

술어 기반 문형 정보를 이용한 한국어의 의미 구조 생성에 관한 연구

박인철^o*, 배우정**, 안동언***, 이용석**
전북산업대학교 전자재산학과*
전북대학교 컴퓨터학과**
전북대학교 컴퓨터공학과***

A Study on the Generation of Semantic Structure of Korean using Sentence Pattern Information based on Predicates

In-Cheol Park^o*, Woo-Jeong Bae**, Dong-Un An*** & Yong-Seok Lee**
Dept. of Computer Science, Chonbuk Sanup University*
Dept. of Computer Science, Chonbuk National University**
Dept. of Computer Engineering, Chonbuk National University***

요 약

대부분 자연 언어 이해 시스템이나 중간 언어 방식을 이용한 기계 번역 시스템에서 자연 언어 문장을 이해하고 번역하기 위해서는 대량의 지식을 이용한 의미 구조의 생성이 요구된다. 따라서 개념 그래프를 이용하여 한국어 문장의 내부 의미 구조를 생성하기 위해서는 각 단어에 해당하는 개념과 개념들 사이의 개념적 관계를 나타내는 지식들이 요구된다. 그러나 이를 위한 의미 구조 생성 방법과 요구되는 지식 베이스를 정확하게 구축하는 것은 어렵고 응용 도메인에 종속한다는 문제가 있다. 본 논문에서 우리는 문형 이론을 도입하여 문형을 중심으로 의미 구조 생성을 위한 변환 규칙을 설계하고 이를 이용하여 방대한 지식 베이스의 구축없이 의미 구조를 생성할 수 있는 방법에 논의한다. 또한 본 논문에서 문형을 이용하면 몇 가지 모호성 문제를 해결할 수 있음을 보이고 문형의 한계에 대해서도 살펴본다.

1. 서 론

대부분의 기계 번역이나 자연 언어 이해 시스템과 같은 자연 언어 처리 분야에서는 자연 언어 문장의 의미 구조를 정확하게 생성할 수 있는 방법과 지식이 필요하다. 그러나 이를 위한 의미 구조 생성 방법과 요구되는 지식 베이스를 정확하게 구축하는 것은 어렵고 응용 도메인에 종속한다는 문제가 있다. 자연 언어 문장의 의미를 나타내는 지식 표현 언어로는 일반적으로 논리 형태(logical form), 의미망(semantic network), 인공 언어를 이용하여 표현하는 데 Sowa에 의해 제안된 개념 그래프(conceptual graph)는 확장된 의미망으로서 문장의 내부 의미를 논리적으로 간결하고 정확하게 표현할 수 있다는 장점을 가지고 있다[Sowa 84 86 87][Velardi 88].

자연 언어 문장의 의미를 나타내는 개념 그래프는 문장을 구성하고 있는 각 단어에 대응하는 개념과 개념 사이의 개념 관계로 표현되어지기 때문에 각 단어에 대응하는 개념의 생성과 개념 사이의 관계를 결정할 수 있는 방법 및 지식이 요구된다. 즉, 자연 언어 문장의 의미 구조를 개념 그래프로 표현하기 위해서는 구문 분석된 파스 트리를 기반으로 주어와 목적어 같은 구문적 역할(표층격)을 행위격(agent)과 대상격(object)과 같은 어떤 의미적 역할(심층격)로 사상할 수 있는 규칙들이 필요하다는 것을

의미한다. 그러나 이러한 지식들을 일일이 각 단어마다 기술하는 것은 어려운 문제 중의 하나이다.

본 논문에서 우리는 한국어 문장으로부터 개념 그래프를 생성할 때 각 술어의 개념과 체언의 개념들 사이의 개념 관계를 결정하기 위해 언어학적 측면에서 분류된 한국어의 기본 문형 구조를 이용한다.

언어학적 측면에서 문형(sentence pattern)은 문장의 기본이 되는 문의 기본 형식으로서 수많은 개의 구체적인 문장으로부터 표층적인 구조적 형식의 공통성에 따라 분류되어 동일한 종류의 문장을 대표할 수 있고 또 생성해 낼 수 있는 필수적 문장 성분만으로 이루어진 문장 구조 유형을 의미한다 [장 95].

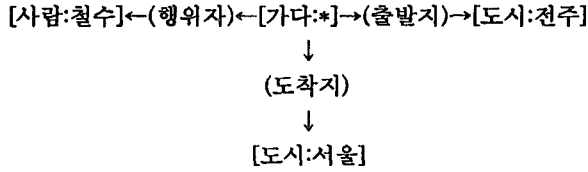
한국어의 문형은 동사와 형용사 같은 술어 자체의 어휘적 의미를 실현시키기 위해 반드시 문장을 구성하는 데 없어서는 안될 필수적 성분(보어 또는 필수격)과 수의적 성분(첨어 또는 임의격)으로 구성되는 특징이 있다. 즉, 문장에서 조사와 결합된 몇 개의 체언들 및 부사와 같은 필수적 성분과 수의적 성분은 한 문장에서 가장 핵심적인 역할을 하는 술어에 의해서 결정된다는 의미이다. 이러한 문형들은 문법적 기능을 나타내는 문장의 표층적인 구조를 나타내기 때문에 본 논문에서는 각 술어들이 가질 수 있는 심층적인 구조를 파악하여 이를 기반으로 한국어 문장의 개념 그래프를 생성할 수 있는 문형 기반 변환 규칙(transformation rules)을 기술한다.

이를 위해 본 논문에서는 문장의 가장 지배적인 위치에서 다른 모든 문장 성분들을 지배하면서 문장 구조를 제약하는 술어를 중심으로 체언의 개념형에 의한 그 의미적 역할의 관계에 따라 각 술어들이 가질 수 있는 심층격들을 설정한다. 이러한 문형들은 격 문법의 격틀(case frame) 또는 개념 그래프의 규범 그래프(canonical graph)와 유사하며 술어의 의미론적 차이에 따라 기본 문장 구조가 달라지는 경우를 고려하여 문형들을 하위범주화(subcategorization)할 수 있다.

본 논문의 구성은 제 2장에서 개념 그래프에 대해 개략적으로 살펴보고 제 3장에서는 한국어의 기본 문형을 고찰한 후 문형을 기반으로 한 변환 규칙에 대해 기술한다. 제 4장에서는 문형 정보를 이용함으로써 여러 가지 모호성을 해결할 수 있음을 보이고 개념 그래프 생성 방법에 대해 기술한다. 마지막으로 제 5장에서 앞으로의 연구 과제를 제시하고 결론을 맺는다.

2. 개념 그래프의 개요

Sowa에 의해 1980년대에 제안된 개념 그래프는 인공 지능 분야에서 자연 언어 처리 및 기계 번역을 위해 구문 분석에 의해 출력된 한 문장의 구문 구조로부터 의미 구조를 정확하게 표현할 수 있는 확장된 의미망(semantic network)에 기반을 둔 논리 체계이다[Sowa 84 86 87][Velardi 88]. 또한 이러한 개념 그래프는 언어학, 심리학, 철학에 기초한 지식 표현 언어로서 사람이 어떤 상황이나 문맥을 이해하는데 필요한 컴퓨터 상에 기술하기 위한 일종의 중간 언어(interlingua)로서 문장에서 각 단어에 해당하는 개념 노드와 그 개념들을 연결해 주는 관계 노드로 구성된 유한 이분 그래프이다. 즉, 개념 노드는 실세계에 존재하는 개체, 속성, 상태, 사건 등을 나타내며 사각형 또는 각 팔호로 표시하며 그 개념 유형(concept type)을 나타내는 유형 필드(type field)와 그 개념이 참조하는 대상을 나타내는 참조대상 필드(referent field)로 구성된다. 또한 문장에서 개념들의 상호 관련성을 나타내는 관계 노드는 심층격이라 말할 수 있으며 형 라벨과 하나 이상의 아크(arc)를 가지고 개념들과 연결되고 원 또는 등근 팔호로 표시하며 일반적으로 이원 관계를 갖지만 일원 관계 또는 다원 관계를 가질 수도 있다. 예를 들어, “철수는 전주에서 서울로 갔다.”라는 문장에 대한 개념 그래프는 그림 1과 같이 표현할 수 있다:



<그림 1> 문장 “철수는 전주에서 서울로 갔다”에 대한 개념 그래프

3. 한국어 문형 구조

3.1 한국어 문형의 개요

언어학적인 측면에서 문형(sentence type)이란 한 언어에서 사용되는 수많은 개개의 구체적인 문장에 대해 구조적 분석을 통해 구조적 형식의 공통성에 따라 공식화한 문장의 틀(templates) 또는 문장의 구조적 유형을 말한다[강 95]. 이러한 문형에 대해 한국어 문형의 분류는 일반적으로 주어, 술어, 목적어, 보어와 같은 서구적 근간 성분을 중심으로 문형 분류가 시도되었으나, [강 95]는 서구적 문법 성분을 근간으로 설정하는 것보다 술어를 중심으로 설정하는 것이 타당함을 주장하였다. 예를 들어, 술어 “읽다”와 “주다”를 동일한 구문적 범주(타동사)로 간주하여 모두 동일한 문형 구조 $N(이)+N(을)+V$ 를 가진다고 생각할 수 있다. 그러나 각 단어의 의미론적 자질을 고려할 때 “읽다”와 “주다”는 각각 다음과 같은 문형 구조를 갖는다고 보는 것이 한국어의 구문/의미 분석 시에 더 적합함을 주장하였다[강 95].

① 학생들이 지금 도서관에서 열심히 책을 읽는다.(SP-11 : $N(이)+N(을)+V$)

N이 N을 V

② 영화는 앞마당에서 꽃나무에 물을 주었다.(SP-12 : $N(이)+N(에)+N(을)+V$)

N이 N에 N을 V

여기서 밑줄친 부분은 문장의 수의적 성분으로서 문장의 상황이나 상태의 의미를 강조하고 구체적으로 나타내기 위해 사용되며, 이러한 성분은 술어의 의미를 실현하는 데 있어 크게 영향을 미치지 않는다. 즉, 한국어 문형은 문장 구조의 특징을 제약하는 동사, 형용사, 서술격 조사(-이다)와 같은 술어를 중심으로 일어나며 문장을 구성하는 데 없어서는 안될 필수적 성분으로서 구문/의미적 자질에 따라 조사와 결합된 몇 개의 체언들(결합가) 및 부사에 몇 개의 수의적 성분을 첨가함으로써 구성된다[김제 88][김일 84][강 95].

이를 기반으로 [강 95]는 한국어 문장이 술어의 의미론적 유사성에 구조적인 공통성을 지니고 있음을 발견하고 11개의 조사(부사 및 -라고, -위해, -대해와 같은 조사 상당어 구를 포함)를 대표 조사로 하여 한국어 문장을 41개의 기본 문형으로 분류하였다.

그러나 [강 95]가 제안한 기본 문형의 구조는 표층적 구조를 나타내기 때문에 술어의 의미적 자질에 따라 심층격을 고려하여 하위범주화할 수 있으며 술어를 중심으로 한 문장을 구성하는 체언들의 의미적 자질은 술어의 의미적 자질과 일치해야 한다는 것을 주목해야 한다. 이는 개념 그래프의 규범 그래프를 작성할 때 고려되어야 하는 선택 제약 사항(selectional constraints)과 동일하다.

본 논문에서는 이러한 기본 문형의 구조가 의미 구조와 유사하기 때문에 한국어 문장에 대한 개념 그래프를 생성하기 위한 틀(templates)로서 이용될 수 있음을 보인다.

3.2 한국어 문형 기반 변환 규칙

자연 언어 문장은 유사한 심층적 구조를 갖지만 발화하는 가운데서 여러 가지 표층 구조로 나타날 수 있으며 심층 구조는 하나의 술어와 하나 또는 그 이상의 체언구로 구성되고 개개의 체언 구는 술어와 일정한 격 관계(case relation)를 갖게 된다. 이는 앞에서 기술한 한국어 문형이 각 술어와 각 체언구가 어떤 의미적 관계로 관련되었는가에 따라 분류되었음을 의미한다.

그러나 [장 95]가 분류한 기본 문형은 표층격(surface case)을 고려하였기 때문에 본 논문에서는 41개의 문형에 대해 각각 문형별로 심층격(deep case)을 고려하여 각 문형들이 가질 수 있는 심층격을 설정한다. 여기서 심층격이란 행위격(agent), 처소격(location), 대상격(object) 등과 같이 각 단어에 대응하는 개념들 사이의 이진 관계를 말한다[김재 88]. 이러한 문형을 기반으로 한 변환 규칙들은 그림 2와 같은 형식으로 기술될 수 있다:

```

<TR> ::= '(' 'Sentence Pattern' { '('<Syntactic Category> <Conceptual Relation>'
      <Default Relation>}' ')
<Syntactic Category> ::= SUBJ | OBJ | NPADV | COMP.....
<Conceptual Relation> ::= {'(<Case Relation> <Concept Type>')}'*
<Case Relation> ::= AGN | OBJ | EXP | INS | REA | .....
<Concept Type> ::= Concept types in conceptual type hierarchy
<Default Relation> ::= {Default conceptual relation according to each sentence type}*
  
```

<그림 2> 문형 정보 기반 변환 규칙

예를 들어, 문장 “나는 전폭으로 자동차를 만들었다.”와 같이 문형 SP-13을 갖는 술어 “만들다”에 대한 변환 규칙은 그림 3과 같이 기술될 수 있다.

```

(TR-13
  ((SUBJ (P-PART ((GR 주격) (FORM 이 가))) (AGN HUMAN)) G13-AGN)
  ((OBJ (P-PART ((GR 목적격) (FORM 을 틀))) (OBJ ALL_ENTITY)) G13-OBJ)
  ((NPADV (P-PART ((GR 부사격) (FORM 로 으로)))
    (MAT PHYSICAL-OBJECT) (STA STATUS)) G13-MAT)
  
```

<그림 3> SP-13에 대한 변환 규칙

이러한 변환 규칙들은 지식 베이스에 있는 각 단어 중 술어가 어떤 기본 문형 구조를 가질 수 있는가를 가리킨다. 따라서 한국어 문장으로부터 개념 그래프를 생성하기 위해서는 변환 규칙에 나타날 수 있는 개념들 사이에 개념 관계(격) 집합을 구성해야만 한다. 그러나 이러한 개념 관계를 설정하는 것은 어렵기 때문에 본 논문에서는 대부분 응용 도메인에서 나타나며 공통적으로 사용되는 개념 관계와 Sowa에서 의해서 제안된 개념 관계를 이용한다.

또한 여러 가지 복잡한 한국어 문장에 대한 개념 그래프를 생성하기 위해서는 접속어, 어미, 조사 상당어 구들에 대해 단항 관계 또는 문맥간의 관계를 결정해야만 한다. 예를 들어, 선어말 어미 ‘았’은 구문 분석 결과의 자질 요소 (tense PAST)를 고려하여 개념 관계 (PAS)를 생성하여 그 문장에 대응하는 개념 그래프에 연결하며 어말 어미 ‘고’는 두 문장을 연결하는 개념 관계 (CON)을 생성한다.

4. 한국어 문장의 개념 그래프 생성

지식 기반 접근에서 개념 그래프를 생성할 수 있으려면 단어와 개념간의 사상은 물론 각 개념간의 관계를 결정할 수 있는 방법이 있어야 하고 이에 필요한 정보를 지식 베이스에 구축해야 한다. 그러나 이러한 구축은 처리 대상 영역이 커질수록 매우 방대한 작업이 될 뿐만 아니라 정확한 관계를 결정하는 것도 쉬운 일이 아니다. 이 장에서는 각 술어와 체언의 개념을 중심으로 각 개념이 근사적으로 가질 수 있는 개념 관계에 대해 문형 기반 변환 규칙을 이용하여 한국어 문장으로부터 개념 그래프를 생성하는 방법과 패턴 매칭 방법의 도입에 대한 이유를 논의한다.

4.1 모호성의 해결

자연 언어 문장의 의미를 정확하게 표현하기 위해서는 여러 가지 모호성을 해결해야만 한다. 한국어에서는 다의적으로 쓰이는 일부 단어들에 그 의미적 차이에 따라 문형 구조를 달리하는 경우가 있기 때문에 구문 정보를 고려할 수 있는 방법이 있어야 한다.

본 논문에서는 술어를 기반으로 분류된 문형 구조는 술어의 의미적 차이에 따라 다른 문형 구조를 가지기 때문에 어휘 또는 구문 모호성을 해결하는 데 이용될 수 있음을 보인다. 예를 들어, 다음 두 문장을 살펴보자.

- ① 나는 철수를 용감하게 만들었다.(SP-17 : N(이)+N(을)+AD+V)
- ② 나는 진흙으로 자동차를 만들었다.(SP-11 : N(이)+N(을)+N(로)+V)

위의 문장에서 첫 번째 문장의 술어 “만들다”는 부사와의 성구론적인 결합 속에서 무엇을 어떻게 하다는 의미를 나타내고 두 번째 “만들다”는 어떤 대상을 제작하는 행동을 의미하는 것으로 각각의 술어의 의미에 따라 문형 구조가 상이하기 때문에 단어의 중의성 문제를 구문 정보 및 문형 정보를 이용하여 해결할 수 있다. 이러한 의미의 구별은 그 술어가 어떠한 기본 문형 구조로 쓰였는가에 따라 달라짐을 의미한다.

또한 문장 “나는 파리에 있는 모자를 보았다.”에서 체언구 “파리에”는 구문 분석 결과 모호성이 발생하게 된다. 즉, 체언구 “파리에”는 술어 “있다”와 “보다”에 개념 관계로 연결될 수 있지만 각 술어의 문형 정보를 이용함으로써 체언구 “파리에”는 술어 “있다”와 개념 관계로 연결됨을 알 수 있다.

- ① 나는 모자를 보았다.(SP-10 : N(이)+N(을)+V)
- ② 모자가 파리에 있다.(SP-2 : N(이)+N(에)+V)

이와 같이 조사 정보와 술어로 쓰인 단어의 문형 정보를 이용하여 이와 같은 단어의 중의성 문제 및 구문 모호성을 해결할 수가 있다.

4.2 문형 기반 개념 그래프의 생성

한국어 문장을 개념 그래프로 표현하기 위해서는 문장을 구성하고 있는 명사, 동사, 형용사 등과 같은 내용 단어(content words)들을 해당 개념으로 사상하고 조사와 어미 등과 같은 기능 단어(function words)들에 대해 개념 관계로 사상할 수 있는 방법 및 지식이 필요하다. 즉, 한국어 문장에 대해 개념 그래프를 생성하기 위해서는 일반적으로 다음 두 단계에 의해 개념과 개념 관계를 표현하는 노드들을 아크로 연결하여 주어진 문장의 내부 의미를 표현해야 한다:

- ① 각 단어로부터 개념 노드를 생성
- ② 각 개념 노드들 사이에 적합한 개념 관계를 결정하고 이를 아크로 연결

이를 위해 본 논문에서는 각 단어에 대응하는 개념은 일반적 개념, 개별적 개념 그리고 특별 개념으로 구분하여 개념 사상 규칙(conceptual mapping rules)[박 95]에 의해 생성할 수 있기 때문에 개념과 개념 사이를 개념 관계로 연결하기 위해서 기본 문형 정보를 이용하여 개념 그래프를 생성하는 방법에 대해 알아본다. 예를 들어, 두 개의 문장 “좋은 생각이 그에게 떠올랐다”와 “그는 집에 갔다”에 대해 살펴보자. 두 문장 모두 조사 “**에게**”가 붙은 체언구를 문장 성분으로 가져야 정확한 의미가 표현되므로 동사 “**떠오르다**”와 “**있다**”는 $N(이)+N(에)+V$ 의 문형을 가진다고 볼 수 있으며 이에 대한 변환 규칙을 아래와 같이 기술할 수 있다.

- ㉑ $SP-11 : N(이)+N(에)+V$ (어떤 결과적 상태를 나타내는 자동사)
- ㉒ (TR-11
 $((SUBJ (P-PART ((GR 주격) (FORM 이 가))) (EXP HUMAN) (OBJ OBJECT)) G11-AGN$
 $((NPADV (P-PART ((GR 부사격) (FORM 에 에게))) (EXP HUMAN) (DST PLACE)) G11-OBJ))$

우리는 이 변환 규칙을 이용하여 각각 아래와 같은 개념 그래프를 생성할 수 있다.

- ㉓ 좋은 생각이 그에게 떠올랐다.
 (PAS) → [PROPOSITION: [RISE] -
 (EXP) → [HE]
 (OBJ) → [IDEA] → (ATT) → [GOOD]].
- ㉔ 그는 집에 갔다.
 (PAS) → [PROPOSITION: [GO] -
 (EXP) → [HE]
 (DST) → [HOUSE]].

위의 예에서 알 수 있는 바와 같이 한 문형 구조에서 하나의 문법 요소가 반드시 하나의 관계만을 가질 수 있는 것이 아니므로 우리는 <Conceptual Relation> 부분에 필요한 개념 패턴을 두어 이

를 해결하였다. 이러한 개념 패턴의 분류가 모호할 때는 우리는 그 변환 규칙을 하위 범주화함으로써 이를 해결할 것이다. 즉, 개념 [HE]가 두 개념 [HUMAN]과 [OBJECT]에 속한다고 가정한다면 위 개념 패턴을 가지고는 적절한 관계를 설정할 수 없을 것이다. 이를 해결하기 위해 우리는 변환 규칙 SP-11를 하위 분류하여 두 동사가 각각 다른 하위 규칙에 속하게 하여 이를 해결할 수 있을 것이다. 그러나 용언이 가지는 모호성을 해결하고 하위 범주화의 단계를 거치지 않고 적절히 관계를 설정할 수 있기 위해서는 각 개념 렉시콘에 개념 패턴을 기술하고 이를 생성에 이용하는 방법을 사용해야 할 것이다.

5. 결론

본 논문에서는 한국어 문장으로부터 개념 그래프를 생성하기 위해 술어의 의미적 자질에 따라 분류된 문형 정보를 기반으로 술어와 각 체언들 사이에 가질 수 있는 심층격을 설정하여 각 개념 사이에 가질 수 있는 개념 관계를 변환 규칙으로 기술하였다. 또한 이러한 변환 규칙을 이용하여 구문/의미 분석 시에 발생하는 여러 가지 모호성을 해결할 수 있음을 보였으며 개념 그래프를 생성할 수 있는 방법론을 논의했다.

그러나 술어의 의미적 차이에 따라 동일한 문형 구조를 갖는 경우는 본 논문에서 기술한 문형 정보 기반 변환 규칙이 개념 그래프를 생성하기 위한 효율적인 지식이라 볼 수 없었다. 따라서 본 논문에서는 패턴 매칭 방법을 이용하여 이를 해결할 수 있음을 보였다.

향후 연구 과제로는 본 논문에서 제안한 문형들을 검증하고 동일한 어휘에 대해 여러 개의 문형 후보 중 적용될 수 있는 변환 규칙을 선택할 수 있는 방법을 연구하는 데 있다.

참고문헌

- [1] John F. Sowa, Eileen C. Way, "Implementing a Semantic Interpreter using Conceptual Graph", IBM J. Res. Develop., Vol.30, No.1, January, pp.57-69, 1986.
- [2] John F. Sowa, Conceptual Structures : Information Processing in Mind, Machine, Addison-Weseley Pub. Com., 1984.
- [3] John F. Sowa, Conceptual Structures : Current Research and Practice, Ellis Com, pp.461-480, 1987.
- [4] Paola Velardi, Maria Teresa Pazienza, Mario De' Giovanetti, "Conceptual Graphs for Analysis and Generation of Sentence", IBM J. Res. Develop, Vol.32, NO.2, March, pp.251-267, 1988.
- [5] 강은국, 조선어 문형에 관한 연구, 박이정 출판사, 1995.
- [6] 김재훈, "중간 언어 방식을 이용한 기계 번역 에서의 한국어 격조사 생성을 위한 한국어 격들 생성", 한국과학기술원 석사 학위 논문, 1988.
- [7] 김일웅, 풀이말의 결합가와 격, 한글 제 186호, pp.35-69, 1984.
- [8] 박인철, 배우정, 이용석, "개념 그래프 생성을 위한 한국어 의미 분석기 설계", 한국정보과학회 춘계 학술 발표 논문집, 제22권 1호, 1995.