

구어체 문장의 한중 기계번역 오류 개선을 위한 프리에디팅 가이드라인 연구

- 중국어 비숙련자를 대상으로

임소영* · 임소정**

목 차

1. 서론
2. 선행연구
3. 연구 대상 및 방법
 - 1) 분석 텍스트 및 번역 프로그램
 - 2) 연구 방법
4. 번역 오류 유형별 프리에디팅 가이드라인
 - 1) 통사 총위
 - 2) 어휘 총위
 - 3) 화용 총위
 - 4) 기타 총위
5. 결론

국문초록

본 연구는 대한민국 청년들이 일상생활에서 사용하는 살아있는 한국어 구어체를 네이버 파파고를 활용하여 중국어로 기계번역하였을 때 발생하는 오류 양상에 주목하여, 구어체 기계번역의 품질 개선을 위한 프리에디팅 규칙을 오류 양상별로 제안하고자 하였다. 기존의 프리에디팅과 관련된 연구는 대부분 한-영, 한-노 등 비중국어를 대상으로 하고 있고, 신문 기사나 연설문 등 정보형 텍스트를 분석 대상으로 삼거

* 한국방송통신대학교 실용중국어학과 강의담당교수(제1저자)

** 한밭대학교 중국어과 부교수(교신저자)

나 텍스트를 특정하지 않았으며, 기계번역 사용자의 수준을 고려하지 않은 채 프리에디팅 가이드라인을 제시하였다. 본 연구는 분석 자료를 한국어 구어체로 특정하고 중국어를 도착어로 삼아 중국어 비숙련자들을 위한 프리에디팅 규칙을 수립하였다는 점에서 차별화된다.

키워드: 한중 기계번역, 프리에디팅, 구어체, 중국어 비숙련자, 네이버 파파고

1. 서론

AI 기술의 발전으로 신경망 기계번역의 품질이 빠른 속도로 발전하면서 네이버 파파고, 구글 번역 등 기계 번역은 외국어 학습자 뿐만 아니라 일반 사용자들에게도 친숙한 도구로 자리매김하고 있다. 물론 현 단계의 기계번역은 아직 완벽하지 않으며 기계가 극복하지 못하는 한계가 분명히 존재하기에 인간의 교정 작업이 수반되는 경우가 대부분이다. 교정 작업은 기계번역 이전의 출발어 텍스트에서 이루어지는 프리에디팅(pre-editing)과 기계번역 이후의 결과물에서 이루어지는 포스트에디팅(post-editing)으로 나뉜다. 전문 번역 업계에서는 1차 번역은 기계 번역을 거치고 2차 번역 및 검수를 인간 번역사에게 맡기는 포스트에디팅 방식을 가장 많이 채택하고 있다. 국내의 선행 연구 또한 포스트에디팅에 대한 연구가 프리에디팅에 비해 수적으로 압도적으로 많다.¹⁾ 언어쌍 기준으로는 한국어-영어(마승혜 2018; 박옥수 2017; 이성화, 김세현 2018 등), 한국어-중국어(김혜림 2022; 한설옥 2020 등), 한국어-노어(한현희 2018) 등 다양한 언어의 포스트에디팅 연구가 이루어졌다.

그러나 포스트에디팅 방식이 모두에게 가용한 것은 아니다. 에디팅을 하려면 대상 언어를 잘 알아야 한다. 한국어가 모국어이고 중국어가 외국어인 사

1) KCI(한국학술지인용색인)에서 '프리에디팅/사전교정'과 '포스트에디팅/사후교정'을 검색어로 검색한 결과, 전자는 총 13편, 후자는 총 66편의 논문이 발표되었다.

람이 기계번역을 활용하여 한-중 번역을 하는 경우, 기계가 출력한 중국어 번역문을 자신있게 포스트에디팅하려면 중국어 실력이 수준급이어야 한다. 초급 학습자나 중국어를 아예 모르는 사람들 입장에서는 외국어를 포스트에디팅하기가 버겁게 느껴질 것이다. 실제로 국내 포스트에디팅 연구의 대다수가 B-A번역(외국어-한국어) 연구²⁾였고, 프리에디팅 연구의 경우 A-B번역(한국어-외국어) 연구³⁾가 상대적으로 많았던 점을 보면 일반적으로 외국어보다는 모국어를 직접 교정하는 쪽이 선호된다는 사실을 알 수 있다.

본 연구는 초급 중국어 학습자 또는 중국어 학습 경험이 전혀 없는 사용자들이 한중 기계번역을 할 때 적용할 수 있는 프리에디팅 가이드라인을 제시하고자 한다. 도착어인 중국어를 전혀 모르더라도 모국어인 한국어 원문에 대한 깊은 이해를 바탕으로 프리에디팅을 하면 중국어 번역문의 품질을 한 단계 더 끌어올릴 수 있다는 점은 충분히 매력적이다. 연구의 궁극적인 목표는 더 많은 중국어 학습자들과 일반 사용자들이 한중 번역을 함에 있어 기계번역을 보다 효과적으로 활용하는 것이다.

2. 선행연구

기계번역은 상당한 정확성을 가지고 신속하게 결과물을 생산하고 있다. 또한, 기계번역을 무조건적으로 수용하는 것이 아니라 비판적이고 창조적으로 활용할 수 있는 능력을 함양하여 인공지능과 공존 관계를 이어나가는 것은 시대적 요구이다.(한설옥 2020) 그러기 위해서는 기계번역의 품질에 대한 평가, 기계번역의 오류 양상에 대한 분석을 통해 기계번역의 현주소를 우선적으로 파악해야 할 것이며, 이를 기반으로 기계번역을 보다 효과적으로 잘 활용

2) 김혜림 2021, 2022; 윤미선 외 2018; 이준호 2021 등

3) 김순미 2017; 박옥수 2016, 2017; 한현희 2018 등

하기 위한 전략을 세워야 할 것이다. 최근 들어 기계번역의 품질 평가, 기계번역의 오류 유형 분석, 포스트에디팅 등 연구가 많이 진행되고 있는데 이는 모두 인간이 기계번역과 공존할 수밖에 없는 이 시대에 어떻게 하면 기계번역을 보다 잘 활용할 수 있을지 모색하는 과정에서 수행된 연구이다. 프리에디팅 연구도 이러한 연구의 일환이다.

프리에디팅 연구 가운데 한국어를 출발어로 하는 프리에디팅에 관한 국내 논문은 다음과 같다. 박옥수(2017)는 한-영 언어쌍의 기계번역의 오류 유형을 통사 구조가 복잡한 긴 문장, 주어 파악이 어려운 경우, 다의어나 동음이의어, 관용어나 특수어 사용 등으로 나누어 출발어인 한국어의 유형적 특징에 착안하여 분석하였다. 이성화·김세현(2018)은 한-영 기계번역의 품질을 향상시키기 위한 프리에디팅 기법으로써 생략된 주어의 명시화, 복합 명사 풀어쓰기를 통한 의미의 명확화, 문장 구조의 단순화 등을 제안하였다. 진용주·서보현·김순영(2018)은 정보성 텍스트 및 문학 텍스트 등 장르를 불문하고 프리에디팅이 한-영 기계번역의 충실성을 일정 수준으로 개선하였고 잠재적인 문법적 오류를 제거하는 데 효과가 있다는 점을 밝혔다. 한현희(2019)는 한-노 기계번역 오류 유형과 원인을 분석한 결과를 바탕으로 일련의 프리에디팅 규칙을 제시하였다. 프리에디팅 관련 연구에서 한국어-중국어 언어쌍에 대한 연구는 전무하다. 이에 본 연구는 한-중 언어쌍으로 프리에디팅 가이드라인을 제안해보고자 한다.

번역 텍스트 유형에 따라 번역의 난이도, 방법, 전략, 기계번역의 정확도 등이 달라지는 만큼, 텍스트 유형을 특정하여 번역 연구를 진행할 필요성도 있다. 현재 실제 일상생활 속에서 사용되는 구어체에 대한 기계번역 관련 연구는 매우 부족한 실정이다. 임소영·임소정(2023)은 일상생활에서 사용되는 생동감 있는 구어체를 대상으로 한중 기계번역을 다룬 거의 유일한 연구이다. 해당 연구는 구어체 문장의 한-중 기계번역에서 다양한 유형의 오류가 발견되고 있음을 보여주고 있다. 특히 본 연구의 분석 대상인 구어체 텍스트는 정형화된 문어체와는 달리 살아 숨 쉬는 언어이기 때문에 그만큼 생동감이 넘

치지만 불완전한 형태라서 기계가 이해하기 쉽지 않으리라는 점을 미루어 짐작할 수 있다. 뿐만 아니라 한국은 고맥락문화(high-context culture)에 속하며, 의사소통을 할 때 직접적인 표현보다는 간접적이고 은유적인 화법으로 청취자가 상황과 맥락 심지어 ‘눈치’까지 동원하여 발화자의 의도를 파악해야 하는 언어권이기 때문에 중국어를 비롯해 외국어로 번역할 때는 원문인 한국어 구어체를 적절히 보정하여 맥락을 추가하고 의미를 명확하게 하는 작업이 필요하다.

구어체를 대상으로 하는 기계번역 연구는 많지 않지만, 일반 번역 연구 가운데 구어체를 대상으로 한 연구는 적지 않게 있다. 황의열(2016)은 한문 문장 속에 있는 대화체 문장의 특징을 분석하고 이에 대한 번역에서의 주의사항에 대하여 논의하고 있다. 대화체 문장은 대체로 짧고 생략이 많으며 생생한 것을 특징을 갖고 있는 만큼 번역자가 변형이나 오해의 소지 없이 전달하기 위해서는 맥락과 말투에 대한 이해가 필요하며, 격에 맞는 준비법을 운용하고, 그리고 무엇보다도 국어에 대한 조예가 뒷받침되어야 함을 강조하였다. 한정은(2014)은 한국어와 중국어의 직시 체계를 비교하고 중국어를 한국어로 번역할 때 대화문에서의 ‘직시’가 어떻게 번역되어야 하는지를 고찰하였다. 해당 연구는 중국어를 한국어로 번역하는 단일 방향에 대해서 직시의 번역을 논하였기 때문에 한국어를 중국어로 번역할 때의 주의 사항에 대해서는 더 논의해볼 필요가 있다. 홍문표(2011)은 한국어와 독일어의 대화체 기계번역에서 주어생략 현상을 처리하는 방안을 논의하였다. 한국어와 같은 주체지향언어에서 문법적 주어는 문맥이나 상황 또는 세상 지식 등을 통해 유추할 수 있는 경우 흔히 생략될 수 있기 때문에, 반드시 문법적 주어가 문장 내에 등장해야 하는 독일어와 같은 주어지향언어로 번역할 때에는 생략된 주어를 복원하는 과정이 필요하다고 하였다.

3. 연구 대상 및 방법

1) 분석 텍스트 및 번역 프로그램

본 연구는 넷플릭스의 연애 리얼리티 프로그램 <솔로지옥> 시즌2 출연진과 방송 패널들의 발화 데이터를 연구 대상으로 하여, 한국을 대표하는 신경망 기계번역 플랫폼인 네이버 파파고의 한-중 기계번역 결과를 분석하고 주요 오류 유형에 따른 프리에디팅 가이드라인을 제안한다.

언어 텍스트로는 <솔로지옥> 시즌2 총10개의 에피소드 중에서, 1화 영상의 한국어 자막을 활용하였으며, 자막 단위로 나누어진 총 문장 개수는 940개이다. 기계번역 플랫폼으로 네이버 파파고를 선정한 기준은 다음과 같다. 기계번역의 성능은 해당 플랫폼의 유저들이 생성하는 언어 데이터를 자동적으로 학습하며 발전하기에, 사용자의 수와 사용 시간이 많은 네이버 파파고는 1순위 분석 대상이었다. 중국인 한국어 학습자들을 L2 수준별(입문/초급/중급/고급)로 분류하여 기계번역 프로그램 사용 빈도를 조사한 공수(2019)에 따르면, 입문·초급 학습자들은 고급 학습자들에 비하여 기계번역에 대한 의존도가 더 높았고, 선호하는 번역 프로그램은 파파고(32%), 바이두(16%), 구글(10%), 파파고/바이두/구글(19%), 기타(23%)였다. 네이버 파파고는 한국어와 중국어의 일상적 표현과 구어체에 강하며 번역의 정확도가 한중 언어 쌍에서 압도적으로 높다.(장애리 2017) 중국어 비숙련자들의 경우, 중국 기업의 기계번역 프로그램인 百度(바이두), 有道(요우다우), 騰訊(텐센트) 등 프로그램에 대해서는 잘 모르므로 네이버 파파고를 사용할 확률이 높다. 이러한 점을 감안하여, 프리에디팅 가이드라인을 수립함에 있어 국내에서 인지도와 활용도가 가장 높은 네이버 파파고를 선정하였다.

2) 연구 방법

본 연구는 구어체(conversational) 한국어 문장의 기계번역 품질을 개선시키는 것을 목표로 삼아, 임소영·임소정(2023)에서 분류한 구어체 문장의 한-중 기계번역 오류 유형에 기반하여 국내 기계번역 플랫폼인 네이버 파파고에서 보이는 각 오류 유형을 고찰하고, 해당 오류 유형에 대하여 프리에디팅을 실험적으로 테스트하여 적절한 가이드라인을 제안하고자 한다.⁴⁾ 중국어 출력문에 대해서는 중국어 원어민의 감수를 통해 검증한다. 오류 유형은 기본적으로 출발어의 관점에서 크게 어휘 층위, 통사 층위, 화용 층위, 기타 층위⁵⁾ 등 네 가지로 분류하고 분석한다.

4. 번역 오류 유형별 프리에디팅 가이드라인

본 장에서는 임소영·임소정(2023)에서 네이버 파파고, 카카오i번역, 바이두번역, 텐센트번역, 유도번역 등 한국과 중국의 대표적인 기계번역 플랫폼 5개에서 공통적으로 나타난 구어체 한-중 기계번역의 각 오류 유형을 참고하여, 초급 중국어 학습자들이 가장 널리 사용하는 네이버 파파고로 범위를 축소하여 오류 유형별 프리에디팅 방안을 집중적으로 탐색하였다. 분석 결과는 아래

-
- 4) 이외에도 세부적으로 여러 오류 양상들이 있겠지만 본 연구는 네이버 파파고가 범하는 오류 가운데 발생빈도가 높은 오류 유형을 위주로 논의하고자 한다. 발생 빈도가 높은 오류 유형을 집중적으로 분석하고 이에 대한 프리에디팅을 진행하는 것만으로도 기계번역의 품질을 향상시키는 데 충분히 기여할 수 있을 것이라고 보기 때문이다.
- 5) 하나의 언어 표현이 동시에 여러 의미를 지니는 경우가 많다. 본고는 단어의 사전적 의미, 지시적 의미를 어휘 층위에서 다루고, 상황 맥락에 의존한 화용적 의미는 화용 층위에서 다루기로 한다.

표1과 같다.

표1. ST 층위별 번역 오류 현황

분류	세부 유형	횟수	비중	발생 빈도	
통사 층위	생략된 문장 성분	18		*****	
	복잡한 구조	3		*	
통사 층위 - 소계		21	38%		
어휘 층위	중의적 표현	9		****	
	다의어	7		***	
		동음이의어	2		*
	관용적 표현	8		****	
	신조어	5		***	
		관용어	2		*
		문화특수어	1		*
어휘 층위 - 소계		17	31%		
화용 층위	맥락 파악	4		**	
	답화 표지	4		**	
	완곡 화법	4		**	
화용 층위 - 소계		12	22%		
기타 층위	문장 표기 등	5		***	
기타 층위 - 소계		5	9%		
총 합		55	100%		

발견된 번역 오류는 총 55개, 이 중에서 가장 많은 오류가 통사 층위에서 발생하였고 그 기저에는 한국어의 대표적인 특징이자 구어체의 특징인 문장 성분 생략(omission)이 있었다. 그 뒤를 어휘 층위, 화용 층위, 기타 층위가 이었다. 비중으로 치면 통사(38%), 어휘(31%), 화용(22%), 기타(9%)로 대략 통사4:어휘3:화용2:기타1의 분포를 보인다.

1) 통사 층위

통사 층위의 오류가 가장 많이 발생한 것은 연구 대상이 출연진들의 자연 발화를 담은 구어체 텍스트였기 때문에 어떻게 보면 충분히 예상 가능한 결과이다. 한-중 기계번역에서 통사 층위의 프리에디팅 가이드라인을 다음과 같이 제안한다. (1) 주어, 목적어, 동사 등 생략된 문장 성분을 다시 추가하고 (2) 장황한 수식어를 삭제하여 간결한 문장으로 만든다. 자세한 내용은 다음과 같다.

1. 생략된 문장 성분을 적절한 위치에 추가하여 문장을 완성한다.

통사 층위에서 발견된 번역 오류의 대부분은 문장 성분 생략(omission)으로 인한 것이다. 문화에 따라 커뮤니케이션 유형이 다르게 나타나는데 Hall(1976)은 의사소통의 의미와 메시지를 다루는 방법에 기초하여 의사소통 방식을 ‘고맥락(High Context)’ 유형과 ‘저맥락(Low Context)’ 유형으로 구분하였다. Hall(1976:91)에 따르면, ‘고맥락’ 의사소통 방식에서는 대부분의 정보가 물리적 맥락 속에 있거나 발화자 개인에 내재되어 있고, 메시지가 코드화되거나 명시적이지 않다. 그에 반해, ‘저맥락’ 소통 방식은 정보의 대부분이 명시적 코드에 포함된다. 물론 같은 문화권 안에서도 늘 맥락도가 동일한 것은 아니다. 상황에 따라 혹은 사람의 성향에 따라 소통의 방식이 달라진다. 예를 들어 객관성과 정확성이 중요시 되는 법정이나 외교 현장에서는 매우 구체적인 메시지로 전달을 하는 저맥락 방식을 택하는 반면에 사적인 대화를 할 때에는 고맥락 방식을 택한다. Bernstein(1964)은 언어 코드를 ‘제한된’(고맥락) 코드와 ‘정교한’(저맥락) 코드로 분류하여 어느 쪽에 속하느냐에 따라 어휘, 구문, 음성 등이 달라진다고 주장하였다. 특히 친근한 사람끼리는 구문이나 음성 등이 상당히 단축된 형태의 ‘제한된’ 코드를 사용한다고 밝혔다.

본 연구는 고맥락 문화권에 속하는 대한민국의 청년들이 실제로 사용하는

대화문을 분석 텍스트로 삼은 만큼, 원천언어 텍스트에서 생략형 구문이 빈번하게 출현하는 것은 당연지사다. 미국 등 서양권은 저맥락 문화에 주로 속하는 데다가 통사적으로 주어 중심의 언어가 많기 때문에, 많은 프리에디팅 연구는 A-B 번역에서 한국어 문장에서 생략되어 있던 주어를 출력문에서 반드시 보충해야 한다는 규칙을 제시하고 있다.(김순미 2017; 박옥수 2017; 이성화, 김세현 2018 등) 중국의 경우는 한국과 더불어 고맥락(high-context) 문화에 속한다.(김은일 2015) 하지만 중국어의 문장은 주어, 목적어, 관형어, 동사 술어 등의 생략 빈도가 한국어에 비해 낮다. 다음은 네이버 파파고 기계번역에서 나타난 성분 생략형 예시이다.

[예시1] 동사 누락

ST 와 소름
TT 哇 鷄皮疙瘩

ST-1 와 소름 돋아
TT-1 哇 起鷄皮疙瘩了⁶⁾

예시1을 일대일로 축어번역을 하였는데 일견 문제없어 보인다. 그러나 문제는 결과물이 중국인 원어민 기준에서 자연스럽게 않다는 점이다. 가독성 기준에서 문제가 되므로 동사를 추가하는 프리에디팅 작업이 필요하다. 이외에 한국어 ‘아이 진짜.’에 적합한 중국어 번역은 ‘哎呀 真是的.’인데, 이 문장에서도 한국어는 술어가 생략되어 있는 반면에 중국어는 동사 술어가 사용되었다. 이처럼 중국어는 한국어에 비하면 동사 술어를 잘 생략하지 않는 경향이 있다.

6) 이하에서는 기계번역 이전의 한국어 출발텍스트(Source Text)를 ST, 기계번역 이후 중국어로 출력된 도착텍스트(Target Text)를 TT로 통일한다. 아울러, 프리에디팅이 적용된 버전은 ST 뒤에 번호를 첨가하여 구분하였고 여러 개의 프리에디팅 버전을 시도한 경우에는 번호를 추가하였다.

한국어는 맥락이 주어지면 주어와 목적어 등의 생략이 잦다. 영어나 독일어 같은 언어는 주어 중심의 언어이기 때문에 주어를 반드시 보충해 넣는다. 중국어는 화제 중심 언어이기 는 하지만 주어의 사용 빈도가 한국어보다 높다. 영어나 독일어처럼 주어가 통사적으로 필수불가결한 성분이라고 볼 수는 없지만 맥락이 주어지더라도 주어를 생략하지 않고 쓰는 것이 자연스러울 때가 많다. 중국어에서는 주어 뿐 아니라 목적어 등도 생략 없이 표현하는 경우가 많다. 한-중 번역에서 프리에디팅을 할 때 아래와 같이 생략된 주어와 목적어를 보충해주면 번역의 정확도가 한층 제고됨을 알 수 있다.

[예시2] 주어 누락

ST 주변에서 저보고 “좀 차가울 거 같다”, “좀 공주 같을 거 같다”라고 말을 많이 하시는데 저랑 친해져 보면 실제로 좀 털털하기도 하고 의외로 유쾌한 면도 있다고 많이 말씀해 주십니다.

TT 周圍的人經常對我說“感覺會有点冷漠”、“感覺會像公主”，但和我親近起來后，實際上會覺得有点洒脱，也有意想不到的愉快的一面。

ST-1 주변에서 저보고 “너는 좀 차가울 거 같다”, “너는 좀 공주 같을 거 같다”라고 말을 많이 하시는데 저랑 친해져 보면 '실제로 네가 털털하고 또 의외로 유쾌한 면도 있다'고 많이 말씀해 주십니다.

TT-1 周圍的人經常對我說“你好像有点冷漠”、“你好像有点像公主”，但是和我親近的話，很多人說‘實際上你很洒脱，也有意想不到的愉快的一面’。

[예시3] 주어 누락

ST 제가 ‘솔로지옥’을 통해서 이상형을 만난다고 하면 떨리긴 하겠지만 저를 많이 표현을 해서 질 마음에 들 수 있게끔 노력을 할 것 같습니다.

TT 如果我通過‘SOLOGION’見到我的理想型會很緊張 但是我會努力多表現自己 讓自己滿意

ST-1 제가 '솔로 지옥'을 통해서 이상형을 만나게 된다면, 떨리긴 하겠지만 저를 많이 표현을 해서 그 분이 절 마음에 들어할 수 있게끔 노력을 할 것 같습니다.

TT-1 如果我通過《單身地獄》見到我的理想型,雖然會緊張,但我會努力多表現自己,讓那個人喜歡我。

[예시4] 목적어와 주어 누락

ST 제가 좋다고 하면 다 좋다고 하던데요

TT 只要我說好,大家都說好。

ST-1 제가 그들을 좋다고 하면 그들도 다 저를 좋다고 하던데요

TT-1 我說喜歡他們,他們都說喜歡我。

한국어와 중국어 직시체계에는 차이가 있다. 한국어의 인칭직시는 구분이 복잡하고 이에 따른 경어법이 발달되어 있지만 인칭직시가 활발하게 사용되지 않고 생략되는 경우가 많다.(이익섭, 이상억, 채완, 1997: 235) 그에 비해, 중국어의 인칭직시는 한국어에 비해 수적으로는 다양하지 않지만 일상대화에서 매우 빈번하게 사용되고 있다.(한정은 2014). 한국어와 중국어는 주어 뿐 아니라 목적어, 관형어 위치에서도 인칭대명사의 사용 빈도상의 차이가 존재한다. 위의 여러 예문을 통해 실제로 <솔로지옥>의 한국어 텍스트에서 주어, 목적어 등의 인칭대명사가 생략되는 경우를 많이 볼 수 있으나 이를 중국어로 옮길 때에는 생략된 인칭대명사를 복원해야 자연스러워지는 경우가 많음을 확인할 수 있다. 그렇기 때문에 프리에디팅을 할 때 생략된 문장 성분을 추가해볼 수 있다.

2. 복잡한 통사 구조 : 어느 특정한 문장 성분이 지나치게 긴 경우, 문장 부호를 사용하여 하나의 단위임을 명확하게 해주고, 군더더기 표현은 삭제하며, 논리 관계의 왜곡이 발생하기 쉬운 긴 문장은 여러 개로 끊는다.

[예시5] 수식어가 지나치게 긴 경우

ST 저는 '솔로지옥'에서 이번 여름 뿐 만 아니라 오랜 시간 볼 수 있는
분을 만나고 싶어요.

TT 我想在《Sology Ok》中見到不僅在這個夏天，還能長時間見到的人。

ST-1 저는 '솔로 지옥'에서 '이번 여름 뿐 만 아니라 오랜 시간 볼 수 있
는 분'을 만나고 싶어요.

TT-1 我想在《單身地獄》中見到不僅是今年夏天，還能長時間見到的人'。

[예시6] 지나치게 긴 수식어 + 군더더기 표현

ST 저는 제가 좀 진지하게 확실하게 좋아해서 만난 사람이 없었어서 그
런 사람을 찾기 위해서 나온 것도 있고요.

TT 我是因為我是認真的喜歡，所以沒有見過的人，所以爲了找那樣的人
才出來的。

ST-1 저는 '제가 좀 진지하게 확실하게 좋아해서 만난 사람'이 없었어서
그런 사람을 찾기 위해서 나온 것도 있고요.

TT-1 因為我沒有非常真摯地確實喜歡，所以交往過的人，所以爲了尋找這樣
的人才出來的。

예시5와 예시6은 관형어가 지나치게 길어 기계번역이 관형어의 범위를 정
확하게 인식하지 못하고 있음을 확인할 수 있다. 이에 관형어와 중심어를 부
호 ' '를 추가하여 하나로 묶어줌으로써 번역의 정확도가 향상되었다.

2) 어휘 층위

1. 의미 축소를 통한 중의성(ambiguity) 해소

다의어나 동음이의어와 같이 여러 의미로 해석될 여지가 있는 중의적 표현
의 경우, 보다 구체적인 의미를 가진 하위어로 대체하거나, 앞뒤에 맥락적 정

보를 추가함으로써 기계가 올바른 선택지를 선택할 확률을 높여야 한다.

기계번역에서는 언어의 중의성으로 인한 오역이 상당히 빈번하게 발생한다. 중의성이란 한 단어 또는 문장이 상황과 맥락에 따라 여러 가지 의미로 해석되는 현상인데, 중의성만 해소시키더라도 기계번역의 품질이 상당 부분 개선될 것으로 본다. 넓은 의미에서 중의성은 다의어와 동음이의어로 구분할 수 있다. 다의어(polysemy)는 서로 연관이 있는 두 가지 이상의 의미를 지닌 단어이고, 동음이의어(homonym)는 음만 같고 의미가 완전히 다른, 사실상 별개의 두 단어이다. 기계번역에서 어휘 번역의 정확성을 높일 수 있는 관건은 다른 뜻으로 해석될 수 있는 확률을 낮추고 해당 문장에 가장 적합한 뜻을 선택하는 확률을 높이는 것이다. 중의성 해소를 위한 프리에디팅 가이드라인은 크게 (1) 애매한 어휘의 경우, 보다 구체적이고 자세한 뜻이 있는 하위어로 교체하는 구체화(clarification) 작업을 하고 (2) 문장 전체적으로 내용과 맥락을 추가하여 의미를 명확하게 하는 것이다. 다음은 다의어에 대한 예시이다.

[예시7] 다의어 ‘몸’

ST 근데 이번에 사실 출연진 대충 아직 얼굴은 보이지 않았는데
몸은 여전하더라고요,

TT 其實這次出演嘉賓雖然還沒有露臉，但身體確實還是老樣子
(근데 이번에 사실 출연진 대충 아직 얼굴은 보이지 않았는데 신체
는 여전하더라고요.)

ST-1 이번에 사실 출연진 대충 아직 얼굴은 보이지 않았는데 몸매는 여전하더라고요,

TT-1 這次出演嘉賓們雖然還沒有露臉，但是身材還是一如既往

위는 <솔로지옥> 출연진들의 모습이 어렴풋이 담긴 선공개 영상을 지켜보던 연예인 패널리 내뱉은 말이다. 이번 시즌 참가자들이 지난 시즌 못지않게 건강미 넘치는 아름다운 몸매를 지녔다며 칭찬하는 대목이다. 여기서 주목해야 할 단어는 ‘몸’이다. ‘몸’의 사전적 의미는 ‘사람이나 동물의 형상을 이루는

전체 또는 그것의 활동 기능이나 상태⁷⁾이다. ‘몸이 튼튼하다’, ‘몸이 허약하다’, ‘몸에 좋다’고 할 때의 그 ‘몸’이다. 그러나 예시7의 ‘몸’은 신체 또는 건강 상태보다는 몸의 맵시나 모양새를 의미하는 ‘몸매’에 더 가깝다. 한국어에서는 ‘몸이 좋다’가 ‘몸매가 좋다’의 의미로 쓰이기도 하나 중국어에서는 ‘몸’을 ‘身体’로, ‘몸매’를 ‘身材’로 명확하게 구분한다. 기계번역은 해당 문장에서 ‘몸’의 여러 의미 가운데 어떤 의미인지 정확히 파악하는 데 실패하여 TT와 같은 오역을 하였다. 한국어의 ‘몸’과 같은 다의어의 중의성을 해소하기 위하여 프리에디팅을 할 때 해당 단어를 의미가 협소한 하위어로 대체하였더니 정확한 번역을 도출할 수 있었다.

다음은 동음이의어의 예시이다.

[예시8] 동음이의어 ‘이성’

ST 중요하게 생각하는 이성의 매력 포인트는 ‘현명함’입니다

TT 我認爲理智的魅力点是‘明智’

(이지의 매력 포인트는 ‘현명함’이라고 생각합니다.)

ST-1 제가 중요하게 생각하는 여성의 매력 포인트는 ‘현명함’입니다

TT-1 我認爲女性的魅力点是‘明智’

(저는 여성의 매력 포인트가 ‘현명함’이라고 생각합니다.)

위는 한 남성 출연자가 본인이 매력적으로 여기는 여성상에 대하여 인터뷰하는 상황이다. 이 맥락에서의 ‘이성’은 성별이 다름을 뜻하는 ‘이성(異性)’이므로 남성인 화자 입장에서는 여성임을 어렵지 않게 추측할 수 있다. 그러나 기계번역은 이러한 정황 정보 없이 자구만을 기준으로 번역을 하다 보니 ‘이성(異性)’의 동음이의어인 ‘이성(理性)’으로 오판하였다.⁸⁾ ‘이성(異性)’과 ‘이성

7) 출처: 국립국어원 『표준국어대사전』

8) ‘이성(理性)’은 개념적으로 사유하는 능력이며 인간을 다른 동물과 구별시켜 주는 인간의 본질적인 특성이다.

(理性)'은 발음은 같지만 한자와 어원이 다르기에 의미가 전혀 다르다. 결과적으로 기계는 '이성(異性)'의 유의어인 '이지(理智)'를 출력하여 '이성과 지혜의 매력 포인트가 '현명함'이라는 다소 엉뚱해보이는 문장으로 완성되었다. 이 문장은 문법적으로나 기능적으로는 문제가 없지만 원천 텍스트의 내용을 오역한 것이다.

이때 비교적 손쉽게 적용할 수 있는 프리에디팅 기법은, 기계의 오판을 유발한 문제의 동음이의어 '이성'을 다른 단어로 교체하는 것이다. 동음이의어 '이성(理性)'으로 출력될 가능성을 차단하기 위해, ST-1에서는 '이성'의 의미를 한 단계 구체화·축소화한 '여성'으로 변경하였다. 그 결과, TT-1에서 동음이의어 오역 없이 화자의 발화 의도가 무난하게 출력되었음을 확인할 수 있다. 물론, '여성(女性)' 또한 한자어인 만큼 '여자의 목소리'를 뜻하는 '여성(女聲)' 등과 같은 동음이의어로 오인될 확률이 아예 없는 것은 아니다. 그러나 '여성(女聲)'과 '이성(理性)'의 사용 빈도를 고려해본다면, '여성(女聲)'으로의 오역 가능성은 프리에디팅 전에 비하여 현저히 낮아졌을 것으로 판단한다.

2. 관용적 표현 오역: 함축된 의미와 맥락 풀어쓰기

관용적 표현은 대개 '두 개 이상의 단어로 이루어져 있으면서 그 단어들의 의미만으로는 전체의 의미를 알 수 없는, 사회 문화적으로 관습화된 특수한 의미로 사용되는 어구(語句)'이다.⁹⁾ '발이 넓다'가 '사교적이어서 아는 사람이 많다.'를 뜻하는 것 따위이다. 송현선(2015)은 한중 번역에서 대중화된 은유로 볼 수 있는 관용 표현을 도착어로 번역할 때 역시 관용적 비유를 사용하는 경우가 많다고 하였다. 예를 들어, 한국어의 관용 표현이 '담대거리'를 중국어의 관용 표현인 '豬腦袋(돼지머리)'로 번역하고 있다. 해당 연구는 출발어와 도착어의 각 언어와 문화에 대한 이해도가 높은 인간 번역가의 번역 결과물을 대상으로 한국어의 은유 표현이 중국어로 어떻게 번역되고 있는지를

9) 본고에서 관용적 표현은 신조어, 관용어, 문화특수어 등을 모두 포괄한다.

Larson(1984)¹⁰⁾의 다섯 가지 번역 전략을 적용하여 분석한 논문이다. 그러나 본 연구는 중국어 비숙련자를 대상으로 기계번역을 활용할 때의 전략을 세워야 하는 만큼 관용어에 대한 번역 전략이 송현선(2015)과는 달라야 한다. ‘닭 대가리’를 프리에디팅할 때 ‘돼지머리’로 교체하는 것은 중국어와 중국 문화에 대한 지식이 많은 기계번역 사용자에게 한해서만 가능할 것이다. 중국어와 중국 문화에 대한 지식이 없거나 초보인 기계번역 사용자의 경우, 관용적 표현의 번역에서 프리에디팅을 할 때 관용 표현은 사전적 의미를 연결시키는 것만으로는 전체 의미를 알 수 없기에, 내재된 의미와 맥락을 성실하게 풀어쓰기(paraphrasing)를 통하여 명시화(explicitation)해야 한다.

[예시9] 관용어 ‘작업 멘트’

ST 이야, 저거 옛날 작업 멘트인데요.

TT 哇 那是以前的工作台词啊

ST-1 이야, 저거 옛날에 여자들 꼬실 때 많이 쓰던 작업 멘트인데요.

TT-1 那是以前勾引女生的时候经常用的台词

‘작업’은 ‘일을 함. 또는 그 일; 일정한 목적과 계획 아래 하는 일’을 뜻하고¹¹⁾, ‘멘트’는 ‘방송이나 예식 등에서 진행자나 사회자가 상황에 맞게 하는 말’을 나타낸다.¹²⁾ 그러나 예시9에서의 ‘작업 멘트’는 ‘마음에 드는 이성을 꼬이기 위해 하는 달콤한 말’로써 한국 사회에서 이미 대중들에게 보편화되어 관용 표현으로 많이 사용되고 있다. 네이버 파파고는 이러한 관습적 의미를 파악하지 못하고 단순히 각 단어의 사전적 의미를 기계적으로 조합하여 ‘工作台词’라는 어색한 번역을 한 것이다. 인공지능 기계 번역이 문맥에 대한 이해

10) Larson(1984)이 제시한 다섯 가지 번역 전략은 다음과 같다. (1)은유에서 은유로 번역, (2)은유에서 직유로 번역, (3)출발어의 은유와는 다른 은유로 번역, (4)은유에 비은유적인 표현을 첨가하는 번역, (5)은유를 비은유적인 표현으로 기술하는 번역

11) 출처: 『표준국어대사전』

12) 출처: 『고려한국어대사전』

도가 높아졌다고는 하지만 아직까지 사회문화적 지식이 동반되어야 하는 관용어 번역에서는 한계를 보여주고 있는 대목이다. 이렇듯 자면 그대로 해석할 경우 완전히 다른 의미로 해석될 여지가 있는 관용어의 경우, 오역의 가능성을 최소화하기 위해서 관용어의 함축의미를 최대한 평이하게 풀어서 쓰는 프리에디팅 과정을 거치는 것이 좋다. ‘작업 멘트’를 ‘여자들 꼬실 때 많이 쓰던 말’로 풀어쓰는 프리에디팅 과정을 거쳤더니 원천 텍스트의 의미를 잘 전달할 수 있게 되었다. 또 다른 관용 표현의 예시를 보자.

[예시10] 신조어 ‘면접프리패스상’

ST 되게 이분 면접 프리패스상 아니에요?

TT 這位不是面試合格相嗎?

‘면접 프리패스상’은 최근 한국 사회에서 많이 쓰이고 있는 신조어라고 할 수 있다. 중국어에는 이 말에 대응하는 적절한 어휘가 존재하지 않는다. 임소영·임소정(2023)에 따르면, 카카오i, 바이두, 텐센트 등 3개의 기계번역이 ‘면접 프리패스상’을 제대로 번역하지 못하였다. 네이버 파파고는 자체적으로 ‘面試合格相’이라는 표현을 고안해내어 그 관용적 의미에 근접하게 표현을 하였다. 하지만 중국인 감수자는 ‘面試合格相’에 대해 무슨 말인지는 대충 이해할 수는 있으나 중국에서는 쓰지 않는 어색한 말이라고 답하였다. 대략적인 의미 전달만을 번역의 목표로 삼고 있다면 ‘面試合格相’도 아주 나쁜 번역은 아닐 수 있다. 다만 좀 더 자연스럽게 수화자로 하여금 단어의 의미를 확실하게 이해시킬 수 있는 수준까지 번역의 효과를 끌어올리기 위해서는 아래와 같이 풀어쓰기를 통하여 명시화하는 프리에디팅 과정을 거칠 필요가 있다.

[예시10]

ST-1 이분은 되게 어떤 면접에 가도 프리패스할 상 아니에요?

TT-1 這位不管參加什麼面試 都能通過吧?

ST-2 이분은 되게 어떤 면접에서도 프리패스할 이미지 아니에요?

TT-2 這位不是在任何面試中都能通過的形象嗎?

ST-3 이분은 되게 어떤 면접에서도 프리패스할 스타일 아니에요?

TT-3 這位不是在任何面試都能自由通過的類型嗎?

만약 위에서 언급한 바와 같이 기계번역 사용자가 번역결과물에 대한 기대치가 높지 않고 어느 정도의 의미전달만을 목표로 하고 있다면 풀어쓰지 않고 ‘面試合格相’과 같은 한국어 관용어를 직역한 형태의 관용어를 도출하도록 그냥 두면 된다. 그러나 이 또한 제반 조건과 주의 사항이 따른다. 우선 관용 표현을 구성하고 있는 각 형태소가 사회 관습적인 함축 의미를 나타내지 않아야 하고, 다의어나 동음이의어가 아니어야 한다. 그러면 관용 표현 단독으로 기계번역을 돌려도 원하는 출력문을 얻을 확률이 높다. 하지만 그렇지 못할 경우에는 맥락 제공이 필수다. ‘면접 프리패스상’은 복합어로서 크게 한자어 ‘면접(面接)’과 영어 ‘프리패스(free pass)’와 ‘한자어 ‘상(相)’으로 결합된 구조이다. 그중 ‘상’은 ‘얼굴의 생김새’를 의미하는데 ‘獎’, ‘上’ 등의 동음이의어가 존재하기 때문에, 동음이의어를 다른 단의어로 대체하거나 맥락을 제공하여 중의성을 해소해야 한다. 현실적으로 단의어의 존재 자체가 많지 않기 때문에 대개 후자의 방법을 선택할 가능성이 높는데, 이러한 맥락은 한 문장 내에서 주어져야 한다. 아래 프리에디팅의 실험 과정을 보자.

[예시10']

ST 면접 프리패스상

TT 面試自由通行獎

ST-1 개는 면접을 보기만 하면 다 붙어. 면접 프리패스상이야.

TT-1 他一面試就考上了。是面試自由通過獎。

ST-2 개는 면접 프리패스상인가봐.

TT-2 看來他是面試自由通過者。

ST-3 개는 얼굴이 면접 프리패스상이야.

TT-3 他的臉是面試通票相。

네이버 파파고에 ‘면접 프리패스상’을 단독으로 입력하면 ‘面試自由通過獎’이 출력된다. 동음이의어에 의한 오역 현상이다. 이를 해소하기 위하여 세 가지 방식으로 맥락을 제공해보았다. ST-1은 ‘면접 프리패스상’이 포함된 문장 외적으로 별개의 단독 문장을 추가하여 맥락을 제공하였는데, 기계번역이 문장 밖으로 벗어나면 맥락을 인지하지 못하고 번역하였다. ST-2는 문장 내부에서 앞에 인칭대명사 주어를 추가하여 ‘면접 프리패스상’이 ‘사람’이라는 정보를 주었다. 그러자 파파고는 ‘獎’에서 ‘者’로 바꾸어 출력하였다. ST-3과 같이 문장 내부에 좀 더 직접적인 맥락 정보(‘얼굴’)를 제공하니 드디어 ‘면접 프리패스상’의 ‘상’을 ‘相’으로 번역하였다. AI번역이 맥락을 파악하고 번역할 수 있는 능력을 갖추었다고 하지만 이는 단일 문장 범위 안에 국한되는 것으로 보인다.

3. 문화특수어 오역 : 특수성 제거를 통한 언어 중립화

[예시11] 문화특수어 ‘돌쇠’

ST 바로 돌쇠로 이용하네요, 와

TT 直接用石鎖了 哇

ST-1 용재씨를 바로 돌쇠로 이용하네요, 와

TT-1 把榮宰當石頭用呢 哇

ST-2 용재씨를 바로 이용하네요, 와

ST-2 直接利用榮宰了 哇

ST-3 용재씨를 바로 짐꾼으로 이용하네요, 와

TT-3 竟然利用榮宰做搬運工 哇

ST-4 바로 짐꾼으로 이용하네요, 와
 TT-4 直接用作搬運工了, 哇

‘돌쇠’는 ‘남자 하인’ 또는 ‘짐꾼’을 일컫는 말이며 대사전에는 수록되어 있지 않다. 문화 특수어에 대한 프리에디팅 가이드라인은 반대로 특수성(specificity)을 제거하여 양국 문화에 통용될 수 있도록 일반화(normalization)하는 것이다. 물론 ‘돌쇠’를 음역하거나 별도 표기를 하여 보류하는 방법도 있다. 대표적인 사례가 BTS 팬덤 번역인데, 해외 팬들이 콘텐츠 내용 뿐 만 아니라 한국 문화를 더 잘 이해할 수 있도록 자국화 요소와 이국화 요소를 모두 남기는 번역 전략을 취한다. 그러나 본고에서는 한국 문화 전파에 대한 목적보다는 의미 전달이라는 기능에 포커스가 맞춰져 있으므로 문화 특수성을 제거하는 일반화 방식을 사용한다.

3) 화용 층위

담화표지(Discourse marker)는 화자의 담화 전략상 사용되며 화자의 감정이나 태도, 뉘앙스 등을 나타내는 화용 층위의 언어 단위이다. 어휘 층위가 아니라서 문장의 명제 의미에는 큰 영향을 미치지 않는다. 담화표지는 앞뒤 문맥의 영향을 많이 받고 구체적인 어휘적 의미가 없기 때문에 기계번역으로 하여금 오역을 양산하게 만드는 요인 중 하나다. 분석 텍스트에서 ‘조금’, ‘좀’, ‘아니’ 등 담화표지로서 기능하는 사례가 적지 않게 있었는데 파파고는 대부분 오역을 하였다. 다음 예시를 보자.

[예시12] 담화표지 ‘조금’

ST 제가 과일은 조금 깎아요
 TT 我會削一些水果。

ST-1 제가 과일은 잘 깎아요

TT-1 我很會削水果。

임소정, 유위(2017)에 따르면 한중 소량 표현인 ‘点’과 ‘좀’이 ‘사물의 소량’이라는 구체적인 기본 의미에서 담화영역까지 확장되어 여러 문맥 속에서 ‘요청 약화’, ‘부정적 평가 약화’ 등의 화용 기능으로도 사용되고 있다. 예시12에서 ‘조금’은 기본 어휘의미인 ‘소량’을 나타내지 않는다. 화자가 자화자찬하는 상황에서 상대방이 느낄 수 있는 부담을 완화하기 위해 ‘좀’을 사용하여 완곡하게 표현한 것이다. 여기서 ‘좀 깎는다’의 실제 의미는 ‘잘 깎는다’의 의미이다. 이러한 상황적 이해를 바탕으로 ‘조금’을 ‘잘’로 교정하여 입력하였더니 상황적 맥락에 적합한 출력문이 나왔다. ‘点’과 ‘좀’과 같이 각 언어에는 유사한 의미 기능을 나타내는 단어가 존재하는데 이들이 의미의 확장 과정까지 완벽하게 일치하는 경우는 없다.¹³⁾ 설사 기본 의미에서 한국어 ‘조금’에 중국어 ‘点’이 대응된다고 하더라도 담화 기능까지 정확하게 일치하는지는 중국어 비숙련자 입장에서는 더더욱 알기가 어렵다. 그렇기 때문에 ‘点’으로 출력될 것을 의도하고 원천언어의 ‘조금’을 그대로 두는 것은 적절치 않으며 프리에디팅을 진행해야 한다. 이제 다소 다른 기능을 지닌 담화표지의 예시를 보자.

[예시13] 담화표지 ‘아니’

ST 아니, 회전목마가 돌아가더라고

TT 不是 旋轉木馬在轉

ST-1 ∅ 회전목마가 돌아가더라고

TT-1 ∅ 旋轉木馬在轉動

ST-2 봤어?, 회전목마가 돌아가더라고

TT-2 看到了嗎? 旋轉木馬在轉

13) 실제로 한국어 ‘좀’은 화용 층위에서 ‘함축/얼버무리기’, ‘요청의 강화’ 등 더욱 세분화된 담화기능으로까지 발전한 데 비해, ‘点儿’의 담화 기능은 한국어만큼 발달되지 않았다. (임소정, 유위 2017)

ST-3 세상에, 회전목마가 돌아가더라고
 TT-3 天啊 旋轉木馬在轉

‘아니’는 본래 부정(否定)을 의미하는 말인데 용법이 확대되어 다양한 담화 기능으로도 사용된다. 담화표지 ‘아니’에 대해 이한규(2012)는 맞장구치기, 시간 끌기, 화제 전환, 말차례 가져오기, 관심 끌기, 자기 수정 등의 기능이 있다고 보았다. 예시13에서 ‘아니’는 부정의 의미가 아닌 담화 용법으로 사용된 것으로, 패널들끼리 대화를 주고 받는 가운데 한 패널이 상대 패널의 말에 맞장구침과 동시에 자기의 말차례를 가져오는 역할을 한다. 그런데 네이버 파파고는 ‘아니’를 모두 중국어의 부정 표현인 ‘不是’로 번역하여 어색한 문장이 되었다. 해당 담화 표지를 ST-3나 ST-4와 같이 다른 비슷한 뉘앙스의 말로 바꿔 표현할 수는 있겠지만, 원문에서 담화표지를 제거해도 원문의 진리조건적 의미에 전혀 영향을 주지 않는다면 중국어 비숙련자에게는 안전하게 담화표지를 삭제하는 방법을 제안한다. 그 이유는 담화표지를 삭제하여도 명제 의미가 달라지지 않는데 굳이 무리해서 프리에디팅을 할 필요가 없기 때문이다. 초급 수준의 중국어 학습자가 다른 유사한 단어로 애매하게 대체하여 프리에디팅을 하였다가 오역이 날 경우 결과물을 포스트에디팅할 능력이 없기 때문에 이러한 보수적인 방법을 택해야 한다.

4) 기타 총위

다음은 문장 표기법을 위주로 한 출발어 입력 시 주의사항이다.

첫째, 한 문장은 엔터 없이 또는 불필요한 공백 없이 연결해서 입력하여야 기계가 한 문장임을 인식한다. 실험 결과, 현 시점의 네이버 파파고가 의미를 인식하고 재구조화하는 기준 단위는 개별 문장이다. 그러므로 문단 하나를 통째로 입력하더라도 기계가 여러 문장의 선후 관계와 맥락을 실시간으로 학습해나가며 번역을 하는 게 아니기 때문에 문장 하나를 입력했을 때와 결과는

대부분 동일하다. 모든 프리에디팅 가이드라인의 적용 범위는 개별 문장이다. 생략 성분의 추가, 수식 구조의 명확화, 하위어로의 의미 구체화, 숨겨진 의미 풀어쓰기, 기계가 파악하지 못하는 맥락 추가 등 다양한 기법은, 문장 전체의 완결성을 높인다는 목표 하에 실행되어야 효과적인 프리에디팅이 된다.

둘째, 정확한 띄어쓰기와 맞춤법을 사용하고 문장부호를 적절히 활용해야 한다. 그 이유는 기계가 의미 구분을 더 쉽게 할 수 있도록 조력하기 위해서다. 일견 원론적인 이야기 같지만, 실제 실험을 해본 결과 띄어쓰기의 유무, 쉼표 유무, 마침표 유무에 따라 기계가 선택하는 단어가 달라지고 번역의 품질이 완전히 뒤바뀌는 경우가 적지 않았다.

5. 결론

본 연구는 중국어 학습 경험이 전무하거나 초급 수준의 학습자 등의 중국어 비숙련자가 한중 기계번역을 보다 효율적으로 활용하고 그 결과물의 품질을 향상시킬 수 있도록 원천 텍스트를 프리에디팅하는 기법을 제안하였다.

본 연구는 크게 세 가지 측면에서 다른 프리에디팅 연구와의 차별성을 찾아볼 수 있다.

첫째, 한국어-중국어 언어쌍에 대한 프리에디팅 논의를 처음 시도하는 것이다. 국내 프리에디팅 가이드라인 연구는 주로 서양언어를 중심으로 이루어져 있다.

둘째, 기계번역을 사용하는 사용자의 언어 수준에 대해 고려를 하여, 본 연구는 초급 수준의 중국어 구사자를 대상으로 프리에디팅 가이드라인을 제안하였다. 프리에디팅 규칙을 제안한 기존의 연구는 대부분 학습자의 외국어 수준에 대한 구분 없이 프리에디팅 규칙을 제안하고 있다. 하지만 기계번역 사용자의 수준에 따라 기계번역의 활용 방법은 다르고 또한 기계번역으로부터

얻고자 하는 결과물에 대한 기대치도 다르다. 기계번역의 품질을 높이기 위하여 우선 기계번역의 한계를 정확하게 인식하고 기계번역 사용자가 얻을 수 있는 현실적인 목표치가 어느 정도인지 가늠해보고, 목표 설정을 한 이후에 국내외 프리에디팅 가이드라인을 비교 분석한 후 이를 실제 한국어-중국어 프리에디팅에 적용하여 가이드라인에 취사선택 및 수정 보완이 필요한 점을 고찰하였다.

셋째, 기계번역 프리에디팅의 텍스트를 구어체로 선정한 점이다. 기존의 연구는 주로 정보형 텍스트를 위주로 논의하거나 여러 텍스트를 두루 다루며 분석 대상에 대한 범위 설정을 하지 않았다. 그러나 텍스트의 유형에 따라 기계번역의 결과도 달라지는 만큼 프리에디팅의 가이드라인 또한 텍스트에 맞추어 분석되어야 한다.

본 연구는 분석 텍스트의 제한된 오류 사례 수로 인하여 본고가 제안한 프리에디팅 규칙을 아직 충분히 검증하지는 못하였다. 이 연구는 한국어-중국어의 프리에디팅에 대한 예비연구적 성격을 지닌 것으로, 추후 해당 프리에디팅 가이드라인을 대량의 데이터에 적용하여 품질의 향상 정도를 조사하여 본 연구 결과의 타당성을 높여줄 필요가 있다. 또한, 기계번역 사용자의 도착어에 대한 이해 부족으로 기계번역을 교정할 수 있는 범위가 제한적이며, 그렇기 때문에 완전무결한 번역과는 일정 부분 거리가 있음을 받아들일 수밖에 없다.

參考文獻

- 공수, 「기계번역 프로그램 품질에 대한 사용자 평가와 사용자의 L2 수준 간 상관관계 고찰」, 『통역과 번역』 21(3), 한국통역번역학회, 2019.
- 기유미, 「기중 기계번역 오류의 문형별 비교분석: 네이버 파파고 번역기와 구글 번역기를 중심으로」, 『중국연구』 74권, 2018.
- 김순미, 「신경망번역기(NMT) 활용 학부 번역교육의 가능성 연구」, 『통번역교육연구』 15(3), 한국통역번역교육학회, 2017.
- 김은일, 「언어유형론 및 어휘화의 관점에서 바라본 영한 번역에 나타난 수동태의 변조」, 『현대문법연구』 제83권, 2015.
- 김혜림, 「중한 기계번역 포스트에디팅 가이드라인에 관한 예비연구: 포스트에디팅 교육을 위한 가이드라인」, 『중국언어연구』 제99권, 2022.
- 김혜림, 「중한 법령 기계번역 포스트에디팅 교육을 위한 예비 연구」, 『번역학연구』 22(3), 2021.
- 마승혜, 「한영 기계번역 포스트 에디팅에 대한 경험적 고찰: 학부 교육 과정 및 결과를 중심으로」, 『통번역학연구』 22(1), 통번역연구소, 2018.
- 박옥수, 「한영 병렬코퍼스와 기계번역에서 의존 명사: '-것'이 포함된 어휘의 번역 방식 연구」, 『동아인문학』 제37권, 동아인문학회, 2016.
- 박옥수, 「한영 기계 번역에서 ST의 유형적 특징에 따른 번역 오류 분석」, 『동아인문학』 제41권, 동아인문학회, 2017.
- 송현선, 「한-중 은유 번역과 번역 전략 연구」, 『중국연구』 제63권, 한국외국어대학교 중국연구소, 2015.
- 윤미선 외, 「영어-한국어 언어쌍에 적합한 포스트에디팅 가이드라인: 미래 포스트에디터를 위한 가이드라인과 그 실패」, 『번역학연구』 19(5), 2018.
- 이성화·김세현, 「영-한 및 한-영 기계번역 품질향상을 위한 프리에디팅 기법 제안」, 『번역학연구』 제19권5호, 한국번역학회, 2018.
- 이익섭·이상익·채완, 『한국의 언어』, 신구문화사, 1997.
- 이준호, 「영한 포스트에디팅 생산성에 대한 고찰- 시간적 노력을 중심으로」, 『

- 통번역학연구』 제25권2호, 통번역연구소, 2021.
- 이한규, 「한국어 담화표지어 ‘아니’의 의미」, 『현대문법연구』 제67권, 2012.
- 임소정·유위, 「중국어 ‘点儿’과 한국어 ‘좀’의 의미 기능 및 의미 확장 과정의 비교」, 『중국문학』 제90집, 2017.
- 임소영·임소정, 「구어체 한-중 AI 번역의 오류 양상 연구 - 넷플릭스 솔로지옥 발화를 중심으로」, 『중국어문학』 제92집, 영남중국어문학회, 2023.
- 장애리, 「국내 기계 통번역의 발전 현황 분석: 한중언어쌍을 중심으로」, 『번역학연구』 제18권2호, 한국번역학회, 2017.
- 진용주·서보현·김순영, 「프리에디팅(pre-editing)이 기계번역 품질에 미치는 영향 고찰:유창성, 충실성, 가독성 비교를 중심으로」, 『통번역학연구』 제22권3호, 통번역연구소, 2018.
- 한설옥, 「인공지능 시대의 한국어 교육에서 기계번역 활용 가능성 - 중한 번역을 중심으로」, 『한중인문학연구』 제69권, 한중인문학회, 2020.
- 한정은, 「중국어 대화체 번역에서 직시체계 번역의 중요성」, 『통번역학연구』 제18권1호, 통번역연구소, 2014.
- 한현희, 「한노 기계번역의 오류 유형화 및 품질 개선을 위한 프리에디팅(pre-editing) 규칙 제안」, 『노어노문학』 31(44), 한국노어노문학회, 2019.
- 한현희, 「한노 기계번역의 문장 부호 처리 양상 및 오류 개선을 위한 실용적 합의:프리에디팅 규칙 수립과 효용성의 관점에서」, 『통번역학연구』 제22권3호, 통번역연구소, 2018.
- 홍문표, 「한-독 대화체 기계번역을 위한 주어생략현상의 처리방안」, 『독어학』 제24권, 한국독어학회, 2011.
- 황의열, 「대화체 문장의 번역에 대하여」, 『태동고전연구』 제37집, 2016.
- Bernstein, Basil. *Elaborated and Restricted Codes: Their Social Origins and Some Consequences*. American Anthropologist 66 (6).1964.
- Hall, E. T. *Beyond culture*. New York:Anchor Press-Doubleday. 1976.
- Larson, M.L. *Meaning-Based Translation: A Guide to Cross Language*. Lanham: University press of America. 1984.

[공구서]

고려대학교 민족문화연구원 『고려대 한국어대사전』

국립국어원 『표준국어대사전』

Abstract

A study on pre-editing guidelines for improving errors in Korean-Chinese
machine translation of spoken sentences
— for unskilled Chinese speakers

Lim, So Young · Lim, So Jeong

This study attempted to propose pre-editing rules for improving the quality of spoken machine translation by paying attention to errors that occur when machine translating live Korean spoken language used by young Koreans into Chinese using Naver Papagos. Most of the existing research related to pre-editing targets non-Chinese such as Korean-English and Korean-Russian, did not analyze informational texts such as newspaper articles and speeches, or did not specify text, and presented pre-editing guidelines without considering the level of machine translation users. This study is differentiated in that it specified analysis data in Korean colloquial style and established pre-editing rules for beginner-level Chinese learners using Chinese as an arrival word.

Key words : Korean-Chinese machine translation, pre-editing, colloquialism, unskilled Chinese speakers, Naver Papago

투 고 일 : 2023. 4. 10. / 심 사 일 : 2023. 4. 15. ~ 2023. 5. 15. / 게재확정일 : 2023. 5. 20.

