

명사의 의미소성의 계층구조

윤종지\* 이주근\* 박차근\*\*  
\*인하대학교 전자계산학과 \*\*청주대학교 전자계산학과

The Hierarchical Structure of Semantic Property

K. J. Yoon\* C. K. Park\*\* J. K. Lee\*  
\*Dept. of Electronics, Inha Univ. \*\*Dept. of Computer Science, Chungju Univ.

Abstract

This paper deals with a semantic properties of Korea noun for semantic process in machine translation. The procedure is carried out as follow;

- 1) 17,000 words of Korean nouns are collected.
- 2) Semantic category is classified into 39 markers.
- 3) We slow the redundancy of semantic properties and improve the efficiency of dictionary by marking the hierarchical concept structure.

I. 서론

인공지능(Artificial Intelligence)에서 취급하는 문제들은 대량의 지식정보가 기본적으로 필요하며, 문제를 해결하기 위해서는 이 지식정보를 효과적으로 처리할수 있어야 한다. 즉 인공지능 문제에서 일차적으로 직면하게 되는 것은 지식의 표현(Knowledge Representation)문제이라 할 것이다.

지식표현의 문제에서 일차적 과제는 자연어 처리의 경우를 표현하는 문제이다. 한국어의 의미표현은 최근에 기계번역 시스템등에서 연구가 진행되고 있으나, 과거에는 없었던 것으로서 자료가 극히 적고, 또 방대한 량의 명사류에서의 미를 부가하기는 대단히 어렵다. 본 논문에서는 한국어의 명사를 39종으로 세분류하고 명사의 의미적 분류를 각 항목에 대해서 명사의 개념적인 계층구조상에서 의미소성을 부가하였다.

저자들의 연구실에서는 한국어의 지식표현에 의한 번역기계및 전문가 시스템에서 연구변바 있으나, 본 연구에서는 보다 효율적이고 광범위한

응용시스템을 목표로 다시 별도로 개념적인 계층구조의 지식표현 방법을 시도해 보았다. 그러기 위해서 한국어의 명사와 동사의 관계를 재조명하여 본다.

II. 동사와 명사와의 관계

본장은 동사와 명사에 의하여 형식화 할 수 있으나[4.5] 이들 사이에 내재하는 관계를 규명하고 형식화 함에는 관점에 따라 여러가지 방법이 존재할 수 있다. 그러나 Production Rule 을 위주로 할때는 의미적 요소가 상실되거나, 각(Case)등의 문법을 중심요소로 할때는 추론(Inference) 기능이 약화될 염려가 있다.[6]

그러므로 동사와 명사 사이의 관계를 문법과 의미의 관계가 합리적으로 표현될수 있도록 하는 것이 바람직하며 특히 한국어는 직관관계가 조사와의 형태로 표면구조상에 나타나므로 문법과 의미의 조화가 요구된다.

1) 명사간의 관계

명사는 동사의 entity로 동사의 일반적 개념을 구체적으로 규정하는 계층개념 역할을 한다. 이들 상호간에는 계층적 집합관계가 형성되고, 개별 속성에 의해 독자적 의미를 규정한다.

그림1에 명사의 계층구조에 대해서 간단히 나타 내었다.

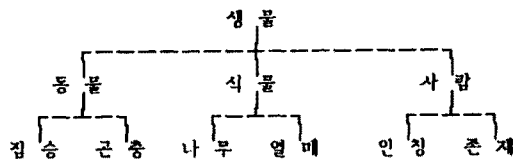


그림1. 명사의 계층구조

2) 명사와 동사의 관계

문장내에서 동사는 일반적 개념을 규정하고 명사는 개별적 개념을 규정한다. 즉 의미의 projection이 이루어지고 있으며, 이들 관계는 고유관계와 문법관계로 구분할 수 있다.

고유관계에서는 동사의 의미복성에 따라 선택된 명사의 관계는 의미적으로 각(case)관계로 형식화 된다. 이 각관계는 동사와 명사를 연결하는 일차적인 관계로 의미형성에 중요한 역할을 한다.

문법관계는 문장내에서 동사는 명사의 의미영역을 한정하는 역할을 한다.[6]

III. 의미소성

1) 의미본의 관점

의미는 인간행위의 많은 개념과 관련되어 있기 때문에 다각적인 관점과 기술에 의하여 표현되어 진다. 이와같은 관점들은 논리학, 심리학, 문학적 비평, 언어학, 또는 통신과학과 같은 것들에서 거론되고 있다. 이러한 관점들은 언어학의 영역에서만도 논리적 관계, 대응어(Substitution), 언어(Collocation), 분포(Distribution), 동의어, 반의어(Antonymy)등의 다양한 구분이 있다. 이와같은 의미의 다양한 접근은 각각각색의 의미를 만들어 낸다.

의미본의 연구는 크게 2가지로 나눌 수 있다. 하나는 대상물에 초점을 두는 것이고 다른 하나는 개념구조에 초점을 두는 것이다. 전자는 언어의 extensionlist view 라 하고 후자는 언어의 intensionlist view 라 부른다.[1]

위의 두가지 방법중 어느 하나만으로는 의미분석을 수행할 수 없다. 즉 두가지다 중요하다. 그러나 지시적 의미(referential meaning) 분석에 기초가 되는 적절한 카테고리를 결정하기 위해 개념적 절차가 기본이 된다.

개념체계에는 상위개념과 하위개념으로 구분되며 두개념은 상대적인 것으로서 전자의 개념은 사물이 되고 후자의 개념은 사물의 각 단어가 된다. [3]

본 논문에서는 개념 카테고리의 설정에 따라 최상위 개념은 다음과 같이 4가지를 설정한다.

- i) 생 물
- ii) 무 생 물
- iii) 추 상
- iv) 행 동

이 4가지의 기본개념을 기본으로 하고 이들

각각의 상위개념에 대한 하위개념은 총 39종류의 의미 소성으로서 구분한다.

2) 품사의 규정

자전을 구성하기 위하여 국어사전으로부터 어휘를 수집하는 것이 쉽지만 사전에 수록된 단어들은 고립된 단어들이므로 구문, 문맥등의 정보는 알수 없다. 따라서 실제문으로부터 각종 품사는 물론 모든 정보를 도출하기 위하여 국민학교, 중,고등학교 국어교과 전과정과 일반사전으로부터 약 4만어의 단어를 수집하고 그 중에서 명사 약 17000개를 도출하여 의미소성및 기타정보를 부가한다.

한국어에서 단어는 최소자립 형식으로서 분리성을 가지고 있고 구문의 최소단위가 된다는 공통된 견해에 따라 언어학 예서는 7품사로부터 12품사로 분리하는등 여러가지가 있으나 본 연구에서는 기계번역과 질문-응답 시스템에 공통으로 사용하기 위하여 수집된 자료를 기초로 12품사로 분류하였다.

한국어에서 용언은 문장의 서술부가 되는 주기능을 갖고며 활용을 위하여 문장에서 주어, 관형어, 부사어, 목적어, 보어가 되고 용언은 형용사, 동사등으로 분류 하였다. 표 1에 분류된 품사표가 있다.

표 1 품사 분류

형태 분류		의미 분류
어휘적 형태소	체언(N)	명사(No)
		대명사(PN)
	용언(V)	수사(NU)
		동사(Vb)
문법적 형태소	관형어(D)	형용사(Va)
		관형사(Det)
	부사어(A)	부사(Ad)
		조사(P)
어미(E)	보조사(Ps)	
		접속조사(Pi)
		종결어미(Es)
		전성어미(Et)
		연결어미(Ec)

또한 수집된 명사의 의미소성을 39개로 분류하였고 표 2에 간단히 나타나 있다.

표 2 의미소성 종류

의미소성	기호
1. 인정	MDC
2. 존재	HUM
3. 관념	THK
4. 근중	INS

5. 병균	VIR
6. 열매	FRU
7. 하늘	SKY
8. 총칭	GEN
9. 인쇄물	PRN
10. 하늘	SKY
11. 총칭	GEN
12. 지리학적	GEO
...	...
36. 지시	MPC
37. 위치	MLC
38. 정적	STT
39. 동적	MOV

사전구성은 표3 과 같다.

표 3. 명사사전의 예

- (SIGAN (TIM))
- (JIJIN (NAP))
- (BONSEOO (MBC))
- (SASAO (MET))
- (MIT (MLC))
- (HYEO (PAR))
- (SOGOS (PRO))
- (CHIR (MQC))
- (SIRHEOM (MOV))
- (YBONGEUG (STT))
- (JEO (HUM))
- (SEOM (GEO))
- (MAGAM (TIM))
- (SAJIN (PRO))
- (SYASSEU (PRO))
- (BAOHAE (MOV))
- (JAYU (STA))

3) 의미소성의 계층구조

i) 저자들이개발한 기계번역 시스템을 전문가 시스템에 응용으로 확장하기 위하여 의미소성을 39개로 증가시켰다.

ii) 일반적으로 의미소성을 부가할때 그 단어가 다의적으로 사용하지 않을 때도 한 카테고리에 복수개의 의미소성을 부가할 필요가 있을 때가 있다. 그렇게되면 사전의 부담이 증가된다. 이 문제를 해결하기 위하여 본 논문에서는 계층구조로 의미소성을 분류했다. 그림3에 분류된 계층구조가 나타나있다.

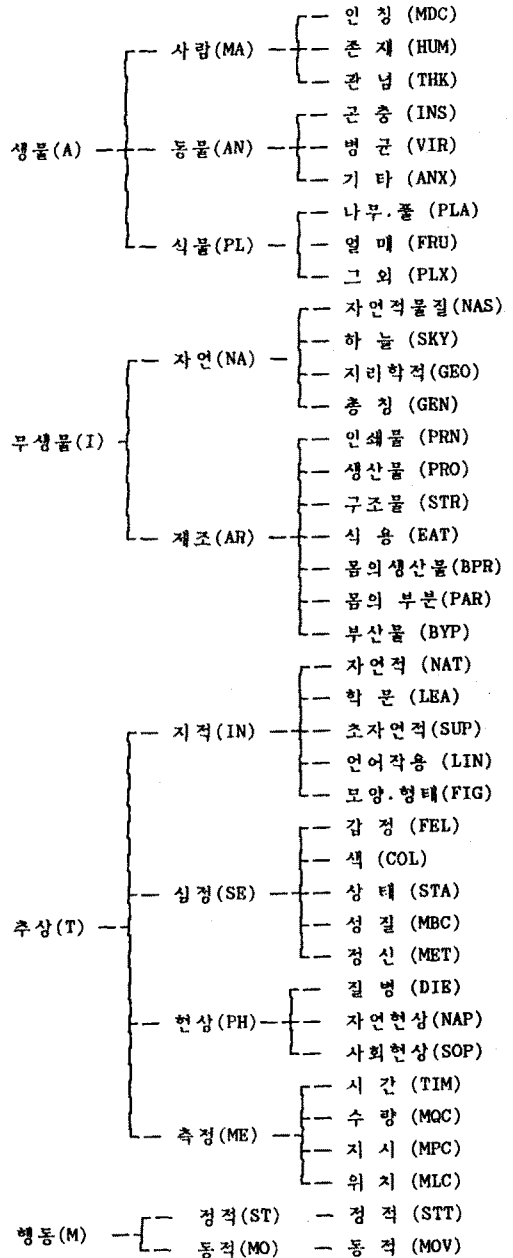


그림 3. 의미소성의 계층구조

IV. 결론

본 논문에서 얻은 성과는 다음과 같다.

- 1) 명사의 사전을 17,000 단어를 수집하여 의미마르를 부가하였다.
- 2) 의미소성은 비교적 세분하여 39종을 규정하였다.
- 3) 의미표현은 계층구조로 분류함으로써 명사의

의미소성의 중복부가를 피할 수 있으며 사전의 효율적 구성을 가능케 했다.

참 고 문 헌

- 1) THOMAS A. SEREOV, "Approaches to semiotics" pp.174-196.
- 2) Sakamoto, Y. , Ishikawa, T. and Satoh, M. "Concept and Structure of Semantic Markers for Machine Translation in Mu - Project, COLING 86, Bonn, 1986.
- 3) Tetsuya Ishikawa, Yoshiyuki sakamoto, Masayuki sato, "Concept and Structure for Semantic Information in Mu-Project", 자연어처리 55-1, 5, 1986
- 4) Chomskyn N., "Aspect of theory of syntax", Cambridge, Mass., 1965.
- 5) Fillmore, C., "The Case for Case," in Universals in Linguistic theory, HRW, NY., 1968
- 6) 이주근, 한성국, "자연어 의미의 Schema 표현" 전자공학회 학술논문집, Vol.7, No.1, pp.322-327, 7, 1984
- 7) 한광록, 손석현, 장두영, 이주근, "의미 frame에 의한 기계번역," 전자공학회 학술논문집, Vol 9, No.1, pp. 184-187, 6, 1986.