

영-조기계번역에서 생략이 일어난 영어비교표현의 해석과 조선어생성에 대한 연구

김광혁, 김현철

경애하는 최고령도자 김정은동지께서는 다음과 같이 말씀하시였다.

《과학기술과 경제의 일체화를 다그치고 나라의 경제를 현대화, 정보화하는데서 과학 기술부문이 주도적인 역할을 하도록 하여야 합니다.》(《조선로동당 제7차대회에서 한 중앙 위원회사업총화보고》 단행본 41페이지)

영문과학기술자료에 대한 번역수요가 끊임없이 늘어나고있고 번역의 고속화를 위한 효과적수단인 영-조기계번역프로그램제품들에 대한 기대와 관심이 계속 높아지고있는 현실은 기계번역결과의 정확성에 대한 더 높은 요구성을 제기하고있다. 기계번역결과를 인간의 번역결과에로 보다 더 가깝게 접근시켜나가기 위해서는 조선어와 언어적차이가 심한 개별적인 영어표현들의 번역방법에 구체적인 주의를 돌릴 필요가 있다.

논문에서는 영-조기계번역에서 출력언어인 조선어와 문장구조적으로 크게 차이나는 생략이 일어난 영어비교표현에 대한 해석 및 조선어생성방법을 제안하였다.

1. 영어비교표현에 대한 선행연구와 문제설정

선행연구[1, 2]들에서는 영어비교표현의 언어적수단에 대한 분석이 광범히 진행되였다. 그러한 연구들에서는 문리문법적측면만이 아니라 형상의미적측면에서까지 비교의 의미를 나타내는 영어표현들을 포괄적으로 논의하였다.

전치사 of나 for를 리용하여 비교의미를 나타내는 표현들(Of the (two) boys, John behaves more politely. He is stupid for a child of his age)과 언어정황과 관습에 의하여 표현되는 비교표현(No two tools are more alike. No night is darker.)들이 그 대표적인 실례로 된다.

논문에서는 영-조기계번역의 견지에서 문제점이 제기되는 문법적생략이 일어난 영어비교표현에 대해서만 논의하기로 한다.

영어비교표현[1]은 크게 동등비교와 우세비교, 열등비교로 나눌수 있다. 생략은 이 모든 비교류형들에서 다 나타날수 있다.

실례 1

I get up as early as I should(get up).

실례 2

She'll enjoy it more than(she enjoyed it) last year.

실례 3

He's not as naughty as he was(naughty).

실례 4

Kim is as stupid as Ri(is).

실례 5

Kim behaves as politely as Ri(does).

번역의 견지에서 보면 언어학적 측면에서 생략현상으로 취급한 이 모든 표현들을 다 생략현상으로 처리할 필요는 없다.

실례로 실례 2, 4, 5는 레외적인 생략처리를 하지 않고서도 원만히 번역할수 있는 문장들이다.

현재 국내에서 실용화되어 리용되고있는 몇가지 기계번역프로그램들을 리용하여 위의 실례문장들을 번역해본 결과는 다음과 같다.

① 기계번역프로그램 1

내가 우에 1년초에 얻는것은 얻어야 한다.

그 녀자는 지난해이상 그것을 즐길것이다.

그는 인것만큼 장난을 좋아하지 못한다.

김은 리로써 하찮다.

김은 레질이 바르게 리만큼 작용한다.

② 기계번역프로그램 2

내가 하여야 하는만큼 일찍 나는 기상한다.

그 녀자는 지난해에보다 더 그것을 즐길것이다.

그는 있었던만큼 버릇이 없다.

김과 마찬가지로 리는 어리석다.

김은 리와 마찬가지로 공손하게 처신한다.

③ 기계번역프로그램 3

나는 내가 하여야 할만큼 일찌기 일어난다.

그 녀자는 지난해보다 그것을 더 많이 즐길것이다.

그는 그가 있던것만큼 간악하지 않다.

김은 리만큼 어리석다.

김은 리만큼 레질바르게 처신한다.

기계번역결과를 놓고 알수 있는것처럼 동등비교와 우세 및 렬등비교의 의미를 나타내는 영어표현들에 대한 번역처리기능은 모든 기계번역프로그램제품들이 직역수준에서 갖추고있다. 그러나 그러한 직역처리를 가지고서는 생략이 일어난 문장들에 대하여 정확하게 번역할수 없다. 실례 1과 3은 레외적인 생략처리를 통해서만 원만히 번역할수 있는 문장류형에 속한다.

론문에서는 이러한 생략현상이 일어난 영어비교표현들을 원만히 번역하기 위한 한가지 방법에 대하여 논의하였다.

2. 생략이 일어난 영어비교표현에 대한 해석과 조선어생성

생략현상이 일어난 영어비교표현들에 대한 인간의 번역에서는 기계번역으로 얻어내기 힘들 정도로 의역되는것들이 적지 않다.

실례 6 Man's life is not yet as comfortable and convenient as it might be. There are still more of nature to exploit.

실례문장에 대한 인간의 번역문은 《인간의 생활은 아직 이상적일만치 편안하고 편리하지 못하다. 개척해야 할 자연의 미지의 세계가 많고도 많다.》이다.

실례 6에 대한 인간의 번역문과 같은 번역결과를 기계번역으로 얻어내는것은 현재 기계번역프로그램들이 도달한 수준을 놓고볼 때 물론 어려운 일이 아니다. 문제는 실례 6만이 아니라 생략현상이 일어난 영어비교표현들에 대한 리해할수 있는 기계번역문을 얻어내기 위한 일반적인 번역방법을 찾아내는것이다.

생략현상이 일어난 영어비교표현을 포함하고있는 영어문장들의 공통된 특징은 생략된 요소에 대응하는 원천요소가 해당 문장속에 존재한다는것이다.

론문에서는 이러한 언어적특징으로부터 추출관계를 해석할수 있는 영어문장해석문법을 리용하여 영어비교표현에서 일어난 생략현상을 인식한 다음 조선어번역문생성단계에서 찾아낸 원천요소를 추출요소와 련결시켜 언어적표현에 맞게 생성하도록 하는 번역방법에 대하여 논의하였다.

① 생략현상이 일어난 영어비교표현의 문장론적해석

영어비교표현들이 가지고있는 다른 하나의 특징은 비교표현을 나타내는 언어적수단이 형용사나 부사의 비교급, more, less, as, than과 같은 몇가지 어휘들에 한정되어있다는것이다. 그리고 생략현상이 일어난 영어비교표현들에서 생략된 요소는 반드시 동사구와 련관되어있다. 이로부터 생략현상이 일어난 영어비교표현의 문장론적해석의 첫단계는 생략요소를 포함하고있는 동사구 및 문장의 추출구조를 해석하는 단계로서 추출구조를 해석하는 문법해석규칙들에 의하여 수행된다. 다음단계는 이 추출구조를 그 원천요소를 포함하고있는 구조와 련결시키는 단계로서 개별적인 어휘(실례로 as나 than)별로 작성된 문형해석규칙에 의하여 구체적인 어휘표현들이 해석된다. 이 어휘별문형해석규칙은 확장된 문맥자유문법형식으로 서술할수 있다.

그 BNF형식은 다음과 같다.

```

RULE::=LHS→CAT[FILL-GAP]
LHS::=LEFT KEY RIGHT
LEFT::=[LEXICALE|VAR]*
RIGHT::=[LEXICALE|VAR]*
LEXICALE::=lexicalE
VAR::=var[{CONSTRAINT-LIST}]
CONSTRAINT-LIST::=[constraint]+
constraint::=+constraint_string-constraint_string
CAT::=category
FILL-GAP::={ε(GAP)=FILLER}
GAP, FILLER∈LHS
lexicalE∈E-Dic, category, var∈CATS, constraint-string∈CONSTS

```

여기서 E-Dic, CATS, CONSTS는 각각 영어단어사전, 구법주목록, 속성제한목록을 의미한다.

실례 3의 비교표현을 해석하는 문형해석규칙을 실례들면 다음과 같다.

as AJP as SENT {+ADJ_EXTRACTION} → AJP { ϵ (SENT)=AJP}

실례 1, 3에 대한 문장론적해석구조를 그림 1과 2에 보여주었다.

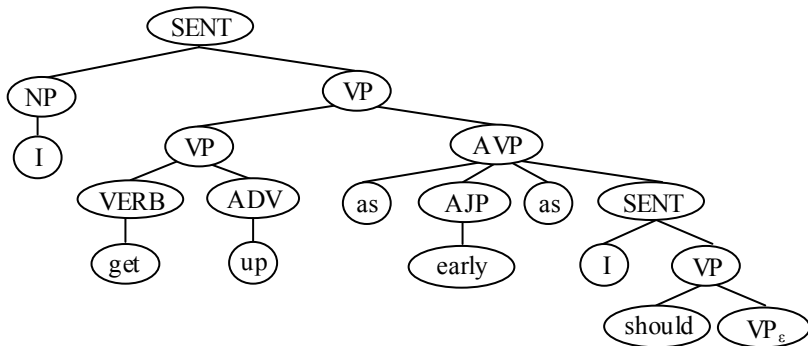


그림 1. 실례 1에 대한 문장론적해석구조

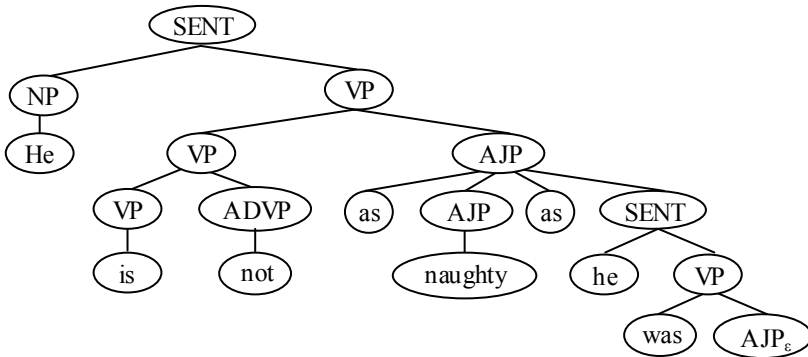


그림 2. 실례 3에 대한 문장론적해석구조

② 생략현상이 일어난 영어비교표현의 조선어생성

생략현상이 일어난 영어비교표현에 대한 영-조기계번역에서 가장 중요한것은 동등 비교와 우세 및 열등비교의 의미를 나타내는 개별적표현들의 의미를 적절하게 나타낼수 있는 조선어번역표현을 리용하는것이다.

우에서 정의한 영어어휘별문형해석규칙을 다음과 같이 확장한다.

RULE::=LHS→CAT[FILL-GAP]:RHS

RHS::=[KOR_{tok}]⁺

KOR_{tok}::=[LEXICAL_K|VAR|VARTO_K]

LEXICAL_K::=[lexical_K]+[TO_K]^{*}

lexical_K∈K-Dic, TO_K∈TO-Dic

여기서 K-Dic, TO-Dic는 조선어단어사전과 조선어토사전을 의미한다.

실례 3의 비교표현을 해석하는 문형규칙은 다음과 같다.

as AJP as SENT {+ADJ_EXTRACTION} → AJP { ϵ (SENT)=AJP}:SENT하는것만큼 AJP하다.

생략현상이 일어난 영어비교표현을 조선어로 번역하는데서 중요한것은 다음으로 조선어와 영어에서 서로 다르게 나타나는 생략과 관련한 언어적특징을 적절하게 반영하는 것이다.

영어비교표현에서 일어나는 생략의 주요한 특징은 그것이 선출현-후생략구조라는 것이다. 다시말하여 생략된 요소에 대응하는 원천요소는 언제나 앞에 존재한다. 이에 대비하여 볼 때 생략현상이 나타난 조선어문장에서는 선행생략-후출현구조가 지배적이다. 즉 생략된 요소에 대응하는 원천요소는 언제나 뒤에 존재한다. 이것은 조선어가 어순상 중심어가 뒤에 놓이는 언어라는 사정과 관련된다.

이와 같이 두 언어에서 생략과 관련한 특성이 서로 다른것으로 하여 생략현상이 일어난 영어표현들에 대한 인간의 번역에서는 원문의 구조보다는 의미를 그대로 보존하는 의역이 보편적인것으로 된다.

논문에서는 기계번역의 견지에서 이해할수 있는 번역문을 얻기 위한 한가지 일반적 방법으로 생략요소와 원천요소를 모두 포함하고있는 영어문장구조에 대한 조선어번역문 생성에서 다음과 같은 생성수법을 적용하기로 한다.

형용사구가 생략된 경우 생략요소를 원천요소로 모두 복원한 상태에서 조선어번역문에서 먼저 출현하는 요소는 그대로 두고 후에 출현하는 반복요소를 《그리하다.》라는 조선어단어로 대치시키도록 한다.

동사구가 생략된 경우 원천요소의 조선어동사를 그대로 복원시켜 모두 출력하도록 한다.

명사구가 생략된 경우 관계절에 대한 조선어생성에서와 마찬가지로 생략요소를 생성하지 않도록 한다.

3. 실현 및 평가

논문에서 제안한 방법을 영-조기계번역프로그램 《룡남산》2.0에 도입하여 실례 1과 3, 6에 대하여 번역한 결과는 다음과 같다.

① 실례 1의 기계번역문

나는 내가 일어나야 하는만큼 일찍 일어난다.

② 실례 3의 기계번역문

그는 그가 장난을 좋아했던만큼 그렇지 않다.

③ 실례 6의 기계번역문

사람의 생활은 아직 그것이 편안하고 편리할수도 있는것만큼 그렇지 않다.

번역결과를 통하여 논문에서 제안한 방법이 생략현상이 일어난 영어표현에 대한 기계번역문의 이해성정도를 개선하는데 효과가 있다는것을 알수 있다.

맺 는 말

영어비교표현의 언어학적특성을 분석한데 기초하여 이해할수 있는 영-조기계번역결과를 얻어내기 위한 영어문장해석 및 조선어생성방법을 제안하고 영-조기계번역프로그램 《룡남산》2.0에 도입하고 실험을 진행하여 제안한 방법의 효과성을 확인하였다.

참 고 문 헌

- [1] 김동일 등; 영어문법사전, 외국문도서출판사, 280~293, 주체102(2013).
- [2] 박철호 등; 영어, 외국문도서출판사, 130~140, 주체106(2017).

주체108(2019)년 2월 5일 원고접수

A Study of the Parsing and Korean Generation of English Comparative Expression with Ellipsis in English-Korean Machine Translation

Kim Kwang Hyok, Kim Hyon Chol

In this paper, we proposed a method of improving the translation of English comparative expression with ellipsis in English-Korean machine translation.

Key words: English-Korean machine translation, English comparative expression, ellipsis