

# 자연언어 처리를 위한 한국어 동사·명사의 개념 분류

○  
박 영 자 , 송 만 석  
연세 대학교 전산과학과

The Classification of Korean Noun and Verb for Natural Language Processing

Young-Ja Park , Man-Seok Song

Department of Computer Science, Yonsei University

## 요 약

본 논문에서는 자연언어 처리 시스템에서 사용할 수 있는 단어의 개념 정보에 대해 연구한다. 그러기 위해 언어의 실생활에서의 쓰임 - 연세 대학교 사전 편찬실의 연세 발음치 -을 바탕으로 한국어의 동사와 명사에 대해 개념을 조사, 분류하였으며 그 개념들이 한국어 문장 분석기에 어떻게 이용되는가를 보인다.

## I. 서론

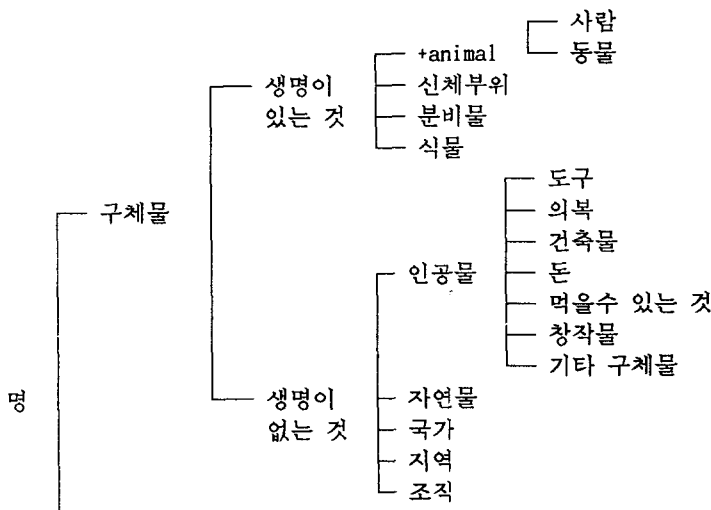
자연언어는 많은 단어들이 중의성을 가지기 때문에 컴퓨터로 자연언어를 처리하는 것은 매우 어렵다. 그러나 인간이 자연언어를 이해할 때는 이러한 중의성이 문제가 되지 않는다. 그것은 인간은 단어를 하나씩 읽을 때 그 단어에 대한 많은 개념 정보를 갖고 있고 그 단어의 개념이 앞 뒤의 다른 단어들의 개념 정보들과 상호 연관작용을 통해 중의성이 자연스럽게 해결되기 때문이다. 그러므로 컴퓨터가 자연언어를 처리할

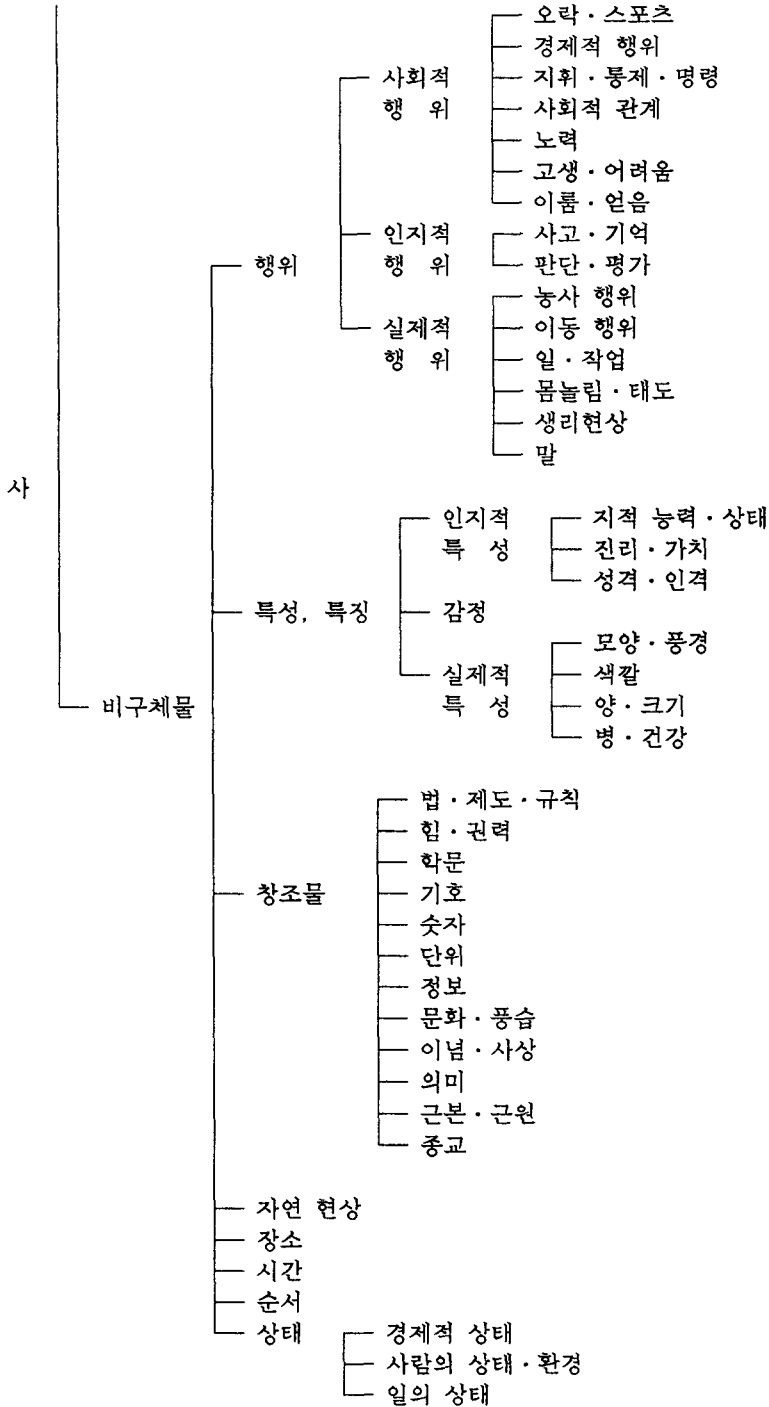
때도 각 단어에 대한 개념 정보를 이용한다면 자연언어의 어휘 중의성 해결이 가능하며 구문 분석이 효율적으로 수행될 수 있다.

최근에 와서 자연언어 처리에 있어서 사전의 중요성이 강조되고 있다. 컴퓨터가 인간이 사용하는 언어를 이해하기 위해서는 그 언어에 관한 지식이 필요하고 이러한 지식은 사전을 통해 컴퓨터에게 제공된다. 그러므로 사전은 처리면에서 효율적으로 구성되어야 하며 내용면에서 언어를 분석, 이해하는데 필요한 많은 구문적, 개념적 정보를 제공하여야 한다. 언어의 구문적 정보에 대해서는 언어학, 전산학에서 많은 연구가 있어 왔으나 개념적 정보에 대한 연구는 많이 부족한 실정이다. 본 논문에서는 한국어 동사와 명사에 대해 자연언어 처리 시스템에서 이용할 수 있도록 개념을 분류하였다.

## I. 명사 개념 분류

본 논문에서는 연세 대학교 한국어 사전 편찬실에서 수집한 말뭉치 가운데 국민학교, 중학교 국어 교과서에서 발췌한 문장에 나타난 명사 1,200개를 대상으로 하였다. 품사가 명사인 한 단어는 보통 여러가지 개념을 가진다. 예를 들면 아버지란 단어는 {사람}이라는 개념이 가질 수 있는 모든 개념을 다 가지는 {가족관계}에 있는 사람이다. 이와 같이 모든 명사들은 잘 검토해 보면 어떤 단어는 좀 더 일반적 개념을 가지고 또 어떤 단어는 좀 더 특별한 개념을 가진다. 그래서 명사 분류는 개념 계층 관계를 가지게 된다. 본 논문에서는 최상위 2개 개념에서 출발하여 최하위 85개의 개념으로 구성된다. 명사의 개념들은 다음과 같다.





<그림 1> 명사 개념 분류표

# 1. 명사 개념 정의

- (1) {사람} : 이 개념은 사람을 나타내는 명사들로 구성되며 4개의 하위범주로 구성된다. 하위범주는 {개인}, {여러사람}, {직업, 신분}, {가족관계} 등이다.
- (2) {동물} : 사람을 제외한 모든 동물이 이 범주에 속한다. {날 수 있는 동물}, {수중동물}, {육지동물}의 3개의 하위범주로 구성된다.
- (3) {신체부위} : 동물의 신체 일부분을 나타내는 명사로 구성되며 {움직일 수 있는 부위}, {움직일 수 없는 부위}의 하위범주가 있다.
- (4) {분비물} : 동물에 의해 분비되는 물질을 나타낸다.
- (5) {식물} : {꽃}, {나무}, {과일}, {농작물}등의 4개의 하위범주로 이루어 진다.
- (6) {도구} : 인간의 실생활에서 사용되는 도구들이 포함되는 범주이다. 이 범주에는 {교통수단}, {전쟁도구}, {건축도구}, {오락도구}등 여러가지 하위범주로 구성된다.
- (7) {의복} : 인간의 몸에 착용되는 모든 물건들이다.
- (8) {건축물} : 인간이 만든 건물을 의미한다.
- (9) {돈} : 각종 화폐가 여기에 속한다.
- (10) {음식} : 먹을 수 있는 것을 의미하는 명사들로서 {마실 수 있는 것}과 {기타 음식}으로 하위 분류된다.
- (11) {창작물} : 미술품, 시집 같은 예술 활동으로 인해 생성된 구체물이 여기에 속한다.
- (12) {기타 구체물} : 인간이 만든 인공물 중에서 위의 다른 범주에 포함되지 않은 명사들이 여기에 속한다.
- (13) {자연물} : 자연을 구성하는 물질로서 {천체}, {지형}, {구성물}로 구성된다
- (14) {국가} : 나라 이름을 나타내는 명사들이 이 범주에 포함된다.
- (15) {지역} : 국가에 속하는 지역, 지방을 의미하는 명사들이 포함된다.
- (16) {조직} : 조직, 기관을 나타내는 명사들이 이 범주에 속한다.
- (17) {행위} : 명사 중에서 활동이나 행동을 나타내는 명사들이 이 범주에 속하며 크게 {사회적 행위}, {인지적 행위}, {실제적 행위} 3개의 하위범주로 구성된다.
- (18) {사회적 행위} : 인간의 사회 활동에서 발생하는 행위 명사들로서 {오락, 스포츠}, {경제적 행위}, {지휘, 통제, 명령}, {사회적 관계}등의 하위범주로 이루어 진다.
- (19) {인지적 행위} : 인간의 사고 활동이나 판단 행위를 나타내는 명사로서 {사고, 기억}, {판단, 평가}의 하위범주가 있다.
- (20) {실제적 행위} : 동물이 실제 자신의 신체를 움직이는 행위를 의미하며 {농사활

등), {이동행위}, {일, 작업}, {태도, 몸놀림}, {생리현상}등의 하위범주로 구성된다.

(21) {특성, 특징} : 이 범주는 {구체물}이 가지는 여러가지 속성이나 특징을 나타내는 명사들로 이루어지며 크게 {인지적 특성}, {감정}, {실제적 특성} 3개의 하위범주로 이루어 진다.

(22) {인지적 특성} : 인간의 지적 능력이나 상태를 의미하는 명사들로서 {지적 능력, 상태}, {진리, 가치}, {성격, 인격} 등의 하위범주가 있다.

(23) {감정} : 인간의 감정 상태를 나타내는 명사들로 이루어 진다.

(24) {실제적 특징} : 구체물의 외적 특성을 나타내는 것으로 {모양}, {색깔}, {병, 건강}, {양, 크기} 등의 하위범주를 가진다.

(25) {창조물} : 인간의 사회 생활에서 필요에 의해 생성된 법, 제도, 전통, 종교등 추상물이 여기에 속한다. {법, 제도, 규칙}, {힘, 권력}, {학문}, {문화, 풍습}, {정보}, {종교} 등의 하위범주가 있다.

(26) {자연 현상} : 자연 현상을 나타내는 명사들이 포함된다.

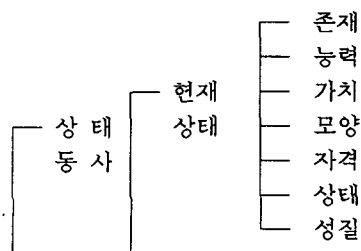
(27) {장소} : 방향이나 추상적 장소를 나타낸다.

(28) {시간} : 계절, 시간, 시대를 나타내는 명사들로 구성된다.

(29) {상태} : 상황이나 환경, 상태를 나타내는 명사들이 이 범주에 속하며 {경제적 상태}, {사람의 상태, 환경}, {일의 상태} 등의 하위범주로 구성된다.

### Ⅲ. 동사 개념 분류

본 논문에서는 연세 대학교 한국어 사전 편찬실에서 수집한 말뭉치 I에서 사용 빈도가 높은 동사 1,000개를 대상으로 하였다. 이 동사들은 연세 말뭉치 I에 나타난 전체 동사 사용의 90.05%를 차지한다. 동사의 개념도 역시 개념 계층 관계를 가지게 된다. 본 논문에서는 최상위 2개 개념에서 출발하여 최하위 76개의 개념으로 구성된다. 동사의 개념들은 다음과 같다.



어

상태  
변화

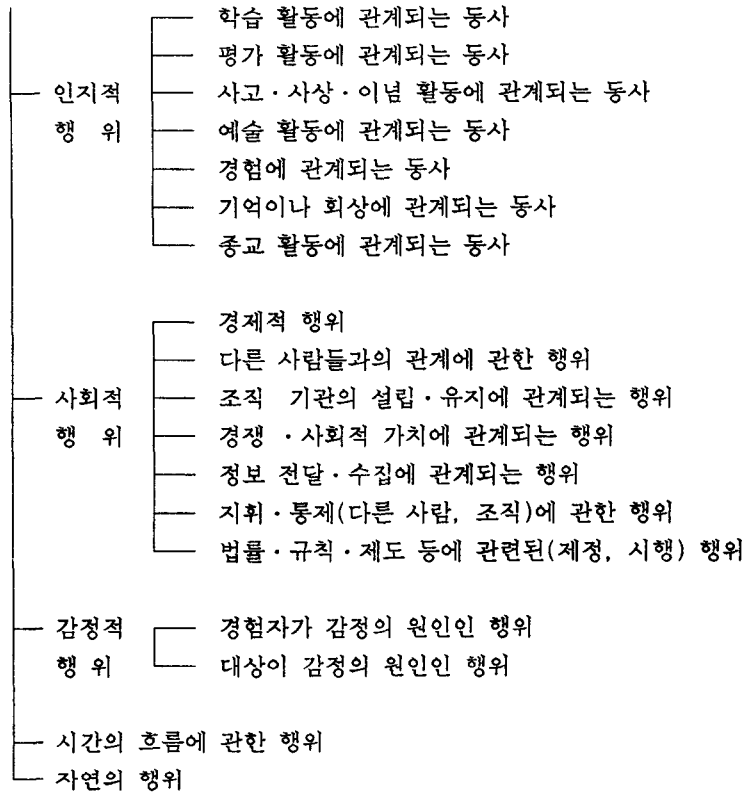
- 생성
- 소멸
- 위치변화
- 양 변화
- 성질 변화
- 신분 변화
- 모양 변화

사

실제적  
행 위

- 위치 이동 동사
  - 스스로 자신의 위치를 이동
  - 다른 것의 위치를 이동
- 생물체가 자신의 신체부분을 움직이는 동사
- 감각 동사
- 새로운 것을 만드는 동사
- 변화 동사
  - 소유권 변화
  - 수정 · 고침
  - 외형 변화
  - 양 · 부피 변화
  - 성질 변화
  - 상태 · 상황 변화
  - 연결 · 끊음
- 어떤 것을 없애는 동사
- 먹는 행위
- 동물의 생리 작용
- 생물체의 성장 · 양육에 관계되는 동사
- 구체적인 일을 수행하는 동사
  - 적는 동사
  - 수치 측정
  - 일을 시키는 동사
  - 농사 행위
  - 치료 행위
  - 직업적 일 수행
- 물체에 힘을 작용하는 동사
  - 힘을 가하는 동사
  - 힘을 맞는 동사
- 다른 대상을 이용하는 동사
- 행위에 관한 행위
  - 행위가 일어나게 하는 동사
  - 행위의 발생을 막는 동사
- 장식에 관한 행위
  - 의복에 관한 행위
  - 꾸밈 · 청소에 관한 행위
- 어려움 · 고생에 관계되는 동사
- 사실의 숨김 · 드러냄

행 위  
동 사



〈그림 2〉 동사 개념 분류표

## 1. 동사 개념 정의

(1) {상태 동사} : 실제적인 움직임보다는 물체의 상태를 나타내는 동사의 범주로서 {현재 상태}와 {상태 변화}의 2개의 하위범주로 구성된다.

(2) {현재 상태} : {구체물}의 현재 상태를 의미하는 동사들의 집합으로 {존재}, {능력}, {상태}, {성질} 등으로 하위 분류 하였다.

(3) {상태 변화} : 현재 상태가 변화됨을 나타내는 동사들로서 {생성}, {소멸}, {위치 변화}, {신분 변화}, {모양 변화}, {양적 변화}, {성질 변화}등의 개념들로 구성된다.

(4) {행위 동사} : 실제 움직임을 나타내는 동사들이 이 개념 범주에 속한다. 행위 동사는 인간의 행동을 중심으로 분류하였다. 인간의 행위를 보면 실제로 육체를 사용

하는 움직임, 두뇌 활동, 감정 발생, 사회 활동으로 이루어 진다. 그래서 이 개념은 {실제적 행위}, {인지적 행위}, {사회적 행위}, {감정적 행위}와 {시간의 흐름}, {자연의 행위}등의 개념으로 구성된다.

(5) {실제적 행위} : 인간이 실제로 육체를 사용하는 행위들로 구성되어 있다. 이 개념이 동사의 개념 중에서 가장 많은 부분을 차지한다. 하위개념에는 {위치 이동 동사}, {감각 동사}, {변화 동사}, {동물의 생리 작용} 등이 있다.

(6) {인지적 행위} : 인간의 두뇌 활동을 의미하는 동사들이 이 범주에 속한다. 여기에는 {학습 활동}, {평가 활동}, {사고, 사상, 이념 활동}, {종교 활동}등이 있다.

(7) {사회적 행위} : 인간의 사회 생활에서 발생하는 활동들을 의미한다. 이 개념은 {경제적 행위}, {다른 사람들과의 교류에 관한 행위}, {경쟁, 사회적 가치 획득에 관한 행위}, {지휘, 통제}, {법률, 제도}등의 하위범주로 구성된다.

(8) {감정적 행위} : 인간(동물)의 감정 상태를 나타내는 동사들로서 감정의 경험자가 생물체이고 그 대상이 감정의 원인이 되는 동사와 감정의 경험자가 생물체이고 그 경험자가 감정의 원인이 되는 동사들로 분류하였다.

(9) {시간의 흐름} : 시간의 변화를 나타내는 동사들로 위의 동사들과는 그 성질이 다르기 때문에 따로 분류하였다.

(10) {자연의 행위} : 기후 변화, 기상등을 나타내는 동사들이다.

## IV. 결론

자연언어 문장을 분석하는데 요구되는 개념정보를 얻기 위해서 동사와 명사에 대해 각 단어의 개념을 분석, 분류하였다. 이들 개념정보는 개념 사전에 저장되어 한국어 문장 분석기에 직접 이용된다. 문장 분석시 단어의 개념정보를 이용함으로써 문법 정보만으로는 해결할 수 없는 많은 어휘 중의성을 해결할 수 있다.

본 논문에서는 제한된 영역의 동사와 명사에 대해 실험적으로 개념 분류를 하였으나 실용적인 문장 분석기를 개발을 위해서는 모든 단어에 대해 이러한 개념 분류가 이루어져야 하며 이 정보를 효율적으로 저장, 관리할 수 있는 사전 구조에 대한 연구가 계속되어야 할 것이다.



## V. 참고 문헌

- [1] 남기심, 고영근, 고교 문법, 탐출판사, 1991
- [2] 이익환, 의미론 개론, 한신 문화사, 1989
- [3] 나동열, 지식 기반 자연언어 파싱에 관한 소고,  
인공지능 소식 제 6권 1호, 1991.3
- [4] 김봉주, 개념학, 한신문화사, 1992
- [5] R.W.Langacker 저, 이기동, 김종도 역, 인지문법, 한신문화사, 1991
- [6] Concept Dictionary, EDR Technical Report, 1990
- [7] International Journal of Lexicography Volume 3 Number 4 Winter 1990,  
Oxford University Press
- [8] John F. Sowa, Knowledge Representation, IBM System Research
- [9] Roger C.Schank, Conceptual Information Processing,  
Amsterdam:North-Holland, 1975
- [10] Small,Cottrell,Tanenhaus, Lexical Ambiguity  
Resolution, Morgan Kaufmann Publishers,Inc, 1988
- [11] J.Fillmore, 격문법이란 무엇인가, 을유문화사, 1988