

정보성 텍스트의 한영 기계번역 포스트에디팅 가이드라인 제시 —신경망 기계번역(NMT)을 사용한 뉴스 기사문 번역의 사례*

박 건 영
(한국외대)

1. 서론

“언어 간 다리 역할을 하던 통 번역가는 이제 인공지능(AI) 기계 번역으로 문화적 의미, 뉘앙스를 추가해 교정해주는 포스트 에디터(post editor)로 바뀔 것입니다. 번역에서 인간의 힘이 계속 필요합니다(김지영 2017. 2. 20).”

구글 번역, 네이버 파파고와 같은 신경망 기계번역(Neural Machine Translation, NMT)이 등장한 이래 급속한 성능 향상을 이루면서 이제 인간 번역사의 역할은 번역이 아닌 포스트에디팅¹⁾이 될 것이라는 의견이 커지고 있다.

* 본 논문의 초고는 2020년 9월 26일에 개최된 한국번역학회 가을 학술대회에서 발표되었음을 밝힙니다. 연구의 근간이 된 자료 사용을 허락한 (주)에버트란에 감사드립니다. 또한 연구 진행 과정에서 조언을 아끼지 않으신 성승은, 이상빈, 최희경 교수님께 진심으로 감사드립니다.

또한 기계번역 품질이 계속 향상하면서 포스트에디팅은 그리 어렵지 않은 작업 일거라는 인식도 크다. 그렇다면 실제로 기계번역 포스트에디팅은 손쉬운 작업이며 포스트에디팅된 문장은 완벽할까?

영한번역의 경우에는 이미 2017년에 신경망 기계번역 기술 도입으로 영어를 타 언어로 바꾼 것에 비해 높은 수준의 품질을 획득한 바 있다(한국경제TV 2017. 2. 9). 그러나 한영번역의 경우에는 사용자 수요가 많음에도 불구하고 데이터가 쌓인 만큼 발전하는 신경망 기계번역의 특성으로 인해 아직 그 완성도가 영한번역에는 미치지 못하고 있는 것이 현실이다. 따라서 한영 기계번역에서 포스트에디팅은 결과물 품질 향상을 위해 더욱 중요한 역할을 할 것으로 예상되며 따라서 기계번역 발전과 함께 포스트에디팅 수준 제고가 당연파제인 것으로 보인다.

연구자는 2019년 하반기에 한 국내 번역회사의 한영 뉴스 기사문 코퍼스 구축 프로젝트에서 감수 프로젝트 매니저로 근무하면서 다수의 한국어 기사 문장(이하 ST), 영문으로 번역된 기계번역 문장(이하 MT), 포스트에디팅된 문장(이하 PE), 최종 감수된 문장(이하 PR)을 접하게 되었다. 연구자는 프로젝트 작업 시작 당시에 참여하는 포스트에디터 전체에게 가이드라인이 배포되었음에도 감수 과정 중 PE 단계에서 나타난 수정 내역에서 전반적으로 개별 번역사 간 빈도 차이는 있으나 범주화 할 수 있는 유사성이 발견된다는 점에 주목했다. 이에 따라 수정된 사항²⁾의 종류를 구체적으로 살펴보고 범주화해 어떤 방향으로 가이드라인³⁾

- 1) 포스트에디팅은 사후 교정이라고도 하며 “기계번역 결과물을 사람이 교정하는 것 (Mossop 2010: 115)”으로 정의할 수 있다. 기계번역의 하위분야에는 기계번역 결과물을 컴퓨터가 자동으로 교정하는 Automatic Post-editing(APE)도 있으나(박찬준, 임희석 2020: 1) 본 연구의 대상인 포스트에디팅은 기계번역을 인간 번역사가 수정한 것이다.
- 2) 기존 선행 연구에서는 기계번역에서 포스트에디팅된 부분을 보통 ‘오류’라고 일컫고 있으나 연구자의 분석 자료에서는 오류가 아닌데도 포스트에디팅한 사례도 다수 발견되었으므로 본 연구에서는 ‘수정 사항’으로 일컫기로 한다.
- 3) 일반적으로 ‘가이드라인’은 보다 폭넓은 전반적 작업 방향을 지칭하고 ‘작업 지침’은 가이드라인 하의 세부사항을 일컫기도 하나 본 연구의 분석 자료를 추출한 프로젝트에서는 세부적 작업 지침을 가이드라인으로 지칭했으므로 본 연구에서도 이러한 세부적 작업 지침을 포괄한 내용을 가이드라인으로 정의하고 해당 용어를 사용하였다.

을 개선하면 추후 한영 포스트에디팅을 효과적으로 수행할 수 있을지 알아보고자 한다.

이를 위해 본 연구는 한국어 원문이 기계번역과 포스트에디팅, 감수를 마친 최종 단계에 이르기까지 수정된 부분을 각 단계별로 정성적으로 분석함으로써 어떤 부분이 수정되었는지 살펴본 후 분석 결과를 토대로 정보성 텍스트의 한영 기계번역 포스트에디팅 가이드라인을 제시하는 것을 목적으로 한다.

본 논문은 아래의 두 가지 연구 질문에 대해 고찰하고자 한다.

1. ST-MT-PE-PR의 병렬 코스에서 번역사가 MT에서 PE, 감수자가 PE에서 PR로 수정하는 과정에서 어떤 부분이 수정되었는가?

2. 실제 프로젝트 진행과정에서 얻은 자료를 토대로 정보성 텍스트의 한영 기계번역 포스트에디팅 가이드라인은 어떻게 제시될 수 있는가?

기존 한영 기계번역 연구는 한영 기계번역에서 특정 어휘의 번역 양상이나 오류 분석을 중심으로 했고, 포스트에디팅 연구는 주로 연구자가 연구를 위해 추출한 문장을 자료로 사용해 외국어 전공 학부생이나 대학원생이 주어진 문장을 어떻게 번역하는지를 분석하여 포스트에디팅의 교육 활용 방안이나 개선 방안 등을 논의해왔다. 그러나 본 연구는 실제 진행된 프로젝트에서 기계번역, 포스트에디팅, 감수를 거친 다량의 문장에 대한 질적 분석을 통해 한영 기계번역 포스트에디팅 실무에서 활용할 수 있는 가이드라인을 제시하고자 한다.

2. 선행연구 분석

국내 번역학에서 기계번역과 관련한 연구 중 주요 연구 흐름인 기계번역의 소비(consumption) 관점의 연구는 주로 기계번역 결과물의 품질 평가 및 오류 유형화, 기계번역 결과물과 인간번역의 비교, 포스트에디팅 프리에디팅, 기계번역 교육, 인간의 고유 영역 모색 등의 주제로 나눌 수 있다(최문선 2019: 285-287). 2017년 이후 기계번역을 다룬 연구는 양적으로 급격히 증가하고 있는 바, 본 장에서는 본 연구와 직접적으로 관련된 한국어-영어 언어 쌍 기계번역 오류 분석과 포스트에디팅 및 포스트에디팅 가이드라인 관련 선행 연구를 살펴보기로 한다.

기계번역의 오류에 관한 연구 중 박옥수(2017, 2018)의 연구는 한영 기계번역에서 출발어인 한국어의 유형적 특징에 따른 번역 오류를 분석했다. 박옥수(2017)는 세종 코퍼스에서 추출한 여러 형태의 문장을 유형적 특징에 따라 긴 문장, 주어 파악이 되지 않는 경우, 동음이의어나 다의어가 포함된 경우, 문장 내에서 의미해석이 되지 않는 경우, 맥락의 이해가 어려운 경우로 분류하여 분석했고 2018년의 연구에서는 오류 분석의 범위를 좀 더 확장해 음역, 시제, 생략, 반복, 관사나 전치사 오류, 한자어 등 다양한 유형의 오류를 다루었다. 본 연구에서는 이들 기계번역의 오류를 한영 MTPE 수정 항목 분류표의 항목에 다수 반영하되 이외 다른 국내외 연구의 오류 분석과 연구자의 분석 자료 사례도 반영했다.

서보현과 김순영(2018)은 영한 기계번역 결과물의 오류유형을 기존 번역품질 평가에서 활용하는 기준을 바탕으로 정확성, 가독성, 통사구조, 오타자로 범주화하고 그 하위 유형으로 어휘누락, 문장구조 오류, 의미 오류 등을 분류했다. 본 연구에서는 이를 수정 항목 분류에 반영하되 실제 작업에서 포스트에디터가 직관적으로 활용할 수 있는 가이드라인을 제시한다는 목적 하에 정확성, 가독성 등의 개념은 항목에서 배제하고 분석 항목의 범주를 더욱 구체화하였다.

김자경(2020)의 연구는 기계번역의 오류뿐만 아니라 포스트에디팅을 마친 결과물에 남아 있는 오류에 주목했다는 점에서 본 연구와 맥락을 같이 한다. 김자경(2020)은 한국정보화진흥원에서 2019년에 공개한 한영 번역 말뭉치의 문장 쌍을 추출해 오류가 발견되는 빈도를 파악했다. 이 연구는 그간의 연구가 기계번역의 오류에 주로 관심을 두었으나 포스트에디팅 결과물의 오류 양상에 대해서도 살펴보아야 하며 이와 같은 오류가 남게 되는 원인에 대해서도 점검해야 한다고 지적했다. 김자경(2020)의 분석 자료는 본 연구가 출발한 프로젝트와 동일한 정부 사업 내에서 진행된 포스트에디팅 결과물이므로 기존 선행 연구가 학생을 대상으로 실험 결과나 연구자가 추출한 자료를 사용한 것에 반해 실제 번역사가 작업한 결과물을 분석했다는 의의가 크다. 그러나 김자경(2020)의 연구가 대중에게 공개된 최종 산출물인 ST-TT에서 나타난 오류의 양상과 빈도를 파악한 반면 본 연구에서는 각 추출 문장 당 ST-MT-PE-PR의 전 과정을 분석해 각 단계 별로 어떤 부분이 수정되었는지를 알아보았다. 그리고 이를 바탕으로 포스트에디팅 작업을 할 번역사와 그 결과물을 수정할 감수자가 실제 작업

에서 활용할 수 있는 가이드라인을 제시하는 것을 목표로 하였다.

한편 아직까지 국내에서 기계번역 오류 연구에 비해 포스트에디팅 가이드라인을 제시한 연구는 그 수가 적은 실정이다. 포스트에디팅 가이드라인을 제안한 기존 연구로는 노한 기계번역의 프리에디팅 및 포스트에디팅 가이드라인(한현희 2019), 한일 기사 헤드라인 번역의 포스트에디팅 가이드라인(이주리에 2019)이 있으며 영어-한국어 쌍 기계번역 포스트에디팅 가이드라인으로는 윤미선 외(2018)의 연구가 유일하다. 윤미선 외(2018)의 연구는 언어구조가 유사한 언어 쌍에만 존재하는 포스트에디팅 가이드라인을 영어-한국어 쌍에 적용하는 데에는 한계가 있다는 문제의식 하에 영한 언어 쌍에 적합한 가이드라인을 제안했다. 이 연구는 해외 선행 연구에서 제시한 가이드라인을 비교해 정리한 후 국내 통번역 석사 과정 학생을 대상으로 실험한 포스트에디팅 과제의 기계번역 오류 분석을 바탕으로 수정한 영한 포스트에디팅 가이드라인을 제시해 한국어를 대상으로 한 포스트에디팅 가이드라인이 부재했던 상황에서 이를 포스트에디터가 실무에서 활용할 수 있도록 하였다. 본 연구에서도 윤미선 외(2018)의 연구에서 활용한 해외 선행 연구의 가이드라인 항목을 참고하였다. 그러나 윤미선 외(2018)의 실험 참여자는 실제 포스트에디팅 작업 종사자가 아닌 학생이며 약 1시간 작업 분량의 텍스트만을 포스트에디팅한 데 반해 본 연구에서는 약 5개월 간 번역 산업 현장에서 진행된 다량의 포스트에디팅 결과물을 바탕으로 가이드라인을 작성하고자 했다.

본 연구는 한영 기계번역의 수요와 사용이 증가하고 있으나 한영 기계번역 포스트에디팅 가이드라인을 제시한 연구가 아직 없다는 점을 인지하고 기계번역 결과물 품질 향상 및 생산성 향상을 위해 가이드라인을 제안할 필요성이 있다고 판단하였다. 그리고 보다 실질적인 가이드라인을 제시하기 위해 기존 포스트에디팅 연구가 주로 학생 등 한정된 실험 대상과 연구자가 연구를 위해 추출한 문장을 분석 자료로 사용했던 것과 달리 실제 프로젝트에서 다수 번역사가 작업한 다량의 결과물을 분석해 이를 토대로 한 가이드라인을 제안하고자 하였다. 이를 위해 기존 기계번역 오류 연구 및 포스트에디팅 가이드라인 연구에서 언급한 오류와 가이드라인을 본 연구의 가이드라인 항목 구성에 참고하는 동시에 연구자의 분석 자료에서 발견한 수정 사항을 반영해 한국어-영어 쌍에 맞게 범주화하여 가이드라인을 작성하였다.

3. 연구 방법

본 연구는 한영 코퍼스 구축 프로젝트에 사용된 한국어 뉴스 기사 문장과 영어 기계번역 문장, 포스트에디팅된 문장, 최종 감수 문장의 병렬 코퍼스를 분석하였다. 먼저 분석 자료에 대한 이해를 돕고자 본 연구의 계기가 된 코퍼스 구축 프로젝트 과정과 이에 따라 취득한 분석 자료를 살펴 본 후 분석방법을 설명하고자 한다.

3.1 분석자료

연구자가 참여한 프로젝트에서 각 번역사는 아래의 그림 1과 같이 태스크의 단위로 신경망 기계번역기로 번역된 문장을 배정받아 포스트에디팅을 했다. 한 개 태스크는 총 200개의 문장으로 구성되었으며 사용한 신경망 기계번역기는 국내에서 가장 많이 사용하는 기계번역기 중 한 종류였다.⁴⁾ 프로젝트 시작과 함께 MS 워드 문서 2쪽 분량의 포스트에디팅 가이드라인이 번역사와 감수자에게 배포되었으며 프로젝트 진행과정에서 감수자가 더 자세한 가이드라인을 작성해 재배포하였다. 재 배포된 가이드라인은 프로젝트 진행 과정에서 번역사의 피드백을 반영해 수정되었다.

그림 1 포스트에디팅 메인 화면⁵⁾

No	태스크명	상태	진행률	시작일	종료일	피드백	리뷰정수
3750	20190918_ ()-1	Proofread	75 / 150 / 200	2019-09-18	2019-09-24		90
3748	20190918_ ()-3	Proofread	31.5 / 100 / 200	2019-09-18	2019-10-12		95
3747	20190918_ ()-2	Proofread	45.5 / 90 / 200	2019-09-18	2019-10-06		93
3746	20190918_ ()-1	Proofread	43 / 86 / 200	2019-09-18	2019-09-29		90
3736	20190917_ ()-1	Proofread	40 / 80 / 200	2019-09-17	2019-11-19		95
3735	20190917_ ()-3	Proofread	100 / 200 / 200	2019-09-17	2019-09-26		80

- 4) 해당 기업의 요청에 따라 신경망 기계번역기의 종류는 밝히지 않기로 한다.
5) 개인정보 보호를 위해 개별 번역사의 이름과 이메일 주소는 삭제하였다.

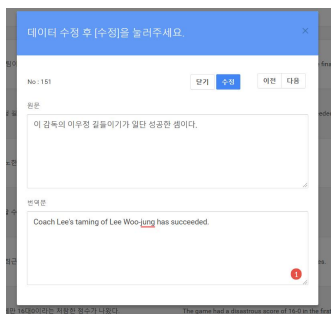
개별 번역사는 그림 1의 메인 화면을 클릭한 후 그림 2의 태스크 내에서 한국어 원문과 대조하면서 신경망 기계번역기로 번역된 번역문을 포스트에디팅 하게 된다.

그림 2 포스트에디팅 태스크 내 화면

No	원문	번역문
1	한화생명 선수들의 2019년 프로필 사진과 슬로건, 팀 컬러가 적용됐다.	The 2019 profile pictures of Hanwha Life's players, slogan and team color were applied.
2	합쳐 찍은 사진들을 결합하면 그 자체가 움직이는 영상이 된다.	Combining the pictures taken together makes itself a moving image.
3	합리적인 소비자가 퇴직연금 시장에서는 비합리적인 선택을 하고 있는 셈이다.	In other words, reasonable consumers are making unreasonable choices in the retirement pension market.
4	합병 이후 콘텐츠 경쟁력 확대 등이 가능할 것으로 나타났다.	After the merger, the company expects to expand its content competitiveness.
5	항암제가 정상세포의 자살 사멸을 유도하는 메커니즘에 대한 국내외 특허도 출원됐다.	It also applied for domestic and international patents on the mechanism by which anti-cancer drugs induce suicide and death of normal cells.

번역사가 그림 2의 문장 중 수정이 필요한 번역문을 클릭하면 화면에 나타나는 그림 3과 같은 팝업창에서 번역문을 포스트에디팅할 수 있다. 한 태스크 내 200개의 기계번역문 중 수정이 필요한 문장을 모두 포스트에디팅 완료하면 완료된 태스크는 감수자의 태스크로 이동한다. 감수자도 그림 1, 그림 2, 그림 3의 흐름에 따라 각 태스크의 번역문을 감수하게 된다.

그림 3 포스트에디팅 작업창



본 연구를 위해 먼저 연구자는 약 20어절 내외의 한국어 뉴스 기사문 중 11명의 번역사에게 배정된 태스크에서 한 번역사당 25개 문장을 추출해 구성된 275개 병렬 세그먼트 중 한글 원문 코퍼스를 ST로 명명했다. 실제 프로젝트에서는 한국어 원문이 7어절, 9어절인 태스크도 있었으나 본 연구에서는 일상에 많이 쓰이는 20어절 내외의 문장으로 구성된 태스크를 택해 다양한 오류 및 수정 사례를 분석하고자 했다. 병렬 코퍼스 문장은 프로젝트 중 IT 스포츠 분야에서 6명 번역사의 태스크와 IT 국제 분야 기사문에서 5명 번역사의 태스크에서 추출하였다. 프로젝트에 사용된 뉴스 기사문의 분야는 다양했지만 본 연구의 분석 대상은 IT, 국제, 스포츠 분야로 한정했다. 스포츠 분야 뉴스 기사에도 선수 이름, 팀 이름, 지역 이름 등의 고유 명사가 포함되나, 한국 국내 경제, 사회 등을 다룬 기사 문장의 경우 기관명을 비롯한 고유명사 및 은유적 표현, 문화특수성을 지닌 어휘 등이 더욱 많이 등장하므로 본 연구에서는 정보성 텍스트 중에서도 이러한 어휘 및 표현이 상대적으로 적은 분야를 분석해 기계번역과 포스트에디팅의 일반적 오류 경향을 분석한 기본적 가이드라인을 제시하고자 했다.

또한 개별 번역사의 번역 능력으로 인한 오류의 가능성을 줄이고자 감수자의 포스트에디팅 결과를 평가에서 일정 수준 이상의 평가를 받은 번역사의 포스트에디팅 문장만을 선별해 코퍼스를 구축했다. 평가는 표기법 등의 가이드라인 미준수, 기계번역 미수정, 오역, 누락 등의 유무를 기준으로 100점 만점에서 감점하는 방식으로 시행되었으며 각 번역사의 결과물에서 주로 반복되는 부분에 대해서는 피드백을 제공했다. 이 점수를 바탕으로 각 분야에서 80점 이상의 점수를 받은 번역사 중 11명이 포스트에디팅한 문장을 원문이 중복되지 않는 범위 내에서 무작위로 각 번역사 당 25개 문장을 추출했다.

<표 1> 분석자료 설명

명칭	내용
ST	한국어 뉴스 기사문 원문 문장
MT	신경망 기계번역기를 사용한 영어 번역문
PE	기계번역문을 포스트에디팅한 문장
PR	포스트에디팅한 문장을 감수한 문장

<표 1>은 분석자료의 명칭과 그 내용을 설명한 것이다. 실제 프로젝트 과정에서 번역사가 포스트에디팅한 문장을 내부 감수자가 감수한 후 다시 외부 감수자가 최종 감수하는 단계를 거쳤으나 본 연구에서는 내부 감수자가 감수한 문장과 외부에서 최종 감수한 문장 간에 큰 차이가 없는 것을 확인하고 최종 감수문을 포함하는 것이 유의미한 변화를 도출하지 않는다는 판단 하에 PR을 ‘포스트에디팅 후 감수를 거친 문장’으로 정의하였다.

3.2 분석방법

분석 자료를 구축한 뒤 ST-MT-PE-PR의 병렬 세그먼트 275개에 대해 정성적 분석을 실시했다.⁶⁾ 본 연구는 풀 포스트에디팅(full post-editing)⁷⁾을 기준으로 오류를 분석했고 포스트에디팅 최종 결과물은 MT를 포스트에디팅한 PE가 아닌 PE를 감수한 PR이므로 바람직한 포스트에디팅 가이드라인을 제시하기 위

6) 본 연구를 진행하기에 앞서 실행한 파일럿 연구에서 11명 번역사의 태스크 중 각 5 문장을 추출한 55개 문장에 대해 ST-MT-PE-PR의 코퍼스를 WordSmith 툴을 사용해 분석한 결과, 문장 당 평균 단어수와 TTR은 아래와 같이 나타났다.

코퍼스	Tokens	Types	TTR	mean(in words)
ST	5,891	4,011	68.09%	20.74
MT	8,587	2,704	31.49%	27.88
PE	9,036	2,953	32.68%	29.92
PR	9,281	2,967	31.97%	30.53

양적 분석 결과 ST-MT-PE-PR의 단계로 원문이 수정되면서 각 코퍼스의 단어 수는 token과 type에서 모두 늘어났고 TTR은 한국어 ST에서는 68.09%였으나 MT, PE, PR에서는 유사한 수치를 보였으며 전반적으로는 PR에서 가장 높은 수치를 나타냈다. 문장 당 평균 단어 수 역시 수정 단계를 거치면서 최종 단계인 PR에서 가장 큰 수치를 보였으나 의미 있는 차이가 나타나지는 않았다. 한편 SPSS 툴을 사용해 ST-MT-PE-PR의 상관관계를 분석하여 어느 코퍼스 간 상관관계가 큰 지, 그리고 Regression 분석으로 ST-MT-PE-PR 간 종속-독립 변수 관계가 있는지 살펴보았으나 각 코퍼스 간 유의미한 결과가 도출되지 않았다. 이에 따라 본 연구에서는 정성적 분석을 통해 수정 과정에서 어떤 점이 수정되었는지 찾아보기로 한다.

7) 2017년에 출범한 포스트에디팅 국제 표준 ISO 18587: 2017에서는 인간 번역물에 상응하는 결과물을 얻기 위한 프로세스를 풀 포스트에디팅(full post-editing), 이해 가능한 번역물을 산출하기 위한 프로세스를 라이트 포스트에디팅(light post-editing)으로 구분한다(곽중철, 한승희 2018: 3).

해서는 PE 뿐 아니라 PR까지 살펴야 한다고 보았다. 연구자는 MT→PE, PE→PR 단계에서 나타난 수정 사항을 분석하기 위해 기계번역 오류 분석 및 포스트에디팅 가이드라인에 대한 선행연구(곽중철, 한승희 2018; 박옥수 2017; 2018; 서보현, 김순영 2018; 윤미선 외 2018; De Almeida, Giselle and O'Brien 2010; Hu and Cadwell 2016; Mossop 2010; Pérez 2012) 및 The LISA QA Model, TAUS Guidelines를 참고하고 연구자가 참여한 프로젝트에서 발견한 수정 사례를 반영해 범주화 했다. 그리고 275개 병렬 세그먼트를 엑셀 파일에 옮겨 정성 분석한 후 분류표의 코딩기호에 따라 코딩하여 어느 항목이 많이 수정되었는지 살펴본 후 가이드라인을 제시했다.

본 연구에서 수정 항목을 분류하는데 사용한 분류표는 <표 2>와 같다.

<표 2> 한영 MTPE 수정 항목 분류표

항목	세부 항목	코딩기호
형식	문장부호	F1
	대 소문자	F2
	단위(통화 거리 무게 비율 온도 등)	F3
	숫자 시간 날짜 요일	F4
	인명 지명(자연 인공 행정)	F5
오탈자	띄어쓰기	T1
	수치	T2
	철자	T3
	중복	T4
어휘	고유명사 공식명칭	L1
	동음의미어 다의어	L2
	단어 오역(직역 포함)	L3
	긍정 부정 표현	L4
	음역	L5
	구절 오역	L6
	단어 변경	L7
	표현 변경	L8
문법	동사 시제	G1
	성 수	G2
	관사(추가 삭제 변경)	G3
	전치사(추가 삭제 변경)	G4
	능동 수동	G5

〈표 3〉 PE와 PR 수정 항목 분석 결과

구조	인과 관계	S1
	수식 관계	S2
	주술 관계	S3
	어순	S4
	구 절 구조 변경	S5
추가	직위 호칭	A1
	고유명사(인명 포함)	A2
	단어	A3
	구 절	A4
	원문에 없는 부분	A5
삭제	어휘에 대한 부가 설명	O1
	원문 의미 전달과 관계없는 부분	O2
	원문에 있는 부분	O3

본 연구에서는 MTPE 분류표의 항목을 형식, 오타자, 어휘, 문법, 구조, 추가, 삭제의 7가지 큰 항목으로 분류한 후 다시 총 35개의 세부 항목으로 나누었다. 선행 연구의 오류 분류에 비해 세부 항목을 자세히 분류한 이유는 비전문 번역사를 포함한 다수 인력이 투입되는 프로젝트이므로 전체적으로 통일된 형태로 포스트에디팅된 문장을 산출할 수 있는 구체적 실무 가이드라인을 제시하기 위해서이다.

4. 분석 결과

4.1 PE와 PR의 수정 내용 분석

〈표 2〉의 수정 항목 분류표에 따라 실제 프로젝트에서 추출한 275개 문장에 대해 코딩 분석으로 MT→PE, PE→PR의 각 단계별로 어떤 오류가 수정되었는지 살펴본 결과는 다음의 〈표 3〉과 같다. 한 문장에서 동일 종류의 수정 내역이 2개 이상 있는 경우에는 개수대로 기재하였다. 〈표 2〉의 7개 대분류 안에서 세부 항목은 PE 기준 수정 빈도가 높은 항목 순으로 기재하였다.

항목	세부 항목	수정 사항(개)	
		PE	PR
형식	대 소문자	63	31
	문장부호	59	59
	인명 지명(자연 인공 행정)	52	20
	숫자 시간 날짜 요일	23	10
	단위(통화 거리 무게 비율 온도 등)	9	25
오타자	띄어쓰기	98	24
	중복	8	2
	철자	2	3
	수치	2	1
어휘	단어 변경	216	6
	표현 변경	240	16
	단어 오역(직역 포함)	75	39
	고유명사 공식명칭	78	55
	구 절 오역	12	4
	동음이의어 다의어	8	5
	긍정 부정 표현	6	1
	음역	4	6
문법	성 수	42	12
	동사 시제	38	4
	관사(추가 삭제 변경)	37	7
	전치사(추가 삭제 변경)	23	3
	능동 수동	21	3
구조	구 절 구조 변경	120	12
	어순	93	7
	주술 관계	33	16
	수식 관계	10	1
	인과 관계	6	0
추가	구 절	106	21
	단어	95	27
	고유명사(인명 포함)	26	10
	직위 호칭	19	14
	원문에 없는 부분	6	0
삭제	원문에 있는 부분	36	1
	원문 의미 전달과 관계없는 부분	18	7
	어휘에 대한 부가 설명	10	2
	총계	1,694	454
문장 당 평균 수정 사항(개)		6.2	1.7

275개 병렬 세그먼트의 PE와 PR 수정 내역을 분석한 결과 PE에서 MT가 수정된 사항의 개수는 총 1,694개였고 PR에서 PE가 수정된 사항의 개수는 총 454개였다. PE에서는 문장 당 평균 6.2개의 사항이 수정되었고 PR에서는 평균 1.7개의 사항이 수정되었다. 따라서 PE에서 MT의 오류 중 많은 부분이 수정되었고 PR에서는 수정된 항목의 수가 훨씬 감소했다고 해석할 수 있다. 총 35개의 세부항목 중 PE에서 가장 많이 수정된 항목은 표현 변경, 단어 변경, 구 절 구조 변경, 구 절 추가, 띄어쓰기 순이었고, PR에서 가장 많이 수정된 항목은 문장부호, 고유명사 공식명칭, 단어 오역, 대소문자, 단어 추가였다. 전반적으로 PE에서 수정된 항목은 PR에서 수정 빈도가 감소하는 경향을 보였으나 단위, 철자, 음역 항목에서는 PR의 수정 빈도가 더 높았다.

<표 4> PE와 PR 고빈도 수정 사항

PE		PR	
세부 항목	빈도	세부 항목	빈도
표현 변경	240	문장부호	59
단어 변경	216	고유명사 공식명칭	55
구 절 구조 변경	120	단어 오역	39
구 절 추가	106	대 소문자	31
띄어쓰기	98	단어 추가	27
단어 추가	95	단위	25
어순	93	띄어쓰기	24
고유명사 공식명칭	78	구 절 추가	21
대 소문자	63	인명 지명	20
문장부호	59	표현 변경, 주술관계(동순위)	16

<표 4>는 <표 3>에 나타난 PE와 PR 각각의 고빈도 수정 사항을 빈도가 높은 순으로 10순위까지 제시한 것이다. PE에서는 표현 변경, 단어 변경, 구 절 구조 변경, 구 절 추가 등 내용면의 수정이 고빈도 수정 사항인 반면 PR에서는 문장부호, 고유명사 공식명칭, 단어 오역, 단어 추가, 단위 등 형태면의 수정도 고빈도 수정 사항에 속해 있음을 알 수 있다.

주목할 점은 PE, PR에서 수정된 내용이 반드시 ‘오류’는 아니라는 점이다. 위의 세부항목 중 인명, 긍정 부정 표현, 단어 변경, 표현 변경, 시제, 성 수, 전

치사, 능동 수동, 어순, 구 절 구조 변경, 어휘에 대한 부가 설명 삭제는 MT, PE의 오류와 비오류에 대한 수정을 모두 포함한다.

<표 5> PE와 PR 수정 내역에 따른 항목 분류

항목 분류(개)		사례 빈도(개)	
		PE	PR
오류 수정 항목	22	790	367
오류 비오류 수정 항목	11	863	86
비오류 수정 항목	2	41	1
총계(개)	35	1,694	454

<표 5>에서는 <표 3>의 세부 항목을 오류 수정 항목, 비오류 수정 항목, 오류와 비오류를 모두 수정한 항목으로 분류하였다. PE에서 수정된 내용 중 MT의 오류만을 수정한 항목은 35개 항목 중 22개로 전체 1,694개 사례 중 절반 정도인 790개로 나타났으나 오류와 비오류를 모두 수정한 11개 항목의 사례가 절반 이상인 863개였다. 또한 명확한 비오류 수정만을 포함한 항목인 원문에 없는 요소가 첨가되거나 원문에 있는 요소가 삭제된 사례도 41개였다. PR의 경우에는 총 454개 수정 사례 중 367건이 오류 수정 사례였으나 86건은 오류와 비오류를 모두 수정한 항목에 속했다. 한편 원문에 있는 요소가 PR에서 삭제된 사례도 1건 있었다.

MTPE의 수정 사례를 구체적으로 살펴보면 아래와 같다. 본고에서는 지면상의 한계를 인해 총 여섯 가지 사례를 대표로 제시하였다. PE와 PR에서 수정된 전체 항목은 각 표의 PE와 PR 문장 하단에 기재하고 표 아래에는 고빈도 수정 항목 위주로 설명하였다.

<표 6> MTPE 수정 사례 1: 표현 변경, 띄어쓰기, 문장부호, 인명, 주술관계, 동음이의어

ST	노규덕 외교부 대변인은 정례브리핑에서 ‘조사 결과가 나오면 진행될 수 있는 미국의 양자제재에 대해 미국 측과 논의된 것이 있다’ 라는 질문에 이같이 답했다.
MT	Foreign Ministry spokesman Roh Kyu-deok answered in a regular briefing, “Has the U.S. discussed with Washington any possible bilateral sanctions that could take place when the investigation results come out?”
PE	Foreign Ministry spokesman Noh Kyu-deok responded to the question

	during the regular briefing: “Are there any discussions with the US about quantum sanctions in the United States that could proceed if the findings come out?”
PE 수정 항목	띄어쓰기, 문장부호, 인명, 표현 변경, 동음이의어, 관사 변경, 전치사 변경, 주술 관계, 구 추가
PR	During the regular briefing, Foreign Ministry spokesman Noh Kyu-deok responded to the question, “Are there any discussions with the US about bilateral sanctions in the United States that could proceed if the findings come out?”
PR 수정 항목	문장부호, 동음이의어, 어순

<표 6>의 사례에서 ST의 “질문에 이같이 답했다”를 MT는 “질문에”는 누락하고 “answered”로 번역했으나 PE에서는 누락 부분이 추가되고 “responded to the question”으로 표현이 변경되었다. 한편 MT에서 큰 따옴표 옆에 생성된 빈 칸은 PE에서 삭제되었다. PE에서는 직접인용문 앞에 ST와 MT에는 없는 구두점인 콜론이 추가되었는데 PR에서 쉼표로 수정되었다. ST의 인명 “노규덕”은 MT에서는 “Roh Kyu-deok”으로 번역되었고 PE에서는 “Noh Kyu-deok”으로 수정되었다. 내용면에서 ST의 “미국의 양자제재에 대해 미국 측과 논의된 것이 있나”라는 부분을 MT는 “Has the U.S. discussed with Washington any possible bilateral sanctions”로 번역해 “한국과 미국이 양자제재를 논의한다”는 원문의 의미를 “미국과 미국”이 논의하는 것으로 오역했다. 이를 PE는 “Are there any discussions with the US”로 수정했으나 “bilateral sanctions”로 번역되어야 하는 “양자제재”를 “quantum sanctions”로 오역했고 PR에서는 올바르게 수정되었다.

<표 7> MTPE 수정 사례 2: 단어 변경, 구조 변경, 고유명사

ST	김선민 대표는 “서울과학기술지주가 투자 차원을 넘어서 자회사와 함께 비즈니스모델을 만들어 사업화를 진행하고 수익을 공유해 좋은 일자리 창출 R&D 사업화 지역동반성장을 <u>동시</u> 이루겠다”고 말했다.
MT	“Beyond investment, we are planning to create business models with our subsidiary and achieve good job creation, R&D and regional growth by sharing profits.” said CEO Kim Sun-min of Seoul National University of Science and Technology.
PE	CEO Kim said, “Seoul and Expected Technology Holdings would create a business model to proceed with commercialization and share profits to simultaneously create good jobs, commercialize R & D,

	and coexist with local growth, with subsidiaries, going beyond investment.”
PE 수정 항목	띄어쓰기, 고유명사, 단어 변경, 표현 변경, 단 복수 변경, 수식 관계, 주술 관계, 어순, 구조 변경, 단어 추가, 구 추가, 원문 내용 삭제
PR	CEO Kim said, “Seoul National University of Science and Technology’s technology holding company would create a business model to proceed with commercialization and share profits to simultaneously create good jobs, commercialize R&D, and coexist with local growth, with subsidiaries, going beyond investment.”
PR 수정 항목	띄어쓰기, 고유명사

<표 7>의 사례에서 ST의 “지역동반성장”을 MT는 “regional growth”로 번역했는데 번역사는 이를 PE에서 “local growth”로 수정했다. 그리고 ST의 “김선민 대표는 “서울과학기술지주가... 이루겠다”고 말했다”를 MT는 “... said CEO Kim Sun-min of Seoul National University of Science and Technology.”로 번역해 “서울과학기술지주의 김선민 대표”를 의미하는 잘못된 수식관계를 구성하였다. 이를 번역사는 “CEO Kim said, “Seoul and Expected Technology Holdings”로 구조는 원문에 맞게 변경하였으나 주어의 이름을 누락하고 성만 기재했고 “서울과학기술지주”를 직역하여 원문과 다른 의미가 되었다. 김수자는 이를 “Seoul National University of Science and Technology’s technology holding company”로 원문의 의미에 맞게 수정했으나 PE에서 누락된 인명의 이름 부분은 추가하지 않았다.

<표 8> MTPE 수정 사례 3: 단어 추가, 구 추가, 어순, 단어 오역 수정

ST	뉴욕타임스(NYT)는 “대통령의 조치(행정명령)가 수정헌법을 무효화할 수 없다”며 수정헌법은 의회나 각 주(州)에서 압도적 다수의 판단에 의해서만 바뀌거나 무효로 할 수 있다고 지적했다.
MT	The New York Times said the president’s action can not override the Constitution, and that the Constitution can be changed or nullified only by the overwhelming majority in Congress and states.
PE	The New York Times(NYT) said the president’s executive order cannot override the Constitution, and that the Constitution can only be changed or invalidated by the overwhelming majority of judges in Congress and states.
PE 수정 항목	띄어쓰기, 단어 변경, 어순, 단어 추가, 구 추가, 원문 내용 삭제

PR	<u>The New York Times (NYT) said, “the presidential measure (executive order) cannot override the amendment,” and pointed out, “the amendment can only be changed or invalidated by the overwhelming majority of judges in Congress and states.”</u>
PR 수정 항목	문장부호, 띄어쓰기, 단어 오역, 단어 추가, 구 추가

<표 8>의 사례에서 ST의 “(NYT)”는 MT에서는 삭제되었는데 PE에서 다시 추가되었고 PR에서는 영어 어법에 맞게 “The New York Times”와의 사이에 빈 칸을 추가했다. ST의 “압도적 다수의 판단에 의해서만 바뀌거나 무효로 할 수 있다”를 MT는 “can be changed or nullified only by the overwhelming majority”로 번역했는데 PE에서는 어순이 “can only be changed or invalidated by the overwhelming majority of judges”로 변경되고 “of judges”가 추가되었다. 또한 ST의 “수정헌법”이 MT에서는 “Constitution”으로 번역되고 PE에서는 수정되지 않았는데 PR에서는 원문의 뜻에 맞는 “amendment”로 수정되었다.

<표 9> MTPE 수정 사례 4: 단·복수 변경, 절 추가

ST	기존에는 원하는 시간의 영상을 보기 위해 스마트폰 화면에 날짜, 시간 등을 손으로 입력해야 하는 불편이 있었는데 이를 해소했다는 설명이다.
MT	Previously, users had to type in the date and time on the smartphone screen in order to see the images of the time they wanted.
PE	In the past, the user had to enter a date, time, etc. on the smartphone screen to view a desired time of video, but they have solved this.
PE 수정 항목	단어 변경, 표현 변경, 단 복수 변경, 단어 추가, 절 추가
PR	In the past, the user had to enter a date, time, etc. on the smartphone screen to view a desired time of the video, but they explained that they have solved this.
PR 수정 항목	절 추가

<표 9>에서 한글 원문에서 명시되지 않은 주어가 MT에서는 복수형인 “users”로 번역되었고 PE에서는 “the user”로 변경되었다. 이와 같이 변경하지 않아도 의미상 문제가 없는 단 복수형의 변경에 대해 가이드라인에서 의미에

영향을 미치지 않는 수정을 지양하도록 명시할 필요가 있다. 한편 ST의 “이를 해소했다는 설명이다”는 MT에서 누락되었으나 PE에서는 “they have solved this”로 일부 추가되었다. 그리고 PR에서는 “they explained that they have solved this”로 MT에서 누락된 부분이 모두 추가되어 각 단계를 거치며 번역문의 수준이 향상되었다.

<표 10> MTPE 수정 사례 5: 고유명사, 숫자, 직위 추가

ST	논란이 확대되자 송 의원은 페이스북에 중국 미세먼지 문제를 언급하며, 노후 석탄화력발전소를 폐쇄하고 신한울 3·4호기 건설을 재개 하자는 주장을 이어갔다.
MT	As the controversy escalated, Song continued to insist on shutting down his aging coal-fired thermal power plant and resuming construction of his new plant, citing China's fine dust problem on Facebook.
PE	As the controversy escalated, Song continued his argument on Facebook to shut down old coal-fired power plants and resume construction of Shin Hanul's 3rd and 4th reactors, referring to China's fine dust.
PE 수정 항목	고유명사, 구 오역, 표현 변경, 단 복수 변경, 어순, 구 추가, 원문 의미 전달과 관계없는 부분 삭제, 원문에 있는 부분 삭제
PR	As the controversy escalated, Rep. Song continued his argument on Facebook to shut down old coal-fired power plants and resume construction of Shin Hanul 3 and 4 reactors, referring to China's fine dust issue.
PR 수정 항목	숫자, 직위 추가, 단어 추가

<표 10>의 예시는 ST의 “송 의원”이 MT에서 “Song”으로 직위가 누락되었는데 PE 단계에서 수정되지 않은 사례이다. 감수자는 누락된 부분을 추가해 “Rep. Song”으로 수정했다. 이와 같이 한국어 원문의 직위가 MT에서 누락된 사례는 분석 결과에서 빈번히 발견되었으므로 PE 단계에서 이에 대해 신경 써야 하며 해당 부분을 가이드라인에서 강조할 필요가 있다. ST의 “신한울 3·4호기”는 MT에서는 “his new plant”로 오역 및 누락되었는데 PE에서는 “Shin Hanul's 3rd and 4th reactors”로 고유명사로 수정하고 “3·4호기” 부분도 추가했다. 이를 PR 단계에서는 원문의 뜻에 맞게 소유격을 삭제하고 서수를 기수로 변경해 “Shin Hanul 3 and 4 reactors”로 수정하였다.

<표 11> MTPE 수정 사례 6: 관사 삭제, 전치사 변경, 동음이의어

ST	김현진 NH선물 연구원은 “미국과 중국의 무역 갈등은 국제 교역량 위축에 대한 우려를 키우며 국가 경제에서 수출 비중이 높은 아시아 통화 약세 압력을 높이는 요인”이라고 말했다.
MT	“The trade conflict between the U.S. and China raises concerns over the possible contraction of international trade and increases the pressure on the weak Asian currency, which has a high portion of exports in the national economy,” said Kim Hyun-jin, a NH Gift researcher.
PE	“The trade conflict between the US and China raises concerns about the contraction of international trade volume and increases pressure on the Asian currency to weaken, which has a high proportion of exports in the national economy,” said NH Futures Research Institute Kim Hyun-jin.
PE 수정 항목	문장부호, 띄어쓰기, 고유명사, 동음이의어, 표현 변경, 관사 삭제, 전치사 변경, 어순, 단어 추가, 원문에 없는 내용 삭제
PR	“The trade conflict between the US and China raises concerns about the contraction of international trade volume and increases pressure on the Asian currency to weaken, which has a high proportion of exports in the national economy,” said NH Futures Researcher Kim Hyun-jin.
PR 수정 항목	동음이의어

<표 11>의 예시에서 MT의 “concerns over”는 PE에서 “concerns about”으로 수정되어 전치사가 변경되었고 “increases the pressure”는 “increases pressure”로 관사가 삭제되었다. 한편 ST의 “NH선물 연구원”을 MT에서는 “NH Gift researcher”로 번역해 “선물”을 동음이의어인 “Gift”로 오역했다. 번역사는 이를 “NH Futures Research Institute”으로 수정해 “선물”은 원문의 뜻인 “Futures”로 수정했으나 “연구원”을 동음이의어인 “Research Institute”로 수정했다. 이를 PR 단계에서는 “NH Futures Researcher”로 수정해 “선물”과 “연구원”을 모두 원문의 뜻에 맞게 수정했다. PE에서 수정되지 못한 기계번역의 오역이 PR에서 수정된 사례이다.

<표 3>과 <표 4>의 분석 결과와 <표 5>의 수정 내역 분류, <표 6>에서 <표 11>까지의 사례를 살펴 본 결과, 원문이 기계번역, 포스트에디팅, 감수를 거치는 과정에서 기존 한영 기계번역 오류 연구에서 지적한 동음이의어, 통사 구조 등으로 인한 오류 수정뿐만 아니라 띄어쓰기, 문장부호, 어순, 누락 추가 등 다양한 수정 사항이 발견되었다. 또한 포스트에디팅과 감수 과정에서 수정

이 필요한 사항을 놓치거나 잘못 수정한 부분, 또한 의미나 형태의 개선 측면에서 수정이 필요하지 않으나 수정한 부분도 있음을 알 수 있었다.

4.2 정보성 텍스트의 한영 기계번역 포스트에디팅 가이드라인 제시

앞에서 살펴본 한국어 원문에서부터 기계번역, 포스트에디팅, 감수를 거치는 과정에서 나타난 수정 사항을 토대로 연구자가 제시하는 정보성 텍스트의 한영 기계번역 포스트에디팅 가이드라인은 다음과 같다.

<표 12> 정보성 텍스트의 한영 기계번역 포스트에디팅 가이드라인

항목	세부 항목	가이드라인
형식	문장 부호	영어 문장부호 표기법에 맞게 전환하고 누락된 문장부호는 추가한다. (ST) 통상교섭본부장이 “일본이 수산물 수입을 요구할 수 있다”고 밝혔다. (MT) Trade Minister said, Japan could demand the import of fisheries products. (PE) Trade Minister said, “Japan could demand the import of fisheries products.”
	대 소 문자	고유명사와 일반명사를 구분하고 직위는 대소문자 용법에 적합하게 표기한다. (ST) 낄슨 장관은 비서실장을 지냈다. (MT) Nielson had served as secretary of state. (PE) Secretary Nielson served as Chief Secretary.
	단위	단위는 앞에 한 칸을 띄고 기호로 표기한다. 단 % (퍼센트), ° (각도), °C, °F (온도)는 붙여서 표기한다. 숫자에는 세 자리마다 쉼표(.)를 찍는다. (ST) 5000미터 (MT) 5000 meter (PE) 5,000 m 통화는 통화 코드 뒤에 금액 수치를 기재하여 표기한다. (ST) 141.4%(4천 100만원) (MT) 141. 4 percent (41 million won) (PE) 141.4% (KRW 41 million)
	숫자 시간 날짜 요일	날짜는 표기법을 지키고 날짜가 요일로 번역되거나 원문에 없는 요소가 추가된 경우 수정한다. 숫자는 통일된 표기법을 사용하고 기수와 서수를 구분한다. (ST) 지난 8일 (MT) on May 8 (PE) on the 8th

인명 지명	한국 인명은 성과 이름의 순서로 띄어 쓰고 성만 번역된 경우 이름을 추가한다. 외국 인명은 정확한 철자를 검색해 기재한다. (ST) 김영권 (MT) Kim (PE) Kim Young-kwon 한국 지명 표기법은 국어의 로마자 표기법과 국립국어원 용례를 따른다. 자연 인공지명은 로마자 표기와 의미역을 병기할 수 있다. (ST) 경상북도 (MT) North Gyeongsang Province (PE) Gyeongsangbuk-do
	문장부호와 단어 간 빈칸은 삭제하고 이의 단어는 한 칸씩 띄어 쓴다. (ST) ‘전신 송금(wire transfer)’ (MT) ‘wire transfer’ (PE) “wire transfer”
	원문의 수치와 다르게 번역된 부분은 원문과 동일하게 수정한다. (ST) IEC61000 4-3 레벨 3 (MT) IEC 61000-3 Level 3 (PE) IEC 61000 4-3 Level 3
오 탈 자	기계번역이 과생성한 부분은 삭제한다. (ST) 화면비 (MT) the ratio of the ratio of the screen (PE) the screen ratio
	고유명사와 공식명칭은 검색하여 올바른 철자와 표기법으로 수정한다. (ST) 2019KBO리그 키움히어로즈와 기아타이거즈의 경기 5회말 (MT) the bottom of the fifth inning of the game between the 2019 KBO League Kium Heroes and the Kia Tigers (PE) the bottom of the fifth round of the 2019 KBO League Kiwoom Heroes and KIA Tigers game
	기계번역의 오역을 원문의 의미에 맞는 단어로 수정한다. (ST) 기존 다공체 유도체 (MT) existing multi-functional induction materials (PE) existing porous derivatives
어 휘	오역된 단어는 원문의 의미에 맞는 단어로 수정하되 오역이 아닌 경우에는 유사한 단어 등으로 변경하지 않는다. (ST) 기존의 먹는 탈모 치료제 (MT) traditional eating drugs (PE) existing edible hair loss treatments
	의미나 문법 오류가 있는 경우에만 수정한다. (ST) 토큰의 추세를 유추하기 쉽지 않지만

표현	(MT) while it is difficult to infer the trend of tokens (PE) it is not easy to infer the drift of token (불필요한 수정)
	고유명사가 아닌 음역된 단어는 적절한 의미역으로 수정한다. (ST) 연내 판호 발급이 가능하다. (MT) Panho can be issued by end of this year. (PE) They could issue a Chinese game version number by end of this year.
	오역된 구 절은 원문의 의미에 맞게 수정한다. (ST) 오씨가 용기를 내 39년 동안 가슴에 묻어뒀던 이야기를 진술하기로 결심했다. (MT) Oh decided to tell the story that he had been brave for 39 years. (PE) Oh decided to tell the story that he had buried in his heart for 39 years.
음역	시제로 인해 번역문의 의미가 원문의 의미와 다른 경우 변경한다. (ST) “EU가 미국에 하는 것들은 믿기 어려울 정도로 나쁘다” (MT) “What the EU is doing to the United States is incredibly bad” (PE) “What the EU does to the United States is unbelievably bad” (불필요한 수정)
	명사 동사의 성 수가 원문의 의미와 문법에 맞지 않는 경우 수정하되 의미상 차이가 없는 변경은 하지 않는다. (ST) 김광현이 반지를 보여준 이후 팀 동료 김태훈은 자신의 SNS에 우승 반지 사진을 올렸고, (MT) After Kim showed the ring, teammates Kim Tae-hoon posed a picture of the winning ring on their SNS, (PE) Since Kim Kwang-hyun showed his ring, teammate Kim Tae-hoon posed a picture of the winning ring on his social media,
	관사가 누락되거나 잘못 쓰인 경우 수정한다. (ST) KT의 목표가 “엣지 클라우드 기반의 초저지연 네트워크를 구현하는 것” (MT) “KT’s goal is to implement ultra-low latency network based on edge cloud.” (PE) KT’s goal is “to implement an ultra-low latency network based on the Edge Cloud.”
구 절 오역	용법에 맞게 수정하되 의미 변화가 없는 변경은 하지 않는다. (ST) 도널드 트럼프 미국 대통령과 두 번째 만남을 앞두고 (MT) With this second meeting with U.S. President Donald Trump (PE) Ahead of his second meeting with U.S. President Donald Trump
	원문의 의미에 맞게 수정하되 오류가 아닌 경우 변경하지 않는다. (ST) 대용량 분리 동시 세탁이 가능한 트윈워시 (MT) Twin Wash, that can be washed at the same time and washed separately

		(PE) Twin Wash that can wash big loads of laundry at the same time or separately
구조	인과 관계	오류나 누락으로 원문의 인과관계가 나타나지 않는 경우 수정한다. (ST) 기상관측위성 ‘천리안 2A호’가 기아나 우주센터에서 발사되기 때문이다. (MT) The weather satellite Chollian 2A will be launched from the Guiana Space Center. (PE) This is because the meteorological satellite “Chollian 2A” is launched from the Guiana Space Center.
	수식 관계	원문 의미와 구조에 맞는 수식관계를 이루도록 수정한다. (ST) 김상식 감독은 2018 자카르타-팔렘방 아시안게임이 끝나고 사임한 허재 전 감독의 빈 자리를 메우며 요르단 원정, 시리아 홈경기를 승리로 이끌었다. (MT) Kim Sang-sik, who resigned after the 2018 Jakarta-Palembang Asian Games, filled an empty seat for former coach Huh Jae, who led the team to victory in away games in Jordan and home games in Syria. (PE) Head coach Kim Sang-sik filled the vacancy of former head coach Huh Jae, who resigned after the 2018 Jakarta-Palembang Asian Games, and led the victories of the away match against Jordan and the home match against Syria.
	주술 관계	주술관계가 원문의 의미와 다르게 형성된 경우 수정한다. (ST) 글쓰이는 카드사가 조만간 혜택 축소와 포인트 소멸예정임을 강조하며 현혹한 후 불법모집인이 만든 신청서에 개인정보를 작성해 보내주면 대신 가입해 주겠다고 했다. (MT) The writer said, “If the credit card company sends personal information on the application made by the illegal recruiter, it will join the card instead.” (PE) Emphasizing that the credit card company is going to reduce benefits and expire points soon, the writer said that he will sign up instead if other people send their personal information on the application form made by an illegal recruiter.
추가	직위 호칭	직위 호칭이 누락된 경우 알맞은 직위 호칭을 추가한다. (ST) 김 교육감은 송승헌 이사장에게 감사의 뜻을 전했다. (MT) Kim told Song Seung-heon (PE) Superintendent Kim thanked Chairman Song Seung-heon
	고유 명사	고유명사의 일부 전체가 누락된 경우 공식명칭을 검색해 추가한다. (ST) 김유석 딜로이트 스타트업 자문그룹 상무 (MT) Kim Yoo-seok, executive vice president of the consulting group for the startup (PE) Managing director of Deloitte Startup Advisory Group, Kim Yoo-seok

삭제	단어 구 절	기계번역에서 누락된 원문의 단어 구 절을 추가한다. 원문의 의미와 관계없는 요소는 추가하지 않는다. (ST) 김승현이 1차 시기에서 3개를 집어넣어 미니농구에서도 최고가 되는 듯 했으나 2차 시기에서 하나도 넣지 못했다. (MT) Kim Seung-hyun seemed to be the best basketball player in the first round, but he failed to put any on the second round. (PE) Kim Seung-hyun seemed to be the best also in mini basketball by putting three in the first stage, but he couldn't put one in the second stage.
	어휘 부가 설명	원문의 어휘에 대한 원문에 없는 부가 설명은 삭제한다. (ST) 깜짝 선발로 나선 롯데 이명우 (MT) Lotte Giants pitcher Lee Myung-woo, who started as a surprise starter (PE) Lotte's Lee Myung-woo, who made a start as a surprise starter
	단어 구 절	원문의 의미를 전달하는 단어 구 절을 삭제하지 않는다. 원문의 의미를 전달하지 않는 단어 구 절은 삭제한다. (ST) 다음 달 중국과 무역협상을 앞두고 있는 미국 도널드 트럼프 정부가 중국 해커들을 무더기로 기소하며 중국에 대한 압박을 강화했다. (MT) U.S. President Donald Trump has stepped up pressure on China by indicating many Chinese hackers ahead of trade talks with China next month. (PE) Set to hold trade talks with China next month, Donald Trump's administration intensified pressure on China by prosecuting a number of Chinese hackers. (잘못된 수정: 원문의 요소 삭제)

<표 12>에서는 기계번역의 오류 수정에 대한 가이드라인뿐 아니라 실제 작업 사례에서 나타난 바와 같이 포스트에디터가 오류가 아니거나 굳이 수정하지 않아도 원문 의미 전달에 지장이 없는 데 수정하는 부분에 대한 주의도 포함한 현실적인 포스트에디팅 가이드라인을 프로젝트 결과물에서 나타난 예시와 함께 제시하였다. 위의 가이드라인에 따른 포스트에디팅 및 감수에 아울러 형식면에서 문장 부호, 띄어쓰기, 인명, 단위 등은 프리에디팅을 통해 전 처리한다면 보다 효율적인 포스트에디팅을 수행할 수 있을 것으로 보인다. 그러나 앞서 제시한 수정 사례에서 볼 수 있듯이 한글 원문의 동음이의어와 다의어, 고유명사 등의 내용적 측면은 기계번역 전처리를 한다 해도 해결될 수 없는 부분이며 이는 포스트에디팅의 몫으로 남게 된다. 따라서 위와 같은 상세한 가이드라인을 제시한다면 포스트에디터의 결과물 수준이 보다 향상될 것으로 기대할 수 있는 것이다.

5. 논의 및 결론

5.1 분석 결과에 대한 논의

본 연구는 실제 번역산업 현장의 프로젝트 결과물에 기반을 두고 MTPE 단계별로 어떤 부분이 수정되었는지를 분석하여 한영 기계번역 포스트에디팅 가이드라인을 제시하였다. 분석 결과 PE 단계에서는 주로 단어, 표현, 구 절 구조가 변경되었으며 구 절 누락이 추가되었고 PR 단계에서는 문장부호, 고유명사 공식명칭, 원문과 다른 의미로 번역한 단어 등이 수정되었음이 나타났다. 또한 PE, PR 단계에서 기계번역의 오류 뿐 아니라 오류가 아닌 부분도 수정되었음을 알 수 있었다. 이에 따라 가이드라인에서는 오류 수정에 대한 지침과 함께 오류가 아닌 부분에 대한 수정을 지양하는 지침 또한 제시함으로써 상세하면서도 효율적인 포스트에디팅 가이드라인을 제안하고자 하였다.

5.2 연구의 한계와 의의

본 연구의 한계는 다음과 같다.

첫째, 전체 프로젝트의 크기에 비해 제한된 수의 문장을 추출해 분석했다.

둘째, 본 연구의 자료를 추출한 프로젝트에서는 2019년에 한 종류의 신경망 기계번역기만을 사용했으므로 현재 다른 종류의 신경망 기계번역기를 사용하는 경우에는 번역 결과물이 달라질 수 있다.

셋째, 연구자 개인의 정성 분석에 기반을 두었다.

넷째, 본 연구에서 제시한 가이드라인은 기사문장 코퍼스 프로젝트의 사례에 바탕을 두고 정보성 텍스트에 사용가능한 가이드라인으로 제시되었으므로 문화특수성을 지닌 어휘 및 표현을 다수 포함한 문학 등 다른 종류의 텍스트에는 전적으로 적용할 수는 없다.

위에서 언급한 한계에도 불구하고 본 연구는 실제 번역 산업 현장에서 진행되는 프로젝트를 토대로 기계번역 포스트에디팅과 감수 단계를 거친 문장 사례를 통해 한영 포스트에디팅 가이드라인을 제시하였다는 데에서 그 의의를 찾을 수 있으며 아래와 같은 함의를 갖는다.

첫째, 기존 기계번역 포스트에디팅 연구는 연구자가 추출한 문장으로 교육

현장에서 학생을 대상으로 하는 등의 연구가 다수였으며 기계번역 오류를 수정하는데 초점을 두었다. 그러나 본 연구는 실제 프로젝트에서 작업한 문장을 자료로 하여 실무에서 직면한 기계번역과 포스트에디팅에서의 어려움을 분석하고 오류 수정과 비오류 수정 양 측면을 모두 고려하였다는 학문적 함의를 갖는다.

둘째, 포스트에디팅 실무에서 나타나는 양상을 보여주고 오류는 수정하고 비오류 수정은 지양하는 것을 목적으로 하여 실제 현장에서 활용 가능한 포스트에디팅 지침을 교육할 수 있다는 교육적 함의를 갖는다.

셋째, 번역 산업 현장에서의 기계번역 및 포스트에디팅 품질 향상을 위해 본 연구에서 제시된 가이드라인을 추후 프로젝트에서 효율적으로 활용할 수 있다는 실무적 함의를 갖는다.

위에서 언급한 연구의 한계를 보완해 후속연구를 진행한다면 더욱 유의미한 연구결과를 기대할 수 있을 것이다. 또한 본 연구에서는 IT, 국제, 스포츠 분야 뉴스 기사문의 사례만을 다루었으나 후속연구에서 다른 종류의 텍스트의 경우를 분석한다면 본 연구에서 나타난 것과는 다른 특징이나 고려사항이 발견되는지 살펴볼 수 있을 것이다.

참고문헌

- 곽중철, 한승희 (2018) 「포스트에디팅 측정지표를 통한 기계번역 오류 유형화 연구」, 『통번역학연구』 22(1): 1-25.
- 김자경 (2020) 「포스트에디팅 결과물에 나타난 오류 고찰 - AI 학습용 한영 번역 말뭉치를 대상으로」, 『통번역학연구』 24(4): 35-62.
- 김지영 (2017. 2. 20) “진화하는 번역 앱... 번역가→에디터 역할로 바뀐다”, 『서울경제』. 2020년 11월 7일 검색.
- 박옥수 (2017) 「한영 기계 번역에서 ST의 유형적 특징에 따른 번역 오류 분석」, 『동아인문학』 41: 155-183.
- 박옥수 (2018) 「원천언어의 관점에서 살펴본 기계번역의 오류분석과 수정 원칙: 비문학 텍스트의 통사적 특징에 근거해서」, 『동아인문학』 44: 151-171.
- 박찬준, 임희석 (2020) 「기계번역 사후교정(Automatic Post Editing) 연구」, 『한

- 국융합학회논문지』 11(5): 1-8.
- 서보현, 김순영 (2018) 「기계번역 결과물의 오류유형 고찰」, 『번역학연구』 19(1): 99-117.
- 윤미선, 김택민, 임진주, 홍승연 (2018) 「영어-한국어 언어쌍에 적합한 포스트에디팅 가이드라인: 미래 포스트에디터를 위한 가이드라인과 그 실패」, 『번역학연구』 19(5): 43-76.
- 이주리에 (2019) 「한일 헤드라인 번역의 포스트에디팅 가이드라인 고찰」, 『통역과 번역』 21(2): 119-144.
- 최문선 (2019) 「국내 번역학 기계번역 연구 동향: 내용 분석과 키워드 분석을 중심으로」, 『언어학연구』 24(1): 275-297.
- 한국경제TV (2012. 2. 9.) 「구글 “AI로 영한 번역 수준 확 높였다”... 8개 언어 중 2위」.
- 한현희 (2018) 「한노 기계번역의 문장 부호 처리 양상 및 오류 개선을 위한 실용적 함의: 프리에디팅 규칙 수립과 효용성의 관점에서」, 『통번역학연구』 22(3): 227-260.
- De Almeida, Giselle and Sharon O'Brien (2010) 'Analysing Post-editing Performance: Correlations with Years of Translation Experience', in *Proceedings of the 14th Annual Conference of the European Association for Machine Translation*, St. Raphael.
- Hu, Ke and Patrick Cadwell (2016) 'A Comparative Study of Post-editing Guidelines', *Baltic Journal of Modern Computing* 4(2): 346-353.
- The LISA QA Model, Available at <https://docs.sdl.com/785465/788071/sdl-multitrans/the-lisa-qa%C2%A0model>
- Mossop, Brian. (2010) *Revising and Editing for Translators*, second edition, Manchester & New York: St. Jerome.
- Pérez, Celia R. (2012) 'A Flexible Decision Tool for Implementing Post-editing Guidelines', *Localisation Focus* 11(1): 54-66.
- TAUS (2016) 'TAUS Post-Editing Guidelines', Available at <https://www.taus.net/think-tank/articles/postedit-articles/taus-post-editing-guidelines>

[Abstract]

Post-editing Guidelines for Korean-English Machine Translation of Informative Texts

Park, Kunyoung
(Hankuk University of Foreign Studies)

This study aims to suggest basic post-editing guidelines for Korean-English machine translation of informative texts. After constructing a parallel corpus composed of sentences extracted from an actual MTPE project, a qualitative analysis was conducted to find out which items were post-edited and how they were revised while Korean source texts went through machine translation, post-editing by translators, and revision by editors. The study categorized post-edited items into seven broad categories of format, typo, lexis, grammar, structure, addition, and omission, classified them into thirty-five sub-categories, and counted the number of items under each sub-category. The result shows that the post-edited items vary from typos and grammatical errors to structural changes and omission of elements. The study also found that post-editors in the project, both translators and editors, post-edited not only errors but also non-errors. Lastly, this study suggests post-editing guidelines based on its findings, which can serve as guidelines for actual MTPE projects to improve post-edited outputs as well as a practical guide for those who want to learn post-editing in action.

▶ Key Words: informative text, machine translation, MTPE, post-editing, post-editing guidelines

▶ 주제어: 기계번역, 정보성 텍스트, 포스트에디팅, 포스트에디팅 가이드라인, 한영번역

박건영

한국외국어대학교 일반대학원 영어번역학과 석박사 통합과정생

kypark1203@hufs.ac.kr

관심분야: 기계번역, 영상번역, 포스트에디팅

논문투고일: 2021년 2월 7일

심사완료일: 2021년 2월 28일

게재확정일: 2021년 3월 4일