

한영자동번역에서의 ‘~ㄴ것은’의 처리: 특허문서를 중심으로

이용훈
(충남대학교 영어영문학과)
ylee@cnu.ac.kr

이숙의
(충남대학교 국어국문학과)
jayou@cnu.ac.kr

류병래
(충남대학교 언어학과)
ryu@cnu.ac.kr

Processing '~n.ket.un' in Machine Translation: A Case Study of Patent Documents

Yong-Hun Lee
(Dept. of English Lang. & Lit.,
Chungnam National University)
ylee@cnu.ac.kr

Suk-Eui Lee
(Dept. of Korean Lang. & Lit.,
Chungnam National University)
jayou@cnu.ac.kr

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

요 약

한영자동번역에서 ‘~ㄴ것은’을 포함하는 문장들은 영어로 번역하기가 용이하지 않을 뿐만 아니라 번역할 때에 많은 오류들을 야기시킨다. 이것은 이러한 문장들이 한국어 원문에 ‘것’과 같은 의존명사를 포함하고 있기 때문인데, 이러한 의존명사들이 한영자동번역을 어렵게 만들고 있다. 본 논문에서는 ‘~ㄴ것은’을 포함하는 문장들을 한영자동번역에서 처리할 수 있는 방법을 제시하고자 한다. 자동번역에는 여러 방식이 있지만, 본 논문에서는 패턴기반 자동번역의 방식을 취한다. 따라서 본 논문에서는 ‘~ㄴ것은’에 대한 한영대역패턴을 어떠한 방식으로 구축하며, 또 구축된 패턴들이 어떻게 한영자동번역에 활용되는 지를 살펴볼 것이다.

1. 서 론

오늘날 과학기술이 발전하고 시장이 세계화됨에 따라 자동번역의 역할과 중요성이 증가하고 있다. 한국어와 영어 사이의 자동번역에서는 일반적으로 한영자동번역이 영한자동번역보다는 훨씬 더 어려운 것으로 알려져 있다. 한영자동번역을 어렵게 만드는 요소들에는 여러 가지가 있겠지만, 그 중에 하나는 ‘~ㄴ것은’과 같이 국어의 의존명사가 조사와 결합되면서 여러 가지 번역상의 문제들을 일으키는 것도 그 중의 하나다.

본 논문에서는 ‘~ㄴ것은’을 중심으로 이러한 문장들을 한영자동번역에서 처리할 수 있는 방법을 제시하고자 한다. 본 논문에서는 패턴기반 자동번역의 방식을 채택하여 ‘~ㄴ것은’에 대한 한영대역패턴을 구축하고, 이를 한영자동번역에 활용하는 방법을 제시하고자 한다.

2. 자동번역의 방식

일반적으로 자동번역의 방식은 크게 규칙기반, 패턴기반, 예제기반, 확률기반 등 네 가지로 나눌

수 있다.¹ 규칙기반 자동번역은 각각의 언어가 가지고 있는 문법규칙을 번역에 활용하여 원시언어(source language)를 목표언어(target language)로 번역하는 방식이다. 규칙기반 자동번역에는 원시언어와 목표언어의 단어와 단어를 직접 매핑시키는 직접방식(direct method)과 통사구조나 중간언어를 이용하여 자동번역을 수행하는 간접방식(indirect method)이 있다.

패턴기반 자동번역은 원시언어와 목표언어 간의 대응관계를 대역(對譯)패턴으로 표현하고, 그 대역패턴에 따라 번역을 수행하는 방식이다. 따라서, 이 방식은 패턴의 수가 증가할수록 번역 성능이 향상되는 특징이 있다.

예제기반 자동번역은 말뭉치(corpus)를 번역에 이용하는 방식이다. 이 방식의 원시언어의 입력문과 가장 유사한 문장을 원시언어 말뭉치에서 찾아 그 문장이 목표언어로 어떻게 번역되는지를 살펴보고 입력문을 최대한 자연스럽게 번역한다.

확률기반 자동번역은 이중언어말뭉치(bilingual corpus)를 분석하여 통계정보를 획득한 후, 이 통계정보를 이용하여 번역을 수행하는 방식이다.

본 논문에서는 이 4가지의 번역방식 중 패턴기반 자동번역 방식을 취한다. 이 방식의 가장 큰 장점중의 하나는 언어구조가 패턴에 반영되어 있기 때문에 개발과 관리가 용이하다는 것이다.

3. 패턴방식에서 ‘~고’ 처리의 문제점

패턴기반 자동번역은 원시언어와 목표언어 간의

1. 자동번역에 대한 개괄적인 설명은 Jurafsky & Martin (2000)의 Chapter 21에 제시되어 있다. 이 절에서 설명되어 있는 방법들은 최승권 (2004), 최승권 & 홍문표 (2005), Somers (2003) 등에도 잘 나와 있다. 특히, 최승권 (2004)과 최승권 & 홍문표 (2005)는 패턴기반 자동번역의 방법이 자세히 설명되어 있으며, Somers (2003)에는 예제기반 자동번역의 방법이 자세히 설명되어 있다.

대응관계를 대역패턴으로 표현한다. 예를 들어, 대등적 연결어미 ‘~고’는 다음과 같은 패턴으로 나타낼 수 있다.²

- (1) 등위적 연결어미 ‘~고’의 대역패턴

[KEY]

고:2_다:1

[CONTENT]

{ VP1[] VP2[] > VP1[ECONJ:[eroot := [COMMA_O_and]]] VP2 }

대역패턴은 크게 [KEY]와 [CONTENT]의 두 부분으로 되어 있다. [KEY]에는 대등적·보조적 연결어미 ‘-고’가 종결어미인 ‘~다’와 결합되어 key 값을 형성한다. Key값에 나오는 숫자는 의존문법(Dependency Grammar)의 의존깊이를 나타낸다. 종결어미인 ‘~다’의 의존깊이는 1값을 가지며, 연결어미들은 2를 값으로 갖는다. [CONTENT]에는 대역패턴이 기술된다. 이 대역패턴은 ‘>’를 중심으로 두 부분으로 나뉘는데, 앞부분은 한국어원문에 적용되는 조건이 기술되어 있으며, 뒷부분에는 해당 연결어미가 영어로 어떻게 번역되는지가 기술되어 있다. 여기서 VP1은 ‘~고’ 앞에 오는 문장을 가리키며, VP2는 ‘~다’ 앞에 오는 문장을 가리킨다. 즉, 국어의 원문이 ‘VP1고 VP2다’의 형태로 되어있다는 것을 나타낸다. 영문 대

2. 최승권 & 홍문표 (2005)의 표기법에 따르면, (1)을 (i)과 같이 기술해야 한다.

(i) 고:2_다:1 | { VP1[] VP2[] > VP1[ECONJ:[eroot := [COMMA_O_and]]] VP2 }

즉, [KEY]값과 [CONTENT]의 사이를 ‘|’로 구분하고 있다. 이는 표기상의 편의를 위한 것이지만, (1)과 (i) 사이에 의미적인 차이는 없다. 본 논문에서 (1)과 같은 표기법을 사용하는 이유는, ‘~고’의 경우는 대역패턴이 ‘~고’처럼 하나의 패턴으로 기술되는 것이 아니라 여러 개의 패턴이 같이 기술된다. 따라서, 패턴기술의 편의를 위하여 편의상 (i)의 표기법보다는 (1)의 표기법을 따르기로 하겠다.

역부의 ECONJ는 원문부의 연결어미를 번역하기 위해 VP1의 마지막에 접속사가 부가되어야 함을 나타낸다. 그러나 연결어미의 번역을 위해 VP1의 처음에 접속사가 부가되어야 함을 나타낼 때에는 SCONJ를 사용하게 된다. 이 패턴은 ‘~고’의 영어 번역(eroot)은 COMMA_O_and가 되어야 함을 나타낸다. 여기서, COMMA_O는 optional comma를 나타낸다

만약, (2)와 같은 문장이 있다면, (1)의 패턴에 따라 (3)과 같이 분석되어 (4)처럼 번역된다.

(2) 철수는 사과를 먹고 영희는 배를 먹었다.

(3) [철수는 사과를 먹]고 [영희는 배를 먹]었다.

(4) [Chelsoo ate an apple], and [Younghee ate a pear].

(3)에서, ‘철수는 사과를 먹’이 VP1에 해당하고, ‘영희는 배를 먹었’이 VP2에 해당한다.

이제 패턴기반 자동번역을 ‘~ㄴ것은’에 적용해보자. 먼저, 다음의 예문을 보자.

(5) 가. 도 14A의 변환패턴에 해당하는 것은 S104와 S105 이다.
나. 어느 한 코일이 방향을 달리하는 것은 본 발명의 범위에 해당된다.

우선 ‘~은것은/는것은’을 ‘~ㄴ것은’의 다른 형태로 본다면 이 문장들은 다음과 같이 분석된다.

(6) 가. [도 14A의 변환패턴에 해당하는 것]은 [S104와 S105 이]다.
나. [어느 한 코일이 방향을 달리하]는 것은 [본 발명의 범위에 해당된]다.

이제 이 문장들을 영어로 자동번역해야 된다.

그러나, ‘~ㄴ것은’과 같은 어구를 포함하고 있는 문장들을 패턴기반 방식으로 자동번역할 때에는 일반적으로 다음과 같은 문제점들이 발견된다.

첫째는 문법구조의 불일치로, 한국어의 구조와 영어 구조의 차이로 인한 불일치이다. 이는 두 언어의 어순을 포함하여 여러 문법 현상들이 포함된다. 예를 들면, 한국어에는 문장의 주어가 종종 생략되는 반면 영어에는 반드시 주어가 존재해야 한다. 따라서, 한국어의 원문에 주어가 없더라도 영어 대역문에는 주어를 복원해야 하는 경우가 발생한다.

둘째는 문법단위의 불일치이다. 이는 한국어의 의존명사 때문에 발생한다. 자동번역에서는 원시언어의 문법단위와 목표언어의 문법단위가 일치하여 서로 일대일 대응이 되는 것이 이상적이다. 그러나, 한국어에는 의존명사가 있는 반면, 영어에는 이에 해당하는 문법단위가 없다. 이런 문법단위의 불일치가 한영자동번역을 어렵게 만든다.

셋째, 한국어 의존명사가 나타내는 이러한 특이성 때문에 이를 영어로 번역하면 그 번역이 다양하게 나온다는 것이다. 따라서, 소수의 패턴으로는 한국어 원문에 의존명사를 포함한 문장들을 영어로 자동번역하기가 매우 어려운 것이 현실이다.

이제, ‘~ㄴ것은’에 대해 생각해 보자. 우선, ‘~ㄴ것은’에 대한 대역패턴을 만들기 위해서는 먼저 ‘~ㄴ것은’이 어떻게 번역되는지를 생각해 보아야 한다. 직관적으로 볼 때에는 한국어의 ‘것’에 해당하는 영어표현은 ‘thing’이다. 그런데, (6가)의 경우에는 thing으로 번역을 해도 큰 문제는 없겠지만, (6나)의 경우에는 ‘thing’으로 번역할 수가 없다. 왜냐하면, (6나)의 ‘달리하는 것’은 사물을 지칭하는 것이 아니라 동작 자체를 지칭하는 것이기 때문이다. 따라서, ‘thing’보다는 to-부정사로 번역하는 것이 더 자연스럽다. 이러한 번역상의 문제가 ‘~ㄴ것은’이 가진 첫 번째 문제이다.

두 번째 문제는, 대등적 연결어미 ‘~고’가

쓰인 문장은 VP1과 VP2가 독립되어 각각 번역될 수 있었지만, ‘~ㄴ것은’의 경우에는 VP1과 VP2가 독립되어 있지 못하다는 것이다. 왜냐하면, (6)의 문장들과 같은 경우 VP1이 VP2의 주어역할을 하고 있기 때문이다.

‘~ㄴ것은’이 가진 세 번째 문제점은 보조사인 ‘~은/는’이 가진 특성에 기인이다. ‘~은/는’은 보조사이기 때문에 (6)의 문장들과 같이 주격조사인 ‘~이/가’를 대신할 수도 있지만, (7)의 문장들과 같이 ‘~을/를’을 대신할 수도 있다.

- (7) 가. 의사가 각 환자에 대한 정보를 열람하는 것은 허용하도록 한다.
 나. 하나로 대두되므로 렌즈를 이용하는 것은 지양하여야 한다.

(6)과 (7)의 문장들을 번역해보면 ‘~은/는’이 주격조사인 ‘~이/가’를 대신할 경우에는 영어대역에서 VP1 VP2의 어순을 유지하는 것이 자연스럽지만, ‘~은/는’이 목적격조사인 ‘~을/를’을 대신할 경우에는 VP1과 VP2의 어순이 바뀌는 것이 자연스러운 번역이 나온다는 것을 알 수 있다.

‘~ㄴ것은’을 패턴기반 자동번역방식으로 처리할 때에 나오는 이러한 문제점들은 의존명사인 ‘것’이 가지는 속성과 조사인 ‘~은/는’이 지닌 고유한 성질 때문이다. 그리고, 이 두 가지 요소(의존명사와 조사)는 한국어에는 존재하지만 영어에는 없는 문법단위들이다.

4. ‘~ㄴ것은’을 위한 패턴 구축

4.1. 패턴구축 방법의 개요

우리가 위에서 살펴본 것과 같이 ‘~ㄴ것은’은 그 자체가 지닌 특성 때문에 한영자동번역에서 이를 처리하기가 쉽지 않다. 따라서, 이에 대한 대역패

턴을 만드는 일 또한 쉬운 일은 아니다. 그러나, 4절에서는 ‘~ㄴ것은’이 지닌 특성을 대역패턴에 반영하고, 이를 자동번역에 활용하는 방법을 제시하고자 한다.

본 논문에서는 이를 위하여 다음과 같은 방법을 취하였다. 먼저, 대규모의 특허문서 말뭉치에서 키워드인 ‘~ㄴ것은’이 들어가는 문장들을 모두 추출하였다. 그런 다음, ‘~ㄴ것은’을 중심으로 VP1과 VP2의 용언의 특징들을 검토하고, 이러한 특징들을 대역패턴의 한국어 원문부에 조건으로 반영하였다. 한국어 원문부 VP1에는 다음과 같은 조건을 사용하였다.

- (8) 국어 원문부 VP1에 사용된 조건
 가. VP1 용언의 품사
 나. 주어가 명시적으로 나타나 있는가?
 다. 목적어가 명시적으로 나타나 있는가?

이를 다음과 같은 표기법으로 나타내었다.

- (9) 국어 원문부 VP1에 사용된 표기법
 가. VERB:[kpos==[ADJ|VI|VT]]
 나. _NOTEXIST(S), _EXIST(S)
 다. _NOTEXIST(O), _EXIST(O)

여기서 (9)의 VERB는 동사를 가리키는 것이 아니라 VP1의 용언을 가리킨다. kpos는 한국어의 품사를 나타내며, 그 값은 ADJ(형용사), VI(자동사), VT(타동사) 중의 하나이다.

한국어 원문부 VP2에 나타난 용언들은 주로 연어(collocation)를 중심으로 파악하고, 이를 다음과 같이 kroot로 나타내기로 하겠다.

- (10) 국어 원문부 VP2에 사용된 표기법
 VERB:[kroot==[용언의원형]]

예를 들어, ‘~것은 가능하다’는 VP2[VERB:[kroot

==[가능하]])와 같이 나타내기로 하겠다.

본 논문에서 특허문서를 말뭉치로 선택한 이유는 다음과 같다. 첫째, 특허문서는 문장의 길이가 길고 복잡하여 패턴기반 자동번역방식이 효율적이다. 둘째, 특허문서는 ‘~ㄴ것은’과 같이 의존명사를 포함하고 있는 문장이 많아 그 특성을 파악하기 쉽다. 셋째, 자동번역의 번역을 한 영역인 특허문서에 한정시킴으로써 번역율의 향상을 꾀할 수 있다.

4.2. VP1의 특성과 패턴처리

위 4.1의 (8)에서 제시된 조건을 사용하여 조합이 가능한 모든 경우의 수를 나열하면 다음과 같다.

- (11) VP1에 적용되는 조건의 결합
- 가. 용언이 형용사, 주어 무
 - 나. 용언이 형용사, 주어 유
 - 다. 용언이 자동사, 주어 무
 - 라. 용언이 자동사, 주어 유
 - 마. 용언이 타동사, 주어 무, 목적어 무
 - 바. 용언이 타동사, 주어 무, 목적어 유
 - 사. 용언이 타동사, 주어 유, 목적어 무
 - 아. 용언이 타동사, 주어 유, 목적어 유

따라서, ‘~ㄴ것은’에 대한 대역패턴도 이러한 조건에 따라 구축하기로 하겠다.

먼저, (11가)는 VP1 용언의 품사가 형용사이고, 주어가 없는 경우로 (12가)와 같은 문장이 이에 속한다. (12가)는 (12나)와 같이 번역된다.

- (12) 가. 수신 하이웨이 보다 적은 것은 슬롯부에 장착된 PP(Peripheral) 보드이다.
 나. What is less than the receiving highways is the PP(Peripheral) board which is attached to the slot.

위의 예문에서 보듯이 이러한 경우에는 ‘~ㄴ것은’이 ‘what’으로 번역되는 것이 자연스럽다. 이를 대역패턴으로 만들어 보면 다음과 같다.

- (13) { VP1[_NOTEXIST(S), VERB:[kpos==[ADJ]]
 VP2[] > VP1[SCONJ:[eroot := [what]]
 VERB:[(eroot := [be:v]) _AND (eform := [notS])]] VP2[VERB:[eform := [notS]]] }

여기서 VERB:[eform := [notS]]는 영어대역에 용언의 주어가 생성되지 말아야 함을 나타낸다. (13)에 제시된 패턴에서는 notS가 두 번 쓰이고 있다. VP2에 notS가 쓰인 이유는 앞에 나오는 VP1전체(즉, what절 전체)가 VP2의 주어로 쓰이기 때문이다. 한편, VP1에 notS가 쓰인 이유는 ‘what’이 VP1의 주어가 되기 때문이다. 이 패턴에서 ‘eroot := [be:v]’는 영어 대역에서 동사는 be-동사가 쓰여야 되며, 그 형태는 변화할 수 있음을 나타낸다. 즉, ‘be:v’는 ‘am’, ‘is’, ‘are’, ‘was’, ‘were’등의 변이형을 모두 포함하는 표기법이다.

다음으로 (11나)를 살펴보자. 이 경우는 VP1 용언의 품사가 형용사이고, 주어가 있는 경우로 다음과 같은 예문이 있다.

- (14) 가. 시이트의 두께가 균일한 것은 활물질 시이트를 세퍼레이트(40)상에 부착시키는 경우에도 적용된다.
 나. 변형이 가능한 것은 3개의 아미노기 및 6개의 카복실기가 있다.

이러한 문장들의 경우에는 VP1에 주어가 존재하기 때문에 that-절로 번역하면 될 것처럼 보인다. 즉, (14가)와 (14나)를 각각 ‘That the thickness of seats is uniform ~’와 ‘That the transform is possible ~’와 같이 번역하면 될 것처럼 보인다. 그러나 문제는 VP1의 번역을 VP2의 번역과 연결시켜 보면 영어 문장이 자연스럽게 못하다는 데 있다.

따라서 that-절을 그대로 사용하기 보다는 다른 형태로 바꾸어서 번역하는 것이 훨씬 더 자연스러울 것 같다. 예를 들어, (14가)의 경우에는 ‘That the thickness of seats is uniform ~’보다는 ‘The property that the thickness of seats is uniform ~’나 ‘To make the thickness of seats uniform ~’와 같이 번역하는 것이 더 자연스럽다. 그러나 이 두 가지 경우들은 모두 문제점을 가지고 있다. 먼저, that-절 앞에 명사를 사용하는 경우에는 이 명사를 몇 개로 한정지을 수 없다는 문제점이 있으며, to-부정사를 사용할 경우에는 한국어 원문과 영어 대역문이 의미가 달라질 수 있다는 문제점이 있다. 즉, (14가)의 원문에는 사역의 의미가 없지만, 이를 to-부정사로 번역하면 사역의 의미를 지니게 된다. 따라서 (11나)와 같은 경우에는 우선 that-절로 번역하기로 하겠다. 그러나, (11나)를 위하여 별도의 패턴을 구축하는 것이 아니라, 전체 패턴 중에서 한국어 원문부에 기술된 조건이 맞지 않는 경우에 적용되는 디폴트패턴(default pattern)에 이를 기술하기로 하겠다.

다음으로 (11다)를 살펴보자. 이 경우는 VP1 용언의 품사가 자동사이고, 주어가 없는 경우로, 다음과 같은 문장이 있다.

(15) 가. 도 14A의 변환패턴에 해당하는 것은 S104와 S105 이다.

나. What corresponds to the transform pattern in Figure 14A is S104 and S105.

이 패턴은 (11가)와 유사하다. 따라서 다음과 같이 패턴을 구축할 수 있다.

(16) { VP1[_NOTEXIST(S), VERB:[kpos==[VI]] VP2[] > VP1[SCONJ:[eroot := [what]] VERB:[(eroot := [be:v]) _AND (eform := [notS])]] VP2[VERB:[eform := [notS]]] }

(13)과 (16)의 차이점은 단지 VP1의 용언의 품사가 형용사인가 자동사인가의 차이만 있을 뿐이다. 따라서 이 둘은 다음과 같이 하나의 패턴으로 단순화시킬 수 있다.

(17) { VP1[_NOTEXIST(S), VERB:[(kpos==[ADJ]) _OR (kpos==[VI])] VP2[] > VP1[SCONJ:[eroot := [what]] VERB:[(eroot := [be:v]) _AND (eform := [notS])] VP2[VERB:[eform := [notS]]] }

즉, 한국어 원문부의 조건을 _OR를 사용하여 하나로 묶어놓은 것이다.

다음으로 (11라)를 살펴보자. 이 경우는 VP1 용언의 품사가 자동사이고, 주어가 있는 경우로 다음과 같은 예문이 있다.

(18) 가. 서로 다른 벨부호가 생성된다는 것은 이로부터 알 수 있을 것이다.

나. We can know that different bell signals are generated.

이런 경우에 대한 패턴은 다음과 같이 만들 수 있다.

(19) { VP1[_EXIST(S), VERB:[kpos==[VI]] VP2[] > VP2 VP1 [SCONJ:[eroot := [that]]] }

즉, VP1은 that절로 번역되고, 영어 대역에서는 VP1과 VP2가 바뀌어야 함을 나타내고 있다.

다음으로 (11마)를 살펴보자. 이 경우는 VP1 용언의 품사가 타동사이고, 주어가 없고 목적어도 없는 경우로 다음과 같은 예문이 있다.

(20) 가. 통신의 기본 수단으로 제공하는 것은 WWW(World Wide Web)표준문서이다.

나. 광고용의 대형 사이즈에 적용하는 것은 가격의 문제로 곤란하다.

이 두 문장을 영어로 번역해보면 다음과 같다.

(21) 가. What is provided as a basic tool of communication is the standard documents in WWW(World Wide Web).

나. To apply them to larger sizes for ADs is difficult because of the the problems in price.

이 예문들에서 보듯이 (11마)와 같은 경우에는 어떤 때에는 (21가)와 같이 'what'으로 번역하는 것이 자연스러운 경우도 있고, 어떤 때에는 (21나)와 같이 to-부정사로 번역하는 것이 자연스러운 경우도 있다. 따라서, (11마)와 같은 경우에는 하나의 패턴으로 만들기가 어렵다. 본 논문에서는 (11나)와 같이 (11마)를 위하여 별도의 패턴을 구축하는 것이 아니라, 전체 패턴 중에서 한국어 원문부에 기술된 조건이 맞지 않는 경우에 적용되는 디폴트패턴(default pattern)에 이를 기술하기로 하겠다.

다음으로 (11바)를 살펴보자. 이 경우는 VP1 용언의 품사가 타동사이고, 주어가 없고 목적어가 있는 경우 다음과 같은 예문이 있다.

(22) 가. 이 디렉토리를 사용하는 것은 현실적인 방안이 될 수 없다.

나. It cannot be a realistic method to use the current directory.

따라서, 다음과 같이 패턴을 구축할 수 있다.

(23) { VP1[_NOTEXIST(S), _EXIST(O), VERB:[kpos==[VT]]] VP2[] > VP2 VP1[SCONJ:[eroot := [to]] VERB:[eform := [sinf]]] }

여기서 *sinf*는 to-부정사의 의미상의 주어가 명시적으로 나타나지 않음을 뜻한다.

다음으로 (11사)를 살펴보자. 이 경우는 VP1 용

언의 품사가 타동사이고, 주어가 있고 목적어는 없는 경우로, 특허문서 말뭉치 안에서는 예문을 찾을 수 없었다. 이에 대한 정확한 이유는 알 수 없으나, (11사)와 같이 용언이 타동사이고 주어가 있고 목적어는 없는 경우는, 목적어가 원래 '것'으로, 이 '것'은 관계절의 선행사가 되지 못하기 때문에 추측된다.

다음으로 (11아)를 살펴보자. 이 경우는 VP1의 용언의 품사가 타동사이고, 주어가 있고 목적어도 있는 경우로 다음과 같은 예문이 있다.

(24) 가. 어느 한 코일이 방향을 달리하는 것은 본 발명의 범위에 해당된다고 할 것이다.

나. It can be included in the scope of this invention that a certain coil changes its direction.

이 예문에서 보듯이 VP1에 해당되는 부분을 *that*-절로 번역했다. 그러나 이러한 문장은 한국어 원문은 자연스러워 보일 수 있으나, 영어 대역문은 비문법적이라고 할 수는 없어도 상당히 부자연스럽다. 문장의 영어 대역부에 *that*-절대신 to-부정사를 사용하고 그 의미상의 주어를 넣는다고 해도 문장이 많이 좋아지지 않는다는. 본 논문에서는 이러한 경우 우선은 to-부정사로 번역하는 것으로 하고, 다음과 같이 패턴을 만들겠다.

(25) { VP1[_EXIST(S), _EXIST(O), VERB:[kpos ==[VT]]] VP2[] > VP2 VP1[SCONJ:[eroot := [to]] VERB:[eform := [inf]]] }

여기서 *inf*는 (23)에 쓰인 *sinf*와는 달리 to-부정사의 주어가 명시적으로 나타남을 의미한다. 이 패턴에서 보면, 영어 대역문의 어순이 VP2 VP1으로 나타나고, VP2에는 아무 것도 명시되어 있지 않음을 알 수 있다. 이는 한국어 원문의 VP2를 그대로 번역

하여 넣으면 된다. (24가)와 같이 한국어 원문에서 VP2의 주어가 명시적으로 나타나 있지 않는 데도 불구하고 영어 대역문에서 VP2의 주어가 필요한 경우는 주어복원 알고리즘이 필요한 주어를 복원하여 넣는 것으로 가정하겠다. 이 패턴에 따르면 (26가)는 (26나)와 같이 번역된다.

(26) 가. 어느 한 코일이 방향을 달리하는 것은 본 발명의 범위에 해당된다고 할 것이다.

나. It can be included in the scope of this invention for a certain coil to change its direction.

이것으로 (11)에 나온 모든 경우에 대해 그 특징을 살펴보고, 해당하는 대역패턴도 만들어 보았다.

그러나, 한 가지 더 살펴보아야 할 것이 남아있다. 3절에서 지적되었듯이 ‘~은/는’은 주격조사인 ‘~이/가’를 대신할 수도 있고 목적격 조사인 ‘~을/를’을 대신할 수도 있다. 그리고 이에 따라 VP1과 VP2의 어순이 바뀔 수도 있다. 그러면 ‘~은/는’이 ‘~이/가’를 대신하는지 ‘~을/를’을 대신하는지를 어떻게 결정하는가? 이와 관련하여 우리는 VP2에 나오는 용언의 문법적 성질에 주목하고자 한다. VP2의 용언이 형용사나 자동사라면 ‘~ㄴ것은’이 붙는 VP1은 VP2의 주어로밖에 쓰이지 않는다. 그러나 VP2의 용언이 타동사라면, 특허문서 말뭉치를 조사해본 결과, ‘~ㄴ것은’이 붙는 VP1은 타동사의 목적으로 쓰이고 있음이 관찰되었다. 그러므로 VP2의 용언이 형용사나 자동사라면 ‘~은/는’이 ‘~이/가’를 대신하고, VP2의 용언이 타동사라면 ‘~은/는’이 ‘~을/를’을 대신한다고 할 수 있다. 또한, VP2의 용언이 형용사나 자동사라면 VP1 VP2의 어순이 자연스럽게, VP2의 용언이 타동사라면 VP1 VP2의 어순이 바뀌어야 함을 알 수 있었다. 우리가 살펴본 (11가) ~ (11아) 중에는 (11가)와 (11다)가 VP2의 용언에 따

라 어순이 바뀌는 것으로 관찰되었다. 따라서, 우리는 (17)의 패턴을 (27)과 같이 바꾸어야 함을 알 수 있다.

(17) { VP1[_NOTEXIST(S), VERB:[(kpos== [ADJ]) _OR (kpos==[VI])]] VP2[] > VP1 [SCONJ:[eroot := [what]] VERB:[(eroot := [be:v]) _AND (eform := [notS])]] VP2 [VERB:[eform := [notS]]] }

(27) 가. { VP1[_NOTEXIST(S), VERB:[(kpos ==[ADJ]) _OR (kpos==[VI])]] VP2 [VERB:[(kpos==[ADJ]) _OR (kpos== [VI])]] > VP1 [SCONJ:[eroot := [what]] VERB:[(eroot := [be:v]) _AND (eform := [notS])]] VP2 [VERB:[eform := [notS]]] }

나. { VP1[_NOTEXIST(S), VERB:[(kpos ==[ADJ]) _OR (kpos==[VI])]] VP2 [VERB:[kpos==[VT]]] > VP2 VP1 [SCONJ:[eroot := [what]] VERB:[(eroot := [be:v]) _AND (eform := [notS])]] }

여기에서 보듯이 (27가)는 VP2의 용언이 형용사나 자동사인 경우로 VP1 VP2의 어순을 취하고 있음을 알 수 있다. 이에 비해, (27나)는 VP2의 용언이 타동사인 경우로 VP1 VP2의 어순이 바뀌어 있음을 알 수 있다.

4.3. VP2의 연어와 패턴처리

이제, ‘~ㄴ것은’과 함께 자주 출현하는 VP2의 용언들을 살펴보고 이들을 대역패턴으로 만들어 보겠다. 먼저, 특허문서 말뭉치에서 ‘~ㄴ것은’과 연어 관계에 있는 용언들을 살펴보니, 다음과 같았다.

- (28) ‘~ㄴ것은’과 자주 연어되는 VP2 용언들
- ㄴ것은 가능하다
 - ㄴ것은 곤란하다
 - ㄴ것은 당연하다
 - ㄴ것은 동일하다
 - ㄴ것은 명백하다
 - ㄴ것은 바람직하다
 - ㄴ것은 불가능하다
 - ㄴ것은 쉽다
 - ㄴ것은 어렵다
 - ㄴ것은 없다
 - ㄴ것은 용이하다
 - ㄴ것은 자명하다
 - ㄴ것은 중요하다

이 용언들에 대한 대역패턴은 (10)에서 제시한 표기법을 사용하기로 하겠다. 예를 들어, ‘~ㄴ것은가능하다’에 대한 패턴은 다음과 같다.

- (29) ‘~ㄴ것은 가능하다’에 대한 패턴
- ```
{ VP1[_NOTEXIST(S)] VP2[VERB:[kroot==[가능하]]] > VP2[SCONJ:[eroot := [it_be:v_possible]] VERB:[eform := [notS]]]
VP1[SCONJ:[eroot := [to]] VERB:[eform := [sinf]]] }
/* 캔틸레버를 진동시키는 것은 가능하다. */
/* It is possible to make Kantil lever to vibrate. */
{ VP1[_EXIST(S)] VP2[VERB:[kroot==[가능하]]] > VP2[SCONJ:[eroot := [it_be:v_possible]] VERB:[eform := [notS]]]
VP1[SCONJ:[eroot := [that]]] }
/* 그들이 명세서에 따라서 실시하는 것은 가능하다. */
/* It is possible that they perform it according to the specification details. */
```

여기에서 알 수 있듯이 VP1의 주어의 유무에 따라,

주어가 없는 경우는 to-부정사로 번역을 하고, 주어가 없는 경우는 that-절로 번역한다. 그러나, 두 경우 모두 VP2 VP1의 어순이 뒤를 알 수 있다.

위의 경우와 같이 ‘~ㄴ것은 가능하다’를 kroot를 사용하여 패턴으로 만들면, ‘~ㄴ것은 가능하다’뿐만 아니라, ‘가능하다’를 수식하는 부사가 들어 있는 문장들도 처리가 가능해진다. 따라서, 이러한 용언들을 패턴으로 처리하는 것이 훨씬 효율적이다.

(28)에 열거되어 있는 용언들 중에는 같은 의미로 번역될 수 있는 것들이 있다. 예를 들면, ‘~ㄴ것은 쉽다’와 ‘~ㄴ것은 용이하다’는 모두 ‘It is easy ~’로 번역될 수 있을 것 같다. 따라서, 이러한 경우는 두개의 용언의 kroot를 \_OR로 연결하여 나타낼 수 있다. 예를 들면, ‘~ㄴ것은 쉽다’와 ‘~ㄴ것은 용이하다’에 대한 패턴은 다음과 같이 구축될 수 있다.

- (30) ‘~ㄴ것은 쉽다/용이하다’에 대한 패턴
- ```
{ VP1[ _NOTEXIST(S) ] VP2[VERB:[(kroot==[쉽]) _OR (kroot==[용이하])] > VP2[SCONJ:[eroot := [it_be:v_easy]] VERB:[eform := [notS]]] VP1 [SCONJ:[eroot := [to]] VERB:[eform := [sinf]]] }
{ VP1[ _EXIST(S) ] VP2[VERB:[(kroot==[쉽]) _OR (kroot==[용이하])] > VP2 [SCONJ:[eroot := [it_be:v_easy]] VERB:[eform := [notS]]] VP1[SCONJ:[eroot := [that]]] }
```

(28)에 있는 다른 용언들도 위와 같은 방식으로 기술할 수 있다.

4.4. 패턴의 종합과 패턴 적용 규약

이제 지금까지 논의된 방법을 종합하여 ‘ㄴ것은:2_다.’의 대역패턴을 만들어 보자. 지금까지 우리가 구축한 대역패턴은 부록에 제시되어 있다. 이 패턴의 가장 위쪽에는 디폴트패턴이 주어의 유무에 따라

_NOTEXIST(S)와 _EXIST(S)로 나뉘어 기술되어 있다. 즉, 다른 조건들이 만족되지 않는다면 주어가 없는 경우는 _NOTEXIST(S) 패턴이, 주어가 있는 경우는 _EXIST(S) 패턴이 적용된다.

그러나, 만약 한국어 원문부에 조건이 둘 이상 기술되어 있다면 다음과 같은 순서로 패턴이 적용된다.

(31) 패턴적용의 순서

- 가. 한국어 원문부에 조건의 개수가 많은 것이 먼저 적용된다.
- 나. 한국어 원문부에 조건의 개수가 같다면 (32)에 제시한 우선순위에 따라 적용되는 패턴을 정한다.
- 다. 한국어 원문부에 조건에 맞지 않는 경우는 디폴트패턴이 적용된다.

(32) 패턴적용의 우선순위

- kroot >> _NOTEXIST(O), _EXIST(O)
- >> _NOTEXIST(S), _EXIST(S)

즉, 패턴에 기술된 조건의 수가 많은 것이 먼저 적용되고, 조건의 수가 같다면, kroot가 있는 것이 먼저 적용된다. 만약 kroot가 없거나 여러 패턴이 동일한 kroot값을 가지고 있다면, 그 다음은 목적어의 유무, 그 다음은 주어의 유무에 따라 적용될 패턴이 결정된다.

5. 결 론

본 논문에서는 ‘~ㄴ것은’을 한영자동번역에서 어떻게 처리할 수 있는지를 알아보았다. 우리는 먼저 ‘~ㄴ것은’이 가진 문제점을 살펴보았다. 그리고 ‘~ㄴ것은’을 중심으로 앞뒤에 나오는 용언들의 특징들을 살펴보고 이 특징들을 대역패턴에 반영하였다. 우리는 또한 특허 문서에서 ‘~ㄴ것은’과

자주 연어 관계에 있는 용언을 찾아 이들을 대역 패턴에 반영하였다. 이렇게 함으로써, 우리는 ‘~ㄴ것은’을 패턴방식의 한영자동번역에서 효과적으로 처리할 수 있게 되었다.

참고문헌

최승권. 2004. “한국어 자동번역 어디까지 왔나?” 제5회 한국어정보화 아카데미 강의 자료집. 233-288. 서울 : 연세대학교.

최승권, 홍문표. 2005. “언어학도를 위한 기계번역 입문”, 제6회 한국어정보화 아카데미 강의 자료집. 215-245. 서울 : 연세대학교.

Jurafsky, D. and S. Martin. 2000. *Speech and Language Processing*. Upper Saddle River, NJ: Prentice Hall.

Somers, Harold. 2003. Machine Translation: Latest Development. In R. Mitkov, ed., *The Oxford Handbook of Computational Linguistics*. 512-528. Oxford : Oxford University Press.

부 록

[KEY]
 ㄴ것은:2_다:1
 [CONTENT]
 /* 디폴트 패턴, 주어 무 */
 { VP1[_NOTEXIST(S)] VP2[] > VP2 VP1[SCONJ:
 {eroot := [to]} VERB:[eform := [sinf]]] }
 /* 아무 준비없이 그곳에 간 것은 영희를 화나게 하였다. */
 /* It makes Younghee angry to go there without any preparation. */
 /* 디폴트 패턴, 주어 유 */
 { VP1[_EXIST(S)] VP2[] > VP2 VP1[SCONJ:
 {eroot := [that]]] }
 /* 철수가 그곳에 간 것은 영희를 화나게 하였다. */
 /* It makes Younghee angry that he went there

```

without any preparation. */
/* VP1: 형용사/자동사, 주어 무; VP2: 형용사/자동사 */
{ VP1[ _NOTEXIST(S), VERB:[(kpos==[ADJ]) _OR
(kpos==[VI])] ] VP2[VERB:[(kpos==[ADJ]) _OR
(kpos==[VI])] ] > VP1[SCONJ:[eroot := [what]]
VERB:[(eroot := [be:v]) _AND (eform := [notS])] ]
VP2[VERB:[eform := [notS]]] }
/* 수신 하이웨이 보다 적은 것은 슬롯부(10)에 장착된
PP(Peripheral) 보드이다. */
/* What is less than the receiving highways is the
PP(Peripheral) board which is attached to the slot. */
/* VP1: 형용사/자동사, 주어 무; VP2: 타동사 */
{ VP1[ _NOTEXIST(S), VERB:[(kpos==[ADJ]) _OR
(kpos==[VI])] ] VP2[VERB:[kpos==[VT]] ] > VP2
VP1[SCONJ:[eroot := [what]] VERB:[(eroot :=
[be:v]) _AND (eform := [notS])] ] }
/* 화로에 남은 것들은 적절히 처리한다. */
/* We process what are left in the fire pot. */
/* VP1: 형용사/자동사, 주어 유; VP2: 타동사 */
{ VP1[ _EXIST(S), VERB:[kpos==[VI]] ] VP2[ ] >
VP2 VP1[SCONJ:[eroot := [that]]] }
/* 서로 다른 벨부호가 생성된다는 것은 이로부터 알 수
있을 것이다. */
/* We can know that different bell signals are generated. */
/* VP1: 타동사, 주어 무, 목적어 유 */
{ VP1[ _NOTEXIST(S), _EXIST(O), VERB:[kpos
==[VT]] ] VP2[ ] > VP2 VP1[SCONJ:[eroot := [to]]
VERB:[eform := [sinf]]] }
/* 이 디렉토리를 사용하는 것은 현실적인 방안이 될 수 없다. */
/* It cannot be a realistic method to use the current
directory. */
/* VP1: 타동사, 주어 유, 목적어 유 */
{ VP1[ _EXIST(S), _EXIST(O), VERB:[kpos==[VT]] ]
VP2[ ] > VP2 VP1[SCONJ:[eroot := [to]] VERB:
[eform := [inf]]] }
/* 어느 한 코일이 방향을 달리하는 것은 본 발명의 범위에
해당된다고 할 것이다. */
/* It can be included in the scope of this invention
for a certain coil to change its direction. */
/* '~ㄴ것은'과 연어되는 용언들의 패턴 */
{ VP1[ _NOTEXIST(S) ] VP2[VERB:[kroot==[가능
하]] ] > VP2[SCONJ:[eroot := [it_be:v_possible]]
VERB:[eform := [notS]] ] VP1[SCONJ:[eroot :=
[to]] VERB:[eform := [sinf]]] }
/* 캔틸레버를 진동시키는 것은 가능하다. */
/* It is possible to make Kantil lever to vibrate. */
{ VP1[ _EXIST(S) ] VP2[VERB:[kroot==[가능하]] ]
> VP2[SCONJ:[eroot := [it_be:v_possible]] VERB:
[eform := [notS]] ] VP1[SCONJ:[eroot := [that]]] }
/* 그들이 명세서에 따라서 실시하는 것은 가능하다. */
/* It is possible that they perform it according to the
specification details. */
{ VP1[ _NOTEXIST(S) ] VP2[VERB:[(kroot==[곤란
하]) _OR (kroot==[어렵])] ] > VP2[SCONJ:[eroot :=
[it_be:v_difficult]] VERB:[eform := [notS]] ] VP1
[SCONJ:[eroot := [to]] VERB:[eform := [sinf]]] }
/* 펄스 음을 특징화하는 것은 곤란하다. */
/* It is difficult to characterize the pulse sound. */
{ VP1[ _EXIST(S) ] VP2[VERB:[(kroot==[곤란하])
_OR (kroot==[어렵])] ] > VP2[SCONJ:[eroot := [it_
be:v_difficult]] VERB:[eform := [notS]] ] VP1
[SCONJ:[eroot := [that]]] }
/* 우리가 복잡한 연결들의 연결 강도를 분석하는 것은
매우 어렵다. */
/* It is very difficult to analyze the connecting
strength of the complex connections. */
{ VP1[ _NOTEXIST(S) ] VP2[VERB:[kroot==[당연
하]] ] > VP2[SCONJ:[eroot := [it_be:v_natural]]
VERB:[eform := [notS]] ] VP1[SCONJ:[eroot :=
[to]] VERB:[eform := [sinf]]] }
/* 스탬프(55)의 표면을 깨끗하게 세정하는 것은 당연하다. */
/* It is natural to clean the surface of the stamp (55). */
{ VP1[ _EXIST(S) ] VP2[VERB:[kroot==[당연하]] ]
> VP2[SCONJ:[eroot := [it_be:v_natural]] VERB:
[eform := [notS]] ] VP1[SCONJ:[eroot := [that]]] }
/* 물체의 변형이 본 발명의 범위에 속하는 것은 당연하다. */
/* It is natural that the transform of the objects is
included in the scope of this invention. */
{ VP1[ _NOTEXIST(S) ] VP2[VERB:[kroot==[가능
하]] ] > VP2[SCONJ:[eroot := [it_be:v_invariable]]
VERB:[eform := [notS]] ] VP1[SCONJ:[eroot :=
[to]] VERB:[eform := [sinf]]] }

```

```

/* 구동모터의 구동력에 의하여 개폐되는 것은 동일하다. */
/* It is invariable to open and/or close by the power
of the driving motor. */
{ VP1[ _EXIST(S) ] VP2[VERB:[kroot==[가능하]]]
> VP2[SCONJ:[eroot := [it_be:v_invariable]] VERB:
[eform := [notS]]] VP1[SCONJ:[eroot := [that]]] }
/* 안창(2)이 부착되는 것은 동일하다. */
/* It is invariable that a shoe liner(2) is attached. */
{ VP1[ _NOTEXIST(S) ] VP2[VERB:[(kroot==[명백
하]) _OR (kroot==[자명하])] ] > VP2[SCONJ:[eroot
:= [it_be:v_clear]] VERB:[eform := [notS]]] VP1
[SCONJ:[eroot := [to]] VERB:[eform := [sinf]]] }
/* CD와 같은 기록매체를 사용하는 것은 명백하다. */
/* It is clear to use the recording media such as CD. */
{ VP1[ _EXIST(S) ] VP2[VERB:[(kroot==[명백하])
_OR (kroot==[자명하])] ] > VP2[SCONJ:[eroot :=
[it_be:v_clear]] VERB:[eform := [notS]]] VP1[
SCONJ:[eroot := [that]]] }
/* 본 장치가 다양한 변형을 가할 수 있다는 것은 자명하다. */
/* It is clear that this equipment can perform various
different transformation. */
{ VP1[ _NOTEXIST(S) ] VP2[VERB:[kroot==[바람직
하]]] > VP2[SCONJ:[eroot := [it_be:v_reasonable]]
VERB:[eform := [notS]]] VP1[SCONJ:[eroot :=
[to]] VERB:[eform := [sinf]]] }
/* 기관의 정확한 위치 및 방향을 아는 것은 바람직하다. */
/* It is reasonable to know the correct position and
direction of the electronic plate. */
{ VP1[ _EXIST(S) ] VP2[VERB:[kroot==[바람직
하]]] > VP2[SCONJ:[eroot := [it_be:v_reasonable]]
VERB:[eform := [notS]]] VP1[SCONJ:[eroot :=
[that]]] }
/* 독립사업자(10)가 개인적으로 소비한 것은 바람직하다. */
/* It is reasonable that independent companies
consume it individually. */
{ VP1[ _NOTEXIST(S) ] VP2[VERB:[kroot==[불가능
하]]] > VP2[SCONJ:[eroot := [it_be:v_impossible]]
VERB:[eform := [notS]]] VP1[SCONJ:[eroot :=
[to]] VERB:[eform := [sinf]]] }
/* 그 게임을 단독으로 실행하는 것은 불가능하다. */
/* It is impossible to run the game individually. */

```

```

{ VP1[ _EXIST(S) ] VP2[VERB:[kroot==[불가능
하]]] > VP2[SCONJ:[eroot := [it_be:v_impossible]]
VERB:[eform := [notS]]] VP1[SCONJ:[eroot :=
[that]]] }
/* 배터리가 조기에 소진되는 것은 불가능하다. */
/* It is impossible that the battery is exhausted early. */
{ VP1[ _NOTEXIST(S) ] VP2[VERB:[(kroot==[쉽])
_OR (kroot==[용이하])] ] > VP2[SCONJ:[eroot :=
[it_be:v_easy]] VERB:[eform := [notS]]] VP1
[SCONJ:[eroot := [to]] VERB:[eform := [sinf]]] }
/* 각종 보험상품을 비교하는 것은 쉽다. */
/* It is easy to compare various insurance products. */
{ VP1[ _EXIST(S) ] VP2[VERB:[(kroot==[쉽])_OR
(kroot==[용이하])] ] > VP2[SCONJ:[eroot := [it_be:v
_easy]] VERB:[eform := [notS]]] VP1[SCONJ:
[eroot := [that]]] }
/* 우리가 하나 하나를 눈으로 보아 확인하는 것은 용이하다. */
/* It is easy that we confirm each one with eyes. */
{ VP1[ _NOTEXIST(S) ] VP2[VERB:[kroot==[없]]]
> VP2[SCONJ:[eroot := [there_be:v_nothing]] VERB
:[eform := [notS]]] VP1[SCONJ:[eroot := [to]]
VERB:[eform := [sinf]]] }
/* 신경 세포에 관련된 것은 없다. */
/* There is nothing to be related to nerve cells. */
{ VP1[ _EXIST(S) ] VP2[VERB:[kroot==[없]]] >
VP2[SCONJ:[eroot := [there_be:v_nothing]] VERB:
[eform := [notS]]] VP1[SCONJ:[eroot := [that]]] }
/* 특별한 치료방법이나 약제가 개발된 것은 없다. */
/* There is nothing that was developed for special
treatment or medicine. */
{ VP1[ _NOTEXIST(S) ] VP2[VERB:[kroot==[중요
하]]] > VP2[SCONJ:[eroot := [it_be:v_important]]
VERB:[eform := [notS]]] VP1[SCONJ:[eroot :=
[to]] VERB:[eform := [sinf]]] }
/* 하나 이상의 성분을 포함하는 것은 중요하다. */
/* It is important to include one or more ingredient. */
{ VP1[ _EXIST(S) ] VP2[VERB:[kroot==[중요하]]]
> VP2[SCONJ:[eroot := [it_be:v_important]] VERB:
[eform := [notS]]] VP1[SCONJ:[eroot := [that]]] }
/* POI가 선택되는 것은 중요하다. */
/* It is important that POI is selected. */

```