

## 번역학계와 언어서비스업체(LSP)간 산학협력연구: ‘포스트에디팅 생산성’과 ‘기계번역 엔진 성능 비교’\*

김순미\*\* · 신호섭 · 이준호  
(숙명여대 · 한샘EUG)

### 1. 서론

2016년 신경망번역기(neural machine translation NMT) 등장 이후 기계번역(machine translation MT) 결과물은 다양한 영역에서 폭넓게 사용되고 있다.

MT번역은 대량 콘텐츠를 번역할 때 가격을 절감하고, 신속하고 일관된 결과를 제공하여 효율성을 올릴 수 있으며(김련희 2011: 15; 김순미 2017: 6) 특히 단순반복적인 텍스트 번역의 효율성을 높이는 번역 메모리(translation

\* 본 연구는 한국번역학회와 한국IT산업세계화학회(KIGO)의 산학연구 결과이다. 제1저자이자 교신저자 김순미가 전체 연구의 협력 지원 및 집필 책임을 맡았고, 공동저자 신호섭은 연구 운영과 데이터 처리, 이준호는 데이터 결과 분석과 집필을 맡았다. 일년간 물심양면으로 지원을 해주신 번역학회 김순영 회장님과 KIGO 박형택 (전)회장님, 연구에 참여해 주신 업계, 학계의 모든 분께 깊은 감사의 뜻을 전한다.

\*\* 제1저자, 교신저자

memory TM)와 함께 쓰일 때 언어 수행 범위를 크게 확장시켜 주기 때문이다. 그러나 상업적 용도, 출판, 대중 유포 등을 목적으로 하며 법적 규제를 준수하는 번역 품질에 달하기 위해서는 인간이 개입하여 기계번역을 전처리하는 과정인 프리에디팅과 후처리하는 과정인 포스트에디팅이 필수적이다(Doherty 2016: 958). 이는 기계를 활용하는 번역량이 증가하는 만큼 기계번역 포스트에디팅(machine translation post-editing MTPE)을 필요로 하는 텍스트도 엄청나게 늘어난다는 의미이다. 언어서비스업체(Language Service Providers LSP)<sup>1)</sup>들은 이미 이같은 변화의 물결을 타고 MTPE나 MT 평가, 개선 등 다양한 업무를 수행하고 있다(Zaretskaya 2017: 117).

그렇다면 국내 언어서비스 시장의 MTPE 관련 현황은 어떤가? 국내 언어서비스업체들의 모임인 한국IT산업세계화학회 키고(Korea IT Globalization Organization KIGO) 회원사들에 의하면 점차 많은 LSP들이 MTPE를 활용한 번역을 고객들에게 제공하고 있으나 아직 도입 초기 단계여서 여러 가지 문제가 불거지는 상황이다. 가장 큰 문제는 번역료 책정의 어려움이다. MTPE의 서비스 및 과금 방식이 구체적이고 세부적으로 정해지지 않아 LSP마다 상황에 따라 다르게 단가를 책정하기도 하고, 고객들은 MT를 돌렸다는 이유만으로 텍스트 난이도나 MT엔진 성능과 특징 등 다른 변수를 고려하지 않은 채 단가를 낮추려고 하여 불이익을 보는 경우가 있다고 한다. 또한 MTPE 단가는 인간이 한 번역과 유사한 수준의 정식 포스트에디팅인지 언어적, 스타일적인 면에서는 완벽하지 않지만 의미 전달은 충분히 가능한 약식 포스트에디팅인지에 따라 차별화 되어야 하는데 번역료는 약식 수준으로 낮게 지불하고 정식 수준의 결과

1) 세계적인 언어서비스업체인 SDL(<https://www.sdl.com/kr/>)과 트랜스퍼펙트(TransPerfect) (<https://kr.transperfect.com/>) 등의 웹사이트와 김련희(2011), 박지영(2017) 등에 의하면, LSP는 번역, 통역 뿐 아니라 트랜스크리에이션, 카피라이팅, 감수 등 다양한 언어서비스, 프로젝트 관리, 웹사이트 현지화, 멀티미디어 서비스(더빙, 보이스 오버), 소프트웨어 로컬리제이션, 글로벌(다문화) 마케팅, 콘텐츠 개발부터 판매까지 전 프로세스 관리, 언어 교육, CAT를 포함 번역 소프트웨어 개발, 언어 컨설팅, 기계번역 서비스 등 다국적 기업들의 글로벌화와 현지화에 있어 장벽을 낮추는 포괄적이고 총체적 서비스를 제공한다. 사업 분야는 소매 및 전자상거래, 생명과학, 법률, 방송 및 엔터테인먼트, 금융, 광고, 마케팅, PR, 산업 및 제조, 여행 및 숙박업, 에너지 및 광업, 하드웨어, 소프트웨어, 기술, 정부 다국어 번역 서비스 등 전문 분야를 망라한다.

물을 원하는 고객사도 있다고 한다. 이에 따라 회사들은 MT 성능과 특징, 초벌 기계번역 품질, 시간과 노력 절약 정도, 고객 요구사항 등 다양한 상황별로 구체적인 근거를 가지고 MTPE 번역료를 요청하는 시스템을 마련하려 노력하고 있다. 결국 업체들이 MT 생산성에 관심을 갖는 중요 이유 중 하나는 번역료 책정 때문이라 할 수 있다(Koponen 2016: 132).

급속히 발전하는 현실에서 번역 교육을 담당하는 교수자와 연구진들 또한 MTPE에 관심이 많다. 학생들 지도를 위해 효과적인 커리큘럼을 구축해야 함과 동시에 MTPE 가이드라인 설정, MT번역 품질 분석, 구급번역기나 파파고, 카카오 등 주요 범용엔진 별 성능 비교 등 연구의 필요성도 느끼고 있다. 서양에서는 MT관련 많은 연구가 나와 있으나 한국어는 서양어와 구조 및 문법에 있어 다양한 차이를 가지고 있는 언어이기 때문에 서양어-서양어를 주로 연구한 해외의 사례를 한국 번역 시장과 교육현장, 실무에 그대로 적용하는 것은 무리가 따른다(윤미선, 김택민, 임진주, 홍승연 2018). 이에 따라 한국의 실정에 맞는 영어-한국어 언어쌍의 MTPE 실무적용 타당성 여부를 평가하는 일은 업계와 학계 모두에 매우 시급한 문제가 되었다.

이와 같은 상황에서 한국번역협회의 연구자들, 전공학생들과 번역업계의 LSP들은 키고의 <2018년 신경망MT 분과>를 통해 MTPE 생산성과 엔진 성능 비교라는 공동 연구 주제를 가지고 총 5회의 산학협력공동연구 모임을 가졌다. 업계는 학계의 연구 틀과 MTPE 실험 참여자 동원 면에서 지원을 받았고, 학계는 현업에서 활용하는 MTPE 운영법을 도입할 수 있었다. 연구주제는 첫째, 기계번역에 적합한 텍스트를 선정한 후, 인간번역과 MTPE를 거친 결과물 간에 소요된 시간과 노력, 수정률 등을 기준으로 생산성이 얼마나 차이가 나는지 비교 분석하는 것, 둘째, 각각 구성이나 학습데이터가 다른 한국어-영어 번역 기준 대표 MT엔진들(구급번역기, 파파고, 카카오)의 성능이나 오류 유형은 어떤 차이가 있는지 살펴보는 것이다. 문제를 풀어가는 과정에서 참가자들은 업계와 학계의 공통 관심사인 ‘과학기술 발전과 통번역’이라는 주제를 본격적으로 탐구한 국내 첫 연구로 산학협력의 모델을 만들고자 했다. 본 연구를 통해 MTPE의 실무적용과 향후 가격 책정 등에 있어 근거를 제시할 수 있는 실마리를 찾는다면 실무자와 교육자 모두에게 유용한 정보를 제공할 수 있을 것이다.

## 2. MT 관련 중요 이슈와 연구의 틀

### 2.1 MT엔진 품질

MT와 관련하여 우리의 관심이 가장 높은 두 가지 이슈는 엔진의 품질이 얼마나 좋아지고 있으며 이것이 번역 생산성에 어떤 영향을 미칠 것인가이다. 엔진 품질에 대해 알아보려면 우선 MT 엔진의 종류에서 시작해야 할 것이다. 언어서비스, 번역업계의 기술발전과 최신 뉴스를 전하는 슬레이터(Slator)의 ‘2018 NMT 보고서(Slator Neural Machine Translation Report 2018)’에 의하면 실제 NMT 등장 이후 기계번역의 품질은 과거 통계기반번역(Statistical Machine Translation SMT) 대비 훨씬 우수한 품질을 보이고 있으며 이에 따라 거대 다국적 기업과 IT, 언어 관련 기업들은 앞다투어 다양한 목적과 용도를 지닌 MT를 제공하고 있다. 실제 LSP와 고객사 모두 더 방대한 콘텐츠를 더욱 신속하고 낮은 단가로 번역한다는 세 가지 궁극적인 지점을 향해 달리고 있다(Koponen 2016: 131).

규모별로 구분해보면 첫째, 대중적으로 가장 널리 알려진 범용엔진들이 있다. 구글(Google), 마이크로소프트(Microsoft), 바이두(Baidu), 아마존(Amazon)과 같은 거대 IT기업들이 제공하는 누구나 사용할 수 있는 클라우드 기반 엔진으로 기업들은 이들 플랫폼을 이용하여 기존 서비스와 고객 영역을 넓혀가고 있다. 이로 인해 일반 유저들도 무료 온라인 기계번역(Free Online Machine Translation FOMT)을 이용하여 상거래, 학습, 번역, 글쓰기 등 영역에서 데이터 검색, 분석, 질문, 요약, 응대, 기술 문서 작성 등 목적을 위해 MT를 사용하게 되었다(Hutchinson 2009: 19; Koponen 2016: 132). 두 번째는, 시스템(Systran), 칸탄MT(KantanMT), 릴트(Lilt)등 언어기술, 머신 러닝 중심 회사나 SDL, 트랜스퍼펙트(TransPerfect) 등 LSP들이 제공하는 특화된 분야에서 철저하게 보안이 이루어진 고객 중심 커스텀 엔진들이다. 세 번째는 세계지적재산권기구(WIPO)나 부킹닷컴(Booking.com) 등 기관이나 회사들이 사업상 필요에 의해 MT를 구축하여 고객의 편의를 위해 웹사이트에서 제공하는 경우이다. 특정 분야에 특화된 MT의 성능은 인간번역 결과물과 MT 결과물의 유사성을 수치화 하여 측정하는 BLEU(Bilingual Evaluation Understudy) 스코어 기준으로

볼 때 구글번역기 같은 범용엔진보다 훨씬 높은 수치를 보여 인간번역과 더 유사한 것으로 나타났다. <표 1>의 연구는 자동번역에 큰 노력을 기울이고 있는 WIPO의 특허번역에 관한 연구로 NMT이전의 구-기반 통계번역기(Phrase Based Statistical Machine Translation PBSMT)와 NMT의 차이, 그리고 분야에 특화된 엔진과 범용엔진의 차이를 잘 보여주고 있다.

<표 1> WIPO MT와 구글 MT 간의 BLEU 스코어 비교(Pouliquen 2017: 7)

Language pair	Training data size		WIPO Translate		Google Translate	
	# Sentences (million)	# English words (million)	WIPO PBSMT	WIPO NMT	Google translate (PBSMT)	GNMT
English → Chinese	62M	1872M	28.57	39.14	20.55	31.06
Japanese → English	112M	4000M	24.82	36.47	25.05	30.75
English → Japanese	112M	4000M	30.54	47.31	n/a	42.93
English → Korean	44M	1480M	26.11	39.20	22.98	32.65
English → French	18M	488M	51.70	62.42	49.58	55.19

네 번째 위치한 영어-한국어 쌍의 경우 특허에 특화된 WIPO NMT와 범용 엔진인 구글 NMT번역기(GNMT)의 BLEU스코어는 각각 39.20과 32.65로 WIPO엔진이 구글보다 인간 번역과 더 유사한 것을 알 수 있다. 또한 이 연구는 NMT등장 바로 이전인 2016년 8월부터 직후인 2017년 1월까지 5개월간을 분석한 것인데 WIPO와 구글번역기 모두 영어-한국어 쌍에서 SMT보다 NMT의 BLEU 스코어가 훨씬 높아진 것 (WIPO 엔진(26.11→39.20), 범용 Google 엔진(22.98→32.65))을 볼 때 짧은 기간 동안에도 기계가 얼마나 크게 발전했는지를 볼 수 있다. 실제 현장에서 LSP들은 IT, 생명과학, 법률 등 고객 특화된 엔진을 활용한다는 점에서 분야 특화된 MT를 실험에 이용한다면 더욱 의미 있는 결과를 도출할 수 있을 것이나 고객사들의 철저한 보안문제로 인해 사용이 어려운 실정이므로 본 연구에서는 범용엔진을 활용했다.

## 2.2 MTPE 생산성

서구에서는 이미 향상된 기계번역 결과물에 대해 적절한 인간의 MTPE를 거친다면, 인간이 수작업으로 처음부터 번역을 진행하는 것과 비교하여 생산성

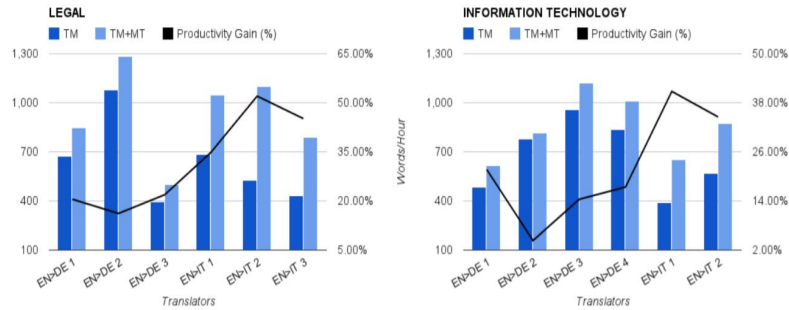
이 높으면서도 품질도 좋은 번역물을 생성할 수 있다는 연구들이 많이 나와 있다(Aranberri et al. 2014; Carl et al. 2011; Federico et al. 2012; Garcia 2010; Guerberof 2009, 2014; Läubli et al. 2013; Martínez 2003; O'Brien 2007, 2011; Plitt and Masselot 2010; Zampieri and Vela 2014). MTPE 생산성은 크게 소요된 시간과 작업에 들인 노력으로 구분해 볼 수 있는데 번역을 의뢰하는 입장에서는 생산성 향상이 매우 매력적이기 때문에 해외의 경우 글로벌 기업 매니저의 28%가 기계번역을 사용하고 있거나 사용할 계획을 가지고 있으며, 응답자의 58%는 MTPE를 긍정적으로 고려하고 있다고 답하고 있다(Carl et al. 2015). 그러나 MTPE 생산성 관련 연구를 하거나 수용하는 과정에서는 주의해야 할 점이 두 가지 있다. 첫 번째는 언어쌍, 번역 방향, 도메인의 시스템 종류에 관계없이 MTPE 생산성을 높이기 위해서는 다음과 같은 환경이 기본적으로 제공되어야 한다는 점이다.

1. MTPE를 하는 환경을 미리 잘 준비해야 한다. 트레이닝이 잘 된 엔진과 수준 높은 어학 사전이 기본적으로 제공되어야 한다.
2. 원문이 깔끔하게 작성되어 있어야 한다. 철자, 맞춤법, 모호성 등을 사전에 체크하고 좋은 MT 결과물을 낼 수 있도록 제작된 텍스트를 준비한다.
3. 원문 제작과정, MT, TM 등 전 과정에서 용어집을 적용하도록 한다.
4. 작업 전 포스트에디터를 사전 교육시킨다.
5. 번역량과 단가 협상 전에 MT 최초 결과물 품질수준을 점검하고 합리적인 MTPE 기대치를 설정한다.
6. 텍스트 사용자의 유형과 텍스트 수용성에 기초하여 MTPE를 하는 텍스트의 최종 품질 기준에 대해 사전 동의를 한다.
7. 포스트에디터가 MT 오류 유형에 대해 구조적인 피드백을 줄 수 있도록 가이드라인을 미리 제공하여 향후 시스템이 개선될 수 있도록 한다. (TAUS 2016: 15)

두 번째는 생산성 향상 연구는 보편적인 상황을 아우르는 것이 아니라 특정 조건을 상정해야 한다는 점이다. 예를 들어 생산성 측정에는 영어-독일어, 영어-일본어 등 특정 언어쌍, 과학기술 문서, 매뉴얼, 법률 문서 등 특정 텍스트, MT 최초 결과물이 좋은 경우와 그렇지 않은 경우, CAT툴과의 결합 유무

와 그 정도, 엔진의 종류 등 다양한 조건이 전제된다. 이와 관련하여 생산성 관련 연구의 예로 <표 2>에 나와 있는 페데리코 외(Federico et al.) 연구를 들 수 있다. 이 연구는 언어쌍은 영어와 독일어, 영어와 이탈리아어 쌍으로, 텍스트 종류는 법률 문서와 기술 문서로, 연구 방법은 과거 구축한 트라도스 스튜디오 TM을 돌리며 번역을 한 경우와 TM과 MT를 함께 사용한 경우를 비교하고 있다. 6명 번역사에게 TM만 제공한 경우 시간당 처리 단어 수와 TM에 MT를 함께 제공한 경우 MTPE 처리 단어 수를 관찰한 결과 TM과 MT를 함께 사용한 경우 모든 번역사의 생산성이 현격하게 증가하는 것을 볼 수 있다.

<표 2> TM만 적용한 경우와 TM+MT 적용한 경우 시간당 MTPE 업무량(Federico et al. 2012)



이에서 보듯 MTPE 생산성 연구는 연구 방법 설정이 매우 중요하다. 위의 연구와 같이 생산성 관련 다수의 선행연구(Guerberof 2009; Läubli et al. 2013; Plitt and Masselot 2010; Zampieri and Vela 2014)는 인간번역과 MTPE 생산성을 직접 비교하는 것보다 인간이 TM을 활용하여 번역을 하는 경우와 MT와 TM을 함께 활용하는 경우를 비교하고 있다. 번역 메모리 포함 각종 기술적 지원을 받으며 작업하는 것이 보편화 된 현업과 조건을 유사하게 하기 위함이다. 반면, 번역사가 수작업으로 번역을 하는 경우와 MT를 이용하여 번역을 하는 경우를 단순 비교하는 경우, 기계번역 도입 시 생산성 향상이 실제보다 훨씬 높게 나타날 가능성이 있다(Läubli et al. 2013). 언어 데이터 서비스 네트워크 기관인 TAUS에서도 이와 관련하여 MTPE 품질과 생산성, 시간, 번역료 관련한 매트릭스를 구축하는 경우 다음과 같은 조건들을 정교하게 분석하여 반영해야

한다고 밝히고 있다.

1. 콘텐츠 종류와 사용 목적
2. MTPE 결과물 품질에 대한 기대 수준과 특정 콘텐츠에서 수용 가능한 오류 수준과 위험도 수준
3. CAT을 함께 사용하는 상황인지, 인간이 직접 번역과 MTPE를 수행하는 경우인지. 그리고 그 차이가 생산성 향상 정도에 미치는 영향 정도
4. 인간번역사의 일일 번역량 (업계 표준 기대 번역량을 기준으로)
5. MTPE 하는 경우 일일 번역량 (생산성 향상 정도를 기준으로)
6. 생산성 향상 (시간당, 일당)
7. 위의 요소들에 근거해 예상하는 번역단가 하락 정도 (TAUS 2016: 12)

앞의 연구들과 TAUS 매트릭스를 기반해 볼 때 연구 조건을 최대한 현업 환경과 유사하게 만드는 것이 결과를 실제에 적용하는데 도움이 될 것임을 알 수 있다. 그러나 국내 번역교육 환경에서는 연구의 중요한 조건인 번역 메모리 활용이나 MTPE 교육이 되지 않고 있어 본 연구와 이후 MTPE 실험에서 제약으로 작용할 수 있음을 밝힌다.

### 2.3 국내 MTPE 연구 동향

서문에서도 밝혔듯이 기계번역 관련 연구는 한국어와 관련된 언어쌍 연구이기 때문에 국내의 연구동향 분석이 매우 중요하다. 그래서 본 연구에서는 2016년 말 NMT 등장 전후로 국내 통번역 관련 주요 등재 학술지인 『번역학연구』, 『통역과 번역』, 『통번역학연구』, 『통번역교육연구』 등에 인공지능과 기계번역 관련하여 실린 연구를 조사하여 주제별로 분류해 보았다. 2019년 2월 현재 총 32편의 연구가 발표되었다. MT 번역의 (오류) 유형과 인간번역과의 차이(9편), 통번역의 변화하는 현황과 미래(8편), 교육(7편) 등이 주를 이루었다. 특이한 점은 본 연구의 첫번째 주제인 MTPE의 시간, 노력 기준 생산성 연구는 한 편도 없다는 점이다. 간접적으로 연구 절차와 관련이 있는 MTPE 가이드라인 관련 연구인 「영어-한국어 언어쌍에 적합한 포스트에디팅 가이드라인」 한편이 있을 뿐이었다. 반면 본 연구의 두 번째 주제인 ‘엔진 성능과 오류’와 직 간접적 관련이 있는 연구는 오류 유형 연구 8편, MT의 한계/가능성 4편, 엔진 품질 연

구 3편 등으로 국내 연구의 큰 부분을 차지했다. MT 품질과 교육, 통번역의 미래에 연구가 치중된 것이 시사하는 점은 MT가 어떤 유형, 품질의 번역을 하고 있으며, 어떻게 MTPE 교육을 시킬지, 미래의 통번역은 어떤 방향으로 움직일지 등에 대한 연구자들의 관심이 지대하다는 것이다. (<표 3>에서 두 가지 연구 분야에 속하는 연구는 양쪽에 중복 포함시켰다.)

<표 3> 국내 MT관련 연구 동향

주제	제목
기계번역과 인간번역 유형과 번역 품질 (9편)	1. 영한번역에서의 시제의 등가성과 기계번역과의 비교 (박영란 2017)
	2. 특허 기계번역 결과물의 평가-KIPRIS 의 무료 한영 기계번역을 중심으로 (최효은, 이지은 2017)
	3. 기계번역 인간번역 트랜스크리에이션의 문체 비교: 광고 번역을 중심으로 (한승희 2017)
	4. 포스트에디팅 측정지표를 통한 기계번역 오류 유형화 연구 (곽중철, 한승희 2018)
	5. 기계번역 결과물의 오류유형 고찰 (서보현, 김순영 2018)
	6. 인공신경망 기계 번역의 한일/일한 번역 품질에 대한 예비연구 - 품질향상 검토와 교열코드 적용 - (이주리에 2018)
	7. 포스트에디팅 교육을 위한 포스트에디팅과 인간번역의 차이 연구(이준호 2018)
	8. 프리에디팅 (pre-editing) 이 기계번역 품질에 미치는 영향 고찰: 유창성, 충실성, 가독성 비교를 중심으로 (진용주, 서보현, 김순영 2018)
	9. 한노 기계번역의 문장 부호 처리 양상 및 오류 개선을 위한 실용적 함의: 프리에디팅 규칙 수립과 효용성의 관점에서 (한현희 2018)
MT 시대 통번역의 지평, 역사 (8편)	1. 통번역의 미래지평 : 인공지능과 소통형 융합통번역 연구 (이노신, 이신재, 이재영, 이주희 2016)
	2. 트랜스크리에이션, 기계번역, 번역교육의 미래 (이상빈 2016)
	3. 언어기술 혁신과 통번역 산업의 미래: 20대 글로벌 LSP의 통번역 서비스분석을 바탕으로 (박지영 2017)
	4. 테크놀로지 패러다임에서의 번역능력 재조명 (신지선 2017)
	5. 국내 기계 통번역의 발전 현황 분석 (장애리 2017)
	6. 4차 산업혁명과 한국의 번역산업 현황 및 통번역 교육의 미래 (전현

	주 2017) 7. AI시대 인간번역과 기계(NMT)번역의 공존-경영학 '확장(Augmentation) 전략' 중심 (김순미 2018) 8. 인공지능 번역 시스템의 출현에 대한 소고 (신지선, 김은미 2018)
MTPE 교육관련 실험 (7편)	1. 신경망번역기(NMT) 활용 학부 번역교육의 가능성 연구 (김순미 2017) 2. 학부번역전공자의 기계번역 포스트에디팅, 무엇이 문제이고, 무엇을 가르쳐야 하는가? (이상빈 2017) 3. 한영 기계번역 포스트 에디팅에 대한 경험적 고찰: 학부 교육 과정 및 결과를 중심으로 (마승혜 2018a) 4. 석사 과정의 기계번역 수업에 대한 소고: 한일번역 전공생의 포스트 에디팅 사례를 통하여 (박해경 2018) 5. 학부 번역전공자의 구글 기계번역 포스트에디팅에 관한 현상학 연구 (이상빈 2018) 6. 영-한 및 한-영 기계번역 품질향상을 위한 프리에디팅 기법 제안 (이성화, 김세현 2018) 7. 포스트에디팅 교육을 위한 포스트에디팅과 인간번역의 차이 연구 (이준호 2018)
MT의 한계와 가능성 (4편)	1.기계번역의 외연 확대-신문기사 특정 은유표현 포스트 에디팅 전략 모색 (마승혜 2018b) 2.문학작품 기계번역의 한계에 대한 상세 고찰 (마승혜 2018c) 3.기계 번역에서의 영어속담 분석 (박영란 2018) 4. 문학번역 적용을 위한 기계번역의 현주소 (이준호 2019)
구글, 파파고 등 엔진 성능비교 (3편)	1.국내 기계 통번역의 발전 현황 분석 (장애리 2017) 2.기계 번역에서의 영어속담 분석 (박영란 2018) 3.인공신경망 기계 번역의 한일/일한 번역 품질에 대한 예비연구 - 품질향상 검토와 교열코드 적용 - (이주리에 2018)
MTPE기준 (2편)	1.영어-한국어 언어쌍에 적합한 포스트에디팅 가이드라인 (윤미선, 김택민, 임진주, 홍승연 2018) 2.번역의 자동평가: 기계번역 평가를 인간번역 평가에 적용해보기 (정혜연 2018)
윤리 담론 (2편)	1. 기계번역 담론에 대한 비판적 고찰 (송연석 2018) 2. AI번역의 속성 및 번역주체에 대한 논의 (이승일 2018)

MTPE 관련 연구를 살펴본 결과 이번 연구 주제로 삼게 된 MTPE 생산성 관련 연구는 전혀 이루어지고 있지 않았다는 점은 이 주제가 참여자 선정과 실험, 결과 분석 등 모든 과정이 연구자 1인이 담당하기에 어려운 분야라는 것을 보여준다. 다시 말해 MT생산성 관련 주제야말로 산학협력을 통해 다루어 볼 수 있는 적절한 영역임을 보여준다. 한편 MTPE를 하게 되면 시간당 생산성이 올라가는 만큼 번역 단가가 내려간다는 것이 업계와 학계의 가장 큰 우려이다. 실제 세계 번역업계 전반적으로 단어당 번역단가는 2008년부터 2013년까지 50% 감소하였는데 전문가들은 경영상의 예산부족과 번역관련 기술 적용에서 단가 하락의 이유를 찾고 있다(DePalma et al., 2013: 8-9). 이런 이유로 기술의 발전이 자칫 언어서비스업계의 불황이나 번역업 자체의 축소 등으로 이어질 것이라는 우려가 많다. 그러나 기하급수적인 정보의 증가와 정보사회의 도래로 번역기술이 발전하더라도 번역가의 일감은 오히려 엄청나게 증가할 것으로 예상된다(Cronin 2013). 실제, 글로벌 시장의 확대와 커뮤니케이션의 중요성으로 인해 2008년 글로벌 금융위기에도 불구하고 2009년부터 2016년까지 LSP 중심 언어시장의 규모는 거의 두 배로 성장했다(DePalma et al., 2016: 10). 즉, 박지영(2017) 주장과 같이 MT로 인해 생산성이 올라간 만큼 번역 단가는 내려가지만 번역 물량의 증가로 이것이 상쇄될 수 있다는 의미이다. 그러나 예상치 못했던 방향으로 번역 단가가 변화될 수 있다는 예측도 나온다. MT 사용이 늘면서 번역가의 업무가 단순 번역에서 감수자, 데이터 생산자, 큐레이터, MT 운영자로 변화하게 되어 번역 단가를 계산할 때 번역한 글자 수가 아니라 업무 시간을 기반으로 번역사의 수당을 책정하는 상황이 올 수도 있다는 것이다(Slator 2018: 34). 이런 상황이야말로 MTPE 생산성과 번역 단가에 대해 산업계와 학계, 교육계 모두 관심을 가질 수밖에 없는 이유이다.

### 3. 연구의 구성

#### 3.1 참여자 모집과 일정

<신경망MT 분과>에는 6개의 KIGO 회원사 (글로벌웨이, 라이온코리아, 매

스웍스, 솔트룩스, 오라클, 한샘EUG) 및 7개 대학(동국대, 서울외대, 숙명여대, 신한대, 이화여대, 평택대, 한국외대)의 교수자 및 번역학 전공자가 참여하여, 연구 참여자 선정에 대해 다양한 의견을 교환했다. LSP의 의견을 반영하여 양측이 합의한 참여자 선정 기준은 첫째, 데이터 객관성을 보충하기 위해서 참여자들의 번역능력과 MTPE 경험을 균등한 수준으로 맞출 것, 둘째, 주 대상은 번역 전공자일 것, 세 번째, MTPE는 자신의 번역에 대한 자신감이 있어야 원활하게 수행할 수 있는 감수 작업이기에 때문에(Wangner 1987), 이를 보장 할 수 있는 수준의 교육을 받은 학생을 선정할 것 등이었다. 최종적으로 본 연구 참여 대상은 번역 전공 학부의 3-4학년 이상 및 대학원생으로 한정했다.

제1차 모임은 2018년 4월 3주차에 주최하여 전반적인 연구 목표 및 절차를 설정했으며, 제2차 모임은 5월 4주차에 열어 참여자 관련 논의를 심화하는 한편 NMT의 발전에 대한 솔트룩스 파트너스 신석환 대표님의 특강을 들었고, 제3차 모임은 6월 3주차에 개최해 본격적인 참여자 모집을 시작했다. 그 결과 6월 4주차 4개 학교에서 총 15명(한국외대 5명, 서울외대 5명, 동국대 4명, 평택대 1명)의 참여자를 모집할 수 있었다.

#### 3.2 텍스트의 설정

2018년 4월에 주최한 제 1회 모임에서 영어-한국어 쌍에 대한 MTPE 생산성 비교 연구의 대상 텍스트는 실무 적용 확률이 높은 텍스트로 한정하자는 의견에 공감했고, 5월 4주차 2회 모임에서 본격적으로 텍스트 선정을 위한 논의를 진행했다. 가장 먼저 고려한 것은 텍스트 성격이다. 정보전달, 미적 표현, 설득 등 텍스트 성격에 따라 기계번역 결과물 품질은 차이가 난다(마승혜a, 2018). 또한 기계번역은 기존의 번역 텍스트를 엔진의 학습데이터로 사용하는 만큼, 번역이 많이 이뤄지는 정형화된 텍스트의 경우 그 품질이 높고(박지영, 2017), 입력 문장의 길이가 길면 품질은 떨어지는 것으로 알려져 있다(Koponen, 2013). 따라서 본 연구에서는 정보전달 유형이며, 기존에 번역이 많이 이뤄졌으며, 문장의 길이가 짧은 텍스트 유형을 선정하고자 했다. 텍스트 주제 선정에 있어서는 LSP 관계자들이 전하는 실제 수주량을 고려했다. 가장 많이 다루는 영역은 IT 및 전자 산업이라는 답변을 얻었으며, 이는 기계번역이 가장 많이 사용되는 영역은

자동차, 전자, IT, 국방 이라고 보고한 그랜드 뷰 리서치<sup>2)</sup>나 박지영(2017) 연구 결과와 크게 다르지 않았다. 따라서 본 연구는 LSP에서 실제로 수주했던 ‘IT 관련 영한 매뉴얼’을 연구 텍스트로 선정했다. 본 연구에 사용된 텍스트는 난이도가 유사한 5개 텍스트로 평균 길이는 1,346 단어였다.

### 3.3 연구의 구성

전술한 바와 마찬가지로 본 연구의 1차적 목적은 제한된 텍스트 유형 및 주제에 대한 MTPE와 인간번역의 생산성 비교에 있다. 이를 위해 모든 참여자는 1개의 텍스트를 처음부터 직접 번역하고 다른 1개의 텍스트는 MTPE 작업하였다. 작업자들을 위한 가이드라인을 제공하기 위해 MTPE의 국제표준인 ISO18587을 참조하였다. ISO가이드라인은 MTPE를 아래 <표 4>와 같이 정식 포스트에디팅과 약식 포스트에디팅으로 분류하고 있다. 정식 포스트에디팅은 인간번역에 가까운 품질을 목표로 하기 때문에 인간번역과 비교할 때 더 객관적 데이터로 사용될 수 있다. 하지만 정식 포스트에디팅 조건 중 “고객 및 특정 영역의 언어를 준수”, “텍스트 유형에 적절한 (고객의) 스타일 사용” 등은 고객이 없는 상황에서 가상의 텍스트를 사용하는 본 연구의 구조상 준수하기 어려운 요건이다. 가상의 고객을 상정하거나 기존에 받았던 조건을 기준으로 실험을 실시할 수도 있었겠지만 이번 실험은 이같은 조건을 상정하지 않았다. 뿐만 아니라 한국 언어서비스 시장에서는 약식 포스트에디팅을 요청하는 고객사가 많다는 LSP와의 인터뷰 결과도 감안하였다. 또한 국내 대학에서 MTPE 전문교육이 거의 전무한 상태에서, 참여자들은 실험을 앞두고 간단한 교육을 받은 후 작업을 하였기 때문에 정식 포스트에디팅을 수행하는 것은 무리였다. 다만 ISO 및 TAUS 가이드라인을 참조하여 LSP에서 제작한 가이드라인을 사용하여 한국 시장의 현실을 충실하게 반영하기 위해 노력하였다.

<표 4> ISO18587 MTPE 가이드라인

정식 포스트에디팅	약식 포스트에디팅
a. 누락 혹은 추가된 정보가 없어야 함 b. 부적절한 콘텐츠 모두를 수정 c. 부정확 혹은 모호한 의미의 경우 문장을 재구성 d. 문법, 통사, 의미적으로 올바른 콘텐츠를 생산 e. 고객 및 특정 영역의 언어를 준수 f. 철자, 구두법 규칙을 적용 g. 텍스트의 유형에 적절한 스타일을 사용하여 고객사가 제시한 스타일 가이드를 준수 h. 형식(formatting) 규칙을 적용	a. 가능한 기계번역을 있는 그대로 사용 b. 누락 혹은 추가된 정보가 없어야 함 c. 부적절한 콘텐츠 모두를 수정 d. 정확 혹은 모호한 의미의 경우 문장을 재구성

(\*연구자의 번역)

본 연구의 2차 목적은 MT 엔진의 차이를 알아보는 것이다. 우선 5 가지 매뉴얼에 대해 범용엔진인 구글번역, 파파고, 카카오를 사용하여 총 15벌의 초벌 번역물을 얻었다. 그리고 <표 5>와 같이 15명의 참여자를 5명씩 각기 다른 MT 결과물을 다루는 3개의 그룹(구글 그룹, 파파고 그룹, 카카오 그룹)에 배정하였다. 각 그룹 참여자는 동일한 엔진 결과물을 할당받아 MTPE 결과물을 산출하는 한편 유사한 텍스트를 번역하는 방식으로 연구를 구성하였다. 한 참여자가 동일한 텍스트에 대해 직접 번역과 MTPE 작업을 하게 설계한다면 인간번역과 기계번역의 비교 신뢰성이 떨어질 것으로 보고, <표 5>와 같이 각각 유사하지만 다른 텍스트에 대해 번역과 MTPE 작업하도록 한 후 결과물을 비교하였다. 예를 들어 참여자 1의 경우 매뉴얼1에 대해 MTPE를 실행하고, 매뉴얼2를 직접 번역하도록 하였다.

2) <https://www.grandviewresearch.com/press-release/global-machine-translation-market>

〈표 5〉 각 엔진별 인원 및 텍스트 배정

파일	워드 수	사용도구	MTPE	인간번역
매뉴얼1	1,377	구글번역	참여자 1	참여자 5
매뉴얼2	1,346	구글번역	참여자 2	참여자 1
매뉴얼3	1,435	구글번역	참여자 3	참여자 2
매뉴얼4	1,285	구글번역	참여자 4	참여자 3
매뉴얼5	1,289	구글번역	참여자 5	참여자 4
매뉴얼1	1,377	파파고	참여자 6	참여자 10
매뉴얼2	1,346	파파고	참여자 7	참여자 6
매뉴얼3	1,435	파파고	참여자 8	참여자 7
매뉴얼4	1,285	파파고	참여자 9	참여자 8
매뉴얼5	1,289	파파고	참여자 10	참여자 9
매뉴얼1	1,377	카카오i	참여자 11	참여자 15
매뉴얼2	1,346	카카오i	참여자 12	참여자 11
매뉴얼3	1,435	카카오i	참여자 12	참여자 12
매뉴얼4	1,285	카카오i	참여자 14	참여자 13
매뉴얼5	1,289	카카오i	참여자 15	참여자 14

본 연구에 참여한 15명의 참여자는 연구에 참여하기 전 학교에서 MTPE 교육을 받은 적이 없어 MTPE 경력에 있어 차이는 없다. 다만 사전에 학년과 전공 등에 제한을 두어 균등한 번역 능력자를 선정하려는 노력을 기울였음에도 개인적 번역 역량에는 차이가 있을 수 있으므로 본 연구 결과를 해석하는데 있어 주의가 필요할 것으로 보인다.

## 4. 연구 결과

### 4.1 생산성

MTPE의 생산성을 이해하기 위해서는 다양한 방식이 사용될 수 있다. MTPE

의 과정을 면밀하게 분석한 것으로 평가받는 크링스(Krings, 2001)는 MTPE에 필요한 노력을 세 가지 측면에서 분석하고 있다. 1) 기술적 노력(technical effort)은 포스트에디팅의 결과물을 생성하기 위해 실제로 필요한 수정의 빈도 및 양을 의미한다. 2) 시간적 노력(temporal effort)은 작업을 완료하는데 소요되는 시간을 의미한다. 3) 인지적 노력(cognitive effort)은 오류는 식별하고 필요한 수정을 인지하는데 필요한 노력을 의미한다.

하지만 한 연구에서 3가지 노력 모두를 살펴보는 것이 꼭 필요한 것은 아니며, 인지적 노력을 신뢰성 있게 측정하게 위해서는 아이트래커(eye tracker) 등 별도의 장비가 필요한 경우가 많다. 따라서 본 연구의 목적에 부합하는 시간적 노력과 기술적 노력을 중심으로 논의를 전개하고자 한다. 시간적 노력 및 기술적 노력을 정량화하기 위해 온라인에서 무료로 제공하고 있는 번역보조 도구인 메이트캣(MateCat) (<https://www.matecat.com/>)을 사용하였다. 연구 시작 전 번역보조 도구의 사용방식을 단계별로 상세한 안내서를 배포하였으며, 연구를 진행하는 과정에 있어 기기 사용과 관련된 기술적 문제를 호소한 참여자는 없었다.

#### 4.1.1 기술적 노력

크링스가 정의한 기술적 노력을 정량화하기 위한 방법은 다양하지만, 다수의 연구 및 LSP가 사용하고 있는 대표적 방법 중 하나는 TER(Snover et al. 2006)이다. TER은 ‘Translation Edit Rate’의 약어로 기계번역 결과물 수정에 필요한 에디팅의 종류 및 작업량을 식별하기 위해 고안되었으며, 삽입(insertion), 삭제(deletion), 대체(substitution), 이동(shift)으로 에디팅 방식을 분류한다. 물론 TER이 장점만 있는 것은 아니다. TER은 MTPE의 수정률을 통해 생산성을 유추해 볼 수 있는 도구이지 절대적 생산성을 나타내는 지표는 아니며, 인간번역의 경우 TER을 계산할 수 없어 MTPE와 인간번역의 TER을 비교하는 것은 불가능하다. 하지만 이러한 단점에도 불구하고 객관적 수정량을 자동화된 방식으로 손쉽게 정량화할 수 있다는 장점 때문에 업계에서는 TER을 사용하고 있다. MT엔진 분야에서 선도적 기술을 보유한 칸탄엠티(KantanMT)는 TER 값이 30% 이상이면 포스트에디팅 작업을 ‘지나치게 많이’ 한 것으로 해석하고 있으

며, 국내 LSP역시 TER을 생산성 및 MTPE 작업 적합성을 판단하는 지표로 지속적으로 사용하고 있다.

참여자들이 약식 에디팅한 결과물의 TER은 구글 3%, 파파고 5.9%, 카카오 5.4%를 기록했다. 평균은 TER은 5.06%로서 칸탄앰티가 제시한 30% 대비 매우 낮은 수치이다. 칸탄앰티가 제시한 30%라는 수치가 서구어와 서구어 언어쌍의 분석의 결과임을 감안하면, 데이터의 양이나 언어간의 거리를 감안하면 더 많은 수정이 있을 것으로 보이는 영어-한국어 언어쌍에서 5%의 TER이 나왔다는 것은 MTPE의 높은 생산성을 가정할 수 있는 근거임이 분명해 보인다. 다만 TER은 수정의 결과만을 보여주는 것이지, 수정에 들어간 인지적인 노력을 반영하지는 못한다는 점을 고려하면 TER이 생산성을 부분적으로 반영하는 지표라고 보는 것이 합리적일 것이다.

#### 4.1.2 시간적 노력

기계번역의 결과물을 수정하는데 필요한 시간적 노력은 MTPE의 생산성을 측정하는 지표로 사용되기도 하며, 기계번역 초벌 결과물의 품질을 평가하는 지표로 사용되기도 한다. 본 장에서는 시간적 노력을 생산성의 지표로 활용하여 논의를 전개코자 한다.

번역보조 도구인 MateCat에는 작업소요시간을 기록하는 기능이 있다. 15명의 작업자에게 작업 중단 시, 타이머를 멈추는 방법 등 번역보조 도구 사용법을 설명한 이후 작업완료 시 작업소요 시간을 기록해 줄 것을 요청하였다. 참여자 별로 번역과 MTPE에 얼마의 시간을 소요했는지는 다음 <표 6>에 정리하였다. 총 15명의 참여자 중 13명의 경우 MTPE에 더 적은 시간이 소요되었다. 다만 ‘참여자 2’와 ‘참여자 8’만이 MTPE에서 더 많은 시간을 소모했는데, 그 원인 및 의미는 다음 장에서 논의하도록 한다.

<표 6> 참여자별 시간적 노력 비교

파일	MTPE 단어 수	인간 번역 단어 수	MTPE 시간 (분)	인간 번역 시간(분)	MTPE/ 인간 번역 시간
참여자 1	1,377 (t1)	1,289 (t5)	45	67	0.67
<b>참여자 2</b>	1,346 (t2)	1,377 (t1)	<b>119</b>	87	1.36
참여자 3	1,435 (t3)	1,346 (t2)	11	23	0.47
참여자 4	1,285 (t4)	1,435 (t3)	43	52	0.82
참여자 5	1,289 (t5)	1,285 (t4)	25	47	0.53
참여자 6	1,377 (t1)	1,289 (t5)	60	122	0.49
참여자 7	1,346 (t2)	1,377 (t1)	35	90	0.38
<b>참여자 8</b>	1,435 (t3)	1,346 (t2)	<b>165</b>	90	1.83
참여자 9	1,285 (t4)	1,435 (t3)	30	80	0.37
참여자 10	1,289 (t5)	1,285 (t4)	59	110	0.53
참여자 11	1,377 (t1)	1,289 (t5)	35	103	0.33
참여자 12	1,346 (t2)	1,377 (t1)	75	122	0.61
참여자 13	1,435 (t3)	1,346 (t2)	68	95	0.71
참여자 14	1,285 (t4)	1,435 (t3)	45	110	0.40
참여자 15	1,289 (t5)	1,285 (t4)	140	193	0.72

각 참여자가 작업한 텍스트의 길이가 다소 차이가 있기는 하지만, 생산성 비교를 위해 인간번역 소요시간을 MTPE 소요시간으로 나누어 보았다. 번역과 MTPE의 소요 시간이 같다면 1이, MTPE 소요시간이 인간번역 대비 적다면 1 이하의 값이 나올 것이다. 본 계산식을 <표 6>의 15명 참여자에게 적용한 결과 0.68의 평균을 기록했는데, 이는 MTPE가 번역보다 소요시간이 ‘적음’을 시사한다. 하지만 비교 대상이 되는 각각의 참여자가 작업한 텍스트의 길이가 완전히 동일하지는 않다는 문제가 제기될 수 있다. 따라서 텍스트 단어수를 작업소요시간으로 나눠서 1분당 처리단어 수를 살펴보고, 위와 유사한 방식으로 MTPE의 분당 처리 단어 수를 인간번역의 분당 처리 단어 수로 나눠보았다.

<표 7> 참여자별 1분당 번역과 MTPE 처리 단어 수

참여자	1	<b>2</b>	3	4	5	6	7	<b>8</b>	9	10	11	12	13	14	15
MTPE	30	<b>11</b>	130	29	51	22	38	<b>8</b>	42	21	39	17	21	28	9
인간번역	19	<b>15</b>	58	27	27	10	15	<b>14</b>	17	11	12	11	14	13	6
MTPE/ 인간번역	1.59	<b>0.71</b>	2.22	1.08	1.88	2.17	2.51	<b>0.58</b>	2.38	1.87	3.14	1.59	1.48	2.18	1.38

<표 7>의 결과도 <표 6>에서 관찰된 시간 생산성 결과와 유사하다. 총 15명의 참여자 중 ‘참여자 2’와 ‘참여자 8’을 제외한 모두가 인간번역 대비 MTPE의 ‘분당 처리 단어 수’가 높다는 것을 볼 수 있다. 생산성 비교를 위해 MTPE의 분당 처리 단어 수를 인간 번역의 분당 처리 단어로 나눈 평균은 1.78을 기록했다. 달리 말하면 인간 번역이 분당 10단어를 처리했다면 MTPE는 17.8단어를 처리하여 인간 번역 대비 78% 높은 생산성을 보였다는 의미이다. 더 나아가 ‘참여자 2’와 ‘참여자 8’을 제외한 13명의 생산성은 인간번역 대비 85%까지 높게 나왔다. 따라서 2명의 참여자를 제외한 13명의 참여자는 MTPE 수행 시 인간번역보다 훨씬 많은 단어를 처리한다는 것을 알 수 있다.

다만 전문적인 기술적 및 시간적 노력만을 보고 MTPE의 생산성 우위를 속단하기에는 어려움이 있을 것으로 보인다. 본 연구는 약식 포스트에디팅으로 진행되었기 때문이다. ISO에서 제시한 약식 포스트에디팅은 MTPE결과의 최소 수정을 통해 의미전달 및 문장의 구조에 문제가 없는 적절한(good enough) 수준의 번역 생산을 목적으로 하기 때문에 위의 결과는 약식 포스트에디팅에 국한된 결과임을 유의해야 할 것이다.

#### 4.1.3 생산성 예외 경우

이제 <표 6>과 <표 7>에서 MTPE가 인간번역보다 느렸던 ‘참여자 2’와 ‘참여자 8’에 대해 그 이유를 분석해 보도록 하겠다. 작업시간 지연의 원인을 파악하기 위해 이들에게 우선으로 다음과 같은 질문을 하였다.

1. 메이트 킷과 같은 도구 사용의 기술적 어려움이 있었는가?

2. 원천 텍스트 이해에 있어 어려움이 있었는가?
3. 기계번역의 오류가 지나치게 많았는가?
4. 작업가이드라인을 잘 이해하지 못했는가?
5. MTPE 작업 방식에 대한 낯설음이 문제인가?

단, 위의 질문 전에 이들의 작업시간 지연이 개인의 번역 속도와 관련된 것이 아닌가를 확인해야 했다. 이를 위해 <표 8>과 같이 두 참여자의 번역 시간을 같은 텍스트를 번역한 다른 참여자와 각각 비교해 보았다.

<표 8> 동일 텍스트를 처리한 참여자 간의 작업 시간 비교

매뉴얼 2	MTPE:인간 번역 (분)	매뉴얼 3	MTPE:인간 번역 (분)	비고
참여자 2	<b>119/87</b>	참여자 3	11/23	구글
참여자 7	35/90	참여자 8	<b>165/90</b>	파파고
참여자 12	75/122	참여자 13	68/95	카카오

‘참여자 2’의 번역 시간은 87분으로 같은 매뉴얼2를 번역을 한 ‘참여자 7’의 90분, ‘참여자 12’의 122분 보다 빨랐다. ‘참여자 8’의 경우 번역 시간은 90분으로 같은 매뉴얼3 번역을 진행한 ‘참여자 3’의 23분 보다는 느렸으나 ‘참여자 13’의 95분 보다 빨랐다. 반면 MTPE 시간을 보면, ‘참여자2’나 ‘참여자8’보다 번역 속도가 느렸던 ‘참여자 12’와 ‘참여자 13’는 번역 대비 MTPE 속도가 빨랐다. 따라서 참여자 2와 8의 MTPE 작업시간 지연은 개인의 번역속도와는 무관하다고 볼 수 있다. 그렇다면 무엇이 작업시간 지연에 가장 큰 요소였을까?

1번 ‘도구 사용의 어려움’은 MTPE 시간 지연과 관련이 없는 것으로 나타났다. 도구 사용의 어려움은 처음 설정에서 시간이 걸릴 뿐 작업 자체에는 영향을 주지 않는다는 답변을 했기 때문이다. 2번 ‘원천 텍스트의 이해’에 대해서는, 원천 텍스트의 이해는 크게 어렵지 않았으며, 작업에 영향을 주지 않았다는 답을 얻을 수 있었다. 이는 원천 텍스트의 난이도(complexity)가 MTPE작업에 주는 영향도가 미미하다는 칼 외(Carl et al. 2015)의 주장과 그 맥락을 같이하

- 3) 첨부 텍스트 참조

는 답변이다. 반면 3번의 ‘기계번역의 많은 오류’에 대해서는 관련이 있다는 답변을 얻었다. 이것은 실제 오류 보다는 MTPE 가이드라인에 대한 개인적인 이해 차이 때문인 것으로 나타났다. 구체적으로 보면, ‘참여자 2’는 총 92개의 오류를 식별했으나 같은 구글번역 그룹에서 MTPE 작업을 한 다른 참여자가 식별한 오류 평균은 26.5개이다. ‘참여자 8’은 총 128개의 오류를 식별했으나, 같은 파파고 그룹에서 유사한 텍스트를 MTPE한 다른 4명이 식별한 오류는 평균 64개였다. ‘참여자 2’와 ‘참여자 8’은 같은 MT엔진을 사용한 그룹에서 유사한 텍스트를 작업한 다른 참여자 보다 많은 오류를 식별한 것이다. 그렇다면 동일한 원문을 MTPE한 다른 참여자들과의 오류 식별 비교는 어떤 결과를 보일까?

〈표 9〉 동일 텍스트를 처리한 참여자간 오류 식별 빈도

매뉴얼2	<b>참여자 2</b>	참여자 7	참여자 12
	<b>92 (구글)</b>	55(파파고)	75(카카오)
매뉴얼 3	참여자 3	<b>참여자 8</b>	참여자 13
	5(구글)	<b>128(파파고)</b>	68(카카오)

〈표 9〉에서 볼 수 있듯이 ‘참여자 2’와 ‘참여자 8’은 동일한 텍스트를 MTPE 작업한 다른 참여자 보다 MT 오류를 훨씬 더 많이 식별했다. 세 가지 다른 번역엔진의 결과물을 비교하고 있는 것을 문제 삼을 수도 있겠으나, 이는 4.2에서 다루게 될 수정 빈도를 보면 해결된다. 구글 번역은 3개의 엔진 중 가장 낮은 수정 빈도를 나타내는데도 ‘참여자 2’의 오류 식별량만 지나치게 높은 점, 파파고와 카카오의 경우 수정 빈도에서 유의미한 차이를 보이지 않는데도 ‘참여자 8’이 ‘참여자 13’보다 훨씬 많은 오류를 식별한 점을 감안하면, ‘참여자 2’와 ‘참여자 8’은 다른 참여자들보다 유독 많은 부분을 수정의 대상으로 삼았다고 할 수 있다. 결론적으로 이들은 동일 엔진으로 작업한 유사 텍스트와 비교해도, 동일 텍스트를 다른 엔진으로 작업한 결과와 비교해도 더 많은 오류를 식별하는 성향을 보였다. 이토록 높은 오류 수정 빈도는 작업시간 지연의 직접적 원인으로 보인다.4)

마지막 질문한 4번과 5번의 ‘가이드라인 이해’와 ‘MTPE 작업 방식’에 대해서는 응답자들은 가이드라인의 내용을 이해하기는 쉬웠지만, 실제 작업에 적용하는 부분에서는 어려움이 많이 있었다고 답했다. 예를 들어 “올바른 의미 전달”이라는 문구는 이해할 수 있었으나, 올바른 의미 전달을 텍스트의 수정에 적용하는 것은 어려운 일 이라는 답변이 돌아왔다. 결국 4번, 5번에서 나타난 실험의 구체적 가이드라인 부족으로 인해 3번과 같이 ‘지나치게 많은 오류 식별’이란 결과가 도출된 것이다. ‘무엇을 어떻게 어느 정도’ 수정하는지 명확히 판단하기 어려운 추상적인 가이드라인으로는 MTPE작업을 의뢰인의 기대에 맞춰 수행하기 쉽지 않다(Annaberri 2014; 이상빈 2018)는 것을 확인 할 수 있는 결과이다.

#### 4.2 수정의 유형과 엔진간 비교

앞에서 살펴본 바와 마찬가지로 무엇을 오류로 인식하는지 역시 중요한 문제이지만, 기계번역 결과물의 오류 정도와 유형 역시 중요한 문제이다(Gueberof 2013). 따라서 본 연구에서는 두 번째 연구 과제로 MTPE 작업에 가장 손쉽게 적용할 수 있는 범용엔진 구글, 파파고, 카카오의 상대적 오류 발생률을 평가하고자 했다.

기계번역 평가는 자동화된 기계기반 평가보다 인간이 정확한 기준에 근거하여 평가하기를 권고하고 있다(Graham et al., 2015). 인간의 기계번역 비교 방식에는 유창성(fluency)·충분성(adequacy), 랭크오더링(Rank ordering), 사용가능성(usability), 포스트에디팅 시간, 주관적 오류 빈도 등이 있다(Sara Stymne, 2017). <신경망MT 분과>는 이 중 참여자의 주관성을 배제하고 정량적 평가를 위해 다음 <표 10>과 같은 KIGO의 내부 MTPE기준을 참여자에게 제시하고 기계번역 오류를 집계하고 수정하도록 설계하였다.

4) 본 연구의 전체 수정빈도와 작업시간을 통계적으로 분석한 결과 두 변수가 매우 상관관계가 높은(피어슨 상관 계수 0.9)것으로 나타났다.

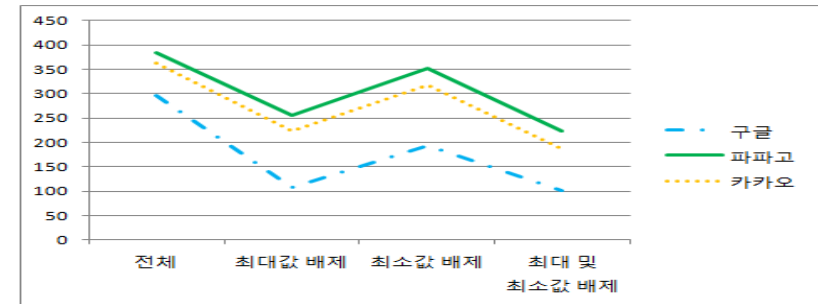
〈표 10〉 오류 평가를 위한 가이드라인

카테고리	정식 포스트에디팅	약식 포스트에디팅
오역	원문의 의미와 유창성을 모두 살린다.	이해가능하고 사용가능한 수준으로 원문의 의미를 전달한다.
정확성	의미를 바르고 유창하게 전달하기 위해 애매한 문장을 수정하고 단어의 순서를 바꾼다. 누락된 부분을 추가하며 중복되는 부분이나 단어는 제거한다.	의미를 바르게 전달하기 위해 애매한 문장을 수정하고 단어 순서를 바꾼다. 누락된 부분을 추가하며 중복되는 부분이나 단어는 제거한다.
용어	용어를 수정하며 단어집 혹은 참조문헌과의 일관성을 유지한다. 신규용어, 브랜드, 제품, 매체명 확인	용어를 수정한다. (단어집 검증 없음) 브랜드, 제품, 매체명 확인
형식 (포맷)	모든 태그를 수정하며 필요할 경우 형식을 수정한다.	의미에 영향을 주는 태그를 수정한다.
철자	철자와 구두법을 수정한다.	철자를 수정한다.
문법	문법을 수정한다.	N/A
스타일	해당 영역 (예시: 법무) 혹은 콘텐츠 유형에 (예시: 마케팅) 맞춰 스타일을 수정 및 변경한다. 고객사의 고유 가이드를 적용한다.	N/A

(\*연구자의 번역)

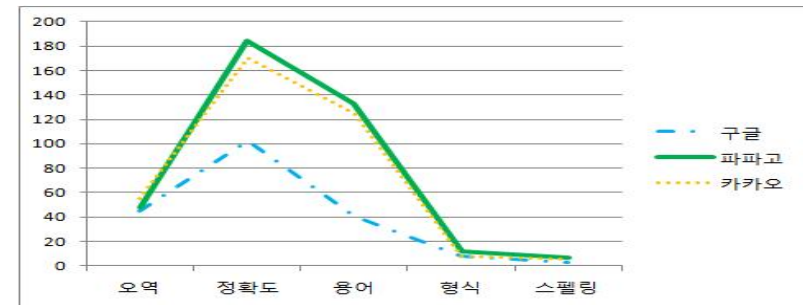
참여자가 수정한 총 오류는 구글번역 199개, 파파고 384개, 카카오 363개로 구글이 가장 적었다. 하지만 이는 4.1.3에서 오류를 과도하게 식별한 것으로 밝혀진 참여자2와 8의 데이터가 포함된 것이다. 따라서, 객관성을 확보하기 위해 가장 많은 오류를 식별한 참여자, 가장 적게 오류를 식별한 참여자 그리고 상기 두 참여자를 모두 배제하고 엔진 간 총 오류 순위가 바뀌는지를 검증하였다. 그림 1의 결과에서 볼 수 있듯, 본 연구에 사용한 텍스트에서는 구글번역이 가장 적은 오류를 지니는 것으로 나타났다.

그림 1 번역엔진 오류 총량 비교 시나리오



그러나 구글번역이 오류가 적다는 위의 분석은 오역, 정확도, 용어, 형식, 스펠링 등 전체 오류를 집계한 결과이기 때문에 그림 2와 같이 본 연구에서 사용된 5가지 영역별 오류를 각 엔진별로 살펴보았으며 그 결과는 아래와 같다.

그림 2 번역엔진 오류 영역별 비교



모든 엔진에서 가장 많은 오류가 나온 영역은 정확도였으며, 그 다음으로 용어와 오역이 뒤를 이었다. 또한 오역, 정확도, 용어, 형식, 스펠링 모든 영역에서 구글번역이 전반적인 우위를 보인다는 점에서 오류 총량 순위와 같은 결과가 도출되었다. 이는 구글의 광범위한 데이터와 선도적인 알고리즘 때문에 구글번역이 가장 우수한 품질을 보일 것이라는 산업계의 생각을 객관적 데이터를 통해 처음으로 증명했다는 점에서 의미가 있다. 그러나 이론적 배경 부분에서도 밝혔듯이 MT 생산성 관련 연구 결과는 연구한 분야에 국한되어 해석해야

하기 때문에, 이 경우도 ‘영한’, ‘IT 매뉴얼’이라는 언어쌍, 텍스트 종류와 분야에서 구글번역이 앞선다는 것으로 협의의 해석이 필요해 보인다. 언어쌍, 텍스트 종류 등 다른 조건에서 연구한 경우 파파고의 우위를 보이는 연구들(박영란(2018), 이성화, 김세현(2018), 이주리에(2018))이 있기 때문이다. 또한 15명의 참여자 간에 번역 능력에 있어 개인차가 있다는 점과 모든 참여자가 학생이라는 점을 감안하면 위의 결과가 엔진성능을 대변한다고 주장하기는 무리가 있다.

## 5. 논의 및 결론

최근 번역시장에서 MTPE에 대한 논의가 점차 가열되고 있다. 하지만 앞의 2.3 국내 연구 동향이 보여주듯 통번역학계에서 한국어의 언어적 특질을 반영한 MTPE 생산성 연구는 지금까지 한편도 이루어지지 않았다. 또한 업계에서도 MTPE 후의 번역료 산정을 위한 과학적이고 체계적인 생산성 측정치가 부족하다는 인식하에 본 산학협력은 시작되었다. 영한 매뉴얼 텍스트를 대상으로 여타 국내 MTPE 연구보다 상대적으로 많은 15명의 참여자를 모집하여 MTPE의 생산성과 MT 엔진 성능 및 오류차이에 집중하여 논의를 전개했다.

본 연구의 첫 번째 성과는 MTPE의 시간적 노력과 기술적 노력을 분석하여 ‘IT 매뉴얼’이라는 분야에 있어 ‘약식 포스트에디팅’의 생산성 향상을 밝혀냈다는 점이다. 구체적으로는, MTPE 수행시 낮은 수준의 기술적 노력(TER=6%)이 들어감을 보여줌과 동시에, 인간번역 대비 최소 30% 이상의 ‘시간절감’을 한다는 결과를 입증하였다. 또한 같은 시간 동안 단어 처리 속도에 있어서도 MTPE는 인간번역 대비 최소 78% 이상 높은 수치를 보여 ‘생산성 증가’ 효과도 있는 것을 입증하였다. 이는 매우 의미 있는 결과이다. 그러나 MTPE의 완전한 생산성 우위를 입증하기 위해서는 약식 뿐 아니라 정식 포스트에디팅도 연구를 하여 인간 번역보다 생산성이 얼마나 높은지를 알아보아야 한다. 이미 언급한 대로 고객사가 기계번역에 대해 약식 수준 번역료를 지불하면서, 실제로는 최고의 품질을 기대하는 점은 MTPE 업무의 난제 중 하나이다. 추후 연구에서 정식 포스트에디팅 생산성 데이터를 확보한다면 이를 약식 포스트에디팅과 비교분석을 진행하여 LSP의 과금정책에 유용한 근거로 활용할 수 있을 것이다. 정식 포

스트에디팅에 들어가는 시간과 노력이 약식보다 얼마나 많은지를 명확하게 설명할 수 있다면 이를 정식 작업 시 가격 향상의 근거로 제시하여 산업계에 의미 있는 기여를 할 것으로 예상된다.

본 연구의 두 번째 의의는 번역보다 MTPE에 더 많은 시간을 소요한 두 참여자들의 사례를 분석하여 그 이유를 밝혔다는 점이다. 그 결과 원문의 난이도 혹은 번역엔진의 성능과는 무관하게 두 참여자들이 MTPE작업에 대한 접근방식과 이해도 차이로 인해 오류를 지나치게 많이 식별하고 수정하게 되었으며, 이로 인해 작업 시간이 지연된 것을 발견할 수 있었다. 이는 본 연구가 제시한 것과 같은 추상적인 작업가이드라인과 약식 교육 정도로는 구체적 MTPE 작업을 성공적으로 수행하기에는 한계가 있음을 여실히 보여준다. 이 문제를 극복하기 위해서는 ‘적정 수준의 수정’이 무엇인지를 사례로 밝히고 체득시키는 구체적인 MTPE 특화교육이 행해져야(O'Brien 2002; 이상빈 2017) 할 것으로 보인다. 하지만 국내 20개 통번역 프로그램을 대상으로 한국번역학회 산학협력위원회가 조사한 바에 따르면 2018년 기준, MTPE 교과목을 운영하는 국내의 통번역 프로그램은 4개에 불과하며, 그 역사도 채 3년이 되지 않았다. 이는 LSP에서 MTPE를 수행하는 인원의 대다수가 MTPE 정규 교육을 단 한 번도 받지 못했다는 것을 의미한다. 이러한 현실을 감안할 때 구체적 MTPE 가이드라인 제공을 위한 연구(Flanagan & Christense 2014; Hu & Cadwell 2016 등)가 영어-한국어 언어쌍에 대해 필요하다 하겠다. 이와 관련, KIGO의 <스타일가이드 분과>에서는 2018년 11월 MTPE에 대한 가이드라인을 출시했으며, 윤미선 외(2018)는 대학원생을 대상으로 사례 중심의 영-한 번역에 대한 가이드라인 구체화 연구를 진행한 바 있다. 이와 같이 시의적절한 노력에도 불구하고 번역사들이 이해하기 쉽고 구체적인 예문을 담은 교육에 최적화된 MTPE 가이드를 만드는 일은 지속적인 개선과 연구가 필요한 작업이라 하겠다.

본 연구의 세 번째 의의는 참여자들의 오류 식별 및 수정 빈도를 활용하여 본 연구의 대상이 된 ‘영한 매뉴얼 번역’에 있어서는 구글번역이 파파고나 카카오 대비 우위를 보인다는 것을 관찰할 수 있었다는 점이다. 이는 2.3 선행연구 부분의 박영란(2018), 이성화 김세현(2018), 이주리에(2018) 등 오류나 정확성에서 네이버의 우위를 보이는 연구와 매우 다른 결과였다. 이를 볼 때 MT성능에 있어서는 텍스트 타입, 주제(IT, 생명과학, 법률, 경제 등), 한국 문화 관련성(속

담, 전통 등) 등 다양한 층위로 구분을 하여 비교 분석이 필요함을 알 수 있다.

본 연구는 협력과 절차적인 면에서도 의의가 있었다. 총5회의 산학협력 모임은 갈등 없이 우호적인 분위기에서 진행되었다. 통번역의 미래에 대한 많은 연구(김련희(2011), 김순미(2018), 박지영(2017), 신지선(2017), 이노신 외(2016), 이상빈(2016), 전현주(2017))에서 볼 수 있듯이 연구자들과 학생들은 변화하는 시대를 배우고 교육체계를 바꾸기 위해서 업계로부터 배워야 한다는 의식이 강했고, 그 동안 숙명여자대학교의 로컬리제이션 대학원 설립, 우수한 통번역 학회들의 ‘인공지능, 기계번역’ 중심 학술제 개최, ‘2017년 숙명여대 로컬리제이션 여름학교’, 번역학회 등 다양한 학회들의 CAT과 MT, 인공지능 관련 웨비나와 워크숍 개최 등 노력(전현주 2017)을 기울이고 있었기 때문에 가능한 일이었다. 또한 KIGO와 한국번역학회는 이미 2017년부터 <신경망 번역기 분과위원회>를 통해 게임, 기술번역, 비기술번역, 마케팅, IT 소프트웨어 등 5개 분야 텍스트에 대한 NMT 성능 연구를 행한 바가 있었기 때문에 2018년의 절차 운영은 더 원활했다. 2017년의 작은 샘플 수와 참여자 수 등을 보완했다는 점은 올해 연구의 성과라 할 수 있다. 또한 연구 결과를 <2018 번역학회 가을학술대회>와 12월 <KIGO 컨퍼런스>에서 각각 교차 발표한 것도 성과였다. 향후 연구에서는 이번의 문제에 기반해 더욱 개선된 연구 모델을 세울 수 있으리라는 기대를 한다.

마지막으로 본 연구의 한계를 평가하고자 한다. 무엇보다 LSP 현업에서 사용하는 특화된 엔진이 아닌 구글이나 네이버 같은 범용엔진을 실험에 사용한 한계가 있었다. 연구 방법론 면에서도 TAUS(2016)에서 제시하고 있는 것과 같이 구체적인 가이드라인을 미리 설정하고 준수하지 못했다. 즉, 연구자들은 CAT 사용, 번역사의 평균 하루 작업량 반영 등 현실의 LSP 작업 환경과 더 유사하며 매우 정교한 조건들을 미리 제시하고 확인하지 못했다. 결국, 참여자들에게 제공한 MTPE 가이드라인과 사전 교육은 충분히 명확하고 구체적이지 못했다. 연구자들 입장에서는 더 구체적인 실험 설계를 미리 해주지 못하고 실험 사후에 이에 대한 분석과 논의에 그쳤다는 것이 아쉬운 점이다. 업체들은 업무에 쫓기는 환경에서 시간을 내서 모임에 참석하고 운영을 하는 입장이라 소수의 열정에 의지하게 된 단점도 드러났다. 그러나 이 한계는 향후 지속될 산학협력에서 해결해야 할 연구 과제를 남겨주었다. 절차적 구체성 부족을 보완하기 위해 향후 산학협력 활동의 일환으로 KIGO MTPE 스타일가이드를 수업 시간

에 활용하여 나온 결과를 분석하여 이를 중심으로 구체적인 예문을 추가하고 학생들이 이해할 수준으로 어휘 및 설명을 개선하는 작업을 시행할 것을 제안한다. 정식 포스트에디팅과 약식 포스트에디팅의 시간과 노력 차이에 근거한 번역료 산출, 언어 전문가와 분야 전문가의 MTPE 속도와 내용 비교, 그리고 더욱 정밀한 엔진 간 성능평가 등도 향후 산학협력 연구가 필요한 분야이다. 본 연구를 기점으로 새로운 산학협력 과제 수행을 통해 4차 산업혁명 시대 과학기술과 번역의 다양한 문제를 다룰 수 있기를 기대한다.

## 참고문헌

- 곽중철, 한승희 (2018) 「포스트에디팅 측정지표를 통한 기계번역 오류 유형화 연구」, 『통번역학연구』 22(1): 1-25.
- 김련희 (2011) 「번역과 언어서비스제공자」, 『인문과학연구』 18: 1-25.
- 김순미 (2017) 「신경망번역기(NMT) 활용 학부 번역교육의 가능성 연구」, 『통번역교육연구』 15: 5-37.
- 김순미 (2018) 「AI시대 인간번역과 기계(NMT)번역의 공존 - 경영학 ‘확장(Augmentation)전략’ 중심」, 『통역과 번역』 20(2): 1-32.
- 마승혜 (2018a) 「한영 기계번역 포스트 에디팅에 대한 경험적 고찰: 학부 교육 과정 및 결과를 중심으로」, 『통번역학연구』 22(1): 53-87.
- 마승혜 (2018b) 「기계번역의 외연 확대 - 신문기사 특정 은유표현 포스트 에디팅 전략 모색」, 『번역학연구』 19(2): 117-145.
- 마승혜 (2018c) 「문학작품 기계번역의 한계에 대한 상세 고찰」, 『통번역학연구』 22(3): 65-88.
- 박영란 (2017) 「영한번역에서의 시제의 등가성과 기계번역과의 비교」, 『통번역교육연구』 15: 165-194.
- 박영란 (2018) 「기계 번역에서의 영어속담 분석」, 『통번역교육연구』 16(1): 131-150.
- 박지영 (2017) 「언어기술 혁신과 통번역 산업의 미래: 20대 글로벌 LSP의 통번역 서비스분석을 바탕으로」, 『통번역학연구』 21(1): 137-68.

- 박혜경 (2018) 「석사 과정의 기계번역 수업에 대한 소고: 한일번역 전공생의 포스트에디팅 사례를 통하여」, 『번역학연구』 19(3): 163-193.
- 송연석 (2018) 「기계번역 담론에 대한 비판적 고찰」, 『번역학연구』 19(1): 119-145.
- 서보현, 김순영 (2018) 「기계번역 결과물의 오류유형 고찰」, 『번역학연구』 19(1): 99-117.
- 신지선 (2017) 「테크놀로지 패러다임에서의 번역능력 재조명」, 『통번역학연구』 21(4): 51-71.
- 신지선, 김은미 (2017) 「인공지능 번역 시스템의 출현에 대한 소고」, 『번역학연구』 18(5): 91-110.
- 윤미선, 김택민, 임진주, 홍승연 (2018) 「영어-한국어 언어쌍에 적합한 포스트에디팅 가이드라인」, 『번역학연구』 19(5): 43-76.
- 이노신, 이신재, 이재영, 이주희 (2016) 「통번역의 미래지평 : 인공지능과 소통형 융합통번역 연구」, 『번역학연구』 17(2): 65-89.
- 이상빈 (2016) 「트랜스크리에이션, 기계번역, 번역교육의 미래」, 『통역과 번역』 18(2): 129-152.
- 이상빈 (2017) 「학부번역전공자의 기계번역 포스트에디팅, 무엇이 문제이고, 무엇을 가르쳐야 하는가?」, 『통역과 번역』 19(3): 37-64.
- 이상빈 (2018) 「학부 번역전공자의 구글 기계번역 포스트에디팅에 관한 현상학 연구」, 『통번역학연구』 22(1): 117-143.
- 이성화, 김세현 (2018) 「영-한 및 한-영 기계번역 품질향상을 위한 프리에디팅 기법 제안」, 『번역학연구』 19(5): 121-154.
- 이승일 (2018) 「AI번역의 속성 및 번역주체에 대한 논의」, 『통번역학연구』 22(4): 184-209.
- 이주리아 (2018) 「인공신경망 기계 번역의 한일/일한 번역 품질에 대한 예비연구 -품질향상 검토와 교열코드 적용-」, 『통역과 번역』 20(1): 32-71.
- 이준호 (2018) 「포스트에디팅 교육을 위한 포스트에디팅과 인간번역의 차이 연구」, 『통역과 번역』 20(1): 73-96.
- 이준호 (2019) 「문학번역 적용을 위한 기계번역의 현주소」, 『통번역학연구』 23(1): 143-68.
- 장애리 (2017) 「국내 기계 통번역의 발전 현황 분석」, 『번역학연구』 18(2):

- 171-206.
- 전현주 (2017) 「4차 산업혁명과 한국의 번역산업 현황 및 통번역 교육의 미래」, 『통번역교육연구』 15(3): 235-261.
- 정혜연 (2018) 「번역의 자동평가: 기계번역 평가를 인간번역 평가에 적용해보기」, 『통번역학연구』 22(4): 265-87.
- 진용주, 서보현, 김순영 (2018) 「프리에디팅 (pre-editing) 이 기계번역 품질에 미치는 영향 고찰: 유창성, 충실성, 가독성 비교를 중심으로」, 『통번역학연구』 22(3): 27-64.
- 최효은, 이지은 (2017) 「특허 기계번역 결과물의 평가 -KIPRIS의 무료 한영 기계번역을 중심으로」, 『통역과 번역』 19(1): 139-78.
- 한승희 (2017) 「기계번역 인간번역 트랜스크리에이션의 문체 비교: 광고 번역을 중심으로」, 『통번역학연구』 21(2): 163-188.
- 한현희 (2018) 「한노 기계번역의 문장 부호 처리 양상 및 오류 개선을 위한 실용적 함의: 프리에디팅 규칙 수립과 효용성의 관점에서」, 『통번역학연구』 22(3): 227-260.
- Aranberri, Nora, et al. (2014) 'Comparison of post-editing productivity between professional translators and layusers', in *Third Workshop on Post-editing Technology and Practice* 20-33.
- Carl, Michael, et al. (2011) 'The process of post-editing: a pilot study', in *Proceedings of the 8th International NLPSC Workshop. Copenhagen Studies in Language* 41: 131 - 42.
- Carl, Michael, Silke Gutermuth, and Silvia Hansen-Schirra (2015) 'Post-editing machine translation', in Aline Ferreira, John W. Schwieter (Eds.), *Psycholinguistic and cognitive inquiries into translation and interpreting*, Amsterdam & New York: John Benjamins Publishing Company, 145-74.
- Cronin, Michael (2003) *Translation and Globalization*, NewYork & London: Routledge.
- DePalma, Donald A., et al. (2013) *The Languages Services Market: 2013*, Lowell, MA: Common Sense Advisory.
- DePalma, Donald A., et al. (2016) *The Language Services Market: 2016*,

- Lowell, MA: Common Sense Advisory.
- Doherty, Stephen (2016) 'The Impact of Translation Technologies on the Process and Product of Translation', *International Journal of Communication* 10: 947-69.
- Federico, Marcello, Alessandro Cattelan, and Marco Trombetti (2012) 'Measuring user productivity in machine translation enhanced computer assisted translation', in *Proceedings of the Tenth Conference of the Association for Machine Translation in the Americas (AMTA)* Madison, WI: AMTA.
- Flanagan, Marian and Christensen, Tina Paulsen (2014) 'Testing post-editing guidelines: how translation trainees interpret them and how to tailor them for translator training purposes', *The Interpreter and Translator Trainer* 8(2): 257-275.
- Garcia, Ignacio (2010) 'Is machine translation ready yet?', *Target International Journal of Translation Studies* 22(1): 7-21.
- Guerberof, Ana (2009) 'Productivity and quality in MT post-editing', in *MT Summit XII-Workshop: Beyond Translation Memories: New Tools for Translators MT* Ottawa: Association for Machine Translation in the Americas.
- Guerberof, Ana (2013) 'What do professional translators think about post-editing', *The Journal of Specialized Translation*, 19.
- Guerberof, Ana (2014) 'Correlations between productivity and quality when post-editing in a professional context', *Machine Translation*, 28(3-4): 165-186.
- Graham, Yvette, Timothy Baldwin, and Nitika Mathur (2015) 'Accurate evaluation of segment-level machine translation metrics', *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
- Hu, Ke and Cadwell, Patrick (2016) 'A comparative study of post-editing guidelines', in *Proceedings of the 19th Annual Conference of the*

- European Association for Machine Translation, Riga: EAMT*.
- Hutchins, John (2009) 'Multiple Uses of Machine Translation and Computerised Translation Tools', *Machine Translation*: 13-20.
- Koponen, Maarit (2013) 'This translation is not too bad: An analysis of post-editor choices in a machine translation post-editing task', in *Workshop Proceeding: Workshop on Post-editing Technology and Practice*, 1-9.
- Koponen, Maarit (2016) 'Is machine translation post-editing worth the effort? A survey of research into post-editing and effort', *JoSTrans* 25: 131-148.
- Krings Hans P. (2001) *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*, Kent: The Kent State University Press.
- Läubli, Samuel, et al. (2013) 'Assessing post-editing efficiency in a realistic translation environment', in Sharon O'Brien, Michel Simard and Lucia Specia (eds.). *Proceedings of MT Summit XIV Workshop on Post-editing Technology and Practice*, 83-91.
- Martínez, Lorena Guerra (2003) 'Human Translation Versus Machine Translation and Full Post-editing of Raw Machine Translation Output', Master's dissertation, Dublin: Dublin City University.
- O'Brien, Sharon (2002) 'Teaching post-editing: a proposal for course content', in *6th EAMT Workshop Teaching Machine Translation*, 99-106.
- O'Brien, Sharon (2007) 'An empirical investigation of temporal and technical post-editing effort', *Translation and Interpreting Studies* 2(1): 83 - 136.
- O'Brien, Sharon (2011) 'Towards predicting post-editing productivity', *Machine translation* 25(3): 197.
- Plitt, Mirko and Francois 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.
- Pouliquen, Bruno (2017) 'WIPO Translate: Patent Neural Machine Translation publicly available in 10 languages', in *Proceedings of the Seventh Workshop*

- on Patent and Scientific Literature Translation Nagoya: Japan, 5-8.
- Schwartz, Lane (2014) 'Monolingual post-editing by a domain expert is highly effective for translation triage', in *Proceedings of the Third Workshop on Post-Editing Technology and Practice*, 34-44.
- Snover, Matthew and Dorr, Bonnie and Schwartz, Richard and Micciulla, Linnea and Makhoul, John (2006) 'A study of translation edit rate with targeted human annotation', in *Proceedings of the Association for Machine Translation in the Americas, Boston: AMTA*.
- Wagner, Emma (1987) 'Post-editing: Practical considerations', in *ITI Conference I: The Business of Translating and Interpreting*, 71-78.
- Zampieri, Marcos and Mihaela Vela (2014) 'Quantifying the influence of MT output in the translators' performance: a case study in technical translation', in *Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation*.
- Zaretskaya, Anna (2017) 'Machine Translation Post-Editing at TransPerfect - the 'Human' Side of the Process', *Tradum tica* 15: 116-23.

#### <인터넷 자료>

- Guerberof, Ana (2008) 'Productivity and quality in the post-editing of outputs from translation memories and machine translation', Available at <http://www.mt-archive.info/MTS-2009-Guerberof.pdf>
- Styme, Sara (2017) 'Machine Translation Evaluation', Available at <http://stp.lingfil.uu.se/~sara/kurser/MT16/slides/f2-eval.pdf>
- Slator (2018) 'Slator Neural Machine Translation Report 2018', Available at <https://slator.com/data-research/slator-neural-machine-translation-report-2018/>
- TAUS (2016) 'TAUS Post-Editing Guidelines', Available at [file:///C:/Users/ss/Downloads/TAUS%20Post-Editing%20Guidelines%20\(3\).pdf](file:///C:/Users/ss/Downloads/TAUS%20Post-Editing%20Guidelines%20(3).pdf)

#### [부록 1] 연구에 사용한 IT 매뉴얼의 일부

Manually scanning your computer

The Scan feature lets you selectively search for viruses and other threats on hard drives, floppy disks, and individual files and folders.

When Scan finds a suspect file, it automatically tries to clean the file, unless it is a Potentially Unwanted Program.

If Scan cannot clean the file, you can quarantine or delete the file.

Manually scanning for viruses and other threats

To scan your computer:

1 Right-click the McAfee icon, point to VirusScan, then click Scan.

The Scan dialog box opens (Figure 2-7).

Figure 2-7 Scan dialog box

2 Click the drive, folder, or file that you want to scan.

3 Select your Scan Options.

By default, all of the Scan Options are pre-selected to provide the most thorough scan possible (Figure 2-7):

Scan subfolders - Use this option to scan files contained in your subfolders.

Deselect this checkbox to allow checking of only the files visible when you open a folder or drive.

[Abstract]

### A University-Industry Joint Study on Machine Translation Post-Editing Productivity and MT Engine Error Rate

Kim, Soon Mi & Lee, Jun-Ho & Shin, Hosup  
(Sookmyung Women's University, HansaemEUG)

As machine translation post-editing (MTPE) has become a common practice among language service providers (LSPs), and universities have an urgent need to incorporate MTPE into their curriculum, a common interest on MTPE practices led to the forming of a university-industry joint research team between The Korean Association for Translation Studies (KATS) and The Korea IT Globalization Organization (KIGO). This collaborative study aims to explore the following issues: First, the productivity of MTPE results gauged in terms of two factors, the amount of time spent on post-editing of raw MT outputs against human translation time, and TER (Translation Edit Rate) of MTPE results; Second, error rates of the most representative free online machine translations (FOMT) for the English-Korean language pair, such as Google Translate, Naver's Papago, and Kakao i in terms of mistranslation, accuracy, terminology, formatting, and spelling. Fifteen undergraduate and graduate students majoring translation participated in the study to produce translation and MTPE results of five IT manuals provided by the participating LSPs. For 13 out of 15 participants, the temporal and technical MTPE efforts were found to be much reduced. Regarding the MT engine error rate, Google Translate was found to have a much lower error rate against Papago and Kakao i in processing IT manuals; however, this result needs further investigation with more variables.

▶ Keywords: MTPE, industry-university joint research, translation and technology, MT productivity, machine translation, language service market

▶ 주제어: MTPE, 산학협력연구, 번역과 기술, 기계번역 생산성, 기계번역, 언어서비스 제  
공업체(LSP)

김순미(제1저자, 교신저자)  
숙명여대 영문학부 조교수  
smikim@sookmyung.ac.kr

관심분야: 과학기술과 번역, 로컬리제이션 번역, 기계번역 포스트에디팅, 커뮤니  
티통역, 번역과 사회학

신호섭(공동저자)  
(주)한샘EUG Language Service Division 대리  
hosup.shin@ezuserguide.com

관심분야: 로컬리제이션, 번역산업, 번역기술

이준호(공동저자)  
숙명여대 영문학부 초빙대우교수  
cuefit@gmail.com

관심분야: 기계번역, 포스트에디팅, 번역교육, 코퍼스

논문투고일: 2019년 1월 20일

심사완료일: 2019년 3월 11일

게재확정일: 2019년 3월 19일