

## 하위범주화에 의한 한국어 파싱 설계

이 호 석  
뉴미디어학과 공과대학 호서대학교  
hslee@office.hoseo.ac.kr

### A Design of Korean Language Parsing based on Subcategorization

Ho Suk Lee  
New Media Dept. College of Engineering Hoseo University

#### 요 약

본 논문은 하위범주화에 의한 한국어 파싱의 설계를 논의한다. 우선, 기계 번역에 중요한 영향을 끼치는 한국어 문법 구성 요소인 통사 범주, 조사, 어미, 통사적 접사, 의존 명사 등을 논의하고 하위 범주화와 표현 패턴을 논의한다. 다음에 간단한 한국어 파서의 기본 골격을 제시한다. 첫 번째 과정은 입력 단계로서 문장을 스캐닝(scanning) 하고 관사, 명사, 수사, 통사적 접사, 조사, 의존 명사, 어미활용, 형용사, 부사, 조동사 등을 처리한다. 두 번째 과정에서는 하위범주화 패턴과 표현 패턴을 처리한다. 세 번째 단계에서는 절을 처리하고, 네 번째 단계에서는 SEA(Sentence Ending+Auxiliary) 를 처리한다.

#### Abstract

This paper discusses a design for Korean language parsing based on subcategorization. First, we discuss some important Korean grammar elements such as syntax category, josa, omi-conjugation, syntactic affix, dependent noun and also discuss subcategorization and expression patterns. Then, we show the basic structure of Korean language parsing process. The first stage scans the input sentence and processes article, noun phrase, numeral, josa, affix, dependent noun, adjective, omi-conjugation, adverb, auxiliary verb. The second stage deals with sub-categorization patterns and expression patterns. The third stage processes the clauses and the fourth stage deals with SEA(Sentence Ending+Auxiliary).

#### 1. 서 론

근래에 들어서, 전산학 분야에서 한국어 분석에 대한 연구가 다시 진행되고 있는 것 같다[1]~[11]. 바람직한 현상이라고 할 수 있다. 참고 문헌 [1]은 대량의 말뭉치로부터 추출한 어휘들의 공기 관계를 바탕으로 문장 내 어절 간의 연관도를 전체적으로 관리할 수 있는 전역적인 연관 테이블을 구성하였다. 그리고 이를 바탕으로 한국어 파싱 방법을 제안하였다. 참고 문헌 [2]는 말뭉치로부터 한국어 구문 분석의 효과성 향상에 적합한 제약 규칙을 자동으로 학습하는 방법을 제시하였다. 참고 문헌 [3]은 기존의 확률적 LR 파싱 기법의 문제점을 극복한 조건부 연산 모델 기법을 제안하였다. 참고 문헌 [4]에서는 한국어 구문 분석 결과로 생성되는 구문 트리의 표현을 위한 복합 레이블 생성 방법을 논의한다. 복합 구문 레이블은 자동으로 생성되며, 두 개의 하위 구성체의 품사 정보를 적절히 구성하여 형성된다고 한다. 참고 문헌 [5]는 명사구와 용언구의 문법 관계 확률을 이용하여 용언구의 술어-논항 관계의 중의성을 해소하는 통계적 방법을 제안하였다. 제안된 방법은 자동으로 문법 관계를 학습한다고 한다. 참고 문헌 [6]은 한국어 구문에서 의존 관계를 결정하기 위하여, 서로 의존 관계에 있는 두 어절인 의존소와 지배소 사이의 수식 거리 확률 모델을 제시하였다. 이 논문에서 논의하는 의존소와 지배소 관계는 한국어 문장에서 매우 보편적인 관계로 보인다. 참고 문헌 [7]은 한국어 문장에 있어서 내포문의 범위를 정하고 이를 이용하여 복문을 내포문과 주절로 분할하여 구문 분석의 중의성을 대폭 감소시켰다고 한

다. 참고 문헌 [8]은 병렬 구조의 해석을 위한 비지도식(unsupervised) 언어 확률 모델을 제시한다. 모델은 병렬 구조의 대칭성과 상호교환성에 근거하여 병렬 구조를 확인한다. 제안된 모델은 “와/과”로 구성된 한국어 명사구 병렬 구조를 해석하는데 사용되었다. 참고 문헌 [9][10]은 핵이 후미에 존재하고 어순이 자유로운 언어의 경우에 의존 문법을 사용하여 구문의 수식 관계를 결정하는 방법에 대하여 논의하고 있다. 참고 문헌 [11]은 의존 문법 개념에 입각하여 개발된 한국어 구문 분석 시스템을 보여 준다. 참고 문헌 [12]는 미국에서 개발된 통계적 방법을 적용한 한국어-영어 기계 번역 시스템을 보여준다. 참고 문헌 중에는 [6][9][10][11]이 주목할 만하다. 특히, [6]번 논문에서 제시한 의존소와 지배소의 개념은 서술어의 하위범주화 개념을 함께 사용하여 완전한 구조를 갖추는 것으로 생각할 수 있다. 문헌 [9][10]은 서술어의 하위범주화 개념은 고려하지 않고 의존 문법 개념만을 사용하여 [6]번 문헌의 생각을 발전시킨 것으로 생각된다. 그러나 의존 문법 개념에는 서술어의 하위범주화 개념이 결합가(valency) 개념으로 함께 구성되어 있다고 할 수 있다[13]. 즉, 의존 관계와 하위범주화 개념은 동일한 문법 현상을 수식적인 관점과 서술어의 결합가 관점에서 서술한 것으로 생각할 수 있다. 본 논문에서는 한국어-영어 기계 번역을 위한 새로운 한국어 파싱 시스템의 기본 구조에 대하여 논의한다. 파서는 주로 의존 문법에서 논의하는 술어(동사와 형용사)의 하위범주화 개념, 한국어의 특징인 조사의 격 개념과 술어 어미의 활용, 그리고 표현 패턴 개념을 기반으로

구축될 것이다. 표현 패턴 개념은 참고 문헌 [13]에서도 논의되었듯이 낱말들의 구성 전체가 하나의 의미 단위를 이루는 표현을 일컫는다. 표현 패턴은 한국어 자체에서도 관형적 표현으로 인식되어 상당수 존재하며, 한국어-영어 기계 번역에 있어서는 한국어와 영어의 표현 대조를 통하여 더 많은 종류의 표현 패턴을 정의할 수 있다.

2. 한국어 문법

기계 번역에서 의미의 전달에 중요한 역할을 하는 한국어 문법의 기본 요소들을 논의하기로 한다.

2.1 통사 범주

한국어 낱말의 통사 범주는 참고 문헌 [13]을 참조하여 기초 통사 범주와 투사 통사 범주로 구분한다. 예를 들어, 문헌 [13]의 4장에서는 X-bar 이론에 입각하여 다음과 같이 통사범주가 정의되어 있다. 일반 명사의 형태 표지는 nng이고, 기초 통사 범주는 N이고, 그리고 최대 투사 통사 범주는 NP이다. 일반 동사의 형태 표지는 vvg이고, 기초 통사 범주는 V이고, 그리고 최대 투사 통사 범주는 VP이다. 일반 형용사의 형태 표지는 vag, 기초 통사 범주는 A, 그리고 최대 투사 통사 범주는 AP이다. 일반 부사의 형태 표지는 mbg, 기초 통사 범주는 ADV이다. 주격조사의 형태 표지는 jks, 기초 통사 범주는 NK, 그리고 최대 투사 통사 범주는 NKP이다. 보격조사의 형태 표지는 jkc, 기초 통사 범주는 BK, 그리고 최대 투사 통사 범주는 BKP로 표시하였다. 어말어미의 형태 표지는 ef, 기초 통사 범주는 F, 그리고 최대 투사 통사 범주는 FP이다. 평서범어미의 형태 표지는 efd, 기초 통사 범주는 DF, 그리고 최대 투사 통사 범주는 DFP이다. ㄴ-관형사형 전성어미의 형태표지는 efds, 기초 통사 범주는 SDF, 그리고 최대 투사 통사 범주는 SDFP로 하였다. 문헌 [13]에 제시된 어휘 부류의 형태 표지 성분의 통사 범주화의 일부 예는 다음과 같다.

- (1) X/nng(일반 명사) => N
- (2) X/nnbg(일반 의존 명사) => N
- (3) X/nr(수사) => NP

2.2 조 사

국어의 조사에는 주격(jks), 보격(jkc), 속격/관형격(jkg), 목적격(jko), 부사격(jkb), 처격/장소격/위격/기준점격(jki), 조격/구격/방법격/재료격/도구격/경로격(jki), 출격/탈격/출발점격(jkso), 달격/향격/방향격/목표격/도달점격(jkt), 공격/공동격/여동격(jkc), 비교격(jkm), 자격격(jkf), 원인격(jku), 변위격/변성격/결과격(jkr), 호격(jkv), 인용격(jkq), 서술격(jkp), 보조사(jx), 앞보조사(jxb), 가운데보조사(jxm), 뒤보조사(jxp), 접속조사(jc) 등이 조사되어 있다 [13]. 기타 특사한 조사로서는 “오늘따라”의 “따라”, “떡은커녕”의 “은커녕” 그리고 관형적인 용법으로 부사 “고사하고”와 함께 쓰이는 “은/는 고사하고” 용법이 있다. 또한 참고 문헌 [14]에서도 조사에 대하여 다음과 같이 설명하고 있다. 조사는 한국어 문장에서 기능적으로나 의미적으로 매우 중요한 역할을 수행하는 품사로서 격조사와 보조조사로 나눌 수 있다. 격조사에는 주격조사(은, 는, 이, 가, 도, 께서), 서술격조사(이다), 목적격조사(을, 를, 도), 보어격조사(이, 가, 과, 와), 관형격조사(의), 부사격조사, 접속격조사(와, 과) 등이 있다. 그리고 부사격조사를 다시 세분하여 살펴보면, (1) 장소에 관한 움직임이나 상태를 나타내는 것, (2) 변동에 관한 움직임을 나타내는 것, (3) 움직임이나 상태의 형편을 꾸미게 하는 것 등이 존재한다. 장소에 관한 움직임이나 상

태를 나타내는 부사격 조사에는 처소격(에), 정처격(에서), 출처격(에서, 에게서, 한테서), 출발격(으로부터), 통과격(으로서), 방향격(에로, 에게로, 한테로, 께로), 부여격(에게, 한테, 께), 순위격(더러, 보고) 등이 존재한다. 변동에 관한 움직임을 나타내는 부사격 조사에는 재료격(으로, 으로서), 기구격(으로, 으로서), 변화격(으로), 원인격(으로, 에) 등이 있다. 그리고 움직임이나 상태의 형편을 꾸미게 하는 부사격 조사에는 자격격(으로, 으로서), 종류격(치고, 쳐놓고), 현상격(대로), 연장격(토록), 비교격(처럼, 같이, 만큼, 보다, 에서), 공동격(과, 와, 이랑, 하고, 으로, 더불어) 등이 존재한다. 거의 동일한 내용이 있지만 국어 학자에 따라서 용어의 사용이 다른 면이 있어서 제시하였다. 문헌 [13]에 제시된 조사(교착소)에 대한 형태 표지 성분의 통사 범주화의 일부 예는 다음과 같다.

- (1) X/jks(주격 조사) => NK
- (2) X/jkg(속격/관형격 조사) => GK
- (3) X/jko(목적격 조사) => OK

2.3 어 미

어미는 형용사와 동사의 활용에서 나타난다. 참고 문헌 [13]은 국어의 어미에 대하여 다음과 같이 분류하고 있다. 국어의 어미에는 어말어미(ef), 종결어미(ef), 평서범어미(efd), 감탄범어미(efx), 의문범어미(efq), 명령범어미(efm), 청유범어미(efo), 연결어미(efc), 명사형전성어미(efn), 관형사형전성어미(efad), ㄴ-관형사형전성어미(efsd), ㄴ-관형사형전성어미(efnd), ㄷ-관형사형전성어미(eftd), ㄹ-관형사형전성어미(efrd), 선어말어미(ep), 주체높임선어말어미(ehn), 시제선어말어미(et), 현재시제선어말어미(ent), 과거시제선어말어미(ept), 대과거시제선어말어미(eppt), 미래시제/양태선어말어미(eft), 회상시제선어말어미(ett), 경양선어말어미(ehm), 공손선어말어미(ers) 등이 있다. 문헌 [13]에 제시된 어미 표지 성분의 통사 범주화의 일부 예는 다음과 같다.

- (1) X/eptn(현재 시제 선어말 어미) => NT
- (2) X/eptp(과거 시제 선어말 어미) => PT
- (3) X/eprs(공손 선어말 어미) => RS

또한 학자에 따라서 종지법, 접속법, 전성법 혹은 종결서술형, 연결형, 관형형, 명사형 혹은 종지 서술형, 접속 서술형, 관형형, 부사형, 명사형 등으로 의미와 문법 기능상 분류된다. 이 중에서 두 번째의 분류가 무난한 것으로 보인다. 이 경우에 종결 서술형 어미는 평서법, 의문법, 청유법, 약속법, 감탄법, 명령법, 선어 말어미로 구분되며, 연결형 어미는 대등법, 종속법, 보조법으로 구분되며, 관형형 어미는 “ㄴ(는), ㄹ(을)”로 구분되고, 명사형 어미는 정법인 “ㄹ(은)”, 미정법인 “기”로 구분된다[14].

2.4 통사적 접사

국어에는 통사적 접사가 존재한다. 접사에는 접두사와 접미사가 있다. 접두사에는 “-대”가 있다. 접미사에는 “-경, -계, -께, -꼴(1000원꼴), -끼리, -답, -당, -들, -말, -별, -분(2인분), -산, -상, -시, -식, -씩, -여(100여명), -가량, -용, -적, -중, -주의, -째, -쯤, -초, -파, -형.” 등이 존재한다[13]. 이 중에서 “-답, -들”을 제외한 접사들은 명사(N)로 간주할 수 있다. 다음은 접두사 “대”와 접미사 “적, 당”의 예이다. “들”은 일반 조사로 그리고 “답”은 서술격 조사로도 간주될 수 있다.

- (1) “대”의 예  
대 일본 수출이 감소하였다.  
(Export to Japan has decreased.)
- (2) “적”의 예  
한국의 시골 양반적인 사고방식.  
(Korean country gentleman's thinking.)
- (3) “당”의 예  
1인당 3만원씩 주었다.  
(I gave 3000won to each person.)
- (4) “들”의 예  
아이들에게 좋은 선물들을 주었다.
- (5) “답”의 예  
그는 우리 회사 사원답다.  
(He is good to be our company's employee.)

2.5 의존 명사

국어의 의존 명사는 “본질의 뜻은 없고 다만 문법적 기능이 명사와 동일하다.”로 생각할 수 있다[15]. 국어의 의존 명사에는 1음절어와 2음절어가 있다. 의존 명사에는 “간, 것, 게, 경, 길, 김, 나름, 나위, 녀석, 년, 녀, 노릇, 놈, 따름, 따위, 탄, 때문, 대로, 덧, 데, 동안, 동, 들, 돛, 등, 등지, 리, 마련, 만, 만큼, 말(씀), 모양, 무렵, 바, 바람, 밖, 빨, 법, 뻘, 뿔, 분, 뿐, 상, 서술, 석, 성, 셈, 손, 수, 심상, 양, 이, 이래, 일췌, 자, 짝, 적, 조, 족족, 쪽쪽, 줄, 즈음, 즉, 지, 지경, 직, 차, 참, 채, 척, 체, 축, 치, 터, 턱, 통, 판, 편, 폭, 품, 한, 해” 등이 존재한다[15]. 그런데 “중”도 의존 명사로 사용되는 것 같은데 문헌 [15]에서는 제시되어 있지 않아서 저자가 새로 추가하였다. 예문은 아래에 (6)에 제시된다. 다음에 “만큼” 경우의 예가 있다[15].

- (1) 생선은 살짝 익을 만큼(만) 찌야 담백하다.
- (2) 그가 먹는 만큼 나도 먹을 수 있다.
- (3) 그 사람만큼은 나를 믿어줄 줄 알았다.
- (4) 기왕 말이 나온 만큼 이 참에 결정을 합시다.
- (5) 산림보호가 이루어지니 만큼 나무들이 굶었다.

위의 예를 보면 “만큼”이 현대국어에서 의존명사로 분류되어 있으나 그 문법 기능과 어휘적 의미는 문장에 따라 다르다는 것을 알 수 있다. 다음은 “중”의 예이다.

- (6) 그는 일하는 중에 사고를 목격하였다.  
(While he was working, he witnessed an accident.)

의존 명사에 선행하는 관형어는 다음과 같다[15].

- (1) 관형어 : 이, 그, 저, 새, 현, 옛, 모든
- (2) 용언의 관형사형 :  
<용언+관형사형 어미 : ‘-ㄴ/-은, -ㄹ/-을’>
- (3) 체언의 관형사형 : <체언+의 : 책의, 내, 네>
- (4) 체언 자체
- (5) 용언의 명사형 : <용언+명사화소 ‘기’ : 먹기, 읽기>

또한 의존 명사는 대상성 의존 명사, 서술성 의존 명사, 부사성 의존 명사로 분류할 수 있다. 대상성 의존 명사는 ‘것’류 의존 명사, ‘쪽’류 의존 명사, ‘나위’류 의존 명사, ‘해’류 의존 명사로 분류된다. 부사성 의존 명사는 ‘바람’류 의존 명사, ‘돛’류 의존 명사로 분류된다. 서술성 의존 명사는 ‘뻘’류 의존 명사로 분류된다[14].

의존 명사의 문맥에 따라서 의미가 다르다. 그러므로 의존 명사를 번역할 때에는 문맥에 따른 의미를 파악하

여 번역하여야 한다. 일반적인 규칙으로는 사물을 의미하는 명사들과 함께 사용된 문장의 경우에는 영어로 번역하는 경우에 “It” 절로 시작하는 것이 비교적 의미의 전달이 자연스럽고, 사람을 의미하는 인칭대명사와 함께 사용된 경우에는 직접 사람을 의미하는 명사로 시작하는 것이 자연스러운 경우가 많다. 다음은 의존명사 ‘것’에 대한 번역 예이다.

- (1) 네가 어제 산 것은 양품이다. (사물)  
What you bought yesterday is good.  
It is good what you bought yesterday.
- (2) 바보 같은 것이 또 운다. (사람)  
You like a fool cry again.
- (3) 토끼를 잡아먹은 것은 호랑이이다. (동물)  
What ate an rabbit is a tiger.  
The animal which ate an rabbit is a tiger.  
It was a tiger which took and ate a rabbit.
- (4) 얼음이 언 것을 보았다. (현상)  
I see a frozen ice.
- (5) 얼음이 어는 것을 본다.  
I see an ice freezing.
- (6) 네가 찾아 온 것은 3일전 이었다. (사실)  
The day you visited me was 3 days ago.  
It was 3 days ago when you visited me.
- (7) 네가 괴로운 것을 알고 있다. (심정)  
I know that you are painful.
- (8) 그가 무사했다는 것이 다행이다.  
It is good that he is safe.
- (9) 철수가 숙제하는 것을 도와주어라.  
Help 철수 do the homework.
- (10) 저 사람이 노래를 하는 것이 이상하다.  
It is strange that he is singing.  
It is strange the way he is singing.
- (11) 빌딩을 가진 것은 흥사장이다.  
The man who owns a building is 흥사장.
- (12) 앞발이 짧은 것은 토끼다.  
The animal with a short forehand is a rabbit.  
It is a rabbit which has a short forehand.

다음은 그 밖에 의존 명사 ‘터, 따름, 뻘, 참, 나름, 김, 뿔’에 대한 번역 예이다[15].

- (13) 나는 내일 LA로 출장을 갈 터이다.  
I am going to LA tomorrow for business.
- (14) 그들은 가난하게 살아오고 있는 터였다.  
It is the fact that they have lived poor.
- (14') 그들은 가난하게 살아오고 있다.  
They have lived poor.
- (15) 송구스러울 따름입니다.  
I think that I am very sorry.
- (16) 요 며칠 흐렸을 뿐예요.  
It has been only cloudy these days.
- (17) 겨우 끝마쳤을 뿐이다.  
I barely finished it.
- (18) 아침을 뜰 참에 신문을 보게 되었다.  
At the moment I begin to have a breakfast,  
I see a newspaper.
- (19) 네가 할 나름이다.  
It is up to you.
- (20) 근처를 지나가는 김에 들렀다.  
I drop by as I pass the neighborhood.

(21) 그는 먼 친척뻘 된다.

He is one of distant relatives.

그러나 일반적으로 의존명사가 사용된 문장을 번역하는 것은 어렵다고 생각된다. 예를 들어서, 의존 명사 '것'의 번역을 위하여 필요한 요소로는 '것'을 수식하는 문장의 의미상 주어의 의미 특질(semantic feature), '것'의 목적어의 의미 특질, 영어 관계대명사 절의 구성 방법 등이 될 것이다. 또한 일부 의존 명사의 경우는 관형적인 표현도 있어서 표현 패턴을 이용한 번역을 적극적으로 고려해 보아야 할 것이다.

2.6 문장의 성분과 종류

문장의 성분은 주성분, 부속 성분, 독립 성분으로 나뉜다. 주성분은 문장의 성립에 필수적인 성분으로서 주어, 목적어, 서술어, 보어가 있다. 부속 성분은 문장의 필수적인 성분을 이루지 못하고 다른 성분에 딸려 있는 성분으로서 관형어, 부사어가 있다. 독립 성분은 문장의 다른 성분과 직접적인 관련이 없이 독립되어 있는 성분으로서, 이에는 독립어가 있다[14].

주어에는 정규적인 주어, 비정규적인 주어, 유사 주어가 있다. 서술어에는 결합가 개념이 있으며 동사나 형용사와 결합하는 논항의 개수를 일컫는다. 논항의 개수는 1가(한 자리 서술어), 2가(두 자리 서술어), 3가(세 자리 서술어), 4가(네 자리 서술어)가 있다. 세 자리 서술어의 예로는 “철수가 영희에게 편지를 보냈다.”가 있고, 네 자리 서술어의 예로는 “철수가 이 책을 영어에서 한국어로 번역했다.”가 있다[13].

목적어는 타동사가 나타내는 동작이나 작용의 대상이 되는 문장 성분을 말한다. 전형적인 목적어는 목적격 조사 “을/를”을 가진 명사구나 목적격 조사구에 의하여 이루어진다. 비정규적인 목적어는 보조사가 목적어로 해석되는 경우이다. 예로는 “철수가 밥을 먹었다.”, “철수가 밥도 먹었다.”, “철수가 밥만 먹었다.”, “철수가 밥까지 먹고 갔다.” 등이다. 그러나 “을/를”이 있다고 해서 항상 목적어가 되는 것은 아니다. 다음은 예이다. “그가 밥을 먹지를 못한다.”와 같이 연결어미 뒤에 나타나는 “을/를” 성분은 목적어로 보지 않는다. 분명히 자동사 구성에 나타나는 “을/를” 성분은 목적어로 보지 않는다. 다음은 예이다. “몸이 아파서 여기에 머물 수를 없다.” 피동사 구성에서 주어를 나타내는 성분 뒤에 나타나는 “을/를” 성분은 목적어로 보지 않는다. 다음은 예이다. “처녀가 사내한테 손을 잡혔다.” 시간, 빈도, 거리를 나타내는 표현에 “을/를”이 연결된 성분은 목적어로 보지 않는다. 다음은 예이다. “철수가 3일을 공부했다.” “철수가 우리 집에 3번을 왔다.” “철수가 10리를 뛰어왔다.”

관형어는 명사나 명사구를 수식하는 문장 성분을 말한다. 관형어의 종류에는 (1) 관형사 : 새 책, 여러 책, 이 책, 그 책, (2) 속격/관형격 조사구 : 나의 책, 우리들의 의무, 근무자로서의 책임, (3) 단순 명사구 : 철수 책, 김 사장, (4) 관형사절 : 큰 나무, 사람들이 보는 책, 철수가 도둑질을 한 사실, (5) 명사절 : 그가 오기 때문에, 그가 오기 전에, (6) 부사 : (6.1) 바로 오늘 그가 왔다, (6.2) 그는 아주 부자이다, (6.3) 오직 그만을 좋아한다. 등이 있다. 여러 개의 관형사가 겹쳐서 나타나는 경우가 있다. (1) “이 여러 새 책”과 같은 경우는 지시 관형사(DD) + 양화 관형사(QD) + 성상 관형사(SD) 순서로 관형사가 겹쳐서 나타나는 경우이다. (2) “저기 저 책”과 같이 장소 표현의 관형사가 반드시 지시 관형사보다 앞에 온다. (3) “눈이 초롱초롱한 어린 소년”에서 긴 관형사 절이 짧은 관형사 절보다 앞에 온다. (4) “철저한 그리고 확실

한 대답”과 같은 예는 관형사형 전성 어미구가 접속된 구성이 명사구를 수식하는 것으로 볼 수 있다[13].

부사어는 주로 서술어를 수식하여 서술어의 의미를 변형시킨다. 부사어는 관형어나 다른 부사를 수식하기도 한다. 부사어의 종류에는 (1) 부사 : 아주 행복하다, 많이 오다. (2) 부사격 조사구 : 마당에서 놀다. (3) 단순 명사구 : 자연 늦는다, 오늘 온다, 내일 간다. (4) 부사성 의존 명사구 : 가는 족족 성공하였다, 먹을 만큼 먹다, 닥치는 대로 읽는다. (5) 부사구 : 아무 소리 없이 비가 온다, 그와 달리 나는 바쁘다. 철수와 함께 토론하다. (6) 연결 어미구 : 철수가 고기를 구워 먹었다. 영수가 밥을 물에 말아서 먹었다. 오늘따라 오늘은 운수가 좋다. 등이 있다. 부사격 조사구에는 (1) 처격/장소격/위격/기준점격 조사구(LKP) : 철수가 집에 있다. 나는 철수를 찾기에 지쳤다. (2) 조격/구격/도구격/재료격/방법격/경로격 조사구(IKP) : 철수가 나무로 집을 짓는다. 나는 연필로 글을 쓴다. (3) 출격/탈격/원격/시원격/출발점격 조사구(SKP) : 철수가 집에서 나왔다. 이 이야기는 중국에서 유래하다. (4) 달격/향격/방향격/목표격/도달점격 조사구(TKP) : 철수가 영희에게 선물을 주었다. (5) 공격/공동격/여동격 조사구(CKP) : 철수가 영희와 논다. 설탕을 커피에 섞다. (6) 비교격 조사구(MKP) : 철수는 영희만 못하다. 얼굴이 꽃처럼 아름답게 생겼다. (7) 자격격 조사구(FKP) : 철수가 대표로 참석하였다. (8) 원인격 조사구(UKP) : 철수가 감기로 많이 아프다. 창수가 과로로 죽었다. (9) 변위격/변성격/결과격 조사구(RKP) : 색깔이 청색으로 변했다. 우리는 집에 가기로 결정하였다. 등이 있다[13].

국어 문장의 종류를 분류하는 기준에는 (1) 문장 성분의 구비 여부에 따른 분류 : 완전문, 불완전문, (2) 종결 형식의 전형성에 따른 분류 : 전형적 종결 어미문, 비전형적 종결 어미문, (3) 화자의 전달 목적에 따른 분류 : 평서문, 감탄문, 의문문, 명령문, 청유문 등이 있다. 완전문은 주어, 목적어, 서술어와 같은 문장의 주요 구성 성분을 완전히 다 갖춘 문장을 일컫는다. 불완전문은 문장의 주요 구성 성분을 완전히 갖추지 못한 문장이다. 전형적 종결 어미문은 예를 들어 “간다, 가느냐, 가는구나” 등과 같은 문장의 종결 형식을 말한다. 비전형적 종결 어미문은 예를 들어 “가거든, 같걸, 간다지” 등과 같은 문장의 종결 형식을 말한다[13].

문장의 전달 목적에 따른 분류에는 평서문, 감탄문, 의문문, 명령문, 청유문 등이 있다.

2.7 하위 범주화

문장 구성에 있어서 동사, 형용사, 부사(“고사하고”는 별도 처리)의 하위범주화 정보는 반드시 필요하다. 하위 범주화 개념은 결합가 개념과 관련이 있으며 의존 문법에서 사용하는 개념이다[13]. 이 개념은 동사나 형용사에 있어서 그 의미 특질을 완성하기 위하여 필요한 보족어(논항)의 존재 및 개수를 말한다.

하위범주화 정보를 활용하여 완전한 문장을 인식하고 그 결과를 의존 관계로 구성하여 한국어 구문을 확인하고 완성한다. 다음은 하위범주화에 대한 예이다[13].

- (1) 철수가 학교에 간다.  
철수가 교실에서 운동장으로 간다.  
-가(동사) : <NKP, (SKP), TKP(TK=에/로)>  
NKP : 주격조사(은/는/이,가),  
SKP : 출격조사(로부터/에서),  
TKP : 달격조사

- (2) 그가 영희에게 선물을 주고 말했다.  
- 주(동사) : <NKP, TKP, OKP>  
- 말(동사) : <NKP, CFP(CF=고)>
- (3) 그것은 남이 보기 심상이다.  
그것은 남이 보기에 심상이다.  
그것은 남이 보기가 심상이다.  
- 심상이(형용사) : <NKP, (NKP), CFP(CF=' '/에/가)>

**2.8 표현 패턴**

다음은 참고 문헌 [13]에서 제시한 한국어의 관형적 표현으로 표현 패턴 개념을 이용하여 번역하여야 한다.

- (1) 아닌게 아니라, 그도 그럴 것이, 그것도 그럴 것이 등
- (2) 바꿔 말하면, 다시 말하면, 달리 말하면, 등
- (3) 뿐만 아니라, 그 위에, 등
- (4) 그렇지 않으면, 그렇지 않아도 등
- (5) 예를 들면, 예를 들자면, 예를 들어, 등
- (6) 그럼에도 불구하고, 그런데도 불구하고, 등
- (7) 그리고 나서, 그리고 나서, 그런 다음, 그리고 나자, 등
- (8) 말할 것도 없이, 더 말할 필요 없이
- (9) 그런 고로, 그런 까닭에, 그런 이유로, 그런 연유로, 등
- (10) 이루 말할 수 없다, 말할 수 없이, 더할 나위 없다
- (11) 아무 쓸모 없다, 아무 쓸데 없다, 할 수 없다,

국어에는 이러한 관형적인 표현이 상당히 많이 존재하며, 한국어-영어 기계 번역의 관점에서는 더 많은 종류의 표현 패턴을 찾을 수 있다.

**2.9 국어 문장의 종결 어미**

국어 문장의 종결 어미는 문장 전체의 의미를 이해하는데 매우 중요하다. 다음 예문을 보자[16].

- (1) 철수는 학교에 갔나 싶다.
- (2) 철수는 숙제를 다 했나 보다.
- (3) 철수가 그 일을 하지 않았나 한다.

국어의 경우에는 이러한 형태의 문장 종결이 매우 많이 사용되며 구문적인 면이나 의미적인 면에서 다루기 어렵다. 참고 문헌 [16]에서는 이러한 형태의 문장 종결을 SEA라고 규정하였다. SEA 구문은 종결어미 다음에 다시 동사가 나와 한 문장을 이루는 구문으로 종결어미와 SEA 동사 사이에 휴지를 제외한 어떠한 요소의 개재도 허용하지 않는 구문을 일컫는다. 이러한 SEA 구문은 국어에서 구문적으로나 의미적으로 중요한 의미를 가지며, 번역을 할 때에 특히 조심스럽게 번역하여야 의미를 제대로 전달할 수 있다.

**2.10 국어의 부사절**

국어에서는 절과 절의 결합으로 문장이 확대되는 현상이 어미에 의존한다. 다시 말하면 어미의 종류에 따라 여러 모습의 합성문이 생성된다[17]. 국어의 어미에 대하여서는 2.3절에서 논의하였다. 문헌 [17]에서는 국어의 어미를 종지법, 자격법, 접속법으로 구분하였다. 또한 어말어미를 종결어미와 비종결어미로 구분하였고 비종결어미는 연결어미와 전성어미로 구분하기도 하였다. 연결어미에는 다시 대등적 연결어미, 종속적 연결어미, 보조적 연결어미로 구분하였다. 전성어미는 관형사형 어미와 명사형 어미로 구분하였다.

국어의 문장은 단순문과 복합문으로 구분하였으며 복합문은 다시 접속문과 내포문으로 구분하였다. 접속문은 다시 대등접속문과 종속접속문으로 구분하였다. 또한 내

포문을 명사형 내포문, 관형사형 내포문, 그리고 부사형 내포문으로 구분하였으며 부사형 내포문을 다시 문장 수식 부사절과 성분 수식 부사절도 구분하였다.

**3. 파서 구성의 기본 원리**

파서는 다음 원리에 의하여 구성한다[13].

- (1) 구문(syntax)과 의미(semantics)는 완전히 분리하여 생각한다.
- (2) 상향 분석 원리 : 구문 분석은 특별한 경우를 제외하고는 '아래에서 위(bottom-up)' 방식으로 이루어진다.
- (3) 어휘 정보 원리 : 구문 분석은 사전의 어휘 정보를 기초로 진행된다.
- (4) 범주 결합 원리 : 구문 분석은 범주와 범주를 결합하여 상위의 범주를 형성하는 방식으로 진행된다.
- (5) 분석 완결 조건 : 모든 통사 범주가 결합을 통하여 하나의 최상위 범주로 귀결되면 구문 분석은 완결된 것으로 본다.
- (6) 필요한 요소의 보충은 구문 분석에 필요한 최소한의 요소에 한정된다.
- (7) SEA 구문 : SEA 구문의 적절한 처리가 필요하다.

국어의 파서는 구문과 의미를 완전히 분리하여 생각하며, 문법과 사전으로부터의 하위범주화 패턴과 표현 패턴 정보를 활용하여 파싱을 수행한다.

**3.1 파서 알고리즘 구성 원리**

파싱 알고리즘은 네 과정으로 구성된다. 첫 번째 과정은 입력 문장을 스캐닝(scanning)하여 하위범주화 패턴을 적용하기 위한 준비를 하는 과정이다. 두 번째 과정은 하위범주화 패턴 정보와 표현 패턴 정보를 적용하여 구문을 구성하고 파싱을 완성하는 과정이다. 세 번째 과정에서는 절을 처리하고 네 번째 과정에서는 SEA를 처리하는 것으로 한다.

스캐닝하는 과정에서는 관사, 명사(구), 수사, 통사적 접사, 조사, 의존 명사, 어미활용(형용사형 어미, 부사형 어미 포함), 형용사, 부사("고사하고" 별도처리) 등을 처리한다. 다음 단계에서는 첫 번째 단계에서 처리한 결과, 하위범주화 패턴 정보, 그리고 표현 패턴 정보를 활용하여 파싱을 수행한다. 세 번째 단계에서는 절을 처리하고 네 번째 단계에서는 SEA를 처리한다.

다음은 파싱 과정에서 사용하는 문법이다. M은 의미 단위를 나타내고, NP는 명사구, Vsaux는 동사 SEA, Asaux는 형용사 SEA를 나타낸다.

- (1) M -> M + M
- (2) M -> NP/주격조사 + ADJ/어미
- (3) M -> NP/주격조사 + NP/주격조사 + ADJ/어미
- (4) M -> NP/주격조사 + NP/목적격조사 + Vt/어미
- (5) M -> NP/주격조사 + NP/주격조사 + Vt/어미
- (6) M -> NP/주격조사 + NP/처소격조사 + Vi/어미
- (7) M -> NP/처소격조사 + NP/주격조사 + Vi/어미
- (8) M -> NP/주격조사 + NP/목적격조사 + NP/자격격조사 + Vi/어미
- (9) N -> Det N
- (10) NP -> N
- (11) NP -> N + N
- (12) NP -> N/와(과) N
- (13) NP -> N(의) N
- (14) NP -> N 그리고 N
- (15) NP -> N + Nr/(의존명사) + N

- (16) Vsea -> Vi/어미+
- (17) Vsea -> Vt/어미+
- (18) Asea -> ADJ/어미+

“나는 학생이다”에서 “이다”를 붙여 쓴 형용사로 간주할 수 있다. “너는 학자답다”에서 “답다”도 일반 국문법에서는 접미사로 분류한다. 하지만 위와 마찬가지로 붙여 쓴 형용사로 간주할 수 있다. 그리고 각 범주와 조사는 생략 가능하며 이때는 술어의 하위범주화 정보로 판단할 수 있다. 또한 조사의 순서도 바뀔 수가 있다. 부사는 동사나 형용사를 수식할 수 있다. 그리고 어미의 활용에 의하여 변형된 동사나 형용사의 부사형은 동사나 형용사를 수식한다.

### 3.2 파싱 알고리즘 구성

다음은 파싱 알고리즘을 C 언어 방식으로 나타낸 것이다. 어휘 분석기와 사전은 존재하는 것으로 가정한다.

```
// Korean parser
parsing() {
    input();
    pattern();
    clause();
    sea();
}

input() {
    int length;
    for(i=0; i<length; i++) {
        switch(words) {
            case word1_pos = noun)
                proc_noun(word1, word2);
            case word1_pos = dep_noun)
                proc_dnoun(word1, word2);
            case word1_pos = adjective)
                proc_adj(word1, word2);
            case word1_pos = adverb)
                proc_adv(word1);
            case word1_pos = iverb)
                proc_verb(word1, word2);
            case word1_pos = tverb)
                proc_tverb(word1, word2);
            case word1_pos = conjunction)
                proc_conn(word1);
        }
    }
}

proc_noun(word1, word2)
{
    josa_k = call josa(word2);
    push_stack(word1, josa_k);
    state(word1, noun);
    if (josa_k = 주격)
    if (josa_k = 목적격)
    ...
}

proc_adj(word1, word2)
{
    adj_conj = call adjomi(word2);
    find_adj_dict(word1, subcat);
    match_subcat(subcat);
    switch(adj_conj) {
        case(adj_conj = ~/은/는) {
            if (noun > 0) {
                pop_stack(nouns);
                build_sent(nouns, adj);
                link_sent_noun();
            } else
                link_adj_noun();
        }
        case(adj_conj = 부사형어미) {
```

```
proc_adv(word1);
    ...
}
}
state(word1, adj);
}

proc_iverb(word1, word2) {
    verb_conj = call verbomi(word2);
    find_verb_dict(word1, subcat);
    match_subcat(subcat);
    ...
}
```

파싱에 필요한 사전에는 명사 사전, 동사 사전, 형용사 사전, 부사 사전, 의존 명사 사전(테이블), 조사 사전(테이블), 어미 사전(테이블) 등이 필요하다. 동사 사전과 형용사 사전은 하위범주화 패턴과 표현 패턴을 수록하고 있어야 한다.

### 5. 결론

본 논문은 한국어 문법을 고찰하고 새로운 한국어 파서의 기본 골격을 설계하였다. 앞으로의 계획으로는 파서를 구현하는 것이다.

한국어에 대한 전산학적인 연구는 한국어 연구의 이론적인 측면에서나 기계 번역과 같은 응용을 위한 측면에서나 매우 중요하다고 생각된다. 특히, 2007년에 마무리된 21세기 세종계획의 뒤를 이어 한국어 처리를 위한 전산학적인 연구는 더욱 발전되어야 할 것이다.

### 참고 문헌

- [1] 윤준태, 김선호, 송만석, “전역적 연관표를 이용한 한국어 구문 분석,” 정보과학회논문지(B), 제 24권, 제 11호, 1997년 11월.
- [2] 박소영, 광용재, 정후중, 황영숙, 임해창, “한국어 구문 분석의 효율성을 개선하기 위한 구문 제약 규칙의 학습,” 한국정보과학회논문지:소프트웨어및 응용, 제 29권, 제 10호, 2002년 10월.
- [3] 광용재, 박소영, 황영숙, 정후중, 이상주, 임해창, “표층 구문 타임을 사용한 조건부 연산 모델의 일반화 LR 파서,” 한국정보과학회논문지:소프트웨어및 응용, 제 30권, 제 1호, 2003년 2월.
- [4] 김성용, 이공주, 최기선, “복합 레이블을 적용한 한국어 구문 규칙,” 한국정보과학회논문지: 소프트웨어및 응용, 제 31권, 제 2호, 2004년 2월.
- [5] 이성욱, 서정연, “한국어 문법 관계에 대한 부분 구문 분석,” 한국정보과학회논문지: 소프트웨어및 응용, 제 32권, 제 10호, 2005년 10월.
- [6] 우연문, 송영인, 박소영, 임해창, “지배가능 경로 문맥을 이용한 의존 구문 분석의 수식 거리 모델,” 한국정보과학회논문지:소프트웨어및 응용, 제 34권, 제 2호, 2007년 2월.
- [7] 이형영, 이용석, “내포문의 단문 분할을 이용한 한국어 구문 분석,” 한국정보과학회논문지: 소프트웨어및 응용, 제 35권, 제 1호, 2008년 1월.
- [8] 최용석, 신지애, 최기선, “확률모형과 수식정보를 이용한 와/과 병렬 명사구 범위결정,” 한국정보과학회논문지: 소프트웨어및 응용, 제35권, 제 2호, 2008년 2월.
- [9] Hoojung Chung, Hae-Chang Rim, “A New Probabilistic Dependency Parsing Model for Head-Final, Free Word Order Languages,” IEICE Transaction on Information & System, Vol. E86-D, No. 11, pp.2490-2493, 2003.
- [10] Hoojung Chung, Hae-Chang Rim, “Unlexicalized Dependency Parser for Variable Word Order Languages based on Local Contextual Pattern,” LNCS 2945, pp.112-123, 2004.
- [11] 고려대학교 컴퓨터학과 자연언어처리연구실, <http://nlp2.korea.ac.kr/~hjchung/parsedemo>.
- [12] Yookyung kim, et al., “Senda S<sup>2</sup>MT : Incorporation of Syntax into Statistical Translation System,” <http://www.fluentialinc.com>, 2005.
- [13] 임홍빈, 이홍식, 한국어 구문 분석 방법론, 한국문화사, 2003.
- [14] 이광정, 국어문법연구(1) 품사, 도서출판 역락, 2003.
- [15] 안홍경, 현대국어의 의존명사 연구, 도서출판 역락, 2001.
- [16] 엄정호, 국어의 보문과 보문자, 도서출판 태학사, 2005.
- [17] 이익섭, 국어 부사절의 성립, 도서출판 태학사, 2004.