

구구조 문법 기반 한국어 파싱: 규칙성과 불규칙성의 구별을 중심으로

이효은⁰¹, 계은진², 최재원³, 안민경², 김경희³, 임은숙⁴, 채희락^{2*}

한국외국어대학교 언어인지학과

¹solsae1201@naver.com, ²{jinakay, avocadorights, hrchae}@hufs.ac.kr,

³{koreanteacherjw, conseigh}@gmail.com, ⁴olinda@hanmail.net

⁰제1저자, *교신저자

Korean Parsing Based on a Phrase Structure Grammar: Focusing on the Distinction Between Regularity and Irregularity

Hyoeun Lee, Eunjin Kay, Jae-Won Choi, Min-kyeong An, Kyunghee Kim, Eunsuk Lim, Hee-Rahk Chae
Department of Linguistics & Cognitive Science, Hankuk University of Foreign Studies

요 약

인간의 언어 표현은 규칙성과 불규칙성을 동시에 가지고 있다. 일반적으로 규칙성이 크다고 생각되는 통사 단위에서도 불규칙성이 나타나며, 불규칙적인 표현이라고 여겨지는 관용표현/숙어도 규칙성을 포함하고 있다. 그러므로 표현들은 규칙적인 것과 불규칙적인 것으로 양분할 수 있는 것이 아니라 그 정도성에 차이가 있는 것으로 보아야 한다. 이렇게 규칙성과 불규칙성은 통사 현상의 본질적 속성임에도 불구하고 대부분의 기존 연구에서는 규칙성에 초점을 맞춘 파싱 방안을 연구하였기 때문에 분석에 한계가 있어 왔다. 본 연구에서는 언어의 규칙성과 불규칙성의 구별을 중심으로 한국어 문장을 분석하는 새로운 방안에 대해 제안한다.

주제어: 구구조문법, 관용표현/숙어, 구문분석, 규칙성, 불규칙성

1. 서론

특정 언어의 모국어 화자는 언어 규칙을 사용해서 무수히 많은 언어 표현을 생성한다. 그러나 개별 표현에 따라 일반적인 언어 규칙으로는 설명할 수 없는 요소들이 존재하며, 대표적으로 관용표현이나 숙어가 여기에 속한다.¹ 예를 들어, '어이/어처구니(-가) 없다'는 일반 언어 규칙인 합성성(compositionality)의 원리를 준수하지 않는다. 국어 대사전에서는 '어이'와 '어처구니'의 의미를 아래와 같이 기술하고 있다.

(1) 가. 어이⁰²: 명사, 엄청나게 큰 사람이나 사물.

=어처구니.

나. 어처구니: 명사, 엄청나게 큰 사람이나 사물.

≡어이.

그 어떤 한국어 모국어 화자도 '어이'나 '어처구니'를 (1)의 의미로 인식하지는 않을 것이다. 대신 '어이가 없다'라는 전체 표현만 'be preposterous'라는 의미를 가진다. 즉, '어이/어처구니(-가) 없다'는 개별 요소들의 의미 결합으로 전체 의미를 파악할 수 없다는 불규칙성

을 보인다.

'어이/어처구니(가) 없다'는 통사적인 측면에서도 일반적인 한국어의 언어 규칙과 다르게 행동한다. 보통 '옷을/*가방을 입다'에서와 같이 술어인 '입다'가 '옷'과 같은 보충어를 선택한다. 하지만 '어이/어처구니(-가) 없다'의 경우에는 보충어인 '어이/어처구니'가 '없다'라는 특정한 술어를 요구한다고 보아야 한다. 왜냐하면 '어이/어처구니'는 '없다'의 주어로만 나타나기 때문이다. 이와 같이 관용표현을 분석 할 때에는 규칙적인 속성뿐만 아니라 불규칙적인 속성도 정확히 파악해야 한다.

지금까지 개발된 한국어 구문분석기에서는 대부분 의존 문법(dependency grammar)에 기반하여 관련 요소들을 서로 묶거나[1] 통사적 제약을 부여하는[2] 원리와 규칙을 활용하였다. 이는 규칙적인 문장의 분석에서는 효과적이겠지만, 숙어 등 불규칙성이 큰 요소를 포함하는 문장에 대해서는 분석 자체를 하지 못하거나 분석 결과의 오류율이 급격히 높아진다는 한계가 있다.

이에 본 연구에서는 관용표현이나 숙어와 같이 불규칙적인 표현을 효과적으로 분석할 수 있는 방안을 제안하고자 한다. 먼저 2장에서는 숙어가 가지는 규칙성과 불규칙성에 대해 살펴보고 이 두 가지를 동시에 포착할 수 있는 이론을 소개한다. 3장에서는 숙어를 포함하는 문장을 분석하는 파싱 절차를 제시한 후, 실제 문장 분석에 적용해 볼 것이다.

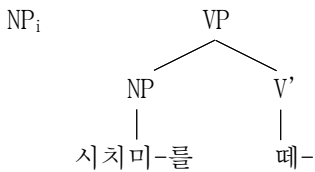
¹ 우리는 관용표현이나 숙어는 관용성/숙어성이 큰 것에서부터 작은 것까지 그 불규칙성의 정도만 있다고 간주하여 관용표현과 숙어를 유사한 의미로 사용한다.

2. 숙어의 분석

2.1. 숙어의 규칙성과 불규칙성

숙어는 대표적인 불규칙적 언어 단위이며 일반적으로 어휘처럼 분석을 한다. 실제로 'all of a sudden'과 '개밥에 도토리'와 같은 일부 숙어는 어휘화되어 내부 구조가 완전히 굳어졌다. 그러나 대부분의 숙어는 고정된 불변의 단위가 아니며 통사적 유연성(flexibility), 즉 규칙적인 측면을 보인다.

(2) 시치미(-를) 떼다



형식적인 측면에서 숙어는 해당 언어가 가지고 있는 구구조 규칙(phrase-structure rules)에 의해 규칙적으로 생성될 수 있다. (2)의 '시치미(-를) 떼다'는 '밥(-을) 먹다'처럼 동사 V/V'와 이 동사가 요구하는 보충어 NP가 결합하여 하나의 동사구가 되었다. 또한 '시치미(-를) 딱 떼다'에서와 같이 숙어 내부 요소가 부사구에 의해 수식을 받을 수 있다. 즉, 숙어는 일반적인 구가 가지는 형식적인 규칙성을 부분적으로는 가질 수 있다. 그러나 '개밥의 도토리'와 '시치미(-를) 떼다'에서 볼 수 있듯이, 숙어의 통사적 유연성은 개별 표현에 따라 그 정도를 달리 한다.

형식적으로 숙어가 가지는 불규칙성은 구가 생성된 이후에 나타나며, 구의 내부와 외부 모두에서 발견된다. 구의 내부에서 나타나는 숙어의 불규칙성에는 확장 불가능성이 있다. 통사적 구는 보충어나 부가어 등을 통해 내부적으로 확장될 수 있다. 반면에 숙어는 일반적인 구와 달리 내부적인 확장에 제약이 있다. 예를 들어, '말이나 행동 따위를 똑똑하게 하다'라는 의미를 가진 숙어 '똑소리(-가) 나다'의 '똑소리'는 '큰 똑소리', '작은 똑소리' 등으로 수식을 받아 확장 될 수 없다.

숙어 구의 외부에서 나타나는 불규칙성에는 대표적으로 이동 제약이 있다. 일반적인 통사 구의 경우, 구구조 규칙이나 일반적 원리/원칙을 제외하고서는 구 내부의 구조나 구성 요소 등의 이동에 제약을 받지 않는다. 그러나 숙어는 그 구성 요소들이 구의 외부로 이동할 때 여러 제약을 받을 수 있다. '꼬리에 꼬리(-를) 물다'는 숙어 내부의 명사구 '꼬리'가 숙어 전체를 이루는 동사구 밖으로 빠져 나갈 수 없다는 특이점을 보인다(예: '*[꼬리에 문] 꼬리'), 동사구 자체가 일종의 "섬 제약(island constraints)"이 작용하는 단위가 되어 내부 요소의 이동을 제약한다.

숙어의 형식적 불규칙성에는 어휘/형태적 불규칙성도 포함된다. 예를 들어, 숙어 '불 보듯 뻔하다'에서 '뻔하다' 대신 '흰하다'라는 단어를 사용하더라도 '의심할 여

지 없이 아주 명백하다'라는 동일한 의미를 가지게 된다. 그러나 '흰하다'가 아닌 다른 비슷한 의미인 '환하다', '밝다' 등의 단어로는 대체 될 수 없다는 점에서 어휘 선택이 제한적이다.

숙어는 형식적 측면뿐만 아니라 의미적인 측면에서도 일반적인 구와 다르게 행동한다. 기존 연구에서는 합성성의 원리에 따라 숙어 내부 요소들을 결합하여 전체 의미를 규칙적으로 도출하려는 시도를 하였다[3-6]. 그러나 숙어의 의미를 분석할 때 규칙성보다 불규칙성이 더 크게 작용한다. 왜냐하면 숙어 내부 요소의 숙어적 의미를 파악하기 위해서는 먼저 전체의 의미를 알고 있어야 하기 때문이다[7]. 숙어 '손(-이) 크다'의 경우, 숙어 내부에 있는 요소 '손'과 '크다'의 개별 의미와 그 의미들이 결합하는 방식을 통해서 '썩썩이가 후하고 크다'라는 전체 의미를 도출해 낼 수 없다. 따라서 본 연구에서는 숙어란 형식적으로는 규칙성과 불규칙성을 모두 가지지만 의미적으로는 불규칙성만을 가지고 있는 구문(construction)으로 가정한다.

2.2. 숙어의 표상 체계

우리는 숙어의 형식적인 규칙성과 불규칙성을 모두 포착할 수 있는 분석 틀로 [8]의 "표상 체계(representation system)"를 사용하기로 한다. 이 분석 체계는 숙어의 불규칙성을 나타내는 네 가지 유형의 기호를 사용한다. [9]는 [8]의 표상 체계에서 제시한 네 가지 유형의 표기법을 한국어의 특성에 맞게 (3)과 같이 추가 및 수정하였다. 이와 더불어 (4)의 자료를 보자.

- (3) 숙어의 형식적 불규칙성을 나타내는 표기법
 - 가. <<XP>>: XP는 통사적 "섬"이며 XP 내부의 요소는 어떠한 이동도 불가능하다.
 - 나. <XP>: XP는 통사적 "섬"이나 화제화와 뒤섞기(scrambling)는 가능하다.
 - 다. /XP/: XP의 내부 요소는 (보충어나 부가어 등에 의해) 추가적으로 확장될 수 없다.
 - 라. !XP!: XP의 내부 요소는 한정사에 의해서만 수식을 받을 수 있다.
 - 마. {...}: 숙어 내부에서 해당 위치에 나타날 수 있는 어휘 목록이다.
 - 바. 대문자(CAPITALIZATION): 문법적 이유로 형태가 변하는 어휘는 대문자²로 표기한다.
- (4) 가. 미나는 분노에 치를 떨었다.
 - 나. 지수가 먼저 꼬리를 쳤다.
 - 다. 부정적인 생각이 꼬리에 꼬리를 물었다.

(4가)에서 숙어 '치(-를) 떨-'은 숙어 내부 요소가 어떤 방법으로도 이동을 할 수 없기 때문에 <<VP>>로 분석된다. 반면에 (4나)의 '꼬리(-를) 치-'는 '꼬리는 지수가

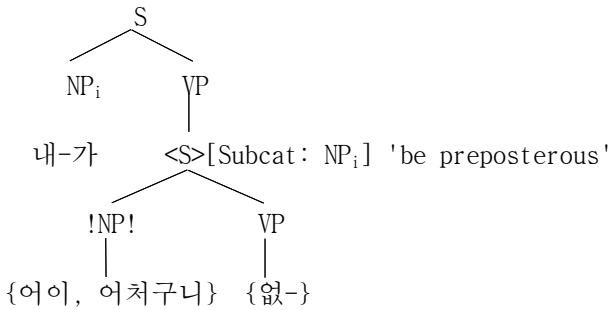
² 한국어에서는 주로 서술어가 다양한 굴절 형태를 가지고 있기 때문에 대문자 대신 '떡-'과 같이 어간 다음 하이픈을 넣어서 표기한다.

먼저 '쳤다'와 같이 화제화될 수 있어 <VP>로 분석되어야 한다. 그리고 (4다)에서 선행하는 '꼬리'는 보충어나 부가어에 의해 확장될 수 없어 /NP/로 표시되지만 후행하는 '꼬리'는 '꼬리'를, '꼬리는'과 같이 한정사의 수식을 받을 수 있으므로 !NP!로 표시된다.

(3가)-(3라)는 속어의 통사적 제약을 나타내며 (3마)와 (3바)는 어휘/형태적 제약을 나타내는 표기법이다. 이 기호들은 구구조 규칙에 의해 규칙적으로 생성된 통사 구조에 적용된다. 즉, 속어의 규칙적인 특성은 통사 구조의 생성 과정에서 포착되며 불규칙적인 특성은 (3)의 표기법에 의해 드러난다.

위의 체계에 따라 속어 '어이/어처구니(-가) 없-'은 다음과 같이 분석할 수 있다.

(5) 어이/어처구니(-가) 없-



분석 (5)에서 속어 부분인 <S>는 화제화와 뒤섞기를 제외하고는 내부 요소가 이동을 할 수 없다(cf. *['미나가 없었던] 어이', '어이가 미나는 없었다'). 속어 내부의 !NP!는 한정사 이외에 추가적인 수식을 허용하지 않는 반면에(cf. *'미나는 큰 어이가 없었다', '미나는 어이는 없었다') VP는 내부적으로 수식을 받을 수 있다('어이가 정말 없다'). 속어 전체의 의미는 'to be preposterous'라는 의미를 가지며 주어를 요구한다. (5)에서 '어이/어처구니'는 속어의 일부이지만 '내-가'는 속어 전체가 요구하는 주어이기 때문에 이 두 NP의 지위는 명확하게 구별된다.

2.3. 삼분 문법 체계

일반적으로 문법이란 어휘 집합체인 사전(lexicon)과 규칙 집합체 이 두 부분으로 이루어져 있다고 본다. 그러나 우리는 문법의 구성 요소가 사전, 규칙, 그리고 구문의 집합체인 "구문집(constructicon)"임을 가정한다 [8, 10]. 모든 관용표현이나 속어는 개별 구문으로 구현되며 구문들은 (5)과 같은 형태로 구문집에 들어 있는 것으로 본다.

삼분 문법 체계에 따라 구문집을 설정하여 한국어 문장을 분석하게 되면 기존에 구축된 사전과 문법 규칙이 현저하게 간단해진다. '어이/어처구니(가) 없-'을 구문집 내에서 처리한다면, '어이'나 '어처구니'를 중복하여 사전 항목으로 넣을 필요가 없기 때문에 사전의 어휘 항목 및 어의 수를 크게 줄일 수 있다. 또한 속어를 억지

로 규칙에 끼워 맞추기 위해 만들었던 불필요한 문법 규칙을 간소화시킬 수도 있다. 그러므로 어휘적 중의성과 규칙 수의 과다로 인해 발생할 수 있는 자연어 처리의 부담도 크게 줄어든 것이다.

3. 구구조 문법 기반 한국어 파싱

3.1. 구문집 구축

우리는 구문집을 구축하기 위하여 [11-16]에서 수집한 약 1200여 개의 구문을 분석하였다. 속어와 관용표현뿐만 아니라 일부 연어(collocations)까지 구문으로 포함하여 분석하였다. 먼저 구구조 규칙에 따라 구문의 수형도를 생성한 후, 표상 체계 기호들을 적용하여 각 구문이 가진 형식적 불규칙성을 나타냈다. 또한 [17]에서 제시한 기준에 따라 보충어와 부가어를 구별하여 개별 구문이 요구하는 보충어를 하위범주화(subcategorization) 틀에 넣었다. 이렇게 작성된 구문 분석 정보는 이후 파싱을 위해 아래와 같이 구문집에 저장되었다.

No.	bracketing	Subcat	meaning
1	[VP/ [<S> [!NP! {어이, 어처구니}]] [VP {없-}]]]	NPO	to be (AdvP) preposterous

그림 1. 구문집 목록 예시

3.2. 파싱 절차

본 연구에서는 사전, 규칙, 구문집의 세 가지 정보를 활용하여 한국어 문장을 분석한다. 이때 사전과 규칙은 문장의 규칙적인 부분을, 구문집은 불규칙적인 부분을 처리하는 데 사용된다. 실제 문장을 분석할 때 먼저 문장 전체 또는 일부에 구문집 항목과 일치하는 부분이 있는지 확인한다. 일치하는 부분은 구문집의 분석 내용을 그대로 가져 온다. 이후 문장의 나머지 부분은 기존 방식대로 사전과 규칙 정보를 활용하여 분석할 수 있다. 구체적인 한국어 파싱 절차는 아래와 같다.

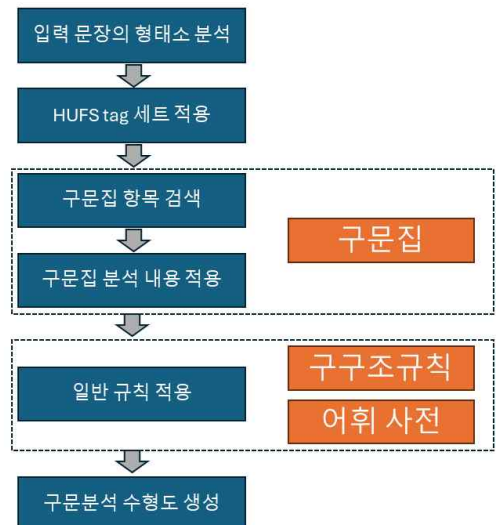


그림 2. 한국어 파싱 절차

첫 번째 단계로 원시 자료의 형태소를 분석하고 결과 값을 보정한다. 우선 울산대학교 한국어처리연구실의 "UTagger: 울산대 형태소/동형이의어 태깅 시스템"를 사용하여 원시 자료에 자동 형태 주석을 한다. UTagger에서 사용하는 형태 주석 태그는 세종 주석 태그 세트를 따르고 있는데, 본 연구에서는 세종 주석 태그에 "HUFS tag" 형식을 적용하여 형태 정보를 수정한다. HUFS tag는 한국어 구문에 대해 더 합리적인 통사 분석을 제시하고자 본 연구진이 고안한 체계로, HUFS tag의 가장 큰 특징은 접어(clitic)의 도입이다. 한국어 표현 중에는 관련된 두 요소 중 최소한 하나가 접어인 경우가 많은데 [18], 기존 연구에서는 이들이 대부분 (복합) 단어로 분석이 되고 있다[19-23]. 본 연구에서는 이렇게 단어로 오분석된 형태 주석 결과 값에 HUFS tag를 적용하여 단어와 접어 단위를 명확하게 구별하였기 때문에 해당 단위가 구로 분석되었다.

두 번째 단계로 분석 대상 문장에 구문집 항목이 존재하는지 검토한다. 이때 검색 범위를 최대한 넓히기 위하여 원시 자료와 형태 주석이 적용된 자료 두 가지를 모두 활용하여 구문집 항목을 스캔한다. 스캔 후 대응되는 항목은 구문집 내의 분석 내용을 그대로 가져와서 분석 대상 문장에 적용한다.

구문 분석 과정에서 어휘나 일반 규칙보다 구문을 먼저 적용하는 이유는 다름 아닌 분석의 효율성에 입각한 것이다[9]. 구문은 예외성을 포함하고 있는 부분으로, 이미 분석이 완료되어 그 자체로 구문집에 저장되어 있다. 그렇기 때문에 통사 구조 분석에서 어려움을 유발할 수 있는 구문을 먼저 처리하고 나면 나머지는 일반적인 절차로 분석할 수 있다.

마지막으로 분석 대상 문장에서 구문집 항목에 해당하지 않는 부분에 대해 일반 규칙을 적용한다. 한국어의 일반적인 통사 현상을 설명하는 구구조 규칙 목록을 작성한 뒤, 전체 규칙을 일정한 순서에 따라 처리하여 해당 문장의 통사 분석을 실시한다. 규칙의 적용 범위는 속어 표현의 외부뿐만 아니라 내부(에서 규칙적인 부분)까지 포함하기 때문에 구문 내에 나타나는 불규칙성과 규칙성을 모두 체계적으로 분석할 수 있는 장점이 있다.

3.3. 파싱 결과

본 절에서는 위에서 기술한 방식에 따라 속어 '어이(-가) 없-'을 포함하는 실제 문장을 분석해 보고자 한다. 먼저 분석 대상 문장을 UTagger로 형태 주석을 한 후 HUFS tag를 적용하여 아래와 같이 분석하였다.

- (6) 가. 철수의 행동은 정말 어이가 전혀 없다.
- 나. 철수/NNP+의/JKG 행동/NGG+은/JX
정말_01/MAG 어이_02/NGG+가/JKS
전혀_01/MAG 없_01/VA+다/EF
- 다. '철수/N:pr', '#의/#Adn', '행동/N:co', '#은/#Del', '정말_01/Adv', '어이_02/N:co', '#가/#Del:nom', '전혀_01/Adv', '없_01/V:a', '#다/#Se'

(6나)는 UTagger의 형태 주석 정보이며 (6다)는 HUFS tag를 적용한 결과 값이다. 본 연구에서는 [18]에 따라 조사 '-의', '-은', '-가'와 어미 '-다'를 접어로 분류하였으며 "#" 기호를 추가하여 해당 요소가 접어임을 나타냈다.

다음으로는 (6가)와 (6다)의 정보를 동시에 사용하여 구문집 항목을 검색하였다.

```

일치하는 값이 있습니다.
번호: 1번
[/VP/[<S> [INP![[어이, 어처구니]]][VP [[없]]]]]
    
```

그림 3. 구문집 항목 매칭 결과

<그림 3>에서처럼 분석 대상 문장 내에 구문집 항목이 들어 있을 경우, 구문집의 통사 정보를 불러와서 문장 분석에 바로 적용한다. 즉, '철수의 행동은 정말 어이가 전혀 없다' 중 '어이(-가) 없-'은 일반 규칙이 아닌 구문집 분석 내용으로 처리한다.

마지막으로 입력 문장의 나머지 부분은 일반 규칙을 적용하여 문장을 분석하였다. '철수의 행동은 정말 어이가 전혀 없다'의 최종 구문 분석 수행도는 다음과 같다.

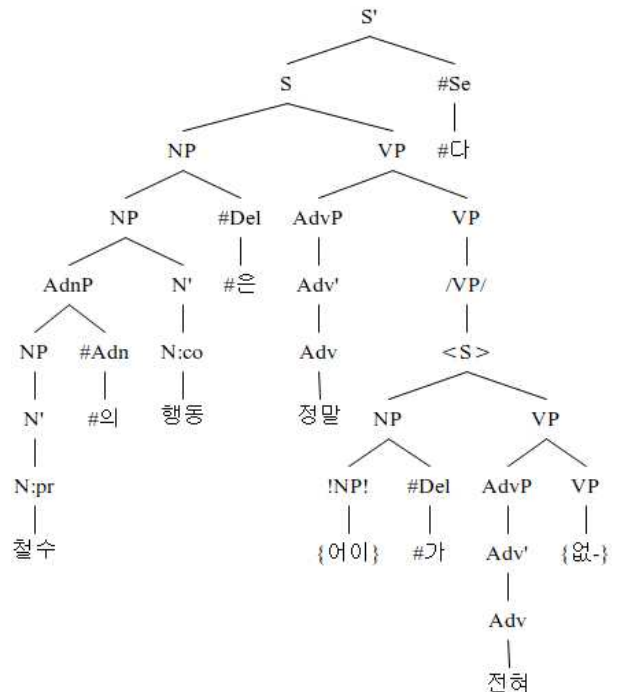


그림 4. 입력 문장의 구문분석 수행도

정리하면, 먼저 속어 표현인 '어이(가) 없-'은 구문집 항목에 존재하기 때문에 구문집 내의 통사 분석 내용과 하위범주화 정보를 그대로 가지고 온다. 구문집 항목의 분석이 끝나면 그 다음에는 문장의 나머지 부분인 '철수의', '행동은', '정말', '전혀', '-다'를 일반적인 통사 규칙에 따라 분석한다.

분석 대상 문장인 '철수의 행동은 정말 어이가 전혀

없다'는 규칙적인 속성과 불규칙적인 속성을 동시에 가지고 있다. 명사구 '어이'는 한정사의 수식에 의해서만 확장될 수 있다는 제약을 받는다. 구구조 규칙으로는 설명할 수 없는 이러한 불규칙성은 구문집의 분석 내용을 활용하여 해결할 수 있다. 이와 동시에 명사구 '어이'가 한정사(로 가정하는) '-가'와 결합하여 다시 명사구가 되는 것은 한정사가 수식을 할 뿐 핵으로서의 역할을 하지 못한다는 일반적인 통사 규칙에 기인한다. 이 경우 "XP Del → XP (XP: NP, PP, VP, AdvP, ...)"와 같은 구구조 규칙을 적용하여 분석한다. 이렇게 삼분 문법 체계를 기반하여 한국어 문장을 분석한다면 구문의 불규칙적인 속성과 규칙적인 속성을 효과적으로 기술할 수 있다.

한국어는 어순이 자유롭고 문장 구성 성분이 쉽게 생략되기 때문에 대부분의 구문 분석 연구에서는 의존구조 분석을 선택한다[2]. 그러나 의존구조 분석 시스템에서는 일반적으로 두 어절 간의 의존 관계만을 결과로 제시하기 때문에[24] 전체적인 문장의 통사적인 구조 정보를 파악하기 어렵다. 본 연구는 이러한 한계점을 보완하기 위해 구구조 문법을 기반으로 하여 한국어 구문을 분석하였다. 또한 개별 구문이 취하는 보충어와 부가어 정보를 모두 분석하여 통사론적으로 더 깊이 있는 구문 분석을 시도하였다.

4. 결론

본 연구에서는 관용표현이 가지고 있는 규칙성과 불규칙성에 대해 살펴보고 이를 체계적으로 구별하여 분석할 수 있는 방안에 대해 알아보았다. 숙어를 포함하는 표현을 체계적이고 효율적으로 분석하기 위하여 사전, 규칙, 그리고 구문집으로 이루어진 삼분 문법 체계를 기반으로 문장을 분석하였다. 문장 전체 또는 부분에 구문집 항목에 해당하는 구문이 존재할 경우, 구문집의 분석 내용을 그대로 적용하여 불규칙성을 기술하였다. 그리고 문장의 나머지 부분에서는 통사 규칙과 사전 정보를 사용하여 형태·통사적 규칙성을 설명하였다. 본 연구는 기존 연구에서 다루어지지 않은 숙어 표현의 구문 분석에 대한 새로운 관점을 제시하였다는 데 의의가 있다.

현재까지는 구문집에 등재된 구문 항목의 수가 충분하지 않기 때문에 지속적으로 더 많은 구문을 분석하여 구문집에 저장하는 작업이 필요하다. 또한 추후에는 대규모 말뭉치 자료에서 추출한 통계 정보를 활용하여 구문 분석 결과의 정확성을 높일 예정이다.

감사의 글

본 연구에서 형태소 분석기로 UTagger를 사용할 수 있도록 허락을 해 주신 울산대학교 옥철영 교수님과 연구원들께 감사의 말을 전한다. 이 연구는 한국연구재단 우수학자 지원 사업의 일환으로 수행되었다 (과제명: "영어와 한국어 통사 현상의 규칙성과 불규칙성", 과제번호: NRF 2022S1A5B1082025).

참고문헌

- [1] 김광백, 박의규, 나동렬, 윤준태, "구간 분할 기반 한국어 구문분석", 한국정보과학회 언어공학연구회 학술발표 논문집, pp. 163-168, 2002.
- [2] 남용, 윤애선, 권혁철, "무제한 한국어 의존 구문분석기 'PNU-KLParse 2.0'의 개발", 정보과학회논문지: 컴퓨터의 실제 및 레터, 제20권, 제6호, pp. 354-358, 2014.
- [3] M. T. Espinal and J. Mateu, "On Classes of Idioms and Their Interpretation," *Journal of Pragmatics*, Vol.42, No.5, pp.1397-1411, 2010.
- [4] T. Wasow, I. Sag and G. Nunberg, "Idioms: An Interim Report," *Proceedings of the XIIIth International Congress of Linguistics*, pp. 102-115, 1983.
- [5] D. Geeraerts, 1995. "Specialization and Reinterpretation in Idioms," *Idioms: Structural and Psychological Perspectives*, M. Everaert, E.-J. Linden, A. Schenk, R. Schreuder, Eds., Lawrence Erlbaum, 1995.
- [6] R. W. Gibbs, "Idiomaticity and Human Cognition," *Idioms: Structural and Psychological Perspectives*, M. Everaert, E.-J. Linden, A. Schenk, R. Schreuder, Eds., Lawrence Erlbaum, 1995.
- [7] H.-R. Chae, "Idioms: Formally Flexible but Semantically Non-transparent," *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, pp.46-54, 2015.
- [8] H.-R. Chae, "A Representational System of Idiomatic Constructions: For the Building of Computational Resources," *Linguistic Research*, Vol.31, No.3, pp.491-518, 2014.
- [9] 채희락, "한국어 [... V... V] 표현의 종합적 통사 분석 III: 관용표현 구문과 합성어", 언어, 제41권, 제1호, pp. 119-163, 2016.
- [10] M. Hilpert, *Construction Grammar and Its Application to English*, Edinburgh University Press, 2014.
- [11] 노용균, 한국어 기본 숙어 사전, 한국문화사, 2002.
- [12] 국립국어원, 외국인을 위한 한국어문법1, 국립국어원, 2005.
- [13] 국립국어원, 외국인을 위한 한국어문법2, 국립국어원, 2005.
- [14] 이종희, 이희자, 어미 조사 사전 : 전문가용, 한국문화사, 2006.
- [15] J. Yeon and L. Brown, *Korean: A comprehensive grammar*, Routledge, 2013.
- [16] 임혜진, 한국어 교육용 숙어 표현의 통사적 속성 연구, 한국외국어대학교 석사학위 논문, 2015.
- [17] H.-R. Chae, "Complements vs. Adjuncts (in

- Korean)," *Studies in Modern Grammar*, Vol.19, pp.69-85, 2000.
- [18] H.-R. Chae, *Korean Morphosyntax: Focusing on Clitics and Their Roles in Syntax*, Routledge, 2020.
- [19] 서정수, "합성어에 관한 문제", 한글, 제173-174호, pp. 367-400, 1981.
- [20] 강현화, 국어의 동사 연결 구성에 대한 연구, 한국문화사, 1998.
- [21] 황화상, "국어 합성 동사의 의미", 한국어학, 제15권, pp. 307-324, 2002.
- [22] 이정훈, "V-어 V 합성 동사 형성 규칙과 범주 통용", 어문학, 제91권, pp. 129-161, 2006.
- [23] 박소영, "한국어 동사 연쇄 구성과 형태-통사론의 접면", 언어, 제39권, 제4호, pp. 773-795, 2014.
- [24] 최맹식, 정석원, 김학수, "CRFs를 이용한 의존구조 분석 및 의존 관계명 부착", 정보과학회논문지: 소프트웨어 및 응용, 제 41권, 제4호, pp.302-308, 2014.