

# 형태소 분석 말뭉치의 파생명사 처리

차준경<sup>1</sup> · 강범모<sup>2</sup>  
고려대학교 국어국문학과<sup>1</sup> · 언어과학과<sup>2</sup>  
{jhall, bmkang}@korea.ac.kr

## Derivational Morphology in a Tagged Corpus

Joon-Kyung Cha<sup>1</sup> Beom-mo Kang<sup>2</sup>  
Dept. of Korean Linguistics<sup>1</sup> Dept. of Linguistics,  
Korea University

### 요 약

이 논문은 형태소 분석 말뭉치를 구축하면서 제기되었던 파생 명사 처리의 문제점을 논의하고 그 해결 방안을 모색한 것이다. 파생 명사의 분석에서 국어학적·전산 언어학적으로 유의미한 분석 범위를 정할 때 몇가지 고려해야 할 사항이 있다. 접두사는 어기가 불규칙하므로 규칙으로 자동처리가 어렵다. 형태소 분석의 대상은 생산성이 높고, 어기와 범주를 변화시키는 서술성 접두사로 그 범위를 정할 수 있을 것이다. 접미사의 분석은 생산성이 높고 규칙적인 굴절 접미사가 분석의 대상이 되며, 또한 서술성을 갖는 한자어계 접미사도 분석 대상이 된다. 파생 명사의 분석에 있어서 접사는 그 위상이 동요되므로 접두사는 관형사와 구별이 어렵고, 접미사는 의존명사와 구별이 어렵다. 그러므로, 대용량의 형태소 분석 말뭉치를 효율적으로 구축하기 위해서는 접사에 대한 다각적인 검토가 필요할 것이다.

### 1. 서론

현재, 분석표지(tag) 부착 말뭉치는 여러 기관<sup>1)</sup>에서 구축되어 왔으며, 분석 표지(tag set)나 표지 부착 방식에 대한 많은 연구<sup>2)</sup>가 있었다. 분석 표지 부착 단위가 ‘품사’이면 ‘품사 표지 부착 말뭉치’로 부르며, ‘형태소’를 단위로 하면 ‘형태소 표지 부착 말뭉치’로 부른다.

말뭉치에 분석 표지의 부착은 일반적으로 형태소 분석을 한 뒤, 가능한 분석 가운데서 정확한 분석을 찾는 과정이다. 이때, 형태소 분석에 대하여 언어학과 전산학에서의 약간의 개념의 차이가 있는 것으로 보인다.

일반적으로 언어학의 관점에서, ‘형태소 분석’은 ‘의미를 가진 최소의 단위’인 형태소 단위로의 분석을 말하며, 이를 통하여 단어의 직접 구성 성분(Immediate Constituent) 분석, 단어 형성 규칙 등을 연구한다.

자연어 처리 관점에서는 ‘형태소 분석’을 단어(어절) 단위로 분리된 입력 문자열로부터 각각의 형태소를 분리하고, 용언의 불규칙 활용이나 굴절 현상이 일어난 단어에 대해서는 원형을 복원하는 과정(김영택 1994: 54)으로 파악하고 있다.

이러한 자연어 처리의 ‘형태소 분석’을 국어학의 관점에서 살펴보면, 입력 어절에 대하여 조사와 어미를 분리하고 난 뒤, 용언에서 어미를 분리하는 과정은 ‘형태소 단위의 분석’이라고 할 수 있으며, 체언과 조사를 분리하는 과정은 ‘품사(어휘) 단위의 분석’이라고 할 수 있을 것이다.

‘세종 형태소 분석 말뭉치’에서는 체언을 접두사와 접미사로 분리하고 어기(어근) 범주를 도입하였다. 이러한 특징으로, 국어학적 개념을 적용한 형태소 분석 말뭉치라고 할 수 있다. 이 형태소 분석 말뭉치를 통해 단순히 품사 정보뿐만 아니라 국어의 단어 형성 원리의 기초를 제공해 줄 수 있을 것이다.

1) 대표적으로 KAIST 구축 말뭉치, 1999년의 ETRI의 ‘품사 부착 말뭉치’, 1999년 21세기 세종 계획의 ‘형태소 분석 말뭉치’ 등이 있다.

2) 자세한 연구사는 임흥빈외(1998)을 참조.

이 논문에서는 형태소 분석 말뭉치를 구축하면서 제기되었던 문제점 중 특히 파생 명사의 분석에 대하여 제기되었던 문제점을 살피고 이에 대한 해결 방안을 모색하려고 한다.

## 2. 파생어

국어의 단어에는 그 단어 구조상 파생어와 합성어, 접어 등으로 분류된다. 논의의 대상이 되는 파생 명사는 ‘접두사+어기’, ‘어기+접미사’, ‘접두사 + 어기 + 접미사’의 구조를 갖는 단어들이다.

자연어 처리의 형태소 분석의 과정에서는 파생 명사 중 일부 접두사 ‘제(第)’와 생산성이 높은 접미사 ‘적(的)’과 여러 굴절적 접미사들을 분석해 왔다. 세종 형태소 분석 말뭉치에서는 국어학적으로 유의미한 결과를 얻기 위해서 체언에서 접두사와 접미사를 분리하였고 이들 접사의 범위를 기존의 분석보다 확장시켰다. 이와 같은 과정에서 형태소 분석의 대상이 되는 접사의 범위에 대하여 고찰이 필요하다.

국어학의 관점에서 파생어의 형태소 분석은 해당하는 모든 접사, 어기 등을 세세히 분석해야 하나, 현실적으로 형태소 분석 말뭉치를 구축하는 과정에서 모든 접사를 낱낱이 분석하는 데는 어려움이 있다<sup>3)</sup>. 국어의 접사는 그 위상이 동요하고 있어서 의존 형식과 접미사 사이에 문법화와 비문법화 현상이 일어나며, 굴절법과 파생법의 동요 현상도 관찰된다. 접미사가 어근과의 유연성을 잃고 어근에 융합되는 것도 있다(고영근 1989).

자연어 처리 관점에서, 파생어 분석에 관심을 두는 이유는 신조어나 사전의 미등록어들이 출현하였을 때, 파생어 분석을 통해서 일정한 규칙으로 해결하기 위한 것이다. 생산성이 높은 접사들은 신조어를 많이 만들어 낼 수 있는 가능성이 높기 때문이다. 자연어 처리의 관점에서는 신조어 형성에 참여하는 접사, 일정한 규칙이 있어서 새로운 어절을 형성할 수 있는 굴절적 접미사들만을 분석하는 이유가 여기에 있다.

이상과 같이 세종 형태소 분석 말뭉치에서 접사의 처리는 기존의 형태소 분석에서 포함되었던 접사와 더불어 국어학적으로 유의미한 접사들을 다수 포함시켰다. 다음 절에서 실제 분석의 예를 통해 파생 명사의 분석

을 살펴보기로 한다.

## 3. 접두 파생어의 분석

국어의 접두사는 어원적으로 고유어 접두사와 한자어 접두사가 있다. 고유어계 접두사는 어기의 범주를 바꾸지 않는 어휘적 접두사이다. 한자어 접두사에는 어휘적 접두사와 어기의 범주를 바꾸는 서술성 접두사가 있다. 또한, 결합하는 어기의 범주에 따라 체언 접두사, 용언 접두사로 분류하기도 한다.

기존의 형태 분석 말뭉치 -ETRI 품사 태그 부착 말뭉치(임해창외 1999)-에서는 수사 접두사 ‘제(第)’만을 분석했다. 고유어 접두사는 결합하는 신조어의 수가 제한되고 대부분 국어 사전에 수록되어 있으므로 이것을 규칙으로 분석하기가 어려웠기 때문이다. 그러나 한자어계 접두사 중 서술성 접두사는 어기의 범주를 바꾸며, 또한 신조어를 형성하기도 한다. 이런 현상을 포착하기 위해, 세종 형태소 분석 말뭉치에서는 다수의 한자어계 접두사를 분석의 대상으로 삼았다. 다음의 세종 형태소 분석 말뭉치의 접두사 분석의 예를 살펴보자.

### (1) 세종 형태소 분석 말뭉치의 접두사 목록

제(第) 815, 불(不) 657, 대(對/大) 571, 무(無) 556, 비(非) 443, 재(再) 354, 반(反) 242, 신(新) 168, 총(總) 128, 부(不) 121, 전(全) 114, 초(超) 109, 한 107, 소(小) 77, 고(高) 71, 피(被) 58, 헛 56, 최(最) 50, 저(低) 37, 생(生) 33, 미(未) 25, 구(舊) 25, 가(假) 24, 맨 22, 주(主) 20, 정(正) 20, 탈(脫) 17, 준(準) 16, 양(洋) 15, 말 15, 친(親) 14, 노(老) 9, 날 7, 꾀 6, 범(汎) 6, 과(過) 5, 왕(王) 4, 첫 3, 숫 3

위에 제시한 것 중 빈도수 상위의 예들은 대부분 한자어계 접두사들이며, 이들 ‘불(不), 대(對/大), 무(無), 비(非), 반(反)’ 등은 서술성 접두사이다. 즉, 이들은 서술성이 있어서 어기의 범주를 바꾸는 것들이다. 특히 ‘대(對)’와 ‘탈(脫)’은 신어 형성에 활발히 참여하고 있다(예: ‘대러시아, 대일본, 대미, 탈소, 탈북’ 등).

또, ‘재(再)’는 1994년의 국립 국어 연구원의 신어 조사 연구에서도 신어 형성에 많이 쓰인다고 보고되었다. 그러나 이들은 서술성 접두사로서 일정한 생산성을 가지고 신조어를 형성하지만 서술성 접두사임에도 결합되는 어휘에 일정한 제약이 있기도 하다.

빈도수 하위의 접두사들은 대부분 고유어 접두사들이고 노명희(1998)<sup>4)</sup>에서 접두사로 분류한 것들(‘생(生), 왕

3) 21세기 세종 계획의 형태소 분석 말뭉치(김홍규 강범모 2000)에서 분석 대상의 접사는 40종의 체언 접두사, 65종의 명사파생접미사, 9종의 형용사 파생접미사, 4종의 동사파생 접미사이다. 국립국어 연구원의 표준국어대사전에 수록된 접사의 수는 약 635종이다.

(王), 양(洋), 친(親), 외(外)' 등 -)이다. 이들은 대부분 어휘화되어 사전에 한 단어로 등재되어 있기 때문이다. 우리의 분석이 어휘 분석이 아니라 형태소 분석이라는 것을 고려하면, 고유어계 접두사와 결합한 파생어도 분석의 대상이 될 수 있으나 이들은 어휘화된 것으로 간주하면, 그 분석의 필요성이 낮을 것으로 본다.

접두사가 관형사와 동형어인 경우에는 접두사와 관형사를 구분하는 작업은 용이하지 않다. 특히, 한자어계 접두사는 관형사와 접사와의 구별뿐만 아니라 어근과 접사의 구별도 쉽지 않으며, 이것은 띄어쓰기의 문제와 관련이 된다.

예를 들어, '신도시'는 '접두사 + 명사'로 분석이 되지만 '신 국제 공항'은 '관형사 + 명사'로 분석이 된다. '신 국제공항'으로 붙여쓴 경우는 '신'이 접두사로 분석이 되기도 한다.

위 목록에는 중의적인 접두사들도 분석의 대상에 포함되었다 (예: '대(對/大) 571'). 그러므로, 정확한 접두사의 빈도수를 측정할 수 없게 된다. 이러한 예는 접미사의 목록에서도 찾아볼 수 있다. 이들은 형태 중의성과 더불어 의미적 중의성을 갖고 있으므로, 이를 자동 태거에 적용했을 때 다중분석을 유발하는 요인이 될 수도 있다. 그러므로 분석 대상의 접두사를 선정할 때 신중한 판단을 기해야 할 것이다.

이상과 같이 파생어에서 접두사 분석은 고유어보다는 한자어 계열 접두사가 신조어를 많이 형성하며, 어휘적 접두사보다는 서술성 접두사가 신조어 형성에 많이 참여하는 것을 알 수 있었다. 이들은 신문 등에서 외국어의 번역에 의한 신조어에 많이 사용되는 것으로 보인다. 그러므로, 신조어 분석에 있어서는 이러한 접두사들을 포함하는 파생어를 고려해야 할 것이다.

#### 4. 접두 파생어의 분석

명사와 결합하는 접미사에는 품사를 파생하지 않는 굴절적인 접미사와 어휘적인 접미사, 명사를 파생하는 접미사로 분류할 수 있다. 굴절 접미사와 품사 파생 접

4) 노명희(1998)에 의하면 다음과 같이 접두성 한자어에 대하여 접두사와 접사성 어근, 어근으로 분류하였다.

- ㄱ. 접두사: 생(生), 왕(王), 양(洋), 친(親), 외(外), 시(媿), 초(超), 범(汎), 최(叢), 준(準)
- ㄴ. 접사성 어근: 고(高), 저(低), 급(急), 무(無), 대(對), 탈(脫), 미(未), 몰(沒), 부(不), 주(駐) 재(在) 맹(猛) 재(再) 항(抗) 피(被) 가(假) 건(乾) 목(木)
- ㄷ. 서술성 접사성 어근: 무(無), 대(對), 탈(脫), 미(未), 몰(沒), 부(不)

미사의 대부분은 고유어 접미사이고, 어휘적인 접미사는 대부분 한자어 접미사이다.

#### 4.1 고유어계 접미사의 분석

세종 형태소 분석 말뭉치에서의 명사 파생 접미사의 분석의 예를 살펴 보자.

##### (2) 세종 형태소 분석 말뭉치의 고유어 접미사

들 22188 + 덜 24, 남 1170, 씩 692, 째 514, 쯤 569, 네 405, 끈 200, 짜리 182, 끼리 156, 질 106, 군 84, 깨 79, 치 31, 쟁이 26, 장이 24, 투성이 21, 직 13, 분 11, 짓 11, 쪽 9, 꿀 9, 배기 8, 뽕 8, 이 6, 깨 5

고유어 접미사의 빈도수 결과를 보면, 빈도수 상위의 접미사들은 대부분 굴절 접미사들로 이들은 규칙적으로 어기와 결합한다. 그러므로 형태소 분석 말뭉치에서는 이들 굴절 접미사들은 일정한 규칙으로 처리할 수 있는 것이다. 그러나 빈도수가 낮은 접미사와 결합된 파생 명사의 분석은 규칙으로 처리한다면 오분석을 양산하게 될 것이므로 최소한으로 제한해야 할 것이다.

또한 국어 형태론에서 대표적인 명사 파생접미사로 연구하는 '-이', '-음', '-기'도 이 목록에서 제외되어 있다. 이들은 중세 국어 이전부터 명사를 파생하는 접미사의 역할을 했기 때문에 이들 파생어는 언중들의 인식에서 단일어와 구별이 되지 않기도 하다(예: 잠, 꿈, 구이 등).

또, '-음'과 '-기'는 명사형 어미와도 형태가 동일하므로, '-음'과 '-기'가 결합한 어휘에 대해서 파생어인지 동사의 명사형인지에 대한 구별을 해주어야 한다. 그러므로 이미 사전에 한 단어로 등재되어 있고, 어휘화가 많이 진전되었다면 굳이 형태소 분석을 할 필요가 없을 것이다.

#### 4.2 한자어계 접미사의 분석

한자어 접미사는 고유어 접미사와 달리 통사적 기능을 바꾸는 접미사가 소수이다. 또한 접사성 한자어에서 접미사, 어근, 의존명사를 구별하는 것도 어렵다. 이들은 그 위상이 끊임없이 변화하므로 중간 단계에 있는 예들도 많기 때문이다. 세종 형태소 분석 말뭉치의 한자어계 접미사의 예를 들어 이를 살펴 보기로 한다.

##### (3) 세종 형태소 분석 말뭉치의 한자어 접미사

적(的) 20783, 성(性) 3117, 화(化) 2445, 여(餘) 1054, 론(論) 815, 간(間) 769, 상(上) 737, 용(用) 484, 측(側) 425, 제(制) 323, 별(別) 287, 식(式) 276, 계(係) 266, 형(型) 211, 시(視) 203, 권(權) 199, 당(當) 169, 층(層) 145, 대(臺) 120, 류(類) 118, 생(生) 107, 경(傾) 96, 설(說) 80, 율(律) 72, 면(面) 65, 차(次) 40, 하(下) 40, 산(産) 39, 륜(率) 39, 급(級) 29, 편(便) 23, 가(哥) 25, 발(發) 14, 행(行) 13, 덕(宅) 11, 내(內) 8

위의 한자어 명사 파생접미사의 빈도수를 보면, 이 목록에서는 국어학계에서 한자어 접미사로 인정하고 있는 ‘적(的)’, ‘성(性)’과 더불어 의존명사로 간주되는 ‘측(側)’, ‘식(式)’ 등이 포함되어 있다.

일단 의존명사는 관형절의 수식을 받아서 자립하여 쓰일 수 있으므로 이런 예들은 제외되어야 할 것이다. 그러나 ‘측(側)’의 경우 ‘남측’, ‘북측’과 같이 선행 명사에 붙여 쓰는 예가 많다. ‘측(側)’의 빈도수 425도 모두 한어절에서 붙여 쓴 경우이다. ‘측(側)’이 의존명사로 분석된 것은 85회밖에 없으므로 의존명사와 접미사가 공존하고 있는 셈이다. 이것은 접미사의 위상이 동요하기 때문으로 보인다.

노명희(1998)에서 접사성 한자어에서 의존명사로 분류한 것 중 ‘측(側), 간(間), 하(下), 차(次), 초(初), 말(末)’ 등은 세종 형태소 분석 말뭉치에서 선행어절과 결합하여 쓰이는 빈도가 높았다. 이들을 띄어쓰기 처리를 하여 모두 의존명사로 처리하는 방법도 있으나 언중들이 이들을 한 어절로 붙여쓰고 있다는 국어의 현실을 존중하여 접미사로 간주한 것이다.

특히 ‘차(次)’는 ‘연구차(研究次)’, ‘방문차(訪問次)’에서는 접미사이지만 ‘연구하던 차에’에서는 의존명사로 분석이 된다. 이와 같이 접미사와 의존명사가 공존하는 예에서는 관형사와 접두사 판별과 같은 방법으로 의미적인 구별과 더불어 자연어처리에서는 띄어쓰기에 따른 한어절에서의 결합양상을 살필 수밖에 없다.

위의 목록에서 빈도수가 가장 높은 ‘-적’, ‘-성’, ‘-화’ 등은 노명희(1998)에서는 ‘원 의미와 유연성을 상실하고, 어기의 범주를 바꾸는 것’이다. ‘-적(的)’은 중국어에서 ‘-tic’을 번역한 것으로, 한국어에는 중국어의 차용어와 더불어 개화기 시기의 일본어의 차용으로 여겨진다.

‘-적(的)’은 어기가 한자어뿐만 아니라 외래어와도 결합이 잘 일어나고 있으나 고유어와는 그렇지 못하다(예: ?마음적 vs 심적, 쉼트적). 국어의 ‘-적’ 파생어는 그 품사가 기능에 따라 명사, 관형사, 부사로 나누어진다.

‘-성(性)’은 영어의 ‘-ity’에 해당하는 접미사로 어기의

성질이나 기질 특성을 나타낸다. ‘-화(化)’는 중국어에서 영어의 ‘-ize’에 해당하는 접미사로 취급되며 국어에서도 이와 비슷한 기능을 한다. 서술성이 없는 명사 어기에 결합하여 명사에 서술성을 부여하는 접미사이다.

이와 같이 빈도수가 높은 ‘-적’, ‘-성’, ‘-화’ 등은 어기의 품사를 바꾸는 기능을 하고 신조어를 만들 가능성도 높기 때문에 형태소 분석의 대상이 된다. 그러나 이들 파생어를 단순히 규칙으로 처리하기에는 몇 가지 어려움이 있다.

특히 1음절 한자어 어근에 접미한 예들이다. 한자어가 차용된 시기에 따라, 어기와 접사의 융합이 1음절 한자어를 하나의 단어로 여기는 시대와 그렇지 않은 시대에 따라 그 양상이 달라진다. 그 예로 다음의 단어를 형태소 분석하면 어기에 해당하는 것에 한자어 어근이라는 표시를 해주어야 한다.

(4) 광적(狂的) 내적(內的) 단적(端的) 사적(私的) 심적(心的) 외적(外的) 지적(知的) 전적(全的) 동적(動的) 정적(靜的)--

이들은 이미 경음화도 일어나고 있으므로 한 단어로 굳어진 것으로 간주하여 형태소 분석에서 제외할 수 있다. 이것은 ‘-성(性)’, ‘-화(化)’나 기타 한자어 접미사에도 해당한다. 공식적으로 1음절 한자어가 명사로 자립하여 쓰일 수 있다면 그것은 형태소 분석을 하고, 자립할 수 없다면 분석을 하지 않는 방법도 있을 것이다. 세종 형태소 분석 말뭉치에서는 1음절 비자립성 한자어 어근과 결합한 파생어들은 단어로 간주하였다.

또 사전의 등재어와 신조어에 대한 구별이 문제점으로 대두되었다. 다음의 접미사 ‘용(用)’이 결합된 예들을 살펴보자

(5) 자가용, 나들이용, 사무용, 어린이용, 성인용, 아버지용--

(5)의 ‘자가용(自家用)’은 의미적으로 어휘화되어 형태소 분석하지 않지만, ‘나들이용’에서는 형태소 분석이 가능하다. ‘자가용(自家用)’, ‘사무용(事務用)’ 등은 사전에 등재되어 있으나 ‘어린이용’, ‘성인용’, ‘아버지용’ 등의 인성 명사와 결합한 것은 사전에 등재되지 않은 어휘로, 형태소 분석의 대상이 된다. 이러한 구별은 분석자가 직관에 의해서 판단할 수밖에 없을 것이다.

즉, 분석 대상이 되는 접사의 목록을 선정한 다음에도 이 접사에 의해 형성된 단어의 어휘화 정도에 따라 분석이 달라질 수 있다는 것이다. 분석의 일관성을 유지하

기 위해 해당하는 모든 파생어를 분석하거나 사전에 수록된 어휘들을 찾아서 어휘화 정도를 파악하여 분석하는 방법이 있다. 세종 형태소 분석 말뭉치에서는 의미적으로 어휘화된 단어는 단어로 취급하였고, 위의 파생어들은 형태소 분석을 하였다.

또 다른 문제점은 세종 형태소 분석 말뭉치에서 분석하는 접사의 수가 많아서 약간의 과잉 분석을 하기도 하였다.

(6) ㄱ. 무신론자들

무/XPN+신/NNG+론/XSN+자/XSN+들/XSN

ㄴ. 개혁화주의론자들

개혁/NNG+화/XSN+주의/NNG+론/XSN+자/XSN+들/XSN

이와 같은 자세한 파생 명사의 분석은 국어학 특히 형태론 연구에는 매우 도움이 될 수 있을 것이나, 일반적인 어휘 연구나 또 다른 목적에서는 불필요한 것이 될 수도 있다. 여기에서 우리는 분석의 대상을 적절히 조절하는 것이 필요하다는 것을 알 수 있다. 세종 형태소 분석 말뭉치에서 국어학적인 관점을 도입하여 세밀한 분석을 시도한 것은 매우 뜻깊은 일이다. 그러나 이러한 세밀한 분석이 대량의 말뭉치를 대상으로 하는 자동 태거에 적용된다면, 분석의 오류를 양산할 수도 있을 것이다.

5. 결론

지금까지 형태소 분석 말뭉치를 구축할 때의 파생 명사 처리의 문제점을 고찰해 보았다. 대량의 말뭉치에서 형태소 분석을 효율적으로 하려면, 국어의 접사의 기능에 대해 다각적인 검토가 필요하다는 것을 강조하였다.

첫째, 체언 접두사는 여기서 불규칙하고 어휘적 개념을 더해주는 기능을 하므로, 이들을 규칙에 의해서 자동으로 처리하기는 어려울 것이다. 그러므로 형태소 분석의 대상을 제한하는 것이 필요하며, 어기의 범주를 변화시키는 접두사로 그 대상을 제한해야 한다.

둘째, 명사 파생 접미사는 고유어 접미사, 한자어 접미사, 어휘적 접미사, 굴절적 접미사로 그 어원과 기능에 따라 분류할 수 있다. 여기서 문제가 되는 것은 고유어 어휘적 접미사이다. 어기의 범주를 바꾸지만 이미 어휘화되어 한 단어로 인식하므로 이것을 분석하는 데 어려움이 있다.

고유어인 굴절성 접미사는 여기에 규칙적으로 결합하고 문법적인 기능을 하므로 형태소 분석의 대상이 된다.

한자어 어휘적 접미사는 고유어 어휘적 접미사와 같은 이유로 분석의 대상에서 제외된다. 그러나 어기의 범주를 바꾸는 서술성 접미사는 분석의 대상이 된다. 생산성도 높고 신조어도 이들 접미사에 의해서 형성되기 때문이다.

셋째, 대량의 형태소 분석 말뭉치를 구축하려면 접사 분석의 범위를 제한하는 것이 필요할 것이다. 세밀한 접사의 분석을 통해 한자어계열의 신조어는 어느 정도 처리가 가능할 것이나, 외래어계열의 신조어는 분석 불가능에 대한 분석 오류를 양산하기 때문이다. 이와 같은 문제점에 대하여 파생 명사의 분석의 범위를 요약하면, 서술성 접두사와 굴절적 접미사, 서술성 접미사가 된다. 일부 명사와 결합하여 쓰이는 관형사도 접두사로 간주될 수 있고, 의존 명사 등도 접미사에 포함된다. 이들에 대해서는 다소 국어학적 고찰이 필요할 것이다.

6. 참고문헌

[1] 고영근 (1989) 「국어 형태론 연구」 서울: 서울대 출판부.  
 [2] 강범모·차준경 (1996) “코퍼스, 파생어, 생산성” 「인지과학」 7-2, 21-37.  
 [3] 국립국어연구원 (1994) 「신어조사연구」 보고서.  
 [4] 김영택 (1994) 「자연 언어 처리」 서울: 교학사.  
 [5] 김홍규외 (1999) “현대 국어 기초 자료 말뭉치 및 형태소 분석 말뭉치 개발”, 「21세기 세종 계획 국어 기초 자료 구축」 보고서, 문화관광부.  
 [6] 김홍규·강범모 (2000) 「한국어 형태소 및 단어 사용 빈도의 분석 1」, 서울: 고려대 민족문화연구원.  
 [7] 노명희 (1998) 「현대 국어 한자어의 단어 구조 연구」 서울대 박사학위 논문.  
 [8] 서상규·한영균 (1999) 「국어 정보학 입문」 서울: 태학사.  
 [9] 시정곤 (1998) 「국어의 단어 형성 원리」 서울: 한국문화사.  
 [10] 송철의 (1992) 「국어의 파생어 형성 연구」 서울: 태학사.  
 [11] 임해창외 (1999) “품사 태그 부착 말뭉치 구축 지침서”, 「자연어 정보 처리 기술 표준화」 보고서, 한국전자통신연구원.  
 [12] 임홍빈외 (1998) “한국어 정보처리를 위한 어절 분석 표지의 표준화 연구”, 「21세기 세종계획 국어 기초 자료 구축」 보고서, 문화관광부.