

# 구 단위 패턴 기반 한영 기계 번역에서의 기능동사 구문의 중심어 선택 모델

김혜경, 채영숙, 최기선  
전문용어 언어공학 연구센터 (KORTERM)

## Selecting Model of Head in Support Verb Constructions for Phrase-Pattern-based Korean-to-English Machine Translation

KIM, Hae-Gyung / Chae, Young-Soog / CHOI, Key-Sun  
KORTERM  
(hgkim,yschae,kschoi)@korterm.kaist.ac.kr

### 요 약

한국어는 잉여성과 중의성의 범 언어적인 특징과 함께 다른 언어에 비해 주어의 생략이 두드러지며 어순이 자유롭기 때문에 구문 형식의 지배를 덜 받는다는 개별적인 특성을 지닌다. 이러한 특성으로 인해 기계번역의 패턴을 추출할 때 서로 유사 가능성이 있는 패턴에 대한 고려가 없이는 같은 의미의 서로 다른 여러 개의 패턴을 모두 하나의 패턴으로 처리하는 오류를 범할 위험이 있다.

본 연구에서 사용되는 구 단위 패턴은 동사구, 명사구, 형용사구 그리고 부사구를 중심으로 한국어 패턴, 패턴 대표 카테고리, 한국어 패턴의 중심어 및 제약조건 대역영어패턴 의미코드로 나뉜다. 범 언어적인 특성의 한국어와 영어간 격차를 해소하기 위해 각각의 명사에 의미코드를 사용하여 다중 언어기반 체계를 구축하였으며, 한국어의 개별적인 특성으로 인해 발생하는 문제를 해소하기 위해 중심어 부과 자질을 사용하였다.

중심어 부과 자질에 있어서, 특히 술어기능명사를 중심어로 하는 기능동사 '하-' 구문은 다른 동사 구문의 형식과는 달리 논항의 수와 형태를 동사가 아닌 명사가 수행하게 된다. 이러한 특징에 대한 변별적인 자질 부여는 구문의 형태-통사적 특징뿐만 아니라 의미적인 고유의 특성까지도 잘 뒷받침하면서 패턴 추출에 월등한 효율성을 제시할 수 있다.

향후 이에 대한 연구는 전반적인 기능동사 구문뿐만 아니라 개별적인 특징을 보이는 모든 구문에 대한 연구로 확대되어 패턴 기반 기계번역의 패턴 추출에 기본적인 정보의 역할을 담당해야 할 것이다.

### 1 들어가는 말

한국어는 다른 언어들과 잉여적이고 중의적인 범 언어적인 특징을 공유하기도 하지만 한국어만이 지니는 개별적인 언어의 자질을 보이기도 한다. 잉여성(redundancy)이라 함은 문장 요소가 생략되는 언어의 특징을 말한다. 한 문맥을 단위로 하여 이미 사용되었거나, 사용되지 않았다 하더라도 주어진 상황(situation)에 의해 청자와 화자 혹은 자료와 그것을 읽는 독자 사이에 짐작 가능한 요소들이다. 이러한 잉여성을 지니는 요소는 작가는 간단한 조사 어휘에서부터 주어와 목적어 같은 품사의 생략, 크기는 문장어구의 생략으로 이어질 수 있다. 중의성(ambiguity)은 문장 또는 어휘에서 나타나는 것으로 나누어 고려할 수 있다. 한국어에서는 이러한 범 언어적인 특징에서의 잉여성에 의한 문장 성분, 특히 주어의 생략이 두드러지게 나타난다. 또한 한

국어는 다른 언어에 비해 어순이 비교적 자유로워서 전체 구문 형식의 지배를 덜 받는다. 이러한 특성으로 인해 다른 기계번역의 방식보다는 번역할 때의 그 수가 비교적 제한적일 수 있고 자료 처리를 효율적으로 할 수 있는 패턴에 기반한 기계번역이 유리하다. 이러한 한국어의 개별적인 특성은 패턴의 추출과 동시에 형태-통사적인(morpho-syntactic) 면뿐만 아니라 의미적(semantic)으로 서로 유사한 패턴에 대한 고려가 선행되어야 한다. 이렇게 함으로써 같은 의미의 서로 다른 여러 개의 문장을 모두 개별적인 패턴으로 인식하여 패턴의 수를 무한대로 증가시키는 오류를 피할 수 있다. 2장에서는 구 단위 패턴 기반 기계번역에서 선택하고자 하는 패턴의 모델에 대해 간략하게 기술하고자 한다. 3장에서는 이 모델을 기반으로 패턴을 구축해 나감에 있어 특히 기능동사 구문의 패턴 처리에서 발생하는 문제점을 찾아보고, 그 문제점을 해결할 수 있는 선택 모델을 제시하

여 그 효율성을 검증하고자 한다.

## 2 구 단위 패턴 기반 기계번역의 패턴 모델

구 단위 패턴 기반 한영 기계번역에서 사용되는 패턴 모델은 패턴의 전체적인 종류를 선택한 다음 세부적인 기술 형식에 따라 구체적인 자질을 표기하는 방식을 취했다.

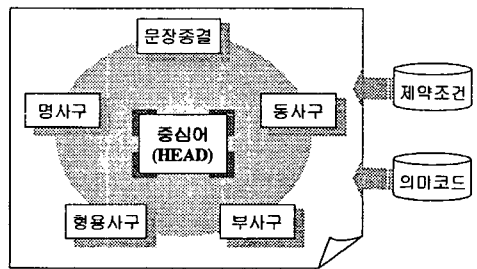
구 단위 패턴 기반 한영 기계번역에서 사용하는 패턴의 종류는 크게 문장 종결 패턴, 접속절 혹은 접속구 패턴, 용언 기본 문형 패턴, 명사구 패턴, 그리고 부사구 패턴의 5가지이다. 문장 종결 패턴은 평서문, 의문문, 청유문, 명령문, 감탄문, 부정문 등에서 쓰이는 문장의 종결어미(., ?, ! 등)를 중심으로, 문장 전체의 개별적인 유형을 결정짓는 패턴이다. 예를 들어, 평서문의 '니니+다+'와 같은 형태이다. 접속절 혹은 접속구 패턴은 두 개 이상의 절 또는 구로 연결된 문장의 연결 유형을 정의한다. 예를 들어, 두 개의 구가 '-었지만'으로 연결되었다면 '구+었지만+구'로 전체 문장의 형태를 기술한다. 용언은 일반적으로 동사, 형용사, 보조용언 등의 품사를 모두 여기에 포함시키지만, 용언 기본 문형 패턴에서는 주로 동사를 기본 단위로 한 구문들을 선택한다. 형용사의 경우, 동사의 보조사 역할을 하는지 명사를 수식하는지에 따라 명사 또는 동사의 수식어구로 처리하며, 보조용언은 동사구에 포함되는 구문 자질로 처리한다. '보내다'라는 동사를 중심으로 '-가 -를 보내다' 등의 문형이 한 예이다. 명사구 패턴은 '-에 관한 편지' 등과 같은 것인데, '편지'라는 명사를 중심으로 하나의 명사를 수식하는 수식어구를 포함한 형태이다. 부사구 패턴의 처리에 있어서는 특히 시간의 부사구는 다른 부사구에 비해 구문정보가 없어도 쉽게 인식될 수 있다. 시간 부사구에 대한 패턴은 패턴 매칭의 전처리 단계에서 따로 처리하도록 한다. 이는 자의적인(arbitrary) 논항(argument)의 특징을 가장 많이 나타내는 부사구의 패턴 매칭에 대한 부담을 줄일 수 있다는 장점이 있다.

위의 5가지 패턴은 패턴의 기술 형식에 의해 세부적인 자질을 포함하는 하나의 패턴의 틀을 갖추게 된다. 패턴의 모든 내용과 형식들은 하나의 구가 중심어(head)와 거기에 부합되는 논항으로 이루어져 있다는 가정을 전제로 한다. 가령 '주다'라는 술어의 의미상으로 볼 때 하나의 주어와 역할이 서로 다른 두 개의 보어를 필요로 한다. 예를 들어, 'N1+가 N2+를 N3+에게 주다'라는 문장은 '주다'라는 술어를 중심으로 술어 행위를 취하는 주체(subject)인 동사에 N2를 N3에게 전달하는 수여자인 N1, N1이 N3에게 전달하는 대상(object)으로서의 N2, 그리고 여격 전치사를 앞서는 수혜자(receiver)의 N3가 그 논항이 된다. 이러한 방식은 하나의 동사가 서로 다른 논항 구조를 지닐 때, 형식이 다른 여러 개의 문장을 하나의 계열로 묶어줄 수 있

다는 장점이 있다. 패턴의 형식은 한국어 패턴, 패턴 대표 카테고리, 한국어 패턴의 중심어, 제약조건 의미코드 대역 영어패턴의 순으로 표현된다.

- (한국어 패턴) . . . . . ①
- (패턴 대표 카테고리) . . . . . ②
- (한국어 패턴의 중심어) . . . . . ③
- ((제약조건1 의미코드1 대역영어패턴1)  
(제약조건2 의미코드2 대역영어패턴2)) . . . . ④[7]

①의 한국어 패턴에는 한영 대역 코퍼스(corpus)의 한국어 문장에서 추출된 개별적인 패턴 각각에 한국어 패턴의 형식이 하나씩 부과된다. 각 항에 기술되는 논항의 쓰임이 필수적이 아니라 자의적일 경우에는 쓰이는 논항의 순서에 따라 n개의 항으로 In(...)이라 표기한다. ②의 패턴 대표 카테고리는 NP, VPO, VPS, ADVP, S로 나누어 표기하는데, NP는 일반적인 명사를 중심으로 하는 명사구를 의미한다. VPO와 VPS는 모두 동사를 중심으로 하는 동사구인데, 한국어의 주어가 생략되는 특징을 감안하여 그 처리를 용이하게 하기 위해 주어를 지니지 않거나 생략된 경우는 VPO로, 주어를 지니는 경우는 VPS로 표기한다. S는 다양한 문장종결 어미를 포함하는 패턴이다. ③은 한국어 패턴에서 쓰이는 중심어에 관한 문제이다. 중심어는 언어학과 전산학에서 모두 쓰이는 개념으로 이미 Jackendoff(1977)의 핵계층이론(X-bar theory)에서 구체적으로 제시된 바 있으며 현재 대부분의 문법 이론들은 이 개념을 수용하고 있다. 또한 중심어는 구 구조를 형성하는데 있어서 가장 핵심적인 역할을 하는 언어 기호로 알려져 있다.[5][9] 구 단위 패턴 기반 한영 기계번역에서도 이 개념을 사용하여 한 패턴내의 전체적인 패턴 대표 카테고리를 결정 짓는다. 일반적으로 명사구인 경우에는 명사, 동사구인 경우에는 동사, 부사구인 경우에는 부사를 의미한다. 중심어에 따라 필수적인 논항의 수나 종류가 결정되며, 그에 따라 번역할 대상 언어의 문장 패턴에서 번역할 목적 언어로의 변환 패턴이 결정된다.



<그림 1: 구 단위 패턴의 구성도>

중심어의 선택은 위의 <그림 1>에서 보는 것처럼 전체 패턴의 틀을 결정하는 절대적인 요소가 된다.

하나의 한국어 패턴이 어떠한 중심어를 지니는가에 따라 대표 패턴과 전체 패턴이 확정되었다면, 다음으로는 ④처럼 그에 따르는 제약조건과 의미코드에 따라 대역영어 패턴을 선택하게 된다. 제약조건은 패턴내에서 해당 예제를 나열하는 방법과 시스러스에서 해당 *nonterminal*을 지정하여 의미코드를 부여하는 방법을 병행한다. 대역 영어 패턴은 그 대표 카테고리에 따라 사용가능한 자질들을 부여하게 된다. 예를 들어 명사구는 \$np(noun='...', cnoun='...', adjp='...', art='...', num=sg/pl, poss='...', capitalized, interrogative)로, 동사구는 \$vp(verb='...', particle='...', subj='...', obj='...', iobj='...', scomp='...', ocomp='...', advp='...', tense='...', aux='...', negative)로 표기한다. 각각의 자질 코드에 대한 의미는 <표 1>과 같다.

\$np		\$vp	
noun	명사(중심어)	verb	동사(중심어)
cnoun	복합명사의 선행어	particle	분사
adjp	수식어구	subj	주어
art	관사	obj	(직접)목적어
num	단수/복수	iobj	간접목적어
poss	소유격	scomp	주격 보어
capitalized	첫문자의 대문자 표기	ocomp	목적격 보어
interrogative	의문사	advp	부사구
		tense	시제
		aux	조동사
		negative	부정형

<표 1 : 대역영어패턴의 자질 표기>

그 외에 조사나 어미를 일관성 있게 처리하기 위해 각각의 이형태(allomorph)는 전처리기에서 동일하게 처리, 패턴의 효율성을 증가시켰다.

위의 전체적인 패턴의 형식을 기술할 때 주의해야 할 점은 순차적으로 한국어 구문을 먼저 분석하여 패턴을 정의하고, 그 다음에 영어의 대역 패턴을 찾아나가는 방식을 사용하면 매끄러운 대역 영어 패턴을 찾기가 어렵다는 것이다. 한국어 패턴을 정의하는 부분에서 영어로의 대역 패턴을 매칭할 때 발생할 수 있는 여러 가지 문제점을 함께 고려해 나가는 것이 패턴 정의에서 훨씬 용이하다.

### 3 기능동사 구문

한영 기계번역의 패턴 추출 단계에서 한영 패턴 매칭의 효율성을 최대화하기 위해, 한국어와 영어에서 나타날 수 있는 범 언어적인 문법 사항들을 동시에 고려하여 분석해 나가는 방식을 취했다. 3장에서는 분석 과정에서 다른 동사구문과 변별적인 특성을 나타내는 기능동사<sup>1)</sup> 구문에 대해 언급하고자 한다. 먼저 기능동사 구문의 범 언어적인 문법 특성에 대해 살펴보고 이러한 특성에 따라 기능동사 구문의 패턴 선택 모델을 제안하고자 한다. 다음으로 기능동사 구문의 중심어인 술어기능명사의 한국어에 나타나는 개별적인 자질에 대해 알아보고, 마지막으로 이러한 다양한 자질을 토대로 추출된 기능동사 구문 모델의 적용 효율성을 분석해 보고자 한다.

#### 3-1. 구문 분석에서의 문제 제기

중심어를 기반으로 한 패턴 추출에서 동사구의 중심어 개념에 대한 한가지 의문이 제기된다. 다음 (1)의 두 문장에 대한 패턴 추출을 가정해 보자.

- (1) a. 그 선수가 운동한다.
- b. 그 선수가 운동을 한다.

위의 (1-a)에서는 명사구 혹은 명사를 NP라 두고 그 패턴을 (NP+가 운동하)로 분석할 수 있으며, '운동하-'라는 동사를 중심어로 한 동사구로 패턴을 추출할 수 있다. 즉, 이 패턴은 중심어 '운동하-'에 주격보어 NP가 선행되는 것이다. (1-b)는 (NP1+가 NP2+를 하)라는 패턴의 VPS이며 '하-'라는 동사가 그 중심어가 된다. 또한 이 패턴을 넓게 인식하고자 할 때에는 NP2의 자리에 올 수 있는 일반적인 명사의 자질을 기술할 수 있을 것이고, 좀 더 세분화하여 자료를 수집하고자 한다면 삽입 가능한 다양한 자질 중에 '운동'의 의미자질을 부여할 수 있을 것이다. 그런데 과연 (1)의 두 문장이 서로 다른 패턴으로 분리되어야 할 다른 의미의 문장들인가 하는 것이다. 위의 두 가지 선택 방식과는 무관하게 (1-a)와 (1-b)는 '운동하-'와 '운동을 하-' 동사를 지니는 것으로, 형식상으로나 의미적으로도 비변별적(indistinctive)이다. 따라서 (1)의 두 문장에 대해서는 일반적으로 다루어지는 중심어와 논항을 기반으로 하는 패턴 추출 방식을 이용한다면 효율적인 패턴을 얻을 수가 없다. (1)의 두 문장은 서로 다른 패턴으로 분류될 뿐만 아니라 서로 다른 자질들을 보유하게 되어 결국 전체적인 패턴의 수를 늘리기만 할 것이며 정확성도

1) 서정수(1975;1996) 등에서는 형식동사(empty verb)라는 명칭으로 사용되고 있지만, 영어와 불어 등에서 사용되는 기능동사(support verb ; verbe support) 개념과 동일한 의미임을 감안, 용어의 일관성을 유지하기 위해 기능동사라는 용어로 대표한다.

떨어질 것이다.

### 3-2. 기능동사 구문의 패턴

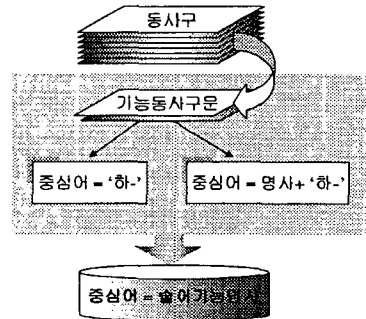
3-1에서 언급한 기능동사 구문의 이러한 특성은 한국어에서 뿐만이 아니라 영어나 불어에서도 나타나는 범 언어적인 특성으로 패턴 선택 모델의 제안시 다국어 기계번역 모델에의 적용 가능성의 장점을 지닌다. 불어에서 흔히 발견할 수 아래 (2)의 세 문장을 서로 비교해 보자.

- (2) a. Max a *désiré* partir.
- b. Max a *le désir* de partir.
- c. Max est *désireux* de partir.

예문 (2-a)에 비해 (2-b)와 (2-c)의 문장에서는 논항의 수나 형태와 분포를 결정짓는 것이 동사 avoir나 être가 아니다. (2-b)에서 논항을 결정짓는 것은 *désiré*라는 명사가 술어가 되며, (2-c)에서는 형용사 *désireux*가 술어적 요소가 된다. 1960년대 말 Maurice Gross를 중심으로 자연언어의 기계적 처리를 위한 언어 자료 축적을 목적으로 시작된 경험주의적 언어 이론인 어휘문법(*lexique-grammaire*)에서는 (2)의 예문에서처럼 특히 술어 역할을 하는 핵요소(*élément noyau*)가 명사인 경우, 그 명사를 술어기능명사(*noms prédictatifs, prédicats nominaux*)<sup>2)</sup>[4]라 부른다. 그리고 이와 같이 동사 이외의 다른 요소가 술어로 작용할 때, avoir나 être와 같이 문장을 이루기 위한 지주(*support*) 역할만을 담당하는 동사는 기능동사(*verbe support*)라 부른다. 기능동사는 동사 이외의 요소가 술어 기능을 담당할 때, 한 문장을 이루기 위해 사용되는 형태, 통사적인 도구이며, 시제(*temps*)나 인칭(*personne*), 수(*nombre*) 그리고 상(*aspect*) 등을 나타내는 표지 역할만을 담당한다. 이러한 술어기능명사와 기능동사로 구성된 문장이 다른 언어로 변환될 때는 동사가 아닌 명사가 술어적 핵요소(*élément prédictatif noyau*)로 피번역어의 기능동사와 논항 구조를 선택한다는 특징이 있다.[1]

한국어의 기능동사인 (1)의 두 문장에 있어서도 마찬가지로, (1-b)에서 중심어는 '하-'라는 동사가 아니라 명사인 '운동'으로 술어기능명사로서의 논항의 수나 형태를 결정짓는다. 그러나, (1-b)의 문장에 술어기능명사 개념을 도입하여 '운동'을 중심어로 선택하고 패턴 추출을 하더라도, (1-a)와 의미가 동일한 두 문장이 서로 다른 패턴으로 분류되어야 한다는 문제점은 여전히 남게 된다. 패턴 선택이 형태-통사적인 측면뿐만이 아니라 의미적인 측면까지 포함하고자 한다면 패턴이 부정확할 뿐만 아니라 추출된 패턴을 처리하는 방식에 있어서도 복잡한 형태를 이루게

된다. 따라서, (1)의 두 문장에 나타나는 동사를 모두 하나의 기능동사로 처리하고 이러한 기능동사로 이루어진 구문을 기능동사 구문(*support verb constructions*)으로 정의하여 술어기능명사를 중심으로 패턴을 추출하는 방식을 선택하면 문제점을 해결할 수가 있다. 즉, (1-a)와 (1-b)를 동일한 하나의 문장이면서 술어기능명사의 분리성의 특징으로 인하여 (1-a)의 문장이 (1-b)의 문장으로 그 형태만을 바꾼 것으로 본다. (1)의 두 문장은 <그림 2>에서 보는 바와 같이 모두 (NP+가 운동하)라는 하나의 패턴으로 처리될 수가 있을 것이다.



<그림 2 : 기능동사 구문의 중심어 선택 모델>

(1)의 두 문장에서 술어적 핵요소는 '운동'이며, 이 '운동'이라는 술어기능명사가 주어와 동사, 두 논항을 부차적인 요소로 필요로 하여 하나의 문장을 형성한다. 즉, '하-'는 하나의 문장을 이루기 위한 형식적인 역할만을 수행하는 기능동사이다.

한국어에 있어서 이러한 기능동사 구문이 '하-' 동사 구문에 모두 적용되는 것은 아닌데, 위 (1)의 문장에 여러 가지 변형 규칙을 적용한 (1')의 문장과 이와 비교되는 (3)의 문장을 함께 살펴해보도록 하자.

- (1') a. 그 선수가 운동을 한다.
- b. 그 선수가 하는 운동
- c. 그 선수의 운동
- (3) a. 그 분이 머리를 한다.
- b. 그 분이 하는 머리
- c. \*그 분의 머리[10]

예문 (1'-a)에서 (1'-b), (3-a)에서 (3-b)로의 관계절화 변형에서는 문장의 의미가 그대로 유지되고 있으나, (1'-c)로의 명사화 변형에서는 그 의미가 같지만, 같은 변형을 거치는 (3-c)에서는 동의관계가 성립되지 않음을 알 수 있다.

다음의 (4)-(5) 문장에서도 (1)의 문장과 마찬가지로 '하-'가 형식적인 기능으로 문법적 형식만을 갖추기 위한

2) 서술명사, 명사적 술어 등의 용어로도 쓰이지만, 명사가 문법적 기능면에서 술어 역할을 한다는 의미에서 '술어기능명사'라는 명칭을 사용한다.

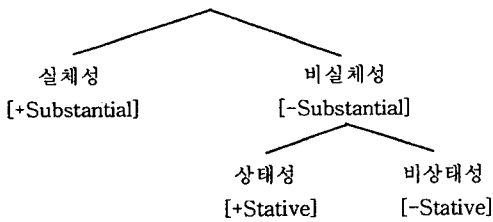
요소라는 것을 쉽게 짐작할 수 있다.

- (4) a. 그렇게 된 것을 참으로 다행하다.  
b. 그렇게 된 것은 참으로 다행이다.
- (5) a. 우리도 무사하고 너희도 무사하다.  
b. 우리도 무사 너희도 무사[11]

그렇다면 한국어에서 기능동사로서의 ‘하-’를 선행하는 요소인 술어기능명사는 어떠한 개별적인 특징을 지나고 하나의 문장에 지배 관계를 이루는가. 다음의 3-3에서는 한국어 기능동사 ‘하-’와 함께 쓰이는 술어기능명사의 중심어로서의 가능한 의미자질에 대해서 살펴보고자 한다.

### 3-3. 한국어 술어기능명사의 특징

명사는 그 의미적 특질에 따라 다음의 <그림 3>과 같이 분류될 수 있다.[10]



<그림 3 : 명사의 분류>

실체성 요소는 흔히 ‘제1종 명사류(first order nominal)’라 불리는 것으로, 명사 고유의 특질만을 가지고 있는 의미적으로 순수한 명사만의 특색을 보이는 실질 명사이다. 이러한 명사들은 술어적 의미를 전혀 내포하고 있지 않으므로 기능동사 역할을 하는 ‘하-’와 함께 나타나지 않는다.

반면 비실체성 명사는 ‘제2종(second-order) 명사류’로, 명사 고유의 특질 외에도 사물의 움직임이나 상태, 성질의 의미를 포함하는 명사이며, (6)-(10)의 예에서 보듯이 의미적으로는 동사나 형용사뿐만이 아니라 심지어 부사의 속성까지도 내포한다.

- (6) a. 절, 곳, 간직, 차지  
b. 놀음, 다짐, 칼부림, 내기, 놀이  
c. 손질, 도둑질, 서방질, 계집질  
d. 현대화, 기계화, 민주화  
e. 운동, 활동, 결정, 거부, 검사, 거수, 식사  
f. 스타트, 스톱, 무브, 미팅
- (7) 사망, 변화, 존재, 건조, 성장, 고정, 액화, 고체화
- (8) 번쩍번쩍, 삐걱삐걱, 달랑달랑, 철렁철렁
- (9) a. 암전, 점잔, 가난, 고요, 조용  
b. 미안, 다정, 간결, 확실, 편리, 열심, 정직

(10) 울긋불긋, 미끌미끌, 반들반들, 동글동글[11]

(6-a)는 순수 우리말 명사이고, (6-b)는 명사화에서 온 명사이다. (6-c)는 명사+‘질’의 형태인데, 실체성 명사 ‘손, 도둑’ 등에 ‘-질’이 첨가되던 비실체성으로 바뀐다. (6-d)는 ‘화(化)’를 첨가하여 한자어 실체성 명사를 비실체성으로 만드는 것이다. (6-e)는 비실체성 한자어 명사의 경우인데, 사실상 비실체성 명사는 한자어에서 온 것이 대다수이다. (6-f)는 ‘스타트’, ‘스톱’ 따위와 같은 외래어로, 그 뜻이 비실체성이다. (7)은 작용성 명사 곧 사물의 수동적 변화 등을 나타내는 말이다. (8)은 동작성이나 과정성에 관련된 부사성 형태이다. (9)는 상태성 명사의 예이다. (9-a)는 순수 국어 명사이고 (9-b)는 한자어 명사이다. (10)은 부사 형태의 상태성 어기이다.

일반적으로 (6)-(10)의 비실체성 명사들은 (11)에서처럼 립 동사 ‘하-’와 어울리는 선행요소이다.

- (11) a. 운동+하다, 연구+하다, 노래+하다, 썸+하다  
b. 성장+하다, 변질+하다, 생각+하다, 발전+하다  
c. 번쩍번쩍+하다, 삐걱삐걱+하다  
d. 가난+하다, 행복+하다, 불행+하다, 친절+하다  
e. 울긋불긋+하다, 반들반들+하다[11]

그렇다면, 위의 (1)과 비교하여 (6)-(10)의 예로 나타나는 명사 중에서 명사 ‘운동’과 ‘하-’ 동사간의 관계와 마찬가지로 명사 뒤에 ‘-을/를’ 등의 목적격 조사를 부과하여 ‘하-’ 동사 구문과 결합과 분리가 가능한 요소들은 어떤 것들인가. (6)과 (7)의 명사들은 이러한 ‘하-’의 결합성과 더불어 분리성이 가능한 요소이거나 넓은 의미에서 가능하다고 짐작되어 지는 명사류이다. 구 단위 패턴 기반 한영 기계번역에서 요구되는 ‘하-’ 기능동사 구문의 중심어로 선택되는 명사류는 이와 같이 비실체성의 특징을 지닐 뿐 아니라 기능동사와의 분리성의 특징까지 포함하는 명사류가 된다. 반면 (8)-(10)은 ‘하-’와 결합 가능하긴 하지만 목적격 어미로 분리가 불가능한 명사류로 본 연구에서 제시하고자 하는 중심어 명사류에는 속하지 않는다.

### 3-4. 기능동사 구문 모델의 적용 효율성 분석

위에서 제안한 ‘하-’ 기능동사 구문의 효율성을 분석하기 위해 사용된 코퍼스는 상업분야에서 편지글로 쓰여진 한영 대역 코퍼스 224개 문장을 기반으로 한다. 전체 코퍼스에서 발췌한 한영 대역 패턴 중 동사구문 패턴은 총 457개였으며, 그 중 ‘하-’ 어절을 포함한 동사구문의 패턴은 138개 구문이었다. ‘하-’ 어절을 지닌 동사구문 패턴 중에 기능동사로 쓰인 ‘하-’를 포함하는 기능동사 구문은 <표 2>에서 보는 것처럼 102개 문장으로 전체 ‘하-’를 포함하는 동사구문의 약 73.9%에 해당되는 높은 적용 가능성을 보였다. 다음으로, 코퍼스에서 나타나는 기능동사

'하-' 구문에 본 연구에서 제안하는 중심어 선택 모델의 적용 후, 'NP+하-' 구문과 'NP+를 하-' 구문 모두를 포함할 수 있는지의 적용 가능성과 이 모델을 적용함으로써의 문법적 용인가능성(acceptability)을 검토하였다. 그 결과 <표 2>에서 보는 바와 같이 전체 기능동사 '하-' 구문에서 98.1%가 용인됨을 알 수 있었다.

전체 한영 대역 코퍼스 수	224개
동사구문 패턴	457개
'하-' 동사구문 패턴 수	138개
기능동사 '하-' 구문 패턴 수	102개
기능동사 '하-' 구문 모델의 적용 가능성	73.9%
용인 가능성(acceptability)	98.1%

<표 2 : 기능동사 '하-' 구문 모델의 전체 본포와 용인가능성>

<표 2>에서 98.1%의 용인가능성으로 포함되지 않는 1.9%의 문장도 비문으로 판단되는 문장은 아니다. 다음 (12)의 문장에서처럼 직관적으로 약간의 어색함을 내포한 문장으로 효율성의 정확도를 위해 제외시켰다.

- (12) a. (NP+를 동봉하)  
 b. ?(NP+를 동봉+를 하)

문장 패턴 분석 중 실제로 '연설을 하-'와 '연설하-', '부탁을 하-'와 '부탁하-', '예약을 하-'와 '예약하-' 등의 많은 동사 구문들이 중복 사용되고 있었다. 중복되어 나타나는 구문은 본 논문이 제시한 기능동사 구문의 선택 모델로 처리하여 술어기능명사를 중심으로 하나의 기본 패턴으로 처리하였다. 본 연구에서 제안한 선택 모델을 사용하였을 때와 기능동사 구문의 중복 처리하지 않은 경우를 비교한 결과 38.2%의 패턴 처리의 절감 및 효율성을 나타내었다. 기능동사 구문의 용인 가능성이 98.1%의 높은 수치를 보인다는 사실을 감안하여 볼 때 코퍼스의 수가 증가하여 중복 사용되는 경우를 더 많이 추출하면 할수록 중복 패턴의 추출 용례는 늘어날 것이고 그에 따라 자연히 패턴 처리의 절감도 및 효율성은 증가할 것이다.

#### 4 맺는 말

한국어가 다른 언어에 비해 주어의 생략이 잦다는 점과 어순이 자유롭다는 특징은 구 단위 패턴 기반 기계번역의 패턴 추출에 있어서 패턴을 설정하는 단계에서 먼저 고려되어야 할 대상이다. 특히 한 문장이어의 논항의 수와 형태를 결정짓는 중심어 개념을 도입한 패턴 기반 번역

방식에 있어서 기능동사 구문과 같이 중심어의 선택에 특성을 갖는 경우는 패턴 추출 이전에 고려되어 효율적으로 분석되어야 한다. 기능동사 구문을 술어기능명사를 중심으로 패턴 분석한 경우는 그렇지 않은 경우와 비교하여 패턴 추출 단계에서 월등한 효율성을 나타내었다.

향후 이러한 연구는 기능동사 구문뿐만이 아니라 동사 이외의 다른 품사 구문에 있어서는 물론 특수성을 나타내는 다른 모든 구문에 대한 연구로 이어져야 할 것이다. 또한 다국어로의 기계번역 시스템을 위해 범 언어적인 특성을 가지는 경우와 개별 언어적인 특성을 보이는 자질로 분류되어 정보를 제공할 수 있어야 할 것이다. 이러한 정보들이 기계번역의 패턴 추출 작업에 필수적인 정보베이스로 쌓여감으로서 더욱 정확하고 효율적이 시스템을 뒷받침할 수 있을 것이다.

#### 참고문헌

- [1] Danlos, L. et Laurens, O. (1991), Eurotra: Présentation de projet EUROTRA et des grammaires d'EUROTRA-France, Rapport Technique n°1, CNRS-Paris 7.
- [2] Giry-Schneider, J. (1978), Les constructions du verbe faire.(L'opérateur faire dans le lexique), Genève : Droz.
- [3] Grevisse, M. (1969), Le Bon Usage, Gembloux : Duculot.
- [4] Gross, M. (1981), Les bases empiriques de la notion de prédicat sémantique, Langages 63, pp.7-52.
- [5] Jackendoff, R. (1977), X-bar Syntax: A study of Phrase Structure, Cambridge, Mass. : MIT Press.
- [6] 김경원(1995), 한국어 문장 분석을 위한 동사 패턴의 인식에 관한 연구, 서울대학교 석사학위 논문.
- [7] 김정재(1999), 두단계 대역어선택 방식을 이용한 구단위 패턴기반 한영 기계번역 시스템, 제11회 한글 및 한국어 정보처리 학회(발표예정).
- [8] 김혜경(1996), 불어의 술어기능명사와 한국어 변환, 부산대학교 석사학위 논문.
- [9] 박효명 (1998), 핵어문법론 I, 한국문화사.
- [10] 서정수 (1975), 동사 "하-"의 문법, 형설출판사.
- [11] 서정수 (1996), 개정 현대 한국어 문법 연구의 개관, 한국문화사.
- [12] 우형식(1998), 국어 동사 구문의 분석, 태학사.
- [13] 이상조(1994), 한.영 기계번역을 위한 중심어 기반 구조 변환, 서울대학교 박사학위 논문.
- [14] 장석진(1993), 정보기반 한국어 문법, 언어와 정보.
- [15] 홍재성 외(1997), 현대 한국어 동사 구문 사전, 두산동아.