며, 전문 번역가의 수급 또한 매우 한정적이다. AI를 활용하면 이러한 자원의 한계를 극복하고 보다 효율적인 번역 작업을 수행할 수 있다. 2) 문화유산의 보존 및 전파: 『조선왕조실록』은 유네스코 세계기록문화유산으로, 사

11) <https://ebook.itkc.or.kr/Viewer/HOX4N52B466C>

료로서의 가치에만 머무르지 않고 다양한 콘텐츠로 활용될 수 있는 매우 중요한 문화유산이다. AI 번역은 이를 세계의 다양한 언어로 광범위하게 보급할 수 있는 기틀을 마련함으로써 한국의 문화와 역사를 전 세계에 전파할 수 있는 가능성을 열어줄 수 있다. 3) 학문적 가치의 확장: 고전 자료는 해당 언어를 구사하는 학자들에게만 접근 가능한 경향이 있다. AI를 활용한 번역은 다양한 언어로의 접근성을 향상시켜 전 세계 학자들이 연구에 활용할 수 있는 기회를 넓히는데 기여할 수 있다.

이와 같은 당위성은 물론 『조선왕조실록』은 다음과 같은 측면에서 AI 기반의 고전 번역을 수행해야 할 타당성을 찾아볼 수 있다: 1) 기술적인 진보: 최근 AI 번역 기술의 진보는 매우 빠르게 진행되고 있으며, 특히 문맥 이해와 번역 품질 측면에서 상당한 성과를 보이고 있다. 이는 고전 문서 번역의 정확도와 번역의 품질을 향상시킬 수 있는 잠재력을 보여준다. 2) 프로젝트 수행 기간 단축: 『조선왕조실록』과 같은 대규모 문서의 번역에는 기본적으로 수십 년이 소요될 수 있다. AI 번역을 활용하면 번역 프로세스를 가속화하여 소요시간을 대폭 줄일 수 있다. 3) 비용 절감: AI 기계 번역은 초기 투자 이후 지속적인 번역 작업에 투입되는 비용을 대폭 줄일 수 있다. 『조선왕조실록』번역과 같은 장기적인 프로젝트에서 이를 활용하면 상당한 규모의 경제적인 이점을 얻을 수 있다. 4) 번역품질 관리: AI 번역 시스템은 지속적인 학습과 개선을 통해 번역품질을 체계적이며 지속적으로 유지할 수 있다. 이는 다수의 번역가를 투입하여 수작업 번역을 수행할 때 발생할 수 있는 번역품질의 변동성을 최소화하는데 기여할 수 있다.

이러한 당위성과 타당성을 바탕으로, AI 번역 기술의 혁신적인 발전이 이루어지고 있는 현재 상황에서 AI를 활용한 『조선왕조실록』의 영역 및 다국어 번역 프로젝트 수행은 반드시 수행해야 할 시대적인 요청이자 사명이라 할 수 있다.

3. 다국어 번역 프로젝트 수행의 가능성

3.1. 사전 고려사항

『조선왕조실록』의 다국어 번역 프로젝트를 수행하기 위해서는 우선적으로 언어, 규모, 시차 및 소요기간, 과업의 성격, 참조용 메타텍스트 활용 등 원전과 번

역본 사이의 근본적인 차이점 및 현황에 대한 충분한 인식과 고려가 선행되어야 한다. 주요 고려대상을 살펴보면 다음과 같다.

3.1.1. 언어적인 측면

우선, 언어적인 측면에서 『조선왕조실록』의 영역 완간을 위해서는 최소한 세 가지 언어와 관련된다는 점을 고려해야 한다. 우선 한자어로 기록된 『조선왕조실록』의 저본을 적극 참조하되 영역 및 다국어 번역 프로젝트 수행을 위해서는 한국어 번역본을 원전으로 삼을 것을 제안한다. 그리고 원전으로 삼을 국역본 역시 앞서 언급한 바와 같이 한글로 완역된 국역본 대신 현재 번역작업이 진행 중이지만 신국역본을 우선적으로 그리고 최대한 활용해야 한다는 점 또한 제안한다.

신국역본을 우선적으로 활용해야 하는 이유는 무엇보다도 완역된 국역본에 비하여 21세기 현재 독자들이 사용하는 언어 표현과 어법에 가깝기 때문이다. 그리고 향후 고전 번역의 지식자산 인프라로서의 활용성을 확장하려면 인간과 기계번역의 협업을 기반으로 전문용어 및 어휘 등의 TM 구축부터 맞춤형 AI 학습용 번역 코퍼스 구축에 이르기까지 일련의 작업 수행 과정에서의 생산성과 효율성 그리고 신뢰성 또한 우선적으로 전제되어야 하기 때문이다. 뿐만 아니라 고전번역 프로젝트 수행 과정에서 수반되는 한국어와 영어의 이중어 병렬 코퍼스 구축 작업에 국역 번역본에 사용되거나 병기된 한자어 처리 작업 관련 부분도 고려해야 하기 때문이다. 정부의 고전국역 정책에 따른 번역본 간행량, 번역률, 그리고 한자 사용률의 변화 분석(서정문 2011: 94-95)¹²⁾에 의하면 『조선왕조실록』의 국역본에 포함된 한자어 사용비율은 수행기간 및 수행기관별로 차이는 있지만 상당한 비율을 차지하고 있다는 점도 참고할 필요가 있다. 단언할 수는 없지만 신국역본에서는 기존의 번역본에 비하여 한자어 사용비율이 한층 낮아질 것으로 전망된다. 이는 코퍼스 구축 과정을 실질적으로 훨씬 용이하게 이루어지도록 작용하는 요소가 될 것이므로 신국역본을 적극적으로 활용할 필요가 있다.

3.1.2. 규모적인 측면

『조선왕조실록』은 조선의 시조인 태조부터 철종까지 25대 472년간(1392~1803)

12) 1기인 1968-1976년 기간 동안 민족문화추진회와 세종대왕기념사업회에서 각각 11책과 36책의 번역을 수행하면서 원문의 글자수 대비 한자 사용률이 각각 6.41%와 6.74%로 나타났다. 2기인 1977년부터 1984년에는 6.05%와 10.05%, 그리고 3기인 1985년부터 1993년까지 4.84%와 8.74%로 나타났다.

에 걸친 172,000여 일의 역사를 무려 6,400만자의 문자로 총 1,893권 888책으로 기록한 방대한 양의 기록물로서 그 자체가 빅데이터라 할 수 있다. 국역 『조선왕조실록』번역본 서지정보 총괄표(서정문 2011: 86, 그림5)에 의하면 한역 작업시 태백산사고본을 국사편찬위원회에서 영인본으로 발간한 서책을 저본으로 활용하였다고 한다. 한역이 이루어진 실록의 수록 범위 역시 1392년 태조원년 7월부터 철종 14년 12월까지의 472년에 이르는 기간이며, 한역 작업은 1968년부터 1995년(색인 발행 포함)까지 28년에 걸쳐서 이루어졌다. 원문은 30,904면이며, 번역문은 139,206면, 그리고 색인은 17,748면으로 전체 쪽수는 187,934면에 이른다. 원문의 글자수는 47,699,848자¹³⁾이며 번역문의 글자수는 173,312,517자에 이른다. 번역문 내에 포함된 한자 또한 12,188,818자로 규모면에서의 방대함 또한 기계번역과의 협업이 반드시 수반되어야 할 당위성에 보다 강력한 설득력을 제공한다.

3.1.3. 시간적인 격차

『조선왕조실록』과 같은 고전을 현대 언어로 번역하는 작업은 단순히 언어의 장벽을 넘어 시간적인 격차까지 고려해야 하는 복잡한 과제이다. 원본이 작성된 시기부터 현재에 이르기까지의 긴 세월 동안 언어는 어휘, 문법, 사용법 등에서 큰 변화가 수반된다. 그리고 시간이 흐르면서 문화적 배경과 사회적 관습 또한 변화한다. 때문에 고전 문헌에서 사용하는 언어는 현대 독자에게 낯설게 느껴질 수 있으며, 번역가는 원본이 생성된 당대의 문화적 맥락을 이해하고, 이를 현대 언어로 풀어내야 하는 동시에 원본의 뉘앙스와 의미를 충실히 보존해야 한다. 이처럼 시간적인 격차는 원천문화권에서 텍스트를 이해하고 수용하는 번역가와 독자 모두에게 다양한 언어적, 문화적 도전을 제공한다. 따라서 『조선왕조실록』의 다국어 번역 프로젝트를 기획할 때 원전과 번역본 사이의 시간적인 격차는 반드시 고려해야 하는 요소이다.

3.1.4. 고전 번역의 특수성

『조선왕조실록』과 같은 고전 번역은 단순한 언어 변환을 넘어서, 당대의 문화적, 사회적 맥락을 현대의 독자에게 전달하는 복잡한 과정이 수반된다. 특히, 기록

13) 고종실록과 순종실록도 조선왕조실록의 일부로서 편찬되었으나, 이 실록을 편찬할 때는 이미 일제강점기였으므로 북한 번역본에는 고종과 순종황제 실록도 포함되어 있으나, 남한에서는 두 실록은 일제 강점기에 일본인이 편찬한 것으로 사실의 왜곡이 심하여 엄격한 실록 편찬 규례에 맞지 않는다고 판단하여 번역 대상에서 제외하였다. <https://db.itkc.or.kr/etc/desc?type=dbList>

의 정확성을 유지하면서 동시에 독자들의 이해도를 높이기 위한 주석과 해제의 추가는 번역 과정에서 매우 중요한 요소로 작용한다. 이는 인간 번역가의 깊은 이해와 섬세한 판단이 필요한 부분이며, AI 기반 번역 시스템 개발과정에도 일종의 큰 도전 과제에 해당한다.

그러므로 고전 번역의 특수성에 입각하여 AI 번역 프로젝트를 수행할 때 무엇보다도 전문용어 및 고유명사 데이터베이스 구축을 염두에 두어야 한다. 고유명사와 전문용어는 번역에서 가장 빈번한 오류가 발생하는 요소이기 때문이다. 이를 위해 『조선왕조실록』에 특화된 전문용어와 고유명사에 대한 방대한 데이터베이스를 구축하고, 이를 AI가 학습할 수 있도록 해야 한다. 그리고 AI 번역이 직역투와 같은 오류를 범하지 않도록, 문맥 이해 및 해석의 정확성, 그리고 가독성을 동시에 만족시킬 수 있는 알고리즘의 개발이 요구된다. 뿐만 아니라 기계 번역문을 생성한 후 인간 번역가가 검토하고 수정하는 MTPE 과정을 필수적으로 포함해야 한다.

그리고 AI 시스템이 주석과 해제를 자동으로 생성할 수 있도록, 관련 정보를 추출하고 체계적으로 제공하는 메커니즘을 개발해야 한다. 이를 위해선 역사학, 문헌학 등 관련 학문의 전문가들과의 협업이 필수적이다. 이와 더불어 번역된 결과물에 대한 학계의 피드백을 시스템에 반영하여, AI가 지속적으로 학습하고 개선할 수 있는 구조를 마련해야 한다. 여기에 오류 수정과 어법의 현대화, 내용의 주석 추가 등도 포함되어야 한다.

이와 같이 『조선왕조실록』의 영역/다국어 번역 프로젝트에서 AI의 적용은 이러한 특수성을 고려하여 진행되어야 한다. AI 기술의 발전과 인문학적 접근의 균형을 맞추으로써, 고전의 정확하고 깊이 있는 번역이 가능해질 것이다. 이는 고전을 현대 독자에게 접근 가능하게 만들 뿐만 아니라, 원전의 의미를 충실히 전달하는 데 있어서도 중요한 역할을 할 것이다.

3.1.5. 생산성 및 효율성 측면

『조선왕조실록』의 1차 국역본은 앞서 언급한 바와 같이 약 30년의 기간이 소요되었다. 그리고 이 작업에 투입된 연도별/기간별 역자현황에 관한 데이터에 따르면, 인적 자원 또한 역자 164명, 해제자 18명, 교열자 50명, 그리고 운문자 80명 등 300여명이 참여(서정문 2011: 89, 표7)하였고 이를 30년으로 환산하면 연인원 9,000여명이 투입되었다고 할 수 있다. 투입된 인적자원의 규모 대비 당시 수작업

에 전적으로 의존할 수밖에 없었던 번역환경을 고려하면, 번역의 품질은 물론 국역본 완간 자체가 실로 기적에 가까운 대장정이었다고 할 수 있다. 따라서 『조선왕조실록』과 같은 방대한 분량의 고전을 번역할 때, 작업의 생산성 및 효율성 측면에서 소요 시간은 프로젝트의 전반적인 계획에 중대한 영향을 미치므로 이에 대한 사전 고려도 반드시 필요하다.

그리고 AI 번역 기술을 최적화한 『조선왕조실록』의 영역 및 다국어 번역 프로젝트 수행을 위한 맞춤형 번역 플랫폼을 구축하여 역자와 해제자, 교열자, 운문자, 그리고 용어추출 및 구축 전문가 등이 작업수행 지침에 따라 순차적으로 그리고 동시다발적으로 맡은바 역할을 수행할 수 있는 번역환경 구축이 최우선적으로 이루어져야 한다. 그리고 프로젝트 운영의 효율성과 인적자원의 수급 및 관리, 그리고 일정관리 및 원활한 예산 운용 등을 포괄하는 종합계획서 마련도 필요하다.

3.2. 프로젝트 수행방식

인간과 기계번역의 협업을 기반으로 다국어 『조선왕조실록』번역 프로젝트를 수행하기 위해서는 무엇보다도 맞춤형 시스템의 구축이 필요하다. AI 자동번역시스템을 적용한 『승정원일기』의 번역 프로젝트 설계 개요(김우정 외 2021: 1)에서와 마찬가지로 본 연구에서도 프로젝트 수행방식을 대체로 다음과 같이 제안한다. 먼저, 기 번역된 『태조실록』과 『세종실록』의 한글과 영어의 언어 쌍에서 일정 규모로 추출한 데이터를 AI에 적용하여 고도화된 학습을 수행할 수 있는 『조선왕조실록』자동번역시스템의 기반을 우선적으로 마련한다. 그리고 보다 정교한 작업을 수행하기 위하여 관련 메타텍스트와 분야별 고유명사와 전문용어집 등을 포함한 병렬데이터를 적용한 맞춤형 AI 학습용 코퍼스를 구축하여 자동번역시스템에 적용한다. 이렇게 추출한 AI 번역 결과물을 토대로 『조선왕조실록』의 원문과 비교하면서 전문번역가의 MTPE 작업을 포함한 교열, 윤색, 주석 등의 과정을 단계별로 거친다. 이러한 결과물은 번역 품질평가 과정을 거친 후 다시 AI 학습용 코퍼스로 활용하여 번역품질의 고도화를 위한 데이터베이스에 추가시키는 선순환 구조를 지속적으로 반복 유지하면서 영어 『조선왕조실록』의 번역작업을 마무리한다. 이렇게 해서 구축된 영어 『조선왕조실록』AI 자동 번역시스템을 다국어 번역 과정에서도 동일하게 적용하여 프로젝트를 수행하면 다양한 언어 쌍의 학습용 코퍼스가 상호보완 관계를 유지하는 선순환 구조 속에서 번역의 품질 또한 지속적

으로 향상될 수 있다.

3.2.1. 맞춤형 AI 학습용 번역 코퍼스 구축

이제 앞서 논의한 프로젝트 수행방식을 토대로 『조선왕조실록』의 AI 자동 번역 시스템 구축을 위하여 우선적으로 맞춤형 AI 학습용 번역 코퍼스 구축과 해당 코퍼스 구축에 적용할 메타텍스트의 범주에 관하여 논의하기로 한다. 먼저, 2012년 당시 국사편찬위원회가 주관하는 『조선왕조실록』영역 작업과 관련된 공식적인 입장을 살펴보면, 앞의 3.1에서 논의한 사전 고려사항 이외에도 프로젝트 수행과 관련하여 수반되는 다양한 측면의 준비사항을 추정할 수 있다. 무엇보다도 국사편찬위원회에서 우리나라 고전에 대한 영역 작업을 수행한 경험이 없었기 때문에, 『조선왕조실록』번역의 추진방향 설정을 목표로 당시 실록에 대한 샘플 번역작업을 먼저 시작하였다. 말하자면, 정종, 태종, 세종, 세조, 성종실록에서 모두 230건, 원문 기준 5만자 분량의 기사를 추출하여 번역을 수행하였다. 이 과정에서 용어표기 문제, 본문 가운데 한자 병기 여부, 특히 고유명사의 로마자 표기 방안 등 사업 기획단계에서는 명확하게 드러나지 않았던 각종 문제를 확인할 수 있었고, 이에 따라 해결방안을 모색할 수 있었다. 이러한 경험을 바탕으로 2013년부터 『세종실록』에서 즉위년 이래의 원문 기사를 대상으로 본격적인 번역작업을 개시하면서 매년 그 분량을 늘려가고 있다고 한다. 여기서 주목해야 할 점은 10여 년 전의 상황과는 달리 샘플 번역과 번역 품질 향상을 위한 다양한 해결방법 모색을 AI 기반의 번역 플랫폼을 활용하여 인간 번역가와 협업하면 상당 부분의 난제를 극복할 수 있다는 것이다.

본 연구는 규모의 방대함과 연구수행의 중요성을 감안하여 무엇보다도 『조선왕조실록』번역을 위한 맞춤형 AI 학습용 코퍼스 구축의 필요성과 시급성을 제안한다. 맞춤형 코퍼스 구축은 『조선왕조실록』의 영역 완간은 물론이며, 나아가 다국어 번역 프로젝트 수행을 위한 선행요건이기 때문이다. 이러한 AI 학습용 코퍼스 구축을 위하여 무엇보다도 인프라 데이터로 활용할 수 있는 기존의 성과물의 종류와 내용을 점검하고 목적에 부합하는 선별기준과 활용방안에 대한 연구 또한 시급하게 요청된다. 때문에 메타텍스트¹⁴⁾의 범주에 포괄할 수 있는 원본과 번역본

14) 메타텍스트는 한 텍스트가 다른 텍스트에 대해 언급하거나 참조하는 텍스트를 말한다. 이러한 메타텍스트는 원본 텍스트의 이해를 돕고, 분석, 비판, 해석에 필요한 다양한 관점을 제공할 수 있다. 메타텍스트의 활용은 원본 텍스트를 더 깊이 있고 다방면으로 이해하는 데 도움을 줄 수 있으며, 특히 『조선왕조실록』과 같은 역사적인 문서에서는 다양한 시각과 접근법을 통한 새로운 해석과

의 종류를 제안하고, 이의 구체적인 활용방안을 우선적으로 모색할 필요가 있다.

3.2.2. 활용 가능한 메타텍스트

그럼, 『조선왕조실록』맞춤형 AI 학습용 코퍼스 구축을 위해서 먼저 활용가능한 한국어와 영어 언어 쌍의 매칭가능한 메타텍스트의 종류를 살펴보기로 한다. 메타텍스트 선별방식은 저본(한자본)인 『조선왕조실록』의 한역본을 원전으로 삼고 원전과의 상관성에 따라 다음과 같이 분류하였다. 1차 메타텍스트는 원전인 『조선왕조실록』을 모태로 직접적으로 생산된 번역본을 지칭하며, 해당 목록에는 『태조실록』과 『세종실록』이 있으며, 참고로 영어 이외의 다른 언어로 이루어진 번역 사례는 아직 찾아볼 수 없다. 2차 메타텍스트는 원전을 직접 번역하지는 않았으나, 조선시대 당대의 주요 저서나 기록물을 영어로 번역한 도서 혹은 영어로 출간된 저서 중에서 사료적인 가치를 지닌 도서를 대상으로 다음과 같이 제시하였다.

표 1. 『조선왕조실록』AI 학습용 코퍼스 구축용 메타텍스트 분석 대상 도서

| 1차 메타텍스트 | | |
|--------------------------------------|--|--|
| ST | TT | 주요내용 및 활용가치 |
| 『조선왕조실록』 한역본 & 신역본 ¹⁵⁾ | 『태조실록』 <i>The Annals of King T'aejo: Founder of Korea's Chosŏn Dynasty</i> (2014) | - 『조선왕조실록』의 최초의 공식적인 영어 번역 도서 - 조선왕조(1392~1910 CE)를 건국한 태조 이성계의 통치에 대한 공식적인 역사 기록이자, 14세기 후반 한국 및 한국인의 삶을 재구성할 수 있는 중요 내용 수록 |
| | 『세종장헌대왕실록』 <i>The Veritable Records of King Sejong</i> (2017) ¹⁶⁾ | - 조선의 4대 국왕 세종(1418~1450)의 재위기간의 역사를 기록한 조선왕조실록 - 세종의 업적이 방대하여 실록의 원문 외에도 《세종실록지리지》, 《세종실록오례》, 《세종실록약보》, 《질정산》내외편이 별도로 수록되어 있음 - 때문에 세종 때만이 아니라 조선 초기의 정치·사회·경제·군사·문화를 전반적으로 이해하는 데 가장 기본이 되는 사료에 해당 |

발견을 가능하게 한다(전현주 2017b).

2차 메타텍스트

| | | |
|--|---|---|
| <p><i>Ryu Sŏngnyong, Chancellor of Chosŏn Korea: On the Battlefield and in Memory</i> (2022)¹⁷⁾</p> | | <ul style="list-style-type: none"> - 조선의 정치가, 전략가, 학자로서 류성룡(1542-1607)은 한국과 아시아 전체의 역사에서 중요한 역할 수행. 그의 전기는 16세기 조선의 정치와 외교관계, 그리고 특히 임진왜란과 관련된 당시의 시대상 고찰 가능 - 조선과 일본 사이의 임진왜란과 중국 제국과의 협상에 대한 도요토미 히데요시의 설명 포함 |
| <p>『정선 목민심서』 (2005)</p> | <p><i>Admonitions on Governing the People: Manual for All Administrators</i> (2010)</p> | <ul style="list-style-type: none"> - 『목민심서』는 다산의 대표적인 저작으로, 지방 수령이 백성을 다스리기 위해 꼭 알아야 할 원칙 및 지침과 세부 사항 수록 - 19세기 초 한국사회와 정세, 그리고 당시의 세계관 조망 가능 |
| <p>『징비록』(1994)</p> | <p><i>The Book of Corrections: Reflections on the National Crisis During the Japanese Invasion of Korea, 1592-1598</i> (2002)</p> | <ul style="list-style-type: none"> - 임진왜란 때 영의정으로서 군국정부의 중책을 짊어지고 외교·군무·민정 등에서 정치적 역량을 발휘해 충무공 이순신과 함께 국난을 극복하게 한 서애 유성룡의 기록인 『징비록』은 우리 역사에 드물게 보존되어온 기록문학에 해당 |
| <p>『북학의』(2011)</p> | <p><i>A Korean Scholar's Rude Awakening in Qing China: Pak Chega's Discourse on Northern Learning</i> (2019)</p> | <ul style="list-style-type: none"> - 박제가는 정조 2년 1778년에 종사관으로 청나라에 가게 되었다. 그곳에서 몇 달 머물면서 농잠, 축목, 성곽, 궁실, 배, 수레에서부터 심지어 붓, 자[尺] 등의 제도에 이르기까지 청나라의 습속 중 조선에서 시행하거나 적용할 만한 내용 기록 - 『북학의』의 정신은 청나라의 선진 문물을 배워 조선의 궁핍과 척박함을 타파하는 데 있다. 여기에는 18세기 후반 당대의 아웃사이더 지식인으로서의 치열한 고민이 녹아 있음 |

15) <https://db.itkc.or.kr/dir/item?itemId=JT#/dir/list?itemId=JT&gubun=book>

<https://db.itkc.or.kr/dir/item?itemId=JR#/dir/list?itemId=JR&gubun=book>

16) <http://esillok.history.go.kr/>

그리고 고전번역연구원이 주관하여 구축한 다양한 종류의 한국고전 DB와 부가 열람메뉴 등도 적극적으로 활용할 필요가 있다.

표 2. 『조선왕조실록』 AI 학습용 코퍼스 구축 참고용 부가자료 목록

| 구축기관명 | 유형 | 자료명 |
|-------------|------------------------|---|
| 한국 고전번역원 | 고전번역연구감 ¹⁸⁾ | 웹페이지 |
| | 한국고전DB ¹⁹⁾ | 고전번역서 |
| | | 『조선왕조실록』 |
| | | 『승정원일기』 |
| | | 『일성록』 |
| | | 『신역 조선왕조실록』 |
| | | 고전픽션 쇼츠 콘텐츠, 고전명구 |
| | 부가열람메뉴 | 사전류: 고전번역서 각주정보, 고전용어 시소러스, 이체자 정보 |
| | | 해제류: 고전번역서 해제, 『신역 조선왕조실록』 해제, 고전원문 해제 한국문집총간 해제, 한국고전총간 해제 |
| | | 부가서비스: 고전번역서 서지정보, 한문고전자동번역, 경서성독, 한중일연호검색, 전자책 라이브러리 |

3.3. 프로젝트 수행 가능성

3.3.1. AI 번역 접목 사례

양질의 기계번역 결과를 얻기 위해서는 AI를 활용한 특화된 플랫폼 구축도 중요하지만 무엇보다도 AI 학습용 코퍼스 구축을 위한 양질의 데이터를 충분히 확보하는 것도 매우 중요하다. 양질의 데이터란 곧 양질의 코퍼스를 의미하기 때문이다. 다음은 고전번역에 특화된 AI 번역 시스템을 구축하고자 시도한 다양한 사

17) 해당 도서는 저서에 해당하지만 그리고 한국어-영어 매칭 쌍에 해당하지는 않지만 조선시대의 선조가 재임했던 시기의 중요 인물로서 특히 임진왜란과 이순신, 그리고 『징비록』 등 류성룡 관련 메타텍스트를 참조하여 집필한 저서이므로 『조선왕조실록』의 메타텍스트 목록에 포함하였다.

https://www.amazon.com/Ryu-S-%C5%8Fngnyong-Chancellor-Chos%C5%8Fn-Korea/dp/1557291993/ref=sr_1_2?qid=1699314152&refinements=p_27%3AByonghyon+Choi&s=books&sr=1-2&text=Byonghyon+Choi

18) <https://www.itkc.or.kr/bbs/board.do?id=76&menuId=86#1>

19) <https://db.itkc.or.kr/>

례들이다. 시스템 구축을 주관하는 기관에 따라 적용하는 기술의 종류는 다르지만 본질적으로 양질의 병렬 코퍼스를 구축하고 AI 기반의 번역 기계학습을 진행하며 번역품질 개선을 위한 고도화 작업을 거친 후 고전문헌 번역에 적합한 모델을 개발하는 단계는 공통적으로 적용되고 있음을 알 수 있다. 고전번역에 적합한 AI 번역 시스템을 개발하기 위한 사례들을 통하여, 기계번역을 활용한 고전번역의 실현가능성은 더욱 높아지고 있음을 알 수 있다.

표 3. AI 번역 기술을 적용한 고전 번역 사례

| 시기 | 대상텍스트 | 주관기관 | 특징 |
|---------|---------------------|---------------------------------|--|
| 2018.03 | 『승정원일기』 | 한국고전번역원; 시스템란 ²⁰⁾ | <ul style="list-style-type: none"> - “1994년부터 『승정원일기』번역이 시작됐지만 양이 워낙 방대해 전문 번역사들은 아직 20%밖에 번역하지 못했다”며 “2062년으로 예상한 완역 시점을 AI 번역과 병행하면 30년 이상 앞당길 것으로 기대한다”고 말했다. AI가 초벌 번역을, 전문 번역가가 감수를 맡는 식의 분업이 되면 작업 소요시간이 확 줄기 때문 - 영조 즉위년~4년치 기록에서 추출된 ‘한글·한자’ 단어쌍 35만개를 기반으로 6개월간 AI 번역 기계학습 진행 - “단어와 구문을 쪼개어 번역하는 기존의 통계기반 번역이 아니라 문장을 통째로 파악해 번역하는 인공지능경망 번역” 적용 - “문장 맥락의 이해도와 정확도가 전문번역사 대비 70% 수준 |
| 2018.04 | 『승정원일기』 『조선왕조실록』 | 한국고전번역원; 시스템란 ²¹⁾ | <ul style="list-style-type: none"> - 2017년 인공지능경망 기계번역 기술을 고전문헌 번역에 활용한 ‘AI 기반 고전문헌 자동번역시스템’ 구축 - NMT 기술을 활용, 번역원이 제공한 승정원일기 번역본과 원문 간 연관 패턴을 찾아내 모델 구축 - 『승정원일기』 적용 모델을 고도화해 『조선왕조실록』번역 모델 구축 계획 |

20) https://biz.chosun.com/site/data/html_dir/2018/03/02/2018030200233.html

21) <https://www.etnews.com/20180416000283>

22) <https://www.sedaily.com/NewsView/1VODCP03Q9>

| | | | |
|---------|---------------------|--|--|
| 2019.09 | 『태조실록』 『정조실록』 | 한국번역학회 ²²⁾ | <ul style="list-style-type: none"> - 『태조실록』의 한국어 원문과 영어 번역본 <i>The Annals of King T'aejo</i>에서 발췌해 구축한 한영 병렬 37,736문장에 일반 문어체 약 600만 문장을 더해 기계 학습을 수행하고 그 결과로 만들어진 ‘고전 번역 모델’을 『태조실록』과 『정조실록』의 영어 번역에 적용하는 방식으로 진행 - 기계번역 결과물의 품질에 대한 회의론이 존재하는 상황에서 한국 고전과 같이 고도로 복잡한 텍스트에 대한 기계번역 적용 가능성을 진단함으로써 기계번역의 미래를 예측할 수 있는 방향 제시 목적 - 역사문화적 측면에서 필요성이 크지만 규모의 방대성 때문에 속도를 내지 못하고 있는 『조선왕조실록』, 『승정원일기』, 『일성록』, 『난중일기』 등 한국 고전의 번역 작업에 기계번역을 적용할 수 있는 근거를 제공 - AI를 효과적으로 학습시킬 데이터 확보와 학습 데이터 처리 인력 양성 필요성을 강조 |
| 2022.09 | 『조선왕조실록』 | 엘솔루 ²³⁾ (구 시스트란 인터내셔널) | <ul style="list-style-type: none"> - 2019년 한국고전번역원에 『조선왕조실록』, 『승정원일기』, 『천문』고서에 특화한 인공지능 자동 번역 시스템을 개발 및 구축한 실증 사례 소개 - 2017년 『승정원일기』를 시작으로 조선왕조실록, 천문 고문헌 『제가역상집』, 『천동상위고』 등에도 인공지능 번역 솔루션이 적용돼 100여 명 정도 남은 고문헌 전문번역가의 수고를 덜게 |
| 2021.03 | 『조선왕조실록』 『승정원일기』 | 중앙대학교 AI 학과·영상학과 김영빈 교수 연구팀; 스캐터랩; KAIST AI 대학원 ²⁴⁾ | <ul style="list-style-type: none"> - AI 모델 기반 『승정원일기』 복원·번역 성공 - 완성된 『조선왕조실록』 데이터를 활용해 인공지능(AI) 모델 학습시킨 결과 완성하는데 수십 년이 넘게 걸릴 것으로 예상됐던 『승정원일기』를 단 며칠 사이에 복원·번역할 수 있음을 입증. 복원·번역 소요시간 획기적으로 단축 가능 |

3.3.2. 인간과 기계번역의 협업

위의 AI 번역 접목 사례에서 살펴보았듯이 고전번역 프로젝트는 이제 더 이상 인간 번역사의 수작업에만 의존할 수 없는 영역으로 선회하였음을 알 수 있다. 『조선왕조실록』과 같은 대규모의 프로젝트는 장기적으로 다수의 번역사가 동시 다발적으로 작업에 투입되어야 하는 작업이므로 특히 인간과 기계번역의 효율적인 협업을 수행할 수 있는 번역 플랫폼 활용은 더 이상 선택의 문제가 아니다. 관련 프로젝트에 투입되는 다수의 번역사뿐만 아니라 그들의 역량과 자질, 원문에 대한 이해도, 수행방식, 번역지침 준수 정도, 담당 분야, 작업 분량, 참여 시기 및 기간의 차이 등 각종 상황이 번역 프로젝트의 관리 및 결과물 도출 과정에서 다양한 변수로 작용한다. 이 때문에 적극적인 AI 번역 기술을 적용한 맞춤형 플랫폼 환경을 구축하지 않으면 번역 품질의 관건을 좌우하는 여타의 수많은 요소들은 차치하고라도 전문용어 번역의 일관성조차도 담보하기 어려워진다.

따라서 신뢰할만한 메타텍스트를 기반으로 원어와 번역어 표현의 정확성과 적절성 등 충분한 검증을 거친 기계번역 수행용 전문용어집을 번역 메모리(TM: translation memory)로 활용하면 다수의 번역사가 협업하는 상황에서도 번역의 일관성 유지는 상대적으로 수월해질 것이다. 뿐만 아니라 이러한 TM을 더욱 확장하여 『조선왕조실록』의 영역을 위한 맞춤형 AI 번역 코퍼스를 구축하면 인간번역사와 기계번역의 협업을 기반으로 하는 번역의 생산성과 품질평가, 그리고 감수에 이르는 일련의 공동작업의 효율성이 적극적으로 강화될 수 있다. 그리고 이러한 결과물을 재가공한 AI 번역 코퍼스를 기반으로 심화 및 강화학습 과정을 거쳐서 후속 작업에 투입하면, 번역 결과물의 품질유지 및 향상을 도모할 수 있게 된다. 뿐만 아니라 이러한 과정에서 추출되는 코퍼스의 중간결과물은 그 자체로서 고전 분야는 물론 관련 콘텐츠의 다국어 번역과 교육, 자연어 처리 등의 영역으로 적용 가능성을 확장할 수 있으므로 언어자산으로서의 가치 또한 상승시킬 수 있는 승수효과를 거둘 수 있을 것이다.

지금까지 논의한 사항을 포괄하여 실제로 번역작업 수행 과정과 관련하여 인간과 기계번역의 협업이 요구되는 과정은 대체로 다음과 같다:

- 1) 『조선왕조실록』 텍스트 전처리 작업: 텍스트 전처리는 텍스트를 분할하고,

23) <https://www.koya-culture.com/news/article.html?no=135973>

24) <http://www.edpl.co.kr>

언어 감지 및 형태소 분석, 단어 정제 등;

2) 전문용어 및 어휘목록 번역 메모리 생성 과정: 번역지침에 따른 주제별, 분야별 전문용어, 어휘 목록을 구축하여 정확성, 일관성, 효율성 도모;

3) 맞춤형 기계번역 엔진 구축: 『조선왕조실록』의 다양한 특성과 문체, 정제된 전문용어 및 병렬 텍스트 코퍼스 등을 적용하여 기계학습, 딥러닝, 자연어 처리 등 강화학습 과정을 거친 맞춤형 번역 시스템 구축;

4) 사전 편집 작업 (pre-editing): 기계 번역 결과물에 대한 파일럿 스터디 후 기계번역의 한계점을 보완하는 사전 편집 작업을 수행, 그리고 기계번역 결과물에 대한 유형별 오류 특성 분석 후 사전편집 작업을 통한 기계번역 결과물의 품질향상 도모;

5) 사후 편집 작업(MTPE): 기계번역 기반의 번역 품질 검사 수행, 그리고 원문과 대조하여 전문용어, 맥락오류, 정확성, 오역, 누락, 의역 등 인간번역사의 기계번역 결과물 보완 및 수정, 편집 작업을 통한 번역 품질 향상 도모;

6) 감수 및 수정: 기계번역 검수 프로그램 활용한 표준 평가 척도 기반의 번역 품질 평가 및 인간번역사/감수자의 기계번역 결과물에 대한 평가, 그리고 그 결과를 반영한 감수 및 평가 작업 반복 수행;

7) 종합 감수 및 번역품질 평가 후 프로젝트 마무리 등.

4. 연구 요약 및 제언

4.1. 연구요약

지금까지 『조선왕조실록』을 인간과 기계번역의 협업 패러다임을 기반으로 다국어 번역 프로젝트 수행의 타당성과 가능성에 관하여 논의하였다. 인공지능망 기계번역과 생성형 AI 등에 적용되는 기술을 폭넓게 적용하여 고전 번역에 최적화된 자동번역 시스템을 구축하고, AI 번역 기계학습을 통하여 기계번역의 품질 향상에 주력하는 다양한 사례들은 매우 고무적이다. 그리고 이러한 AI 번역 기술의 발전과 연구에 힘입어 번역가들이 고전 작품의 번역에 있어서 오랫동안 직면해온 시대적, 언어적, 문화적 장벽을 뛰어넘을 수 있는 가능성은 더욱 현실화되어 가고 있다.

이처럼 혁신적인 AI 번역 기술을 접목하는 시대에 언어적인 한계 때문에 원천 문화권에만 머물러있는 『조선왕조실록』을 비롯한 우리나라 고전 문헌과 문학 작품의 활용성이 제한되는 현실을 타개하고 이들이 지닌 콘텐츠로서의 가치를 본격적으로 확장할 수 있는 대안을 적극적으로 마련할 필요가 있다. 이를 위한 가장 현실적인 대안은 바로 글로벌 커뮤니티에서 보다 폭넓은 독자군을 확보하는 것이며, 이를 위해서 『조선왕조실록』에 최적화된 AI 번역 기술을 접목하여 다국어 번역 프로젝트를 수행하는 방안을 실행하는 것이다. 그리고 이렇게 수행하는 성과는 『조선왕조실록』에만 머무르지 않고 고전 문헌과 문학 전반에 걸친 다양한 역사적인 기록물들이 다언어로 거듭날 수 있는 파급효과도 기대할 수 있을 것이다.

이러한 현실적인 활용가능성과 이에 대한 연구의 필요성을 동인으로 본고는 『조선왕조실록』의 다국어 번역 프로젝트 수행의 가능성을 구체화하기 위하여 고전번역의 특수성을 감안한 사전 고려사항과 맞춤형 AI 번역 코퍼스 구축, 이에 수반되는 관련 메타 텍스트의 범주, AI 번역 접목 사례와 인간과 기계번역의 협업방식 등에 관하여 논의하였다. 특히, 『조선왕조실록』에 특화된 맞춤형 AI 학습용 코퍼스 구축과 이에 적용할 메타텍스트의 선정 및 범주에 관한 논의는 고품질의 학습용 코퍼스 구축이 프로젝트 수행의 성패를 가늠하는 가장 중요한 단계로 인식하기 때문이다.

그리고 『조선왕조실록』과 같은 고전을 현대 언어로 번역하는 작업은 단순히 언어의 장벽을 넘어 프로젝트를 기획하는 단계에서 사전에 고려해야 할 다양한 변수들을 언어적인 측면, 규모적인 측면, 저본과 번역본 저술의 시간적인 격차 그리고 생산성 및 효율성 측면에서 논의하였다. 그리고 프로젝트 수행방식은 맞춤형 AI 학습용 번역 코퍼스 구축과 메타텍스트로 활용할 수 있는 『조선왕조실록』의 영어 번역 도서와 조선시대를 배경으로 사료의 가치를 지닌 저서의 번역도서와 저서를 중심으로 선정하였다. 프로젝트의 수행가능성과 관련해서는 AI 번역을 접목한 고전 번역의 사례를 살펴보고 인간과 기계번역의 협업 방식에 관하여 논의하였다.

4.2. 제언 및 후속 연구의 필요성

먼저 『조선왕조실록』의 다국어 번역 프로젝트를 기획할 때 종합적으로 고려해야 할 사항을 제안하면 다음과 같다:

www.kci.go.kr

1) 장기적 계획 수립의 필요성: 해당 번역 프로젝트는 단기간에 완료될 수 있는 작업이 아니다. 그러므로 장기적인 계획과 체계적인 관리가 필요하며, 이는 작업에 투입되는 예상 기간과 세부적인 일정을 마련하여 인적자원과 물적 자원 등을 배분하고 관리해야 한다.

2) 현대어 번역 전략 구축: 원문의 의미를 존중하면서 동시에 현대의 독자들의 가독성을 확보하는 전략을 개발해야 한다. 이는 원본의 언어와 문화적 맥락을 현대적으로 재해석하는 작업도 포함된다. 그리고 글로벌 커뮤니티의 독자층 확산을 위하여 언어 및 문화권별 맞춤형 번역전략도 반드시 수립해야 한다.

3) 다양한 분야의 전문가 인력풀 확충 및 협업 시스템 마련: 언어학자, 문화학자, 역사학자, AI 번역 프로그램 개발 전문가 등 다양한 분야의 전문가들과의 협업을 통해 안정적인 AI 번역 플랫폼 환경에서 작업을 수행할 수 있고 번역품질을 담보하는 결과물을 도출할 수 있는 지킴 마련에 전문가 집단의 적극적인 참여를 도모해야 한다.

4) 기술의 활용: 인공지능과 같은 최신 기술을 활용하여 번역의 효율성과 정확성을 높일 수 있는 방안을 모색해야 한다. 이는 작업에 투입하는 시간과 비용의 부담을 경감시킬 뿐만 아니라, 지속적인 업데이트와 갱신을 통하여 프로젝트 전반의 원활한 운용에도 영향을 미치기 때문이다.

5) 번역품질 관리와 검토: 번역의 품질을 지속적으로 관리하고 검증하는 체계적인 관리 방안이 가동되어야 한다. 번역은 초기 작업 이후에도 지속적인 수정과 보완을 요구하는 작업이므로, 체계적인 품질 관리 시스템을 구축하는 것이 중요하다.

6) 주석과 해제의 중요성: 600여 년의 시간적인 격차를 극복하고 원본의 문화적 맥락을 이해하는 데 있어 주석과 해제 작업은 필수적이다. 독자가 번역본을 통해 원전의 의미와 문화적 배경 이해를 돕는 역할은 물론 학문적인 측면에서도 사료에 기반한 주석과 해제는 고전 번역 연구의 중요한 지침을 제공하는 역할도 수행하기 때문이다.

7) 교육적 가치와 함의성 인식: 고전 번역 프로젝트는 단순히 문서를 다른 언어로 옮기는 작업을 넘어, 원전에 포괄된 문화와 역사를 전파하고 교육 부문에서 활용할 수 있는 무한한 가치와 함의성도 지닌다. 특히 고전 번역을 수행할 번역, 해제, 주석, 전문용어 구축, 번역품질 감수 및 측정 등에 투입할 인재 양성과도 직결되므로 프로젝트의 설계와 실행 과정에서부터 인식이 필요하다.

8) 지속 가능한 자원 배분: 장기 프로젝트의 특성상, 안정적이고 지속 가능한 자원 배분이 필수적이다. 특히 번역에 필요한 인적자원, 재정, 시간 등의 자원이 효율적으로 배분되어야 한다.

9) 역동적인 프로젝트 관리: 프로젝트를 진행하는 시간이 경과하면서 발생할 수 있는 여러 변수들에 대처하기 위해 유연하고 역동적인 프로젝트 관리가 요구된다. 이는 변화하는 학계의 기준, 기술의 발전, 투입 자원과 재원의 변화 등을 포함한다.

마지막으로 본 연구는 『조선왕조실록』의 다국어 번역 프로젝트 수행의 가능성을 인간과 기계번역의 협업 패러다임을 기반으로 AI 기계 번역 기술의 적용 효과를 확장할 수 있는 맞춤형 AI 학습용 코퍼스 구축과 적극적인 메타텍스트의 활용, 그리고 수행 방식에 역점을 두었다. 때문에 AI 번역을 활용한 구체적인 번역 방식이나 전략 등 세부적인 사항에 관해서는 논의의 범주에 포괄하지 않았다. 그리고 실질적인 번역작업 수행 과정과 관련하여 인간과 기계번역의 협업이 요구되는 과정을 포함한 프로젝트 수행 가능성에 관한 논의 또한 번역 프로젝트 기획 시 종합적인 고려사항을 제시하는 수준을 벗어나지 못한 연구의 한계가 있다.

이러한 한계점을 보완하기 위하여 다음과 같은 주제를 포함하여 관련 연구를 장기적인 후속 과제로 남겨 두기로 한다: 1) 『조선왕조실록』을 포함한 고전번역에 관한 종합 담론; 2) AI 번역 기반 『조선왕조실록』의 영어/다국어 번역 프로젝트 수행을 위한 종합기획서 구축; 3) 맞춤형 AI 학습용 번역 코퍼스 기반 『조선왕조실록』의 영어/다국어 번역 품질평가; 4) 맞춤형 AI 학습용 번역 코퍼스 구축용 번역 메모리 구축 방법론; 5) 영어/다국어 고전 번역 인적자원 관리 및 수급 방안 등.

그럼에도 본 연구는 그동안 축적된 번역학 분야의 AI 번역 관련 연구 성과를 토대로 고전번역 부문에서 활용할 수 있는 가능성을 적극적으로 모색하였다. 이는 번역학의 지평을 고전번역 분야로 확장하고 한국(학) 다국어 콘텐츠 구축의 생산자로서 보다 적극적인 역할을 담당할 수 있는 계기를 마련하는데 일조할 수 있을 것이다. 뿐만 아니라 한국의 고전 문헌이 로컬의 영역을 넘어서 글로벌 커뮤니티로 잠재적인 독자층을 확장하고 나아가서 다양한 콘텐츠로 활용될 수 있는 계기가 마련되기를 기대한다. 그리고 이를 위하여 인간 번역사와 AI 기계번역의 협업 패러다임을 기반으로 번역학이 인문학의 전통적인 영역을 보완하고 강화하는 방향으로 진행되기를 기대한다.

참고문헌

- 권오현. (2021). 「『조선왕조실록』 번역사업과 남북한 체제경쟁: 1970년대와 1980년대를 중심으로」. 『문화와 사회』 29(3): 73-115.
- 권정만, 김학만. (2014). 「『조선왕조실록』에 나타난 통상(通商)의 용례와 언어로서의 분석」. 『한국동북아논총』 19(3): 307-326.
- 김성문. (2018). 「인공지능 시대와 고전문학」. 『문화와융합』 40(6): 129-154.
- 김용철. (2013). 「고전번역학의 학적대상 시론」. 『코기토』(74): 159-189.
- 김우정. (2021). 「고전문언문(古典文言文) 기계번역의 현황과 과제」. 『중국문학』 109(0): 21-51.
- 김우정 외. (2021). 「승정원일기 자동번역 시스템의 활용방안 연구」. 한국고전번역원.
- 김주원. (2006). 「『조선왕조실록』의 번역에 나타난 오류」. 『알타이학보』 16: 1-33.
- 김준연. (2021). 「AI 시대 중국 고전문학 연구의 새로운 방법론 탐색」. 『중국문학』 106(0): 1-28.
- 박제가. (2011). 『북학의』. 을유문화사.
- 박현주. (2018). 「『조선왕조실록』번역-번역 성과물의 연계성 및 활용 방안 고찰」. 『번역학연구』 19(1): 61-97.
- 신성운. (2009). 「고전번역사업 추진의 효율화 및 성과활용 극대화 방안 기획연구」. 한국고전번역원.
- 서정문. (2011). 「국역조선왕조실록 현대화 사업계획 수립을 위한 기초조사 연구(I) & (II)」. 한국고전번역원.
- 서정화. (2015). 「한국고전번역원의 주석DB 활용 및 개선방안 연구」. 『민족문화』 45: 69-96.
- 연규동, 이전경, 김은희, 김남사. (2012). 「『조선왕조실록』국역본에 나타난 文字의 번역 문제」. 『인문과학』 96: 45-65.
- 유성룡. 이민수 역. (1994). 『징비록』. 을유문화사.
- 이근우. (2014). 「『조선왕조실록』 일본 관련 자료 번역의 문제점」. 『동북아 문화연구』 1(40): 5-26.
- 이영훈. (2011). 「한국에서의 번역 개념의 역사: 『조선왕조실록』에서 본 ‘번역」. 『통번역학연구』 15(1): 129-151.
- 이지은·최효은. (2022). 「인공신경망 특허 기계번역 성능에 관한 연구: Patent Translate와 WIPO Translate 한영 번역 결과물의 누락과 통사 오류 분석을 중심으로」. 『T&I REVIEW』 12(2): 105-130.
- 전현주. (2022). 「인공지능 번역플랫폼 기반 번역가의 직명 및 직무기술의 분화에 관한 연구」. 『통번역학연구』 26(1): 167-193.
- 전현주. (2020). 「인간과 기계번역의 공존 패러다임 모색」. 『통번역교육연구』 18(4): 59-96.
- 전현주. (2017a). 「4차 산업혁명과 한국의 번역산업 현황 및 통번역 교육의 미래」. 『통번역교육연구』 15(3): 235-261.
- 전현주. (2017b). 「『목민심서』의 영역본(英譯本) Admonitions on Governing the People 소고(小考): 한국학의 메타텍스트 관점에서」. 『통번역교육연구』 15(1): 233-255.
- 정만조. (2010). 「승정원일기 번역종합계획 수립을 위한 기초조사 연구」. 한국고전번역원.
- 정성훈, 하지영, 김우정. (2021). 「한문고전문헌의 기계번역 평가방안 탐색」. 『한문학논집(漢文

- 學論集』 60: 105-155.
- 정성훈. (2019). 「언어학과 기계 번역-한문학 텍스트의 기계 번역과 관련하여」. 『한문학논집 (漢文學論集)』 53: 7-38.
- 정약용. (2005). 『정선 목민심서』. 창비.
- 정영미. (2016). 「남북한 『조선왕조실록』 번역 비교 - 『현종실록』을 중심으로」. 『민족문화』 48: 299-342.
- 진재교. (2008). 「한문고전 번역의 특수성의 안과 밖」. 『민족문화』 32(32): 3-42.
- 최미경. (2023). 「메타텍스트, 상호텍스트, 하이퍼 텍스트- 한불고전번역 예를 통해서 본 텍스트성과 번역의 쟁점」. 『T&I REVIEW』 13(1): 81-106.
- 최성희. (2019). 『AI시대의 번역: 이론과 실제』. 서울: 세창출판사.
- 한용진. (2017). 「『조선왕조실록』에 나타난 교육 관련 용어 분석」. 『민족문화연구』 74: 357-380.
- 한국고전번역원. (2022). 『2022 한국고전연감』. 한국고전번역원.
- 허철. (2022). 「고전문헌(한문) 기계번역용 말뭉치 구축 방안 제안」. 『중국문학연구』 88: 31 9-340.
- 허철, 배은한, 박용범. (2019). 「한문고전 인공지능 번역 연구의 필요성과 선결 과제」. 『한문학논집』 53: 39-54.
- 허철. (2018). 「지식확산의 새로운 경로, 漢文 기계번역의 가능성과 구축 방향에 대한 일견」. 『민족문화』 51: 75-113.
- Byonghyon, C. (2022). *Ryu Sŏngnyŏng, Chancellor of Chosŏn Korea: On the Battlefield and in Memory* (Korea Research Monograph 39). Institute of East Asian Studies. University of California, Berkeley.
- Byonghyon, C. Translator and Annotator. (2014). *The Annals of King T'aejo: Founder of Korea's Chosŏn Dynasty*. Cambridge, Massachusetts, London: Harvard University Press.
- Chega, P. (Author), Byonghyon, C. (Translator), Seung B. K. (Translator), Timothy V. A. (Translator). (2019). *A Korean Scholar's Rude Awakening in Qing China: Pak Chega's Discourse on Northern Learning* (Korean Classics Library: Historical Materials Book 6). University of Hawaii Press.
- NATIONAL INSTITUTE OF KOREAN HISTORY. (2017). *The Veritable Records of King Sejong*. <http://esillok.history.go.kr/>
- Song-Nyong, Y. (Author), Byonghyon, C. (Author). (2002). *The Book of Corrections: Reflections on the National Crisis During the Japanese Invasion of Korea, 1592-1598* (Korea Research Monograph). Institute of East Asian Studies.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.
- Yagyong, C. (Author), Byonghyon, C. (Translator). (2010). *Admonitions on Governing the People: Manual for All Administrators First Edition*. University of California Press.

[인터넷 자료]

- 교육플러스. (2021.03.26.). 2억4250만자 '승정원일기' AI로 복원·초벌 번역 성공...고문서 연구 탄력 받나. [https://www.edpl.co.kr/news/articleView.html?idxno=488\(2023.11.01.검색\)](https://www.edpl.co.kr/news/articleView.html?idxno=488(2023.11.01.검색)).
- 국민일보. (2012.01.10.). 『조선왕조실록』번역.

www.kci.go.kr

- <https://news.kmib.co.kr/article/view.asp?arcid=0005725734&code=11171211>(2023.10.25. 검색).
- 국사편찬위원회. 『조선왕조실록』. <https://sillok.history.go.kr/main/main.do?jsessionid=3FD8C5E1E9EC50021BAE74F2000A7DB0>
<http://esillok.history.go.kr/about/veritableRecordsInfo.do?lang=ko>
<http://esillok.history.go.kr/about/veritableRecordsInfo.do?lang=en>(2023.9.25. 검색).
- 국사편찬위원회. 『세종장헌대왕실록』. <http://esillok.history.go.kr/>(2023.10.23. 검색).
『조선왕조실록』 원본보기.
https://sillok.history.go.kr/popup/viewer.do?id=kaa_000#(2023.11.05. 검색).
- 국사편찬위원회. (2021.12.27.). 보도자료: 한국고전번역원 신역 조선왕조실록 발간. <https://www.itkc.or.kr/bbs/boardView.do?id=44&blIdx=159029&page=1&menuId=102&bc=0>(2023.11.04. 검색)
- 나무위키. 『조선왕조실록』. <https://namu.wiki/w/%EC%A1%B0%EC%84%A0%EC%99%95%EC%A1%B0%EC%8B%A4%EB%A1%9D>(2023.9.25. 검색).
- 뉴스팀. (2017.07.20). 시스템, 인공지능 기반 번역기술 PNMT 통해 세계 최초 고전문헌-헌대문자동번역시스템 구축.
<http://www.newstap.co.kr/news/articleView.html?idxno=57102>(2023.2.25. 검색).
- 다음 정책브리핑. (2017.03.30). 전통의 진화, '세종실록'의 영역(英譯)과 간행.
<https://v.daum.net/v/20170330133517058>(2023.9.27. 검색).
- 다음 정책브리핑. (2005.07.18.) 유네스코 등록, 조선왕조실록(1)
<https://v.daum.net/v/20050718124028846?f=o>(2023.11.04. 검색)
- 서울경제. (2019.09.24.). ‘조선왕조실록’ AI 기술로 번역된다.
<https://www.sedaily.com/NewsView/1VODCP03Q9>(2023.09.15. 검색).
- 아마존. The Annals of King T'aejo: Founder of Korea's Chosŏn Dynasty. <https://www.amazon.com/Annals-King-Taejo-Founder-Dynasty/dp/0674281306>(2022.01.03. 검색)
- 연합뉴스. (2021.01.14). 승정원일기 등 한문고전 AI가 번역한다…시범서비스 시작.
<https://www.yna.co.kr/view/AKR20210114072200005>(2023.11.04. 검색).
- 우리문화신문. (2022.06.29). 인공지능 자동 번역시스템으로 고문헌 번역한다. <https://www.koya-culture.com/news/article.html?no=135973>(2023.11.01. 검색).
- 인공지능신문. (2023.09.21.). ETRI, 인공지능으로 한문고서 번역해 문화유산 접근성 높인다.
<https://www.aetimes.kr/news/articleView.html?idxno=28968>(2023.11.01. 검색).
- 전자신문. (2018.04.16). 『조선왕조실록』도 AI로 자동번역... 고전문헌 활용 높다. <https://www.etnews.com/20180416000283>(2023.11.5. 검색).
- 조선일보. (2018.03.02). 한문 AI 났다, 승정원일기 3245권 번역 30년 빨라진다. https://biz.chosun.com/site/data/html_dir/2018/03/02/2018030200233.html(2023.11.5. 검색)
- 한국고전번역원. 『고전번역연감』. <https://www.itkc.or.kr/bbs/board.do?id=76&menuId=86#1>(2023.03.15. 검색).
- 한국고전번역원. 「고전종합DB」, 「부가열람메뉴」. <https://db.itkc.or.kr/>(2023.11.02. 검색).
- 한국의 세계기록유산. 『조선왕조실록』. https://www.cha.go.kr/korea/heritage/world_heritage/record_treasure_02.jsp?mc=KS_01_03_03(2023.3.27. 검색).
- 행정안전부 국가기록원. 『조선왕조실록』. <https://theme.archives.go.kr/next/sillok/sub2.do> (2023.3.27. 검색).

- 헬로디디. (2021.01.14.). 읽기 어려운 한문고전, AI가 번역한다. <https://www.hellodd.com/news/articleView.html?idxno=91377>(2023.11.01.검색).
- ChatGPT. (2022.11.10.). <https://openai.com/blog/chatgpt>(2023.11.01.검색).
- KF한국교류재단. (Winter, 2016). INTERVIEW Choi Byong-hyon: Bringing to Life Heroes from Korean Classics. <https://www.koreana.or.kr/koreana/na/ntt/selectNttInfo.do?mi=1079&nttSn=50491&bbsId=1121&langTy=KOR>(2023.04.30.검색).
- Yun-jöng, S. (2016). *The Veritable Records of King Sejong*. Volume 1. National Institute of Korean History. <https://books.google.co.kr/books?id=7ky7swEACAAJ&dq=The+Veritable+Records+of+King+Sejong&hl=en&sa=X&ved=2ahUKewjMntq46cb9AhWDCN4KHT-1AME4ChDoAXoECAIQ>AQ(2023.3.25.검색).

This paper was received on 05 November 2023; revised on 20 November 2023; and accepted on 20 December 2023.

Author's email address

transju@shinhan.ac.kr; wisepearl33@gmail.com

About the author

Chun Hyunju is an Associate Professor at Shinhan University. Her research interests include translation project planning for Korean classics based on AI translation platform and constructing its relative AI learning corpus.

www.kci.go.kr