

한국어 속담의 영어 기계번역에 미치는 격조사의 영향 — 구글번역기를 중심으로 —

김경숙
(승실대학교)

Kim, Kyungsook. (2018). The influence of Case markers on the machine-translation of Korean proverbs into English. *The Linguistic Association of Korea Journal*, 26(3), 139-157. This study investigates the influence of case markers on the machine-translation of Korean proverbs into English. While grammatical roles are determined by the word order in English, they are indicated by case markers in Korean. The problem is that Korean case markers are optionally deleted. It is, therefore, hypothesized that the presence of case markers in the original text is to be crucially important to the success rate of machine-translation from Korean into English. For this study, the characteristics of Google Translate, the translation system used here, were first surveyed. Then, sentences from three different categories of Korean texts (spoken, written, and proverb texts) were chosen and machine-translated using Google Translate. The findings are as follows. First, the Korean-English machine-translation of three different text types shows that there were more syntactic errors than lexical errors across the categories. Second, there were more errors in complex sentences than in simplex sentences. Third, the error rate was highest in the proverb category and lowest in the written text category. Fourth, across the three categories, the error rate was the highest when deletion of case markers or multi-case markers was involved. Lastly, the supplement of missing case markers to the original Korean proverbs substantially enhanced the success rate of machine-translation. This shows that the presence of case markers plays a crucial role in machine-translation into English, which, in turn, suggests that we can improve the success rate of machine-translation from Korean into English if we add the missing case markers to the Korean input sentences.

주제어(Key Words): 구글번역기(Google Translate), 한영 기계번역(Korean-English machine translation), 격조사(case markers), 번역 오류(translation error), 통사오류(syntactic error), 속담(proverbs)

1. 머리말

기계번역 연구를 주도하고 있는 구글번역기의 번역 방식은 기본적으로 통계적 기계번역(위키백과)으로, 최근에 인공신경망 기술을 접목하여 한층 개선된 번역 서비스를 제공하면서 국내에서도 기계번역기의 이용 빈도가 높아지고 있다. 그런데 기계번역기의 번역 결과가 예전에 비해 어색하지 않고 자연스러우며 정확도가 높아졌다는 평가를 받고 있지만, 적지 않은 경우에 여전히 부정확한 번역을 출력시키거나 심지어는 당황스러운 정도로 엉뚱한 번역을 내놓기도 한다. 기계번역이 만족스럽게 이루어지지 못하는 원인은 매우 복잡적이어서 다방면의 연구와 극복 방안이 나와야 하지만, 그 원인 중의 하나는 한국어 격조사가 가지는 독특한 성격과 깊은 관련이 있어 보인다.

주지하다시피, 영어는 기본적으로 어순에 따라 주어, 목적어와 같은 문법적 역할(grammatical roles)이 결정되고, 한국어는 명사에 후치되는 격조사에 의해 문법관계가 표시되며 의미역이 결정된다. 즉, 영어는 문장 내의 위치에 따라 주어와 목적어의 통사적인 기능이 결정되는 반면에, 한국어는 문법관계를 나타내는 격표지의 형태에 의해 의미가 결정되는 것이다. 그런데 여기에서 결정적인 차이점이 하나 더 개입되는데, 즉 한국어의 격조사는 상황에 따라서 종종 생략될 수 있다는 것이다. 이와 별도로, 한국어에는 동일한 조사가, 예를 들어서 ‘~은/~는’이, 주어의 역할을 하기도 하고 목적어의 역할을 하기도 한다.

격조사의 표면상 존재 여부에 따라 한영 기계번역이 매우 상이한 결과, 즉 번역 오류를 범하는 것은 아주 흔한 일이다. 일례로 구글번역기에 ‘좋은 아빠 되세요’를 입력해보면, ‘Have a good father’로 원문의 의도와는 다른 번역이 되지만, 같은 예문의 ‘아빠’에 격조사 ‘~가’를 붙인 후 ‘좋은 아빠가 되세요’를 입력하면, ‘Be a good father’로 원문의 의미가 정상적으로 번역된다. 이러한 예는 별다른 의미 차이가 없지만 격조사가 있는 경우와 생략된 경우의 한국어 데이터를 기계번역기에 입력하고 영어로 번역할 때, 번역의 결과가 원문의 의미와는 다르게 나타날 수 있음을 보여준다.

기계번역기를 통한 한영 번역에서 나타나는 오류는 크게 어휘오류와 통사오류로 나누어 볼 수 있다. 어휘오류는 한국어 고유명사, 의성어, 동음이의어의 다의성이나 중의성(ambiguity), 또는 한자어휘의 사용이나 띄어쓰기의 잘못으로 인해 발생하는 오류로, 영어 번역 후 문맥(context)에서 어휘 의미를 바로잡지 않으면 안 되는 경우이다. 반면, 통사오류는 한국어와 영어의 통사적 불일치인 어순의 차이나 한국어 특성인 격조사로 인한 것이 대표적인 것으로, 이로 인해 번역의 정확성이 심각한 수준으로 떨어지기도 한다.

최근에 이러한 통사오류의 문제를 인식하고 한영 기계번역에서의 오역을 줄이기 위해 기계번역 과정의 개선 방법을 모색하는 연구가 이루어지고 있다.(배지연·김수연, 2010; 김원경, 2011; 박영란, 2017) 그러나 기계번역에서의 통사오류를 알아보기 위한 연구 중에서, 특정 유형의 텍스트를 대상으로 오류의 실례와 원인을 객관적인 수치로 비교하고, 그 결과를 분석

한 연구는 많지 않다. 이에 본고에서는 가장 먼저 기계번역을 시작하고 세계적으로 100여개 언어의 번역 서비스를 제공하고 있는 구글번역기를 활용하여, 한국어 세 가지 유형의 문어, 구어, 속담 텍스트를 입력하여 기계번역을 시도한 후 그 결과를 분석하고자 한다.

기계번역의 결과를 가지고, 먼저 번역 오류의 유형을 어휘오류와 통사오류로 분류한다. (그런데 본 연구에서는 한국어의 특성상 나타나는 어휘오류는 분석의 대상에서 제외하고, 통사오류에 초점을 두기로 한다.) 통사오류는 두 언어 간의 구조적 차이로 인해 주로 발생하지만, 여기서는 한국어 격조사 표지의 유무와 다중 유표¹⁾로 인한 오류에 집중하고, 텍스트 유형별로 통사오류의 빈도를 계량적으로 비교한다. 특별히 한국어 속담을 영어 기계번역기로 번역할 때 격조사가 미치는 영향에 대하여 구체적으로 살펴보고, 그 결과를 바탕으로 한영 기계번역에서 통사오류를 발생시키는 원인을 조사한 후 번역 오류를 줄일 효율적인 방안을 모색한다.

이를 위해 다음과 같은 단계로 연구를 진행한다. 첫째, 구글번역기의 번역 방식과 번역 시스템을 알아본다. 둘째, 한국어 데이터를 구어, 문어, 그리고 한국어 고유의 향토적 속성을 보이는 속담 텍스트로 나누어 각각 100개 문장씩 수집하고 단문과 복문으로 표시한 후 구글번역기에 입력한다. 셋째, 텍스트 유형별로 영어 번역 결과에서 발생한 오류를 어휘오류와 통사오류로 분류한다. 넷째, 통사오류에 초점을 두고 오류의 원인을 격조사 표지의 유표, 무표, 다중 유표로 나누어 텍스트 유형별로 빈도를 조사한다. 다섯째, 통사오류 빈도가 좀 더 높은 것으로 나타난 속담을 대상으로 오류 속담에 격조사 표지를 복원한 후 다시 번역하고 그 결과를 조사한다. 마지막으로, 이러한 결과를 바탕으로 한영 기계번역 시 번역 오류를 최소화하고 번역의 완성도를 높일 방법을 논의하고자 한다.

2. 연구 배경

2장에서는 본 연구를 위해 구글번역기의 번역 방식과 한국어 격조사에 대해 알아보고, 격조사와 한영 기계번역과 관련한 선행연구를 살펴보았다.

2.1. 구글번역기의 번역 방식과 한영 기계번역에 대한 기존 연구

먼저 본고의 한영 기계번역의 도구로 활용한 구글번역기의 번역 방식을 알아보고 기계번역에 대한 기존의 연구를 살펴보도록 한다. 나무위키에 의하면, 구글번역은 무료 번역 서비스로서 2006년 처음 만들어졌을 때는 4개 외국어인 영어, 중국어, 프랑스어, 스페인어로 시작

1) 본고에서는 다중 유표를 하나의 조사가 두 개 이상의 격 역할을 한다는 ‘문법적’ 의미로 사용하는데, 이는 전통적으로 보조사라고 불리며 일정 부분 담화표지의 기능도 수행한다.

했다. 2012년에는 63개 언어, 2013년에는 71개 언어를 지원했다. 통계적 기계번역으로 번역하기 때문에 정확도가 높다고 했지만 사실은 그렇지 못했다. 그러나 2016년 인공지능명을 기반으로 한 번역 서비스가 도입되었으며, 딥러닝 기반 번역 지원으로 직역도 자연스럽게 정확도가 높아지고 있다고 한다. 또한 구글번역 프로그램은 국제기구나 기업 등에서 동일한 문서를 A언어, B언어, C언어 등의 여러 언어로 만들어 놓은 데이터를 검색 엔진이 검색하여, 많은 결과를 바탕으로 이루어진다는 것이다. 이러한 번역 원리를 말뭉치 기반 기계번역(corpus-based machine translation; CBMT)이라고 일컬으며, 두 언어 간의 병렬 말뭉치(bilingual parallel corpus)를 기반으로 비교분석하여 번역 결과물을 출력하는 과정을 거친다. 구글번역기의 말뭉치를 만드는 원리는 통계적 기계번역(statistical machine translation; SMT)에 해당하는데, 단어 대 단어, 구절 대 구절의 사용 빈도를 나타내는 모델을 만들어서 어떤 단어 및 구절의 번역으로 가장 많이 나타나는 표현을 번역의 결과로 보여주기 때문에 확률적으로 정확하다는 것이다. 즉 정확한 표현이 선택되면 배열되는 순서도 가장 많이 나타나는 순서가 정확하다고 판단하여 번역물을 출력한다는 것인데, 여전히 구글번역은 번역 대상 언어의 격조사와 같은 문법적 지식을 바탕으로 한 번역이라기보다는 말뭉치 기반의 번역 방식이기에 오류가 있을 수 있다.

최근에는 구글번역기 외에도 여러 종류의 기계번역기가 상용되고 있는 가운데 기계번역에 대한 연구들이 많이 이루어지고 있다. 한국어를 영어로 번역하는 기계번역에 관한 연구들은 입력 데이터인 한국어와 번역 결과인 영어 간의 통사 불일치와 같은 문제들을 조사한 것들이 대부분이다. 먼저, 박홍원·심재석 등(2001)은 한영 번역 시스템에서 불특정 조사를 포함한 용언구 처리 문제를 연구했는데, ‘공부를 하다’와 ‘잠이나 자다’와 같은 어구는 명사와 동사의 결합이 자유로워 기계번역 시스템에서 이들을 처리할 때 문제가 생긴다는 것이다. 따라서 명사 부분과 용언 부분을 각각의 품사 사전에서 탐색하여 품사별로 인식한 후에, 다시 통사적으로 하나의 용언으로 인식시켜 번역문을 생성시켜야 한다고 주장하였다. 그리고 최재혁·김권양 등(1989)은 한영 기계번역을 위한 조사 사전 구성에 관한 연구에서, 한국어의 조사에 대응하는 영어 전치사의 수가 많아서 어려울 수 있기 때문에, 정확히 번역하기 위한 방법으로 한국어 조사의 격 분류와 체언의 속성 정보를 세분화할 것을 제안하였다. 또한 조준모·이상조(1994)는 한영 기계번역 시스템의 목적에 맞게 부사의 특징을 조사하여 분류했고, 이것을 기초로 하여 부사의 위치와 순서를 처리하는 기틀을 마련했다. 그리고 배지연·김수연(2010)은 한영 기계번역에 있어서 두 언어의 격 실현 체계의 차이에 따른 문제점을 밝히고 한국어와 영어의 격 체계의 차이를 고려한 기계번역이 요구된다고 지적하였다.

이외에도 이하규·김연택(1994)은 한영 기계번역 시스템에서의 한국어 속어의 표현 및 숙어인식 방법을 기술했으며, 박영란(2017)은 기계번역기에서 영어 속담을 번역할 때 구글과 네이버 번역기에서 어느 번역기가 더 잘 해내는지 평가하고 오역의 결과를 비교하였다. 또한 이병희(1994)는 한영 기계번역 시스템의 오역 중에서도 시제 처리 때문에 발생하는 오류에

중점을 두고, 한국어의 시제와 영어의 시제를 비교하여 시제의 구조를 개념그라프에 기반하여 기술했다.

2.2. 한국어 격조사 연구

기계번역기를 통해 한국어를 영어로 번역하는 과정에서 발생하는 오류를 조사할 때 한국어의 격에 대한 논의는 꼭 필요한 것으로, 본 절에서는 격조사에 대하여 알아보고 이와 관련한 연구들을 살펴본다.

먼저 한국어 조사에 대해 알아보면, 『표준국어문법론』에서는 조사를 “자립성이 있는 말에 붙어 그 말과 다른 말과의 관계를 표시하는 품사”로 정의하고 있다. 전통문법에서 조사는 크게 격조사와 보조사로 구분해왔는데, 격조사는 원칙적으로 체언에 붙어 그 말의 다른 말에 대한 관계를 표시하는 것이다. 곧 체언으로 하여금 일정한 자격을 갖도록 하는 것으로, 격조사에는 주격, 서술격, 목적격, 보격, 관형격, 부사격, 호격의 7격이 있다. 이러한 격조사에서 주격과 목적격 조사는 생략된 경우가 있는데 주로 구어체에서 현저하게 목격된다. 가령 ‘그 사람(이) 무슨 책(을) 읽었니?’에서는 주격과 목적격 조사가 생략된 경우로, 이런 생략현상은 체언만으로도 격 관계가 분명할 때 나타난다, 또한 조사에는 단순한 격표시 이외에 특수한 뜻을 더하는 조사가 있는데 이것이 보조사로 ‘만’과 ‘도’가 있다. 예를 들어, ‘철수는 좋은 아이다.’, ‘철수도 좋은 아이다.’, ‘철수만 점심을 먹는다.’에서 ‘는, 도, 만’과 같은 조사는 주격의 자리에서 그 의미는 ‘대조’, ‘역시’와 ‘단독’으로 설정되어, 체언의 격을 표시하는 격조사와는 달리 뜻을 더해준다는 것이다.(남기심·고영근, 1985, pp. 99-105)

한국어의 격조사와 관련한 연구들도 다양한데, 이종근(2018)은 한국어 조사인 ‘~을/를/이/가’는 단순한 격표지라기 보다는 좀 더 넓은 개념인 논항과 논항을 변별하고, 개체/개념과 다른 대안 개체/개념을 변별하는 변별 표지로 해석해야 한다고 주장하고, 엄정호(2011)는 격의 정의와 기능에 대한 연구에서 ‘~가’와 ‘~를’에 대하여 격조사가 아니라는 견해도 있지만, 기능으로 격을 이해한다면 이 조사들 역시 여전히 격조사로 분류될 수 있다고 하였다.

격조사의 기능에 관한 다른 연구에서 이호승(2012)은, 국어의 격조사는 통사기능이나 의미역을 표시해 주는 기능을 수행하는데, 국어 격 범주의 이러한 사실은 ‘격=격 조사’라는 등식을 성립시키는 바탕이라고 하였다. 그는 또 다른 논문(2006)인 격조사 없는 명사구의 격 문제에 관한 연구에서, 격조사를 갖는 명사구는 해당 격을 갖는 것이며, 격조사를 갖지 않는 명사구는 격을 갖지 않는 것으로 보고, 이 명사구들도 문장성분으로 처리될 수 있다고 하였다.

또한 한국어 조사의 격과 의미에 관한 연구에서 홍기선(2017)은, 영어의 문법격은 문법 기능을 통해 간접적으로 의미역 정보와 연결되는데 반해, 한국어의 격조사는 절 내에서 주어와 목적어에만 제한되지 않고, 상황의미 조건이 충족되는 모든 위치에 나타날 수 있으며 담

화의미를 수반한다고 밝히고 있다.

마지막으로, 한국어의 격 범주에 관한 연구에서 김원경(2009)은, 최근에 기계번역과 같은 한국어정보처리에 적용하기 위하여, 대상 언어의 분석과 목표 언어의 생성을 연결하는 언어 보편적인 격체계에 대한 연구가 활발히 진행되고 있는데, 언어 유형적으로 격 범주는 ‘위치’에 의존하는 유형과 ‘표지’에 의존하는 유형으로 나누어 살필 수 있다고 주장하였다. 또한 그는 한국어와 같이 조사인 접미사의 부착으로 격을 표시하고 격 표지에 의존하여 격을 표시하는 언어를 ‘M언어(Marker-Language)’라 하고, 영어와 같이 위치에 의해 어순이 갖는 정보가 매우 유용하게 활용되는 언어를 ‘L언어(Location-Language)’로 분류하였다.

지금까지 구글번역기의 번역 방식, 한국어 격조사, 한영 기계번역 시스템에서 이들을 어떻게 처리하는지에 관한 연구들을 살펴보았다. 그런데 기계번역에서 격조사의 격 기능이나 실현과 관련한 연구는 어느 정도 이루어지고 있지만, 기계번역에서의 번역 오류 요인을 한국어 격조사의 유무로 특정하여, 텍스트 유형별로 나누어 조사하고 객관화된 수치로 비교한 연구는 부족하였음을 알 수 있다. 이런 점에 착안하여, 본 연구에서는 한국어 텍스트 세 가지 유형을 기계번역기에서 영어로 번역한 후, 텍스트 유형별로 격조사관련 통사오류의 양상과 원인을 조사 비교하고, 그 결과를 바탕으로 기계번역기에서의 번역 오류를 줄이기 위한 방안을 논의하고자 한다.

3. 연구 목표와 연구 방법

최근에 많이 이용되고 있는 구글번역기는 100여개 이상의 언어를 번역해주고 있는데, 인공신경망 기술을 이용한 통계적 번역 시스템으로 한층 개선된 번역 서비스를 제공하여 번역 결과가 상당히 만족스럽게 나온다는 평가를 받고 있다. 하지만 여전히 번역 오류가 발생된다는 점에 주목하여, 본 연구에서는 구글번역기에서의 한영 번역에 초점을 두고 오류의 양상과 원인을 조사하였다. 먼저 한국어 데이터는 세 가지 유형의 구어, 문어, 속담 텍스트로 나누어 수집하였는데, 구어와 문어 외에 속담을 추가한 것은 속담이 옛 사람들의 생각과 문화를 잘 반영한 특유의 문체를 갖고 있다는 점에서 영어 번역의 결과가 어떠할지 궁금하였기 때문이다. 그런데 여기서 **pro-drop** 현상으로 주어(**understood subject**) 전체가 생략된 경우는 명사 자체가 드러나 있지 않으므로, 격조사의 유무를 따지는 것이 무의미하여 분석대상에서 제외하였다. 그리고 이 세 유형의 한국어 데이터를 구글번역기에 입력한 후 발생한 번역 오류를 어휘오류와 통사오류로 분류하고, 이 통사오류의 원인을 격조사의 유무로 나누어 결과를 비교하였다. 또한 세 유형 중에서 통사오류 빈도가 좀 더 높게 나타난 속담을 대상으로, 격조사 표지가 생략된 속담에 적절한 격조사 표지를 복원한 후 영어로 다시 번역하고 그 결과를 살펴보았다. 마지막으로, 현재의 한영 기계번역기 시스템에서 번역의 완성도를 높이기 위한

방안으로, 한국어 원문을 입력할 때 오해의 소지가 있는 격조사 표지의 생략된 부분을 확인하고 통제할 필요가 있음을 제안한다.

본 연구는 다음과 같은 방법으로 진행되었다. 먼저, 한국어 데이터를 구어, 문어, 속담의 세 가지 유형으로 나누어 수집하였는데, 구어 텍스트는 2013년 10월부터 KBS에서 방영된 드라마 굿닥터에서 1회 차부터 5회 차까지 각 회당 20 문장씩 총 100 개의 발화(utterance)를 수집하였고, 문어 텍스트는 배명진·김명숙의 『소리로 읽는 세상』에서 각 장당 20 문장씩 총 100문장을 수집하였다. 그리고 속담 텍스트는 인터넷 다음 블로그의 <재미있는 속담 1000개>에서 발화 수준의 속담을 가나다의 순서로 각 10여개씩 총 100 문장을 수집하였다.

위와 같은 방법으로 수집한 세 가지 유형의 한국어 데이터를 단문과 복문으로 구분하여 구글번역기에 입력한 후, 번역된 영어에서 나타난 오류를 어휘오류와 통사오류로 분류하고, 한국어 고유명사, 의성어, 동음이의어와 같이 언어 고유의 특성으로 인해 발생하는 어휘오류는 분석대상에서 배제하였다. 통사오류는 다시 격조사 표지가 있는 것을 유표로, 생략된 것을 무표로, ‘~는, ~만, ~도’와 같은 조사의 사용은 다중 유표로 분류하여 유형별로 빈도를 조사하고 비교하였다. ‘~는, ~만, ~도’를 다중 유표라는 명칭으로 별도로 분류한 이유는 이 조사들이 경우에 따라서 주격도 될 수 있고 목적격이 될 수 있기 때문이다. 이렇게 다중적인 기능을 하는 이 조사들은 주격이나 목적격을 확정적으로 표시하는 유표적 조사나, 조사가 아예 표시되지 않은 무표적 상황과 다르기 때문에 다중 유표라는 명칭으로 독립적으로 취급하였다.

다음은, 세 유형의 텍스트에서 통사오류 빈도가 좀 더 높게 나타났던 속담을 대상으로 오류가 발생한 속담 텍스트에 격조사 표지를 적절한 것으로 복원한 후, 기계번역기에 다시 입력하고 번역 결과를 조사하였다. 이때 속담의 원문에 다른 어휘를 추가하지 않고 격조사만을 복원하였다.

본고에서는 이들 자료와 연구방법을 통해 다음과 같은 주제로 연구를 진행하였다. 먼저 한국어 데이터를 세 가지 유형의 텍스트로 나누어 수집하고 기계번역기에 입력한 후, 발생된 오류를 어휘오류와 격조사관련 통사오류로 구분하여 텍스트 유형별로 조사하고 둘째, 통사오류가 좀 더 많이 발생한 속담에 격조사 표지를 복원한 후 다시 번역하고, 그 결과를 바탕으로 한국어 속담을 기계번역기로 번역할 때 격조사가 어떤 영향을 미치는지 살펴보고 오류를 줄일 방안에 대해 논의하였다.

4. 텍스트 유형별 기계번역의 통사오류

4.1. 격조사관련 오류

본 절에서는 한영 기계번역기에서 발생하는 오류를 텍스트 유형별로 어휘오류와 통사오

류로 나누어 빈도를 조사하고, 특별히 통사오류의 원인을 격조사와 관련하여 격표지가 있는 경우는 유표로, 생략된 경우는 무표로, 그리고 다중 유표로 분류하고 텍스트 유형별로 오류를 조사하여 비교하였다.

4.1.1. 구어 텍스트

표준국어대사전에 의하면 구어체는 문어체와 대응하는 문체로서 글에서 쓰는 말투가 아닌 일상적인 대화에서 주로 쓰는 말투이다. 구어체의 특징은 일상생활에서 사용되는 비교적 쉬운 표현으로 문어체보다 비교적 격식을 차리지 않고 생생하고 자연스런 표현이 많으며, 문장이 짧고, 격조사나 단어가 생략되기도 하며, 상황에 따라서 높임말이나 사투리가 사용되기도 하고 비문법적인 경우도 많이 있다는 것이다.

본 절에서는 구어 텍스트의 한영 기계번역 시 발생하는 오류 양상을 알아보기로 한다. 이를 위해 입력 데이터로 2013년 10월부터 20부작으로 방영된 KBS 드라마 ‘굿닥터’를 선정하여 1회 차에서 5회 차까지 각 회당 20문장씩 총 100문장을 수집하였다. 이 드라마는 서번트 증후군을 갖고 있는 주인공이 소아과 의사가 되고 많은 사람들을 만나면서 사람들에게 진정으로 무엇이 필요한지를 보여준 힐링 의학 드라마로, 등장인물들이 나누는 대화에는 사투리도 자주 나타난다. 이렇게 수집한 드라마 대사를 각각 단문과 복문으로 표시하여 구글번역기에 입력하고, 영어 번역 결과 나타난 오류를 어휘오류와 통사오류로 나누어 빈도를 조사했다. 그리고 격조사로 인한 통사오류를 격조사 표지의 유표, 무표, 그리고 다중 유표로 분류하여 결과를 비교하였다.

다음의 표 1은 구어 텍스트의 한영 기계번역 시 발생한 어휘오류와 격조사관련 통사오류의 빈도를 비교한 것이다.

표 1. 구어 텍스트에서의 어휘오류와 격조사관련 통사오류 빈도

	어휘오류	통사오류(격조사)		
		유표	무표	다중 유표
단문(51)	8	2	7	2
복문(49)	5	0	19	3
계	13	2	26	5
합계	13	33		

위의 표를 보면, 구어 텍스트의 단문과 복문의 비율은 51개와 49개로 비슷하다. 번역의 결과 나타난 어휘오류와 통사오류는 각각 13개와 33개로, 어휘오류보다는 통사오류의 빈도가 상당히 높다는 것을 알 수 있다. 통사오류의 빈도를 자세히 살펴보면, 단문에서 11개 복문에서 22개의 오류로, 복문에서 오류가 두 배 정도 많다. 또한 요인별 통사오류 빈도에서, 격

조사가 있는 유표의 경우에 2개, 생략된 경우에 26개, 다중 유표의 경우에 5개의 오류가 나타났다는데, 격조사가 생략된 경우에 오류가 훨씬 많다. 이 결과는 구어 텍스트에 격조사 표지가 생략되는 무표의 경우가 많고, 이로 인해 구어 텍스트의 기계번역 시 오역이 종종 발생할 수 있다는 것을 보여준다.

다음의 표 2는 구어 텍스트의 기계번역 시 발생한 격조사관련 오류의 예시이다.

표 2. 구어 텍스트에서의 격조사관련 오류 예시

한국어(입력)	영어(결과)
형이 용돈 모아 샀어!!	My brother bought us some money.!!
너 그런 거 없잖아.	You do not have to do that.
같이 놀자 그래야지.	Let's play together like that.
니가 계속 놀아 달라 그러면 동구 새끼도 받아 줄 거야.	If you keep playing and you will get a puppy.
아주 교육들 한 번 제대로 시켰구만.	It was very educational once I got it right.
하늘나라 가기 전에 잘 치료하면 안 갈 수도 있어.	Heaven You can not go to healing before going to the country.
의사가 되면 하늘나라 안 가게 할 수 있어요?	When Zion is a doctor. can you let the sky go?
나중에 엄마랑 형아 아프면 니가 다 고쳐 줘야 돼. 알았지?	If you're sick and you're sick later, you I have to fix everything. know?
갖고 나오면 시온이 니네 편에 껴주는 거다.	If you come out with a sword, Zion will put it on your side.
죽은 애 부모, 의료사고니 뭐니 하면서 100% 소송 건다고!	Dead parents, medical accidents, whatever you say, 100% suits!

위의 예에서 보듯이, ‘너 그런 거 없잖아’는 ‘You do not have to do that’으로 번역되어, 원문의 의미와는 다른 번역임을 알 수 있다. 그런데 원문의 ‘너’에 ‘~는’ 이라는 조사를 복원하여 다시 입력하면, ‘You do not have that’으로 자연스러운 번역이 된다. 또한, ‘죽은 애 부모, 의료사고니 뭐니 하면서 100% 소송 건다고!’는 ‘Dead parents, medical accidents, whatever you say, 100% suits!’로 번역되었는데, 여기서 죽은 애의 부모를 의미하는 ‘죽은 애 부모’가 ‘Dead parents’로 번역이 된 것이다. 그러나 원문의 ‘죽은 애’에 관형격 조사인 ‘~의’와 ‘부모’에 주격 조사인 ‘~가’를 붙여서 다시 입력해보면, ‘Parents of

dead children say they are 100% sued for medical care!’로 원문과 동일한 의미의 정상적인 번역이 된다. 이러한 오역의 예시는 한국어 텍스트의 기계번역 시 격조사가 미치는 영향이 크다는 점을 보여준다.

4.1.2. 문어 텍스트

두산백과에서는 문어체를 일반적으로 현대의 일상생활에서 사용되는 구어체에 비하여 잘 쓰이지 않는 보다 오래된 시대의 말투가 많이 들어가는 문체로 구어에 대응한다고 설명하고 있다. 여기서는 이러한 특징의 문어체를 기계번역 할 때 나타나는 오류 양상을 살펴보기로 한다. 입력 데이터로 『소리로 읽는 세상』의 각 장에서 20 문장씩 총 100 문장을 수집하여 각각 단문과 복문으로 표시하고 구글번역기에 입력한 후, 발생한 번역 오류를 어휘오류와 통사오류로 나누어 빈도를 조사하였다. 그리고 격조사로 인한 통사오류를 격조사 표지의 유표, 무표, 그리고 다중 유표로 나누어 빈도를 조사하였다.

다음의 표 3은 문어 텍스트의 한영 기계번역 시 발생한 어휘오류와 격조사관련 통사오류의 빈도를 비교한 것이다.

표 3. 문어 텍스트에서의 어휘오류와 격조사관련 통사오류 빈도

	어휘오류	통사오류(격조사)		
		유표	무표	다중 유표
단문(22)	0	1	1	0
복문(78)	7	4	7	2
계	7	5	8	2
합계	7	15		

위의 표를 보면, 문어 텍스트의 단문과 복문 비율은 22개와 78개로 복문이 많고, 번역의 결과 나타난 어휘오류와 통사오류는 각각 7개와 15개로, 어휘오류보다는 통사오류의 빈도가 상당히 높다는 것을 알 수 있다. 통사오류의 빈도를 좀 더 자세히 살펴보면, 단문과 복문의 경우에 각각 2개와 13개의 오류로, 복문에서 통사오류가 훨씬 많다. 또한 요인별 통사오류의 빈도에서, 격조사 표지가 있는 경우에 5개, 생략된 경우에 8개, 그리고 다중 유표의 경우에 2개의 오류가 나타났는데, 이러한 문어 텍스트에서의 결과 역시 격조사 표지가 생략된 경우에 오류가 더 많이 발생한다는 것을 보여준다.

다음의 표 4는 문어 텍스트의 기계번역 시 나타난 격조사관련 오류의 예시이다.

표 4. 문어 텍스트에서의 격조사관련 오류 예시

한국어(입력)	영어(결과)
대부분의 사람들은 엄마 배 속에서 소리를 처음 듣는다.	Most people hear the sound in their belly for the first time.
엄마와 아빠가 내 이름을 불러주는 소리, 가족들의 웃음소리, 즐거운 노랫소리, 잔잔한 바람 소리나 시원한 파도 소리일 수도 있지만.	Mom and dad call my name, family laughter, joyful song, gentle wind or cool wave sound.
동굴 안에서는 모든 소리가 울린다.	Every sound in the cave.
내가 정의하는 소리공학은 우리 주변의 모든 소리를 분석하고 규명해서 실생활에 도움이 될 수 있도록 만드는 기술이다.	I define is a technique that analyzes and identifies all the sounds around us to help us in real life.
이렇게 소리의 종류가 많다 보니 소리를 연구하는 학문 분야도 다양하다.	There may be more sounds that are not mentioned here, or you may know more kinds of sounds than I am.
캘리포니아대학교의 한 교수가 초음파를 이용해 계란을 삶았다는 실험논문을 발표했는데, 이에 근거하여 소리를 모아 라면을 끓일 수 있는지 소리공학연구소에서 실험해달라고 제안한 것이다.	A professor at the University of California published an experimental paper on ultrasound boiling eggs and suggested that the sound engineering researchers should experiment with boiling noodles to collect the sounds.
즉시 일산에 있는 공연장 음향 전문회사가 섭외되었고, 그곳에서 소리에너지 변환실험을 하게 되었다.	Immediately there was a sound company specializing in a concert hall in Ilsan, where he experimented with sound energy conversion.
길을 걷다보면 어떤 소리가 유달리 귀에 거슬리거나 또는 아련한 추억을 떠올리는 유행가로 들리는 경우가 종종 있다.	Walking the streets often sounds like a buzzing sound or a faded memory.
개가 정말 웃는다면 그 소리를 녹음해서 다른 개들에게 들려주자는 이야기였다.	If the dog really laughed, it would have been possible to record the sound and tell the other dogs.
방송사의 제안대로 개들이 웃는 소리를 녹음해서 잡음을 제거한 후 음반에 녹음하여 다른 개들에게 들려주어 웃음소리의 효능을 실험해보기로 했다.	As suggested by the broadcaster, the dogs recorded a laughing sound, removed the noise, recorded it on the music, and told the other dogs to try out the efficacy of laughter.

위의 예를 보면, ‘대부분의 사람들은 엄마 배 속에서 소리를 처음 듣는다.’는 ‘Most

people hear the sound in their belly for the first time.’으로 번역되었는데, 원문의 의미와는 다른 번역임을 알 수 있다. 그런데 원문의 ‘엄마 배’에 ‘~의’라는 관형격 조사를 붙여서 다시 번역해보면, ‘Most people hear the sound in their mother’s belly for the first time’으로, 원문과 동일한 의미의 정상적인 번역이 된다. 또한 ‘이렇게 소리의 종류가 많다 보니 소리를 연구하는 학문 분야도 다양하다’는 ‘There may be more sounds that are not mentioned here, or you may know more kinds of sounds than I am’으로 원문과는 상이한 번역이 되었다. 그러나 원문의 ‘학문 분야도’에 다중 유표인 ‘도’를 주격 조사인 ‘~가’로 바꾸어 번역하면, ‘There are so many kinds of sounds, so there are a variety of disciplines that study sound.’으로 정상적인 번역이 된다. 위의 결과들은 격표지가 생략된 문장이나 다중 유표의 문장을 기계번역 할 경우, 원문과는 전혀 다른 오역이 발생할 수 있음을 보여준다.

4.1.3. 속담 텍스트

속담은 네이버 지식백과에 따르면, ‘개팔자’, ‘억지춘향’과 같이 단어로 이루어진 경우와, ‘고생 끝에 낙이 있다’와 같은 단문이나, ‘열길 물속은 알아도 한 치 사람 속은 모른다’와 같은 중문 또는 복문으로 된 경우, 그리고 ‘귀에 걸면 귀거리, 코에 걸면 코걸이와 같이 나열형이 있다. 본 절에서는 속담을 기계번역 할 때 발생하는 통사오류를 살펴보기 위해, 인터넷 다음 블로그의 <재미있는 속담 1000개>에서 단문이나 복문 형태의 속담을 가나다 순서로 각 10여개씩 총 100 문장을 수집하였다. 이렇게 수집된 텍스트에 각각 단문과 복문으로 표시하여 구글번역기에 입력하고, 영어 번역 결과 발생한 오류를 어휘오류와 통사오류로 나누어 빈도를 조사했다. 그리고 격조사관련 통사오류의 원인을 격조사 표지의 유표, 무표, 그리고 다중 유표로 나누어 빈도를 비교하였다.

다음의 표 5는 속담 텍스트의 한영 기계번역 시 어휘오류와 격조사관련 통사오류의 빈도를 비교한 것이다.

표 5. 속담 텍스트에서의 어휘오류와 격조사관련 통사오류 빈도

	어휘오류	통사오류(격조사)		
		유표	무표	다중 유표
단문(14)	6	0	3	1
복문(86)	18	5	15	16
계	24	5	18	17
합계	24	40		

위의 표를 보면, 속담 텍스트의 단문과 복문 비율은 14개와 86개로 복문이 훨씬 많고, 번역의 결과 나타난 어휘오류와 통사오류는 각각 24개와 40개로 어휘오류보다 통사오류의 빈도가 상당히 높다는 것을 알 수 있다. 통사오류의 빈도를 좀 더 자세히 살펴보면, 단문과 복문의 경우에 각각 4개와 36개의 오류로, 복문의 경우에서 통사오류가 훨씬 많다. 또한 요인별 통사오류의 빈도에서, 격조사 표지가 있는 경우에 5개, 생략된 무표의 경우에 18개, 그리고 다중 유표의 경우에 17개의 오류가 나타났는데, 특별히 격조사의 생략이나 다중 유표의 사용으로 인한 오류가 많다는 것을 알 수 있다. 이 결과는, 속담 텍스트에 격조사의 생략이나 다중 유표의 활용 구문이 많고, 이로 인해 통사오류가 발생한다는 것을 보여주는 것이다.

다음의 표 6은 속담 텍스트의 기계번역에서의 격조사관련 번역 오류 예시이다.

표 6. 속담 텍스트에서의 격조사관련 오류 예시

한국어(입력)	영어(결과)
가난 구제는 나라도 못 한다.	Poverty relief is not even a nation.
가는 말이 고와야 오는 말도 곱다.	The word goes well, and the word comes.
나 부를 노래를 사돈 집에서 부른다.	Sing a song to the family at home.
나는 새도 깃을 쳐야 날아간다.	I fly with a bird feathers.
남이 떡 먹는데 팔고물 떨어지는 걱정한다.	I'm worried about falling asleep when I eat rice cake.
대문턱 높은 집에 정강이 높은 머느리 들어온다.	A high-jaw high house with a shin creeps in.
될성부른 나무는 떡잎부터 알아본다.	Trees that call to be called are known from cotyledons.
늦게 배운 도둑이 날 새는 줄 모른다.	I do not think the late learned thief leaks
범도 제 소리 하면 오고 사람도 제 말하면 온다.	I will come to you if you come to me, and the person will speak to me
새 잡아 잔치할 것을 소 잡아 잔치한다.	Grab a bird and have a feast to celebrate.

위의 예를 보면, '가난 구제는 나라도 못 한다'는 'Poverty relief is not even a nation'으로, '늦게 배운 도둑이 날 새는 줄 모른다'는 'I do not think the late learned thief leaks'으로 번역되었는데, 원문과는 다른 오역임을 알 수 있다. 이러한 실례를 통해, 속담의 기계번역 시 생략된 격조사나 다중 유표로 인해 다소 황당한 번역의 결과가 초래된다는 것을 알 수 있는데, 이에 대해서는 4.2에서 좀 더 자세히 논의하도록 한다.

4.1.4. 텍스트 유형별 통사오류 비교

지금까지 세 가지 유형의 한국어 텍스트를 구글번역기에서 영어로 번역하고, 그 결과 받

생한 통사오류를 텍스트 유형별로 격조사 표지의 유표, 무표, 다중 유표로 나누어 빈도를 조사하고, 오류의 예를 구체적으로 살펴보았다. 본 절에서는 텍스트 기계번역 시 나타난 격조사 관련 통사오류를 텍스트 유형별로 비교하기로 한다.

다음의 표 7은 텍스트 유형별 격조사관련 통사오류의 빈도를 비교한 것이다.

표 7. 텍스트 유형별 격조사관련 통사오류 빈도 비교

	구어 텍스트			문어 텍스트			속담 텍스트		
	유표	무표	다중유표	유표	무표	다중유표	유표	무표	다중유표
단문(17)	2	7	2	1	1	0	0	3	1
복문(71)	0	19	3	4	7	2	5	15	16
계	2	26	5	5	8	2	5	18	17
합계(88)	33 개			15 개			40 개		

위의 표에서, 먼저 텍스트 유형별로 단문과 복문의 통사오류 빈도를 보면, 단문(17) 보다는 복문(71)에서 오류가 훨씬 많다는 것을 알 수 있는데, 이는 단문이 아닌 복문을 영어로 번역할 경우 오역의 가능성이 훨씬 높다는 점을 말해준다. 또한, 텍스트 유형별로 격조사관련 통사오류의 빈도를 보면, 속담(40개) > 구어(33개) > 문어(15개)로 속담 텍스트에서 오류가 가장 많다는 것을 알 수 있다. 이 결과는 비교적 어법이 잘 정돈된 문어 텍스트보다는 속담의 번역 시 오류가 더 많다는 것을 명시적으로 보여주는 것이다. 마지막으로, 텍스트 유형별로 통사오류의 원인을 격조사 표지의 유표, 무표, 다중 유표로 나누어 조사한 결과, 무표(52) > 다중 유표(24) > 유표(12)로 나타났는데, 격조사의 생략으로 인한 번역 오류가 가장 많다는 것을 알 수 있다.

이러한 결과를 통해 복문의 속담을, 특별히 격조사 표지가 생략된 복문의 속담을 기계번역기로 번역할 때 오역이 가장 많이 발생할 수 있다는 것을 알 수 있는데, 앞으로의 기계번역기는 원문의 생략된 격 구문에 대응할 영어 구문을 어떻게 효과적으로 처리할 것인지에 중점을 두고 개선되어야 할 것이다.

4.2. 속담 텍스트에서의 격조사 복원 효과

본 절에서는 텍스트 유형별 통사오류의 결과를 바탕으로, 속담에서 오류가 가장 많이 나타난 점에 집중하여, 오역으로 나타난 속담에 격조사 표지를 복원한 후 다시 번역하고 그 결과를 살펴보았다.

다음의 표 8은 오역으로 나타난 속담 텍스트에 격조사 표지를 붙이고 다시 번역한 결과이다.

표 8. 속담 텍스트에 격조사 복원 후 변화

	유표		무표		다중 유표		성공률
	오류	정상	오류	정상	오류	정상	
단문(21개)	0	0	3	3	1	1	
복문(79개)	5	1	15	9	16	10	
계	5	1	18	12	17	11	60%

위의 표에서 보듯이, 생략된 격조사를 복원하거나 다중 유표의 조사를 적절하게 바꾸어 다시 번역한 결과, 40개의 오류 속담 중에서 24개의 속담이 정상적으로 번역되었다. 이런 결과는 속담에 격조사가 생략된 관용적 표현이 많기 때문에, 기계번역과정에서 원문의 입력 시 격조사를 표기하는 인위적 개입이 오류를 줄일 대안이 될 수 있음을 보여준다.

다음의 표 9는 번역 오류가 발생한 속담에 격조사 표지를 복원하거나 적절한 격조사로 바꾸어 다시 번역한 예시이다.

표 9. 속담 텍스트에 격조사 표지 복원 후 번역의 예시

한국어 (입력)	영어(결과)
가난 구제는 나라도 못 한다. → 가난 구제는 나라가 못 한다.	Poverty relief can not be achieved either. → The country can not afford poverty relief.
가는 말이 고와야 오는 말도 곱다. → 가는 말이 고와야 오는 말이 곱다.	The word goes well, and the word comes. → If the word you say is good, then the word coming back at you is good.
늦게 배운 도둑이 날 새는 줄 모른다. → 늦게 배운 도둑이 날이 새는 줄을 모른다.	I do not think the late learned thief leaks. → A late learned thief does not know that the day is leaking.
대문턱 높은 집에 정강이 높은 며느리 들어온다. → 대문턱이 높은 집에 정강이가 높은 며느리가 들어온다.	A high-jaw high house with a shin creeps in. → A high-winged daughter-in-law comes into the house with a high threshold.
동아 속 썩는 것은 밭 입자도 모른다. → 동아 속이 썩는 것을 밭 입자가 모른다.	The owner of the field does not know that rotting in the East. → The owner of the field does not know that Dong-a is rotten.
될성부른 나무는 떡잎부터 알아본다. → 될성부른 나무를 떡잎부터 알아본다.	The tree that is called is known from the cotyledons. → I learn the tree from the cotyledon.
세 살 버릇 여든까지 간다. → 세 살 버릇이 여든까지 간다.	Three years old. → A three-year old habit goes up to eighty.
소문난 잔치에 먹을 것 없다.	Do not eat on a celebratory feast.

→ 소문난 잔치에 먹을 것이 없다.	→ There is nothing to eat in a celebratory feast.
소 잃고 외양간 고친다. → 소를 잃고 외양간을 고친다.	Losing a cow and fixing a cow. → Lose cattle and fix the stall.
열 길 물속은 알아도 한 길 사람 속은 모른다. → 열 길 물속을 알아도 한 길 사람 속을 모른다.	The way of the road I know the water I do not know the road person. → I know the way in the water, I do not know the person inside.

위의 예시에서 속담 원문에 생략된 격조사 표지를 복원시키거나 다중 유표의 조사를 적절한 형태의 격조사로 바꾸어 다시 번역할 때, 영어 표현이 다소 어색하지만 통사오류 없이 속담이 주는 의미를 전달하고 있음을 알 수 있다. 이러한 결과는 한국어 속담을 기계번역기에서 영어로 번역할 때, 상당수의 속담이 표면상으로 나타나지 있지 않는 격조사로 인해 원문의 의미와는 거리가 먼 오역이 될 수 있음을 인지하고 인위적 개입이 필요함을 말해준다.

4.3. 결과 논의

지금까지 텍스트 유형별 통사오류의 양상과 속담에서의 격조사 복원 후 변화의 실태를 살펴본 결과, 다음과 같은 사실을 알 수 있다. 첫째, 세 가지 유형의 구어, 문어, 속담 텍스트를 구글번역기에 입력하고 번역을 시도한 결과, 모든 유형에서 어휘오류보다는 통사오류가 더 많았다. 이 결과는 두 언어간의 통사적 불일치로 인한 오역이 더 많다는 것을 객관적으로 확인시켜준 것으로, 한국어 구문에 대한 기계번역기의 통사 정보처리 기능이 아직은 완벽하지 않다는 것을 보여주었다는 점에서 의의가 있다.

둘째, 텍스트 유형별로 단문과 복문으로 나누어 통사오류를 조사한 결과, 모든 유형에서 복문에 오류가 더 많았다. 이 결과는 단문보다는 복문의 경우에 격 정보가 더 많이 생략되거나 다중 유표가 많이 활용되기 때문에, 이로 인해 복문에서 오류가 더 많은 것으로 설명할 수 있다.

셋째, 텍스트 유형별로 통사오류의 빈도를 조사한 결과, 속담 텍스트에서 통사오류가 가장 많았다. 이 결과는 속담에 격조사가 생략된 관용적 표현이 많기 때문에, 규범적인 어법을 비교적 잘 지키는 문어보다 속담에서 통사오류가 더 많은 것으로 해석할 수 있다.

넷째, 텍스트 유형별로 통사오류의 원인을 비교한 결과, 모든 유형의 텍스트에서 격조사 표지의 유표에 의한 오류보다 격조사 표지의 생략이나 다중 유표로 인한 오류가 훨씬 많았다. 이 결과는 격조사 자체가 아예 생략되거나 격조사 표지가 뒤섞이는 경우와 관련하여 한국어와 영어의 격 체계 차이로 인해 번역 오류가 생긴다(배지연, 2010; 배지연·김수연, 2010)는 견해로 설명할 수 있다. 따라서 격조사가 생략되었거나 중의적 의미를 유발하는 다중 유표의

조사가 사용된 한국어 텍스트를 기계번역기로 번역하려면, 텍스트 입력 시 격조사 생략 구문의 처리에 신중을 기해야한다.

마지막으로, 기계번역 시 오역으로 나타난 속담 텍스트에 격조사 표지를 복원하거나 적절한 격조사로 바꾼 후 다시 번역한 결과, 다수의 속담이 정상적으로 번역되었다. 이러한 결과는, 격조사의 명시적인 표기가 기계번역의 정확성을 높여주는 여러 방법 중 한 가지 대안이 될 수 있음을 말해준다. 이 점은 임홍빈(2007)의, “한국어에서 조사를 가지지 않는 무조사 명사구는 제시어와 엄격히 구별되지 않는 것으로, 어휘부에 주어진 논항 정보에 의해서는 분석되지 않는 잉여 성분의 성격을 띤다. 이러한 무조사 명사구는 정도상의 차이를 가지고, 강한 제시어와 약한 제시어 사이에 걸치게 되는데, 속담과 같은 관용 표현에서는 그 성격이 추상화되어 구체성을 상실하기도 한다.”라는 무조사 명사구에 대한 설명에서도 명확해진다. 또한 격조사를 갖지 않는 명사구는 문맥에 따라 다양한 의미역을 가질 수 있다는 이호승(2006)의 설명을 이에 적용해보면, 격조사가 생략되거나 다중 유표의 조사가 사용된 속담의 원문은 통계적 번역 방식의 기계번역기에서 여러 의미역으로 출력될 수 있기 때문에, 오역을 피하기 위해서는 격조사 표지의 유무를 먼저 확인하는 것이 최선이라고 할 수 있다. 그 이후 격조사가 없는 경우에는 그것을 복원하여 표시해 주고, 다중 유표의 경우에는 적절한 격조사를 지정해줄 때 좀 더 완성도 있는 번역을 얻을 수 있다고 본다.

5. 결론

본 연구는 한국어 텍스트를 영어로 기계번역 할 때 격조사가 번역의 정확성에 미치는 영향을 분석한 것이다. 연구에 앞서서 먼저 기계번역의 선두 주자인 구글번역기의 번역시스템에 대해 알아본 후, 세 가지 유형의 구어, 문어, 속담 텍스트를 각각 단문과 복문으로 분류하여 구글번역기에서 번역하고, 그 결과 발생한 오류를 텍스트 유형별로 어휘오류와 통사오류로 나누어 빈도를 조사하고, 통사오류의 원인을 격조사의 유무와 다중 유표의 개입으로 분류하여 텍스트 유형별로 빈도를 비교하였다. 또한 속담에서 통사오류의 빈도가 좀 더 높은 것으로 나타났으므로, 오역된 원문에 격조사 표지를 복원한 후 다시 기계번역을 시도하고 그 결과를 조사하여, 과연 격조사 표지의 명시적인 개입이 번역 오류를 줄일 수 있는지 살펴보았다.

그 결과 텍스트 유형별 번역 오류 조사에서는, 모든 유형의 텍스트에서 어휘오류보다 통사오류가 더 많았고, 단문보다는 복문에서 오류가 훨씬 더 많았다. 또한 텍스트 유형별 통사오류의 빈도는 속담 > 구어 > 문어의 순서로 속담에서 통사오류가 좀 더 많았다. 그리고 모든 텍스트에서 격조사가 있는 경우보다 격조사 표지가 생략되었거나 ‘-은, 만, 도’와 같이 전체나 내포된 의미를 담은 다중 유표가 사용된 경우에 통사오류가 상당히 더 많았다. 마지막

으로, 번역 오류가 좀 더 많이 발생한 속담에 격조사 표지를 복원한 후 다시 번역했을 때 다수의 속담이 정상적으로 번역되었다. 이 결과는 영어가 어순으로 문법적 역할(grammatical roles)을 결정하는 데에 비하여, 한국어는 격조사에 의해 의미역이 표시될 뿐 아니라 종종 격조사가 생략되기도 하기 때문에, 한국어 원문의 격조사 유무에 따라 기계번역의 결과가 다르게 나타날 수 있음을 보여준 것이다. 따라서 현재로선 기계번역기의 번역 방식이 번역 대상 언어의 어법을 기반으로 한 것이 아니기에, 한영 기계번역 시 번역의 정확성을 기하려면 이를 반영하여 격조사 표지를 명시적으로 표시할 것을 대안으로 제시한다.

이번 연구에서는 한국어 데이터를 영어로 기계번역 할 때 격조사가 번역에 미치는 영향을 살펴보았는데, 원문의 입력을 개인이 통제할 수 있는 경우에는, 격조사가 나타나 있지 않은 원문에 격조사를 복원하거나 다중 유표의 조사를 적절하게 바꾸어 번역한다면, 좀 더 정확한 결과로 번역의 완성도를 높일 수 있음을 확인하였다. 따라서 본 연구는 기존의 생성된 텍스트가 아닌 새로운 텍스트를 입력하여 기계번역기로 번역하고자 할 때, 문장 단위로 입력하면서 격조사 표지를 모두 표시하는 노력이 필요하고, 나아가 데이터 입력자와 기계번역기 간의 상호작용(interaction)이 전제되어야함을 시사한다.

최근에 학문적으로나 상업적 목적으로 기계번역기가 많이 활용되고 있는 상황에서, 본고는 기계번역기의 번역이 아직은 상업성을 담보할 만큼 완벽한 단계가 아니라는 것을 보여준 실험적 성격의 연구이다. 앞으로의 연구는 좀 더 많은 한국어 텍스트를 대상으로 구글번역기 외에 다른 기계번역기를 활용하여 번역의 결과를 비교해보고, 나아가 기계번역기의 번역 오류를 줄일 수 있는 효율적인 방안을 제시하는 방향으로 이어질 필요가 있다.

참고문헌

- 김원경. (2009). *한국어의 격*. 서울: 박문사.
- 김원경. (2011). 격조사의 정보와 기계번역: 한-영 번역중심으로. *한국어의미학*, 35, 23-43.
- 남기심, 고영근. (2010). *표준국어문법론*. 서울: 탑출판사.
- 박영란. (2017). 기계번역에서의 영어 속담 분석. *한국통번역교육회 학술대회 자료집*, 58-61.
- 박홍원, 심재석, 이수진, 석영민, 오승훈. (2001). 한영 번역 시스템에서의 불특정 조사를 포함한 용언구 처리. *한국정보과학회 언어공학연구회 학술발표 논문집*, 523-529.
- 배명진, 김명숙. (2013). *소리로 읽는 세상*. 김영사.
- 배지연. (2011). *한영 기계번역의 격 실현*. 세종대학교 대학원 박사학위논문.
- 배지연, 김수연. (2010). 한영 기계번역에 있어서 격 실현 현상 고찰. *언어과학*, 17(4), 99-120.
- 엄정호. (2011). 격의 개념과 한국어의 조사. *국어학*, 62, 199-223.
- 이병희. (1994). 한/영 기계번역 시스템을 위한 시제 도우미의 설계와 구현. *인터넷정보학회*

논문지, 2(4), 55-67.

- 이종근. (2018). 한국어 격표지의 본질과 초점. *언어연구*, 33(4), 733-755.
- 이하규, 김영택. (1994). 기계번역을 위한 한국어 속어의 표현 및 인식. *정보과학회논문지*, 21(1), 139-149.
- 이호승. (2006). 격조사 없는 명사구의 격 문제에 대하여. *어문학*, 93, 139-159.
- 이호승. (2012). 국어 격 범주의 성격과 격 체계의 설정. *어문학*, 117, 39-67.
- 임흥빈. (2007). 한국어 무조사 명사구의 통사와 의미. *국어학*, 49, 69-106.
- 조준모, 이상조. (1994). 한·영 기계 번역을 위한 부사의 위치 및 순서제약 해결의 방안 및 구현. *한국정보과학회 언어공학연구회 학술발표 논문집*, 163-167.
- 최재혁, 김권양, 박상규, 이상조. (1989). 한영 기계번역을 위한 조사 사전구성에 관한 연구. *한국정보과학회 언어공학연구회 학술발표 논문집*, 219-224.
- 홍기선. (2017). 한국어의 조사 '가'와 '를': 격과 의미. *언어*, 42(4), 961-988.

김경숙

06978 서울시 동작구 상도로 369
숭실대학교 베어드교양대학 부교수
전화: (02)828-7484
이메일: chris812@ssu.ac.kr

Received on July 13, 2018

Revised version received on September 26, 2018

Accepted on September 30, 2018