

다단계 기계번역을 활용한 비전문가 번역 품질 향상 방안 모색 - 학부 실습 과정 및 결과 분석을 중심으로* **

마 승 혜 · 성 승 은***
(한국외대)

1. 서론

최근 사회적·기술적 변화와 발전으로 비전문가의 번역 참여가 확대되는 상황이다. 대표적으로 신경망 기계번역(Neural Machine Translation) 시스템, 전자책(electronic books), 플랫폼의 발전과 확산으로 클라우드소싱 번역, 네티즌 번역 등 비전문가 번역¹⁾이 더욱 활발해지고 있다(김순미 2016a; Lee 2017). 특히,

* 이 연구는 2019학년도 한국외국어대학교 교내학술연구비의 지원에 의하여 이루어진 것임.

** 이 연구는 2019년 한국번역학회 창립 20주년 기념 학술대회 발표 자료를 바탕으로 보완·확대한 것임.

*** 성승은(교신저자)

1) 비전문가 번역은 참여 번역(participatory translation), 자원봉사 번역(volunteer translation), 사용자 번역(user-generated translation), 공개번역(open translation), 온라

비전문가 번역 확산에 기여한 것이 기계번역 시스템의 발전이다. 언어서비스기업(Language Service Providers, LSP)의 경우, 고객 중심 커스텀 엔진을 구축하여 번역 품질 향상을 꾀하지만, 이러한 엔진은 철저한 보안 하에 사용되기 때문에 일반인들이 사용할 수는 없다. 결국, 비전문가 개인은 구글, 파파고와 같은 범용 엔진을 사용해야 한다.

범용 엔진을 활용한 신경망 기계번역 결과물은 언어쌍, 번역 방향에 따라 품질 차이가 있다. 본고에서는 비전문가 그룹에 속하는 영어통번역학을 공부하는 학부생들이 좀 더 어려워하는 한-영 번역(이하 설문조사 결과 참조)을 할 때 범용 엔진을 활용하여 어떻게 좀 더 효율적으로 더 나은 결과물을 생성해낼 수 있을지 살펴보고, 향후 비전문가의 기계번역 활용 및 번역 참여에 도움이 될 수 있는 일 방향을 모색해보고자 한다.

2. 선행 연구 및 연구 문제

2.1 비전문가 번역 확산

최근 번역 업계와 학계 모두 큰 화두 중 하나는 ‘비전문가 번역’이다(Pérez-González & Susam-Saraeva 2012). 확산 일로에 있는 팬번역, 자원봉사 번역, 커뮤니티 번역, 온라인으로 대중을 모집하여 번역하는 크라우드소싱 번역 등은 크게 비전문가 번역으로 아우를 수 있다. 번역에서 통상적으로 ‘전문가’는 공식적으로 전문가에게 요구되는 통번역 기술을 교육받은 경우로 인식되고 있다(Pérez-González & Susam-Saraeva 2012: 150). 김순미(2016b)는 전문가에 대해, 거시적 차원에서 제도에 편입되어 있고 미시적 차원에서 자신의 번역품질을 보장할 수 있는 경우로 보고 논한다. 본고에서의 ‘비전문가’는 전문 대학원 이상의 공식 통번역교육을 받지 않는 경우로 보기로 하며, 따라서 통번역을 공부하는 학부 학생은 비전문가 범주에 포함된다.

한편 비전문가라는 용어 사용의 적절성에 대해 문제가 제기되기도 한다. 현

인 커뮤니티 번역(on-line community translation) 등으로 지칭되기도 한다(박지영 2017; DePalma & Kelly 2008; O'Hagan 2009; O'Hagan 2011).

재 비전문가로 통용되는 이들의 번역 활동이 매우 활발한 가운데, 상호 분업과 협업, 감수, 번역 툴의 사용 등으로 전문가의 역량을 갯출 소지가 충분하기 때문이다(김순미 2016b). 또한 디지털 시대에는 기존의 ‘전문가’들이 담당하던 분야 이외의 다양한 콘텐츠가 급증하면서 이들이 주류가 될 여지도 충분하다. 다만 본고의 논의상 번역학 저널 *The Translator*의 2012년 특별호에서 사용되고 학계에서 통용되는 ‘비전문가’라는 용어를 그대로 사용하기로 한다.

페레즈 곤잘레스와 수삼 사레바(2012)는 비전문가 번역 확산의 배경으로 아파두라이(Appadurai 1990: 8-9)가 제시한 민족, 금융, 이데올로기, 미디어, 기술 지형(ethnoscape, financescape, ideoscape, mediascape, technoscape) 간의 상호작용을 들어 설명한다. 민족 지형은 빈곤·정치·종교 등으로 인해 세계적으로 이주민의 이동 증가와 관련된 것으로, 이러한 현상에 따라 이주민을 수용하는 국가에서 이들을 위한 통번역 서비스를 재정적으로 지원하느냐 여부가 비전문가 통번역의 부상과 관련이 깊다. 이러한 참여 비전문가들의 활동은 특정 집단의 정치적 가치나 내러티브를 지지 및 저항하는 노력의 일환이 되기도 한다. 또한 새로운 디지털 기술의 확산으로 소비자들은 이제 미디어 콘텐츠를 관리하고 필요에 따라 전용하며 유통할 수도 있게 되었으며, 이는 경제활동과 문화, 소비와 생산, 미디어의 생산과 사용 사이의 경계를 무너뜨리는 미디어 컨버전스(media convergence)로 이어졌다.

이러한 지형의 변화는 번역 활동에 직접적인 영향을 미쳤다. 크로닌(Cronin 2010)의 주장과 같이 번역의 ‘기술적 전환’이 일어나면서 얼마 전까지도 상상하기 어려운 속도와 규모로 번역이 이루어지고, 웹 2.0의 쌍방 소통 인터넷 문화를 바탕으로 사용자가 곧 생산자가 되는 참여 문화가 확산되면서(Jenkins 2006), 다양한 번역 플랫폼, CAT(Computer-Assisted Translation) 툴, 기계번역 활용이 폭발적으로 증가하게 된 것이다. 디지털 시대의 번역은 번역 테크놀로지와 함께 번역 주체로서 비전문가가 부각되는 현상을 수반하였고, 비전문가 번역 활동은 자신이 속한 온라인 커뮤니티 번역, 팬서빙, 뉴스 번역, 종교 번역 등 매우 다양한 분야의 내용을 망라하고 있고 있다. 이들에 대한 명칭은 아래 표처럼 다양하다.

〈표 1〉 비전문가에 대한 다양한 명칭

| 활동의 초점 | 예시 |
|---|--|
| 참여 번역(Participatory translation) | 보수를 받지 않는 번역, 취미로 하는 번역 -위키피디아 번역(McDonough Dolmaya 2012) |
| 자원봉사 번역 (Volunteer translation) | -보수를 받지 않고 자원하는 번역 -자원 통번역 활동가들의 국제 네트워크 바벨(Babels)의 통번역(Boéri 2012) |
| 팬 번역(Fan translation) | -팬으로서 수행하는 번역 -팬 자막 번역(김호영·홍남희 2012) |
| 사용자 제공 번역(User-generated translation) | -사용자가 직접 번역을 제공 -사용자 제공 사전(Perrino 2009) |
| 협업 번역(Collaborative translation), 커뮤니티 번역(Community translation) | -온라인 커뮤니티에서 협력하여 수행하는 번역 -번역가와 해당 기관의 스텝이 협업하는 박물관 번역(Neather, 2012) |
| 클라우드소싱 번역(Crowd-sourcing translation) | -온라인으로 대중에 아웃소싱하여 번역을 의뢰 -비키(Viki), TED, 페이스북(Facebook), 위키피디아(Wikipedia) 번역 (O’Hagan 2017) |

위의 각 분류가 서로 배타적인 것을 물론 아니다. 예를 들어 위키피디아 번역은 위의 모든 분류에 해당할 수 있으며, 어떤 부분에 초점을 두느냐에 따라 다른 명칭으로도 불린다는 것이다. 이 외에도 아마추어 번역(Amateur translation), 네티즌 번역(Netizens’ translation), 공개 번역(Open translation)으로 부르기도 한다.

엄격한 바와 같이 비전문가 번역 활동은 디지털 환경의 대두가 중요한 배경이다. 엄청난 양의 정보가 쏟아지면서 우리가 텍스트를 읽고 쓰는 관행이 변하고 있으며, 이는 번역에도 영향을 미치게 되었다(Cronin 2013). 종이로 보던 책을 전자 매체로 보고 소셜 미디어의 사용이 일상화되면서 이제는 자신에게 필요한 내용을 신속히 찾아서 훑어 내려가는 경우가 많다.2) 동영상, 소프트웨어, 블로그, 게임, 오락물 등 시각 텍스트를 포함한 디지털 텍스트는 문체가 단 순화되고 문장이 짧아지는 특징이 있다(박지영 2018). 시간 제약과 텍스트 양의 급증으로 인해 혼자보다는 여럿이 신속하게 대량의 번역을 수행하는 경우가 많

2) 이는 문학텍스트 읽기에도 영향을 끼쳐 최근의 독자들은 긴 문장 보다는 일정 길이의 문장을 선호한다는 연구도 제시되었다(정희연 2018).

으며, 이 같은 속도와 규모의 번역에 대해 기존에 적용하던 번역 품질 기준을 그대로 사용하기에는 무리가 있다. 즉, 번역 상황에 따라 다른 품질이 요구될 수 있어 때문에 일률적 기준을 적용할 수 없게 되었다(Cronin 2013).

디지털 환경의 대두가 비전문가 번역의 부상과 관련이 있다는 것은 곧 비전문가 번역에 참여하고자 하는 이들에게 디지털 리터러시가 필수적임을 의미한다(김순미 2017; 박지영 2017; Pym 2013). 클라우드소싱 번역, 팬번역, 온라인 커뮤니티 번역 등은 번역 플랫폼을 활용하면서 체계적인 협업 및 분업 등을 이루어서 원천 텍스트의 세그멘테이션, 용어집 관리, 번역, 감수, 편집을 효율적으로 수행한다. 교육 콘텐츠 TED와 자막 번역 스트리밍 서비스 비키(Viki) 등에서는 자체 번역 툴을 지원하고 있다(박지영 2017: 142). 영어 통번역을 공부하는 학부 학생들은 통번역 실력이 아직 전문가의 수준은 아니지만 디지털 리터러시를 갖추고 있으며, 기계번역 및 CAT 툴 등을 활용하여 비전문가 번역에 적극적으로 참여할 수 있을 것이다.

2.2 한·영 기계번역 연구 및 현황

기계번역의 품질은 언어쌍과 방향에 따라 차이가 크다(Ráido 2016: 982; Cadwell, O'Brien & Teixeira 2018). 본고에서는 한국어-일어-영어의 다단계 번역을 고찰하기 위해 한영 기계번역과 일영 기계번역의 연구 및 현황을 살펴 보면서 각각에 대한 기계번역의 품질을 비교해보고자 한다. 영어의 링구아 프랑카로서 역할로 인해 수용 언어가 영어가 되는 방향의 기계번역은 번역 교육은 물론 영어교육 및 기타 실제적인 측면에서 다양하게 활용되고 있다.

우선 L2에 대한 학생들의 자신감 결여 해소에 도움이 된다는 고찰이 있다(조인희 2018; Krings 2001). 크링즈(Krings 2001)는 기계번역이 학부 학생들의 L1과 L2 간의 간극에 대한 부담을 줄여줄 수 있다고 주장하였다. 조인희(2018) 역시 학부 영작문 수업에서 기계번역을 활용한 결과 학생들의 자기효능감이 통계적으로 유의미하게 증가하였다. 이 연구는 외국어 능력이 부족한 한국 EFL 학습자들이 기계번역을 활용함으로써 영어로 의사소통하는 기회를 획기적으로 확대할 수 있다는 측면에서 기계번역을 긍정적으로 평가하고 있다.

영어가 모국어가 아닌 학생들을 위한 영어 학술적 글쓰기 수업에 기계 번

역을 활용한 사례들도 있다. 카이스트 바이오및뇌공학과 정재승 교수는 지난 2017년 5월 27일에 이화여대에서 열린 '인공지능과 번역의 미래'라는 주제의 특강에서 자신이 지도하는 대학원 학생들에게 처음부터 영어로 논문을 쓰기보다 한국어로 쓴 후에 기계번역을 돌릴 것을 권한다고 하였다. 이렇게 하는 것이 영어 논문 작성에 더 효율적이라는 것이다. 또한 향후 기계번역 성능 향상으로 영어 글쓰기 수업에서 기계번역 활용 여지가 많아질 것임을 고찰한 연구도 있다(Groves & Mundt 2015).

이외에도 업무에서 원활한 의사소통을 위해 기계번역을 활용한 다양한 사례가 제시되고 있다. 설문지를 영어로 작성하는 데에 기계번역이 도움이 된 경우(Guo 2016), 영어로 원활한 의사소통이 힘든 화자들이 회의에서 실시간 기계번역을 활용한 결과, 대화 흐름이 끊기지 않고 토론이 가능했다는 연구(Calefato, Lanubile, Conte & Prikladnicki 2016)도 있다.

한영 기계번역은 번역 교육에서도 다각도로 검토되고 있다. 한영 기계번역 전반과 품질(최효은·이지은 2017; 마승혜 2018a), 포스트 에디팅(박옥수 2016 & 2017; 김순미 2017; 이성화·김세현 2018) 및 그 외 특정 내용에 집중한 논의도 있다. 예컨대 기계번역에서 나타나는 문체(이창수 2019), 기계번역이 수행한 문학번역의 문제점(마승혜 2018b; 이준호 2019), 텍스트의 모호성과 관련한 기계번역의 문제점(이미경 2019)에 대한 논의 등이 있다.

이들 연구를 통해 한영 기계번역은 번역 품질에서 아직 많은 문제점을 안고 있음을 알 수 있다. 최효은·이지은(2017)은 한영 특허 번역에 대한 기계번역을 평가한 결과 자동평가인 BLEU나 인간 평가 모두 낮은 점수를 받았다고 보고하였다. 한영 기계번역의 문제점 및 포스트에디팅과 프리에디팅의 필요성이 구체적으로 제시된 박옥수(2017), 김순미(2017), 이성화·김세현(2018)의 내용을 종합하면 한영 기계번역은 영어 원문에서 주어가 생략된 경우, 이해하기 어려운 의미가 포함된 경우(복합명사, 모호한 의미, 맥락의 이해, 다의어), 길고 복잡한 문장 구조가 나타나는 경우에 오역과 누락 등이 빈번히 발생한다.

이처럼 한영 기계번역은 급속도로 발전하고 있지만 여전히 많은 문제를 보이고 있다. 한영 기계번역 품질 향상을 위해서는 많은 양의 번역 데이터를 축적하여 기계번역을 학습시키는 것이 중요하다. 신경망 기계번역은 인공지능(AI)이 데이터 학습을 통해 문장 단위로 언어를 번역하는 기술로, 사용자가 번역 엔진

에 문장을 입력하면 맥락을 파악한 후 이를 단어, 구문, 어순 등의 정보가 담겨 있는 벡터(좌표값)로 전환하여 번역을 한다. 개발자가 신경망 구조를 결정해 주면 머신러닝, 딥러닝 등을 통해 AI가 스스로 학습하여 번역을 수행하므로 입력한 데이터가 많을수록 정교한 번역이 가능하다(이주리에 2018). 그러나 한영번역은 데이터가 충분히 축적되어 있지 않다. 구글 번역 관리를 총괄하는 바락 투로프스키(Barak Turovsky)는 한국어-영어 번역 쌍의 부족을 언급하면서, 다중 언어 신경망 모델 학습으로 인해 이전에 비해 성능이 개선되고 있다고 설명한 바 있다. 2016년 12월 조선비즈와의 인터뷰에서 그는 한국어-영어 번역 쌍은 일본어-영어나 터키어-영어에 비해 데이터가 훨씬 부족하지만, 한국어와 유사한 문법과 어순을 가진 일본어, 터키어와 함께 동시 학습이 되어서 일본어-영어와 터키어-영어의 패턴이 한국어-영어에도 적용되어 개선되고 있다는 것이다.

2.3 일-영 기계번역 발전 과정 및 현황

일본의 경우 과거부터 번역 데이터 축적을 위해 부단히 노력해 왔다. 일찍이 1811년 막부가 번역국(翻譯局)이라고도 하는 ‘만서화해어용(蠻書和解御用)’을 ‘천문방(天文方)’ 내에 설치하여 이곳에서 외교문서 조사 및 번역 업무를 수행하였다(이근희 2007). 이렇듯 정부 산하에 번역국을 두고 체계적인 번역작업을 꾸준히 벌인 결과, 일본은 ‘번역왕국’이라는 애칭을 얻었고, 이는 기록적인 노벨상 수상으로 이어졌다(박상익 2017 & 2018; 박성용 2019). 즉, 방대한 번역 데이터를 바탕으로 일본어 기계번역 결과물이 정교해질 수 있는 기반을 과거부터 다져 온 셈이다. 일례로 최근 영어-일본어 기계번역에 관한 선행연구(Greenstein & Penner 2015)에서는 특정 언어쌍의 번역 코퍼스 양이 성공적인 기계번역 결과물 산출에 큰 영향을 끼친다고 언급하며 교토 위키 코퍼스(Koyto Wiki Corpus) 50만 영-일 번역문장, TED Talks의 번역 자막, 학생 번 코퍼스인 타나카 코퍼스(Tanaka Corpus) 15만 영-일 번역문장을 데이터로 기계번역을 학습시킨 결과 문법적으로 다소 복잡한 문장도 성공적으로 번역된 사례를 소개한다.

반면 한국은 주변국가와의 교섭에 필요한 외국어 학습 기구이자, 역관의 양성기관인 사역원을 설치해 활발한 활동을 했으나, 점차 그 역할이 축소되었고, 개화기에는 통리기무아문 내에 어학사를 설치해 번역전담기구의 역할을 하도록

했으나 실질적인 역할은 없었다(이근희 2007). 기계번역 발전을 위한 데이터 축적은 중요성은 아시아 언어 프로세싱에 관한 국제 컨퍼런스(International Conference on Asian Language Processing)에서도 언급되었다. 컨퍼런스에서 제기된 아시아 지역 언어기반 신경망 기계번역 관련 문제점 중 하나가 기계번역을 훈련할 데이터가 부족하다는 것 이었다.3)

따라서 결국 축적된 번역 데이터양이 많을수록 기계번역 학습과 발전에 도움이 되고, 과거부터 번역 데이터를 축적해온 일-영 기계번역이 한-영 기계번역 결과물에 비해 품질이 더 나올 수 있다는 가설이 성립된다.

그렇다면 실제로 비전문가가 한영번역을 수행할 때 한영 기계번역을 활용하는 것보다 한일영 기계번역을 활용하는 것이 좀 더 정확도를 높이는 방안이 될 수 있을지 실제 비전문가 번역에 적용하여 살펴보겠다.

3. 학부 실습 적용 과정 및 결과

본 장에서는 비전문가에 해당하는 영어통번역을 공부하는 학부생을 대상으로 기계번역을 활용한 한-영 번역을 실습해보고, 좀 더 효율적으로 더 나은 번역 결과물을 생성해내는 방안을 모색해보고자 한다.

3.1 실습 참여자

실습에 참여한 학습자들은 총51명의 영어통번역을 공부하는 학부생으로 1학년(35명)이 대부분이지만 2학년(2명), 3학년(8명), 4학년(6명)도 있다. 한영번

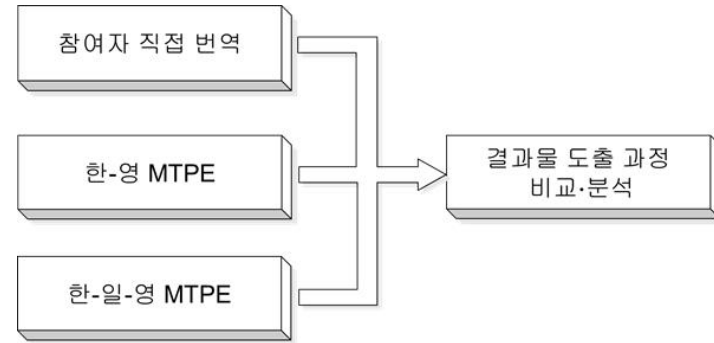
3) 2018년 인도네시아 반둥에서 개최된 컨퍼런스에서 “Machine Translation & Language Resources” 세션을 주재한 Wuying Liu는 다음과 같이 말한다. “사실, 동남아시아 언어에 대한 데이터는 상당히 부족하다. 필리핀어를 예를 들어보면, 문장별 이중언어 병렬 코퍼스를 찾기 힘들다. 따라서 이중언어 병렬 코퍼스 구축이 매우 중요하고 이를 위해 노력 중이다(To be honest, the data for SEA languages is fairly scarce. Take Filipino as an example. We can hardly find any bilingual sentence-aligned corpus. So it is quite meaningful to build corpus for languages like that and we are still working hard on this)”(Diño 2019).

역 실습을 선택한 이유는 영한번역과 한영번역 중 어떤 것이 더 어려운지 물어 보는 설문조사⁴⁾에서 참여자 약 84%에 해당하는 43명의 한영번역이 더 어렵다고 응답했기 때문이다.⁵⁾ 번역 시 기계번역을 활용해본 적이 있는지에 관한 문항에서는 41명(약 80%)이 기계번역을 종종 활용한다고 응답하였고, 나머지 10명(약 20%)은 기계번역을 활용하지 않는다고 응답하였다. 기계번역 활용 응답자 모두 구글과 네이버 파파고를 사용한다고 답하였고, 이들 중 29명(약 67%)은 한영번역 시 좀 더 기계번역을 많이 활용한다고 한다. 이렇듯 사전조사 결과 실습에 참여한 학부생 상당수는 이미 기계번역을 활용하여 번역을 하고 있고, 영한번역과 한영번역 중 한영번역이 좀 더 어렵기 때문에 한영번역을 할 때 기계번역을 더 많이 활용하는 것을 알 수 있다. 따라서 비전문가인 이들이 기계번역을 활용하여 한영번역을 할 때 정확성과 효율성을 제고할 수 있는 방안을 모색해보고자 다음과 같이 실습을 진행하였다.

3.2 실습 과정

실습 참여자들이 기계번역을 활용하여 번역 실습을 한 텍스트는 기계번역이 가장 많이 활용되는 기술적 문서에 가까운 정보적 텍스트(informative text)이다. 기계번역과 포스트 에디팅은 은유적 표현이나 용어 변화가 많지 않은 기술적 문서에서 주로 활용된다는 연구(Koglin 2015: 129)에 기반하여 정보전달 텍스트를 선정하였다. 참여자들은 두 가지 텍스트를 실제 번역해보고, 기계번역을 활용하지 않고 처음부터 직접 한 번역(참여자 직접 번역)과 한-영 기계번역 결과물을 포스트 에디팅(한-영 MTPE), 한-일-영 기계번역 포스트 에디팅(한-일-영 MTPE)을 해본 후 이 세 과정 중 어떤 것이 효율성과 정확성 측면에서 가장 효과가 좋은지 비교해보고 자신들의 생각을 기록해보았다.

〈그림 1〉 실습 과정 요약



첫 번째 텍스트는 인간 복제에 대한 정보 전달 텍스트이고, 두 번째 텍스트는 해운대를 소개하는 텍스트로 부산관광정보를 제공하는 부산광역시 홈페이지에서 발췌한 것이다.

<텍스트 1>

인간 복제는 전 세계적으로 논쟁이 되고 있는 문제이며 많은 생명과학자 팀이 과학적 연구 목적으로 배아를 만드는 것과 관련한 연구에 종사하고 있다. 줄기세포는 여러 다른 종류의 세포로 변형될 수 있으며 손상된 기관이나 조직을 재생시킬 수 있는 가능성을 제공한다. 많은 과학자는 줄기세포가 당뇨병, 암, 심장발작 뿐만 아니라 알츠하이머나 파킨슨병과 같은 뇌 관련 질환을 치료하는데 희망을 준다고 믿는다.⁶⁾

<텍스트 2>

부산 해운대구에 위치한 해운대 해수욕장은 넓은 백사장과 아름다운 해안선을 자랑하고 있으며 얇은 수심과 잔잔한 물결로 해수욕장의 최적 조건을 갖추고 있다. ‘부산하면 제일 먼저 떠올리는 곳이 해운대 해수욕장’이라고 할 만큼 부산을 대표하는 명소이며, 해마다 여름철 피서객의 규모를 가능하는 척도로 이용될 만큼 국내 최대 인파가 몰리는 곳이기도 하다.⁷⁾

4) 첨부 1 번역 및 기계번역에 대한 사전 조사 참조.

5) 응답자 중 5명(약 10%)은 영한번역이 더 어렵다고 응답하였고, 3명(약 6%)은 둘 다 비슷하다고 답하였다.

6) 시사 뉴스 통번역 학습서 뉴스 잉글리시 파워딕(곽영섭·양승진 2008) 기사에서 발췌 및 편집

7) 부산광역시 홈페이지 부산관광정보 해운대 해수욕장 소개글에서 발췌 (<https://www.busan.go.kr/geopark/tm040401>)

실습 참여자는 위 두 텍스트를 직접 번역해보고, 한-영 MT를 포스트 에디팅과 한-일-영 MT를 포스트 에디팅 해보며 어느 과정이 비전문가인 자신들에게 가장 효과적인 방법인지 살펴보았다.

우선, 한-영 MT와 한-일-영 MT 결과물을 비교·분석해보자.

3.3 한-영 MT 와 한-일-영 MT 비교 분석

<텍스트 1>에 대한 <한-영 MT>와 <한-일-영 MT>⁸⁾는 다음과 같다.

<한-영 MT>

Human cloning is a global dispute issue and many teams of biotechnology scientists are involved in research into embryos for scientific research purposes. Stem cells can be transformed into different kinds of cells and provide the possibility to regenerate damaged organs or tissues. Many scientists believe that stem cells hope to treat diabetes, cancer, heart attacks as well as brain-related diseases such as Alzheimer's and Parkinson's.

<한-일-영 MT>

Cloned humans are an issue that has been discussed around the world, and many teams of scientists in life are engaged in research in connection with creating embryos for scientific research purposes. Stem cells can be transformed into several other types of cells, offering the potential to regenerate damaged organs and tissues. Many scientists believe that stem cells give hope to treat diabetes, cancer, heart attacks, as well as brain related diseases like Alzheimer's and Parkinson's disease.

한국어-영어 기계번역 오류 유형을 정리한 선행 연구⁹⁾(윤미선 외 2018: 52)를 기반으로 위 기계번역 결과물의 오류를 찾아보면, <한-영 MT>에서 누락 오류가 발생한 것을 볼 수 있다. 원문은 “많은 과학자는 줄기세포가 당뇨병, 암, 심장발작 뿐만 아니라 알츠하이머나 파킨슨 병과 같은 뇌 관련 질환을 치료하

8) <한-영 MT>와 <한-일-영 MT>는 4월 30일자 구글 번역 결과물

9) <첨부 3>의 선행연구 기계번역 오류 유형 참조

는데 희망을 준다고 믿는다”인데 이 문장을 <한-영 MT>에서는 “Many scientists believe that stem cells hope to treat diabetes, cancer, heart attacks as well as brain-related diseases such as Alzheimer's and Parkinson's”라고 번역하였다. ‘줄기세포가 희망을 준다’는 의미를 ‘stem cells hope’이라고 번역하였고, ‘준다’는 의미가 누락된 것이다. 이는 원문의 복잡한 문장구조에 기인한 오류이다. 복문으로 구성된 이 문장에서는 ‘많은 과학자는... 믿는다’라는 문장 안에 ‘줄기세포가... 희망을 준다’라는 문장이 삽입되어 있는데, 기계번역은 이 구조를 제대로 파악하지 못한 것이다. 반면, <한-일-영 MT>에서는 기계번역 오류 유형에 해당할만한 오류가 발생하지 않았고, 위 문장의 경우도 “Many scientists believe that stem cells give hope to treat diabetes...”라고 의미 누락 없는 번역 결과물을 제시하였다.

<텍스트 2>에 대한 <한-영 MT>와 <한-일-영 MT>는 다음과 같다.

<한-영 MT>

Haeundae Beach, located in Haeundae-gu, Busan, boasts a large white sand beach and beautiful coastline, and is equipped with a shallow water depth and calm waves for optimal beaches. Busan is the most famous spot in Haeundae Beach, which is the most famous place in Busan. It is also the place where the biggest crowd in Korea is used as a scale to measure the size of summer passengers every year.

<한-일-영 MT>

Located in Busan Haeundae-gu, Haeundae Beach boasts a large sandy beach and a beautiful coastline, with shallow water depths and gentle waves providing optimum conditions for the beach. Haundae Beach, the first place to think of "Pusan," is a landmark that represents Busan, and it is used as a measure to measure the size of summer vacationers every year, even when the nation's largest people are flooded there.

<텍스트 2>의 경우 두 기계번역 모두 오류가 등장했다. 우선, <한-영 MT>의 경우, “‘부산’하면 제일 먼저 떠올리는 곳이 해운대 해수욕장이라고 할 만큼 부산을 대표하는 명소이며”라는 원문 문장에 대해 “Busan is the most famous

spot in Haeundae Beach, which is the most famous place in Busan”라고 번역하여 통사구조 오류를 범하고 있다. 원문에서는 해운대 해수욕장이 ‘부산’하면 제일 먼저 떠올리는 곳이라고 하였는데, 기계번역에서는 ‘Busan is the most famous spot in Haeundae Beach’라고 하여 단어 차원의 통사구조 오류를 보이고 있다. 반면 <한-일-영 MT>에서는 이 부분에 대해 “Haundae Beach, the first place to think of “Pusan,” is a landmark that represents Busan”라고 번역하였다. ‘부산’이라는 지명에 대해 ‘Busan’과 ‘Pusan’이라고 번역하여 일관적이지 않은 단어 사용 오류를 보이고는 있으나, 이 오류는 단순한 지명 철자법 오류이므로 통사구조 오류에 비해서는 포스트 에디팅을 할 때 그만큼 인지적 노력이 투입되지는 않는다.

두 번째 두 기계번역 결과물을 비교해볼 부분은 원문에서 “해마다 여름철 피서객의 규모를 가늠하는 척도로 이용될 만큼 국내 최대 인파가 몰리는 곳”이라고 언급한 부분이다. 이 부분에 대해 <한-영 MT>에서는 “It is also (1)the place where the biggest crowd in Korea is used as a scale to measure the size of (2)summer passengers every year”이라고 하였고, <한-일-영 MT>에서는 “it is used as a measure to measure the size of (2-1)summer vacationers every year, even when (1-1)the nation's largest people are flooded there”라고 하였다. 일단 주의 깊게 살펴볼 부분은 밑줄 친 두 부분이다. “최대 인파가 몰리는 곳”이라는 부분을 <한-영 MT>에서는 “(1)the place where the biggest crowd in Korea”라고 번역하여 ‘인파가 몰리다’는 의미가 누락되는 오류가 발생한 반면 <한-일-영 MT>에서는 “(1-1)the nation's largest people are flooded there”라고 하여 ‘몰리다’는 의미가 누락되지 않았다. 다음 ‘여름철 피서객’을 <한-영 MT>에서는 ‘(2)summer passengers’라고 하여 오역이 발생했지만, <한-일-영 MT>에서는 ‘(2-1)summer vacationers’라고 정확하게 번역하였다.

물론 <한-일-영 MT> 결과물에 오류가 전혀 없는 것은 아니다. 앞서 살펴보았듯이 첫 번째 비교 부분에서는 용어 사용 일관성 문제도 있고, 두 번째 비교 부분에서는 문장 연결 시 논리적으로 어색한 곳도 있다¹⁰⁾. 그러나 <한-영 MT>

10) 두 번째 비교 부분인 “it is used as a measure to measure the size of summer vacationers every year, even when the nation's largest people are flooded there” 문장에서 ‘even when’ 대신 인과관계를 나타내는 접속사 ‘since’ 등을 사용하면 논리

오류와 경중을 비교해볼 때 <한-일-영 MT>의 오류는 매우 심각하거나 명백한 오류는 아니라고 할 수 있다. 그렇다면 비전문가인 통번역학부생이 이러한 특징을 보이는 두 기계번역 결과물을 포스트에디팅 해보며 본인들은 향후 한-영 번역 시 어떤 기계번역 결과물을 좀 더 활용할 것인지, 그 이유는 무엇인지 실습 참여자의 의견을 취합·정리해보았다.

3.4 참여자 인식 조사

기계번역 결과물에 어떤 오류 유형이 등장하는지(박옥수 2017; 곽중철·한승희 2018; 서보현·김순영 2018) 또는 포스트에디팅 할 경우 어떤 부분을 주로 에디팅 하는지(김순미 2017; 이상빈 2017; 마승혜 2018a; 이상빈 2018)에 대한 선행연구는 다수 진행되었기 때문에 본고에서는 비전문가인 학부생들이 기계번역 포스트 에디팅을 하며 어떤 부분을 수정하는지에 초점을 맞추기 보다는 기계번역 활용에 대한 사용자로서의 인식에 초점을 맞추어 살펴보았다¹¹⁾.

우선 텍스트를 <직접 번역>, <한-영 MTPE>, <한-일-영 MTPE>를 진행한 후 실습 참여자가 답한 질문은 다음과 같다.

<질문>

1. 정보전달 텍스트 한영번역 경우 직접번역과 기계번역 포스트 에디팅 중 어떤 것이 더 효율적이라고 생각하는가?
2. 한영 기계번역에서 한국어→영어 기계번역과 한국어→일본어→영어 기계번역 중 어떤 것이 더 정확도가 높다고 생각하는가?
3. 향후 정보전달 텍스트 한영번역 업무를 수행한다면, 직접번역, 한→영 기계번역, 한→일→영 기계번역 중 어떤 과정을 선택하겠는가? 그 이유는?

실습 참여자들은 ‘효율적’이라는 단어에 대해서 ‘들인 노력에 비해 얻는 결과가 큼’이라는 사전적 정의를 숙지한 후 위 질문에 응답하였다. 먼저 직접 번역해보고 그 다음 기계번역 포스트 에디팅도 해본 후 <직접 번역>과 <기계번

적으로 연결이 좀 더 매끄러울 것임.
11) [첨부 2] 사후조사 참조

역 포스트 에디팅> 중 어느 과정이 최종 번역물을 완성하는데 더 효율적이냐는 질문에 응답하였다. ‘효율성’에 대한 질문에서는 실습 참여자 51명 전원이 직접 번역하는 것보다 기계번역으로 1차 결과물을 얻은 후 이를 포스트 에디팅 하는 것이 더 효율적이라고 응답하였다. 이 질문에 대한 응답 사례는 다음과 같다.

<효율성에 대한 참여자 응답 사례>

“기계번역 포스트 에디팅이 압도적 효율성을 가지고 있다. 정보전달 텍스트의 경우 해당 분야 필드에서 자주 사용되는 단어만 제 뜻으로 입력해 놓는다면 오역의 가능성이 낮기 때문이다.”

“기계번역 포스트 에디팅이 더 효율적이라고 생각한다. 정보전달 텍스트의 경우 글의 문맥이나 뉘앙스 같은 세밀한 표현 보다는 글의 핵심내용을 전달하는 것이 주목적이다. 그러므로 수초면 바로 결과가 나오는 기계번역을 적극 활용하는 것이 시간과 노력을 대폭 줄여줄 것이다.”

실습 참여자 전원이 위와 같은 설명을 바탕으로 정보적 텍스트 한영번역은 기계번역 포스트 에디팅을 활용하는 것이 좀 더 효율적이라고 응답하였다.

다음, 두 번째 질문인 ‘정확도’에 대한 응답 사례를 살펴보자. 참여자 전원이 정보적 텍스트 한영 번역의 경우 직접번역 보다는 기계번역 포스트 에디팅이 더 효율적이기 때문에 후자를 택하겠다고 응답한 상황에서 그렇다면 어떻게 기계번역 포스트 에디팅의 정확도를 높일 수 있는지 모색해 보았다. 그 중 하나가 다단계 기계번역 활용이었고, 따라서 한-영 기계번역과 한-일-영 기계번역을 비교해보았다. 정확도에 대한 질문에서는 참여자 51명중 39명(약 76%)이 한-일-영 기계번역이 더 정확하고 자연스러운 문장을 만들어 낸다고 응답하였고, 9명(약 18%)는 한-영 기계번역이 더 낫다, 3명(약 6%)은 둘 다 비슷하다는 의견을 냈다.

그렇다면 최종적으로 향후 정보적 텍스트 한영번역을 한다면 <직접번역>과 <한-영 MTPE>, <한-일-영 MTPE> 중 어떤 과정을 선택하겠는가 하는 질문에 대해서는 51명 중 36명(약 70%)가 <한-일-영 MTPE>를 활용하겠다고 응답하였고, 10명(약 20%)은 <한-영 MTPE>, 4명(약 8%)은 <직접번역>, 1명(약 2%)은 그때그때 상황 봐서 선택하겠다고 응답하였다.

정보적 텍스트 한영번역을 할 때 <한-일-영 MTPE>를 활용하겠다고 선택한

참여자의 이유는 크게 번역 결과물의 ‘정확성’과 ‘(문장 구성의) 자연스러움’으로 나뉜다. 구체적으로 살펴보면 다음과 같다.

“‘한국어→일본어→영어’로 기계번역을 한 문단이 더 풍성한 어휘들과 구조를 사용하며 더 정확한 의미를 표현한다고 확인할 수 있다. 나도 이러한 번역과정을 보는 것은 처음인데 ‘한국어→영어’로 바로 번역하는 것보다 훨씬 더 정확한 번역본을 보여주어서 깜짝 놀랐다.”

위 학습자는 기계번역을 활용해보기는 해도 다단계 기계번역은 활용해 본 것은 처음이고, 다단계로 기계번역을 활용하는 것이 좀 더 정확한 번역 결과물을 보여주어서 놀랐다고 언급한다. 또 다른 응답자는 비전문가로서 번역할 때 문법 오류가 많이 생기는데 이를 보완하기 위해 다단계 기계번역을 활용하겠다고 하였다.

“나는 한→일→영 기계번역을 선택하겠다. 우선 전문 번역가가 아닌 상황에서 문장을 만들어내고 문법적인 오류를 찾아내는 것이 부족하기 때문에 후자를 선택하여 번역한 후 어휘 위주로 포스트 에디팅을 하겠다.”

“한→일→영 기계번역이 문법구조 측면에서 볼 때, 더 정교하고 수정할 부분이 적어서 후자를 선택하겠다.”

실제로 비전문 번역가로서 일을 해본 경험이 있는 참여자의 경우 다음과 같은 응답을 하였다.

“일본어로 되어있는 만화, 소설, 애니메이션 자막 등을 한국어나 영어로 옮겨진 횟수가 한국어로 된 글이나 자막을 영어로 옮기거나 그 반대의 경우보다 훨씬 많다. 이게 어느 정도나 하면, 신경망 번역이 있기 전, 신뢰도가 그렇게 떨어지던 구글 번역기를 가장 정확하게 사용하는 방법이 일본어로 옮겼다가 다른 언어로 옮기는 것이라는 말이 나올 정도였고 실제로 그러했다.”

결론적으로 <한-일-영 MTPE>를 선택하는 이유를 종합적으로 설명한 참여자의 의견이다.

“정보전달 텍스트의 경우 정확한 정보전달이 우선이 되어야 한다. 실제로 ‘내가 알지 못했던 구글 번역기의 기능’을 바로 한-영 기계번역을 돌려 보았을 때 ‘Features of Google Translator that I did not know’로 번역이 된다. 어느 정도 번역이 되기는 했지만 약간 부족하다고 느껴서 일본어로 돌려 보게 되었고 ‘私知らなかつたGoogleの翻訳機能’ 이렇게 번역이 된 것을 바로 영어로 돌려보았다. 그러자 ‘Google translation function that I did not know’로 번역이 되어 한-영으로 돌린 것과 꽤 차이가 나는 모습을 볼 수 있었다. 이 두 개의 문장을 비교해 보았을 때 한국어에서 영어로 바로 번역을 한 것 보다 일본어를 중간에 끼었을 때가 더 깔끔한 문장으로 번역 되었다고 볼 수 있었다. 한국어에서 영어로 기계번역을 하는 것 보다 한국어에서 일본어, 일본어에서 영어로 기계번역을 하는 것이 더 정확도가 높아지고 많은 문장들이 보다 더 매끄럽게 번역이 된다는 것을 위에서 조사하면서 알 수 있었기 때문에 한국어-영어 기계번역 보다는 한국어-일본어-영어 기계번역을 사용하게 될 것 같다고 판단했다.”

성을 차지하고 있다. 하지만 기계는 공감할 수 없는 인간의 높은 차원의 감정을 사람들끼리 공유하면서 살아가고 있다. 번역 시간은 느릴지라도 노력 부분에서는 기계보다 더 낮고, 공감도가 높은 인간번역 과정을 선택할 것이다.”

“인간의 감정표현과 같이 세밀한 번역이 필요한 경우나, 동음이의어의 구별·처리와 같은 부분은 기계번역을 하더라도 인간번역의 수정 과정을 거쳐야 완벽한 번역물이 나올 수 있다고 생각한다.”

마지막으로 기계번역을 여러 방식으로 활용해본 후 향후 기계번역이 인간번역가를 대체할 만큼 위협적이라는 생각이 드는지 아니면 기계번역을 활용한 인간번역의 업무가 더 많아지고 다양해질 거라고 생각하는지에 대한 실습자의 인식에 대해 조사해보았다.¹²⁾ 그 결과 41명(약 80%)은 기계번역을 활용한 인간 업무가 더 다양해질 거라고 답하였고, 10명(약 20%)은 기계번역이 인간번역을 대체할 만큼 위협적이라고 응답하였다. 인간번역을 대체하지는 못할 것이라고 응답한 참여자의 경우 주된 이유는 감정과 맥락을 전달하기에는 기계번역이 아직 한계가 있기 때문이라고 생각하기 때문이고, 향후 번역 업무가 더 세분화되어 기계번역만을 활용한 번역, 기계번역 포스트 에디팅을 통한 번역, 전적으로 인간이 담당하는 번역으로 나눌 수 있고, 더 나아가 기계번역 훈련을 위한 번역 데이터 생산자, 관리자와 같은 새로운 직종이 등장할 것 같다는 의견이 제시되었다.

“한-영 기계번역을 선택하겠다. 왜냐하면 모국어인 영어로 기계번역 되었을 때 만들어지는 오역을 감지하는 것이 비교적 쉽기 때문이다.”

“기계번역을 활용한 인간번역의 업무가 많아지는 않아도 다양해질 것이다. 아직까지는 5000원(장당)이라는 썩 가격에 초벌번역가를 모집하지만 이는 무급인 기계번역이 대체할 것이다. 초벌번역 이후 해당분야의 전문 지식을 가진 번역가들이 이를 에디팅 하는 직업이 증가할 것이다. 반면 감정을 전달하는 텍스트(영화자막, 노래, 시, 소설 등)는 지금까지 그래왔고 앞으로도 계속 기계가 인간을 대체하기는 어려울 것이다.”

“한-일-영 중역으로 인한 오류는 더 커질 수도 있기 때문에 한-영 기계번역을 선택하겠다.”

“기계번역은 데이터에 의존해야하기 때문에 데이터를 보완하고 더 정확히 만들며 관리하는 일을 담당할 인간 업무가 더 많아질 것이다”

참여자 중 소수이기는 하지만 4명은 <직접번역>을 선택하였다. 그러나 이들은 주로 ‘정보적 텍스트 번역에 한정하여 논의 한다’는 본 실습의 취지에 벗어난 답변을 하였다. 즉, 감정을 전달하고 공감할 수 있는 번역을 위해서 <직접번역> 과정을 선택할 것이라고는 하나, 이는 정보적 텍스트 외에 설득적 또는 문학적 텍스트 번역 시 고려해 보아야 하는 점이다.

“물론 기계번역은 인간번역과 비교해서 번역 시간이 짧다는 점에서 우월

반면 기계번역이 인간번역가를 대체할 정도로 위협적이라고 답한 참여자는

12) [첨부 2] 사후조사 문항 4 참조

다음과 같은 의견을 제시하였다.

“시간이 지날수록 기계번역이 위협적으로 다가올 것 같다. 아직 문화, 감정 전달에 대한 데이터가 충분하지 않아 오류가 있기는 하지만 더욱 발전될수록 데이터들이 많이 쌓여 인간의 검토가 필요 없을 수준에 도달할 수도 있을 것이다.”

3.4 실습 의의 및 한계

본 실습에서는 어떤 기계번역 결과물이 더 나은지 비교하거나 기계번역 결과물의 오류 지적에 초점을 맞추기 보다는 실습 참여자들이 실습 과정을 경험하며 깨닫게 되는 인식 변화에 초점을 맞추었고, 여기에서 연구 의의를 찾고 있다. 우선 다단계 기계번역 활용 실습 후 얻을 수 있는 의의로는 첫째, 실습 참여자들은 비전문가로서 어렵게 느끼는 한영 번역을 좀 더 정확하게, 좀 더 효율적으로 진행할 수 있는 한 방안을 모색해보았다는 점이다. 실습 후 한영 번역 시 기계번역을 활용하겠다는 응답자 수가 소폭 더 증가하였고(실습 전 80% → 실습 후 90%), 그 중에서도 다단계 기계번역 활용 방안에 대해 생각해본 적 없던 참여자들이 새로운 방법을 모색해볼 수 있는 계기가 되었다.

“상당히 미숙하고 경험이 없는 번역가로서 처음부터 번역을 진행하는 것보다 기계번역을 활용하면 시간적으로나 노력 상으로나 더 효과적일 수 있다고 생각한다.”

둘째, 실습 참여자들은 향후 기계번역을 더 잘 활용하기 위해서라도 근본적인 번역 역량을 키우는 것이 중요하다는 것을 깨달을 수 있었다. 어떤 엔진을 사용하든지, 어떤 방식으로 사용하든지 기계번역의 오류를 파악해내고 좀 더 정확하게 포스트 에디팅하기 위해 기초 번역 역량 증진의 필요성에 대해 깨닫게 된 것이다.

“결국에는 직접 스스로의 번역 수준을 높여 기계번역보다 조금은 느릴지라도 번역 실력을 단련하여 간단하고 명료하게 내용을 전달하고 포스트 에디팅 할 수 있도록 실력을 늘리기 위해 노력하여야겠다고 다짐하는 계기가 되었다.”

마지막으로, 미래에 사라질 직업 1위에 번역가가 선정¹³⁾되며 통번역 전공에 대한 회의가 증가하고 있던 상황에서 참여자들은 직접 기계번역을 해보고 포스트 에디팅도 해보며 막연한 두려움에서 벗어나 오히려 번역 관련 인간이 할 일이 다양해질 수 있겠다는 긍정적인 인식을 갖는 계기가 되기도 하였다.

단기간 짧은 텍스트를 바탕으로 진행한 실습이기 때문에 실습 한계도 상당히 많다. 우선, 실습 시 사용한 텍스트가 모든 정보적 텍스트를 대표하는 것은 아니기 때문에 모든 정보적 텍스트 기계번역 활용 시 <한-영 MTPE>보다 <한-일-영 MTPE>가 더 낫다고 일반화 할 수 있는 것은 아니다. 정보를 전달하는 텍스트라고 하더라도 주제에 따라, 문장구성 방식에 따라 결과물에는 차이가 있을 수 있다. 또한 기계번역 결과물은 데이터 학습에 따라 시시각각 달라지기도 한다. 따라서 실습 진행 시점(2019년 4월) 이후 결과물은 또 달라질 수 있고, 그에 따라 분석 결과도 달라질 수 있다.

다만 비전문가들의 기계번역 활용도가 점점 높아지고 있으므로 경우에 따라 다양한 방식으로 기계번역을 활용할 수 있는 방안을 소개해보고자 하였다. 이를 바탕으로 기계번역 훈련을 위한 데이터가 충분히 누적되지 않은 언어쌍 번역을 할 경우에도 데이터가 충분히 쌓인 언어쌍 기계번역을 중간 과정으로 하여 번역 정확도를 높일 수 있을 것이다.

4. 결론

기계번역을 활용한 비전문가 활동이 점차 증가하고 있는 상황이다. 이러한 시점에 기계번역을 좀 더 효율적으로 정확하게 활용해보는 여러 방안에 대해 모색해 본다면 비전문가 번역 품질 향상에 기여할 수 있을 것이다. 특히 언어적 주변국(linguistically-peripheral countries)¹⁴⁾ 언어 번역 시 기계번역을 활용하고

13) 잡코리아와 아르바이트 포털 알바몬이 직장인 및 취업생 4147명을 대상으로 ‘미래에 사라질 직업 vs 살아남을 직업’에 대해 설문조사를 실시한 결과...직장인 및 취업생들이 꼽은 미래에 사라질 것으로 생각되는 직업(*복수응답) 1위에 ▲번역가(31.0%)가 뽑혔다(강인귀 2018).

14) 언어적 주변국(linguistically-peripheral countries)을 “소수언어 커뮤니티

자 할 때 다단계 기계번역을 적용해 보면 좀 더 정확도를 높일 수 있을 것이다. 예를 들어, 한국어-카탈로니아어(Catalan)¹⁵⁾번역을 하고자 한다면 한국어-카탈로니아어 번역 데이터는 많이 축적되어 있지 않기 때문에 한국어-카탈로니아어 기계번역을 바로 하지 않고 한국어-프랑스어-카탈로니아어 기계번역을 한다면 좀 더 정확도를 높일 수 있다. 한국어-프랑스어 번역 데이터는 한국어-카탈로니아어에 비해 좀 더 많이 축적되어 있고, 프랑스어-카탈로니아어는 동일한 인도유럽어족이고 번역 데이터도 비교적 많기 때문에 기계번역 정확도가 제고될 수 있기 때문이다.

더 나아가 번역 데이터 축적 및 관리의 중요성도 간과해서는 안 된다. 우리나라는 IT 강국이라 불리고 전자정부가 매우 발달한 나라이지만, 양질의 데이터는 적고, 이를 체계적으로 관리하는 시스템도 정교하게 구축되어 있지 않다. 그나마 축적된 데이터를 효율적으로 분석하거나 거래하는 일은 더더욱 부족하다(정재승 2019). 기계번역 활용 시에도 결국 데이터 규모가 번역기 성능을 좌우하기 때문에 번역 데이터를 수집, 축적, 거래 및 활용이 더욱 활발해져야 하고 번역 데이터의 중요성에 대한 인식 또한 제고되어야 한다.

아울러 번역 학습자를 포함한 비전문가의 번역 역량 자체 향상에도 계속 힘써야 한다. 앞서 실습 참여자들의 실습 일지에서도 보았듯이, 탄탄한 번역 실력이 뒷받침되어야만 기계번역을 제대로 활용할 수 있다. 번역은 이제 과거보다 훨씬 폭넓은 개념과 범위를 아우르게 되어, 페레즈 곤잘레스와 수삼 사레바(2012)는 이제 더 이상 전문가 번역이 주류가 아니며 향후에는 오히려 예외가 될 수도 있다고 역설한다. 디지털 환경에서 이루어지는 번역은 매우 다양한 배경의 사람들이 함께 지식을 공유하면서 협업과 분업을 통한 집단지성의 힘을 보이며 품질 향상을 이루고 있다. 이러한 새로운 번역 지형을 준비하기 위해 번역 역량을 강화하고 기계번역과 같은 번역 기술을 잘 활용한다면 앞으로 많은 기회를 맞이하게 될 것이다.

(minor-language communities)"(Pokorn 1995: 35) 또는 "언어사용 분포 및 확산이 제한적인 언어(a language of restricted distribution or limited diffusion)"(ibid.)라고도 명한다.

15) 인도유럽어족 로망스어 중 하나로 사용 인구는 500~700만 명이며 절반 이상이 카탈루냐 지방에 산다.

참고문헌

- 곽영섭, 양승진 (2008) 뉴스 잉글리시 파워워드. 서울: 도서출판 넥서스.
- 곽중철, 한승희 (2018) 「포스트에디팅 측정지표를 통한 기계번역 오류 유형화 연구」, 『통번역학연구』 22(1): 1-25.
- 김순미 (2016a) 「디지털 시대 비전문가들의 참여 번역 현상-그 의미와 번역계의 미래에 주는 시사점」, 『번역학연구』 17(3): 7-32.
- 김순미 (2016b) 「번역가의 전문성 개념 변화의 필요성-디지털 시대 ‘비전문가’번역을 중심으로」, 『통역과 번역』 18(특별호) 1-29.
- 김순미 (2017) 「신경망번역기(NMT) 활용 학부 번역교육의 가능성 연구」, 『통번역교육연구』 15(3): 5-37.
- 김호영, 홍남희 (2012) 「전지구적 미디어 텍스트의 온라인 유통과 자막 제작자의 역할: 미국드라마 팬자막(fansub) 제작자를 중심으로」, 『미디어, 젠더 & 문화』 23: 43-77.
- 마승혜 (2018a) 「한영 기계번역 포스트에디팅에 대한 경험적 고찰: 학부 교육 과정 및 결과를 중심으로」, 『통번역학연구』 22(1): 53-87.
- 마승혜 (2018b) 「문학작품 기계번역의 한계에 대한 상세 고찰」, 『통번역학연구』 22(3): 65-88.
- 박옥수 (2016) 「한영 병렬 코퍼스와 기계번역에서 의존 명사: ‘-것’이 포함된 어휘의 번역 방식 연구」, 『동아인문학』 37: 469-492.
- 박옥수 (2017) 「한영 기계번역에서 ST의 유형적 특성에 따른 번역 오류 분석」, 『동아인문학』 41: 155-183.
- 박지영 (2017) 「언어 기술 혁신과 통번역 산업의 미래-20대 글로벌 LSP의 통번역 서비스 분석을 바탕으로」, 『통번역학연구』 21(1): 137-168.
- 서보현, 김순영 (2018) 「기계번역 결과물의 오류유형 고찰」, 『번역학연구』 19(1): 99-117.
- 이근희 (2007) 「번역과 한국 및 일본의 근대화」, 『번역학연구』 8(2): 103-132.
- 이미경 (2019) 「기계번역 vs. 인간번역, 텍스트 모호성 해소를 위한 전략 고찰 -연설문 번역을 중심으로」, 『인문사회 21』 10(2): 687-700.
- 이상빈 (2017) 「학부번역전공자의 기계번역 포스트에디팅, 무엇이 문제이고, 무

- 있을 가르쳐야 하는가?], 『통역과 번역』 19(3): 37-64.
- 이상빈 (2018) 「국내학부 번역전공자의 구글 기계번역 포스트에디팅 과정 (process) 및 행위 연구」, 『번역학연구』 19(3): 259-286.
- 이성화, 김세현 (2019) 「영-한 및 한-영 기계번역 품질향상을 위한프리에디팅 기법 제안」, 『번역학연구』 19(5): 121-154.
- 이창수 (2019) 「문학번역에서의 기계번역과 인간번역 문체에 대한 전산문체학적 비교 연구」, 『번역학연구』 20(2): 111-130.
- 이주리에 (2018) 「인공신경망 기계 번역의 한일/일한 번역 품질에 대한 예비연구 - 품질향상 검토와 교열코드 적용」, 『통역과 번역』 20(1): 43-71.
- 이준호 (2019) 「문학번역 적용을 위한 기계번역의 현주소」, 『통번역학연구』 23(1): 143-167.
- 정희연 (2018) 「전자매체 시대를 고려한 문학번역 방법」, 『번역학연구』 19(4): 235-276.
- 조인희 (2018) 「기계번역을 활용한 영어작문 수업의 효과: 자기 효능감과 글쓰기의 양적 증가를 중심으로」, 『영어영문학연구』 60(2): 253-279.
- 최효은, 이지은 (2017) 「특허 기계번역 결과물의 평가 - KIPRIS의 무료 한영 기계번역을 중심으로」, 『번역학연구』 19(1): 259-286.
- Boéri, Julie (2012) 'Translation/Interpreting Politics and Praxis,' *The Translator* 18(2): 269-290.
- Cadwell, Patrick, Sharon O'Brien, & Carlos S.C. Teixeira (2017) 'Resistance and accommodation: factors for the (non-) adoption of machine translation among professional translators', *Perspectives* 26(3): 301-321.
- Cronin, Michael (2013) *Translation in the Digital Age*. New York: Routledge.
- DePalma, Donald and Nataly Kelly (2008) *Translation of, for, and by the people*. Lowell, MA: Common Sense Advisory
- Groves, Michael and Klaus Mundt (2015) 'Friend or Foe? Google Translate in Language for Academic Purposes', *English for Specific Purposes* 37: 112-121.
- Guo, Jia-Wen (2016) 'Is Google Translate Adequate for Facilitating Instrument Translation from English to Mandarin?', *CIN: Computers Informatics, Nursing* 34(9): 377-383.
- Calefato, Fabio, Filippo Lanubile, Tayana Conte and Rafael Prikladnicki (2016) 'Assessing the Impact of Real-time Machine Translation on Multilingual Meetings in Global Software Projects', *Empirical Software Engineering* 21(3): 1002-1034.
- Jenkins, Henry (2006) *Fans, Bloggers, and Gamers: Exploring Participatory Culture*. New York: New York University Press.
- 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): 126-143.
- Krings, Hans P. (2001). *Repairing Texts: Empirical Investigations of Machine Translation Post-editing Processes*. Kent, Ohio & London: Kent State University Press.
- Lee, Ji-min (2017) Crowdsourced translation and machine translation and their implications in the digital age: A case study of Flitto, 『번역학연구』 18(4): 91-120.
- McDonough Dolmaya, Julie (2012) 'Analyzing the Crowdsourcing Model and Its Impact on Public Perceptions of Translation', *The Translator*, 18(2): 167-191.
- Neather, Robert (2012) "'Non-Expert' Translators in a Professional Community: Identity, Anxiety and Perceptions of Translator Expertise in the Chinese Museum Community', *The Translator* 12(2): 245-268.
- O'Hagan, Minako (2009) 'Evolution of User-generated Translation: Fansubs, Translation Hacking and Crowdsourcing', *Journal of Internationalization and Localization* 1(1): 94-121.
- O'Hagan, Minako (2011) Community translation: Translation as a social activity and its possible consequences in the advent of Web 2.0 and beyond. *Linguistica Antverpiensia*, 10, 11-23.
- O'Hagan, Minako (2017) Deconstructing translation crowdsourcing with the case of a Facebook Initiative. *Human issues in translation technology*.

Routledge.

Pym, Anthony (2013) ‘Translation Skill-Sets in a Machine-Translation Age’, *Meta* 58(3): 479-672.

Pokorn, Nike Kocijančič (2005) *Challenging the traditional axioms: Translation into a non-mother tongue*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Pérez-González, Luis and Şebnem Susam-Saraeva (2012) ‘Non-professionals Translating and Interpreting: Participatory and Engaged Perspectives’, *The Translator* 18(2): 149-165.

Raído, Vanessa Enríquez (2016) ‘Translators as Adaptive Experts in a Flat World: From Globalization 1.0 to Globalization 4.0?’, *International Journal of Communication* 10: 970-988.

<인터넷 자료>

강인귀 (2018) 「미래에 사라질 직업 1위 ‘번역가’, 살아남을 직업 1위는?」, 『머니S』 2019년 8월 17일 검색.

노자운 (2016) 「[한국어와 인공지능]⑨ 투로프스키 구글 번역 총괄 “中 바이두도 구글의 중국어 번역 속도 못 따라와...스케일 차이”」, 『조선비즈』 2019년 6월 5일 검색.

박상익 (2017) 「번역청 설립을 공약하라」, 『아시아경제』 2019년 8월 17일 검색.

박상익 (2018) 「번역청을 설립하라」, 『신동아』 2019년 8월 17일 검색.

박성용 (2019) 「일본, 왜 노벨상 수상자 다수 배출국 됐나...번역국과 탄탄한 인문학 기반」, 『BBS 뉴스』 2019년 8월 17일 검색.

정재승 (2019) 「데이터청과 데이터거재소를 설립하자」, 『중앙선데이』 2019년 8월 17일 검색.

Cronin, Michael (2010) ‘The Translation Crowd’, *Revista Tradum tica*, 8. Retrieved July 7, 2019, from <https://ddd.uab.cat/pub/tradumatica/15787559n8/15787559n8a4.pdf>

Diño, Gino (2019) The State of Neural Machine Translation for Asian Languages. Retrieved Aug. 17, 2019 from

<https://slator.com/features/the-state-of-neural-machine-translation-for-asian-languages/>

Greenstein, Eric and Daniel Penner (2015) Japanese-to-English Machine Translation Using Recurrent Neural Networks. Retrieved Aug. 19, 2019 from <https://cs224d.stanford.edu/reports/GreensteinEric.pdf>

Perrino, Saverio (2009) ‘User-generated Translation: The Future of Translation in a Web 2.0 Environment’, *The Journal of Specialised Translation* 12. Retrieved June 22, 2019, from https://www.jostrans.org/issue12/art_perrino.pdf.

[Abstract]

**A Search for Quality Improvement of Non-professional Translator
Translations through Relay Machine Translation:
A Case Study of an Undergraduate Course Practice**

Mah, Seung-Hye · Sung, Seung-eun
(Hankuk University of Foreign Studies)

The rapid development of machine translation(MT) has recently facilitated the spread of non-professional translations. Non-professionals or amateur translators participate in translation works in a wider range of contexts, emerging as a new distinctive social and cultural phenomenon. Therefore, this research explores ways to help non-professional translators make full use of MT systems. One of the ways suggested in this research is to run MT twice or a relay MT. It is applied to a Korean-English translation practice of 51 undergraduate students majoring in English interpreting and translation. The hypothesis at the practice is that the MT would produce better translation output in English if it goes through Korean-Japanese-English translation process because the quality of MT output depends heavily on accumulated parallel data. It has long been known that Japanese-English translation parallel data have been systematically stored, whereas the history of collecting Korean-English translation data is not as long as that between Japanese and English. The participants compare and analyze two MT outputs, and many of them conclude that the Korean-Japanese-English MT is better to post-edit than the Korean-English MT. They also emphasize that building their translation capacity is as important as knowing how to make use of MT efficiently because ultimately, they are the ones who judge the MT quality and make final decisions.

▶ Key Words: Machine Translation, Non-professional Translator, Kor-Eng Translation Education, Machine Translation Post-editing

▶ 주제어: 기계번역, 비전문가 번역, 한영번역, 기계번역 포스트 에디팅

마승혜

한국의국어대학교 영어통번역학부 조교수

shm213@hufs.ac.kr

관심분야: 번역교육, 문학번역, 기계번역 포스트 에디팅

성승은

한국의국어대학교 영어통번역학부 조교수

michsung@hufs.ac.kr

관심분야: 번역교육, 번역과 이데올로기, 아동문학번역

논문투고일: 2019년 8월 30일

심사완료일: 2019년 9월 30일

게재확정일: 2019년 10월 8일

[첨부 1] 사전조사

번역 및 기계번역에 대한 사전 조사

[기본 정보]

- 1. 이름: _____
- 2. 학년: _____
- 3. 해외(영어권) 거주 경험(장소 및 기간): _____

[번역 관련 조사]

1. 영→한 번역 관련 수업을 들은 적 있다.
 예 _____ 아니오 _____
 ‘예’라고 응답한 경우 어떤 수업을 어느 정도 기간 동안? _____

2. 한→영 번역 관련 수업을 들은 적 있다.
 예 _____ 아니오 _____
 ‘예’라고 응답한 경우 어떤 수업을 어느 정도 기간 동안? _____

3. 영한번역과 한영번역 중 본인이 좀 더 어렵게 느끼는 것은? _____

4. 기계번역을 활용해본 적이 있는지?
 예 _____ 아니오 _____
 - ‘예’라고 응답한 경우 어떤 기계번역 도구를 활용해 보았는지?
 (예를 들어, 구글번역, 파파고, 그 외 번역보조도구)

- ‘예’라고 응답한 경우 영한번역과 한영번역 중 어떤 번역에서 좀 더 많이 활용해보았는지?

5. 4번 문항에서 ‘아니오’라고 응답한 경우 번역할 때 기계번역을 활용해볼 의사가 있는 지?
 예 _____ 아니오 _____

- ‘예’라고 응답한 경우 영한번역과 한영번역 중 어떤 번역에 좀 더 활용하고자 하는지?

[첨부 2] 사후조사

기계번역 활용 후 인식 조사

[기본 정보]

- 1. 이름: _____
- 2. 학년: _____

[번역 관련 조사]

1. 정보전달 텍스트 한영번역 경우 직접번역과 기계번역 포스트 에디팅 중 어떤 것이 더 효율적이라고 생각하는가?
 한→영 _____ 한→일→영 _____

2. 한영 기계번역에서 한국어→영어 기계번역과 한국어→일본어→영어 기계번역 중 어떤 것이 더 정확도가 높다고 생각하는가?
 한→영 _____ 한→일→영 _____

3. 향후 정보전달 텍스트 한영번역 업무를 수행한다면, 직접번역, 한→영 기계번역, 한→일→영 기계번역 중 어떤 과정을 선택하겠는가? 그 이유는?
 예 _____ 아니오 _____

이유: _____

4. 기계번역을 활용해보니 기계번역이 인간번역가를 대체할 만큼 위협적이라고 생각하는지? 아니면 기계번역을 활용한 인간번역의 업무가 더 많아지고 다양해질 거라고 생각하는지? (답변 및 간단한 이유 설명)

[첨부 3] 선행연구 기계번역 오류 유형(윤미선 외 2018: 52)

| 영한 기계번역 오류 유형 | |
|---------------|---|
| 부정확한 단어 | 의미 전달에 문제가 있는 단어 (semantic error) |
| | 의미는 통하지만 형태적으로 부정확한 단어 (incorrect form) |
| | 불필요하게 추가된 단어 (added words) |
| | 전반적인 문체와 어울리지 않는 단어 (style) |
| | 부정확한 관용어 인식 (idioms) |
| 통사구조 | 단어 차원 어순 문제 (word order) |
| | 구(phrase) 차원 어순 문제 (phrase order) |
| | 기타 통사구조 오류 (other syntactical errors) |
| 문법 | 구두점 관련 오류 (punctuation) |
| | 기타 문법 오류 (other syntactical errors) |
| 일관성 | 일관적이지 않은 단어 사용(non-consistent word use) |
| 누락 | 단어의 누락 (missing words) |
| 무번역 | 단어, 구 차원에서의 무번역 (non-translation) |
| 알려지지 않은 단어 | TT에는 대응어가 없는 단어 (unknown words) |