

영-한 및 한-영 기계번역 품질향상을 위한 프리에디팅 기법 제안*

이성화 · 김세현
(부산대)

1. 서론

최근 기계번역에 신경망 기계번역(Neural Machine Translation) 시스템이 도입되면서 기계번역의 품질이 급속히 개선되고 있다. 기계번역은 반복적이면서도 일관성이 요구되는 대규모의 비문학 기술 텍스트의 번역에 특히 유용하게 사용되고 있으며 비용 효율적인 장점이 있지만 실제 기계번역 결과물(output)을 바로 사용하기에는 그 품질이 불충분하며(Hutchison 2003: 9-10) 아직까지 인간 번역가의 수준을 따라잡는 것은 무리라는 지적이 있다(김성수 2017). 기계번역 결과물이 출판 가능한 수준으로 개선되려면 인간이 검토하고 편집하는 등 추가적인 개입이 요구된다. 본 연구는 기계번역 결과물의 품질을 향상시키기 위한

방안의 일환으로 효율적인 프리에디팅(pre-editing) 기법을 정리·제안하는 것을 목적으로 한다.

기계번역의 효율성은 결국 기계번역 결과물 품질에 의해 결정되기(Hutchins & Somers 1992:2) 때문에 번역학계에서도 실제 기계번역 결과물을 바탕으로 오류를 파악하여 품질을 개선하는 것에 목적을 둔 실증적인(empirical) 연구가 많이 이뤄졌다. 기계번역에 대한 실증적¹⁾ 성격의 국내 연구 현황을 살펴보면, 한국어-영어 조합에서는 기계번역 결과물의 오류 분석(박옥수 2016, 2017, 2018; 신지선·김은미 2017; 광중철·한승희 2018; 서보현·김순영 2018; 최동익 2013 등)이나 포스트에디팅(마승혜 2018a, b; 신지선 2017; 이상빈 2017, 2018a, b; 이준호 2018 등)에 치중²⁾하고 있는데 이러한 연구들이 궁극적으로 지향하는 바는 기계번역의 생산성 향상으로 귀결된다. 하지만 기계번역 결과물의 낮은 품질을 포스트에디팅(만)으로 품질을 향상시키려는 것은 비생산적일(Koponen 2016: 132) 뿐만 아니라 원문을 적절히 편집하면 여러 번의 (사후)편집을 피할 수 있다(Forcada 2010: 217)는 연구결과도 있다. 따라서 고품질의 기계번역 결과물을 얻기 위해서는 포스트에디팅에만 의존할 것이 아니라 포스트에디팅과 프리에디팅이라는 두 가지 상호보완적인 과정을 모두 고려할 필요가 있다.³⁾

따라서 기계번역의 효율성을 최대화하기 위한 프리에디팅 기법에 대한 연구가 필요하다 하겠다. 특히 영-한 언어 쌍의 프리에디팅 전략에 대한 연구가 부족한 상황이므로, 본 논문은 한-영 조합뿐만 아니라 영-한 언어조합에 대한 프리에디팅 기법을 고찰함으로써 기계번역 결과물의 품질을 개선하는 초석을 다지고자 한다. 또한 기계번역 중 구글 번역(Google Translate)과 네이버의 파파고(Naver Papago)의 장단점을 파악하고 그에 따른 차별적인 프리에디팅 전략을 제시하고자 한다.

- 1) 본 연구에서는 기계번역에 관한 교육적·철학적 관점에서의 연구는 제외하고 기계번역의 결과물을 논의하는 연구만을 선행 연구의 범위에 포함하였다.
- 2) 박옥수(2016, 2017, 2018), 김은미(2017), 신지선(2017), 이상빈(2018a, b)에서 프리에디팅의 개념이 언급되고 있으나 프리에디팅을 논문 전체의 주요 쟁점으로 다루고 있지 않다.
- 3) 프리에디팅 & 포스트에디팅 (제네바대학교 통번역학부):

<https://www.unige.ch/fti/en/faculte/departements/dtim/recherches/ta/>

* 2018년 1학기 ‘사회문화번역워크숍’ 수업 중 기계번역 프리에디팅에 참여하고 수집한 데이터 사용을 기꺼이 허락한 김나영, 김세미, 김정임 선생님께 감사한다.

본 연구는 부산대학교 번역학 대학원 수업에서 수행한 과제 및 토론 결과를 분석 정리하였다. 연구대상인 텍스트로는 신문에 게재된 사회현상에 관한 연설문, 칼럼 및 기사를 사용하였으며, 구글 번역과 파파고에서 각각 결과물을 도출하고 오류를 분석하여 적절한 에디팅을 수행하였다. 프리에디팅한 텍스트를 다시 두 기계번역에 입력하여 이전보다 더 충실한⁴⁾ 결과물을 얻게 되고, 동일한 에디팅이 3회 이상 발생할 때 그 전략을 프리에디팅 기법으로 제안한다.

이후의 논문 전개는 다음과 같다. 2장에서는 기계번역에 대한 연구 동향을 선행연구 중심으로 고찰하며, 3장에서는 텍스트 정보를 비롯한 연구방법에 대해 살펴본다. 4장 연구결과에서는 프리에디팅 기법을 영·한과 한·영 언어 조합 별로 제시하며, 5장에서는 본 연구를 정리하며 결론을 맺는다.

2. 기존 연구 동향

2.1 프리에디팅 관련 선행 연구

프리에디팅이란 기계번역에 적용하기 전에 원문 텍스트를 수정하는 것⁵⁾으로 크게 어휘 정규화(lexical normalisation)(Han *et al.* 2013), 텍스트 단순화 및 재배열(Bisazza & Federico *et al.* 2012 등), 통제언어(O'Brien 2006 등) 관점에서 접근할 수 있다. 이 가운데 통제언어 관점은 원문 텍스트 작성 단계에서부터 사용하는 언어를 제한하는 개념이므로 본고에서는 논외로 한다.

기계번역 시스템에 입력하는 텍스트를 단순화하여 기계번역 결과물의 품질을 개선하는 것은 포스트에디팅 과정에 직접적인 영향을 줄 수 있으며(Krings 2001: 556) 프리에디팅 규칙이 기계번역 결과물의 품질을 유의미하게 개선시킨다(Gerlach *et al.* 2013). 프리에디팅이 기계번역의 성능에 긍정적인 영향을 준다는 연구 결과는 기계번역 시스템의 유형별로 다음과 같이 다수 존재한다. 규

칙기반(O'Brien & Roturier 2007), 통계기반(Aikawa *et al.* 2007), 사례기반(Way & Gough 2005) 기계번역 체제에서 프리에디팅의 유용성이 입증된 바 있다.

지금까지 프리에디팅을 주요 쟁점으로 다룬 국내 연구는 포스트에디팅 연구에 비하면 대단히 부족한 실정이지만 프리에디팅이라는 개념을 암묵적으로 언급한 연구까지 포함하면 유의미한 연구들이 다수 존재한다. 가령 박옥수(2016: 489)는 의존 명사 '-것'이 포함된 어휘에 대한 기계번역의 번역 오류에 대하여 '도출한 규칙에 따라 원문을 정리한 후 재번역'할 것을 제안한다. 또한 박옥수(2017: 176)에서는 기계번역의 오류를 파악하여 '문장 나누기, 주어 첨가, 컴퓨터가 이해하기 쉬운 표현 사용하기, 상위어나 단순한 문장으로 표현하기, 지시어 대신 구체적인 지시 대상 넣기, 보편적 어휘 사용하기, 장황한 부분의 생략'의 전략을 활용해 원문을 고쳐 쓴 다음 재번역하였다. 해당 연구자가 제안한 이러한 '고쳐쓰기' 전략은 프리에디팅 기법으로 보더라도 무방할 것이다.

보다 명시적으로 프리에디팅을 논한 연구는 다음과 같다. 이상빈(2018a, b)의 영·한 포스트에디팅에 관한 현상학적 연구와 번역 프로세스 연구에서는 포스트에디팅을 경험한 학생들이 프리에디팅에 대한 개념을 학습하지 않았음에도 불구하고 프리에디팅의 필요성과 당위성 등을 스스로 인식하였다고 밝히고 있다. 김순미(2017)의 연구에서는 한·영 프리에디팅에 대한 명확한 개념을 제시한 후에 학부생들에게 번역을 실시하도록 하였고 이러한 과정을 통해 몇 가지 프리에디팅 기법을 도출해내기도 했다. 그리고 신지선(2017)의 정리에 따르면, 포스트에디팅의 능력에는 통제언어 규칙을 적용해 텍스트를 사전 편집한 다음 기계번역을 수행하는 프리에디팅 기술이 포함된다. 한편, 문장부호에 국한하여 프리에디팅 기법을 수립하거나(한현희 2018), 기법 도출보다는 프리에디팅 후 번역품질 평가에 초점을 둔 연구(진용주 외 2018)도 시도되었다.

이러한 기존 연구들을 통해, 프리에디팅은 기계번역의 품질을 높이는데 필요한 과정의 일부라는 공통된 인식이 존재하며, 기계번역을 경험해 본 연구자나 피험자는 그 필요성과 방법을 스스로 체득하고 있음을 확인할 수 있었다. 기존 연구에서 암묵적으로나 명시적으로 제안된 프리에디팅의 기법을 정리해보면 다음 <표 1>과 같다.

4) 본고에서는 충실성을 원문의 내용을 충분히 반영한 경우를 뜻하는 포괄적인 개념으로 사용하였다.

5) 모라비아 사이트:

info.moravia.com/blog/machine-translaion-pre-editing-to-boost-output-quality

〈표 1〉 기존 연구에서의 프리에디팅 기법

프리에디팅 기법	박옥수 (2016)	박옥수 (2017) 김순미 (2017)	박옥수 (2018)	신지선 (2017)	이상빈 (2018a) (2018b)	한현희 (2018)
생략된 표현은 첨가하기		✓	✓	✓		
모호한 표현을 명확하게		✓	✓			
은유를 일상적인 표현으로		✓				
복잡·난해한 표현을 쉽고 단순하게	✓	✓	✓	✓	✓	
텍스트 재구성하기 (분할/재배치/통합)		✓	✓	✓	✓	
문장부호 바꾸기					✓	✓

하지만 기존의 프리에디팅 연구는 주로 한·영(박옥수 2016, 2017, 2018; 김순미 2017) 기계번역 사례에 국한되었고 영·한 번역(이상빈 2018 a, b)을 다룬 연구라 할지라도 구체적인 사례 제시나 기법 도출 없이 프리에디팅의 필요성 강조에 중점을 두고 있다. 이에, 한·영과 영·한을 동시에 분석하여 프리에디팅 기법들은 비교·제시하는 연구가 필요하다.

2.2 기계번역 엔진별 선행 연구

기계번역의 효용성을 높이기 위해서는 기계번역의 엔진의 특성을 파악하여 이를 텍스트 유형이나 언어 쌍에 맞춰 적절히 사용할 필요가 있다. 기계번역 결과물의 품질이 낮거나 주어진 텍스트 영역(domain)에 맞지 않는 기계번역 엔진을 사용한 경우 그 대가는 고객이 아니라 포스트에디터가 부담하게 되므로, 적절한 기계번역 엔진을 선택하는 것은 전체 기계번역 과정에 있어 매우 중요하다(Thicke 2013). 그런데 한국어를 다루는 이전의 국내 연구에서는 번역하는 언어 쌍이나 콘텐츠, 기계번역별 장·단점에 대한 고려 없이 단일 혹은 복수의 기

계번역 엔진을 연구자 임의로 선정하여 사용한 경우가 많았다.

국내에서 구글 번역만을 사용한 연구에는 임순정(2012), 조수연·박경리·전종섭(2013), 임순정·한미선(2014), 마승혜(2018a), 이상빈(2018)이 있다. 이들 연구에서는 해당 번역 엔진의 특성이나 그 번역 엔진을 선택한 이유가 언급되지 않고 기계번역의 결과물만 논하고 있다. 한승희(2017)의 경우, 예비연구 단계에서는 구글 번역, 마이크로소프트 Bing 번역기(Microsoft Bing Translator), 네이버 파파고를 사용하였지만, 본연구에서는 구글 번역의 결과물에서 문법적 오류가 가장 적었다고 밝히면서 구글 번역만 활용하였다. 진실희(2017) 역시 한승희(2017)를 인용하면서 같은 이유로 구글 번역만 사용하였다. 이와 같이 단일 번역 엔진을 사용한 경우 구글 번역이 압도적으로 많이 채택되었다.

복수의 번역 엔진을 사용한 연구는 다음과 같다. 구글 번역과 파파고를 비교한 연구에는 신지선·김은미(2017), 박옥수(2016), 박영란(2018)이 있다. 신지선·김은미(2017)는 서로 다른 텍스트 세 유형을 대상으로 구글 번역과 파파고의 번역 결과를 정성적으로 분석하였는데 네이버 파파고가 2016년 10월부터 신경망 기반 번역으로 전환(김준석 2017: 71)하였으나 해당 연구에서는 통계기반 체제로 상정하고 있다. 박옥수(2016)와 박영란(2018)은 각 번역 엔진 결과물의 오류를 정량적으로 비교하였다. 박옥수(2016)에서는 문장이 성립되지 않는 경우가 구글 번역보다 파파고가 더 많았으나, 박영란(2018)에서는 파파고보다 구글 번역의 오역률이 높은 것으로 확인됐다. 이러한 상반된 결과가 나타난 것은 같은 기간 동안 파파고의 품질이 월등히 향상되었기 때문이라고 추정된다. 이상빈(2018b: 274-275)의 연구에서는 피험자가 구글 번역기의 결과물에 만족하지 못해 파파고 기계번역을 시도하였고 감탄사를 내뱉을 만큼 만족스러운 결과를 찾아내는 상황이 언급된다. 이러한 배경 하에 구글 번역과 이에 준하는 대조군(reference)으로써 파파고 기계번역 결과물을 비교하는 것은 의미있는 시도라고 판단된다.

본고는 구글 번역과 파파고의 기계번역 결과를 논의의 대상으로 삼되 두

6) 이상빈(2018)의 경우 연구자의 임의적인 선택이 아니라 번역 의뢰자의 번역 브리프 요청에 따라 구글 번역을 사용하였다. 특히 해당 연구자는 기계번역의 여러 변인들—텍스트 유형, 언어 조합, 번역기의 종류, 사용자 역량, 번역 기술 수준 등을 고려해 기계번역의 사용 수준을 선별적으로 결정해야 한다고 명시하고 있다.

번역엔진의 우열을 가능하는 것이 연구의 목적은 아님을 밝혀두는 바이다. 한편, 두 번역엔진 모두 신경망 기계번역 방식을 채택하고 있는데, 이러한 신경망 번역 체제하에서는 번역기가 계속해서 새로운 패턴을 스스로 학습하므로 특히 프리에디팅을 적용한 후의 기계번역 결과물을 검토하려는 연구라면 박옥수(2017)의 연구와 같이 데이터 확보 시점을 명시할 필요가 있다. 본고는 데이터 수집 기간을 2018년 3월에서 6월 사이로 명시하였다. 또한 데이터 수집 기간 이후에 같은 문장을 재번역하여 도출된 유의미한 결과물은 주석으로 보충 설명하였다.

2.3 연구대상별 선행 연구

이번에는 기존 연구들이 어떤 텍스트를 연구 대상으로 삼았는지 유형별로 살펴보겠다. 구체적으로, 앞서 언급한 대로 기계번역은 비문학 텍스트의 번역에 많이 활용되고 있는데, 국내의 연구 현황도 과연 그러한지 알아보겠다. 텍스트 유형별로는 특허(최효은·이지은 2017), 기관의 정관(신지선·김은미 2017), 의료 텍스트(이상빈 2018), 광고문(한승희 2017), 신문기사의 은유표현(마승혜 2018b), 속담(박영란 2018), 영상번역(최수연 2017), 정책브리핑 전사자료(진실희 2017), 자기소개서(김순미 2017) 등 기술적이고 정보성 위주의 텍스트에서 점차 그렇지 않은 텍스트로 연구가 확장되고 있다. 한편, 홀문장, 안은 문장, 이어진 문장 등 문장의 종류별(박옥수 2017)로 비교·분석하거나 특정 문법적 표현(의존명사 ‘것’)(박옥수 2016), 문장부호(한현희 2018)에 주목한 연구도 있다. 본고의 경우 기사문과 연설문, 칼럼을 연구 대상으로 삼아 정보적 텍스트, 설득적·표현적 텍스트를 두루 고찰하고자 한다.

기존 연구와 비교해 본 연구가 지니는 차별점은 다음과 같다. 첫째, 프리에디팅 기법을 도출함에 있어 기존의 기법을 원용하지 않고 실제 번역수업 과정에서 확보된 기계번역 결과물을 토대로 기술론적이고 상향식(bottom-up) 관점에서 접근한다. 이 과정에서 도출하는 프리에디팅 기법은 일반인도 내재화하여 손쉽게 활용할 수 있도록 가급적 포괄적이고 단순하게 만든다. 둘째, 하나의 특정 번역엔진의 결과만을 논한 기존 연구와는 달리 한국어 조합에서 활용도가 높은 구글 번역과 파파고의 결과물을 비교·분석하여 각 번역엔진의 장·단점을

파악하고 활용방안을 모색한다. 또한 각 번역엔진에서 데이터를 확보한 시점을 명시하고 데이터 추출 기간 이후에 발견된 개선된 번역 결과물의 경우는 주석에 추가 설명하였다. 셋째, 주로 특정 텍스트 유형에만 국한해 주로 한·영 번역 결과물에 대해 논의한 기존 연구와는 달리 정보적, 설득적, 표현적 텍스트를 모두 다루며 한·영과 영·한이라는 양방향 기계번역에 대한 프리에디팅 기법을 제안하고자 한다.

3. 연구 대상 및 연구 방법

3.1 연구에 사용된 텍스트

본고에서 사용된 프리에디팅 데이터는 2018년 3월부터 6월까지 15주간 동안 부산대학교 번역학 대학원 과목인 ‘사회문화번역워크숍’에서 얻어진 것이다. 이 수업에서 네 명의 대학원생들(석사과정생 2명, 박사과정생 2명) 및 교수자가 함께 구글 번역기와 네이버의 파파고를 효율적으로 사용하기 위한 프리에디팅 기법을 만들어 정리하였다. 이 연구에서는 한국어 원천언어(Source Text: ST)에서 영어 도착언어(Target Text: TT)로 번역할 때 필요한 프리에디팅 기법과 ST를 영어로 하여 한국어로 번역하는 경우 등 원천언어에 따른 두 종류의 프리에디팅 기법을 고찰·분석하여 결과를 보고한다.

텍스트 선정 기준으로는 주제 측면에서는 세 가지의 사회적 이슈—페미니즘, 인종차별, 난민문제—에 대한 영어와 한국어 텍스트를 각각 하나씩 선택하였다. 텍스트 종류에 따른 차이점을 관찰하기 위해서 설득적·표현적 텍스트인 사설 및 연설문을 비롯해 정보적 텍스트인 기사문을 선택하였다⁷⁾. 먼저 ST가 영어인 텍스트로는 글래머(*Glamour*)지에 게재된 미국의 전 대통령 버락 오바마(Barack Obama)의 연설문, 뉴욕타임즈에 게재된 인종차별에 대한 칼럼, 그리고 UN 뉴스에 게재된 예멘 난민에 대한 기사 등 세 종류이다. 영어 ST 텍스트는 모두 2,650의 단어로 이루어져 있다. 한국어 원천언어 텍스트는 역시 세 종류로

7) 한·영 언어 쌍의 경우 한국어 원문기사와 번역 기사를 각각 선택하여 유사점과 차이점을 살펴보려 하였으나, 데이터양의 한계로 유의미한 연구가 가능하지 않았다.

총 4,473자로 이루어졌으며, 한겨레신문에 게재된 미투운동에 대한 노혜경의 칼럼, 연합뉴스에 게재된 북한이탈주민 실태조사에 대한 기사, 및 뉴스 페퍼민트지에 실린 세계 인구에 관한 번역기사가 그것이다. 각 텍스트의 제목 및 세부사항은 <표 2>에 제시하였다.

<표 2> 텍스트 세부 정보

ST	제목	종류	게재지	단어수 (글자수)
영어	President Barack Obama Says, "This is What a Feminist Looks Like"	연설문	Glamour	1,551
	We're sick of racism, literally	칼럼	New York Times	802
	Yemen: Raging violence displaces more than 85,000 civilians, says UN refugee agency	기사	UN News	297
	총 단어수			2,650
한국어	새로운 말이 새로운 시대를 연다	칼럼	한겨레신문	494 (1,724)
	탈북민, 자영업 선호... '2017 북한이탈주민 실태조사'	기사	연합뉴스	208 (847)
	세계 인구에 관한 11가지 주요 사실	번역 기사	뉴스 페퍼민트	551 (1,902)
	총 단어수 (글자수)			1,253 (4,473)

3.2 연구 수행 방법

수업을 통해 프리에디팅 기법을 정하는 절차는 다음과 같다. 구글 번역과 파파고의 차이점 및 유사점을 살펴보기 위해서 동일한 ST에 대하여 두 명은 구글 번역을, 나머지 두 명은 파파고를 사용하였다. 우선 개인적으로 과제를 수행하였다. 과제 수행 시에는 다음의 절차를 따랐다. 주어진 텍스트를 각각의 MT에 실행시킨 후 그 결과물을 살펴본다. 이때 기계번역된 텍스트가 오류가 있거나 부자연스러우면 그 부분의 ST를 수정한 후 다시 MT를 실행한다. ST의 내

용을 충실히 반영하는 번역 결과물이 도출될 때까지 이 과정을 반복하여, 만족스러운 결과물을 도출한 에디팅을 프리에디팅 기법으로 제시한다. 프리에디팅된 텍스트를 MT에 실행하여 나온 결과물을 검토하여, 즉 포스트에디팅하여 최종 번역 결과물을 생산해낸다⁸⁾. 이러한 과정을 거쳐 수행된 과제를 수업 전에 온라인 학습관리 시스템(PLMS)에 탑재하여 서로 공유하였고 수업 중 프리에디팅 기법과 포스트에디팅된 결과물의 품질을 토론·평가하였다. 수업이 진행되어 MT와 프리에디팅에 익숙해짐에 따라 개인 과제 없이 수업 중 구글 드라이브의 스프레드시트를 사용하여 교수자와 학생 네 명이 토론을 병행하며 공동작업으로 기법을 만들고 적용하며 최종 결과물을 만들어냈다.

본고는 수업을 통해 얻어진 프리에디팅 기법을 언어 조합과 기계번역 종류로 분류하였다. 또한 수집된 프리에디팅 전략을 분석한 결과 크게 문장 구조를 변화시키는 경우와 의미의 모호성을 명확하게 하기 위한 경우로 분류할 수 있었다. 번역 결과물을 논문 작성 시점에서 새로이 추출한 경우는 날짜를 표기하였다. 별도의 날짜 표기가 없는 데이터는 2018년 3월 중순에서 6월 초 사이에 추출되었다.

4. 연구결과

4.1 프리에디팅 기법: 영어-한국어

이번 절에서는 원천언어인 영어를 한국어로 번역할 때 유용한 기법들을 구글 번역과 파파고 두 경우를 비교하며 살펴보고자 한다. 영-한 번역의 프리에디팅에 대한 가이드라인을 제시하는 선행연구가 거의 전무한 상태이므로 연구방법에서 설명하였듯이 특정한 가정이나 원칙을 정하지 않고, 실제로 기계번역을 사용하여 번역을 수행하는 과정에서 ST의 내용을 충실히 반영하는 결과물 도출에 필요한 프리에디팅 기법을 만들어 나가는 것을 원칙으로 하였다. 이 절에서는 네 가지의 영-한 프리에디팅 기법을 제안한다.

8) 본고에서는 포스트에디팅 과정 및 결과는 다루지 않는다.

4.1.1 주절+종속절은 종속절+주절로 변환

복문에서 주절과 종속절이 차례로 오는 경우는 종속절을 주절 앞에 두는 것이 더 충실한 결과물을 낳았다. (1a)에서처럼 when 이하의 종속절이 주절 뒤에 나올 경우, 구글 번역은 (1b)에서 보듯이 종속절 전문이 누락 되었으나, 파파고는 구글 번역에 비해 개선된 결과물을 내었다(1c). (예문에 사용된 밑줄 등 편집 보조 도구의 의미는 각주8 참조)

- (1) a. ST: Life became a lot easier when I simply started being myself.⁹⁾
- b. TT(G): 삶이 훨씬 쉬워졌습니다. Ø
- c. TT(P): 제가 제 모습을 찾기 시작했을 때, 인생은 훨씬 쉬워졌습니다.

(1a)를 (2a)와 같이 종속절을 문두로 옮기면, 구글 번역(2b)은 종속절 번역이 이루어져 프리에디팅 전보다 개선된 번역을 보였고, 파파고(2c)는 유사한 품질의 결과물을 산출했다.

- (2) a. PE_ST: When I simply started being myself, life became a lot easier.
- b. PE_TT(G): 내가 단순히 자신이 되기 시작했을 때 인생은 훨씬 쉬워졌습니다.
- c. PE_TT(P): 내가 나 자신이 되기 시작했을 때, 삶은 훨씬 더 쉬워졌다.

복문의 또 다른 예인 (3)은 파파고의 빠른 개선을 보여주는 실례가 되겠다. (3a)는 주절과 종속절인 부사구로 이루어진 문장이다. (3b) 구글 번역 결과물에서는 종속절에 속한 **daring**이 주절의 **women**을 수식하여 번역 오류가 발생했다. 2018년 5월 1일에 접속한 파파고 결과물(3c)은 종속절의 번역에 역시 오류가 없다.

9) 예문에 사용된 편집 보조 도구의 의미는 다음과 같다. 밑줄() = 프리에디팅과 관련된 텍스트 일부; Ø = 원천텍스트에는 있으나 번역이 되지 않은 경우; **볼드와 이탤릭체** = 오역; **볼드** = 프리에디팅 후 개선된 텍스트

- (3) a. ST: We need to keep changing the attitude that permits the routine harassment of women, whether they're walking down the street or daring to go online.
- b. TT(G): **우리는 거리를 걸어가는 중이거나 온라인에서 대담한 여성의 일상적인 괴롭힘** 허용하는 태도를 계속 변경해야 합니다.
- c. TT(P)¹⁰⁾: **우리는 여성들이 길을 걷거나 인터넷에 접속하는 대담한 방법으로** 일상적인 괴롭힘을 당하는 것을 허락하는 태도를 계속 바꿔 나갈 필요가 있다.

(4a)에서는 (3a)의 영어 원천텍스트를 종속절을 문두로 보내 편집하였다. 그 결과 (4b)에서 볼 수 있는 바와 같이 구글 번역은 오류가 없을 뿐 아니라 비교적 다듬어진 번역물을 생산하였다.

- (4) a. PE_ST: Whether they're walking down the street or daring to go online, we need to keep changing the attitude that permits the routine harassment of women.
- b. PE_TT(G): 그들이 거리를 걸어가는 중이거나 온라인에서 대담하게 행동하든, 우리는 여성의 일상적인 괴롭힘을 허용하는 태도를 계속 변경해야 합니다.
- c. PE_TT(P): 그들이 거리를 걷고 있든, 인터넷에 접속하고 싶어 하든, 우리는 여성들의 일상적인 괴롭힘을 허용하는 태도를 계속해서 바꿀 필요가 있다.

정리하자면 종속절을 문두에 두는 이 기법은 구글 번역에 적용하여 충실한 번역을 생성한 반면, 파파고의 경우 프리에디팅 적용 없이도 충실한 번역을 보이는 경우와 프리에디팅 적용 후에도 정확하지 못한 번역을 도출해 내는 경우 고루 존재했다.¹¹⁾

10) 세 달 뒤인 7월 24일에 접속하였을 때는 프리에디팅된 ST의 번역오류가 개선된 결과물을 보여주었다. 세 달 사이에 개선된 파파고의 번역물은 기계번역이 얼마나 빠르게 개선되고 있는지를 시사한다. “우리는 여성들이 길을 걷고 있든, 인터넷에 접속하려고 하든 간에 일상적인 괴롭힘을 허용하는 태도를 계속 바꿀 필요가 있다.” (2018. 7. 24)

11) 주석 9의 결과를 반영하면 파파고는 기법적용 없이 충실한 번역을 생산하거나 기법

4.1.2 생략된 표현 명시화

두 번째 기법으로는 생략된 표현을 살려내어 반복을 통해 명시화해 주는 전략이다. 영어에서는 한 번 나온 표현은 문법이 허용하는 내에서 생략하는 것을 선호한다. 예를 들어 (5a)에는 and와 Muslim 사이에 that he wants to ban이 생략된 문장임을 알 수 있다. (5b)와 (5c)에서 볼 수 있듯이 구글 번역과 파파고 모두 서로 다른 두 가지 금지 내용(성전환자 입대 금지, 이슬람교인 입국 금지)을 바르게 번역하지 못한다.

- (5) a. Our commander in chief has said he wants to ban transgender people from our military and Muslim people from our nation.
 b. TT(G): 우리 사령관은 **우리 민족과 이슬람교도의 사람들로부터 트랜스젠더(transgender) 사람들을 금지하고 싶다고 말했습니다.**
 c. TT(P): 우리의 최고 사령관은 **우리나라의 군대와 이슬람 교인들에 대한 트랜스젠더들의 입국을 금지하고 싶다고 말했다**

(6a)에서 생략된 부분을 추가해서 에디팅했을 때 구글 번역(6b)은 ST의 내용을 충실히 반영한 번역을 한 반면, 파파고(6c)는 (5c)에 비해 개선은 되었으나 정확하지는 않았다¹²⁾.

- (6) a. PE_ST: Our commander in chief has said that he wants to ban transgender people from our military and that he wants to ban Muslim people from our nation.
 b. PE_TT(G): 우리의 수석 사령관은 **트랜스젠더(transgender) 인원을 우리 군대에서 금지하고 그가 이슬람교도를 우리 민족에게서 금지하고 싶다고 말했습니다.**
 c. PE_TT(P): 우리 최고 사령관은 **성 전환을 금지하고 싶고** 우리나라의 이슬람교도들을 금지하고 싶다고 말했다.

적용하여 충실한 번역을 생산하는 경우가 각각 절반이 된다.

- 12) 같은 해 10월에 프리에디팅된 ST를 파파고에 번역 수행시켰을 때 충실히 번역된 결과물을 얻을 수 있었다. 그러나 프리에디팅 되지 않은 ST는 종전의 결과물과 동일하였다. “우리의 최고 사령관은 그가 **우리 군대에서 트랜스젠더들을 금지하고 싶고** 우리 나라에서 이슬람인들을 금지하고 싶다고 말했다.” (2018. 10. 24)

또 다른 실례로, (8a)는 (7a)에서 생략된 I say that을 명시화하였다. 구글 번역은 생략된 단어나 문구에 대해 민감하게 반응한다. (7b)는 명백한 오역을 생산해 내었지만, 생략된 구문을 반복함으로써 명시화했을 경우 (8b)에서는 정확한 번역을 보여준다. 반면 파파고는 이러한 문구의 생략이나 반복에 따른 영향이 적었고 (7c)와 (8c)에서 동일하게 번역하였다. 하지만 ‘말합니다’를 ‘말입니다’로 번역하여 프리에디팅된 구글 번역에 비해 정확도는 떨어진다¹³⁾.

- (7) a. ST: And I say that not just as President but also as a feminist.
 b. TT(G): 그리고 나는 그것을 **대통령뿐만 아니라 페미니스트라고도 합니다**
 c. TT(P): 대통령으로서뿐만 아니라 페미니스트로서 **말입니다.**
- (8) a. PE_ST: And I say that not just as President but also I say that as a feminist.
 b. PE_TT(G): **대통령으로서뿐만 아니라 페미니스트로서도 그렇게 말합니다.**
 c. PE_TT(P): 대통령으로서뿐만 아니라 페미니스트로서 **말입니다.**

따라서 (5)-(8)에서 내린 결론으로는 파파고에는 생략된 것을 명시화하는 프리에디팅은 불필요한 것으로 보인다. (6)에서처럼 에디팅하여도 더 좋은 결과를 내지 못하거나, (8)처럼 프리에디팅이 필요하지 않은 번역을 생산하였다.¹⁴⁾

4.1.3 대시(-): 생략 후 별도 문장 생성

대시는 신문기사나 칼럼 등과 같은 가벼운 글쓰기에서 쉼표(,)나 콜론(:) 대신 빈번히 사용된다. 이러한 대시에는 삭제 후 별도의 문장을 생성하는 전략이

13) 10월에 파파고에 프리에디팅하지 않은 ST를 재가동하였을 때, 원문에 충실한 번역으로 개선되어 있었다. “그리고 나는 대통령으로서 뿐만 아니라 페미니스트로서도 **그렇게 말한다.**”(2018. 10. 24)

14) 파파고의 경우는 10월에 업데이트된 결과물을 기준으로 결론을 내리자면, 명시화 기법을 적용하여 개선된 경우와 명시화 기법 적용이 필요하지 않은 경우가 각각 절반이었다.

유용했다. (9)에서는 대시가 콜론과 같은 역할로 뒷문장이 앞에서 정의한 것이 무엇인지 부연 설명한다. 대시가 있는 문장 번역은 두 기계번역에서 현저한 차이를 보였는데, 구글 번역은 (9b)처럼 대시 이후의 문장 해석이 전혀 되지 않는다. 반면 파파고는 대시를 마침표로 끝난 문장과 같이 두 단문으로 번역하였다.

- (9) a. ST: In fact, the most important change may be the toughest of all —and that's changing ourselves.
 b. TT(G): 사실 가장 중요한 변화는 모든 것 중에서 가장 힘든 것일 수 있습니다. 0
 c. TT(P): 사실, 가장 중요한 변화는 아마도 가장 어려운 것일 것입니다. 그것은 바로 우리 자신을 변화시키는 것입니다.

(10a)과 같이 대시를 생략하고 두 문장으로 만들었을 때 뒷 문장의 해석이 제시되는 개선을 보여주었다. 그러나 **changing**을 동명사가 아닌 현재분사로 간주해 진행형으로 번역하는 오류를 보인다. 반면 (9c)와 (10c)에서 알 수 있듯 파파고는 원문의 대시의 사용 의도대로 정확히 번역함을 알 수 있다.

- (10) a. PE_ST: In fact, the most important change may be the toughest of all. And that's changing ourselves.
 b. PE_TT(G): 실제로, 가장 중요한 변화는 가장 힘든 것일 수 있습니다. 그리고 그것은 우리 자신을 변화시키고 있습니다.
 c. PE_TT(P): 사실, 가장 중요한 변화는 가장 어려운 것일 수도 있습니다. 그것은 바로 우리 자신을 변화시키는 것입니다.

대시에 대한 또 다른 편집 사례로 (11a)에서는 대시 이하에서 대시 이전에 나온 통근에 대한 정확한 거리를 설명하고 있다. 구글 번역(11b)은 **consumed**를 ‘인명을 잃었다’는 중대한 오역을 보여준다. 파파고에는 **as a state senator**(주 상원 의원으로) 번역이 누락되었다.

- (11) a. ST: For many years my life was consumed by long commutes—from my home in Chicago to Springfield, Illinois, as a state senator, and then to Washington, D.C, as a United States senator.

- b. TT(G): 수년 동안 시카고의 한 가정에서 일리노이 주 스프링필드에 이르기까지 주 상원 의원으로, 그리고 미국 상원 의원으로 워싱턴 DC에 오랜 통근으로 **인명을 잃었습니다**.
 c. TT(P): 수년 간 저는 시카고에 있는 제 집에서부터 0 일리노이 주 스프링필드, 그리고 미국 상원 의원으로서 워싱턴 DC까지 긴 통근 시간을 보냈습니다.

(12a)처럼 대시를 삭제하고 새 문장을 형성했을 때 구글 번역(12b)은 개선된 번역을 보여준다. 파파고에서도 두 문장으로 나누었을 때 (11c)에서는 생략되었던 부분이 (12c)에서는 바르게 번역되었다.

- (12) a. PE_ST: For many years my life was consumed by long commutes. My commutes were from my home in Chicago to Springfield, Illinois, as a state senator, and to Washington, D.C, as a United States senator.
 b. PE_TT(G): 오랫동안 내 인생은 오랫동안 **통근으로 소비되었습니다**. 나의 통근 통학은 시카고의 집에서 일리노이 주 스프링필드까지 주 상원 의원으로, 그리고 워싱턴 D.C에 미국 상원 의원으로 있었다.
 c. PE_TT(P): 여러 해 동안 나의 삶은 긴 통근에 의해 소비되었다. 저는 시카고에 있는 제 집에서 일리노이 주 스프링필드, **주 상원 의원으로서** 그리고 미국 상원 의원으로서 워싱턴 DC로 통근했습니다.

대시 관련 프리에디팅 기법은 구글 번역에서는 중요한 요인으로 작용함을 알 수 있다. 파파고는 대체로 대시유무에 민감하지 않으나 복잡한 문장의 경우 개선점을 보여주는 사례도 있었다.

4.1.4 직접화법을 유도하는 주어+동사는 직접화법 뒤에 위치

직접화법을 유도하는 주어와 동사가 앞쪽에 있을 때는 이를 직접화법 뒤로 옮겨주는 것이 번역품질을 개선시킨다. (13a)의 **President Barack Obama says**, 가 문두에 있을 때 구글 번역은 ‘말합니다’를 파파고는 직접화법 전체를 번역하

지 않는 오류를 보였다.

- (13) a. ST: President Barack Obama Says, “This Is What a Feminist Looks Like”
- b. TT(G): 버락 오바마 대통령은 “이것은 여권 신장론자의 모습입니다” ∅
- c. TT(P): 버락 오바마 대통령은 이렇게 말한다. ∅

(14a)에서처럼 이를 직접의문문 뒤로 보냈을 때 (14b)와 (14c)에서 볼 수 있듯이 오류 없이 번역 처리되었다.

- (14) a. PE_ST: “This Is What a Feminist Looks Like”, President Barack Obama says.
- b. PE_TT(G): 버락 오바마 대통령은 “이것은 여성주의자의 모습입니다”라고 말합니다.
- c. PE_TT(P): 버락 오바마 대통령은 “이것은 페미니스트의 모습이다”라고 말한다.

같은 맥락에 있는 (15)의 예에서 (13b)와 마찬가지로 구글(15b)은 ‘말했다’라는 서술어 번역을 하지 않았다. 반면 파파고는 (13c)와는 달리 (15c)와 (16c)에서 볼 수 있듯이 유도문의 위치와 상관없이 오류 없는 번역을 생성하였다.

- (15) a. ST: She once said, “The emotional, sexual, and psychological stereotyping of females begins when the doctor says, ‘It’s a girl.’”
 - b. TT(G): 그녀는 한 때 “여성의 정서적, 성적, 심리적 고정 관념은 의사가 ‘소녀’라고 말하면서 시작됩니다.” ∅
 - c. TT(P): 그녀는 “여성들에 대한 감정적, 성적, 심리적 고정 관념은 의사가 ‘여자 아이’라고 말할 때 시작됩니다.”라고 말했다.
- (16) a. PE_ST: “The emotional, sexual, and psychological stereotyping of females begins when the doctor says, ‘It’s a girl.’”, she once said.
 - b. PE_TT(G): “여성의 정서적, 성적, 심리적 고정 관념은 의사가 ‘소녀입니다.’라고 말하면서 시작됩니다.”라고 그녀는 말했습니다.

- c. PE_TT(P): “여성들에 대한 감정적, 성적, 심리적 고정 관념은 의사가 ‘여자 아이’라고 말할 때 시작됩니다.”라고 한때 그녀는 말했다.

직접화법에 대한 프리에디팅 기법도 두 기계번역에서 적용 차이를 보여주었다. 구글 번역은 이 기법을 적용하여 100% 번역품질이 개선되었으며, 파파고는 적용하여 결과물 품질이 개선되는 경우와 적용이 필요 없는 경우가 공존했다.

4.1.5 기타 기법

이 외의 다양한 제안들이 있었는데 크게 구조변화와 관련된 것과 의미 명시화를 위한 에디팅 전략으로 나눌 수 있다. 이는 모두 텍스트에서 그 용례가 2회 이하로 나타난 경우이다. 텍스트에서 나타난 빈도수에 따라 분류하였으므로 정확한 중요도를 가늠할 수 없는 한계가 있다. 텍스트 분량을 늘려서 조사를 실시할 경우 적용 빈도수에 따른 규칙화를 논할 수 있을 것이다. 이는 차후 연구 과제로 남겨둔다. 지면 관계상 예문 없이 기법만 나열하였다.

구조와 관련해서는 다음과 같은 일곱 가지의 제안이 있다. 첫째, 도치구문은 일반어순으로 바꿀 것, 둘째, 다소 길이가 긴 삽입구는 문두나 문미로 이동할 것. 셋째, 축약된 단어(예: you’re)는 풀어쓸 것 넷째, 접속사로 연결된 중문은 두 개의 단문으로 나눌 것. 다섯째, 주어가 무생물 추상명사인 경우는 There is/are N(주어) 구조로 전환할 것. 여섯째, 문두에 오는 긴 부사구는 문미로 이동할 것. 일곱째, 분사구문은 접속사가 있는 문장으로 전환할 것 등이다.

명시화를 위한 전략으로는 여섯 개의 제안이 있다. 먼저, 관계대명사 계속적 용법은 접속사+주어+동사로 전환할 것. 이는 기계번역이 쉽표를 인지하지 못하고 관계대명사 앞의 명사를 수식하는 형태로 번역하기 때문이다. 둘째, so..that 구문은 so..and therefore로 전환할 것. 이는 that의 다양한 용법으로 인해 so..that의 상관구문을 인식하지 못하는 것으로 짐작된다. 셋째, 소유격+명사는 소유격 뒤에 ‘own’을 삽입하여 명시화할 것. 넷째, 관용적 표현은 홀따옴표를 추가할 것. 다섯째, ‘only to+동사’는 ‘so eventually become’으로 명시화할 것. 여섯째, 다의어는 하위어로 구체화할 것 등이 있다.

이상의 제안들은 구글 번역과 파파고 중 2회 이하로 나타난 현상들이라 두 기계번역상의 차이점을 고찰하지 못하는 한계가 있었다. 이 또한 후속 연구과

제로 남겨놓는다.

4.2 프리에디팅 기법: 한국어-영어

이번 절에서는 한국어를 원천언어로 하는 텍스트를 영어로 번역할 경우 사용되는 기법을 살펴보고자 한다. 한국어 원천언어의 프리에디팅은 선행연구에서 다루어진 바가 있다. 그 중 김순미(2017)과 박옥수(2017)가 공통으로 제안한 것을 살펴보면, 주어의 명시화, 의미해석이 되지 않는 경우¹⁵⁾, 길고 복잡한 문장 구조 변화 등이 있다¹⁶⁾. 본 연구에서는 위의 선행연구를 기반으로 크게 세 개의 기법으로 분류하여 살펴보고자 한다. 첫째, 생략된 주어의 명시화. 둘째, 의미의 명확화. 셋째, 구조의 단순화 전략 등으로 나누어 구글 번역과 네이버의 파파고를 비교·고찰한다.

4.2.1 생략된 주어의 명시화

고맥락(high-context) 언어인 한국어의 특성상 주어의 생략이 잦고, 이는 주어가 반드시 명시되는 영어로 번역할 때 장애물로 여겨진다. 하지만 선행연구의 결과나 일반적인 예상과는 달리 주어의 존재 여부가 기계번역의 결과에 영향을 미치는 정도에는 문장에 따라 차이가 있었다.

우선 주어의 부재로 번역 오류가 생기는 경우를 살펴보자. (17a)의 마지막 문장인 중문에서는 앞에서 설명한 남성이 각 단문의 주어가 된다. 밑줄 부분은 생략된 주어의 서술어이다. 이 문장의 경우는 추상적으로 맥락을 알지 못하고는 주체가 누구인지 짐작할 수 없는 경우이다. 예측 가능한 대로 구글 번역(17b)과 파파고(17c) 모두 번역 오류가 나타났다.

15) 김순미(2017)는 이를 세분화하여 의미의 모호함, 은유적 표현, 복잡하거나 어려운 표현 등 세 가지로 나누어 분류하였다.

16) 이 외에도 김순미(2017)에서는 텍스트를 전면적으로 재구성한 사례를, 박옥수(2017)는 맥락의 이해가 어려운 문장의 사례를 들고 있으나, 본 논문에서는 다음과 같은 이유로 두 경우를 다루지 않는다. 첫째, 텍스트의 특성상(칼럼 및 기사) 내용의 재구성은 논의 밖이다. 둘째, 단편 문장이 아닌 칼럼과 기사 전문을 번역하였으므로 맥락이 문제가 되지 않는다.

(17) a. ST: [‘미투에서 자유로운 남성’이란, 성폭력 가해를 하지 않는 남성뿐만 아니라 이미 자신의 잘못을 깨닫고 피해자에게 사과하고 반성한 남성이다.] 당연히 **Ø** 있어야 하고 또 **Ø** 있다.

b. TT(G): *Of course there is and there is another.*

c. TT(P): *Of course there is and there is.*

(18a)에서 볼 수 있듯이 각 단문에 주어를 삽입하면 이상적인 번역 결과물을 얻을 수 있다.

(18) a. PE_ST: 당연히 그러한 남성들이 있어야 하고, 또 그러한 남성이 있다¹⁷⁾.

b. PE_TT(G): *Of course there must be such men, and there are such men.*

c. PE_TT(P): *Of course there should be such men and there are such men.*

아래의 예에서는 주어의 존재 여부가 두 기계번역에서 다르게 반응하는 경우를 살펴보자. (19a)의 복문에서 종속절에 일반주어에 해당되는 ‘사람들은’이 빠져있다. (19b)의 구글 번역은 주어를 ‘I’로 하는 오역을 보여주었고, 반면 (19c)의 파파고는 종속절을 분사구문으로 처리하여 주절의 주어와 의미가 연결되게 올바르게 번역하였다.

(19) a. ST: **Ø** 전기밥솥 하나 살 때도 사용설명서 보고 열심히¹⁸⁾ 공부하면서, 폭력이라는 단어 하나의 개념도 모르고 있는 사람이 얼마나 많은가.

17) 중문의 쉽표(,) 사용에 대해 구글 번역과 파파고의 민감성에 차이가 있다. 구글 번역은 단문 사이에 쉽표가 없을 경우 다음과 같은 오역을 생산했다.

PE_ST: ‘당연히 그러한 남성들이 있어야 하고 또 그러한 남성이 있다.’

PE_TT(G): ‘Of course there must be such men and such men.’

따라서 접속사 뒤 쉽표는 필수적이며, 영어 결과물에도 쉽표가 있다. 반면 파파고는 쉽표의 여부와 상관없이 동일한 결과물을 내며 영어 문장에도 쉽표가 없다.

18) 원문은 ‘깡깡대며’로 되어있으나 사용 빈도수가 높은 ‘열심히’로 바꾸는 의미 명확화 전략을 사용하여 프리에디팅 한 후 기계번역을 수행하였다.

- b. TT(G): *When I buy an electric rice cooker*, while studying hard to see the manual, there are many people who do not even know the concept of one word of sexual violence.
- c. TT(P): How many people are unaware of the concept of sexual violence while studying hard through instructions even **when purchasing** an electric rice cooker?

(20a)에 종속절 문두에 주어 ‘사람들은’을 삽입하였을 때 구글 번역(20b)는 충실한 결과물을 낸다. 파과고(20c)는 주어가 없었던 (19c)와 비교할 때 오히려 오류는 없으나, 원문에서는 종속절에 해당하는 문장을 주절의 주어로 표현함으로써 덜 효과적인 문장을 만들었다.

- (20) a. PE_ST: 사람들은 전기밥솥 하나 살 때도 사용설명서 보고 열심히 공부하면서, 성폭력이라는 단어 하나의 개념도 모르고 있는 사람이 얼마나 많은가.
- b. PE_TT(G): **When people buy an electric rice cooker** and study hard to see the manual, how many people do not even know the concept of one word of sexual violence.
- c. PE_TT(P): When people buy an electric rice cooker, they study hard using instructions, and do not know the concept of sexual violence.

주어의 생략이 두 기계번역 모두에게 문제가 되지 않는 경우도 자주 나타났다. 이 경우는 주로 주어가 일반적인 경우에 해당되었다. (21)에서 볼 수 있듯이 ‘그것은’에 해당되는 주어의 생략은 ‘it is’ 혹은 ‘that is’ 등으로 잘 대체 되는 것을 알 수 있다.

- (21) a. ST: **Ø** 저질 담론 투쟁이다. **Ø** 새로운 말이 가리키는 새로운 현상을 제대로 이해하고 올바르게 사용하기 위한 노력이 언론과 독자의 의무가 되는 이유다.
- b. TT(G): **It is a struggle** for low quality discourse. **That is why** the press and readers are obliged to understand the new phenomena that the new words point to and to use them properly.

- c. TT(P): **It is a struggle** for low discourse. **That is why** efforts to fully understand and use the new word's new phenomenon are the duty of the press and readers.

요약하면, 주어의 명시화 필요성은 문장의 의미 명료화의 정도와 기계번역 종류에 따라 다르게 나타났다. 추상적인 문장에서 생략된 주어는 명시화를 통해 구글 번역과 파과고에서 개선된 번역을 얻을 수 있었다. 복문의 종속절에 주어가 생략된 경우는 구글 번역에서는 명시화가 요구되었으나 파과고에서는 분사구문으로 처리함으로써 주어 명시화가 필요 없었다. 일반적인 뜻을 가지는 주어는 두 기계번역 모두에서 명시화 없이도 대명사로 처리되었다.

4.2.2 의미의 명확화 기법: 복합명사 풀어쓰기

기계번역은 텍스트의 의미가 모호할 경우 오역을 생산하는 결과를 낳는다. 세 개 이상의 명사로 구성된 복합명사가 그러한 경우이며, 전치사 등을 덧붙여 풀어쓰서 명확한 의미를 전달할 필요가 있다. 아래의 (22a)에 제시된 복합명사 ‘경제활동 참가율’은 그 단어 자체보다는 전체 문장 구조 파악을 어렵게 만들어 오역을 만들어 낸다. (22b)는 복합명사 번역은 괜찮으나 수치 변화를 반대로 기술하였으며, (22c)는 복합명사를 풀어쓰 표현하였으나 오류가 있다.

- (22) a. ST: 2017년 탈북민의 경제활동 참여율은 61.2%로 전년(57.9%) 보다 높아졌지만 역시 일반 국민(63.3%)보다는 낮은 수치를 보였다.
- b. TT(G): In 2017, the labor force participation rate of the refugees **increased from 61.9% to 57.9%** in the previous year, but was also lower than that of the general public (63.3%).
- c. TT(P): North Korean defectors **participated in 61.2 percent of their economic activities** in 2017, up from 57.9 percent last year, but the figure was also lower than that of the general population (63.3 percent).

아래 (23)에서 복합명사 ‘경제활동 참가율’을 전치사를 사용하여 ‘경제활동에 대한 참여율’로 바꾸면, 문장의 구조가 완벽한 번역 결과물을 도출한다.

- (23) a. PE_ST: 2017년 탈북민의 경제활동에 대한 참여율은 61.2%로 전년(57.9%)보다 높아졌지만 역시 일반 국민(63.3%)보다는 낮은 수치를 보였다.
- b. PE_TT(G): In 2017, the participation rate of the refugees in economic activities was 61.2%, which is higher than the previous year (57.9%) but also lower than that of ordinary people (63.3%).
- c. PE_TT(P): In 2017, 61.2 percent of North Korean defectors participated in economic activities, up from 57.9 percent last year, but the figure was also lower than that of the general population (63.3 percent).

복합명사 및 또 다른 명사를 나열한 조합을 조사와 관형사를 사용하여 풀어쓰면 더 나은 결과를 가져온다. (24a)에서 연결고리 없이 명사와 복합명사가 나열된 ‘재단 소속 탈북민 전문상담사’는 (24b)의 구글 번역에서는 ‘탈북민’이 번역에서 누락되었으며 (24c)의 파파고에서는 ‘재단 소속’ 번역이 누락되었을 뿐 아니라 ‘북한(인)상담사(North Korean counselor)’로 오역되었다.

- (24) a. ST: 이번 실태조사는 지난해 6월 7일부터 8월 18일까지 재단 소속 탈북민 전문상담사가 탈북민을 방문조사하는 방식으로 이뤄졌다.
- b. TT(G): The survey was conducted from June 7 to August 18 last year by a *specialist counselor from the Foundation who visited the North Korean defectors*.
- c. TT(P): The investigation was conducted by a *North Korean counselor* from June 7 to August 18 of last year who visited North Korean defectors.

이를 (25a)에서는 조사와 관형사 등을 사용하여 풀어썼다. 구글 번역과 파파고 모두 충실한 번역을 수행한 것을 볼 수 있다.

- (25) a. PE_ST: 이번 실태조사는 작년 6월 7일부터 8월 18일까지 재단에 소속된 탈북민을 전담하는 상담사가 탈북민을 방문조사하는 방식으로 이뤄졌다.

- b. PE_TT(G): This survey was conducted by a **counselor dedicated to the refugees who belonged to the Foundation** from June 7 to August 18 last year to investigate the defectors. (2018. 7. 20)
- c. PE_TT(P): The survey was conducted by a **consultant specializing in North Korean defectors who were affiliated with the foundation** from June 7 to August 18 last year to examine North Korean defectors. (2018. 7. 20)

복합명사 및 나열된 명사를 풀어쓰는 기법은 의미를 명확하게 해주며 이는 구글 번역과 파파고에 적용하여 번역품질을 향상시키는 결과를 가져왔다.

4.2.3 문장 구조의 단순화

긴 문장의 번역은 기계번역뿐 아니라 인간 번역사에게도 시간이 걸리는 부담되는 구조이다. 기계번역을 위해서 이러한 긴 문장은 두 개 이상의 문장으로 나눈다. (26a)의 복합문은 긴 문장으로 인해 독해가 용이하지가 않다. (26b)의 구글 번역은 *미투에서... 사라져버리고*에 이르는 두 줄의 글이 번역에서 누락되었다. (26c)의 파파고는 번역누락이 극적으로 나타났다.

- (26) a. ST: 권력자의 은폐된 성폭력을 드러낸 용기 있는 여성들을 따라 사회에 만연한 성폭력을 공론화하고자 하는 운동을 가리키던 이 말이, “미투에서 자유로운 남성이 어디 있나”라는 식으로 남성 물귀신의 언어가 되면서 성폭력의 공론화라는 의미가 슬그머니 사라져버리고 마치¹⁹⁾ 여성이 폭로하면 당할 수밖에 없다는 식으로 왜곡된다.
- b. TT(G): This movement, which refers to the movement to publicize sexual violence in society along with courageous women who reveal the hidden sexual violence of the power, *is the expression of “sexual freedom”* \emptyset It is distorted by the way that it disappears sadly and it can not be helped if a woman exposes it. (2018. 7. 20)
- c. TT(P): What kind of man is this? (2018. 7. 20)

19) 원문에서는 ‘흡사’로 되어있으나 빈도수가 높은 단어인 ‘마치’로 에디팅하여 의미를 명확하게 하였다.

(27a)에서는 (26a)를 세 개의 문장으로 나누고 문장 사이에 적절한 연결사를 사용하였다. 이러한 단순한 프리에디팅은 (27b)와 (27c)에서 볼 수 있듯이 번역품질이 혁신적으로 개선되는 결과를 가져온다.

- (27) a. PE_ST: 20)권력자의 은폐된 성폭력을 드러낸 용기있는 여성들을 따라 사회에 만연한 성폭력을 공론화하고자 하는 운동을 가리킨다. 하지만 이 말이, “미투에서 자유로운 남성이 어디 있나”라는 식으로 남성 물귀신의 언어가 되면서 성폭력의 공론화라는 의미가 슬그머니 사라져버린다. 그리고 마치 여성이 폭로하면 남성은 당할 수밖에 없다는 식으로 왜곡된다.
- b. PE_TT(G): It refers to the movement to publicize sexual violence that is prevalent in society along with courageous women who reveal concealed sexual violence by the power. **But this means, “Where is the free man in Mitu”, the expression of sexual violence as the language of the male empire is disappearing meaninglessly.** And as if a woman is exposed, it is distorted by the way men are forced to suffer. (2018. 7. 20)
- c. PE_TT(P): It refers to a campaign to make sexual violence prevalent in society public along with courageous women who reveal the hidden sexual harassment of those in power. However, when this phrase becomes the language of the male disappearance in “Where is a free man from Mito,” the meaning of sexual assault disappears. And if a woman exposes a man, he can not help but suffer. (2018. 7. 20)

또 다른 용례로 긴 문장의 복문은 연결사를 사용하여 두 개의 단문으로 나누는 기법을 들 수 있다. (28a)는 종속절로 부사절을 갖는 복문이다. (28b)의 구글 번역은 주절에 대한 번역을 전혀 생성하지 않았다. (28c)의 파파고는 구글에 비해 내용면에서 좀 더 충실한 번역 결과를 보였으나 종속접속사가 없으며 종속절이 비문인 문장을 생성했다.

- (28) a. ST: 1990년대 들어 예전의 여성해방투쟁과는 결이 다른 페미니즘 운동이 생겨나면서, 대학가를 중심으로 21)성폭력반대 운동이 거세게 일어났다
- b. TT(G): In the 990s, a feminist movement was created that was different from the old women's liberation struggle. **Ø** (2018. 7. 20)
- c. TT(P): In the 1990s, **Ø** the former women's liberation of the struggle **be different** from feminism movements have sprung up around college town, the fight against sexual violence happens. (2018. 7. 20)

(29a)는 복문을 두 개의 단문으로 나누고 접속부사로 연결하였다. (29b)의 구글 번역은 누락되었던 주절을 바르게 생성하였고, (29c)의 파파고의 번역은 개선된 결과물을 보여주었으나 첫 번째 문장이 여전히 비문으로 남아있었다.

- (29) a. PE_ST: 1990년대 들어 예전의 여성해방투쟁과는 결이 다른 페미니즘 운동이 생겨났다. 이에 따라 대학가를 중심으로 성폭력반대 운동이 거세게 일어났다. (2018. 7. 20)
- b. PE_TT(G): In the 1990s, a feminist movement emerged that was different from the old women 's liberation struggle. **As a result, the movement against sexual violence has been taking place in university camps.** (2018. 7. 20)
- c. PE_TT(P): In the 1990s, the former women's liberation of the struggle **be different** from feminism movements have sprung up. **As a result, there has been a strong movement against sexual violence in the college districts.** (2018. 7. 20)

문장 구조를 단순하게 만드는 기법은 구글 번역과 네이버의 파파고 두 기계번역에 공통적으로 적용되는 유용한 기법을 알 수 있다.

20) 생략된 주어인 ‘이 말은’의 추가 여부는 번역에 영향을 미치지 않았다.

21) 원문에는 ‘반성폭력 운동’이나 의미 명확화를 위해 프리에디팅하였다.

5. 결론

본 연구는 기계번역을 효율적으로 활용하고 그 결과물의 품질을 높이기 위해 텍스트를 프리에디팅하는 기법을 연구·제안하였다. 기사문, 칼럼, 연설문 원문 텍스트에 대하여 기계번역을 실시한 결과 구글 번역과 네이버의 파파고를 사용할 때 유용하게 활용될 수 있는 프리에디팅 기법은 원천어별로 다음과 같이 정리할 수 있다. 영어가 원천어일 경우 첫째, 주절+종속절로 이루어진 복문은 종속절을 문두로 옮긴다. 둘째, 영어 원문에서 중복된 표현이 생략된 경우 생략된 표현을 반복함으로써 명시화해 준다. 셋째, 대시(-)가 포함된 문장은 대시를 기준으로 두 개의 문장으로 분리한다. 넷째, 직접화법을 유도하는 주어+동사는 직접화법 뒤로 배치시킨다. 이 밖에도 다수의 기타 기법들이 예비 기법으로 도출되었다. 한국어가 원천어일 경우 활용될 수 있는 프리에디팅 기법으로는 첫째, 생략된 주어의 명시화, 둘째, 복합명사 풀어쓰기를 통한 의미 명확화, 셋째, 문장 구조의 단순화로 요약될 수 있었다. 도출된 기법들은 일견 단순히 보일 수 있으나, 이러한 기법들은 인간 사용자가 쉽게 숙지하여 기계번역을 활용하기 위한 것이므로 그 효용가치는 충분하다고 판단된다.

기계번역의 효율적인 활용을 위해서는 언어 조합에 따른 기계번역 종류의 선택과 텍스트 종류에 따른 기계번역 종류의 선택을 아는 것이 중요하겠다. 우선 언어 조합에 따른 기계번역의 선택을 고려해 보면, 영·한 언어쌍의 경우는 파파고가 구글 번역에 비해 텍스트 종류와 관계없이 프리에디팅의 필요성이 현저히 낮아 시간을 적게 들이고 좋은 결과물을 얻을 수 있었다. 한·영의 경우는 텍스트의 영역에 따라 파파고와 구글 번역의 선택을 고려해 볼 필요가 있다. 본 연구에서 살펴본 바로는 파파고와 구글 번역은 각각의 장점과 단점이 있다. 파파고의 경우 구글 번역에 비해 문장 구조를 잘 읽어내는 것으로 나타났다. 따라서 칼럼과 연설문과 같이 독자를 설득하기 위해 평이한 어휘를 사용하면서도 한편으로는 문장이 복잡한 구조를 보여주는 설득적·표현적 텍스트에 적절한 것으로 판단된다. 구글 번역의 경우, 파파고에 비해 복잡한 문장구조를 파악하는 데는 약한 반면에 전문적이고 난이도 높은 어휘를 적절히 사용하였다. 따라서 사회현상을 기술하고 설명하는 보도문 형식인 기사 번역에는 구글 번역 사용을 권장한다. 기사의 문장구조는 비교적 단순하나 전문적 어휘가 요구되는 까닭이다.

신경망 번역 시스템의 도입 이후 기계번역의 품질은 빠른 속도로 발전하고 있다. 연구결과에서 볼 수 있었듯이 특히 네이버의 파파고는 빠르게 그 품질을 향상시켜 나가고 있어, 불과 몇 개월 전에 보였던 번역 오류를 개선하는 것을 볼 수 있었다. 이에 인간 번역사는 기계번역의 효용성을 받아들이고 효율적으로 활용한다면 쉽고 빠르게 그러면서도 좋은 결과물을 얻을 수 있을 것이다.

본 연구는 제한된 분석텍스트의 양으로 인해 동일한 패턴을 충분히 추출할 수가 없었던 한계가 있다. 동일한 패턴의 문장에서 같은 프리에디팅 기법의 적용률이 얼마나 높은지 밝히는 것이 연구결과의 타당성을 높여줄 것이다. 따라서 코퍼스를 활용하여 같은 패턴의 문장을 다량 수집하여 동일한 프리에디팅 기법을 적용하여 결과에 따라 기법을 규범화하는 과정이 필요할 것이다. 또한 본고에서 제안한 프리에디팅 기법은 기계번역 원자료(raw MT output)의 품질을 개선하는데 유용하므로 프리에디팅과 포스트에디팅을 병행하여 최종 번역물의 품질이 얼마나 효율적이고 효과적으로 향상되었는지에 대해서는 추가적인 논의가 필요할 것이다. 이는 후속 과제로 남겨두기로 한다.

참고문헌

- 곽중철·한승희 (2018) 「포스트에디팅 측정지표를 통한 기계번역 오류 유형화 연구」, 『통번역학연구』 22(1): 1-25.
- 김순미 (2017) 「신경망번역기(NMT) 활용 학부 번역교육의 가능성 연구」, 『통번역교육연구』 15(3): 5-37.
- 김준석 (2017) 「일상생활 속으로 들어온 기계 번역」, 『새국어생활』 27(4): 63-79.
- 마승혜 (2018a) 「한영 기계번역 포스트 에디팅에 대한 경험적 고찰: 학부 교육 과정 및 결과를 중심으로」, 『통번역학연구』 22(1): 53-88.
- 마승혜 (2018b) 「기계번역의 외연 확대 - 신문기사 특정 은유표현 포스트 에디팅 전략 모색」, 『번역학연구』 19(2): 117-144.
- 박영란 (2018) 「기계 번역에서의 영어속담 분석」, 『통번역교육연구』 16(1): 131-150.

- 박옥수 (2016) 「한영 병렬 코퍼스과 기계번역에서 의존 명사: ‘-것’이 포함된 어휘의 번역 방식 연구」, 『동아인문학』 37: 469-492.
- 박옥수 (2017) 「한영 기계 번역에서 ST의 유형적 특징에 따른 번역 오류 분석」, 『동아인문학』 41: 155-183.
- 박옥수 (2018) 「원천언어의 관점에서 살펴본 기계번역의 오류분석과 수정 원칙: 비문학 텍스트의 통사적 특징에 근거해서」, 『동아인문학』 44: 151-171.
- 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.
- 이상빈 (2017) 「학부번역전공자의 기계번역 포스트에디팅, 무엇이 문제이고, 무엇을 가르쳐야 하는가?」, 『통역과 번역』 14(5): 39-66.
- 이상빈 (2018a) 「학부 번역전공자의 구글 기계번역 포스트에디팅에 관한 현상학 연구」, 『통번역학연구』 22(1): 117-143.
- 이상빈 (2018b) 「국내학부 번역전공자의 구글 기계번역 포스트에디팅 과정(process) 및 행위 연구」, 『번역학연구』 19(3): 259-286.
- 이준호 (2018) 「포스트에디팅 교육을 위한 포스트에디팅과 인간번역의 차이 연구」, 『통역과 번역』 20(1): 73-96.
- 임순정 (2012) 「구글번역의 사례연구 - 번역품질과 번역단위와의 관계를 중심으로」, 『번역학연구』 15(1): 177-209.
- 임순정·한미선 (2014) 「구글번역의 사례연구 - 번역품질과 번역단위와의 관계를 중심으로」, 『번역학연구』 15(1): 177-209.
- 조수연·박경리·전종섭 (2013) 「타동성 교체 구문을 활용한 구글 번역기의 영한 번역 평가」, 『언어와언어학』 61: 111-140.
- 진용주·서보현·김순영 (2018) 「프리에디팅(pre-editing)이 기계번역 품질에 미치는 영향 고찰: 유창성, 충실성, 가독성 비교를 중심으로」, 『통번역학연구』 22(3): 27-64.
- 최동익 (2013) 「무생물 주어 구문에 대한 영한 기계 번역 오류분석」, 『언어학연구』 29: 279-299.
- 최수연 (2017) 「번역학의 ‘기술적 전환’(Technological Turn): AI시대의 영상번역」, 『번역학연구』 18(2): 207-228.
- 최효은·이지은 (2017) 「특허 기계번역 결과물의 평가-KIPRIS의 무료 한영 기계번역을 중심으로」, 『통역과 번역』 19(1): 139-178.
- 한승희 (2017) 「기계번역·인간번역·트랜스크리에이션의 문체 비교」, 『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.
- Aikawa, Takako, Lee Schwartz, Ronit King, Monica Corston-Oliver & Carmen Lozano (2007) ‘Impact of Controlled Language on Translation Quality and Post-editing in Statistical Machine Translation Environment’, *Proceedings of the MT Summit XI*: 1-7.
- Bisazza, Arianna & Marcello Federico (2012) ‘Modified distortion matrices for phrase-based statistical machine translation’, *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. Association for Computational Linguistics*: 478-487.
- Forcada, Mikel L. (2010) ‘Machine translation today’, in Yves Gambier and Luc Van Doorslaer (eds) *Handbook of translation studies* (Vol. 1). Amsterdam: John Benjamins, 215-223.
- Gerlach, Johanna, Victoria Porro, Pierrette Bouillon & Sabine Lehmann (2013) ‘Combining Pre-editing and Post-editing to Improve SMT of User-generated Content’, *Proceedings of MT Summit XIV Workshop on Post-editing Technology and Practice*: 45-53.
- Han, Bo, Paul Cook, & Timothy Baldwin (2013) ‘Lexical normalization for social media text.’ *ACM Transactions on Intelligent Systems and Technology (TIST)*, 4(1): 5.
- Hutchison, John & Harold Somers (1992) *An Introduction to Machine Translation*. London: Academic Press.

- Hutchison, John (2003) 'The development and use of machine translation systems and computer-based translation tools', *International journal of translation*, 15(1): 5-26.
- Jin, Sil-hee (2017) 'Limitations of Machine Translation: Focuese on the Role of Cognitive Complements in Human Communication', *통번역학연구* 21(4): 167-199.
- Krings, Hans P. (2001) *Repairing Texts: Empirical Investigations of Machine Translation Post-editing processes* Vol. 5. Kent State UP.
- Koponen, Maarit (2016) 'Is machine translation post-editing worth the effort? A survey of research into post-editing and effort', *JoSTrans* 25: 131-148.
- O'Brien, Sharon (2006) 'Controlled Language and Post-Editing', *Multilingual* October/November Issue: 17-19.
- O'Brien, Sharon & Johann Roturier (2007) 'How portable are controlled language rules? A comparison of two empirical MT studies', *Proceedings of MT summit XI*: 345-352.
- Way, Andy & Nano Gough (2005) 'Comparing example-based and statistical machine translation', *Natural Language Engineering* 11(3): 295-309.

<인터넷 자료>

- 김성수 (2017. 5. 31) 「AI 통번역의 진화」, 『컴퓨터월드』, from <http://www.comworld.co.kr/news/articleView.html?idxno=49216> (접속일자: 2018년 7월 31일)
- Densmer, Lee (2015) 'Machine Translation Pre-Editing to Boost Output Quality', from <http://info.moravia.com/blog/machine-translation-pre-editing-to-boost-output-quality> (접속일자: 2018년 6월 12일)
- Thicke, Lori (2013) 'The Industrial Process for Quality Machine Translation', in *JoSTrans*, Issue 19, from http://www.jostrans.org/issue19/art_thicke.php#backto_note_1 (접속일자: 2018년 7월 1일)
- Pre-editing & Post-editing (제네바 대학교 통번역학부): <https://www.unige.ch/fti/en/faculte/departements/dtim/recherches/ta/> (접속일

자: 2018년 7월 17일)

<https://news.un.org/en/story/2018/02/1002441>

<https://www.nytimes.com/2017/11/11/opinion/sunday/sick-of-racism-literally.html>

<https://www.glamour.com/story/glamour-exclusive-president-barack-obama-says-this-is-what-a-feminist-looks-like>

<http://newspeppermint.com/2017/07/18/worldpopulation/>

[Abstract]

**Pre-editing Rules to Enhance Output Quality of Machine Translation:
English-Korean and Korean-English**

Lee, Sunghwa · Kim, Sehyoun
(Pusan National University)

This paper studies pre-editing rules for English-Korean and Korean-English language sets when using machine translation. A total of 2,650 words and 1,253 words of English and Korean texts published in newspapers and a magazine were examined by entering them into Google Translate and Naver Papago. Four rules to perform efficient and consistent pre-editing for English-Korean text are proposed as follows: (1) Move a subordinate clause to the beginning of the sentence when a main clause is followed by the subordinate clause. (2) Retrieve the omitted expressions from the original English text. (3) Make two sentences out of a sentence that contains a dash (—). (4) Place a subject + verb after a direct speech. Also, three pre-editing rules are suggested for Korean-English: (1) Retrieve omitted subjects. (2) Clarify the meaning of compound nouns. (3) Simplify the sentence structure. This study also compares Google Translate and Naver Papago. Results show that as for the English-Korean language set, Papago requires less pre-editing compared to Google Translate. In the case of Korean-English, on the one hand, Naver Papago seems to be more appropriate for persuasive texts that use plain vocabulary but show complicated structures. On the other hand, Google Translate seems to be proper for news of social context where the vocabulary selection is formal and professional.

▶ Key Words: machine translation, pre-editing, Google Translate, Papago, English-Korean, Korean-English

▶ 주제어: 기계번역, 프리에디팅, 구글 번역, 파파고, 영·한, 한·영

이성화(제1저자)

부산대학교 글로벌 영상번역 창의인재 교육-연구 혁신 사업단 초빙교수
esunghwa@pusan.ac.kr

관심분야: 영상번역, 형태론, 연구방법론

김세현(제2저자)

부산대학교 영어영문학과 박사과정
sehyoun@pusan.ac.kr

관심분야: 영상번역, 번역과 젠더, 기계번역

논문투고일: 2018년 11월 4일

심사완료일: 2018년 11월 23일

게재확정일: 2018년 12월 5일