

# 개념그래프를 이용한 의미사전의 개발에 관한 연구

김 은 회<sup>0</sup>, 양 기 철  
목포대학교 전산통계학과

## A Study on Development of Semantic Lexicon Based on Conceptual Graphs

Eun-Hee Kim , Gi-Chul Yang

### 요 약

자연언어는 사람이 쉽게 이해할 수 있지만 이를 컴퓨터가 이해하는데는 많은 문제점들이 발생한다. 이러한 문제의 해결을 위해 의미처리가 수행되어야 하며 효율적인 의미처리를 위해서는 정확한 의미사전의 개발이 요구된다. 본 논문에서는 개념그래프를 이용한 의미사전의 개발과정에서 고려되어야 할 몇가지 문제점들을 살펴본다.

### I. 서론

사전은 자연언어처리에서 중요한 역할을 담당하고 있다. 특히 문장의 의미해석을 위한 의미사전에 관한 관심이 높아지고 있다. 본 논문에서는 한국어 문장해석시스템을 위한 기초작업으로서 개념그래프를 사용한 의미사전의 개발에 관한 몇가지 문제점을 살펴본다. 2장에서는 개념그래프를 이용한 의미사전에 관하여 기술하고 3장에서는 본 논문에서 개발하고자 하는 의미사전의 구성상 발생하는 문제들과 그에 대한 최소한의 해결책들을 살펴보고 4장에서 결론을 맺는다.

### II. 개념그래프와 의미사전

본 논문에서 개발하고자 하는 의미사전은 개념, 개념관계, 규범그래프, 그리고 유형계층으로 구성되어 있다.

#### 2-1. 개념

개념은 속성, 상태, 사건, 실체 등을 나타낸다.

사전상에서 다음과 같은 형태로 단어에 대한 개념을 표현한다.

단어" {문법적인 범주}"/{[개념]  
| "개념없음"}"

예) 사람. 명사;[사람].

어떤. 관형사;개념없음.

개념은 다양한 표현기능들을 가지고 있으며 개념유형(concept type)과참고지대(referent field)로 구성된다. 개념유형은 단어의 유형을 표현하고참고지대는 대명사, 고유명사, 관형사와 같은 품사를 표현한다. 즉, '어떤 사람'이라는 어절을 개념으로 나타낼때 '어떤'은 관형사로서 별도의 개념표현없이 '?'라는참고지대를 이용하여 의미를 표현할 수 있다. <표 1>에서 나타나는 바와 같이 '[사람:?]'으로 각 단어의 의미를 표현한다. <표 1>은 자연언어의 의미를 표현할 수 있는 언어인 개념그래프를 고안한 Sowa의 개념표현들의 예이다.

## 2-2. 개념관계

개념 관계는 연결할 수 있는 최대한의 의미를 갖는 개념들을 연결한 연결 양식을 말한다. 다음은 개념 관계의 정의와 표현 예, 그리고 표현 예에 따른 개념 그래프와 개념 관계

<표 1> 여러 가지 개념표현

개념의 예	의 미
[사람] 또는 [사람:*]	사람 또는 임의의 사람
[사람:#1234]	#1234라는 사람
[사람:순희]	순희 또는 순희라는 사람
[사람:@1]	한사람
[사람:?]	어떤 사람?
[사람:{순희,철수}]	순희와 철수
[사람:{*}]	사람들
[사람:{*}@3]	세명의 사람들
[사람:{*}#]	그 사람들
[사람:A]	모든 사람
[사람:-]	사람이 아님

의 활용이다.

개념 관계(축약된 개념 관계 표현)

개념 관계의 연결 양식

예문장

예문장에 대응하는 개념 그래프와

개념 관계의 활용

예) content. (CONT)

[개체:\*x]→(CONT)→[개체:\*y].

. 필통안에 볼펜이 있다.

[필통]→(CONT)→[볼펜].

개념 관계에는 일원 관계와 이원 관계가 있는데 일원관계는 하나의 아크로만 구성되어 있으며 이원관계는 두 개의 아크로 구성되어 있다. 대부분의 개념관계는 이원관계로 표현된다. 다음은 일원관계와 이원관계를 보여주는 예이다.

일원관계의 예) past (PAST)

(PAST)→[명제:\*x]

. 순희가 떠났다.

(PAST)→[명제:[사람:순희]←

(AGNT)←[떠나다]].

이원관계의 예) duration (DUR)

[상태]→(DUR)→[시간간격].

.비가 3일간 계속왔다.

[시간간격:@3일]←(DUR)←

[오다]→(OBJ)→[비].

## 2-3. 규범그래프

규범그래프는 의미있는 개념그래프를 구성하기 위한 선택제약으로서 본 사전에서는 용언만 규범그래프를 갖도록 제한한다. 개념 '[갖다]'의 규범 그래프는 다음과 같다.

[갖다] -

(AGNT)→[개체]

(OBJ)→[사물].

개념 '[갖다]'는 행위의 주체로서 개념 '[개체]'의 하위 개념들만이 개념 관계 '(AGNT)'로 연결이 가능하다. 또한 '[갖다]'는 타동사로서 목적을 갖는데 그 목적으로는 개념 '[사물]'의 하위 개념만이 개념 관계 '(OBJ)'로 연결이 가능하다. 이 규범 그래프를 통하여 동사나 형용사에 대한 선택 제약을 가한다. 다음은 '[만들다]'와 '[있다]'의 규범그래프이다.

[만들다] -

(AGNT)→[개체]

(MATR)→[물질]

(RSLT)→[사물].

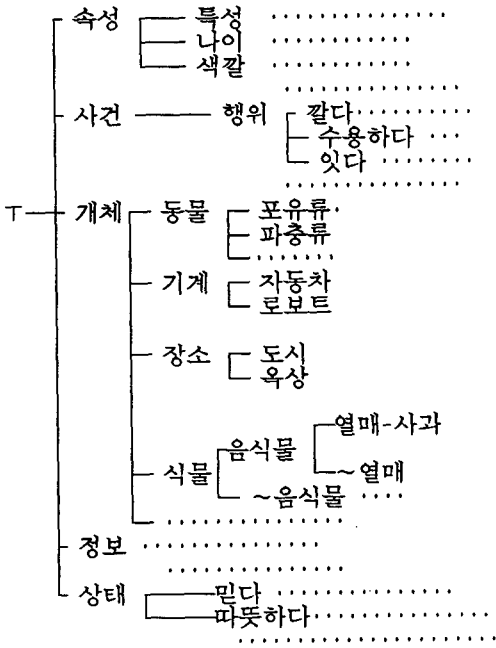
[있다] -

(AGNT)→[동물]

(OBJ)→[사물].

2-4. 유형계층

유형계층은 개념의 의미범주에 관한 정보를 제공하는 것으로서 유형계층을 구성하기 위해서는 먼저 일반적으로 널리 쓰이는 상위 개념들부터 고려해야한다. 본 논문에서는 일반적인 상위 개념들은 어느 시스템에서나 거의 동일하므로 Sowa가 정의한 상위 개념을 한국어로 바꾸어 첨가 또는 수정하고 본 논문에서 이용한 문장들로부터 하위 개념들을 설정하며 상속성을 갖는다. <그림 1>은 유형계층의 예이다.



(그림 1) 유형계층의 예

III. 의미사전 구성의 문제점

본 장에서는 개념그래프를 사용하여 의미사전을 구성할 때 발생하는 몇가지 문제점과 그에대한 최소한의 해결책을 살펴본다.

첫번째 문제점은 규범그래프에 어떤 요소들을 정의할 것인가를 결정하는 것이 어렵다는 것이다. 예를들어 '[펼럭이다]'의 규범그래프가 다음과 같다고 하자.

- 1) [펼럭이다] -
  - (AGNT)→[개체]
  - (LOC)→[장소].

이때, '태극기가 옥상에서 펼럭입니다'라는 문장은 1)의 개념그래프로 표현되지만,

- 1-1) [펼럭이다] -
  - (AGNT)→[태극기]
  - (LOC)→[옥상].

'태극기가 바람에 펼럭입니다'라는 문장은 주어진 '[펼럭이다]'의 규범그래프에 '(CAUS)'라는 관계를 추가해서 해석하여야 바람이 장소가 아니라 펼럭이게하는 원인이라는 의미를 나타낼 수가 있다. 결과적으로 2)의 개념그래프로 표현되어야한다.

- 2) [펼럭이다]-
  - (AGNT)→[태극기]
  - (CAUS)→[바람].

2)의 개념그래프를 생성하기 하기 위해서는 2-1)과 같은 규범그래프가 필요하다. 이러한 문제들의 해결을 위해서는 먼저 규범그래프를 작성할 때 필요한 최소한의 개념과 개념관계만으로 규범그래프를 구성하고 문장분석시 처음의 규범그래프로는 해결하지 못하는 문장이 자주 나타나게 되면 추가로 필요한 개념과 개념관계를 추가하여 규범그래프를 수정하여 나가도록 한다. 어떤 개념관계를 규범그래프에 포함시켜야 되는가를 신중히 고려해야한다.

- 2-1) [펼럭이다]-
  - (AGNT)→[개체]
  - (LOC)→[장소]
  - (CAUS)→[원인].

두번째 문제점은 유형계층과 규범그래프상의 의미범주 지정이 어렵다는 점이다. '[먹다]'의 개념을 위한 규범그래프가 다음과 같다고 할때,

[먹다] -  
 (AGNT)→[동물]  
 (OBJ)→[음식물].

‘철수가 사과를 먹는다’라는 문장은 다음과 같은 개념그래프로 해석될 것이다.

[먹다] -  
 (AGNT)→[사람:철수]  
 (OBJ)→[사과].

하지만, 위의 규범그래프에서 ‘[음식물]’대신 ‘[개체]’를 쓰게된다면 유형계층에서 ‘[개체]’는 먹는 대상뿐만아니라 먹지 못하는 것들까지도 포함하므로 동물이 먹을 수 있는 대상이 ‘음식물’로 한정되지 못하고 ‘개체’로 확대된다. 즉, ‘[먹다]’의 규범그래프가 다음과 같다고 할때,

[먹다] -  
 (AGNT)→[동물]  
 (OBJ)→[개체].

‘철수가 자동차를 먹는다’라는 문장의 경우 의미상으로 틀린 문장이지만 위의 규범그래프를 적용시키면 옳게 해석되어 다음과 같은 개념그래프로 해석이 가능하게 된다.

[먹다] -  
 (AGNT)→[사람:철수]  
 (OBJ)→[자동차].

이러한 오류를 방지하기 위해서는 규범그래프의 각 개념들을 설정할 때 개념관계 ‘(OBJ)’에 대응될 수 있는 의미범주로 제한해야 한다. 하지만 문학적인 작품에 나오는 문장은 의미범주를 제한하면 해석되지 못하는 경우가 있다. 예를 들어 ‘[말하다]’의 ‘(AGNT)’를 ‘사람’으로 제한하면 ‘낙엽이 가을을 말한다’는 문장은 표현할 수 없게 된다. 따라서, 사전구성자가 응용분야를 철저히 파악하여 분야에 맞게 의미범주를 어느정도로 제한하고 확장해야 할 것인가를 결정하여 유형계층을 세분화하고 규범그래프상의 개념에 대한 의미범주를 지정해야만 좀 더 정확한 해석을 기대할 수 있다.

세번째 문제점은 전치사나 조사가 아닌 용

어들도 개념보다는 개념관계로 분류하는 것이 문장을 개념그래프로 표현하는 데 있어서 더 용이한 경우가 있어서 단어를 개념으로 분류를 해야할지 개념관계로 분석해야할지를 결정하는 것이 어렵다는 것이다. 예를 들어서 ‘철수와 순희는 사랑하다’라는 문장이 있을 때 일반적인 규칙에 따라 ‘사랑하다’라는 단어가 개념 ‘[사랑하다]’로 분석되어 개념그래프를 생성한다면 다음과 같이 표현될 것이다.

[사랑하다] -  
 (EXPR)→[철수]  
 (OBJ)→[순희].

그러나, ‘(사랑)’이라는 개념관계를 사용하면 의미적으로 더욱 분명하게 표현되며 그래프도 간결하다.

[철수]→(사랑하다)→[순희].

용언이지만 개념보다는 개념관계로 분석되는 경우가 의미적으로 더 정확히 표현되는 장점을 갖는다면 해석대상이 되는 문장들에서 얼마나 잦은 횟수로 이러한 경우가 발생하는지를 고려하여 새로운 개념관계를 지정하는 것이 바람직할 것이다.

마지막으로 네 번째 문제점은 복합명사의 표현문제이다. 응용분야에서 이용되는 문장중에서 자주 나오는 어절들과 복합명사는 각 단어에 대해서 개념을 설정하고 개념관계로 이들을 연결할 필요없이 하나의 개념으로 정의하여 사용할 수 있으며 이러한 표현의 사용은 정보검색시 효율을 높이는 결과를 보였다[5].

즉, ‘정보검색’이라는 어절이 여러 문장에서 쓰인다고 할때 [정보]←(OBJ)←[검색]으로 하지 않고 ‘[정보검색]’이라는 하나의 개념으로 표현하여 쓰면 그래프를 간결화할 수 있으며 읽기도 더 쉽다. 새로운 개념을 정의할때는 반드시 정의된 개념유형이 유형계층에 나타나야하는데 개념‘[정보검색]’은 정보의 의미와 검색의 의미를 상속받는다. 개념그래프를 축약하여 새로운 개념을 정의할 경우에도 얼마나 잦은 횟수로 발생하는지를 고려하여 개념을 지정해야한다.

#### IV. 결론

효율적이 의미처리를 위해서는 정확한 의미사전의 구축이 필요하다. 본 논문에서는 개념그래프를 이용한 의미사전의 개발과정에서 생기는 몇가지 고려될 문제점들과 그에 대한 최소한의 해결책을 제시하였다. 앞으로 남아있는 문제점들이 해결되어 의미사전을 통한 자연언어처리시스템의 질을 향상시킬 수 있기를 바라는 바이다.

#### V. 참고문헌

- [1] 김현효, “영어 준동사 구문의 개념구조”, 충남대학교 박사학위 논문, 1995.
- [2] 정미애, “영어 조용표현의 개념구조”, 충남대학교 박사학위 논문, 1995.
- [3] 조길호, “영어 관계구문의 개념구조적 기술”, 충남대학교 박사학위논문, 1994.
- [4] 조성원, 송만석, “사전에 기반한 한국어 문장 해석 시스템 원형의 복구”, 한국정보과학회 논문집 Vol. 18, No 2, 1991.
- [5] 김문석, 남세진, 신동욱, “복합명사 통계자료를 이용한 한글 자동색인시스템 개발”, 한국정보과학회 논문집 Vol. 23, No. 1, 1996
- [6] 김은희, “한국어 문장분석을 위한 의미사전의 설계 및 구현”, 목포대학교 석사학위 논문, 1996.
- [7] J. F. Sowa, Conceptual Structures : Information Processing in Mind and Machine, Addison-Wesley Publishing Company, 1983.
- [8] Timothy E. NAGLE, Janice A. Nagle, Laurie L. Gerholz, “Conceptual structures Current Research and Practice”, Ellis Horwood, 1992
- [9] Electronic Dictionary Research Institute, Ltd(EDR), “EDR Electronic Dictionary”, EDR Institute, 1993.