번역학연구
2018 봄 제19권1호

기계번역 결과물의 오류유형 고찰

서보현 · 김순영*
(동국대-서울)

1. 서론

근래 들어 '기계번역, 인간번역' 이라는 말들을 빈번하게 접한다. 국립국어원 표준국어대사전의 정의에 따르면 번역은 '어떤 언어로 된 글을 다른 언어의 글로 옮김'이다. 이 정의에 행위의 주체가 명시적으로 드러나 있지는 않지만 일반적으로 행위자가 사람인 경우, 요즘의 용어로 하자면 인간번역을 의미하는 것으로 받아들여져 왔다. 이제는 번역행위를 수행하는 주체가 기계인지 인간인지를 명시적으로 밝히는 것이 당연하게 여겨질 만큼 기계번역에 대한 관심이 확대되고 있다. 이러한 관심은 최근 들어 활발하게 이루어지고 있는 연구들(김순미(2017), 마승혜(2017), 신지선·김은미(2017), 장애리(2017), 조재범(2017) 등)을 통해서도 확인할 수 있다. 하지만 기계번역이라는 개념이 최근에 등장한 것은 아니다.

* 교신저자, 동국대학교 영어통번역학과 교수, 전자우편: kimsy@dongguk.edu

기계번역의 역사는 1950년대부터 시작된다(신호섭 2017: 28-37). 기계번역은 원문과 번역문의 구문을 분석하여 만든 문법적 알고리즘을 활용하는 규칙기반 기계번역(Rule-Based Machine Translation: RBMT)에서 출발하였다. 규칙기반 기계번역은 문법적 요소를 기반으로 하기에 오류 패턴이 일정하고 수정이 간단하다는 장점이 있었으나 끊임없이 변화하는 언어에 맞추기 위해 문법 규칙을 지속적으로 수정 및 추가해주어야 했으며 해당 작업에 적지 않은 시간이 들어간다는 단점이 있었다. 규칙기반 기계번역의 한계를 극복하기 위해 개발된 것이 통계기반 기계번역(Statistical Machine Translation: SMT)이다. 통계기반 기계번역은 원문과 번역문의 조합으로 이루어진 말뭉치를 사용하여 두 텍스트의 관계를 통계적으로 분석한 후 해당 결과를 적용하는 방식이다. 많은 양의 말뭉치를 활용하기에 기계번역 알고리즘을 확립하는데 많은 시간이 걸리지만 기술의 발전으로 빅 데이터 운용이 용이해짐에 따라 통계기반 기계번역은 빠르게 발전하게 된다. 하지만 텍스트의 맥락을 고려하지 못하므로 출발어와 도착어의 문법구조 및 어순이 다른 경우에는 품질이 크게 저하된다는 단점 또한 존재한다. 2010년대에 들어서 컴퓨터 하드웨어의 성능이 향상되면서 컴퓨터에 인간의 신경망과 비슷한 구조의 네트워크를 만들어 활용하는 신경망 이론을 구현할 수 있는 길이 열렸다. 이런 신경망 기술을 기계번역에 적용한 것이 신경망 기계번역(Neural Machine Translation: NMT)이다.

신경망 기계번역이 등장하면서, 여전히 만족할만한 수준이 아니라고는 하지만 기계번역 결과물의 품질은 비약적으로 향상되고 있고 앞으로도 발전 속도는 더욱 빨라질 것으로 보인다. 기계번역의 발전을 바라보는 우리의 시선은 대량의 번역을 신속하게 처리해줌으로써 업무 효율성과 경제성을 높여줄 것에 대한 기대감과 동시에 머지않아 기계가 인간을 대체할지도 모른다는 불안감이 혼재되어 있다. 하지만 신호섭(2017: 28)이 언급하고 있듯이 인간의 역할을 빼앗아 갈 경쟁자로만 바라본다면 “새로운 기술을 인간의 동반자로 맞이할 수 있는 기회”를 잃어버리게 될 수도 있다.

인간과 기계의 협력이라는 관점에서 기계번역 결과물을 수정 및 개선하여 활용도를 높이기 위한 방안으로 대표적인 것이 포스트에디팅 작업이다. 포스트에디팅은 기계번역 결과물에 나타난 오류를 수정하여 더욱 개선된 최종 결과물을 만들어내는 작업이다. 그렇기에 기계번역에서 빈번하게 출현하는 오류를 유

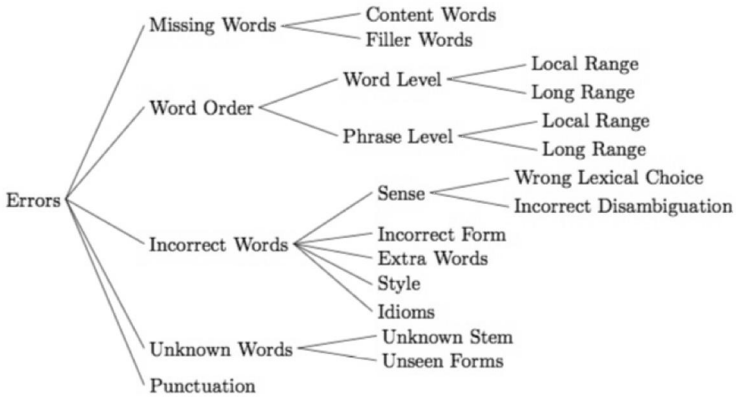
형화할 수 있다면 포스트에디팅 작업에 도움이 될 것으로 생각된다. 이미 기계번역기를 개발하는 연구분야에서는 기계번역 모델과 알고리즘을 개선하기 위해 자주 나타나는 오류를 유형화하고 궁극적으로는 오류 발생을 방지하기 위한 연구가 시도되고 있다. 기계번역 오류유형의 파악은 기계번역 품질개선 뿐만 아니라 포스트에디터와 같이 기계번역 결과물을 가지고 작업하는 사람들에게 가이드라인을 제시하는 등 도움이 될 수 있다. 또한 장애리(2017)에서 제안하는 신경망 기계번역의 오류를 바로잡을 수 있는 보완적 기제로서 오류유형을 활용할 수도 있다.

본 연구에서는 영어-한국어 언어쌍의 신경망 기계번역 결과물에서 나타나는 오류를 파악해보고 오류유형화 가능성을 고찰해보고자 한다. 이를 위해 먼저 지금까지 기계번역의 오류를 다룬 기존의 연구들을 살펴보고 영어-한국어 언어쌍에 적용할 수 있는 오류유형을 파악해 보기로 한다. 다음으로 기계번역 결과물을 오류유형 분류에 적용하여 분석하고 그 결과에 대해 논의해볼 것이다.

2. 기계번역 오류유형에 대한 선행연구 검토

기계번역 오류에 관한 연구는 주로 기계번역 평가의 일환으로 다루어져왔다. 코헨(Kohen 2009)은 기계번역 평가를 크게 두 가지 방식으로 나눈다. 첫째로는 사람이 직접 기계번역 결과물을 평가하는 방식이며 둘째는 컴퓨터를 사용하여 기계번역 결과물을 사람이 번역한 결과물인 참고번역과 자동으로 비교대조하여 평가하는 방식이다. 빌라(Vilar 2016)는 자동으로 오류를 평가하는 기준과 실제 번역 결과물에서 나타나는 오류 사이의 관계를 밝혀내기 어렵다고 한계를 지적하고 있다. 자동평가기준과 실제 번역 결과물 사이의 연관성이 모호하기 때문에 사람이 직접 분석하는 방식에 적용할 수 있는 오류 유형을 정립하기 위한 연구가 활발하게 진행되고 있다. 빌라(2016)는 사람이 진행하는 기계번역 오류분석을 위한 오류유형을 제시했다(그림 1 참조).

그림 1 빌라(2016)의 기계번역 오류유형



빌라는 기계번역 오류를 크게 5가지 유형으로 분류하고 각 유형별로 하위 유형을 상세히 기술하였다. 이 연구는 영어-스페인어, 스페인어-영어, 중국어-영어 언어쌍을 기반으로 진행되었는데 기계번역 유형 기준을 매우 자세하게 제시하였다. 어휘 누락(Missing Words)은 기계번역으로 생성된 번역문에서 어휘가 누락되어 발생하는 오류이며, 누락된 어휘가 의미 전달에 필수적인지 아닌지에 따라 두 가지 하위유형으로 분류된다. 어휘 배열(Word Order)오류는 번역문에서 어휘의 배열 순서로 인해 발생되는 오류이다. 부정확한 어휘(Incorrect Words)는 기계번역이 원천텍스트(Source Text: ST)에서 주어진 어휘에 대응되는 올바른 번역을 찾지 못해서 발생하는 오류이다. 부정확한 어휘로 발생되는 오류는 다섯 개의 하위유형으로 구분되는데, 첫 번째는 문장의 의미를 해치는 경우, 두 번째는 기본 의미는 통하지만 어휘 형태가 적절하지 못한 경우, 세 번째는 다른 어휘가 추가된 경우, 네 번째는 번역문의 어휘사용 방식이 문장의 형식에 어긋나거나 가독성을 해치는 경우, 마지막으로 관용어를 인식하지 못해 오류가 일어난 경우이다. 불분명한 어휘(Unknown Words)는 중국어-영어 언어쌍의 특성상 자주 나타나는 오류 유형이다. 중국어와 영어는 언어체계가 다르고 사용하는 철자도 다르기 때문에 한 쪽 언어에 존재하지 않는 어휘를 그대로 표기할 수 없어 인명, 지명, 단체명 등을 잘못 번역하는 오류를 의미한다. 구두점 오류(Punctuation Error)는 맞춤법 규정이 정해지지 않은 언어가 포함된 언어

쌍에서 발생하는 오류이다. 이렇게 분류한 유형을 바탕으로 코퍼스 데이터를 분석하여 각 유형이 나타나는 빈도를 통계적으로 제시하였으며 주요하게 나타나는 오류 유형이 언어쌍 별로 다르다는 결과를 도출하였다.

자오(Zhao 2013)은 기계번역 오류를 부정확한 어휘(incorrect words), 내용어 누락(missing content words), 잘못된 어휘 배열(wrong word order), 원천텍스트에서와 다른 의미로 번역(translation with meaning contrary to the original), 대상의 이름 표기 오류(errors of named entity), 숫자나 수량형용사/시간 표시 어휘 오류(errors of numeral and quantifiers/temporal words), 기타 오류 등 총 7가지 유형으로 분류한다. 영어-중국어 언어쌍을 중심으로 한 이 연구에서는 총 7가지의 오류유형 중 부정확한 어휘, 내용어 생략, 잘못된 어휘 배열이 가장 많이 나타난 것으로 보고되었다.

한편 팡(Fang 2016)은 기계번역의 오류유형을 3가지로 구분하였다. 오류를 부정확한 어휘 선택, 구조 오류, 구성요소 생략으로 나누었으며 영어-중국어 언어쌍 텍스트에서 길이가 긴 문장을 분석대상으로 하였다. 그의 분석에 따르면 구조 오류가 많이 나타났고, 주로 긴 문장에 포함된 주어-동사가 없는 절에서 오류가 발생한다는 결과가 도출되었다. 팡은 이러한 분석결과를 바탕으로 기계번역의 품질을 향상시키기 위해서는 기계번역을 진행하기 전에 긴 문장을 나눌 것을 제안하고 있다.

기계번역 결과물 전반에 나타나는 오류를 유형화하려는 해외 연구와는 다르게 국내 연구들은 특정 오류양상을 중심으로 한 분석이 주를 이룬다. 양승헌과 김영택(1997)의 연구에서는 영어-한국어 언어쌍에서 나타나는 통사관계 오류를 검출하고 이에 대한 교정 방법을 제시하였다. 많은 양의 한국어 코퍼스에서 추출한 통사관계 오류를 분석하여 주로 구문 분석, 변환, 처리 오류 및 데이터 부족이 오류 발생의 원인임을 통계적 실험을 통해 밝히고 있다. 하지만 통사관계 오류 중에서도 교정하기 쉬운 오류만을 대상으로 한 교정 모듈을 설계했고 다른 유형의 오류는 연구에 포함되지 않았다. 최동익(2012)은 무생물 주어 구문에 대한 영어-한국어 언어쌍 기계번역에 대한 오류를 분석하였다. 다수의 기계번역기를 사용하여 여러 의미역을 지닌 무생물 주어 구문을 번역 후 오류를 확인하였다. 기계번역기가 영어와 한국어에 다르게 나타나는 무생물 주어 사용을 고려하지 않아서 오류를 범한다는 결론을 도출하며 기계번역 프로그램

을 보완할 수 있는 방법을 제시하고 있다. 하지만 분석에 사용된 데이터의 양이 총 22문장으로 적고 통사 구조 중에서도 한 가지에만 집중하여 다른 요소들을 고려하지 않았다는 제한점이 있다.

앞서 검토한 해외 연구들은 모두 통계기반 기계번역을 활용하여 진행되었으며, 국내 연구들 역시 명시적으로 언급되지는 않았지만 시기상 통계기반 기계번역을 활용한 것으로 보인다. 2016년 이후로 많이 사용되는 기계번역기에는 신경망 기계번역 시스템이 적용되어있다. 코헨(2017)에 따르면 신경망 기계번역에 사용되는 신경망에는 숨겨진 층(hidden layer)이 존재한다. 이는 입력 문장과 산출 문장은 알 수 있지만 그 둘 사이에 존재하는 과정에 대해서 알 수 없다는 의미이다. 번역과정에서 발생하는 오류의 원인 및 개선 방안을 알 수 없기 때문에 오히려 결과물을 바탕으로 오류유형을 파악하고 검토하는 작업이 이루어져야만 품질 개선 및 향상이 가능할 것으로 판단된다. 본 연구에서는 현재 가장 많이 사용되는 기계번역기 중 하나이자 신경망 기계번역이 적용되어있는 구글 번역(Google Translate)의 번역결과물을 활용하여 번역오류를 파악해보고 오류들을 유형화해보고자 한다. 분석을 위해 기존의 기계번역 오류유형 연구에서 공통적으로 제시된 오류항목들을 중심으로 오류유형을 분류하고, 영어-한국어 언어쌍 텍스트를 선정 및 분석하며, 그 결과에 대해 논의해볼 것이다.

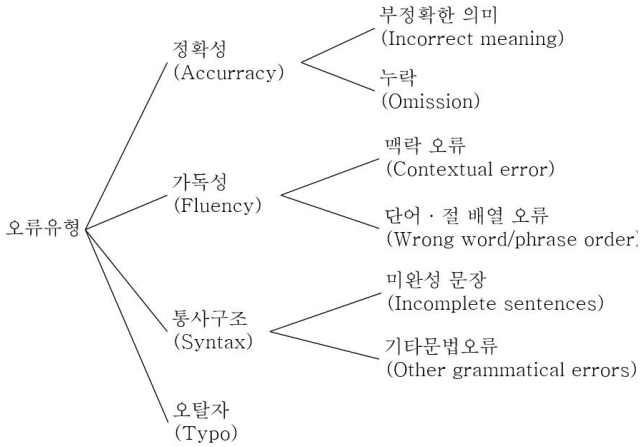
3. 오류유형 분류 및 분석 자료

3.1. 기계번역 오류유형 분류

선행연구에서 살펴본 기계번역 오류유형은 주로 기계번역기를 개발하는 분야에서 논의 및 제시된 것들이다. 때문에 번역학 분야에서 번역의 품질 및 평가와 관련하여 대표적으로 활용되는 기준인 충실성 및 가독성의 개념과는 부합되지 않는 부분이 있으며, 따라서 이들 기준을 그대로 적용하는 것은 적절치 않아 보인다. 또한 통계기반 기계번역과 신경망 기계번역의 결과물을 비교하는 연구(신지선·김은미 2017)나 신경망 기계번역을 번역교육에 어떻게 적용할지 고찰하는 연구(김순미 2017) 등이 이루어졌지만 신경망 기계번역에서 나타나는 오

류 유형에 집중한 연구는 찾아보기 어렵다. 이에 본 논문에서는 기존의 번역학 분야에서 번역품질 평가에 활용하는 기준을 바탕으로 오류유형을 분류해보기로 하였다. 먼저 오류유형을 크게 정확성(Accuracy), 가독성(Fluency), 통사구조(Syntax), 오탈자(Typo)의 네 가지로 분류하였다. 이는 번역결과물을 분석하기 위한 틀로써 번역학 분야에서 대표적으로 활용되는 기준인 충실성, 가독성 개념을 포함하는 동시에 기계번역에서 지속적으로 발생하는 문법 및 기타 오탈자 오류 또한 포함하는 것으로 판단하였다. 각각의 하위유형은 앞에서 소개한 선행연구에서 언급된 공통항목들인 어휘누락, 문장구조오류, 의미오류 등을 유형별 개념에 맞추어 분류하기로 하였다. 물론 이러한 분류방식이 기계번역에서 나타나는 오류를 모두 포함하지는 못하지만 현재 단계에서는 먼저 포괄적으로 오류항목들을 파악하고 오류의 유형화를 시도해보는데 의의를 두기로 한다. (아래 그림 2 참조).

그림 2 기계번역 오류유형 분류



정확성 항목은 번역평가에서 이야기하는 충실성의 개념에 해당하는 것으로 결과물이 원천텍스트의 의미와 달라짐으로써 발생되는 오류를 나타낸다. 주로 어휘의 의미가 틀리게 번역된 경우가 대부분이며 고유명사, 관용어 등의 의미를 다르게 번역하는 경우가 대표적이다. 또한 문장의 구성요소가 누락되어 의

미가 부정확해지는 경우도 해당된다. 여기서는 누락이 ST의 의미를 변화시킨 경우만 포함하였다. 문장 속 구성요소가 삭제되었음에도 문장의 의미가 변화하지 않은 경우는 의미전달에 무리가 없기에 오류에 포함시키지 않았다. 예를 들면 영어와 다르게 한국어는 주어가 없어도 의미가 전달되는 경우가 많은데, ST의 의미가 변화하지 않았기 때문에 오류로 보기에는 무리가 있다고 판단하였다.

가독성 항목은 의미가 틀리지는 않았지만 읽기에 난해하게 번역된 오류를 나타낸다. 단어나 절 차원에서 의미가 맞게 번역되었지만 문장 단위 및 맥락 차원에서 난해하게 번역된 경우를 포함하였다. 하위항목으로는 해당 텍스트의 상황적 맥락에 맞지 않는 경우와 단어 및 절의 순서만 바꾸면 의미가 자연스럽게 통할 수 있는 경우를 제시하였다.

통사구조 항목은 주로 문법오류를 의미한다. 문법오류를 하나하나 하위유형으로 나누면 그 수가 엄청나기 때문에 여기서는 기계번역 결과물에서 두드러지게 나타나는 미완성 문장 오류와 기타문법오류로 크게 구분하였다. 기타문법오류에는 도착텍스트(Target Text: TT)를 기준으로 주술관계, 수식관계 등이 잘못 설정되었거나 종결어미 등이 잘못 사용되는 오류를 포함하였다. 통사구조 유형은 기계번역이 문장구조를 파싱한 후 학습하고 그에 기반하여 번역을 진행하기에 오류유형 분류에 포함시켰다.

오탈자 유형은 기계적으로 발생하는 오타, 맞춤법, 문장부호, 띄어쓰기와 같은 오류를 의미한다. 해당 유형은 인간이 처리하기엔 매우 간단한 오류이나 기계번역에 있어서는 처리하기 어렵다. 기계번역이 각 언어권별로 다르게 사용되는 문장부호 및 기타 맞춤법 개념을 처리하지 못하여 오류가 발생하기에 오류유형의 하나로 포함시켰다.

3.2. 분석 자료 선정 및 분석 방법

기계번역기의 성능은 여러 요인에 따라 편차를 보이는 것으로 알려져 있다. 라이도(Raido 2016)에 따르면 기계번역의 품질은 언어쌍, 텍스트 유형, 주제 등의 여러 요인에 의해 영향을 받는다. 그 중 텍스트 유형에 따른 성능 차이가 두드러지게 드러난다. 마승혜(2017)는 포스트에디팅에 대한 경험적 연구를 진행하면서 텍스트 유형별로 기계번역 결과물의 품질이 달랐고 포스트에디팅의 효율성

또한 달라진다는 결과를 제시하였다. 조재범(2017)에서는 기계번역 독자수용성을 연구하면서 실용 장르의 번역품질이 많이 개선되었고 정보 전달이 목적인 장르와 심미적 표현이 관건인 장르 간 품질과 수용성에 차이가 있음을 제시하였다.

본 연구에서는 영어-한국어 언어쌍의 기계번역에서 반복적 습관적으로 발생되는 오류의 유형화 시도를 목적으로 하기에 가능한 품질 편차가 적은 기계번역 결과물을 분석대상으로 선정할 필요가 있을 것으로 판단하였다. 앞에서 언급한 텍스트 유형과 포스트에디팅 관련 두 연구 모두 정보적 텍스트가 기계번역품질이 가장 높았다는 결과가 도출되었다. 이에 따라 기계번역의 성능이 가장 안정적으로 나타나는 정보적 텍스트 중에서 분석대상을 선정키로 하였다. 정보적 텍스트 중에는 해당 텍스트가 이미 번역되었거나 병렬코퍼스 형태의 기계번역 트레이닝 데이터로 존재하는 경우가 있으며, 이 경우 기계번역기만의 결과물이라고 보기 힘들기에 제외하였다. 이러한 조건을 고려하여 한국지사가 없는 외국 기업의 연례보고서를 분석대상으로 선택하였다. 기업의 연례보고서는 한 해 동안 기업이 낸 성과를 주주 및 다른 사람에게 전달하는 것을 목적으로 하는 정보적 텍스트이고 한국지사가 존재하지 않는다면 영-한 번역이 이루어졌을 가능성이 희박하기 때문이다. 이러한 조건을 만족하면서 동시에 사업 분야별 특성으로 인해 발생하는 기계번역 품질 변화를 최대한 방지하기 위해서 각기 다른 업종의 연례보고서를 선정하였다. 각 보고서의 분량이 방대했기에 보고서 전반부의 기업 소개 및 사업 내용 소개 부분을 발췌하여 분석하였다. 분석에 사용된 데이터는 총 914문장이며 기계번역은 구글번역을 이용해 진행하였다.

〈표 1〉 분석 텍스트 별 업종 및 문장 수

	A사	B사	C사
업종	생활가전용품 판매 및 리스	버스 제작	석유제품 채취 및 판매
문장 수	303	242	369

분석에 있어 번역단위는 문장으로 설정하였는데, 이는 기계번역의 번역단위가 아직은 문장단위에 머물러 있기 때문이다. 각 문장에서 앞에 제시한 기계번역 오류유형이 얼마나 나타났는지 분석하고 특정 경향성이 나타나는지 확인하였다.

4. 분석 및 분석결과 논의

분석은 문장단위로 나타나는 오류유형을 확인하고 각 텍스트에서 유형별로 나타나는 횟수를 측정하는 방식으로 이루어졌다. 하나의 문장에서 같은 오류가 여러 번 나타나는 경우에는 1회로 계산했으며, 반면에 여러 오류가 동시에 나타나는 경우는 횟수를 중복해서 계산하였다. 각 텍스트에서 나타난 기계번역 오류유형 별 횟수는 아래 표와 같다.

〈표 2〉 오류유형별 검출 횟수

오류유형		A사	B사	C사
정확성	부정확한 의미	47	23	107
	누락	18	17	27
가독성	맥락 오류	58	23	64
	단어·절 배열 오류	10	17	31
통사구조	미완성 문장	9	11	28
	기타문법오류	22	16	37
오탈자		104	82	133

정확성 유형에서는 부정확한 의미 오류가 많이 나타났다. 특히 C사 텍스트에서 다른 텍스트보다 의미 오류가 많이 나타났는데, 이는 C사 텍스트의 주제가 석유화학 분야이기 때문에 고유명사인 지명 및 전문용어가 많이 등장하여서 번역에 어려움이 있었던 것으로 보인다. 아래 예시 1에서 지명인 'the Permian Basin'을 각 어휘의 사전적 의미로 번역하면서 오류를 범하고 있다.

예시 1)
ST: Currently, the majority of the rigs running in the Permian Basin are drilling horizontal wells.
TT: 현재 페름기 분지에서 운영되는 대부분의 굴착 장치는 수평 한 우물을 굴착하고 있습니다.

예시 2는 B사 텍스트의 버스 제작에 대한 정보를 제공하는 부분이다. 자동

차 변속장치를 의미하는 ‘transmission’이 ‘전송장치’로 번역되면서 다른 의미를 전달하고 있다. 위의 예시들에서 볼 수 있듯 어휘를 완전히 다른 의미로 번역하는 경우는 주로 고유명사 및 다의어 번역에서 주로 나타났다.

예시 2)
ST: In addition, we are first to market with the new Cummins V8 ISV diesel engine in a school bus application, as well as the dual-clutch seven-speed Eaton transmission.
TT: 또한, 우리는 스쿨 버스 응용 분야의 새로운 Cummins V8 ISV 디젤 엔진과 듀얼 클러치 7 단 이튼 전송 장치를 시장에 최초로 선보였습니다.

누락 오류는 다른 오류들에 비해서 상대적으로 적게 나타났다. 이를 통해서 기계번역 알고리즘이 주로 ST에 존재하는 구성요소를 잘못 번역하더라도 전부 번역하는 방식을 택하고 있다고 가정할 수 있다. 아래 예시에서는 ‘per day’가 TT에서 누락되어 문장 전체의 의미를 왜곡시키는 모습을 볼 수 있다.

예시 3)
ST: EPAct 2005 also gives the FERC authority to impose civil penalties for violations of the Natural Gas Act or Natural Gas Policy Act up to $1 million per day per violation.
TT: 또한 EPAct 2005는 FERC가 천연 가스 법 또는 천연 가스 정책 법 위반에 대해 위반 시마다 최대 1 백만 달러까지 민사 처벌을 가할 수있는 권한을 부여합니다.

가독성 유형에서는 맥락상 부자연스럽게 번역된 경우가 많이 나타났고 단어 및 절 배열 오류는 상대적으로 적게 나타났다. 맥락상 부자연스러운 경우는 앞의 정확성 유형에서 부정확한 의미로 번역된 경우와 맥을 같이 하는 것으로 보이는데, 이는 두 오류 모두 다의어 번역에서 나타나기 때문이다. 하지만 맥락상 부자연스러운 경우는 텍스트의 의미가 통한다는 점에서 정확성 유형과는 다른 양상을 보인다. 아래 예시 4는 ST의 ‘term’이 ‘임기’로 번역되었는데, 기간이라는 큰 의미는 통하지만 앞의 ‘계약’이 언급된 부분과 맥락상 부자연스러워서 수정을 필요로 하는 것을 볼 수 있다.

예시 4)
ST: Our standard agreement is for a term of ten years, with one ten-year renewal option.
TT: 우리의 표준 계약은 10 년의 갱신 기간을 가진 10 년 임기입니다.

단어·절 배열 오류는 의미적으로 번역은 맞게 되었지만 구성요소의 순서가 잘못되어 이해를 어렵게 만드는 경우이다. 이 유형은 주로 길이가 길거나 등위 접속사가 많이 등장하는 문장에서 나타났다. 한국어와는 다르게 영어는 접속사 및 콜론, 세미콜론을 활용한 긴 문장을 자주 사용하여 기계번역기가 번역과정에서 문장 및 구성 요소를 분할하면서 오류를 발생시키는 것으로 보인다. 아래 예시 5에서 TT의 '작업장 노출'을 '유해 물질'쪽으로 재배열하면 ST와 의미가 통하는 모습을 볼 수 있다.

예시 5)
ST: Also, pursuant to OSHA, the Occupational Safety and Health Administration has established a variety of standards relating to workplace exposure to hazardous substances and employee health and safety.
TT: 또한, OSHA에 따라, 산업 안전 보건 청 (Industrial Safety and Health Administration)은 유해 물질 및 근로자의 건강과 안전에 대한 작업장 노출과 관련된 다양한 표준을 제정했습니다.

통사구조 유형은 주로 문법오류에 집중되어 있다. 앞에서도 언급했듯 나타나는 문법오류의 양상이 다양하여 미완성 문장 오류와 주술관계, 수식관계, 종결어미 오류 등을 합한 기타문법오류 유형으로 나누었다. 미완성 문장의 경우에는 거의 대부분이 단어 및 절 배열 오류와 같이 나타났는데, 이는 배열이 잘못된 단어 및 절이 문장 맨 뒤에 위치시키면서 문장을 미완성 상태로 만드는 것으로 보인다. 아래의 예시 6에서는 'and fiscal 2020' 부분의 배열 오류가 나타나는 동시에 맨 뒤에 위치하여 문장을 미완성 상태로 남겨두는 경우가 나타났다.

예시 6)
ST: Our management believes, based on our industry forecast model developed using R.L. Polk school bus registration data and considering

population changes of school age children, that Type C and Type D school bus registrations are expected to grow by 2% to 3% annually between fiscal 2017 and fiscal 2020.
TT: 우리 경영진은 RL Polk 학교 버스 등록 데이터를 사용하고 학령기 아동의 인구 변화를 고려하여 개발 된 업계 예측 모델을 기반으로 2017 회계 연도에 C 유형 및 D 유형 학교 버스 등록 건수가 매년 2 %에서 3 % 증가할 것으로 예상합니다 그리고 2020 회계 연도.

아래의 예시 7, 8, 9에서는 각각 주술관계, 수식관계, 종결어미 오류가 나타난다.

예시 7)
ST: To fulfill demand for parts that are not maintained at the distribution center, we are linked to approximately 35 suppliers that ship directly to dealers and independent service centers.
TT: 유통 센터에서 유지 관리되지 않는 부품에 대한 수요를 충족시키기 위해 당사는 약 35 개의 공급 업체와 연계되어 있으며 판매 업체 및 독립 서비스 센터로 직접 배송됩니다.

예시 8)
ST: • Higher consumer credit quality - DAMI primarily serves customers with FICO scores between 600 and 700, which make up approximately a quarter of the U.S. population.
TT: • 높은 소비자 신용도- DAMI는 주로 미국 인구의 약 4 분의 1을 차지하는 600 ~ 700 사이의 FICO 점수를 고객에게 제공합니다.

예시 9)
ST: The final rule also requires operators to pay royalties to the BLM on flared gas from wells already connected to gas capture infrastructure, and allows the agency to set royalty rates at or above 12.5 percent of the value of production.
TT: 마지막 규칙은 또한 운영자가 이미 가스 포집 인프라에 연결된 우물에서 플레어 가스(flared gas)에 대한 BLM에 로열티를 지불하도록 요구하고, 해당 기관이 생산 가치의 12.5 % 이상을 로열티로 설정하도록 허용한다.

오탈자 유형은 기계적인 오류가 발생한 경우이다. 모든 텍스트에서 Typo 유형이 많이 나타나는 양상을 보이며, 띄어쓰기 오류가 주를 이루었다. 주로 ~인, ~된 ~한으로 끝나는 부분에서 오류가 자주 나타나는 경향을 보였다. 하지만 한글 맞춤법 규정 상 활용형에 따른 예외 규정도 존재하기에 항상 오류가 나타난다고 보기에는 어렵다.

예시 10)[1)]
ST: Blue Bird is also a leading manufacturer of school buses fueled by CNG.
TT: Blue Bird는 또한 CNG가 연료를 공급하는 스쿨 버스의 선도적∨인 제조업체이기도합니다.

예시 11)
ST: Substantially all produced items continue to be leased or sold through Aaron's stores, including franchised stores.
TT: 실질적으로 생산∨된 모든 품목은 프랜차이즈 점포를 포함한 Aaron 매장을 통해 계속해서 임대되거나 판매됩니다.

예시 12)
ST: During 2016, all of the wells we commenced, or participated in, drilling were drilled horizontally.
TT: 2016 년 동안 우리가 시추∨한 모든 우물은 수평으로 천공되었습니다.

이외에는 특정한 경향성을 보이지 않고 무작위로 오류가 발생하였다. 띄어쓰기 오류 다음으로는 영어와 한국어의 차이 때문에 문장부호 오류가 많이 나타났다. 철자 오류는 나타나지 않았다.

앞에서 제안한 오류유형 분류를 통해 기계번역 결과물을 분석하여 다음과 같은 사항을 확인할 수 있었다. 첫째, 의미 중심 오류가 많았으며 주로 고유명사, 다의어가 있는 문장에서 나타났다. 이는 신경망 기계번역이 아직까지는 글의 맥락을 고려하여 상황에 맞는 의미를 구별하여 선택하는 능력이 떨어지기 때문에 발생하는 것으로 보인다. 둘째, 누락 오류가 상대적으로 적게 나타난 것

1) 예시 10, 11 ,12의 ∨기호는 띄어쓰기 구분을 위해 연구자가 삽입한 것이며 실제 텍스트에서는 공백으로 나타났다.

을 보아 현재 기계번역은 주어진 구성요소를 전부 번역하려고 하는 특성을 가지고 있다고 볼 수 있다. 셋째, 가독성 유형의 단어 및 절 배열오류가 발생하면서 통사구조 유형의 미완성 문장 오류를 동반했다. 이는 영어와 한국어의 문장 구조 및 어순이 다르기에 많이 발생하는 것으로 보인다. 넷째, 오탈자 유형 오류는 특정 경향성을 보이지 않고 무작위로 나타났다.

분석한 데이터의 양이 적었기에 위에 언급한 오류들이 모두 규칙성을 보이며 나타난다고 보기는 어렵다. 그러나 전반적인 오류발생의 경향을 볼 수는 있었으며, 비록 제한적이기는 하지만 이러한 오류발생의 경향 연구를 통하여 포스트에디팅 작업에 도움이 될 만한 지침 마련 등 기계번역 품질 향상에 기여할 수 있으리라고 본다.

5. 결론

기술의 발달과 함께 기계번역 또한 빠르게 발전하고 있다. 기계번역의 품질이 과거보다 많이 좋아진 것은 사실이지만 아직 완벽한 결과물을 생산하지는 못하고 있다. 그에 맞추어 기계번역 결과물을 수정 및 개선하는 포스트에디팅과 같은 새로운 형태의 번역도 이루어지고 있다. 하지만 수정 및 개선에 도움이 될 만한 연구들, 예를 들어 기계번역이 어떤 오류를 범하는지에 대해서 유형화하는 방식의 연구는 아직 본격적으로 이루어지지 않았다. 본 연구는 기계번역 오류를 유형화하고자 하는 시도로써, 우선 포괄적으로 기계번역 오류 항목을 분류하고, 이에 기반하여 실제 텍스트를 분석, 오류의 경향성을 탐구하고 논의하였다. 분류한 유형은 선행 연구들에서 제시된 분류에서 공통적으로 드러난 요소들을 중심으로 구성하였으며 약 900문장의 영어-한국어 기계번역 텍스트를 분석한 결과로 의미 중심의 오류가 많이 나타나고, 누락 오류는 상대적으로 적게 나타나며, 단어·절 배열 오류가 미완성 문장 오류를 동반해서 나타나고, 마지막으로 오탈자 유형은 특정 경향성을 보이지 않고 무작위로 나타난다는 결과를 도출할 수 있었다.

본 분석을 일반화하기에는 한계가 여럿 존재한다. 먼저 본 연구에서는 분석 텍스트를 정보성 텍스트 중에서도 기업 연례보고서만으로 한정하였기에 추후 다

른 종류의 텍스트 분석 또한 필요하다. 본 연구는 오류유형을 먼저 분류하고, 이에 맞추어 데이터를 분석하는 연역적 방식으로 진행되었는데, 이는 분석에 사용된 텍스트의 양이 매우 제한적이었던 데에 가장 큰 이유가 있다. 대량의 데이터를 확보하여 발생된 오류양상을 파악하고, 파악된 오류를 유형화하는 귀납적 방식의 연구를 통해 좀 더 유의미한 결과를 얻을 수 있을 것이라 판단되며, 이는 양적 질적으로 대표성을 확보할만한 데이터를 확보하여 오류유형을 검증하는 후속연구를 통해 확인하고자 한다. 마지막으로 분석이 연구자의 주관에 기반을 두기에 다른 연구자들의 후속연구 및 실제 기계번역 업계 전문가들과의 협업을 통해 추가 검증이 이루어져야할 것이다. 기계번역 오류를 유형화하려고 하는 첫 시도였기에 크게 유의미한 경향성을 찾지는 못했지만 이런 연구들이 축적되면 기계번역 결과물을 활용하는 작업에 대한 가이드라인 수립 및 포스트에디팅 교육에도 도움이 될 수 있으리라 본다.

참고문헌

김순미 (2017)「신경망번역기(NMT) 활용 학부 번역교육의 가능성 연구」,『통번역교육연구』15(3): 5-37.

마승혜 (2017)「한영 기계번역 포스트 에디팅에 대한 경험적 고찰」,『한국번역학회 2017 가을 학술대회 발표논문집』: 84-98.

신지선, 김은미 (2017)「인공지능 번역 시스템 출현에 대한 소고」,『번역학연구』18(5): 91-110.

신지선 (2017)「포스트에디팅 －최적의 협업 방식을 찾아서」,『KIGO 저널』8호: 38-53.

신호섭 (2017)「신경망 번역 연구」,『KIGO 저널』8호: 28-37.

양승현, 김영택 (1997)「영한 기계번역 시스템에서 출력된 한국어 문장의 후편집을 위한 통사관계 오류의 검출 및 교정 방법」,『정보과학회논문지』24(7): 759-766.

장애리 (2017)「국내 기계 통번역의 발전 현황 분석 －한 · 중 언어 쌍을 중심으로」,『번역학연구』18(2): 171-206.

조재범 (2017)「기계번역 독자 수용성: 제4차 산업혁명 번역사례」,『한국번역학회 2017 가을 학술대회 발표논문집』: 78-83.

최동익 (2013)「무생물 주어 구문에 대한 영한 기계 번역 오류분석」,『언어학연구』29: 279-299.

Fang F. & Ge S. & Song (2016) 'Error Analysis of English-Chinese Machine Translation' in Sun M. & Huang X. & Lin H. & Liu Z. & Liu Y. (eds) *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*: 35-49.

Kohen, P (2009) *Statistical machine translation*, Cambridge & New york: Cambridge University Press.

Raido, V. E. (2016) 'Translators as Adaptive Experts in a Flat World: From Globalization 1.0 to Globalization 4.0?', *International Journal of Communication* Vol.10: 970-988.

Vilar, D & Xu, J & D'Haro, L & Haro, D & Ney, H (2006) 'Error analysis of statistical machine translation output', *Proceedings of the Fifth International Conference on Language Resources and Evaluation*: 697-702.

Zhao, H & Xie, J & Lv, Y & Yu, H & Zhang, H & Liu, Q (2013) 'Common error analysis of machine translation output', *The Ninth China Workshop on Machine Translation*, Kunming.

[인터넷 자료]

김태균 (2017.02.21.) '인간 vs AI' 번역대결서 인간 압승…AI, 문학 번역 취약」,『연합뉴스』 http://www.yonhapnews.co.kr/bulletin/2017/02/21/0200000000AKR20170221160600017.HTML

표준국어대사전 http://stdweb2.korean.go.kr/search/List_dic.jsp

Barak Turovsky (2016.11.15.) 'Found in translation: More accurate, fluent sentences in Google Translate' https://blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/

Kohen, P (2017) 'Neural Machine Translation' http://mt-class.org/jhu/assets/nmt-book.pdf

[Abstract]

An Analysis of Errors in Machine Translation

Seo, Bo-Hyun · Kim, Soonyoung
(Dongguk University-Seoul)

Advancement in technology leads to rapid development of machine translation. On account of such development, a new way of translating like post-editing is emerging. Post-editing is fixing errors in machine translation output, hence enhancing the quality of machine translation. Effort required by post-editing may vary by genre/type of text and the patterns of errors specific to source text. This pilot study intends to classify machine translation errors in informative text. Firstly, the study provides classification of machine translation errors based on four broad classes: Accuracy, Fluency, Syntax, and Typo. Secondly, errors from English-Korean machine translation of informative texts are analysed with the proposed classification. Lastly, the paper explores occurrence frequency of each error classes and deduces tendencies from the analysis: Incorrect meaning error occurs rather frequently while omission error is found relatively few; Wrong word/phrase order error comes with the incomplete sentence error; Typo errors occur randomly without any patterns. The findings fall short of presenting error patterns due to relatively small size of sample. Future research could look into more predictable error patterns in machine translation that could not be investigated here and might contribute to reducing efforts required by post-editing.

▸ Key Words: Error classification, Informative text, Machine translation, Neural Machine Translation, Patterns of error, Post-editing
▸ 주제어: 오류유형화, 정보적 텍스트, 기계번역, 신경망 기계번역, 오류 패턴, 포스트에디팅

서보현
동국대학교 대학원 영어영문학과 번역학전공
bhyun412@gmail.com
관심분야: 기계번역, 포스트에디팅, 번역교육

김순영
동국대학교 문과대학 영어통번역학과
kimsy@dongguk.edu
관심분야: 문학번역, 번역품질평가, 기계번역

논문투고일: 2018년 1월 31일
심사완료일: 2018년 3월 13일
게재확정일: 2018년 3월 20일