

인간번역과 기계번역 및 포스트에디팅의 오류 비교: 한영 법률 번역 사례를 중심으로*

이지은(이화여자대학교)

최효은(이화여자대학교)

1. 서론

인공신경망 기계번역(NMT) 기술 개발 이후 기계번역 및 포스트에디팅 연구가 활발히 전개되는 가운데 품질과 생산성은 핵심 주제라 할 수 있다 (신지선, 2020; 이상빈, 2020; 최문선, 2020). 시간과 비용 면에서 기계번역이 경제적이지만 양질의 결과물을 생산하기 위해서는 포스트에디팅이 필요하다. 기계번역 품질이 떨어질수록 포스트에디팅에 투입되는 노력이 증가하므로 일정 수준 이상의 기계번역을 대상으로 할 때만 포스트에디팅은 효율적인 작업 방식이 될 수 있다. 따라서 기계번역 활용 여부는 기계번역의 품질과 텍스트 종류, 최종납품 번역의 목적 및 용도 등에 따라 고려해야 할 것이다.

* 본고는 한국법제연구원의 용역연구과제 ‘번역 효율성 제고를 위한 영한 법령용어 번역메모리 구축 및 포스트에디팅에 관한 연구’의 일부를 바탕으로 한 후속연구이다.

포스트에디팅과 인간번역을 비교한 연구는 관련 변수가 많고, 그로 인한 영향도 크기 때문에 상반된 연구결과를 얻기도 했다. 하지만 최신 연구 상당수는 포스트에디팅 소요 시간을 인간번역과 비교했을 때 작업 생산성이 높고 품질도 나쁘지 않다는 결과를 보인다(김순미 등, 2019; 이준호와 김순미, 2022; Jia et al., 2019). 대표적으로 영한 IT 매뉴얼 텍스트를 대상으로 한 이준호와 김순미(2022)에 의하면, 포스트에디팅은 인간번역 대비 생산성 우위를 보이며 결과물의 품질도 인간번역에 뒤지지 않은 것으로 평가되었다. 많은 선행 연구가 일부 사례에 기반한 한계점이 있지만 전문성이 강하지 않은 보편적인 텍스트를 대상으로 한 최신 번역학 연구는 공통적으로 기계번역 활용 확대 가능성을 보여준다.

그렇지만 기계번역과 포스트에디팅의 다양한 측면에 대해서 아직 살펴 보지 않은 부분이 많다. 텍스트 유형이나 주제와 텍스트 난이도에 따른 기계번역 및 포스트에디팅 결과물의 품질과 생산성에 대한 연구는 활발하게 진행되고 있지만 번역 수요가 높은 법률 텍스트를 대상으로 한 연구는 국내 외적으로 미진한 편이다. 한국어 법률 텍스트의 기계번역에 관한 연구는 여전히 손에 꼽을 정도로 적다(이준호, 2022; 이지은과 최효은, 2023a, 2023b; Lee & Choi, 2023). 또한 이제까지 구글, 파파고, 카카오 i 등 범용 번역엔진의 번역 품질 관련 국내 연구가 활발히 진행되었는데(서보현과 김순영, 2018; 이준호, 2019; 한승희와 강동희, 2022; 한현희, 2020), 2023년 한국어 서비스를 시작한 DeepL Translator에 대한 연구는 아직 이뤄지지 않았다. 이러한 연구 공백을 메우고 법률 기계번역 활용에 대해 보다 구체적으로 고찰하기 위해 본 연구는 처음부터 사람이 번역한 결과물과 DeepL의 기계번역, 그리고 포스트에디팅 결과물의 오류 분석에 초점을 맞추고자 한다. 작업 방식이 다른 세 종류의 한영 법률 번역 결과물에 나타나는 오류의 유형과 빈도, 그리고 포스트에디팅에서 오류 수정의 여부를 분석함으로써 인간번역과 비교되는 기계번역을 활용한 결과물의 특징을 파악할 수 있을 것이다. 어느 때보다 번역 업계에서 인간과 기계의 협업 가능성이 높아진 상황에서 본 연구는 법률 번역 실무 및 교육에 대한 함의가 있을 것으로 기대된다.

2. 관련 선행 연구

2.1 포스트에디팅 품질과 오류

인간번역과 기계번역 모두 핵심 평가 기준은 원문에 대한 충실성과 목표 언어의 유창성 두 가지로 압축된다. 전자는 내용 보존과 의미 전달의 정확성(다른 말로 충실성, 충분성, 적정성)을 뜻하며 후자는 목표 언어의 적절성 내지 가독성을 가리킨다(김자경, 2020; 이지은과 최효은, 2022, 2023a; Lommel, 2018; Vieira, 2017). 대체로 정확성 오류에는 누락, 왜곡, 추가 등이 포함되며, 유창성 오류에는 문법, 통사, 맞춤법 관련 오류 등이 있는데, 연구자마다 적용하는 기준과 분류의 차이가 존재하며 구체적인 평가 목적에 따라 용어와 스타일 준수를 별도 기준으로 적용하거나 다양한 세부 기준을 마련한다(김자경, 2020; 남보리, 2021; 이준호, 2022; 이지은과 최효은, 2023a; Jia et al., 2019; Lee & Choi, 2023).

번역과 마찬가지로 포스트에디팅에서도 언어능력과 번역능력, 전문 지식 등을 포함한 포스트에디터의 역량과 성향, 그밖에 가이드라인, 서비스 유형, 효율, 텍스트 주제 등을 비롯한 여러 가지 요소가 결과물의 품질에 영향을 미친다(김자경, 2020, 2022a, 2022b; 서보현과 김순영, 2020; 이상빈, 2017; 마승혜, 2018; Aranberri et al., 2014; Green et al., 2013; Vieira, 2017). 따라서 포스트에디팅을 거쳐도 수정이 미흡할 수 있고, 수정에 오류가 있을 수도 있다. 모국어 방향의 포스트에디팅 결과물 오류를 분석한 Mitchell(2015)에서는 전체 기계번역 오류의 73% 정도가 수정되었는데, 미수정 오류 중 오역과 정보 누락 등 정확성 관련 오류가 가장 많았고 그다음으로 유창성에 해당하는 언어적 오류가 많았다. Vardaro 등(2019)에 의하면 유럽연합집행위원회 번역총국(EU DGT)에서 사용하는 다양한 주제의 문건에 대한 기계번역 결과물에서는 어휘 오류가 가장 많고, 그다음으로 정보 배열, 누락 등의 오류가 많았다. 포스트에디팅 결과물의 오류도 어휘 오류, 특히 스타일/어역, 기능어, 오역, 용어 오류가 주를 이루었다(Vardaro et al., 2019, p. 25). 흥미로운 점은 DGT 번역사들의 손을 거쳤음에도 기계번역 오류 가운데 약 1/3가량은 포스트에디팅을 통해 수정되었으나 절반을 넘는 61.5%는 수정되지 않았다.

오류 유형별로 문체와 관련된 문제가 가장 적게 수정된 반면 오역은 가장 많이 수정되었다(Vardaro et al., 2019, p. 17).

선행연구를 종합해 보면 기계번역과 포스트에디팅 결과물 양자 모두 정확성 오류 비중이 높은 것으로 보인다. 정확성 오류는 포스트에디팅 노력을 많이 필요로 하는 중대 오류이나 수정되지 않을 가능성이 있다(김자경, 2020, 2021, 2022a, 2022b; 윤미선 등, 2018; 이준호, 2021). 특히 복잡한 문장이나 내용 난이도가 높을 때 그러하다(김자경, 2022b; 정재혁, 2023). 김자경(2020)이 분석한 한영 뉴스 텍스트의 포스트에디팅 코퍼스에서는 정확성 오류 중 누락이 가장 많았고, 그 다음으로 오역, 용어, 시제, 다의어 순이었으며, 가독성 오류로는 문체, 문법, 직역, 은유, 맞춤법 오류, 정보배열 오류 순이었다. 한편 학생들의 한영 포스트에디팅 결과물에 나타난 정확성 오류를 분석한 김자경(2022b)에 의하면 미수정 오류 유형 중 구조 오류가 가장 많았고, 그 다음으로 어휘 오류, 누락 오류 순으로 나타났다. 학생 포스트에디팅에서 용어 오류는 미수정될 가능성이 가장 높았고, 구조 오역의 경우에 내용 전달 문제를 논리적으로 쉽게 판단할 수 있는 오류는 수정률이 높으나 원문 주제 관련 지식이 부족할 경우 수정률이 낮은 것으로 분석되었다(김자경, 2022b, p. 112).

기계번역이 품질이나 스타일면에서 인간번역과 차이가 있듯이 기계번역을 수정한 포스트에디팅도 인간번역과 차이가 있다. 영한 IT 텍스트의 인간번역과 포스트에디팅 결과물을 분석한 이준호(2018)에 의하면 포스트에디팅 결과물은 어절 수로 본 양적 확장과 어휘 다양성 감소의 특징이 있었다. 이 같이 다른 방식으로 생산한 결과물의 품질과 오류도 다른 양상을 보일 수 있다. 품질과 생산성 관점에서 과학기술 텍스트의 영한 인간번역과 구글번역 포스트에디팅 결과물을 분석한 정재혁(2023)에 의하면, 인간번역의 경우 작업 시간은 길었지만 정확성과 용어 및 구조분석 오류로 본 품질은 포스트에디팅보다 우위에 있었다. 정확한 정보전달이 중요한 과학기술 텍스트의 포스트에디팅 과정에서 정확성 오류를 식별하기 쉽지 않다는 점이 재확인되었고, 그만큼 주제지식과 전문용어에 대한 이해를 바탕으로 한 번역 및 포스트에디팅 작업을 수행하는 사람의 역량과 숙련도가 중요한 문제임을 보여주는 결과였다(정재혁, 2023, pp. 196, 198).

2.2 법률 번역 기계번역과 포스트에디팅

본 연구는 정확한 정보 전달이 중요한 법률 텍스트의 번역 사례를 통해 기계번역을 활용하는 정도가 다른 작업 결과물의 오류 양상을 파악하고자 한다. 법률 분야도 법령, 계약서, 학술서 등 텍스트 유형과 주제 등에 따라 텍스트 특성이 다르고, 그밖에 사용한 번역엔진, 언어쌍과 평가 방식의 차이 등으로 인해 직접적인 비교가 어렵지만 법률 기계번역 관련 선행연구를 간략히 살펴보면 다음과 같다. 다만 자동평가 기반 연구(Koehn & Knowles, 2017; Defauw et al., 2019)는 오류 분석이 불가하고 NMT를 이용한 법률 기계번역 수동평가 연구가 많지 않다는 점을 밝혀둔다.

2010년대 중반까지만 해도 통계기반 번역엔진(SMT)을 활용한 법률 텍스트 기계번역의 품질은 낮은 편이었다(이지은과 최효은, 2023a, p. 13). 영어-크로아티아어 간의 구글 번역 결과물을 평가한 Brkić 등(2014, p. 4)에 의하면 유창성 문제에 해당하는 언어적 오류 중 형태적(morphology) 오류가 가장 많았고, 그다음으로 어휘 오류, 비번역, 통사 오류, 과잉 번역 순이었다(Brkić et al., 2014, p. 4).

현재 DGT에서 사용하는 NMT는 법률 텍스트를 포함한 유럽연합 문건에 특화된 ‘eTranslation’으로 지속적 품질 관리가 이뤄지고 있다(Horváth, 2024). 언어방향이나 언어쌍에 따라 오류 양상이 다를 수 있다는 점을 감안해야 하지만 슬로베니아어 번역팀이 사용하는 ‘eTranslation’의 법률 기계번역 결과물을 분석한 결과, 용어 정확성과 일관성이 가장 심각한 문제로 확인되었다(Arnejšek & Unk, 2020, p. 7).

DeepL Translator를 활용한 법률 기계번역 평가를 다룬 해외 연구도 소수에 불과하다. Wiesman(2019)이 DeepL Translator와 CAT 툴을 이용하여 이탈리아어 법률 텍스트를 독일어로 번역한 결과물의 품질을 분석한 바 있다. 당시 연구 결과에 의하면 DeepL이 상대적으로 우수했으나 전반적으로 정확성과 유창성 점수 모두 낮았고, 정확성은 상당한 수정을 요하는 낮은 수준으로 평가받았다. 독일어-이탈리아어쌍에 관한 Heiss와 Soffritti(2018)의 연구와 화란어-영어쌍에 관한 Dik(2020)의 연구는 DeepL의 잠재력을 긍정적으로 평가하였다. 반면, 독일어-스페인어쌍에 대한 DeepL 결과물을 분석한

Roiss(2021)는 어휘, 용어, 어역 측면에서 부정적인 평가를 내렸다. 번역학 연구는 아니지만 DeMattee 등(2022)은 사회과학연구를 위한 법률 기계번역 활용을 모색할 목적으로 DeepL, 구글, 마이크로소프트의 번역엔진으로 얻은 기계번역과 인간번역 결과물을 비교하여 평가한 바 있다. 기계번역과 달리 인간번역은 중대한 오류가 없었으며, 오류 개수를 기준으로 볼 때 구글이 가장 적고, 그 다음이 마이크로소프트, DeepL 순으로 나타났다. DeMattee 등(2022, p. 19)은 품질과 비용면에서 기계번역 결과물을 번역사들이 수정하는 법률 번역 포스트에디팅을 제안하였다.

법률 텍스트의 기계번역 포스트에디팅 결과물과 인간번역을 비교한 Killman과 Rodríguez-Castro(2022)에 의하면 구글 SMT로 산출한 영어-스페인어 기계번역 포스트에디팅이 생산성과 품질면에서 모두 인간번역을 앞섰다. 인간번역 대비 오류가 적고 16%가량 작업 시간 절감 효과가 있었는데 이 결과는 통계적으로 유의미한 차이를 보였다(Killman & Rodríguez-Castro, 2022, p. 66).

한편 자체 개발한 NMT를 이용하여 그리스어-영어 간 법률 번역 결과물을 평가한 Sosoni 등(2022)에서는 포스트에디팅의 생산성 효과가 확인되지 않았다. 비록 단 두 명의 번역사의 작업 수행에 기반한 제한적인 연구 결과이지만 결과를 볼 때 포스트에디팅 작업 시간이 더 길었고 오류 개수로 본 품질 역시 인간번역보다 우세하다고 할 수 없었다. 두 가지 다른 작업 방식의 결과물에서 정확성과 유창성 오류는 별 차이가 없었다. 큰 차이는 아니지만 인간번역에서는 정확성 오류가, 포스트에디팅 결과물에서는 스타일 오류가 좀 더 많았다(Sosoni et al., 2022, p. 103).

앞서 언급했듯이 법률 기계번역 품질 및 포스트에디팅 관련 국내 연구는 극소수에 불과하다. 김혜림(2021)은 교육 관점에서 중한 법령 기계번역(파과고) 결과물에 대한 번역 전공 석사생들의 포스트에디팅 오류를 분석한 연구다. 오류를 크게 의미(용어 일관성, 정확성)와 언어 문법(어휘, 통사, 문장부호, 맞춤법) 범주로 구분하여 오류 사례를 살펴보았으나 양적 분석은 실시하지 않았다. 김혜림(2021)에 의하면 학생들은 법령의 언어적 특성이나 내용 이해가 부족하여 가독성 수정에 치우치는 경향이 있었고, 정확성, 완결성, 일관성 등 법률 기계번역 오류를 충분히 수정하지 못하였다. 이러한 연

구 결과는 법률 번역 기술이 법률 포스트에디팅에 필수 요건임을 시사한다.

한영 법률 기계번역에 관한 선행연구는 NMT 기반 범용 번역엔진과 특화 번역엔진에 대한 평가 연구로 포스트에디팅은 다루지 않았다. 특화 번역엔진과 구글번역이 산출한 한영 계약 문건 번역 결과물을 평가한 이준호(2022)에서는 정확성, 유창성, 사용 적합성을 기준으로 본 난제 문장의 번역 품질은 일반 문장 대비 낮았지만 정확성과 유창성 품질면에서 특화 번역엔진이 범용 번역엔진보다 더 좋은 평가를 받았다. Lee와 Choi(2023)에 의하면 자동 평가와 수동 평가의 결과 두 가지 모두 특화 번역엔진(오토란)이 범용 번역엔진(구글, 파파고)에 비해 우수한 성능으로 평가되었다. 엔진 종류에 따라 오류 양상의 차이도 드러났는데, 특화 번역엔진의 경우는 용어 오류가 가장 많았고 범용 번역엔진은 정확성 오류가 많았다. 범용엔진에 비해 특화 번역엔진의 경우 정확성 오류가 절반도 안 되었고, 유창성 오류도 절반 수준으로 훨씬 적었다(Lee & Choi, 2023, p. 84). 한편 동일한 특화 번역엔진의 법률 텍스트와 특허 텍스트 번역 결과물 품질을 비교했을 때 법률 번역이 더 좋은 평가를 받았는데, 이는 길고 복잡한 문장으로 구성된 특허 텍스트의 특성과 밀접한 관계가 있는 것으로 보인다(이지은과 최효은, 2023a, 2023b).

3. 연구 방법

인간번역, 기계번역, 포스트에디팅과 같이 다른 작업 방식으로 도출한 결과물에 대한 오류 분석을 목적으로 한 본 연구의 참가자들은 한영 번역사들과 번역 전공생들, 그리고 이들이 작업한 결과물을 평가할 평가자들로 구성된다. 번역 및 포스트에디팅은 로펌 소속 전문번역사 6명(이하 ‘전문가’)로 칭하며, P1~P6으로 구분함)과 법률번역 과목을 수강 중인 번역 전공 석사과정 학생 6명(이하 ‘학생’)으로 칭하며, S1~S6으로 구분함)이 수행하였다. 본 연구에 사용한 국문 법령은 『스토킹범죄의 처벌 등에 관한 법률』(이하 스토킹법)과 『수상레저기구의 등록 및 검사에 관한 법률』(이하 수상레저기구등록법) 2종이며, 비슷한 난이도로 평가된 240단어 가량 발췌한 분량을 각각

90분 이내에 작업하게 하였다. 포스트에디팅은 2023년 5월 초 산출한 DeepL의 기계번역 결과물을 대상으로 하였으며, 인간번역 수준에 부합하도록 오류 수정 중심으로 작업하고 개인 선호에 따른 스타일 수정은 지양하도록 가이드라인을 제시하였다. 참가자들은 트랜스로그를 설치한 연구자 PC에 작업 지침에 따라 작업한 후 중간 휴식 시간을 10분 정도 가진 후 다른 방식의 작업을 수행하였다. 이때 작업 순서를 교차로 하여 참가자 중 일부는 인간번역을 먼저 수행하고, 나머지는 기계번역 후 포스트에디팅을 먼저 수행하여 작업 순서에 따른 영향을 최소화하고자 하였다.

평가는 법률 번역 경험이 10년 이상인 평가자 2인과 번역교수 1인이 참여하였다. 평가자 3명은 번역 주체와 작업 방식이나 소요 시간 등에 대한 정보 없이 익명화된 24건의 텍스트(참가자 12명 x 2가지 결과물)와 DeepL 기계번역 결과물을 원문인 국문 법령 텍스트와 함께 받고, 평가 가이드라인에 따라 비대면 방식으로 각자 평가를 진행하였다. 결과물 평가는 정확성, 유창성, 용어 기준에 따라 세그먼트별로 0~4점 척도를 이용하여 평가하도록 하였고, 전체 총점 또한 0~4점 사이를 기입하도록 하였다.¹⁾ 평가 가이드라인을 통해 0점에서 4가지 대략적인 수준에 대해 기술하여 평가자들이 참고할 수 있도록 하였다. 가급적 일관성 있는 평가를 위해 한 가지 작업 방식으로 얻은 텍스트 평가를 완료한 후 다른 작업 방식의 텍스트를 평가 진행하도록 하고, 작업 방식이 동일한 12건은 가능한 한 단시간 내에 한 번에 평가를 완료해줄 것으로 요청하였다.

오류 분석을 위해 평가자들이 작성한 오류 코멘트를 분류하고 분석하였다. 본 연구에 사용한 평가 기준은 법률 기계번역 관련 선행연구(이지은과 최효은, 2023a)를 참고하였다. 크게 정확성, 유창성, 용어 세 가지 기준 아래 추가, 누락, 오역은 정확성 오류로, 문법, 통사 문제는 유창성 오류에 포함시켰다.²⁾ 용어가 결과물 품질에 중요한 비중을 차지하기 때문에 본 연구에서

1) 3인의 평가자가 시행한 평가는 61.8% 일치하며(ICC=0.618), 이는 통계적으로 유의미하다($p=0.002$).

2) 번역 및 포스트에디팅 과정 기록을 위해 참가자들이 트랜스로그를 사용하여 작업하였기 때문에 트랜스로그에서 맞춤법 검사가 되지 않는 점을 감안하여 맞춤법 오류는 평가 대상에서 제외하였다.

별도 평가 항목으로 두고 정확한 용어 사용과 일관된 용어 사용 두 가지 관점에서 평가하도록 하였다. 평가자들의 오류 코멘트를 기준으로 오류 개수를 계산하였는데, 먼저 대분류에 해당하는 정확성, 유창성, 용어의 오류를 결과물별로 분석한 다음, 하위의 소분류에 해당하는 추가, 누락, 오역 오류와 문법, 통사 오류, 그리고 용어 정확성과 일관성 오류를 분석하였다. 또한 기계번역 오류가 포스트에디팅 과정에서 수정이 이루어졌는지 평가자들이 공통적으로 지적한 오류를 파악하여 포스트에디팅 결과물에서 오류 유형별 수정 여부를 분석하였다.

4. 연구 결과

본 장에서는 논의의 편의를 위해 인간번역은 HT, 기계번역은 MT, 포스트에디팅은 MTPE로 칭한다.

4.1 대분류별 오류 분석

먼저 HT와 MTPE 결과물에 나타난 정확성, 유창성, 용어의 대분류별 오류 개수를 집단별로 집계하면 아래 <표 1>과 같다. 다만 본 연구의 초점이 전문가와 학생의 비교에 있지 않기 때문에 이하 소분류별 오류에 대한 전문가 집단과 학생 집단의 비교는 간략히 기술하겠다.

표 1

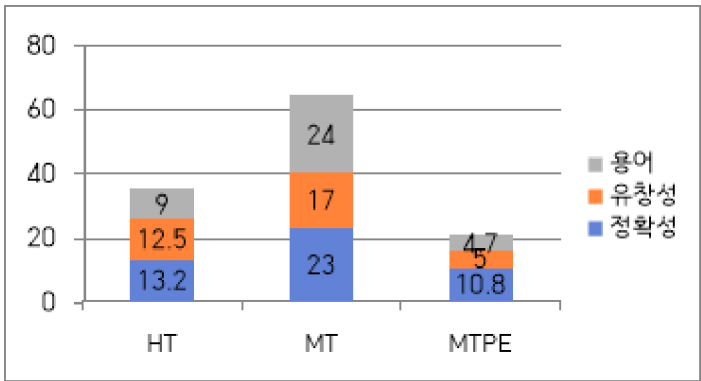
HT와 MTPE 대분류 오류 비교(단위: 개)

	HT		MTPE	
	전문가	학생	전문가	학생
정확성	49	109	60	70
유창성	60	90	23	37
용어	47	61	20	36
총합	156	260	103	143

<표 1>에서 확인할 수 있듯이 전체적으로 HT에 비해 MTPE의 오류 개수가 훨씬 적고, 세 가지 오류 모두 전문가의 오류가 학생보다 적었다. 작업 방식별로 보자면, HT의 경우에는 정확성에서 전문가와 학생 간의 오류 개수가 가장 큰 차이를 보였고 그다음이 유창성, 그리고 용어에서 가장 차이가 적었다. 한편 MTPE는 HT에 비해 두 집단 간 오류 개수의 차이가 적으며, 정확성의 경우 집단 간 격차가 크게 줄었다.

MT는 DeepL 결과물 단 1건이기 때문에 절대 비교를 위해 각기 12건인 HT와 MTPE의 대분류별 오류 개수를 모두 환산하여 건당 오류를 살펴보았다.

그림 1
HT, MT, MTPE 건당 대분류별 오류 개수



위 <그림 1>에 제시한 바와 같이 건당 오류 개수는 MT가 가장 많고, HT가 그다음으로 많으며 MTPE가 가장 적다. 이와 같은 절대적인 오류 개수는 결국 품질과 직결된다. 즉, 품질로 보면 MTPE > HT > MT 순이라 할 수 있다. 작업 방식에 따른 대분류별 오류의 개수를 보면, HT의 경우 정확성과 유창성이 비슷한 비율로 문제가 된 반면, MTPE는 정확성 문제가 유창성과 용어 오류에 비해 상대적으로 비중이 높다. HT와 MTPE 모두 가장 비중이 높은 오류는 정확성이다. 정확성 오류의 경우, 세 가지 다른 결과물 중 MT 결과물이 가장 많고, 비중이 높다. MT에서 가장 많은 용어 오류도 정확

성과 밀접한 관련이 있다는 점을 고려할 때 MT에서 유창성보다 정확성과 용어 오류가 두드러진다. MTPE의 정확성 오류가 세 가지 중 가장 적다. 내용 및 정보의 정확한 보존과 전달이 중요한 법률 번역에서 MTPE가 효율적인 방식이 될 수 있음을 보여주는 결과이다.

MTPE의 경우, MT에 비해 오류의 절대적인 수가 눈에 띄게 줄어 포스트에디팅을 통해 그 품질이 향상된 것을 쉽게 확인할 수 있다. 한편 MTPE 결과물에서 정확성 오류가 비중이 높다는 것은 MT의 정확성 오류 수정이 비교적 덜 이루어졌을 가능성을 시사하며, 정확성 오류는 포스트에디팅 노력을 많이 요하기 때문에 수정되지 않을 가능성이 있다는 선행연구(김자경, 2020, 2021, 2022a, 2022b; 윤미선 등, 2018; 이준호, 2021; 정재혁, 2023) 결과와 일치한다.

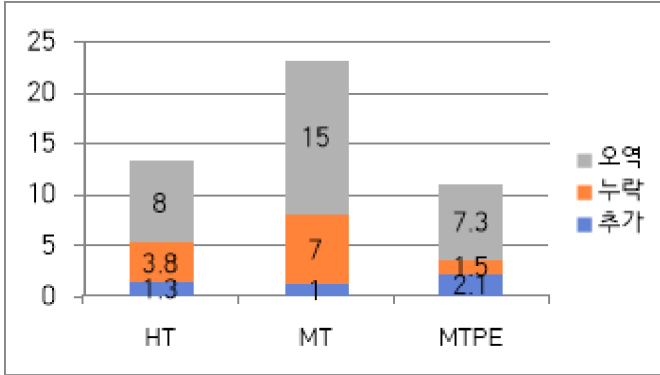
또 주목할 만한 점으로 MTPE가 MT를 수정한 결과물임에도 불구하고 오류 빈도가 정확성 > 유창성 > 용어 순으로 HT와 같고 용어 > 정확성 > 유창성 순인 MT와 다른 점이다. 이는 인간 개입으로 인해 MTPE 결과물이 HT와 유사한 오류 양상을 보일 수 있다는 점을 시사한다.

4.2 소분류별 오류 분석

4.2.1 정확성

본 소절에서는 정확성 오류의 하위 분류인 추가, 누락, 오역의 비중이 작업 방식에 따라 어떠한지 살펴보고, MT의 정확성 오류가 MTPE에서 어느 정도 수정되었는지를 분석하겠다.

그림 2
HT, MT, MTPE에 나타난 정확성 오류



작업 방식별 건당 정확성의 하위 분류를 보면, 위 <그림 2>와 같이 HT, MT, MTPE 모두 오역 비중이 가장 많았고, 품질 순위는 위와 마찬가지로 MTPE > HT > MT 순이다. 전체 정확성 오류에서 오역이 차지하는 비중을 살펴보면 HT가 61%로 가장 낮았고, MT는 65%, MTPE는 67%의 비중을 차지했다. 따라서 오역을 최소화하여 의미 전달의 정확성을 요하는 번역에는 인간번역이 가장 적합하다고 할 수 있다. 이는 <표 1>에서 보이듯이 전문가 집단의 HT 정확성 오류가 MTPE보다 적다는 결과에서도 알 수 있다.

HT에서 오역의 경우 전문가보다는 학생 집단의 비중이 높았는데 모호하게 기술되어 있는 원문을 제대로 이해하지 못하거나 잘못 해석한 경우가 대부분이었다. 예를 들어, 스토킹법 제3조제3호는 ‘피해자등에 대한 긴급응급조치 및 잠정조치 요청의 절차 등 안내’이다. 이 중 ‘잠정조치 요청의 절차’를 S6은 ‘potential measures and requested procedures(잠정조치 및 요청 절차)’로 오역하였다.

MTPE에서 오역은 MT에서 잘못된 부분을 수정하면서 오히려 제대로 되어 있는 부분까지 수정한 경우가 많았으며, 이러한 현상은 학생 결과물에서 자주 나타났다. 예를 들어, 수상레저기구등록법 제4조제1호에 나오는 ‘유선 및 도선 사업법」에 따른 유·도선사업 및 그 사업과 관련된 수상에서의 행위를 하는 경우’를 MT는 ‘the wire and cable business pursuant to the Wire

and Cable Business Act and acts on the water related to such business’로 처리하여 용어 문제는 있지만 ‘the business... and acts on the water’의 구조는 문제가 없다고 할 수 있다. 그런데 S4와 S5는 ‘유·도선사업’ 용어 오역을 수정하면서 ‘In case of excursion ship and ferry business and activities on the water related to such businesses pursuant to the Excursion Ship and Ferry Business Act’로 구문을 수정하여 ‘~법에 따른 수상에서의 사업 및 행위’로 오역하였다.

한편 누락의 경우 HT와 MTPE에 비해 MT에서 훨씬 빈번하게 나타난다는 점 또한 눈여겨볼 만하다. HT와 MTPE는 추가와 누락 빈도의 차이가 크게 없는 반면, MT에서는 추가가 단 1건인데 반해 누락은 7건으로 무려 7배 차이를 기록했다.

다음으로 MT의 정확성 오류가 MTPE에서 얼마나 수정되었는지를 살펴 보겠다. 이를 위해 전체 MTPE 12건을 대상으로 평가자들이 공통으로 지적한 MT 오류가 MTPE에서 수정되었는지 여부를 분석하였다. 평가자 3인이 공통으로 지적한 정확성 오류는 총 14건으로 이 중 누락이 6건, 오역이 8건이었으며, 추가 오류는 없었다. 이에 대한 수정 여부를 살펴본 결과는 아래 <표 2> 및 <표 3>에 제시하였다.

표 2

MT의 누락 오류에 대한 MTPE 수정 건수

	항목1	항목2	항목3	항목4	항목5	항목6
수정 건수	8	8	10	10	12	12

표 3

MT의 오역 오류에 대한 MTPE 수정 건수

	항목1	항목2	항목3	항목4	항목5	항목6	항목7	항목8
수정 건수	7	11	10	11	2	12	12	7

<표 2>를 보면, MTPE 12건 중 최소 8건 이상이 MT의 누락 오류를 수정하여 평균 수정률이 83.33%인데 비해, <표 3>에서 알 수 있듯이 오역은 7

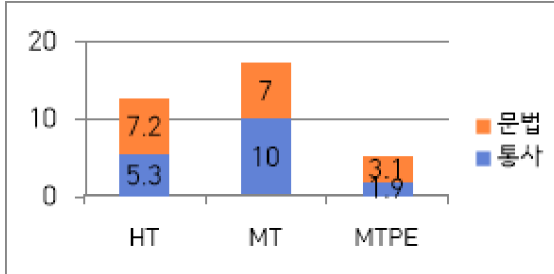
건 이하인 경우가 세 항목에서 나타나 평균 수정률이 75%이다. 특히 항목5의 경우, 수정이 2건에 그쳐 평균 수정률이 16.67%에 불과하다. 이와 같이 필요한 수정이 이루어지지 않은 경우는 대부분 학생의 MTPE였다. 누락의 경우 전문가의 미수정 건수는 전체 6개 항목 중 단 1건이었으며, 오역의 경우도 항목5를 제외하고 전문가의 미수정 건수는 항목당 0~2사이다. 전문가의 미수정률은 항목5를 제외하면 16.67%에서 33.33% 사이인 반면 학생의 미수정률은 16.67%에서 83.33%까지 상당히 넓은 폭으로 분포하여 개별 차이가 큰 점을 보여준다.

오역에서 특히 수정률이 낮은 항목들은 대개 문장 구조 전체의 변형이 필요하여 인지적 노력을 상당히 요하는 경우였다. 오역 오류 중 항목1을 예로 들면, 수상레저기구등록법 제3조 후단 ‘다만, 동력수상레저기구의 총톤수, 출력 등을 고려하여 대통령령으로 정하는 경우에는 그러하지 아니하다.’가 MT에서 ‘However, this shall not apply if the Presidential Decree determines the total tonnage, power output, etc. of the motorized water recreation device. (대통령령이 동력수상레저기구의 총톤수, 출력 등을 정하는 경우에는 그러하지 아니하다.)’로 번역되어 그 의미가 왜곡되었다. 이를 바로잡기 위해서는 문장 구조 전체를 바꾸어야 하므로 포스트에디팅에 들어가는 시간과 노력이 상당히 요구된다. 항목1의 경우 올바르게 수정하지 않은 5명 모두 학생이었다.

4.2.2 유창성

본 소절에서는 유창성과 관련해서 유창성을 구성하는 하위 분류인 통사와 문법 오류 비중을 작업 방식에 따라 살펴보고, MT의 유창성 오류가 MTPE에서 어느 정도 수정되었는지 살펴보고자 하겠다.

그림 3
HT, MT, MTPE에 나타난 유창성 오류



<그림 3>에 작업 방식별 건당 유창성의 하위 분류를 제시하였다. 전체 유창성 오류의 수는 정확성 오류의 수와 동일하게 MT > HT > MTPE 순으로 나타났다. 한편 HT와 MTPE에서는 통사와 문법이 차지하는 비중이 비슷한 가운데 문법 오류의 비중이 조금 높은 반면, MT에서는 통사 오류의 비중이 더욱 높은 것으로 나타났다. 이러한 결과는 MT의 유창성 품질이 아직도 취약하며 특히 통사 문제가 HT에 비해 훨씬 심각하다고 해석할 수 있다.

한편 문법과 관련해서 MT 오류로 대부분 세미콜론이나 콜론의 사용 문제, 어순을 고려하지 않은 지나친 직역 등이 거론되었다면 HT에서는 부정확한 관사의 사용, 부정확한 전치사의 사용, 타동사의 목적어 부재 등이 문법 오류로 꼽혔다.

HT의 문법 오류 예를 보면, 스토킹법 제3조의 ‘사법경찰관리는 진행 중인 스토킹행위에 대하여 신고를 받은 경우 즉시 현장에 나가 다음 각 호의 조치를 하여야 한다.’에서 ‘다음 각 호의 조치’는 특정되므로 정관사를 수반해야 함에도 불구하고 일부 학생은 ‘the’ 없이 ‘following actions’로 번역하여 유창성 오류가 발생했다. HT에서의 문법 오류 비중은 전문가와 학생 간 질적 및 양적으로 큰 차이 없이 비슷한 수준을 보였다. 한편 MTPE에서는 MT의 오류를 수정하지 않은 사례 외에 정확하게 번역된 MT 세그먼트를 수정하면서 오류를 초래한 경우 또한 살펴볼 수 있었다. 마찬가지로 학생 MTPE에서 이러한 오류 양상이 좀 더 두드러지게 나타났다. 예를 들어, 수상레저기구등록법 제3조의 후단은 ‘다만, 동력수상레저기구의 총톤수, 출력 등을

고려하여 대통령령으로 정하는 경우에는 그러하지 아니하다.’로 되어 있는데, MT는 이를 ‘However, this shall not apply if the Presidential Decree determines the total tonnage, power output, etc. of the motorized water recreation device.’로 오역하였다. 이에 대해 S1은 ‘However, this shall not apply if its total tonnage, power output, etc. is as prescribed by Presidential Decree.’로 수정하면서 ‘as’를 삭제하지 않아 유창성 오류를 초래하였다.

다음으로 MT의 유창성 오류가 MTPE에서 얼마나 수정되었는지를 살펴 보면, MT에서 평가자 3인이 공통으로 지적한 유창성 오류는 총 8건으로 그 중 문법 오류가 7건, 통사 오류가 1건이었다.

표 4

MT의 유창성 오류 중 통사 오류에 대한 MTPE 수정 건수

	항목1
수정 건수	12

표 5

MT의 유창성 오류 중 문법 오류에 대한 MTPE 수정 건수

	항목1	항목2	항목3	항목4	항목5	항목6	항목7
수정 건수	11	8	11	11	11	9	10

위의 <표 4>와 <표 5>를 보면, MTPE 12건에서 평가자들이 지적한 통사와 문법 문제는 MTPE에서 대체로 모두 잘 수정되었음을 알 수 있다. 통사 오류 1건은 수상레저기구등록법 제6조제2항제2호의 ‘동력수상레저기구의 구조, 설비 및 장치가 제15조제1항제1호 및 같은 조 제5항에 따른 신규검사 기준에 맞지 아니한 경우’가 MT에서 ‘the structure, equipment, and devices of the motorized water recreation device do not meet the standards for new inspection pursuant to Article (이하 생략).’로 번역됨으로써 ‘If’ 접속사가 누락되고 평서문으로 처리된 사례였다. 참가자 전원이 해당 오류를 수정하여 100%의 수정률을 보였다.

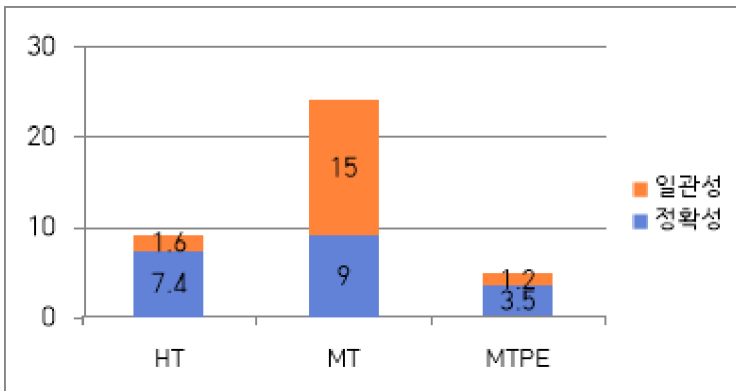
문법 오류는 대개 세미콜론과 콜론의 사용에 대한 문제였으며 이 역시

전문가와 학생 간 큰 차이 없이 대부분 수월하게 수정하여 84.52%의 높은 수정률을 보였다. 하지만 전문가와 학생 집단 간 미수정률에는 차이가 있었는데, 전문가 집단의 경우 미수정률이 16.67%에서 33.33%로 전문가 간 차이가 적은 반면, 학생 집단의 경우 미수정률이 16.67%에서 50%로 정확성에서와 마찬가지로 개별 차이가 컸다. 이 가운데 수정 건수가 낮은 항목6을 구체적으로 살펴보면, 수상레저기구등록법 제6조제2항으로 ‘시장·군수·구청장은 다음 각 호의 어느 하나에 해당하는 경우 등록신청을 거부할 수 있다.’가 MT에서 ‘The mayor, county governor, and head of the district office may reject the registration application if any of the following applies.’로 번역된 사례다. 2항 이하 1호와 2호에서 ‘다음 각 호’를 설명하므로 ‘applies’ 다음의 마침표(.)를 콜론(:)으로 수정해야 하나 전문가 2인, 학생 1인이 수정하지 않은 사례다.

4.2.3 용어

본 소절에서는 용어와 관련해 용어를 구성하는 하위 분류인 용어의 정확성과 일관성 오류의 비중을 작업 방식에 따라 살펴보고, MT의 용어 오류가 MTPE에서 어느 정도 수정되었는지를 살펴보고자 하겠다.

그림 4
HT, MT, MTPE에 나타난 용어 오류



<그림 4>에 작업 방식별 건당 용어의 하위 분류 오류를 제시하였다. 전체 용어 오류의 수는 정확성 및 유창성 오류와 동일하게 $MT > HT > MTPE$ 순으로 나타났다. 한편 HT와 MTPE에서는 용어의 정확성이 차지하는 비중이 높고 용어의 일관성은 거의 문제가 되지 않은 반면, MT에서는 일관성 문제가 두드러지게 나타났다. 유창성 오류와 마찬가지로 용어의 일관성 문제 또한 사람이 개입하는 HT와 MTPE에서 보다 MT에서 부각될 수 있음을 시사하는 부분이다.

예를 들면 주요 용어에 해당하는 ‘동력 수상레저기구’가 MT에서 조항에 따라 ‘a motorized watercraft’, ‘motorized water recreational devices’, ‘motorized water recreational vehicles’와 같이 다양하게 표현되었다. 이같이 빈도가 높은 단어나 어휘에 대한 MT가 다양한 표현을 사용하여 일관성을 저해하는 양상은 Lee와 Choi(2023)가 지적한 바와 같이 법률에서 사용하는 용어에 따라 지칭하는 바가 다르게 해석될 수 있으므로 주의를 기울여 수정해야 한다. 본 연구에서 법률 번역 경험이 많은 전문가들은 용어의 정확성과 일관성에 유의하여 작업하였으나 학생들은 그렇지 못하여 절반이 ‘동력 수상레저기구’를 일관성 없이 수정하였다. 예를 들면, S5의 경우는 ‘motorized water recreational devices’와 ‘motorized water recreational equipment’를, S3와 S7은 ‘motorized water recreational devices’와 ‘motorized water recreational vehicles’를 혼용하였다.

한편 용어의 정확성 오류 사례로 ‘시장·군수·구청장’의 예를 들어보겠다. MT는 이를 ‘mayor, county governor, and ward chief’로 번역하였고, 작업 가이드라인에는 동일한 용어는 아니지만 ‘시·군·구’를 ‘Si/Gun/Gu’로 제시하였다. 참가자 중 전문가 4명과 학생 3명은 지침을 고려하여 ‘시장·군수·구청장’을 ‘head(s) of Si/Gun/Gu’로 수정하였으나 나머지 참가자들은 지침을 반영하지 않고 부적절하게 수정하였다. 부적절하게 수정한 예로는 ‘mayor, county governor, head of district office’ 또는 ‘mayor of Si, major of Gun, mayor of Gu’ 등을 들 수 있다. 이러한 사례는 해당 용어가 국문 법령에 상당히 자주 사용되고, 가이드라인에 관련 용어를 제시하였음에도 이를 충실히 따르지 않았음을 보여준다. 이와 같은 용어의 정확성 오류는 법령 및 법률문서에서 사용하는 용어에 비교적 서투른 학생 집단에서 더 빈번하게 일

어났다.

MT에서 평가자 3인이 공통으로 지적한 용어 오류는 총 4건으로 그 중 정확성 오류가 2건, 일관성 오류가 2건이었다. 용어 오류에 대한 수정 여부를 아래 <표 6>, <표 7>에 제시하였다.

표 6

MT의 용어 오류 중 정확성 오류에 대한 MTPE 수정 건수

	항목1	항목2
수정 건수	8	11

표 7

MT의 용어 오류 중 일관성 오류에 대한 MTPE 수정 건수

	항목1	항목2
수정 건수	8	7

평가자들이 지적한 용어의 정확성과 일관성 문제는 MTPE에서 대체로 잘 수정되었다. 하지만 전문가와 학생 집단의 수정률 차이는 있었다. 용어의 정확성 오류에 관해서는 항목 1의 경우 전문가와 학생의 미수정률이 33.33%로 동일한 반면 용어의 일관성 오류 중 전문가의 미수정률은 0%였으나 학생의 미수정률은 항목1과 2가 각각 66.67%, 83.33%로 높았다. 위 <표 7>의 일관성 오류 중 항목1의 미수정 4건과 항목2의 미수정 5건은 모두 학생들이 수정하지 않은 사례였다. 가장 수정률이 낮은 일관성 오류의 항목2를 살펴보면, 수상레저기구등록법 제6조제2항제2호의 ‘동력수상레저기구’에 관한 것으로 다른 조 및 항과의 일관성을 맞추는 필요가 있지만 학생 대부분이 이 부분을 미수정하여 수정률이 저조하였다. 해당 호가 포스트에디팅 대상 텍스트의 가장 마지막 세그먼트임을 감안할 때 학생 참가자들의 집중력 저하, 부주의 등이 영향을 미쳤을 가능성이 있다.

5. 논의 및 결론

지금까지 법률 텍스트의 인간번역과 포스트에디팅 결과물 총 24건 및 기계번역 결과물 1건에 대한 평가를 통해 세 가지 다른 방식으로 산출한 결과물에 나타나는 번역 오류를 분석했다. 정확성, 유창성, 용어의 대분류별 오류의 개수를 건당으로 환산하여 살펴본 결과, 품질은 포스트에디팅 결과물이 가장 양호하고, 그다음이 인간번역, 그리고 기계번역이 가장 낮았다. 특히 본 연구에서는 무료 범용 엔진 DeepL의 기계번역 결과물에 대한 포스트에디팅 오류가 인간번역 결과물보다 적었다. 이는 충분한 학습과 용어 관리가 이뤄진 특화 번역엔진 사용 시 포스트에디팅 결과물의 품질이 더 우수할 수 있다는 가능성을 시사한다. 이를 고려할 때 법률 번역에서 기계번역 활용에 대한 가능성을 충분히 기대할 수 있다.

대분류별 주요 연구 결과를 요약하자면, 인간번역의 경우 정확성과 유창성이 비슷한 비율로 문제가 된 반면, 기계번역에서는 정확성과 용어의 문제가 부각되었다. 인간번역과 기계번역은 정확성, 유창성, 용어 가운데 정확성 오류의 비중이 높다는 점에서 공통점을 가지나 각기 오류 유형별 비율로 볼 때 인간번역은 유창성 문제, 기계번역은 용어 문제가 두드러지는 차이가 있었다. 이점은 텍스트 유형과 언어쌍은 다르지만 한승희와 강동희(2022)를 비롯한 선행연구 결과와 유사하다.

한편 포스트에디팅의 경우, 기계번역에 비해 오류의 절대적인 수가 눈에 띄게 줄었으므로 기계번역의 오류가 다수 수정되어 기계번역 결과물의 품질이 향상된 것을 볼 수 있다. 다만 포스트에디팅에서도 정확성 오류의 비중이 가장 높았다는 점을 주목할 만하다. 이는 기계번역 오류 중 정확성 오류가 다른 오류에 비해 미수정 비율이 가장 높을 가능성을 시사하기 때문이다.

결과를 오류별로 살펴보면 다음과 같다. 본 연구에서 작업 방식과 관계 없이 전체 결과물을 통틀어 가장 큰 비중을 차지하는 오류 유형은 정확성이고, 구체적으로는 추가와 누락에 비해 오역의 비중이 가장 높았다. 특히 기계번역에서 심각한 오역의 비중이 높았다. 한편 포스트에디팅에서 수정 여부를 분석하였을 때, 누락에 비해 오역의 수정률이 낮았으며, 무엇보다 인지

적 노력을 많이 필요로 하는 오역의 수정률이 낮아 미수정되는 경향을 보였다. 이는 선행연구 결과를 뒷받침하며 인간번역에 비해 포스트에디팅에서 전문가의 정확성 오류가 더 많았던 점을 고려할 때, 의미 전달의 정확성이 중요한 번역에서는 특히 정확성과 관련된 문제에 더욱 주의를 기울여야 할 것으로 보인다.

유창성은 인간번역과 포스트에디팅 결과물에서 두 번째로 높은 비중을 차지하는 오류 유형이다. 작업 방식에 따라 유창성 오류의 양상이 차이를 보였는데 인간번역과 포스트에디팅의 경우 통사와 문법 오류의 비중이 비슷하며 그 중 문법 오류의 비중이 소폭 높은 반면, 기계번역의 경우 통사 오류 비중이 좀 더 높은 경향을 보인 것이다. 특히 인간번역에서는 기계번역과 달리 직접적으로 유창성에 영향을 미칠 수 있는 문법 문제들이 상당수 관찰되었다. 한편 유창성과 관련된 기계번역 오류는 포스트에디팅 과정에서 대부분 잘 수정된 것으로 나타났다. 결과적으로 포스트에디팅의 오류 개수가 인간번역보다 적었다.

용어는 특히 기계번역에서 가장 높은 비중을 차지하는 오류 유형이다. 세부적으로 살펴보았을 때, 유창성과 마찬가지로 작업 방식에 따라 오류 양상이 조금 다른 경향을 보였다. 인간번역과 포스트에디팅의 경우에는 용어 정확성 오류의 비중이 높고 용어 일관성의 오류는 거의 찾아보기 힘든 반면, 기계번역의 경우 용어 일관성의 오류 비중이 월등하게 높았다. 번역과 포스트에디팅의 경우 법률에서 사용된 주요 용어를 일관성 있게 번역하는데 작업자들이 주의를 기울인 것과 달리 기계번역의 경우 동일한 국문 용어를 다양한 표현으로 처리했다. 이러한 결과는 범용 번역엔진 DeepL이 법률 텍스트의 용어 번역에 한계가 있음을 보여준다. 같은 언어쌍으로 특화 번역 엔진을 연구한 Lee와 Choi(2023)와 다른 언어쌍 및 엔진을 대상으로 한 Arnejšek과 Unk(2020)에서도 용어와 관련하여 유사한 결과가 확인된 바 있다. 이는 법률 기계번역의 취약점을 고려한 포스트에디팅이 필요하며, 법률 용어 학습과 관리가 필수적임을 뜻한다. 한편 인간번역에서는 기계번역에 비해 용어의 정확성이 문제가 되었는데, 이는 주로 번역 과정에서 인간의 해석이 개입되거나 학생들의 실수 혹은 무지로 인해 오류가 발생한 경우였다. 용어와 관련된 기계번역의 오류는 포스트에디팅을 통해 대부분 잘 수정

되었다.

본 연구는 법률 번역 기술을 갖춘 전문가의 포스트에디팅이 학생과 어떠한 차이를 보이는지 살펴보았기 때문에 교육적 함의가 있다. 오류 수정 결과를 전문가와 학생으로 구분해서 살펴보면 다음과 같다. 학생들이 수행한 포스트에디팅의 경우 오수정이나 미수정이 전문가들에 비해 많았지만 정확성 오류와 관련해서는 전문가 집단과의 격차가 크게 좁혀졌다. 오류 개수로 보았을 때 인간번역에 비해 포스트에디팅에서 전문가와 학생 집단의 작업 결과물 품질 차이가 좁혀졌다. 이러한 연구 결과는 전문가에 비해 번역 기술과 경험이 부족한 사람에게 보다 나은 품질의 결과물을 생산하는 데 포스트에디팅이 유용하다는 점을 시사한다.

본 연구는 제한적인 연구 자료와 사례를 기반으로 했기 때문에 연구 결과를 일반화할 수 없다. 하지만 전문 텍스트 중 법률 텍스트를 대상으로 인간번역과 기계번역 및 포스트에디팅 결과물의 오류 및 수정 양상을 분석함으로써 법률 번역에 기계번역을 활용할 경우 품질 관리에 유의해야 할 부분을 파악할 수 있었고, 이를 통해 포스트에디팅 노력을 효율화하고 최종 결과물의 품질을 높이는 데 조금이나마 기여할 수 있으리라 여겨진다. 이러한 연구 결과가 더 축적되면 법률 번역 및 포스트에디팅 교육과 실무에 참고할 수 있을 것이다. 향후 연구 범위를 보다 확대하여 실무에 사용되는 법률 특화엔진을 이용한 기계번역 및 포스트에디팅 결과물을 살펴보고, 법령뿐 아니라 다양한 법률 텍스트를 대상으로 하여 고찰할 필요도 있다. 또한 후속 연구에서는 인간번역과 포스트에디팅의 작업 과정에 대해 살펴봄으로써 품질과 생산성에 관한 또 다른 시사점을 얻을 수 있을 것으로 기대한다.

참고문헌

김순미, 신호섭과 이준호. (2019). 번역학계와 언어서비스업체(LSP)간 산학협력연구: 포스트에디팅 생산성과 기계번역 엔진 성능 비교. 번역학연구, 20(1), 41-76.

김자경. (2020). 포스트에디팅 결과물에 나타난 오류 고찰: AI 학습용 한영

- 번역 말뭉치를 대상으로. 통번역학연구, 24(4), 35-62.
- 김자경. (2021). 포스트에디팅 결과물의 정확성 오류 고찰: AI 학습용 금융/증시 분야 한-영 번역 말뭉치를 대상으로. 통번역학연구, 23(3), 29-58.
- 김자경. (2022a). 한영 포스트에디팅 과정에서의 노력 탐색 - 시간, 기술적 노력, 검색을 중심으로. 통번역학연구, 26(2), 1-24.
- 김자경. (2022b). 한영 포스트에디팅의 정확성 오류의 수정 양상 고찰. 번역학연구, 23(5), 91-117.
- 김혜림. (2021). 중한 법령 기계번역 포스트에디팅 교육을 위한 예비 연구. 번역학연구, 22(3), 65-98.
- 남보리. (2021). LSP 관점에서 바라본 기계번역 포스트에디팅의 전략적 지향점 - 스타일 가이드 지침 연구 및 LSP 기계번역 서비스 사례. 번역·언어·기술, 2, 5-43.
- 마승혜. (2018). 한영 기계번역 포스트 에디팅에 대한 경험적 고찰: 학부 교육 과정 및 결과를 중심으로. 통번역학연구, 22(1), 53-87.
- 서보현과 김순영. (2018). 기계번역 결과물의 오류유형 고찰. 번역학연구 19(1), 99-117.
- 서보현과 김순영. (2020). 번역사의 번역교육 경험이 포스트에디팅 결과물의 품질에 미치는 영향에 관한 사례연구. 번역학연구, 21(3), 63-91.
- 신지선. (2020). 기계번역 포스트에디팅에 관한 해외 연구 동향. 번역학연구, 21(4), 87-114.
- 윤미선, 김택민, 임진주와 홍승연. (2018). 영어 - 한국어 언어쌍에 적합한 포스트에디팅 가이드라인: 미래 포스트에디터를 위한 가이드라인과 그 사례. 번역학연구, 19(5), 43-76.
- 이상빈. (2017). 학부번역 전공자의 기계번역 포스트에디팅, 무엇이 문제이고, 무엇을 가르쳐야 하는가?. 통역과 번역, 19(3), 37-64.
- 이상빈. (2020). 기계번역에 관한 KCI 연구논문 리뷰: 인문학 저널 논문 (2011~2020년 초)의 논의내용과 연구방법을 중심으로. 통역과 번역, 22(2), 75-104.
- 이준호. (2018) 포스트에디팅 교육을 위한 포스트에디팅과 인간번역의 차이 연구. 통역과번역, 20(1), 73-96.

- 이준호. (2019). 신경망기계번역의 객관적 평가를 위한 예비연구: 자동평가와 수동평가의 균형점. *통번역학연구*, 23(3), 171-202.
- 이준호. (2021). 한영 포스트에디팅 노력 예비연구: 트랜스로그II를 활용한 한영 인간번역과 포스트에디팅의 차이분석. *번역학연구*, 22(5), 271-296.
- 이준호. (2022). 법률 특화 번역엔진 성능 평가: 한영 계약서 번역을 중심으로. *T&I Review*, 12(1), 169-192.
- 이준호와 김순미. (2022). 풀 포스트에디팅에 대한 고찰: 풀 포스트에디팅 생산성에 영향을 주는 요소를 중심으로. *번역학연구*, 23(5), 119-144.
- 이지은과 최효은. (2022). 한영 법령 번역 품질 제고를 위한 번역 방식에 관한 소고: 직역 문제를 중심으로. *통역과 번역*, 24(1), 243-276.
- 이지은과 최효은. (2023a). 인공신경망 기반 맞춤형 기계번역엔진의 성능 평가: 법률 및 특허 한영 번역 결과물 평가 사례를 중심으로. *번역학연구*, 24(1), 9-37.
- 이지은과 최효은. (2023b). 원천텍스트 난이도와 기계번역 품질에 대한 고찰: 구글 한영 법령번역 사례를 중심으로. *언어학연구*, 28(1), 77-101.
- 정재혁. (2023). 영한 포스트에디팅과 인간 번역의 정확성 비교 - 정확성 오류 검증을 중심으로. *통번역학연구*, 27(3), 177-201.
- 최문선. (2020). 국내 번역학 기계번역 연구 동향: 내용 분석과 키워드 분석을 중심으로. *언어학연구*, 24(1), 275-297.
- 한승희와 강동희. (2022). 기계번역과 인간번역, 기계번역 사후편집본과 인간번역 감수본의 품질평가 비교 연구 - 한-노 번역 사례를 중심으로. *통번역학연구*, 26(1), 195-219.
- 한현희(2020). 한-노 기계 번역, 어디까지 왔나?: Google과 Papago 번역 성능 비교를 기반으로. *노어노문학*, 32(3), 63-93.
- Arnejšek, M., & Unk, A. (2020). Multidimensional assessment of the eTranslation output for English-Slovene. In A. Martins, H. Moniz, S. Fumega, B. Martins, F. Batista, L. Coheur, C. Parra, I. Trancoso, M. Turchi, A. Bisazza, J. Moorkens, A. Guerberof, M. Nurminen, L. Marg & M. L. Forcada (Eds.), *Proceedings of the 22nd Annual Conference of the*

- European Association for Machine Translation* (pp. 383-392).
- Aranberri, N., Labaka, G., de Ilaraza, A. D., & Sarasola, K. (2014). Comparison of post-editing productivity between professional translators and lay users. In S. O'Brien, M. Simard & L. Specia (Eds.), *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas* (pp. 20-33).
- Brkić, M., Seljan, S., & Vičić, T. (2013). Automatic and human evaluation on English-Croatian legislative test set. *Computer Science*, 7816, 311-378.
- Defauw, A., Szoc, S., Vanallemeersch, T., Bardadym, A., Brabers, J., Everaert, F., Scholte, K., Van Winkel, K., & Van den Bogaert, J. (2019). Developing a neural machine translation system for Irish. In A. Karakanta, A. K. Ojha, C. H. Liu, J. Washington, N. Oco, S. M. Lakew, V. Malykh & X. Zhao (Eds.), *Proceedings of the 2nd Workshop on Technologies for MT of Low Resource Languages* (pp. 32-38). European Association for Machine Translation.
- DeMattee, A. J., Shibaike, T., Gertler, N., Bloodgood, E. A. (2022). Overcoming the laws-in-translation problem: Comparing techniques for translating legal texts. *Qualitative & Multi-Method Research*, 25(2), 13-21.
- Dik, H. (2020). *CTRL+V for Verdict: An analysis of Dutch to English legal machine translation* [MA thesis, Leiden University].
- Green, S., Heer, J., & Manning, C. D. (2013). The efficacy of human post-editing for language translation. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 439-448).
- Heiss, C. A., & Soffritti, M. (2018). DeepL traduttore e didattica della traduzione dall'italiano in Tedesco-Alcune valutazioni preliminari. *INTRALINEA TRAlinea Special Issue: Translation and Interpreting for Language Learners (TAIL)*, 20, 1-11. Available at <https://www.intraline.org/specials/article/2294>
- Horváth, I. (2024, January 12). *AI-Driven transformations in translation at the CdT* [Oral presentation]. 2024년 1월 12일 한국외국어대학교 통번역연

구소 제23회 국제학술대회, 한국.

- Jia, Y., Carl, M., & Wang, X. (2019). How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study. *The Journal of Specialised Translation*, 31(1), 60-86.
- Killman, J., & Rodríguez-Castro, M. (2022). PE vs translating in the legal context: Quality and time effects from English to Spanish. *Revista de Llengua I Dret/Journal of Language and Law*, 78, 56-72.
- Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. In T. Luong, A. Birch, G. Neubig & A. Finch (Eds.), *Proceedings of the First Workshop on Neural Machine Translation* (pp. 28-39). Association for Computational Linguistics.
- Lee, J., & Choi, H. (2023). A case study on the evaluation of Korean-English legal translations by generic and custom neural machine translation engines. *Interpretation and Translation*, 25(1), 75-98.
- Lommel, A. (2018). Metrics for translation quality assessment: A case for standardising error typologies. In J. Moorkens, S. Castilho, F. Gaspari & S. Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 109-128). Springer.
- Mitchell, L. (2015). The Potential and Limits of Lay Post-editing in an Online Community. In Í. D. El-Kahlout, M. Özkan, F. Sánchez-Martínez, G. Ramírez-Sánchez, F. Hollowood & A. Way (Eds.), *Proceedings of the 18th Annual Conference of the European Associations for Machine Translation*, European Associations for Machine Translation, 67-74.
- Roiss, S. (2021). Y las máquinas rompieron a traducir... Consideraciones didácticas en relación con la traducción automática de referencias culturales en el ámbito jurídico. *TRANS: Revista de Traductología*, 25, 491-505.
- Sosoni, V., O'Shea, J., & Stasimioti, M. (2022). Translating law: A comparison of human and post-edited translations from Greek to English. *Revista de Llengua I Dret, Journal of Language and Law*, 78, 92-120.

- Vardaro, J., Schaeffer, M., & Hansen-Schirra, S. (2019). Translation quality and error recognition in professional neural machine translation post-editing. *Informatics*, 6(3), 1-29.
- Vieira, L. N. (2017). From process to product: links between post-editing effort and post-edited quality. In A. L. Jakobsen & B. Mesa-Lao (Eds.), *Translation in Transition: Between Cognition, Computing and Technology* (pp. 161-186). John Benjamins.
- Wiesmann, E. (2019). Machine translation in the field of law: A study of the translation of Italian legal texts into German. *Comparative Legilinguistics*, 37(1), 117-153.

An analysis of errors in human translation, machine translation and post-edited machine translation outputs: A case of Korean-English legal translation

Jeun Lee (corresponding author, jieun.lee@ewha.ac.kr)

Graduate School of Translation and Interpretation, Ewha Womans University

Hyoem Choi

Graduate School of Translation and Interpretation, Ewha Womans University

Abstract

This paper addresses the quality of machine translation (MT) and post-editing (MTPE) outputs in comparison with human translation (HT) in the context of Korean-English legal translation, focusing on error analysis. Using the three major criteria of accuracy, fluency, and terminology, the study examines errors in twelve from-scratch translations and twelve post-edited MT outputs as well as the MT output generated by DeepL Translator. Error comments provided by three human evaluators were analyzed to identify error types and occurrences in the three different output types. Our findings reveal that MTPE generated the fewest errors, while MT exhibited the highest error count. Across all output types (HT, MT, and MTPE), accuracy errors were the most prevalent, with distortion being the predominant type. Fluency errors ranked next in frequency for HT and MTPE, while terminology errors were more common in MT than in HT and MTPE. Further analysis indicates that fluency and terminology errors were generally corrected in MTPE, but accuracy errors, particularly distortion, requiring substantial cognitive effort, often remained uncorrected. In terms of prevalent error types, MTPE demonstrated a closer resemblance to HT than MT. These findings underscore the potential of MTPE to yield high-quality legal text translations with reduced errors. However, HT remained advantageous in terms of producing accurate translations, particularly with regard to meaning transfer. This study contributes valuable insights to the ongoing discourse on the efficacy of different translation methods in the legal domain.

Keywords: Error analysis; human translation; legal translation; machine translation; post-edited machine translation

키워드: 오류 분석, 인간번역, 법률 번역, 기계번역, 포스트에디팅

이지은(1저자/교신저자, <https://orcid.org/0000-0002-6011-8786>)

이화여자대학교 통역번역대학원 교수

jieun.lee@ewha.ac.kr

최효은(공동저자, <https://orcid.org/0000-0002-7195-0682>)

이화여자대학교 통역번역대학원 초빙교수

hyoeun.choi@ewha.ac.kr

논문 투고일: 2024년 1월 30일

1차 심사 완료일: 2024년 3월 1일

2차 심사 완료일: 2024년 3월 10일

게재 확정일: 2024년 3월 15일

