

한영자동번역에서의 관형절 처리

이용훈
충남대학교 영어영문학과
ylee@cnu.ac.kr

류병래
충남대학교 언어학과
ryu@cnu.ac.kr

Processing Noun-Modifying Clauses in Korean-English Machine Translation

Yong-hun Lee
Dept. of English Lang. and Lit.
Chungnam National University
ylee@cnu.ac.kr

Byong-Rae Ryu
Dept. of Linguistics
Chungnam National University
ryu@cnu.ac.kr

요약

한영 자동번역 시스템에서 관형절의 처리는 시스템의 효율에 영향을 미치는 중요한 요소 중의 하나로 인식되어져 왔다. 본 논문의 목적은 기존의 선행 연구들에서는 주목을 받지 못해 왔던 명사보어관형절을 제시함으로써 한국어 관형절의 종류를 재분류하고, 이들 관형절을 처리하는 알고리즘을 자동번역의 관점에서 제시하는 데 있다. 본 논문에서 다루게 될 명사보어관형절은 기존의 관계관형절이나 동격관형절과는 다른 언어학적 특성을 보이고 있으나, 자동번역에서는 그들과 비슷한 알고리즘으로 처리될 수 있음을 보이고자 한다.

1 서론

한국어의 관형절은 한국어문장에서 가장 많이 나타나는 구문형태 중의 하나이고, 형상구조적 언어, 예를 들어 영어를 대상으로 개발된 처리 방법으로는 정확한 분석이 되지 않는 면이 많다. 이러한 이유로 한국어 자연언어처리 분야에서 관형절의 처리는 매우 중요하고 가장 어려운 부분의 하나로 인식되어 온 것이 사실이다. 한영자동번역에서도 한국어 관형절의 처리는 전체 시스템의 효율에 영향을 미치는 중요한 요소로 알려져 있다.

자동번역의 관점에서 볼 때, 관형절을 처리하는 방법에는 분절(segment)의 유무에 따라 두 가지로 나뉜다. 하나는 관형절 부분을 분절하여 별도로 번역한 후 다시 결합하는 방식이고 (윤승, 2001), 다른 하나는 관형절을 분절하지 않고 통사분석기(syntactic analyzer)로 분석하는 방식이다. 전자와는 달리 후자는 성능좋은 통사분석기를 요구한다.

본 논문의 목적은 자연언어처리분야의 선행

연구들에서는 별로 주목을 받지 못해 왔던 또 다른 형태의 관형절인 명사보어관형절을 제시함으로써 관형절의 종류를 재분류하고, 이들 관형절을 처리하는 알고리즘을 자동번역의 관점에서 제시하는 데 있다. 언어학에서는 90년대 초 이후로 명사보어관형절에 대한 연구와 논의가 계속 이루어져 왔으나, 이러한 연구들을 자연언어처리 분야에 적용시켜 보려는 시도는 미약했다. 본 논문에서 제시하는 관형절처리 알고리즘은 이런 명사보어관형절을 포함하며 기본적으로 분절방식을 사용하고 있기 때문에, 통사분석기의 성능이 뛰어나지 못한 시스템에서도 쉽게 응용이 가능하리라고 생각된다.

본 논문은 다음과 같이 구성되어 있다. 2절에서는 관형절의 종류를 관계관형절, 동격관형절, 그리고 명사보어관형절로 나누어 특징을 간단히 살펴본다. 3절에서는 이 세가지 관형절 중 관계관형절과 동격관형절이 자연언어처리에서 어떻게 다루어 질 수 있는지를 살펴본다. 4절에서는 이들 관형절이 어떤 과정을 거쳐 영어로 번역되는지 분석과 번역 알고리즘을 제시하며,

2절에서 살펴본 관형절들이 자동번역에서는 어떻게 처리되는 지를 살펴본다. 5절에서는 본 논문을 요약한다.

2 관형절의 종류와 기존연구의 문제점

2.1 관형절의 정의

한국어의 관형절에는 여러 종류가 존재한다. 다음 문장들을 보자.

- (1) a. 철수가 아름다운 소녀를 보았다.
- b. 철수가 뛰어가는 소녀를 보았다.
- c. 철수가 사과를 먹는 소녀를 보았다.

이 문장들에서 명사구인 ‘소녀를’을 수식하고 있는 관형절은 각각 ‘아름다운’, ‘뛰어가는’, ‘사과를 먹는’이 해당한다.¹ 각 관형절의 용언에 해당하는 (1a)의 ‘아름다운’은 형용사이며, (1b)의 ‘뛰어가는’은 자동사이고, (1c)의 ‘먹는’은 타동사이다. 본 논문에서는 관형절의 범위를 (1b)와 (1c)로 제한한다. 즉, 관형절의 용언이 동사일 경우만을 다루기로 한다. (1a)와 같이 관형절의 용언이 형용사일 경우에는 본 논문에서는 관형절로 보지 않는다. 그 이유는, 용언이 형용사일 경우에는 주로 단독으로 쓰여서 다음에 나오는 명사를 꾸미는 경우가 많기 때문이다. 따라서, 본 논문에서는 (1a)와 같은 문장을 관형절로 보기보다는 ‘형용사 + 명사’로 본다. 이러한 접근법이 가지는 장점은 (1a)와 같은 문장은 (1b)나 (1c)와는 달리 분절을 하지 않으므로, 자연언어 처리 시스템의 효율을 향상시킬 수 있다는 것이다. 또 한가지 문제가 되는 문장들은 의존명사를 포함하고 있는 문장들이다. 다음의 문장들을 보자.

- (2) a. 철수가 가는 것이 바람직하다.
- b. 이 논문의 목적은 관형절의 구조를 밝히는 데 있다.
- c. 철수는 빨리 떨어 수 있다.

이 문장들은 각각 ‘것’, ‘데’, ‘수’ 라는 의존명사를 포함하고 있으며, 각각의 의존명사들을 관형절들이 수식하고 있다. 이러한 문장들을 자동번역 할 경우에는 각각의 관형절들을 분절하여

처리하는 것 보다는 ‘~ㄴ/는 것이 바람직하다’, ‘~ㄴ/는 데 있다’, ‘~ㄴ 수 있다’ 등을 하나의 패턴으로 처리하는 것이 훨씬 효율적이다. 예를 들면, (2가)의 경우에는 ‘~ㄴ/는 것이 바람직하다’를 하나의 패턴으로 처리하여 ‘It is reasonable that ~’으로 번역한 후, ‘철수는 빨리 뛰다’를 번역하여 ‘~’의 자리에 넣어 번역을 완성하는 것이다. 따라서, 본 논문에서는 (2)에 있는 문장들과 같이 의존명사를 포함한 문장들의 경우에는, 관형절을 분절하지 않고 패턴으로 처리하도록 할 것이다.

2.2 관형절의 종류

2.2.1 관계관형절

관계관형절은 한국어문장에서 가장 많이 쓰이는 관형절의 형태 중의 하나이다. (3)의 문장들은 몇가지 예를 보여 주고 있다.

- (3) a. 철수가 사과를 먹는 영희를 보았다.
- b. 철수가 영희가 잃어버린 지갑을 찾았다.
- c. 철수가 마음씨가 착한 영희를 사랑한다.

(3a)는 ‘철수가 영희를 보았다’와 ‘영희가 사과를 먹었다’라는 두개의 문장이 결합되어 이루어진 복문이다. 이 문장에서 ‘사과를 먹는’이 ‘영희’를 수식하는 관계관형절이다. 마찬가지로, (3b)는 ‘영희가 잃어버린’이 ‘지갑’을 수식하는 관계관형절이고, (3c)는 ‘마음씨가 착한’이 ‘영희’를 수식하는 관계관형절이다.

관계관형절의 가장 큰 특징 중의 하나는 수식받는 명사가 관형절 안에서 일정한 역할을 하고 있다는 것이다. 문장 (3a)의 경우, ‘영희’가 문장전체의 술어인 ‘보았다’의 목적어도 되지만, 관형절의 술어인 ‘먹다’의 주어도 된다. 즉, 수식을 받는 명사는 관형절 안에서 하나의 성분으로서의 역할을 하며, 관형절 자체는 그 성분이 비어있게 된다. 이는 관계관형절을 다른 종류의 관형절과 구별하는 가장 큰 특징이 된다. (3b)에서는 ‘지갑’이 관형절과 주절 모두의 목적어로 쓰인다. (3c)의 경우에는 ‘영희’가 문장전체의 술어인 ‘사랑한다’의 목적어이며 동시에 관형절의 술어인 ‘착하다’의 주어도 된다.

¹학자에 따라서는 ‘소녀를’과 같은 구를 ‘명사구’가 아닌 ‘후치사구’로 간주하기도 한다. 이 문제가 본 논문의 범위 밖에 있음은 분명하다. 그러나 어느 견해를 택하더라도 본 논문의 핵심 내용은 적용이 가능한 것은 확실하다.

2.2.2 동격관형절

동격관형절은 관형절의 내용과 관형절에 의해서 수식받는 명사가 동격(apposition)임을 나타내는 것으로, 다음과 같은 예문이 대표적이다.

- (4) a. 철수가 영희가 예쁘다는 사실을 알았다.
 b. 철수가 영희가 아름답다는 얘기를 들었다.
 c. 철수가 하느님이 계시다는 믿음을 갖고 있다.

(4a)에서는, ‘영희가 예쁘다’라는 관형절이 ‘사실’이라는 명사와 동격을 이루고 있다. 즉, ‘사실’의 내용이 ‘영희가 예쁘다’라는 것이다. 이와 비슷하게, (4b)에서는, ‘영희가 아름답다’라는 관형절이 ‘얘기’라는 명사와 동격을 이루고 있고, (4c)에서는, ‘하느님이 계시다’라는 관형절이 ‘믿음’이라는 명사와 동격을 이루고 있다.

관계관형절과 달리 동격관형절이 갖는 특징은 수식을 받는 명사가 관형절 안에서 비어있는 성분이 아니라는 것이다. 즉, 관형절 자체가 하나의 독립된 문장으로, 그 안에는 비어있는 성분이 없다는 것이 가장 큰 특징이다.

2.2.3 명사보어관형절

명사보어관형절은 관계관형절이나 동격관형절과는 다른 통사적 특성을 보인다. 먼저, 다음의 문장을 보자.

- (5) a. 철수가 생선이 타는 냄새를 맡았다.
 b. 철수가 타조가 걷는 모양을 관찰했다.
 c. 철수가 물이 흐르는 소리를 들었다.

(5a)에서는 관형절인 ‘생선이 타는’은 명사인 ‘냄새’를 수식한다. 그래서, 이 문장은 ‘생선이 타다’와 ‘철수가 냄새를 맡았다’라는 두개의 문장으로 이루어진 복문이 된다. 마찬가지로, (5b)에서는 ‘타조가 걷는’이 명사인 ‘모양’을 수식하고 있으며, (5c)에서는 ‘물이 흐르는’이 명사인 ‘소리’를 수식한다. 그런데, 이 문장들의 형태를 보자. 우선, 이 문장들은 관계관형절은 아니다. (5a)의 관형절인 ‘생선이 타다’는 비어있는 문장성분이 없는 하나의 완벽한 하나의 문장이다. 그렇다고 이 문장을 동격관형절이라고 할 수도 없다. 왜냐하면, 관형절인 ‘생선이 타는’과

수식을 받는 명사인 ‘냄새’가 동격의 의미를 가지고 있지 않기 때문이다. 따라서, 이 문장은 관계관형절도 아니고 동격관형절도 아닌 또 다른 관형절의 형태를 갖게 된다. (5b)와 (5c)의 경우도 마찬가지이다. 이러한 형태의 관형절을 명사보어관형절이라 하며, Yoon(1993), Park(1993), Cha(1998, 1999) 등에서는 이러한 명사동격관형절을 언어학적으로 분석하고 있다.

2.3 기존 연구

윤승(2001)에서는 관계관형절과 동격관형절에 대한 복문 분할 알고리즘이 제시되어 있다. 그러나, 이 알고리즘은 명시적으로 제시된 것은 아니며, 분석 예를 통하여 간접적으로 알 수 있다. 이 절에 제시되는 알고리즘은 윤승(2001)에 제시된 예를 기본으로 만들어진 것이다.

2.3.1 관계관형절

먼저 관계관형절을 포함하고 있는 문장 (3a)를 분석해보면 (6)과 같다. 여기서 주의를 기울여 보아야 할 것은 관계관형절 안에 비어 있는 성분을 어떻게 복원하느냐 하는 것이다.

- (3) a. 철수가 사과를 먹는 영희를 보았다.

윤승(2001: 16)의 알고리즘에 의하면, 관형사형 전성어미 또는 종결어미 + ‘(고 하)는’이 발견되면 관형절이 되고, 이런 어미의 다음을 분절점으로 한다. 즉, ‘먹는’ 다음이 분절점이 되는 것이다. (6)에서 보듯이, 문장 (3a)는 Step 1과 같이 분절되어 Step 2와 같이 분할된다.

- (6) 관계관형절의 분할
 Step 1: 철수가 사과를 먹는 & 영희를 보았다.
 Step 2:
 (1) 영희가 사과를 먹었다.
 (2) 철수가 그런 영희를 보았다.

Step 2에서는 다음과 같은 세가지의 작용이 일어난다. 먼저, 관형절의 어미를 종결어미로 바꾼다. 그리고, 관형절이 있던 자리에 ‘그런’을 넣는다. 마지막으로 관형절 안에 비어있는 성분을 복원한다.

이 문장에서 비어있는 성분을 복원하는 방법은 다음과 같다. 먼저, 관계관형절 안에서 술어를 찾는다. (3a)에서는 ‘먹는’이 술어이다. 그 다음 이 동사의 하위범주화사전을 검색한다. 동사

‘먹다’는 타동사이므로 완전한 문장이 되기 위해서는 주어와 목적어가 필요하다. 해당하는 관형절에서 목적어는 ‘사과를’이 되지만, 이 관형절 안에서는 주어가 없다. 따라서, 관형절의 수식을 받는 명사인 ‘영희’에 주격조사 ‘가’를 붙여서 주어를 복원한다.

2.3.2 동격관형절

동격관형절에 대한 분절 알고리즘도 관계관형절과 비슷하다. 그러나 동격관형절에서는 관형절 안에 비어있는 성분이 존재하지 않으므로 비어있는 성분을 복원하는 과정이 필요로 하지 않는 것이 가장 큰 차이점이다. (7)는 동격관형절을 포함하고 있는 문장 (4a)를 분석한 것이다.

- (4) a. 철수가 영희가 예쁘다는 사실을 알았다.
- (7) 관계관형절의 분할
Step 1: 철수가 영희가 예쁘다는 ≍ 사실을 알았다.
Step 2:
(1) 영희가 예쁘다.
(2) 철수가 그런 사실을 알았다.

먼저, 분절점의 정의에 따라 ‘예쁘다는’의 다음이 분절점이 된다. Step 2에서는 관형절의 어미를 종결어미로 바꾸고, 관형절이 있었던 원래 위치에 ‘그런’을 넣음으로써, 문장 (4a)에 대한 복문 분할이 끝나게 된다.

2.3.3 기존연구의 문제점

윤승(2001)의 관형절 분할 알고리즘은 관형절의 원래의 위치를 명시해 주고, 비어있는 성분들을 복원할 수 있는 알고리즘을 제공하고 있다는 점에서 그 의의가 크다고 할 수 있다. 그러나, 이 알고리즘에도 몇가지 문제점이 있다. 본 논문에서는 그 중 두가지를 지적하고자 한다.

가장 먼저 지적될 수 있는 것은 2.2.3절에서 본 명사보어관형절은 윤승(2001)의 알고리즘으로는 (5)와 같은 문장들을 분석해낼 수 없다.

- (5) a. 철수가 생선이 타는 냄새를 맡았다.
b. 철수가 타조가 걷는 모양을 관찰했다.
c. 철수가 물이 흐르는 소리를 들었다.

(5)는 관계관형절도 동격관형절도 아니기 때문에 2.3.1이나 2.3.2에 제시된 알고리즘으로는 처리할 수 없다.

두번째 문제점은 둘 이상의 관형절이 한 문장안에 존재할 때에 일어난다. 다음의 문장을 보자

- (8) 사과를 먹는 철수는 영희가 예쁘다는 사실을 알았다.

이 문장은 두개의 관형절을 포함하고 있다. 하나는 ‘사과를 먹는’이라는 관계관형절이고, 다른 하나는 ‘영희가 예쁘다’라는 동격관형절이다. 이 문장은 윤승(2001)에 따르면 다음과 같이 분절된다.

- (9) 문장(8)의 분절
Step 1: 사과를 먹는 ≍ 철수는 영희가 예쁘다는 ≍ 사실을 알았다.
Step 2:
(1) 철수가 사과를 먹었다.
(2) 그런 철수가 영희가 예쁘다는 ≍ 사실을 알았다.
Step 3:
(1) 철수가 사과를 먹었다.
(2) 영희가 예쁘다.
(3) 그런 철수가 그런 사실을 사실을 알았다.

먼저 관계관형절인 ‘사과를 먹는’이 먼저 분절되어 어미가 종결어미로 바뀌고, 주어인 ‘철수가’가 복원된다. 관형절이 있었던 자리에는 ‘그런’이 삽입된다. 다음은 동격관형절인 ‘영희가 예쁘다는’이 분석된다. 그 결과 이 동격관형절이 있던 자리에도 ‘그런’이 삽입된다. 문제는 분석결과에서 ‘그런’이 두번 나오게 되는데, 어느 ‘그런’이 어느 관형절을 가리키는 지를 알 수가 없다는 것이다. 즉, 분절의 과정에서 정보의 손실(information loss)이 생긴 것이다. 이런 문제점은 관계관형절과 동격관형절이 섞여 있는 문장들뿐만 아니라, 관계관형절이나 동격관형절이 한 문장에서 여러 번 나타날 때에도 같은 문제가 발생한다.

3 관형절의 자동번역

이 절에서는 2절에서 본 세가지의 관형절에 대한 알고리즘을 한영자동번역의 관점에서 제시한다. 이 절에서 제시되는 알고리즘은 2.3.3에

서 논의된 윤승(2001)의 문제점을 보완한 것이다.

3.1 기술규약과 기본 알고리즘

2.3.3절에서 지적된 문제점인 분절시의 정보의 손실을 최소화하기 위해서 우리는 관형절의 위치를 나타내는 표지와 함께 관형절의 종류를 인덱스로 기술할 수 있다. 먼저, 우리가 다루게 될 관형절의 종류는 다음과 같다.

(10) 관형절의 종류

- a. 관계관형절 (Relative Clause)
- b. 동격관형절 (Appositive Clause)
- c. 명사보어관형절 (Noun Complement clause)

따라서, 관형절의 종류를 기술할 때에는 영문 이름의 첫자만을 따서 R, A, N으로 각각 기술한다. 한 문장안에 같은 종류의 관형절이 두 번 이상 나오는 경우를 위하여 관형절의 종류 뒤에 번호를 붙여 구분한다. 즉, 관계관형절은 R_1, R_2, \dots, R_n 순으로 기술하고, 동격관형절은 A_1, A_2, \dots, A_n 순으로 기술하며, 명사보어관형절은 N_1, N_2, \dots, N_n 순으로 기술한다.

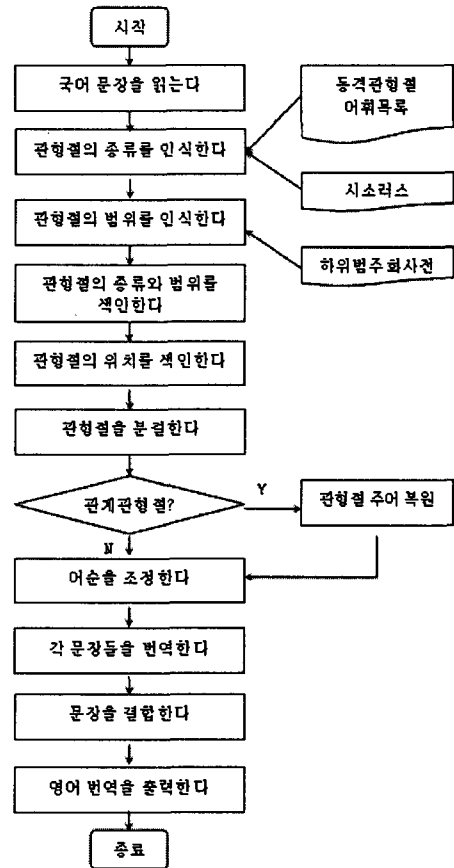
다음으로서는 관형절의 시작과 끝, 그리고 관형절이 들어갈 위치에 대한 규약은 다음과 같다.

(11) 기술규약

- a. 관형절부분
 - 관계관형절 - $R_n: [\dots]$
 - 동격관형절 - $A_n: [\dots]$
 - 명사보어관형절 - $N_n: [\dots]$
- b. 관형절이 들어갈 부분
 - 관계관형절 - $\dots [R_n] \dots$
 - 동격관형절 - $\dots [A_n] \dots$
 - 명사보어관형절 - $\dots [N_n] \dots$

대괄호로 관형절의 시작과 끝을 각각 나타내고, 관형절의 원 위치는 $[R_n], [A_n], [N_n]$ 과 같이 나타내며, 번호는 관형절의 번호를 나타낸다. 본 논문에서 제시하는 알고리즘의 전체 흐름도는 (12)와 같다.

(12) 전체 흐름도



먼저, 번역해야 할 한국어문장을 읽어 들인 후, 그 문장에서 관형절을 찾아내고, 어휘목록과 시소러스를 활용하여 그 관형절의 종류를 결정한다. 다음은, 관형절의 술어를 찾아내어 그 술어를 중심으로 하위범주화사전을 탐색하여 관형절의 범위를 결정한다. 관형절의 종류와 범위가 결정되면, (11a)의 기술규약을 사용하여 해당 관형절에 색인을 달아 놓고 관형절을 분절한다. 이때, 관형절이 있었던 원래의 자리에는 (11b)에 있는 기술 규약을 사용하여 위치를 색인하여 둔다. 만약, 분절된 관형절이 관계관형절인 경우에는 관형절 안에 비어있는 성분을 찾아서 복원하는 과정이 필요하므로 이를 수행한다. 그 다음으로서는, 어순을 조정한다. 한국어문장에서는 관형절이 수식하는 명사보다 선행하지만 영어로 번역된 문장에서는 관형절이 명사의 다음에 나오게 되므로, 수식을 받는 명사와 관형절의 위치 — 여기서는, 관형절의 위치정보 — 를 바꾸어준다. 그 다음, 분절된 문장들을 각각 영어로 번역한 후, 색인된 정보를 이용하여

번역된 문장을 결합하여 문장을 완성한다.

관형절을 분절할 때에 일어날 수 있는 정보의 손실을 최소화 시키기 위하여, 우리는 분절하기 이전에 (11)의 기술규약을 사용할 것이며, 이 정보를 사용하여 번역된 문장을 복원할 것이다. 즉, 번역된 주절에 있는 $[R_n]$, $[A_n]$, $[N_n]$ 의 위치에 관형절인 $R_n: [\dots]$, $A_n: [\dots]$, $N_n: [\dots]$ 의 내용을 번역하여 치환하여 번역을 완성한다.

3.2 관계관형절

관계관형절도 관형절의 하나이므로 (12)에서 열거한 알고리즘에 따라 번역된다. 그러나, 관계관형절에서 추가하여 고려해야 될 것은, 2.2.1절에서 논의한 바와 같은 비어있는 성분을 복원하는 알고리즘이다. (13)은 관계관형절이 포함되어 있는 문장 (3a)가 어떻게 자동번역되는 지를 보여주고 있다.

(3) a. 철수가 사과를 먹는 영희를 보았다.

(13) 관계관형절의 자동번역

Step 1: 철수가 R_1 : [사과를 먹는] 영희를 보았다.

Step 2: 철수가 R_1 : [사과를 먹는] \asymp 영희를 보았다.

Step 3:

(1) R_1 : [영희가 사과를 먹었다.]

(2) 철수가 $[R_1]$ 영희를 보았다.

Step 4:

(1) R_1 : [영희가 사과를 먹었다.]

(2) 철수가 영희를 $[R_1]$ 보았다.

Step 5:

(1) R_1 : [who ate an apple.]

(2) Chelsoo saw Younghee $[R_1]$.

Step 6: Chelsoo saw Younghee, who ate an apple.

관계관형절의 자동번역은 다음과 같이 진행된다. Step 1에서는, ‘사과를 먹는’이 관형절이므로, 이 관형절의 처음과 끝을 ‘[’와 ‘]’로 각각 표시하고, 관형절의 종류와 번호를 ‘ R_1 :’과 같이 붙여준다. Step 2에서는, 관형절 다음을 분절점으로 하여 문장을 분절한다. Step 3에서는, 관형절을 주절로부터 분리해낸다. 이때, 관형절에서는 주어인 ‘영희가’를 복원하고, 관형절이 있었던 주절의 원래 자리에는 표지로 $[R_1]$ 을 표시해 둔다.

Step 4에서는, 표지인 $[R_1]$ 의 위치를 바꾸어 준다. 위 (13)에서 알 수 있듯이, Step 3에서는 $[R_1]$ 이 수식을 받는 ‘영희’의 앞에 있지만, Step 4에서는 $[R_1]$ 이 ‘영희’의 뒤에 놓여 있다. 이것은 번역되기 전의 한글문장과 번역된 후의 영어문장의 어순상의 차이를 반영한 것이다. 즉, 한국어에서는 핵심어(head)가 보어의 뒤에 오지만, 영어에서는 핵심어가 보어의 앞에 온다. Step 3과 Step 4의 사이에 일어나는 작용은 이러한 어순상의 차이점을 반영한 것이다. Step 5에서는, 주절과 관형절을 각각 영어로 번역한다. Step 6에서는, 관형절과 주절을 결합하여 번역을 완성한다. 이때, 주절의 $[R_1]$ 의 자리에 R_1 : $[\dots]$ 을 치환해 넣음으로써 번역을 완성한다.

3.3 동격관형절

동격관형절도 또한 관형절이므로 (12)에서 열거한 절차 값아 번역된다. 그러나, 관계관형절과는 달리 비어있는 성분을 복원하는 알고리즘이 필요로 하지 않는다. (14)는 동격관형절이 포함되어 있는 문장 (4a)가 어떻게 분석되는 지를 보여주고 있다.

(4) a. 철수가 영희가 예쁘다는 사실을 알았다.

(14) 동격관형절의 자동번역

Step 1: 철수가 A_1 : [영희가 예쁘다는] 사실을 알았다.

Step 2: 철수가 A_1 : [영희가 예쁘다는] \asymp 사실을 알았다.

Step 3:

(1) A_1 : [영희가 예쁘다.]

(2) 철수가 $[A_1]$ 사실을 알았다.

Step 4:

(1) A_1 : [영희가 예쁘다.]

(2) 철수가 사실을 $[A_1]$ 알았다.

Step 5:

(1) A_1 : [Younghee is pretty.]

(2) Chelsoo knew the fact that $[A_1]$.

Step 6: Chelsoo knew the fact that Younghee is pretty.

관계관형절의 자동번역은 다음과 같이 진행된다. Step 1에서는, ‘영희가 예쁘다는’이 관형절이므로, 이 관형절의 처음과 끝을 ‘[’와 ‘]’로 각각 표시하고, 관형절의 종류와 번호를 ‘ A_1 :’과 같이 붙여준다. Step 2에서는, 관형절 다음을 분절점으로 하여 문장을 분절한다. Step 3에서는, 관

형절을 주절로부터 분리해낸다. 이때, 관형절이 있었던 주절의 원래 자리에는 표지로 [A₁]을 표시해준다.

Step 4에서는, 표지인 [A₁]의 위치를 바꾸어 준다. Step 5에서는, 주절과 관형절을 각각 영어로 번역하여 준다. Step 6에서는, 관형절과 주절을 결합하여 번역을 완성한다. 이때, 주절의 [A₁]의 자리에 A₁:[...]을 치환해 넣음으로써 번역을 완성한다.

3.4 명사보어관형절

명사보어관형절도 역시 관형절이므로 (12)에서 열거한 과정을 밟아 번역된다. 명사보어관형절의 알고리즘은 기본적으로 동격관형절과 같다. (15)은 명사보어관형절이 포함되어 있는 문장 (5a)가 어떻게 분석되는지를 보여주고 있다.

(5) a. 철철수가 생선이 타는 냄새를 맡았다.

(15) 명사보어관형절의 자동번역

Step 1: 철수가 N₁:[생선이 타는] 냄새를 맡았다.

Step 2: 철수가 N₁:[생선이 타는] ≪ 냄새를 맡았다.

Step 3:

(1) N₁:[생선이 탄다.]

(2) 철수가 [N₁] 냄새를 맡았다.

Step 4:

(1) N₁:[생선이 탄다.]

(2) 철수가 냄새를 [N₁] 맡았다.

Step 5:

(1) N₁:[a fish is burning.]

(2) Chelsoo has the smell that [N₁].

Step 6: Chelsoo has the smell that a fish is burning.

명사보어관형절의 자동번역은 다음과 같이 진행된다.

Step 1에서는, ‘생선이 타는’이 관형절이므로, 이 관형절의 처음과 끝을 ‘[’와 ‘]’로 각각 표시하고, 관형절의 종류와 번호를 ‘N₁:’과 같이 붙여준다. Step 2에서는, 관형절 다음을 분절점으로 하여 문장을 분절한다. Step 3에서는, 관형절을 주절로부터 분리해낸다. 이때, 관형절이 있었던 주절의 원래 자리에는 표지로 [N₁]을 표시해준다.

Step 4에서는, 표지인 [N₁]의 위치를 바꾸어 준다. Step 5에서는, 주절과 관형절을 각각 영

어로 번역하여 준다. Step 6에서는, 관형절과 주절을 결합하여 번역을 완성한다. 이때, 주절의 [N₁]의 자리에 N₁:[...]을 치환해 넣음으로써 번역을 완성한다.

3.5 관형절의 구분

지금까지 한국어의 세가지 관형절에 대해 살펴 보았다. 그러나, 아직 언급되지 않은 문제가 남아있다. 그것은 어떻게 세가지의 관형절을 자동번역에서 구분해낼 수 있는냐는 것이다.

우선, 관계관형절을 다른 종류의 관형절로부터 구분해 내는 것은 그리 어렵지가 않다. 관형절의 술어를 찾아내고, 그 술어의 하위범주화사전을 검색하여, 술어를 중심으로 관형절 안에 비어있는 성분이 있으면 관계관형절이고, 비어있는 성분이 없는 경우에는 관계관형절이 아니다.

그러나, 동격관형절과 명사보어관형절을 구분해 내는 것은 그렇게 쉽지는 않다. 동격관형절에서 명사보어관형절을 구분해 낼 수 있는 근거는 수식을 받는 명사에서 찾을 수 있다. 동격관형절의 경우에는 수식받는 명사가 ‘사실’이나 ‘믿음’ 등의 몇가지 명사로 제한된다. 따라서, 이러한 어휘적 차이를 단서로 잡을 수 있을 것이다. 이에 비하여, 명사보어관형절의 경우에는, 관형절에 의해 수식을 받는 명사가 다음과 같이 주로 인간의 오감과 관련이 있는 명사들이다.

- (5) a. 철수가 생선이 타는 냄새를 맡았다.
b. 철수가 타조가 걷는 모양을 관찰했다.
c. 철수가 물이 흐르는 소리를 들었다.

따라서, 명사보어관형절의 경우에는, 시소러스(서영훈 외, 1998)를 사용하여 인간의 오감과 관련이 있는 명사들을 가려낼 수 있다. 이와 같이, 하위범주화사전, 시소러스, 또는 어휘적 단서를 사용하여 세가지 종류의 관형절을 각각 구분해 낼 수 있다.

4 결론

한국어의 관형절처리는 자동번역의 효율을 결정하는 중요한 요소 중의 하나이다. 본 논문에서는 관형절의 종류를 관계관형절, 동격관형절, 명사보어관형절로 구분하고, 그 특징을 간단히 살펴 보았다. 그리고, 윤승(2001)에서는 관형절

을 어떻게 처리하는 지를 살펴본 후, 이 알고리즘을 바탕으로 한국어의 세가지 관형절을 영어로 자동으로 번역하는 알고리즘을 제시했다. 우리는 세가지의 관형절을 정보의 손실을 최소화하며 효율적으로 자동번역할 수 있음을 알았다.

References

- Cha, Jong-Yul. 1998. 'Relative Clause or Noun Complement Clause: Some Diagnoses'. Selected Papers from the 11th International Conference on Korean Linguistics, eds. by Byung-Soo Park & James Hye-Suk Yoon, 73-82.
- Cha, Jong-Yul. 1999. 'Semantics of Korean Gapless Relative Clause Constructions'. *AS-studies in the Linguistic Sciences*, Vol. 29, No. 1, 25-41, University of Illinois at Urbana-Champaign.
- Park, Byung-Soo. 1993. 'A Lexical Approach to Korean Relative Clause Constructions'. Paper Presented at the International Workshop on HPSG. The Ohio State University.
- Yoon, Jae-Hak. 1993. 'Different Semantics for Different Syntax'. *OSU Working Papers in Linguistics*, Vol. 42, The Ohio State University. 73-82.
- 서영훈 외. 1998. '토큰 기반 한국어 분석기 개발 - 한국어 의미 분석 사전 및 하위범주화 사전 구축'. 한국전자통신연구소.
- 윤승. 2001. '한국어 복합문 분할방안 연구'. 연세대학교, 석사학위논문.