

한영 포스트에디팅 노력 예비연구: 트랜스로그 II를 활용한 한영 인간번역과 포스트에디팅의 차이 분석

이 준 호
(한국외대)

1. 서론

번역 기술 발전과 함께 다수의 국내외 연구에서 기계번역 결과물의 포스트 에디팅(Machine Translation Post-Editing)을 번역 수행의 새로운 방식 중 하나로 제시하며 포스트에디팅의 긍정적 기능을 논의해 왔다. 국내에서도 이미 포스트에디팅을 교육하는 학교 기관 및 사기관이 존재하며, 외국어에서 한국어 그리고 한국어에서 외국어로의 포스트에디팅 서비스가 이뤄지고 있기에 포스트 에디팅이 먼 미래의 이야기라고 보기만은 어렵다.

이처럼 하나의 서비스가 존재하기 위해서는 서비스 의뢰인과 서비스 제공자가 존재해야 하며, 서비스를 통한 양측의 가치교환이 이뤄져야 한다. 하지만 합리적 가치교환을 위해서는 포스트에디팅이라는 행위가 어떠한 속성을 가지는지 먼저 이해할 필요가 있다. 그렇다면 우리는 포스트에디팅에 대해 얼마나 알

고 있을까? 포스트에디팅이라는 활동의 정의, 작업의 생산성과 품질, 그리고 생산성과 품질에 기여하는 작업자들의 물리적 그리고 정신적 노력에 대해 알고 있을까? 이 질문에 “그렇다”라고 답하기는 어려울 것이다.

하지만 이처럼 포스트에디팅의 속성과 본질에 대한 정보가 부족하지만, 국내 서비스 제공사나 언론의 보도는 포스트에디팅의 “생산성”과 “비용 절감”이라는 장점만을 언급하는 경우가 많다.

“기업의 입장에서서는 번역에 걸리는 시간을 절약할 수 있습니다. 그뿐만 아니라 포스트 에디팅 작업은 전문 번역가의 작업량이 대폭 줄어들기 때문에 번역에 필요한 인건비용을 크게 절감할 수 있습니다.” (스마트라이온 2019)

“프리랜스 번역사들이 인공지능 번역 엔진과 보조 번역 CAT Tool을 사용한 결과 번역 시간이 50% 이상 줄었다. 이는 번역사들이 인공지능 번역 엔진을 통해 번역 업무를 진행하면 동일한 시간 동안 기존 대비 2배 이상의 수입을 얻을 수 있다는 의미이기도 하다.” (신수정 2021)

과연 작업량이 “대폭” 줄고 인건비용이 “크게” 절감이 가능하다는 것은 어느 정도의 효익을 의미하는 것일까? 번역 시간이 “50%” 이상 줄어들었다는 것은 어떠한 데이터에 근거한 것일까? 그리고 기존 대비 “2배 이상의 수입”이라는 주장은 어떠한 생산성 데이터와 효율 산정 기준에 따른 것일까?

물론 작업 시간을 기준으로 포스트에디팅의 생산성을 설명하는 방식이 잘못되었다는 뜻은 아니다. 아주 오랫동안 작업 시간에 근거한 보상 체계가 우리 사회에서 광범위하게 사용됐기에, 일반 대중에게 포스트에디팅의 장점을 강조하기 위해 작업 시간에 근거하여 설명하는 것은 이해 가능한 측면이 있다.

하지만 아이러니하게도 번역시장에서는 작업 시간에 기초하여 효율을 산정하는 경우는 많지 않다. 번역 제공업체들은 작업 단어당 기본 효율에 기초하여 번역 견적을 생성해왔다. 그렇지만 기계번역의 초기 번역 결과물을 수정하는 포스트에디팅에는 기존의 견적 방식이 적용될 수 없다. 따라서 많은 업체가 기계번역 결과물에 수정이 이뤄진 양을 근거로 견적을 생성하고 있다(김순미, 신호섭과 이준호 2019; 남보리 2021). 그러나 텍스트의 삭제, 추가, 대체 등을 통

해 수정량을 표현하는 TER(Translation Edit Rate) 등의 방식은 몇 가지 단점을 지닌다. 첫째, 수정 과정에 들어간 노력을 모두 반영하지 못한다. 예를 들어 하나의 단어에 대해 몇 번의 수정을 했어도 최종 수정 행위만 계산에 포함된다. 둘째, 작업자의 인지적 노력을 반영하지 못한다. 예를 들어 작업자가 원천 텍스트와 기계번역 결과물을 읽고, 품질에 문제가 없다고 판단하고, 수정만 하지 않았다면 TER은 0으로 계산된다. 따라서 포스트에디팅 작업자들의 작업 과정과 노력에 대해서 더 많은 정보를 파악하고, 궁극적으로 합당한 가치교환 체계를 만들고 이를 교육할 필요성이 있다.

이러한 고민은 해외에서 이미 오래전에 시작되었다. 크링스(Krings 2001)는 포스트에디팅에 필요한 노력을 시간적 노력, 기술적 노력, 인지적 노력으로 구분하여 포스트에디팅의 본질에 접근할 필요성을 주장하였다. 또한, 기계번역과 포스트에디팅에 대한 썹크랭크인 TAUS(2013)는 기계번역을 사용하기 전에 자동화된 품질 점수(automated quality score), 인간의 품질 검토(human quality review) 그리고 생산성 평가(productivity assessment) 등을 종합적으로 고려해야 함을 제안하였다.

하지만 한국어와 외국어 언어쌍의 포스트에디팅에 대한 인간 평가, 생산성 연구는 극히 제한적이어서 추가적 포스트에디팅 연구 및 포스트에디팅의 실무적 사용을 위한 객관적 정보 전파가 어려운 것이 사실이다. 더욱이 포스트에디팅 작업자가 어떠한 기술적 노력 및 인지적 노력을 하는지를 논의한 연구는 매우 찾기 어려운 실정이어서 포스트에디팅의 본질을 깊이 있게 이해하기는 더욱 어렵다. 이러한 정보의 부재는 실무 차원에서는 시장의 왜곡으로 이어질 수 있으며, 번역학계에는 미래 번역 교육 설계에 걸림돌이 될 수 있다.

이에 본 연구는 인간번역과 포스트에디팅을 비교하는 접근을 통해 서비스 제공자들의 시간적, 기술적, 인지적 노력을 분석하고자 한다. 먼저 시간적 노력 분석을 통해 포스트에디팅은 인간번역보다 생산성이 높다는 통념을 데이터에 기반하여 고찰할 것이다. 기술적 노력 분석을 통해 인간번역과 포스트에디팅의 키스트로크(keystroke) 양상 차이가 있는지를 분석하고, 인간번역과 포스트에디팅의 휴지(pause) 양상 차이가 있는지를 분석하여 두 가지 번역방식의 인지적 노력을 새롭게 조명한다. 그리고 마지막으로 기술적 노력과 인지적 노력이 시간적 노력과 연관성이 있는지 분석한다. 이상의 분석과 논의를 통해 포스트에

디팅이 인간번역과 어떤 차이가 있는지를 설명하고, 서비스 제공자를 위한 적절한 보상 메커니즘을 고민하고, 포스트에디팅을 더욱 체계적으로 이해하는 데 필요한 접근법과 미래 연구 과제를 제안할 것이다.

2. 선행연구 분석

2.1 포스트에디팅 생산성 연구 현황과 진화

포스트에디팅 연구의 쟁점 중 하나는 생산성과 품질에 대한 기초적 논의를 통해 포스트에디팅이 하나의 서비스로서 존재할 수 있는 공고한 근거를 마련하는 것이었다. 이러한 맥락에서 해외에서는 생산성과 품질에 관한 연구가 다수 이뤄진 것이 사실이다. 물론 언어쌍 등 다양한 변수에 의해 연구 결과는 상이했지만, 포스트에디팅이 인간번역보다 생산성이 높다는 연구 결과가 그렇지 않다는 연구 결과보다 많았다(Gueberof-Arenas 2008; Plitt and Masselot 2010 등). 하지만 포스트에디팅의 생산성이 높다고 주장한 학자들도 다양한 변수에 의해 생산성 증가율이 영향을 받을 수 있음을 인정하며 생산성 증가를 정량적으로 일반화하기보다는 생산성을 중심으로 연구 주제를 확장해 나가려 노력하였다. 생산성과 품질의 상관관계를 연구하거나(Gueberof-Arenas 2014), 텍스트의 난이도와 기계번역의 품질은 연관성이 있으며 생산성에 악영향을 줄 수 있음을 보고한 연구(Gaspari, Toral, Naskar, Groves and Waym 2014), 번역 경험 등이 생산성에 미치는 영향을 분석한 연구도 존재한다(Aranberri, Labaka, de Ilarraza and Sarasola 2014). 이러한 흐름에서 볼 수 있듯 생산성에 대한 기초데이터 축적은 중요하지만, 기초데이터만을 활용하여 모든 상황을 일반화할 수는 없다. 따라서 생산성 증가에 도움 혹은 방해가 되는 요소를 탐구하는 것이 포스트에디팅의 이해 및 활용을 위한 바람직한 연구 방향이라 하겠다.

하지만 국내의 경우 생산성과 품질에 대한 논의가 번역학계에서 본격화된 것은 2018년으로 다소 늦은 감이 있다. 번역학회의 산학협력 위원회가 발표한 영한 라이트 포스트에디팅의 생산성 연구가 그 시발점이라 볼 수 있기 때문이

다(김순미, 신호섭과 이준호 2019). 물론 이후 이준호(2021b)의 영한 풀포스트 에디팅의 시간적 노력을 분석한 연구가 발표되었지만, 그 외에 포스트에디팅의 생산성과 품질 연구가 활발하게 이뤄지지 않는 것이다. 또한, 영어와 한국어 언어쌍의 포스트에디팅 생산성 연구는 영어에서 한국어 방향으로만 이뤄졌고, 연구자가 조사한 범위 내에서는 한국어에서 영어 방향의 포스트에디팅 생산성 연구는 아직 이뤄지지 않았다.

따라서 포스트에디팅 서비스의 적용 가능성 여부를 판단하는데 핵심이라고 할 수 있는 영한 및 한영 포스트에디팅 생산성에 대한 기초 데이터를 확보하고 이에 영향을 주는 요인을 분석하는 것은 시급한 과제라 하겠다. 더욱이 모국어 가 아닌 영어로 진행되는 한영 포스트에디팅의 경우 영한 포스트에디팅과 비교하여 생산성 증가율이 다를 수 있다. 따라서 한영 포스트에디팅 생산성과 영한 포스트에디팅의 생산성을 비교 분석하고, 한영 포스트에디팅의 생산성에 영향을 주는 요소를 미래에 더욱 심도있게 연구할 필요성이 높다.

생산성과 품질에 대한 논의가 추가 연구를 위한 기초데이터라고 한다면, 기초데이터를 사용하여 탐구의 범위를 확장하고 다양한 접근법을 시도하는 연구가 증가하고 있다. 예를 들어 포스트에디팅의 생산성과 품질을 인지적 노력의 관점으로 확장한 연구(O'Brien 2011), 그리고 스크린레코딩(screen recording)과 키로깅(keylogging)을 사용하여 인간번역과 포스트에디팅을 비교한 연구(Jia, Carl and Wang 2019) 등 단순히 생산성과 품질의 논의를 뛰어넘어 포스트에디팅 자체의 과정이나 인지적 노력과 연계한 연구들이 존재한다. 여기에 더해 작업자의 노력을 계량화하기 위한 대리 지표의 유효성을 평가하고 이를 연구에 적용한 사례도 다수 존재한다. 기계번역의 품질과 시선의 고정(fixation)을 아이트래킹(eye tracking)을 통해 분석한 연구(Vieira 2017), 그리고 휴지(pause)를 인지적 노력의 대리 지표로 사용하여 다양한 분석을 진행한 연구(Lacruz, Denkowski and Lavie 2014; Vieira 2016 등) 등이 대표적인 예시이다. 이처럼 결과물 중심의 기초데이터에 기반하여 과정 및 인지 중심의 연구가 이뤄지면서, 포스트에디팅의 특질을 체계적으로 설명하기 위한 시도가 이뤄지고 있다.

하지만 영어와 한국어 언어쌍의 경우 포스트에디팅의 결과물이 아닌 과정과 노력을 연구한 사례는 얼마 되지 않는다. 스크린레코딩과 TAP(think-aloud protocol)을 사용하여 학부생들의 포스트에디팅 과정을 분석한 이상빈(2018)의

연구와 석사과정생을 대상으로 번역저널, TAP, 스크린레코딩 데이터를 분석한 이준호(2021a)의 연구를 제외하면, 포스트에디팅 과정을 집중 조명한 연구는 거의 없다. 더욱이 그간 영한 포스트에디팅 연구에 사용된 도구나 연구 방법의 특성상 포스트에디팅의 시간적 노력에만 집중하여 연구가 이뤄졌지, 기술적 노력과 인지적 노력에 대해 논의가 거의 이뤄지지 않은 것이 사실이다. 그 결과 영한 포스트에디팅의 특질을 설명하기 위한 충분한 데이터가 부족한 것이 사실이다. 여기에 더해 한영 포스트에디팅 연구는 결과물 중심의 연구가 대부분이지, 과정이나 노력을 논의한 연구가 거의 없다. 따라서 한영 포스트에디팅이 인간번역과 어떠한 차이점과 공통점을 지니는지 답하기가 더욱 어려운 실정이다.

이에 본 연구는 먼저 시간적 노력을 분석하여 한영 포스트에디팅과 인간번역의 생산성에 대한 초기 자료를 확보하고자 한다. 둘째, 키스트로크 분석을 통해 인간번역과 포스트에디팅의 작업 방식에 있어 기술적 노력의 차이가 있는지를 살펴볼 것이다. 그리고 기술적 노력의 대소 여부가 생산성에 어느 정도 영향을 주는지 논의할 것이다. 셋째, 인지적 노력의 대리 지표로 추정되는 휴지(pause) 데이터 분석을 통해 한영 포스트에디팅과 인간번역의 인지적 과정을 새롭게 조명하고, 인지적 노력이 생산성에 어느 정도 영향을 주는지 살펴볼 것이다. 이상의 분석을 통해 한영 포스트에디팅에 대한 기초적 데이터를 확보하고, 인간번역과 포스트에디팅 작업의 절차적 차이점을 파악하고, 생산성에 영향을 줄 수 있는 기초적 요소를 식별하고자 한다.

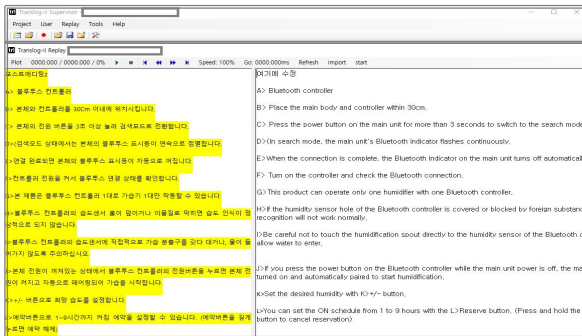
2.2 트랜스로그(Translog) 소개 및 관련 연구

시간적, 기술적, 인지적 노력을 보여주는 데이터 확보를 위해 본 연구는 트랜스로그II를 선택하였다. 키스트로크 데이터 확보를 위해 널리 사용되어온 트랜스로그의 후속 도구이며, 번역과 번역 기술 연구 및 혁신 센터(Center for Research and Innovation in Translation and Translation Technology) 홈페이지에서 무료로 다운로드가 가능하다¹⁾. 글의 읽기 및 쓰기 과정을 기록할 수 있는 도구로서 번역학 및 포스트에디팅 연구에서 해당 도구를 사용한 사례를 찾는 것은 어렵지 않다(Carl, Aizawa and Yamada 2016; Koglin 2015 등).

1) <https://sites.google.com/site/centrtranslationinnovation/translog-ii?authuser=0>

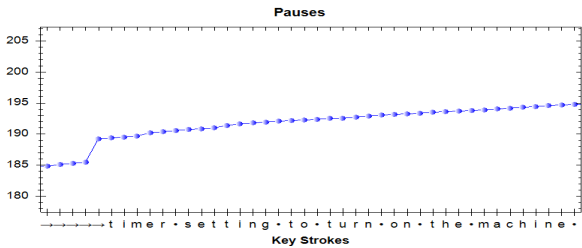
트랜스로그II의 기능과 연구 활용법은 다운로드 페이지의 사용자 메뉴얼과 칼(Carl 2012)의 연구에 상세하게 소개되어 있다. 과정 연구를 위해 녹화기능을 제공하고 있어 연구 참여자의 모든 작업을 녹화할 수 있다. 분석 과정에서는 녹화된 영상을 원하는 시간대별로 재생이 가능하며, 영상과 텍스트 결과물이 연계되어 있어 결과물 텍스트를 클릭하면 해당하는 영상의 재생이 가능하다. 녹화 및 작업 시간 기록 이외에 기초 통계를 자동으로 제공하는 기능이 있어, 키보드 상의 활동을 전체 이벤트, 텍스트 추가 이벤트, 삭제 이벤트, 기타 이벤트 등으로 구분하여 정량화된 리스트를 제공한다.

그림 1 참여자의 트랜스로그 분석 화면



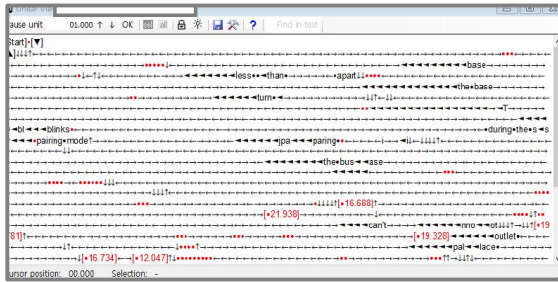
또한, 분석 도구 탭에는 휴지 플롯(Pause plot)과 선형적 뷰(Linear view) 메뉴가 있는데, 휴지 플롯을 선택하면 인지적 노력의 대리 지표라 할 수 있는 휴지의 누계치를 시각적으로 볼 수 있다

그림 2 휴지 플롯의 예시



X축에는 키스트로크 정보가 나열되어 있고 Y축에는 초 단위의 휴지 발생량과 누계치를 볼 수 있다. 따라서 어떤 작업 구간에서 휴지가 발생했는지 분석이 쉽다. 반면 선형적 뷰는 타이핑이 이뤄지는 모습을 상세하게 보여주고, 휴지 발생의 길이를 점으로 보여주고 있다. 다만 읽기가 어렵고 휴지를 하나씩 클릭해서 초 단위의 휴지를 파악해야 한다는 단점이 있다.

그림 3 선형적 뷰의 예시



이처럼 트랜스로그를 사용하여 실험을 진행할 경우 본 연구에서 목표로 하는 시간적 노력, 기술적 노력, 인지적 노력에 대한 데이터를 한 번에 얻을 수 있다는 장점이 있다. 이러한 유용성 때문에 트랜스로그는 이미 번역 및 포스트에디팅 연구에서 다양한 목적으로 사용되어 왔다. 작업자의 키스트로크 로깅(keystroke logging)을 위한 도구로 사용한 사례(Koponen, Salmi and Nikulin 2019; Nitzke 2016; Screen 2016 등), 트랜스로그와 아이트래커를 연동한 사례(Carl, Gutermuth and Hansen-Schirra 2015; Carter et al. 2020 등), 트랜스로그를 사용해 휴지를 포착한 사례(O'Brien 2006; Yildiz 2020 등) 등이 존재한다. 또한 TAP의 유무에 따른 시간적, 기술적, 휴지의 분석을 트랜스로그로 분석한 사례도 있다(Fonseca 2019).

특히 본 연구와 유사하게 인간번역과 포스트에디팅을 트랜스로그와 다양한 도구를 사용하여 비교한 연구가 존재한다(Carl et al. 2011). 그들의 연구는 포스트에디팅을 처음 체험한 번역자들의 포스트에디팅은 인간번역 대비 생산성이 다소나마 높으며, 인간 전문가가 참여한 품질 평가를 통해 기계번역을 사용한다고 해서 인간번역 대비 품질이 저하되지 않는다는 초기 데이터를 도출하였다.

여기에 추가적 분석을 통해 인간번역에서 포스트에디팅 대비 원천 텍스트에 더 많은 시선 고정(fixation)이 있음을 보고하여 인지적 노력에 차이가 있음을 주장하였다. 또한, 키스트로크 양상 역시 큰 차이가 있음을 보고하였다. 인간번역에서 텍스트의 추가(insert)가 포스트에디팅 대비 훨씬 높았으며, 커서의 움직임(navigation)은 포스트에디팅에서 더 높게 나타났다. 이들은 상기 데이터를 모두 종합하여 포스트에디팅의 생산성이 높을 가능성을 최종적으로 제시하였다. 이처럼 시간이란 단위로 표현된 포스트에디팅의 생산성을 다각적 데이터를 사용하여 분석적으로 설명한 연구는, 신뢰성이 높으며 포스트에디팅의 특질에 대한 이해를 확장할 수 있다는 측면에서 그 의미가 크다. 다만 10여 년 전의 연구이고 사용된 번역 엔진이 현재 사용되는 신경망 번역과는 다르다는 점에서 다시 한번 연구가 필요하다 하겠다.

반면 국내의 경우 트랜스로그를 사용한 번역학 연구가 많지는 않다. 김자경(2020a)은 전문가와 학생의 번역 과정에 나타난 검색 및 자기교정 관찰을 위해 트랜스로그를 사용해 키스트로크 데이터를 확보하였다. 또한 김자경(2020b)은 트랜스로그II 등의 도구를 사용하여 전문가와 학생의 은유 번역 전략을 다양한 각도에서 분석한 바 있다. 하지만 한국어-영어 포스트에디팅의 시간적, 기술적, 인지적 노력을 트랜스로그를 사용하여 분석한 사례는 찾을 수 없었다.

이에 본 연구는 트랜스로그II를 사용하여 한영 포스트에디팅의 시간적, 기술적, 인지적 노력에 대한 새로운 정보를 제공할 수 있을 것이다. 새롭게 발견한 정보는 한영 포스트에디팅과 인간번역의 차이를 더욱 명확하게 이해하는 데 기여할 것이며, 앞으로 전개될 한영 포스트에디팅 연구의 근거 혹은 반박의 대상이 되는 초석 역할을 할 수 있을 것으로 기대한다.

3. 연구 방법

인간번역과 포스트에디팅의 시간적, 기술적, 인지적 노력의 비교 분석을 위해서는 두 가지 번역방식을 모두 경험한 참여자의 모집이 필수적이다. 인간번역 수행에 필요한 기초적 수행 역량이 부족할 경우, 참가자들의 노력이 포스트에디팅이라는 번역 수행 방식 때문인지 아니면 기초적 번역 수행 역량 부족 때

문인지 파악이 어렵기 때문이다. 또한, 인간번역만 경험한 참여자들을 대상으로 연구를 진행한다면 포스트에디팅 수행 방식에 대한 이해 부족으로 인해 참여자들의 시간적, 기술적, 인지적 노력이 과도하게 높게 나타날 우려가 있다. 이에 인간번역 경험 및 포스트에디팅 경험 부재로 발생하는 데이터 왜곡을 최소화하고자 통번역대학원 석사과정을 수료하고 포스트에디팅 수업을 1학기 이상 수강했으며, 한국어가 모국어인 5명의 참여자를 모집하였다.

모든 참여자는 온라인으로 연구에 참여했으며 본 연구를 위해 트랜스로그II 사용방법을 사전에 교육받았다. 텍스트의 선정에서 있어, 기계번역이 용이한 정보 전달성 텍스트를 선정했으며(마승혜 2018), 그중 포스트에디팅의 수요가 많은 전자제품 매뉴얼을 선택했다(김순미, 신호섭과 이준호 2019). 하나의 문서에서 2개의 텍스트를 추출했으며, 텍스트1에 대해 인간번역을 트랜스로그 프로그램에서 10분간 진행했으며(103단어) 이후 구글번역 결과물에 대한 포스트에디팅을 10분간 진행하였다(109단어). 이후 유사한 방식으로 텍스트2에 대해 인간번역을 트랜스로그 프로그램에서 10분간 진행했으며(103단어) 이후 구글번역 결과물에 대한 포스트에디팅을 10분간 진행하였다(102단어). 하지만 일부 참여자의 경우 트랜스로그 프로그램의 오류가 있어 텍스트1에 대한 인간번역-포스트에디팅 로그 4건, 텍스트2에 대한 인간번역-포스트에디팅 로그 3건만을 분석하였다.

시간적 노력을 측정하기 위해 인간번역과 포스트에디팅의 총 작업 시간 및 단위 시간당 처리 단어 수를 비교했으며, 기술적 노력을 측정하기 위해 키보드 위에서 이뤄진 텍스트의 추가, 삭제, 기타 이벤트를 모두 정량화하여 살펴보았다. 그리고 마지막으로 인지적 노력을 비교하기 위해 작업 중에 나타난 휴지의 길이와 빈도를 정량적으로 산출했으며, 추가적으로 휴지와 인지적 노력의 상관관계를 다룬 기존의 연구에서 제시한 공식을 사용하여 인간번역과 포스트에디팅의 인지적 노력을 비교하였다.

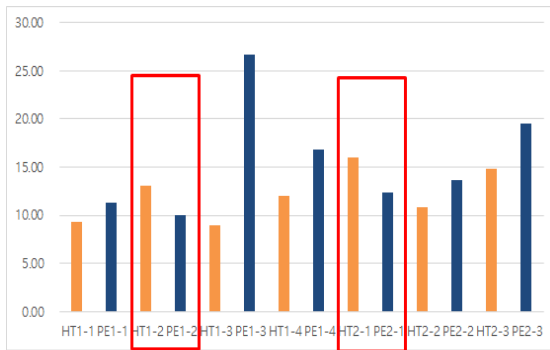
4. 분석 결과

4.1 시간적 노력

포스트에디팅을 사용하는 가장 큰 목적은 작업자의 시간적 노력을 줄이는 것이다. 본 연구 역시 포스트에디팅에 용이한 텍스트를 선택한 만큼 다수의 참여자가 인간번역 대비 포스트에디팅의 생산성이 높을 것으로 가정하였다.

총 7쌍의 결과물을 비교하기 위해 1분당 처리한 단어수를 비교했으며, 그 결과 인간번역의 작업 속도가 포스트에디팅보다 빠른 경우가 2건 있었다 (HT1-2: PE1-2, HT2-1: PE2-1). 하지만 다른 5건의 로그에서는 개인차는 있지만, 모두 포스트에디팅에서 1분당 더 많은 단어를 처리하였다.

그림 4 인간번역과 포스트에디팅의 생산성 비교



전체 평균을 살펴보면, 포스트에디팅은 인간번역 대비 1분당 29.8% 더 많은 단어를 처리했으며, 예외적으로 인간번역에서 1분당 단어 처리 수가 더 높았던 로그 두 건을 제외하면 포스트에디팅에서 인간번역 대비 1분당 57% 더 많은 단어를 처리하였다. 반면 텍스트1의 경우 포스트에디팅이 인간번역 대비 1분당 49.5% 더 많은 단어를 처리했고, 텍스트2의 경우 인간번역 대비 1분당 9.2% 더 많은 단어를 처리하는 데 그쳤다.

짧은 텍스트를 번역한 결과이기도 하지만 본 데이터는 한영 포스트에디팅이 인간번역 대비 더 적은 시간적 노력이 필요한 경우가 그렇지 않은 경우보다 더 많음을 보여주었다. 하지만 이보다 더욱 중요한 결과는 개인 간의 생산성 증가 편차가 매우 크다는 점, 그리고 동일한 문서에서 추출한 두 개의 텍스트에 대해서 생산성의 차이가 크게 차이가 날 수 있다는 점이다. 또한, 모든 참여자가 포스트에디팅에서 생산성이 향상되는 것은 아님을 볼 수 있었다. 따라서 포

스트에디팅이 높은 생산성을 항상 보증하는 것은 아니며, 포스트에디팅의 생산성 증가율을 보편화하여 주장하기가 어렵다는 것을 잘 보여주는 데이터라고 할 수 있다.

포스트에디팅의 또 다른 목적은 인간번역과 유사한 품질을 확보하는 것이다. 포스트에디팅의 생산성이 높고, 품질 역시 인간번역과 유사하다면 생산성 데이터를 더욱 긍정적으로 해석하는 것이 바람직할 것이다. 하지만 생산성은 높지만, 품질이 인간번역과 유사하지 못하다면 포스트에디팅의 높은 생산성은 재평가가 필요할 것이다. 다만 본 연구의 목적이 인간번역과 포스트에디팅 결과물의 품질 상세 분석을 목적으로 하지는 않기에, 번역 평가 메타 연구에서 일반적인 방법으로 제시한 평가 지표인 충분성과 유창성만을 평가 변수로 사용하였다(Kohen 2009). 평가를 위해 통번역대학원을 졸업, 현장 경험 4년 이상, 영어권 국가에서 4년 이상 교육을 받은 평가자 2인을 초빙하여 어떠한 결과물을 평가하는지 알 수 없는 맹검(blind) 상황에서 평가를 진행하였다.

충분성: 결과물이 출발언어 텍스트의 의미를 충실하게 전달하고 있다.

수용성: 결과물이 자연스러운 영어를 구사하며, 구성 및 문법의 어색함이 없다.

〈표 1〉 충분성 평가 기준

10점	출발언어 텍스트의 의미를 매우 충실하게 반영하고 있다.
8점	출발언어 텍스트의 의미를 충실하게 반영하고 있으나, 미미한 오류가 있다.
6점	출발언어 텍스트의 전반적인 의미를 반영하고 있으나, 중대 오류가 있다.
4점	출발언어 텍스트의 의미를 반영하고 있지 못하며, 중대 오류가 많다.
2점	출발언어 텍스트의 의미를 전혀 반영하지 못하고 있다.

〈표 2〉 수용성 평가 기준

10점	자연스러운 영어를 구사하며, 문장의 구성 및 문법에 있어 문제가 전혀 없다.
8점	전반적으로 자연스러운 영어를 구사하나, 문장의 구성 및 문법에 있어 사소한 문제가 있다.
6점	영어 구사가 다소 부자연스럽고, 문장의 구성 및 문법에 있어 몇 가지 문제가 있다.
4점	영어 구사가 부자연스럽고, 문장의 구성 및 문법에 중대한 문제가 있다.
2점	매우 부자연스러운 영어의 구사 및 다수의 중대한 오류가 있다.

텍스트1에 대한 총 16개의 상세 평가에서 두 평가자 간의 평가 점수 차이가 2점을 초과한 경우는 단 1회였다. 또한, 16개의 항목에서 한 평가자가 포스트에디팅의 상대적 우위로 평가한 경우, 다른 평가자가 인간번역의 우위로 평가한 적은 없었다. 따라서 인간평가의 근본적 문제라고 할 수 있는 평가자의 주관에 기반한 평가가 이뤄지지 않는다고 볼 수 있다.

〈표 3〉 텍스트1 결과물에 대한 전문가 평가

		HT	PE	HT	PE	HT	PE	HT	PE
		1-1	1-1	1-2	1-2	1-3	1-3	1-4	1-4
평가자 1	충분	4	5	4	7	5	7	4	4
	수용	2	6	5	8	5	7	4	6
평가자 2	충분	5	6	6	8	7	7	5	7
	수용	6	6	7	8	6	6	5	6

정량적 분석에서는 포스트에디팅 우세 12회, 동률 4회의 결과를 보였다. 또한, 인간번역의 품질이 상대적으로 더 높다고 평가한 경우는 없었다. 따라서 본 실험에서 획득한 포스트에디팅 결과물은 인간번역 결과물 대비 우수하다고 해석할 수 있다. 하지만 텍스트2에서는 평가 결과가 엇갈렸다.

〈표 4〉 텍스트2 결과물에 대한 전문가 평가

		HT2-1	PE2-1	HT2-2	PE2-2	HT2-3	PE2-3
		평가자1	충분	4	8	4	8
	수용	2	6	4	5	4	6
평가자2	충분	8	8	8	8	6	3
	수용	5	4	9	8	4	4

포스트에디팅이 품질이 높다고 평가한 경우가 5회, 동률 4회, 인간번역이 품질이 높다고 평가한 경우가 3회였다. 또한, 두 평가자는 다수의 평가 항목에서 다른 평가를 보였다. 결국, 텍스트2에 대한 전문가 평가 결과는 신뢰성이 높다고 보기 어렵다. 따라서 텍스트1의 평균 생산성 증가율 49.5%는 비교적 객관성이 높은 데이터이지만, 텍스트2의 평균 생산성 증가율 9.2%는 판단을 유보하는 것이 바람직할 것이다.

물론 이상의 데이터가 한영 포스트에디팅이 인간번역보다 동일한 시간 내

에 더 많은 작업을 소화해 낼 수 있다는 가능성을 부정하는 것은 아니다. 다만 작업 생산성 증가를 정량적으로 일반화하여 주장하는 것에 주의가 필요함을 보여주는 데이터라 하겠다. 본 데이터에서 관찰된 것처럼 개인의 역량 혹은 성향의 차이, 텍스트의 미묘한 차이 등이 생산성에 영향을 줄 수 있기 때문이다. 또한, 생산성 향상을 주장하기 이전에 포스트에디팅 결과물의 품질 등 다양한 요소를 고려할 필요가 있기 때문이다. 따라서 시간적 노력 자체도 중요하지만, 어떠한 과정과 어떠한 노력의 조합이 최종 작업 시간을 구성했느냐를 탐구하는 것이 더욱 중요할 것이다.

4.2 기술적 노력

기술적 노력을 측정하기 위하여 트랜스로그에서 제공하는 텍스트 추가, 텍스트 삭제, 기타 이벤트 데이터를 사용하였다. 인간번역과 포스트에디팅의 수행 방식이 다르고, 키스트로크 역시 차이가 있다는 선행연구가 있기에(Carl, Dragsted, Elming, Hardt and Jakobsen 2011), 결과물 텍스트를 직접 작성하는 인간번역은 텍스트 추가가 포스트에디팅 대비 많을 것으로 가정하였고, 포스트에디팅은 주어진 기계번역 결과물을 수정하는 행위이기 때문에 인간번역 대비 텍스트 추가는 적고 텍스트 삭제는 상대적으로 많을 것으로 가정할 수 있었다.

〈표 5〉 트랜스로그의 이벤트 데이터

	전체 이벤트	텍스트 추가	텍스트 제거	기타 이벤트
HT1-1	1250	848	312	30
PE1-1	745	431	194	120
HT1-2	1011	648	186	127
PE1-2	709	265	162	282
HT1-3	1443	1009	283	151
PE1-3	1101	256	167	678
HT1-4	1474	1100	127	247
PE1-4	2095	554	317	1164
HT2-1	1326	1051	172	103
PE2-1	2218	199	212	1807
HT2-2	1026	756	51	209
PE2-2	334	33	43	258

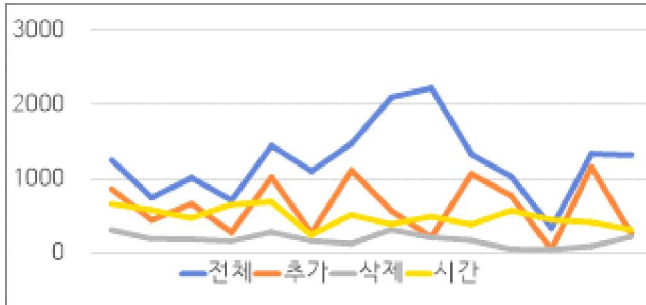
HT2-3	1333	1147	90	96
PE2-3	1316	242	221	853

예상대로 텍스트 추가에 대해서는 모든 참여자가 인간번역 결과물에서 더 많은 추가 이벤트를 보였다. 하지만 예상과 달리 HT1-1, 2, 3과 HT 2-2는 인간번역 결과물에서 더 많은 텍스트 제거를 보였다. 따라서 인간번역을 진행하면서 번역 결과물을 쓰고 지우는 작업이 많이 있었다고 추론할 수 있다.

반면 포스트에디팅 결과물에서 더 많은 기타 이벤트가 있음을 모든 참여자에게서 예외 없이 관찰할 수 있었다. 이는 기계번역 결과물과 포스트에디팅 결과물을 비교하는 과정에서 다양한 마우스나 커서의 움직임 있었기 때문으로 추정된다. 물론 기타 이벤트가 작업자의 노력에 있어 매우 큰 의미를 지니는 데이터는 아니지만, 인간번역과 포스트에디팅 수행에 있어 다른 패턴이 있다는 것을 시사하는 데이터로 이 부분에 관해서는 별도의 연구를 설계하여 추가적 조사가 필요해 보인다.

또 한 가지 흥미로운 점은 전체 이벤트가 많다고 해서 작업 시간의 증가로 이어진다고 해석하기는 어렵다는 점이다. 모든 참여자의 전체 작업 시간을 초 단위로 누적한 세로축은 모든 작업자의 개인별 작업 시간을 초 단위로 표시하였고, 가로축은 모든 작업자의 전체 이벤트, 텍스트 추가, 삭제 이벤트 횟수를 표현하였다.

그림 5 작업 시간과 전체 이벤트의 상관관계



그림에서 확인할 수 있듯 작업 시간의 총량과 전체 이벤트 수의 상관관계

가 높아 보이지 않는다. 피어슨 상관관계 분석에서도($R=-0.156$, $P=0.592$) 전체 이벤트 수와 작업 시간관의 상관관계가 있다고 보기는 어렵다. 하지만 텍스트의 추가만 따로 분리해서 분석한 경우는 작업 시간과 텍스트 추가 사이에는 중간 정도의 양의(*medium positive*) 상관관계가 있다고 할 수 있었고, 통계적으로 유의미한 수준은 아니었다($R=0.394$, $P=0.163$). 반면 텍스트의 삭제만 따로 분리해서 분석한 결과($R=0.169$, $P=0.562$) 작업 시간과 양의 상관관계가 약하고 통계적으로도 유의미하지 않았다. 마지막으로 텍스트의 추가와 삭제 빈도를 합산하여 작업 시간과의 상관관계를 살펴본 결과 중간 정도의 양의 상관관계가 있음을 알 수 있었지만, 이 역시 통계적으로 유의미하지는 않았다($R=0.409$, $P=0.146$). 따라서 기술적 노력이 작업 시간에 현격한 영향을 주는 요소라 보기는 어렵다. 하지만 텍스트 추가에 필요한 노력은 작업 시간에 어느 정도 영향을 준다고 해석할 수 있다.

상기 데이터의 함의를 요약하자면 표면적으로 나타난 노력은 무시할 수 없는 요소이지만, 결코 해당 요소만으로 작업 시간을 모두 설명할 수 있는 것은 아니라는 점이다. 따라서 표면적으로 나타난 기술적 노력 이외에, 인지적 노력을 추가적으로 분석할 필요성이 있어 보인다.

4.3 인지적 노력

휴지를 어떻게 분석해야 인지적 노력을 대변할 수 있을지는 다수의 학자가 연구해 온 주제이다. 본 연구는 선행 연구에서 언급된 휴지의 총량적 접근을 먼저 시도했으며, 그 이후 상세한 분석을 위해 별도의 공식을 사용하였다.

먼저 상세 분석을 위해 선형 뷰에 있는 모든 휴지의 빈도와 길이를 수기로 엑셀 파일에 입력하여 분석하였다. 포스트에디팅은 기계번역 결과물이 주어지기 때문에 원천 텍스트와 기계번역 결과물을 모두 읽어야 한다는 점에서 인간번역과 큰 차이가 있다. 이에 텍스트를 읽는데 걸리는 시간이 휴지로 관찰될 수 있으며, 인간번역 대비 긴 시간의 휴지가 다수 발생할 것이라 가설을 수립하였다. 하지만 다양한 관점에서 휴지의 양상을 정량화했지만, 해당 가설을 뒷받침하지는 못했다.

〈표 6〉 다양한 휴지 정량화 데이터

	휴지 빈도	휴지 길이 합	휴지 평균 길이	2초 이하	5초 이상	10초 이상	20초 이상
HT1-1	88	213.5	2.426	55	8	1	0
PE1-1	53	134.5	2.539	26	4	0	0
HT1-2	67	301.7	4.503	29	14	7	4
PE1-2	94	512.2	5.449	29	24	16	6
HT1-3	82	423.6	5.166	41	19	11	4
PE1-3	67	436.7	6.518	26	20	12	4
HT1-4	97	368.2	3.795	29	16	8	2
PE1-4	44	98	2.228	23	2	1	0
HT2-1	67	183.7	2.743	22	5	0	0
PE2-1	48	270.4	5.633	15	11	8	1
HT2-2	85	408.3	4.803	17	13	9	3
PE2-2	21	446.5	21.262	5	14	9	4
HT2-3	69	187.4	2.716	49	2	0	0
PE2-3	52	159.5	3.069	27	4	1	0
더 높은 비율	HT:4 PE:3	HT:4 PE:3	HT:2 PE:5	HT:5 PE:1	HT:4 PE:3	HT:3 PE:3	HT:2 PE:4

*인간번역 포스트에디팅 휴지가 동물이면 정량화에서 제외

표에서 확인할 수 있듯 휴지의 빈도, 휴지 길이의 총합으로는 인간번역과 포스트에디팅의 작업 방식이 상이한지를 파악하기 어려워 보인다. 다만 휴지 길이의 평균에 있어 포스트에디팅에서 더 긴 휴지를 보인 로그가 5건인 만큼 포스트에디팅의 휴지가 길다고 해석할 수도 있다. 하지만 상대적으로 장기간의 휴지라고 할 수 있는 5초, 10초, 20초 휴지의 정량적 비교 결과, 포스트에디팅의 휴지가 인간번역보다 길다는 가설은 성립하지 않았다. 여기에 더해 오브리언(O'Brien 2006) 역시 기존의 연구를 요약하며 휴지의 길이와 빈도 자체는 개인의 성향과 연관성이 있다는 보고를 했던 것을 감안하면, 단순 휴지의 빈도와 길이의 정량화를 통해 인간번역과 포스트에디팅의 인지적 노력을 파악하기는 어려워 보인다.

조금 더 상세한 분석을 위해 ‘휴지의 수/단어의 수’로 인지적 노력을 추정하는 PWR(Pause-to-word ratio)을 산출해 보았으며, PWR은 수치가 높을수록 더 많은 인지적 노력이 필요했음을 시사한다(Vieira 2016). 하지만 얼마만큼의

길이를 휴지로 정의할지의 문제가 여전히 남는다. 기존의 연구에서는 밀리세컨드 단위를 분석해야 함을 주장한 연구가 다수 존재한다(Green, Heer and Manning 2013; Lacruz, Shreve and Angelone 2012 등). 하지만 본 연구자가 수집한 데이터는 1.00초가 최소 단위이고 1.00초 이후에만 밀리세컨드 단위로 기록이 되어있다. 따라서 1초 이상의 휴지를 인지적 노력의 지표로 분석한 연구 설계에 기초하여(Lacruz and Shreve 2014) 1.00초 이상 2.00초 미만의 휴지를 PWR의 분석 대상으로 설정하였다.

〈표 7〉 PWR을 통한 인지적 노력 비교

구분	PWR	PWR 높은 번역방식
HT1-1	0.534	HT
PE1-1	0.239	
HT1-2	0.282	HT
PE1-2	0.266	
HT1-3	0.398	HT
PE1-3	0.239	
HT1-4	0.282	HT
PE1-4	0.211	
HT2-1	0.216	HT
PE2-1	0.146	
HT2-2	0.165	HT
PE2-2	0.049	
HT2-3	0.476	HT
PE2-3	0.265	

휴지 길이 1.00초 이상 2.00초 미만 PWR 분석에 따르면 인간번역에 더 많은 인지적 노력이 필요한 것으로 보인다. 하지만 PWR은 특정 휴지 길이에 한정된 데이터 분석이기에 전체 휴지를 아우르는 또 다른 분석을 시도하였다. 휴지당 평균시간/단어당 평균시간을 계산하는 APR(average pause ratio)을 적용했으며, APR은 수치가 낮을수록 인지적 노력이 높은 것으로 해석한다(Lacruz, Shreve and Angelone 2012).

〈표 8〉 APR을 통한 인지적 노력 비교

구분	APR	APR 낮은 번역방식
HT1-1	0.377	HT
PE1-1	0.477	
HT1-2	0.943	HT
PE1-2	1.057	
HT1-3	0.768	HT
PE1-3	2.900	
HT1-4	0.759	PE
PE1-4	0.624	
HT2-1	0.618	HT
PE2-1	1.023	
HT2-2	0.776	HT
PE2-2	4.474	
HT2-3	0.496	HT
PE2-3	0.828	

APR 계산에 따르면, HT1-4를 제외하고는 인간번역이 포스트에디팅보다 더 많은 인지적 노력을 필요로 하는 것으로 보인다. 이상의 분석에 따르면 상기 PWR과 APR 데이터는 한영 포스트에디팅 수행에 필요한 인지적 노력이 인간번역 대비 적다는 가설을 수립할 수 있는 데이터이기에 의미가 있고, 앞으로 추가적으로 연구가 더 필요한 부분이다.

다음으로 휴지를 통해 알아본 인지적 노력과 시간적 노력의 상관관계를 분석해 보고자 한다. 상관관계 분석을 위해 시각적 그래프 및 통계적 분석을 실행한 결과 PWR(P=0.200, R=0.364)과 APR(P=0.1984, R=-0.3657)이 시간적 노력과 상관관계가 높다고 보기는 어렵다. 그 원인을 파악하기 위해 인지적 노력과 시간적 노력의 상관관계를 낮추는데 기여한 참여자의 데이터를 추가적으로 분석할 필요가 있다. 즉 인간번역에서 인지적 노력이 더 필요한 것으로 나타났지만, 시간적 노력은 포스트에디팅에서 더 많이 필요했던 참가자를 추가적으로 분석할 필요가 있었다.

〈표 9〉 포스트에디팅에서 시간적 노력은 높으나 인지적 노력은 낮은 참여자

	작업 시간(초)	APR	PWR	전체 이벤트	텍스트 추가	텍스트 삭제	기타 이벤트
HT 1-2	580	0.943	0.282	1011	648	186	127
PE 1-2	693	1.057	0.266	709	265	162	282
HT 2-1	386	0.618	0.216	1326	1051	172	103
PE 2-1	494	1.023	0.146	2218	199	212	1807

두 참여자 모두 공통적으로 텍스트 추가가 인간번역에서 포스트에디팅보다 압도적으로 많다는 공통점을 보였다. 작업 시간을 보면 텍스트의 추가가 작업 시간에 큰 영향을 준다고 보기는 인지적 노력에 오히려 큰 영향을 주는 것으로 보인다. 통계적으로도 APR($P=0.023$, $R=-0.599$)과 PWR($P=0.017$, $R=0.620$) 모두 기술적 노력의 텍스트 추가와 유의미한 상관관계를 보였다. 반면 APR과 PWR 모두 기술적 노력의 텍스트의 삭제 및 기타 이벤트와 통계적으로 유의미한 상관관계를 보이지 못했다.

요약하자면 휴지에 대한 상세 분석을 통해, 인간번역이 포스트에디팅보다 더 많은 인지적 노력이 필요함을 보여주는 새로운 데이터를 확보할 수 있었다. 하지만 인지적 노력만으로도 시간적 노력을 설명하기 어려워 보인다. 또한, 인지적 노력은 기술적 노력의 텍스트 추가와 통계적으로 유의미한 상관관계를 보였다. 따라서 텍스트 추가와 삭제 등 기술적 노력을 상세 분류하고 행위별로 필요한 노력의 정도가 다른지를 추가 분석할 필요성이 있다.

5. 논의 및 결론

본 연구는 한영 포스트에디팅의 시간적, 기술적, 인지적 노력에 관한 기초적인 연구로서 트랜스로그를 사용하여 데이터를 수집 및 분석하였다. 분석 결과 전반적으로 포스트에디팅이 인간번역보다 적은 시간적 노력을 보였다. 하지

만 모든 참여자가 포스트에디팅에서 시간적 노력이 적은 것은 아니었으며, 텍스트마다 개인마다 생산성 증가 비율이 다를 수 있다는 결과를 얻었다. 여기에 더해 품질 평가에서 포스트에디팅 결과물의 품질이 인간번역 품질보다 낮은 경우도 있었다. 따라서 이상의 분석 결과는 포스트에디팅의 생산성이 인간번역의 생산성보다 높을 가능성을 일부 지지한다고 요약할 수 있다. 하지만 동시에 시간적 노력을 정량적으로 일반화하는 접근은 정보 왜곡의 가능성이 있음을 시사한다고 하겠다.

기술적 노력의 분석 결과, 키스트로크와 시간적 노력이 높은 상관관계를 지니지 않는 것으로 나타났다. 다만 텍스트의 추가 행위 및 텍스트의 삭제 행위의 합은 시간적 노력과 중간(*medium*) 정도의 상관관계가 있는 것으로 나타났다(단, 통계적으로 유의하지는 않음). 이는 기술적 노력이 시간적 노력을 설명하는 데 있어 의미가 없는 것은 아니지만, 시간적 노력을 설명하는 유일한 요소는 아니라고 풀이된다. 따라서 기술적 노력만을 측정만 보상 체계는 합리적인 보상 체계라 할 수 없다.

인지적 노력의 분석 결과, 적어도 본 데이터에서는 인간번역과 포스트에디팅의 인지적 노력 패턴이 뚜렷하게 다르지 않았다. 다만 거의 모든 경우 인간번역에서 더 많은 인지적 노력이 필요한 것으로 나타났다. 또한, 추가 분석 결과 인지적 노력은 시간적 노력과 높은 상관관계를 보이지 않았다. 결국, 인지적 노력만으로도 포스트에디팅의 시간적 노력을 모두 설명하기는 어려워 보인다.

반면 기술적 노력의 텍스트 추가 이벤트와 인지적 노력과 상관관계가 높은 것으로 나타났다. 이는 모든 기술적 노력이 동일한 인지적 노력이 필요하지 않음을 시사하는 데이터로, TER처럼 모든 기술적 노력 이벤트를 동일한 가중치에 기반해 계산하는 방식은 작업자의 인지적 노력을 충분히 반영할 수 없을 것이다.

결론적으로 시간적 노력을 설명하기 위해서는 최소한 기술적 노력과 인지적 노력을 모두 고려할 필요가 있으며, 둘 중 어느 하나의 노력만을 고려하는 것은 논리적인 접근이라 보기 어렵다. 여기에 추가적으로 어떠한 다른 요소가 시간적 노력을 구성하는지 식별할 필요가 있으며, 어떠한 노력이 어느 정도 비중으로 시간적 노력을 구성하는지 지속적으로 연구할 필요가 있어 보인다. 그리고 해당 연구의 결실이 나오는 순간 포스트에디팅의 진정한 속성을 이해하고

또한 합리적인 보상 체계 마련이 가능할 것이다.

본 연구는 한영 포스트에디팅 연구에 필요한 생산성 기초데이터를 제공했으며, 기술적 노력 분석을 통해 인간번역과 포스트에디팅의 작업 방식이 다를 수 있으며, 인간번역 대비 포스트에디팅의 인지적 노력이 적을 수 있다는 가설을 수립했다는 점에서 의미가 있다.

하지만 매우 제한적인 데이터로 분석을 진행한 것이기에 데이터를 추가로 확보하여 연구를 진행한다면, 본고에서 보고한 내용이 달라질 가능성이 있다. 구체적으로는 대학원을 졸업한 지 얼마 되지 않은 번역사들을 대상으로 했기에, 번역 경력이 많은 전문 번역사를 대상으로 연구가 진행될 경우 다른 결과가 나올 수 있다. 또한, 확보한 데이터에 대해서도 정량적 분석만을 실행했기에 다른 관점에서 추가적으로 분석할 경우 더 유의미한 통찰이 가능할 수 있다. 따라서 본 예비연구에서 관찰된 사항에 기반하여 포스트에디팅 수행에 필요한 노력에 관한 연구 범위를 확대하고 연구 방법론을 더욱 고도화할 필요성이 있다.

참고문헌

- 김순미, 신호섭, 이준호 (2019) 「번역학계와 언어서비스업체 (LSP) 간 산학협력 연구: ‘포스트에디팅 생산성’과 ‘기계번역 엔진 성능 비교」, 『번역학연구』 20(1): 41-76.
- 김자경 (2020a) 「전문가와 학생의 번역 과정에 나타난 검색 및 자기교정 관찰 연구」, 『T&I REVIEW』 10(1): 219-232.
- 김자경 (2020b) 「영한 은유 번역 과정 고찰-전문가와 학생의 비교 분석을 중심으로」, 『번역학연구』 21(1): 31-59.
- 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.
- 이상빈 (2018) 「국내학부 번역전공자의 구글 기계번역 포스트에디팅 과정

- (process) 및 행위 연구」, 『번역학연구』 19(3): 259-286.
- 이준호 (2021a) 『포스트에디팅 교육 실행연구: 석사과정 영한 포스트에디팅 수업 최적화 방안에 대한 고찰』, 한국외국어대학교 대학원 박사학위논문
- 이준호 (2021b) 「영한 포스트에디팅 생산성에 대한 고찰-시간적 노력을 중심으로」, 『통번역학연구』 25(2): 55-83.
- Aranberri, Nora, Gorka Labaka, Arantza Diaz de Ilarraza and Kepa Sarasola (2014) ‘Comparison of post-editing productivity between professional translators and lay users’, *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas*: 20-33.
- Carl Michael, Barbara Dragsted, Jakob Elming, Daniel Hardt and Arnt Lykke Jakobsen (2011) ‘The process of post-editing: A pilot study’, *Copenhagen Studies in Language*, 41: 131-142.
- Carl, Michael (2012) ‘Translog-II: a Program for Recording User Activity Data for Empirical Reading and Writing Research’, *LREC* 12: 4108-4112.
- Carl, Michael, Silke Gutermuth and Silvia Hansen-Schirra (2015) ‘Post-editing machine translation’, in Aline Ferreira and John Schwieter(eds) *Psycholinguistic and cognitive inquiries into translation and interpreting*, Amsterdam: John Benjamins, 115-145.
- Carl, Michael, Akiko Aizawa and Masaru Yamada (2016) ‘English-to-Japanese translation vs. dictation vs. post-editing: Comparing translation modes in a multilingual setting’, *Proceedings of the Tenth International Conference on Language Resources and Evaluation*: 4024-4031.
- Carter, Meagan, Samanta de Frutos García, Alexandra López Vera, Karina Ornelas, Aline Ferreira, Giselle Barbosa and Daniela Guerra (2020) ‘Translating the Violent Content of Grim Brother’s Frairy Tales: An Eye-tracker Exepreiment’, *Cadernos de Tradução*, 40: 17-33.
- Fonseca, Norma Barbosa de Lima (2019) ‘Analysing the impact of TAPs on temporal, technical and cognitive effort in monolingual post-editing’, *Perspectives* 27(4): 552-588.
- Gaspari, Federico, Antonio Toral, Sudip Kumar Naskar, Declan Groves and

- Andy Way (2014) 'Perception vs reality: Measuring machine translation post-editing productivity', *Third Workshop on Post-Editing Technology and Practice* 60: 60-72.
- Green, Spence, Jeffrey Heer and Christopher Manning (2013) 'The efficacy of human post-editing for language translation', *Proceedings of the SIGCHI conference on human factors in computing systems*: 439-448.
- Guerberof-Arenas, Ana (2008) 'Productivity and quality in the post-editing of outputs from translation memories and machine translation', *Localisation Focus* 7(1): 11-21.
- Guerberof-Arenas, Ana (2014) 'Correlations between productivity and quality when post-editing in a professional context', *Machine Translation* 28.3-4: 165-186.
- Jia, Yanfang, Michael Carl and Xiangling Wang (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: 60-86.
- Koglin, Arlene (2015) 'An empirical investigation of cognitive effort required to post-edit machine translated metaphors compared to the translation of metaphors', *Translation & Interpreting* 7(1): 126-141.
- Koponen, Maarit, Leena Salmi and Markku Nikulin (2019) 'A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output', *Machine Translation* 33(1): 61-90.
- Koehn, Philipp (2009) *Statistical machine translation*, New York: Cambridge UP.
- Krings, Hans (2001) *Repairing texts: Empirical investigations of machine translation post-editing processes*, Kent, Ohio: Kent State UP.
- Lacruz, Isabel, Gregory Shreve and Erik Angelone (2012) 'Average pause ratio as an indicator of cognitive effort in post-editing: A case study', *Workshop on Post-Editing Technology and Practice*: NA.
- Lacruz, Isabel and Gregory Shreve (2014) 'Pauses and Cognitive Effort in

- Post-Editing’, in Sharon O’Brien, Laura Winther Balling, Michael Carl, Michel Simard and Lucia Specia (eds) *Post-editing of Machine Translation: Processes and Applications*, Newcastle upon Tyne: Cambridge Scholars Publishing, 246-272.
- Lacruz Isabel, Michael Denkowski and Alon Lavie (2014) ‘Cognitive demand and cognitive effort in post-editing’, *Proceedings of the third workshop on post-editing technology and practice, AMTA*: 73-84
- Nitzke, Jean (2016) ‘Monolingual post-editing: An exploratory study on research behaviour and target text quality’, *Eyetracking and Applied Linguistics* 2: 83-108.
- O’Brien, Sharon (2006) ‘Pauses as indicators of cognitive effort in post-editing machine translation output’, *Across Languages and Cultures* 7(1): 1-21.
- O’Brien, Sharon (2011) ‘Towards predicting post-editing productivity’, *Machine translation* 25(3): 197-215.
- Plitt, Mirko and François Masselot (2010) ‘A productivity test of statistical machine translation post-editing in a typical localisation context’, *The Prague bulletin of mathematical linguistics* 93: 7-16.
- Screen, Benjamin Alun (2016) ‘What does Translation Memory do to translation?: The effect of Translation Memory output on specific aspects of the translation process’, *Translation & Interpreting* 8(1): 1-18.
- Vieira, Lucas Nunes (2016) ‘How do measures of cognitive effort relate to each other? A multivariate analysis of post-editing process data’, *Machine Translation* 30(1-2): 41-62.
- Vieira, Lucas Nunes (2017) ‘From process to product: links between post-editing effort and post-edited quality’, in Arnt Lykke Jakobsen and Bartolomé Mesa-Lao (eds) *Translation in transition: Between cognition, computing and technology*, Amsterdam: John Benjamins, 162-186.
- Yildiz, Mehmet (2020) ‘How Do Translation Students’ Cognitive Efforts Vary?-An Answer in Consideration of Pauses’, *Journal of Education and Practice* 11(2): 48-55.

<인터넷 자료>

신수정 (2021.04.16.) 「인공지능(AI), 전문성과 정확성 두 마리 토끼 잡는다」 『산
업융합저널』, 2021년11월 5일 검색

스마트라이온 (2019.09.17). 2021년 11월 5일 검색

([https://smartlion.co.kr/news-%EA%B8%B0%EA%B3%84%EB%B2%88%
EC%97%ADtrends/](https://smartlion.co.kr/news-%EA%B8%B0%EA%B3%84%EB%B2%88%EC%97%ADtrends/))

TAUS (2013) ‘TAUS Best Practice Guidelines’, Available at
[https://www.taus.net/index.php?option=com_rsfiles&layout=preview&tmpl=
component&path=Articles%2FFinal-TAUS-PricingMTPE-Guidelines-Sept
ember2013.pdf](https://www.taus.net/index.php?option=com_rsfiles&layout=preview&tmpl=component&path=Articles%2FFinal-TAUS-PricingMTPE-Guidelines-September2013.pdf)

[Abstract]

A Pilot Investigation into Korean to English Post-editing Efforts

Lee, Jun-ho

(Hankuk Univervisty of Foreign Studies)

Machine translation post-editing(MTPE) service is already available in the Korean market. However, it is still challenging to have access to reliable Korean to English MTPE productivity and quality data. Therefore, it is premature to assume that Korean to English MTPE service will be more productive and of similar quality to human translation. To make matters worse, the lack of understanding about MTPE can misguide the market and future translation training. To that end, this research tries to shed light on crucial efforts required for MTPE: temporal, technical, and cognitive. First, the temporal data and expert quality assessment will provide basic insight about productivity between human translation and MTPE. Second, the keystroke data will show the difference in keystrokes between the two translation modes, if there are any. Third, the pause data will tell us if participants made different levels or patterns of cognitive efforts. Finally, based on these data sets, this research will suggest what can be a reasonable way to understand Korean to English MTPE.

- ▶ Key Words: machine translation, post-editing, MTPE, post-editing training, cognitive effort
- ▶ 주제어: 기계 번역, 포스트에디팅, MTPE, 포스트에디팅 교육, 인지적 노력

이준호

한국외대 강사

cuefit@gmail.com

관심분야: 기계번역, 포스트에디팅, 번역교육

논문투고일: 2021년 11월 7일

심사완료일: 2021년 11월 28일

게재확정일: 2021년 12월 12일