

# 통사론적 단어를 기준으로 하는 형태 분석 지침에 대하여

— 세종 형태 분석 지침의 수정 방안을 중심으로 —

이진병(전북대 강사)

## 〈 목 차 〉

- |                        |                          |
|------------------------|--------------------------|
| 1. 서론                  | 3. 통사론적 단어를 기준으로 한 이원화   |
| 2. 기존 지침의 개념 체계와 문제점   | 3.1. 개선 방향의 검토           |
| 2.1. 지침의 전제 조건         | 3.2. 통사론적 단어 중심의 분석 표지 안 |
| 2.2. 문법 층위 및 문법 단위의 혼재 | 4. 결론                    |

## 국문초록

본고의 문제 제기는, 21세기 세종계획 지침을 위시한 기존의 형태 분석 지침이 개념적인 체계성과 활용상의 효용성 측면에서 부족하다는 것이다. 문제점으로 조어론의 층위와 통사론의 층위가 혼재되어 있다는 점, 분류 단계의 독립성이 떨어진다는 점 등을 들었다.

형태 분석 작업은 중요하다. 원시 말뭉치를 형태·의미 분석 말뭉치, 구문 분석 말뭉치, 기타 담화 분석 말뭉치 등으로 연계하는 밑 작업이자 기본 과정이기 때문이다. 세종 형태 분석 지침은 다음과 같이 개선할 수 있다. 첫째, 통사론적 단어의 개념을 분석 기준으로 삼는다. 아울러 어휘 범주와 기능 범주의 구분을 고려한다. 둘째, 통사론적 단어를 중심으로 상위의 분석 단계와 하위의 분석 단계를 두어 통사론적 층위(통사 단위 구성의 층위)와 형태론적 층위(단어 형성의 층위)를 구분한다. 셋째, 통사론적 층위를 기본 분석 단계

로 삼고 형태론적 층위는 부가적인 분석 단계로 삼아서, 즉 분석의 이원화를 취해서 말뭉치 구축 및 이용에 있어서 유연성과 실질성을 제고한다. 형태론적 층위를 부가적인 단계로 설정한 주요 이유는 조어의 성분 분석이 어려울 뿐더러 효용성도 적다는 점이다. 한편, 말단의 태그는 기존의 것을 되도록 그대로 사용하여서 기구축 말뭉치와의 연계도 취한다. 본고는 이러한 논의의 결과를 통사론적 단어 중심의 분석 표지 안으로 제시하였다.

말뭉치 구축의 이론적 원리는 말뭉치를 구축할 때도, 이용할 때도 매우 중요하다. 본고의 논의는, 말뭉치 구축 작업에 대한 하나의 제안이면서, 형태 분석이라는 문법론적 연구 절차에 대한 고찰이기도 하다.

주제어 : 형태 분석 지침, 분석 표지, 통사론적 단어, 문법 층위, 통사  
층위, 조어 층위, 기능 범주, 어휘 범주

## 1. 서론

본고는, 온전한 우리말의 대규모 말뭉치를 위하여 형태 분석 작업을 할 때, 문법적으로 어떤 기준을 가지고 기존의 문제점을 개선해 나가야 할지를 논의하고자 한다. 특히 문법 단위와 문법 층위에 대한 논의를 형태 분석 지침에 적용하고자 한다.<sup>1)</sup> 형태 분석 작업은 문법론적으로도 말뭉치언어학적으로도 중요하다. 왜냐하면 형태·의미 분석, 구문 분석, 담화 분석 등으로 이어지는 밑 작업으로서 기본 과정에 해당하기 때문이다. 이 연구의 목적은, 궁극적으로는 말뭉치 구축의 문법적 체계성과 아울러 활용상의 효율성을 높

1) 진정한 ‘형태 분석’의 ‘분석’은 동음이의어를 구분하는 어깨번호까지 붙이는 작업을 일컫기도 하나, 본고는 그러한 작업은 ‘형태·의미 분석’으로 보고 형태를 나누고 태그를 붙이는 작업도 ‘형태 분석’이라 부르기로 한다. 대체로 21세기 세종 계획 당시의 용어를 따르고자 한다.

이는 데 기여하는 것이며, 당면하여서는 그러한 기여를 위해 근거를 정리하는 것이다.

21세기 세종계획 말뭉치를 기반으로 현재 다양한 말뭉치가 광범위하게 구축되고 있다. 이제 말뭉치의 대규모화는 부정할 수 없는 추세이다.<sup>2)</sup> 그런데 21세기 세종계획의 문어 말뭉치를 전형으로 하는 주요 말뭉치들은 언어의 다양성을 충분히 고려했다고 보기 어렵다. 구어 말뭉치나 특수 말뭉치 등을 따로 두어 언어의 다양성을 일종의 격리를 통해 처리한 측면이 있다. 그런데 현실적으로 21세기 세종계획의 지침이 중심이 되어 가고 있다. 대개의 형태 분석기가 이 지침에 의해 구축된 말뭉치를 기본으로 삼고 있기 때문이다. 이러한 상황에서 다시 한 번 문제시해야 하는 것은, 언어의 다양성을 충분히 고려하지 않은 형태 분석 방법이 기본 지침이 되어 가는 것이 과연 바람직한가 하는 점이다.

한편, 21세기 세종계획 말뭉치는 공식적으로 완전히 공개된 것이 아니다. 저작권 문제 때문이다. 이러저러한 상황들을 고려할 때 이제는 저작권의 제약을 피하면서 언어의 다양성을 반영할 수 있도록 구어나 방언 등을 수용하는 말뭉치의 구축이 필요하다. 구어나 방언은 직접 채록이 가능하며 우리말의 다양성을 잘 보여 준다. 전형적인 문어이자 남한어이자 표준어인 것에 국한하지 않고 우리의 실체를 온전하게 모아서 말 그대로의 ‘우리말 말뭉치’를 구축하기 위해서는 대규모화와 함께 하위 언어의 다양성도 고려해야 한다.

말뭉치의 대규모화와 다양화를 위해서는 무엇보다 형태 분석이 분명한 문법적 원리를 가지고 있어야 한다. 형태 분석은 말뭉치 구축 및 연계의 밑 작

---

2) 흔히 ‘빅 데이터(big data)’라는 말을 쓴다. 하지만 언어학자의 입장에서 언어 자료를 ‘빅 데이터’라고 부르는 것은 조심스럽다. 이기황(2016) 등에 따르면, 빅 데이터는 고전적으로 ‘3V’를 기본 속성을 가져야 하는데, 그 중의 하나가 ‘Variety’이다. 이를 번역하면 ‘비정형적인 다양성’이라고 할 수 있다. 언어학자의 역할이 언어의 원리와 규칙을 설명해 내는 것이라 할 때, 언어를 비정형적인 대상으로 칭하는 것은 주의해야 할 일이다. 본고에서는 ‘빅 데이터’를 사용하지 않고 ‘대규모’라는 일반어를 사용하기로 한다. 그 아무리 다양하더라도 언어는 유한한 규칙으로 다양하게 운용되는 관찰 가능한 실체로 간주하여야 한다. 이것이 본고에서 말뭉치의 처리를 대하는 기본 관점, 혹은 기대이다.

업이기 때문이다. 본고는 이에 대한 고민이다.

2장에서는 기존 지침의 문제점을 다룰 것이다. 이제 가장 널리 기준이 되고 있는 21세기 세종계획의 ‘형태분석 지침’을 중심으로 대표적으로 세 가지 문제점을 지적할 것이다. 3장에서는 2장에서 문제 제기한 바에 대하여 대안을 제시하고 검토할 것이다. 논의의 핵심은, 한국어의 형태 분석을 위하여 단어를 어떻게 설정하고 기준으로 활용해야 할지에 대한 것이다.

## 2. 기존 분석 표지의 개념 체계와 문제점

### 2.1. 지침의 전제 조건

지침은 말뭉치의 구축의 근거이다. 지침에 비추어 파악할 수 없거나, 파악하기 어려운 말뭉치는 양질의 말뭉치라 할 수 없다. 본고는 대용량의 우리말 말뭉치를 효율적으로 구축하기 위하여 개선해야 할 점을 논의하고자 한다. 그리고 개선의 방향성은 지침의 체계와 맞물려 있다.

#### (1) 말뭉치 구축 지침이 갖추어야 할 조건

- ㉠. 개념의 정연성: 논리가 정연해야 하며, 이러한 개념 체계는 용어 등으로 최대한 반영되어 있어야 한다.
- ㉡. 이해의 용이성: 되도록 단순해야 한다. 그리하여 이해하기에 용이해야 한다.
- ㉢. 적용의 유연성: 다양한 발화 형식이나 사용 목적에 되도록 유연하게 적용할 수 있어야 한다.
- ㉣. 활용의 실질성: 말뭉치 구축 및 활용의 실질적인 환경도 고려해야 한다.

지침이 체계적이고 효율적이라면 (1)과 같은 조건을 가져야 한다. (1㉠)의 조건을 갖추지 못한다면 말뭉치 구축의 준거로서 부족하게 되어 말뭉치의 질적 하락을 낳게 된다. (1㉡)과 같이 지침이 단순하지 못하고, 지침의 양이

많아지거나 복잡해지거나 세부 항목들이 일반화되지 못하고 목록만으로 명시되거나 한다면 이는 효과적인 지침이라 할 수 없다. 단적으로 이러한 지침에 의해 구축된 말뭉치는 다른 이용자의 접근성을 막는다. 예를 들어 분석해야 할 파생 접사를 원리적으로 정리하지 못하고 임의적으로 목록화하여 이를 기준으로 말뭉치를 구축하게 되면, 말뭉치의 구축자도 이 지침을 따르면서 혼동을 일으키기 쉬울 뿐만 아니라, 구축된 말뭉치를 이용하는 입장에서도 사용을 어려워하게 된다. 온전한 우리말 말뭉치를 다양하게 통합하여 처리하기 위해서는 (1c)의 조건도 필요하다. 하나의 말뭉치는 필요에 따라 다음 단계의 말뭉치로 연계되어 추가로 분석되어야 한다는 점에서도 (1c)의 조건은 유의미하다. (1c)은 (1a)과 상충되기 쉬운 조건이다. (1a)을 추구한다고 하여 말뭉치가 너무 추상적이거나 복잡해지면 여러 구축자 사이에 불일치가 발생하기 쉬워 말뭉치 전체의 질이 떨어지는 역효과가 발생하기도 한다. (1c)에 문제가 발생할 경우, 기본적으로 해당 말뭉치를 구축하는 데도 어려움이 발생하지만, 이후로 말뭉치를 추가·유지·보수하기도 어렵게 된다. (1a)과 (1c)은 적절하게 조율되어야 한다.

본고는 (1)을 전제로 하여 기존 지침의 개선점을 논의하고자 한다. 특히 세종 지침을 기준으로 구축된 기본적 말뭉치와의 연계성을 고려하는 동시에 개선 방안을 취해야 할 것이다. 그리하여 형태 분석의 효율성을 여러 차원에서 높일 수 있는 방안을 제시하고자 한다. 이를 위하여 다음 절에서는 기존 지침의 문제점부터 진단하고자 한다.

## 2.2. 문법 층위 및 문법 단위의 혼재

현재 말뭉치 처리에 가장 많이 통용되고 있는 분석 기준은 ‘21세기 세종 계획’의 ‘국어 기초자료 구축’ 분과에서 보고한 최종 지침이다.<sup>3)</sup> 바로 21세기

3) 또는 그에 따라 이루어진 말뭉치 자체일 수도 있다. 자연어 처리나 딥러닝 기술을 이용한 형태소 분석에서는 지침의 중요성에 대한 인식이 크지 않다. 말뭉치 구축상 말단의 결과물이라 할 수 있는 말뭉치 자체의 사례만을 고려하여 처리하

세종계획(2007: 183-223)에 수록된 「형태분석 말뭉치 구축 지침 (ver. 2005-1)」(이하 ‘세종 지침’)이다. 그 중에서도 특히 아래 (1)의 ‘세종 분석 표지’가 핵심이다. 세종 지침과 관련된 논의는 여러 가지 측면에서 지속적으로 이루어져야 할 것으로 보인다. 본고에서는 그러한 여러 측면을 모두 다루기보다 개념 체계의 문제를 중심으로 다루기로 한다.

(2) 세종 분석 표지(21세기 세종계획 2007: 185)<sup>4)</sup>

대분류	소분류	세분류	대분류	소분류	세분류	
(1) 체언	명사NN	일반명사NNG	(5) 관계언	격조사JK	주격조사JKS	
		고유명사NNP			보격조사JKC	
		의존명사NNB			관형격조사JKG	
	대명사NP	대명사NP			목적격조사JKO	
수사NR	수사NR	부사격조사JKB				
(2) 용언	동사VV	동사VV			호격조사JKV	
	형용사VA	형용사VA			인용격조사JKQ	
	보조용언VX	보조용언VX		보조사JX		
(3) 수식언	관형사MM			(6) 의존형태	접속조사JC	접속조사JC
		부사MA				일반부사MAG
(4) 독립언	감탄사HC	감탄사HC	접두사XP	중결어미EF		
				접속부사MAJ	연결어미EC	
			접미사XS	명사형전성어미ETN		
				관형형전성어미ETM		
			어근XR	체인접두사XPN		
				명사과생접미사XSN		
				동사과생접미사XSV		
				형용사과생접미사XSA		

위의 표는 분류 체계상 여러 문제를 안고 있다. 문법 층위, 문법 단위, 분류의 단계가 체계성의 문제를 안고 있다. 문제점을 대략 세 가지로 정리하고자 한다.

(3) 세종 분석 표지의 문제점

가. 문법적으로 단어 형성의 층위(또는 ‘조어 층위’)와 통사 단위 구성의

는 분석기를 개발하는 입장일수록 더욱 그러하다.

4) 원문의 표에서 논의상 “(7) 기호” 부분을 제외하고 문법적 해석이 꼭 필요한 나머지 부분을 가져왔다.

층위(또는 ‘통사 층위’)가 혼재되어 있다. 조어의 파생 접사와 이른바 굴절의 어미가 함께 묶여 있지만 굴절의 조사는 따로 처리하고 있다.<sup>5)</sup>

- ㉸. 단어의 기준이 분명하지 않아 형태소/단어/어절의 문법 단위가 혼재되어 있다. “체언/용언/수식언/독립언/관계언/의존형태”는 단어를 대상으로 하는 품사 분류도 형태소를 대상으로 하는 분류도 어절을 대상으로 하는 기능 분류도 아니다.<sup>6)</sup>
- ㉹. 각 분류 단계의 독립성이 없다. 대분류 단계와 소분류 단계가, 필요에 따라서, 세분류 단계와 별도로 사용될 수 없다.

(3㉿)의 문제는 이론적으로도 문제가 될 수 있지만, 실제로 말뭉치를 구축할 때 더욱 문제가 된다. 조어론적 성분을 뚜렷한 원리 없이 분석을 하게 되면 말뭉치의 구축자(작업자)와 구축 후의 이용자(특히 연구자) 모두 어려움을 겪기 마련이다. 앞서 제시한 지침이 갖추어야 할 (1)의 조건을 모두 여기기 쉽다.

조사나 어미는 어휘부에 저장되어 있다가 발화 형성 과정에서 통사적 결

- 5) ‘굴절’은 어형 변화를 뜻하는 “어구성(WORD FORMATION)의 두 가지 주된 과정의 하나”(조성식 외 편 1990: 588)이다. 국어 문법의 경우, 조사나 어미를 단어의 구성 성분으로서 볼 때 성립하는 개념으로서 조사 결합이나 어미 결합을 굴절로 보기도 한다. 특히 어미의 결합은 어형 변화의 관점에서 ‘활용’으로 불려 왔다. 활용을 인정하는 관점에서 볼 때, 국어 문법론은 형태론과 통사론으로 구분되고, 형태론은 다시 조어론과 굴절론으로 구분된다. 사실 이러한 용어 체계가 일반적이다. 그러다 보니 형태론과 통사론의 층위 구분 문제가 국어의 특징처럼 거론된다. 이와 달리 본고에서는, 특히 3장에 이르러, 조사와 어미를 통사론적 단어로 인정한다. 조사와 어미를 단어로 인정한다면, 국어에서 문법 단위가 통합되는 기제는 조어와 통사로 양분되기 때문에, 국어 문법론도 조어론과 통사론으로 양분된다 하겠다. 이때의 조어론은 곧 형태론이 된다. 3장에서는 이런 용어 체계를 따를 것이다. 단, 2장에서는 기존 논의와의 연속성을 위하여 ‘조어’와 대조하는 용도로 ‘굴절’을 사용할 수는 있다.
- 6) 본고에서 ‘어절’이라는 문법 단위를 인정하는 것은 아니다. 다만 띄어쓰기 단위라는 표기법상의 형태로 일컫기 위하여 사용할 뿐이다. 띄어쓰기 단위로서의 어절은 문어 말뭉치를 처리할 때 필요하다. 또는 기존 지침이나 문법 설명을 거론하기 위하여 필요할 수 있다.

함에 쓰이는 요소이다. 하지만 파생 접사는 “선언어적 개념 구조로부터 문법 구조를 생성하거나 문법 구조로부터 음운 표시를 만드는 과정 또는 음운 표시를 음성 표시로 전화하는 과정처럼 자동적이고 무의식적으로 빈번하게 이루어지는 인지 과정이 아니라 필요할 때만 간헐적으로 일어나는 매우 의식적인 인지 과정”(송원용 2016: 21)이다. 통사 층위와 조어 층위는 구분되어야 한다. 굴절이라 일컬어지는 통사 단위 구성의 재료는 “빈번하게 이루어지는 인지 과정”(같은 글)에 참여하는 재료로서 분석이 용이한 편이며, 형태 분석 말뭉치 이후로 발전시켜야 하는 형태·의미 분석 말뭉치나 구문 분석 말뭉치에도 필요한 기초 성분이다. 그러나 조어의 재료는 분석의 기준을 잡기에 매우 어려운, 어휘부를 해체해야만 온전하게 원리를 이해할 수 있을 법한 분석 대상이다. 분석의 결과물도 이후로 이어질 말뭉치 구축에 그리 큰 효용이 되지 않는다. 기계 번역이나 딥러닝 등의 전산 처리의 측면에서도 분석의 효용성이 크지 않다.

이론적인 체계성은 더욱 문제이다. 파생 접사를 조사와 어미와 나란히 계열화해 놓았다는 것도 문제인데, 거기에 어근도 나란히 배치되어 있다. 그런데 또 합성어의 어기나 어근 분석에 대한 것은 언급조차 안 되어 있다.<sup>7)</sup> 21세기 세종계획(2007: 223)에 이르러 “합성어”에 대한 언급이 있기는 하나 구와 합성어의 구분 문제를 다루고 있을 뿐이다. 이는 구와 단어의 구분 문제에 지나지 않는 것으로, 조어론적 분석으로 파생법의 성분은 분석하지만 합성법의 성분은 왜 분석하지 않는지에 대한 언급이 없다. 세종 분석 표지는 합성어 분석을 제대로 고려하지 않았다고 할 수 있다. 설령 합성어를 분석한 다손 치더라도 세종 분석 표지는 합성어 형성에 쓰인 어근이나 어기와 통사에 쓰인 단어를 구별해 주지 못한다. 예를 들어 ‘손이 오가다’와 ‘손님이 오다’를 분석하면 ‘손’은 모두 ‘NNG’가 되고, ‘오-’는 모두 ‘VV’가 될 뿐이다.

7) 본고는 세종 분석 표지를 설명하기 위하여, ‘불규칙적 어근’ 및 ‘규칙적 어근’과 그 둘을 아우르는 ‘어기’의 용어 체계를 사용하지 않고, ‘불규칙적 어근’ 대신에 ‘어근’을 ‘규칙적 어근’ 대신에 ‘어기’를 사용하고자 한다.

- (4) ㄱ. 어른+스럽다→아른스럽다, 크+다랗다→커다랗다  
 ㄴ. 막+압+묻+엄→마감/무덤, 감+장/검+정→감장/검정  
 ㄷ. 개+아지→강아지, 무겁+이→무게(이상, 고영근 1995: 19)

세종 지침에서 분석의 대상으로 삼고 있는 파생의 경우도 문제가 크다. 예를 들어 고영근(1995: 19-20)에서는 파생 시에 어기·어근과 접사 교체가 다양하게 일어남을 위 (4)의 예로 보여 주고 있다. (4ㄱ)에서는 어기가 교체되고 있고 (4ㄴ)에서는 파생 접사가 교체되고 있으며 (4ㄷ)에서는 어기와 파생 접사가 모두 교체되고 있다. 특히 (4ㄷ)의 예는 형태 분석의 어려움을 여실히 보여 준다. ‘-아지’[adʑi]는 ‘-아지’[ɲadʑi]로 교체되어 있어서 분석된 형태의 표기법조차 문제가 된다. 이러한 형태의 교체는 조어론적 성분을 구분할 때 발생하는 여러 문제점 중의 하나이다.

- (5) ㄱ. 개똥(≡‘개의 똥’):똥개(≠‘똥의 개’), 날뛰다:흠날다  
 ㄴ. 개꾼:강아지, 길다랗다:커다랗다  
 ㄷ. 개꾼:개다리:개다리소반  
 ㄹ. 막이:마감  
 ㅁ. 시답다 : 정답다 : 마음이 아름다운 사람답다

형태 표기도 문제이지만, 더 큰 문제는 ‘어디까지 분석할 것이냐’이다. 위 (5ㄱ)은 의미 합성성의 차이를 보이며, (5ㄴ)은 형태 변이의 차이를 보이며, (5ㄷ) 분석되는 성분이 형태소·단일어·합성어로 차이를 보이며, (5ㄹ)은 조어의 생산성에서 차이를 보이며, (5ㅁ)은 어기 및 어근의 문법 범주가 차이를 보인다. 이러한 복잡한 사정 속에서 조어론적 성분을 분석한다는 것은 어휘부를 규칙적으로 설명하는 것만큼이나 논란 거리이다.<sup>8)</sup>

8) 그 밖에도 조어론적 성분의 분석은 여러 가지 문제를 낳는다. 극단적인 경우에는 대명사의 분석에 대하여 생각해 볼 수 있다. 단어 단위 분석에서는 대명사의 태그(NP)가 유의미할지 모르지만 합성법의 성분을 분석해 낸다면 대명사 태그는 1·2인칭 대명사를 제외하고는 대부분 해체되어 버릴 것이다. 이렇듯 단어의 분석과 형태소의 분석은 양립하기 힘든 경우도 존재한다. 또한 널리 알려진 형태소

조어론적 성분을 형태 분석하려면 두 가지 방법이 가능하다. 첫째, 파생 접사 사전이나 복합어 사전을 만들어 기준으로 삼는 방법이다.<sup>9)</sup> 둘째, 개별 연구자의 견해에 따라 복잡한 지침을 만들어 처리하는 방법이다. 이 두 방법은 지침을 매우 복잡하고 숙지하기 어렵게 만든다. 또한 이에 따라 이루어진 말뭉치는 다른 국어 연구자들이(심지어 구축자마저) 다루기에도 매우 비효율적이다. 연구자 간의 견해 차이나 용도 차이가 발생하기 쉬워서 대개는 재가공을 해야만 쓸 수 있게 된다. 심지어 분석을 안 해 놓는 것이 나은 경우도 많다. 예를 들어 어휘 통계를 내야 한다면 조어론적 성분의 분석을 모두 없애고 처리를 해야 한다. 그런데 조어론적 형태 분석이 이루어져 있는 것을 원어절의 모습으로 복원하기가 쉽지만은 않다. ‘커다랗다’나 ‘강아지’와 같은 경우 ‘크/VA+다랗/XSA+다/EF’를 특별한 처리 없이 일괄적으로 복원하면 ‘커다랗/VA+다/EF’가 아니라 ‘크다랗/VA+다/EF’가 될 것이기 때문이다. 합성 명사의 경우에는 사이시옷의 문제가 개입되어 복원하기가 쉽지 않게 된다.

- (6) 명사 접두사를 분석할 때에는 어디까지를 접사로 인정하고 어디까지를 완전히 어휘화되었는가에 대해 작업자마다 일관성을 유지하기 어려운 면이 있다. 일단 여기서는 비교적 생산성이 높다고 인정되는 접두사와 접두사를 분리했을 경우 단일한 표제어로 등재되기 어려운 경우에 한해서 접두사 분석을 하기로 한다.(21세기 세종계획 1999: 163)[밑줄 필자]

---

분석의 문제로 이른바 통사적 접사의 문제(김창섭 1984, 임홍빈 1989, 고창수 1992, 시정곤 1998 등)와 구성소와 형성소의 구분 문제(고영근 1993: 13-15)가 표준 안 및 세종 분석 지침과 관련성이 크다. 기타 영접사, 내적 변화, 중첩어, 통사 구성의 어휘화, 혼성어, 축약어, 역형성어, 한자어의 조어 등 산적한 문제들이 있다. 막상 분석할 목록이 정해진다 하더라도 ‘해돋이’류에 대한 논의들(고영근·구본관 2018: 243의 관련 논저 참고)을 통해 보아온 것처럼 직접 구성 성분이 합성어나 파생어나도 가려내기 쉽지 않을 것이다.

- 9) 유혜원(2004: 71-72)에서는 형태소를 분석할 때 참조하게 되는 사전을 크게 어휘 사전과 접사 사전으로 나누어 볼 수 있다고 하였다. 그러나 이때의 “접사”는 파생 접사에 조사나 어미까지 포함하는 넓은 의미의 접사이다. 본고에서는 별도로 밝히지 않는 이상, ‘접사’를 파생 접사를 가리키는 것으로 사용한다.

(7) 명사파생접미사의 선정 기준(21세기 세종계획 1999: 164-165)

- ① 생산성: 해당 접미사가 현재 생산적으로 사용되고 있는지의 여부
- ② 표제어선정 가능성: 접미사가 결합된 형태를 표제어로 선정하는 데 무리가 없느냐의 여부

임흥빈·송철의(1998)의 ‘어절 분석 표지 표준안’ 이후, 최초의 세종 지침은 종전의 여러 표준 안과 조율을 거쳐 상당 부분 축소·수정되어 나왔다. ‘형태소 분석 말뭉치 구축 지침(1999-1)’이 그것인데, 여기에는 위 (2)의 세종 분석 표지의 초판이 확정되어 있다.<sup>10)</sup> 그리고 21세기 세종계획에서는 파생 접사의 처리 문제에 대하여 (6-7)과 같이 언급하고 있다. 파생법적 성분 분석에 국한되기는 하지만, (6)의 밑줄 부분처럼 문제점을 인식하면서 (7)처럼 “생산성”과 “표제어선정 가능성”을 구분의 원리로 제시하고 있기는 하다. 그러나 이는 또 다른 통계나 연구 결과를 기준으로 해야 하는 기준으로서 임시적인 원리이다. 실상 복잡하게 목록 제시에 의존하게 되는 불완전한 방법일 뿐이다. 21세기 세종계획(1999: 140-167)을 보면 지침 내용의 상당 부분을 목록 제시에 할애하고 있다. 파생 접사로 한정하여 본다면 한자어 체언 접두사 목록, 고유어 체언 접두사 목록, 한자어 명사 파생 접미사 목록, 고유어 명사 파생 접미사 목록, 동사 파생 접미사 목록, 형용사 파생 접미사 목록 등이 일일이 열거되고 있다.

이러한 지침 아래서 말뭉치의 구축자는 작업 과정에서 이 목록을 기준으로 삼게 될 것이며 일일이 숙지하기가 어려워져 일일이 확인을 해 가며 작업을 해야 하기 때문에 작업의 효율성을 잃게 될 것이다. 구축된 말뭉치를 이용하는 입장에서도 지침의 이 목록들과 그 선정 취지를 일일이 확인하고 숙지를 해야만 해당 말뭉치를 제대로 이용할 수 있게 된다. 그뿐만 아니라

---

10) 세종 지침의 최종판은 21세기 세종계획(2007: 183-224)의 ‘형태분석 말뭉치 구축 지침(2005-1)’이다. 초판과 최종판 사이에 세부적인 기술에 차이가 있기도 하지만, 분석 표지(21세기 세종계획 1999: 141)에는 큰 차이가 없다. “어기”가 “어근”으로 바뀐 정도이다. 본고에서는 초판부터 최종판에 이르기까지의 차이점을 세세히 다루지는 않는다. 대체로 변함이 없었던 전체 개념 체계를 문제 삼는다.

형태론 연구자들은 본인의 견해와 차이가 있는 부분들까지 일일이 확인을 하고 재가공을 해야만 온전하게 그 말뭉치를 활용할 수 있게 된다. 만약 분석된 대로 활용을 하려면 연구 자료로서 활용할 수 있는 폭이 상당히 협소해질 수밖에 없다. 지침이 체계적이지 못하고 복잡하면 분석할 때도 어려움을 초래할 뿐만 아니라 연구의 실효성마저 떨어지게 한다.<sup>11)</sup> 이러한 단점을 가진 조어론적 분석을 기본적으로 취할 필요가 있는가 하는 점이 문제의 핵심이다.

결국 (2)의 세종 분석 표지는 (1)의 전제 조건에 비추어 볼 때 여러 문제를 안고 있는 셈이다. 통사 층위와 조어 층위가 혼재되어 있어서 (1ㄱ)의 개념의 정연성도 문제가 되고 있으며, 목록화의 방법을 취하게 되어 (1L)의 용이성도 (1ㄴ) 활용의 실질성도 크지 않다. 또한 조어 성분이 한데 섞여 있어서 (1ㄷ)의 유연성도 떨어지게 된다. 예를 들어 세종 지침을 따라서 구축한 형태 분석 말뭉치를 가지고 그 다음에 구문 분석 말뭉치를 만들려면 재가공을 거쳐야 할 것이다. 축약이 빈번한 구어나 방언에 적용하려 해도 어디까지 분석을 해야 할지 개념적으로 기준을 잡기가 어려워질 것이다.

이상의 문제점들은 통사 층위, 즉 통사 단위 구성의 층위를 기본 분석 층위로 삼고 조어 층위, 즉 단어 형성의 층위는 필요에 따라 부가 분석 층위로 삼으면 상당히 해소할 수 있다. 이러한 방법이 대규모의 말뭉치를 구축하기에도 효율적이다.<sup>12)</sup>

(3L), 세종 분석 표지의 두 번째 문제점은 문법 단위에 대한 체계가 가지는 문제점이다. 주지하다시피 “체언/용언/수식언/독립언/관계언/의존형태”의 수평적 “대분류”는 체계적이지 않다. 조사를 따로 “관계언”이라는 단어로서

11) 다소 극단적이기는 하나, 이진병(2017: 239)에서는, “분석 표지의 세밀화가 더 많은 것을 제공해 주리라는 것이 그간의 전제였다. 그러나 그렇지 않을 수도 있다는 사실을 살펴보았다. 실질적으로 말뭉치를 대하는 대다수의 일반 연구자들에게, 너무 재단된 자료는 오히려 쓸모가 없을 수도 있는 것이다.”라고 평가하기도 하였다.

12) 이에 대해서는 다음 2.3.에서 자세히 다룰 것이다.

인정을 하는 동시에, 어미와 과생 접사를 한데 묶어서 품사 분류 체계와 나란히 분류하는 것은 국어학적으로 보기 드문 체계이다. 용어상으로도 “의존 형태”에 조사가 포함될 수 있어서 문제이다. 이는 (3ㄱ)의 문제와도 관련이 될 터이다. (3ㄱ)이 굴절과 조어라는 문법 층위의 구분에 대한 문제라면, 이 (3ㄴ)의 지적은 문법 단위의 구분에 대한 문제이다. 문법 층위에 대한 것이냐, 문법 단위에 대한 것이냐를 차치하고 본다면 (3ㄱ)과 (3ㄴ)의 발단은 모두 문법 형태소, 즉 조사와 어미와 접사의 처리에 있다. 일단 (3ㄱ)에 대한 논의에서 조어 층위의 분리를 언급한 만큼 여기서 남은 핵심은 조사와 어미이다. 조사와 어미를 어떻게 설정할 것이냐는 단어관으로 직결된다. “체언/용언/수식언/독립언/관계언”까지의 구분도, 기존의 표준 문법 체계가 단어와 어절의 구분에 대하여 여전히 논란을 내포하고 있다는 점을 고려할 때, 말뭉치의 분석 체계로 그대로 수용할 필요가 있었는가 하는 문제에 걸린다. 더 나은 분류 체계가 있다면 적극적으로 수용할 필요가 있었을 것이다. 그런데 기존의 단어/어절 구분 문제에 더하여, 어미와 접사를, 심지어 어근까지를 한데 묶어서 “의존형태”로 “체언/용언/수식언/독립언/관계언”에 병치시키고 있다. 예를 들어 표준 문법은 ‘기준을 없애다’에 대하여 ‘기준을/없애다’나 ‘기준/을/없애다’나 ‘기준/을/없/애/다’로 단위 분석을 하여 왔고 품사 분류는 ‘기준/을/없애다’의 분석 결과에 부여해 왔다. 그런데 세종 분석 표지는 그 어디에 부여되는 것인지 기준을 잡을 수 없다. 이로써 세종 분석 표지의 분류는 기존 표준 문법 중심의 단위 체계인 ‘어절(조사 불인정, 어미 불인정, 접사 불인정) : 단어(조사 인정, 어미 불인정, 접사 불인정) : 형태소(조사 인정, 어미 인정, 접사 인정)’의 구분을 취한 것도 아니다. 그렇다고 통합 관계에 따라 ‘통사 단위 구성 : 단어 형성’의 구분을 취한 것도 아니다. 결과적으로 납득하기 힘든 단어관을 취하고 있는 셈이다.<sup>13)</sup>

한 가지 분명한 사실은 조사와 어미를 분석하지 않으면 형태 분석 말뭉치는 물론이고 구문 분석 말뭉치에 이르기까지 그 말뭉치의 효용성이 매우 떨

13) ‘어절’을 개념적으로도 중시하여서 지침에서 사용하는 것은 되도록 피하는 것이 좋다. 어절 설정의 문제점은 유현경(2018: 396-398, 403) 등을 참고할 수 있다.

어진다는 점이다. 반면 파생 접사의 분석은 앞서 논의했듯이 부정적인 측면도 존재한다. 이러한 문제점을 해결해 줄 수 있는 이론적인 방안은 ‘통사론적 단어’에 대한 설정이다. 21세기 세종계획 당시보다 논의가 더 무르익은 작금의 상황에 비추어 볼 때, ‘음운론적 단어’, ‘형태론적 단어’, ‘통사론적 단어’ 등의 개념에 대해 충분히 고려하고 수용할 필요가 있다. 이른바 개정 표준 문법 연구에 대한 결과물인 유현경(2018: 198)에서도 “최근 가장 많이 논의된 통사 원자(統辭原子, syntactic atom)가 이에 해당하는데, 이에 따르면 국어의 조사나 어미 등은 통사 구성에 참여하는 단위이므로 당연히 단어가 된다.”라고 수용하고 있다.<sup>14)</sup>

형태 분석 표지는 통합 관계의 종류에 따라 통사 층위와 조어 층위를 구분할 필요가 있다. 문법 단위의 측면에서도 단어 형성의 결과물로서의 단어와 통사 단위 구성에 참여하는 구성 성분으로서의 단어를 일관되게 파악할 수 있도록 하여야 한다. 그런데 이를 실현할 수 있는 단어의 개념이 바로 ‘통사론적 단어’이다. 이는 조사와 어미를 “체인/용언/수식언/독립언”과 같은 문법 층위이자 단위로 취급하면서, 파생 접사를 제외하여 별도로 처리할 수 있게끔 하는 개념이 된다. 통사론적 단어를 형태 분석의 기본 단위로 개념짓는다면, 조어의 층위 및 관련 단위와 통사의 층위 및 관련 단위를 체계적으로 구분할 수 있다.<sup>15)</sup>

(3ㄷ), 세종 분석 표지의 세 번째 문제점은 “대분류/소분류/세분류”의 구분이 결과적으로 유의미한 구분이 되지 못하였다는 점이다. 세종 분석 표지 이전에 제안된 원안이라 할 수 있는 임흥빈·송철의(1998)에 따르면 다음과 같은 설명들이 보인다.

14) 통사 원자나 통사론적 단어에 대해서는 Sciallo & Williams(1987), 박진호(1994), 최형용(2002, 2010), 한정환(2011), 김용하 외(2018)등을 참고할 수 있다. 3.1.에서는 이를 주요하게 다룰 것이다.

15) 자세한 내용은 3장에서 다시 다룰 것이다.

- (8) ㄱ. 기존의 연구와 본 표준안이 크게 다른 것은 본 표준안이 단계적 분석 표지 집합을 설정한 것이다. 단계적 분석 표지 집합이란 품사-형태 분류의 세분 정도를 단계화하여 사용자가 자신의 목적이나 용도에 맞는 분석 표지 집합을 선택할 수 있도록 한 것을 말한다.(임흥빈·송철의 1998: 193)
- ㄴ. 본 표준안은 분석의 단계가 높아질수록 분석 표지는 더 세분되는 양상을 보이게 된다.(임흥빈·송철의 1998: 285)
- ㄷ. 파생 접미사를 분석할 때에는 원말 뒤에 어기와 접사를 분석하여 괄호 속에 묶는다.(임흥빈·송철의 1998: 286)
- (9) ㄱ. 제1단계 분석 (텍스트 1)  
 기자묘/n 근처/n+에서/j 무슨/md 슬픈/v 음률/n+이/j 봄/n 공기/n+를/j 진동시키며/v 날아오는/v 것/n+이/j 들렸다/v ./s
- ㄴ. 제5단계 분석 (텍스트 1)  
 기자묘/nnp 근처/nng + 에서/jkb 무슨/mdd 슬프/vag + ㄴ/efd 음률/nng + 이/jks 봄/nng 공기/nng + 를/jko 진동시키(진동(진동nng))/wb=시키/xsv/vvgt + 며/efcs 날아오/vvgi + 는/efd 것/nnb + 이/jks 들리(들(들/vvgt)/wb=리/xsv)/vvgi + 었/ep + 다/eft+./sf(이상 임흥빈·송철의 1998: 287)

위의 (8)을 보면 임흥빈·송철의(1998)의 원안(이하, ‘표준 안’)은 단계적 분석, 즉 필요에 따라서 1단계만 분석을 하거나 2단계까지만 분석을 하거나 최대 5단계까지 분석할 수도 있는 방안을 고려하고 있다. 그리고 그러한 설정은 (9)의 분석 예시를 통해서도 확인할 수 있다. 표준 안을 따랐다면 지침이 갖추어야 할 유연성((1ㄷ))을 더 가질 수 있었을 것이다. 그러나 세종 말뭉치는 물론이고 이후의 형태 분석 말뭉치는 말단의 분석만이 취해지고 있는 실정이다. 물론 표준 안이 5단계를 취하는 등으로 다소 복잡하다는 문제점이 있었을 것이다. 그러나 세종 분석 표지는 (1ㄴ)의 단순성을 늘리는 대신, (1ㄷ)의 유연성을 잃는 결과를 낳았다. 위 (9ㄴ)에서 괄호 안을 제거하면 상위 단계의 결과물을 얻을 수 있다. 통사 층위에 대한 분석 결과물과 조어

층위에 대한 분석 결과물을 어느 정도 호환시킬 수 있다. 그러나 세종 말뭉치는 그러하지 못하다. 이러한 모습은 세종 분석 표지가 가지는 이 문제점이 앞서 논의한 (37-L)의 문제와도 연결됨을 보여 준다.

### 3. 통사론적 단어를 기준으로 한 이원화

#### 3.1. 개선 방향의 검토

지금까지 세종 분석 지침의 문제점을 논의하였다. 문법적 개념이 문법 층위의 구분, 문법 단위의 구분에서 체계적이지 못하다는 문제점을 지적하였다. 또한 분류 단계 구분의 실효성도 떨어진다고 보았다. 이러한 문제점은 결국 지침이 갖추어야 할 여러 조건이 미비했음을 뜻한다. 기존 지침을 문제시 삼은 이상, 이제 이 절에서는 세종 분석 지침에 대하여 다음의 사항을 쟁점으로 삼아야 할 것이다.

- (10) ㄱ. 문제점을 개선할 수 있는 대안이 있는가?
  - ㄴ. 그 대안은 이론적 근거를 가지고 일반성을 획득할 수 있는가?
  - ㄷ. 그 대안은 기구축 말뭉치와의 호환성에 문제가 없는가?

2.2.에서는 기존 지침의 문제점을 정리하면서 그 문제점에 대한 대안도 제시하였다. 그때 제시하였던 대안을 정리하여 다음과 같이 설정할 수 있다.

- (11) ㄱ. 문법 단위의 설정: 통사론적 단어를 기준 문법 단위로 한다. 즉 조사와 어미를 단어로 취급한다.
  - ㄴ. 문법 층위의 설정: 통사론적 단어를 기준으로 하여서 다음과 같이 중층적으로 설정한다.
    - ① 기본 층위: 통사 층위(또는 ‘통사론적 층위’)
    - ② 부가 층위: 조어 층위(또는 ‘형태론적 층위’)

ㄷ. 위 ㄴ의 두 층위에 따라서 분석의 단계 및 해당 지침을 이원화한다.

(11ㄱ)은 (3ㄴ)의 문제점에 대한 대안 설정이며, (11ㄴ)은 (3ㄱ)에 대한 대안 설정이다. 논의의 흐름에 따라 문제점을 지적하였던 것을 여기서는 개념 설정의 절차를 따라서 재배치하였다. (11ㄱ)의 통사론적 단어를 기준으로 상위 통사 구성으로부터 이 통사론적 단어를 구성 성분으로 분석해 내는 과정이 (11ㄴ①)의 기본 층위이다. 2.2.에서는 이를 ‘통사 단위 구성의 층위’라 하여 ‘단어 형성의 층위’와 대조하였는데, 이제 통사 단위 구성의 층위는 통사 층위 또는 통사론적 층위라 부를 수 있다. 단어 형성의 층위는 조어 층위 또는 조어론적 층위라 부를 수 있을 것인데, 조사와 어미를 단어로 보는 이상 굴절론이 사라지게 되므로 조어론적 층위는 곧 형태론적 층위라 부를 수 있게 된다. 형태론적 층위는 (11ㄴ②)의 부가 층위이다. 이 층위는 통사론적 단어를 구성체로 삼아서 그 하위의 내부 구성 성분으로 세분화하는 과정이다. (3ㄷ)에서는 세종 분석 지침이 다원화되지 못하여 유연성이 떨어진다는 문제점을 지적하였다. 위 (11ㄷ)과 같이 기본 층위를 1단계 분석으로 설정하고 부가 층위를 2단계 분석으로 설정하여서, 기본적인 대용량의 형태 분석은 1단계 위주를 진행하면서 필요에 따라서 2단계의 부가 층위 분석을 추가해 갈 필요가 있다. 이렇게 이원적으로 말뭉치를 구축해 가는 것이 대용량의 말뭉치 구축에도, 다양한 언어 자료를 품고 있는 말뭉치 분석에도 유연하게 적용할 수 있는 방안이 될 것이다.

그렇다면 (11)과 같이 선정한 이론적 근거는 무엇이며 일반성을 획득할 수 있는가? 이는 (10ㄴ)의 질문이다. 이에 대하여 살펴보면 다음과 같다. 먼저, 조사와 어미를 품사 분류의 대상과 동등하게 단어로 인정한 견해를 확인할 필요가 있다. 박진호(1994), 임홍빈(1997), 이훈(2003), 최형용(2002, 2010), 한정환(2011), 김용하 외(2018) 등을 비롯한 다수의 논의가 이에 해당한다.

박진호(1994: 4)에서는 Bloomfield(1933)의 자립성(freeness), Hockett(1954)의 휴지라는 전통적인 단어의 정의가 음운론적임을 지적하였다. 그러면서

“통사론의 기본 단위로서의 단어”인 “통사원자(syntactic atom)”을 제안하였다.

이훈(2003: 133-135)에서는, 최현배(1937/1971: 147)의 “따로설만함(分立成)” 내지 “여러 가지의 자리에 설 수 있는 성질”도 단어의 기준으로 적절치 않음을 지적하면서, 단어의 기준으로 남는 것은 “낱말의 구성 부분인가 아닌가”에 있다고 하였다. “어떤 요소가 그 외재적 실현상 다른 요소에 기대어 나타날지라도 그것이 낱말의 구성요소가 아니라면, 분리되어야 하고 통사적 지위를 부여받아야 한다.”라고 분명한 형식적 기준을 제시하였다. 그리고 품사를 어휘적/문법적 기능의 차이에 따라 크게 “내용어”와 “기능어”로 양분하고 있다. 이때 조사와 어미는 기능어에 포함되며, 파생 접사는 “낱말의 구성 요소”이기 때문에 단어가 아니며 품사의 자격도 부여받지 못한다.

최형용(2002: 13-90)에서는 단어를 “음운론적 단어 / 어휘적 단어 / 문법적 단어”로 구분하면서 조사와 어미를 “문법적 단어”로 취급하였다. 최형용(2010: 85)에서는 “단어 형성과 문장 형성의 층위를 구분해야 하고 품사는 가장 작은 통사 단위를 대상으로 해야 한다”라고 역설하고 있다. 이는 이훈(2003)과 동계의 주장이다. 그리고 “접미사는 통사적 구성에 참여하는 일이 없다.”라고 선을 그었다.

김용하 외(2018: 36-40)에서는 기존 품사(parts of speech)가 “개념적인 정의에 의거한 언어 요소들의 분류”이기 때문에 문제가 있다고 정리하면서, 이에 대한 대체 분류 체계로 “통사 범주(syntactic category)”를 다음과 같이 밝히고 있다. 이는 용어의 차이가 있을 뿐 이훈(2003: 135)에서 제시한 분류 체계와도 같다.

- (12) ㄱ. 어휘 범주(lexical category): 명사, 동사, 형용사, 부사, 관형사
- ㄴ. 기능 범주(functional category): 조사(particle), 굴절소(inflexion), 보문소(complementizer)

위에서 “동사”와 “형용사”는 기존의 어간만을 말하며, 굴절소는 기존의 선

어말 어미에 해당하고, 보문소는 어말 어미에 해당하는 것이라고 할 수 있다. 구분의 체계를 위의 (12)에 준하여 해석하는 것이 기존 세종 분석 표지보다 그 이론적 바탕에 일관성이 있다. 조사 및 어미는 선행 체언 및 선행 어간과 분리하여 분석해야만 한다. 그렇지 않고 조사와 어미를 선행 성분과 하나로 통합하여 분석하면 이후로 이어지는 구문 분석은 기대할 수 없게 된다. 김용하 외(2018: 56-68)에서도 조사와 어미의 문법적 지위에 대하여 그 근거를 상세히 다루었지만, 핵심은 한국어의 조사와 어미가 통사론적 구에 결합하는 것으로 일반화된다는 사실이다. 널리 알려진 예의 하나인데 아래와 같은 예에서 (13ㄱ)·(14ㄱ)은 틀리고 (13ㄴ)·(14ㄴ)은 맞는 것으로 이해된다.

- (13) ㄱ. \*[[어제 온] [사람이]] 죽었다.  
 ㄴ. [[어제 온 사람]이 죽었다.(이상, 한정환 2011: 45)
- (14) ㄱ. \*[[서울에서는] [철수가, [부산에서는] [영희가] [죽었다]].  
 ㄴ. [[[서울에서는] [철수가, [부산에서는] [영희가 죽]었다]].(이상, 한정환 2011: 45)

한정환(2011)에서는 한국어 단어의 설정 문제에 대하여 천착하고 있다. “국어의 단어는 통사부의 기본 단위라는 그 문법적 지위의 중요성에도 불구하고 아직 그 정의조차 분명하지 않은 혼란스러운 개념이다.”(p. 64)라고 지적하였다. 그러면서 그간 단어 분류(또는 품사 분류)의 기준으로 적용되었던 기준들이 “결코 어우러져서(뒤섞여) 적용되어서는 안 되는 기준들”(p. 65)인데 그 기준들이 너무 많을뿐더러 원칙 없이 뒤섞여 있다는 점을 그 이유로 지적하였다. 한편 어미가 라틴어와 같은 굴절 범주로 볼 수 없음을 정리하기도 하였다(p. 45). 모든 기준, 모든 언어를 충족시킬 수 있는 단어의 설정은 있을 수 없다. 한국어 분석을 위해 기준 단어를 설정할 때, 현재까지의 연구 성과 중에서 가장 적합한 개념은 통사론적 단어이다.

실상 (3)에서 거론한 문제와 관련된 고민들이 어느 정도는 형태 분석 표

준 안을 연구한 임흥빈·송철의(1998)(이하, 별도의 수식이 없는 ‘표준 안’은 이를 말함.)과 구문 분석 표준 안을 연구한 임흥빈·이홍식 외(2002)에서도 엿보인다. 그런데 표준 안의 설정이 ETRI의 ‘태그세트 표준안’, 고려대학교 자연어처리연구실의 ‘KUNLP 표준안’ 등과 함께 협의 과정을 거친 후에 결과적으로 (3)의 문제점을 남기게 되었다.16)

(15) 어절 분석 표지 표준 안(임흥빈·송철의 1998: 271)

어휘- 문법적 범주	품사 단위	기본 품사	확장 품사				
		대단위	소단위	저단위	고단위		
		체언 n	명사 mn	일반 명사		nng	
				고유 명사		nnp	
				의존 명사 nmb	일반 의존 명사		nmbg
					단위 의존 명사		nmbc
		대명사 np	인칭 대명사		npp		
			지시 대명사		npd		
		조사 j	수사 nr	양수사		nrc	
				서수사		nro	
		용언 v	동사 vv	일반 동사	자동사	vvgi	
				vvg	타동사	vvgv	
				보조 동사		vvx	
			형용사 va	일반 형용사		vag	
	존재 형용사			vat			
	지정 형용 사 vac			긍정 지정 형용사	vacp		
		부정 지정 형용사	vacn				
	보조 형용사		vax				
	관형사 md	지시 관형사		mdd			
		수 관형사		mdq			
성상 관형사		mds					
부사 mb	일반 부사	성분 부사	mbgc				
	mbg	양태 부사	mbgm				
	접속 부사		mbc				
감탄사 ic	감탄사		ic				
교착소 단위	지단위	중단위	고단위				
	조사 j	격조사 jk	주격 조사	jks			
			보격 조사	jkc			
			관형격 조사	jkg			
			목적격 조사	jko			
			부사격 조사	jkb			
			호격 조사	jkv			
인용격 조사	jkq						

16) 21세기 세종계획(1999: 140)를 참고함.

				(서술격 조사)	jkp				
				보조사	jx				
				접속 조사	jc				
				어미 e	선어말 어미	ep			
					어말 어미 ef	종결 어미	eft		
						연결 어미 efc	대등적 연결 어미	efcc	
							종속적 연결 어미	efcs	
							보조적 연결 어미	efcx	
						전성 어미	명사형 전성 어미	efn	
				관형사형 전성 어미	efd				
				어기- 접사 단위	접사	어기			wb
						형태적 접사	통사적 접사		
					접두사 xp		명사 접두사	xpn	
							동사 접두사	xpv	
							형용사 접두사	xpa	
접미사 xs	명사 파생 접미사	xsn							
	동사 파생 접미사	xsv							
	형용사 파생 접미사	xsa							
	부사 파생 접미사	xsb							

위의 표는 표준 안의 분석 표지이다.<sup>17)</sup> 위의 분석 체계는 세종 분석 표지와 같은 문제점도 많지만 몇 가지 측면에서 준용할 만한 사항들이 있다. 먼저 어휘 범주와 문법 범주에 대한 인식이다. 다음으로 조사와 어미를 묶어 “교착소”라는 범주를 설정하였다는 점이다. 같은 층위로 병립하여 있다는 점이 문제이기는 하나 조어 층위를 “어기-접사 단위”로 구별하였다는 점도 주목할 만하다. 2.2.에서도 언급하였지만 단계별로 다원화하여 분석할 수 있도록 고려한 점도 수용할 만하다. 물론 분석의 단계가 5단계에 이를 정도로 복잡하다는 점은 문제점일 수 있다.

임홍빈·송철의(1998)에서 특히 주목할 만한 점은 “교착소”의 설정에 대한 것이다. “어기-접사 단위”를 제외한다면 “품사 단위”로 분류된 것은 어휘 범주로, “교착소 단위”로 분류된 것은 문법 범주로 재분류될 수 있다. 이 교착소는 임홍빈(1997)에서 제안되었으며 이선웅(2012: 123-124) 등에서 인정

17) 원안에는 “기타 범주”가 있어서 “어휘-문법적 범주”와 대립이 되지만, 본고에서는 문장 부호 등을 논의하지 않기 때문에 앞서 (2)에서 제시한 세종 분석 표지와 마찬가지로 “기타 범주”를 제외하고 표를 제시하였다.

되고 있다. 임홍빈(1997)과 이선웅(2012: 123-124)의 정리에 따르면, 교착소를 설정해야 하는 핵심은 다음과 같다. 한국어의 조사, 어미는 여러 문법 현상에 비추어 볼 때 의심할 바 없이 확실하게 통사적 요소이다. 그런데 이를 수용할 만한 용어가 ‘굴절’뿐이었다. 그러나 어형변화를 통해서 문법적 기능을 실현하는 것을 뜻하는 이 굴절은 인구어에서 형태론적인 개념이 강하다. 한국어에서 ‘어간+어미’, ‘체언+조사’가 형태론적으로도 인식되는 것은 재구조화(restructuring)에 의한 것일 뿐, 어미와 조사는 통사론적 요소로 취급되어야 한다는 것이다.<sup>18)</sup> 이러한 이유로 교착소를 설정하는 주장이 성립되었던 것이다. 예를 들어 조사는 “체언구교착소”가 되고, 어미는 “용언구교착소”가 된다고 하였다.

(16) ㄱ. 분석의 단위는 품사나 형태소이다. 어휘적 요소는 품사를 단위로 하며, 문법적 요소는 대체로 형태소를 단위로 한다. 그러나 품사 분류와 형태소 분석은 동일한 차원에 있는 것이 아니다. 분석 차원의 구별이 필요하다.(임홍빈·송철의 1998: 212)

ㄴ. 교착소는 조사와 어미를 포괄하여 특별히 따로 부르는 이름이다. (임홍빈·송철의 1998: 268)

ㄷ. 본 표준안에서는 “교착소” 범주 및 단위를 설정하기로 한다. 조사나 어미를 “교착소”로 설정함으로써 그것이 통사적 관련 요소라는 사실을 불완전하게나마 반영하려는 것이다.(임홍빈·송철의 1998: 512)

조사와 어미의 개념적인 처리 문제, 즉 용어 문제에 대한 고민이 바로 (15)의 어절 분석 표지 표준 안에 반영되어 있다. (16ㄱ)과 같은 고민 아래 (16ㄴ-ㄷ)과 같이 설정하고 있는 것이다.

‘교착소’라는 용어는 당시에 널리 정착된 용어가 아니었던 것으로 보인다. 그럼에도 불구하고 본고에서는 조사와 어미를, ‘굴절 접사’와 같이 형태

18) 재구조화(restructuring)에 대하여 서태룡(1988: 26-32)을 참고할 수 있다.

론적 이해가 담겨 있는 개념보다는 ‘교착소’와 같이 같이 통사론적 개념으로 이해하는 것이 언어 분석의 입장에서 효용성이 크다고 본다. 다만 ‘교착소’만으로는 일반성이 부족하다고 보아서 일반적인 용어로 뒷받침하는 것이 수용성이 높다고 본다. 그것이 바로 (12)의 “어휘 범주”와 “기능 범주”이다.<sup>19)</sup>

이러한 견해는 일찍이 자연어 처리 분야에서도 그 필요성이 다루어졌다. 강승식(1993, 2003: 81-83)에서는 “형태소 분석을 위한 품사 체계”로 “대분류/중분류/소분류”의 체계를 제시하였는데 그 중 대분류 체계를 보면 “체언/용언/수식언/접사/조사/어미”로 제시하였다. 그리고 이 분류는 개념적으로 “어휘형태소”와 “문법형태소”라는 두 부류의 결합 관계를 중시한 입장이다. 조사와 어미의 분류가 구문 분석이나 의미 분석 단계에서 중요함도 언급하고 있다. 다만 문제는 “접사”를 품사의 한 종류로 넣은 이유가 명시적이지 않다는 점인데 대개 ‘-하다’와 같이 매우 생산적이고 통사적인 접사의 문제 때문이 아닐까 한다. 파생 접사를 체계적으로 이해하여 반영한 것은 아닌 것으로 보인다. “접두사와 체언접미사는 체언 유형에 영향을 미치지 않는다. 따라서 접두사와 체언접미사를 체언에 부가된 정보로 간주할 수 있다.”라고 하면서 파생 접사를 “체언부”로 포함시키야 한다고 거론한 점을 보면 파생 접사는 대체로 부가적인 사항으로 처리하였다고 볼 수 있다.<sup>20)</sup>

다음으로 (10ㄷ)의 문제에 대하여 논의하고자 한다. 기존 지침에 문제점이 있다 하더라도 그 지침에 의해서 이루어진 말뭉치가 상당수가 존재한다면 일종의 선점 효과를 고려해야 한다. 말뭉치의 실효성도 고려해야 하기 때문

---

19) 조사와 어미를 묶는 또다른 용어로 ‘굴절 접사’나 ‘통사적 접사’가 있다. 그러나 ‘굴절 접사’는 앞서 지적한 것처럼 형태론으로 읽힐 소지가 있기에 피하는 것이 좋다. 유혜원(2004: 71-77)에서는 형태소 분석을 위해 필요한 “접사 사전”을 정리하면서 접사를 “어휘적 접사”와 “통사적 접사”로 분류하는 방안을 제시하고 있다. 그러나 이 또한 형태론적으로 문법 형태소를 분류하려는 개념이다. 조사와 어미를 명사, 동사 등과 함께 1차적인 분석 대상으로 묶어 줄 만한 개념 체계가 아니다.

20) 강승식(2003)을 보면 조사와 어미는 그 분석에 대하여 상당 부분을 할애하여 중요하게 다루고 있는 모습을 볼 수 있다. 하지만 파생 접사는 그렇지 않다.

에, 어떤 문제가 발견될 때 그때그때 새로운 방식을 도입할 수는 없는 노릇이다. 따라서 새로운 방안을 도입하더라도 기구축 말뭉치와의 호환성을 고려하여서 도입해야 한다.

이를 위해서는 기구축 형태 분석 말뭉치의 모습을 고려할 필요가 있다. 개념적인 분류, 중간 또는 상위 층위의 재설정에는 말뭉치 구축자나 구축자에게 중요하다. 하지만 말뭉치 자체의 형상에는 말단의 태그만이 표시될 뿐이다. 예를 들어서 조사와 어미를 단어로 보느냐, 통사론적 층위의 것으로 보느냐와 상관없이 “어미는 중요하다”는 “어미/NGG+는/JX 중요하다/VA+다/EF”로 남아 있을 뿐이다. 따라서 최대한 말단의 태그는 유지하면서 분석 표지의 체계를 개선하면 될 일이다.

한편, 말뭉치 구축의 추세는 양적으로 대규모화시키거나 질적으로 세밀화시켜 가는 다소 양립하기 힘든 두 차원으로 흘러가고 있다. 이 두 방향의 절충안은 층위를 구분하여 구축하는 것이다. 예를 들어 개괄적으로 분석하는 방식의 말뭉치를 1차적으로 구축을 하고 필요한 경우 및 필요한 양에 한하여 질적 세밀화라는 2차적 구축을 취하는 방식이다. 이러한 방식은 일부 제한적인 파생 접사를 분석해 놓은 기구축 말뭉치의 문제 때문이라도 필요하다. 2.2.에서 중요하게 다룬 것은 파생 접사 분석의 문제였다. 다시 한 번 정리하지만, 조어의 불규칙성은 분석의 대규모화에도 저해 요소가 되며, 질적 세밀화의 취지를 충족시키기에도 결코 쉽지 않은 난제를 던져 준다.<sup>21)</sup> 기구축 말뭉치의 제한적인 파생 접사 분석도 이러한 난제를 고스란히 안고 있다. 그러나 기구축 말뭉치의 가치도 무시할 수만은 없다. 따라서 기본적으로는 통사론적 단어까지만 1차 분석을 하고, 필요 시에 부가적으로 조어론적 요소까지 2차 분석을 하는 방식이 필요하다. 이는 기존 지침의 문제점과 기구축 말뭉치의 가치를 고려해서 취할 수 있는 일종의 절충안이다.

덧붙여서 또 한 가지의 호환성을 고려할 필요도 있다. 한 단계의 말뭉치

---

21) 일부의 파생 접사를 접사를 분석한다고 해서 기계 학습 등에 크게 도움이 될 것으로 보이지 않는다. 오히려 의미 분석의 세밀화가 기계 학습이나 인공 지능 개발 등에도 도움이 될 것으로 보인다.

는 그 상태로 그치지 않고 다음 단계 말뭉치 구축의 밑 작업이 되기도 한다. 대표적으로는 구문 분석 말뭉치 구축을 고려하면서 형태 분석 말뭉치 지침을 세워야 한다. 이러한 고려는 이미 임홍빈·송철의(1998)만이 아니라 임홍빈·이홍식(2002)의 구문 분석 표지 표준 안으로도 일관되게 이어져 있던 것이다.

(17) ㄱ. 형태 분석에서와 마찬가지로 구문 분석에서도 어절 단위를 중심으로 분석하려는 강한 경향을 보여 왔다고 할 수 있다. 그러나 기본적으로 한 어절은 어휘 요소와 문법 요소로 이루어진다. 두 이질적인 기능의 복합체이다. 이 둘을 하나의 단위로 다루면 기능의 충돌이 생기게 되는 것은 피할 수 없는 일이다.(임홍빈·이홍식 2002: 42)[밑줄 필자]

ㄴ. 구문 분석은 적어도 어휘 요소와 문법 요소를 분리하여야 하며, 나아가서는 문법 요소 중심의 체계를 구성해야 한다.(임홍빈·이홍식 2002: 42)

ㄷ. 문법 범주의 다양성을 훼손하지 않으면서, 교착소 중심의 국어 구문 구조를 최대한 살릴 수 있는 문법이 요청된다고 할 수 있다.(임홍빈·이홍식 2002: 42)[밑줄 필자]

(17)은 구문 분석 표준 안 작성을 위한 기본 원칙으로 제시된 것 중 일부이다. 특히 (17ㄱ)의 밑줄 그은 부분은 어휘 범주와 기능 범주는 반드시 분리하여 분석해야 하며, 기능 범주로서는 조사 어미가 반드시 분석되어야 함을 선언하고 있다.

### 3.2. 통사론적 단어 분석 표지 안

이상의 논의에 따라 본고에서 제안하는 분석 표지를 보이면 다음의 (18)과 같다. 기준 단위는 통사론적 단어이다.

(18) 통사론적 단어 중심의 분석 표지 안

범주 구분	1단계: 기본 층위 분석			2단계: 부가 층위 분석			
	통사 단위 구성 층위		태그	단어 형성 층위		태그	
어휘 범주	체언	명사	일반 명사	NNG	*	*	
			고유 명사	NNP	*	*	
			의존 명사	NNB	*	*	
		대명사	NP	*	*		
		수사	NR	*	*		
	용언	동사	VV	*	*		
		형용사	VA	*	*		
		보조 용언	VX	*	*		
		지정사	긍정 지정사	VCP	*	*	
	부정 지정사		VCN	*	*		
	수식언	관형사	성상관형사	MMA	*	*	
			지시관형사	MMD	*	*	
			수관형사	MMN	*	*	
	부사	일반 부사	MAG	*	*		
		접속 부사	MAJ	*	*		
	독립언	감탄사	IC	*	*		
			-	어근		XR	
	기능 범주	교착소	조사	주격 조사	JKS	*	*
				보격 조사	JKC	*	*
				격	관형격 조사	JKG	*
조				목적격 조사	JKO	*	*
사				부사격 조사	JKB	*	*
				호격 조사	JKV	*	*
				인용격 조사	JKQ	*	*
			보조사	JX	*	*	
			접속 조사	JC	*	*	
어미			선어말 어미	EP	*	*	
		종결 어미	EF	*	*		
		부사형 어미	EC	*	*		
		명사형 어미	ETN	*	*		
		관형사형 어미	ETM	*	*		
			접두사	체언 파생 접두사	XPN		
			접미사	명사 파생 접미사	XSN		
				동사 파생 접미사	XSV		
				형용사 파생 접미사	XSA		

1단계는 통사구, 즉 문장이나 절을 포함하는 개념인 통사구를 그 구성 성분인 통사론적 단어로 분석해 내는 단계이다. 1단계의 분석에서는 단어 형성

층위를 배제하였음을 확인할 수 있다. 이 점이 세종 분석 표지의 문법 체계와 가장 다른 점이다. 앞서 3.1.에서 표준 안과 관련하여서 언급하였듯이 교착소의 개념을 반영하여서 기능 범주의 하나로 위치시킨 것도 특징이다.<sup>22)</sup>

그러나 말단의 표지에 영향을 줄 만한 구분은 되도록 피하였다. 기구축 말뭉치와의 호환성을 위한 방편이다. 예를 들어 지정사의 설정은 불필요할 수 있다. 밝혀진 바로는 아직 형용사 중에 지정사와 동음이의어가 없기 때문에, 견해 차이에 따라 ‘이/VA’로 처리하여도 ‘이/VCP’로 처리하여도 결과적으로 큰 차이는 없을 수 있다. 상호 변환이 가능하기 때문이다. 특히 ‘이/VCP’와 ‘아니/VCN’의 차이는 큰 의미가 없을 수 있다. 이때 (10c)에 대하여 언급하였던 것처럼 기존의 표지와 연계를 고려한다면 기존의 분류 체계를 따르는 것이 낫다. 그리고 특이한 방언형이 발견되어 지정사와 동음이의 관계를 형성할 가능성을 아예 배제할 수도 없다는 이유도 있다. 보조 용언의 분류도 기존의 것을 그대로 따랐다. 보조 용언을 동사, 형용사 등과 나란히 위치시켰다는 점은 이질적으로 보인다. 하지만 본용언과 보조 용언을 구분하는 열을 추가하는 것도 문제가 발생할 수 있다. 그렇다고 보조 용언을 없애는 것은 기구축 말뭉치와의 호환성에 심각한 문제가 생긴다. 보조 용언과 관련하여 추가적으로 고려할 수 있는 문제는 보조 동사와 보조 형용사를 구분하는 문제이다. 이는 기존의 방식을 따르는 것이 여러모로 이점이 있다. 보조 동사와 보조 형용사를 구분할 경우, (1c)에서 언급하였던 활용의 실질성, 구체적으로는 말뭉치 구축상의 효율성이 상당히 떨어지게 된다. 또한 ‘않-’과 같이 보조 동사로도 보조 형용사로도 분석될 수 있는 형태들을 구분해 놓는다면 말뭉치 활용의 측면에서도 비효율적일 수 있다. 이러한 단점은 ‘EC’ 하나를 두지 않고 이른바 대등적 연결 어미(병렬 어미), 종속적 연결 어미, 보조적 연결 어미, 부사형 어미 따위로 구분할 때도 발생한다. 예를 들어 예를 들어 ‘-고’ 별개의 것으로 일일이 구분하게 되면 활용의 실질성이 매우 떨어지게 된다. 보조적 연결 어미는 보조 용언(VX)를 뒀으로써 변별할

22) 기존의 용어 중에 ‘관계언’을 고려할 수도 있으나, 이는 종결 어미 등을 포함시키기에 적절하지 않아 배제하였다.

수 있다. 나머지의 구분은 기존대로 하나로 유지하는 것이 좋다. 다만, 그 중 하나의 용어를 선택해야 한다면 세종 지침의 용어처럼 ‘연결 어미’로 하는 것보다는 ‘부사형 어미’로 설정하는 것이 나을 것으로 보인다. 통사 구성을 중심으로 하는 개념 체계에서는 구문 구조 중심의 관점이 더 적합하기 때문이다. 이들 어미는 기본적으로 구조적 차이를 보이지 않으며 굳이 그 성분을 들자면 부사가 올 만한 자리를 점하기 때문에 부사형 어미라 칭하기로 한다. ‘전성 어미’의 층위는 생략하고 ‘부사형 어미, 명사형 어미, 관형사형 어미’라 칭하는 분류 체계를 따르기로 한다.<sup>23)</sup> 다만 말단의 표지는 기구축 말뭉치와의 호환성을 위해 ‘T’를 그대로 두기로 한다.

2단계는 통사론적 단어를 대상으로 그 내부 구조를 형태론적으로 분석해 내는 단계이다. 2단계의 설정은 논의가 더 필요하다.<sup>24)</sup> 수정을 할 경우엔 여러 사용 목적에 맞추어 매우 세밀한 형태론적 논의를 거쳐야 할 것이다. 본고는 통사론적 단어의 분석과 그 상위/하위의 구분에 초점을 두므로 통사론적 단어의 하위 분석은 기존 지침과의 연계성을 고려하는 차원에서 제시한다. 결국 위의 분석 표지에 의해 2단계까지 분석한 결과는, 즉 형태 분석 말뭉치의 모습은 기구축 말뭉치와 크게 다르지 않다. 태그는 세종 분석 표지와 동일하다.

단, \* 표시의 형태가 합성어와 파생어에서 어기로 쓰일 때는 별도의 방법을 강구할 필요가 있다. 만약 ‘나무꾼 나무’를 분석했을 때 ‘나무/NGG+꾼/XSN 나무/NGG’와 같이 분석된다면 조어에 관여하지 않은 ‘나무’와 조어에 관여한 어기로서의 ‘나무’가 구분되지 않으며, 결국 1단계와 2단계 구분의 효과를 잃게 될 것이다. 그렇다고 ‘XR’로 처리할 수도 없기 때문에 별도의 표시를 해 주는 것이 좋다. 그런데 한편으로 고려해야 하는 것은 기구축 말뭉치와의 호환성과 함께 1단계 분석과의 연계이다. 따라서 ‘나무/XNNG+꾼

23) 임동훈(2016: 275) 등에서 전성 어미의 층위를 따로 두지 않는 방식으로 어미 체계를 정리하고 있다.

24) \* 표시 부분은 이후 논의를 통해서 권장 안 차원으로 더 채워질 수도 있을 것이다. 본고에서는 기구축 말뭉치와의 연계를 고려하여 (18)과 같이 제시하는 선에서 논의를 마치기로 한다.

/XSN 나무/NNG'와 같이 처리할 것을 제안한다. 이는 1단계 분석의 표지 앞에 단어 형성을 의미하는 'X'를 공통으로 붙이는 방안이다. 그리하면 두 '나무'를 별개로 처리할 수도 있고 'X'를 일괄적으로 제거하여서 하나로 처리할 수도 있다. 또한 파생어뿐만 아니라 경우에 따라서 합성어를 분석할 방법도 생긴다는 이점이 있다. 물론 1단계 층위와 2단계 층위도 체계적으로 구별하여 표시해 준다는 장점도 생길 것이다.

다음으로 문제가 될 만한 것은 이른바 통사적 접사이다. 임홍빈·송철의 (1998)의 표준 안에서는 이른바 통사적 파생에 참여하는 통사적 접사(이하, '통사적 파생 접사')에 대하여 중요하게 다루었다.<sup>25)</sup> 그러나 이는 세종 분석 표지에도 반영이 되어 있지 않다. 세종 분석 표지를 중심으로 이루어져 온 현재 말뭉치에는 어차피 통사적 파생 접사와 형태적 파생 접사의 구분이 되어 있지 않다. 기구축 말뭉치와의 호환성을 고려한다는 측면에서도 통사적 파생 접사의 별도 설정은 부담이 된다.

이론적인 측면에서도 통사적 파생 파생 접사의 설정은 유보를 하는 편이 안정적이다. 통사적 파생 접사를 설정하면 “핵이동 설정에 의한 연산의 과도함, 통사적 어휘 접사의 의미역 배당의 부적절함, 생성이론의 임의적 수용, 어순재배치 구문 설명의 난점, 구절과 자매항인 요소에 대한 부당한 대우, 문장과 단어의 경계 붕괴, 조어론과의 판별 불가능성, 용어상의 이율 배반성 등”(이훈 2003: 121)의 논란마저 수용해야 한다. 최형용(2010: 85)에서는 “접미사는 통사적 구성에 참여하는 일이 없다.”라고 단언하고 있다.

(19) ㄱ. 아름답다 ; [재벌이 사는 집]답다

ㄴ. 軍用, 부인용 ; [청소년이나 성인]用 잠옷[이상, 최형용(2010: 86)에서 재인용]

25) 여기서의 통사적 접사는 일종의 협의의 개념으로 고창수(1992), 시정근(1998)의 광의의 용어를 말하는 것이 아니다. 고창수(1992), 시정근(1998)의 통사적 접사는 조사, 어미까지 모두 포괄한다. 혼동을 피하기 위하여 협의의 통사적 접사를 '통사적 파생 접사'로 기술하려 한다.

형태 분석에서 문제가 될 만한 사례는 위 (19)와 같은 두 가지 경우이다. 최형용(2010: 86)에서는 기존 논의를 예를 두 사례로 정리하여 제시하면서 문제점을 지적하였다. (19ㄱ)의 ‘-답다’ 따위는, 구 이상의 단위와 결합하는 경우가 존재하기는 하지만, “아름답다”처럼 조어에만 나타나는 어근과도 결합된다는 점에서 통사적 구성에 참여하는 요소로 볼 수 없다는 것이다. (19ㄴ)과 같은 사례는 한자어 접사이다. 의미 단위인 한자는 고유어처럼 파악할 수 없다고 보고 있다. 심지어는 “공용, 범용”과 같은 예를 보면 ‘-용(用)’이 어근이 되기도 한다.<sup>26)</sup> 따라서 1단계의 분석에서는 ‘-답-’이나 ‘-용’ 따위를 분석하지 않는다. 2단계 분석과 관련하여 논의하는 것이 효율적이다.

- (20) ㄱ. 몸도 튼튼, 마음도 튼튼.  
 ㄴ. 몸도 튼튼하고, 마음도 튼튼하다.

(20ㄱ)과 같은 예에 나타나는 “튼튼”도 (18)에 1단계 분석 대상으로 설정하지 않았다. 세종 분석 표지와 같이 ‘튼튼/XR’로 분석하더라도 문제가 해결되는 것이 아니다. 세종 분석 표지의 방식대로라면 (20ㄱ)의 ‘튼튼’과 (20ㄴ)의 ‘튼튼’이 동계의 것이 되고 만다. 조어론적 입장에서는 유의미한 통계를 제공할지 모르지만, 단어 이상의 관점에서는 여러 문제를 낳을 뿐이다. 특히 구문 분석의 관점에서 (20ㄱ)의 동사구를 ‘튼튼/XR’로 표시하고 만다면 이는 통사 규칙의 일반화를 어렵게 만들 뿐이다. 1단계 분석에서는 ‘튼튼하/VA’나 ‘튼튼/VA’와 같이 처리하고, 2단계 분석에서는 ‘튼튼/XR’과 같이 처리하는 방안을 제시해 볼 수 있다.<sup>27)</sup> 핵심은 1단계와 2단계의 층위 구분이며, 이를

26) 최형용(2010: 87)에서는 ‘[[do good]-er]’(영어)나 ‘[[vierde klass]-er]’(네덜란드어)처럼 범언어적으로도 매우 생산적인 접사에 한정적으로 나타나는 예외적 현상이 있음을 인용하기도 하였다.

27) 기존에 일부 지침에서는 원 어절의 특성을 고수하려는 경향도 있었다. 그러나 불규칙적인 어간-어미 결합이나 축약형 등을 고려할 때 형태 분석 결과에서 원 어절의 형태를 보존한다는 것은 한계가 분명하다는 점을 관찰할 수 있다. 원 어절의 형태와 분석 형태를 동시에 열람하는 것이 충분히 가능하기 때문에 굳이 원 어절의 형태 특성을 살리고자 지침을 강구할 필요는 없을 것으로 보인다.

통한 말뭉치 처리의 효율성 제고이다.

## 4. 결론

말뭉치 구축 지침은 개념이 정연해야 하고, 이해가 용이해야 한다. 그리고 적용상 유연해야 하며, 활용의 실질성도 갖추어야 한다. 본고의 문제 제기는, 21세기 세종계획 지침을 위시한 기존의 형태 분석 지침이 이러한 조건에 부족함이 있다는 것이었다. 구체적으로는 단어관이 분명하지 않다는 점, 그래서 단어 형성의 층위와 통사 단위 구성의 층위가 혼재되어 있다는 점, 분류 단계의 독립성도 떨어진다는 점 등을 거론하였다.

세종 형태 분석 지침은 개선을 추구해야 한다. 그 방안은 다음과 같았다. 첫째, 통사론적 단어의 개념을 분석 기준으로 삼는다. 아울러 어휘 범주와 기능 범주의 구분을 고려한다. 둘째, 통사론적 단어를 중심으로 상위의 분석 단계와 하위의 분석 단계를 두어 통사론적 층위와 형태론(조어론)적 층위를 구분한다. 셋째, 통사론적 층위(통사 단위 구성의 층위)를 기본 분석 단계로 삼고 형태론적 층위(단어 형성의 층위)는 부가적인 분석 단계로 삼아서, 즉 분석의 이원화를 취해서 말뭉치 구축 및 이용에 있어서 유연성과 실질성을 제고한다. 형태론적 층위를 부가적인 단계로 설정한 주요 이유는 조어의 성분 분석이 어려울뿐더러 효율성도 적다는 점이다. 한편, 말단의 태그는 되도록 기존의 것을 그대로 사용하여서 기구축 말뭉치와의 연계도 취한다. 본고는 이러한 논의의 결과를 (18)에서 통사론적 단어 중심의 분석 표지 안으로 제시하였다.

말뭉치 구축에서 형태 분석은 기본 절차로서 다른 분석 말뭉치의 밑 작업이 된다. 본고의 논의가 간혹 원리적 논의에 그친다고 볼 수도 있으나, 실제로 말뭉치를 분석·구축하는 입장에서나 구축된 말뭉치를 활용하는 입장에서나 말뭉치 구축의 과학적인 원리는 매우 중요하다. 말뭉치 구축의 이론적 원리를 납득하지 못한 작업자는 수많은 불일치를 양산해 낼 것이며, 말뭉치

구축의 배경을 쉽게 이해하지 못하는 말뭉치 이용자는 그 말뭉치의 가치를 결코 높게 평가하지 않을 것이다. 만약 말뭉치 구축의 현실 앞에 선다면, 본고의 주장은 일종의 제안이 될 것이다. 또 한편으로는 형태 분석 자체에 대한 논의로서 문법론적으로도 기여가 되기를 바란다. 형태 분석은 문법론적 설명의 시작이자 끝이기 때문이다.

본고는 분석의 세밀화보다는 효율적이고 원리적인 처리를 우선시하는 논의를 취했다. 말뭉치는 종류에 따라서는 다소 논란거리가 발생하더라도 되도록 분석을 자세히 하는 것이 나올 수도 있다. 본고는 그러한 측면까지는 고려하지 못하였다. 특히 단어 형성 층위에 대한 논의는 부가적인 차원에서 논의하는 것으로 그쳤다. 그리고 지침의 기술 문제 등 구체적인 부분에는 논의가 미치지 못하였다. 현재 형태 분석은, 이미 말뭉치 구축도 상당히 이루어져 있을뿐더러, 대개가 자연어 처리나 딥러닝 기술에 의해 만들어진 프로그램에 의해서 이루어지고 있다. 하지만 문법 규칙 기반보다는 사례 기반이다 보니, 다양한 하위 언어 형식 앞에서 그 한계를 분명히 드러내기도 한다. 문법론적인 논의와 규칙 기반으로 학습시키는 형태소 분석기의 개발이 더 늘어나야 할 것이다. 논의가 미치지 못한 부분은 이후의 과제이다.

## 참 고 문 헌

- 21세기 세종계획(1999), 『국어 기초자료 구축(학술용역 과제 보고서)』, 문화관광부.
- 21세기 세종계획(2007), 『국어 기초자료 구축(연구보고서)』, 국립국어원.
- 강승식(1993), 「음절 정보와 복수어 단위 정보를 이용한 한국어 형태소 분석」, 서울대 컴퓨터공학과 박사학위논문.
- 강승식(2003), 『한국어 형태소 분석과 정보 검색』, 홍릉과학출판사.
- 고영근(1992/1993), 「형태소란 도대체 무엇인가?」, 이병근·채완·김창섭 편, 『형태(국어학강좌 3)』, 태학사, 11-23면.
- 고영근(1995), 『단어·문장·텍스트』, 한국문화사.
- 고영근·구분관(2018), 『우리말 문법론(개정판)』, 집문당.
- 고창수(1992), 「국어의 통사적 어형성」, 『국어학』 22, 국어학회, 259-269면.
- 김용하·박소영·이정훈·최기용(2018), 『한국어 생성 통사론』, 역락.
- 김일환(2013), 『‘물결 21’ 코퍼스의 구축과 활용』, 소명출판.
- 김창섭(1984), 「형용사 파생 접미사들의 기능과 의미: ‘-답-, -스럽-, -롭-, 하-’와 ‘-적’의 경우」, 『진단학보』 58, 진단학회, 145-161면.
- 박진호(1994), 「통사적 결합 관계와 논항구조」, 『국어연구』 123, 국어연구회.
- 송원용(2016), 「언어학 개론」, 강현화 외, 『한국어 교원을 위한 한국어학』, 한국방송통신대학교출판문화원, 1-26면.
- 서태룡(1988), 『국어 활용어미의 형태와 의미』, 탑출판사.
- 시정관(1998), 『국어의 단어형성 원리(수정판)』, 한국문화사.
- 유현경 외(2018), 『한국어 표준 문법』, 집문당.
- 유혜원(2004), 『한국어 정보 처리의 이론과 실제』, 제이앤씨.
- 이기황(2016), 「언어 자료의 보고, 빅데이터」, 『새국어생활』 26-2, 국립국어원, 31-55면.
- 이선웅(2012), 『한국어 문법론의 개념어 연구』, 월인.
- 이진병(2017), 「형태 주석용 표준 지침의 방향성과 세부 쟁점」, 『건지인문학』

- 18, 전북대학교 인문학연구소, 217-242면.
- 이훈(2003), 「통사적 접사 설정에 대한 비판적 검토」, 『어문연구』 31-3, 한국  
어문교육연구회, 121-145면.
- 임동훈(2016), 「한국어의 어미」, 강현화 외, 『한국어 교원을 위한 한국어학』,  
한국방송통신대학교출판문화원, 271-308면.
- 임홍빈(1989), 「통사적 파생에 대하여」, 『어학연구』 25-1, 서울대 어학연구소,  
167-196면.
- 임홍빈(1997), 「국어 굴절의 원리적 성격과 재구조화」, 『관악어문연구』 22,  
서울대 국어국문학과, 17-98면.
- 임홍빈 · 송철의(1998), 「한국어 정보 처리를 위한 어절 분석 표지의 표준화  
연구」, 21세기 세종계획(1998), 『국어 기초자료 구축(학술용역 과제  
보고서)』, 문화관광부, 191-295면.
- 임홍빈 · 이홍식 외(2002), 『한국어 구문 분석 방법론』, 한국문화사.
- 조성식 외 편(1990), 『영어학사전』, 신아사.
- 최현배(1937/1971), 『우리말본』, 정음사.
- 최형용(2002), 「국어 단어의 형태·통사론적 연구: 통사적 결합어를 중심으로」,  
서울대 박사학위논문.
- 최형용(2010), 「품사의 경계: 조사, 어미, 어근, 접사를 중심으로」, 『한국어학』  
47, 한국어학회, 61-92면.
- 한정환(2011), 「통사 단위 단어」, 유현경 외, 『한국어 통사론의 현상과 이론』,  
태학사, 11-49면.
- Bloomfield, L.(1933), *Language*, George Allen & Unwin.
- Hockett, C. F.(1954/1957), Two models of grammatical description」, Joos,  
M. ed.(1957), *Readings in Linguistics I*, The University of  
Chicago Press.
- Sciollo, A.-M. D. & Williams, E.(1987), *On the definition of word*  
(*Linguistic Inquiry Monographs 14*), Cambridge, Massachusetts:  
MIT Press, 10-118.

【Abstracts】

## About Morphological Analysis Guideline Based on Syntactic Words

— Focusing on the Amendment of Sejong Guidelines —

Yi, Jinbyeong

The issue raised in this paper is that existing guidelines for morphological analysis are lacking in terms of conceptual organization and utility. The problem is as follows. A mixture of word-formation theory and syntax, Less classification independence, etc. Morphological analysis, as a base work and basic process, is then linked to the morph-sense tagged corpus, syntactic corpus, and other discourse corpus. So that's important. The Sejong Guidelines can be improved as follows. First, the concept of syntactic words is used as an analysis criterion. In addition, consider the division of lexical and functional categories. Second, Procedurally separates the analysis above and below the syntactic word. Thus it distinguishes between syntactic and morphological levels. Third, the syntactic level is used as the basic analysis step, and the morphological level is used as the additional analysis step. Thus, the analysis is dualized to increase flexibility and substantiveness in corpus construction and use. The main reasons for setting the morphological level as additional steps are as follows. Difficulties in component analysis, lack of utility. On the other hand, the terminal tag can be used as it is. Doing so also allows for linkage with previously established corpus. This article presents the results of this discussion into the word analysis marker.

**Key words** : morphological analysis guideline, analysis maker, syntactic word, grammatical levels, Syntax levels, word formation levels, functional category, lexical category

이 논문은 2020년 02월 16일에 투고되었으며, 2020년 03월 17일에 심사 완료되어 2020년 03월 18일에 게재가 확정되었음.