

한영기계번역을 위한 ‘있다’ 구문의 처리 방법*

이 동 혁
(고려대학교)

Lee, Dong-hyeok. (2004). Processing methods of *issta* constructions for Korean-English machine translation. *Language Research* 40(1), 253-277.

This paper aims to introduce the processing methods of *issta* constructions for Korean-English machine translation. Semantic ambiguities as well as many sentence patterns of *issta* constructions in Korean make them difficult to automatically translate into English ones (target language). My claim, in this paper, is as follows. 1) We need to differentiate *issta* that functions as predicate and one that does not. 2) We need to solve processing *issta* ambiguity in terms of using the case frame of *issta* constructions. 3) To analyze *issta* constructions, after searching features of predicate and analyzing case markers followed by arguments of the construction, we write its results in the form of table (Proposition Table). 4) Regarding multiple nominative (*ka*) constructions, we first search the dictionary information of the nouns following *ka*. In doing so, if identified as nouns [Possessor], then we assign [Agent] features to the nouns following *ka*; and if identified as nouns [Possessed], then we assign [Theme] features to the same ones. We think, the ways of dealing with semantic ambiguities and case analysis of *issta* constructions can surely be extended to other types of similar constructions in Korean.

Key words: machine translation, ambiguity, multiple nominative construction, *issta*-construction, case marker

1. 문제 제기

이 글에서 우리는 한영기계번역을 위해 ‘있다’ 구문을 처리하는 방안을 제시하고자 한다. 그동안 자연언어로서의 ‘있다’는 형태적인 특이성과 중의적인 의미 특성으로 인해 상당히 많은 연구가 있어왔고, 이것이 결국 ‘있다’의 범주 귀속 문제로 귀착된 바 있다.¹⁾

* 이 글을 심사하신 세 분의 심사자들에게 감사할 드린다. 이 글에서 문제가 되는 부분은 모두 필자의 책임이다.

1) ‘있다’의 범주 설정 문제는 ‘있다’의 형태론적 패러다임의 불일치에서 비롯된다.

가. *오늘은 수업이 있다.

나. *그곳에 학교가 있다.

그러나 한영기계번역에서는 번역 대상어(source language)가 되는 ‘있다’는 분석 대상이 되는 문자열(character string)이고, 분석 결과를 목표어(target language)인 영어의 특수성에 부합하도록 생성이 되어야 할 문자열이다. 따라서 ‘있다’가 본유적으로 의미적 중의성을 가지거나 여러 가지 구문 패턴을 보이고 있는 것이라면, 분석 과정에서 이러한 의미적 중의성과 구조적 중의성을 해결하여 올바른 번역어를 생성해야 하는 것이 ‘있다’와 관련한 한영기계번역 처리의 궁극적인 목적이다. 다음은 한영기계번역을 위한 ‘있다’의 처리와 관련하여 문제가 되는 예문과 현재 시중에 판매 중인 한영기계번역기가 그 예문을 번역한 문장이다.²⁾

(1) 가. 있다가 먹읍시다.

⇒ Let's eat while is.

나. 칠수가 작년 에 미국에 있었다.

⇒ It was in United States of America at last year Chul-su.

다. 칠수에게 돈이 있다.

⇒ There is money to Chul-su.

라. 칠수가 돈이 있다.

⇒ Chul-su has money.

(1가)의 ‘있다’는 (1나)~(1라)의 ‘있다’와는 달리 문장의 서술어로서의 지위를 가지지 못한 것이다. (1가)의 ‘있다’는 후행하는 어미인 ‘-다가’와 더불어 부사로 전용된 것이지 그 문장의 서술어의 기능은 ‘먹다’가 전적으로 담당한다. 따라서 ‘있다’ 구문을 처리하기 위해서는 서술어로 기능할 수 있는 ‘있다’와 그렇지 못한 ‘있다’를 구분할 수 있는 방법을 마련해야 한다.

(1나)의 ‘있다’가 [존재]인 반면에, (1다)와 (1라)의 ‘있다’는 [소유]라는 점에서 다르다. 한국어에서는 ‘있다’가 두 의미를 함께 포함하고 있지만, 영어에서는 각 의미가 다른 번역어로 생성되어야 한다. 결국 이러한 차이는 (1나)와 (1다, 라)를 다르게 분석해야 할 당위성을 제공해 주는데, 이 두 의미를 기계번역 과정에서 어떻게 구별해 낼 것인가가 문제다. 이것은 비단 ‘있다’ 구문에서만 문제가 되는 것이 아니고 기계번역의 격 정보 분석의 핵심적인 문제다. 본고에서는 ‘있다’의 구문 패턴에서

다. 여기에 있는 학생은 내가 찾는 사람이 아니다.

라. 그 일이 있을 후에 아무도 그를 찾지 않았다.

(가)와 (나)에서 어미 ‘-는다’가 어간 ‘있-’에 결합하지 못하지만, (다)에서는 관형형어미 ‘-는’이 어간 ‘있-’에 결합할 수 있다. 두 어미 모두 동작성 동사와 결합한다는 제약을 갖지만, (가)에서는 비문법적이고 (다)에서는 문법적이다. 같은 단어에 대해 이러한 모순적인 현상이 벌어지는 것은 상당히 특이한 일인데, (가, 나)와 (다)에서 볼 수 있는 불일치 때문에 ‘있다’를 동사나 형용사 범주에 편입시키지 말고 새로운 범주를 제시하자는 의견(예: 이희승(1956), 성광수(1976))도 대두되었다.

2) 본문 (1)의 번역문은 (주)엘엔아이소프트에서 2001년 7월에 출시된 한영기계번역기 한가이드 2.0 버전에서 뽑은 것이다.

그 해답을 찾을 수 있으리라고 본다.

한편, (1라)는 주격조사 '이/가'가 중출되어 나온 구문이다. 이는 문장에서 주격조사가 나오면 그 조사에 선행하는 NP는 주어로 확인된다는 단순한 처리가 불가능함을 보여주는 예라고 할 수 있다. 따라서 '있다' 구문의 처리를 위해서는 소위 격중출 구문에 대한 고려가 필요하다.

이로써 (1)의 예문을 통해 한영기계번역에서 제기될 수 있는 문제를 정리하면 다음과 같다.

(2) 가. 서술어인 '있다'와 서술어가 아닌 '있다'는 어떻게 구분하여 처리할 것인가?

나. '있다' 구문의 격 정보는 어떻게 분석되어야 할 것인가?

다. '있다'의 의미를 어떻게 구분하여 처리할 것인가?

라. '있다'의 격중출 구문을 어떻게 처리할 것인가?

위에서 제기된 문제들을 해결하기 위하여 이 글의 논의는 다음과 같이 진행될 것이다. 2장에서는 한영기계번역 시스템의 처리 원리를 자질연산문법을 중심으로 살펴보고 4장과 5장에서의 '있다' 구문의 처리와 영어 생성에 이론적 도움이 되도록 할 것이다. 3장에서는 서술어 '있다'를 확인할 것이다. 4장에서는 1장에서 제기한 문제를 중심으로 '있다' 구문의 분석 방법을 제시할 것이다. 그리고 5장에서는 '있다' 구문이 영어로 생성되는 모습을 보일 것이다. 마지막으로 6장에서는 4장과 5장에서 설정한 분석과 생성 절차를 요약하고, 본고에서 해결하지 못한 문제점을 밝힐 것이다.³⁾

2. 한영기계번역의 처리 방법론: 자질연산문법을 중심으로

2.1. 자질연산문법의 개관

본고에서 채택하고 있는 한영기계번역을 위한 문법 모델은 자질연산문법(Feature Computational Grammar, 줄여서 FCG)이다.⁴⁾ FCG의 목적은 분해기 문법이라는 정보처리를 위한 언어 분석의 이론이다. FCG는 언어 자료체를 중심으로 하여 연구를 진행하며, 이 자료체들에서 귀납되는 규칙과 반복적으로 등장하는 원리를 인간의 언어 직관, 즉 언어 능력의 반영으로 간주한다(김원경, 2000, p. 31). 이와 같은 성격의 FCG는 다음과 같은 체계를 가진다.

3) 이 글에서 사용되는 자질의 목록은 <부록>에 설명되어 있다.

4) FCG에 대한 자세한 설명은 김원경(2000), 유혜원(2001), 고창수(2002)를 참고하기 바란다.

(3) FCG의 체계

어휘부(포화된 자질 집합) ⇕ 연산부(연산규칙)	← 연산 원리
시스템 구성	시스템 이론

어휘부는 어휘항목의 형태적 정보, 통사적 정보, 의미적 정보, 시스템 정보가 포화된 자질의 집합으로 구성되어 사전을 이룬다.⁵⁾ 사전은 형태소분석, 구문분석의 각 분석 과정과 생성의 과정에서 참조하는 정보를 자질과 어휘항목으로 저장하고 있는 셈이고, 각 분석 과정에서 처리한 정보들을 사전에 반영하기 때문에 사전은 폐쇄적인 집합체가 아니며, 자질의 첨가와 삭제가 가능한 개방적인 집합체이므로 유연한 성격을 지닌다. 그리고 하나의 어휘항목에 대하여 대역어가 여러 개인 다의어 구성이라면, 사전 내부에 다의성을 해결하는 연산 절차가 포함되어 있고, 하나의 어휘항목이 독립적인 기능을 갖지 못하고 다른 단어와 복합 구성을 이루어서 독립적인 기능을 갖게 되면, 이에 대한 처리도 사전 내부에서 하게 된다.⁶⁾

연산부에는 형태소분석과 구문분석과 의미해석을 담당하는 분석부와 단어를 형성하고 구문을 생성하는 생성부가 포함되어 있으며, 각 분석 과정의 연산 결과로 도출된 정보는 FCG에서 별도의 부문으로 구축한 분석결과표에 저장된다.⁷⁾

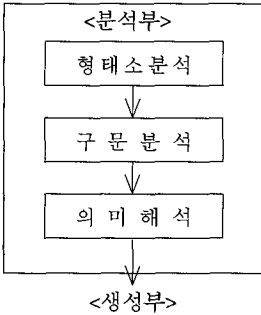
5) 김원경(2000)에서는 FCG에서 상정한 자질의 특성을 아래와 같이 설명하였다.

- 자질: 특정한 어휘항목의 외연을 표시하기 위한 내항
 - ㄱ. 어휘항목 α 는 항상 자질을 가지며, 이러한 자질 집합은 $\{f_1, f_2, f_3, \dots\}$ 로 표시된다.
 - ㄴ. 어휘항목 α 와 구별되는 어휘항목 β 는 적어도 하나의 자질에서 α 와 구별된다.
 - ㄷ. 어휘부 내 어휘항목의 자질체계는 개방적이다.
 - ㄹ. 어휘항목의 자질은 $[\pm]$ 로 표시되지 않는다.
 - ㅁ. 어휘항목들은 어휘 그물로 연결되어 있다.
 - ㅂ. 연결된 어휘항목들은 특정 자질을 공유한다.

6) 후자의 경우는, 언어 표현(collocational expression)과 관용 표현(idiom)이 대표적이다. 이와 같은 복합 구성을 사전에서 처리하는 방법에 대해서는 3장에서 서술하기로 한다.

7) 본 연구가 형태소 분석이나 생성이 목적이 아니나, 한 심사위원이 목적적 조사인 '을'과 '를'의 선택 조건과 '늑차를'이 융합된 '늑찰'의 형태소 분석에 대해 질문하였다. 형태소 분석은 음절 단위로 분석될 뿐만 아니라 때로는 음소 단위로 확대되기도 한다. 그리고 주변 형태소가 모음으로 끝나거나 자음으로 끝나는 것이냐는 것도 형태소 분석의 근거가 되기도 한다. 예컨대, '을'은 자음으로 끝나는 명사에만 결합한다는 제약이 있지만, '를'은 모음으로 끝나는 명사에만 결합한다는 제약이 있어서 이를 토대로 형태소 분석을 한다. 그리고 '늑찰'의 형태소 분석을 위해서는 둘 이상의 형태소(혹은 형태소의 일부)가 하나의 음절에 융합될 수 있어서 어떤 형태소는 음소 단위로 실현될 수 있다는 한국어의 형태소 특성을 이해하여야 한다. 황화상(2003)에서는 형태소 융합 가능성이 있는 음절을 모아 음절 사전을 구축하고 이를 통해 '늑찰'과 같은 어절을 형태소 분석을 시도한 바가 있다. 자세한 논의는 황화상(2003)을 참고하기 바란다.

(4) 연산부의 조직



정보 테이블은 분석의 결과를 저장하는 공간이고, 이 공간에 저장된 정보는 다음 분석에 참조가 될 수 있으며, 사전에서 대역어를 선택하는 근거가 된다. FCG에서는 이 정보 테이블을 명제 테이블(propositional table)과 양태 테이블(modal table)로 나눈다. 명제 테이블에 표시되는 정보는 문장성분의 정체다. 문장의 성분들이 문장에서 기능할 수 있는 최대의 후보는 테이블에 기능을 표시하여 칸으로 나누어져 있다. 그리고 각 칸에는 핵어가 되는 단어가 올 수 있으며, 핵을 꾸미는 수식어는 핵어의 아래에 수식어열로 저장된다. 한편, 양태 테이블에는 Fillmore(1968)에 따라 부정, 시제, 법, 상 등의 정보가 표시된다. 양태 테이블도 각 기능마다 칸이 나누어져 있으나 명제 테이블과 달리 그 칸에는 문자열이 쓰이지는 것이 아니라 기능이 문장에서 실현되고 있는지의 여부만 체크한다.

(5) 정보 테이블

가. 명제 테이블⁸⁾

		S	V	DO	P	
Head		L1				
Mod1		L2				
Mod2		L3				
.....					

나. 양태 테이블

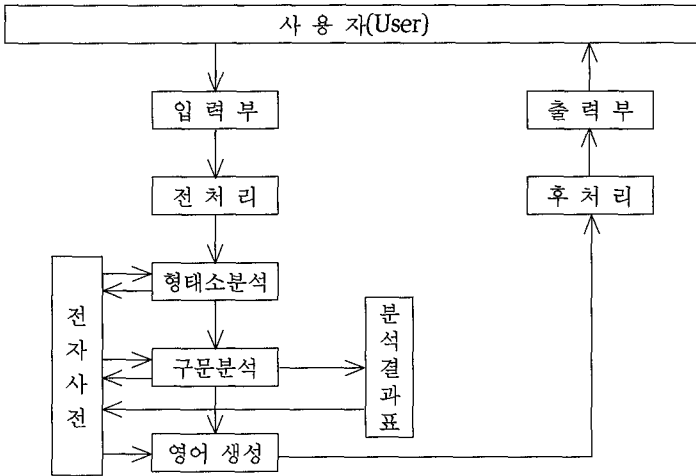
	NEG	TENSE			ST			
		Past	Present	Future	Dcl	Imp	Que	Excl
	√	√			√			

8) X'-이론에 근거한 수형도 표상 방법과는 달리 이 명제 테이블의 표상 방법은, 수식과 피수식 관계를 제외하면 위계적이지 않고 선형적이다. 그리고 목표어인 영어의 이순 구조를 고려한 테이블이다.

2.2. 한영기계번역 시스템 구조도

본고의 한영기계번역 시스템은 다음과 같은 구조를 보이고 화살표에 따른 처리 순서를 보인다.

(6) 한영기계번역 시스템 구조도



3. 서술어 ‘있다’의 확인

분석 대상 문자열에 ‘있다’라는 문자열이 포함되어 있다고 해서 반드시 이 문자열이 문장의 서술어가 되는 것은 아니다. ‘있다’가 단독으로는 문장에서 특정한 문법적·의미적 기능을 가지지 못하고 선행하는 문자열과 연합하여 문장에서 문법적·의미적 기능을 가지게 되면, ‘있다’가 서술어가 되지 못한다. 이와 같이 한국어 내적인 필요성에 의하여 ‘있다’가 서술어에서 제외되는 예가 있는가 하면, 대상어(한국어)와 목표어(영어)의 언어 차이 때문에 ‘있다’가 서술어가 되지 못하는 경우가 있다. 본 기계번역 시스템에서 문자열 ‘있다’가 서술어가 되지 못하는 때는, 다음의 (7) 예와 같이 ‘있다’가 보조용언, 양태적 요소, 연어 표현 요소(관용 표현 포함), 부사로 기능할 경우다. 아래의 예는 그 중 대표적인 것만 든 것이다.

- (7) 가. 철수가 집에 가고 있다.
- 나. 철수는 이 일을 할 수 있다.
- 다. 철수는 영화를 때린 적이 있다.
- 라. 철수는 영화에게 관심이 있다.
- 마. 있다가 먹읍시다.

위 다섯 문장에 모두 '있다'가 포함되어 있다. 그러나 독자적으로 서술어로 기능한다고 보기에는 어려운 구석이 있다.9) (7가)의 '있다'는 보조용언이고, (7나)의 '있다'는 선행하는 '을 수'와 더불어 '능력'을 나타내는 양태적(modal) 요소다. 그리고 (7다)는 영어 대역문에 비추어 보아 '은 적이 있다' 전체를 부사 'sometimes'로 처리하면 효과적인 구성이다. (7라)는 '관심'과 '있다'가 연어 표현을 이루고 있기 때문에 '관심이 있다' 전체를 하나의 서술어로 간주하는 것이 효율적이다. 마지막으로 (7마)는 어미 '-다가'와 통합하여 부사로 기능이 전용된 경우다.

따라서 (7)의 문장에서 각각의 '있다'들은 분석 대상 문자열에서 서술어가 아닌 것으로 처리를 해 주어야 한다. 이것은 '있다'에 국한된 개별적인 현상이기 때문에 구문분석 과정에서 처리하는 것보다 '있다'를 어휘항목으로 가지고 있는 사전에서 별도로 처리하는 것이 더 낫다. (7)의 '있다'는 선행하거나 후행하는 특정한 형태 단위들이 있고, 그 형태 단위들과 '있다'가 합쳐져 특별한 의미 기능을 담당하는 것들이다. 사전에서는 특정 형태 단위들이나 그 형태 단위의 자질들을 하나 하나씩 확인하여 문자열 삭제와 자질 주기를 하여 처리한다.10) 여기서는 (7나)와 (7라)의 사전 처리만 보이겠다.

(8) 인접 선행하는 [n, lex] '수'가 있고, 그 '수'를 인접 선행하는 [v, aff, unc] '을'이 있는가? (단, [n, aff]는 뛰어넘어 탐색하라.)

ly : 현문자열에서 '수'와 '을'을 삭제하라. 그리고 '있다'의 현재 자질을 모두 삭제하고, [aux] 자질을 명시하고, 대역어로 'can'을 명시하라.

ln : ……

……

5n : 선행하는 [n, lex]인 '관심'이 있는가?

6y : 현문자열에서 '관심'을 삭제하라. 그리고 '있다'의 현재 자질을 모두 삭제하고, [v, lex, t1] 자질을 명시하고, 대역어로 'be interested in'을

9) 만일 이 번역기가 한국어 대 영어의 일대일 번역 방식이 아니라 중간언어 기반 번역 방식이라면 다르게 처리되어야 할 것이다.

10) 보조용언이라든지, 양태적 요소라든지, 관용 표현이라든지 두 어절 이상이 통합하여 하나의 의미를 담당하는 요소를 사전에서 통계로 등록할 수도 있다. 그러나 한국어에서는 한 어절 이상이 통합하여 하나의 의미 기능을 담당할 수 있으나, 통합 구성 내부에 있는 조사가 개제되거나 심지어 부사까지도 개제될 수 있어 분리성이 강하다고 할 수 있다.

(가) 나는 집에 가고만 싶다.

(나) 칠수는 그 시험에서 미역국을 많이도 먹었지.

따라서 사전에 어휘항목으로 등재하여 강하게 처리하는 것보다 분리성을 최대한 인정하기로 하고, 통합 구성의 핵어(head)가 되는 단어의 사전에서 따로 처리하는 것이 좋다. 이 때는 핵어를 중심으로 통합 구성의 요소를 분포적으로 확인하고 핵어를 제외한 나머지 단어들은 삭제하고, 통합 구성이 갖는 기능을 핵어의 자질로 복사하는 처리가 필요하다. 물론 핵어가 갖는 어간의 자질은 통합 구성의 자질이 복사되기 이전에 모두 삭제되어야 한다.

(다) 먹다 [v, v3, lex, act]

(라) 미역국을 먹다 → (미역국을)삭제 먹다 [v, v1, lex, act]

명시하라.11)

6n : ……

이처럼 사전에서 별도로 처리할 수 있는 ‘있다’ 구성이 제외되면, 순수하게 서술어로 처리되어야 할 ‘있다’가 [v, lex] 자질을 가진 채로 현문자열에 남아있게 된다. 그리고 아래와 같은 서술어 확인 수순을 거쳐 PT의 V항에 쓰지며, 다음 수순을 대기한다.

(9) 서술어의 확인

현문자열에서 [v, lex]의 자질을 가진 문자열을 찾아 PT의 V항에 쓰려.

	V	
	있	

4. ‘있다’ 구문의 분석 방법론

4.1. 개관

‘있다’의 구문 구조는 곧, ‘있다’ 서술어가 문장에서 실현하는 격 범주다. 속격을 격에 포함하기 위하여 ‘격이란 종속 명사(dependent nouns)가 그 핵어(heads)에 대해 지니는 관계 유형(type of relationship)을 표시하는 체계’라고 정의한 Blake(1994, p. 1)의 격에 대한 정의를 따른다면, ‘있다’의 구문 구조의 특성을 밝히는 일은, 핵어가 되는 ‘있다’와 그 핵어의 종속 명사의 관계 유형과 그것을 표시하는 체계를 밝히는 것과 같다.12) 핵어와 종속 명사 사이에는 특수한 통사·의미적인 관계가 있고, 이러한 관계를 여러 가지 언어적 수단을 사용하여 표시한다. 한국어는 김원경(2000)의 표현을 빌리자면, 표지에 의존하여 격을 표시하는 M언어다.13)

가시적인 문자열을 근거로 하여 분석해야 하는 한영기계번역 시스템에서는 조사를 이용하여 격을 분석할 수 있겠다. 그러나 하나의 조사 형태가 담당하는 기능이

11) 한 심사위원이 ‘관심이 있다’에 대한 대응 번역으로 ‘be interested in’만 있는 것이 아니고 ‘have an interest in’도 있어서, 이 경우의 ‘있다’를 서술어와 구분할 필요가 있는지 검토되어야 한다고 지적하였다. 이 경우, 심사위원의 지적대로 ‘have an interest in’으로 번역하여 ‘있다’를 서술어로 처리할 수도 있겠지만, 번역기를 사용하는 사용자가 가장 자연스럽게 받아들이는 번역어를 최우선으로 선택하여야 한다는 조건을 고려한다면 ‘be interested in’으로 번역하는 편이 나으며 이에 따라 ‘있다’도 서술어가 되지 못한다.

12) 핵어가 명사이어서 종속 명사와 맺는 관계 유형이 수식이라면 그 체계를 속격 조사가 표시할 것이다. 한영기계번역 시스템에서는 PT의 V에 쓰여진 [v, lex]의 문자열을 가지고 논항 분석을 하기 이전에 이미 ‘명사구 한정’의 분석 수순을 거치기 때문에 본 절에서 말하는 격이란 서술어와 종속 명사가 맺는 관계 유형을 표시하는 체계라고 구체화시켜도 무방할 것이다.

13) 위치에 의해 격 표시가 되는 언어를 ‘Location-Language’(줄여서 L언어)라고 하고, 표지에 의존하여 격을 표시하는 언어를 ‘Marker-Language’(M언어)라고 한다.

단지 하나만이 아니라 여러 개일 수 있어 격을 분석하는 일이 여간 어려운 것이 아니다. 하나의 조사가 통사격뿐만 아니라 의미격, 나아가 화용격까지 표시할 수 있으니 말이다.¹⁴⁾ 따라서 현문자열을 분석하여 각 명사에 격을 할당하려면 상당히 복잡한 과정을 거쳐야 한다. 이에 FCG에서는 다음과 같은 순서로 문장의 격 정보를 분석할 것을 제안한다.

(10) 격 정보 분석 절차

- ① PT의 V항 자질 점검
- ② 구조격 분석
- ③ 본유격 분석
- ④ 내면격 분석
- ⑤ PT에 분석 결과 쓰기

- ①에서는 서술어가 능동술어인지, 피동술어인지 확인한다.¹⁵⁾ 그리고 서술어가 문장에서 취할 수 있는 논항의 개수와 각 논항의 의미역을 점검한다.
- ②에서는 '가' 조사의 인접선행 명사에 [sub]를 할당하고, '를' 조사의 인접선행 명사에 [obj]를 할당한다(단, 문장에 '가' 조사가 중출하거나 '를' 조사가 중출하는 경우는 따로 처리한다). 그리고 '가' 조사와 '를' 조사를 현문자열에서 삭제한다.
- ③에서는 '로', '와', '에', '에게'의 조사를 찾아 그 조사를 PT의 P 자리에 쓴다. (단, 서술어가 특수한 자질을 갖는 것이라면 따로 처리한다)
- ④에서는 ②와 ③에서 유보된 처리를, 명사의 내면격 정보를 이용해 다시 분석한다. 그리고 보조사가 인접후행하거나 조사가 아예 인접후행하지 않아서 구조격과 본유격에서 처리하지 못한 명사를 처리한다.¹⁶⁾
- ⑤ ①에서 ④의 단계로 확인된 명사의 [sub], [obj] 자질을 점검하고, 그 결과를 PT에 쓴다.¹⁷⁾

- 14) 여기서 의미격은, 김원경(2000)의 용어로는 '내면격'에 해당한다. 내면격이란 핵어인 서술어에 대해 종속 명사의 관계를 의미적인 면으로 바라본 것이다. 그리고 화용격에 대해서는 한정현(1999)를 참고하기 바란다.
- 15) 본고에서는 술어를 내적 상 특성에 따라 분류하였다. 각주 1)에서 언급한 것처럼 '있다'는 기존의 동사나 형용사의 성질을 모두 가지고 있어서 이 두 범주 어디에도 분명하게 귀속시키기가 어렵기 때문에 술어의 상적 개념을 도입하는 게 정당화될 수 있을 것이다. 피동술어는 [+stative]의 특성을 가진 술어이고, 반면에 능동술어는 [-stative]의 특성을 가진 술어라고 규정된다. 따라서 본고에서 상정하고 있는 자질 목록을 통해 보면, 능동술어는 서술어가 동작동사 [act]이거나 심리동사 [psy]이거나 소유동사 [pos] 자질을 가질 때 확인되며, 피동술어는 서술어가 상태동사 [state] 자질을 가질 때 확인된다. 능동술어와 피동술어의 구분 테스트는 Van Valin *et al.* (1997, p. 94)를 참고하기 바란다.
- 16) 보조사가 인접후행하거나 조사가 아예 인접후행하지 않아서 ④의 단계로 넘어온 구문에 대한 처리는 이 글의 논의의 대상에서 제외하였다.
- 17) 지금까지 설명한 격 정보 분석 절차는 서술어의 의미역을 실현할 수 있는 논항이 문장에 모두 있는 경우만을 가정한 것이었다. 하지만 한 심사위원이 지적한 대로 서술어의 고유한 의미역 정보는 있

‘내 아들이 집에서 밥을 먹었다.’의 예를 가지고 위의 격 정보 분석 절차를 따라가 보자. 분석 대상이 되는 문자열 ‘내 아들이 집에서 밥을 먹었다.’를 형태소 분석한 이후에 주요 자질을 확인하면 아래와 같다.

나	의	아들	이	집	에서	밥	을	먹	었	다
n, lex, pro	n, aff, gen	n, lex, hum	n, aff, agt	n, lex, plc	n, aff, sour	n, lex	n, aff, thm	v, v2, lex	v, aff, pt	v, aff, dcl, fin

각 문자열과 그 문자열의 자질을 이용해 명사구를 한정하고, 문장의 유형을 파악한 후, 격 정보를 처리하기 이전 단계의 모습은 다음과 같다.

(11) PT

	V	
	먹	

(12) 현문자열

	아들	이	집	에서	밥	을	먹
자질	n, lex, hum	n, aff, agt	n, lex, plc	n, aff, sour	n, lex	n, aff, thm	v, v2, lex, act, pst

- ①의 수준에 따라, PT의 V항의 자질을 점검한다. ‘먹’은 [v2, act]이기 때문에 능동술어로 확인된다.
- ②의 수준에 따라, 현문자열에서 [n, aff, agt] 문자열을 탐색한다. 그리고 이 문자열의 인접선행 [n, lex] 문자열에 [sub] 자질을 할당한다. 또 현문자열에 [n, aff, thm] 문자열을 탐색한다. 그리고 이 문자열의 인접선행 [n, lex] 문자열에 [obj] 자질을 할당한다.
- ③의 수준에 따라, 현문자열에서 [n, aff, sour] 문자열을 탐색하여 발견되는 문자열을 PT의 P칸에 쓴다. 그리고 이 문자열의 인접선행 [n, lex] 문자열을 [n, aff, sour] 문자열의 Mod에 쓴다.
- 서술어에 ④의 처리를 거쳐야 할 특수한 자질이 발견되지 않으므로, 곧바로 ⑤

나 그 의미역을 실현할 수 있는 논항이 문장에 없는 경우도 기계번역 시스템에서는 처리해야 한다. 그런데 필수논항이 생략된 문장을 처리하는 일은 매우 힘든 작업이다. 왜냐하면, 현재의 기계번역기 수준이 문장의 범위를 넘어서 텍스트 단위의 정보를 이용하거나 화용론적인 정보를 이용할 수 없고, 거의 전적으로 문장 내부의 정보만을 이용하고 있기 때문이다. 본 기계번역 시스템에서는 문장 내부의 정보를 최대한 이용하여 필수논항이 생략된 문장을 처리하고 있다. 격 정보 분석 절차 중 ⑤단계에서 [sub]나 [obj] 등의 자질을 발견하지 못하면, PT의 빈 논항 자리에 [n, lex] 자질을 갖는 임시 문자열 a를 쓰고, 생성 과정에서 MT의 정보와 서술어의 어말어미 정보 등을 활용하여 a를 최대한 복원할 수 있다. 특히 생략된 주어 성분의 복원에 대해서는 최정혜(2003)를 참고하기 바란다.

의 수순으로 이동한다.
 지금까지의 처리를 통해 남아있는 현문자열과 PT를 보이면 다음과 같다.

(13) PT

		V		P	
Head		먹		에서	
Mod				집	

(14) 현문자열

	아들	밥	먹
자질	n, lex, hum, sub	n, lex, obj	v, v2, lex, act, pst

- ⑤의 수순에 따라, 현문자열에서 [sub] 문자열을 탐색하여, 그 문자열을 PT의 S에 쓴다. 그리고 현문자열에서 [obj] 문자열을 탐색하여, 그 문자열을 PT의 DO에 쓴다. 이로써 격 정보의 분석이 끝나고 그 결과가 표시된 PT를 보이면 다음과 같다.

(15) PT

		S	V	DO		P	
Head		아들	먹	밥		에서	
Mod		나				집	

‘내 아들이 집에서 밥을 먹었다.’의 예는 (10)의 격 정보 분석 절차에서 ④의 과정을 거치지 않고도 격 정보가 제대로 분석된 경우다. 그러나 ④의 과정을 거치지 않고 격 정보가 분석되는 예는 실제 용례를 보면 그리 많지 않다. 이것은 ④의 과정이 그만큼 중요하다는 것을 나타낸다. ‘가’ 조사가 문장에서 중출하는 경우를 예로 해서 ④의 과정에서 어떻게 처리하는지 알아보자. ‘영희가 얼굴이 예쁘다.’의 예로 이 과정을 설명해 보기로 한다.

(10)의 ①의 수순에 따라, PT의 V항인 ‘예쁘’의 [state] 자질을 통하여 ‘예쁘’가 피동술어임을 확정한다. 그리고 ‘예쁘’의 [v1, Theme] 자질로 논항의 개수와 논항의 의미역을 점검한다. (10)의 분석 절차에서 ②와 ③의 과정은 이미 ‘내 아들이 집에서 밥을 먹었다.’의 예로써 설명한 바이기 때문에 생략하기로 한다. 이제 ③의 과정이 끝나고 ④의 과정을 앞두고 있는 상태의 현문자열을 보이면 아래와 같다.

	영희	얼굴	예쁘
자질	n, lex, hum, sub	n, lex, sub	v, v1, lex, state, pst

위의 현문자열이 ④의 수순을 거쳐야 하는 이유는 현문자열에 [sub] 자질이 포함된 명사가 두 개 있기 때문이다. PT의 S칸에 두 개의 명사가 동시에 써질 수 없기 때문에 두 개의 [sub] 자질의 명사 가운데 진짜 주어가 될 단어는 하나다. 진짜 주어를 찾기 위해 ①에서 점점한 서술어의 고유한 의미역 정보를 확인해야 한다. 서술어 '예쁘다'는 의미역 관련 정보로 [v1, Theme]을 가지고 있다. 따라서 의미역 계층(thematic hierarchy)에 따라 [Theme] 자질의 의미역을 갖는 명사가 주어가 된다.¹⁸⁾ 이미 ②의 처리 결과, [Theme]의 자질을 가질 수 있는 명사는 [sub] 자질을 가진 '영희'와 '얼굴'로 압축되었다. 결국 두 개의 [sub] 자질의 명사 중에서 [Theme] 자질을 가진 명사를 찾아내는 것이 ④의 수순에서 해야 할 일이다.

의미역 정보가 명사에 고유하게 저장된 정보가 아니라 해석되어야 할 것이라면 [Theme] 자질의 명사를 찾아내는 알고리즘이 또한 필요하다.¹⁹⁾ 현문자열의 서술어 '예쁘'는 [d3]의 자질을 가지고 있다. 이 자질로 [sub] 자질의 명사들이 부분-전체 관계임을 알 수 있는데, [Theme]의 자질을 명세화할 수 있는 것은 부분에 속하

- 18) 의미역의 정밀하게 설정하고 세분화하는 것은 언어 일반 이론에서나 자연언어처리에서 매우 필요하고 의미가 있는 일이다. 그러나 본고에서는 의미역 자체보다는 의미역이 통사부와 관계하면서 통사부의 어떤 자리에 위치하느냐를 더욱 중요시하기 때문에 의미역 자체에 대한 정밀한 탐구와 모색을 하지는 않았다. 본고의 의미역은 Van Valin et. al.(1997, p. 85)을 따라 다음과 같이 설정한다.

agent, effector, experiencer, instrument, force, patient, theme, benefactive, recipient, goal, source, location, path

의미역에 대해 논하는 연구자마다 의미역 개수가 다르고, 의미역의 명칭도 달라 매우 혼란스러운 것이 사실이다. 예컨대, 김원경(2000)에서는 총 11개의 의미역을 설정하였으나, 최호철외(1999)에서는 39개의 의미역을 설정하고 있다. 이와 같은 혼란스러운 의미역을 이용해 절대적으로 격 정보를 분석하는 것은 문제가 있을 수밖에 없다는 것을 예측하게 해 준다(Koenig & Davis, 2001). 그리고 서로 다른 의미역이 주격이나 대격으로 실현될 수 있기 때문에 의미역과 격 할당 사이에는 일대일 관계가 성립하지 않는다. 이 문제를 적절하게 해결할 수 있는 방법 중 하나가 의미역 계층 이론에 기대는 것이다. 의미역은 Van Valin et. al.(1997)과 한정현(2002)에 따라 그룹이 만들어지고, 이 그룹은 계층을 이루게 된다.

(1그룹)	(2그룹)	(3그룹)	(4그룹))
agent	source	goal	theme
force	location	instrument	patient
	path		
	experience		
	extent		

이제 주격(본 번역기 시스템에서는 [sub])을 확인하고, 대격([obj])을 확인하는 일은 개별적인 의미역에 의하지 않고 그룹의 순위에 의한다. (10)의 ①의 수순에 따라 능동술어로 확인되면, 1그룹에 가장 가까이 속한 의미역 자질을 가진 명사가 [sub] 자질을 할당 받고, 4그룹에 가장 가까이 속한 의미역 자질을 가진 명사가 [obj] 자질을 할당 받는다. 반면에 피동술어로 확인되면, [sub] 자질은 4그룹에 가장 가까이 속한 의미역 자질을 가진 명사에 할당되고, [obj] 자질은 1그룹에 가장 가까이 속한 의미역 자질을 가진 명사에 할당되어 그 순서가 능동술어와 정반대가 된다. 의미역계층 이론이 한국어의 격 정보 분석에 유용하게 이용될 수 있음은 이미 한정현(2002)에서 증명된 적이 있다.

- 19) '가' 조사가 증출되는 경우를 몇 가지 유형으로 정리한 유형선(1999)의 연구가 참고가 된다. 이 중에서 '예쁘다'는 제3유형에 속한다. '예쁘다'가 제3유형에 속하는 특성은 '예쁘다'의 사전에 [d3]이라는 자질로 저장되어 있다. 그리고 제3유형의 서술어가 '가' 조사가 붙은 명사에 요구하는 것은, 그 명사들 사이의 의미 관계가 부분-전체 관계를 유지해야 한다는 제약이다. 따라서 [d3]의 자질은 '가' 조사 증출 구분이 나올 수 있다는 예측할 수 있으며, '가' 조사(본 기계번역 시스템에서는 [n, aff, agt]의 자질)의 인접선행 명사들 사이의 관계가 부분-전체 관계임을 나타내는 것이다.

는 명사다. 부분-전체 관계는 현재 [sub] 자질을 가진 각 명사의 사전에서 확인할 수 있다. 이로써 '얼굴'이 [Theme] 자질을 획득하게 되어 [sub] 자질을 유지할 것이며, '영희'는 [sub] 자질을 잃게 된다. 대신 '영희'는 '얼굴'과의 관계에 의해 [gen] 자질을 얻어서 '얼굴'을 수식한다. ④의 과정을 간단히 정리하면, 다음과 같다.

(16) '가' 조사 중출 구문의 알고리즘 일부

현문자열에 [sub]항이 두 개 이상인가?

1y : PT의 V항의 자질에 [d1]이 있는가?

2y : ……

2n : ……

3y : ……

3n : PT의 V항의 자질에 [d3]이 있는가?

4y : [sub]항의 Word_Dic를 탐색하여 Whole>Part의 순서로 나열하라. Part의 방향에 가장 가까운 항에 [Theme] 자질을 추가하라. 나머지 [sub]항의 자질 목록에서 [sub]를 삭제하고, [gen] 자질을 추가하라. 그리고 [Theme]항의 수식어열에 [gen]항을 써라.

4n : ……

……

다음으로 [sub] 자질을 점검하여 PT의 S에 쓰는 ⑤의 수순을 거치고 나면, 분석결과표에는 다음과 같은 정보들이 저장된다.

		S	V	
Head		얼굴	예쁘	
Mod		영희		

4.2. '있다' 구문의 격 정보 분석

'있다' 구문의 격 정보 분석은 4.1절의 두 예의 격 정보 분석보다 상당히 복잡한 과정을 거쳐야 한다. 그 이유는 다음과 같다. 첫째, '있다'는 최호철(1993)의 뜻풀이에서 알 수 있듯이 다의어다. 둘째, '있다'는 그 의미에 따라 다양한 구문 패턴이 나타난다.

(17) '있다'의 의미(최호철, 1993, p. 358)

1-1. 일정한 곳에 자리를 차지하다(책상 위에 책이 있다)

1-2. 존재하다(사랑과 믿음이 있는 사회 / 몸에 열이 있다 / 그는 나한테 거짓말을 한 적이 있다)

- 1-3. 발생하거나 벌어지다(모임이 있다 / 곧 시험이 있다)
- 1-4. 일정한 곳에 머물러 살거나 지내다(형은 영국에 있다)
- 2-1. 머무른 상태를 지속하다(어디 가지 말고 여기 있어라)
- 1-5. 어떤 직장에 근무하다(지금 있는 데는 대우가 괜찮다)
- 1-6. 갖추어지다(용기가 있다 / 양심이 있다)
- 1-7. 소유되다. 또는, 넉넉하게 갖추어지다(나에게 1000원이 있다 / 돈이 있는 집안 / 있는 집 아이)
- 1-8. 시간적으로 경과하다(1년만 있으면 졸업이다)
- 1-9. 어떠한 상태에 놓이거나 처하다(공사는 마무리 단계에 있다)
- 1-10. <‘에(게) 있어서’의 꼴로 쓰이어> 무엇을 화제로 삼을 때 사용하는 말로, ‘에’, ‘에게’, ‘에서’를 문어체로 이르는 말(한국 문학에 있어서의 모더니즘 / 인간에게 있어서 자유란 소중한 것이다)
- 1-11. <‘-르 (을)수 있다’의 꼴로 쓰이어> 가능성을 나타내는 말(나는 무슨 일 이든 할 수 있다)

그리고 ‘있다’의 구문 패턴은 고석주(1996)을 보이기로 한다.²⁰⁾

(18) ‘있다’의 구문 패턴(고석주, 1996)

- 가. 있다1: NP1-이 (NP-에) 있다 / 귀신이 (이 세상에) 정말 있을까?
- 나. 있다2: NP1-이 NP2-에 있다 / 보육원이 모텔 옆에 있다.
- 다. 있다3: NP1-이 NP2-에 있다 / 네가 집에 있어라.
- 라. 있다4: NP1-에/에게 NP2-가 있다 / 철수에게 돈이 있다.
- 마. 있다5: NP1-이 NP2-가 있다 / 철수가 돈이 있다.
- 바. 있다6: NP1-이 (NP-에) NP2-로 있다 / 그가 (연세대학교에) 학장으로 있다.
- 사. 있다7: NP1-이 NP2-에서 있다 / 국어학 강의가 213호에서 있었다.
- 아. 있다8: NP1-이 부사어+있다 / 행복은 멀리 있는 거라고 여겼었다, 나는 이제 당신과 함께 있겠오.

‘있다’의 의미와 그 의미에 따른 구문 패턴을 대응시켜 보면 아래와 같다.²¹⁾

20) 본고에서는 고석주(1996)이 설정한 구문 패턴을 받아들이고 있다. 고석주(1996)에서는 ‘있다’의 논항을 실현하는 통사적 특징에 따라 우선적으로 구분하였고, 논항 실현 양상이 같다고 하더라도 ‘있다’의 의미가 구분되면 이를 다른 패턴으로 처리하였다.

21) (17)의 의미 중에서 (1-6), (1-10), (1-11)의 의미를 구문 패턴에 대응시키지 않은 것은, 3장에서 설명한 대로 구문분석 과정에서 해결할 일이 아니기 때문이다. 즉, ‘있다’의 사전 내부에서 처리를 하면 되는 것이고, 사전에서 처리된 이후에는 ‘있다’와는 전혀 다른 단어로 기능한다.

- (19) 가. NP1-이 NP2-에 있다 → 1-1, 1-2, 1-4, 2-1, 1-9
 나. NP1-에게 NP2-가 있다 → 1-7
 다. NP1-이 NP2-가 있다 → 1-7
 라. NP1-이 (NP-에) NP2-로 있다 → 1-5
 마. NP1-이 NP2-에서 있다 → 1-3
 바. NP1-이 부사어+있다 → 없음

(19바)의 구문 패턴에 대응하는 의미가 없는 것은 아니다. 다만 고석주(1996)에서는 부사어를 필수성분으로 보았기 때문에 별도의 구문 패턴으로 독립되어 있지만, 의미적으로는 (17)의 (1-2)에 해당한다. (17)을 보면 한 구문 패턴에 여러 의미가 대응되기도 하고 한 의미가 복수의 구문 패턴으로 실현되기도 함을 알 수 있다. 결국, '있다'가 일반적인 서술어와 다르게 처리되어야 할 부분은 다음과 같다.

- 첫째, 능동술어와 피동술어를 어떻게 결정할 것인가?
- 둘째, '에게'를 인접선행하는 명사 논항을 어떻게 처리할 것인가?
- 셋째, '가' 조사 중출 구문을 어떻게 처리할 것인가?

첫째 문제는 (19)의 관계를 통해서 '있다'의 사전에 저장되어 있는 자질을 확인해 본 결과 아래와 같이 능동술어와 피동술어를 결정할 자질이 동시에 들어 있기 때문에 생긴 것이다.

- (19가)와 (19라)와 (19마)와 (19바)의 '있다'는 [v1, state, Theme]의 자질을 가진다.
- (19나)와 (19다)의 '있다'는 [v2, poss, Agent, Theme]의 자질을 가진다.

①의 수순에 따라, PT의 V항인 '있'의 자질을 점진한다. '있'은 [v1, state]와 [v2, poss]의 자질이 동시에 있어서 중의적이다. '있'이 능동술어인지 피동술어인지 정하기 위해서는 구문 패턴의 정보를 이용하는 것이 최선의 방법이다. 따라서 현문자열에 '에게' 조사가 있거나 '가' 조사가 둘 이상이 있으면 능동술어로 해결하고,²²⁾ 그 환경이 아니면 피동술어로 해결한다.

22) 본고에서는 '있다'가 능동술어로 판별될 경우에는 [소유]의 의미를 가진 것으로 보았다. 이에 대해 한 심사위원이 '있다'가 능동술어로 판별되었더라도 모두 [소유]의 의미로 판단할 수 없다고 하면서 '철수가/에게 책임이 있다'의 예를 들었다. 그러나 이 예의 '있다'를 [소유]의 의미로 결정하기가 꺼려지는 이유는 [Theme]의 의미역을 받는 '책임'이 구체물이 아니라 추상물이기 때문일 것이다. 이를 [內在] 따위로 의미 설정하여 [소유]와 구분할 수 있겠지만, 본고에서는 두 의미의 차이가 [Theme] 논항의 차이로 인한 것이지 본유적인 의미의 차이는 없다고 본다.

【능동술어이면】

- ②의 수순에 따라, 현문자열에서 [n, aff, agt] 문자열을 탐색한다. 그리고 이 문자열의 인접선행 [n, lex] 문자열에 [sub] 자질을 할당한다. 단, 현문자열에서 [n, aff, agt] 문자열이 둘 이상이면, ④의 수순을 거친다.
- ③의 수순에 따라, 현문자열에서 [n, aff, des] 문자열이나 [n, aff, sour] 문자열이나 [n, aff, inst] 문자열을 탐색하여 발견되는 문자열을 PT의 P칸에 쓴다. 단, 현문자열에서 ‘에게’ 조사가 발견되면, ③의 수순을 통과한다. 그리고 현문자열에서 이 문자열이 둘 이상 발견되면, ④의 수순을 거친다.
- ④의 수순에서는 명사의 의미역 정보를 이용해 최종적으로 격 정보를 부여하는 절차를 수행한다. 아울러 위에서 둘째 문제와 셋째 문제를 해결할 수가 있다.
 - 1) ②의 수순에 의해 현문자열에 [sub] 자질의 문자열이 둘 이상이면, ‘가’ 조사 중층 구문의 알고리즘을 가동한다. 서술어 ‘있’은 [d4]이기 때문에 [sub] 자질의 명사들은 소유주-소유물의 관계가 된다. 각각 [sub] 명사의 사전을 검색하여 소유주와 소유물에 해당하는 명사를 확정하고, 소유주에 해당하는 명사에 [Agent] 자질을 할당하고 소유물에 해당하는 명사에는 [Theme] 자질을 할당하는 대신 [sub] 자질을 삭제한다.
 - 2) ③의 수순에 의해 현문자열에 ‘에게’(본 기계번역 시스템에서는 [n, aff, dat] 자질)가 발견되면, ‘에게’의 인접선행 명사에 [Agent] 자질을 할당한다. 그리고 [sub] 자질의 명사에는 [Theme] 자질을 할당하는 대신 [sub] 자질을 삭제한다. 이후에 현문자열에서 ‘에게’ 조사를 삭제한다.
 - 3) ③의 수순에 의해 현문자열에 [n, aff, des] 문자열이나 [n, aff, sour] 문자열이나 [n, aff, dat] 문자열이나 [n, aff, inst] 문자열이 둘 이상이면, [n, aff, dat] > [n, aff, sour] > [n, aff, inst] > [n, aff, des]의 순서로 나열하고, 그 순서에 따라 PT의 P1에서 P4까지 차례대로 쓴다.²³⁾
- ⑤의 수순에 따라, 현문자열에서 [sub] 문자열을 탐색하여, 그 문자열을 PT의 S에 쓴다. 그리고 현문자열에서 [obj] 문자열을 탐색하여, 그 문자열을 PT의 DO에 쓴다.

【피동술어이면】

- ②의 수순에 따라, 현문자열에서 [n, aff, agt] 문자열을 탐색한다. 그리고 이 문자열의 인접선행 [n, lex] 문자열에 [sub] 자질을 할당한다.

23) PT에서 P에 해당하는 칸은 모두 네 개다. 현문자열에서 P에 쓰여질 문자열이 하나면 P1에 쓴다. 그런데 현문자열에서 P에 쓰여질 문자열이 두 개 이상이면, 별도의 알고리즘을 거쳐야 한다. 본 기계번역 시스템은 목표어가 영어이어서 PT에서 목표어의 어순을 반영한다고 이미 밝힌 바 있다. 본 기계번역 시스템의 개발을 위해 약 10,000문장 정도의 코퍼스를 구축하여 P에 쓰여질 문자열의 순서를 조사한 결과, 대체로 아래의 순서를 지키는 것으로 나타났다. 만일 현문자열에서 P에 쓰여질 문자열이 두 개 이상이면, 아래의 순서대로 P1부터 차례대로 쓴다.

[n, aff, dat] > [n, aff, sour] > [n, aff, inst] > [n, aff, des]

- ③의 수순에 따라, 현문자열에서 [n, aff, des] 문자열이나 [n, aff, sour] 문자열이나 [n, aff, inst] 문자열을 탐색하여 발견되는 문자열을 PT의 P칸에 쓴다. 단, 현문자열에서 이 문자열이 둘 이상 발견되면, ④의 수순을 거친다.
- ④의 수순에 따라, ②와 ③에서 유보된 처리를 명사의 의미역 정보를 이용해 다시 분석한다. ③의 수순에 의해 현문자열에 [n, aff, des] 문자열이나 [n, aff, sour] 문자열이나 [n, aff, dat] 문자열이나 [n, aff, inst] 문자열이 둘 이상이면, [n, aff, dat] > [n, aff, sour] > [n, aff, inst] > [n, aff, des]의 순서로 나열하고, 그 순서에 따라 PT의 P1에서 P4까지 차례대로 쓴다.
- ⑤의 수순에 따라, 현문자열에서 [sub] 문자열을 탐색하여, 그 문자열을 PT의 S에 쓴다. 그리고 현문자열에서 [obj] 문자열을 탐색하여, 그 문자열을 PT의 DO에 쓴다.

지금까지의 '있다' 구문의 격 정보 분석 절차로 다음의 예가 처리되는 과정을 살펴보자.

(20) 가. 학교가 신림동에 있다.

	학교	가	신림동	에	있
자질	n, lex, plc	n, aff, agt	n, lex, plc	n, aff, des	v, lex, v1, v2, Agent, Theme

나. 철수에게 책이 있다.

	철수	에게	책	이	있
자질	n, lex, human	n, aff, dat	n, lex	n, aff, agt	v, lex, v1, v2, Agent, Theme

다. 철수가 책이 있다.

	철수	가	책	이	있
자질	n, lex, human	n, aff, agt	n, lex	n, aff, agt	v, lex, v1, v2, Agent, Theme

- ①의 수순에 따라, 현문자열에서 [n, aff, dat]항이 있는 (20나)와 [n, aff, agt]항이 중출된 (20다)의 서술어 '있'은 능동술어로 선택된다. 그리고 (20가)는 피동술어로 확인된다. 능동술어로 선택된 (20나)와 (20다)의 '있' 문자열의 자질 중에서 [v1, state] 자질을 삭제한다. 그리고 피동술어로 선택된 (20가)의 '있' 문자열의 자질 중에서 [v2, poss, Agent] 자질을 삭제한다.
- ②의 수순에 따라, 현문자열에서 [n, aff, agt]항을 찾아, 그 항을 인접선행하는 [n, lex]항에 [sub] 자질을 추가한다. 그리고 [n, aff, agt]항은 현문자열에서 삭제한다.

- 능동술어의 ③ 수순에서는, (20나)의 [n, aff, dat]를 탐색한다. 탐색된 [n, aff, dat]항의 인접선행 [n, lex]항에 [sub, Agent] 자질을 추가한다. 그리고 ②의 수순에 의해 [sub] 자질이 주어진 [n, lex]항에서 [sub]를 삭제하고 대신 [Theme] 자질을 추가한다.
- 피동술어의 ③ 수순에서는, (20가)의 [n, aff, des]를 탐색한다. 탐색된 [n, aff, des]항을 PT의 P에 쓴다. 그리고 [n, aff, des]항의 인접선행 [n, lex]항을 [n, aff, des]항의 Mod에 쓴다.
- (20다)의 예가 능동술어의 ④ 수순이 필요한 까닭은 현문자열에 [sub]항이 들어 있기 때문이다. 능동술어 ④ 수순에서는, 서술어 ‘있’의 자질이 [d4]임을 확인하고 [sub]항들의 사전을 각각 탐색한 후 소유주와 소유물의 [n, lex]항을 정한다. 소유자로 확인되는 ‘철수’에 [Agent] 자질이 추가된다. 그리고 소유물로 확인되는 ‘책’에 [Theme] 자질이 추가되는 대신 [sub] 자질이 삭제된다.
- ⑤ 수순에서는 현문자열에서 [sub]항을 찾아 그 항을 PT의 S에 쓴다. 그리고 현문자열에서 [obj]항을 찾아 그 항을 PT의 DO에 쓴다. ⑤의 수순까지 끝난 분석 결과의 모습이 아래의 테이블로 정리된다.

(21) 가. 학교가 신림동에 있다.

		S	V		P	
Head		학교	있		에	
Mod					신림동	

나. 철수에게 책이 있다.

		S	V	DO	
Head		철수	있	책	
Mod					

다. 철수가 책이 있다.

		S	V	DO	
Head		철수	있	책	
Mod					

5. ‘있다’ 구문의 영어 생성

이제 (20)의 세 예를 처리하여 PT에 저장된 문자열은, 각각 사전에 들어가 내부의 대역어 선택 조건을 거치게 된다.²⁴⁾ 그리고 아래의 영어 생성 절차를 거친다.

24) 한 어휘항목의 의미적 중의성은 그 어휘항목의 사전 내부에 표상되어 있다. 본 시스템은 한영기계번

(22) 영어의 생성 절차

- ① 대역어 치환
- ② 동사 일치
- ③ 대역어 배열
- ④ 대역어 재배열
- ⑤ 관사 생성

- ①에서는 대역어 선택 조건을 거쳐서 출력된 영어 대역어로 해당 PT의 한국어 문자열을 치환한다.
- ②에서는 PT의 S항의 자질과 MT의 시제 정보를 참고하여 PT의 V항을 일치시킨다.
- ③에서는 PT의 문자열을 다시 텍스트 형식의 문자열로 나열한다.
- ④에서는 MT의 문형 정보를 참고하고 PT의 V항을 참고하여, ③의 결과 나열된 문자열을 재배열한다.
- ⑤에서는 관사를 생성한다.

본 절에서는 4.2절에서 분석한 (20)의 예를 들어 보이기로 한다. 먼저 ①의 수순을 거치면 PT의 한국어 문자열은 아래의 영어 대역어로 치환된다.

(23) 가. 학교가 신림동에 있다.

		S	V		P	
Head		School	be		at	
Mod					Sillimdong	

나. 철수에게 책이 있다.

		S	V	DO	
Head		Cheolsu	have	book	
Mod					

역하기기 때문에 어휘항목의 중의성은 목표어인 영어로 표상되어 있다. 그리고 사진 내부에 문맥에 적합한 대역어를 선택하는 조건이 알고리즘으로 명세화되어 있다. 알고리즘으로 명세화된 대역어 선택 조건은 다음과 같은 문자열 탐색 방식을 취한다.

- (i) 현문자열에서 특정 문자열 탐색
- (ii) 현문자열에서 특정 자질을 가진 문자열 탐색
- (iii) PT의 특정 성분에 해당하는 문자열 탐색
- (iv) PT의 특정 성분에 해당하는 특정 자질의 문자열 탐색

즉, (의미적 중의성을 해결해야 할) 분석문자열이 포함되어 있는 현문자열에서 탐색하는 방법, 구문 분석이 끝난 후 그 결과로 저장된 PT의 정보를 이용하는 방법, 특정 문자열을 직접 탐색하는 방법, 특정 자질을 가진 문자열을 탐색하는 방법에 의하여 위와 같은 네 가지 탐색 방식이 가능하다.

다. 철수가 책이 있다.

		S	V	DO	
Head		Cheolsu	have	book	
Mod					

- ②의 수순에 따라, PT의 S항의 자질을 점검해 본다. 세 예 모두 3인칭의 자질을 가지고 있다. 그리고 MT의 시제 정보를 점검하니 현재시제로 체크되어 있다. 따라서 PT의 V항은 (23가)에서는 'is', (23나)와 (23다)에서는 'has'로 변환한다.
- ③의 수순에 따라, PT의 문자열을 텍스트 형식의 문자열로 나열한다. PT는 좌우 배열과 상하 배열의 형식을 갖추고 있다. 좌우 배열의 경우, 좌측에서 순서대로 문자열을 텍스트로 출력한다. 그리고 상하 배열의 경우, Mod항의 자질을 보아 Head의 앞으로 출력할 것인지, 아니면 뒤로 출력할 것인지를 결정한다.²⁵⁾ (23가)에서만 Mod항이 보이는데, 'Sillimdong'은 'at'의 후행 문자열로 출력된다.

(24) 가. School is at Sillimdong.

나. Cheolsu has book.

다. Cheolsu has book.

- ④의 수순에 따라, MT의 문형 정보를 점검한다. 세 예 모두 평서문으로 확인되어 일단은 문형으로 인한 문자열 이동은 없다. 그런데 (23가)의 예에서 PT의 V항 자질과 S항의 자질을 점검한 결과 'there'가 삽입되어야 할 조건이 발견되면, 'there' 삽입 조건을 거쳐 텍스트로 출력된 문자열을 재배열한다.²⁶⁾

25) 영어에서 대명사나 고유명사가 Mod항이면, 'my' 따위의 소유격으로 변환되거나 'Cheolsu's' 따위의 형식으로 변환된다. 반면에 대부분의 일반명사가 Mod항이면, 'of'를 삽입해야 하는 경우가 있다. 따라서 이 두 경우는 Head항을 중심으로 텍스트로 출력된 방향이 달라진다. 바로 이러한 현상을 염두에 둔 것이다.

26) 영어에서 'there'는 청자에게 어떤 개체의 존재를 알리는 기능을 갖는 존재문에 쓰인다. 영어에는 문장 구성에 있어 어순상 구정보(old information)에 속하는 요소는 상대적으로 앞자리에 오고, 신정보(new information)에 속하는 요소는 뒷자리에 오는 경향성이 있다(Kim, 1985). 따라서 문장의 의미가 존재를 나타내고, PT의 S에 있는 단어가 신정보에 속하는 것이라면 PT의 V보다 뒷자리로 이동해야 하며 빈칸으로 남겨진 S에 'there'를 삽입하게 된다. 그러나 PT의 S에 있는 모든 요소가 이와 같은 과정을 거치는 것이 아니다. 결국 신정보에 속하는 단어가 PT의 S 자리에 있을 때 이 과정을 거치도록 알고리즘을 구성하여야 한다. 그러나 신정보는 화용적 개념이기 때문에 문장에서 그 정체를 판별하기가 상당히 어렵다. 본 기계번역 시스템에서는 신정보가 될 수 없는 요소만 자질로 찾아내어 제약하는 정도의 알고리즘을 구성하였다. 예컨대, 고유명사나 인칭대명사는 이 알고리즘으로 제약될 수 있다.

가. PT의 V항의 자질 중에 [exist]('존재성'의 의미 자질)가 있는지 탐색한다.

나. PT의 S항이 인칭대명사(본 기계번역 시스템에서는 [human, pro])나 고유명사(본 기계번역 시스템에서는 [prp])가 있는지 탐색한다.

(25) There is school at Sillimdong.

- ⑤의 수순에 따라, 일반명사 자질을 가진 'school'과 'book'에만 관사 'a'를 생성한다. 이로써 최종적으로 번역된 영어 문장을 보이면 아래와 같다.

(26) 가. There is a school at Sillimdong.

나. Cheolsu has a book.

다. Cheolsu has a book.

6. 요약 및 남은 문제

지금까지 우리는 한영기계번역기에서 '있다' 구문이 어떻게 처리될 수 있는지를 자질연산문법(FCG)의 체계에 따라 살펴보았다. 김홍규·강범모(2000)에 따르면, '있다'는 전체 용언 가운데 두 번째의 사용 빈도를 보일 정도로 일상 언어 생활에서 자주 쓰이는 단어이어서 실제 한영기계번역의 성공률을 크게 좌우할 수 있다. 그러나 '있다'는 형태론적 패러다임의 특이성과 중의적인 의미 특성으로 인해 한영기계번역을 비롯한 자연언어처리 부분에서 상당히 처리가 까다로운 단어로 알려져 있다. 더구나 이론적인 접근법이 아니라 한영기계번역을 위한 접근법에서는 '있다' 구문이 분석되고 영어 대역어로 생성되는 과정을 하나 하나 선형적인 절차로 보여주어야 하기 때문에 처리의 어려움이 더욱 커진다.

본고에서는 '있다' 구문분석 중에서 격 정보 분석 절차와 그 분석의 결과가 영어로 생성되는 과정을 한영기계번역의 알고리즘으로 보였다. '있다' 구문은 능동성과 피동성을 동시에 가지고 있기에, 이 중의성을 해결하기 위하여 '있다' 구문의 패턴 정보를 이용하였다. 그리고 '가' 조사 중출 구문이면 일반적인 격 정보 분석 절차와는 구별되는 별도의 알고리즘이 수행됨을 보였다. '있다'의 의미역 정보인 [Agent, Theme]를 점검하여 각각의 의미역 자질이 실현되는 문자열을 찾는 과정이 이 알고리즘에 포함된다. 이 때 '가' 조사가 붙은 명사들은 '있다'의 의미 자질에 의해 '소유자-소유물' 관계로 정해지는데, 이 관계는 '가' 조사가 붙은 명사의 사전 정보로 확인할 수 있었다.

그렇지만 본고에서 아직까지 설명하지 못한 부분들이 있다. 이 글에서는 '있다' 구문의 분석에 관한 설명을 영어의 생성에 관한 설명보다 더 자세하게 하였기 때문에 상대적으로 한영기계번역에서 발생할 수 있는 생성의 문제를 여러 각도로 되짚어 보지 못했다. 그리고 '있다'의 구문분석 과정에서도 구조격 분석에만 초점을 맞추었기 때문에 보조사가 붙은 명사 논항의 분석이나 부가어의 분석이 소홀한 느

다. (7)과 (8)의 조건을 만족하면, [v, lex]항에 인접선행하도록 there를 삽입하여 출력하라.
라. PT의 S항은 [v, lex]항에 인접후행하여 출력하라.

김도 있다.²⁷⁾ 이렇게 남은 문제는 본고에서 자세하게 설명하기에는 너무 방대하였기 때문이다. 후고에서 이 문제들을 본격적으로 다룰 것을 다짐하며 글을 맺는다.

참 고 문 헌

- 고석주. (1996). '있다' 구문에 대한 연구. 국어 문법의 탐구 III: 국어 통사론의 문제와 전망. 서울: 태학사, 99-127.
- 고창수. (2002). 자질연산문법이론. 서울: 도서출판 월인.
- 김원경. (2000). 한국어 격 정보와 자질 연산 문법. 고려대 박사학위논문.
- 김차균. (1982). 「있다」의 의미 연구. 언어학 5, 55-82.
- 김홍규·강범모. (2000). 한국어 형태소 및 어휘 사용 빈도의 분석 1. 서울: 고려대 민족문화연구원.
- 성광수. (1976). 존재(동)사 '있다'에 대한 재고. 한국어문논총. <成光秀(2001)에 再錄> 成光秀. (2001). 국어의 단어형성과 의미해석. 서울: 도서출판 월인.
- 유형선. (1999). 이중주격 구문의 논항구조에 대한 연구. 국어의 격과 조사(한국어학회). 서울: 도서출판 월인, 717-732.
- 유혜원. (2002). 국어의 격 교체 구문의 연구: 한영 기계 번역 시스템을 중심으로. 고려대 박사학위논문.
- 윤재학. (1999). '존재동사'에 관하여. 언어 24(1), 87-97.
- 李安九. (2002). '있다'와 '없다'에 대한 通時的 研究. 서울대 석사학위논문.
- 이희승. (1956). 존재사 '있다'에 대하여. 서울대 논문집 3.
- 최승권. (2002). 기계번역의 가능성과 전망. 2002년 인문과학연구소 초청강연회 자료집(성신여대).
- 최정혜. (2003). 한영번역기에서의 주어 복원 문제에 대하여. 언어과학연구 27, 385-406.
- 최호철. (1993). 현대 국어 敍述語의 의미 연구: 義素 設定을 中心으로. 고려대 박사학위논문.
- 최호철 · 홍종선 · 조일영 · 송향근 · 고창수. (1998). 기계 번역을 위한 한국어 논항 체계 연구. 한국어의미학 3, 1-39.
- 한정한. (1999). 의미격과 화용격 어떻게 다른가? 국어의 격과 조사(한국어학회). 서울: 도서출판 월인, 361-392.
- 한정한. (2001). 다국어 기계번역 시스템을 위한 의미 표시 방법. 민족문화연구(고려

27) 한 심사위원이 표면적으로 동일한 듯이 보이는 형태들(예: 예게/한테/더러/보고) 사이에도 용법의 미세한 차이가 있을 수 있는데, 이에 대한 반영 여부를 물었다. 본 연구는 한국어 '있다'를 대상으로 구문을 분석하는 것이 목적이기 때문에 이들의 의미 분석은 반영하지 않았다. 의미 분석은 구문 분석과 다른 독립된 알고리즘을 요구하는 것이고, 그 처리 절차가 매우 복잡하여 본 연구에서 자세히 서술하기에는 무리가 있었기 때문이다. 다른 문제들과 더불어 후고를 기약하도록 한다.

대) 35, 49-71.

한정환. (2002). '가', '를'의 연결이론. 語學研究(서울대) 38(3), 827-850.

홍종선. (1999). 생성문법과 국어의 격. 국어의 격과 조사(한국어학회). 서울: 도서출판 월인, 83-112.

황화상. (2003). 한국어 형태소 분석에서 형태소 분리의 방법에 대하여. 언어정보(고려대 언어정보연구소) 5, 115-137.

Blake, B. J. (1994). *Case*. Cambridge and New York: Cambridge University Press.

Dorr, B. J., Pamela W. J., and W. B. John. (1999). A survey of current research in machine translation. *Advances in Computers* 49, 1-70.

Fillmore, C. J. (1968). The Case for Case. *Universals in Linguistics Theory*. 1-88.

Kim, A. H.-O. (1985). *The Grammar of Focus of Korean Syntax and its Typological Implications*. Dissertation. University of Southern California.

Koenig, J. and A. R. Davis. (2001). Sublexical modality and the structure of lexical semantic representation. *Linguistics and Philosophy* 24, 71-124.

Van Valin, R. D. and R. J. LaPolla. (1997). *Syntax: Structure, Meaning and Function*. Cambridge and New York: Cambridge University Press.

< 부 록 >

아래에 나열하는 자질의 목록은 본 한영기계번역 시스템에서 사용하고 있는 자질을 모두 나열한 것이 아니라 이 글의 논의 과정에서 사용한 자질만 목록화한 것이다.

- (1) 명사 어간 [n, lex]
 - 사람명사 [n, lex, hum]
 - 장소명사 [n, lex, plc]
 - 대명사 [n, lex, pro]
 - 고유명사 [n, lex, prp]

- (2) 동사 어간 [v, lex]
 - 동작동사 [v, lex, act]
 - 상태동사 [v, lex, state]
 - 심리동사 [v, lex, psy]
 - 소유동사 [v, lex, poss]

- (3) 조사 [n, aff]
 - ‘가’류 조사 [n, aff, agt]
 - ‘를’류 조사 [n, aff, thm]
 - ‘의’류 조사 [n, aff, gen]
 - ‘에게’류 조사 [n, aff, dat]
 - ‘에’류 조사 [n, aff, des]
 - ‘로’류 조사 [n, aff, inst]
 - ‘에서’류 조사 [n, aff, sour]

- (4) 어미 [v, aff]
 - 관형형어미 [v, aff, unc]
 - 과거시제 [v, aff, pt]
 - 평서형 종결어미 [v, aff, dcl, fin]

- (5) 조동사 [aux]

이 동 혁

136-701

서울시 성북구 안암동 5가

고려대학교 민족문화연구원 기계번역연구실

전자우편 : metacog@ikc.korea.ac.kr

접수일자 : 2002. 11. 27.

수정본 접수 : 2004. 2. 13.

게재결정 : 2004. 2. 21.