

한국어 무주어 구문의 영어 번역 양상: 인간번역, 구글번역, 챗GPT 간의 차이를 중심으로

임진

(이화여자대학교)

Yim, Jin. (2024). Strategies for translating zero-subject Korean sentences into English: A focus on the differences between Human, NMT, and LLM Translations. *The Linguistic Association of Korea Journal*, 32(3), 1-22. This paper investigates the translation strategies employed in human and machine translations of Korean zero-subject sentences into English. The author translated 343 zero-subject segments from management forewords in business reports using Google Translate (NMT) and GPT-3.5 (LLM) and compared the results with quality human translation, seeking to investigate the patterns of three corpora's translation strategies—subject restoration or structural modification. It was found that all three corpora-human translation (HT), NMT, and LLM translation-the dropped subject was most commonly replaced by personal pronouns rather than other nouns. Two statistically significant differences emerged among the corpora. First, HT exhibited a higher frequency of proper or general noun subjects, likely reflecting translators' efforts to avoid repetitive use of the first-person plural pronoun "we" in adjacent sentences. In contrast, NMT and LLM translations frequently adopted "we," leveraging it as a safe choice to enhance reader engagement in this genre. Second, NMT showed an overuse of short passive constructions without an agent, a choice underrepresented in LLM translations. While short passives can be effective when the subject is omitted in the source text, they may weaken the connection between action and agent, thereby altering the original discourse effect. This study contributes to the MT literature by expanding the scope to include genre-specific features, LLM translation tendencies, and particular translation challenges.

주제어(Key Words): 기계번역(machine translation), 신경망기계번역(neural machine translation), 대규모언어모델(large language model), 무주어(zero subject), 한영번역(Korean-English translation)

1. 들어가는 말

한국어는 동사 중심 언어로, 문장 구조가 자유롭고 문장의 주요 성분인 주어, 목적어 등이 빈번히 생략되는 언어이다(박창희, 2012, p. 172). 주요 성분이 생략되더라도 모국어 사용자의 경우 종결어미 또한 맥락을 통하여 그 의미를 유추할 수 있다. 그러나 주요 성분이 생략된 한국어 문장을 주요 성분의 생략을 허용하지 않는 언어로 번역해야 하는 경우, 번역자는 언어 간 차이를 극복하기 위해 원문에 일정 부분 개입해야 한다. 우선 생략된 문장 성분의 내용을 유추하여 원문에 없는 내용을 추가할 수 있을 것이다. 원문에 없는 내용을 추가하는 위험을 감수하지 않으려면, 문장의 구조를 수동태 등 다른 형태로 전환해야 할 것이다.

번역자의 개입 방식과 이에 따른 번역 선택은 해당 문장이 포함된 텍스트의 장르와 맥락, 번역가의 선호도 등 다양한 변수에 따라 달라질 수 있다. 이 때 이루어지는 번역 선택은 번역의 정확성뿐만 아니라 텍스트의 기능, 수용 측면에서도 영향을 미칠 수 있기에 매우 중요하다. 그러나 실증적 차원에서 주어가 생략된 문장을 번역하는 양상에 주목한 연구는 김한식 외(2019), 김경희(2020) 등 극히 소수에 불과하다. 이 문제는 최근 일상화되고 있는 기계번역에서도 매우 중요하지만, 인간 번역과 기계번역 간에 주어가 없는 문장을 다루는 방식에 어떠한 차이가 있는지에 대해서도 알려진 바가 거의 없다.

한국어에서 영어 방향으로 이루어진 기계번역의 결과물에 대해 축적된 연구는 대부분 신경망기계번역(Neural Machine Translation, NMT) 번역 서비스의 전반적인 품질 평가에 초점을 맞추고 있기에, 특정 번역 문제나 전략, 원문의 언어적 특성에 초점을 맞춘 고찰 또한 부족한 실정이다(이지은 & 최효은, 2023, p. 78; Yim, 2024, p. 70). 최근 챗GPT를 필두로 한 대규모언어모델(Large Language Model, LLM)이 대중화되고 기계번역 품질이 언어쌍, 번역 방향, 텍스트 장르에 따라 상이한 점을 고려할 때 보다 다양한 텍스트와 언어쌍, 언어적 특성에 주목하여 NMT와 LLM 번역의 차이점을 고찰할 필요성이 대두되는 시점이다.

이러한 연구 공백에 착안하여 본고는 무주어 형태의 한국어 문장을 주어 생략이 허용되지 않는 영어 방향으로 번역할 때 인간번역, NMT, 대규모언어모델(Large Language Model, LLM) 간의 번역 양상에 차이가 존재하는지 알아보고, 번역 양상이 해당 텍스트 장르 특성에 부합하는지 논의한다. 이를 통해 기계번역에 대한 논의를 일반적인 품질 평가에서 특정 번역 문제로 세분화하고 기계번역 연구에서 텍스트의 장르 특성을 고려해야 할 필요성을 제시하는 것이 본 연구의 궁극적 목적이다.

본 연구에서 제기하는 연구 질문은 첫째 인간 번역, NMT, LLM 간 한국어 무주어 구문의 주어 복원 여부에는 차이가 있는가, 둘째, 세 코퍼스 간 한국어 무주어 구문 번역 전략 간 차이가 있는가이다. 연구 질문에 답하기 위해 본고는 우선 무주어 한국어 문장을 영어로 번역할 때 적용할 수 있는 번역 선택과 그 의미에 대해 기존 연구에서 알려진 점을 개괄하고(2장), 분석대상 코퍼스 수집 및 분석 방법(3장) 및 분석 결과(4장)를 기술한다. 결

론(5장)에서는 연구 결과를 분석 텍스트가 속한 장르의 언어적 특성과 연계하여 분석 결과가 갖는 의미와 시사점을 도출한다.

2. 이론적 배경

본 장에서는 무주어 한국어 문장의 영어 번역과 관련된 문제를 기술하고 분석 틀을 확립하기 위한 이론적 배경과 선행연구의 결과를 개괄한다. 우선 한국어와 영어 간 나타나는 주어 생략 현상의 차이에 대해 약술하고(2.1절) 한국어 무주어 문장의 번역 선택에 대해 살펴본 다음(2.2절) 기계번역의 무주어 문장 번역에 대한 선행연구를 살펴본다(2.3절). 마지막으로 본고의 분석 대상 텍스트에서 주어 복원 번역 전략이 갖는 의미를 살펴본다(2.4절).

2.1. 한국어와 영어의 주어 생략 양상의 차이

한국어 문법의 가장 큰 특징 두 가지로 고영근과 구본관(2018, p. 15)은 “형태론적으로는 교착어”이며 “통사론적으로는 주어-목적어-동사 유형”이라는 점을 꼽는다. 본고에서 다루는 주어란 “주요 문장 성분의 하나로, 술어가 나타내는 동작이나 상태의 주체가 되는 말”(국립국어원, n.d.)을 의미한다. 주어의 기본 형태는 체언에 속하는 명사, 대명사, 수사에 주격조사가 결합되어 구성되며(고영근 & 구본관, 2018, p. 264), 문장의 맨 앞에 위치한다(고영근 & 구본관, 2018, p. 268).

교착어의 특성 상 한국어는 어순이 유연하고, 이에 따라 문장의 주요 성분인 주어나 목적어가 생략되는 현상이 일반적이다(문귀선, 2010; 박청희, 2012, p. 174). 의미 추측이 가능하다면 대명사의 생략이 허용된다는 점에서 대명사탈락언어(pro-drop language)라고 통칭한다(김한식 외, 2019, p. 34). 주어 생략은 보편문법 측면에서도 매우 중요한 요소이다.

그러나 한국어에서도 주어가 항상 생략되는 것은 아니다. 주어를 생략하기 위하여 가장 중요한 조건은 “확인성(identifiability, unambiguity, 또는 recoverability)”(정연창, 2007, p. 102)이다. 즉, 문맥이나 상황을 통해서 생략된 주어의 의미를 확인할 수 있어야 한다는 것이다. (1)의 대화문에서 화자와 청자 모두 지칭하는 대상이 주격 인칭대명사인 “당신”(1a)과 “나”(1b)인 것을 알 수 있으므로, 주격 인칭대명사가 생략되었다.

(1) a. (Ø) 지금 어데 가세요?

b. (Ø) 학교에 갑니다. (오준규, 1971, p. 195; 정연창, 2007, p. 103에서 재인용)

특히 한국어에서는 한 번 언급되어 상대방이 이미 알고 있는 정보가 생략되는 특성이

있다(진실로 & 곽은주, 2013, p. 173). (2)와 같이 선행 문장의 주격 명사 “한국”이 후행 문장에서도 주격 명사가 되는 경우, 독자는 후행 문장의 주격 명사 역시 “한국”임을 유추할 수 있으므로 주어가 생략되었다.

- (2) 한국은 그 한복판에 있다. 정세 변화에 따라선 (Ø) 한·미·일 3각 동맹과 중국 사이에서 [국가적 선택을 재촉받는 순간을 맞을] 수도 있다. (조선일보, 2013.11.22. 사설, 김수정, 2016, p. 46에서 재인용)

상기와 같은 무주어 형태의 문장이 한국어에서는 빈번하게 사용되는 것으로 알려져 있다. 한국어 소설에서의 인칭대명사 주어 생략을 연구한 김한식 외(2019)에 따르면 분석 대상 문장 542개 중 40% 가량인 221개 문장에서 인칭대명사 주어가 생략되었다.

반면 굴절어인 영어는 격이 어순에 따라서 정해지기 때문에 한국어에 비해서 문장 구조가 경직적이며, 명령문 등 동사로 시작하는 문장의 예외를 제외하고는 주어의 생략이 거의 이루어지지 않는다(박청희, 2012, pp. 174-175). 영어는 대명사의 탈락이 허용되지 않는 언어(non-pro-drop language)이므로, 주격으로 사용된 인칭대명사의 경우 생략할 수 없다(김한식 외, 2019, pp. 174-75). 이러한 차이는 (3)에서 두드러진다. 이 예에서는 선행문장의 주어가 후행문장과 일치하지만 생략되지 않고 인칭대명사 “he”가 사용되었다.

- (3) Hercule Poirot sat on the white sand and looked out across the sparkling blue water. He was carefully dressed in a dandified fashion in white flannels... (Christie, 1936, p. 196, Baker, 1992, p. 182에서 재인용)

위와 같은 생략 가능성은 한국어와 영어의 중요한 차이점 중 하나이다(Kiaer, 2018, p. 1). 양적인 측면에서 문장 주요 성분인 주어와 목적어의 생략을 연구한 박청희(2012, p. 180)에 따르면 한국어와 영어 시나리오에서 주어가 생략되는 비율은 한국어는 65%, 영어는 35%인 것으로 나타났다. 양 언어의 이러한 차이 탓에 주어가 생략된 한국어 문장을 타 영어로 번역하는 번역자는 생략된 주어를 유추하여 복원하거나, 복원이 필요하지 않은 구조로 원문을 변형해야 하는 선택에 직면한다. 구체적인 번역 양상에 대해서는 다음 절에서 다룬다.

2.2. 한국어 무주어 구문의 번역 양상

번역학에서 출발어 원문에 가해지는 변화는 변이(shifts)라고 볼 수 있다(Bakker 외, 2009, p. 269). 변이의 분류 방식과 범주는 학자별로 다양하지만, 본고는 주어 생략 문장의 번역이라는 특정 현상만을 목적으로 하기에, 이와 연관된 분류만을 소개한다.

번역 변이가 나타나는 원인을 중심으로 변이를 범주화한 Catford(1965, p. 73)는 번역 과정에서 나타나는 원문의 변화 중 출발어와 도착어의 언어적 차이로 인해 번역 과정에서 이루어지는 필수적인 것을 의무적 변이(obligatory shifts), 필수적이지는 않지만 여러 가지 이유로 인해 나타나는 선택적 변이(optional shifts)로 구분한 바 있다. 한글에서 생략된 주어를 영어에서 복원하는 것은 두 문법적 차이로 인해 나타나는 필수적인 변화로, 의무적 변이에 해당한다고 볼 수 있다.

본고에서 주목하는 의무적 변이 양상을 고찰하는 데 유용한 분류는 원문과 번역 결과물 간의 차이라는 관점을 기준으로 한 Nida(1964)의 범주이다. Nida는 번역문과 원문 사이의 등가를 달성하기 위해 원문을 조정(adjustment)하는 기법(techniques)으로 세 가지를 제시한 바 있다. 원문에 없는 내용을 추가하는 추가(addition), 원문의 정보를 번역에서 삭제하는 차감(subtraction), 원문의 품사, 어순, 구조, 의미 등의 변화를 가져오는 변경(alteration)이 여기에 해당된다(Nida, 1964, pp. 226-239).

2.2.1. 생략된 주어를 복원하는 추가 전략

한국어에서 생략된 주어를 영어에서 복원하는 방식은 원문에 명시되어 있지 않은 영어 문법 범주의 필수 요소를 첨가하는 과정이므로 추가(Nida, 1964) 범주에 해당된다.

이렇게 원문에 없는 정보를 번역문에서 추가하는 현상은 번역문의 특성인 명시화(explicitation, Klaudy, 2009)로도 설명할 수 있다. 명시화에는 세 가지 종류가 있다. 첫째, 의무적(obligatory) 명시화, 둘째, 선택적(optional) 명시화, 셋째, 화용적(pragmatic) 명시화이다. 의무적 명시화란 구문적, 의미적 명시화를 뜻하는데 언어의 구조적 차이로 인하여 도착어에서 문법적으로 문제가 없는 문장을 구사하기 위하여 원문에 없는 정보가 반드시 추가되어야 하는 경우를 뜻한다. 선택적 명시화는 번역 전략의 차이나 문체 상의 선호도에 따라 원문에 없던 정보가 추가되는 것을 뜻한다(Klaudy, 2009, p. 106). 화용적 명시화는(Klaudy, 2009, pp. 106-107) 문화적 차이로 인해 도착어 독자에게 출발어 문화에서 일반적인 지식으로 공유되는 정보를 번역에 명시적으로 추가하는 것이다. 주어탈락 언어를 무주어탈락 언어로 번역하는 경우 일어나는 주어의 복원은 의무적 명시화라고 볼 수 있다.

한국어 무주어 문장을 영어로 번역할 때에는 대부분의 문장에서 생략된 주어가 복원되는 것으로 알려져 있다. 인칭대명사가 주어에서 생략된 한국어 문장 221개의 주어 복원 양상을 조사한 김한식 외(2019, p. 41)에 따르면 76%인 168문장에서 주어가 복원되었다.

원문에서 생략된 주어를 번역문에서 복원하는 방식은 (4), (5)와 같다. 번역자는 원문에서 생략된 주어가 어떠한 것인지 선행 문장이나 전체 문맥을 통해 파악한 다음 번역문에서 이를 복원하였다. (4)의 경우 생략된 주격 인칭대명사 'you'가 추가되었다. (5)는 기사문인데, 해당 문장에서 생략된 주어는 기사문의 선행 단락에 언급된 특정 기업이기에 번역 과정에서 이를 'Korean Air'라는 고유명사로 복원하였다.

(4) ST: (Ø) 체크인을 한다. (Ø) 호텔 레스토랑에서 저녁을 먹는다.

TT: You check in. You eat in the hotel restaurant. (김한식 외, 2019, p. 42)

(5) ST: 이에 따라 오는 15일부터 (Ø) 한국에서 출발하는 장거리 노선에서는 콘덕 또는 피자, 해외 출발편에서는 핫포켓(파이 겹질 속에 다양한 속을 채운 음식) 등을 제공할 계획이다. (서유근, 2024)

TT: Accordingly, Korean Air plans to offer corndogs or pizza on long-haul routes departing from Korea, and Hot Pockets on outbound flights from overseas. (Suh and Lee, 2024)

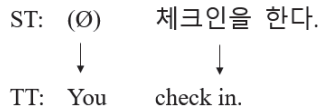


그림 1. (4) 무주어 문장 주어를 복원하는 번역

상기 두 예문은 원문의 문장 형태가 번역문에서도 그대로 유지된다. 즉, 원문에서 생략된 주어가 번역문에서도 주어로 복원되었다. 그러나 원문에서 생략된 주어가 목적어 등으로 복원되는 경우 등 다양하게 원문의 형태와 구조를 변경하여 생략된 주어를 다루는 방식도 존재할 것이다. 그러나 본고에서는 한영 번역의 경우 주어 생략 문장의 대부분이 주어로 복원된다는 점에서 이러한 경우만을 주어 복원으로 간주한 선행연구(김한식 외, 2019)의 구분을 따르기로 한다. 정리하면 원문의 문장구조를 그대로 유지한 채 생략된 주어를 번역문에서 주어로 복원하는 경우를 추가 전략에 포함하고, 이보다 적극적으로 원문을 변경하는 것을 변경 전략으로 구분한다. 다만, 추가 전략의 경우 2.4절에서 논의할 본고의 데이터가 갖는 장르적 특성을 고려하여, 주어를 인칭대명사로 번역하는 경우와 기타 명사로 번역하는 경우를 구분하여 분석한다.

2.2.2. 변경 전략

원문의 주어가 생략되어 있는 경우 이를 문맥으로 유추하여 번역문에 추가해야 하는 어려움이 따르므로 종종 번역 오류로 이어진다(김경희, 2020, pp. 111-113). 추가 전략 없이 원문의 정보만을 활용하여 무주어 문장을 주어가 필수 문장 성분인 언어로 번역하기 위해 광범위하게 활용되는 것은 태(voice)의 전환이다. 변길자(2024, p. 28)에 따르면 주어 생략이 가능한 한국어에서는 행위자 정보를 명시하지 않아도 능동문으로 표현이 가능하기에 수동 표현에 대한 수요가 낮은 반면 영어는 수동문 수요가 높다. 영어에서 수동태는 학술, 뉴스 장르에서 그 빈도가 높으며 대화에서 빈도가 가장 낮은 것으로 알려져 있다(Biber 외, 1999,

p. 476). 수동태에는 ‘by’와 함께 행위자를 명시하는 장형 수동문과 행위자를 명시하지 않는 단형 수동문이 있다(Biber 외, 1999, p. 477). 주어가 명시되지 않은 한국어 문장을 영어로 번역할 때 단형 수동문을 활용하면 생략된 주어를 추정하여 추가하지 않고 원문의 정보만으로 번역할 수 있다. (6)의 원문에서 생략된 주어는 “모든 사람”이지만, “we” 등의 인칭대명사로 주어를 복원하는 대신 수동태가 사용되었다.

(6) ST: (Ø) 권리는 남용하지 못한다.

TT: No abuse of rights shall be permitted. (이지은 & 최효은, 2020, p. 267)

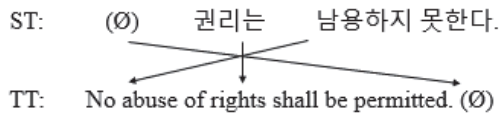


그림 2. (6) 무주어 문장 주어를 복원하지 않고 변경하는 번역

태(voice)의 변화를 활용하면 추가 전략 없이도 무주어 한국어 문장을 영어로 번역할 수 있지만, 태의 변화는 원문이 지닌 화용적 효과의 변화를 초래할 가능성도 존재한다. 수동문에서는 행위자에 대한 정보가 생략되거나 어순 상 문장의 끝에 위치하게 되므로 능동문에서보다 행위자가 덜 부각된다(Biber 외, 1999, p. 476). 또한 학술 장르에서 수동태가 쓰이게 되면 문장에서 묘사되는 행위가 행위자가 분리되기 때문에, 해당 행위가 보다 객관성을 띠게 된다(Biber 외, 1999, p. 477). 반면 뉴스의 경우 선행 문장에서 이미 행위자가 언급되어 유추하기 쉬운 경우 단형 수동문을 통해 행위자를 반복해서 언급하지 않으면서도 효과적으로 행위를 강조할 수 있다(Biber 외, 1999, p. 477). 따라서 번역 과정에서 태를 변경할 때에는 텍스트의 장르 특성을 고려할 필요가 있다.

(7) ST: 밤에도 밝으니 혼자 있어도 (Ø) 덜 외로운 것처럼 느껴졌다.

TT: And because it was light at night, it made me feel less lonely. (김한식 외, 2019, p. 41)

주어가 없는 한국어 원문을 변경하여 영어로 번역하는 전략에는 (7)과 같이 주어를 목적어 등 다른 문장성분으로 변경하거나, 주어 복원 없이 사물주어를 활용하여 번역하는 등의 변경 전략도 존재한다. 그러나 선행연구(김한식 외, 2019)를 참고할 때 세부 항목의 빈도가 매우 제한적이어서 통계 검정이 불가능할 것으로 사료되며, 선행연구와의 비교 가능성을

높이기 위하여 본고에서도 이를 모두 ‘기타’ 범주로 구분한다. 정리하면, 변경 전략은 ‘수동태’ 및 ‘기타’ 두 가지로 구분한다.

2.3. 무주어 한국어 구문의 기계번역

상기에 논의된 바와 같이, 무주어 문장을 주어 탈락을 허용하지 않는 언어로 번역할 때 수반되는 번역 선택은 번역의 정확성뿐만 아니라 화용적 효과 측면에서 영향을 미칠 수 있으며(강동희, 2023; 임소정 & 이아형, 2024), 한국어가 모국어인 학습자들에게도 어려움을 초래하는 것으로 알려져 있다(왕청동, 2024; 전설주, 2024). 따라서 초기 신경망기계번역 시스템 개발 측면에서 관심의 대상이 되어 왔으며, 언어학적 관점에서는 기계번역이나 포스트에디팅 결과의 품질평가를 다룬 연구(김자경, 2021; 박옥수, 2017, 2018; 한현희, 2020)에서 오류 사례로 언급되어 왔다.

한국어 무주어 구문은 한국어에서 영어 방향으로 이루어지는 기계번역에서 문제를 야기하는 것으로 보고된다(박옥수, 2017, 2018). 보고된 오류는 대부분 (8)과 같이 생략된 주어를 예측하지 못하여 발생한 오류였다. 원문에서 생략된 주어는 문맥상 특정 증권사인데, 번역문에서는 ‘investors’를 추가하여 원문과 의미가 달라졌고, 포스트에디팅에서도 이러한 오류가 수정되지 않았다.

(8) ST: 증권사 신탁계좌에서 투자자가 주식 매매 지시를 과도하게 할 경우 (Ø) 수수료
를 부과할 수 있도록 관련 규정이 바뀐다.

TT: Related regulations will be changed to allow investors to impose fees if they
order excessive stock trading in trust accounts of securities
firms. (김자경, 2021, p. 50)

무주어 한국어 구문의 영어 번역 문제를 보고한 기존 연구들은 NMT 결과물의 품질평가를 목적으로 하고 있기에 그 초점을 오류 발견에 두고 있으며, 인간번역과 기계번역의 주어복원 양상에 대해서는 다루고 있지 않다. 무주어 구문 기계번역 연구 동향을 조사한 Wang 외(2023)는 네 가지 결론 및 제언을 제시하였다. 첫째, 그간 많은 연구가 진척되어 왔음에도 불구하고 여전히 주어가 생략된 형태의 문장의 번역은 기계번역에 문제를 야기하며 기존에 제시된 보편적인 자동 및 수동 품질평가 방식이 무주어 구문 번역의 품질을 평가하는 데 유효하지 않기에 인간이 개입하는 수동 평가를 통한 번역품질 평가 방식이 개선될 필요가 있다. 둘째, 장르와 담화 특성을 고려할 필요성을 제시하였다. 주어가 생략되고 복원되는 방식은 장르별로 상이하며, 이를 어떻게 복원하는지에 따라 담화 특성에도 영향을 미치기 때문이다. 셋째, 무주어 구문의 번역에 대한 실증적 연구는 주로 영어, 중국어에 집중

되어 있다. 넷째, 품질평가의 일부로 다루어지던 무주어 구문 번역은 최근 점차 독립된 연구 주제로 자리잡고 있다.

그럼에도 불구하고 한국어 무주어 구문은 아직까지 기계번역에서 주요 연구 주제로 자리잡지 못하고 있는 것으로 보인다. 주어가 생략된 문장이 일반적으로 한국어에서 통용되며 한국어와 영어 간 번역도 그 어떤 언어보다 빈번히 일어나고 있음에도 불구하고, 한국어를 출발어로 하는 기계번역에서 이 문제에 집중한 연구는 찾아보기 어렵다. 품질평가 항목의 일부로 주어 생략 구문의 번역을 다룬 소수의 연구(박옥수, 2017, 2018) 또한 LLM 출시 이전에 수행된 것이기에 NMT 번역만을 대상으로 하고 있다. 챗GPT의 번역 기능에 대한 관심이 증대되면서 시(이유정, 2023), 소설(박수정 & 최은실, 2023; 신천 & 조혜진, 2023, 조승연, 2023), 과학 기사(최효은, 2024), 비즈니스 텍스트(Yim, 2024) 등 다양한 장르의 번역 결과에 대한 연구가 속속 이루어지고 있는 시점이기 때문에 NMT와 LLM 간의 번역 양상 차이를 고찰하는 시도는 시의적절한 것으로 사료된다.

2.4. 기업보고 장르에서 번역 전략의 중요성

본고의 분석 데이터 원문은 국내 기업보고서에 수록된 한국어 대표이사 서한이다. 이를 분석 데이터로 선정한 원인은 크게 세 가지이다. 첫째, 다수의 인간 번역자가 작업한 양질의 번역 데이터를 구하고자 했다. 기업보고서는 그 중요성 때문에 아직 전문성을 가진 인간 번역자가 작업하는 영역이지만 번역가 정보가 공개되어 있지 않으므로 번역가에 대한 정보를 알 수 없다. 그러나 다수 기업의 보고서를 수집하는 방식을 취하면 특정 번역가가 선호하는 번역 방식으로 분석 결과가 편중될 가능성을 효과적으로 낮출 수 있다고 판단하였다. 둘째, 대표이사 서한은 기업보고 장르이지만 대표이사가 기업의 이해관계자에게 보내는 서한의 형태로 쓰여 있어 주어가 빈번하게 생략되어 있었다. 이에 본 연구의 데이터로 적합하다고 판단하였다. 셋째, 주어 복원 전략에서 주로 쓰이는 인칭대명사의 경우 대표이사 서한 장르의 기능과 수용에 중요한 의미를 갖기에, 인간 번역과 NMT, LLM 번역에서 이를 다루는 양상을 고찰함으로써 향후 해당 장르에 대한 기계번역 품질의 발전과 개선에 간접적으로 기여할 수 있을 것이라 사료된다.

기업보고서의 대표이사 서한은 발신자는 보고서 발행 기업의 대표이사이며, 수신자는 기업의 이해관계자이다. 한국어의 경우 술어가 타동형 행위동사일 때 인간 행위자만이 주어가 될 수 있고, 무생물 행위자는 주어가 될 수 없다(이창수, 2014, p. 124). 따라서 발신자와 수신자가 고정되어 있는 분석대상 텍스트에 속한 대부분의 문장에서 주어는 대표이사 또는 해당 기업 자체를 의미하기 때문에, 생략되더라도 화용적으로 이를 추론하기가 용이한 것으로 추정할 수 있다. 따라서 생략된 주어를 복원하는 전략에 초점을 맞출 수 있을 것으로 판단하였다.

한국어 대표이사 서한의 주어 형태는 크게 세 가지로 추정해볼 수 있다. 우선 1) 인칭대명사(저, 우리), 2) 고유명사 및 일반명사를 포함하는 기타 명사(기업명, 제품 또는 서비스), 3) 생략이다. 본고의 분석 대상은 주어가 생략된 문장이며, 장르 특성 상 대부분이 평서문의 형태를 띠고 있다. 무주어 문장을 번역하는 경우 2.3에서 논의한 바와 같이 원문에 없는 주어를 복원하거나, 문장 형태를 변형해야 한다. 주어를 복원하는 경우 주격 인칭대명사인 “I”나 “we”, 또는 기업명, 제품명이나 “company/firm/service/product” 등의 일반명사를 활용할 수 있다. 또는 2.2.2항에 제시된 변경 전략을 활용할 수 있다.

그런데 어떤 번역 전략을 선택하는지는 의사소통의 효과 측면에서 영향을 미친다. 대표이사 서한 장르에서 인칭대명사의 빈도는 번역된 텍스트와 원래 해당 언어로 쓰인 비번역 텍스트를 구분하는 역할을 한다(김현아, 2012; Huang & Rose, 2018; Junge, 2011; Yim & Lee, 2024). 인칭대명사를 사용하면 독자와의 대인관계적 요소가 강화되어 독자와 저자 간의 거리가 좁혀지기 때문에 영어로 작성된 텍스트에서는 인칭대명사가 활발하게 사용되는데, 외국어로 작성되어 영어로 번역된 텍스트에서는 비번역 텍스트에서보다는 인칭대명사 빈도가 낮아지기 때문이다. 다시 말하면 주어 복원 시 인칭대명사를 사용하는 것이 기타 명사를 사용하는 것보다 서한의 대인관계적 기능을 살릴 수 있는 선택이라고 상정할 수 있다.

또한 변경 전략을 사용하여 생략된 주어를 복원하지 않고 수동태로 번역하는 경우 수동문의 일반적인 특성 상 행위자의 존재가 덜 부각된다. 저조한 실적을 기록한 기업의 기업 보고서에서 대표이사는 행위자인 자신과 실적 간의 거리를 두기 위하여 수동태를 더 많이 사용하는 것으로 알려져 있다(Hyland, 2015; Asay 외, 2018). 따라서 주어를 생략하는 한국어 문장에서 주어를 복원하지 않고 수동태를 사용한다면 자칫 해당 행위와 행위자와의 관계가 약화될 것이다.

표 1. 분석 틀

번역 전략		비고
추가	인칭대명사 추가	주격 인칭대명사 (I, we, you, he, she, it, they)
	기타 명사 추가	고유명사, 일반명사
변경	수동태	장형 수동문/단형 수동문
	기타	물주구문, 주어를 목적으로 복원

이러한 장르적 특성을 감안하여, 본고는 표 1과 같은 분석 틀을 정립하고, HT, NMT, LLM 세 종류의 코퍼스에서 나타나는 번역 전략 양상의 차이를 양적, 질적으로 분석하고자 한다. 추가 전략에서는 무주어 구문의 생략된 주어를 번역 과정에서 인칭대명사 또는 기타 명사 주어로 복원하는 빈도를 조사한다. 변경 전략에서는 생략된 주어를 주어로 복원하지

않고 다양한 방식으로 변경하는 빈도를 조사하되, 기업보고 장르에서 의미 있는 언어적 특성으로 간주되는 수동태로 번역되는 빈도와 기타 변경의 빈도를 조사한다.

3. 연구방법

본 장에서는 본고의 연구목적을 달성하기 위한 데이터 수집 및 분석 방법과 절차를 소개한다. 3.1절에서는 본고의 분석 코퍼스 정보 및 기계번역 생성 절차에 대해 소개한다. 3.2절에서는 수집된 데이터의 양적, 질적 분석 절차를 설명한다.

3.1. 분석 데이터

본고의 분석데이터 표본은 원문이 한국어인 한국 기업의 연차보고서의 대표이사 서한으로, 상장기업을 대표할 수 있게 다양한 규모와 업종의 30개 기업으로 구성된 지수인 KTOP30(한국거래소, n.d.) 기업에서 발간한 보고서를 수집하였다. 2023년 7월 온라인 홈페이지를 통해 수집된 원문은 총 11,865 단어(형태소 분석 전 기준)였으며, 공식 영어 번역본은 총 18,342 단어였다. 소규모 코퍼스이지만 총 KTOP 29개 기업이 발행한 원문 및 번역문으로, 대표성을 띤다고 볼 수 있다.

2.4절에서 기술한 바와 같이 다수의 기업 웹사이트에 공개된 영어 연차보고서를 인간 번역으로 간주하였다. 각사 연차보고서의 한국어 버전에서 원문을, 영어 버전에서 번역본 코퍼스를 수집하였다. 이를 대표적인 NMT와 LLM 서비스인 구글번역과 챗GPT(GPT-3.5)로 번역하였다. 기계번역은 2023년 7월 6일과 7일 양일에 걸쳐 텍스트 단위로 생성되었으며, 번역 시작 전 GPT-3.5에 “Tell me your version information”이라는 프롬프트를 제시하자 “Version: 2021-09 (Knowledge cutoff: September 2021)”이라는 응답을 얻을 수 있었다. 챗GPT 번역은 선행연구(Jiao 외, 2023)를 참고하여 “Translate the text into English.”를 프롬프트로 사용하였다. 주어가 생략된 문장을 다루는 NMT와 LLM의 양상을 관찰하는 본고의 특성 상, 특정 번역 전략을 활용하여 텍스트를 번역하라는 프롬프트를 추가하지 않았다. 챗GPT를 통한 번역은 시간차를 두고 실시할 경우 그 결과가 달라질 가능성이 존재하기에 본고는 상기 프롬프트로 얻은 최초의 번역만을 수집하였다. 또한, NMT와 LLM의 번역 전략을 비교하는 본고의 목적 상 프롬프트를 변경하여 LLM에서 최상의 번역 결과를 얻으려는 시도는 하지 않았다. 이와 같은 과정으로 얻은 인간 번역과 NMT, LLM 번역을 각각 원문 세그먼트에 따라 정렬하여 병렬코퍼스를 구축하였다.

3.2. 분석 절차

본고의 분석은 양적분석과 질적분석으로 두 단계로 진행되었다. 분석을 위해 원문과 인간 번역을 문장 단위로 정렬하여 병렬코퍼스를 구축하였다. 번역 과정에서는 문장이 분할 또는 병합되는 경우가 빈번한데, 원문이나 번역문에 분할된 문장이 포함되지 않고 완전한 문장이 포함되도록 조정이 필요했다. 본 코퍼스의 번역문은 세 종류이므로, 이 중 한 번역이라도 문장이 두 개 이상으로 분할된 경우에는 하나의 세그먼트로 구분하여 병렬코퍼스를 구축할 필요가 있었다. 이렇게 조정을 거쳐 총 700개의 세그먼트로 구성된 병렬코퍼스를 구축할 수 있었다. 이렇게 완성된 원문(ST), 인간번역(HT), NMT 번역, LLM 번역 병렬코퍼스의 한국어 원문 총 716문장 중에서 주어가 생략된 346개의 문장과 이에 대한 번역을 추출하였다. 각 코퍼스의 세부적인 문장 수는 표 2를 참조한다. HT, NMT, LLM 세 코퍼스의 각 문장에서 생략된 주어를 어떻게 복원하였는지 표 1의 분석 틀에 따라 4가지로 구분하고 계수하였다.

표 2. 코퍼스별 문장 수

코퍼스	ST	HT	NMT	LLM
총 문장 수	716	785	740	791
주어 생략 원문 및 번역	346	392	365	397

양적분석에서는 표 1을 기준으로 번역 전략을 계수하고, 코퍼스별로 그 양상에 어떠한 차이가 있는지 통계 검정을 실시하였다. 통계 검정에는 Python SciPy 1.14.0 패키지를 활용하였다. 질적분석에서는 양적분석의 결과를 토대로, 통계적으로 유의한 차이가 나타난 번역전략의 병렬코퍼스를 검토하여 코퍼스별로 주어 복원 전략에 어떠한 차이가 나타났는지 고찰하였다.

4. 연구결과

본 장에서는 3장에서 기술한 절차와 방법에 따라 수행한 데이터 분석 결과를 소개한다. 4.1절에서는 표 1의 분석 틀에 따라 HT, NMT, LLM 코퍼스의 번역 전략 빈도를 계수하고 통계 검정을 실시한다. 4.2절에서는 세 코퍼스의 번역 전략 차이를 예문을 통해 상세히 고찰한다.

4.1. 양적분석 결과

각 코퍼스에서 무주어 문장의 번역 전략을 집계한 결과는 표 3과 같다. 세 코퍼스 모두 번역문에서 생략된 주어를 복원하는 전략을 가장 많이 사용한 것으로 나타났다. 인칭대명사

주어로 복원이 이루어진 경우는 NMT가 84.1%, LLM이 84.9%로 높은 비중을 차지했고 인 HT가 78.8%로 가장 낮았다. 기타 명사로 복원된 경우는 HT가 15.8%로 가장 높았고 NMT, LLM에서 각각 8.5%, 11.1%였다. 주어 복원 없이 수동태로 번역된 문장 비중은 NMT가 6.3%로 가장 높았고, 그 다음이 HT(3.6%), LLM(3.3%)에서는 소수 발견되었다.

표 3. 코퍼스별 무주어 문장 번역 전략 빈도/비중 및 기대값

코퍼스 코퍼스		추가		소계	변경		소계
		인칭대명사	기타 명사		수동태	기타	
HT	빈도	309 (78.8%)	62 (15.8%)	94.6%	7 (1.8%)	14 (3.6%)	5.4%
	기대값	323.72	46.53	-	11.20	10.53	-
NMT	빈도	307 (84.1%)	31 (8.5%)*	92.6%	23 (6.3%)*	4 (1.1%)	7.4%
	기대값	301.42	43.33	-	10.43	9.80	-
LLM	빈도	337 (84.9%)	44 (11.1%)	96.0%	3 (0.8%)*	13 (3.3%)	4.0%
	기대값	327.85	47.13	-	11.35	10.66	-

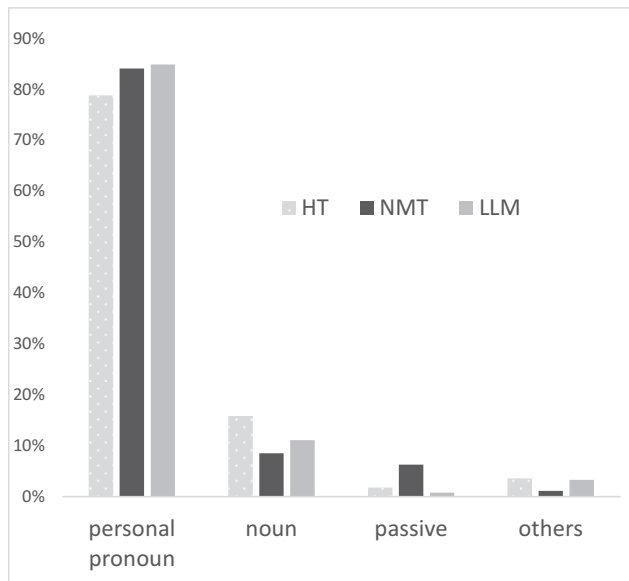


그림 2. 코퍼스별 무주어 문장 번역 전략 비중 비교

인간 번역과 NMT, LLM 간의 무주어 구문 번역 양상에 통계적으로 유의미한 차이가 있는지 확인하기 위하여 카이검정(Chi-squared test)을 실시하였다. 검정은 연구질문에 따라 두 단계로 실시하였다. 우선 세 코퍼스의 주어 복원 양상, 즉 추가 및 변경 전략 간 차이를 검정하였다. 그 후 세 코퍼스 간 세부 번역전략 간 양상을 검정하였다.

우선 세 코퍼스가 추가 또는 변경 전략을 선택하는 데에 있어 차이가 있는지 확인하기 위해 카이검정을 실시하였다. 이 때 영 가설(H_0)은 세 코퍼스가 추가, 변경 전략 빈도에 차이가 없다는 것이며, p 값이 0.05 이하이면 기각된다. 분석 결과 p 값은 0.125($df=2$, Chi-square statistics: 4.15)로, 영 가설을 기각할 수 없었다. 즉 세 코퍼스의 추가, 변경 전략 빈도 간에는 통계적으로 유의미한 차이가 발견되지 않았다.

다음은 코퍼스별 세부전략, 즉 인칭대명사, 기타 명사, 수동태, 기타 전략 빈도에 차이가 있는지를 검정하였다. 세 코퍼스 간 인칭대명사와 기타 명사를 주어로 추가하는 전략 간 차이가 없다는 영 가설에 대해 검정을 실시한 결과, p 값은 $1.21e-06(<0.01)$ 로($df=6$, Chi-square statistics: 37.82), 영가설이 기각되어 세 번역가가 무주어 문장을 번역하는 세부 전략 간에는 유의미한 차이가 존재하는 것으로 나타났다. 그러나 표 3의 데이터 빈도 간 차이가 커서 카이제곱 검정의 신뢰도가 약화될 수 있기에, 이를 보완하기 위해 각 전략 범주별로 나누어서 유의미한 차이가 있는지 추가 검정을 실시하였다. 우선 주어를 인칭대명사와 일반명사로 복원하는 양상에 대한 코퍼스별 차이가 없다는 영가설을 검정한 결과 p 값은 $0.0078(<0.01)$ 로($df=2$, Chi-square statistics: 9.70)였으며, 주어 복원을 하지 않고 변경 전략을 활용할 때 수동태와 기타 전략을 활용하는 빈도에 대한 코퍼스별 차이가 없다는 영가설을 검정한 결과 p 값은 $1.74e-05(<0.01)$ 로($df=2$, Chi-square statistics: 21.91)로 영가설이 기각되어 세 코퍼스 간 차이를 확인할 수 있었다.

코퍼스 간 차이가 어디에서 나타났는지를 살펴보기 위하여 관측치와 기대치의 차를 기대값의 제곱근으로 나눈 표준잔차(standardized residuals)를 살펴보았다. 여기에서 기대값은 카이검정에서 사용된 표 3의 기대값으로, 각 셀이 위치한 행의 합, 열의 합을 모두 더한 값을 전체 빈도로 나누는 방식으로 계산되었다. 관측치는 표 3의 관측치를 사용하였다. 통상 잔차가 2 이상일 때 유의미한 차이가 있는 것으로 해석된다.

표 4. 코퍼스별 무주어 문장 번역 전략 잔차

코퍼스	추가		변경	
	인칭대명사	기타 명사	수동태	기타
HT	-0.8182	2.2666*	-1.2573	1.0692
NMT	0.3210	-1.8733	3.8884*	-1.8538
LLM	0.5052	-0.4560	-2.4790*	0.7151

표 3의 잔차 값에 따르면 코퍼스 간 차이가 통계적으로 유의미했던 세부 전략은 HT에서 기타 명사의 과다 사용, NMT에서 수동문의 과다 사용, LLM에서 수동문의 과소 사용인 것으로 나타났다. 각 코퍼스에서 네 가지 번역 선택이 구현된 방식은 4.2절에서 상세히 살펴본다.

4.2. 질적분석 결과

본 절에서는 양적분석 결과 나타난 세 코퍼스 간 번역 전략의 차이 양상을 고찰한다. 추가 전략의 경우 세 코퍼스에서 모두 인칭대명사를 사용한 주어 복원이 가장 높은 비중을 차지했으며, HT의 경우 기타 명사로 주어를 복원하는 비중이 15.8%로, 나머지 코퍼스(NMT 8.5%, LLM 11.1%)보다 높았다. 변경 전략의 경우, 무주어 문장이 수동태로 번역되는 빈도는 NMT가 가장 높았고(6.3%), LLM의 경우 가장 낮았으며(0.8%) 이러한 차이는 모두 통계적으로 유의했다. 본 절에서는 유의하게 나타난 차이를 중심으로 각 코퍼스의 번역 전략을 살펴본다.

4.2.1. 추가 전략

생략된 주어를 복원할 때 인칭대명사를 사용하는 전략은 세 코퍼스에서 모두 가장 광범위하게 나타났다. LLM의 경우 주어 복원 시 1인칭 복수 형태의 인칭대명사를 활용하는 빈도가 가장 높은 것으로 나타났다. 이는 NMT와 인간번역에서도 가장 높은 비중으로 나타난 번역 선택이었다. (9)와 같이 기업을 뜻하는 주어가 생략된 경우, 세 코퍼스에서 모두 인칭대명사 ‘we’가 사용되었다. 통계적으로 유의하지는 않았지만, HT의 경우 잔차가 음수(-)여서 기계번역(NMT, LLM)보다 인칭대명사 주어가 적게 사용되었다는 점이 주목할 만 하다. 이는 주어 복원 시 HT가 일반 명사를 과대사용하는 경향 때문으로 추정되는 바, 4.2.2항에서 상술한다.

(9) ST: 셋째, (Ø) ESG 경영을 더욱 강화해 나가겠습니다.

HT: Thirdly, we will step up our ESG management.

NMT: Third, we will further strengthen ESG management.

LLM: Third, we will further enhance ESG management.

생략된 주어를 기타 명사 및 고유명사로 복원하는 추가 전략은 HT에서 통계적으로 유의하게 높은 것으로 나타났다. 이는 통상 기업보고서서 장르에서 독자와의 유대를 강화하기 위해 번역 과정에서 인칭대명사를 추가하는 양상(김현아, 2012; Huang & Rose, 2018; Junge, 2011)과는 정반대의 경향이었다. (11)의 이어지는 원문 두 문장에서 생략된 주어는 모두 보고서를 발간한 기업이다. NMT와 LLM은 모두 ‘we’라는 주격 인칭대명사를 추가하

는 전략을 택했다. 반면 HT는 첫 문장은 ‘project’라는 사물주어로 문장의 형태를 변경하는 전략을 취했으며, 두 번째 문장에서는 인칭대명사 대신 고유명사인 사명을 주어로 추가하였다. 이러한 선택은 텍스트의 독자 친화적 성격을 약화시킬 가능성이 존재한다.

- (10) ST: 또한 (Ø) 전 국민 디지털 역량 강화 교육 사업을 통해 취약계층의 디지털 역량을 높여주었습니다. (Ø) 거버넌스 측면에서는 이사회의 구성 및 운영을 개선했습니다.

HT(변경, 기타 명사 추가): Also, a digital capacity building project for citizens contributed to strengthening the digital capabilities of vulnerable classes. In terms of governance, SK Telecom improved its Board composition and operation.

NMT(인칭대명사 추가): In addition, through the national digital competency strengthening education project, we have enhanced the digital competency of the underprivileged. In terms of governance, we have improved the composition and operation of the board of directors.

LLM(인칭대명사 추가): Through the nationwide digital competence enhancement education project, we have also improved the digital capabilities of vulnerable groups. In terms of governance, we have improved the composition and operation of our board of directors.

비록 목표어 장르의 언어적 특성과는 다소 상반되는 선택이라고 볼 수 있지만, HT의 번역 선택이 번역자의 개입일 가능성도 존재한다. NMT와 LLM 번역을 살펴보면 이어지는 두 문장이 모두 ‘부사구+we’의 형태를 띠고 있다. 이렇게 동일한 문장 구조가 이어지면 번역문이 단조로운 인상을 줄 수 있으므로 인간 번역자가 원문에 개입했을 것이라는 추정이 가능하다. NMT와 LLM의 경우 ‘we’를 주어로 사용한 문장구조가 반복되고 있어 독자 수용성 측면에서 부정적인 영향으로 작용할 가능성이 있다.

4.2.2. 변경 전략

양적분석 결과, 무주어 구문 번역에서 수동태 전략 빈도는 NMT에서 높고 LLM에서는 낮았으며 이러한 차이는 통계적으로 유의하였다. 분석 데이터에서 발견된 수동태는 모두 단형 수동태로, 행위자 정보가 생략되어 있었다. (11)의 예에서 HT와 LLM에서는 인칭대명사를 추가하는 전략이 사용되었으나 NMT에서는 주어가 복원되지 않아 행위자가 드러나지 않는 단형 수동태가 사용되었다.

- (11) ST: (Ø) 사회(S) 부문에서는 포용적 사회를 위한 인권경영과 임직원의 안전보건 활동 강화, 책임있는 비즈니스 체계 확립을 위한 공급망의 ESG 관리 및 지원 확대 등의 노력을 기울였습니다.

HT(인칭대명사 추가): In the area of society, we strengthened human rights management; improved the safety and health of our employees; and bolstered ESG-oriented management of our supply chains and expanded support for the supply chains.

NMT(수동태): In the social (S) sector, efforts were made to strengthen human rights management for an inclusive society, safety and health activities for employees, and expand ESG management and support in the supply chain to establish a responsible business system.

LLM(인칭대명사 추가): In the social aspect, we have strengthened human rights management, occupational health and safety activities, and established a responsible business system by expanding ESG management and support within our supply chain.

2.4절에서 기술한 바와 같이 행위자와 행위 간의 거리를 부여하는 수동문은 영어로 작성된 기업보고 장르에서 통상 기업 실적이나 경영성과의 부진을 설명할 때 나타나는 언어적 특성이다. 그러나 (11)은 기업이 사회공헌 활동을 위해 어떠한 노력을 기울이고 있는지를 홍보하는 문장이다. HT와 LLM의 번역은 해당 기업을 뜻하는 생략된 주어를 ‘we’라는 인칭대명사와 능동문으로 번역하고 있다. 이는 행위자와 사회공헌 활동을 연결하는 역할을 할 뿐 아니라, 독자와의 거리를 좁혀주는 번역 선택이라고 볼 수 있다. 그러나 NMT의 수동문은 사회공헌 활동과 그 행위자인 기업 간의 연결 고리를 약화하고 행위의 능동성을 약화하는 번역 선택이라고 사료된다.

분석 결과를 요약하면 인간 번역, NMT, LLM 번역 코퍼스는 무주어 한국어 구문을 번역할 때 생략된 주어를 번역문에서도 주어로 복원하는 양상이 각각 94.6%, 92.6%, 96.0%로 나타났으며, 세 코퍼스 간 차이는 통계적으로 유의하지 않았다. 그러나 세부 전략 간에는 통계적으로 유의미한 차이가 발견되었다. 인간 번역자는 생략된 원문의 주어를 고유명사 또는 일반명사로 복원하는 빈도가 유의미하게 높은 것으로 나타났다. 또한, NMT에서는 원문에서 생략된 행위자 정보를 명시하지 않는 단형 수동태로 번역하는 빈도가 다른 코퍼스보다 유의미하게 높았으며, LLM의 경우 단형 수동태의 빈도가 타 코퍼스보다 유의미하게 낮았다.

5. 결론

본고는 한국어에서 광범위하게 나타나는 주어 생략 구문이 주어가 필수적으로 명시되어야 하는 영어로 번역될 때 인간 번역자, NMT, LLM을 통한 번역 양상에 나타나는 공통점과 차이점을 고찰하였다. 대표성 있는 소규모 병렬코퍼스를 구축하고, 원문을 대표적인 NMT, LLM 서비스인 구글번역, 챗GPT로 생산한 다음 비교, 분석한 결과 다음과 같은 결과를 얻을 수 있었다.

첫째, 본고에서 분석한 기업보고서의 경우 전체 원문 세그먼트의 약 50% 정도에서 주어가 생략되는 현상이 관찰되었다. 이는 소설 40%(김한식 외, 2019)과 시나리오 66%(박청희, 2012)의 중간 정도였다. 통상 문어체의 경우 구어체보다 주어 생략이 덜 이루어지는 것으로 알려져 있는데, 본고에서 분석 대상으로 삼은 기업보고서의 경우 서한 형태를 취하고 있었기에 빈번한 생략이 이루어진 것으로 보인다. 향후 다양한 장르에 대한 후속연구가 진행될 필요가 있다.

둘째, HT, NMT, LLM 세 코퍼스 모두 생략된 주어의 90% 이상을 번역문에서 주어로 복원하였다. 인칭대명사로 주어를 복원한 경우는 각각 81.2%(HT), 85.1%(NMT), 86.2%(LLM)였다. 이는 인간 번역자가 번역한 소설에서 나타난 76.3%(김한식 외, 2019)보다 높은 수치였다. 고유명사나 일반명사로 주어를 복원한 경우는 각각 13.9%(HT), 7.2%(NMT), 10.1%(LLM)였다. 인간 번역자의 경우 인칭대명사 대신 기타 명사로 주어를 복원하는 경우가 빈번했으며, 이는 통계적으로 유의미했다. 인칭대명사는 영어 기업보고 장르에서 담화의 대인관계적 기능을 강화하는 언어적 자원으로 간주된다. NMT와 LLM의 경우 이와 같은 장르적 특성이 잘 반영되고 있는 것으로 보였지만, 동일한 문형이 반복되어 원문 텍스트의 표현적 기능에 영향을 줄 가능성도 있는 것으로 나타났다. 반면 HT에서는 동일 문형의 반복을 피하려는 시도가 이루어졌던 것으로 추정할 수 있었다. 이렇게 전체적인 맥락에서 번역의 수용성을 높이려는 시도는 인간 번역자의 차별화된 전략이라고도 볼 수 있을 것이기에, 향후 이에 대한 체계적인 고찰이 필요할 것으로 사료된다.

셋째, 생략된 주어를 주어로 복원하지 않고 문장의 형태와 구조를 변경하는 비중은 4.9%(HT), 7.8%(NMT), 3.7%(LLM)였다. 변경 중에서는 행위자 정보를 추가하지 않고 단형수동태로 변경하는 비중이 2.0%(HT), 6.0%(NMT), 0.9%(LLM)였다. NMT의 경우 수동태가 과다사용 되었으며, LLM에서는 과소사용 되었고 이는 통계적으로 유의미한 차이로 나타났다. 원문에서 생략된 주어를 단형수동태로 변경하면 행위자 정보를 추가하지 않다고 된다는 점에서 번역 오류를 줄일 수 있는 안전한 선택일 수 있다. 그러나 기업보고 장르에서 수동태는 주로 부진한 성과에서 행위자인 기업이나 기업 대표를 분리하는 역할을 하기 때문에 담화의 수용 측면에서 수동태를 사용하는 데에는 주의가 필요하다.

본 연구의 한계점은 다음과 같다. 첫째, 본고에서 보고된 번역 양상은 기업보고서라는

특정 장르에 국한된다. 대표성 있는 코퍼스 설계를 통해 이를 보완하고자 했지만, 이를 일반화하는 데에는 무리가 있을 것이다. 그러나 타 장르의 기계번역 탐색을 위한 출발점 역할을 할 수 있을 것으로 사료된다. 둘째, 본고는 원문에서 생략된 주어가 복원되는 양상 중 대표적인 사례에만 집중했다는 한계가 있다. 2장에서 언급된 바와 같이, 원문에서 생략된 주어가 격변화를 통해 다양한 방식으로 복원될 가능성이 존재한다. 그러나 본고는 선행연구를 토대로 이 중 빈도가 절대적으로 높은 번역 전략이라고 볼 수 있는 생략된 주어 주어로 복원하는 경우만을 고려했다는 한계가 있다. 셋째, 본 연구의 실험은 GPT 모델의 최근 업데이트 전에 이루어진 것이기에, 최근 모델에서 결과가 달라질 수 있다.

상기 언급된 한계에도 불구하고, 본 연구는 세 가지 측면에서 기존 기계번역 연구에 기여한다. 첫째, 전반적인 품질 평가에 집중하여 번역 오류가 발생한 문장에 주목하는 기존 기계번역 연구의 초점을 특정 언어 현상으로 돌리고, 오류가 발생하지 않은 전체 문장의 번역 양상에 분석에 포함해야 할 필요성을 제시한다. 둘째, 한국어의 주어 생략 현상에 대한 후속 연구 필요성을 제시한다. 주어 생략은 외국어로서의 한국어 교육과 한국어를 출발어로 하는 번역에 큰 영향을 주는 문제임에도 불구하고, 장르별로 주어가 생략되는 세부적인 양상에 대한 고찰이 부족한 실정인 바, 향후 다양한 텍스트 장르별 주어 생략 양상과 주어가 생략되고 번역 과정에서 원문이 변경되는 요인과 맥락에 대한 다각적인 추가 고찰이 필요하다. 이러한 연구는 외국어로서 한국어를 학습하는 학습자와 한국어를 매개로 하는 번역자들이 특정 장르의 텍스트를 작성할 때 유용한 지침을 제공할 수 있을 것이다. 셋째, 본고는 텍스트 장르의 담화 특성, 해당 장르의 번역에서 나타나는 다양한 문학적 중재 양상을 충분히 고려하여 기계번역 결과물을 고찰할 필요가 있다는 점을 제안한다. 예를 들어, 문장을 분석 단위로 삼을 때는 생략된 주어 수동태로 번역하는 것이 충분히 수용 가능한 번역이지만, 대표이사 서한이라는 본고의 분석 텍스트 장르에서는 이것이 행위의 적극성을 약화할 수 있어, 담화의 수용과 효과에 예상치 못한 영향을 미칠 수 있다. 해당 문장의 맥락과 장르를 종합적으로 고려한 번역 선택은 적어도 현재 시점에서는 인간 번역가 고유의 특성이라 볼 수 있을 것이다. 이를 적극적으로 고찰할 필요가 있다고 사료된다. 마지막으로 본고는 대규모 언어모델 번역 결과물의 특성에 주목한 소수 연구 중 하나로, 대표성 있는 코퍼스를 대중적으로 가장 널리 활용되는 신경망기계번역과 대규모언어모델로 번역한 결과물에 존재하는 공통점과 차이점을 제시함으로써 향후 기계번역 품질 개선에 간접적으로 기여할 것으로 기대된다.

참고문헌

- 강동희. (2023). 노한 문학 번역에서 나타난 생략과 환원의 양상 고찰: 문체와 번역 전략의 관점에서. *번역학연구*, 24(1), 253-278.

- 고영근, 구본관. (2018). *우리말 문법론*. 집문당.
- 국립국어원. (n.d.) 표준국어대사전. Retrieved July 19, 2024, from <https://stdict.korean.go.kr/main/main.do>
- 김경희. (2020). 주격 인칭대명사의 생략과 번역. *통번역교육연구*, 18(4), 97-118.
- 김수정. (2016). 문어 텍스트의 장르별 주어의 실현 양상. *한민족어문학*, 72, 25-62.
- 김자경. (2021). 포스트에디팅 결과물의 정확성 오류 고찰. *통역과 번역*, 23(3), 29-58.
- 김한식, 강동희, 남슬기, 서승희, 석주희, 송신애, 최지수, & 홍승빈. (2019). 인칭대명사 생략 문장의 A-B 번역 양상. *통역과 번역*, 21(3), 31-54.
- 김현아. (2012). 번역 및 비번역<주주에게 보내는 편지>에 구축된 저자와 독자의 상호작용: 텍스트적 메타담화 분석을 중심으로. *통번역학연구*, 16(2), 115-137.
- 문귀선. (2010). 영논항의 속성 재조명. *언어학*, 18(1), 67-92.
- 박수정, 최은실. (2023). 챗GPT의 아이러니 번역 활용 가능성 고찰. *번역학연구*, 24(2), 131-160.
- 박옥수. (2018). 원천언어의 관점에서 살펴본 기계번역의 오류분석과 수정 원칙: 비문학 텍스트의 통사적 특징에 근거해서. *동아인문학*, 44, 151-171.
- 박창희. (2012). 한국어와 영어의 생략 현상에 대한 통계적 접근. *어문논집*, 66, 171-191.
- 변길자. (2024). 영어수동문의 한국어 번역에 대한 비교 분석: 기계번역과 학부생 및 통번역전공대학원생의 포스트에디팅 중심으로. *언어학연구*, 29(1), 25-50.
- 서유근. (2024). 대한항공 일반석에서 컵라면 안 주기로...이유는? *조선일보*. Retrieved August 5, 2024, from <https://www.chosun.com/economy/industry-company/2024/08/01/C2B3CZAVEVBY7EZA72GMKGVWFE/>
- 신천, 조혜진. (2024). 위화 소설 『허삼관 매혈기』 중한번역을 통해 본 문학번역에서의 챗GPT의 활용 가능성. *언어학연구*, 29(1), 51-65.
- 왕칭둥. (2024). 대만인 한국어 학습자의 한중 문장 대역 양상-대화문 주어를 중심으로. *아시아문화학지*, 3, 115-141.
- 이유정. (2023). 현대시 인공지능(AI) 번역의 오류 양상 연구. *문화와융합*, 45(10), 97-110.
- 이지은, 최효은. (2020). 코퍼스 연구를 통해 살펴본 법령 번역 텍스트의 언어적 특성: 수동태 구문을 중심으로. *번역학연구*, 21(2), 251-284.
- 이지은, 최효은. (2023). 원천텍스트 난이도와 기계번역 품질에 대한 고찰-구글 한영 법령번역 사례를 중심으로. *언어학연구*, 28(1), 77-101.
- 이창수. (2014). 번역한국어와 비번역한국어 간 타동형 감정구문에서의 무생물 주어 사용 차이 연구. *통번역학연구*, 18(1), 123-141.
- 임소정, 이아형. (2024). 한중 번역에서의 인칭대명사 주어 처리 양상 대조—소설 《프라자호텔》을 중심으로. *중국문학*, 118, 137-159.

- 전설주. (2024). 한국어 모어 화자와 중국인 한국어 학습자의 영주어 의미 해석 양상. *사회언어학*, 32(2), 79-107.
- 정연창. (2007). 주어 생략에 대한 소고. *언어과학*, 14(2), 101-120.
- 조승연. (2023). 인간번역과 기계번역의 여성어·남성어 비교 연구: 한일 번역본 [82년생 김지영]의 문말(文末) 표현을 중심으로. *언어학*, 31(1), 45-63.
- 진실로, 곽은주. (2013). *번역수업* 101. 한국문화사.
- 최효은. (2024). 대중과학의 영한 번역에 있어서 인간 번역과 ChatGPT 번역의 명시화(explicitation) 양상 고찰 - <BBC Science>의 표제 분석을 중심으로 -. *T&I Review*, 14(1), 149-175.
- 한국거래소. (n.d.). KTOP30. Retrieved August 8, 2024, from <http://data.krx.co.kr/contents/MDC/EASY/visualController/MDCEASY500.cmd>
- 한현희. (2020). 한-노 기계 번역, 어디까지 왔나?: Google과 Papago 번역 성능 비교를 기반으로. *노어노문학*, 32(3), 63-93.
- Asay, H. S., Libby, R., & Rennekamp, K. (2018). Firm performance, reporting goals, and language choices in narrative disclosures. *Journal of Accounting and Economics*, 65(2-3), 380-398.
- Baker, M. (1992). *In other words: A coursebook on translation*. Routledge.
- Bakker, M., Koster, C., & Van Leuven-Zwart, K. (2009). Shifts. In M. Baker & G. Saldanha (Eds.), *Routledge encyclopedia of translation studies* (2nd edition) (pp. 269-273). Routledge.
- Biber, D., Stig Johansson, Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English* (6th ed). Longman.
- Catford, J. (1965). *A linguistic theory of translation: An essay in applied linguistics*. Oxford University Press.
- Huang, Y., & Rose, K. (2018). You, our shareholders: Metadiscourse in CEO letters from Chinese and Western banks. *Text and Talk*, 38(2), 167-190.
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.
- Junge, S. (2011). Corporate rhetoric in English and Japanese business reports. In S. Kranich, V. Becher, S. Hoder, & J. House (Eds.), *Multilingual discourse production: Diachronic and synchronic perspectives* (pp. 209-232). John Benjamins Publishing Company.
- Kiaer, J. (2018). *Korean Translation*. Routledge.
- Klaudy, K. (2009). Explicitation. In M. Baker & G. Saldanha (Eds.), *Encyclopedia of Translation Studies* (pp. 104-108). Routledge.

- Nida, E. A. (1964). *Toward a science of translating: With special reference to principles and procedures involved in Bible translating*. Brill.
- SciPy 1.14.0, Eric Jones, Travis Oliphant, Pearu Peterson and others. Title: SciPy: Open Source Scientific Tools for Python.
- Seo, Y., & Lee, J. (2024). Korean Air to stop serving cup noodles in economy class. *Chosun Daily*. Retrieved August 5, 2024, from <https://www.chosun.com/english/national-en/2024/08/01/26MU2JWNV5G4VOH256OM7NUKIQ/>
- Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- Wang, K., Zhao, X., Li, Y., & Peng, W. (2023). Prose: A pronoun omission solution for Chinese-English spoken language translation. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2297-2311.
- Yim, J. (2024). Sentence length and translation: A comparative review of human, NMT, and LLM translations. *T&I Review*, 14(1), 69-93.
- Yim, J., & Lee, Y. (2024). Sensitivity of translation universals to genre/ register variations: Focused on corporate reporting. *The Linguistic Association of Korea Journal*, 32(2), 153-173.

임진

서울특별시 서대문구 이화여대길 52
이화여자대학교 통변역대학원 강사
전화: +82-2-3277-3662
이메일: jy2812@gmail.com

Received on August 11, 2024

Revised version received on September 23, 2024

Accepted on September 30, 2024