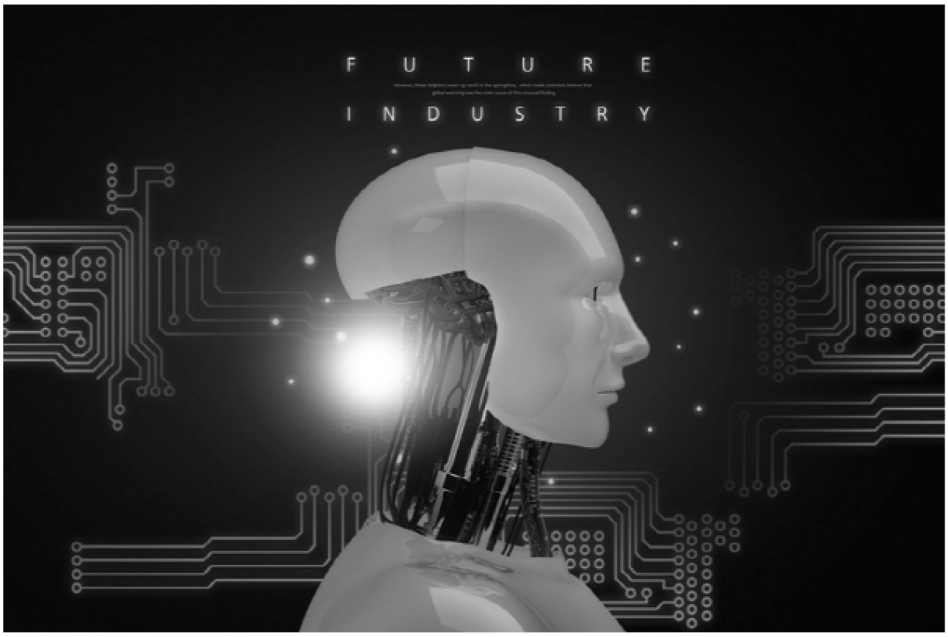
2023년 전통문화연구회 학술대회

東洋古典飜譯의 AI活用과 미래 그리고 傳統文化硏究會의 과제



일시 : 2023년 11월 15일 (수) 14시
장소 : YouTube 및 facebook 생중계
주최 : 社團法人 傳統文化硏究會

모시는 글

아름다운 결실의 계절 만추에 본회에서는 “東洋古典飜譯의 AI活用과 미래 그리고 傳統文化硏究會의 과제”라는 주제로 학술 발표회를 개최합니다.

이번 傳統文化硏究會 학술 발표회는 古典 飜譯의 AI 활용 사례와 성과를 살펴보고, 이에 따른 미래 전망과 도전을 탐구하려고 합니다. 어떻게 하면 고전 문학의 본질을 훼손하지 않으면서도 AI 기술을 효과적으로 활용할 수 있을까요? 그 해답을 찾기 위해 AI가 번역 분야에 미치는 영향과 함께 미래 번역가들의 역할이 어떻게 변화할지에 대해 이야기하고자 합니다.

이번 학술 발표회를 통해 우리는 인간과 기술의 협업이 어떻게 새로운 문화 생성 및 이해를 가능케 하는지를 살펴보고 더 나아가서는 우리의 문화유산을 보전하고, 현대 독자들에게 더 가깝게 전달할 수 있는 방안을 모색하게 될 것입니다.

금년도 국제학술대회는 2021년 11월 15일(수)에 온라인 생중계로 진행됩니다.
많은 참여와 성원 바랍니다. 감사합니다.

社團法人 傳統文化硏究會

이 사 장 李 啓 晃
회 장 郭 成 文

<2023년 학술대회 일정표>

東洋古典飜譯의 AI活用과 미래 그리고 傳統文化硏究會의 과제

사회 : 朴相水(본회 연구위원)

▣ 1부 : 개회선언 및 축사 ····· 14:00~14:05

- ■ 개회선언 朴相水(본회 연구위원)
- ■ 축　　사 金哲玟(국회의원, 국회 교육위원장)

▣ 2부 : 주제발표 ····· 14:05~17:00

1. 人工知能시대의 漢文敎育 - 傳統文化硏究會의 과제 -
   | 발 표 : 金 炫(본회 부회장, 한국학중앙연구원 인문정보학 교수)
2. 漢文古典AI飜譯支援시스템 開發과 活用
   | 발 표 : 李淸浩((주)에버트란 이사)
3. 東洋古典 漢韓말뭉치 구축과 활용
   | 발 표 : 尹鍾雄(인문정보연구소 소장)
4. '한문고전 자동번역' 개발의 의미와 과제
   | 발 표 : 金愚政(단국대 한문교육과 교수)

▣ 폐회 ····· 17:00~17:10

<목 차>

▣ 주제발표

人工知能 시대의 漢文敎育
- 傳統文化硏究會의 과제 -

金　炫
(본회 부회장, 한국학중앙연구원 인문정보학 교수)

1. 한문(漢文)의 세계에 다가온 인공지능
2. 인공지능과 동행하는 한문 공부
3. 인공지능을 교육하는 동양고전 연구

<논문요약>

Chat GPT와 같은 대화형 AI가 대변하는 오늘날의 인공지능은 처음 보는 한문 문장도 번역해낼 뿐 아니라, 그 뜻이 무엇인지 부연 설명해 줄 수 있는 수준으로 진화했다. 아직까지는 미흡한 점이 많다 하더라도 그 능력이 점점 더 증대되어 갈 것이고, 결국 동양고전(東洋古典)의 세계도 '인공지능'과의 동행하는 길을 찾지 않고는 더 이상 존립할 수 없는 시대에 이를 것이다.

우리는 미래 세대가 동양 고전을 공부하는 방식은 우리가 해 온 공부법과는 완전히 달라질 것이라는 사실을 인정하고, 그것에 대비하는 새로운 교육 프로그램을 준비해야 한다. 디지털 원어민이라고 해야 할 젊은 세대들은 새로운 한문 문장을 만나면 자신이 터득한 문리로써 그것을 해석하거나 선생에게 묻기보다는 인공지능에게 먼저 그 해석을 구할 것이다. 미래의 한문 선생은 모든 것을 스스로 가르치려 하기보다는 인공지능을 조력자(助力者)로 활용할 수 있어야 한다.

동양고전의 지식 세계에서 유의미한 학술적 지식의 중계자 역할을 할 인공지능 응용 시스템의 개발은 전통문화연구회의 새로운 미래과제이다. 지금까지 인간들만

이 이해할 수 있는 언어로 탐구되었던 동양고전의 지식을 '데이터'로 전환하고, 그것을 컴퓨터에게 학습시킴으로써 인공지능이 동양고전 분야에서도 신뢰할 수 있는 지식 탐구의 동반자 역할을 할 수 있게 하는 것. 이제 우리는 이를 위한 실천 전략과 구체적인 로드맵을 마련해야 할 시점에 서 있다.

키워드 : Chat GPT, 인공지능, 데이터 기반 지도학습, 강화학습, 인공지능과 협업하는 동양고전 교육

1. 한문(漢文)의 세계에 다가온 인공지능

대화형 인공지능 Chat GPT(Generative Pre-traine d Transformer)에 대한 논의로 세계가 떠들썩하다. 2022년 12월에 출시된 이후 그로부터 반년도 지나지 않아 학습 데이터량을 획기적으로 증대시킨 새 버전이 발표되었다. 여기에서는 한국어를 비롯한 다국어(多國語) 처리 기능 향상이 특별히 눈에 뜨인다.

도대체 이 소란이 무엇인지 이해가 안 되는 사람도 많겠지만, IT 산업계(産業界)는 물론이고 학계(學界)와 교육계(敎育界)도 이것으로부터 받은 충격이 이만저만하지 않다.

오늘날의 인공지능 프로그램이 우리의 관심사인 "한문(漢文) 공부(工夫)"와 무슨 관계가 있는지 이야기하기 위해 문제의 Chat GPT의 간단한 사용 예시를 보이도록 하겠다.[1)]

- 나의 질문 :

"子曰學而時習之, 不亦說乎!" 이 글에서 '학(學)'과 '습(習)'은 무슨 뜻이고 어떠한 차이가 있습니까?

1) Chat GPT와의 대화할 수 있는 사이트의 주소 : https : //chat.openai.com/

- Chat GPT의 대답 :

이 문장은 고대 중국의 철학자인 공자(孔子)의 말씀입니다.

“子曰學而時習之, 不亦說乎!” 라는 문장에서 ‘학(學)’은 ‘배우다’라는 뜻이며, ‘습(習)’은 ‘습득하다’라는 뜻입니다. 이 둘은 비슷한 의미를 가지고 있지만, ‘학(學)’은 지식을 습득하는 과정 자체를 가리키고, ‘습(習)’은 습득한 지식을 꾸준히 복습하고 익히는 것을 말합니다.

즉, 이 문장에서 ‘학이시습지(學而時習之)’는 "지식을 배우고 꾸준히 복습하여 습득하라"는 뜻입니다. 이것이 공자의 교육 철학 중 하나인 "학습의 반복성"을 강조하는 말씀입니다. 이 말은 공자가 학습에 대해 단순한 지식 습득을 넘어서 꾸준한 복습과 연습이 중요하다는 것을 강조하고 있습니다.

- 나의 질문 :

공자의 이 말씀에 대해 조선의 유학자 다산 정약용은 “學所以知也. 習所以行也. 後世之學, 學而不習, 所以無可悅也.”라고 풀이하였습니다.(정약용(丁若鏞), ≪논어고금주(論語古今註)≫) 이 문장을 한국어로 번역하고 정약용의 주장은 무엇인지 설명해 주세요.

- Chat GPT의 대답

다산 정약용은 "공부는 지식을 얻기 위한 것이며, 연습은 실천하기 위한 것이다. 그래서 나중 사람들이 공부할 때, 공부만 하고 연습을 하지 않으면 기쁨을 느낄 수 없다." 라고 번역할 수 있습니다.

정약용은 이 말씀을 통해 지식을 얻는 것이 중요하지만, 지식만 갖고 있어서는 실제로 그것을 활용하는 데에는 한계가 있다는 것을 말하고 있습니다. 따라서, 얻은 지식을 실제로 활용하기 위해서는 연습과 실행이 필요하다는 것을 강조하고 있습니다. 즉, 공부와 실천이 모두 중요하다는 것을 말하고 있는 것입니다.

Chat GPT가 어느 수준의 한문 문장을 번역할 수 있고, 그 내용을 풀어서 설명할 수 있는지 알아보고자 여러 가지 문헌의 문장을 가지고 이런 식의 질문을 던져 보았다. 한국어 번역은 전체적으로 어색했지만, 영어로 번역하라고 하면 놀라우리만큼 정확한 답을 주는 경우가 많았다. 영문 번역이 존재하는 중국 고전은 당연히 인공지능이 그것을 학습했을 터이니 정확한 번역이 가능할 것이다. 그러나 영어 번역이 있을 것 같지 않은 우리나라의 한문 텍스트도 꽤 정확하게 번역하는 것은 이 인공지능이 상당한 응용 능력이 있음을 알게 한다. 이것은 한국한문이나 중국한문이나 "한문"으로서의 공통성이 있는 것이고, 인공지능이 그 공통성에 기초해서 유의미한 예측을 할 수 있게 된 때문이다. ≪여유당전서(與猶堂全書)≫나 ≪조선왕조실록(朝鮮王朝實錄)≫의 문장 하나하나는 중국 고전에 없는 표현일 수 있어도 그런 식의 표현을 만들어 내는 용어와 용법은 중국 문헌에도 적지 않다는 것이다. 물론, 한국의 고유한 제도나 개념어 등 '공통성'에서 벗어난 단어가 들어간 텍스트의 경우, 그 부분은 틀리거나 이상한 번역으로 표현되기도 한다.

현재의 Chat GPT는 고전 한문 번역을 위해 특별히 훈련된 AI가 아니다. 전지구의 인터넷 서버에 존재하는 방대한 양의 데이터(500TB, 책 3천만권이 넘는 분량)를 망라적으로 학습하는 가운데 고전 한문에 관한 지식이라고 할 수 있는 것도 더불어 얻게 된 것으로 보인다. 아직까지는 미흡한 점이 많다 하더라도 그 능력이 점점 더 증대되어 갈 것이고, 결국 동양고전(東洋古典)의 세계도 '인공지능'과의 동행하는 길을 찾지 않고는 더 이상 존립할 수 없는 시대에 이를 것이다. 그리 머지않아서.....

2. 인공지능과 동행하는 한문 공부

그동안 한문 고전의 교육과 연구가 주된 관심사였던 전통문화연구회의 가족들도 이제는 이 '인공지능과의 동행'에 관해 보다 분명한 입장 정리와 실천이 있어야 할 것이다. 무엇부터 시작해야 할까?

우선, 미래 세대가 동양 고전을 공부하는 방식은 우리가 해 온 공부법과는 완전히 달라질 것이라는 사실을 인정하고, 그것에 대비하는 새로운 교육 프로그램을 준비해야 한다. 디지털 원어민이라고 해야 할 젊은 세대들은 새로운 한문 문장을 만나면 자신이 터득한 문리로써 그것을 해석하거나 선생에게 묻기보다는 인공지능에게 먼저 그 해석을 구할 것이다. 미래의 한문 선생은 모든 것을 스스로 가르치려 하기보다는 인공지능을 조력자(助力者)로 활용할 수 있어야 한다.

상업적인 현대 외국어 번역 시장에서는 이미 이런 식의 해법이 낯설지 않다. 영어 번역의 경우, 전문 번역가는 원고를 처음부터 직접 번역하기보다 인공지능(기계번역 프로그램)에게 초벌 번역을 맡기고, 자신은 그 결과 속의 오류를 수정하거나 그 뜻이 사람들에게 잘 읽혀질 수 있도록 수정하는 일을 전문적으로 수행한다. 이런 식의 일처리가 시간과 비용을 절약하는 방법이라고 인식하는 번역 회사들이 점점 늘어나고 있다. 하지만 이러한 변화에 대해 비용 절감을 위해 사람의 일을 기계가 대신하게 하는 것이라고만 이해해서는 안 된다. 인공지능의 번역결과를 검토하고 그 속의 오류를 수정하는 "새로운 전문성"이 강도 높게 요구되는 일이다. 예전의 기계번역 결과물은 그 표현이 투박했다. 그래서 문장을 매끄럽게 만드는 윤문이 사람의 손으로 하는 후처리 작업이라고 인식하곤 했다. 그러나 요즈음 Chat GPT가 보이는 상황은 그렇지가 않다. 적어도 영어문장의 경우, 그 표현이 상당히 자연스럽다는 것이 중론이다. 그러나 그 매끄러운 문장 속에 아는 사람이 아니면 속아 넘어갈 틀린 이야기가 섞여 있는 경우가 종종 발견되곤 한다. 인공지능이 만든 결과물을 검토할 전문가는 언어를 다듬는 윤문가가 아니라 그 내용의 사실성을 검증할 수 있는 해당 분야의 지식 전문가여야 한다.

미래의 한문 선생은 학생들에게 한문 구문의 분석 방법을 가르칠 필요가 없을지도 모른다. 학생들은 이미 훨씬 쉽게 접근할 수 있는 인공지능으로부터 한문 텍스트에 대응하는 영어나 한국어 대역문(對譯文)을 얻을 수 있고, 또 그 내용을 친절하게 설명해 주는 해설문(解說文)도 얻었을 수 있다. 선생의 역할은 학생들로 하여금 인공지능이 준 대답 중에서 신뢰할 수 있는 것과 그렇지 않은 것을 가려낼 수 있게 하는 안목을 갖게 하는 것, 그리고 인공지능이 도울 수 있는 것의 한계를 넘어서는 새로운 지식의 탐구 방법을 알려 주는 것이다. 인공지능 시대의 한문 선

생은 동양고전의 지식세계에서 깊은 통찰력과 구체적인 정보력을 지닌 지식전문가(知識專門家)여야 한다.

손에서 휴대전화를 놓고서는 살 수 없는 현대인들은 이미 일상생활에서 인공지능과 동행하는 존재가 되어버렸다. 정보 검색창에서 찾아지는 검색 결과 하나도 포털이 운영하는 인공지능 시스템에 의해 걸러진 것이다. 유튜브가 알아서 보여주는 연주 영상과 교양 프로, 온라인 쇼핑몰에서 구매를 권유하는 물품 목록이 모두 나의 취향을 데이터화한 인공지능의 선택이다.

동양고전(東洋古典)의 교육 현장에서도 '인공지능'과의 동행은 불가피하다. '인공지능이 한문을 해석하고 설명해 주니, 더 이상 동양 고전 교육은 불필요하다'는 생각은 '동행'의 필요성과 방법을 알지 못하는 이들의 푸념일 뿐이다. 미래의 한문 선생은 '주제넘게 한문까지도 해석하는' 인공지능을 무시하거나 적대시하기보다는 그것이 학생들의 고전 공부에 도움을 줄 수 있다는 사실을 인정하고, 인공지능의 교육적 기능이 더 유효하고 건실해질 수 있도록 하는 역할을 해야 한다.

3. 인공지능을 교육하는 동양고전 연구

Chat GPT와 같은 대화형 AI가 대변하는 오늘날의 인공지능은 처음 보는 한문 문장도 번역해낼 뿐 아니라, 그 뜻이 무엇인지 부연 설명해 줄 수 있는 수준으로 진화했다. 그것은 GPT라는 인공지능 엔진이 지난 수년간 우리의 상상을 뛰어넘는 방대한 분량의 텍스트를 학습하면서 스스로 터득한 '말주변'이다.

내가 Chat GPT의 놀라운 능력에 감탄하면서도 그 능력을 '말주변'이라고 폄하하는 이유는 그것이 주는 정보는 신뢰할 수 있는 '근거'로 '검증'할 수 있는 지식이 아니기 때문이다. Chat GPT가 둘러대는 이야기들이 그럴듯하게 들려도, 그 안에는 비슷하지만 틀리거나 관계가 없는 이야기도 적지 않다. Chat GPT는 기본적으로 무엇인가 실마리가 주어졌을 때 그것과 유관하면서 말이 되는 문장을 만들어내는 데 특화된 기계이다. 그러다보니 15세기의 역사적 사실을 이야기 하다가 유사한 용어로 설명되는 19세기의 일을 갖다 붙이는 식의 사고도 일어난다. 역사

적인 인물 여러 사람의 이름이 거론되는 한문 문장을 번역하라고 시키면, 그 문장 자체는 하자 없는 영문이나 한국어 현대문으로 번역한다. 그러고 나서 그 사람들이 누군가고 물으면 맞는 이야기와 무관한 이야기를 섞어서 누군지 모를 가공의 인물을 만들어내기도 한다.

사실상 지식 동반자로서의 인공지능은 아직까지 극복해야 할 많은 과제를 안고 있다. Chat GPT의 가장 심각한 문제는 틀린 이야기를 하면서도 스스로 그것이 틀렸다는 것을 모른다는 것이다. 하지만 여기서 맞거나 틀리다고 하는 것은 질문의 의도를 알고 있는 인간의 관점에서 하는 이야기이지 인공지능에게 그 책임을 물을 수 있는 것이 아니다. 인공지능의 모든 대답은 자신이 학습한 데이터를 가지고 스스로 만들어낸 사고의 논리에 따라 충실하게 답변한 것이라고 할 수 있다. 만일 어떤 특별한 지식 영역 안에서 '옳다'고 인정되는 답을 인공지능이 그대로, 또는 가깝게 찾아주기를 원한다면, 우리는 그 분야의 정리된 지식을 집중적으로 교육하는 방식으로 인공지능의 '전문성'을 제고시킬 수 있다. 이런 일을 '지도학습(supervised learning)'이라고 한다. '지도학습'보다는 느슨하게 '주어진 지식'과 인공지능 스스로 습득한 지적능력을 균형적으로 융합할 수 있게 하는 것을 '강화학습(reinforcement learning)'이라고 한다.[2)]

인공지능이 오답을 주기도 한다고 해서 그것을 외면하는 것은 길이 아님을 우리는 알고 있다. 그렇다고 해서 오류가 계속 확대재생산되는 것을 방관해서도 안 된다. 인공지능이 동양고전 교육과 연구의 동반자가 될 수 있게 하기 위해서는 그것이 이 세계의 지식에 대해서도 보다 건실한 지식과 지능을 쌓을 수 있도록 도와야 한다. 그 방법은 동양고전 연구자 스스로 인공지능에게 동양고전 지식을 전수하는 '지도 학습'이나 '강화 학습'에 참여하는 것이다. 인공지능 관련 기술의 특징은 그것을 어느 특정 조직이 폐쇄적이거나 독점적으로 운용하기보다는 다양한 외부 조직에서 그 기반 기술을 응용할 수 있도록 허용하고 협업을 촉진하는 문화 속에서 발전하고 있다는 점이다. GPT의 경우도 마찬가지로 GPT API(Application

2) 현재 세계인들에게 선보인 Chat GPT는 언어적인 관점에서 '자연스럽게' 대답하는 능력을 이 강화학습의 과정을 통해 습득했다고 한다. 지식의 정확성보다는 언어구사의 유창함을 우선시하는 부분도 있었을 것이다.

Programming Interface)를 제공함으로써 특정 목적으로 활용될 수 있는 응용 시스템 개발의 문을 열어 두고 있다. 이것을 이용하면, 우리 스스로 동양고전의 지식세계에서 보다 의미 있고 정확한 지식을 전달할 수 있는 특수 목적으로 GPT 응용 프로그램을 개발하는 것도 가능하다. 물론 이를 위해서는 동양고전 데이터의 특수성에 맞춰 GPT API의 기본 기능을 보완하는 맞춤형 알고리즘을 개발해야 하고 또 '지도학습'이나 '강화학습'에서 충분한 효과를 거둘 수 있도록 충분한 규모의 학습 데이터를 준비해야 한다.

동양고전의 지식 세계에서 유의미한 학술적 지식의 중계자 역할을 할 인공지능 응용 시스템의 개발은 전통문화연구회의 새로운 미래과제이다. 지금까지 인간들만이 이해할 수 있는 언어로 탐구되었던 동양고전의 지식을 '데이터'로 전환하고, 그것을 컴퓨터에게 학습시킴으로써 인공지능이 동양고전 분야에서도 신뢰할 수 있는 지식 탐구의 동반자 역할을 할 수 있게 하는 것. 이제 우리는 이를 위한 실천 전략과 구체적인 로드맵을 마련해야 할 시점에 서 있다.

漢文古典AI飜譯支援시스템 開發과 活用

李 淸 浩
((주)에버트란 이사)

<논문요약>

21세기에 이르러 기계번역 기술이 급속히 발전되는 상황에서, 이를 무시하고 넘어갈 수 없는 것은 자명하다. 기계번역 기술 이전에 번역학은 다양한 번역 현상을 분석하고 정리하는 것이 목적이며, 이를 통해 언어간, 문화간의 소통을 이루고, 폭넓은 인문학의 발전에 기여해 왔다.

산업계에서는 이미 기계번역을 다양한 방식으로 비즈니스화하여 그 발전을 위한 토대를 마련하고 있다. 이로 인해 번역이 인간 번역에 의존하던 시대가 곧 종료될 것이라는 강한 인상을 준다. 실제로 번역이 더 이상 고통받는 인간의 작업이 아니라는 희망을 심어주는 상황이 되고 있다.

국제적인 표준을 제정, 공포하고 있는 ISO(International Organization for Standardization)는 MTPE(Machine Translation Post Editing)에 관한 규격 ISO 18587 (Translation services – Post-editing of machine translation output)을 번역서비스로 공포함으로써 기계번역이 더 이상 인간번역에 대항하는 기술이 아니고, 번역서비스가 가능한 완성된 방법임을 전세계에 공포하고 있는 것이다.

아직 기계번역이 미완의 과제로 남겨두고 있는 영역은 고전과 문학 분야이다. 그래서 문학과 고전 분야의 번역이 기계번역으로 어느 정도 가능한지가 초미의 관심사인 것은 틀림없는 일이지만 최근 초거대 AI와 생성 AI 기술의 발전을 살펴보면 이 또한 불가능한 일이 아님을 쉽게 알 수 있을 것이다.

2022년 11월 30일에 CHATGPT 3.5가 나왔고, 2023년 3월 15일에는 CHATGPT 4.0이 세상에 등장하면서 이전에 챗봇, 음성비서 등 앞으로 10여년 이상의 AI 기술 발전이 필요하다고 생각했던 영역에서 혁명적인 발전으로 당장 실현 가능 영역으로 빠르게 변하고 있다.

이런 점을 감안해 볼 때 이전에 기계번역 기술의 적용이 불가능한 영역으로 간주되었던 고전 및 문학 번역 분야도 머지 않아 더욱 발전된 AI 번역기술에 의해 점령될 것은 명확한 일이다.

그렇다면 앞으로 더욱 발전할 AI 번역기술에 필요한 것은 무엇일까? 과학기술정보통신부가 2023년 4월 15일에 '초거대 AI 경쟁력 강화방안'을 발표하였다. 이 내용을 살펴보면 초거대 언어모델에 대한 국가적인 투자를 통해서 혁신적으로 다가오는 생성AI 시대에 적극적으로 대응하고, 이를 통해 AI의 선도적인 국가로 발전해 가는 비전을 제시하고 있다.

고전번역 분야에서 있어서도 초거대AI와 생성 AI 시대의 도래는 자명한 일이라고 보면 미래를 위해서 현재 무엇을 해야 할지를 쉽게 예측해 볼 수 있을 것이다. 그것은 바로 고전번역에 있어서 언어모델 생성이 가능토록 하는 빅데이터의 구축과 수집이라고 할 수 있다. 이는 전통문화연구회가 지난 2012년부터 현재까지 매년 투자해 온 정보화사업의 결과물이라 할 수 있을 것이다.

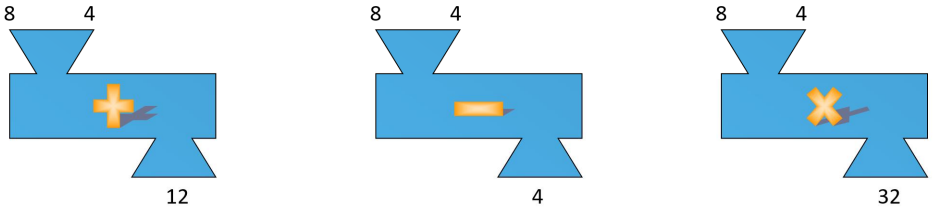
이처럼 지속적인 동양고전 정보화에 대한 투자는 전세계적으로 유일한 고전 데이터셋의 구축이 될 것이며, 빠르게 변하고 있는 AI 시대에 능동적으로 대처하고, 가까운 미래에 생성AI 기술이 일반화됨에 따른 핵심적인 준비가 될 것이다. 이는 고전분야에서 국가가 지향하는 초거대 AI의 경쟁력 강화 방안이 될 것이다.

1. AI의 理解

1) AI 技術의 特徵

AI 기술 이전에 소프트웨어의 개발은 입력과 출력에 맞는 알고리즘을 개발하는 방식으로 수행되었다. 그림에서 보는 바와 같이 8과 4과 입력되어 결과로 12가 나와야 한다면 "+" 기능을 처리하는 프로그램을 개발했다. 이처럼 입력과 출력 데이터에 따라 기능이 결정되는 방식을 프로그램 개발에 적용한 것이다.

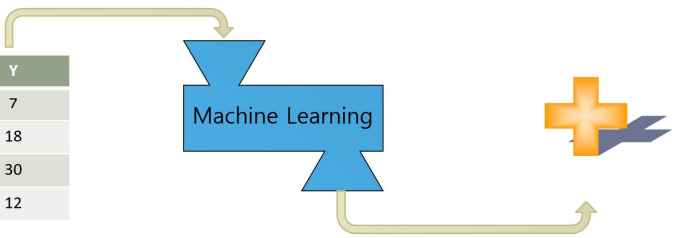
■ 전통적인 프로그래밍 – 프로그래머에 의한 개발 방식



이에 반해 AI 기술은 입력과 출력 데이터를 학습데이터로 해서 데이터들의 관계를 학습함으로써 프로그램(모델)이 개발되는 방식이다. 그림에서와 같이 입력인 x1, x2와 출력인 y 데이터를 학습하여 모든 입력과 출력의 조건에 맞는 "+" 모델을 결과로 만드는 과정인 것이다.

■ 인공지능 – 데이터에 의한 개발 방식

x1	x2	Y
3	4	7
8	10	18
10	20	30
5	7	12



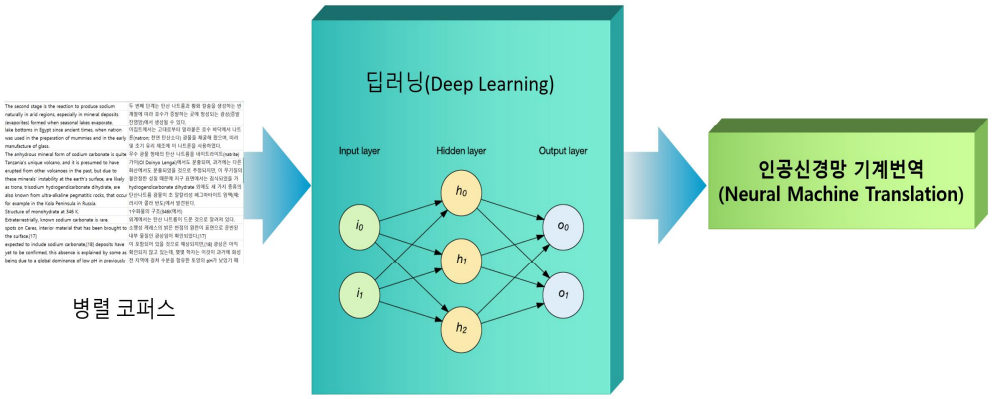
이러한 특징은 AI 기술이 데이터에 따라 필요한 기능을 개발하기에 매우 적합한 개발 방식이라는 것이다.

2) AI의 學習 方法

AI 기계번역모델이 만들어지는 과정을 살펴보면 그림과 같이 원문과 번역문으로만 구성된 병렬코퍼스(말뭉치)를 AI 기술의 핵심이라고 할 수 있는 딥러닝(Deep Learning)으로 학습시키면 인공신경망 기계번역(NMT; Neural Machine Translation) 모델이 생성된다. 이것의 특징은 번역 언어의 방향과 무관하게 딥러닝 기술은 동일한 것이 사용된다는 것이다.

즉 영한 번역모델을 만들려고 한다면 원문이 영어, 번역문이 한국어인 병렬말뭉치를 학습시키면 되고, 일한 번역모델을 만든다면 원문 일본어와 한국어 번역문의 병렬말뭉치를 학습시키면 된다는 것이다.

한한(漢韓) 번역모델을 만들려면 한자 원문과 한국어 번역문을 쌍으로 하는 병렬코퍼스를 딥러닝하면 한한(漢韓) 번역모델을 만들 수 있다.

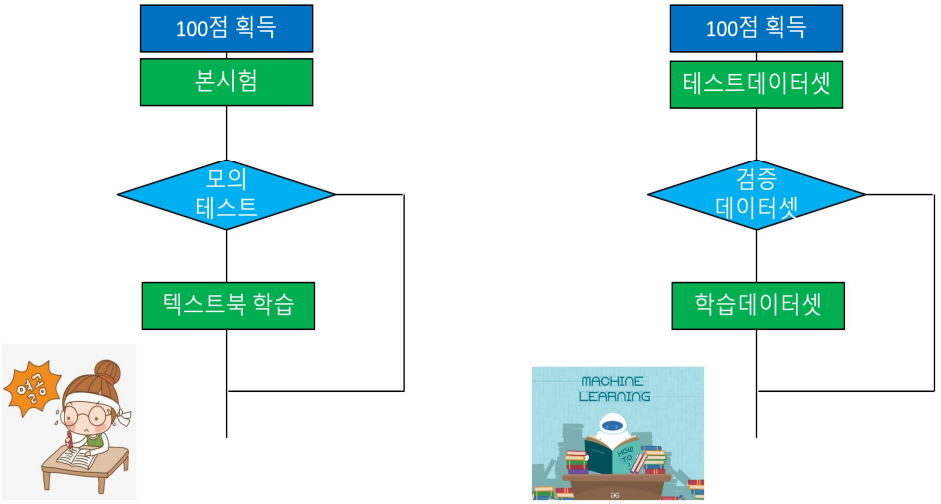


3) 딥러닝의 學習 方法

그렇다면 AI 기술의 꽃이라고 할 수 있는 딥러닝은 어떻게 학습을 하는 것일까? 이러한 학습의 과정을 이해하면 아무리 어려운 문제도 해결이 가능하다는 가능성을 알게 될 것이며, 이로써 향후 AI 기술의 발전은 지금까지는 불가능한 문제들이 해결 가능하게 될 것임을 깨닫는데 도움을 줄 것이다.

그림과 같이 딥러닝의 학습 방식은 인간이 학습을 하는 방식과 매우 흡사하다고 할 수 있다.

대학 입학을 위한 수능시험을 준비하는 학생은 먼저 교과서를 학습하고, 수능시험에 준하는 다양한 문제집을 통해 자신의 학습을 평가하는 과정을 반복하면서 학습이 이루어진다. 따라서 학습 능력의 완성도 여부는 문제집의 다양성과 난이도에 따라 결정되게 된다.



이처럼 딥러닝도 학습데이터를 이용하여 학습을 하고, 그 결과로 만들어진 모델을 자신이 평가하는 과정을 통해서 최적의 모델을 생성하게 된다.

기계번역 모델을 딥러닝하기 위해 100만 문장의 영한 병렬코퍼스가 있다고 가정해 보자. 딥러닝은 주어진 100만 문장을 모두 학습하는 것이 아니고, 학습데이터셋(80%), 검증데이터셋(10%), 테스트데이터셋(10%)으로 데이터셋을 분할한다.

딥러닝은 1차적으로 학습데이터셋(80만 문장)으로 학습을 한 후 모델을 생성하고, 학습된 번역모델로 검증데이터셋(10만 문장)의 원문을 번역하고, 번역문을 정답문장과 비교하여 자신이 학습한 모델을 평가한다. 평가결과가 흡족하지 않으면 다시 학습을 반복하게 된다. 이와 같은 과정으로 학습을 지속하여 10만 문장의 검증데이터셋을 모두 만족할 때까지 학습을 반복한다.

2. AI 技術의 發展

1) AI 技術의 登場

산업혁명의 단계로 보면 지금은 지능혁명인 4차 산업혁명의 시대라고 한다. 2010년대 들면서 컴퓨터가 지능을 갖추는 지능혁명은 빠르게 발전했다. 2016년 3월 이세돌과 알파고의 바둑 대결을 기억할 것이다.

알파고는 딥마인드(DeepMind)라는 영국의 인공지능 기업이 개발한 AI 프로그램이다. 총 5번의 대결이 이루어졌으며, 알파고가 4승 1패로 이세돌을 이겼다. 이 결과는 전 세계적으로 큰 반향을 일으켰는데, 그 이유는 바둑이라는 게임이 그동안 인공 지능이 해결하기 어려웠던 복잡한 문제 중 하나로 여겨졌기 때문이다.

인공 지능이 바둑의 세계 챔피언을 이기는 것은 기존의 기술 발전 속도 예측을 크게 앞당긴 이벤트로 평가받았다. 이 대결 이후, 바둑 뿐만 아니라 다양한 분야에서 인공 지능의 가능성과 미래에 대한 논의가 활발히 이루어졌고 특히 인공 지능의 발전이 인간의 직업과 능력, 그리고 미래 사회에 어떤 영향을 미칠 것인지에 대한 논의가 주요한 관심사로 부상하게 되었다.

2) 文脈을 이해하는 機械飜譯으로 發展

알파고가 나온 2016년 10월에는 구글이 세계 최초로 딥러닝 기술을 이용한 인공신경망 기계번역(NMT) 기술을 공개했다. 공개 이후 7년 간 기계번역 기술은 빠르게 발전하여 최근에는 문맥을 이해하는 기계번역으로 발전하고 있다.

문장단위로 번역을 수행하는 것과 여러 문장을 번역할 경우 다른 번역문이 나온다는 것이다.

원문	문장별 번역	문단별 번역
에밀리는 파리에 처음 왔다.	Emily first came to Paris.	It was Emily's first time in Paris.
그녀는 에펠탑을 보고 싶어했다.	She wanted to see the Eiffel Tower.	She wanted to see the Eiffel Tower.
그래서 지하철을 타고 트로카데로 역에 도착했다.	So I arrived at Trocadero station by subway.	So she took the metro and arrived at the Trocadero station.
역에서 내려 에펠탑을 향해 걸었다.	I got off the station and walked toward the Eiffel Tower.	She got off and walked toward the Eiffel Tower.

주어가 누락된 경우에 앞의 문맥에서 주어를 가져와야 정확한 번역이 되는데 문장단위로 할 때에는 이전 문맥 정보의 활용이 안되지만 문단별로 번역을 수행

하게 되면 문맥을 이해하여 정확한 주어로 번역한다. 이처럼 AI는 사람처럼 문맥을 이해하는 수준으로 발전해 가고 있다.

3) 生成型 AI의 登場

2023년은 2016년 이어 역사적인 지능혁명의 해로 기억될 것이다. 그 이유는 바로 ChatGPT 4.0의 등장이다. AI 발전 분야에서 이전까지는 기계번역 기술이 사람에게 유용한 AI 활용 분야로 손꼽혀 왔다. 그러나 ChatGPT의 등장은 이제 새로운 AI 시대를 만들어 가고 있다.

ChatGPT는 번역을 포함하여 사람이 사용하고 있는 언어를 이해하고, 생성하는 능력이 매우 뛰어나 AI 기술의 활용 분야 확대에 크게 기여하고 있다. 최근에는 Dalle-3의 개발로 디자인 분야에서도 AI 기술은 영향력을 확대해 가고 있고, 지금까기 AI 기술의 불완전함으로 인해 갖게 되었던 불신을 완전히 말끔하게 제거해 주는 획기적인 전환점이 되고 있다는 것이다.

3. 古典飜譯과 AI 技術

1) 고전 분야의 기계번역 기술 활용

대개의 경우, 인문학자라면, 특히 평생 고전번역에 종사한 번역가라면, AI를 통한 고전번역에 대해 거부반응 내지는 상반된 감정(Ambivalence)를 느낄지 모른다.

기계에 대한 인간의 불신과 신뢰는 어제 오늘의 일이 아니지만, 우리는 AI를 통한 고전번역 문제를 필요성과 당위성의 관점에서 접근할 시점에 도달했다. 왜냐하면 사람과는 비교도 할 수 없을 정도의 언어 문제 해결 능력을 갖춘 ChatGPT가 실용화되어가고 있기 때문이다.

2016년 인공신경망 기계번역(NMT) 기술이 나온 이후 고전번역원 등에서 딥러닝을 이용한 고전번역모델 개발에 투자하고 있고, 느리기는 하지만 점차 그 유용성을 확대해 가고 있다. 전통문화연구회도 2022년에 지금까지 구축된 동양고전 번역

말뭉치를 학습데이터로하여 한문 번역 모델을 생성하여 평가를 진행하였다.

2) 向後 古典飜譯 分野에서의 AI 技術 活用

앞에서 살펴 본 바와 같이 AI 번역기술은 사람의 능력을 추월하여 빠르게 발전하고 있고, ChatGPT의 등장으로 번역 이외에 언어 문제를 종합적으로 해결해 주는 AI 기술이 일반화되어 가고 있다.

AI 기술이 언어 문제를 해결하고 있는 발전 속도를 고려해 볼 때 고전번역에도 큰 변화가 일어날 것을 분명한 일이다.

GPT의 “T”는 Transformer의 약어이다. Transformer는 딥러닝의 핵심 알고리즘으로 2017년 구글에 의해서 세상에 공포되었다. ChatGPT를 개발한 OpenAI는 Transformer 기술을 이용하여 2018년도부터 GPT 1.0을 시작으로 꾸준히 업그레이드 하여 2023년 4.0를 공개하였다. Transformer의 원 소유자인 구글도 바드(Bard)를 런칭하였지만 ChatGPT에 비하면 부족함이 많이 보인다.

OpenAI는 2023년 말 또는 2024년 상반기 중에 5.0을 런칭할 계획이다. 4.0도 매우 훌륭한데 5.0이 1년도 안되어 론칭된다고 하니 그야 말로 AI 기술의 발전은 초초스피드라고 할 수 있을 것이다.

그렇다면 왜 ChatGPT 4.0의 등장에 전세계가 열광하는 이유는 무엇일까를 생각해 봐야 한다. 최근 ChatGPT에 푹 빠져 다양한 연구를 진행하고 있는 본 발표자가 ChatGPT을 사용하면서 생기는 느낌을 한마디로 요약한다면 “대단하고 매우 훌륭하다”이다. 활용 분야에 따라 느끼는 강도의 차이는 있겠지만 우수한 능력을 소유한 엔지니어 또는 전문가를 비서로 활용하고 있다는 느낌이다.

이처럼 AI 기술은 인간의 능력을 뛰어 넘어 초경지에 들어가고 있다는 것이다. 지금까지 고전 번역이라는 하는 특수한 영역에 AI 기술이 활용되기 어렵다는 것이 일반적이었다면 ChatGPT는 이전의 모든 부정을 말끔히 제거하고, 가까운 미래에 AI 기술이 활용될 것이 분명하기 때문이다. 이제는 기술의 개발을 기다리는 시간이 아니라 현재까지 나와 있는 기술을 어떻게 활용해서 적용할지를 고민해야 하는 시간이 되었다는 것이다.

3) AI 技術의 發展과 飜譯家의 役割

다양한 매체를 통해서 보도된 바와 같이 AI 기술이 발전하면 사라질 직업군에 통번역사가 들어 있다는 것은 많이 알려진 사실이다. 사실 직업이 없어진다고 하는 것은 새로운 시대의 도래로 직업이 변화된다는 것을 의미한다.

향후 가까운 미래에 AI 기계번역 기술은 일반 번역분야는 물론 고전번역에도 큰 기여를 하게 될 것이며, 이렇게 되면 번역사의 역할도 변화될 수밖에 없을 것이다.

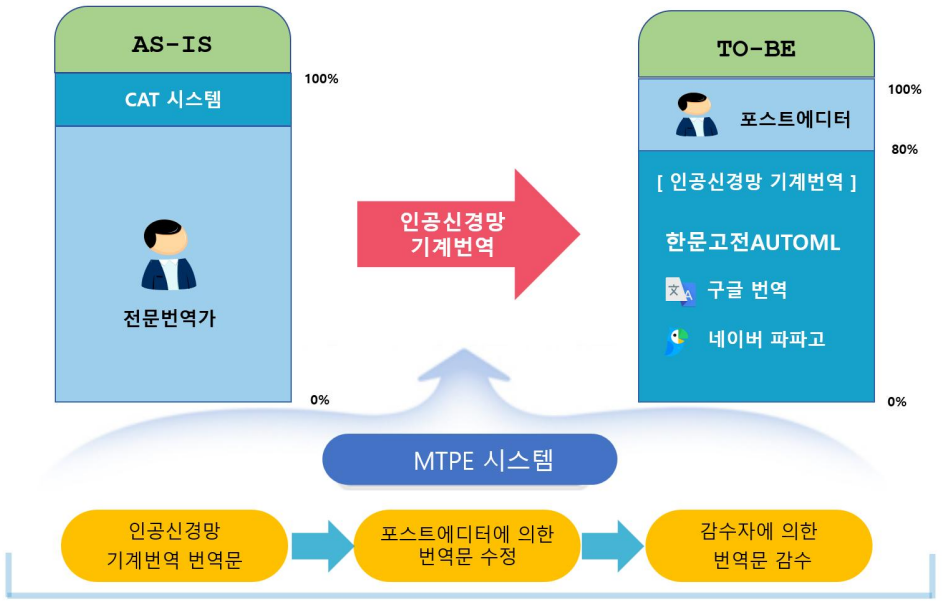
최근 산업번역 시장에서 번역사의 역할이 포스트에디터 역할로 변화되고 있다. 즉 1차 기계번역을 한 이후에 이를 수정하는 전문가 번역 서비스가 일반화되어 가고 있다는 것이다. 이는 그 만큼 기계번역의 1차 번역 품질이 우수하여 약간의 수정을 통해서 전문 번역 서비가 가능하게 됨을 의미하는 것이다.

고전번역 분야도 일반 산업분야보다 시간은 더 걸리겠지만 이처럼 기계번역 이후 포스트에디팅을 수행하는 것이 번역사의 역할이 될 것이다.

4. 漢文 古典 AI 飜譯支援시스템

1) 한문 고전 번역지원시스템 개발 배경

한문 고전번역에 MTPE(Machine Translation Post Editing)가 가능한 시스템을 개발하기 위해 전통문화연구회가 지난 10년간 고전DB 사업으로 축적한 한문/한글 병렬 말뭉치와 윤즈정보개발이 보유하고 있는 DB 구축 및 정제기술, 그리고 에버트란이 보유한 NMT 기술을 활용해서 2022년도에 고전번역 시스템을 개발하게 되었다. 이 시스템의 기본적인 목표는 한문 번역을 수행하는 번역사에게 AI 번역기술을 활용하여 1차 기계번역을 수행하고, 이를 편리하게 수정할 수 있도록 지원하는 것이다.



2) 향후 한문 고전 번역지원시스템 발전

본 시스템의 목적은 전문번역사에만 의존해서 수행하던 고전 번역업무를 AI기술의 활용을 통한 1차 번역과 전문번역가의 번역감수 업무를 협업하여 고전 번역업무의 생산성을 제고하자는 것이다.

아울러 번역모델 개발에 필수적인 번역말뭉치를 체계적으로 수집/구축하여 지속적으로 한자 번역모델의 번역품질 향상에 기여한다는 것이다. 향후에는 번역말뭉치 축적은 AI 기술의 개발 및 활용에 있어서 매우 중요한 요소가 된다. 번역지원시스템을 이용해서 번역을 수행하면 번역메모리가 축적되어 번역모델 학습을 위한 학습데이터로 바로 활용이 가능하게 된다.

따라서 현재 개발된 한자 번역모델의 번역 품질은 향후 지속적으로 개선될 것이며, 까까운 미래에는 한문 번역에 있어서 AI 번역 모델이 많은 부분을 담당하게 될 것이다.

東洋古典 漢韓 말뭉치 구축과 활용

尹 鍾 雄
(인문정보연구소 소장)

<논문요약>

(사)전통문화연구회는 일찍부터 기계 번역의 핵심은 말뭉치에 있다는 것을 인지하고 2018년부터 漢韓 병렬 말뭉치와 형태소 분석 말뭉치를 구축해 왔다.

필자가 재직하고 있는 (주)윤즈정보개발은 전통문화연구회에서 발주한 사업을 오랫동안 수행해오면서 기계 번역에 관심을 두고 연구를 계속해 왔다. 그 과정에서 AI 바우처 사업의 지원을 받아 AI 기반 번역 지원 시스템을 갖추게 되었고 '2023년 산업맞춤형 혁신 바우처 지원 사업'으로 기계 번역 모델에 대해 교육을 지원 받을 수 있는 기회를 얻게 되었다.

2023년 초에 필자가 재직하는 회사는 전통문화연구회, 에버트란과 MOU를 통해서 기계 번역 모델을 공동으로 연구하기로 협의를 하였다. 본 연구도 그러한 연구의 일환으로 기계 번역 모델을 개발하기 위해서 학습에 영향을 미치는 여러 가지 요소를 살펴보고 기계 번역의 성능을 최대로 발휘할 수 있는 조합을 찾아내려는 것이다.

우선 전통문화연구회에서 한한 병렬 말뭉치를 어떻게 구축해 왔는지 살펴보고, 기계 번역 품질에 영향을 미치는 요소가 무엇이 있는지 살펴 보았다.

기계 번역의 품질에 영향을 미치는 요소는 크게 AI 라이브러리와 딥러닝 모델, 데이터의 품질, 학습용 데이터의 운영, 하이퍼파라미터 등으로 구분할 수 있다.

Tensorflow나 Pytorch는 구글과 메타라는 초거대 기업이 막대한 예산을 투입해서 만든 라이브러리로 개인이나 일반 기업이 연구를 통해서 이들을 뛰어 넘는 라이브러리를 개발하는 것은 사실상 불가능하다고 봐야 한다. 결국 기계 번역 모델은 이들 라이브러리 중 하나를 선택하여 그 라이브러리를 기반으로 데이터를 정제하고 하이퍼 파라미터를 조정하여 모델의 성능을 향상시킬 최적의 조합을 찾아내야만 한다.

딥러닝계에서 괄목할 만한 성과를 낸 것으로 유명한 컴퓨터 과학자 앤드류 응은 “실질적으로 AI 시스템의 성능을 높이는 것은 코드의 대한 개선이 아니라 데이터에 대한 개선이다”라고 말했듯이 다른 기술적인 면보다 정제된 데이터가 모델의 성능에 차지하는 비중이 큼을 알 수 있다.

우리는 일반적으로 언어 생활을 하는데 있어 같은 표현을 반복해서 쓰기보다 같은 의미의 다른 표현을 다양하게 사용하는 것이 좋다고 배워왔다. 그러나 인간의 언어생활과 달리 AI를 학습시키는데 있어서 다양성과 일관성이라는 관점으로 접근하면 절대적으로 일관성이 중요하다. 이 일관성을 해치는 요소를 중심으로 살펴 보았다.

번역문에서 의미 전달을 위해 사용되는 의역 또한 데이터의 일관성을 저해하는 요소이다. 전문 번역가의 번역 스타일이 제각각 다르기 때문에 같은 문장이라도 다르게 번역하는 경우가 허다하다. 다행히도 전통문화연구회에서 한한 병렬 말뭉치를 구축할 때 원문에 사용된 문자가 번역문에 나타나는지 확인하여 번역문에 나와 있지 않으면 원문에 해당 부분에 [-]로 표시해 주었다. 반대로 원문에 없는 내용이 번역문에 추가되었으면 그 부분을 [-]로 표시해 주었다. 의역 문제를 해결한 정제된 말뭉치를 전통문화연구회에서 확보하고 있는 것이다.

문체나 글의 유형도 중요한 요소이다. 현재 구축한 말뭉치는 자치통감강목과 같은 역사서, 시경이나 당송팔대가의 시문, 일반 문집 등 다양한 문헌이 병렬 말뭉

치로 구축이 되었다. 이러한 분야의 다양성은 학습의 성능을 떨어뜨리므로 시대별, 문체 또는 장르별로 별도의 학습용 말뭉치 셋을 만들어 번역에 활용할 수도 있다. 그러나 문체별로 분류하면 자칫 학습에 활용할 말뭉치의 수가 부족해질 수 있으므로 주의해야 한다.

전통문화연구회에서 학습용 말뭉치를 구축할 때 구, 절, 문장 단위의 3세트의 말뭉치를 구축하였다. 이러한 3종류의 말뭉치를 학습했을 때 성능에 어떤 영향을 미치는지 확인할 필요가 있다.

실제 학습을 시킬 때 조정하는 하이퍼파라미터도 중요하다. 에포크, vocab 사이즈, 배치 사이즈 등의 하이퍼 파라미터를 조정하여 모델의 성능을 향상시키는 방법도 있으나, 성능을 높이기 위해 선택할 수 있는 변수가 그리 많지 않다.

학습 단계에서는 데이터 셔플(shuffle)을 통해서 랜덤하게 데이터를 섞어 주어야 한다. 이 셔플은 학습용 데이터를 생성할 때도 사용하지만 실제 학습을 할 때에서 매 애포크(epoch)마다 Training data를 셔플한다. 데이터셋을 섞지 않고 바로 split한다면 train data와 test data의 분포가 너무 다를 수가 있다.

ChatGPT는 초거대 언어 모델로 학습된 범용 모델을 기반으로 실행 단계에서 샘플 제공(few shot learning)만으로 어느 정도 영역별 과업을 수행할 수 있게 되었다. 이렇게 초거대 언어 모델을 이용하면 한문 번역의 수준도 어느 정도 개선이 될 것이다. 더 이상 AI를 이용하여 사람의 번역 수준에 근접하거나 뛰어 넘는 번역 모델을 만드는 것은 꿈이 아니다. 문제는 데이터를 수집하는데 많은 예산이 들고, 수집한 데이터를 학습시키는데에도 어마어마한 비용이 들어간다는 것이다. 그런데 한문 텍스트를 대량으로 수집해서 정제하는 것이 쉬운 작업이 아니다. ChatGPT는 데이터를 한 번 학습할 때마다 어마어마한 비용이 들어간다.

이때부터 기계 번역은 기술적인 문제에서 경제적인 논리로 전환된다. 한글과 특히 규모가 크지 않은 한문 시장에서 데이터를 수집하고 많은 학습 비용을 들여서 학습을 시킨다는 것은 기대하기 어렵다.

최근 도메인 파인튜닝이라는 개념이 도입되어 기존에 초거대 언어 모델을 기반으로 특정 분야에 전문적으로 학습시키는 방법이 도입이 되었다.

이 개념은 몇 개의 문장을 학습시키는 퓨샷러닝과 달리 몇 십만 단위의 어느

정도 규모가 있는 문장을 학습시켜 해당 도메인에 최적화시키는 방법이다.

전통문화연구회에서 구축한 데이터는 트랜스포머 기반으로 기계 번역기에 학습할 때 활용할 수 있으며, 초거대 언어 모델의 성능이 향상이 되면 그 바탕에서 도메인 파인 튜닝 데이터로 활용할 수 있을 것이다. 어떤 경우이든 전통문화연구회에서 구축한 한한 병렬 말뭉치는 그 가치는 잃지 않을 것이다.

1. 서문

1) 연구 배경

傳統文化硏究會는 일찍부터 기계 번역의 핵심이 말뭉치에 있다는 것을 인지하고 2017년부터 漢韓 竝列 말뭉치를 구축해 왔다. 필자는 한한 병렬 말뭉치를 구축하는 업무를 수행하면서 자연스럽게 구축된 한한 병렬 말뭉치를 활용할 방법을 찾게 되었고 특히 기계 번역에 관심을 갖게 되어 AI 기반 기계 번역 시스템을 갖추기 위해 많은 노력을 기울여 왔다.

그 결과 필자가 재직하고 있는 (주)윤즈정보개발은 2022년 정보통신산업진흥원에서 공모한 AI 바우처 사업의 지원을 받아 AI 기반 번역 지원 시스템을 갖추게 되었다. 그러나 한적 자료의 번역 특성상 필요한 원문과 번역문 텍스트의 검색, 각주의 처리 등의 기술 지원과 기계 번역의 품질이 역자들의 수준에 미치지 못하는 관계로 실제 번역 업무에 적용하기가 어려웠다.

2023년 과학기술정보통신부와 정보통신산업진흥원, 서울창조경제혁신센터에서 지원하는 ‘2023년 산업맞춤형 혁신 바우처 지원 사업’에 지원하여 AI 학습을 통해서 기계 번역 모델과 개체명 인식, 병렬 말뭉치 자동 생성 등의 분야에서 기계 번역 모델의 성능을 향상시킬 수 있는 교육을 지원받을 수 있게 되었다.

이러한 노력과 더불어 2023년 (주)윤즈정보개발은 전통문화연구회, 에버트란과 MOU를 맺고 공동으로 기계 번역 모델 성능을 개선하기로 협의를 하였다.

본 연구는 이런 성능 개선의 일환으로 전통문화연구회에서 구축한 한한 병렬 말뭉치의 역사와 그 내용을 살펴보고 기계 번역의 성능에 영향을 미치는 요소를

찾아보았다.

그리고 전통문화연구회에서 구축한 말뭉치에 AI 바우처 사업을 통해서 자체적으로 구축한 자치통감 병렬 말뭉치를 활용하여 AI 바우처 사업의 기계 번역 모델[1]과 혁신 바우처 사업의 OpenNMT 기반 기계 번역 모델[2]의 성능을 가능한 수준에서 테스트해 보았다.

2) 연구 목적

본 연구의 목적은 기계 번역 학습에 영향을 미치는 요소를 분석하고 성능을 최대로 발휘할 수 있는 요소를 찾아내는 데 있다. 기계 번역의 품질에 영향을 미치는 요소는 크게 AI 프레임워크와 딥러닝 모델, 데이터의 품질, 학습용 데이터의 운영, 하이퍼파라미터로 구분할 수 있다.

최근 AI 학습에 사용하는 프레임워크는 대부분 Tensorflow[3]나 Pytorch를 기반으로 하고 있다. Tensorflow나 Pytorch는 초거대 기업이 막대한 예산을 투입해서 만든 라이브러리로 개인이나 일반 기업이 연구를 통해서 이들을 뛰어넘는 라이브러리를 개발하는 것은 사실상 불가능하다고 봐야 한다. 결국, 기계 번역 모델은 이들 라이브러리 중 하나를 선택하여 데이터를 정제하고 하이퍼 파라미터를 조정하여 모델의 성능을 향상시킬 최적의 조합을 찾아내야만 한다.

2. 한한 병렬 말뭉치 구축 현황

현재 전통문화연구회에서 구축한 한한 병렬 말뭉치는 ≪論語≫, ≪孟子≫, ≪大學≫, ≪中庸≫ 등의 經典과 ≪資治通鑑綱目≫과 같은 역사서, 唐宋八大家의 시문, ≪莊子≫ 등의 제자백가 등 그 범주가 다양하다. 지금까지 전통문화연구회에

1) (주)에버트란이 개발한 기계 번역 모델로 파이토치 기반에 transformer 모델을 사용하였다.
2) (주)디엠티랩스에서 개발한 기계 번역 모델로 OpenNMT에서 공개한 오픈소스를 기반으로 최적화 한 것이다.
3) 구글에서 공개한 라이브러리로 11월 현재 2.14 버전까지

서 구축한 한한 병렬 말뭉치의 현황을 살펴보면 다음과 같다.

<표 1> 전통문화연구회에서 구축한 한한 병렬 말뭉치 연차별 현황

연차	구축 대상
2017년	격몽요결, 계몽편, 고문진보(전집~후집), 국어(1~2), 근사록집해(1~3), 논어주소(1~3), 논어집주, 당송팔대가문초 구양수(1~3), 당송팔대가문초 소식(1~5), 당송팔대가문초 소철(1~3), 당송팔대가문초 왕안석(1~2), 당송팔대가문초 유종원(1~2), 당송팔대가문초 한유(1), 당시삼백수(1~3), 대학연의(1), 대학장구, 동래박의(1), 동몽선습, 맹자집주, 명심보감, 사마법직해, 사자소학, 삼략직해, 상서정의(1), 서경집전(상~하), 소학집주, 손무자직해, 손자수, 순자집해(1), 시경집전(상~하), 심경부주, 안씨가훈, 예기정의, 예기집설대전(1), 오자직해, 울료자직해, 육도직해, 이위공문대직해, 자치통감강목(1~5), 장자(1~4), 전국책(1~2), 정관정요집론(1), 주역전의(상~하), 주역정의(1), 주해천자문, 중용장구, 추구, 춘추좌씨전(1~8), 통감절요(1~9), 효경대의
2018년	연산군일기, 중종실록 28년까지, 노자도덕경주, 당송팔대가문초 소순, 당송팔대가문초 증공, 설원(1~2), 안씨가훈(2), 양자법언, 자치통감강목(6), 정관정요집론(2~4)
2019년	중종실록 29년~39년, 인조실록, 효종실록, 현종실록 7년까지, 대학연의(2~4), 동래박의(2), 모시정의(1~2), 상서정의(2~3), 순자집해(2~3), 자치통감강목(7~10), 주역정의(2~3), 한비자집해(1), 효경주소(1), 고문서 523개
2020년	현종실록 7년~15년, 현종개수실록, 명종실록, 순조실록, 선조실록 10년까지, 공자가어(1~2), 당송팔대가문초 한유(2~3), 당송팔대가문초 구양수(4~5), 동래박의(3), 묵자간고(1~2), 자치통감강목(11~13), 상서정의(4~5), 순자집해(4~5), 한비자집해(2~3), 고문서 515개
2021년	인종실록, 선조실록 11년~41년, 선조수정실록, 광해군일기 3년까지(정초본), 현종실록, 철종실록, 당송팔대가문초구양수(6), 대학연의(5), 사정전훈의자치통감강목(14, 17, 18), 상서정의(6~7), 순자집해(6), 정경목민심감(1), 주역정의(4)
2022년	문종실록, 단종실록, 세조실록, 예종실록, 성종실록, 당육선공주의(1),당육선공주의(2),동래박의(4),사정전훈의자치통감강목(15),사정전훈의자치통감강목(19), 사정전훈의자치통감강목(20), 역대군감(1), 예기집설대전(2), 한비자집해(4)

주목할 점은 전통문화연구회에서 구축한 한한 병렬 말뭉치는 일반적인 말뭉치와 달리 구, 절, 문장 단위의 3세트의 말뭉치로 구성이 되었다는 것이다. 대상 자료가 부족한 상황에서 학습용 말뭉치를 최대한 효율적으로 활용하기 위한 고안한 방법이다.

이를 통해서 2022년도까지 최종적으로 구축한 한한 병렬 말뭉치의 수는 아래와 같다.

<표 2> 한한 병렬 말뭉치 구축 현황

구분	문장수	비고
동양 고전	421,845	전통문화연구회에서 번역한 번역서
조선왕조실록	1,307,578	국사편찬위원회의 원문과 고전 번역원의 번역문
고문서	19,517	한국학중앙연구원에서 제공하는 고문서 원문, 번역문
전체	1,748,940	

3. 한한 병렬 말뭉치 구축 내용

1) 전처리

AI로 학습을 시키기 위해서는 원문과 번역문이 쌍을 이루는 병렬 말뭉치가 있어야 한다. 전통문화연구회에서 병렬 말뭉치를 구축하면서 학습 효과를 높이기 위해서 구축 단계에서 미리 여러 가지 전처리를 하였다.

전통문화연구회에서 간행한 번역서에는 번역문에 한자가 그대로 노출이 되어 있고, 원문에는 토를 달았다. 간혹 원문에서 讀音이 특수한 글자나 僻字의 音은 한글음을 넣어 주었다. 원문과 번역문에는 '●, ·, , ·, ≪, 《, 「, ' 처럼 비슷한 역할을 하는 다양한 부호가 사용되었다. 한자도 중국의 간체자[4]와 한중일 호환영역의 한자가 사용되었다.

번역문에는 각주와 間註, 원주의 번역문 등이 부가되어 있고 이것들은 모두 딥러닝 학습에 부정적인 영향을 미친다.

■ 栗谷 李珥가 42세 때(1577, 宣祖 10)
■ 2년 신미(B.C.710)
■ 맹자께서는 인간의 본성이 착하다는 것을 말씀하셨고 말씀하실 때마다 요임금과 순임금을 칭술하셨다.〔滕文公爲世子 將之楚 過宋而見孟子 孟子道性善 言必稱堯舜〕"
■ 舊本엔 題漢孔安國傳이라 其書 至晉에 豫章內史梅賾(색)이 始奏於朝[1]하여 唐貞觀十六年에 孔穎達等이 爲之疏하고 永徽四年에 長孫無忌等이 又加刊定하다
■ 孔安國의 〈書序〉가 ≪昭明文選≫에 기록되자, 세상 사람들은 모두 독실하게 믿었으나, 오직 南宋의 朱熹만은 문체가 西漢 文字 같지 않다고 하여 후세 사람의 假託임을 의심하였다.

4) 아래한글의 신명조 폰트는 중국의 간체자를 번체자로 제작하였기 때문에 화면상에서는 구분할 수가 없다.

전통문화연구회는 애초부터 AI 기계 번역을 염두에 두고 말뭉치를 구축했기 때문에 위에서 언급한 학습 효과를 떨어뜨리는 여러 가지 요소를 미리 제거했다. 우선 같은 의미를 지닌 문장부호를 통일시켰으며, '豈, 更' 등의 한중일 한자 호환영역의 한자를 모두 통일시켜 주었다.

원문과 번역문에 포함된 原註와 原註의 번역문은 모두 제거하였고, 독자의 이해를 돕기 위해 추가한 간주나 보충역, 본문 중에 포함된 각주를 모두 제거하여 문맥이 단절되지 않도록 하였다.

2) 원문과 번역문의 불일치 표시

한한 병렬 말뭉치를 구축할 때 원문에 사용된 문자나 어휘가 번역문에 나타나 있지 않은 경우에 해당 부분을 [-]로 표시해 주었다.

	예시	
1	且近日被虜者, 若聞刷還人不免於死,	또 근일에 포로된 자들이 〔-그들이〕 만일 쇄환된 자가 죽음을 면하지 못하였다는 말을 들으면,
2	情蹤窮蹙,	〔-신의〕 마음과 자취가 위축(萎縮)되어
3	重起於皇朝宋濂, 解縉等諸儒,	명나라 송렴(宋濂)·해진(解縉) 등 여러 유신들에게서 거듭 일어났으며,
4	牢約下鄕,	굳게 약속한 다음 향리로 내려가고,
5	可也."	옳겠다."

위의 예시에서 '그들이', '신의' 등은 문맥을 정확하게 전달하기 위해서 역자가 주어를 추가한 것이다. 이 경우에 원문과 번역문이 일대일 대응이 되지 않기 때문에 해당 부분을[-]로 표시해 주였다.

3번을 보면 원문의 '皇朝'와 번역문의 '명나라'는 완전히 일대일 대응이 될 수 없지만 같은 행 내의 정보(송렴, 해진은 명나라 사람이다.)로 의미가 분명해질 시에는 의역이나 보충역 또는 사전에 없을지라도 일대일 대응으로 보았다.

4번은 전후 문맥에 따라 자연스럽게 의미가 생겨나는 예이다. '다음'은 문맥에 따라 '약속하고, 약속을 한 뒤에, 약속을 하고서'처럼 다양하게 번역이 될 수 있고

앞뒤 문맥에 따라 자연스럽게 추가되는 부분이다. 이런 경우에는 별도로 표시하지 않았다.

3) 원문 및 번역문의 오류 수정

원문 오자로 인해 수정이 필요하다면 해당 부분을 수정한 후 그 부분에 #으로 표시해 주었다. 번역문도 이처럼 한다. 이 부분은 학습 단계에서는 일률적으로 제거한다.

위와 같이 오자, 오타, 띄어쓰기의 명백한 오류(ex 세 종대왕, 승정원일 기) 사례를 제외하고, 단순 띄어쓰기나 맞춤법, 표준국어대사전에 등재된 옛말은 수정하지 않았다.

수정 전		수정 후
命只罷.	지 파직(罷職)시키도록 하였다.	단지# 파직(罷職)시키도록 하였다.
法當處絞,	법에 비추어 마따히 교형(絞刑)에 처해야 하옵니다마는,	법에 비추어 마땅히# 교형(絞刑)에 처해야 하옵니다마는,

수정 전	수정 후	
臣亦與〔-子〕焉,	臣亦與予焉,	신도 이 일에 참여하였으므로
郭王■有言	郭璞有言	곽박이 말하기를
不足以卽地刺擧.	明不足以卽地刺擧.	현지에서 조사하여 조처하지 못하였습니다.

4. 기계 번역의 성능에 영향을 미치는 요소

기계 번역의 품질에 영향을 미치는 요소는 AI 프레임워크와 딥러닝 모델, 데이터의 품질, 학습용 데이터의 운영, 하이퍼파라미터로 구분할 수 있다. 데이터의 중심으로 학습에 영향을 미치는 요소에 대한 연구는 ≪한문고전 기계 번역의 이해≫[5]에 잘 정리되어 있으므로 필자는 기계 번역에 영향을 미치는 요소를 아래의 내용을 중심으로 살펴보았다.

5) 배은한 외, ≪한문고전 기계 번역의 이해≫, 문예원, 2022.

■ 학습에 사용할 딥러닝 프레임워크(텐서플로우, 파이토치)
■ 번역 모델(OpenNMT, AI 바우처 사업 결과물)
■ 학습 데이터양의 증가
■ 학습에 영향을 미치는 데이터의 요소(문장 부호, 현토, 한자, 의역)
■ 한적 자료에 적합한 토크나이져
■ 주요 하이퍼파라미터(에포크, 배치 사이즈, vocab 사이즈 등)

1) AI **프레임워크**

딥러닝에 있어 프레임워크는 절대적인 역할을 한다. 구글(Google)이나 메타(Meta)에서 공개한 텐서플로(Tensorflow)와 파이토치(Pytorch)가 없었더라면 현재와 같이 인공지능이 활성화되지 않았을 것이다. 현재까지 공개된 대표적인 AI 프레임워크는 Tensorflow와 Pytorch가 있다. OpenAI의 챗GPT나 메타의 LAMA 등에서 도메인 파인튜닝을 하는 방법이 최근 연구되고 있으나 한문 기계 번역에 어느 정도 성능을 보여줄 지는 아직 확인하기 어렵다.

현재 Open 소스로 공개된 번역 라이브러리인 OpenNMT도 Tensorflow와 Pytorch 기반의 두 개의 버전을 공개하고 있다.

2) **데이터 구조**

학습에 활용한 데이터 셋에는 학습에 부정적인 영향을 미치거나 학습을 할 수 없게 하는 요소들이 많이 있다. 우선 병렬 말뭉치에서 원문과 번역문이 짝이 안 맞거나 원문과 번역문의 길이의 비율이 비정상적으로 차이가 나는 경우 본문 중에 탭이 포함된 경우 등이 특히 그러하다.

대량의 데이터를 학습하면서 가장 빈번하게 부딪치는 문제는 본문 중에 탭과 하나 이상의 개행이 포함되는 문제이다. 정제 과정을 거치다 보면 원문과 번역문 중에 탭과 여러 개의 개행이 삽입되고 학습 단계에서 에러를 발생시킨다. 이런 것들을 전처리 단계에서 모두 제거해 주어야 한다.

원문과 번역문 쌍이 중복되는 경우도 있다. 말뭉치 구축 단계에서 문장부호나

한중일 호환영역의 한자를 모두 통일시켰다면 이 단계에서는 병렬 말뭉치의 중복이나 기계적으로 의역 가능성이 큰 문장을 걸러 주어야 한다.

3) 데이터양

데이터양이 학습에 절대적인 영향을 미치는 것은 의심할 여지가 없다. 그렇다고 해서 데이터를 무한적으로 확보할 수는 없다. 학습용 데이터의 증가에 따른 학습 성능을 분석할 필요가 있다.

전체 데이터 190만 개의 문장에서 10만 개를 테스트용 문장으로 별도로 뽑아 놓은 후에 나머지 각각 30만, 60만, 90만, 120만, 150만, 180만 개의 문장을 6개의 그룹으로 나누어서 양의 증가에 따른 기계 번역의 품질 향상을 확인할 필요가 있다.

4) 데이터의 일관성과 다양성

보통 글을 쓰는데 같은 표현의 반복을 피하라고 배웠다. 같은 단어나 표현이 반복되면 읽기 불편하고 지루해지기 때문이다. 그러나 인간의 언어생활과 달리 AI를 학습시키는 데 있어서 다양성과 일관성이라는 관점으로 접근하면 절대적으로 일관성이 중요하다.

앤드류 응(Andrew Ng) 스탠포드 대학 교수는 “좋은 데이터를 수집하고 가공하는 것이 인공지능을 만드는 과정의 80%를 차지하는데, 이는 데이터가 인공지능의 핵심적인 부분임을 의미한다.”라고 말했다. 그는 데이터 중심 인공지능(Data-centric AI) 개발의 중요성을 강조했고, 인공지능 개발자들이 코드 수정을 통한 모델 하이퍼파라미터 변경에 매몰되지 않고 좋은 데이터를 확보하고 유지하려고 노력해야 한다고 덧붙였다. 그리고 데이터로 성능 개선을 이루어 냈는지 실제 사례를 통해 보여주었다.

앤드류 응은 하나의 음성을 전사할 때 ‘Um, today's weather, Um... today's weather, Today's weather’에서와 같이 ‘음(Um)..’이라는 부분을 세 가지 정도로 라벨링 할 수 있는데 어떤 것이 옳을까요 라는 질문에 “셋 다 괜찮지만 모든 데이터가 일관적(consistent)이여야 한다"라고 답변하였다.

기계 학습에서 일관성 확보는 아무리 강조해도 지나치지 않다.

(1) 표현의 다양성

과거에 '君子曰'는 '군자가 말하기를, 군자가 논평하기를' 등으로 번역하던 것이 최근에는 대부분 번역 문장을 단문으로 끊어서 번역한다. 그 결과 '君子曰'이라는 단문은 아래와 같이 다양한 번역이 나타나게 되었다. 이러한 요소를 어떻게 해결해야 할지 고민해야 한다. 보통 원문과 번역문 쌍이 중복되는 것을 제거하지만 원문과 번역문 쌍이 다를 때는 제거하지 않는다. 하나의 문장이 다양하게 번역되는 사례는 딥러닝 학습 단계에서 검토해 볼 필요가 있다.

원문	번역문	빈도
君子曰	이에 대해 군자(君子)는 다음과 같이 논평(論評)하였다.	2
君子曰	이에 대해 군자(君子)는 다음과 같이 논평하였다.	2
君子曰	군자(君子)가 다음과 같이 논평하였다.	1
君子曰	군자(君子)가 이에 대해 다음과 같이 논평(論評)하였다.	1
君子曰	군자(君子)는 이에 대해 다음과 같이 논평(論評)하였다.	3
君子曰	군자(君子)는 이에 대해 다음과 같이 논평하였다.	1
君子曰	군자(君子)는 논평(論評)하였다.	1
君子曰	군자(君子)는 다음과 같이 논평(論評)하였다.	2
君子曰	군자(君子)는 이에 대해 다음과 같이 논평(論評)하였다.	19
君子曰	군자(君子)는 이에 대해 다음과 같이 논평하였다.	1
君子曰	군자(君子)들은 말하기를	1
君子曰	군자가 이에 대해 다음과 같이 논평하였다.	1
君子曰	군자가 말하였다.	1
君子曰	군자가 이에 대해 다음과 같이 논평하였다.	1
君子曰	군자는 다음과 같이 말한다.	1
君子曰	그러면 군자는 말할 것이다.	1
君子曰	이 일에 대해 군자(君子)는 다음과 같이 논평(論評)하였다.	1
君子曰	이에 대해 군자(君子)가 다음과 같이 논평(論評)하였다.	1
君子曰	이에 대해 군자(君子)가 다음과 같이 논평하였다.	1
君子曰	이에 대해 군자(君子)는 다음과 같이 논평(論評)하였다.	7
君子曰	이에 대해 군자(君子)는 다음과 같이 논평하였다.	4
君子曰	이에 대해 군자(君子)는 다음과 같이 평론(評論)하였다.	1
君子曰	이에 대해 군자(君子)는 다음과 같이 평론하였다.	1
君子曰	이에 대해 군자는 다음과 같이 논평(論評)하였다.	2
君子曰	이에 대해 군자는 다음과 같이 논평하였다.	5
합계		62

(2) 漢字 倂記, 懸吐

전통문화연구회에서 번역한 번역서에는 한자를 그대로 노출을 시킨다. 반면에 ≪국역 조선왕조실록≫은 한자를 병기하고 있다. 그렇다고 모든 한자어에 대해서 일률적으로 한자로만 표기한다거나 병기하지는 않고 있다. 같은 번역서에서도 모든 한자어에 대해 일률적으로 한자[6]로만 표기하지 않았다. 한자의 표기나 병기는 자의적이기 때문에 잘못되었다고 할 수도 없는 것이다. 그렇지만 일관성이 없이 사용된 한자 병기나 한자어는 분명하게 딥러닝 학습에 부정적인 영향을 미친는 것은 어쩔 수가 없다.

한자 병기를 제거하고 학습을 시키면 그 결과도 당연히 한자 병기가 되지 않을 것이다. 그렇기 때문에 번역이 끝난 다음에 이를 다시 복원해 주어야 한다. 즉 한자 병기를 해 주어야 한다.

후처리는 번역된 TU를 이용하여 사용자가 선택한 단위에 따라 한자를 병기해 줄 수 있다. 한자를 병기하는 기준이 되는 단위는 책 전체, 페이지, 기사(국역서의 경우 번역된 단락), 문장이 있을 수 있다. 이때 출현하는 위치, 난이도, 의미의 모호성 등을 고려할 수 있다. 출현 위치는 정해진 단위에서 맨 처음에 나오는 어휘에만 한자를 병기해 주는 것이고, 난이도는 정해진 난이도에 따라 일률적으로 병기해 주는 것이다. 의미의 모호성에 따라 병기해 주는 것은 동음이의어 사전을 이용하여 본문 중의 의미를 구분해 주어야 한다고 판단 되는 어휘를 한자를 병기해 주는 것으로 해결할 수 있다. 여러 가지 방법을 복합적으로 적용할 수도 있다.

옵션에 의해 모든 한자 어휘를 한글음 없이 한자를 그대로 노출시킬 수도 있다. 한자 병기는 AI 기술을 사용하면 어렵지 않게 구현할 수 있을 것이다.

(3) 意譯, 補充譯, 間註

直譯은 외국어로 된 말이나 글을 단어 하나하나의 의미에 충실하게 번역하는 것이다. 말 그대로 원문의 글자 내지는 구절의 본래 의미를 하나하나 충실하게 옮

6) 대학연의권4에서 '王翦을, 왕전이, 왕전이' 등으로 표기되었다.

기는 것이다. 반면에 意譯은 원문의 단어나 구절에 지나치게 얽매이지 않고 전체의 뜻을 살려 번역하는 것 또는 그런 번역을 말한다. 直譯과 意譯은 좋고 나쁘고의 문제가 아니라 단지 각각의 쓰임새가 다른 번역의 한 종류일 뿐이다.

補充譯은 역자가 의미 전달을 명확하게 하기 위해 원문에 없는 내용을 추가해 넣은 것으로 독자의 이해를 돕는다는 측면에서 바람직한 번역이다. 전통문화연구회는 보충역을 구분하기 위해 < >기호를 사용하여 표시하였다.

間註는 괄호 안에 짧게 부연 설명을 하는 것으로 주석의 분량이 짧을 때 주로 사용한다.

그런데 인공 지능에서는 意譯과 補充譯, 間註는 모두 학습에 부정적인 영향을 미친다. 일관성의 원칙에 위배되기 쉽기 때문이다. 이 문제를 해결하기 위해 전통문화연구회에서는 말뭉치를 구축할 때 원문에 사용된 문자가 번역문에 나타나는지 확인하여 번역문에 나와 있지 않으면 원문에 해당 부분에 [-]로 표시해 주었다. 반대로 원문에 없는 내용이 번역문에 추가되었으면 그 부분을 [-]로 표시해 주었다.

원문과 번역문에 [-]이 없는 문장은 원문에 사용된 문자가 번역문에 그대로 나타나 있는 것이다. 즉 의역이 전혀 포함되어 있지 않은 문장이다. 이런 문장이 모두 92만 개나 구축이 되었다. 이 92만 개의 문장만 선택해서 학습시킬 수도 있고, 전체 문장에서 [-] 부호를 제거하고 학습시킬 수 있다. 이때에는 본문 중에 사용된 [ ] 부호 적지 않으므로 정규식을 사용해서 [-] 만을 제거해야 한다.

(4) 分野, 文體

학습용 데이터를 정제하는 것이 기계 번역의 품질을 높이는 지름길이다. 그렇지만 학습용 말뭉치를 도메인별로 나누어서 학습을 시키는 방법도 고려해 보아야 한다. 전통문화연구회에서 번역한 도서는 선진 시대의 문헌으로부터 당송 이후의 문헌까지 다양하며 문체별로는 시경 등의 운문과 제자백가와 같은 산문 또한 존재한다. 국가별로 우리나라의 문헌과 중국의 문헌으로 구분할 수도 있다.

≪承政院日記≫나 ≪朝鮮王朝實錄≫과 같은 단일한 아이템을 대상으로 학습을

시킬 때는 작성된 시기가 큰 문제가 되지 않지만 한국의 고전과 중국 고전을 모두 포괄하는 AI 모델을 생성할 때 이 문제를 짚어 보아야 한다.

이러한 문체의 다양성은 학습의 성능을 떨어뜨리므로 시대별, 문체 또는 장르별로 별도의 학습용 말뭉치 셋을 만들어 번역 품질을 확인할 필요가 있다. 학습 모델을 역사, 시 등의 운문, 문집 등 분야 별로 만들어 학습에 활용할 수도 있을 것이다. 실제 번역 단계에서는 분야를 미리 분석하여 해당 분야의 학습 모델을 적용해서 번역을 한다.

그러나 문체별로 분류하면 자칫 학습에 활용할 말뭉치의 수가 부족해질 수 있으므로 말뭉치가 충분히 축적이 된 후에 적용해야 한다.

(5) 注釋의 종류

經典에는 다양한 注釋이 존재한다. 아래는 ≪詩經集傳≫의 일부이다. 맨 위에 제목이 있고 그 아래 경전 원문이 있고 그 아래에 경문을 주석한 傳이 있다. 經文은 완전 운문이라면 傳은 산문에 가깝다. 딥러닝으로 학습을 할 때 이런 부분은 학습에 당연히 고려해야 할 요소이다.

〈關雎〉
關關雎鳩ㅣ 在河之洲ㅣ로다
窈窕淑女ㅣ 君子好逑ㅣ로다
傳
興也라 關關은 雌雄相應之和聲也라 雎鳩는 水鳥니 一名王雎라 狀類鳧鷖하니 -- 이하 생략
○ 興者는 先言他物하여 以引起所詠之詞也라 周之文王이 生有聖德하고 又-- 이하 생략

다행스럽게도 전통문화연구회는 한한 병렬 말뭉치를 구축하면서 학습용 말뭉치의 활용성을 높이기 위해 경문과 주석을 종류별로 구분해 주었다. 이를 활용하면 좀더 효율적인 학습이 가능할 것이다.

5) 셔플링(shuffling)

병렬 말뭉치를 딥러닝으로 학습할 때 셔플(shuffling)은 성능에 큰 영향을 미친다. 셔플은 데이터를 무작위로 선택하거나 처리하는 것은 말한다. 데이터셋을 섞지 않

고 바로 분할한다면 train data와 test data의 분포가 너무 다를 수가 있다.

셔플은 두 가지 방향으로 진행이 된다. 먼저 모델을 학습시킬 때 데이터셋을 셔플한다. 데이터셋을 트레이닝 데이터(training data)와 밸리데이션 데이터(validation data), 테스트 데이터(test data)로 나눌 때도 셔플하고, 학습을 시킬 때는 매 에폭(epoch)마다 training data를 shuffle하기도 한다. AI 개발 실무자는 셔플링에 따라 성능이 큰 차이를 보인다 하여 셔플링의 중요성을 여러 번 강조하였다.

6) 분절(Tokenization)

AI에 학습을 시키기 위해서는 텍스트를 그대로 사용할 수 없기 때문에 적당한 단위로 나누어 주어야 한다. 이렇게 텍스트를 나누는 단위를 토큰(token)이라고 한다. 이 토큰의 단위는 단어, 문자, n-gram 등이 있다. 그리고 텍스트를 토큰(token)으로 나누는 것을 토큰화(tokenization)라고 하고 토큰화하는 것을 토크나이징(Tokenizing)이라고 한다. 즉 토그나이징은 텍스트를 자연어 처리를 위한 모델에 적용할 수 있게 '언어적인 특성을 반영해 단어를 수치화'하는 것을 말한다.

토큰의 단위는 자모 단위, 글자 단위, 어절 단위, 형태소 단위, 서브워드(subword) 단위 등 다양한 방법으로 토큰을 할 수 있다.

7) 토크나이저(Tokenizer)의 선택

딥러닝 학습에 어떤 토크나이저를 사용하느냐에 따라 성능에 큰 영향을 미친다. 기계가 훈련 단계에서 학습한 단어들을 모아놓은 것을 단어 집합(vocabulary)이라고 한다.

이 단어 집합은 토크나이징 결과로 생성이 된다. Vocab size를 보통 32,000개에서 20만 개 정도 추출을 하는데 원문과 번역문에 대해서 각각 단어 집합을 생성해야 한다.

단어 집합를 생성하는 기본 단위인 토큰(token)은 자모, 문자, 형태소 등이 사용되었으나 최근에는 서브워드(subword) 방식을 주로 사용한다. 그런데 테스트 단계에서 기계가 미처 배우지 못한 단어가 등장할 수 있다. 이를 단어 집합에 없는 단

어란 의미로 OOV(Out-Of-Vocabulary)[7]라고 한다. OOV 문제를 해결하기 위해서 등장한 방법이 Subword 분절 방식이다. 서브워드 방식엔 BPE, SentencePiece, WordPiece 등 작은 변형들이 존재한다. 서브워드 분절은 OOV를 문제를 어느 정도 해결하였으나 완전한 것은 아니다. 단어집합의 수가 너무 많으면 OOV 해결이 어렵게 되며, 반대로 너무 적으면 모호성이 증가한다.

이들 중에 한문 기계 번역에는 Sentencepiece의 model_type unigram이 좋은 성능을 보이고 있다. Unigram tokenizer는 BPE와 같이 문장 혹은 단어 안에 있는 글자들을 적절한 단위로 나누는 subword tokenizer의 하나로, 큰 vocabulary로부터 필요하지 않은 토큰을 하나씩 줄여나감으로써 최종적인 token들을 도출하는 알고리즘이다.

8) **하이퍼 파라미터**(Hyper parameter)

기계 번역에는 데이터뿐만 아니라 하이퍼 파라미터도 학습에 많은 영향을 미친다.

배치 사이즈(batch size)는 학습에 사용하는 데이터를 한 번에 몇 건씩 처리하고 가중치를 갱신할 것인지 선언, 즉 몇 문항을 풀고 업데이트할 것인가를 지정하는 것이다. 이때 정답과 비교하고 그 결과를 이용하여 가중치를 수정한다.

에포크(epochs)는 전체 데이터에 대하여 반복할 학습 횟수를 말한다. 이 때 에포크를 무한대로 늘인다고 좋은 모델이 만들어지는 것은 아니다. 보통 과적합을 방지하기 위해서는 EarlyStopping 이라는 콜백 함수를 사용하여 적절한 시점에 학습을 조기 종료시킨다.

9) **평가용 데이터 셋의 구성**

모델을 설계할 때 가장 중요한 것 중 하나는 '모델의 성능을 어떻게 평가할 것인가'이다. 자연어 처리 분야에서 객관적인 성능을 평가하는 가장 유명한 자연어 처리 벤치마크가 GLUE이다.

---

7) 또는 이를 UNK(Unknown Token)라고 표현하기도 한다.

벤치마크 데이터셋은 공통된 기준으로 인공 지능 정확도를 평가하고 경쟁할 수 있는 기반이며, 인공지능 발전에 핵심 역할을 담당하고 있다.

그러나 GLUE는 BERT가 출현한 이후로 인공 지능이 사람보다 더 뛰어난 성능을 보이면서 제 역할을 못하게 되었다. 그로 인해 2019년에 페이스북, 딥마인드, 뉴욕대, 워싱턴대가 참여하여 SuperGLUE를 새로 만들었다.

국내에는 KLUE라는 벤치마크를 구축하였으나 안타깝게도 한문을 정확하게 번역했는지 평가할 수 있는 데이터셋은 존재하지 않는다.

한문 번역 분야에서도 번역 모델의 성능을 평가할 수 있는 벤치마크 데이터 셋이 필요하다. 벤치마크 데이터 셋은 모델의 성능을 높이기 위한 용도로 사용하는 것이지 벤치마크 데이터를 학습에 포함시켜 평가가 높아지는 것을 목표로 하지 않는다.

벤치마크 데이터셋은 한문 번역의 대표성을 띄어야 한다. 시대별, 분야별, 문체별로 표준 번역을 선정해야 한다.

5. 평가

학습 결과에 대한 평가는 BLEU Score(Bilingual Evaluation Understudy Score), 로 하였다. BLEU는 기계 번역 결과와 사람이 직접 번역한 결과가 얼마나 유사한지 비교하여 번역에 대한 성능을 측정하는 방법이다. 측정 기준은 n-gram에 기반한다. 언어에 구애받지 않고 사용할 수 있으며, 계산 속도가 빠르다는 장점이 있다. 이때 BLEU Score[8]가 클수록 성능이 좋다.

기계 번역 모델의 성능을 평가하려면 평가 기준이 되는 데이터셋의 품질이 높아야 한다. 평가셋이 의역이 심하거나 오역이 많을 경우에는 평가 자체가 신뢰하기 어렵다. 그렇기 때문에 객관적인 평가를 위해서는 잘 정리된 평가 데이터셋이 필요하다.

8) 공개된 소스 multi-bleu-detok를 이용하여 평가하였다.

6. 맺음말

초거대 언어 모델이라는 용어는 OpenAI의 GPT-3가 나오면서부터 쓰이게 되었다. 이전의 딥러닝 언어 모델 학습이 대규모 데이터를 학습한 사전 학습 언어 모델(pre-training language models)에 기반하여 각 영역별 데이터를 정밀 학습(fine-tuning)하는 두 단계 언어 모델 개발 방식이라면 초거대 언어 모델은 초거대 데이터를 학습한 언어 모델 하나로 추가 정밀 훈련 없이 여러 서비스에서든 활용하자는 것이다. ChatGPT의 전후로 엄청나게 많은 데이터와 엄청나게 큰 모델을 만들어 학습을 많이 시키니까 좋은 결과가 나타난다는 것을 알게 되었다. 그 결과를 어떻게 활용할지 고민하다가 해결책으로 나온 것이 ChatGPT이다.

초거대 언어 모델을 이용하면 한문 번역의 수준도 어느 정도 개선이 될 것이다. 더 이상 AI를 이용하여 사람의 번역 수준에 근접하거나 뛰어 넘는 번역 모델을 만드는 것은 꿈이 아니다.

그런데 ChatGPT는 세계 거의 모든 언어를 수집해서 학습을 시켰으나 대부분 영어권의 데이터를 위주로 했기 때문에 영어로 문장을 생성하거나 요약하는 일은 잘 할 수 있지만 우리말이나 한문 같은 경우에 영어에 비하여 품질이 기대만큼 미치지 못하는 것이 사실이다.

초거대 언어모델에 한글 텍스트와 한문 텍스트를 대량으로 넣고 학습을 시키면 당연히 한문 번역 능력도 크게 향상이 될 것이다.

문제는 데이터를 수집하는데 많은 예산이 들고, 수집한 데이터를 학습시키는데에도 어마어마한 비용이 들어간다는 것이다. 이때부터 기계 번역은 기술적인 문제에서 경제적인 논리로 전환된다.

ChatGPT는 데이터를 한 번 학습할 때마다 어마어마한 비용이 들어간다. 한문 텍스트를 대량으로 수집해서 정제하는 것이 쉬운 작업이 아니다. 한글과 특히 규모가 크지 않은 한문 시장에서 데이터를 수집하고 많은 학습 비용을 들여서 학습을 시키기에는 경제적 부담이 너무나도 크다.

최근 도메인 파인튜닝이라는 개념이 도입되어 기존에 초거대 언어 모델을 기반

으로 특정 분야에 전문적으로 학습시키는 방법이 도입이 되었다. 이 개념은 몇 개의 문장을 학습시키는 퓨샷러닝과 달리 몇 십만 단위의 어느 정도 규모가 있는 문장을 학습시켜 해당 도메인에 최적화시키는 방법이다. ChatGPT를 학습하는것에 비해 훨씬 저렴한 비용으로 비슷한 성능을 내는 것으로 알려졌다.

전통문화연구회에서 구축한 데이터는 트랜스포머 기반으로 기계 번역기에 학습할 때 활용할 수 있으며, 초거대 언어 모델의 성능이 향상이 되면 그 바탕에서 도메인 파인 튜닝 데이터로 활용할 수 있을 것이다. 어떤 경우이든 전통문화연구회에서 구축한 한한 병렬 말뭉치는 그 가치는 잃지 않을 것이다.

마지막으로 AI를 학습시키는 데에는 많은 시간이 소요된다. 또한 시행 착오의 반복이다. 돈으로 시계를 살 수 있어도 시간을 살 수 없다는 말이 있지만 딥러인 학습에서는 돈이면 시간을 살 수 있다. AI를 학습시키는 데에는 고사양의 GPU 장비가 필요하고 많은 시간을 쏟아 부어야 한다. 보통 엔비디아 GPU3080을 이용하여 학습할 때 100만 개의 문장을 학습시키는데 평균 8시간 이상의 시간이 소요되었다. 그런데 이런 작업을 끊임없이 반복하면서 최적의 조합을 찾아내야 한다.

시간의 제약으로 기계 번역의 성능을 충분히 테스트하지 못한 것이 못내 아쉽다. 다만 한문 기계 번역에 필요한 여러 가지 요소를 집어 보는데 만족해야 했다.

‘한문고전 자동번역’ 개발의 의미와 과제

金 愚 政
(단국대 한문교육과 교수)

<논문요약>

이 글은 현재 국내에서 개발된 유일한 한문고전 기계번역기인 ‘한문고전 자동번역’의 의의와 개발 과정 및 결과를 통해 얻게 된 과제를 점검해 보고자 하는 목적에서 작성되었다. 개발에 착수했을 당시, 전문역자들의 번역에 큰 도움을 줄 것이라는 기대와 아울러 전문역자들의 일자리를 위협할 것이라는 우려마저 있었지만, 막상 예상에 못 미치는 성능으로 활용 가치를 고민하는 상황에 놓여있다. 그러나 이는 기계번역에 대한 지나친 기대와 이해 부족이 만들어낸 평가다. ‘한문고전 자동번역’은 기계번역 분야에서 난제로 꼽히는 고전 문언문 기계번역에 도전하여 실제적 성과를 낸 사업으로 기억되어야 마땅하다. 번역 모델 구축이나 평가 방법 등에서 다소 아쉬운 점이 있었고, ≪승정원일기≫라는 특정 유형의 도메인에서만 성능을 발휘하는 한계를 지녔지만, 양질의 병렬 코퍼스가 부족한 상황에서 성과를 낸 점은 인정해야 할 것이다. 이와 같은 진단을 바탕으로 본고에서는 도메인 다양화, 클라우드 기반 번역 플랫폼 구축, 데이터 구축 및 정제 고도화, 사전

정보 및 부가 시스템 개발, 평가 방법 개선 등을 제안하였다.

키워드 : 한문고전, 기계번역, 번역 모델, 데이터, 코퍼스, 자동평가, 수동평가

1. 들어가는 말

이 글은 현재 국내에서 개발된 유일한 한문고전 기계번역기인 '한문고전 자동번역'의 의의와 개발 과정 및 결과를 통해 얻게 된 과제를 점검해 보고자 하는 목적에서 작성되었다.

한국고전번역원에서는 2017년 3월부터 2019년 12월까지 인공지능 기반 고전문헌 자동번역시스템 개발을 진행하였으며, 2021년 1월에 그 성과를 '승정원일기 자동번역기'[1]라는 이름으로 공개하였다. 그런데 개발에 착수했을 당시, 알파고가 보여준 충격적인 성능과 신경망 기계번역에 대한 이해 부족이 겹치며 이 번역기가 전문역자들의 번역에 큰 도움을 줄 것이라는 기대는 물론, 역자들의 일자리를 위협할 것이라고 우려하는 이들도 있었으며, 사업 주체의 적절성 문제라든지 성공 여부를 가늠하기 어렵다는 점 등을 이유로 부정적인 반응을 보인 이들도 있었다. 이런 세간의 관심 속에 진행된 사업이 결과적으로 기대에 못 미치는 성능을 보이며, 지금은 비록 서비스는 제공되고 있지만 고전번역에 소요되는 시간을 획기적으로 줄일 것이라던 약속이 실현되려면 아직 많은 인내와 노력이 필요해 보인다.

그렇다면 '한문고전 자동번역'은 실패한 사업인가? 번역 결과만 두고 평가한다면 그렇다고 할 수 있을지 모르지만 신경망 기계번역의 원리를 아는 사람이라면 동의할 수 없을 것이다. 고작 100만 어절 남짓한 학습 코퍼스로 이루어지는 한문고전 기계번역이 매일 1,400억 단어 이상을 처리하는 구글 번역기에 필적하는 성능을 발휘할 수 있을 것이라는 믿음과 이러한 믿음을 실현하는 데 그리 많은 시간이 소요되지 않을 것이라는 속단만 제외하면 '한문고전 자동번역'은 기계번역 분야에서 난제로 꼽히는 고전 문언문 기계번역에 도전하여 실제적 성과를 냈다고 평가할 수 있는 사업이었다. 다만 사업 착수 시점부터 상기한 바와 같은 막연한

1) 2021년 1월 최초 공개 당시의 명칭으로, 현재는 '한문고전 자동번역'으로 개칭하였다.

믿음과 속단에 힘을 보태는 발언들이 많았던 데다 사업 진행 중에 이루어진 성능 평가에서도 상당히 양호한 성능을 보이고 있는 것으로 보고된 점이 오히려 실망감을 더 키우는 데 큰 역할을 했다고 본다.

본고에서는 ‘한문고전 자동번역’의 성능에 대한 논란이 일게 된 여러 원인 중 하나인 평가 방법 및 결과를 중심으로 ‘한문고전 자동번역’ 사업의 성과와 의의를 다시금 살펴보고, 더 나은 결과를 얻기 위해 힘써야 할 점은 무엇인지도 아울러 살펴보고자 한다.

2. ‘한문고전 자동번역’ 개발과 그 의미

‘한문고전 자동번역’은 2017년 한국고전번역원이 충남대학교와 공동으로 ICT기반 공공서비스 촉진사업의 하나로 수주한 ‘인공지능 기반 고전문헌 자동번역시스템 구축사업’의 결과물이다. 이 사업은 총 5년에 걸쳐 진행되었지만 2020년 이후는 대국민서비스를 목표로 한 것이기에 번역기 개발만으로 한정하면 약 3년이 소요되었다. ≪승정원일기≫ 자동번역 모델을 개발하였고, 번역자를 위한 번역공정 관리시스템, 대국민 서비스를 위한 ‘한문고전 자동번역 서비스’ 사이트를 개발하였다. 또한 번역 학습에 사용하기 위해 129만 쌍의 병렬 코퍼스를 구축하였으며, 기계번역 모델도 개발하였다.[2] 이 모델은 우선 다양한 조합의 병렬 코퍼스(parallel corpus)를 원문과 번역문 기준으로 전처리하고, 토큰을 어떤 단위로 분석할지에 대한 검증 작업을 완료한 다음, 병렬 코퍼스를 데이터 기반으로 샘플링하여 인공신경망에 입력한 후 번역 데이터를 공급하는 방식으로 이루어져 있다.

한문고전의 경우, 시기・지역・문체에 따라 현격히 다른 특징을 띠고 있으며, 번역문(reference)이 비교적 풍부한 텍스트가 있는가 하면 번역본이 아예 없는 경우도 흔해 양질의 병렬 코퍼스를 구축하기가 어렵다. 다수의 번역문이 존재하는 경우에도 어떤 번역문을 코퍼스 대상 자료로 선정할 것인가에 따라 번역 품질에 영향을 주게 되며, 특정 도메인의 자료만을 많이 학습할 경우 과적합(overfitting) 현상

2) 한국고전번역원(2020), 17쪽.

이 발생하기도 한다. 또한 원시 코퍼스(raw corpus)를 그대로 이용하면 좋은 결과를 기대할 수 없기 때문에 대부분 전처리를 하게 되는데, 어떤 방식으로 하는가에 따라 번역 품질에 큰 영향을 끼치게 된다. 예를 들어, 기계번역의 품질을 단기간에 향상시키기 위해서는 학습에 사용되는 문장을 가급적 짧게 처리하는 것이 유리하나, 장문 번역에 있어서는 불리할 수 있다. 따라서 번역 품질을 전반적으로 향상시키기 위해서는 문장을 어떻게 분절하여야 할 것인지 결정해야 하는데, 시간과 비용 문제도 결부되어 있기 때문에 판단을 내리기 쉽지 않다. 이러한 이유로 한국고전번역원에서는 단일한 번역물이 있고, 문체가 비교적 일관되며, 다량의 원시 코퍼스를 확보할 수 있는 ≪승정원일기≫를 기계번역 대상 텍스트로 선정한 후, 코퍼스 제작 규칙을 자체적으로 정하고 학습용 병렬 코퍼스를 제작하여 번역모델에 투입하였는데, 제반 여건을 고려하면 합리적인 선택이었다고 할 수 있다. 따라서 '한문고전 자동번역'은 ≪승정원일기≫라는 특정 도메인에 특화된 전문 번역기라는 점을 기억할 필요가 있다. 지금까지 개발된 신경망 기계번역이 그렇듯이 이 번역기도 다른 유형의 한문고전 자료에 대해서는 양질의 번역 결과를 내기 어렵다는 뜻이다.

기술적인 면에서 '한문고전 자동번역'은 기본 도메인(target domain)인 ≪승정원일기≫ 외에도 ≪실록≫과 천문고전 등의 부가 도메인(out-of-domain)과 합성 데이터를 사용하여 학습 효율을 높이고자 하였다.[3] 각 학습 단계별로 학습되는 기본 도메인과 부가 도메인, 합성 데이터의 비율을 조정해 가면서 학습이 잘 될 수 있도록 각 학습 단계별 데이터를 조정하였는데, 발표된 바에 따르면 각 데이터의 비율은 번역모델마다 다르지만 대체로 in-domain 75%, out-of-domain 9%, 합성 데이터 16%의 비율로 사용되었다. 또한 기본 도메인의 역방향 번역모델(back-translation)을 구축하여 병렬 코퍼스를 확장하였으며, 하이퍼파라미터(hyperparameter)[4] 튜닝을 비롯해 기계 학습에서 사용되는 여러 가지 기법을 적용하였다.

---

3) 부가 도메인은 기본 도메인의 데이터만으로는 충분치 않을 때 사용되는 데이터이며, 합성 데이터는 다수의 기본 도메인을 학습함에 있어서 특정 번역에 치우치는 현상을 막고 번역의 유연성을 확보하기 위해 추가되는 데이터다.

4) 최적의 훈련 모델을 구현하기 위해 모델에 설정하는 변수로 학습률(Learning Rate), epoch 수(훈련 반복 횟수), 가중치 초기화 등을 포함한다.

이처럼 ‘한문고전 자동번역’은 번역 모델의 성능을 높이기 위해 다각적인 노력을 기울인 결과 기본 도메인만 학습한 번역모델보다 더 나은 성능을 이끌어냈으며, 저사양 하드웨어 환경에서도 사용 가능한 기계번역 모델링을 구성하였다.

한편 한국고전번역원은 이 번역기를 개발하는 과정에서 자동평가와 수동평가를 수차례 병행 실시하였다. 기계번역은 예상하지 못한 번역 오류를 포함하고 있으며 동일한 기계번역 모델에서도 투입된 데이터나 병렬 코퍼스 등에 따라 다양한 수준의 번역문들이 생성되는 문제도 나타나므로 품질에 대한 정확한 평가와 확인 과정을 거쳐야 한다. 기계번역의 품질에 대한 정확한 평가와 확인은 좋은 기계번역 모델을 만들게 하는 기초 작업이며, 이를 통해 기계번역 모델의 피드백을 주어 해당 기계번역 모델을 적절히 수정할 수 있기 때문이다.[5)]

자동평가는 참조번역(reference translation)과 후보번역(candidate translation)을 고려하여 기계번역의 품질을 평가한다. 자동평가는 명료성(intelligibility), 정확성(accuracy), 유창성(fluency) 등을 기준으로, 기계가 번역한 일부 텍스트의 품질을 평가하는 방법이다. 명료성은 번역된 문장이 원문 텍스트를 잘 나타내고 있는가를 나타내는 것으로 번역의 타당성과 관련된 기준이고, 정확성은 참조번역의 내용이 기계가 번역한 후보번역에 얼마만큼 잘 유지되어 있는가를 나타내는 것으로, 번역의 신뢰성과 연관된다. 유창성은 기계가 번역한 텍스트가 해당 언어로 얼마나 자연스럽게 번역되었는지를 나타내는 지표이다.

‘한문고전 자동번역’에서 채택한 자동평가 지표는 BLEU(BiLingual Evaluation Understudy) 점수다. BLEU는 주로 범용 목적의 기계번역 품질평가에 사용되며, 기계번역을 거친 후보번역과 전문가가 직접 번역한 참조번역 사이의 유사성을 기반으로 한다. BLEU는 전문가의 번역과 유사할수록 기계번역의 품질이 좋을 것이라는 점을 전제로 하여, 전문가의 번역에 대한 유사성(closeness)을 평가할 정밀도(precision)와 전문가가 번역한 높은 수준의 병렬 코퍼스를 필수적으로 요구한다.[6)]

BLEU의 알고리즘은 일치하는 단어의 수(n)를 계산하는 일종의 n-gram 방식인데, 기계가 번역한 후보번역과 번역 전문가가 미리 번역해 놓은 참조번역을 1-gram,

5) 이하 기계번역 평가방식에 관한 내용은 정성훈・하지영・김우정(2021)을 요약한 것이다.
6) Papineni et al.(2002)

2-gram, 3-gram, 4-gram 순으로 비교하여 정밀도를 계산한다. 즉 후보번역과 참조번역 간 일치하는 n-gram의 수가 많을수록 기계번역의 품질이 좋은 것으로 평가한다. BLEU 평가방식은 후보번역 속의 단어나 구 중에서 참조번역에도 사용된 단어나 구가 몇 개인지를 측정하여 후보번역의 n-gram이 얼마나 유사한지를 평가하는 방식이다. BLEU 점수는 짧은 번역에 벌점을 부여하는 장치가 있음에도 불구하고 번역문이 짧을수록 상대적으로 높은 점수를 받을 수 있는 것으로 알려져 있다. 또한 BLEU 점수는 대체로 하나의 참조번역만을 반영하는데, 유의어나 반의어(단어 관계), 어순(구의 배열순서)에 따른 적절하고 다양한 참조번역을 반영하기 어려우며, '번역투 효과'에 의해 그 평가결과가 왜곡될 수도 있다. 극단적인 경우에는 원문의 의미와 반대인 후보번역에도 높은 BLEU 점수를 부여하는 문제를 노출하기도 한다. 이와 같은 한계를 극복하기 위해 최근에는 참조번역을 여러 개 사용하여 n-gram의 다양성과 통사구조를 반영하거나 반복적으로 번역되는 동일한 통사구조를 보정하여 n-gram의 정밀도를 수정하는 방식을 사용하기도 한다.[7] 그러나 이 경우에도 참조번역이 많을수록 BLEU 점수가 높아지기는 경향이 존재하고, n-gram의 수를 짧게 하지 않는다고 해서 반드시 문법적으로 적절한 번역을 생성한다는 보장도 없기 때문에 BLEU 점수가 높아졌다고 해서 기계번역의 품질이 꼭 향상되었다고 할 수 없다. 실제 연구에서도 수동평가에서 1등을 한 번역이 BLEU 점수로는 6등에 그친 경우도 있었다.[8]

이런 문제를 해결하기 위해 학계에서는 METEOR(Metric for Ecaluation of Translation with Explicit ORdering)를 비롯해 ROUGE, LePOR, BLEUmod., WER 등 다양한 자동평가 방법이 제안되기도 하였지만, 언어적 특성이나 번역 모델의 특성에 따라 결과가 달라질 수 있으므로 특정 방식만을 고집할 것이 아니라 '한문고전 자동번역'에 적합한 평가 방법을 찾아내는 것이 중요하다.

수동평가에는 모두 8명의 내・외부 전문가가 참여한 것으로 확인된다. 수동평가를 위한 데이터셋은 구축된 코퍼스를 대표할 수 있도록 코퍼스 전반에서 추출하였으며, 훈련 데이터로 사용되지 않은 문장 중에서 번역문 길이가 300자 이내이

7) Callison-Burch et al.(2006; 2007)

8) 상게문.

며 하나의 종결 기호를 갖는 데이터를 선정하였다. 2017년 ≪승정원일기≫ 번역 모델은 54 또는 60문장으로 수동평가 데이터셋을 구성하였으며, 2018년과 2019년에는 단문 60문장과 장문 40문장으로 평가 데이터셋을 구성하였다. 2018년과 2019년에 개발된 조선왕조실록과 특수고전(천문분야)도 각각 50문장, 60문장으로 이루어진 데이터셋을 구성하였다.

수동평가는 5점 척도로 외부 전문가 평가위원(번역 전문가)에 의한 수동평가 방식으로 실시되었다. 평가의 기준이 되는 ‘정답문’(즉, 참조번역)과 테스트 모델별 자동번역 결과문을 비교하여 평가하였으며, 참조번역문에 오류가 있을 경우에는 원문을 참고하여 평가위원이 직접 평가하도록 하였다. 이러한 과정을 통해 얻어진 평가위원들의 점수는 그 평균을 계산하여 수동평가 점수로 삼았다. 수동평가에서 평가자에게 제시된 평가기준은 아래와 같다.

<표 1> 한국고전번역원 수동평가 자체 평가 기준

평가 점수	평가기준	보조 설명	오류의 중요성 및 비율
※ 평가요소 : 번역율누락 여부), 표현 및 문법, 의미 전달의 명확성			
5	의미 전달이 명확하고 문법적 오류가 없는 경우		0%
4	표현 및 문법에 사소한 오류가 있으나 비교적 의미가 이해 가능한 경우	문맥에 큰 영향을 주지 않는 오류	10~20%
3	주요 단어는 번역했으나 의미 전달이 일부 부정확한 경우	번역어휘 선택의 오류, 문법적 오류(비문), 사소한 결역 및 중복번역	30~50%
2	주요 단어를 번역하지 못해 의미 전달이 전반적으로 부정확한 경우	번역 누락 및 중복의 오류(심각한 결역 및 중복번역)	50% 이상
1	번역은 하였으나 이해가 전혀 불가능한 경우	전문 오역(종합 오류)	90%
0	번역 결과가 없는 경우		100%

또한, 2019년 수동평가에서는 2018년 최종모델의 자동번역 결과와 비교 후 번역 품질의 “높아짐(상), 비슷함(중), 낮아짐(하)”을 평가하여 번역기의 성능 향상을 확인할 수 있는 평가 항목을 추가하였다. 이와 같은 방식으로 진행된 평가 결과는 <표 2>와 같다.[9]

<표 2> 2019년 인공지능 기반 고문헌 자동번역 베이스모델(≪승정원일기≫) 성능 고도화 결과

구분	1차년도(2017) 결과	2차년도(2018) 결과	3차년도(2019) 결과
	(대표모델 : #33)	(대표모델 : 2018-11_1)	(대표모델 : BS_SJ_26)
수동평가 (5점 만점. 4점 이상 목표)	3.00점	3.51점	4.20점
BLEU평가 (100점 만점. 30점 이상 목표)	26.35점	29.08점	30.72점

이를 보면 자동평가는 1차년도에는 26.35점에 그쳤으나 3차년도에는 30.72점을 획득해 목표 점수인 30점을 근소하게 넘어선 것으로 되어 있다. 흥미로운 점은 수동평가다. 1차년도 3.00점에서 3차년도 4.20점으로 크게 향상되었는데, 설문 조사에서 흔히 쓰는 리커트 척도(likert scale)로 되어 있어 100점 만점으로 되어 있는 BLEU 점수로 환산하면 무려 84점이나 받은 셈이 된다. 이는 천문고전 특화 모델 평가에서도 마찬가지여서 BLEU 점수는 40.02점이고, 수동평가 점수는 4.05점(환산점수 81점)이었다.[10]

전문 역자들에 의해 이루어진 수동 평가 점수가 이처럼 기형적으로 높게 나타난 데서 다음과 같은 문제점을 확인할 수 있다.

첫째, ≪승정원일기≫ 자동번역시스템 수동평가 방식은 일반적인 수동평가 방식, 즉 유창성(가독성)/정확성(적절성)으로 분류하지 않고 번역의 정확성만을 기준으로 하여 5점 척도로 단순화하였다. 이로 인해 번역 품질 평가의 또 다른 주요 요소인 가독성 부분에 대한 평가가 제대로 이루어지지 못했다.

둘째, 각각의 척도에 해당하는 기준이 명확하게 제시되지 않아 평가자의 주관에 따라 평가결과가 크게 달라지는 경우가 나타났다. 평가의 신뢰성을 높이기 위해서는 평가자 훈련과 사전 조율도 필수적인데, 한국고전번역원 수동평가에서는 이 과정이 생략되었다.[11]

---

9) <2019년 클라우드 기반 고문헌 자동번역 확산 서비스 구축 : 사업추진결과보고서>

10) 2019년도 신규 개발 모델인 DM_AS_19를 기준으로 평가한 점수다.

11) 2019년에 수행된 수동평가를 예로 들어 보기로 한다. 이 당시 ≪승정원일기≫ 원문 100문장에 대해 테스트 모델 별 자동번역 결과문을 5점 척도로 평가하였는데, 2018년 모델과

셋째, 한국고전번역원의 수동평가는 한국고전번역원의 번역 방식에 익숙한 전문 역자로만 평가를 진행하였다. 전문 번역자가 수동평가를 수행할 경우 지나치게 엄격하게 평가할 우려가 있기 때문에 기계번역에 대한 기본적인 이해를 가지고 있는 인력이 평가를 수행할 필요가 있다.[12] 기계번역은 알고리즘의 선택, 코퍼스의 유형과 가공방법, 학습 절차 등에 따라 상이한 결과가 나올 수 있으므로, 이러한 특성을 이해하고 있는 사람이 평가에 참여하여야 여타 번역기와의 차이도 더 정확히 감별할 수 있고, 기계번역기의 성능을 향상시키는 데에도 활용할 수 있다.

넷째, 평가 결과에 대한 신뢰도 분석을 별도로 수행하지 않았다. 일반적으로 수동평가를 수행할 때는 평가의 신뢰도를 확보하기 위해 카파 상관계수(Cohen’s Kappa Coefficient) 등을 활용하여 평가 일치도를 측정하는 과정을 거친다. 이는 평가 결과가 얼마나 객관적이고 합리적으로 이루어졌는가를 검증하기 위한 것이기도 하지만, 성능 개선에 유용하기 때문이기도 하다.

이러한 제반 문제들이 중첩되어 실제보다 과장된 평가 결과가 도출되었으며, ‘한문고전 자동번역’에 대해 실제 이상의 기대감을 품게 하는 원인의 한 가지가 되었다.

3. 과제

1) 번역시스템 고도화

(1) 도메인 다양화

‘한문고전 자동번역’의 번역 성능에 대해서는 2021년에 검증한 바 있다. 하지만 이로부터 상당한 시간이 지난 만큼 그간 성능의 변화가 있었는지를 다시 살펴보았다. 대조를 위해 2021년 검증 당시와 동일한 문장을 사용했는바, ≪승정원일기≫

비교하였을 때 판단이 엇갈리는 문장, 즉 동일 번역문에 대해 어떤 평가자는 개선되었다고 평가한 반면 비슷하다거나 낮아졌다고 본 평가자도 존재하는 사례가 10건이나 보이며, 최고점과 최저점의 차가 3점 이상인 사례도 7건이었다.

12) Rossi & Wiggins(2013)

15문장(미번역 문장 3건 포함), ≪실록≫ 3문장(숙종, 현종, 효종), ≪만기요람≫, ≪대전회통≫ 각 1문장으로 구성되어 있다. 그 결과 2021년과 다른 결과가 나온 것은 3건에 불과했으며, 이 가운데 의미가 확연히 달라진 것은 없었다. 따라서 그간 성능의 변화가 거의 없었다고 볼 수 있다([첨부 1] 참조).

또한 ≪승정원일기≫와 같은 역사 문헌 외의 자료 번역 성능을 알아보기 위해 바이두 번역기와도 대조해 보았다. 바이두 번역기는 '한문고전 자동번역'과 더불어 한문고전 기계번역을 제공하는 유이한 번역기로, 2019년 성능 평가[13] 당시 사용했던 문장 중에서 10문장을 추출하여 한문고전 자동번역'과 대조해 보았다. 검토에 사용된 문장은 ≪春秋繁露≫·≪荀子≫·≪孝經≫·≪論衡≫·≪論語≫ 등 동양고전 자료와 ≪太祖實錄≫에서 임의로 추출하였는데, 바이두 번역기의 경우 2019년과 2023년 조사 결과 중 서로 달리 번역된 것이 3건이었으며, 이 가운데 의미가 확연히 달라진 것은 없었다.

한편 2020년에 두 번역기의 성능을 비교하는 수동평가를 진행한 바 있었는데, '한문고전 자동번역'이 더 나은 것으로 조사된 경우가 3건(≪春秋繁露≫·≪荀子≫·≪孫子兵法≫)이었고, 나머지 7건은 바이두 번역기가 더 나았으며, 전체 평점에서도 근소하나마 우위에 있는 것(14.5 : 15.9)으로 조사된 바 있었다(<표 3>).[14] 2023년 10월 24일에 진행된 이번 조사에서도 의미 있는 변화가 나타나지 않았으므로 여전히 당시 조사 결과와 비슷한 양상을 띠고 있다고 할 수 있다. '한문고전 자동번역'은 내용이나 문체가 ≪승정원일기≫와 유사한 자료를 번역하는 데는 양호한 성능을 발휘하지만 다른 유형의 자료를 번역하는 데는 명확한 한계를 지니고 있음을 확인케 한다.

13) 조성덕·박종훈·이종웅·김우정(2019)

14) 김우정 외(2021). '한문고전 자동번역'과 바이두 번역기에 각각 투입하여 산출된 번역문을 대조하는 방식으로 진행하였다. 평가자는 ≪승정원일기≫ 기계번역 전문가 평가를 수행했던 평가자 3인으로 구성하였으며, 조사일시는 2020년 12월 4일이다.

<표 3> ‘한문고전 기계번역’과‘바이두 기계번역’의 성능 비교(김우정 외, 2021)

Set No.	한문고전 자동번역			바이두 번역기		
	유창성	명확성	총점	유창성	명확성	총점
1	4.3	11.0	15.3	3.3	11.0	14.3
2	5.0	13.3	18.3	3.7	12.0	15.7
3	4.7	10.0	14.7	5.0	13.0	18.0
4	2.7	10.0	12.7	5.7	9.7	15.3
5	2.7	9.7	12.3	2.7	10.7	13.3
6	3.7	12.3	16.0	3.3	10.3	13.7
7	2.7	8.3	11.0	6.3	11.0	17.3
8	4.7	13.3	18.0	6.0	13.3	19.3
9	3.0	6.0	9.0	4.0	9.7	13.7
10	5.7	12.3	18.0	5.7	12.7	18.3
평균	3.9	10.6	**14.5**	4.6	11.3	**15.9**

앞에서도 지적하였다시피 ‘한문고전 자동번역’은 ≪승정원일기≫ 및 이와 유사한 특성을 지닌 문헌 번역에 특화된 모델일 뿐이다. 따라서 ‘한문고전 자동번역’의 성능을 향상시키기 위해서는 다양한 입력 데이터 환경에서 번역이 가능한 범용적 모델을 개발하는 데 더 힘을 쏟아야 한다.

이를 위해서는 다양한 유형의 도메인을 적절한 규모와 비율로 학습시키는 것이 관건이 되겠지만, MTPE(Machine Translation Post-Editing)을 활용해 번역 성능을 높이는 환류형 번역 체계를 구축해보는 것도 고려해 볼 일이다. MTPE는 기계가 번역한 텍스트를 인간이 교정(edit)하는 작업으로, 이미 기계가 초벌 번역한 텍스트를 기반으로 하여 새로운 텍스트를 번역하는 작업이다. 기계번역의 도움 없이 평균적인 전문 번역가가 제공하는 번역 품질 수준으로 번역하려면 시간과 비용이 많이 드는데, MTPE는 이 시간과 비용을 줄이는 데 도움을 준다. 또한 MTPE는 기계번역의 품질 평가에서도 활용될 수 있는데, 기계번역을 수행한 후보 번역들을 대상으로 전문 번역가가 MTPE를 수행했을 때 어느 번역모델이 시간 당 생산성이 높은지를 측정하거나 시간 당 평균 몇 단어 정도의 오류를 수정하였는지 등을 검토하는 방식으로 번역 품질을 평가하고 성능 향상에 활용할 수 있다.

(2) 클라우드 기반 한문고전 번역 플랫폼 구축

‘한문고전 자동번역’ 사업을 테스트 수준에서 마무리하고자 한다면 그만이지만 장기적인 목표를 가지고 추진해나가고자 한다면 클라우드 기반 번역 플랫폼 구축도 고려해봄직하다. 고전문헌을 특화와 범용의 두 축에서 기계번역 모델을 통해 학습하고 이를 클라우드 환경의 플랫폼으로 제공하면, 일반적인 사용자가 번역하기 힘든 고전문헌을 여러 곳에서 번역하는 것이 아닌 하나의 플랫폼에서 손쉽게 번역할 수 있다는 장점이 있다. 일반 사용자 또는 개발자는 고전문헌 자료 수집을 위해서 여러 플랫폼을 사용할 필요 없이 단일 플랫폼을 사용하여서 사용 또는 개발을 수행할 수 있고, 특수하거나 특화된 기계번역이 필요한 경우 플랫폼 학습데이터에 누적된 데이터를 제공 받아서 효과적인 사용자 번역시스템을 구성할 수 있을 것이다. 또한 기계번역된 모델을 생성하게 됨에 따라서 차후 비슷한 유형의 데이터 또는 동일한 추가 데이터를 학습함에 있어서도, 기존의 기계번역이 누적 데이터로써 도움이 되기 때문에 기계학습에 도움을 줄 수도 있다. 그리고 방대한 양의 양질의 고전문헌 데이터를 클라우드 환경에 저장하여 확보할 수 있고 이를 번역해 일반 사용자에게 제공하고 사용자는 클라우드를 통한 개발을 기대해 볼 수 있으므로 고문헌 번역 및 개발 비용 절감에 기여할 수 있다. 마지막으로 누적된 양질의 고전문헌 데이터, 특화된 사용자 사전, 번역 메모리 정보 제공을 통해서 추가적인 산업 창출과 동시에 다른 부가 서비스와의 연동, 플랫폼 확장을 통한 번역 서비스를 제공할 수도 있다. 아울러 플랫폼을 데이터를 위한 클라우드 플랫폼으로 개발한다면 컴퓨터 프로그램 소스를 공유하고 협업하여 개발할 수 있는 버전 관리 시스템인 깃허브(Github)나 네이버 오픈소스, 카카오테크와 같은 형태로 발전시킬 수도 있을 것이다.

2) 번역 품질 제고 방안

(1) 데이터 구축 및 정제 고도화

지극히 당연한 말이지만 양질의 번역 결과를 얻으려면 양질의 데이터가 필요하

다. 우리는 이미 많은 번역 데이터를 가지고 있다고 생각하지만 기계번역 학습에 적합한 데이터는 막상 그렇게 많지 않다. 돌이켜보건대 2000년부터 시작된 고전적 데이터베이스 사업은 대량의 원시 코퍼스 제작을 가능케 한 동력이 되었으며, 그 결과 그 어떤 나라보다 많은 고전적 디지털 자원을 확보하게 되었다. 하지만 집적과 검색에 초점을 맞추고 진행되어 왔기 때문에 데이터의 활용 측면에서는 개선해야 할 점이 상당히 존재한다.

기계번역 학습을 위한 양질의 번역 데이터를 구축하려면 이에 적합한 번역 방법 및 번역 규칙부터 마련해야 하는 바, 문종이 전혀 다른 예를 실험하는 경우만이 아니라, 다른 언어 연구에서 이미 밝힌 바 있는 문장의 길이나 POS 태깅의 유무, 문장부호의 유무 등 다양한 경우와의 영향 관계에 대한 연구가 필요하다. 또한 입력 데이터가 학습하기에 효과적인 데이터였는지를 검증하는 일도 중요하다. 합성 데이터, out-of-domain 데이터 등등이 실제 번역 학습에 있어서 도움을 주는 데이터인지에 대한 연구, 혹은 원문 문장을 어떤 규칙의 코퍼스로 구성을 하면 같은 번역 모델에서 가장 효과적인 학습을 수행하고 좋은 번역 결과를 발생할 수 있는지, 코퍼스를 구성하는 과정에서 최대한의 원문 입력 데이터에서 수정을 덜 하면서 최대한의 번역 학습이 가능하게 하는 것은 어떤 것인지에 대한 개별 연구도 필요하다. 이러한 연구가 진행된다면 최소한의 비용으로서 코퍼스를 구성할 수 있을 것이며, 가장 보편적인 코퍼스 데이터를 사용해 학습을 진행할 수 있을 것이므로 실제 번역에 있어서도 더 나은 결과를 얻을 수 있다.

(2) 사전 정보 및 부가 시스템 개발

한문이라는 언어의 특성상 다양한 참고 자료가 필요하다. 따라서 기계번역의 성능을 개선하기 위해서는 코퍼스, 기계번역 알고리즘 못지않게 자연어 처리를 위한 사전정보, 예를 들어 한자의 이체자・통가자 처리와 다양한 자의군의 통합과 연계, 특수 용어 시소러스를 포함한 어휘망, 더 나아가 컨텍스트적 해석을 가능하게 하는 감성어휘사전 등이 제공될 필요가 있다.

이러한 다양한 사전정보를 학습한 알고리즘 간의 상호 비교는 한국 한문 기계번역 최적화를 위한 선행 연구 성격을 지니고 있다. 만일 이러한 부가정보가 타

언어 번역시스템에서 유의미한 결과를 보여주고 있다면 우리는 이를 어떻게 구축할 것인가에 대한 논의가 절대적으로 필요하다. 이러한 연구는 한문전적과 관련된 대량의 자료를 손쉽게 검색하고 통계화함으로써 좁게는 문자학 및 문법학, 사전학, 문학의 수사학 및 언어학 연구 등에 활용할 수 있으며, 넓게는 번역 모델 개발, 문화학, 역사학, 사회학 등 다방면에 적용할 수 있다. 이런 점에서 전산언어학을 접목한 한문자료 코퍼스 구축은 예측 불가능할 만큼 다양한 파생 효과를 창출할 수 있는 종합적이고 융합적인 미래 과제이다.

전처리 과정에서 필수적인 토큰을 어떤 단위로 분석할지 검증하기 위한 시스템 구축도 요긴한 과제다. 현재의 기계번역 시스템에서 사용하는 토큰 분석은 무엇을 기반으로 하는지 알 수 없으나, 현재까지 한문을 대상으로 한 토큰분석기는 개발된 바 없다. 토큰분석기는 쉽게 표현하면 자동 표점기와 형태소 분석기로, 자연언어 처리에서는 가장 기본적인 작업이다. 이 분석에 따라서 모든 언어의 특성 분석이 달라지기 때문이다. 추측컨대 현재 시스템에 한자의 자의 정보가 번역 결과에 큰 영향을 미치지 않는다면 이는 토큰을 단어 위주로 구성하기 때문일 수 있으며, 반대로 일정 정도 영향을 미친다면 이는 토큰이 낱글자 단위로 구성되었기 때문이라고 유추할 수 있다.

근래 많이 다뤄지고 있는 언어모델의 일종인 BERT(Bidirectional Encoder Representation from Transformers) 등을 기계번역 품질 향상에 활용하는 방안에도 관심을 가질 필요가 있다. BERT는 Transformer 기술 중 인코더 부분만 사용한 사전학습(pre-training) 모델로, 상대적으로 적은 데이터로도 고성능을 낼 수 있다는 장점이 있다. 고전문언문의 경우, 殆知閣(http：//www.daizhige.org)과 ≪四庫全書≫ 자료를 기반으로 구축한 GuwenBert(https：//github.com/Ethan-yt/guwenbert), Siku-Bert(https：//huggingface.co/SIKU-BERT/sikubert)이 있는데, 문맥을 학습하고 단어 시퀀스에 확률을 할당하여 이후 출현할 단어를 예측하고, 개체명을 인식하고 형태소를 분석하는 기능을 갖췄다. 이밖에 다언어 의존관계 말뭉치(Universal Dependencies)와 BERT・RoBERTa・DeBERTa 등의 사전학습 모델을 활용해 ≪孟子≫ 등 몇 종의 고전 문언문에 품사 태깅을 한 연구도 있다.[15] 이와 같은 모델이 중요한 이유는 언어 특성에 따라 기본 알고리즘에 변화만 주면 기존의 인간 혹은

15) 安岡孝一(2022), 127~163면.

형태소 분석기에 비해 비교할 수 없는 성능을 보여줄 수 있기 때문이다.

3) 평가 방법 개선

번역 품질을 객관적으로 평가함으로써 궁극적으로 번역 성능 향상에 기여할 수 있는 방법과 절차도 마련하여야 한다. 우선 자동평가의 경우, 평가방식마다 장단점이 있으므로 특정한 평가방식만 고집할 일이 아니라 번역 텍스트의 성격과 목표에 맞는 방식을 찾아 적용하는 것이 바람직하다. 예를 들어, 일반적으로 BLEU 점수는 30점이 넘으면 번역이 잘 되었다고 평가하는데, BLEU 점수의 한계 등을 고려하면, 이런 절대평가 방식보다는 여러 다양한 평가 시스템과 비교한 상대평가 방식을 활용하는 것이 더 나을 것이다. 짧은 문장 위주로 평가되는 방식도 재고해 볼 일이다. 일반적으로 한 문장의 음절을 최대 100음절 등으로 제한하거나 15어절 등으로 제한하는 번역방식을 채택한다. 즉 대부분의 자동평가 시스템에서는 짧은 문장이 높은 점수를 받도록 설계되어 있기 때문에 평가 테이터셋도 상대적으로 짧은 문장이 선호된다. 그만큼 누락 등의 오류가 나타날 가능성이 적고 확률적으로 정확도가 일정 수준 이상 유지될 가능성이 있기 때문이다. 그러나 고전번역의 특성상, 문장의 유형을 언어학적으로 분류하여 기계번역의 품질을 평가해 보아야 할 필요성도 있어 보인다. 문장의 유형에 따라 단문과 복문으로 구분하고 복문은 다시 내포문, 대등문, 조건문 등으로 세분화하여 각각의 번역 품질이 어떻게 나타나는지를 세부적으로 평가해 볼 필요도 있다. 또한 문장의 길이가 길어질수록 복잡한 수식구조가 나타날 수 있는데 이에 대한 세부지침도 마련할 필요가 있다.

다음으로 수동평가 방식도 개선해야 한다. 선행연구에서는 평가지표 개발과 관련하여 평가자의 지식수준, 환경 등 정량화하기 어려운 변수들도 반영할 수 있는가, 평가지표는 번역의 명확성과 유창성을 고루 평가할 수 있도록 구성되었는가, 누락되거나 중복된 평가요소는 없는가, 배점은 평가지표의 특성이나 중요도에 알맞게 부여되었는가, 배점의 총합이 번역품질의 차이를 확인할 수 있도록 적절하게 부여되었는가를 살펴야 한다고 하였다. 또한 평가방식 개선과 관련하여 기계번역문과 인간번역문 차이를 감별할 수 있는지를 평가할 수 있는가,[16] 평가자의 주관

에 따른 편차를 줄일 수 있는 절차가 마련되어 있는가, 번역품질 평가집단은 합리적이고 객관적으로 구성되었는가, 번역기의 활용 가능성(역자들의 번역 지원, 대국민 서비스 등)을 판단하는 절차가 포함되었는가를 고려해야 한다고 하였다.[17]

또한 한국고전번역원에서 수동평가에 사용한 리커트 척도와 달리 새로운 평가 기준을 개발하고(<표 4>),[18] 이를 적용한 평가를 진행해 평가 지표의 적절성을 검토하였다.

<표 4> 김우정 외(2021)의 수동평가 평가 기준

평가 영역	평가 지표	평가 요소	배점 (총25점)
명확성	오류 오역	• 어휘의 의미를 바르게 표현하였는가 (인명, 지명, 관명, 서명의 이해, 漢字竝記의 적합성, 관용구의 풀이 따위 포함)	3
		• 어구의 연결이 옳은가 (ex : 軍職實職 : '군직과 실직', '군직의 실직')	3
		• 문장의 구조, 구와 구의 관계에 맞게 번역하였는가 (역접/순접, 인과관계 등)	3
		• 문형(평서문/감탄문/의문문 따위), 화자와 청자의 관계, 글의 어조 등에 맞게 번역하였는가	3
	추가 누락 미번역	• 원문에 없는 내용이 추가되었는가	3
		• 원문에 있는 내용이 누락되었는가	
		• 원문을 번역하지 못하고 번역문에 노출하였는가	

16) 자동평가와 수동평가는 단지 평가방식만 다른 것이 아니라 그 목적과 성격도 다르다. 원문, 정답문, 번역문을 모두 제시한 후 평가하는 기존의 평가방식, 즉 기계학습의 결과물임을 인지한 상태에서 진행되는 평가는 '원문 VS 번역문'에 대한 평가 외에 '정답문 VS 번역문'에 대한 평가도 동시에 이루어지는 셈이어서, 정답문에 의해 왜곡될 가능성이 있다. 한국고전번역원의 수동평가 지침에도 밝혀져 있다시피 인간이 작성한 정답문이 반드시 정답이라고 할 수는 없다.

17) 김우정 외(2021)

18) 명확성 영역은 '오류・오역'과 '추가・누락・미번역' 2가지 지표로 나누었다. 이 2가지 지표는 번역 품질평가에서 차지하는 비중이 높음을 감안하여 평가요소를 세분하여 제시하되 각 요소 간 중복이 없도록 예시를 들어 설명하였다. 요소별 배점은 3점으로 하고 결점이 발견될 경우 1점 감점을 원칙으로 하되 복수로 나타나거나 단수로 나타나더라도 심각한 오류・오역에 해당할 경우 추가 감점이 가능하도록 하였다. 추가・누락・미번역은 통합하여 3점을 부여하였다. 유창성 영역은 원문과의 대조로 인한 영향을 배제하기 위해 번역문만 제시하였으며, 맞춤법과 띄어쓰기를 포함하여 평가하도록 하였다. 배점은 10점 만점으로 하되 주관성을 완화하기 위해 구간별로 점수를 부여하도록 하였다.

평가 영역	평가 지표	평가 요소	배점 (총25점)
유창성	유창성 및 가독성	• (번역의 정확성과 무관하게) 번역문이 자연스러운가 • 맞춤법, 띄어쓰기의 오류는 없는가 - 10~9점 : 거의 완벽하거나 사소한 오류가 있을 경우 - 8~7점 : 번역투는 있지만 의미는 이해 가능한 경우 - 6~5점 : 맞춤법, 띄어쓰기를 포함한 문법적인 오류, 의미 파악이 어려운 경우 - 4~1점 : 내용 이해가 불가한 경우	10

평가는 ≪승정원일기≫를 비롯한 한문고전 경험을 갖춘 3인의 전문가에게 위촉하였으며,[19] 그 결과는 상기 지표에 따라 ‘한문고전 자동번역’의 성능 평가를 수행하였다.

<표 5> 새로운 평가 척도를 적용한 ‘한문고전 자동번역’ 평가 결과(김우정 외, 2021)

유형		Set No.	평가자A			평가자B			평가자C			종합		
			유창성	명확성	합계	유창성	명확성	합계	유창성	명확성	합계	유창성	명확성	합계
단문	承-人	1	8	15	23	9	14	23	10	14	24	9.00	14.33	23.33
		2	10	15	25	9	14	23	10	15	25	9.67	14.67	24.33
	承-機	3	10	15	25	9	14	23	10	14	24	9.67	14.33	24.00
		4	10	15	25	9	15	24	9	15	23	9.33	15.00	24.00
		5	6	13	19	8	13	21	8	14	22	7.33	13.33	20.67
		6	6	14	20	9	15	24	9	13	22	8.00	14.00	22.00
	其-機	7	8	14	22	9	14	23	10	14	23	9.00	14.00	22.67
		8	8	15	23	8	15	23	10	13	24	8.67	14.33	23.33
		9	10	15	25	8	14	22	8	15	23	8.67	14.67	23.33
		10	8	15	23	10	15	25	9	15	24	9.00	15.00	24.00
		11	6	12	18	6	13	19	8	11	19	6.67	12.00	18.67
		12	10	15	25	10	15	25	8	14	22	9.33	14.67	24.00
전후문장맥락	承-人	13	10	14	24	9	14	23	10	13	23	9.67	13.67	23.33
	承-機	14	6	14	20	6	13	19	5	14	19	5.67	13.67	19.33
		15	6	11	17	8	13	21	7	13	20	7.00	12.33	19.33
		16	6	14	20	8	13	21	6	13	19	6.67	13.33	20.00
	其-機	17	6	12	18	6	14	20	6	12	18	6.00	12.67	18.67
		18	6	14	20	8	13	21	6	13	19	6.67	13.33	20.00
		19	10	14	24	8	13	21	7	14	21	8.33	13.67	22.00
		20	8	12	20	8	12	20	5	10	15	7.00	11.33	18.33
평균			7.9	13.9	21.8	8.25	13.8	22.05	8.05	13.45	21.45	8.07	13.72	21.77

* st1, 2, 13은 ≪승정원일기≫ 인간번역문(承-人), set3~6, 14~16는 ≪승정원일기≫ 기계번역문(承-機), set7~12, 17~20은 기타 자료의 기계번역문(其-機)임.

19) 평가자A는 ≪승정원일기≫ 역자, 평가자B는 개인 문집을 주로 번역한 한국고전번역원 외

한국고전번역원에서 실시한 평가(2019년 BS_SJ_26 모델 대상 평가)의 경우, 자체 수행 평가자별 점수 평균점이 3.54점부터 4.50점까지 편차가 컸던데 반해, 새로운 척도를 적용한 평가에서는 명확성 (13.45~13.90)과 유창성(7.9~8.25)에서 모두 큰 편차가 발생하지 않았다. 이에 한국고전번역원에서 수행한 수동평가 결과에 비해 연구팀이 개발한 수동평가 방식이 '한문고전 자동번역' 성능평가에서 더 유의미한 결과를 얻을 수 있다고 판단된다. 다만 수동평가는 번역 텍스트의 성격에 맞는 지표를 개발하여 적용하는 것이 바람직하므로, 이에 관한 깊이 있는 분석과 후속 연구가 수반되어야 할 것이다.

4. 전문 역자들의 반응

'한문고전 자동번역'의 일차적 목표는 ≪승정원일기≫ 번역 업무의 편의성 및 효율성을 증진하는 데 있다. 이를 가늠해보기 위한 선행연구에서는 전문 역자들을 대상으로 설문조사를 실시한 바 있는데, '한문고전 자동번역'을 활용해본 후, 번역 품질, 편리성, 유용성에 대한 의견을 묻는 방식으로 진행되었다.[20]

<표 6> '한문고전 자동번역' 역자 평가(김우정 외, 2021)

평가자	연령대	번역 경력	귀하는 구글, 파파고, 바이두 등 인공지능 번역기를 사용해 본 경험이 있습니까?	귀하는 새로운 기술을 사용해 보는 것에 적극적인 편입니까?
A	50대	10년 이상 ~ 15년 미만	사용해 본 적이 있다	그렇지 않다
B	40대	5년 이상 ~ 10년 미만	주 1회 이상 사용	매우 그렇다
C	30대	5년 이상 ~ 10년 미만	사용해 본 적이 전혀 없다	보통이다
D	50대	10년 이상 ~ 15년 미만	사용해 본 적이 있다	보통이다
E	50대	10년 이상 ~ 15년 미만	사용해 본 적이 있다	그렇다
F	30대	5년 미만	가끔 사용한다	보통이다
G	30대	5년 미만	사용해 본 적이 있다	그렇다
H	50대	15년 이상 ~ 20년 미만	주 1회 이상 사용	매우 그렇다
I	40대	10년 이상 ~ 15년 미만	가끔 사용한다	그렇다

부 역자, 평가자C는 번역거점연구소 공동연구원이다. 조사용 문장은 <첨부 1>과 같으며, 조사일은 2020년 11월 14일이다.

20) 조사 대상자는 ≪승정원일기≫ 번역이나 평가 경험이 있는 역자 9명이며, 조사일은 2021년 1월 8일~1월 14일이다.

설문조사에 참여한 역자는 총 9명으로 연령과 번역 경력을 고려하여 선정하였다. 이 중 5인(A~E)은 ≪승정원일기≫를 번역한 경험이 있으며, 나머지 4인(F~I)은 ≪승정원일기≫ 번역 평가 경험이 있는 역자이다. 본 설문에는 번역기에 대한 평가에 앞서 개인의 혁신성도 함께 조사하였다. 혁신성은 기존의 것과는 다른 ‘새로운 아이디어를 빨리 채택하는 정도’로 규정할 수 있는데,[21] 역자들의 ≪승정원일기≫ 번역기 인식에 변인으로 작용할 수 있다고 판단하였기 때문이다. 조사 결과를 요약하면 다음과 같다.

① 번역 품질 : 번역기를 활용한 후 번역 품질을 맥락, 어휘, 가독성 부분으로 나누어 평가하게 하였다. 전반적으로 “보통”으로 평가한 의견이 많았다. 세부 의견을 살펴보면 공통적으로 상투적인 구문의 번역 품질은 상당히 뛰어나지만 그밖의 경우에는 번역 품질이 낮다는 의견, 어휘 부분은 고유명사 번역에서 오류가 많다는 의견이 많았다. 또 컨텍스트를 충분히 파악하지 못한다는 점을 한계로 꼽았다.[22]

② 편리성 : 편리성 부분에서는 의견이 갈렸다. 다만 ‘그렇지 않다’고 답한 평가자는 기존의 다른 기계 번역기를 사용한 경험이 없기에 승정원일기 번역기에 대한 평가도 긍정적이지 않았던 것으로 보인다.

③ 유용성 : 평가자들은 현재 번역기의 번역 품질에 대해서는 다소 유보적인 태도를 보였으나 업무에 활용 가능성에 대해서는 대체로 긍정적으로 인식하였다. 9명 중 8명이 번역기가 실제 번역 시간을 단축시킬 수 있다고 평가하였다. 9명 중 6명이 업무에 유용하다고 평가하였으며 추후 활용 여부에 있어서도 상당히 긍정적으로 답변하였다. 다만 평가자 A는 전반적으로 ‘한문고전 자동번역’의 유용성에 대해서 긍정적으로 평가하지 않았는데, 이는 혁신성 부분에서 보수적인 태도가 ‘한문고전 자동번역’의 평가에 영향을 미쳤을 가능성도 생각할 수 있다. 세부 의

21) 천종성(2020), 283면.
22) 자세한 내용은 김우정 외(2021) 참조.

견을 살펴보면 평가자들은 번역기의 번역 결과를 초벌 번역 형태로 활용할 수 있는 가능성에 대해 긍정적으로 인식하였다. 특히 특정 어휘나 형식이 반복적으로 나열되는 저맥락 텍스트의 경우 번역 속도를 향상시킬 수 있다고 보았다, 그리고 번역 과정에서 원문 텍스트의 전후 기사의 내용을 번역기를 통해 신속하게 확인할 수 있다는 점, 번역의 규식을 확인할 수 있다는 점에서 번역기의 활용 가능성을 높게 평가하는 의견이 있었다.

위의 설문조사를 통해 번역기의 번역 품질뿐 아니라 편리성과 효용성 부분에서도 의미 있는 평가와 제언을 확인할 수 있었다. '한문고전 자동번역'은 충분히 신뢰할 만한 수준은 아니지만, 전문 역자들의 업무 효율을 높여주는 유용한 도구로 활용될 수 있는 가능성을 지닌다. 이로 볼 때, '한문고전 자동번역' 모델 개발의 애초의 목적은 어느 정도 달성한 것으로 판단된다. 그러나 빈번하게 확인되는 번역 오류를 개선하고 편리성과 효용성을 제고하기 위해, 번역기의 성능을 고도화하는 한편 사용자의 번역 습관・ 번역기 활용 방식을 고려한 개선이 필요하다는 것을 본 설문조사를 통해 확인할 수 있다. 이상의 심층 설문조사는 자동번역 시스템의 구체적인 피드백을 제공하기에, 자동/수동평가와 함께 수행한다면 향후 '한문고전 자동번역' 모델을 고도화화는 데 도움을 주리라 기대한다.

5. 나가는 말

오늘날 디지털 기술은 물리적 공간을 이동하는 차원을 넘어서서 삶과 사유의 지평에서 창조와 탈주를 가능케 하는 단계로 나아가고 있다. 그러나 불행스럽게도 인문학, 그중에서도 교양처럼 취급되는 고전학은 디지털이 만들어내는 놀랍고 아름다운 세계를 증명하는 콘텐츠로만 소비되는 느낌이다. 이제 인문학은 정보의 저장력과 확산력이 더욱 강해질 디지털 시대 속에서 존재의 의의를 찾아내야 하는 상황에 직면해 있다.

이런 점에서 정부 산하기관인 한국고전번역원이 막강한 자본과 기술을 가진 거

대 IT 기업이 아니고는 엄두를 못 낼 한문고전 기계번역에 도전한 일은, 사업의 성패를 떠나 디지털 전환 시대에 능동적으로 대응하고자 하는 의지를 보였다는 점만으로도 큰 의의를 지닌다. 공개된 번역 성능이 사업 착수 단계에서 제시된 기대와 전망에 못 미친다는 비판도 없지 않으나 기계번역의 현황에 견줘보면 비교적 단기간에 상당한 수준의 성과를 올린 것으로 평가할 만하며, 이는 ≪승정원일기≫를 비롯한 한문고전 번역에 참여했던 전문가들의 반응을 통해서도 확인되었다.

다만 실제로 번역 현장에 적용하기에는 아직 보완되어야 할 점이 많고, ≪승정원일기≫라는 특수 도메인에만 한정해 성과를 발휘하는 번역기라는 사실을 이해하지 못하는 일반인에게 공개하기에도 부적합한 것 또한 사실이다. 어떤 상품이든 아이디어로부터 개발 단계를 거쳐 상품화되기까지는 많은 시간과 노력이 필요하다. 일반적인 기술성숙도(TRL)로 볼 때 ‘한문고전 자동번역’은 1~5단계를 건너뛰고, 기존의 다른 언어모델과 코퍼스 구축을 통해 7단계에 해당하는 데모 버전을 개발한 후, 또다시 단계를 뛰어넘어 9단계에 해당하는 ‘상용제품 생산’에 도전하는 형국이다. 이처럼 단기적인 성과에 집착할 경우 기초 연구 부족으로 인한 문제들이 발생할 뿐 아니라 최종 성과 역시 과소평가되거나 부정될 수 있다는 점을 유의해야 할 것이다. 이러한 점을 고려하여 다음과 같은 점을 추가로 제안하는 것으로 이 글을 마무리하고자 한다.

첫째, 한문고전 데이터 구축을 단지 사업으로만 취급할 일이 아니라 전산언어학, 데이터사이언스 등 유관 학문과 연계해 추진함으로써 학적 기초를 두텁게 하고 활용 범위를 확대할 수 있도록 할 필요가 있다. 양질의 코퍼스 구축은 문헌의 해독에 기울어 있는 동아시아 고전학의 연구 범위를 크게 넓힐 뿐만 아니라 인공지능 기술과 접목하여 다양한 연구를 촉발하는 기초가 되기 때문이다.

둘째, 기계번역에는 대용량의 전문사전, 이중언어 사전, 특수용어집 등 다양한 부가정보도 필요하다. 그런데 이와 같은 사전류들은 그저 번역 성능을 제고하는 도구에 그치는 것이 아니라 문자학, 문법학, 사전학, 문학의 수사학 및 언어학 연

구 등에 활용할 수 있으므로 궁극적으로는 인문고전학의 토대를 굳건히 하는 기초자료가 되기도 하므로, 꾸준한 연구와 개발이 요구된다.

셋째, 한문고전 기계번역에 대한 인식의 변화가 필요하다. 기계번역은 상당한 시간과 노력이 필요한 일이므로 당장의 효용만 요구할 일이 아니라 중장기적 목표를 세우고 단계적으로 추진해나갈 수 있도록 하여야 한다. 디지털 자원에 대한 이해를 넓힘으로써 번역 이상의 다양한 효과를 창출할 수 있다는 믿음과 인내심을 가지고 지원하여야 한다. 한문고전 기계번역에 대한 투자가 학문 이상의 의미를 지니고 있다는 점도 강조할 필요가 있다. 우리 선조의 손에서 나온 문헌을 우리 손으로 번역하지 못해 중화권 국가에 의존하는 일이 생기도록 하지 않으려면, 인력 양성만 바라볼 일이 아니라 우수한 번역기를 개발하여 대응할 수 있도록 하는 것이 바람직하다. 정보화시대의 특성상 데이터와 그 결과로 생성되는 정보는 '옳음'으로 간주되어 설사 잘못된 정보라도 다중의 사용자가 믿음을 가지고 신뢰하는 경우가 있다. 특히 알고리즘과 데이터가 조작되거나 편향되었을 때의 문제는 더욱 심각하다. 이는 현실 언어의 번역보다 한문 기계번역이 더 절실하게 요구되는 이유이기도 하다.

끝으로 이 글에서 다루지 않은 문제들, 예를 들어 데이터 정제에 필요한 행·재정적 지원 문제, 데이터의 저작·배포·사용권을 포함한 법적 문제, 연구용 프로그램이나 연구 결과 공유를 위한 한국형 플랫폼 구축 문제 등에도 많은 관심과 지원이 이루어져야 할 것이다.

참고문헌

김우정·박용범·배은한·백한기·변구일·안광호·정성훈·최지연·하지영·허철(2021), <≪승정원일기≫ 자동번역시스템의 활용방안 연구>, 한국고전번역원 정책연구과제.

단국대학교 한문교육연구소 편(2023), ≪한국과 중국의 디지털인문학≫, 단국대학교 한문교육

연구소.

이준호(2019), <신경망기계번역의 객관적 평가를 위한 예비연구: 자동평가와 수동평가의 균형점>, ≪통번역학연구≫ 23권 5호, 한국외국어대학교 통번역연구소, 171~202면.

정성훈・하지영・김우정(2021), <한문고전문헌의 기계번역 평가방안 탐색>, ≪한문학논집≫ 60, 105~156면.

조성덕・박종훈・이종웅・김우정(2019), <바이두(百度) 번역기의 한문고전 번역 수준과 향후의 과제>, ≪한문학논집≫ 53, 89~126면.

천종성(2020), <전문번역사들의 기계번역 수용에 관한 연구>, ≪한국융합학회논문지≫ 제11권 제6호, 한국융합학회, 281~288면.

최효은・이지은(2017), <특허 기계번역 결과물의 평가: KIPRIS 의 무료 한영 기계번역을 중심으로>, ≪통역과 번역≫, 19(1), 139~178면.

한국고전번역원(2019), <2019년 클라우드 기반 고문헌 자동번역 확산 서비스 구축사업 추진결과보고서>

한국고전번역원(2020), <한국고전번역원 제31회 연구집담회 발표자료집>

安岡孝一(2022), <Universal Dependencies와 BERT/RoBERTa 모델을 통한 고전 중국어 정보처리>, ≪한자한문응용연구≫ 1, 단국대학교 한문교육연구소, 127~163면.

Bazrafshan. M(2014), *Semantic Features for Statistical Machine Translation*, Unpublished doctoral thesis, University of Rochester, New York, US.

Banerjee, S. & A, Lavie(2005), *METEOR: An automatic metric for MT evaluation with improved correlation with human judgments*, Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization, pp.65~72.

Chung, Hye-Yeon(2020), *Automatische Evaluation der Humanübersetzung: BLEU vs. METEOR*, Lebende Sprachen 65(1), pp.181~205.

Denkowski, M, & A, Lavie(2011), *Meteor 1.3: Automatic metric for reliable optimization and evaluation of machine translation systems*, Proceedings of the sixth workshop on statistical machine translation, pp.85~91.

Papineni, Kishore, et al(2002), *Bleu: a method for automatic evaluation of machine translation*, Proceedings of the 40th annual meeting of the Association for Computational Linguistics, pp.311~318.

Callison-Burch, C, et al(2006), *Re-evaluating the role of BLEU in machine translation research*, 11th Conference of the European Chapter of the Association for Computational Linguistics, pp.249~256.

Callison-Burch, C, et al(2007), *(Meta-) evaluation of machine translation*, Proceedings of the Second Workshop on Statistical Machine Translation, pp.136~158.

S. Hampshire & Porta Salvia, C(2010), *Translation and the internet：Evaluating the quality of free online machine translators*, Quaderns：Revista de Traduccio, 17, pp.197~209.

Han. L(2016), *Machine translation evaluation resources and methods.* A survey. arXiv：1605.04515v8, Cornell University Library.

J. Hutchins, and S. Harold(1992), *An Introduction to Machine Translation*, London：Academic Press.

John B. Carroll(1966), *An experiment in evaluating the quality of translation*, Mechanical Translation and Computational Linguistics, 9(3-4), pp.67~75.

Rossi, L., & Wiggins, D(2013), *Applicability and application of machine translation quality metrics in the patent field*, World Patent Information, 35, pp.115~125.

Snover, M, et al(2006), *A study of translation edit rate with targeted human annotation*, Proceedings of the 7th Conference of the Association for Machine Translation in the Americas：Technical Papers, pp.223~231.

Van Slype, G(1982), *Economic aspects of machine translation*, Practical Experience of Machine Translation, Amsterdam：North-Holland, pp.79~93.

Woojeong KIM(2021), *Artificial Intelligence-based Recognition of Chinese Characters and Translation of Classical Chinese：Focusing on the Current State of Korea*, IJCCS(世界漢字通報), Vol.4, No.2. pp.1~62.

<첨부 1> ≪승정원일기≫ 기계번역 품질 조사용 문장(전문가 집단)

연번	문 및 번역문(참조 번역) ① 한문고전 자동번역(2019년) ② 한문고전 자동번역(2023년 10월 24일)
1	▪ 卿其祗復休命, 勉恢令圖, 必誠必勤而盡聽斷之道, 惟公惟正而嚴黜陟之方. ≪승정원일기≫ 영조 2년 7월 2일 경은 왕명을 공경히 받들어 좋은 계책을 널리 펴서 반드시 성실함과 부지런함으로 청송(聽訟)과 단옥(斷獄)의 도리를 다하고, 공명함과 정대함으로 인사(人事)의 방도를 엄히 하라. ① 경은 왕명을 공손히 받들어 좋은 계책을 널리 펴서 꼭 성실함과 부지런함으로 청송(聽訟)과 단옥(斷獄)의 도리를 다하고, 공명함과 정대함으로 인사(人事)의 방도를 엄히 하라. ② 상동
2	▪ 故朝家爲慮其弊, 兩人給一保, 謂之合得保, 使之責徵二疋於其保, 以充三疋之納, 然而水軍之未得合保者, 實爲强半. ≪승정원일기≫ 영조 3년 8월 20일 그래서 조정에서는 그러한 폐단을 염려하여 수군 2인에게 한 명의 보인(保人)을 붙여 합득보(合得保)라고 하고, 그 보인에게 2필을 징수하여 3필의 공납을 채우도록 하고 있지만, 수군 가운데 보인을 얻지 못한 자들이 실제로는 절반이 넘습니다. ① 그러므로 조정에서는 그 폐단을 염려하여 두 사람에게 보인(保人)을 1 보(保) 씩 주는 것을 합득보(合得保)라고 하고, 그로 하여금 보인(保人)에게 2필을 징수하여 3필(疋)의 납부를 충당하게 하였는데, 수군(水軍)에 합보(合保)를 얻지 못한 자는 실로 반이나 됩니다. ② 상동
3	▪ 且古語云, 君使臣以禮, 況大臣, 豈可强迫乎? ≪승정원일기≫ 영조 2년 5월 6일 또 옛말에 임금은 신하를 예로 부려야 한다고 하였으니, 더구나 대신을 어떻게 억지로 따르게 할 수 있겠는가? ① 또 옛말에 임금은 신하를 예(禮)로 부린다고 하였는데, 하물며 대신을 어찌 강요할 수 있겠는가. ② 상동
4	▪ 方其欲行己意之處, 則不暇察是非之分, 故事多不中, 每有顚錯之患, 當其偏主己見之時, 則不能恢聽納之道, 故言未見採, 反加詘詘之色. ≪승정원일기≫ 영조 2년 5월 16일 자신의 뜻을 행하려는 곳에서는 시비의 분별을 살필 틈이 없으므로 중도에 맞지 않는 일이 많아 매번 잘못되는 문제가 생기고, 자신의 의견을 치우치게 주장할 때에는 받아들이는 도리를 넓히지 못하므로 말을 받아들여 주지 않고 도리어 언짢은 기색을 보이십니다. ① 상동 ② 자신의 뜻을 행하려는 곳에서는 시비의 분별을 살필 틈이 없으므로 중도에 맞지 않는 일이 많아 매번 잘못되는 문제가 생기고, 자신의 의견을 치우치게 주장할 때에는 받아들이는 도리를 넓히지 못하므로 말을 받아들여 주지 않고 오히려 못마땅한 기색을 보이십니다.
5	▪ 至於經書, 則一句一字, 莫非聖賢傳心之法, 其味至深, 其義至大. ≪승정원일기≫ 영조 1년 8월 4일

	경서의 경우는 한 구절 한 글자가 모두 성현이 심법을 전한 것이니, 그 의미가 지극히 깊고 큽니다.
	① 경서(經書)의 경우, 한 구(句) 라는 글자는 모두 성현이 전수한 심법(心法)이고 그 맛이 지극히 깊으니 그 뜻이 지극히 큽니다. ② 상동
6	▪ 嚴畏分義, 强疾參坐, 此豈有因仍從仕之意? ≪승정원일기≫ 영조 2년 6월 2일 분수와 의리를 두려워하여 병을 무릅쓰고 좌기(坐起)에 참석하기는 했지만 이것이 어찌 그대로 눌러앉아 벼슬하려는 생각이 있어서였겠습니까.
	① 신하로서의 도리가 엄하고 두려워 병을 무릅쓰고 좌기에 참석하였으니, 이 어찌 그대로 관직에 종사할 생각을 하겠습니까? ② 상동
7	▪ 臣罪山積, 萬隕猶輕. ≪승정원일기≫ 영조 38년 10월 14일 미번역
	① 신의 죄는 만번 죽어도 가볍습니다. ② 신의 죄가 산처럼 쌓였으니 만번 죽어도 오히려 가볍습니다.
8	▪ 臣於昨筵, 已有所仰奏, 而此非大臣請對之事, 如是相率仰籲者, 已違按例擧行之本意矣. ≪승정원일기≫ 순조 2년 5월 12일 신이 어제 연석에서 이미 우러러 아뢰었습니다만 이것은 대신(大臣)이 청대(請對)할 일이 아닌데, 이와 같이 서로 이끌고 우러러 호소하는 것은 이미 전례를 살펴 거행하는 본뜻에 위배됩니다.
	① 신이 어제 연석에서 이미 우러러 아뢰었습니다만, 이는 대신이 청대한 일이 아닌데 이렇게 서로 이끌고 와서 우러러 호소하는 것은 규례를 살펴 거행한 본뜻을 이미 어긴 것입니다. ② 상동
9	▪ 上曰, 旣無所懷, 則何以入來? 竝命退出. ≪승정원일기≫ 영조 51년 윤 10월 29일 미번역
	① 상이 이르기를, 소회가 없다면 어떻게 들어오겠는가? 모두 물러가라고 명하였다. ② 상동
10	▪ 眇予小子, 當國家危亂之秋, 承祖宗艱大之統, 托乎兆民之上, 九年于玆. ≪숙종실록≫ 9년 2월 10일 작은 나 소자(小子)가 국가의 위태하고 어지러운 때를 당하여 조종(祖宗)의 어렵고 큰 왕통(王統)을 이어서 많은 백성들의 위에 있은 지 이에 아홉 해가 된다.
	① 보잘것없는 소자(小子)가 나라가 위태롭고 어지러운 때를 당하여 조종(祖宗)의 어렵고 큰 왕통을 계승하여 만민의 위에 가탁한 지 이제 9년이 되었다. ② 상동
11	▪ 府北二百里, 有三十里大野, 中有二大澤, 澤邊有臺, 高數百丈. ≪만기요람≫ 부(府) 북쪽 200리 지점에 30리 넓이의 큰 들이 있고, 그 가운데 큰 못 둘이 있으며, 못가에 대(臺)가 있는데 높이가 수백 길이나 된다.
	① 부(府)와 북(北) 200리에 30리 큰 들판이 있는데, 가운데에 두 개의 큰 은택이 있어 못에 대(臺)가 있고 백 장(丈)의 높이가 있습니다. ② 상동
12	▪ 逆賊父年八十者減律絶島定配, 二三歲兒應在放流者勿爲定配. ≪대전회통≫

	역적의 아비로서 나이 팔십이 된 자는 율을 감하여 절도(絶島)에 정배(定配)하고 2,3세된 아이로서 방류(放流)함에 이른 자는 정배하지 아니한다. ①역적의 아비로서 나이가 80세가 된 자는 형률을 감하여 절도에 정배하고, 2, 3세의 아이로서 응당 유배 중에 있는 자는 정배하지 말라. ②상동
13	▪新陵外假家, 大轝假家外, 竝勿掇, 馬槽等物, 亦爲留置陵所, 使其本官次知事, 分付都監畿營. ≪승정원일기≫ 영조 7년 8월 28일 새 능소 밖 가가(假家)와 대여(大轝)의 가가(假家) 외에는 모두 철거하지 말고 말구유 등의 물건도 능소에 남겨 두었다가 본관(本官)으로 하여금 차지(次知)하게 하도록 도감과 경기 감영에 분부하라. ①신릉(新陵)의 외가가(外假家)와 대여(大轝)의 가가(假家) 외에는 모두 철거하지 말고, 말구유 등의 물건도 능소에 유치(留置) 하여 본관의 지경연사로 하여금 도감의 기영(畿營)에 분부하도록 하소서. ②상동
14	▪守法不撓, 明示邦禁, 使其民知其處心行事, 一出於奉公, 而有所懲戢, 不敢以利相干, 然後可以爲法, 則民而盡心於國也. ≪승정원일기≫ 인조 13년 11월 22일 법을 지키는 데에 흔들리지 않고 나라에서 금지하는 것을 분명하게 보여 그 백성들로 하여금, 자신의 마음가짐과 일 처리가 한결같이 공무를 봉행하는 데에서 나온다는 것을 알게 하고, 징계하고 금지하는 것이 있어 감히 이익을 가지고 간여하지 못하게 한 뒤에야 법이 될 수 있으니, 그렇게 되면 백성으로서도 나라에 마음을 다하게 될 것입니다. ①법을 지키는 데에 흔들리지 않고 나라에서 금지하는 것을 분명히 보여 그 백성으로 하여금 자신의 마음가짐과 일 처리가 한결같이 공무를 봉행하는 데에서 나와 징계되는 바가 있어 감히 이익을 가지고 간여하지 못하게 한 뒤에야 법이 될 수 있으니, 그렇게 되면 백성으로서도 나라에 마음을 다하게 될 것입니다. ②상동
15	▪出入銀臺·玉署, 蔚然賢大夫名稱, 歷敭柏府薇垣, 綽有古諍臣風彩. ≪승정원일기≫ 인조 13년 7월 22일 은대(銀臺)와 옥서(玉署)에 출입하게 되어서는 성대하게 현대부(賢大夫)로 불리었고, 백부(柏府)와 미원(薇垣)을 두루 역임해서는 옛날 간쟁(諫爭)하는 신하의 풍채가 넉넉하게 있었다. ①승정원과 홍문관에 출입하면서 훌륭한 대부가 이름이 나와 백부(柏府)와 사간원을 두루 거쳤는데, 옛 쟁신(諍臣)의 풍채(風彩)가 충분히 있었습니다. ②상동
16	▪人無禮, 比之禽獸, 痛切之辭, 使一國之人風動, 一世, 皆知好禮, 可也. ≪승정원일기≫ 영조 7년 8월 9일 사람에게 예가 없는 것을 금수에 비유하여 아주 적절한 말로 온 나라 사람을 고무시키고 온 세상이 모두 예를 좋아할 줄을 알도록 하는 것이 좋습니다. ①사람의 무례(無禮)는 금수(禽獸)와 견주어 통절하고 절절한 말로 온 나라 사람으로 하여금 감화되게 하여 온 세상 사람들이 모두 예를 좋아하는 것을 알게 해야 합니다. ②상동

17	▪京外耗穀, 鬪鬨殺人, 俱由於此. 前後禁酒之請, 予亦豈不是, 而每稱迂闊而不聽, 反禁其禁者, 意蓋在焉. ≪승정원일기≫ 영조 31년 9월 10일 경외에서 곡식을 소모하고 싸우다 살인(殺人)을 하는 것도 모두 이 술에 말미암은 것이니 전후로 금주(禁酒)하자는 청을 내 어찌 옳지 않다고 여겼겠냐마는 매양 오활하다 하여 들어주지 않고, 도리어 금하는 것을 금하였으니 의도가 있어서이다.
	① 경외(京外)의 모곡(耗穀)이 싸움에 싸우다가 사람을 죽인 것이 모두 여기에서 말미암습니다. 그동안 술을 금하라는 청은 나 또한 어찌 옳지 않겠는가마는 매번 우활하다고 말하면서 들어주지 않고 도리어 금하는 것을 금한 것은 뜻이 있었다. ② 상동
18	▪不必深責之有, 然咫尺之地, 每有無嚴之擧, 考喧之察, 或事抗拒之習. ≪승정원일기≫ 순조 7년 3월 17일 미번역
	① 심하게 책망할 필요는 없지만, 지척의 거리에서 매번 무엄한 일이 있고 고훤(考喧)하는 일을 살피면서 항거하는 습속이 있기도 합니다. ② 상동
19	▪復陽又曰 : 自古人君, 行幸則凡有孝行表著之人, 必加嘉奬, 亦或有召見矣. ≪현종실록≫ 7년 4월 16일 복양이 또 아뢰기를, 예로부터 임금이 거둥할 때면 효행이 두드러진 사람에게 반드시 상을 내렸고, 불러서 만나보기도 하였습니다.
	① 양복양이 또 말하기를 예로부터 임금이 행행할 때는 효행(孝行)이 두드러진 사람이 있으면 반드시 가상히 여겨 장려하였고, 또 혹 소견(召見) 하기도 하였다. ② 양복양이 또 아뢰기를, 예로부터 임금이 행행할 때는 무릇 효행(孝行)이 두드러진 사람이 있으면 반드시 가상히 여겨 장려하였고, 또 혹 소견(召見) 하기도 하였습니다.
20	▪臣民之尤所深悲者, 十一年間, 勵精修省, 不遑寧息, 無一歲受享供之時, 無一旬不憂勞之日, 方圖至治, 未及風動, 皇天不弔, 終靳必得之壽, 抱恨鬱伊, 齎志莫伸, 斯乃我東方無窮之痛也. ≪효종실록≫〈효종행장〉 신민들이 더욱 깊이 슬퍼하는 것은, 11년 동안 정신을 가다듬고 수성(修省)하느라고 편안히 쉴 겨를이 없었던 탓으로 한 해도 향공(享供)을 받은 때가 없었고 단 열흘이라도 근심하여 괴로워하지 않은 날이 없었으며, 바야흐로 훌륭한 통치를 도모하여 미처 풍동(風動)시키기도 전에 하늘이 도와주지 않아 어진이라면 반드시 얻어야 할 수명(壽命)을 끝내 하늘이 아낌으로써 답답한 한을 품은 채 지닌 웅지를 펼 수 없게 되었다고 하는 바로 그 점이었다. 이는 그야말로 우리 동방의 끝없는 통한인 것이다.
	① 신하와 백성들이 더욱 깊이 슬퍼하는 것은 11년 동안 정신을 가다듬고 수성(修省)하여 편안히 쉴 겨를이 없었고, 한 해도 공복(供服)을 받을 때가 없어 한 달도 근심하지 않는 날이 없었고 한창 지극한 정치를 도모하려다가 미처 바람이 북받치지 않아 황천(皇天)이 불쌍히 여기지 않아 끝내 반드시 장수하려는 뜻을 이루지 못한 채 한을 품고 답답한 마음을 품은 채 뜻을 펴지 못한 것이니, 이는 우리 동방의 끝없는 아픔입니다. ② 상동

<첨부 2> 승정원 VS 바이두 성능 대조용 자료

연번	원문 ① 바이두 번역문(2019년 3월 4~10일) ② 바이두 번역문(2023년 10월 24일) ③ 한문고전 자동번역문(2023년 10월 24일)
1	▪謂一元者, 大始也. 知元年誌者, 大人之所重, 小人之所輕. 是故治國之端在正名. ≪春秋繁露≫ ① 對一元的, 非常開始的. 目的了解元年, 偉大的人是重要的, 普通人輕視. 所以治理國家的端在正确的名称. ② 상동 ③ 이일원을 말한 것은 큰 시작입니다. / 지원지원지(知元年誌) 라는 것은 대인(大人)의 소중함과 소인이 경시하는 바입니다. / 이 때문에 나라를 다스리는 단서는 정명(正名)에 있습니다.
2	▪治國者, 敬其寶, 愛其器, 任其用, 除其妖. ≪荀子≫ ① 理國家的人對他的宝貝, 愛他的工具, 任用, 除了那些妖怪. ② 治理國家的人, 尊敬的宝, 愛他的工具, 任用, 除了那些妖怪. ③ 나라를 다스리는 자는 그 보배를 공경하고 그 그릇을 사랑하여 그 용(用)을 맡기고 그 요점을 제거합니다.
3	▪子曰, 君子之事親孝, 故忠可移於君. 事兄悌, 故順可移於長. 居家理, 故治可移於官. 是以行成於內, 而名立於後世矣. ≪孝經≫ ① 孔子說, 君子侍奉双親孝順, 所以忠臣可以轉移給你. 侍奉兄長友愛, 所以李順可以轉移到長. 居家理, 所以治理可以轉移到官. 因此行成于內, 而名聲建立在后世了. ② 상동 ③ 공자(孔子)가 말하기를, 군자가 어버이를 섬기는 효도를 일삼기 때문에 충(忠)은 임금에게 옮길 수 있다고 하였습니다. / 형 섬기기를 섬김에 공손하기 때문에 순리에 옮길 수 있다. / 집안의 이치를 살았기 때문에 잘 다스려지는 것을 관직에 옮길 수 있었던 것이다. / 이 때문에 행동이 안에서 이루어지고 이름이 후세에 확립되는 것입니다.
4	▪詩人之賦麗以則, 辭人之賦麗以淫. 如孔氏之門用賦也, 則賈誼升堂, 相如入室矣. 如其不用何? ≪揚子法言≫ ① 詩人之賦麗以則, 詩人的賦麗以淫. 如孔子的門用賦的, 那么賈誼堂, 相如進入房間了. 如果不使用什么? ② 상동 ③ 시인(詩人)이 고려를 가르칠 때에는 인사(人事)를 사양하고 음탕함으로 간음합니다. / 예컨대 공씨(孔氏)의 문하에서 부(賦)를 쓰면 가의(賈誼)가 당(堂)에 오르기를 마치 입방(入室) 하듯이 하였습니다. / 그를 기용하지 않는다면 어찌하겠는가?

5	▪ 無駭帥師入極. 無駭者何? 展無駭也. 何以不氏? 貶. 曷爲貶? 疾始滅也. 始滅昉於此乎? ≪春秋公羊傳≫ ① 无駭率領軍隊進入极. 不要害怕的是什么? 展无駭啊. 爲什么不氏? 貶. 爲什么要貶低? 有病才消滅了. 開始了防在這嗎? ② 상동 ③ 놀라서 병사가 들어온 것은 극도에 달하였다. / 놀랄 것이 없다는 것은 무엇인가? / 전알하는 것은 놀라울 것이 없다. / 어찌하여 불씨가 있는가? / 폄하하였다. / 어찌 폄하되었겠는가. / 병이 비로소 없어졌다. / 처음에는 윤방을 섬멸하였는가?
6	▪ 三軍之衆, 可使必受敵而無敗者, 奇正是也. 兵之所加, 如以碬投卵者, 虛實是也. ≪孫子兵法≫ ① 全軍, 可以使必受敵而不會失敗的, 正是這樣的. 軍隊的增加, 如以碬把鷄蛋的人, 事實是這樣的. ② 상동 ③ 삼군(三軍)의 백성들이 반드시 적을 맞아서 패할 수 없게 할 수 있는 것이 기특하고 올바릅니다. / 군대가 늘어난 것은 마치 쑥으로 알을 던지는 것과 같은 것이니, 허실이 그것입니다.
7	▪ 扈江離與辟芷兮, 紉秋蘭以爲佩. 汩余若將不及兮, 恐年歲之不吾與. 〈離騷〉 ① 我身披幽香的江离白芷啊, 縫紉秋蘭作爲佩帶. 巾, 担心年齡的不与我. ② 상동 ③ 강호(江扈)와 이척(離斥), 지붕(地莫)을 꿰고, 추란초(秋蘭炒)를 꿰어 찼다. / 나를 골몰하면 장차 미치지 못할 듯이 한다라고 하였으니, 해가 나에 미치지 못할까 두렵다.
8	▪ 世之論事者, 以才高當爲將相, 能下者宜爲農商. ≪論衡≫ ① 社會上的議論事情的人, 認爲才高應爲將相, 能下的應該是農商. ② 상동 ③ 세상에서 일을 논하는 자들은 재주가 뛰어난 사람을 장상(將相)으로 삼고, 잘 내려가는 자는 농상(農商)으로 삼아야 합니다.
9	▪ 子張學干祿. 子曰, 多聞闕疑, 愼言其餘, 則寡尤, 多見闕殆, 愼行其餘, 則寡悔. 言寡尤, 行寡悔, 祿在其中矣. ≪論語≫ ① 子張學習求職. 孔子說：听說朝廷怀疑, 愼言其余, 那么我們尤其; 多見朝廷危險, 愼行其余, 那么我們后悔. 說話少過失, 行寡悔, 祿在其中矣. ② 子張學習求職. 孔子說, 博聞网怀疑, 愼言其余, 那么我們尤其, 多見朝廷危險, 愼行其余, 那么我們后悔. 說話少過失, 行寡悔, 祿在其中矣. ③ 자장(子張)이 장수록(張壽祿)을 구하였다. / 공자(孔子)는 «서경(書經)»에 의문이 많이 들었고, 신언(申言)의 그 나머지는 과오(寡誤)를 많이 드러냈는데, 빠짐없이 신행(新行) 하고 그 밖에는 후회를 적게 한다고 하였습니다. / 말이 허물이 적고 행실이 후회가 없으면 녹이 그 안에 있는 것입니다.

10	▪ 侍中裵克廉等白王大妃曰, 今王昏暗, 君道已失, 人心已去, 不可爲社稷生靈主, 請廢之. 遂奉妃敎廢恭讓. ≪太祖實錄≫
	① 侍中裵克廉等白王大妃說：現在大王昏暗, 君道已經失去, 人的心已經离開了, 不可爲國家百姓主, 請求廢除他. 于是奉妃敎廢弃恭敬謙讓.
	② 侍中裵克廉等白王大妃說, 現在大王昏暗, 君道已經失去, 人的心已經离開了, 不可爲國家百姓主, 請求廢除他. 于是, 奉太妃敎廢弃恭敬謙讓.
	③ 시중(侍中) 배극렴(裵克廉) 등의 백왕대비(白王大妃)가 말하기를, 지금 왕이 혼미하여 임금의 도리가 이미 잘못되었고 인심이 이미 떠났으므로 사직과 생령(生靈)의 군주가 될 수 없으니, 폐지하소서. / 마침내 비(妃)가 공양왕(恭讓王)의 가르침을 폐하였다.

<첨부 3> ≪승정원일기≫ 기계번역 품질 조사용 문장(비전문가 집단)

Set No.	원문/번역문
1	卿其祗復休命, 勉恢令圖, 必誠必勤而盡聽斷之道, 惟公惟正而嚴黜陟之方. 경은 왕명을 공경히 받들어 좋은 계책을 널리 펴서 반드시 성실함과 부지런함으로 청송(聽訟)과 단옥(斷獄)의 도리를 다하고, 공명함과 정대함으로 인사(人事)의 방도를 엄히 하라.
2	新陵外假家, 大轝假家外, 竝勿掇, 馬槽等物, 亦爲留置陵所, 使其本官次知事, 分付都監畿營. 새 능소 밖 가가(假家)와 대여(大轝)의 가가(假家) 외에는 모두 철거하지 말고 말구유 등의 물건도 능소에 남겨 두었다가 본관(本官)으로 하여금 차지(次知)하게 하도록 도감과 경기 감영에 분부하라.
3	故朝家爲慮其弊, 兩人給一保, 謂之合得保, 使之責徵二疋於其保, 以充三疋之納, 然而水軍之未得合保者, 實爲强半. 그래서 조정에서는 그러한 폐단을 염려하여 수군 2인에게 한 명의 보인(保人)을 붙여 합득보(合得保)라고 하고, 그 보인에게 2필을 징수하여 3필의 공납을 채우도록 하고 있지만, 수군 가운데 보인을 얻지 못한 자들이 실제로는 절반이 넘습니다.
4	人無禮, 比之禽獸, 痛切之辭, 使一國之人風動, 一世, 皆知好禮, 可也. 사람의 무례(無禮)는 금수(禽獸)와 견주어 통절하고 절절한 말로 온 나라 사람으로 하여금 감화되게 하여 온 세상 사람들이 모두 예를 좋아하는 것을 알게 해야 합니다.
5	至於經書, 則一句一字, 莫非聖賢傳心之法, 其味至深, 其義至大. 경서(經書)의 경우, 한 구(句) 라는 글자는 모두 성현이 전수한 심법(心法)이고 그 맛이 지극히 깊으니 그 뜻이 지극히 큽니다.
6	嚴畏分義, 强疾參坐, 此豈有因仍從仕之意? 신하로서의 도리가 엄하고 두려워 병을 무릅쓰고 좌기에 참석하였으니, 이 어찌 그대로 관직에 종사할 생각을 하겠습니까?
7	守法不撓, 明示邦禁, 使其民知其處心行事, 一出於奉公, 而有所懲戢, 不敢以利相干, 然後可以爲法, 則民而盡心於國也. 법을 지키는 데에 흔들리지 않고 나라에서 금지하는 것을 분명히 보여 그 백성으로 하여금 자신의 마음가짐과 일 처리가 한결같이 공무를 봉행하는 데에서 나와 징계되는 바가 있어 감히 이익을 가지고 간여하지 못하게 한 뒤에야 법이 될 수 있으니, 그렇게 되면 백성으로서도 나라에 마음을 다하게 될 것입니다.
8	方其欲行己意之處, 則不暇察是非之分, 故事多不中, 每有顚錯之患, 當其偏主己見之時, 則不能恢聽納之道, 故言未見採, 反加詘詘之色. 자신의 뜻을 행하려는 곳에서는 시비의 분별을 살필 틈이 없으므로 중도에 맞지 않는 일이 많아 매번 잘못되는 문제가 생기고, 자신의 의견을 치우치게 주장할 때에는 받아들이는 도리를 넓히지 못하므로 말을 받아들여 주지 않고 도리어 언짢은 기색을 보이십니다.
9	出入銀臺・玉署, 蔚然賢大夫名稱, 歷敭柏府薇垣, 綽有古諍臣風彩. 승정원과 홍문관에 출입하면서 훌륭한 대부가 이름이 나와 백부(柏府)와 사간원을 두루 거쳤는데, 옛 쟁신(諍臣)의 풍채(風彩)가 충분히 있었습니다.
10	且古語云, 君使臣以禮, 況大臣, 豈可强迫乎? 또 옛말에 임금은 신하를 예(禮)로 부린다고 하였는데, 하물며 대신을 어찌 강요할 수 있겠는가.