

# 법률 도메인 특화를 위한 중한 기계번역 오류의 귀납적 분석: 중국 민법전의 파파고 번역을 중심으로

최 승 혁  
(서강대)

## 1. 문제 제기

본 연구는 법률 도메인 특화 기계번역을 위해 중국 민법전을 중심으로 중한 법률 텍스트 기계번역에서 나타나는 오류를 귀납하고 공학 연구자와의 협력 모색을 목적으로 한다.

2016년 구글 신경망 기계번역(GNMT)이 등장한 이후 10여 년이 지난 지금, 기계를 통한 빠르고 정확한 번역으로 외국어 의사소통과 텍스트 수용에 있어 장벽이 크게 낮아졌다. 외국어로 된 정보의 검색이나 여행 등 일상생활에서 구글 번역(Google Translate), 네이버 파파고(Naver Papago), 바이두(Baidu) 등 기계번역은 컴퓨터와 모바일에서 광범위하게 활용되고 있다. 기계번역<sup>1)</sup> 사용자가 늘어나면서 데이터가 방대해지고 이 데이터가 다시 기계번역의 학습용 데이터로 사용되고 있다. 이러한 순환은 데이터 증강 외에도 기계번역 결과물의 품질

1) 본고에서 말하는 기계번역은 모두 신경망 기계번역(NMT)을 가리킨다.

을 향상하는 데 도움을 주고 있다.

사용자의 일상에는 단순한 대화나 메시지 교환 외에도 금융, 의료, 법률 등 보다 전문적인 분야도 포함된다. 하지만 이러한 분야에서 사용하는 언어는 일상 언어와 다르기 마련이다. 전문분야에서 기계번역을 활용하려면 여전히 많은 문제가 존재한다. 가장 큰 문제는 해당 분야에서 사용하는 전문용어가 정확하게 번역되지 않는다거나 어순, 조사 사용 등의 오류와 같이 도착언어의 언어 규범에 맞지 않아 어색하고 이해하기 어려운 문장이 도출되는 것이다. 본 연구는 다음과 같은 연구 문제에서 출발하였다.

첫째, 범용 기계번역의 성능이 개선되어 수용도가 높아진 지금, 의료, 금융, 법률 등 전문분야에서도 기계번역을 수월하게 사용할 수 있을까? 그렇지 않다면 어떤 솔루션이 있을까? 이에 대해 도메인 특화 신경망 기계번역(Domain Specific NMT)을 제안한다.

도메인(Domain)의 사전적 의미는 지식과 활동의 영역이나 분야이다. 본고에서 말하는 ‘도메인’은 출발언어 텍스트(Source Text, 이하 ST)가 사용되는 사용역(register)를 의미한다. 사용역은 언어의 사용과 관련한 변이로 종교, 법률, 신문, 의학, 전문어 등 직업 분야에서 나타난다(Trosborg 2016: 34). 번역자는 언어 변이에 대한 판단과 선택을 통해 텍스트를 효율적으로 번역할 수 있다. 사용역 관점에서 본고는 도메인을 범용 도메인과 특정 도메인으로 구분한다.

특정 도메인의 가장 큰 특징은 전문용어의 비중이 높다는 점이다. 전문용어가 포함된 텍스트를 정확히 번역하려면 도메인 지식이 필요하다. 여기서 도메인 지식은 보편지식의 의미와 상반되는 전문적 학문이나 분야의 지식을 뜻하며, ‘도메인 특화’는 특정 도메인 지식을 Pretrain-Finetuning Approach(PFA)와 같은 전이학습을 통해 기계번역에 학습시키는 것을 말한다(김세린, 장칭하오와 권혁철 2022: 678). 법률은 법률 언어로 이루어진 대표적인 특정 도메인 텍스트로 명사나 명사구 형태의 전문용어, 긴 문장, 복잡한 통사구조 등이 사용되는 특징을 가진다(이해운 2020: 39-54).

둘째, 여러 전문분야 가운데 법률에 관한 특화가 필요한 이유는 무엇인가? 이에 대해 법률 번역 수요의 증가에서 대답을 찾을 수 있다. 현재 외국계 로펌이나 번역 에이전시에서는 국내 법률 번역 매출액 규모를 연간 2,000~3,000억 원 정도로 추산한다.<sup>2)</sup> 이뿐 아니라 중국에서 상표 분쟁과 같은 법률 분쟁 증가

하고 있고<sup>3)</sup> 국내 기업 중국법인의 매각이나 청산 건수가 6년간 46곳에 달할 정도로 탈중국화 속도가 빨라지면서<sup>4)</sup> 이에 필요한 법률 번역의 수요도 늘어날 것으로 예상할 수 있다.

셋째, 중한 법률 기계번역의 활용도와 수용도가 아직 낮다면 번역학 관점에서 문제의 해결을 위해 어떤 역할을 할 수 있을까? 이에 대한 대답으로 법률 텍스트 기계번역에서 출현하는 오류 항목을 귀납하여 공학 연구자들에게 제시함으로써 공학과 번역학의 협업을 촉진하고자 한다.

기계번역은 기계적 메커니즘을 통해 언어 간 전환을 처리하는 분야로 공학과 번역학은 밀접한 연관성을 갖는다. 하지만 현실에서 공학 전공자는 언어적 특징에 대한 인식이 부족하고, 반대로 언어 전공자는 기계번역의 원리를 이해하지 못하는 등 괴리가 존재하는 것이 사실이다(남철진 2019: 2). 이는 개별 전공 분야가 서로를 배타적으로 대하는 것이 아니라 현실적인 어려움에서 기인한 역할상의 괴리로 볼 수 있다. 다시 말해, 언어 전공자가 기계번역과 관련한 공학적 지식을 단시간에 이해하고 적용하는 것은 불가능하며 그 반대도 마찬가지이기 때문이다. 이런 간극을 좁히기 위해 공학적 지식을 잘 이해하도록 언어학자들에게 기술적 개념을 설명하려는 매우 고무적인 시도가 있었다(김희동과 지인영 2017). 따라서 가장 현실적인 대안은 공학 전공자는 번역 엔진의 기계적 메커니즘을 조정하고, 어학이나 번역학 전공자는 오류가 적은 양질의 데이터를 제공하며 역할의 접점을 찾는 것이다(이준호 2022: 171).

이러한 관점에서 번역학 전공자는 공학 전공자에게 ‘기계적 메커니즘으로 이런 부분이 바뀌었으면 좋겠다’라는 오류 수정 항목을 제시할 수 있다. 하지만 기존의 범용 기계번역의 오류에 관한 많은 평가와 분석은 연역적 방법을 취하고 있다(권도경 외 2019; 기유미 2018; 김나영 2021; 류수린, 국지연과 정동규 2021; 서보현과 김순영 2018 등). 즉, 평가와 분석 기준을 사전에 설정하고 해당 기준에 부합하지 않는 오류를 찾아 정리한 후에 정확성, 충실성, 가독성을

2) 법률신문(2019. 10. 24.)

3) 한국지식재산보호원에 따르면, 2019년 기준 중국에서 한국 기업의 상표 출원은 720건, 등록은 1,200건이며, 현재 이의신청 중인 건은 106건, 무효 심판 중인 건도 30건에 이른다.

4) 한국경제(2023. 7. 30.)

논하는 것에 그치고 있다. 연역적 방식의 평가나 분석에서 나타나는 공통적 문제로는 오류 관련 예문에서 한가지 오류만을 제시한다는 점이다. 하지만 실제로 기계번역 결과물을 분석하다 보면 하나의 분석 단위에서 한가지 오류만 출현하는 것이 아니라 둘 또는 그 이상의 오류가 복합적으로 나타나기도 한다. 아래 예문을 보자.

ST: 第一百三十四條 民事法律行爲可以基于双方或者多方的意思表示①一致成立, ②也可以基于③單方的意思表示成立。法人④、非法人組織依照法律或者章程規定的議事方式⑤和表決程序作出⑥決議的, ⑦該決議行爲成立。

MT-P: 제134조 민사법률행위는 쌍방 또는 다자간 의사표시에 기초하여 성립하거나 일방적 의사표시에 기초하여 성립할 수 있다. 법인 또는 비법인 조직이 법률 또는 정관에 규정된 의사 방법 및 의결 절차에 따라 결의를 하는 경우 결의행위가 성립한다.5)

상기 예문을 보면 ST의 밑줄 친 부분이 MT-P에서 ① 누락 ② 논리적 연결 오류 ③ 어휘 해석 오류 ④ 문장부호 ⑤ 어휘 오류 ⑥ 어휘 해석 오류 ⑦ 누락 이렇게 총 7가지 오류로 복합되어 나타났다. 이러한 문제를 해결하기 위해 일부 연구에서는 항목별로 가중치를 부여해 번역 결과물의 정확성을 평가하였다. 하지만 어떤 항목이 더 중요하고 덜 중요한 것으로 오류의 중요도를 설정하는 것은 주관성이 높고 설득력이 낮다.

이에 보고는 중한 법률 특화 기계번역의 구현을 위해 귀납적 방식을 통해 중한 법률 기계번역에서 출현하는 오류 가운데 개선이 시급한 주요 오류들을 귀납함으로써 공학 전공자들이 오류 수정 항목들을 참고하도록 제시할 것이다.

5) 2023. 3. 15. 파파고 중국어(간체) → 한국어 기계번역 실시

## 2. 선행연구

### 2.1 기계번역 오류에 관한 연구

기계번역의 결과물에 관하여 ‘번역 품질 평가(TQA)’, ‘성능 평가’, ‘오류 분석’ 등을 주제로 이미 많은 연구가 진행되었다.<sup>6)</sup> 비록 축사적 표현은 다르지만, 기계번역이 출발언어의 언어적 규범, 요소, 관습 등을 얼마나 정확히 이해하여 도착언어로 전환하는지에 대한 분석이라는 점에서 연구의 방향과 취지는 유사하다. 번역학에서 ‘기계번역 평가’에 관한 논문은 이미 많은 연구에서 정리하였으므로(최문선 2019: 284-286) 본 소절에서는 중한 기계번역 오류에 관한 선행 연구를 중심으로 살펴본다.

강병규와 이지은(2018)은 범용 기계번역인 구글, 네이버 파파고, 바이두의 신경망 기계번역을 활용해 의사소통 기본표현과 실용문에 대한 기계번역의 정확률을 평가하였다. 더 나아가 텍스트 유형에 따른 기계번역의 선택적인 활용과 특화된 신경망 기계번역의 개발과 활용 등 기계번역의 효율적인 활용방안을 제시했다(강병규와 이지은 2018: 286-287). 이는 단순한 정확률 측정을 넘어서 기계번역 발전 방향에 대한 시사점을 제시한 연구이다. 권도경 외(2019)는 중국의 연설문, 논설문, 기사문, 대화체, 수필을 선정해 구글과 네이버 파파고 기계번역의 오류 유형을 비교, 분석하였다. 이 다섯 가지 출발언어 텍스트는 뚜렷한 유형적 특징을 가지며, 텍스트 유형별로 오류가 상이하게 나타났다. 해당 연구는 기계번역에서 출발언어 텍스트의 유형이 갖는 중요성을 강조하였다. 임형재와 왕침(2018)은 문화번역이론 관점에서 중국 외교연설문을 대상으로 구글, 파파고, 바이두 번역기에서 문화소 처리 양상을 분석하고, 기계번역이 번역 누락, 문법적 오류, 맥락 오류 등과 같이 언어 외적 요소의 처리에 있어 취약성을 지

6) 본고는 기계번역 결과물에서 나타나는 문제점에 초점을 두고 ‘오류 분석’이라는 표현을 사용한다. ‘번역 품질’은 무엇이 좋은 번역인가에 대한 거대 담론이다. ‘번역 품질 평가’는 문학작품을 포함하여 번역 학습자의 능력 평가, 번역상의 오류, 통역의 품질 평가를 아우르는 광의의 개념으로(이향 2010: 108-109) 대부분 점수를 부여하는 방식을 취한다. ‘성능 평가’는 논문 제목에서 ‘기계번역 엔진의 성능 평가’와 같이 사용되어 마치 기계번역의 공학적 메커니즘의 문제점과 개선방안을 연상시키지만 결국 기계번역 결과물에서 나타나는 오류 분석인 경우가 많다.

적하였다. 김혜립(2021)은 네이버 파파고를 이용해 ‘中華人民共和國個人所得稅法(중화인민공화국 개인소득세법)’을 한국어로 번역한 다음 포스트에디팅(MTPE)을 실시하였다. 그 결과, 기계번역 번역문은 법률 언어와 법률 번역의 특징을 반영하지 못했으며, 정확성, 완결성, 일관성, 어휘 측면에서 복합적인 문제점을 드러냈다. 해당 연구는 비록 MTPE 관점에서 진행되었지만, 번역의 관점을 넘어서서 법률 언어와 법률 텍스트의 고유한 특징을 종합적으로 고려해야 정확한 번역이 가능하다는 메타적 인식을 제시하였다는 점에서 중요한 의미를 지닌다. 채종린(2022)은 네이버 파파고로 대만 민법을 번역한 다음 어휘와 문장 층위에서 오류 유형을 분류하였다. 대만 법률을 분석 대상으로 한 것은 충분한 희소성이 있다. 하지만, 어휘 차원에서 법률 텍스트의 특성에 따라 법률 전문용어의 오류인지, 일반 어휘의 오류인지 구분이 미흡하고, 문장 차원에서도 오역 유형을 상세하게 제시하지 않았다.

이 밖에도 기계번역의 정확성 제고 관점에서 중국어 복문(夏句)처리 양상을 분석한 연구(한양양 2019), 문학 텍스트인 뻬수민(畢淑敏) 단편소설 『1cm』의 기계번역에서 나타나는 어휘 처리 양상을 분석한 연구(賈楊帆 2019), 기계번역에서 전문용어 처리와 정확한 대응쌍 구축의 필요성을 강조한 연구(최혜령 2020) 등이 있다.

상기 연구를 살펴보면, 분석 방법에 있어서 신경망 기계번역 등장 초기에 둘 이상의 범용 기계번역 플랫폼에서 얻은 결과물을 비교하는 방식이었다. 분석 기준 측면에서는 ‘정확성’을 기준으로 내세웠다. 이후에 다양한 언어 모델이나 메커니즘의 개발에 따라 기계번역의 성능이 크게 향상되면서 어휘, 통사, 문화소 등 분석 대상과 관점에 대한 미시적 접근이 이루어졌다.

## 2.2 도메인 특화 기계번역에 관한 연구

도메인 특화는 전술한 바와 같이 특정 도메인의 언어적 특징이나 도메인 지식을 기계번역에 학습하는 과정을 수반한다. 이러한 관점에서 도메인 특화 기계번역 및 법률 텍스트를 중심으로 한 도메인 특화 기계번역에 관한 선행연구를 전반적으로 살펴보자.

강병규(2006)는 중국어 전문 서적(專著)에 대한 중한 번역의 정확도를 높이

기 위해 코퍼스 언어학과 컴퓨터 보조 번역(Computer Aided Translation, CAT)을 활용해 고빈도 어휘와 정형 표현 등을 추출하였으며, 해당 데이터를 CAT Tool에 적용해 자동화 처리하는 방안을 연구하였다. 신경망 기계번역이 없었던 시절에 이루어진 연구로 도메인 특화의 관점에서 번역에 기계를 활용한 초기 연구이다. 해당 연구는 전문 서적과 같이 전문성이 높은 특정 분야에서 고정적이고 관습적인 표현의 추출이 번역 효율에 미치는 영향에 대한 중요한 시사점을 주었으며, 공학적 관점과 언어학적 관점에서 균형을 이루고 있다. 최승권 외(2007)는 공학적 측면에서 영한 특허에 관한 자동번역 시스템 구축을 통해 도메인 특화의 초기 단계 방법을 제시하였다. 이후 최승권 외(2012)는 기술문서를 대상으로 번역 사전(dictionary) 튜닝, 패턴 확장, 정답 셋, 번역률 자동 측정 방법 BLEU 도입 등 초기의 도메인 특화보다 개선된 방법을 소개하였다. 박찬준 외(2020)는 COVID-19라는 특정 도메인에 대해 Seq2Seq 모델을 이용해 도메인 특화 번역 모델을 개발해 성능 면에서 구글 번역 엔진보다 우수한 결과를 얻었다.

상기 선행연구의 흐름을 보면 국내에서도 도메인 특화 기계번역에 관한 관심과 연구가 꾸준히 증가하고 있음을 확인할 수 있다. 초기에는 코퍼스 데이터를 중심으로 한 기계번역이 주를 이루었다. 이후 신경망 기계번역의 등장과 트랜스포머 메커니즘을 적용한 다양한 언어 모델이 고안되면서 도메인 특화 기계번역과 관련하여 새로운 언어 모델의 개발, 병렬 데이터의 증강, 특정 도메인 고유의 언어적 특징에 대한 전이학습 등 다양한 연구가 시도되고 있다.

국내에서 법률 도메인 특화 기계번역에 관한 연구로는 이진철(2021), 이준호(2022), 이준호, 이청호와 최효은(2023), 이지은과 최효은(2023)이 있다.

이진철(2021)은 NMT-PT(Neural Machine Translation - Patent Translation) 모델을 구축하고 특허 데이터를 학습시킨 후에 학습 전후 모델과 범용 기계번역(Google Translate, Naver Papago)의 성능을 비교하였다. 그 결과 특정 도메인에서의 대량의 데이터를 통한 추가학습(fine tuning)이 번역 성능을 개선하는 데 효과적이라는 점을 검증하였다. 특허는 넓은 의미에서 법률 도메인에 속하고, 그 뚜렷한 언어적 특징으로 인해 관련 선행연구가 많은 편이다. 해당 연구를 통해 특정 도메인 지식 학습으로 번역 품질이 개선될 수 있음을 예상해 볼 수 있다.

이준호(2022)는 전문 번역업체(시스트란)의 법률 특화 번역 엔진을 활용해 계약서(법률 문서)의 한영 기계번역 결과를 분석하였고, 전체 평가항목에서 구글 번역보다 우수한 결과를 얻었다. 이준호, 이청호와 최효은(2023)은 자동화된 기계학습(AutoML)과 구글 번역을 비교해 AutoML이 도메인 특화 기계번역의 유용한 도구로 활용할 수 있다는 가능성을 시사하였다. 이는 기존의 연구와 차별성을 이루는 유의미한 방법론으로 참고의 가치가 크다. 특히, 기계번역의 실제 사용자가 각자의 업무상 필요에 따라 기계번역 엔진을 맞춤형(Customizing)으로 설계할 수 있음을 보여준 굉장히 실용적인 연구이다. 해당 연구는 실험 데이터가 적어 일반화하기에는 어렵다는 한계가 있지만, 국내에서 도메인 특화 엔진을 직접 사용해 분석했다는 점에서 의미가 매우 크다. 도메인 특화 기계번역에서 엔진 개발의 공학적 측면과 번역학적 협업(이준호 2022: 190) 필요성을 실천적으로 보여준 중요한 사례로 평가할 수 있다.

이지은과 최효은(2023)은 범용 기계번역을 활용한 연구에서 벗어나 법률과 특허에 특화된 NMT 기반의 자동 번역 엔진인 오토란(Otran)<sup>7)</sup>을 활용해 법률과 특허의 한영번역 결과를 분석했다. 평가에는 자동평가와 수동평가를 모두 사용하였다. 법률은 두 가지 평가에서 모두 양호한 점수를 받았지만, 특허는 비교적 낮은 점수를 받았다. 특히, 정확한 용어의 사용에 있어서 개선의 필요성이 제기되었다. 해당 연구는 비록 특정 도메인이라고 해도 사용역에 따라 번역 품질에 차이가 존재한다는 점을 보였고, 도메인 특화에 있어서 용어 학습이 중요하다는 사실을 밝혔다는 점에서 중요한 시사점을 준다. 무엇보다 범용 기계번역과 비교해 도메인 특화 기계번역의 성능이 비교 우위에 있음을 확인하면서 향후 기계번역의 발전과 연구의 방향을 제시하였다.

7) 시스트란(systran)의 번역 시스템에 AI 학습을 통해 법률 도메인에 특화된 법률 번역 서비스



### 3. 연구 방법

#### 3.1 민법전 및 파파고 기계번역 선정 이유

본고는 네이버 파파고 기계번역을 활용해 중화인민공화국 민법전(中華人民共和國民法典, 이하 민법전) 1,258개 조항(부칙 제외)을 번역했을 때 나타나는 오류를 귀납, 분석하였다.

민법전을 분석 대상으로 선정한 이유는 다음과 같다. 한 국가의 법은 크게 헌법, 민법, 형법으로 나눌 수 있는데, 민사(民事)는 국민의 삶에 직접적으로 영향을 끼치는 중요한 부분이다. 중화인민공화국 민법전은 법전의 형태로 명명된 중국 최초의 법률로 상징적 의미가 크다. 또한, 민사 사법과 민사 활동의 근거가 되는 기본법인 동시에 민법전을 기반으로 다양한 하위 법률이 파생된다. 이에 민법전 분석은 거시적 차원에서 법률 도메인 특화를 위한 기계번역의 오류를 귀납하는데 적절한 텍스트라고 판단하였다. 민법전은 총 7편(각 편은 총칙, 물권, 계약, 인격권, 혼인가정, 상속, 권리 침해에 관한 책임, 부칙), 1,260개 조항으로 이루어져 있다. 부칙을 제외하고 1,258개 조항을 분석하였다.

이뿐 아니라 민법전은 법률 텍스트로서 모호한 의미를 나타내는 어휘(其他, 等),<sup>8)</sup> 독특한 통사적 요소(的字短語), 반복적으로 출현하는 정형 표현(有下列情形之一的) 등이 사용된다(최승혁 2023). 이는 언어학적인 측면에서 일반 텍스트와 구별되는 차이점을 형성한다.

분석에서 범용 기계번역 플랫폼으로 네이버 파파고를 사용한 이유는 아래와 같다. 첫째, 2023년 기준 월간 활성사용자수(MAU)가 1,400만 명으로 2018년 6월 이후 국내 번역 서비스 1위를 유지하고 있다.<sup>9)</sup> 따라서 외국어에서 한국어로의 번역 데이터가 구글 등 다른 범용 기계번역 플랫폼보다 더 풍부하다. 둘째, 특정 서비스 및 업계에서 사용하는 고유한 용어를 일관성 있게 번역하도록 도와주는 커스텀(custom) 용어 사전 API를 제공한다. 법률 번역에서 용어의 일관성은 정확성과 함께 매우 중요한 요소이다. 커스텀 용어 사전 기능을 제공하

8) 언어 사용 측면에서 법률 언어는 ‘모호성’을 원칙으로 한다. 입법자는 제한된 법률만으로 발생 가능한 모든 법적 사실을 포괄할 수 없다. 이러한 이유로 입법자는 모호한 표현을 사용하게 되고 법률은 모호성을 갖게 된다.

9) 전자신문(2022. 8. 31.)

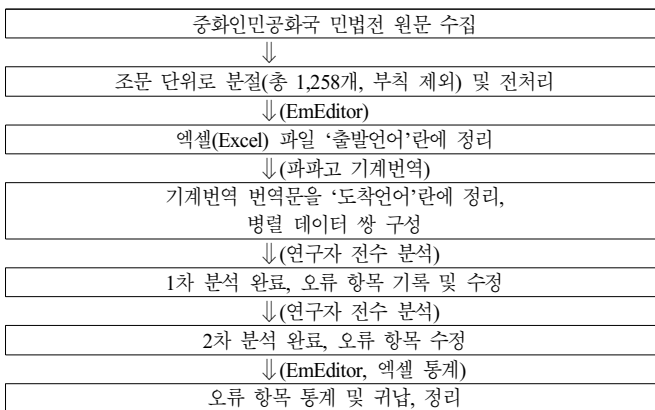
는 점에서 법률 전문용어의 일관성과 정확성 평가에 적합하다. 셋째, 파파고는 사용자로부터 번역에 관한 오류나 수정 의견을 받아 개선해 왔다. 최근에 품질 평가 과정을 자동화한 QE 모델을 적용해 번역 품질을 계속 개선하고 있다.<sup>10)</sup>

이처럼 방대한 사용자 규모, 전문용어 사전 제공, 지속적인 번역 품질 개선이 범용 기계번역 플랫폼으로서 대표성이 충분하다고 판단하였다.

### 3.2 오류 분석 절차

오류 분석 절차는 다음과 같다. 먼저 온라인에 공개된 중화인민공화국 민법전 원문을 수집하였다. 문서 편집기 EmEditor로 불필요한 행 삭제, 조항별 행 구분 등 간단한 전처리를 한 후에 엑셀(Excel) 파일에 출발언어로 입력한다. 각 조항을 파파고에서 한국어로 번역하고, 원문 옆에 도착언어로 입력해 출발언어와 병렬 데이터 쌍을 구성하였다(2023. 3. 15). 이렇게 구성한 데이터를 한 달간(2023. 3. 20. ~ 2023. 4. 20.) 일일이 비교, 대조하면서 오류 항목을 기록, 수정하였고 이것이 1차 분석이다. 1차 분석을 마친 후에 다시 한 달간(2023. 4. 21. ~ 2023. 5. 21.) 민법전 관련 서적, 논문을 참고해 2차 분석을 진행하면서 분석의 정확성을 높였다. 이 과정을 아래 <표 1>과 같이 정리할 수 있다.

<표 1> 민법전 기계번역 오류 분석 절차



10) ZDNET Korea (2023. 2. 28.)

앞에서 기계번역 오류 분석에 관한 선행연구는 대부분 사전에 분석 기준을 설정하고 그에 따른 정량적, 정성적 분석을 진행하는 연역적 방식이 위주였음을 지적했다. 다시 말해, 사전에 기준을 제시하고 그 기준에 어긋나면 감점을 하는 방식이었다. 하지만 하나의 분석 단위에 하나의 오류만 출현하는 것은 아니며 언어적 요소에 가중치를 부여하는 기준과 방법도 모호하고 상이하다.

본 연구는 오류 항목을 분석하기는 하지만 점수를 부여하거나 충실성과 가독성과 같은 번역평가의 상위 개념으로 접근하지 않음을 재차 강조한다. 기존 연구와 차별을 두기 위해 귀납적 방식을 취하여 분석 과정에서 나타나는 오류를 의미, 언어와 문법, 구조 세 가지 층위로 귀납한 다음 상세한 오류 항목을 제시할 것이다. 이것이 법률 기계번역의 오류를 줄이기 위해 번역 연구자가 저야 할 1차 역할이라고 생각한다. 이렇게 제시된 상세 오류 항목을 기계번역 언어 모델에 적용해 엔진 성능을 높이는 것은 공학 연구자나 엔지니어의 2차 역할일 것이다.

### 3.3 오류 유형 귀납 기준

본고에서 오류 항목 분석의 목적은 오류 유형을 정리하고 귀납하는 것이 아니지만, 해당 상세 오류 항목을 단순히 나열하는 것보다 그것들이 어떤 층위에 속하는지 간단히 제시하고자 한다.

이를 위해 본고는 김혜림(2023)에서 제시한 ‘중한 포스트에디팅 가이드라인 수정안’을 따른다. 김혜림(2021)은 윤미선 외(2018)의 포스트에디팅 가이드라인을 참고해 중국어 법령의 중한 기계번역 포스트에디팅에 적용하였다. 이후 김혜림(2023)에서는 기계번역 품질 평가 기준에 대한 평가자의 인식을 반영해 김혜림(2021)의 가이드라인을 정교하게 수정하였다. 해당 가이드라인은 법령 중한 번역에서 나타나는 오류를 유형화한 후에 수정 방안을 제시하였으며, 본고의 상세 오류 항목을 모두 커버한다는 점에서 귀납의 근거로 충분하다고 판단하였다.

본고의 상세 오류 항목과 ‘수정된 중한 포스트에디팅 가이드라인 수정안’을 아래 <표 2>와 같이 대응시켰다.

〈표 2〉 수정된 중한 포스트에디팅 가이드라인(김혜림 2023: 389-390)과 상세 오류 항목 대응

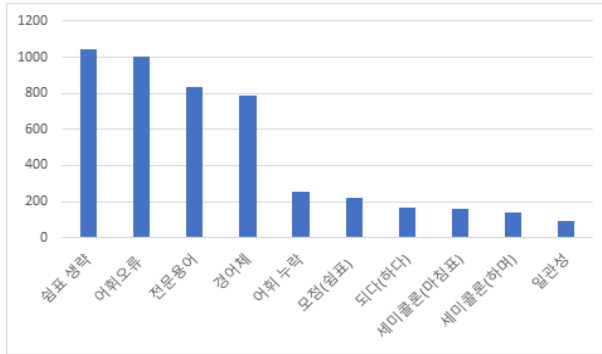
항목	세부항목	가이드라인
의미	정확성	<ul style="list-style-type: none"> <li>원문에서 전달하고자 하는 의미를 정확히 전달해야 한다. (<b>주체-대상</b>)</li> <li>원문의 중국어 문장부호 용법에 주의하여 의미를 정확히 전달해야 한다. (<b>문장부호</b>)</li> <li>원문의 절과 절 사이의 논리 관계가 적절한 한국어 연결표지로 제시되어야 한다.</li> <li>원문의 상황에 적절한 한국어 시제로 제시되어야 한다.</li> </ul>
	완결성	<ul style="list-style-type: none"> <li>의도적으로 원문의 정보를 누락하거나 추가해서는 안 된다. (<b>어휘 누락 및 추가</b>)</li> </ul>
	일관성	<ul style="list-style-type: none"> <li>텍스트 전체에서 용어나 표현 등이 통일성 있게 제시되어야 한다. (<b>어휘 사용 일관성</b>)</li> <li>텍스트의 응결성을 고려하여 내용이 유기적이면서 논리적으로 록 한다.</li> </ul>
언어와 문법	어휘	<ul style="list-style-type: none"> <li>맥락에 적합한 어휘를 사용해야 한다. (<b>법률 어휘</b>)</li> <li>함께 자주 쓰이는 어휘, 관용구와 숙어, 문화적 용어 등의 사용이 올바르게 제시되어야 한다. (<b>정형 표현</b>)</li> <li>해당 분야에서 사용하는 전문용어를 사용한다. (<b>전문용어</b>)</li> <li>텍스트를 읽는 독자의 목적과 상황을 고려한다.</li> <li>한국 한자음대로 번역된 어휘가 맥락에 적합해야 하며, 함께 자주 쓰이는 어휘, 전문용어 등 사용이 올바르게 제시되어야 한다. (<b>정형 표현</b>)</li> </ul>
	통사 구조	<ul style="list-style-type: none"> <li>조사의 사용과 문장성분 호응이 올바르게 제시되어야 하는 등 문장 구조에 어색함이 없어야 한다. (<b>피동, 도치, 통사 성분</b>)</li> </ul>
	맞춤법	<ul style="list-style-type: none"> <li>오타자가 없고 띄어쓰기가 정확하게 되어야 한다.</li> </ul>
	문장부호	<ul style="list-style-type: none"> <li>문장부호 등에 오류가 없어야 하고 올바르게 사용되어야 한다.</li> </ul>
구조	문체	<ul style="list-style-type: none"> <li>언어사용역(register)에 적합한 문체를 사용해야 한다. (<b>중결 어미</b>)</li> <li>보고서는 명확한 정보 전달이 이루어질 수 있도록 만연체 문장을 간결체로 수정한다.</li> <li>연설문은 독자 설득을 위한 강조, 반복 등 수사적 효과에 유의하여야 한다.</li> <li>한국어로 된 유사한 텍스트에서 쓰이는 표현이나 어말어미로 수정한다. (<b>중결 어미</b>)</li> </ul>
	형식	<ul style="list-style-type: none"> <li>들여쓰기, 여백, 제목 및 소제목 설정 등에 문체가 없어야 한다.</li> <li>폰트, 글자크기, 볼드체, 밑줄설정, 각주, 목차, 괄호사용 등을 원문과 비교하여 동일한 효과를 내도록 한다.</li> </ul>

## 4. 분석 결과

### 4.1 민법전 기계번역의 오류 통계 분석

민법전 기계번역에서 나타나는 오류를 귀납하기 전에 오류 항목과 빈도 등을 먼저 전체적으로 고찰해보자. 아래 <그림 1>은 민법전 기계번역에서 나타난 상위 10가지 상세 오류를 출현 빈도 순서로 나타낸 그래프이다. 문장부호(씽표 생략), 어휘, 종결어미, 일관성 등 매우 다양한 오류가 나타났음을 알 수 있다.

<그림 1> 민법전 기계번역 상위 10가지 상세 오류 항목



출현 빈도가 가장 높은 상세 오류 항목은 문장부호 중 씽표의 생략이었다. 민법전에는 씽표 외에도 모점(頓號), 세미콜론, 콜론 등 문장부호가 사용되는데, 이는 의미의 적절한 분절이나 법률의 서술 구조와 관련이 있다. <그림 1>을 보면 씽표 관련 오류 외에도 문장부호와 관련한 상세 오류 항목이 매우 다양한 양상으로 출현했다.

두 번째로 출현 빈도가 높은 것은 법률 어휘 및 전문용어 등 어휘 사용 오류였다. 어휘 사용 오류 출현 빈도가 높다는 것은 어휘 데이터 학습이 강화되어야 함을 보여주는 대목이다. 특히, 전문용어는 사용역에 따라 다양한 변이를 보이는 요소이자 도메인의 특성을 드러내는 중요한 언어적 자질이다. 법률 전문용어는 ‘일관성’을 중요시하는 법률 언어의 특징으로 인해 어휘의 의미 투명도가 높다. 따라서 1:1 대응쌍 DB를 구축하기 용이하므로 이를 활용하면 어휘 관

런 오류는 쉽게 해결할 수 있다.

세 번째로는 종결어미 오류인 경어체의 빈도가 높게 나타났다. 한국어는 어미가 발달한 언어로 법률 텍스트에서는 모두 ‘-하다’ 형태의 종결어미를 사용한다. 하지만 파파고 번역에서 ‘-합니다’, ‘-습니다’의 경어체로 전환되는 빈도가 높는데, 이는 한국어 법률 텍스트의 문체적 특징과 맞지 않는 부분이므로 개선이 필요하다.

민법전 기계번역에는 전체적으로 총 135가지 오류 항목이 출현했다. 하지만 출현 빈도가 10회 이하인 것이 104가지로 전체의 77%를 차지하였다. 세세한 오류 항목들을 모두 귀납하고 열거하는 데는 한계가 있으므로 <표 2>의 기준에 따라 의미, 언어와 문법, 구조 층위에서 출현 빈도가 높은 주요 오류 항목 중심으로 살펴볼 것이다. 민법전 중한 기계번역에 출현한 주요 상세 오류 항목을 <표 3>으로 정리하였다.

<표 3> 민법전 기계번역에 출현한 상세 오류 항목

오류 유형 귀납	상세 오류 항목 <sup>11)</sup>
의미 층위	문장부호(쉼표 생략 1,043), 어휘 누락(257), 어휘 사용 일관성(95), 어휘 추가(42), 주체-대상(28) 등
언어와 문법 층위	법률 어휘(1,004)와 전문용어(835)의 부정확한 사용, 불필요한 피동(166), 도치(73), 통사 성분의 위치(58), 정형 표현(26) 등
구조 층위	종결어미(786)

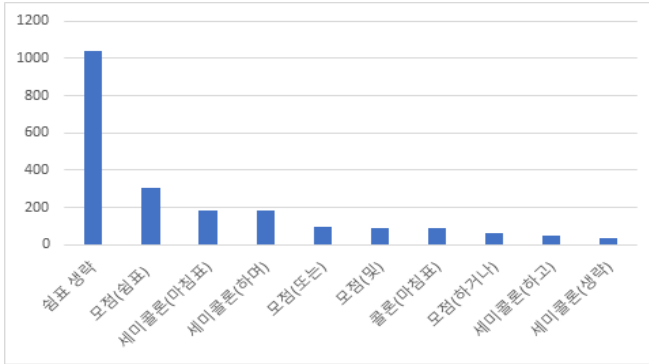
## 4.2 의미 층위의 상세 오류 항목

### 4.2.1 문장부호

민법전에는 문장부호가 총 5,164회 사용되었다. 사용 빈도에 따라 쉼표(3,115) > 마침표(2,150) > 모점(1,399) > 세미콜론(555) > 콜론(95) 순이었다. 쉼표, 모점, 세미콜론, 콜론은 문장 중간에서 휴지를 표시하는 동시에 나름의 기능에 따라 조문을 분절하기 때문에 법률 해석에 영향을 준다. 기계번역을 법률에 특화해 정확한 결과물을 얻기 위해서는 중국어 법률 텍스트의 문장부호에 관한 면밀한 연구가 필요하다(최승혁 2023: 357-360).

11) 괄호 안의 숫자는 빈도를 나타낸다.

<그림 2> 민법전 기계번역의 문장부호 처리 양상



<그림 2>는 민법전 기계번역의 문장부호 처리 양상을 보여준다. 가장 높은 비중을 차지한 것이 ‘쉼표 생략(1,043회)’으로 1,258개 조문이 한국어로 번역되는 과정 중 조문 82%에서 쉼표가 사라졌다. 중국과 달리, 한국어 법률에서는 문장부호의 사용을 제한하는 경향이 있지만, 최근에는 법률의 정확한 이해를 위해 문장부호 사용을 적절하게 권장하고 있다(송덕수 2018: 400-401).

쉼표가 번역문에서 사라지는 단순한 오류보다 심각한 문제는 처리에 일관성이 없다는 점이다. 모점의 경우, 모점→쉼표(306회), 모점→및(89회), 모점→또는(98회), 모점→하거나(61회) 등과 같이 하나의 문장부호가 각기 다르게 번역되고 있는데 어떠한 일관성이나 규칙성도 없다. 심지어 한국어 법률에서 잘 사용하지 않는 가운데맺점(·)으로 바뀐 경우도 10회에 달했다. 아래 예문을 보자.

(1)

[ST] 第一千零二十條 合理實施下列行為的, 可以不經肖像權人同意:

(一) 爲個人學習、藝術欣賞、課堂教學或者科學研究, 在必要範圍內使用肖像權人已經公開的肖像; (二) 爲實施新聞報道, 不可避免地制作、使用、公開肖像權人的肖像; (三) 爲依法履行職責, 國家機關在必要範圍內制作、使用、公開肖像權人的肖像; (생략)

[MT-P] 제1020조 다음 각 호의 행위를 합리적으로 행하는 자는 초상권자의 동의를 받지 아니할 수 있습니다. (1) 개인학습, 예술감상, 교실수업 또는 과학연구를 위하여 필요한 범위 내에서 초상권자가 이미 공개한 초상화를 사용합니다. (2) 신문보도를 실시하기 위하여 불가피하게 초상권자의

초상을 제작·사용·공개하는 경우 (3) 법에 따라 직무를 수행하기 위하여 국가기관은 필요한 범위 내에서 초상권자의 초상을 제작, 사용 및 공개합니다. (생략)

예문(1)은 조(條) 아래에 5개의 ‘항’이 부가된 구조로 다양한 문장부호가 여러 차례 사용되었다. ST와 MT-P를 비교해 보면, 먼저 첫 번째 문장에서 쉼표가 생략되었다. 중국어 법률에서 항을 부가할 때 콜론으로 문장의 경계를 짓는데 이 부분 역시 마침표로 처리되었고, 모점의 경우 (1)항에서는 쉼표로 처리되었다. 주목할 점은 (2), (3)항에서 ‘제작, 사용, 공개’를 나열하는 내용이 반복해서 등장하는데 (2)항에서는 가운뎃점, (3)항에서는 쉼표와 접속 부사 ‘및’으로 각각 다르게 처리되었다. 세미콜론은 생략되거나 마침표로 바뀌었다.

#### 4.2.2 어휘 누락 및 추가

누락은 ST 어휘의 일부가 기계번역에서 번역되지 않는 것이다. 어휘 층위에서 누락의 출현 빈도는 총 257회였다. 그중 ‘其’, ‘該’과 같은 대응어(anaphor)의 누락이 총 99회로 39%를 차지했다. 대응어 누락 부분을 (Ø)로 표시하였다. 다음의 예문을 보자.

(2)

[ST] 第一百九十條 无民事行為能力人或者限制民事行為能力人對其法定代理人的請求權的訴訟時效期間，自該法定代理終止之日起計算。

[MT-P] 제190조 민사행위능력이 없는 자 또는 민사행위능력이 제한된 자의 (Ø)법정대리인에 대한 청구권의 소멸시효기간은 (Ø)법정대리인이 종료된 날부터 계산한다.

(3)

[ST] 第五百一十條 合同生效后，當事人就質量、價款或者報酬、履行地點等內容沒有約定或者約定不明確的，可以協議補充；不能達成補充協議的，按照合同相關條款或者交易習慣確定。

[MT-P] 제510조 계약이 발효된 후 당사자가 품질, 가격 또는 보수, 이행 장소 및 기타 내용에 대해 합의가 없거나 합의가 명확하지 않은 경우 합의에 따라 보완할 수 있으며, 추가 합의에 도달할 수 없는 경우 계약 관련 조항 또는 거래 관습에 따라 결정됩니다.



예문(2)는 대응어 오류 예시이다. ‘其’는 그 앞에 선행하는 ‘민사행위무능력자’와 ‘민사행위제한능력자’를 지칭하므로 ‘그’ 또는 ‘해당’으로 번역해야 한다. 하지만 MT-P는 ‘其’를 누락하고, 선행하는 두 주체와 한꺼번에 묶어서 법정대리인을 수식하는 오류를 범했다. ‘該’ 역시 선행하는 내용을 다시 한번 지칭하고 있지만 MT-P에서는 누락되었다. 예문(3)은 추가 오류의 예시이다. 추가 오류 총 42회 중에서 ‘等’과 관련한 오류가 총 21회(50%)로 비중이 높았다. 예문(3)에서 ST의 ‘等’이 MT-P에서 ‘및 기타’로 번역되면서 ST에 없던 내용이 추가된 것을 확인할 수 있다.

#### 4.2.3 어휘 사용 일관성

법률 언어의 사용에 있어서 ‘일관성’은 매우 중요한 원칙이다. 일관성은 하나의 법령이나 내용을 같이하는 법령 간에 최대한 동일한 용어와 표현을 사용하는 것이다(이상모 2021: 85). 기본 법률에서 사용한 전문용어나 어휘가 관련 법률에서 다른 의미로 사용된다면 법률 해석에 혼란을 줄 뿐 아니라 법에 대한 수법자의 신뢰와 권위를 저하할 수 있다. 아래 예문을 보자.

(4)

[ST] 第三十四條 監護人的職責是代理被監護人實施民事法律行爲，保護被監護人的人身權利、財產權利以及其他合法權益等。監護人依法履行監護職責產生的權利，受法律保護。監護人不履行監護職責或者侵害被監護人合法權益的，應當承擔法律責任。因發生突發事件等緊急情況，監護人暫時無法履行監護職責，被監護人的生活處於無人照料狀態的，被監護人住所地的居民委員會、村民委員會或者民政部門應當爲被監護人安排必要的臨時生活照料措施。

[MT-P] 제34조 후견인의 직무는 피후견인을 대리하여 민사법률행위를 수행하고 피후견인의 인적권, 재산권 및 기타 정당한 권익을 보호하는 것입니다. 후견인이 법에 따라 후견 업무를 수행함에 따라 발생하는 권리는 법률의 보호를 받습니다. 후견인이 후견업무를 수행하지 않거나 피후견인의 정당한 권익을 침해한 경우 법적 책임을 진다. 돌발사태 등 긴급상황으로 보호자가 일시적으로 보호자의 직무를 수행할 수 없고 피보호자의 생활이 돌봄이 없는 상태인 경우 피보호자의 주소지 주민위원회, 촌민위원회 또는 민정부서는 피보호자를 위해 필요한 임시 생활 돌봄 조치를 마련해야

한다.

예문(4)의 ‘監護’는 한국어 법률 전문용어 ‘후견’에 대응한다. ‘후견’이란 ‘친권에 의하여 보호를 받을 수 없는 미성년자, 피성년후견인(금치산자)을 보호하기 위하여 마련한 민법상의 직무’를 말한다.<sup>12)</sup> 장애아동 복지지원법, 소년법, 유아교육법 등 관련 법률마다 다소 차이는 있지만 ‘보호자’는 후견인을 포함하는 상위 개념으로 본다.<sup>13)</sup> 즉, ‘후견인’과 ‘보호자’의 범주가 다르다. 하나의 조문 내에서 같은 어휘가 다르게 번역되는 것은 법률 번역에 있어서 치명적인 오류이므로 ‘監護’를 ‘후견’으로 통일하는 것이 좋다.

#### 4.2.4 주체-대상

민법은 사람들 간의 민사행위를 규정하는 기본법으로 다양한 계약관리를 포함한다. 계약의 구성에는 최소한 두 명의 주체가 필요하다. 이들은 서로 대척점에 서게 되고, 그에 따라 동작이나 행위에 방향성을 갖게 된다. 예를 들어, 사업을 발주하는 측이 있으면 그 발주를 받는 측이 있다. 그리고 이런 관계는 보통 발주와 수주, 도급과 수급과 같이 쌍을 이루는 용어들이 있다. 이런 상반된 관계를 처리할 때 누가 누구에게 무엇을 하는지 동작이나 행위의 방향을 정확히 옮겨야 한다. 다음 예문을 보자.

(5)

[ST] 第七百七十三條 承攬人可以將其承攬的輔助工作交由第三人完成。承攬人將其承攬的輔助工作交由第三人完成的，應當就該第三人完成的工作成果向定作人負責。

[MT-P] 제773조 발주자는 그 발주된 보조업무를 제3자에게 인계하여 수행할 수 있다. 계약자가 계약한 보조 업무를 제3자에게 인계하여 완성하는 경우, 제3자가 완성한 업무 성과에 대해 주문자에게 책임을 져야 한다.

예문(5)에서 ST의 ‘承攬人(수급인)’과 ‘定作人(도급인)’은 행위의 주체와 대상으로 계약에서 방향성을 가진다. 하지만 MT-P에서는 ‘발주자’와 ‘주문자’

12) 법제처 국가법령정보센터 - 그 밖의 정보 - 법률 용어

13) 상동

로 번역했다. ‘발주’는 계약을 통해 주문을 신청하는 것으로 이미 주문의 의미를 담고 있다. 즉, 예문(5)에서는 다른 어휘를 같은 의미로 번역하면서 주체와 대상 간 동작의 방향을 부정확하게 처리하였다.

### 4.3 언어와 문법 층위의 상세 오류 항목

#### 4.3.1 법률 어휘 및 전문용어의 부정확한 사용

(6)

[ST] 第六百七十六條 借款人未按照約定的期限返還借款的，應當按照約定或者國家有關規定支付逾期利息。

[MT-P] 제676조 대출자가 약정된 기한 내에 대출을 반환하지 않는 경우 약정 또는 관련 국가 규정에 따라 연체 이자를 지불해야 합니다.

예문(6)에서 ‘借款人’, ‘借款’, ‘支付’ 등 어휘 오류를 발견할 수 있다. ‘借款人’, ‘借款’은 일반적으로 ‘대출자’, ‘대출’과 대응하는 전문용어다. 하지만 ‘대출자’라고 하면 대출을 시행하는 측인지 대출을 신청해 받는 측인지 모호하다. 따라서 대출을 받는 측의 입장을 살린 ‘차주’로 수정하는 것이 좋다. ‘借款’은 ‘돈을 빌리다’ 또는 ‘대출(하다)’을 의미한다. ST에서는 ‘대출금’이라는 명사로 사용되었지만 정확하게 번역되지 않았다. ‘支付’는 ‘支付宝(알리페이)’, ‘電子支付(전자결제)’ 등과 같이 일상에서도 상용되는 어휘이다. 하지만 법률 도메인에서는 채무 변제를 위해 금전이나 어음 등을 채권자에게 준다는 의미를 나타내어 ‘지급’<sup>14)</sup>이 정확한 번역이나 ‘지불’로 잘못 번역되었다.

#### 4.3.2 불필요한 피동

한국어 법령문에는 일상생활에서 자주 사용하지 않는 부자연스러운 표현이 자주 사용되어 수범자들의 이해를 떨어뜨린다. 대표적으로 일본어의 영향을 받은 피동문을 꼽을 수 있다. 우리말은 사물이나 무생물보다는 사람을 주어로 삼기 때문에 다른 외국어에 비해 피동문이 적은 편이다. 법제처에서 발간한 ‘알기 쉬운 법령 정비 기준(제10판)’에 따르면 ‘피동문은 자연스러운 문장이 아니므로

14) 네이버 국어사전

주체를 드러내어 능동문으로 고쳐 쓸 것’을 제안하며 법률에 나타나는 불필요한 피동문을 줄이도록 권고한다. 다음의 예문을 보자.

(7)

[ST] 第三百零六條 按份共有人轉讓其享有的共有的不動產或者動產份額的，應當將轉讓條件及時通知其他共有人。其他共有人應當在合理期限內行使優先購買權。兩個以上其他共有人主張行使優先購買權的，協商確定各自的購買比例；協商不成的，按照轉讓時各自的共有份額比例行使優先購買權。

[MT-P] 제306조 공유자가 향유하는 공유 부동산 또는 동산의 몫을 양도하는 경우 양도조건은 즉시 다른 공유자에게 통지되어야 합니다. 기타 공유자는 합리적인 기간 내에 우선 구매권을 행사해야 합니다. 2인 이상의 다른 공유자가 우선매수권 행사를 주장하는 경우 각자의 구매 비율을 협상하여 결정하고, 협상이 실패할 경우 양도 시 각자의 공유지분 비율에 따라 우선매수권을 행사한다.

예문(7)에서 ST의 첫 번째 문장을 보면 ‘應當’을 기준으로 선행절이 주어부, 후행절이 술어부이다. 선행절과 후행절은 모두 사람(按份共有人)을 주어로 한다. MT-P에서는 ‘공유자’, ‘양도 조건’ 이렇게 두 개의 주어를 설정하였고, 무생물 주어인 ‘양도조건’의 서술어를 ‘통지되다’라는 피동문으로 호응시켰다. 한국어 경동사 ‘-하다’와 ‘-되다’의 적절한 구분, 즉 피동과 능동을 적절하게 구분하여 사용하면 가독성이 높아지기 때문에(구명철과 정수정 2018: 21) ‘양도조건을 통지한다’라고 수정할 것을 제안한다.

#### 4.3.3 도치

많은 언어에서 어순은 다양한 기능을 수행한다. 따라서 어순 조정은 문장을 구성하는 어휘의 단순 조정이 아니라 주제구조나 정보구조의 변화로 강조점이 바뀐다(이은용 2011: 98-100). ‘도치’는 문장 성분의 순서를 바꾸어 내용을 강조하거나 두드러지게 하는 표현 방법이다. 내용을 강조한다는 것은 서술자의 의도가 개입되어 있음을 의미한다. 입법자는 입법 과정에서 자신의 의도에 따라 어순을 구성하기 때문에 법률 번역 시에 최대한 ST의 어순을 배열을 유지하는 것이 좋다. 다음 예문을 보자.

(8)

[ST] 第三百六十八條 居住權無償設立，但是當事人另有約定的除外。設立居住權的，應當向登記机构申請居住權登記。居住權自登記時設立。

[MT-P] 제368조 거주권은 당사자가 따로 약속한 경우를 제외하고는 무상으로 설립한다. 거주권을 설정하는 경우 등록 기관에 거주권 등록을 신청해야 합니다. 거주권은 등기 시에 설정된다.

예문(8)에서 ST의 첫 번째 문장은 두 개의 절이 접속사 ‘但是’로 연결되어 역접 관계를 이룬다. 그러나 MT-P에서는 ‘但是’의 후행절이 앞으로 도치되었음을 볼 수 있다. ST는 거주권 성립의 일반 조건을 먼저 제시하고 예외의 경우를 부연 설명하였지만, 도치의 결과로 ST의 강조점이 달라진 것이다. 주목할 점은 도치 오류 총 87회가 모두 ‘XP+的除外’ 구조에서 발생했다. 민법전에서 어떠한 예외사항을 부연할 때 ‘XP+的除外’의 통사구조의 사용 빈도가 높다. ‘的除外’ 부분이 고정적, 반복적으로 나타나기 때문에 이를 ‘경우에 그렇지 않는다’로 대응시켜 기계번역에 학습하면 해결할 수 있을 것이다.

#### 4.3.4 통사 성분의 위치

민법전 기계번역에서는 ST의 ‘其他(기타)’, ‘或者(또는)’, ‘也(도)’, ‘僅(만)’, ‘自(로부터)’ 등의 통사적 위치가 잘못 처리되는 오류가 빈번히 발생했다. 특히, 해당 조사나 부사의 통사 위치는 의미역에 영향을 주어 다른 뉘앙스를 만들기 때문에 정확하게 번역해야 한다.

(9)

[ST] 第二百一十三條 登記机构不得有下列行爲：

- (一) 要求對不動產進行評估；
- (二) 以年檢等名義進行重復登記；
- (三) 超出登記職責範圍的其他行爲。

[MT-P] 제2013조 등록기관은 다음 각 호의 행위를 하여서는 아니 된다.

- (1) 부동산에 대한 평가를 요구한다.
- (2) 연차검사 등의 명목으로 중복등록한다.
- (3) 기타 등록업무 범위를 초과하는 행위이다.

(10)

[ST] 第一千零一十二條 自然人享有姓名權，有權依法決定、使用、變更或者許可他人使用自己的姓名，但是不得違背公序良俗。

[MT-P] 제1011조 자연인은 성명권을 가지며 법에 따라 타인이 자신의 이름을 사용, 사용, 변경 또는 허가할 권리가 있지만 공서양속을 위반해서는 안 된다.

예문(9)에서 ST의 ‘其他’는 行爲(행위)를 수식한다. 하지만 MT-P에서는 제 3호의 문두에 놓여 통사적으로 ‘등록업무’와 거리가 가장 가깝다. 따라서 ‘기타’가 ‘등록업무’를 수식하는 것으로 오해할 수 있다. 이는 ‘등기 직무의 범위를 초과하는 그 밖의 행위’로 수정하는 것이 정확하다. 예문(10)에서 ST의 ‘或者’의 앞에는 결정하다(決定), 사용하다(使用), 변경하다(變更) 3개의 서술어가, 뒤에는 허가하다(許可) 1개의 서술어가 등장한다. ‘或者’ 앞의 3개 서술어와 공통적으로 호응하는 목적어는 ‘姓名權’이고, 서술어 ‘허기하다’의 목적어는 ‘他人使用自己的姓名’이다. 다시 말해, MT-P에서 ‘或者’를 정확하게 번역하지 못하고, 앞의 서술어 3개와 뒤의 서술어 1개의 목적어를 한꺼번에 처리하는 오류를 범했다.

#### 4.3.5 정형 표현

한 텍스트에서 반복적, 관습적으로 사용되는 단어의 결합을 정형 표현(formulaic expressions)이라고 한다. 법률에도 각각의 정형화된 구문이 있는데, 이것이 서로 상이하게 번역되면 다른 해석으로 오인될 수 있어서 정형화된 표현은 반드시 일관되게 번역해야 한다(이상모 2021: 88). 다음 예문을 보자.

(11)

[ST] 第三百一十三條 善意受讓人取得動產后，該動產上的原有權利消滅。但是，善意受讓人在受讓時知道或者應當知道該權利的除外。

[MT-P] 제313조 선의의 양수인이 동산을 취득한 후 해당 동산에 대한 원래 권리는 소멸됩니다. 다만, 선의의 양수인이 양수 당시 그 권리를 알았거나 알았어야 하는 경우는 예외로 한다.

(12)

[ST] 第一百零六條 有下列情形之一的，非法人組織解散：（一）章程規定的存續期間屆滿或者章程規定的其他解散事由出現；（二）出資人或者設立人決定解散；（三）法律規定的其他情形。

[MT-P] 제106조 다음 상황 중 하나에 해당하는 경우 비법인 조직은 해산됩니다. (1) 정관에 규정된 존속기간의 만료 또는 정관에 규정된 기타 해산 사유가 발생한 경우 (2) 출자자 또는 설립자가 해산을 결정한다. (3) 기타 법률에 규정된 상황입니다.

예문(11), (12)에는 법률 텍스트의 관습적인 정형 표현이 사용되었다. 예문(11)에서 ST의 ‘知道或者應當知道’를 MT-P는 과거 시제와 의무 양태를 결합해 ‘알았거나 알았어야 하는’으로 번역했다. 하지만 한국어 법률 텍스트에서는 ‘알았거나 알 수 있었을’과 같이 과거 시제와 가능 양태를 결합한 정형 표현을 사용한다. 예문(12)에서 ST의 ‘有下列情形之一的’는 법률 외에도 계약서나 기타 정본 문서에서 사용 빈도가 높은 대표적 정형 표현이다. MT-P의 번역이 크게 틀리지는 않았지만 ‘다음 각 호의 어느 하나에 해당하는 사유가 있는 경우에’로 번역하는 것이 한국 법률 텍스트 정형 표현 관습에 부합한다.

#### 4.4 구조 층위의 상세 오류 항목

##### 4.4.1 종결어미

한국어는 어미가 발달한 언어로 연결어미나 종결어미가 굉장히 다양하고 풍부하다. 이는 중국어와의 큰 차이이다. 법률 텍스트의 문장은 일반적으로 평서문으로 사용하고, 한국어의 경우 ‘-하다’ 형태의 종결어미를 사용해 사실에 대한 인식, 진술, 서술을 나타낸다. 하지만 민법전의 범용 기계번역에서는 한국 법률 텍스트의 특징에 부합하지 않는 경어체 종결어미, 명사 형태의 문장 종결 등의 문제가 발생했다.

(13)

[ST] 第一百七十三條 有下列情形之一的，委託代理終止：（一）代理期限屆滿或者代理事務完成；（二）被代理人取消委託或者代理人辭去委託；（三）代理人喪失民事行為能力；（四）代理人或者被代理

人死亡；(五) 作為代理人或者被代理人的法人、非法人組織終止。

[MT-P] 다음 상황 중 하나에 해당하는 경우 위탁 대행은 종료됩니다.

(1) 대리 기간이 만료되거나 대리 사무가 완료됩니다. (2) 대리인에 의해 위탁이 취소되거나 대리인이 위탁을 사임한다. (3) 대리인은 민사행위능력을 상실한다. (4) 대리인 또는 피대리인의 사망 (5) 대리인 또는 피대리인으로서의 법인 또는 비법인 조직을 종료한다.

예문(13)은 종결어미의 다양한 오류가 복합적으로 드러난 예시이다. 특히 조문 이하의 항 단위에서 아주 높임체, 아주 낮춤체, 명사형이 혼용되었다. 한국 법률 텍스트에서 항 단위는 일반적으로 명사형으로 종결하거나 부득이 문장 형태로 종결해야 할 때는 ‘경우’로 종결한다. 제(1)항을 예로 들면 ‘대리 기간이 만료되거나 대리 사무의 완료’ 또는 ‘대리 기간이 만료되거나 대리 사무를 완료한 경우’로 번역하는 것이 적절하다.

## 5. 결론

### 5.1 연구 요약

본 연구는 중화인민공화국 민법전 1,258개 조문을 네이버 파파고 기계번역으로 번역한 다음, 각각의 조문에서 나타나는 오류를 일일이 정리하고, 의미, 언어와 문법, 구조 층위로 귀납, 정리하였다.

분석 결과, 민법전 기계번역에서는 문장부호, 법률 어휘와 전문용어의 부정확한 사용, 불필요한 피동, 도치, 법률 텍스트 고유의 정형 표현, 어휘의 누락과 추가 등 총 135가지 오류가 출현했다. 출현 빈도가 가장 높은 것은 문장부호 오류였으며 그중 ‘섬표 생략’이 가장 많았다. 두 번째로 출현 빈도가 높은 것은 법률 어휘 및 전문용어가 부정확하게 사용된 오류였다. 세 번째로는 법률 텍스트에서 사용하지 않는 종결어미인 경어체 오류의 빈도가 높게 나타났다.

상기 오류를 의미, 언어와 문법, 구조 층위로 귀납한 후에 예문을 통해 상세한 오류 항목을 살펴보았다. 한 조문에 여러 오류 유형이 복합적으로 출현하기 때문에 괄호 안의 백분율은 개별 오류 유형이 전체 1,258개 조문에서 차지



하는 비율이다. 따라서 오류 항목 비율의 총합은 100%가 아니다.

의미 층위에서 나타난 상세 오류 항목은 문장부호 오류 중 쉼표 생략(1,043회/83%), 어휘 누락(257회/20%), 어휘 사용 일관성(95회/8%), 어휘 추가(37회/3%), 주체-대상(28회/2%) 순이었다.

언어와 문법 층위에서 나타난 상세 오류 항목은 법률 어휘(1,004회/80%) 및 전문용어(835회/66%)의 부정확한 사용, 불필요한 피동(166회/13%), 도치(73회/6%), 통사 성분의 위치(58회/5%), 정형 표현(26회/2%) 순이었다.

구조 층위에서는 종결어미(786회/62%) 오류가 나타났다.

분석 결과를 살펴보면 각각의 층위별로 다양한 오류가 출현하였다. 그중 법률 어휘의 정확한 사용과 문장부호의 일관된 번역이 가장 시급하게 해결해야 할 오류로 확인되었다. 어휘의 경우, 법률 어휘와 전문용어를 출발언어와 도착 언어 DB로 구축해 기계번역에 학습시키면 신속하게 개선할 수 있다. 다만, 기계번역 학습용 법률 어휘 DB의 구축 과정에서 일상 언어와 차이, 국가 간 법률 체계의 차이, 사회문화적 차이, 법적 효력의 등가 등을 면밀하게 고려해 DB의 정확성을 확보하는 것이 관건이다. 문장부호의 경우 오류 양상의 일관성이 결여된 것이 가장 큰 문제점이다. 따라서 법률 텍스트에 사용되는 5가지 문장부호(마침표, 쉼표, 콜론, 세미콜론, 모점)가 정확하고 일관성 있게 처리되도록 공학적 도움을 빌릴 필요가 있다.

## 5.2 연구의 한계와 의의

본고는 민법전문만을 대상으로 했기 때문에 중국 법률 텍스트에 대한 대표성을 확보하기 어렵다는 한계가 있다. 민법전은 중국에서 ‘법전’으로 명명된 첫 번째 법률이다. 하지만 이외에도 헌법, 형법, 행정법, 상법 등 여러 법률이 있고, 각각 어휘적, 통사적 층위에서 많은 차이를 보인다. 따라서 민법전 기계번역의 오류 분석을 통해 중한 법률 기계번역 특화를 100% 구현하는 것은 어렵다. 이보다 더욱 방대한 법률 데이터를 구축해 기계번역을 학습, 특화하면 보다 뚜렷한 도메인 특화의 효과를 얻을 수 있을 것이다.

또한, 본고는 번역 연구의 관점에서 상세한 오류 항목을 제시함으로써 공학 전공자들이 현재 중한 법률 기계번역에서 나타나는 오류를 파악하도록 하였다.

이를 통해 번역학과 공학의 역할 괴리를 좁히며 협업의 접점을 찾고자 했다는 의의가 있다.

법률시장 개방 등 대외적 환경이 변화로 대량의 법률 번역 수요를 소화하기 위해 법률 도메인에 기계번역을 활용하려는 노력이 꾸준히 이어지고 있다. 이 과정에서 번역연구자가 기계번역의 공학적 메커니즘 개선보다는 기계번역이 정확하게 학습할 수 있는 데이터를 만드는 데 기여하고, 공학계에 좋은 시사점을 제공할 수 있다면 번역학의 가치가 더 돋보일 것이다.

법률 도메인 특화 기계번역에 관한 관심과 연구가 부족한 현실이 아쉽지만, 본 연구를 계기로 오류 분석 이외에도 언어학적, 번역학적, 공학적 측면에서 다양한 후속 연구가 활발히 이루어지길 기대한다.

## 참고문헌

- 송덕수 (2018) 『민법전의 용어와 문장구조』, 서울: 박영사.
- 이혜운 (2020) 『법언어학의 이해』, 서울: 역락.
- Anna Trosborg 편저 (2021), 이혜승·권영교 옮김, 『텍스트유형론』, 서울: 한국외국어대학교 지식출판콘텐츠원.
- 강병규 (2006) 「面向中文專著的漢韓機器輔助翻譯研究」, 박사학위논문, 북경대학.
- 강병규, 이지은 (2018) 「신경망 기계번역의 작동 원리와 번역의 정확률 — 중한 번역을 실례로」, 『중어중문학』 73: 253-295.
- 구명철, 정수정 (2018) 「가독성의 관점에서 본 법률 언어의 어휘론적인 문제점 및 개선 방안」, 『독어학』 37: 1-27.
- 권도경, 문수정, 김정수, 허윤정 (2019) 「2018년 인공지능경망 기계번역의 중한 번역 현황 — 구글 번역과 네이버 파파고를 중심으로」, 『중국문학』 98: 233-266.
- 기유미 (2018) 「한중 기계번역 오류의 문형별 비교분석 — 네이버 파파고 번역기와 구글 번역기를 중심으로」, 『중국연구』 74: 3-32.
- 김나영 (2021) 「러-한 기계 번역의 현황 및 지향점: 형식적 등가(formal

- equivalence)와 역동적 등가(dynamic equivalence)에 따른 번역 소프트웨어 결과물의 품질 연구를 중심으로, 『슬라브학보』 36(1): 35-54.
- 김세린, 권혁철, 장칭하오(Qinghao Zhang) (2022) 「도메인 지식에 특화된 신경망 기계번역」, 『한국정보과학회 학술발표논문집』, 678-680.
- 김혜림 (2019) 「중한 법령 기계번역 포스트에디팅 교육을 위한 예비 연구」, 『번역학연구』 22(3): 65-99.
- 김희동, 지인영 (2017) 「심층학습을 이용한 기계번역 기술과 정확도 연구」, 『인문언어』 19(2): 23-46.
- 남철진 (2018) 「中韓(韓中) 번역을 통해 본 구글번역(GNMT)의 문제」, 『중국과 중국학』 34: 1-27.
- 류수린, 국지연, 정동규 (2021) 「독한 기계번역 정확도 평가의 언어학적 접근 — 구글번역기와 네이버 파파고를 중심으로」, 『독어독문학』 62(3): 119-140.
- 박찬준, 김경희, 박기남, 임희석 (2020) 「COVID-19에 특화된 인공지능경망 기계번역기」, 『한국융합학회논문지』 11(9): 7-13.
- 서보현, 김순영 (2018) 「기계번역 결과물의 오류유형 고찰」, 『번역학연구』 19(1): 99-117.
- 이건철 (2022) 「특허데이터로 학습한 트랜스포머 기반의 인공지능경망 번역모델」, 석사학위논문, 국민대학교.
- 이상모 (2021) 「패러다임의 전환을 통한 법령번역 접근방법」, 『T&I Review』 11(2): 75-100.
- 이은용 (2011) 「일한 출판 번역물에 나타나는 번역 양상과 번역전략 고찰」, 박사학위논문, 이화여자대학교.
- 이준호 (2022) 「법률 특화 번역엔진 성능 평가 — 한영 계약서 번역을 중심으로」, 『T&I Review』 12: 169.
- 이준호, 이청호, 최효은 (2023) 「자동화된 기계학습(AutoML)을 활용한 특허 특화 번역엔진의 영한번역 성능 평가」, 『번역학연구』 24(2): 101-130.
- 이지은, 최효은 (2017) 「특허 기계번역 결과물의 평가 — KIPRIS의 무료 한영 기계번역을 중심으로」, 『통역과 번역』 19(1): 139-178.
- 이지은, 최효은 (2023) 「인공지능경망 기반 맞춤형 기계번역엔진의 성능 평가: 법

를 및 특히 한영번역 결과물 평가 사례를 중심으로, 『번역학연구』 24(1): 9-37.

- 이향 (2010) 「번역품질평가의 현황과 과제」, 『번역학연구』 11(2): 107-125.
- 임형재, 왕첸 [Wang, Tian] (2019) 「한국어 번역을 위한 문화소의 기계번역 연구 — 중국 외교연설문 번역의 정확도 평가를 중심으로」, 『현대사회와다문화』 8(2): 68-94.
- 채종린 (2022) 「법령 텍스트 기계번역에서 나타나는 중한 번역 오류 분석: 대만의 민법을 중심으로」, 석사학위논문, 한국외국어대학교.
- 최문선 (2019) 「국내 번역학 기계번역 연구 동향: 내용 분석과 키워드 분석을 중심으로」, 『언어학연구』 24(1): 275-297.
- 최승권, 권오욱, 이기영, 노윤희, 박상규 (2007) 「도메인 특화 방법에 의한 영한 특허 자동 번역 시스템의 구축」, 『소프트웨어 및 응용』 34(2): 95-103.
- 최승권, 이기영, 노윤희, 권오욱, 김영길 (2012) 「패턴기반 영한 자동번역 시스템의 상용화를 위한 도메인 특화 방법」, 『소프트웨어 및 응용』 39(4): 253-260.
- 최승혁 (2023) 「법률 특화 기계번역을 위한 중국 법률 텍스트의 어휘적 특징 분석 — 중화인민공화국 민법전을 예로」, 『중어중문학』 92: 349-380.
- 최혜령 (2020) 「중한 기계번역의 어휘 오류분석: 『人工智能革命(인공지능혁명, 王天一)』 번역의 정확성과 이해가능성을 중심으로」, 석사학위논문, 한국외국어대학교.
- 한양양 (2019) 「정보적 텍스트의 기계번역에 나타나는 중한 번역 오류 분석」, 석사학위논문, 한국외국어대학교.
- 賈揚帆 [Jia, YangFan] (2019) 「기계번역에 나타나는 중한 번역 오류 분석: 畢淑敏『1cm』 어휘를 중심으로」, 석사학위논문, 한국외국어대학교.

<보고서>

법제처 (2021) 『알기 쉬운 법령 정비 기준』 (제10판).

<인터넷 자료>

가장 우수한 한국어 번역 품질을 제공하는 기계번역 API.

<https://www.ncloud.com/product/aiService/papagoTranslation>

네이버 국어사전

<https://ko.dict.naver.com/#/entry/koko/dab253264ed84e5781a10d28cb598298>

매일경제 (2023) LG이노텍, 中카메라모듈 공장 단계적 축소.

<https://www.mk.co.kr/news/business/10796891>

법률신문 (2019) 연 매출 2000~3000억 규모… ‘법률번역 시장’ 선점 경쟁.

<https://www.lawtimes.co.kr/news/156645>

법제처 국가법령정보센터 - 그 밖의 정보 - 법령용어

<https://www.law.go.kr/lsTrmSc.do?menuId=13&subMenuId=65>

전자신문 (2022) 네이버 파파고 “이제 찍지 않고 들고만 있어도 번역해 준다.”

<https://www.etnews.com/20220831000185>

중화인민공화국 민법전

<https://www.12371.cn/2020/06/01/ARTI1591021670041266.shtml>

한국경제 (2020) ‘한국 로펌 중국서 활동 힘든데…중국 로펌은 잇따라 서울 상륙.’ <https://www.hankyung.com/society/article/2020060798091>

ZDNET Korea (2023) [데뷰23] 네이버 ‘파파고’ 높은 번역 품질…“빨간펜 선생님 덕분이죠.” <https://zdnet.co.kr/view/?no=20230228170421>

[Abstract]

**An Inductive Analysis of Machine Translation Errors for  
Specializing Machine Translation in the Chinese-to-Korean  
Legal Domain: Focusing on Papago Translations of  
the Chinese Civil Code**

Seung-hyuk Choi  
(Sogang University)

The Civil Code is China's primary legislation and the first law designated as the "code" within China's legal system. This study selected 1,258 articles from the Chinese Civil Code to create a machine translation system specialized for the Chinese-to-Korean legal domain. First, the selected articles were translated using the Naver Papago machine translation system. After the translation, errors found in each article were systematically compiled and categorized by type, including message, language and grammar, structure. The analysis showed a significant frequency of errors, especially in legal terminology and punctuation usage. Although the study is limited by insufficient data for thorough analysis constructing an extensive corpus of legal texts and scrutinizing their characteristics for a highly specialized machine translation in the legal domain would result in a more apparent domain-specific effect. The changing landscape of the legal market, resulting from factors like legal market liberalization, emphasizes the need for specialized machine translation within the legal domain to keep up with significant demand for legal translation. Considering the situation, this study identifies specific error instances from the perspective of translation research, enabling engineering scholars to recognize the prevalent errors in Chinese-to-Korean legal machine translation, thereby reducing the gap between the roles played in research by translation and engineering scholars and identifying areas for cooperation and collaboration between the two studies.

Keywords: neural network machine translation, domain-specific machine translation, legal translation, translation errors, inductive analysis

주제어: 신경망 기계번역, 도메인 특화 기계번역, 법률 번역, 번역 오류, 귀납적 분석

최승혁

서강대학교 중국문화학과 박사과정 수료

choi00joy@hanmail.net

관심 분야: 신경망 기계번역, 법언어학, 법률번역, 현대 중국어 어법

논문 투고: 2023년 8월 11일

1차 심사 완료: 2023년 8월 31일

2차 심사 완료: 2023년 9월 10일

게재 확정: 2023년 9월 19일