

# 한국어 명사 의미 부류 체계의 구축과 활용

강범모 박동호 이성현 박진호  
고려대학교 서울대학교 (주)언어과학  
bmkang@mail.korea.ac.kr tilri24@snu.ac.kr  
lsh0717@hananet.net synpjh@eoneo.co.kr

## Construction and application of semantic classes of Korean nouns

Beom-mo Kang Dong-Ho Pak Seong-Heon Lee Jin-ho Park  
Korea University Seoul National University Eoneo, Inc.

### 요 약

명사 의미 부류 체계는 언어 처리의 다양한 분야에서 그 필요성이 부각되고 있다. 예를 들어, 기계 번역에 있어서의 단어 의미의 중의성 해소(word sense disambiguation), 정보검색 시스템에서도 재현율과 정확률의 향상, 추론 시스템 등을 위하여 명사 의미 부류는 중요한 역할을 한다.

명사 의미 부류 체계의 이러한 중요성 때문에 여러 온톨로지(ontology)가 기존에 구축되어 있다. 그런데 이러한 온톨로지들은 대개 순수한 개념적 기준에 입각한 것이며 단어의 통사적 특성을 별로 고려하고 있지 않다. 정보검색 시스템이나 추론 시스템의 경우에는 통사적 고려가 별로 중요하지 않을 수 있으나 기계번역의 경우 통사적 특성에 대한 고려가 매우 중요하다.

이러한 점에 주목하여 21세기 세종계획 전자사전 분과에서는 개념적 기준과 통사적 기준을 모두 고려하여 명사 의미 부류 체계를 구축하고 있다. 즉, 해당 부류에 속하는 명사들이 결합할 수 있는 술어(적정 술어) 등의 통사적 요인을 중요시하여 명사들을 분류하고 있는 것이다. 이에 따라 세종 체언 사전의 모든 명사들에 대해 의미부류 정보가 주어지고, 용언 사전의 용언의 각 논항에 대한 선택제약 정보도 이 명사 의미 부류 체계를 이용하여 제시되고 있다. 이러한 정보들은 한국어 처리에 중요한 자료로 이용될 것이다.

## 1. 명사 의미 부류 체계의 필요성

명사 의미 부류 체계의 필요성은 언어 처리의 여러 분야에서 부각되고 있다.

우선 단어 의미의 중의성 해소(word sense disambiguation)를 위하여 명사 의미 부류가 중요한 단서가 되는 일이 많이 있다. 이 문제는 기계 번역에서 특히 중요하게 제기된다. 예를 들어 영어의 put on 이나 wear를 한국어로 번역할 때 '입다', '쓰다', '신다', '끼다', '걸치다' 등 중에서 어느 것을 선택할지는 목적어가 어떤 명사 부류에 속하느냐에 따라 결정된다. 일본어의 예를 들면, 동사 'つける'의 default 의미는 '붙이다'이지만 전기제품을 가리키는 명사가 목적어로 올 때에는 '켜다'가 된다. 또한 일본어의 조사 'に'를 한국어로 번역할 때 '-에'로 할지 '-에게'로 할지 역시 선행 명사의 의미 부류에 따라 결정된다.

정보검색 시스템에서도 재현율과 정확률의 향상을 위하여 명사 의미 부류 정보가 이용될 수 있다. 예를 들어, 사용자가 제시한 질의어를 포함한 문서뿐 아니라 이와 동의 관계에 있는 단어나 기타 관련 단어를 포함한 문서도 사용자에게 유용할 수 있다(질의어 확장). 반면 사용자가 제시한 질의어를 포함한 문서가 너무 많을 때에는 질의어의 하위 부류에 속하는 단어들을 사용자에게 제시하여 범위를 좁히도록 유도할 수도 있다(질의어 정교화).

추론 시스템에서도 명사 의미 부류 체계의 중요성은 명확하다. '철수가 고등어를 먹었다'라는 문장이 '철수가 생선을 먹었다'라는 문장을 함의(entail)하는 것은 '고등어'가 생선이라는 의미 부류에 속한다는 것으로 포착할 수밖에 없다.

명사 의미 부류 체계의 이러한 중요성 때문에 세계 여러 기업, 학교, 연구소에서 소위 온톨로지(ontology)를 구축하고 있다. 그런데 이러한 온톨로지들은 대개 순수한 개념적 기준에 입각해서 구축되고 있으며 단어의 통사적 특성을 별로 고려하지 않고 있다. 정보검색 시스템이나 추론 시스템의 경우에는 통사적 고려가 별로 중요하지 않을 수 있으나 기계번역의 경우 통사적 특성에 대한 고려가 매우 중요하다. 이러한 점에 주목하여 21세기 세종계획 전자사전 분과에서는 개념적 기준과 통사적 기준을 모두 고려하여 명사 의미 부류 체

계를 구축하고 있다. 프랑스 LLI(Laboratoire de Linguistique Informatique)의 Gaston Gross 교수가 주축이 되어 연구하고 있는 대상부류(classes d'objet)의 방법론을 이용하여, 해당 부류에 속하는 명사들이 결합할 수 있는 술어(적정 술어) 등의 통사적 요인을 중요시하여 명사들을 분류하고 있는 것이다. 이에 따라 체언 사전의 모든 명사들에 대해 의미부류 정보가 주어지고, 용언 사전의 용언의 각 논항에 대한 선택제약 정보도 이 명사 의미 부류 체계를 이용하여 제시되고 있다. 이러한 정보들이 축적되면 한국어 처리에 중요한 자료로 이용될 것이다.

## 2. 명사 의미 부류 체계 구축 방법론

### 2.1. 개념적 접근법과 언어학적 접근법

지금까지 온톨로지는 주로 개념적 기준에 따라 구축되어 왔다. 개념적 접근법 자체가 잘못된 것은 아니지만, 이 접근법에만 의존할 때 해결하기 어려운 문제들이 있다. 그 중 하나가 교차 분류(cross-classification)의 문제이다. 예를 들어, 악기라는 부류를 상정했다고 가정하고 이 부류를 어떻게 하위분류할 것인가 생각해 보자. 한 가지 방법은 연주 방법을 기준으로 하여 현악기, 관악기, 타악기 등으로 분류하는 것일 것이고, 다른 방법은 악기의 유래를 기준으로 하여 서양악기, 동양악기 등으로 분류하는 것일 것이다. 그런데 서양악기도 다시 연주 방법에 따라 현악기, 관악기, 타악기 등으로 하위분류할 수 있고, 동양악기도 마찬가지다. 반면에, 현악기에도 서양악기와 동양악기가 있고 관악기, 타악기도 마찬가지다. 즉 연주 방법이라는 기준과 유래라는 기준은 서로 독립적이어서 두 기준이 맞물려서 악기의 많은 하위부류를 낳는데, 두 기준 중 어느 기준을 우선시할 것인가가 문제인 것이다. 이와 같이 개념적인 면만 고려하면 어느 기준이 더 중요하다고 판단하기 어려운 경우가 많다. 이럴 경우, 언어학적 기준이 도움이 될 수 있다. 한국어에는 연주를 의미하는 동사가 세분되어 있는데('켜다', '붙다', '치다', ...), 목적어 명사가 현악기, 관악기, 타악기나에 따라 동사가 달라지지만 서양악기, 동양악기나에 따라 달라지지는 않는다. 따라서 연주 방법의 기

준이 유래라는 기준보다 언어학적으로 더 중요함을 알 수 있다.

모든 분류 체계가 그렇듯이 온톨로지도 어디까지 나눌 것인가의 문제가 있다. 개념적인 기준만 고려한다면 부류들을 계속 세분해 나가다가 어디에서 멈출 것인지 결정하기 어려울 수 있다. 예를 들어, 교통기관, 무기, 신발 같은 부류를 더 세분할 것인가 말 것인가? 개념적 측면에서 보면 이들 모두 세분할 근거가 충분히 있다. 그러나 언어학적 측면에서 보면 차이가 있다. 교통기관의 경우 '이륙하다', '착륙하다', '발사하다' 등의 동사는 공중 교통기관 명사와만, '정박하다', '표류하다', '좌초되다' 등의 동사는 수상 교통기관 명사와만 결합할 수 있기 때문에 교통기관 명사를 세분할 필요가 있다. 무기의 경우에도 '쏘다'는 총포류나 활과는 결합할 수 있지만 칼, 창 부류와는 결합할 수 없다. 반면에 '찌르다'는 후자와만 결합한다. 따라서 무기류도 세분할 근거가 있다. 신발은 개념적으로는 구두, 운동화, 샌들, 장화 등으로 세분할 수 있을지는 몰라도 언어학적 근거를 찾기는 어렵다. 요컨대, 언어학적 근거는 어디까지 세분할 것인가 하는 문제에 대해 유용한 지침을 제공한다.

## 2.2. 다중 계승

모든 분류 체계를 따라다니는 골치 아픈 문제 중의 하나가 다중 계승(multiple inheritance)의 문제다. 개념적 깔끔함을 위해서나 실제적 이용을 위해서나 분류 체계는 tree 구조를 갖는 것이 좋다. 즉, 하위 노드(child node)는 단 하나의 상위 노드(parent node)만을 갖도록 하는 것이다. 이런 제약을 두어야만, '사과는 과일이 다', '과일과 채소는 동위 노드(sister node)이다'라는 사실들로부터 '사과는 채소가 아니다'라는 결론을 도출해 내는 식의 추론이 가능하다. 그러나, 실제로 많은 명사들을 관찰해 보면 하나의 명사가 둘 이상의 부류에 속하는 것처럼 보이는 경우가 많이 있다. 이것은 개념적 측면뿐 아니라 언어적 측면에서도 드러난다. 대표적인 예로 텍스트, 기관 등을 들 수 있다. 책, 신문, 잡지, 소설, 시 등의 텍스트 명사는 '찢다', '태우다' 등의 동사와 결합할 수 있는 점을 고려하면 구체물 명사처럼 행동하지만 '책을 쓰다', '시를 짓다', '소설

을 구상하다'와 같은 예를 보면 추상적 대상의 속성도 지닌다. 학교, 은행, 병원, 대사관 같은 기관 명사도 '은행이 친절하다', '미 대사관에서 성명을 발표했다'와 같은 예를 보면 인간 내지 인간들의 집단처럼 행동하지만, '학교에 가다', '병원에 있다'와 같은 예에서는 장소 명사처럼 행동한다. 이런 현상은 특정 명사에 한정된 특이성이 아니라 텍스트 명사와 기관 명사 전반에 걸쳐, 그리고 다른 많은 부류에서 두루 나타나는 일반적인 현상이다. 이런 현상을 처리하는 방식에는 여러 가지가 있을 수 있다. 현상 자체를 중시하여 온톨로지가 나무(tree) 구조가 되는 것을 포기하고 격자(lattice) 구조를 채택할 수도 있다. 그럴 경우 나무 구조에 비해 자료를 처리하고 이용하는 방식이 훨씬 복잡해질 것이다. 반대로, 자료 처리와 이용의 편의를 중시하여 나무 구조를 유지하고 다중 속성을 지닌 부류를 어느 한 쪽으로 몰아서 처리하는 방법도 있을 수 있다. 그런데 이 경우, 해당 부류가 편의상 어느 한 쪽으로 처리되기는 했어도, 다른 부류의 속성도 지니고 있음을 특별한 기제로 명시해 주고, 나중에 자료 이용시에 이 정보를 참고할 수 있어야 할 것이다. 세종 전자 사전의 명사 의미 부류 체계는 후자의 방법을 채택하고 있다.

## 2.3. 적정 술어

예를 들어 인간 명사 의미 부류가 주어 위치에 올 때 쓸 수 있는 술어는 '이야기하다', '말하다', '생각하다', '쓰다' 등이며 목적어 자리에 올 때 쓸 수 있는 술어는 '격려하다', '설득하다' 등이다. 이 술어들은 은유적 표현이나 문체적 효과의 경우를 제외하고는 인간 의미 부류를 효과적으로 특징지어 준다. 이런 술어들을 적정술어라 한다. 동일한 방식으로 '밀', '포도', '오이' 같은 명사의 식물 의미 부류도 적정술어를 사용하여 형식적으로 특징지을 수 있다. '씩이 나다', '말라죽다' 등의 술어의 주어 자리에, 그리고 '재배하다' 등의 동사의 목적어 자리에 오는 명사가 식물의 의미 부류에 속하는 것이다.

## 3. 세종 사전의 명사 의미 부류 체계

세종 전자 사전의 의미 체계의 일부를

보이면 다음과 같다<sup>1)</sup>. 우선 최상위 층위는 구체자연물, 구체인공물, 관계구체물, 집단, 장소, 추상적 대상, 상태, 사건|현상, 행위의 일급 가지로 구성이 되어 있다. 다음은 최상위 층위들과 그 각각의 주요한 하위 층위를 보여주고 있다.

**top-level**

- <1:구체자연물>
  - <2:무생물> 석탄
  - <2:생물> 소나무
  - <2:식품> 배추
- <1:구체인공물>
  - <2:신체작용물> 모자
  - <2:교통기관> 버스
  - <2:무기> 총
  - <2:음식> 떡
  - .....
- <1:관계구체물>
  - <2:부분> 손잡이
  - <2:재료|성분> 원료
  - <2:속성구체물> 고물
- <1:집단>
  - <2:인간집단> 회사
  - .....
- <1:장소>
  - <2:자연적 장소> 그늘
  - <2:인공적 장소> 마을
  - <2:관계 장소> 안
  - <2:속성 장소> 어촌
- <1:추상적 대상>
  - <2:시간> 때
  - <2:관계> 상이
  - <2:역할> 노릇
  - <2:범주> 부류
  - <2:속성> 모양
  - .....
- <1:상태>
  - <2:감정> 연민
- <1:사건|현상>
  - <2:사고> 화재
  - <2:사회갈등> 폭동
  - <2:기상현상> 소나기

1) 이 논문에서 제시하는 명사 의미 부류 체계는 현재까지 구축된 세종전자사전의 항목들에 표시된 의미부류 체계와 똑같지 않다. 후자의 의미부류를 확장, 세분, 정교화한 것이 이 논문에서 제시하는 의미 부류 체계이다. 그리고 여기서 제시하는 체계가 전체적인 윤곽을 갖춘 어느 정도 완성된 것이지만, 세부적인 조정을 거칠 가능성이 있다.

<2:상태변화> 증가

.....

<1:행위>

- <2:물리적 행위> 땀뻘기
- <2:인지적|추상적 행위> 생각
- <2:언어행위> 대화
- <2:감정행위> 질투

최상위 층위의 9개 부류는 크게 보면, 주로 서술명사(predicative noun)의 의미를 분류하는데 사용되는 상태(‘슬픔’), 사건|현상(‘바람’), 행위(‘공부’)들과, 비서술명사의 의미를 분류하는 데 사용되는 나머지 부류로 대별된다. 비서술명사 명사들 중에는 여러 가지 종류의 추상적 대상 의미를 갖는 것들과 비추상적 대상의 것들이 있다. 추상적 대상은, 위의 표에서 모두 명시하지는 않았지만 시간, 관계, 역할, 범주, 속성, 방향, 수학적 대상, 생각, 기호 등 수십 가지의 하위 부류를 가지는 것으로 상정되어 있다. 이와 대조적으로 구체적 대상의 의미 부류는 넓게 퍼진 형태가 아니라 단계적 묶음이 가능한 깊이가 깊은 구조이다. 원래 최상위 층위를 구체물로 상정하고 그 하위 부류를 자연물, 인공물, 관계구체물로 세분하려고 하였으나, 의미 부류 체계의 전체적 체계를 고려할 때 구체물 부류의 지나친 깊이(depth)를 완화하고자 구체자연물, 구체인공물, 관계구체물을 최상위 층위로 채택하게 되었다. 그렇게 하더라도 전체적으로 구체물에 나타는 최대 6단계의 층위는 다른 부류의 최대 4단계 층위보다는 깊다. 집단 부류 중 인간집단(‘국가’, ‘민족’, ‘부부’)은 인간의 행위를 나타내는 술어(‘생각하다’, ‘일하다’, ...) 등과 결합할 수 있다는 점에서는 생물로서의 개체 인간과 공통점을 일부 가지고 있으나 생명체의 생로병사를 나타내는 술어(‘죽다’, ‘아프다’, ...) 등과는 자연스럽게 결합하지 못한다는 점에서 집단의 하위부류로 배치되었다. 또한 인간 집단은 대부분의 경우, 장소의 의미를 갖는다.

<1:구체인공물>의 하위부류인 <2:교통기관>은 다음과 같이 세분된다.

<2:교통기관>

- 적정 술어: 타다, 내리다
- <3:육상 교통기관>
  - <4:자가용 교통기관>
    - 오토바이, 대형차, 트럭
  - 적정 술어: 몰다
  - <4:대중 교통기관>

적정 술어: 운행하다,  
 — 노선, — 정거장, — 정류장,  
 — 종점

<5:도로 대중 교통기관>  
 버스, 택시

<5:철도 교통기관>  
 기차, 열차, 지하철  
 적정 술어: 탈선하다  
 — 역, — 종착역

<3:수상 교통기관>  
 잠수함, 잠수정, 여객선, 화물선, 유조선,  
 쾌속정; 전함, 군함, 구축함; 탐사선  
 적정 술어: 정박하다, 표류하다, 좌초하다,  
 좌초되다, 승선하다, 항해하다

<3:공중 교통기관>  
 비행선, 우주선, 탐사선, 스페이스셔틀,  
 인공위성, 비행기, 여객기, 전투기, 수송기,  
 전용기, 제트기, F16, 미그기  
 적정 술어: 운항하다, 이륙하다, 착륙하다,  
 분시착하다, 탑승하다, 격추하다

교통기관은 적정 술어에 따라 우선 육상, 수상, 공중 교통기관으로 나뉜다. 육상 교통기관은 자가용 교통기관과 대중 교통기관으로 나뉜다. 대중 교통기관은 다시 도로 대중 교통기관과 철도 교통기관으로 세분된다.

#### 4. 세종 사전에서 명사 의미 부류 정보를 표상하는 방법

세종 전자 사전의 체언 사전에서는 각 표제어에 대해 뜻(sense) 구분을 하고, 각각의 뜻에 대해 의미 부류 정보를 제공하고 있다. 뜻을 구분하는 독자적인 기준이 마련되어 있지만, 의미 부류 체계 자체도 뜻 구분을 위한 단서를 제공할 수 있다. 즉, 두 뜻 후보가 각각 서로 다른 의미 부류에 속한다면 이 둘을 별개의 뜻으로 나눈다는 것이다.

```
<superEntry>
●가자미
<entry>
  %writer={}
  %dic=[]
  %crd=[]
  %mdd=[]
  %note=[]
  <toplevel>
    @form={가자미}
    @pos={nn}
    @see=[]
    <morph_a>
      @var={xs={}; xd={}; xx={가재미}}
      @abb=[]
      @lng=[]
      @str=[]
```

```
@org=[]
</morph_a>
</morph_b>
@hom=[]
@der=[]
@comp={n}~식해|것|눈]
@metc =[]
</morph_b>
</toplevel>
<sense n=1>
@eg={이번 낚시에서 가자미가 많이 잡
했다]
@trans={flatfish}
@domain=[]
@reg=[]
@con=[]
@kurs={C]
@sem={동물}
</lr>
@syn=[]
@ant=[]
@hyper={물고기}
@hypon=[]
@coord=[]
@holo=[]
@mero=[]
@rel=[]
</lr>
<synt_a>
  @cl={uni=마리;grp=;div=;qnt=}
  @prt=[]
  @av=[]
  @ds=[]
</synt_a>
<synt_b>
  @comb_aj=[]
  @magn=[]
  @comb_v={~를 잡다|낚시하다]
  @comb_n=[]
```