

# 자연언어 처리를 위한 한국어 품사 태그

## A Korean Part-of-Speech Tag Set for Natural Language Processing

(Version 1.0)

김재훈, 서정연  
대전시 유성구 구성동 한국과학기술원  
전산학과, 인공지능연구센터

전화: 82-42-869-5551, 3534

팩스: 82-42-869-3510

전자우편: {jhoon | seo}@csking.kaist.ac.kr

1994년 6월 30일

### 요 약

규칙을 기반으로 한 자연언어 처리는 여러 분야에서 그 한계를 보이고 있다. 최근에 와서 이와 같은 문제점을 해결하기 위한 한 분야로서 확률 및 통계를 이용한다. 확률과 통계를 이용하기 위해서는 일반적으로 대량의 말뭉치가 필요하게 된다.

본 연구는 대량의 말뭉치를 이용한 자연언어 처리 연구의 가장 기초적인 단계로서 단어의 품사 태그(tag)를 결정하는 것이다. 품사 태그는 언어학자들이 소위 말하는 문법 범주와 비슷한 개념으로 사용된다. 그러나, 정확한 확률 및 통계 정보를 추출하기 위해서 사용형태와 기능에 따라 부가적으로 좀더 세분되었다. 한국어 품사 분류(품사 태그 결정)에 관한 연구는 언어학자에 의해서는 물론이고, 전산학자에 의해서도 이미 많은 연구가 되었다. 그러나, 이들 연구들은 일반적으로 언어학적인 성질과 특정 시스템에 적합하도록 분류되었다. 본 연구는 전산언어학적인 측면에서 모호한 부분을 분명하게 하여 품사 분류를 위한 어떤 지침서를 제안하고자 한다. 또한, 본 연구는 가능한한 정확한 확률 정보를 추출하기 위해서 품사를 세분류하고 이들을 계층적으로 재구성하여 기존의 여러 자연언어 처리 시스템에 자연스럽게 이용할 수 있도록 하였다. 전체 분류된 품사 태그의 수는 공백(space)을 포함해서 53개이다.

---

이 연구는 한국과학기술원 전산학과에서 1993년 봄학기에 개설한 “자연언어 처리 특강 - 문서처리 - (CS772, 담당교수: 최기선, 한영균)” 과목의 term project로 수행된 연구 결과를 수정·보완한 것이다. 이 연구의 초안을 검토해 주신 김성용, 김창현, 이상호에게 깊은 감사를 드리며, 그 밖에도 많은 토의에 함께 참여해 주신 많은 동료들에게도 감사드립니다.

이 연구의 결과는 앞으로도 수정·보완되어야 할 많은 여지를 가지고 있다고 생각합니다. 독자 여러분들은 조언 및 문의 사항이 있을 시에는 저자에게로 연락주시기 바랍니다.

# 목 차

1	소개	4
2	한국어의 품사 태그	5
2.1	기호(s)	5
2.1.1	s,(comma, 쉼표)	5
2.1.2	s.(sentence closer, 문장의 종결)	6
2.1.3	s' (left quotation and parenthesis mark, 여는 따옴표 및 묶음표)	6
2.1.4	s'(right quotation and parenthesis mark, 닫는 따옴표 및 묶음표)	6
2.1.5	s-(connection mark, 이음표)	7
2.1.6	su(unit, 단위)	7
2.1.7	sw(currency, 화폐 단위)	7
2.1.8	sp(space, 공백)	7
2.1.9	sy(other symbols, 기타 여러 기호들)	7
2.2	외국어(f)	7
2.2.1	f(foreign word, 외국어)	7
2.3	체언(n)	8
2.3.1	nct(temporal common noun, 시간성 보통 명사)	8
2.3.2	nca(active common noun, 동작성 보통 명사)	8
2.3.3	ncs(stative common noun, 상태성 보통명사)	9
2.3.4	nc(common noun, 보통명사)	9
2.3.5	nq(proper noun, 고유명사)	10
2.3.6	nbu(unit bound noun, 단위성 의존명사)	11
2.3.7	nb(bound noun, 의존명사)	11
2.3.8	npp(personal pronoun, 인칭 대명사)	11
2.3.9	npd(demonstrative pronoun, 지시 대명사)	12
2.3.10	nnn(number, 숫자)	12
2.3.11	nn(numeral, 수사)	13
2.4	용언(p)	13
2.4.1	pv(verb, 동사)	13
2.4.2	pad(demonstrative adjective, 지시 형용사)	14
2.4.3	pa(adjective, 형용사)	14
2.4.4	px(auxiliary verb, 보조 용언)	15
2.5	관형어(m)	15
2.5.1	md(demonstrative adnoun, 지시 관형사)	15
2.5.2	mn(numeral adnoun, 수 관형사)	16
2.5.3	m(adnoun, 관형사)	16
2.6	부사어(a)	16
2.6.1	at(time adverb, 시간 부사)	16
2.6.2	ad(demonstrative adverb, 지시 부사)	17
2.6.3	ajw(word-conjunctive adverb, 단어 접속 부사)	17
2.6.4	ajs(sentence-conjunctive adverb, 문장 접속 부사)	17

2.6.5	a(adverb, 부사)	17
2.7	독립언(i)	18
2.7.1	i(interjection, 감탄사)	18
2.8	조사(j)	18
2.8.1	jc(case particle, 격 조사)	18
2.8.2	jcm(adnominal case particle, 관형격 조사)	19
2.8.3	jcv(vocative case particle, 호격 조사)	19
2.8.4	jca(adverbial case particle, 부사격 조사)	19
2.8.5	jcp(predicative case particle, 서술격조사)	20
2.8.6	jx(auxiliary particle, 보조사)	20
2.8.7	jj(conjunctive particle, 접속조사)	20
2.9	어미(E)	21
2.9.1	ecq(equal conjunctive ending, 대등적 연결 어미)	21
2.9.2	ecs(subordinative conjunctive ending, 종속적 연결 어미)	21
2.9.3	ecx(auxiliary conjunctive ending, 보조적 연결 어미)	22
2.9.4	exm(adnominal ending, 관형사형 전성 어미)	22
2.9.5	exn(nominal ending, 명사형 전성 어미)	22
2.9.6	exa(adverbial ending, 부사형 전성 어미)	23
2.9.7	efp(prefinal ending, 선어말 어미)	23
2.9.8	ef(final ending, 종결 어미)	23
2.10	접사(x)	24
2.10.1	xn(noun suffix, 명사 접미사)	24
2.10.2	xpv(verb-derived suffix, 동사 파생 접미사)	25
2.10.3	xpa(adjective-derived suffix, 형용사 파생 접미사)	26
2.10.4	xa(adverb-derived suffix, 부사 파생 접미사)	26
<b>3</b>	<b>토의 및 결론</b>	<b>27</b>

# 1 소개

규칙을 기반으로 한 자연언어 처리는 여러 분야에서 그 한계를 보이고 있다. 최근에 와서 이와 같은 문제점을 해결하기 위해서 자연언어 처리의 거의 모든 분야에서 여러 가지 형태의 부가적인 정보를 사용한다. 이들 부가적인 정보의 대부분은 확률 및 통계 정보이다. 이와 같은 정보는 어떤 종류의 태그가 부착된 대량의 말뭉치로부터 추출하게 된다. 그와 같은 말뭉치를 형성하기 위해서는 먼저 각 성분(태그가 부착될 단위)에 어떤 종류의 태그를 붙여서 그들을 서로 구별할 지를 결정해야 한다. 본 연구에서 한국어 문장에 나타나는 형태소(언어학자들이 말하는 형태소와는 조금 차이가 있다.)를 기본 단위로 그들에 부착될 적절한 태그를 결정하는 것이며, 확률 및 통계를 이용한 자연언어 처리의 가장 기본적인 단계이다. 본 연구에서 말하는 품사의 태그는 언어학자들이 말하는 품사(part-of-speech) 혹은 문법 범주(grammatical category)와 아주 유사하다. 품사는 문법적 성질이 공통된 단어끼리 모아 놓은 단어의 갈래를 말한다[5]. 언어학자들은 품사분류 기준으로 일반적으로 단어의 의미, 기능, 형식을 든다.

의미란 개별적인 어휘의 의미가 아니라, 형식적인 의미로서 어떤 단어가 사물의 이름을 나타내느냐 그렇지 않으면 움직임이나 성질 상태를 나타내느냐 하는 것이다. 예를 들면, ‘무게’, ‘묵직하다’, ‘무겁다’는 어휘적인 의미를 볼 때는 하나의 덩어리로 생각할 수 있다. 그러나, 이들을 하나의 품사로 보기는 매우 어렵다. 따라서, ‘무게’, ‘높이’, ‘길이’ 등 어떤 사물의 이름을 표시하는 이들을 하나의 덩어리로 생각할 수 있다.

기능이란 한 단어가 문장 가운데서 다른 단어와 맺는 관계를 가리킨다. 이를 조사하는 방법으로 어떤 문장의 틀을 만들어 그 틀 내에서 같은 위치에 올 수 있는 단어들의 무리가 같은 품사를 가질 수 있다. 예를 들면, ‘X가 얼마냐?’라는 틀에서 X가 위치하는 곳에 ‘높이’, ‘무게’, ‘길이’ 등이 올 수 있으므로 이들은 같은 품사로 간주할 수 있다.

형식이란 단어의 형태적인 특징을 의미한다. 이것은 단어와 단어가 결합될 때 어떠한 변화를 가지느냐에 따라 분류되는 것이다. 예를 들면, ‘읽{었다, 는다, 는구나, 어라, ... }’, ‘회{었다, 다, 구나, \*어라, ... }’, ‘(책)이{었다, 는다, 로구나, \*어라, ... }’ 등과 같이 서로의 변화 양상이 조금씩 다르다. 따라서 이들을 서로 다른 품사를 부여한다. 즉, ‘읽’는 동사, ‘회’는 형용사, ‘(책)이’는 서술격 조사이다.

전통적인 품사분류에서는 의미, 기능, 형식 가운데서 기능이 우선적으로 고려되었다. 따라서 학교문법에서 주로 통용되는 9품사 체계는 먼저 기능을 중심으로 국어의 품사를 가르고(체언, 용언, 관계언, 수식언, 독립언) 다음으로 형식에 따른 분류를 제시한 것이다. 본 연구에서는 이를 바탕으로 품사 태그를 더 세분한다. 예를 들면, 같은 수사라고 할지라고 숫자 형태로 나타난 수사는 순서적으로 다음에 나타나는 의존 명사와 붙어서 쓴다. 예를 들면, “모두 10명이다.”와 “모두 십 명이다.”에서 ‘10’과 ‘십’을 서로 다른 품사를 갖도록 한다.

한국어 문법 범주의 분류에 관한 연구는 언어학자에 의해서는 물론이고, 전산학자에 의해서도 이미 많은 연구들이 이루어졌다[9, 12]. 본 연구에서는 가능한 한 정확한 확률 정보를 추출하기 위해서 이들 연구 결과를 참조하여 품사 태그를 재조정한다. 또한 품사 태그는 계층적으로 구성하여 가능한 한 문법 범주의 성질을 유지하면서 여러 응용 분야에서도 이를 쉽게 이용할 수 있도록 하였다. 예를 들면 정보 검색에서 색인어가 되는 품사는 주로 명사이다. 따라서 명사 이외의 다른 품사에 대해서는 특별한 세분류가 필요없을 것이다.

## 2 한국어의 품사 태그

언어학자들의 품사 분류 기준인 통사론적 기준(기능), 형태론적 기준(형태), 의미론적 기준(의미)에 따라서 크게 품사 태그를 체언(명사, 대명사, 수사), 용언(동사, 형용사), 수식언(관형사, 부사), 관계언(조사), 독립언(감탄사) 외에도 기호, 외국어, 접사로 나뉘었다. 용언은 자연언어 처리에 적합하도록 어근과 어미로 분리하여 다루고 수식언도 관형사와 부사로 구성되었으나 실질적으로 수식을 한다는 성질 이외에는 어떠한 공통점이 없기 때문에 이들을 각각 분리하여 다룬다.

본 연구에서는 아래와 같은 몇 가지의 고려사항을 고려하여 가능하면 문법 범주를 벗어나지 않는 범위에서 품사 태그를 정하였다.

1. 정확한 문서(사람이 보아서)만을 고려한다.
2. 화상(image)과 수식은 고려하지 않는다.
3. 기본적으로 형태소 단위로 한다. 그러나, 경우에 따라서는 두 개 이상의 형태소가 어울려 하나의 개념을 가리키는 말은 하나의 덩어리로 간주한다.
4. 공백(space)도 처리 상의 하나의 태그로 생각한다. 한국어에 있어서 공백은 띄어쓰기의 중요한 정보가 된다.
5. 접두어는 고려하지 않는다. 즉, “강추위”, “개살구” 등은 하나의 단어로 간주한다.
6. 조사끼리의 결합은 하나의 조사로 본다. 이때 문제는 여러 가지의 결합에 따라 어떻게 품사를 부착할 것인가에 있다. 예를 들면, 격조사+격조사⇒ 격조사, 보조사+격조사⇒ 격조사, 보조사+보조사⇒ 보조사, 격조사+보조사⇒ 보조사

### 2.1 기호(s)

기호는 [4, 8]을 주로 참조하였다. 영어나 다른 외국어와는 달리 한국어 처리에서는 공백(space)이 유용한 정보이므로 이를 하나의 태그로 간주한다.

#### 2.1.1 s,(comma, 쉼표)

쉼표는 같은 자격을 지닌 어구의 나열, 대등절과 종속절이 이어질 때, 부르는 말이나 대답하는 말 뒤 등에서 사용된다.

- 예문:  
“명사와 동사<sub>2</sub>, 동사와 동사<sub>2</sub>, 부사와 동사가 서로 어울려 말끝이 ‘기’로 끝나는 말로서, 하나의 동작<sub>2</sub>, 작업<sub>2</sub>, 상태<sub>2</sub>, 놀이를 나타내는 술어는 붙여 쓴다.”
- 이 부류에 속하는 기호들:  
,(쉼표), :(가운데점), :(쌍점), /(빗금), ...

- 참고사항:  
쌍점은 어말 어미로 끝나지 않고 간단한 설명을 추가할 때에 사용된다. 예를 들면, 내포되는 종류를 적을 때에 사용되는 것으로 “문장 부호: 마침표, 쉼표, 따옴표 등”에서와 같은 쌍점을 의미한다.

### 2.1.2 s.(sentence closer, 문장의 종결)

문장의 종결을 나타내는 기호를 나타낸다.

- 예문:  
“아! 달이 떠오른다.”  
“별써, 가을이 왔구나!”  
“아프니?”
- 이 부류에 속하는 기호들:  
.( 마침표), ! (느낌표), ? (물음표), ,( 쌍점), ; ...
- 참고사항:  
부연 설명을 위해서 사용되는 쌍점을 말한다. 예를 들면, “사회가 안정되고, 생활은 넉넉했다: 그래서, 국민은 행복하게 살 수 있었다.”에서와 같은 쌍점을 의미한다.

### 2.1.3 s‘ (left quotation and parenthesis mark, 여는 따옴표 및 묶음표)

대화, 인용, 특별어귀 따위의 시작을 표시하거나, 주석 및 부가적인 설명을 시작할 때에 사용된다.

- 예문:  
“아이가 어제 처음으로 “맘마”라고 했다.”
- 이 부류에 속하는 기호들:  
“( 여는 큰따옴표), ‘ ( 여는 작은 큰따옴표), ( ( 여는 소괄호), { ( 여는 중괄호), [ ( 여는 대괄호) ...

### 2.1.4 s’(right quotation and parenthesis mark, 닫는 따옴표 및 묶음표)

대화, 인용, 특별어귀 따위의 마지막을 표시하거나, 주석 및 부가적인 설명을 끝낼 때에 사용된다.

- 예문:  
“아이가 어제 처음으로 “맘마”라고 했다.”
- 이 부류에 속하는 기호들:  
” ( 닫는 큰따옴표), ’ ( 닫는 작은 큰따옴표), ) ( 닫는 소괄호), } ( 닫는 중괄호), ] ( 닫는 대괄호) ...

### 2.1.5 s-(connection mark, 이음표)

말을 붙이거나 ‘내지’의 뜻 등으로 사용되는 기호이다.

- 예문:  
“1990년 9월 15일 ~ 1993년 9월 25일”  
“어머님께 말했다가 - 아니, 말씀드렸다가 - 꾸중만 들었다.”
- 이 부류에 속하는 기호들:  
-(출표 및 붙임표), ~(물결표),

### 2.1.6 su(unit, 단위)

일반적으로 단위를 나타내는 기호로서 그 량을 나타내는 숫자 다음에 온다.

- 예문:  
100%, 10Kg
- 이 부류에 속하는 기호들:  
mm, cm, %, ...

### 2.1.7 sw(currency, 화폐 단위)

화폐 단위를 나타내는 기호로서 일반적으로 해당하는 금액을 나타내는 말 앞에 온다.

- 예문:  
\$100, £100, W100, ...

### 2.1.8 sp(space, 공백)

### 2.1.9 sy(other symbols, 기타 여러 기호들)

기타 여러 다른 종류의 특수기호를 의미한다.

## 2.2 외국어(f)

### 2.2.1 f(foreign word, 외국어)

외국어는 문장 상에 나타난 모든 외국어를 말한다.

- 예문:  
“흔히 미티(MITI)라고 불리는 통산성은 ‘일본주식회사’를 이끄는 기관사로 ...”
- 참고사항:  
한국어 처리에 있어서 이들은 특별히 자주 사용되는 것을 제외하고는 사전에 일일이 등록하지 않고 여러 가지 방법을 동원해서 적절한 품사를 예측해서 사용한다.



## 2.3 체언(n)

체언은 사물의 실체를 가리키는 말로서 문장에서 조사의 도움을 받아서 주어, 목적어, 보어의 구실을 하는 말이다.

### 2.3.1 nct(temporal common noun, 시간성 보통 명사)

시간성 보통 명사는 시간을 나타내는 많은 말로서 단독으로 사용될 경우에는 부사적인 역할을 하고, 조사와 함께 어울려 명사적인 역할을 한다.

- 예문:  
“어제는 학교 가는 날이다.”(시간성 보통 명사)  
“어제 학교에서 놀았다.”(시간성 부사)
- 이 부류에 속하는 단어들:  
오늘, 요즈음, 이래, 잠시, 저녁, 지금 ,,,
- 참고사항:  
시간성을 가지지 않더라도 보통 명사와 부사로 함께 사용되면서 의미적으로 서로 같을 경우에는 이 부류에 속한다.
  - 예문:  
“그 사건의 **대강**은 이러하다.”  
“형편은 **대강** 알고 있다.”
  - 예:  
대략, 대충, ...

### 2.3.2 nca(active common noun, 동작성 보통 명사)

동작성 보통 명사는 행위의 뜻을 내포하고 있는 명사를 의미한다. 한자어 뿐 아니라 외래어의 경우에도 같은 취급을 한다. 이 부류에 속하는 보통 명사는 일반적으로 ‘-하다’, ‘-되다’와 같은 동사 파생 접미사와 어울려 동사가 된다.

- 예문:  
“**공부**하는 것은 참으로 어려운 일이다.”  
“**다이나믹**하게 움직이는 공이 저기 보인다.”
- 이 부류에 속하는 단어들:  
가공, 가담, 가정, 계약, 박탈, 방송, 병행, 정의, 평가, 표현, 향상, ...
- 참고사항:
  1. ‘-하다’와 함께 어울려 형용사가 되는 명사(상태성 보통 명사)는 여기서 제외된다. 이들은 상태성 보통 명사(ncs)로 등록한다. 2.3.3절을 참조하기 바란다.

2. **동사 파생 접미사(xpv)와 함께 어울려 동사(pv)가 된다.** 동작성 보통명사와 동사 파생 접미사(xpv)가 함께 어울리면 구문해석 이상에서 이용하기 위해서는 이들 둘을 결합해서 하나로 간주해야 한다. 예를 들면, ‘-하다’와 ‘-되다’와 같은 와 결합하면 동사(pv)가 된다. 2.10절을 참조하기 바란다.
3. 일반적으로 ‘-하다’와 결합되면 타동사가 되고, ‘-되다’와 결합되면 자동사된다. 그러나, 그렇지 않은 경우도 많다. 예를 들면, ‘가입하다’는 자동사와 타동사가 모두 가능하고, ‘경유하다’는 타동사만 가능하고, ‘귀속하다’는 자동사로만 가능하다. 본 연구에서는 이를 구별하지 않고 하나의 품사로 사용한다. 이것에 관한 정보는 격틀을 이용하여 해결하고자 한다. 더 자세한 정보는 격틀에 관한 설명을 참조하기 바란다[2].

### 2.3.3 ncs(stative common noun, 상태성 보통명사)

상태성 보통 명사는 어떤 상태를 가리키는 말로서 동사 파생 접미사 ‘-하다’와 함께 어울려 형용사가 된다.

- 예문:  
“이 공사는 한 달이면 **만족**하리라고 생각한다.”
- 이 부류에 속하는 단어들:  
가난, 가능, 간곡, 간단, 감사, 강력, 거대, 건강, 건장, 건전, 격렬, 견고, 고상, 곤란, 공정, 과감, 과격, 과분, 광범, 동일, 마땅, 만족, ...
- 참고사항:  
상태성 보통명사(ncs)는 부사 파생 접미사 ‘-히’와 함께 어울려 부사가 되기도 한다. 예를 들면, 간곡**히**, 간단**히**, 감사**히**, 강력**히** 등이다.

### 2.3.4 nc(common noun, 보통명사)

보통명사 중에서 시간성, 동작성, 상태성이 아닌 보통명사가 여기에 속한다.

- 예문:  
“**사람**은 만물의 영장이다.”
- 이 부류에 속하는 단어들:  
가게, 가격, 가관, 가구, 가내, 가닥, 가락, 가랑비, 가랭이, 가로등, 가로, 가마솥, 가망, 가명, 가방, 가속도, 가수, 가스, 가습, 가시, 가운데, 가족, 가족
- 참고사항:  
1. **명사 파생 접미사와 결합된 명사는 사전에 등록한다.** 국어 사전 상에 존재하는 명사형 접미사는 한자어로 된 것이 많다. 이들이 한글로 표기될 경우에 많은 모호성을 가지고 있다. 이와 같은 모호성은 구문 및 의미 해석에 의해서 해석할 수 없다. 예를 들면, 명사 파생 접미사 ‘-사’는 사찰(temple), 역사(history), 전문가(expert), 회사(company), 사건(event), 건물(house) 등의 의

미를 갖는다. 이들이 ‘기숙+사’에서 ‘-사’가 어떤 의미로 사용되었는지 알 수 없으므로 ‘기숙사’로 사전에 등록한다. 2.10를 참조하기 바란다.

2. ‘-음’, ‘-고’, ‘-기’는 **명사형 전성어미로 간주한다**. 이들은 용언의 어근과 결합해서 명사를 파생시킨다. 그러나 문장의 의미를 분석함에 있어서 이와 같은 파생명사는 단순한 명사와는 달리 앞에 나오는 성분의 용언적인 역할을 수행한다. 2.9.3절을 참조하기 바란다.
3. **명사형 굴급 접미사와 결합된 명사는 사전에 등록하지 않는다**. 그러나, 명사 굴급 접미사(inflexional suffix)는 보통 명사만 사전에 등록하고, 명사 굴급 접미사(-들, -님, 등)는 명사형 접미사로 등록된다. 이들은 형태소 해석 후에 적절한 처리에 의해서 서로 결합하여 하나의 단어로 만들어야 한다. 2.10.1절을 참조하기 바란다.
4. **체언을 만드는 합성 명사는 사전에 등록한다**. 한국어에서는 체언을 만드는 여러 가지의 방법(조어법)이 있다. 예를 들면, 명사와 명사의 결합(산올림, 가로등, 손발 등), 관형사와 명사의 결합(새마을, 큰물, 빈주먹 등), 용언의 어간과 명사의 결합(늦잠, 묵발 등) 부사와 명사의 결합(혼잣말, 부슬비, 혈떡고개 등) 등이 있다. 이들은 사전에 등록한다. 이들의 관계를 구문적 혹은 의미적으로 해석하기는 대단히 어려운 일이다.

### 2.3.5 nq(proper noun, 고유명사)

특정한 사람이나 사물에 대해서 붙여진 이름, 사람 이름, 땅 이름, 나라 이름, 책 이름, 사건 이름 등을 나타내는 단어가 여기에 속한다.

- 예문:  
“인수는 오늘도 **동대문**에서 지하철을 탔다.”
- 이 부류에 속하는 단어들:  
경기, 고성, 김포, 남대문, 남영동, 뉴욕, 대전, 동남, 런던, 서귀포, 삼국유사, ...
- 참고사항:
  1. **고유명사의 무한성을 갖는다**. 고유명사는 새로이 만들어 지는 것이 너무 많기 때문에 이를 해결하기 위한 여러 형태의 견고한 처리를 필요로 한다.
  2. **고유명사는 띄어 쓸 수도 있고 붙여 쓸 수도 있다**. 여기서는 단어 단위로 띄어 쓰기를 원칙으로 한다. 예를 들면, ‘대한 중학교’, ‘한국 과학 기술원’, ‘서울 대학교’ 등과 같이 각 단어를 띄어 쓴다. 이유는 이렇게 할 경우에 새로이 만들어지는 많은 고유명사에 대해서 형태소 해석까지는 처리가 가능할 것이다. 그리고 고유명사는 처리하고자 하는 영역에 따라서 많은 차이를 보이기 때문에 응용 분야에 따라 고유명사 사전을 따로 가지고 있어야 할 것이다. 따라서 기본 사전에는 모든 고유 명사를 등록하지 않는다.  
그러나, 실질적으로 이들은 하나의 객체를 가리키는 말이기 때문에 하나의 품사로 사용되어야 한다. 따라서 이는 구문해석을 처리하기 전에 띄어 쓰기가 된 여러 개의 단어를 하나의 묶어리로 묶어 주는 작업이 필요하다.

### 2.3.6 nbu(unit bound noun, 단위성 의존명사)

수 관형사 밑에서 의존하여 그 수의 단위를 표시해 주는 의존명사이다.

- 예문:  
“쌀 한 **자루**만 배달해 주세요.”
- 이 부류에 속하는 단어들:  
집 한 **채**, 두 **개**, 세 **분**, 한 **그루**, 100**원**, 차 한 **대**, 개 두 **마리**, 18**개월**, 두 **번째**, ...

### 2.3.7 nb(bound noun, 의존명사)

명사의 성격을 띠고 있으면서도 그 의미가 형식적이어서 제 홀로 자립하여 쓰이지 못하고 반드시 앞에서 꾸미는 관형어 상당어구(관형사, 관형격 조사)의 도움을 받아 온전하게 쓰이는 명사이다. 보편성, 주어성, 서술성, 부사성 의존 명사들 모두가 여기에 속한다.

- 예문:  
“저기 가시는 **분**이 언제 이곳에 다시 올 **지** 알 수 없다.”  
“모르는 **체** 시치미를 떤다.”
- 이 부류에 속하는 단어들:  
것, 나위, 데, 듯, 등, 따름, 따위, 때문, 리, 만큼, 만, 바, 분, 뿐, 뿐, 수, 양, 줄, 지, 채, 척, 체, 터, ...
- 참고사항:
  1. **의존 명사는 개개의 단어에 따라서 여러 형태의 모호성을 가지고 있다.** 예를 들면, ‘뿐’, ‘대로’, ‘만큼’ 등은 조사이면서 의존명사이다. 이들의 구별은 좌우의 문맥정보에 의해서 결정될 수 있다. 용언의 관형형 다음에 이들 단어가 올 경우에는 의존 명사이고, 체언 다음에 올 경우에는 조사이다.
  2. 그밖에도 ‘양’, ‘체’, ‘척’ 등은 동사 파생 접미사(xpv) ‘-하다’와 어울려 그럴 듯하게 꾸미는 거짓 태도라는 뜻의 보조용언(실제로 보조동사)(px)로 사용된다. 그러나 보조용언의 경우에는 ‘양하’, ‘척하’, ‘체하’ 형태로 등록한다. ‘듯’이 ‘-하다’와 어울려 의 경우도 위의 경우와 같은 방법으로 처리한다. 이 경우에는 일반적으로 앞 어절은 관형절이다. 예를 들면, “겨울이 온 **듯하다**”에서 ‘온’은 관형어이다.

### 2.3.8 npp(personal pronoun, 인칭 대명사)

사람을 가리키는 대명사이다. 미지의 사람을 가리키는 인칭 의문 대명사도 여기에 속한다.

- 예문:  
“나는 영희가 **좋다**.”  
“**자네**가 **그**이보다 낫네.”

- 이 부류에 속하는 단어들:  
나, 우리, 저, 저희, 너, 너희, 자네, 당신, 그대, 저놈, 그애, 그이, 그분, 누구, 자기, 당신, ...
- 참고사항:  
**재귀 대명사는 단순한 대명사로 간주한다.** 한 문장 안에서 주어를 도로 가리키는 대명사를 재귀 대명사라고 한다. 재귀 대명사라고 하는 ‘자기’, ‘자신’ 등은 한 문장 안에서 재귀적으로 사용되기는 하나, 일반적인 대명사 처럼 단순한 대명사로서의 역할을 한다. 예를 들면, “그 친구 정말 건망증이 심하군! **자**기가 그런 짓을 해놓고 누구를 탓한담.”에서 ‘자기’는 3인칭 대명사이다.

### 2.3.9 npd(demonstrative pronoun, 지시 대명사)

사물이나 장소를 대신하여 지시하는 대명사이다. 미지의 사물을 가리키는 의문 대명사도 여기에 속한다.

- 예문:  
“거기서 무엇을 하느냐?”
- 이 부류에 속하는 단어들:  
이것, 그것, 저것, 여기, 저기, 거기, 그때, 이때, 저때, 이쪽, 그쪽, 저쪽, 무엇, 어디, ...
- 참고사항:  
의문의 뜻으로 사용되는 단어는 여러 가지의 품사(대명사, 관형사, 부사 등)로 나타난다. 예를 들면, ‘누구’와 ‘무엇’ 등은 대명사 형태로 ‘무슨’, ‘어느’, ‘몇’ 등은 관형사 형태로, ‘얼마나’, ‘왜’ 등은 부사 형태로 나타난다. 따라서 의문을 나타내는 품사를 사용하는 것은 효율적인 방법이 아니기 때문에 어떤 자질(feature)에 의해서 표현된다. 예를 들면, ‘언제’에 대한 사전에 대한 품사는 ‘npd’이고, 의문의 종류를 나타내는 자질(whq)을 추가한다.

### 2.3.10 nnn(number, 숫자)

수량을 나타내는데 있어서 우리말이 아닌 숫자를 직접 이용하여 나타내는 경우의 수사이다. 일반적으로 이와 같이 사용될 경우에는 다음에 오는 말과 붙여서 쓴다.

- 예문:  
“신경제 **5**개년 계획에 제시된 공정거래부문은 ...”
- 이 부류에 속하는 단어들:  
**1, 12.4, 1,234, ...**
- 참고사항:  
1. **숫자는 사전에 등록하지 않는다.** 정수, 소수, 분수 등 숫자 자체를 표현하는 단어들이다.

- 2. 숫자는 일반적으로 단위성 의존명사와 붙여 쓴다. 예를 들면, ‘80원’, ‘2대대’, ‘7미터’, ‘10개’ 등이다.

### 2.3.11 nn(numeral, 수사)

사물의 수량이나 순서를 가리키는 말로서 한글로 표현된 단어가 여기에 속한다. 둘 이상의 수사가 결합하여 그 수를 어림잡아 나타내는 부정수도 여기에 속한다.

- 예문:  
“서울에 동생 **하나**가 있다.”
- 이 부류에 속하는 단어들:  
하나, 둘, 셋, 넷, 다섯, 여섯, 일곱, 여덟, 아홉, 열, 스물, 마흔, 서넛, 예닐곱, 일, ....
- 참고사항:  
**서수사도 수사에 포함된다.** 한글로 표기되는 서수사들은 수관형사 혹은 수사와 접사와 결합한다. 예를 들면, ‘첫째’는 ‘첫’(수관형사)+‘째’(접미사)와 결합된 것이다. 이와 같은 두 종류의 품사에 의해서 결합되어 하나의 품사를 나타내는 경우는 구문해석 전에 전처리에 의해서 처리된다. 그러나, 한자로 표기되는 서수사 ‘제일’, ‘제이’ 등은 수사로 표기한다.

## 2.4 용언(p)

### 2.4.1 pv(verb, 동사)

사물의 움직임이나 작용을 나타내는 말을 말하며 문장의 서술어가 되는 것이 주기능이다.

- 예문:  
“강물이 유유히 흐르**다**.”  
“철수가 노래를 부르**다**.”
- 이 부류에 속하는 단어들:  
가다, 건너다, 눕다, 늙다, 닳다, 뜨다, 마시다, 먹다, 새다, 오다, 웃다, 읽다, 있다, 잡다, 죽다, 짓다, 피다, 흐르다, ...
- 참고사항:  
1. **사동 및 피동 형태를 사전에 등록한다.** 사동 파생 접미어 ‘-이-’, ‘-히-’, ‘-리-’, ‘-기-’, ‘-우-’, ‘-추-’와 피동 파생 접미어 ‘-이-’, ‘-히-’, ‘-리-’, ‘-기-’는 모호성을 가진다. 이들은 동사에 따라서 모호성을 갖는 접미어와 함께 사용될 때 사동사와 피동사가 결정된다. 예를 들면, ‘속이다’는 ‘속다’의 사동사이고, ‘덮이다’는 ‘덮다’의 피동사이며, ‘갈리다’는 ‘땅을 갈다’의 피동사인 동시에 ‘자리를 갈다’의 사동사이다. 따라서 이들은 접미어는 접미어로 취급하지 않고 어근과 사동/피동 파생접사가 결합되어 하나의 동사로 간주하고 용언에 따른

개별적인 정보는 자질 정보로 추가한다. 예를 들면, 동사 ‘간히다’의 개별적인 자질 정보 [[root,가두], [voice,passive]]를 추가적으로 기입하여 구문 및 의미 해석에서 이를 이용한다.

2. 부사와 어울리는 동사 파생 접미사 ‘-거리다’ 등은 동사로 등록한다. 예를 들면, ‘만지작거리다’, ‘반짝거리다’, ‘출렁이다’ 등이다.
3. 본 연구에서는 **자동사와 타동사를 합하여 하나의 동사로 처리한다**. 형태소 해석에서는 사동 및 피동형태가 모두 사전에 등재되기 때문에 타동사와 자동사를 구별하는 것은 별다른 의미를 주지 않는다. 그러나 구문 및 의미 해석에서 다소 영향을 초래할 수 있다. 이는 격률 정보를 이용해서 해소할 수 있다. 또한 많은 동사가 타동사와 자동사이므로 모호성의 너무 많이 증가할 뿐 아니라 그 문장의 의미를 이용하지 않고 서로 구별할 수 있는 별다른 정보가 없는 실정이다.

#### 2.4.2 pad(demonstrative adjective, 지시 형용사)

사물의 성질, 모양, 상태 등을 지시하는 형용사이다.

- 예문:  
“내 생각은 역시 그러하+다.”
- 이 부류에 속하는 단어들:  
그렇다, 아무렇다, 어떠하다, 이렇다, 저렇다, 조렇다, ...
- 참고사항:  
의문 형용사도 여기에 포함되며, 2.3.9절의 참고사항을 참고 바란다.

#### 2.4.3 pa(adjective, 형용사)

사람이나 사물의 성질과 상태가 어떠한가를 형용하거나 그 존재를 나타내면서 문장 안에서 주로 서술어 기능을 가지는 단어이다.

- 예문:  
“이 꽃이 매우 아름답+다.”
- 이 부류에 속하는 단어들:  
끔찍하다, 기쁘다, 깨끗하다, 높다, 맑다, 뻑뻑하다, 조용하다, 좋다, 크다, ...
- 참고사항:  
‘이히’ 형용사<sup>1</sup>를 별도로 구별하지 않는다. 종전에 ‘이히’ 형용사는 그 자체가 어떤 형태소 역할을 하지 못하기 때문에 사전의 수가 조금 늘어 나더라도 사전에 첨가하기로 결정하였다.

---

<sup>1</sup>이히 형용사는 부사 파생 접미사 ‘-이’나 ‘-히’와 어울려 형용사가 되는 말은 한단다[3].

#### 2.4.4 px(auxiliary verb, 보조 용언)

보조 용언은 겉모습은 일반적인 동사나 형용사와 다름이 없지만, 자립성이 없거나 약하여 앞에 오는 본 용언에 기대어 쓰이는 무리의 용언으로 본 용언의 의미를 도와 주는 용언을 말한다.

- 예문:  
“이 꽃은 아름답지 **아니하다.**”  
“영숙이는 어제 떠나 **버리었다.**”
- 이 부류에 속하는 단어들:  
-게 되다, -게 만들다, -게 지다, -게 하다, -고 있다, -고 보다, -고 싶다, -나 보다, -는가 보다, -아/어 가다, -아/어 나다, -아/어 내다, -아/어 놓다, -아/어 대다, -아/어 두다, -아/어 드리다, -아/어 되다, -아/어 버리다, -아/어 보다, -아/어 있다, -아/어 오다, -아/어 주다, -아/어 지다, -아/어 하다, -지 말다, -지 못하다, -지 아니하다, -지 않다,
- 참고사항:
  1. **보조 용언의 띄어쓰기는 원칙적으로 띄어 쓴다.** 붙여 쓰는 경우는 견고한 처리에 의해서 이를 처리할 수 있도록 한다. 그러나 두 동사가 나란히 자신의 뜻을 가지고 함께 사용되는 합성동사는 하나의 덩어리로 간주한다. 예를 들면, ‘들어가다’, ‘들어내다’, ‘따라가다’, ‘뛰어가다’, ‘뛰어들다’ 등이다.
  2. **보조 동사와 보조 형용사는 모두 본 용언의 의미를 더해 주는 역할을 수행한다.** 그러나 보조 형용사는 형용사적인 성질을 가지고 있기 때문에 현재 진행의 의미를 지닌 ‘-다’와 ‘-는다’가 붙을 수 없다. 예를 들면 “호르지 **아니하다.**”에서 ‘아니하’는 보조 동사이고 “깊지 **아니하다.**”에서 ‘아니하’는 보조 형용사이다. 그러나 이들은 부정의 의미를 전달하는 것은 동일하다. 따라서 이들을 서로 구별할 필요가 없이 모두 본 용언의 의미를 더하는 것으로 간주한다.

## 2.5 관형어(m)

관형어는 체언 앞에서 그 체언의 내용을 구체적으로 어떠한다고 꾸며 주는 말이다.

### 2.5.1 md(demonstrative adnoun, 지시 관형사)

지시 관형사는 말하는 이가 주관적으로 사물의 성질이나, 상태 등을 가리키어 꾸며 주는 구실을 하는 관형사이다.

- 예문:  
“**이** 어른이 저 사람을 도와 준 바로 **그** 선생님이시다.”
- 이 부류에 속하는 단어들:  
그, 그런, 무슨, 본, 아무, 어느, 어떤, 요, 이, 이런, 저, 저런, ...



- 참고사항: 의문 관형사(무슨, 어느 등)도 이 부류에 포함되며, 2.3.9절의 참고사항을 참고 바란다.

### 2.5.2 mn(numeral adnoun, 수 관형사)

수 관형사는 뒤에 오는 명사의 수량을 표시하건, 단위성 의존 명사와 어울려 앞에 오는 명사의 수량을 표시하는 관형사이다.

- 예문:  
“사과 **한** 개로 시장기를 면했다.”
- 이 부류에 속하는 단어들:  
한, 두, 세, 네, 다섯, 여섯, 스무, 여러, 모든, 몇, ...
- 참고사항: 숫자는 묻는 의문 관형사(몇, 등)도 이 부류에 포함되며 2.3.9절의 참고사항을 참고 바란다.

### 2.5.3 m(adnoun, 관형사)

관형사는 체언이 가리키는 사물이나 성질을 어떠한 방식으로 꾸며 주는 말이다.

- 예문:  
“**새** 책은 내 것이다.”
- 이 부류에 속하는 단어들:  
새, 옛, 오른, 온갖, 왼, 현, ...

## 2.6 부사어(a)

부사는 주로 용언 앞에 놓여서 뒤에 오는 용언을 꾸밈으로써 의미를 더욱 분명히 해 주는 말이다.

### 2.6.1 at(time adverb, 시간 부사)

시간 부사는 시간을 지시하여 한정하는 부사이다.

- 예문:  
“**내일** 다시 만납시다.”
- 이 부류에 속하는 단어들:  
그때, 접때, 가끔, 이따금, 그제, 어제, 오늘, 일찍, 이미, 이제, 내일, 모레, 언제, ...
- 참고사항:  
시간성 명사에 관한 설명은 2.3.1절의 참고사항을 참조하기 바란다.

### 2.6.2 ad(demonstrative adverb, 지시 부사)

지시 부사는 장소나 앞에 나온 이야기의 내용을 지시하는 부사이다.

- 예문:  
“이리 와서 자세히 이야기합시다.”  
“그들은 어디 있습니까?”
- 이 부류에 속하는 단어들:  
이리, 그리, 저리, 어디, 요리, 고리, 조리, 여기, 저기, 요기, 고기, 어찌, 아무리, ...
- 참고사항:  
의문 부사(어디 등)도 이 부류에 포함되며 2.3.9절의 참고사항을 참고 바란다.

### 2.6.3 ajw(word-conjunctive adverb, 단어 접속 부사)

단어 접속부사는 성분과 성분을 이어주는 역할을 수행하는 부사이다.

- 예문:  
“연필 **또는** 공책을 사야 했다.”
- 이 부류에 속하는 단어들:  
또, 또는, 곧, 빛, 혹은, 그리고, ...

### 2.6.4 ajs(sentence-conjunctive adverb, 문장 접속 부사)

문장 접속 부사는 문장과 문장을 이어 주면서 뒤의 말을 꾸며 주는 부사이다. 이들은 일반적으로 문두에서 문장을 수식하는 기능을 가진다.

- 예문:  
“그러나, 아무도 그것을 믿지 않았다.”
- 이 부류에 속하는 단어들:  
그래도, 그래서, 그러나, 그러면, 그러면, 그러므로, 그런데, 그렇지마는, 그리고, 더우기, 더욱, 따라서, 또는, 또한, 또, 왜냐하면, 이른바, 특히, 하지만, 한편, ...

### 2.6.5 a(adverb, 부사)

위에서 설명한 이외의 부사로서 성상 부사와 부정 부사와 문장 양태 부사가 여기에 포함된다. 문장 양태 부사는 일반적으로 문두에서 문장 전체를 수식한다. 그러나 이들은 대부분은 문장의 어느 위치에나 자연스럽게 나타나서 단순한 부사적인 역할을 한다.

- 예문:  
“올해는 날씨가 **매우** 청명하다.”(성상 부사)  
“그분은 **과연** 위대한 정치가이다.”(문장 부사)

- 이 부류에 속하는 단어들:  
가령, 갑자기, 거의, 결코, 과연, 글썄, 기어이, 기필코, 꼭, 꽤, 너무, 단연코, 더욱, 마치, 만약, 만일, 만일, 모름지기, 못, 물론, 바로, 반드시, 부디, 비록, 빨리, 설령, 설마, 설사, 설혹, 실로, 아니, 아마, 아무리, 아무쪼록, 아주, 안, 어찌, 얼마나, 왜, 응당, 잘 정녕, 정말, 정작, 제발, 조금도, 조금, 조용히, 진실로, 천천히, 철썩철썩, 펍, 하물며, 흡사, ...
- 참고사항:  
부정 부사(못, 아니 등)도 이 부류에 포함되며 2.3.9절의 참고사항과 같이 자질에 의해서 구별된다.

## 2.7 독립언(i)

### 2.7.1 i(interjection, 감탄사)

감탄사는 말하는 이의 본능적인 놀람이나 느낌을 표시하는 말이나, 부르고 대답하는 말 또는 입버릇으로 내는 말 등의 말이다.

- 예문:  
“예, 오늘 오후 세 시 정각에 모이기로 했대요.”  
“글썄요, 저도 잘 모르겠습니다.”
- 이 부류에 속하는 단어들:  
그래, 아이구, 아, 암, 애, 예, 여보세요, 여보시오, 여보, 오냐, 응, 저, ...

## 2.8 조사(j)

조사는 스스로 자립할 수 없고 자립성을 가진 다른 말(주로 체언)에 붙여서 다른 말과의 관계를 나타내거나(격조사, 접속조사) 뜻을 더해 주는 단어(보조사)이다. 조사는 체언 뿐만 아니라 체언의 상당어구(의문문, 인용문, 인용된 단어 등)와도 함께 어울려 상용된다. 예를 들면, “빵이냐 자유냐가 우리의 당면 과제이다.”(의문문), “나는 살고 싶다.”가 그의 외침이다.(인용문). ‘잡다’가 아니라 ‘먹다’이다.(인용된 단어), 등이다.

조사에서 가장 큰 문제를 복합 조사를 어떻게 처리할 것인가 하는 것이다. 복합조사는 하나의 조사로 표기하고 품사는 제일 끝의 조사에 의존한다.

### 2.8.1 jc(case particle, 격 조사)

스스로 자립성을 가지지 못하고 자립성을 가진 다른 말에 붙여서 다른 말과의 관계를 나타내는 역할을 한다. 주격 조사, 목적격 조사, 보격 조사가 여기에 속한다.

- 예문:  
“나무가 무럭무럭 자란다.”
- 이 부류에 속하는 단어들:  
-이, -가, -께서(높인 말), -에서(단체 표시어), -서, -을, -를, ...

- 참고사항:  
보격조사와 주격조사는 형태적으로 같기 때문에 이를 품사적으로 구별하지 않고 구문해석에서 격틀에 의해서 구별된다. 특히 부격조사를 취하는 용언은 ‘되다’, ‘아니다’로 극히 제한된다.

### 2.8.2 jcm(adnominal case particle, 관형격 조사)

체언이나 용언의 명사형 아래에 붙어 그 말을 관형어가 되게 하는 조사이다. 실질적으로 ‘-의’와 ‘-의’로 끝나는 복합조사들이 여기에 속한다.

- 예문:  
“우리**의** 처리를 약진**의** 발판으로 삼아, ...”
- 이 부류에 속하는 단어들:  
-의
- 참고사항:  
‘-의’로 끝나는 복합조사도 여기에 속한다. 예를 들면, -과의, -까지의, -께로의, -께만의, -대로의, -로부터의, -로서의, -로써의, -로의, -마다의 등이다.

### 2.8.3 jcv(vocative case particle, 호격 조사)

주로 사람을 가리키는 체언에 연결되어 호칭의 대상이 되게 하는 조사이다.

- 예문:  
“**철수아**, 어디로 가니?”
- 이 부류에 속하는 단어들:  
-야, -아, -여, -시여, -이시여, -이야, -이여,

### 2.8.4 jca(adverbial case particle, 부사격 조사)

체언과 용언의 명사형 아래에 붙어 그 말을 부사어로 만드는 조사이다.

- 예문:  
“이 편지를 우체통**에** 넣어라.”
- 이 부류에 속하는 단어들:  
-에게, -같이, -한테, -으로, -로, -에서, -처럼, -대로, -하고, -과, -와, -에, -한데서, -에게로, -으로써, -으로서, -로서, -로써, -더러, -보고, -처럼, -보다, -만큼, -랑, -라고, -이라
- 참고사항:  
기본 부사격 조사로 끝나는 복합조사도 여기에 속한다. 예를 들면, -까지로, -께로부터, -께서로부터, -라고까지, -로부터, -마다로부터, -마다에게, -만으로서 등이다.

### 2.8.5 jcp(predicative case particle, 서술격조사)

체언이나 용언 명사형 아래에서 연결되어 그 말로 서술어가 되게 하는 조사로서 ‘이다’ 하나뿐이다. ‘이다’는 다른 조사와 달리 형태소 ‘-이-’가 그 앞의 체언으로 하여금 용언적인 기능을 가지게 하며, 그 뒤에 활용 어미를 취하는 특이한 형태의 말이다.

- 예문:  
“이것은 책**이다**.”
- 이 부류에 속하는 단어:  
-이-
- 참고사항:  
서술격조사 ‘-이-’와 어울려 복합어미를 이루는 어미는 복합어미로 간주하지 않는다. 많은 국어 사전에는 서술격조사 ‘-이-’와 어울려 복합어미를 이루는 어미들은 하나의 어미로 등록하고 있다. 예를 들면, ‘-이로고’, ‘-이로구나’, ‘-이로구료’, ‘-이로다’, ‘-이며’, ‘-이지마는’, ‘-인지라’, ‘-일결’, ‘-일라’, ‘-일러나’, ‘-일레’, ‘-일레라’, 등은 ‘-이 + 로고’, ‘-이 + 로구나’, ‘-이 + 로구료’, ‘-이 + 로다’, ‘-이 + 며’, ‘-이 + 지마는’, ‘-이 + 짜라’, ‘-이 + 결’, ‘-이 + 라’, ‘-이 + 러나’, ‘-이 + 레’, ‘-이 + 레라’ 등으로 “서술격 조사 + 어미”로 분리해서 처리한다.

### 2.8.6 jx(auxiliary particle, 보조사)

보조사는 스스로 자립성을 가지지 못하고 자립성을 가진 다른 말에 붙어서 다른 말과의 뜻을 더해주는 역할을 하고, 연결되는 체언에 일정한 격을 규정하지 하지 않고 여러 격에 두루 쓰이게(그 문장이 요구하는 격을 가지게) 된다.

- 예문:  
“우리**도** 자유를 원한다.”(주격)  
“우리는 젊음**도** 구가한다.”(목적격)  
“올바른 생각**도** 아니다.”(보격)  
“우리는 학교**도** 갔었다.”(부사격)
- 이 부류에 속하는 단어:  
-곧, -그러, -까지, -탕, -나, -나, -는, -는커녕, -도, -든지, -라고, -라도, -랑은, -랑, -마는, -마다, -마저, -만, -부터, -뿐, -서낀, -야, -야말로, -요, -은, -은랑, -은커, -이나, -이든지, -이라도, -이야말로, -이야, -조차, ...

### 2.8.7 jj(conjunctive particle, 접속조사)

성분과 성분을 서로 연결하여 문장에서 동등한 자격을 가지게 하는 조사이다. 여기서 말하는 성분은 단어 혹은 문장이 될 수 있다.

- 예문:  
“**철수와 영희**가 찾아왔다.”  
“그는 **영어와 독일어와 프랑스어**를 할 줄 안다.”

- 이 부류에 속하는 단어들:  
-과, -다, -랑, -며, -에다, -와, -이랑, -이며, -하고, ...
- 참고사항:  
**접속조사 ‘-하고’, ‘-과’, ‘-랑’ 등은 보조사와 모호성을 갖는다.** 문장 내에서 서로 동등한 자격으로 사용되지 않는 경우에는 보조사로 사용되는 경우이다. 예를 들면, “철수는 나**하고** 같이 갔다.”에서 ‘-하고’는 보조사로 사용된 경우이고, “철수**하고** 나는 학생이다.”에서 ‘-하고’는 접속조사이다.

## 2.9 어미(E)

어미와 조사가 결합될 경우에 어떻게 할 것인지를 결정하는 것은 매우 어려운 일이다. 보조사의 경우에는 어미에 포함되도록하고 격조사는 분리하도록 한다.

### 2.9.1 ecq(equal conjunctive ending, 대등적 연결 어미)

대등적 연결 어미는 같은 종류의 두 용언을 맞서게 하거나 상반되는 두 용언을 맞서게 하여 이어 주는 말로서 나열, 선택, 반복, 동작의 계속해서 이어짐 등의 의미로 서로 연결된다.

- 예문:  
“노래**하며** 춤춘다.”,  
“가**든지** 말**든지** 마음대로 해라.”  
“하늘은 맑**고** 푸르다.”
- 이 부류에 속하는 단어들:  
-거나 -거나, -거니 -거니, -고, -느니 -느니, -든지 -든지, -며, -면서, -자, -락 -락, ...

### 2.9.2 ecs(subordinative conjunctive ending, 종속적 연결 어미)

두 문장을 주종 관계로 연결시키는 말로서 구속, 가정, 이유, 필연, 방임, 양보, 설명, 도급, 추정, 의도, 비교 등의 의미로 서로 연결된다.

- 예문:  
“입이 닳**도록** 타일렀다.”(도급)  
봄이 오**면** 꽃이 핀다.”(가정)  
고기를 잡**으려** 그물을 쳤다.(의도)
- 이 부류에 속하는 단어들:  
-거든, -거늘, -거든, -건마는, -고자, -다손치더라도, -데, -뜰, -족, -피라, -나, -나니, -나마, -노니, -는데, -는들, -는바, -는지라, -니, -니까, -다가, -더라도, -던들, -도록, -되, -디, -뿐더러, -추록, -피라도, -피언정, -라고, -러, -려, -려니와, -련마는, -매, -면, -므로, -아도, -아서, -아야만, -아야, -어도, -어서, -어야, -어야만, -으나, -으나마, -으니까, -으려, -으려, -으려니와, -으련마는, -으매, -으면, -으므로, -은데, -은들, -은바, -는지라, -은즉, -을뿐더러, -을수록, -을지라도, -을지언정, -자, -지라, -지마는, -지만은, ...

### 2.9.3 ecx(auxiliary conjunctive ending, 보조적 연결 어미)

보조적 연결 어미는 스스로 완전한 서술 능력을 가지지 못하고 보조 용언과 잘 어울려 완전한 서술 능력을 발휘한다. 즉, 보조적 연결 어미는 보조 용언의 도움을 받아 본 용언에 일정한 의미를 부여한다.

- 예문:  
“나비를 잡아 오너라.”  
“그 동안 정말 보고 싶었다.”  
“아이들에게 나비를 잡게 하여라.”
- 이 부류에 속하는 단어들:  
-게, -고, -아, -어, -지
- 참고사항:  
함께 어울려 사용하는 보조 용언은 2.4.4절을 참조하기 바란다.

### 2.9.4 exm(adnominal ending, 관형사형 전성 어미)

관형사형 전성 어미는 앞말에 대해서는 서술어적인 역할을 하고, 뒤따르는 말에 대해서는 관형어 역할을 하는 말이다. 또 관형사형 전성 어미는 현재, 과거, 미래 등의 시제를 동반한다.

- 예문:  
“내가 읽는 책은 만화책이다.”(현재)  
“내가 읽은 책은 만화책이다.”(과거)  
“내가 읽을 책은 만화책이다.”(미래)
- 이 부류에 속하는 단어들:  
동사 + **는**(현재), 동사 + (**은**)(과거), 동사 + (**을**)(미래), 형용사 + (**은**)**는**(현재),  
형용사 + **던**(과거), 형용사 + (**은**)**던**(미래), 서술격 조사 + **[**(현재), 서술격 조사 +  
**던**(과거), 서술격 조사 + **던**(미래),

### 2.9.5 exn(nominal ending, 명사형 전성 어미)

명사형 전성 어미는 용언을 명사처럼 활용시키므로, 조사를 취하여 문장의 여러 성분(주어, 목적어, 보어, 부사어 등)으로 사용된다. 이것은 단순한 명사와는 차이가 있다. 왜냐하면, 앞말에 대해서는 서술어적인 역할을 하고 뒤따르는 말에는 체언 역할을 한다. 그러나, 단순한 파생 명사는 체언 구실 밖에는 하지 못한다.

- 예문:  
“그의 걸음이 무척 빠르다.” (단순 파생 명사) “차도를 걸음은 교통 위반이다.” (용언의 명사형)
- 이 부류에 속하는 단어들:  
-기, -음

### 2.9.6 exa(adverbial ending, 부사형 전성 어미)

부사형 전성 어미도 다른 전성 어미와 마찬가지로 앞말에 대해서는 서술어적인 역할을 하고, 뒤따르는 말에는 부사어 역할을 하는 전성 어미이다.

- 예문:  
“그놈이 발음이 똑똑하게 말한다.”
- 이 부류에 속하는 단어들:  
-게,

### 2.9.7 efp(prefinal ending, 선어말 어미)

선어말 어미는 실질 형태소인 어간과 형식 형태소인 어말 어미 사이에 끼이어, 높임, 공손, 시제를 표시하는 어미이다.

- 예문:  
“순이가 과자를 먹었다.”(시제)  
“아버지는 조끼를 입으시었다.”(높임)  
“아버지께서는 만수무강하옵소서.”(공손)
- 이 부류에 속하는 단어들:  
-겠-(미래), -더-(회상), -시-(높임), -으시-(높임), -옵-(공손), -오-(공손), -았-(과거), -었-(과거), ...
- 참고사항:  
현재 시제의 선어말 어미 ‘-고’, ‘-는-’은 어말 어미에 포함한다. 예를 들면, ‘-다’, ‘-는다’, ‘-파’, ‘-는가’ 등은 어말 어미이다.

### 2.9.8 ef(final ending, 종결 어미)

종결 어미는 한 문장을 끝마치게 하는 어미로서 평서형, 감탄형, 의문형, 명령형, 청유형 어미가 있다.

- 예문:  
“철수는 과자를 먹는다.”(평서형)  
“철수는 과자를 먹는다.”(감탄형)  
“철수는 과자를 먹느냐.”(의문형)  
“철수야, 과자를 먹어라.”(명령형)  
“철수야, 과자를 먹자.”(청유형)
- 이 부류에 속하는 단어들:  
-게(명령형), -구나(감탄형), 까(의문형), -다(평서형), -나(의문형), -냐(의문형), -네(감탄형), -네(평서형), -느냐(의문형), -는가(의문형), -는구나(감탄형), -는구료(감탄형), -는구먼(감탄형), -는다(평서형), -는도다(감탄형), -다(평서형), -비까(의문형), -비다(평서형), -세(청유형), -습니까(의문형), -습니다(평서형), -십시



다 (칭유형), -십시오(명령형), -아라(명령형), -어라(명령형), -으나(의문형), -으십시다(칭유형), -으십시오(명령형), -은가(의문형), -읍니까(의문형), -읍니다(평서형), -자(칭유형), ...

- 참고사항:  
2.9.7절의 참고사항을 참조하기 바란다.

## 2.10 접사(x)

접사의 기능에는 크게 두 가지가 있다. 하나는 어근에 붙어 어근의 품사를 그대로 두고 뜻을 더하여 그 뜻을 한정하는 한정적 접사이고, 다른 하나는 어근에 붙어 어근의 품사를 바꿈으로써 품사를 파생하는 지배적 접사이다.

또한 접사의 위치에 따라 접미사와 접두사가 있는데 본 연구에서 접미사만 다루고, 접두사에 의해서 파생되는 모든 단어는 사전에 등록한다. 왜냐 하면 접두사는 뜻을 더하는 기능(한정적 접사)만 가지므로 접두사가 붙을 경우에 대부분의 단어는 새로운 개념을 가지므로 기계적으로 이들의 의미를 생성하기는 너무 어려운 일이다. 예를 들면, 접두사 ‘개-’(a false one)과 ‘떡’(rice-cake)이 어울려 ‘개떡’(bran cake)이라는 새로운 단어를 형성할 수 있으나, 전산 언어학적인 측면에서 이들을 관찰할 때 ‘개’와 ‘떡’의 의미를 더해서 ‘개떡’의 의미를 생성할 방법이 없다. 이외에도 ‘갓+스물’, ‘날+고기’, ‘맨+손’, ‘한+겨울’ 등이 있다. (본 연구에서 특별한 언급 없이 접사라고 하면 접미사를 의미한다.)

접미사는 한정적이며 지배적인 두 가지의 기능을 모두 가지고 있다. 본 연구에서는 한정적 접미사에 의해서 파생된 단어는 모두 사전에 등록하는 것을 원칙으로 한다. 다만 ‘들’(명사의 복수), ‘님’(호칭을 나타내는 명사의 높임) 등과 같은 대부분의 단어에 자유롭게 어울릴 수 있는 접미어는 한정적 접미사이더라도 별도로 구별한다.

어근과 접미사가 어울릴 때 어근이나 어미가 변할 경우에는 파생어를 사전에 등록한다(일종의 불규칙 현상이다). 예를 들면, ‘빠르 + 이 → 빨리’, ‘크 + 다랗다 → 커다랗다’ 등이다.

접사는 태깅을 할 때 이들을 하나의 품사로 간주하지 않고 파생된 단어를 하나의 덩어리로 간주한다.

### 2.10.1 xn(noun suffix, 명사 접미사)

명사 접미사는 명사에 붙어 주로 뜻을 더는 일을 하거나 굴곡 접사와 같이 말들의 사이의 관계를 표시하는 접미사이다.

- 예문:  
“사람들이 많이 모여 있다.”  
“김씨는 그림을 잘 그린다.”
- 이 부류에 속하는 단어들:  
-군, -님, -들, -씨, -적, ...
- 참고사항:

1. 명사 파생 접미사에 의해서 파생된 대부분의 파생어는 사전에 등록한다. 그러나, 앞에서 언급한 것처럼 큰 제약없이 명사과 잘 어울리는 접미어만 별도로 취급한다. 이들에 의해서 파생되는 파생어는 일반적인 한국어 사전에도 올라 있지 않다.
2. 명사 접미사 ‘-적’에 의해서 파생된 단어는 명사이면 관형사이다. 이것은 문맥 정보에 의해서 해결되어야 한다. 즉, 조사를 동반할 경우에는 명사이고 그렇지 않을 경우에는 관형사이다. 다른 명사 접미사와 좀 다른 형태로 처리되어야 한다.
3. ‘국제’, ‘획기’ 등과 같이 스스로는 명사적 구실을 하지 않고 명사 접미사 ‘-적’과 잘 어울리는 단어는 파생어를 사전에 등록한다.

### 2.10.2 xpv(verb-derived suffix, 동사 파생 접미사)

동사 파생 접미사는 명사, 형용사의 어근과 부사에 붙어서 뜻을 더하지 않고 단지 품사를 동사로 변경시키는 접미사이다.

- 예문:

“교통 규칙을 위반하는 사람이 해가 갈수록 줄어든다.”

“한 반은 60명으로 구성된다.”

- 이 부류에 속하는 단어들:

- 동작성 명사 + -하- → 동사(자동사, 타동사)  
가동하다, 가입하다, 결합하다, 경쟁하다, 공급하다,
- 동작성 명사 + -되- → 동사(자동사) 가동되다, 가입되다, 결합되다, 경쟁되다, 공급되다,
- 동작성 명사 + -시키- → 동사(사역 동사)[5]  
가동시키다, 가입시키다, 결합시키다, 경쟁시키다, 공급시키다,

- 참고사항:

1. 동사의 어근에 붙어 어근의 뜻을 제한하는 피동 접미사, 사동 접미사 등은 한정적 접미사로서 본 연구에서는 제외한다.
2. 피동 및 사동 접미사(이,히,리,기 등)은 형태 자체에서 모호성을 가지고 있다. 예를 들면, ‘끌리다’는 ‘차를 끌다’에서 ‘끌다’의 사역 동사이고 ‘인기를 끌다’에서 ‘끌다’의 피동 동사이다. 이와 같은 모호성은 어떤 기계적인 방법에 방법에 의해서 쉽게 해결될 수 있을 것으로 보지는 않는다. 따라서 사동 및 피동에 의한 사/피동사는 사전에 등록하고 개별적인 정보를 추가적으로 부여한다. 예를 들면, ‘끌리다’는 사동사이고 피동사임을 명시한다.
3. 첩어성을 띤 부사에 붙는 동사 파생 접미사 -거리다, -이다 등에 의해서 파생되는 파생어는 사전에 등록한다. 예를 들면, 바둥거리다, 철렁거리다, 킁킁거리다, 출렁이다 등이 있다.

### 2.10.3 xpa(adjective-derived suffix, 형용사 파생 접미사)

형용사 파생 접미사는 명사, 동사의 어근, 관형사와 부사에 붙어서 뜻을 더하지 않고 단지 품사를 형용사로 변경시키는 접미사이다.

- 예문:

“그는 가난**하**집안에서 자랐다.”

“그는 학생**답**게 행동한다.”

- 이 부류에 속하는 단어들:

- 상태성 보통명사 + **하** → 형용사  
가난**하**다, 간단**하**다, 중요**하**다, 총명**하**다, 희미**하**다, ...
- 명사 + **답** → 형용사  
남자**답**다, 신사**답**다, 아이**답**다, 어른**답**다, 학생**답**다,
- 명사 + **스럽** → 형용사  
사랑**스럽**다, 어른**스럽**다, 정성**스럽**다, ...
- (추상성)명사 + **롭** → 형용사  
공교**롭**다, 슬기**롭**다, ...

- 참고사항:

1. 상태성 보통명사에 대한 설명은 2.3.3을 참조바란다.
2. 형용사 어근에 붙어 그 어근의 뜻을 제한하는 형용사적 한정 접미사 -**다랄다** 등은 여기서 제외한다.
3. 관형사에 붙는 형용사 파생 접미사 -**롭다** 등에 파생되는 파생어는 사전에 등록한다. 예를 들면, 새**롭다** 등이다.
4. 첩어성을 띤 부사에 붙는 형용사 파생 접미사 -**하다** 등에 의해서 파생되는 파생어는 사전에 등록한다. 예를 들면, 울긋불긋**하다**, 차근차근**하다** 등이 있다.

### 2.10.4 xa(adverb-derived suffix, 부사 파생 접미사)

부사 파생 접미사는 명사, 형용사의 어근, 동사의 어근, 관형사에 붙어서 뜻을 더하지 않고 단지 품사를 부사로 변경시키는 접미사이다.

- 예문:

“그는 조용**하**게 책을 읽었다.”

“이 문제를 더 간단**히** 풀어라.”

- 이 부류에 속하는 단어들:

- 상태성 보통 명사+ **히**, **이** → 부사  
‘간단**히**’, ‘감사**히**’, ...

- 참고사항:

부사 파생 접미사는 위에서 언급한 것 이외에는 대체로 사전에 등록한다.

### 3 토의 및 결론

본 연구에서는 한국어 말뭉치로부터 정확한 통계정보를 추출하기 위한 일환으로 한국어 품사 태그를 결정하였다. 이를 선정하기 위해 사전 연구로써 한국어 처리에서 사용하고 있는 몇몇 품사 분류[3, 7, 9, 12]와 외국의 경우 흔히 널리 사용되고 있는 Penn TreeBank[14], Brown 말뭉치[15], LOB 말뭉치[15]에서 사용하는 태그와 일본 EDR의 품사 분류[13] 등이 검토·분석되었다.

본 연구에서 제안된 한국어의 품사 체계가 완전한 것은 아니라고 믿는다. 적용시키는 응용에 따라 더 세분화시킬 필요도 있고, 때로는 더 일반화시킬 필요도 있을 것이다. 다만, 전산 언어학적으로 한국어 정보처리에 필요한 여러 보편적인 품사정보에 관한 국내외의 관련 연구를 참조하여 나름대로 정리해 보았다. 이 품사들은 현재 계산 언어학의 중요한 연구 방법으로 떠오르는 말뭉치를 이용한 통계적 언어처리에 가장 기본이 되는 필수적인 정보로서 이 정보가 부가된 말뭉치(tagged corpus)를 구축해 나가기 위한 첫 단계로 신중하게 결정된 것이다.

실질적으로 이 품사 태그가 유용한지를 검토하는 과정으로 약 10만 여 단어의 말뭉치에 대해서 손으로 태깅을 해 보았다. 그 결과 큰 문제는 발견할 수 없었다. 그러나, 아직도 더 많은 실험을 거쳐야만이 얼마나 유용한지를 평가할 수 있을 것으로 예측되며, 여러 형태의 실험 과정을 거쳐 이를 수정·보완할 계획이다.

## 참고 문헌

- [1] 강승식, **음절 정보와 복수어 단위 정보를 이용한 한국어 형태소 분석**, 서울대학교 컴퓨터 공학과 박사학위 논문, 1993.
- [2] 김재훈, 김철호, 최기선, 김길창, “PIVOT 방식을 이용한 기계번역에서 한국어 격조사 설정과 중간언어로부터 조사 생성,” **한국정보과학회 가을학술발표 논문집**, 제14권, 제2호, pp. 216-219, 1987. 10.
- [3] 김재훈, **한국어 말뭉치로부터 통계 정보를 위한 한국어 품사 태그의 결정**, 한국과학기술원, 전산학과, “자연언어 처리 특강 - 문서 처리 - ” 과목의 term project, 컴퓨터 실험실 내부 메모, 1993년, 7월 19일.
- [4] 김희보, **한글 바로까**, 종로서적, 1987.
- [5] 남기심, 고영근, **표준 국어 문법론**, 탐출판사, 1987.
- [6] 이운재, **한국어 문서 태깅 시스템의 설계 및 구현**, 한국과학기술원, 전산학과, 석사학위 논문, 1993.
- [7] 이성진, **Two-level 한국어 형태소 해석**, 한국과학기술원 석사학위 논문, 1992.
- [8] 이희승, 안병희, **한글 맞춤법 강의 - 한글 맞춤법/표준어 규정/외래어 표기법 -**, 신구문화사, 1989.
- [9] 장석진, **한국어 문법 - NLP를 위한 HPSG/K -**, 한국과학기술원 인공지능연구센터 기술보고서(CAIR-TR-92-33), 1992.
- [10] 조규빈, **하이라이트 고교문법**,지 학사, 1986.
- [11] 한국과학기술원, **한국어 특질 및 대화체 기계번역에 관한 연구**, 한국전기통신 공사 연구개발단 최종보고서, 1991.
- [12] 한국과학기술원, **언어 모델을 위한 한국어 Text DB 구성에 관한 연구**, 한국전자통신연구소 최종보고서, 1993.
- [13] EDR, *EDR Electronic Dictionary Technical Guide*, Japan Electronic Dictionary research Institute Ltd., 1993.
- [14] M. P. Marcus, M. A. Marcinkiewicz, and B. Santorini, “Building Very Large Natural Language Corpora: the Penn Treebank,” *Computatiional Linguistics*, vol. 19, no. 2, pp. 313-330.
- [15] R. Garside, G. Leech, and G. Sampson, *The Computational Analysis of English: a Corpus-Based Approach*, LONGMAN, 1987.