

영-조기계번역에서 종속형생략구조번역의 한가지 방법

김광혁, 김현철

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《시대의 발전추세에 맞게 나라의 과학기술을 높은 수준에 올려세워야 우리의 정치군사적위력을 더욱 강화할수 있고 경제강국도 건설할수 있으며 인민들에게 세상에 부러운것 없는 유족하고 문명한 생활을 마련해줄수 있습니다.》《김정일선집》증보판 제22권 19~20페이지)

일반적으로 생략이라고 할 때 그것은 문장속에서 하나 또는 그 이상의 구성요소가 빠지는것을 말한다. 생략현상은 어느 언어를 막론하고 입말과 글말이라는 두가지 형태의 언어생활모두에서 나타나는 일반적인 언어적현상의 하나이다.

영어에서 나타나는 생략현상에 대한 언어학적연구는 여러가지 측면에서 비교적 많이 진행되었으나 영-조기계번역에서 생략구조들을 어떻게 번역할것인가 하는데 대한 체계화된 연구는 아직 진행되지 못하였다. 그 리유는 기계번역체계의 개발에서 다양한 형태의 생략현상들을 사람만큼 번역할데 대한 요구가 제기되지 않은데도 있지만 주요하게는 기계번역체계에서 생략현상에 대한 명백한 판정기준과 번역방법을 제기하지 못한데 있다.

본문에서는 현재 실용화되고있는 영-조기계번역체계들의 번역률이 일정한 수준에 올라선 조건에서 기계번역체계의 번역정확성과 유순성을 더욱 높이기 위하여 영어에서 나타나는 생략현상에 대한 연구를 진행하고 그 번역을 개선하기 위한 한가지 방법에 대하여 제기하였다.

1. 영어에서 나타나는 생략현상에 대한 분류

영어에서 나타나는 생략현상에 대한 언어학적연구들을 분석해보면 생략현상은 그 분류기준에 따라 단순생략과 복합생략, 앞부분생략과 뒤부분생략, 관용어적생략과 정황적생략 등으로 구분할수 있다.[1, 5-8]

그러나 언어학적인 측면만을 고려하여 진행된 선행연구들에서 제기한 분류기준들은 기계번역의 특성에는 잘 맞지 않는것들이 적지 않으며 보다 중요하게는 생략현상이 서로 다른 문법구조적특성을 가지는 영어와 조선어를 입력 및 출력언어로 하는 영-조기계번역체계에서는 생략현상을 원만하게 번역하도록 하는데 별로 도움을 주지 못한다.

영어에서 중단문이나 불완전문과 구별되는 생략문의 주요한 본질적특성은 자기의 해당 문맥을 리용하여 생략된 요소를 찾을수 있는가 없는가 하는데 있다.

편리상 생략문에서 생략된 요소를 생략요소, 생략문의 문맥속에 존재하는 생략요소와 동일한 요소를 원천요소라고 정의하면 영어의 생략현상을 생략요소와 원천요소를 연결시켜주는 수단의 종류에 따라서 병렬형생략과 종속형생략으로 구분할수 있다.

병렬형생략에서는 생략요소와 원천요소사이의 연결이 병렬접속수단에 의하여 이루어지며 종속형생략에서는 그 연결이 병렬접속수단이 아닌 다른 수단에 의하여 이루어진다. 논문에서는 종속형생략에 대하여서만 구체적으로 언급한다.

종속형생략의 대표적인 몇가지 실례들을 아래에 주었다.

실례 1 I am sorry (that) you can't come.

실례 2 I've been told (that) they're a nice colour.

실례 3 That's all (that) I can do.

실례 4 She is not so young as she looks (young).

실례 5 They needed high utilization to get as much as they would (need) from their investments.

실례 6 Thus, resource utilization may be lower than it would be (low) without a deadlock-avoidance algorithm.

종속형생략은 실례 1, 2, 3에서 나타난 종속접속사 that의 생략과 같이 원천요소가 없는 경우와 실례 4, 5, 6에서처럼 원천요소가 존재하는 경우로 나누어볼수 있다.

2. 종속형생략구조의 해석과 합성

1) 원천요소가 존재하지 않는 종속형생략에 대한 해석과 합성

영어에서 일어나는 종속접속사 that의 생략특성은 그것이 구체적인 형용사나 동사어휘에 관계된다는것이다. 영조기계번역체계 《룡담산》 2.0에서는 종속접속사 that가 생략된 that-절을 이끄는 동사나 형용사에 대한 패턴을 정의하고 해당 사전올림말에 그 패턴정보와 합성정보를 함께 입력하였다. 그리고 문법해석체계에서 해당 패턴정보를 리용하여 접속사 that의 생략을 처리하도록 하였다. 실례 1, 2에 대한 해석구조를 그림 1, 2에 주었다.

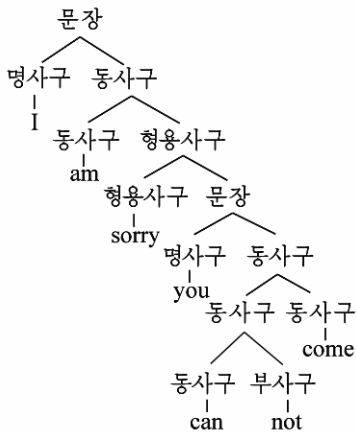


그림 1. 실례 1에 대한 영-조기계
번역체계 《룡담산》 2.0의
해석구조

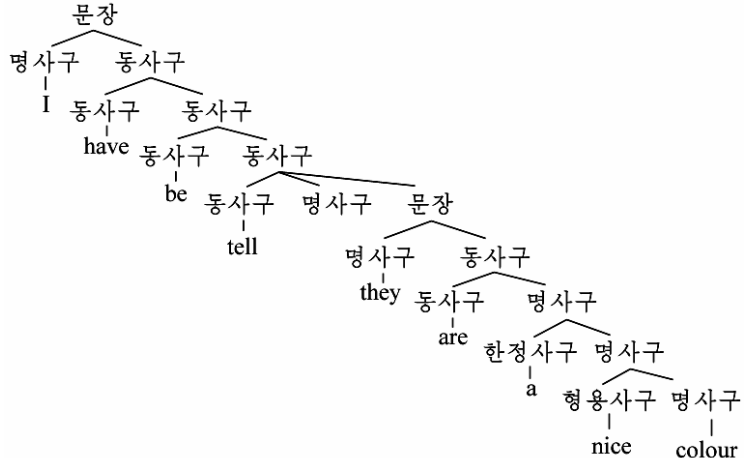


그림 2. 실례 2에 대한 영-조기계번역체계
《룡담산》 2.0의 해석구조

영-조기계번역체계 《룡담산》 2.0에서는 또한 관계대명사가 생략된 경우를 포함하여 관계절의 규정어적쓰임을 해석하기 위하여 동사패턴을 중심으로 추출구조를 정의하고 관계절속에 들어있는 추출된 요소와 선행사를 호상 연결시키도록 하였다. 그리고 관계대명사가 생략된 경우 관계절의 합성에서는 추출요소와 함께 추출요소에 붙어야 할 토도 없애버리도록 하였다.

2) 원천요소가 존재하는 종속형생략에 대한 해석과 합성

비교를 나타내는 영어표현과 같이 원천요소가 존재하는 종속형생략의 주요한 특징은 그것이 구체적인 어휘표현과 관련되어있으며 또한 생략요소가 반드시 동사구와 련관된다는 것이다.

영-조기계번역체계 《룡남산》2.0에서는 원천요소가 존재하는 종속형생략에 대한 해석을 두 단계로 나누어 진행한다.

첫 단계는 생략요소를 포함하고있는 동사구 및 문장의 추출구조를 해석하는 단계로서 이 단계에서는 일반적인 문법해석규칙들에 추출구조가 해석된다.

두번째 단계는 이 추출구조를 그 원천요소를 포함하고있는 구조와 련결시키는 단계로서 여기서는 문형해석규칙들에 의하여 구체적인 어휘표현이 해석된다.

영-조기계번역체계 《룡남산》2.0에서는 실례 4, 5, 6에서 나타나는 영어비교표현들을 as 와 than의 문형규칙을 리용하여 처리한다. 리용되는 문형규칙의 BNF형식은 다음과 같다.

RULE ::= LHS → CAT : RHS

LHS ::= [LEXICAL_E | VAR]* KEY [LEXICAL_E | VAR]*

LEXICAL_E ::= lexical_E

VAR ::= var [CONSTRAINT-LIST]

CONSTRAINT-LIST ::= [CONSTRAINT]⁺

CONSTRAINT ::= constraint

constraint ::= +constraint-string | -constraint-string

CAT ::= category

RHS ::= [LEXICAL_K | VAR]* KEY [LEXICAL_K | VAR]*

LEXICAL_K ::= [lexical_K]⁺ [toK]*

lexical_E ∈ E-Dic, category ∈ CATS, constraint-string ∈ CONSTS, lexical_K ∈ K-Dic, toK ∈ TO-Dic, 여기서 E-Dic, CATS, CONSTS, K-Dic, TO-Dic는 각각 영어단어사전, 구법주목록, 속성제한 목록, 조선어단어사전, 조선어토사전을 의미한다. 실례 4의 비교표현을 해석하는 문형규칙을 실례로 들면 다음과 같다.

as ^AJP as SENT {+ADJ_EXTRACTION} → AJP : SENT하는것만큼 AJP하다.

여기서 +ADJ_EXTRACTION은 형용사구추출이 일어났는가를 검사하는 속성제한을 의미한다.

실례 4에 대한 해석구조는 그림 3과 같다.

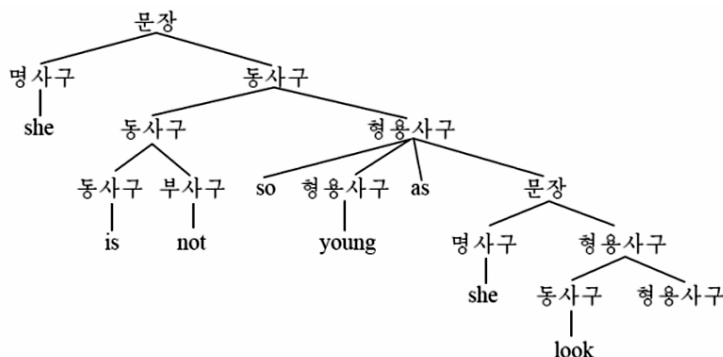


그림 3. 실례 4에 대한 영-조기계번역체계 《룡남산》2.0의 해석구조

원천요소가 존재하는 종속형생략구조의 조선어합성에서는 원천요소가 존재하지 않는 종속형생략구조의 번역에서와는 달리 원천요소를 찾아 추출요소의 자리에 복사해주도록 한다.

3. 종속형생략구조번역방법의 실현 및 평가

영-조기제번역체계 《룡남산》 2.0에 제안한 방법을 적용하여 선행한 방법의 번역결과와 비교분석하였다.

원천요소가 존재하는 종속형생략현상을 대상으로 하여 이미 실용화되어있는 영-조기제번역프로그램[2-4]들의 번역결과는 다음과 같다.

실례 7 She is not so young as she looks(young).

그 녀자가 보는것만큼 그 녀자는 그리 젊지 않다.[2]

그 녀자는 그 녀자가 볼 때처럼 젊지 못하다.[3]

그 녀자는 그 녀자가 봄에 따라 젊지 않다.[4]

실례 8 They needed high utilization to get as much as they would (need) from their investments.

그것들은 그것들이 그들의 투자들로부터 한만큼 그만큼 얻는데 높은 사용효률이 필요되었다.[2]

그것들은 그 투자들로부터 만큼 구하기 위한 높은 유용성을 구하였다.[3]

그(것)들은 그(것)들이 그(것)들의 투자들을 fit하게 될것임에 따라 많이 도착하기 위하여 높은 리용을 요구하였다.[4]

실례 9 Thus, resource utilization may be lower than it would be (low) without a deadlock-avoidance algorithm.

이렇게 자원리용은 교착상태회피알고리즘이 없을것인것보다 더 낮을 수 있다.[2]

그래서, 그것이 교착회피알고리즘이 없이 있을것인것보다 자원사용효률은 더 낮을 수도 있다.[3]

이리하여 자원리용은 그것이 막힘회피알고리즘이 없을것인것보다 더 낮을 수 있다.[4]

우의 번역문들에서 보는바와 같이 영어비교표현들의 조선어번역이 유순하게 얻어지지 못하고있는것은 생략된 요소에 대한 처리방식에 기본원인이 있다.

우의 실례문장들에 대하여 제안한 방법을 적용한 영-조기제번역체계 《룡남산》 2.0의 번역결과는 다음과 같다.

그 녀자는 젊어보이는것만큼 그리 젊지 않다.

그것들은 그들의 투자들로부터 그것들이 받을만큼 많이 받기 위하여 높은 리용을 필요로 하였다.

이와 같이 자원리용은 그것이 교착회피알고리즘이 없이 낮을것보다 더 낮을수도 있다.

이상의 비교분석을 통하여 논문에서 제안한 방법이 원천요소가 존재하는 종속형생략현상을 처리하는데서 효과가 있다고 볼 수 있다.

참 고 문 헌

- [1] 김일성종합대학학보(어문학), 54, 2, 78, 주체97(2008).
- [2] 영-조기제번역프로그램 《룡남산》 1.0, 2008.
- [3] 영-조기제번역프로그램 《아리랑》 1.0, 2006.
- [4] 영-조기제번역프로그램 《광명》 2.0, 2007.
- [5] 리명환; 외국어연구, 21, 131, 주체94(2005).
- [6] 리명환; 외국어연구, 22, 139, 주체95(2006).
- [7] 리명환; 외국어연구, 23, 119, 주체95(2006).
- [8] 조현; 외국어연구, 32, 113, 주체99(2010).

주체106(2017)년 8월 5일 원고접수

**An Approach to Translating the Subordinate Ellipsis Structure
in English-Korean Machine Translation**

Kim Kwang Hyok, Kim Hyon Chol

We have classified English ellipsis phenomena into 2 kinds, coordinate and subordinate, and conducted a research work for improving the translation of subordinate ellipsis in EKMTS.(English-Korean Machine Translation Software)

Key words: EKMTS(English-Korean Machine Translation Software), ellipsis