

『韓国語教育研究』(第11号)別刷

ISSN 2186-2044

【研究論文】

한일 번역의 품질 평가 기준에 대한 일고찰
-새로운 평가 모델의 제안-

朴 敏瑛

日本韓国語教育学会

2021年9月

한일 번역의 품질 평가 기준에 대한 일고찰*

-새로운 평가 모델의 제안-

朴 敏瑛

本稿では、日韓翻訳の平準化された品質評価を目的に、これまでの様々な評価基準を具体的に検討し、新たな品質評価モデルを設計した。本稿の翻訳の品質評価モデル(KJTQA)の特徴は、1) マクロ的な観点とミクロ的な観定の折衝、2) 減点制度に加算点を導入、3) 誤謬の測定方式の多元化の3つが挙げられる。本稿の考察を通じて、翻訳の品質評価は、単なる誤りの減点採点ではなく、究極的には翻訳というコミュニケーションの特徴及び翻訳家の能力と努力、それにテキストと言語の構造などに関する言語学的な知識などと関連付けながら、包括的かつ綿密に議論されるべきであることを改めて確認できた。本稿で提示した新しい韓日翻訳の品質評価(KJTQA)モデルは今後様々な実験を通してより平準化された評価システムとして機能できるように修正・補完を続けていくつもりである。本稿で設計された品質評価モデルが、今後機械翻訳の誤り分析のみならず、翻訳教育の現場やポストエディティングなどで客観的かつ体系的な評価基準として活用できることを期待したい。

1. 연구의 목적

본 논문의 목적은 기계번역을 포함하여 한국어-일본어(이하 ‘한일’) 번역물의 품질 평가를 위한 객관적이며 평준화된 평가 기준을 마련하여 향후 한일 번역의 품질평가 및 품질 향상에 유익한 정보를 제공하고자 하는 것이다.

언어의 4기능을 언어 활동의 처리 과정과 연계해보면 듣기와 읽기는 ‘수용적 언어활동(reception)’이며 말하기와 쓰기는 ‘산출적 언어활동(production)’이라고 말할 수 있다. 이와 관련하여 번역과 통역은 직접적인 커뮤니케이션 능력을 갖지 못한 언어 사용자 간의 커뮤니케이션을 가능하게 하는 이른바 ‘중개활동(mediation)’으로서 외국어 능력에 있어서는 최고 레벨의 언어 수용 능력과 언어 산출 능력이 함께 요구된다.

* 이 연구는 2021 학년도 한국외국어대학교 교내학술연구비의 지원에 의하여 이루어진 것임.

최근 딥러닝을 통한 기계번역의 눈부신 발전과 상용화로 인하여 번역 현장에서는 엄청나게 빠른 속도로 대량의 정보 처리가 가능하게 되었다. 특히 매뉴얼 번역과 같은 일반 업무용 문서의 경우에는 비교적 한정된 어휘와 단순한 구문 덕분에 기계번역만으로도 어느정도 정밀도가 높은 번역물을 제공받을 수 있으며 특히 한국어와 일본어처럼 구조상 유사한 언어 간 번역의 경우에도 상대적으로 완성도가 높은 결과물을 얻을 수 있게 되었다.

그러나 작금의 기계번역은 장래 무한한 가능성을 내포하고 있으면서도 동시에 극복해야 하는 많은 오류와 문제점을 안고 있는 것 또한 사실이다. 선행연구에서도 기계번역의 등장 이래 다양한 관점과 텍스트를 통해 기계번역의 오용 및 품질을 검토해 왔으나 가장 큰 문제점은 번역 품질을 객관적으로 측정 가능한 평준화된 ‘평가 기준’이 부재하다는 점이다.

다시 말해서 번역 품질에 대한 평가는 전적으로 평가 기준에 의해 좌우되는데 선행연구에서는 언어에 따라서 또는 연구자에 따라서 각각 번역물을 평가하는 평가 기준 자체가 다르고 또한 평가 항목에 있어서도 분류 기준 자체가 명확하지 않기 때문에 같은 번역물이라 할지라도 평가 기준에 따라 평가 결과가 달라지는 경우가 종종 있다.

일례로 기계번역에 대한 오류를 고찰한 연구에서도 오류를 분류하는 기준은 논문을 집필하는 연구자에 따라 다르며, 대학이나 대학원의 번역 수업에서도 번역물을 평가하는 기준은 강의를 담당하는 교수에 따라 달라진다. 최근 주목받고 있는 ‘포스트에디팅(Post-editing)’의 경우에도 에디터의 역량이 곧 번역의 품질이라고 해도 과언이 아닐 만큼 에디터 개인의 능력에 따라 번역의 품질이나 완성도에는 큰 차이가 있다.

한일 번역 관련 연구에서도 번역물의 품질 평가 및 오류 분석을 위한 다양한 평가 기준을 볼 수 있지만 여기서도 연구자에 따라 평가 기준이 제각기 다르기 때문에 투명성과 타당성을 담보하기가 어렵다. 또한 평가 기준의 설정에 있어서도 ‘번역’이라는 커뮤니케이션의 특성을 충분히 반영하지 못하고 있거나 각각의 평가 기준 자체가 애매모호하여 실제 분류에는 어려움이 많다는 점 등의 문제점을 지적할 수 있다.

따라서 본고에서는 기계번역을 포함하여 모든 유형의 한일 번역물에 대한 객관적이고 투명한 품질 평가 기준을 마련하는 것을 목적으로 먼저 선행연구의 다양한 평가 기준을 살펴보고 각각의 문제점에 대한 검토를 토대로 평가를 받는 번역 주체뿐만 아니라 평가를 시행하는 평가 주체에게도 상호 납득할만한 객관적이며 평준화된 번역의 품질 평가

¹ 주관적인 평가와 관련하여 장혜선(2012;145)에서는 김진아 외(2002)에서 기술된 흥미로운 내용이 소개되고 있다. 두 명의 평가자에게 한 학생의 번역물의 채점을 의뢰한 결과, 평가자 A 에게는 2 위에 해당하는 점수를 받았지만 평가자 B 에게는 34 위라는 매우 저조한 점수를 받았다는 것이다. 이와 같이 번역의 품질 평가에 있어서 평준화된 평가 기준이 없다면 객관적인 평가 결과를 기대하기 힘들다.

기준을 제시하고자 한다.

2. 선행연구

먼저 선행연구의 개관에 앞서 번역의 평가와 관련된 용어들을 간단히 정리해보고자 한다. 번역과 관련된 평가 분야에서는 ‘품질 보증, 품질 검사, 품질 평가’ 등 다양한 용어가 혼재되어 사용되고 있는데 일반적인 어휘의 사용 의미, 더불어 행위를 둘러싼 주체와 객체와의 관계를 고려하면 각 용어 간에는 다소 차이가 있다고 생각된다.

먼저 ‘품질 보증(Quality Assurance)’이란 말 그대로 제품 또는 시설이 정상적으로 가동한다는 충분한 확증을 얻는 데 필요한 모든 작업을 지칭하는 것으로서 통상 생산자가 주체가 되어 실시하는 활동을 뜻한다. 따라서 번역 분야에 있어서도 품질 보증이라 함은 번역을 하는 주체(생산자)의 입장에서 번역을 의뢰한 대상(소비자)의 요구에 대한 만족도를 검토하는 의미, 즉 번역가가 자체적으로 실시하는 검증 행위를 가리키는 것으로 해석할 수 있다.

반면에 ‘품질 검사(Quality Check)’는 품질의 요구 조건을 확인하기 위하여 직접 제품을 관찰, 관측하는 활동이며, ‘품질 평가(Quality Assessment/Evaluation)’는 품질을 평가하는 활동으로서 양자는 모두 평가의 주체가 반드시 생산자에 국한되지 않는다는 점에서 위의 ‘품질 보증’과는 다른 의미로 해석된다. ‘품질 검사’도 ‘품질 평가’도 공통적으로 정량, 정성평가를 포함하면서 대상의 가치를 판단하는 행위이지만, 번역이라는 활동은 단순히 정해진 기준에 대한 수치 측정이 아니라 번역을 수행하는 언어 능력에 대한 보다 포괄적인 상대 평가이기 때문에 본고에서는 ‘품질 검사’가 아닌 ‘품질 평가’라는 용어를 사용하기로 하겠다.

2016년 11월 15일 구글이 한국어, 영어, 중국어, 프랑스어, 스페인어, 일본어, 터키어 등 8개 언어에 대하여 기존의 통계기반 기계번역에서 인공지능망 기계번역으로 적용하겠다고 발표한 이후 기계번역의 품질에 대한 개별 연구의 패러다임이 변화하고 있다. 먼저 한일 기계번역의 품질 평가와 관련된 연구 동향을 살펴보면 다음과 같다. 표1의 연구 목록은 기초학문연구센터 사이트에서 ‘기계번역, 번역 품질, 품질 평가’ 등의 키워드로 검색된 논문 중에서 한일 번역 관련 연구 논문들을 연도별로 정리한 것이다.

<표 1> 한국어와 일본어의 기계번역의 오류 및 품질에 관한 선행연구

신석기	2008	한국과 일본의 기계번역 예에 관한 비교연구 -한국어의 일본어 역을 중심으로-	日本文化學報39, pp.5-20, 한국일본문화학회
이길원	2010	日本語「～ていただく」の韓国語機械翻訳における問題の実際と代案 -ビジネス日本語のE-mailの内容の分析を中心として-	일어일문학48, pp.265-283, 대한일어일문학회
조남성	2011	한국인 일본어 학습자와 한일 기계번역의 어휘 오류 비교 연구	일본연구3, pp.197-216, 일본연구소
이하자	2011	한·일 기계번역시스템의 문제점 고찰 -번역가 번역문과의 대조분석을 통해-	日本語教育58, pp.151-166, 한국일본어교육학회
장혜선	2011	목적별 번역 평가 모델에 관한 소고_국내외 번역품질평가 분석을 중심으로-	외국어교육연구25권(1호), pp.107-136, 외국어교육연구소
최영수	2012	기계번역에 나타난 한일 번역 양상 -부정표현을 중심으로-	번역학연구13권(4호), pp.211-233, 한국번역학회
장혜선	2012	진단적 번역품질평가 모델 구축을 위한 오류 항목 분류에 관한 연구	일본어문학59, pp.143-178, 일본어문학회
김옥순	2014	翻訳教育の効率性向上のための一考察-機械翻訳文の誤訳分析を中心に-	일본어문학66, pp.23-44, 일본어문학회
박순애	2014	일·한 기계번역과 일본어교육	일본어문학67, pp.97-122, 일본어문학회
장혜선	2014	번역평가를 위한 번역 오류 기술 및 분류에 관한 소(小)연구 -통번역대학원 수강생들의 일한(JK) 번역물 오류 분석을 중심으로-	日本語文學62, pp.99-132, 한국일본어문학회
김옥순	2015	韓・日翻訳授業に関する少考-現代小説の誤訳関連論を中心に-	일본어문학71, pp.71-94, 일본어문학회
윤호숙	2018	일한기계번역의 오류 유형에 관한 고찰-일본 소설을 중심으로-	일어일문학연구107, pp.3-25, 한국일어일문학회
이주리아	2018	인공신경망 기계 번역의 한일/일한 번역 품질에 대한 예비 연구-품질향상 검토와 교열 코드 적용-	통역과번역20(1), pp.43-71, 한국통역번역학회
박혜경	2018	석사 과정의 기계번역 수업에 대한 소고: 한일 번역 전공생의 포스트에디팅 사례를 통하여	번역학연구19(3), pp.163-193, 한국번역학회
이주리아	2020	피드백 항목에 따른 일한 번역 피드백 고찰-교수자와 졸업생 평가를 중심으로-	日本語教育93, pp.75-92, 한국일본어교육학회

먼저 전체적인 연구 동향을 살펴보면 한일 기계번역에 관한 연구는 2008년부터 시작되고 있는데 이것은 온라인 상의 무료 번역 서비스가 시작된 시기(네이버:2002년부터, 구글:2007년부터)를 계기로, 다시 말해서 기계번역의 상용화와 더불어 본격적으로 연구가 개시되었다고 말할 수 있다.

또한 한일/일한 번역과 관련된 연구는 <표1>에서 볼 수 있듯이 다른 어학 연구 분야에 비해 수적으로 많지 않은데 전체적인 흐름을 개관해보면 구글이 인공지능망 기계번역으로 전환한 2016년을 기점으로 양적으로나 질적으로 그 양상이 달라지고 있음을 알 수 있다.

다시 말해서 온라인 상의 무료 서비스가 개시된 2007년부터 2016년까지의 10년 동안은 텍스트의 유형별, 언어 표현별로 다양하게 기계번역의 오류와 문제점을 고찰하고 있을 뿐만 아니라 번역의 품질 평가에 대한 기준 검토 및 번역 교육이나 일본어교육과의 관련성에 이르기까지 폭넓게 확장되고 있었다. 그러나 인공지능망 기계번역이 등장한 2016년 이후부터는 예상과 달리 기계번역을 테마로 한 연구 자체가 감소하였으며 연구 내용에 있어서도 포스트에디팅과 같이 새로운 주제 연구가 등장하고는 있지만 아쉽게도 2016년 이전과 비교하여 연구 주제도 한정적이며 제기되었던 논제들에 대한 지속적인 심화 연구도 이루어지지 않고 있는 실정이다².

다음으로 선행연구의 기계번역에 대한 오류 분석 및 품질 평가의 연구 동향은 다음의 2가지로 요약할 수 있다.

첫번째로 선행연구의 오류 분석은 주로 자연언어 처리 연구에서 사용되는 방식과 동일하게 각각의 기계번역의 결과물로부터 오류를 추출, 분석하는 연구 방법이 주류를 이루어왔다. 그러나 이와 같은 용례 중심의 ‘보텀업(bottom-up)’ 방식은 실제의 사용을 토대로 구체적이고 세밀하게 오류의 양상을 제시할 수 있다는 장점이 있는 반면 텍스트의 종류나 연구자 개인의 기준에 따라 오류의 양상이 균일하지 않기 때문에 한편으로는 사례 연구라는 비난을 벗어나기 어려운 면이 있다.

두번째로 선행연구의 오류 분석이나 품질 평가에서는 품사나 문법과 같이 언어의 구조나 형식을 기준으로 삼는 경우가 많았다. 특히 기계번역은 딥러닝을 통하여 끊임없이 번역 능력을 향상시키고 있다는 점에서 오류 분석에 있어서도 언어 습득 과정이나 외국어 교육에서 실시되는 오용 분석 방식을 그대로 적용하는 경우가 많았다. 그러나 앞서 기술

² 참고로 기초학문연구센터에서 기계번역, 번역품질, 품질평가 등의 키워드로 전체 번역 관련 논문을 검색해보면 다른 언어권의 번역 연구에서는 2016년 이전과 이후를 비교해도 양적으로나 질적으로 별다른 차이가 없다는 것을 알 수 있다.

³ 일본어 학습자의 오용에 대한 분류 방법으로서 가장 일반적인 것은 이치카와야스코(市川保子 2001)이다. 이치카와(2001:15)는 오용을 크게 ①탈락(脱落, omission), ②부가(附加, addition), ③어형성(語研成, misformation), ④

한 바와 같이 번역은 이문화 간의 커뮤니케이션을 가능하게 하는 ‘중개 활동(mediation)’으로써 일반적인 작문 평가와는 다르기 때문에 이러한 특성을 무시한 채 언어의 구조나 형식 위주로만 평가하는 것은 종합적인 번역 능력을 평가하기에는 한계가 있다고 생각된다. 따라서 무엇보다도 번역의 품질 평가를 위한 객관적이며 일관성이 있는 평가 기준의 마련이 시급하다고 말할 수 있다.

3. 한일 번역의 품질 평가 기준의 검토 및 문제점

앞서 <표1>에서 살펴본 바와 같이 한일 번역 분야의 선행연구에서 품질 평가 시스템 자체를 대상으로 한 연구는 그리 많지 않다. 선행연구 중에서 번역물의 구체적인 평가 모델을 제시한 연구로는 장혜선(2011, 2012, 2014)과 이주리에(2018, 2020)를 들 수 있다. 두 연구는 모두 기존의 평가 기준을 한국어와 일본어의 특징에 맞게 수정 보완하여 적용하고 있다는 공통점이 있다.

먼저 장혜선(2011, 2012, 2014)은 일본의 (주)히타치테크니컬커뮤니케이션즈에서 20년 이상 번역 감수를 해 온 에디터인 케빈모리세리 (Kevin Morrissey)가 다년간의 경험을 토대로 개발한 TES (Translation Evaluation System) 라는 평가 시스템을 활용하고 있다. 이 평가 시스템은 총 9개의 범주와 27개의 에러 코드(번역오류항목)로 구성되어 있으며 각각의 에러 코드에 해당되는 오류의 수를 세어 감점하는 감점 방식의 번역 평가 시스템이다.

아래의 평가 항목에서 볼 수 있듯이 TES 는 원래 영일 번역을 대상으로 한 평가 기준이기 때문에 장혜선(2012:154-169)은 학생들의 실제 번역물을 토대로 한일 번역 평가와 부합하지 않는 항목을 삭제하거나 일부 변경하는 등의 수정을 거쳐 활용하고 있다. 구체적인 평가 항목과 장혜선(2012:154-169)에서 제시된 수정 내역은 다음과 같다.

혼동(混同, alternating form), ⑤위치(位置, misordering), ⑥기타의 6 가지로 분류하고 있다. 이 분류 방법은 주로 일본어 학습자의 오용 분석에 활용되지만 기계번역의 번역 오류를 분석하는 경우에도 그대로 활용되는 경우가 적지 않다.

<표 2> TES 의 평가항목와 장혜선(2012)의 수정 보안 내용

分類	Code	説明	장혜선(2012)
文法	01	不正確な冠詞: 冠詞(a, an, the)가間違っている。	삭제
	02	主語と動詞の不一致: 主語と動詞が一致していない。	삭제
	03	単数、複数の間違い: 複数であるべき名詞が単数、またはその逆	일부수정: 間違 트 나 名 詞 의 複 수 형
	04	前置詞の間違い: 前置詞が間違っているか、不要、または不自然	일부수정: 助詞의 間 違 이
	05	句読点の間違い: 句読点が間違っている。	일부수정: 記号 와 符 号 의 間 違 이
	06	文法の間違い: 上記以外の文法エラーがある。	그대로 적용
綴り	07	綴りの間違い: 単語の綴りが間違っている。	그대로 적용
不自然な 英語	08	不自然な英文: 英語가不自然	일부수정: 不自然 な 和 文
意味	09	誤訳された意味: 誤訳されているため、英文が事実に関して不正解であり、和文で意図されている内容と矛盾している。	Code09+10+11 통합
	10	事実に関して不正確な意味: 実エラー、または矛盾がある。	
	11	不明瞭な意味: 文章全体または一部が、曖昧か不明瞭である、または理解できない。	
	12	繰り返される意味: 類語反復および回りくどい定義、不必要な繰り返し、あるいは冗長さが含まれている。	Code12+13 통합하 여
テクニカル ライテ ィングの 問題など	13	長過ぎる文: 文が長すぎて理解しにくい。	테크니컬(기술)범 주 로 이 동
	14	動作の主体不明: 誰または何か動作を行うのか不明であり、明示する必要がある。	그대로 적용
	15	用語가不明瞭: 指示語(通常, it, they, this, those)などの代名詞가曖昧または間違っている。	그대로 적용
	16	修飾語가不明瞭: 被修飾語가不明瞭または間違っている。おそらく修飾語と被修飾語가離れすぎているため。	Code 16+17+18 통합
	17	語順가不明瞭: 語順가読み手を混乱させる。用語または句가どこから始まり、どこで終わっているのか가不明瞭	
	18	目的語가不明瞭: 動詞의目的語가不明瞭。おそらく動詞と目的語가離れすぎているため。	
19	受動態가不適切: 受動態가不必要に使われている。能動態가望ましい。		
不適切な 動詞의使 い	19	受動態가不適切: 受動態가不必要に使われている。能動態가望ましい。	Code 19+20 통합
	20	動詞構文가不明瞭: 選択部分に、不要な助動詞、動作の名	

		詞形、不要な否定形などが含まれている可能性がある。	
用語	21	長過ぎる用語：用語の修飾語が多過ぎる。または、用語が、理解しにくい単語の羅列で構成されている。	그대로 적용
	22	未定義の用語：用語または略語の定義が然るべき場所に明記されていない。	Code 22+23 통합
	23	不適切な用語：用語が不明瞭、不正確または不適切。	
	24	辞書と矛盾する用語：用語が、奨励された翻訳辞書と矛盾する。	삭제
不統一	25	句の不統一：二つ以上の異なる用語または句が、同じ意味のことを指している。	Code 25+26 통합
	26	構文の不統一：並列文（見出し、リスト項目、クロスリファレンスなど）に複数の類似形式が使われていて、一貫性がない。	
스타일 가이드	27	선택部分가 스타일 가이드에相反: 선택部分가, 奨励された 스타일 가이드 또는 체크리스트에相反する。	삭제

<표2>에서 볼 수 있듯이 영어와 한국어의 언어유형적인 차이로 인하여 TES의 상당 부분은 삭제되거나 통합되었다. 영일·일영 번역 평가를 위해 개발된 TES는 번역 평가 모델 중에서는 가장 평가 항목이 세분화된 평가 시스템이지만 한일 번역 평가를 위해 수정 보완된 장혜선(2012)의 모델에서는 ‘主語と述語の不一致(주어와 술어의 불일치)’와 같이 새로 만들어진 항목도 있지만 평가 항목 간의 통합과 삭제를 거쳐서 결과적으로는 평가 항목이 대폭 축소되었으며 이후 장혜선(2014;113-121)에서는 상기 오류들의 항목을 보다 더 포괄적으로 ‘문법, 의미, 번역 기술’의 3가지로만 분류하고 있다.

이와 같이 장혜선(2011, 2012, 2014)의 일련의 연구는 한일 번역의 품질 평가 기준 자체를 고찰한 첫번째 본격적인 시도로서 번역의 다양한 특징을 평가 항목으로 도입하고 있다는 점에서도 높이 평가할 수 있다. 다만 장혜선(2012)은 적극적인 비판 없이 TES를 절충적으로 활용하고 있지만 TES 자체는 체득한 경험을 토대로 구체적인 사례 중심으로만 구성이 되어 있기 때문에 전체적으로 개별 항목 간의 관계성 및 분류의 체계성에 있어서 다소 문제가 있다고 생각된다.

가장 문제가 되는 것은 문법 범주의 항목들이다. <표2>의 ‘テクニカルライティングの問題など(테크니컬한 쓰기 문제)’의 대부분은 ‘술어와 대상어, 수식어, 목적어, 어순’과 같이 대부분 문법 문제에 해당하지만 제일 먼저 열거된 ‘文法(문법)’ 항목과는 별도로 설정되어 있기 때문에 분류의 체계성에 대한 의문이 생긴다. 또한 ‘不適切な動詞の使い(부적절한 동사의 사용)’은 일반적으로 문법의 범주에 해당함에도 불구하고 왜 별도로 분류되고 있는지, 왜 유독 태(voice)와 동사 구문에 관련된 내용만 ‘不適切な動詞の使い(부적절한 동사

의 사용)에 해당되는지, 그럼 다른 동사의 오용은 어디에 분류되는지 쉽게 납득하기 어렵다.

또한 어휘론의 관점에서 용어에 대한 평가 항목은 설정이 되어 있지만 기본적으로 어휘 자체에 대한 평가 항목은 없기 때문에 한일 번역에서 오류가 많이 발생하는 ‘동음이의어’와 같은 오류의 처리는 분류하기 어렵다는 점, 그리고 코드 06의 ‘不自然な和文：日本語が不自然(부자연스런 일본 문: 일본어가 부자연스러움)’은 그 내용이 광범위하며 다른 평가 항목과의 중복을 피할 수 없다는 점 등을 들 수 있다.

위와 같은 평가 기준 자체의 모순으로 인하여 실제로 분류를 실행해보면 코드의 중복 처리 문제와 코드에 분류되지 않은 오류가 적지 않기 때문에 한일 번역 평가에 활용하려면 무엇보다도 먼저 분류의 기준과 체계를 재정비할 필요가 있다고 생각된다.

다음으로 이주리아(2018,2020)는 豊島知恵의(2016)의 ‘MNH-TT校閲カテゴリ体系(교열 카테고리 체계)’를 수정 보완한 평가 기준을 도입하고 있다.

‘MNH-TT校閲カテゴリ体系(교열 카테고리 체계)’는 협동 번역 교육의 관점에서 영한 번역 학습자들에게 일관성 있는 피드백을 제공하고자 豊島知恵의(2016)에 의해 개발된 에러의 분류 체계로서, 특히 협동 번역 교육의 일환으로서 공동 번역 플랫폼인 ‘みんなの翻訳実習(모두의 번역 실습, 이하 MNH-TT)’을 통하여 번역 학습자와 교수뿐만 아니라 번역 학습자 상호 간에도 번역에 관한 지식을 공유하고 번역 학습을 효율적으로 하기 위하여 고안되었다.

‘MNH-TT校閲カテゴリ体系(교열 카테고리 체계)’의 특징은 오류를 분류함에 있어서 번역 학습자의 레벨(언어 능력)뿐만 아니라 어휘, 문법, 의미와 같은 언어 구조와의 연계성을 유기적으로 구현하고 있다는 점이다. 학습자의 언어 능력 별 분류 내용과 ‘校閲事例(검열사례)’로 언급된 유형을 함께 제시하면 다음과 같다.

⁴ ‘みんなの翻訳実習(MNH-TT)’ 사이트(<https://mt-auto-minhon-mltucrjgn-x.jp/a/edu/>)는 2018년 5월 30일 10시부터 서버 이전에 따라 서비스가 일시 중단된다는 공지만 있을 뿐 2021년 8월 현재까지도 운영이 중단된 상태이다.

<표 3> 豊島知恵 외(2016)의 MNH-TT 校閲カテゴリ体系

レベル	ID	校閲カテゴリ	校閲事例
Lv1 訳が未完成である	X4a	未翻訳	CONTENT
	X6	曖昧さ未解消	CONTENT
Lv2 起点言語文書の要素に対して 過不足や誤解がある	X7	用語の訳語誤り	LEXIS
	X1	原文内容の欠落	CONTENT
	X2	原文にない要素の付加	CONTENT
	X3	原文内容の歪曲	CONTENT
Lv3 目標言語の文法的・統語的な問 題がある	X8	コロケーションの誤り	LEXIS
	X10	前置詞や助詞の誤り	GRAMMAR
	X11	活用の誤りや数・性などの不一致	GRAMMAR
	X12	綴り誤り・誤変換	GRAMMAR
	X13	句読法に関する誤り	GRAMMAR
	X9	その他の文法的・統語的な誤り	GRAMMAR
Lv4 目標言語文書に質的な問題が ある	X16	結末性違反	TEXT
	X4b	直訳調	CONTENT
	X15	表現のぎこちなさ	TEXT
Lv5 納品・公表するプロダクトとし ての問題がある	X14	レジスタ違反	TEXT

현재 ‘みんなの翻訳実習(모두의 번역 실습)’ 사이트는 중단이 되어 구체적인 절차와 내용은 확인할 수 없지만 번역 학습자들은 상기의 MNH-TT를 기준으로 교수자 및 번역 학습자 간의 상호 협동을 통하여 번역 오류를 진단하고 수정 지도를 받았을 것으로 추정된다. ‘MNH-TT校閲カテゴリ体系(교열 카테고리 체계)’는 오류 분류에 있어서 의사결정나무(Decision Tree)도 함께 제공하고 있기 때문에 비교적 평준화된 평가가 가능한 구조로 설계되어 있다는 점에서 번역 오류의 판정 및 분류를 위한 기초 자료로서 매우 유용하다고 판단된다.

이주리에(2018:64)는 한일 기계번역의 결과를 포스트에디팅해서 납품 단계에 이른다

⁵ 豊島知恵 외(2016:53) 참조

는 가정 하에 품질 평가를 목적으로 <표3>의 ‘MNH-TT校閲カテゴリ体系(교열 카테고리 체계)’를 아래의 <표4>와 같이 재구성하였다. 특히 <표3>의‘校閲事例(교열 사례)’를 학습자의 레벨과 연동시켜서 애초부터 레벨 자체를‘Lv1(단어), Lv2(구), Lv3(문장), Lv4(텍스트)’의 4가지 언어 형식으로 대치하고 있다는 점에서 특징적이다.

<표 4> 이주리아(2018)의 ‘MNH-TT 校閲カテゴリ体系’ 수정안

분류		코드	내용	
Lv 1 :단어	인식	X1	어휘	①누락: ST에 있는 단어가 TT에는 없음
			인식 오류	②첨가: ST에 없는 의미불명의 단어 및 음절이 TT에 있음
				③미번역: ST 문자가 그대로 TT에 남아있음
				④완성도 미흡: 번역된 단어가 미완성이거나 이해 불가능한 형태임
				⑤미학습 어휘: 외래어 및 고유명사가 아닌데도 음역하거나 이해 불가능한 단어로 번역됨
				⑥뭉음 인식 오류(복합어 등)
				⑦기타 인식 오류
	X2	외래어 및 외국 인명, 지명 등의 음역 오류 (외래어 표기법 위반, 음역 안 하고 번역 등)		
	X3	띄어쓰기 오류		
	의미	X4	의미 선택 오류에 따른 오역 (다의어, 동음이의어, 기타 어휘 선택 오류)	
Lv 2 :구	관계성	X5	연어관계, 관용표현 파악 오류	
	문법	X6	조사, 조동사 오류(오역, 누락 등)	
		X7	용언처리 오류(활용, 시제, 태, 상 등)	
		X8	기타 문법 오류	
Lv 3 :문장		X9	문장 내 연결(연결어미 등)	
		X10	문장구성 및 표현, 문체	
Lv 4 :텍스트		X11	일관성(텍스트 내 같은 어휘, 문체의 통일성)	
		X12	독자에 맞춘 어휘의 명확화, 전환 등	

위와 같이 이주리아(2018)가 제시한 품질평가 시스템은 Lv1(단어), Lv2(구), Lv3(문장), Lv4(텍스트)와 같이 언어학적인 관점을 기준으로 분류가 되어 있기 때문에 비교적 알기

쉽고 기계번역의 오류 경향을 파악하기 좋다는 장점이 있다. 그러나 이 평가 시스템을 활용하여 실제로 필자가 수집한 기계번역의 오류⁶를 분석해 본 결과, 다음과 같이 중복되거나 분류할 수 없는 예가 적지 않았다.

(1)	ST	사람들은 은을 <u>밀수출했다</u> .
	HT	人々は銀を <u>密輸出した</u> .
(마)	MT	人々は銀を <u>密輸出した</u> .

(2)	ST	지금 한국은 <u>갈라파고스</u> 신세다.
	HT	韓国は今、 <u>ガラパゴスのような状態だ</u> .
(의)	MT	今、韓国は <u>ガラパゴスの身の上だ</u> .

먼저 (1)은 일차적으로 형태소 누락으로 ‘X1①’의 ‘어휘 인식 오류’에 분류할 수 있지만 ‘密輸出’와 ‘密輸’의 어휘 선택 오류로서 ‘X4’의 ‘의미’문제도 될 수도 있으며 구문론의 관점에서는 술어로 사용되고 있으므로 ‘X7’의 ‘용언 처리 오류’로도 분류될 가능성도 있다. 이와 관련하여 위와 같이 오류 분석표에서는 ‘의미’의 오류가 ‘X4’의 단어 레벨로만 한정되어 있으나 일반적으로 ‘의미’의 문제는 단어와 구, 절, 문장 전체 등 모든 언어 형식과 관련이 있으므로 ‘의미’범주 다음으로 (2)의 경우는 직역을 한 것이 오히려 오류가 되는 경우로서 이 경우는 단순한 어휘 선택의 오류나 용언 처리의 오류라기보다는 직역과 의역이라는 번역 기술과 관련되어 있다. 기계번역의 경우에는 이와 같이 완벽한 직역에서 유래한 오류의 패턴이 많다는 점에서 특징적이지만 이주리아(2018)에 따르면 ‘X10:문장 구성 및 표현, 문체’에 분류할 수 밖에 없기 때문에 이 분류로는 기계번역에 있어서 가장 특징적인 오류 패턴을 파악하기 어렵다.

이상과 같이 이주리아(2018)의 품질 평가 시스템은 직관적으로 알기 쉬운 ‘단어, 구, 문장, 텍스트’를 기준으로 하고 있지만 번역의 특성을 고려한 평가 항목이 충분하지 못한 탓에 언어지식적인 평가 결과가 되기 쉽다. ‘X11:일관성(텍스트 내 같은 어휘, 문체의 통일성)’이나 ‘X12:독자에 맞춘 어휘의 명확화, 전환 등’과 같이 보다 다양하게 번역의 특징이나 기술과 관련된 평가 항목의 도입이 필요하다고 생각된다.

⁶ 용례 자료는 2020년 2월에서 3월까지 조선일보와 중앙일보의 홈페이지에 게재된 12개의 신문 논설문, 칼럼 등이며 해당 신문사가 제공하는 같은 글의 일본어판 번역문을 HT의 샘플로 하고 있다. 총 글자수는 17,700이다.

이후 이주리에(2020)는 번역 교육의 현장에서 학습자의 번역 과제물에 대한 일관성 있는 피드백을 목적으로 상기한 MNH-TT校閲カテゴリ体系(교열 카테고리 체계)의 16개 항목 중 발견되지 않은 항목을 삭제하고 유사한 항목을 통합하여 ‘①용어 오역 ②누락 ③첨가 ④내용 왜곡 ⑤표기/맞춤법 ⑥문법 오류 ⑦문맥 연결 및 일관성 ⑧직역 및 어색함 ⑨문체 부적합’의 9개 분류 기준을 적용하였으나 피드백 항목에 대한 평가자의 이해 부족 및 적당한 피드백 항목의 부재, 다시 말해서 평가 항목에 대한 문제점으로 인하여 아쉽게도 평준화된 평가 결과를 도출하지 못하였다고 자체 결론짓고 있다. 커뮤니케이션 활동으로서 번역의 특성을 고려한 보다 객관적이고 체계적인 기준 마련이 시급하다.

4. 한일 번역의 품질 평가(KJTQA) 모델

이상과 같이 번역의 품질 평가에 있어서는 무엇보다도 어떤 항목들을 기준으로 삼고 있는지 평가 지표의 설계에 따라 모든 것이 좌우된다. 이제까지 한일 번역을 대상으로 한 선행연구에서는 주로 미시적인 관점에서 언어학적인 구조나 형식을 중심으로 평가 지표가 설계되어 왔다. 그러나 본고에서는 ‘완결성(Completeness), 정확성(Accuracy), 가독성(Readability)’과 같이 일반적인 언어 이해 및 산출 능력과 관련된 텍스트의 요소를 상위 카테고리로서 설정함으로써 매개활동이라는 번역의 커뮤니케이션 특성을 최대한 고려하고자 하였다. 이러한 요소들은 기본적으로 언어 및 외국어 학습에 있어서 ‘읽기, 쓰기, 말하기, 듣기’의 언어 이해 및 산출 과정에서 필수불가결한 요소로서 일찍부터 번역 감수의 기준으로도 활용되었던 것이다. 구체적인 평가 지표 및 평가 기준은 다음과 같다.

4.1 평가지표

여기서는 구체적인 평가 모델을 설계하기에 앞서 몇 가지 중요한 평가지표에 대하여 명시해 두고자 한다.

1) 대상 언어

선행연구를 통하여 대상 언어가 달라지면 평가 항목도 달라질 수 밖에 없다는 사실, 즉 영어를 대상으로 하는 평가 항목을 그대로 한국어에 적용시킬 수는 없다는 것을 확인하였다. 본고의 평가 모델은 기본적으로 한국어와 일본어의 언어쌍을 대상으로 하고 있

⁷ 장혜선(2014:110-112)에서 조사된 번역 감수의 기준을 살펴보면 의미 전달에 있어서의 정확성과 완결성, 가독성 등을 공통으로 들고 있다.

다. 다른 언어의 번역을 대상으로 하는 경우에는 상위 개념에 의거하여 각각의 하위 평가 항목을 조정하여 활용할 수 있다.

2) 평가 항목의 재배치

본고에서는 거시적인 측면에서 ‘완결성, 정확성, 가독성’과 같은 텍스트의 특성을 상위 카테고리로 설정하고 미시적인 측면에서 언어의 구조와 형식에 기반한 언어지식적인 평가 항목을 하위 카테고리에 배치함으로써 번역이라는 커뮤니케이션의 특성과 언어학적인 지식 및 능력을 함께 고려하고자 하였다.

3) 감점제와 가점제의 융합

대부분의 품질 평가 모델과 마찬가지로 본고에서도 기본적으로는 오류의 개수를 세어 가는 정량평가 방식을 기본으로 하여 감점의 수치가 높으면 높을수록 낮은 품질을 의미하게 된다. 더불어 독자를 배려한 훌륭한 번역 전략의 경우에는 가점을 줄 수 있도록 설계하였는데 이것은 매개활동으로서 번역이 갖는 특성을 평가 항목에 반영한 것으로서 다른 품질평가표에서는 거의 찾아볼 수 없다.

4) 오류 측정의 다원화

어떤 번역이라도 실제 평가를 해보면 오류에도 정도가 있다는 사실을 알 수 있다. 따라서 심각한 오류에 대해서는 확실하게 가중 벌점을 부과하기 위하여 본고에서는 오류의 개수에 대한 합산과 더불어 오류의 정도에 따라 ‘심각(Critical), 중도(Major), 경도(Minor), 없음(None)’의 4 단계로 나누고 각각 ‘-2 점, -1 점, -0.5 점, 0 점’의 기본값을 설정하여 평균화된 평가 결과가 나올 수 있도록 설계하였다.

오류의 심각도를 판정할 때는 먼저 ‘중도(-1 점)’를 기준으로 오류의 심각도에 따라 가중하거나 경감하는 방식을 취한다. 오류의 심각도는 ST의 의미를 기준으로 판정하는데 ST의 의미가 심각할 정도로 손상, 변형되거나 의미를 아예 알 수 없는 경우에는 ‘심각’으로 가중 감점하고, 다소 부자연스럽지만 ST의 의미에 손상이 없는 경우에는 ‘경도’로 판정하여 -0.5 점의 감점을 부가할 수 있다.

4.2 오류 카테고리

오류 카테고리에서 가장 문제가 되는 것은 분류 항목에서 누락된 오류들과 중복 분류의 가능성을 가진 오류들의 처리이다. 본고에서는 가능한 한 다양한 관점과 분류를 포섭하여 미분류 및 중복 분류의 예가 발생하지 않도록 각각의 평가 기준과 항목을 구체적으로 명시하였다. 여기서는 실제로 수집한 기계번역(이하 ‘MT’)의 오류⁸에 대한 분류를 통

⁸ 본 연구에서는 2020년 3월에 조선일보와 중앙일보의 홈페이지에 게재된 신문 사설, 칼럼 등의 ‘오피니언’을

하여 ‘완결성, 정확성, 가독성’의 기준을 검토하고 각각의 세부 오류 내용을 구체적으로 검토해보고자 한다.

4.2.1 완결성

여기서 ‘완결성’이란 번역의 품질을 논하기 이전에 번역 자체가 평가의 대상으로서 완성되었는지 여부를 판단하는 중요한 요소이다. 즉 출발어 원문을 충분히 이해하고 충실하게 번역하였는지를 평가하는 항목이다. 따라서 ‘완결성’에 심각한 문제가 있는 경우라면 품질 불량으로 판정한 후 평가를 반려하고 재작업을 통하여 번역물 자체를 완성하도록 요구할 수 있다.

‘완결성’에 위배된 구체적인 오류 내용은 ‘1) 인식 오류, 2) 미번역, 3)누락, 4) 부가’와 같은 항목을 들 수 있으며 기타 ‘완결성’을 위배한 오류 가운데 상기 분류에 해당하지 않는 것이 있으면 기타에서 추가로 설정이 가능하다.

1) 인식 오류

‘인식 오류’은 (3)과 같이 문자가 깨지는 경우를 가리키는데 이런 현상은 주로 기계번역의 오류에서 발생한다. (3)은 고유명사(인명)에 대한 번역 오류이며 특히 오류의 배경에는 ‘동음이의어’와의 혼동을 예측해 볼 수 있다. 같은 인명이라도 ‘조국’에서는 문자가 깨지는 오류가 발생했지만 ‘조민’에서는 제대로 번역이 이루어지고 있는데 이것은 고유명사 ‘조국’과 일반명사 ‘조국(祖國)’의 혼동이 문제가 된 것으로 보여진다. 참고로 논문을 집필하는 현재(2021년 8월)에도 이 문장은 여전히 같은 ‘인식 오류’로 구현되고 있다.

(3) <인식 오류>의 예

(의)	ST	<u>조국</u> 전 법무부 장관의 딸 조민씨의 부산대 의전원 부정 입학 의혹 사례에서 보듯(생략)
	HT	<u>チョ・グク</u> 前法務部長官の娘、チョ・ミン氏の釜山(プサン)大学医学専門大学院不正入学疑惑の事例で分かるよう、
	MT	<u>国(チョ・グク)</u> 前法務部長官の娘、チョ・ミン氏が釜山(プサン)大学医学部への不正入学疑惑の事例からも分かるように、

분석 대상으로 삼았다. 대상 신문사는 모두 동일한 내용의 일본어 번역본을 함께 게재하고 있기 때문에 인간번역(Human translation)과의 대역 대조가 용이하며 일반 보도문에 비해 속담이나 관용표현, 비유표현과 같이 언어 표현이 다양하므로 상대적으로 오류의 양상도 다양하게 관찰될 것으로 예측되었기 때문이다.

2) 미번역

‘미번역’은 ST의 일부분이 도착어의 언어 형식으로 변환되지 못하고 그대로 남아 있어서 결과적으로 번역물 안에 출발어와 도착어가 함께 병기되어 있는 경우를 말한다. 번역자의 실수로 남겨진 경우도 있겠지만 번역 능력의 문제로 너무 어려워서 번역하지 못한 채 남겨진 경우 등을 예측할 수 있다. 이 오류는 주로 HT에서는 자주 관찰이 되지만 이번 에 수집한 MT의 오류 조사에서는 한 예도 발견되지 않았다.

3) 누락

‘누락’은 주로 단어 레벨에서 관찰되는 오류로서 ST에 있는 형식이 TT에는 번역이 되지 않고 그냥 없어진 경우이다. 원래 있던 형식이 누락됨으로 인하여 필연적으로 TT의 의미에는 변형이 생기게 되고 따라서 아래와 같이 텍스트 전개 방식에 있어서 부자연스러운 의미가 될 확률이 높다.

(4) <누락>의 예

(을)	ST	1896년 첫 근대 올림픽이 시작된 이후 대회가 취소된 건 다섯 차례(하계 3번, 동계 2번)뿐이다. 모두 제1·2차 세계대전 때였다.
	HT	1896年に最初の近代オリンピックが始まってから、大会が中止になったのは5回(夏季3回、冬季2回)だけだ。すべて第1・2次世界大戦の時だった。
	MT	1896年、初の近代オリンピックが始まって以来、大会が取り消されたのは5回(夏季3回、冬季2回)だけだ。いずれも第1・2次世界大戦だった。

4) 부가

전체 오류 중에서 ‘부가’의 오류는 12개 논설문, 총 346문 중에서 단 1예였다. 앞서 기술한 바와 같이 MT는 이미 완성된 ST를 전제로 번역물을 산출하기 때문에 ST에는 전혀 없었던 내용이나 형식이 갑자기 부가될 확률은 거의 없기 때문에 이러한 오류가 관찰되는 것은 이례적이다. (5)의 용례를 보면 ST에서는 ‘서넛’이라고만 제시되어 있는 부분에 대하여 HT에서는 당연히 ‘확진자’의 숫자로 판단하여 ‘3, 4인(서너사람)’이라고 번역하고 있지만 MT에서는 전혀 상상하지도 못한 ‘3, 4歳(서너살)’의 나이로 번역하는 심각한 오류를 범하고 있다. 이와 같이 ‘부가’의 오류는 표면적으로는 단어 레벨의 문제로 보이지만 실상은 문맥상 유추가 필요한 경우로서 아직 텍스트의 일관적인 지시에 취약한

⁹ 총 346개의 문 중에서 오류가 발생한 문의 개수는 86문(24.85%)으로서 이것은 대략 4문장 중 1문장 비율로 오류가 발생하고 있다는 것을 알 수 있다. 더불어 오류의 전체 횟수는 90개로서 2개 이상의 중복 오류가 출현한 문장은 4문장뿐이었다.

MT의 한계를 보여주고 있다.

(5) <부가>의 예

(예)	ST	설 연휴 직후 우한 코로나 확진자가 <u>서넛</u> 일 무렵이다.
	HT	旧正月の連休直後、武漢コロナ（新型コロナウイルス感染症）の感染者が <u>3、4人</u> だったころのことだ。
	MT	旧正月の連休明け、武漢コロナ確定者が <u>3、4歳</u> の時だ。

4.2.2 정확성

‘정확성’은 출발텍스트의 형식이나 내용이 과부족없이 충실하게 번역되어 있는가에 대한 평가 항목이다. ‘정확성’의 평가를 통하여 번역가에게 있어서 필수불가결한 출발어에 대한 지식 및 독해력, 그리고 도착어에 대한 지식 및 표현력을 함께 평가할 수 있다. 따라서 ‘정확성’의 세부 내용은 번역가가 가진 언어 지식과 언어 능력 자체를 최대한 평가할 수 있는 언어지식적인 평가 항목이 많아질 수 밖에 없다. 특히 MT의 오류의 빈도를 살펴보면 ‘어휘>문법>표현’의 순으로 어휘의 오류가 가장 많이 관찰된다¹⁰. 여기에서는 ‘정확성’의 항목을 크게 어휘의 오류와 문법의 오류로 나누어 각각의 오류의 유형을 살펴 보도록 하겠다.

1) 어휘의 오류

선행연구의 오류 분석 결과와 마찬가지로 MT 오류 중에서 가장 많이 부분을 차지하는 것은 ‘어휘’선택의 오류이다. 오류의 빈도가 높은 만큼 오류의 유형 또한 다양하다. 먼저 ‘어휘’의 오류 중에서 가장 많은 오류는 다음과 같은 ‘동음이의어’의 혼동이다.

(6)<동음이의어>의 예

(6-1) (팅)	ST	그러나 이제 한국 <u>여권</u> 은 이름 모를 아프리카 섬나라에서도 문전박대당하는 신세로 전락했다.
	HT	しかし今や韓国の <u>旅券</u> は名前も知らないアフリカの島国よりも厳しく門前払いされる状況にまで転落した。
	MT	しかし、もはや韓国 <u>与党</u> は、名も知らぬアフリカの島国でも門前払いされる羽目に転落した。

¹⁰ 박민영(2020;169)참조

(6-2) (의)	ST	사정이 이런데도 굳이 의사와 택배기사 공을 가로채고 싶은 사람들이 최근 눈에 많이 띈다.
	HT	このような状況でもあえて医師と宅配配達員の <u>功績</u> を横取りしたがる人々が最近目に立つ。
	MT	事情がこうであるにもかかわらず、あえて医師と宅配ドライバーの <u>ボール</u> を横取りしたい人たちが最近多い。

‘동음이의어’의 오류가 발생한 경우에는 위의 예문과 같이 원문과는 완전히 다른 의미가 되어 버리기 때문에 오류의 정도는 상당히 심각하다고 볼 수 있다. (6-1)에서는 ‘HT : 旅券(여권)vs. MT:与党(여권)’과 같이 완전히 다른 의미의 단어가 사용되고 있으며 (6-2)에서는 ‘공작’을 의미하는 ‘ST:공’이 MT에서는 물체의 ‘ボール(볼)’로 번역되는 심각한 오류를 확인할 수 있다¹¹. ‘동음이의어’의 오류는 선행연구에서도 하나의 평가 항목으로 자리매김할 정도로 어휘의 오류 중에서는 가장 많고 눈에 띄는 유형인데 이 오류가 발생하게 되면 ST의 의미와 완전 불일치하거나 모순이 되는 경우가 많아지기 때문에 결과적으로 심각한 오류로 판정될 가능성이 매우 높다.

더불어 어휘의 오류 중에서는 ‘용어’에 대한 오류도 자주 관찰된다. ‘용어’ 중에는 단어의 일부를 생략한 형태의 약자나 영어의 이니셜만으로 표기되는 경우도 많기 때문에 번역가에게 요구되는 특정 전문 분야에 대한 지식 범주에는 이러한 용어들에 대한 지식과 이해도 포함되어 있으며 이에 대한 평가 항목 또한 반드시 필요하다.

이번에 수집한 MT 오류 중에는 번역 표현 자체는 틀리지 않았지만 도착어가 사용되는 지역의 관습과 일치하지 않는 오류가 관찰되었다. 이러한 오류는 주로 날짜나 시간, 통화 단위, 주소 형식 등을 나타내는 어휘 레벨에서 관찰되는 경우가 많다. 이 경우는 그 대로 직역하는 것이 오히려 오역이 되는 경우가 많다.

¹¹ 그럼 빅데이터로 학습하는 신경망 기반 기계번역에 있어서 이와 같은 오류가 발생하는 이유는 무엇일까? 이와 관련하여 한 가지 흥미로운 사실은 ‘동음이의어’의 오류는 주어진 ST에 대해서는 틀림없는 오역이지만 MT 그 자체로만 보면 현실 세계에서 벌어지거나 생각할 수 있는 사실로서 성립할 수 있다는 점이다. 예를 들어 (6-1)에서와 같이 ‘한국 여권이 섬나라에서 문전박대당하는 일도 벌어질 수 있으며 (6-2)과 같이 실제로 의사와 택배기사가 손에 든 공을 뺏는 일 또한 벌어지지 않으리라는 보장이 없다. 즉 ST에 대한 번역 결과로 평가하면 생각지도 못한 치명적인 오류가 되지만 MT의 결과물 자체로만 생각해 보면 최소한의 문의 성립은 담보하고 있다고 말할 수 있다.(박민영(2020) 참조)

(7) <지역 관습의 오류>의 예

(상)	ST	<u>엓그제</u> 하루 일본 땅을 밟은 한국인이 겨우 5명이었다고 한다.
	HT	<u>数日前</u> 、日本の地を踏んだ韓国人は1日でわずか5人だったという。
	MT	<u>昨日の一日</u> 、日本の地を踏んだ韓国人がわずか5人だったという。

(7)에서는 ST의 ‘엓그제’라는 단어가 HT에서는 ‘数日前(수일전)’로 의역이 되어 있는데 MT에서는 “昨日の一日(어제 하루)”로 오역이 되어 있다. HT에서 ‘数日前(수일전)’로 의역을 하고 있는 것은 일본인 독자들이 일본어로 번역된 이 기사를 읽게 되는 시간이 항상 당초 기사가 나온 날 보다 늦어진다는 사실을 전제로 한 것으로 매우 적절하다고 말할 수 있다. 그러나 MT의 경우에는 “昨日の一日(어제 하루)”로 번역되어 있어서 선택한 단어 자체도 틀렸지만 이것을 ‘一昨日(그저께)’로 바르게 직역한다 해도 일본인 독자들이 읽는 시간에는 이미 엓그제 일이 아니게 되므로 이 경우는 지역 관습에 따라 시간 표현을 조정할 필요가 있다.

그 밖에도 출발어에 대한 지식과 독해력 및 도착어에 대한 지식과 표현력이라는 관점에서 ‘정확성’을 위배한 어휘의 오류 가운데 상기 분류에 해당하지 않는 것이 있으면 기타 항목에서 추가로 설정 및 평가가 가능하다.

2) 문법의 오류

본고에서는 문장 안의 법칙과 관련된 모든 종류의 오류를 ‘문법’의 카테고리로 분류하였다. 선행연구의 품질평가에서는 ‘어휘, 문법, 표현’의 경계선이 애매모호하여 오류의 중복을 야기하는 경우가 많았으나 본고에서는 형태론적 관점에서 품사의 형식과 관련된 오류 전체와 구문론적인 관점에서 문제가 되는 절(주절, 종속절)과 술어 형식(시제, 상, 태, 모델리티, 대우표현...) 등 문장의 형식에 관련된 모든 오류를 ‘문법’의 범주로 분류하였다.

이번에 수집한 MT의 오류 중에서 ‘문법’의 오류에 해당하는 용례는 ‘어휘’ 오류에 비해 상대적으로 적었으며 ‘문법’ 오류의 대부분은 다음과 같은 ‘격조사’의 오류였다.

(8) <문법> 오류의 예

(팅)	ST	지난 1일 <u>네덜란드의</u> 한국인 유학생이 동유럽 국가 조지아에 갔다가 공항에서 강제 추방당했다.
	HT	今月1日、 <u>オランダの</u> 韓国人留学生が東欧の国ジョージアに行ったところ、空港で強制追放された。
	MT	1日、 <u>オランダから</u> 韓国人留学生が東欧国家のジョージアに行き、空港で強制追放された。

(8) MT 의 번역에서는 ‘~から’가 사용되어 있기 때문에 ‘네덜란드’가 ‘한국인 유학생’이 유학하고 있는 거주지가 아니라 단순한 출발 장소로 표시되어 결과적으로 ST 의 의미에 변형이 일어나고 있다. 참고로 논문을 집필하고 있는 2021년 8월 현재 파과고에서는 이러한 오류가 수정되어 HT 와 마찬가지로 ‘オランダの’로 신출되고 있다.

4.2.3 가독성

일반적으로 ‘가독성’이란 텍스트가 잘 읽히는 정도, 읽고 이해하기 쉬운 정도를 가리키는 말로 가독성이 좋은 문장은 읽어 내려가기 쉬울 뿐만 아니라 이해나 기억이 용이하며 막힘없이 계속해서 읽을 수 있는 특성을 갖고 있다.

이러한 ‘가독성’에 영향을 주는 요소로는 디자인이나 서체, 일러스트와 같은 시각적인 요소도 포함되지만 번역에 있어서의 가독성이란 ‘번역투 없이 자연스럽게 도착언어의 규범과 관습에 맞아 도착독자가 거부감 없이 쉽게 읽고 잘 이해할 수 있는’¹²요소를 가리킨다. ‘가독성’의 가장 큰 특징은 글을 읽는 독자의 시점에서 텍스트 전체를 고찰한다는 점에 있으며 이것은 번역가에게 꼭 필요한 능력 중 하나인 ‘번역을 객관적으로 평가, 고찰하는 능력’¹³과도 무관하지 않다고 생각된다.

앞서 살펴본 ‘정확성’이 문 이하의 단위와 관련된 미시적인 평가 항목이었다면 ‘가독성’은 보다 거시적인 관점에서 텍스트 전체의 일관성이나 논리성, 문체의 통일성 등을 평가하는 항목이라고 말할 수 있다. 따라서 본고에서는 도착텍스트 전체의 완성도나 일관성 측면에서 다음과 같은 지표 항목을 통해 ‘가독성’을 평가하고자 하였다.

- ① 지시어의 일관성 오류
- ② 직역 표현 오류
- ② 문체의 통일성 오류
- ③ 문장 부호의 통일성 오류
- ④ 입력 오류
- ⑤ 독자를 배려한 번역 전략

위의 ① ‘지시어의 일관성 오류’에서부터 ⑤ ‘입력 오류’까지는 -1 점씩의 감점 방식을 취하며 ⑤ ‘독자를 배려한 번역 전략’에서는 +1 점의 가산 점수를 줄 수 있다. 번역에 있

¹² 안인경(2015;150) 참조

¹³ 望月良浩(2014;75)의 ‘번역가에게 필수불가결한 6가지 능력·스킬’ 참조

어서 100 점이라는 절대점수가 만들어지기 어려운 것처럼 좋은 번역의 경우도 ‘좋다’는 기준점을 설정하는 것은 결코 쉽지 않다. 그러나 “번역투 없이 자연스럽게 도착언어의 규범과 관습에 맞아 도착독자가 거부감 없이 쉽게 읽고 잘 이해할 수 있는’ 훌륭한 번역의 경우에는 어떤 방식을 취하더라도 ‘도착어 독자를 배려한 번역 전략’을 높이 평가하는 것은 매우 당연한 일이라고 생각된다.

‘가독성’의 평가 항목에서 ① ‘지시어의 일관성 오류’는 다음과 같이 텍스트의 일관적인 지시가 서툰 MT의 번역에서 자주 관찰되는 오류이다.

(9) <지시어의 일관성> 오류의 예

(상)	ST	" <u>강경화</u> 외교의 첫 쾌거" "방역 선진국 상투메를 미처 몰라봤다" "감비아와 상투메가 있는 한 두려울 게 없다"는 식이다.
	HT	「 <u>康京和</u> 外交初の快挙」 「防疫先進国サントメを今まで認識していなかった」 「ガンビアとサントメがあるかぎり、怖いものはない」といった具合だ。
	MT	" <u>強硬</u> 外交初の快挙" "防疫先進国サントメを見誤った" "ガンビアとサントメがある限り、恐れることはない"といった具合だ。

(9)의 MT의 번역에서는 고유명사인 ‘강경화’에 대한 오역을 볼 수 있다. 이 경우 일차적으로는 인명과 일반명사를 혼동한 ‘동음이의어’의 문제이지만 텍스트 전체로 보면 ‘지시어의 일관성’에 대한 오류를 보여주고 있다. 텍스트 전체를 통하여 ‘강경화, 강장관’에 대한 지시가 수차례 등장함에도 불구하고 MT에서는 동일 인물을 ‘強硬, 姜長官, 康長官’과 같이 매번 다르게 번역이 되고 있었다.

다음으로 ② ‘직역 표현의 오류’는 ST를 그대로 직역한 것이 오히려 오류가 되는 경우로서 ‘자연스러움’에 위배되기 때문에 ‘가독성’의 지표로 설정하였다. MT의 특징 중 하나가 ‘직역 표현’이 많다는 것인데 과도한 직역은 다음과 같이 심각한 오류가 될 수 있다.

(10) <직역표현> 오류의 예

(의)	ST	지금 대한민국이 나름 정상적으로 굴러가는 건 <u>‘샤’ 자 달린 두 직업</u> , 그러니까 <u>의 ‘샤’ 와 택배기 ‘샤’</u> 덕분이 아닐까 싶다.
	HT	今の大韓民国がそれなりに正常に動いているのは <u>二つの職業</u> 、つまり <u>医師と宅配便配達員</u> のおかげではないかと思う。
	MT	現在、大韓民国が正常に進んでいるのは" <u>士</u> "の字がついた <u>二つの職業</u> 、つまり" <u>士</u> "と <u>宅配機士</u> のおかげではないだろうか。

(10) 의 ST에서는 ‘사’자로 끝나는 2 개의 직업, ‘의사’와 ‘택배기사’가 대구를 이루고 있지만 도착어인 일본어에서 ‘医師’와 ‘宅配配達員’는 끝말이 다르기 때문에 같은 대구법을 사용할 수 없다. 이 때문에 HT에서는 의역을 선택했지만 MT에서는 원무 그대로 ‘직역 표현’으로 번역함으로써 형식과 의미가 모두 어긋난 심각한 오류를 초래하고 있다.

마지막으로 가점을 매길 수 있는 ⑤ ‘독자를 배려한 번역 전략’의 대표적인 예로는 먼저 자연스러운 의역 표현을 들 수 있다. 본고에서 파과고를 이용하여 번역한 MT 중에는 HT 보다 부자연스러운 번역도 많았지만 HT 와 완전히 일치하거나 다음과 같이 HT 보다 잘된 번역이라도 생각되는 용례 또한 적지 않았다.

(11) <독자를 배려한 번역전략> 가점의 예

(상)	ST	여행이야 나중에라도 갈 수 있지만 <u>당장 발등에 불이 떨어진</u> 비즈니스맨과 유학생은 <u>피가 마르는 심정</u> 일 것이다.
	HT	旅行なら先延ばしにもできるが、 <u>急を要する</u> ビジネスマンや留学生は <u>非常に</u> <u>つらい</u> 思いをしているだろう。
	MT	旅行は後でも行けるが、 <u>足元に火がついた</u> ビジネスマンと留学生は、 <u>血のにじむ</u> <u>思い</u> だろう。

용례 (11)의 ST ‘당장 발등에 불이 떨어진, 피가 마르는 심정’과 같은 관용적인 표현에 대하여 HT 는 평범한 표현을 사용하고 있는 반면에 MT 는 ‘足元に火がついた, 血のにじむ思い’와 같이 최대한 뉘앙스를 살리는 표현을 사용함으로써 출발텍스트가 가진 긴박감이나 절박함을 훌륭하게 표현해내고 있다. 이러한 좋은 번역에는 +1 점의 가산점을 부여하는 것은 당연하다는 생각이 든다.

이와 같이 ‘지시어의 일관성 오류’를 포함하여 ‘가독성’과 관련된 지표들은 ‘문장 부호의 통일성’ 및 ‘오입력’등과 같이 번역의 기술적인 부분도 포함되어 있으므로 번역의 품질 평가에 있어서는 문 이하의 단위로 ‘정확성’을 측정하는 항목들과 별도로 추가적으로 설계가 되어야 한다고 생각된다. 따라서 실제 평가표에서는 (1) ‘가독성’의 항목을 임의로 구성하여 해당 항목이 관찰되는 경우에만 감점이 가능하도록 하였으며 (2) 가점 항목인 ‘독자를 배려한 번역 전략’을 별도로 배치함으로써 각 점수의 비교 평가가 용이하도록 구성하였다.

4.3 새로운 한일 번역의 품질평가 (KJ TQA) 평가지와 채점 결과

이상과 같은 평가 지표 및 평가 항목을 토대로 본고에서 제안하는 한일 번역의 품질

평가 (Korean Japanese Translation Quality Assessment, 이하 'KJTQA') 모델은 다음과 같다.

<표 6> KJ TQA 평가표

KJTQA 모델 평가 용지							
							평가자
평가 대상							
번역물 정보	출발어-도착어	한국어-일본어	텍스트종류				
	글자수						
평가 기준	평가 내용	에러 코드	오류의 심각도 판정				가독성
			심각 -2	중도 -1	경도 -0.5	없음 0	좋은 번역전략 가점 +1
완결성	인식오류	E1					
	미번역	E2					
	누락 오류	E3					
	부가 오류	E4					
	기타 완결성 오류	E5					
정확성	어휘의 오류	동음이의어	E6				
		용어 및 고유명사	E7				
		지역관습	E8				
		관용 표현	E9				
		기타 어휘 오류	E10				
	문법의 오류	조사 사용	E11				
		술어 형식 (태, 상, 시제 등)	E12				
		구문 구성	E13				
		기타 문법 오류	E14				
기타 정확성 관련 오류	E15						
가독성 (임의적)	지시어의 일관성 오류	E16					
	직역 표현 오류	E17					
	문체의 통일성 오류	E18					
	문장 부호의 통일성 오류	E19					
	입력 오류	E20					
	기타 가독성 관련 오류	E21					
(합계)							

위의 품질평가표를 토대로 실효성과 객관성을 테스트하기 위하여 샘플 평가 실험을 진행하였다. 평가 대상은 본고에서 수집한 신문의 논설문에 대한 MT이며 평가자 A와 B는 모두 다년간 번역 감수를 해온 전문가이다. 평가자 A와 B에게 신문의 논설문에 대한 MT를 제공하고 위의 품질 평가 모델을 설명한 후 각각 평가를 실행하였다.

실험 결과, 각각의 평가 점수는 A는 -35.5 점, B는 -28.5 점으로 점수의 차이가 있었는데 이것은 주로 ‘오류의 심각도’에서 기인한 것으로 평가자 A 쪽이 ‘심각(-2)’으로 판정한 경우가 많았기 때문이다.

그럼에도 불구하고 ‘오류의 심각도’를 삭제하고 일괄적으로 모든 오류에 -1 점의 감점을 하는 방식을 취하는 것은 바람직하지 않다고 생각된다. 실제 평가를 해 보면 오류에는 반드시 중대도의 차이가 있어서 심각한 오류와 그렇지 않은 오류를 동일하게 취급하는 것은 오히려 객관성을 저해할 가능성이 있기 때문이다. 따라서 현상황에서는 최대한 ‘오류의 심각도’에 대한 기준을 구체적으로 명시하고 향후 다양한 방법과 실험을 통하여 한일 번역의 세부 평가 항목을 수정 보완해 나갈 필요가 있다.

5. 맺음말

이상으로 본고에서는 한일 번역의 평준화된 품질 평가를 목적으로 기존의 평가 기준에 대한 면밀한 검토를 토대로 새로운 품질 평가 모델을 설계하였다. 본고의 품질 평가 모델의 특징은 다음과 같다.

1) 거시적인 관점과 미시적인 관점의 절충

선행연구의 한일 번역의 품질 평가는 주로 언어학적인 관점을 기준으로 이루어져왔으나 본고에서는 매개활동이라는 번역의 커뮤니케이션 특성을 고려하여 ‘완결성, 정확성, 가독성’이라는 텍스트의 유형과 목적을 상위 카테고리로 설정함으로써 미시적인 측면의 한계점을 극복하고자 하였다.

2) 감점제와 가점제의 융합

본고에서도 기본적으로는 오류의 개수를 세어 가는 정량적 평가 방식을 취함과 동시에 독자를 배려한 훌륭한 번역 전략의 경우에는 +1점의 가점을 줄 수 있도록 설계하였다. 이것 역시 번역의 품질 평가에 있어서 중요한 요소인 ‘가독성’을 도입한 결과이다.

3) 오류 측정의 다원화

어떤 번역물이라도 실제 평가를 해보면 오류에도 정도가 있다는 사실을 알 수 있다.

본고에서는 오류의 개수에 대한 합산과 더불어 오류의 심각도 정도에 따라 ‘심각(Critical), 중도(Major), 경도(Minor), 없음(None)’의 4단계로 구분하고 각각 ‘-2점, -1점, -0.5점, 0점’의 기본값을 설정하여 평준화된 평가 결과가 나올 수 있도록 설계하였다.

이상의 고찰을 통하여 번역의 품질 평가는 단순한 오류의 감점 채점이 아니라 궁극적으로 번역이라는 커뮤니케이션의 특징 및 번역가의 능력 및 노력, 그리고 텍스트와 언어의 구조에 관한 언어 지식 등과 관련해서 포괄적인 관점에서 다각적으로 논의되어야 한다는 사실을 새삼 확인할 수 있었다. 본고에서 제시한 새로운 품질평가 기준은 향후 다양한 실험과 방법을 통하여 보다 평준화된 평가 시스템으로 기능할 수 있도록 지속적으로 수정 보완해갈 계획이다. 본고에서 설계된 품질 평가 모델이 기계번역의 오류 분석을 포함하여 번역 교육의 현장 등에서도 객관적이며 일관적인 평가 기준으로서 일조하기를 기대해본다.

용례출전

- (마) 조선일보<만물상> ‘마스크 기축통화론?’ (2020.3.4): 「マスク基軸通貨論」
 (메) 조선일보<만물상> ‘메르스 때와 너무 다른 서울시장’(2020.2.27): 「MERSのとときとあまりに異なるソウル市長」
 (을) 중앙일보<분수대> ‘올림픽의 주인공’(2020.03.20): 「オリンピックの主人公」
 (의) 중앙일보<시선> ‘의사와 택배기사가 한국을 살렸다’ (2020.0.20): 「医師と宅急便配達員が韓国を救った」
 (상) 조선일보<사설> ‘상투메프린시페’(2020.03.11): 「サントメ・プリンシペ」
 (텅) 조선일보<만물상> ‘텅 빈 인천공항에 선 줄’(2020.3.9): 「ガラガラの仁川空港にできた行列」 2020.03.09

참고문헌

- 김혜림(2020) 「기계번역과 인간번역의 문체 비교-신문 사설 한중번역을 중심으로-」 『통번역학연구』 제 24 권, 한국외국어대학교 통번역연구소 pp.21-47
 박민영(2020) 「신문 논설문에 대한 한일 기계번역의 오류 분석-쓰기 평가의 관점에서-」 『韓國語教育研究』 第 10 号, pp.152-171, 日本韓國語教育学会
 안인경(2015) 「번역 가독성 개념에 대하여」 『언어과학연구』 제 74 집, pp.147-168, 한국언어과학회,
 이주리에(2018) 「인공신경망 기계 번역의 한일/일한 번역 품질에 대한 예비연구-품질향상 검토와 공역코드 적용」 『통역과 번역』 (20)1, pp.21-47, 한국통역번역학회
 이주리에(2020) 「피드백 항목에 따른 일한 번역 피드백 고찰-교수자와 졸업생 평가를 중심으로-」 『日本語教育』 93, pp.75-92, 한국일본어교육학회
 장혜선(2011) 「목적별 번역 평가 모델에 관한 소고-국내의 번역품질평가 분석을 중심으로-」 『외국어교육연구』 25 권(1 호), pp.107-136, 외국어교육연구소

- 장혜선(2012) 「진단적 번역품질평가 모델 구축을 위한 오류항목 분류에 관한 연구」 『일본어문학』 59, pp.143-178, 일본어문학회
- 장혜선(2014) 「번역평가를 위한 번역오류 기술 및 분류에 관한 소(小)연구 -통번역대학원 수강생들의일한 (JK) 번역물 오류 분석을 중심으로」 『日本語文学』 62(1), pp.99-132, 한국일본어문학학회
- 市川保子(2005) 「誤用研究と日本語教育」 『開かれた日本語教育の扉4』 スリーエーネットワーク pp.109-121
- 金澤裕之(2006) 「誤用分析研究の可能性」 『横浜国大言語研究』 24号 横浜国立大学国語・日本語教育学会 p.1-18,
- 豊島知恵, 藤田篤, 田辺希久子, 影浦峯, Hartley, A. (2016). 「校閲カテゴリ体系に基づく翻訳学習者の誤り傾向の分析」 『通訳研究への招待』 16, 47-65.

(한국외국어대학교 일본어통번역학과)

mypark@hufs.ac.kr

韓国語教育研究（第11号）

2021年9月15日 発行

発行者 文 慶喆

発行所 日本韓国語教育学会

〒577-8052 大阪府東大阪市小若江3-4-1

近畿大学 国際学部 酒匂康裕 研究室気付

編集者 『韓国語教育研究』編集委員会

金世徳、柳朱燕

印刷所 株式会社 仙台共同印刷
