

한국어 문장이해를 위한 가변패턴네트의 구성과 응용

한 광 록[†]

요 약

문장의 개념세계는 체언과 용언으로서 이루어진다. 문장의 의미적 중심은 용언이고, 체언은 용언의 구속을 받으며, 이들은 다양하게 결합한다. 본 논문은 한국어의 문에서 체언과 용언의 결합관계를 분석하고, 형태소 해석의 결과로 부터 도출한 구 단위문을 가변패턴네트를 구성함으로써 구문 및 의미해석을 하여 절형식의 개념단위를 추출한다. 또한 종래의 제한된 혼절 이론을 일반문에 확장하고, 장문을 단문으로 자동 분리하여 논리적인 개념단위의 절 형식으로서 지식베이스를 구축하고 질문응답시스템으로의 응용을 시도하였다.

Construction of Variable Pattern Net for Korean Sentence Understanding and Its Application

Kwang Rok Han[†]

ABSTRACT

The conceptual world of sentence is composed of substantives (nouns) and verbal. The verbal is a semantic center of sentence, the substantives are placed under control of verbal, and they are combined in a various way. In this paper, the structural relation of verbal and substantives are analyzed and the phrase unit sentence which is derived from the result of morphological analysis is interpreted by a variable pattern net. This variable pattern net analyzes the phrases syntactically and semantically and extracts conceptual units of clausal form. This paper expands the traditionally restricted Horn clause theory to the general sentence, separates a simple sentence from a complex sentence automatically, constructs knowledge base by clausal form of logical conceptual units, and applies it to a question-answering system.

1. 서 론

인간이 생각하고 판단하는 것과 같은 과정을 컴퓨터에 의하여 실현하려는 것이 인공지능의 목표이다. 즉 인공지능은 전통적인 알고리즘에 기초를 둔 고정된 순서를 위주로 하는 처리 (procedure-oriented processing)의 영역을 넘어서 자기발견적인 방법 (heuristic method)으로 인간의 사고능력을 컴퓨터로 하여금 실현시키려 하는 것이다. 인간의 지적행위의 대상이 되는 것이 지식(knowledge)인 것 처럼 인공지능의 주된 대상도 지식이다. 인간의 지식에는 공학의 여러가지 법칙이나 정리 뿐만 아니라 인체의 감각기관을 통하여 얻어지는 유형, 무형의 지식과 경험적인

지식도 포함된다.

그런데 인간의 지적행위나 사고는 자연언어에 의하여 표출되므로 인공지능화의 핵심은 자연언어의 이해로 귀착된다고 할 수 있으며 따라서 인간이 컴퓨터를 사용함에 있어서 프로그램을 제거하고 자연언어로 대체하여 컴퓨터를 인공지능화하려는 것이 세계적인 추세이고 이것은 21세기 문명의 유일한 도구로 지목되고 있다[1]. 즉 컴퓨터와 인간이 자연언어로 대화하고, 컴퓨터는 인간의 자연언어를 이해함으로써 주어진 문제를 이해하여 해결하도록 하는 인간의 지적인 능력을 컴퓨터에 이식시키려는 것이다. 이를 위한 주요한 단계로서 자연언어를 이해하여 지식베이스를 구성하기 위한 연구가 미국이나 일본등에서 구문 및 의미해석등을 중심으로 활발히 진행되고 있으며, 신문의 요약과 정보추출, 전화자동통역등 실

[†] 종신회원 : 호서대학교 컴퓨터공학과 조교수

논문접수 : 1994년 12월 12일, 심사완료 : 1995년 3월 25일

용적인 시스템의 상용화에 박차를 가하고 있다 [5, 6, 7, 10, 11].

그러나 한국어를 모국어로 사용하는 우리에게 한국어를 의하여 모든 지식을 표현하고 전달하기 때문에 한국어 문장을 이해하여 그것을 컴퓨터가 이용할 수 있는 지식베이스로 구성하는 일이 무엇보다도 필요하며 한국어의 인공지능화에 관심을 갖는 것은 당연한 일이며 필수적인 것이지만, 한국어는 형태소적으로 탈락과 변형이 심하며, 또 구문구조가 대단히 복잡하고 다양하게 변하므로 컴퓨터로 처리하기에는 많은 어려움이 따르기 때문에 한국어 문장으로부터 지식표현을 표현하는 일은 매우 어렵고, 또한 언어학자들의 언어학적인 연구를 컴퓨터 상에서 직접 구현하기에는 많은 문제점이 있다.

이와 같은 어려움을 극복하기 위한 한국어 처리에 대한 연구는 특정한 영역에 있어서 영-한 기계번역 시스템, 한-일, 일-한 기계번역 시스템 등 주로 기계번역 시스템 개발을 위한 형태소 해석, 구문해석등의 기초연구로 진행하고 있으며, 문장을 지식표현하는 연구는 아직 초보적인 단계로서 의미네트워크를 이용한 연구등 몇 가지 사례가 있지만[15, 16, 17, 18, 19, 20], 이들 연구는 문장요약, 문맥이해, 휴먼 인터페이스(human-interface)등 한국어를 대상으로하는 자연언어처리에 본격적으로 적용하기에는 구문 및 의미해석을 위한 정해진 도구(tool)들이 마련되어 있지 않고, 정형적인 처리방법이 미흡하기 때문에 보다 세밀한 의미분석과 정형화된 해석도구가 필요하다.

따라서 본 논문에서는 한국어 문장을 이해하고 컴퓨터가 이용할 수 있도록 하기 위하여 문장의 동사를 중심으로 의존관계가 있는 명사구들의 의미속성과 그들의 역할에 기반을 둔 가변패턴네트를 구성하였다. 이 가변패턴네트를 이용하여 한국어 문장으로부터 절형식의 개념단위 정보를 추출하여 지식베이스를 구축하는 방법과 질문응답 시스템으로의 적용을 시도하였다.

2. 가변패턴네트의 구성

자연언어의 구문 및 의미해석은 여러가지 방법

이 있으나, 현재 자연언어처리에서 많이 사용하고 있는 방법으로는 격문법과 의존문법등이 있다. 격문법과 의존문법은 많은 공통적인 요소를 포함하고 있으며, 가장 유사한점으로는 문의 구조를 용언중심으로 다루고 있고 각 명사구는 이 용언에 대하여 어떤 역할을 하고 있다는 점에 착안하여[4, 14], 본 논문에서는 이들 격문법과 의존문법의 개념을 기초로 하여 명사의 의미속성과 격과 술어와의 관계를 해석하여 가변패턴네트를 구성하고, 이것을 이용하여 구문, 의미해석을 행한다.

2.1 가변패턴네트의 구성 배경

츨스키(N. Chomsky)의 이론구성은 말하는 사람과 듣는 사람 사이의 언어표현에 대한 이해 그 자체의 모델을 만드는 것을 직접 목표로 하지 않았지만 이와 같은 언어행동을 위한 실제적인 모델을 만들 필요가 있다. 츨스키의 변형생성문법에서는 변형의 일반법칙을 기술하였지만[3], 본 논문에서는 변형문법을 사용하지 않고, 구구조 규칙에서 표층구조를 만드는 방법과 구구조규칙에서 만들어진 문은 구구조규칙으로 인식할 수 있다는 생각을 반영한 것이다. 자연언어처리에 있어서 변형규칙을 사용하지 않고 문제를 해결한다면 그 잇점은 매우 크다. 츨스키의 표준이론에서는 심층구조에 의미가 연계되고 의미가 같아도 다른 표층구조를 갖는 것은 다른 변형규칙이 적용된 것이라는 설명원리가 있다. 본 논문에서는 변형규칙을 사용하지 않고 구구조 규칙으로만 문을 해석하기 위하여 격문법과 결합가문법에 의미속성과 격 역할을 확장시킨 가변적인 패턴네트를 구성한다. 이것은 하나의 용언에 대하여 몇 개의 명사구가 필수적으로 연결되는가, 그 명사구가 각각의 용언에 담당하는 주제관계는 어떤 것인가(Agent, Topic, Goal등)를 기술한 것으로 이점에서 가변패턴네트는 Fillmore의 격문법의 일부를 도입한다고 할 수 있다[4].

그러나 본 논문에서는 실제 수집한 문장을 중심으로이제까지의 문법이론과는 달리 용언을 중심으로 표층적으로 나타나는 격과 그 격이 필요로 하는 명사의 의미소성과의 관계도 규명함으로

써 용언이 지배할 수 있는 새로운 패턴을 구성하였다. 즉 문장의 생성에 있어서는 구구조 표시에 문법관계인 주어, 목적어등이 부여된다. 이와 같은 관점에서 기능구조를 나타내는 격구조가 유도되고, 이 기능구조에서 파생된 표층구조에 의미소성을 결합시켜 가변패턴네트를 구성함으로써 의미구조를 표현하고 여기서 구문과 의미가 해석된다.

2.2 패턴네트의 구성

한국어의 구문 및 의미를 해석하기 위한 가변패턴네트는 앞 절에서 기술한 바와 같이 문장을 용언을 중심으로 생각하고 이 용언과 결합관계에 있는 조사를 수반한 명사구와 명사의 의미속성의 짝으로서 기술한다. 의미속성이란 용언에 종속되는 명사구, 즉 actant의 의미특징을 나타냄으로써 용언과 이 actant의 선택을 제한하게 된다. 따라서 명사구인 actant가 용언에 대하여 필요불가결한 요소인가, 임의적인 요소인가에 따라 필수패턴과 임의패턴으로 구분하고 입력문을 해석하여 이 패턴을 조합하여 패턴네트를 구성한다. actant 구성요소에 대한 필수성분, 임의성분의 구별은 소거법을 이용하고 있다. 이것은 용언을 중심으로 가능한 모든 명사구 actant 중에서 각 요소를 제거하면서 문장의 성립여부를 판단한다. 필수성분은 주어나 목적어 그리고 일부 부사적 성질을 갖는 명사구이고, 임의성분은 부사적 성분이 많지만 격을 수반하는 명사가 여기에 속하는 것도 있다.

[정의1] 필수기본패턴 V,

$$V_i = \{N_i(\text{semp}) : P(X_i)\}^*$$

$$N_i(\text{semp}) = \text{명사의 의미소성}$$

$$P(X_i) = \text{조사}$$

필수패턴 V_i의 설정은 실제 문장을 수집하여 문장 중에서 나타나는 용언과 반드시 공기하는 명사구들을 조사하여 구성한 패턴이며 이것은 용언에 내재한 본질적인 구문의미에 대한 성분형성을 의미한다. 그런데 문장형성에는 필수기본패턴에 임의적으로 확장해가는 성분구조가 존재한다. 이들 임의적 성분은 용언에 선택적으로 첨가되어 세부 의미표현을 보충하는 기능을 한다. 이와 같

은 선택적인 확장기능을 임의패턴으로 정한다.

[정의2] 임의패턴 Q_i

$$Q_i = \{N_i(\text{semp}) : P(X_i)\}^*$$

임의패턴 Q_i는 필수기본패턴과 구조가 동일하지만 선택적으로 적용되어 세부적인 의미표현을 지원하는 것에 차이가 있다. 필수기본패턴과 임의패턴은 용언의 결합능력을 결정하는 것으로 이들의 상호조합에 의하여 무수히 다른 구조를 생성할 수 있다. 이러한 결합생성 능력은 자연언어의 심층구조를 면밀하게 표현하고 효율적인 자연언어처리 시스템을 구성하는 중요한 요소가 된다. 용언과 체언의 심층 결합관계로부터 표면구조로 투사하여 표면구조 상에서 체언의 역할을 규정하는 구조가 존재해야 문장의 생성이 용이해진다. 즉 심층구조와 표층구조의 인터페이스가 요구되는데 이것을 의미범주와 격결합패턴으로 정의한다.

[정의3] 격결합패턴 P_k

$$P_k = \{N_k(\text{semp}) : C_k\}^*$$

$$C_k = \text{심층격}$$

2.3 패턴네트의 모델화

패턴네트의 기본형은 노드(node)와 이들 노드를 상호연결하는 연관된 의미마크 (associative semantic arc)로 구성한다. 이때 각 노드는 개념의 총칭적인 의미를 나타내며 아크는 노드간의 상호 연결관계를 나타낸다.

한국어 문장의 의미구조를 분석하기 위하여 도입된 패턴네트는 다음과 같이 모델화된다.

$$\langle \text{Pattern Net} \rangle := \langle \text{Semantic node} \rangle \mid \langle \text{Associative arc} \rangle$$

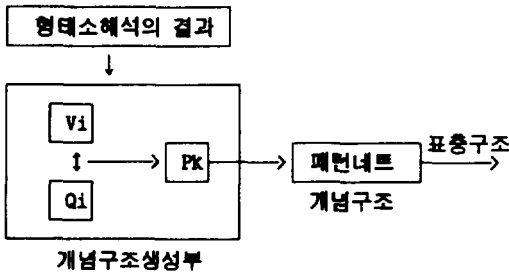
$$\langle \text{Semantic node} \rangle := \langle \text{Concept node} (V, Q_i) \mid \langle \text{Predicate node} \rangle \mid \langle \text{Argument node} \rangle$$

$$\langle \text{Associative node} \rangle := \langle \text{Concept relation} \rangle \mid \langle \text{Case relation} \rangle \mid \langle \text{Post relation} \rangle$$

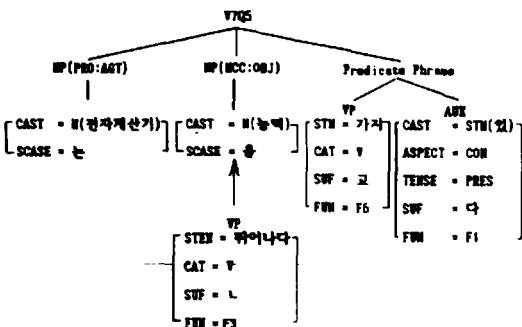
$$\langle \text{Case relation} \rangle := \langle \text{AGT} \rangle \mid \langle \text{REC} \rangle \mid \langle \text{CAU} \rangle \mid \langle \text{TOO} \rangle \mid \langle \text{PRT} \rangle \mid$$

(SPC) | (TIM) | (COR)
 | (ATT) | (SFR) |
 (STO) | (ROL) | (TOP)
 | (OBJ) | (GOA) |
 (FRE)
 (Post relation) := (CONT) | (DISJ) |
 (COND) | (MODI)

여기서 (Post relation)은 문장이 복문으로 구성될 경우 용언의 어미에 의하여 다음 문과 상호연관된 관계를 표시하는 아크이다. 이것을 적용하여 V_i, Q_i, P_i 를 상호조합하여 용언에 대한 패턴의 각 노드와 아크를 연결하여 의미 및 문법적으로 호응하는 패턴네트를 모델화한다. 이 과정은(그림 1(a))와 같으며, 이 모델에 의해 표현된 패턴네트의 표현 예들(그림 1(b))에 보인다 전자계산기는 뛰어난 능력을 갖고 있다.



(a) 가변패턴네트 모델



(b) 가변패턴네트의 예

(그림 1) 가변패턴네트의 생성과정
(Fig. 1) The procedure of a variable pattern net construction

3. 문장의 지식표현 및 실험

3.1 개념단위의 추출과 논리표현

2장에서 구성한 가변패턴네트를 이용하여 문장의 의미중심이되는 용언과 용언에 의존관계가 있는 명사구에 대한 관계를 해석함으로써 용언과 체언 사이의 결합관계를 파악한다. 이를 처리하기 위한 과정은 다음과 같다.

1단계 : 입력문의 형태소 해석의 결과로부터 문장을 구 단위문으로 분리한다.

2단계 : 가변패턴네트 해석기에 의하여 구문, 의미를 해석한 다음에 절형식의 개념단위를 도출한다.

3단계 : 대명사의 조응관계에 의해 구조적인 데이터 베이스를 구축한다

- (a) 국민들은 정직한 지도자를 존경한다.
 - (C1) 지도자는 정직하다.
 - (C2) 국민들은 지도자를 존경한다.
- (b) 예쁜 꽃들이 넓은 마당에 피어있다.
 - (C3) 꽃들이 예쁘다.
 - (C4) 마당이 넓다.
 - (C5) 꽃들이 마당에 피어있다.

(그림 2) 문장의 개념단위 분리
(Fig. 2) Separation of sentence by conceptual units.

(그림 2(a))는 두개의 용언 [정직하다]와 [존경하다]에 의해 개념단위(C1)과 (C2)로 분리되며, (그림 2(b))는 세개의 용언 [예쁘다], [넓다], [피어있다]에 의한 개념단위인 (C3)와 (C4) 그리고 (C5)로 분리할 수 있다. 이와 같은 분리는 가변패턴네트를 적용함으로써 실현한다. 이와 같이 문의 결합관계를 패턴네트를 이용하여 개념단위의 절형식으로 분리할 수 있고 특히, (C1), (C3), (C4)는 한정적 결합관계로 체언과 관형어, 또는 용언과 부사 사이에서 형성되는 2차적인 결합관계를 나타내는 것으로 개념단위의 두 주요 요소인 용언과 체언의 의미적 범위를 명확하게 제한하려는 의미표현이다. (그림 2)에 나타난 바와 같이 한국어의 문장의 기본적인 결합관계는 구조적인 결합과 한정적인 결합으로 이루어지고, 이들은 주로 문법 형태소의 변형으로 나

타난다. 따라서 본 논문에서는 가변패턴네트에 의해서 복문을 개념단위로 분리해냄과 동시에 효과적인 데이터베이스 구성과 컴퓨터처리로의 응용을 위하여 용언의 지배를 받는 개념단위들이 연결되는 논리적인 관계를 다음과 같은 규칙으로 설정한다.

[규칙] 개념의 논리관계

- (1) 하나의 용언에 대한 논리적 관계는 긍정과 부정으로 한다.
- (2) 두 용언 사이에서 논리적 결합관계는 AND관계, OR관계, IMPLIES관계로 한다.
- (3) 전성어미에 의해 삽입되는 하위문은 AND관계로 한다.
- (4) 연결어미의 논리적 결합관계는 <표 1>과 같이 분류한다.

(표 1) 연결어미의 논리적 결합관계
(Table 1) Logical relation of verbal ending

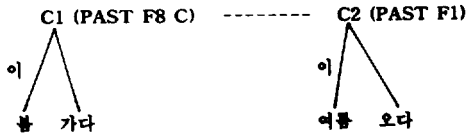
논리관계	연결어미의 종류
OR	방임형(F7), 선택형(F11)
AND	별립형(F8), 심명형(F6), 연발형(F12), 의태어미형(F13), 첨가형(F14), 의인형(F15), 의동어미형(F16), 목적형(F17), 도입어미형(F18), 반복형(F19), 관습형(F20)
IMPLIES	구속형(F9), 비교형(F10)

(그림 3)은 문장에 대한 개념단위 절형식의 논리적 관계를 보인 것이며, (그림 3(a))의 ---은 AND관계를, (b)의 ←--→는 OR관계를, (c)의 ⇒는 IMPLIES관계를 의미한다. 따라서 (그림 3(a))의 [봄이가고]와 [여름이왔다]의 두개념 사이에는 AND관계가 존재하고, (그림 3(b))의 [너는 그림을 그리거나]와 [책을 읽어라]의 사이에는 OR관계가 존재하며, (그림 3(c))의 [내일 비가 오면]은 IMPLIES관계로 연결되어 있다. 또한 (그림 3(c))의 개념단위는 부정음을 의미한다.

3.2 개념단위의 지식표현

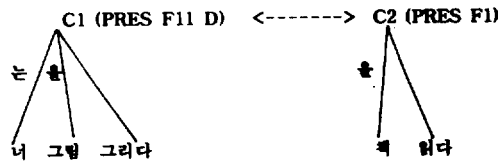
본 논문에서는 한국어의 일반문을 대상으로 하고 있기 때문에 보통명사를 수식하는 관형어들은 관형어 리스트로 구성하여 기본형의 한 인수로

봄이 가고 여름이 왔다.



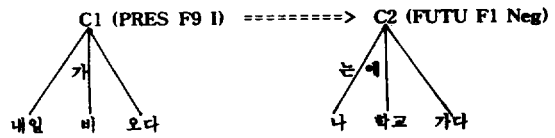
(a) AND 관계

너는 그림을 그리거나 책을 읽어라.



(b) OR 관계

내일 비가 오면 학교에 가지 않겠다.



(c) IMPLIES 관계

(그림 3) 개념의 논리적인 관계
(Fig. 3) Logical relations of concept

(a) 그 소녀는 장미꽃을 아주 좋아한다.

- (ISA (X1 소녀 (HUM (그))))
- (ISA (X2 장미꽃 (PLA)))
- ((C1)(좋아하다 (는다(T1 F1))
- (X1 (AGT는) X2 (OBJ을) (아주)))

(b) 철수는 어려운 책을 빨리 읽다.

- (C2) 책이 어렵다 (는 (T1 F3))
- (C3) 철수는 책(C2)을 빨리 읽다(는다 (T1 F1))
- (ISA (X3 책 (LIN)))
- ((C2) (어렵다 (는 (T1 F3))
- (X3 (AGT이)))
- ((C3)(읽다 (는다(T1 F1))

(철수 (HUM AGT)는 X3 (C2 OBJ 을) (빨리))

(그림 4) 절형식의 지식표현

(Fig. 4) Knowledge representation of clausal form

표현한다. 또한 용언을 수식하는 부사어들도 부사어 리스트로 구성하여 기본형의 한 인수로 표현한다. 개념단위가 절의 기본형식으로 표현된

논리적 지식베이스는 (그림 4)과 같다. (그림 4)에서 (a)는 [좋아하다]를 술어로하는 개념단위 C1이 X1인 AGT와 X2인 OBJ를 격요소로 취하고 있고, (b)에서는 술어 [어렵다]와 [읽다]에 의한 C2와 C3의 개념단위가 추출되고, 다시 C2는 C3의 X3인 OBJ를 내포적으로 수식하는 지식표현이 된다. 본 논문에서 사용된 HUM, LIN, PLA 등은 의미속성을 나타내는 것으로 참고문헌 [8,12,13]을 참조로 재분류한 것이다.

3.3 실험 및 결과고찰

모든 실험은 IF/Prolog를 이용하여 IBM PC와 SUN workstation 상에서 행하였으며, 이와 같이 구성된 지식 베이스를 이용하면, 컴퓨터에 입력된 문장을 해석하여 데이터베이스를 구축하고, 이 지식베이스를 기초로 패턴 매칭과 추론에 의해 질문에 대한 답을 찾는 질문응답 시스템을 구성하였다. 실험대상의 문장은 주로 국민학교 교과서와 정형적인 신문 기사를 중심으로 발췌한 860 문장의 200 Kbyte의 데이터를 대상으로 실험을 행하였다.

따라서 아직 충분한 양의 실험이 되었다고 할 수 없지만, 정형적인 문장에서는 가변패턴네트의 적용에 의한 절형식의 지식표현이 효과적으로 이루어진다는 것을 알 수 있었으나, 중복된 수식관계로 인한 복잡한 내포문과 도치 또는 생략등이 빈번하게 발생하는 문장등은 처리가 불가능하였다. 이는 보다 많은 문장을 대상으로 패턴네트를 작성하지 않았기 때문에 보다 많은 데이터 수집과 실험이 요구되기 때문에 한국전자통신연구소의 협조하에 480만 어절의 말뭉치(corpus)를 수집하여 계속 실험 중이다. 현재 총 860 문장에 대해서 약 86%의 성공율을 나타냈다. 참고적으로 본 연구에 사용된 사전(lexicon)은 총 엔트리(entry)는 8만 개로 크게 명사와 대명사를 포함하는 체언부, 동사와 형용사를 포함하는 용언부, 조사, 어미, 부사, 관형사 등의 6가지 부분으로 나누었고 각 부분은 필요에 따라 세분된 표식을 사용하였으며 각 용언의 엔트리마다 가변패턴네트의 형이 기록되어 있다. 그리고 이 사전의 엔트리도 12만 개로 확장 중에 있다. (그림 5)의 실험문장은 동아일보 1994년

12월 6일 33면에 나타난 기사의 일부를 발췌한 것이다.

input sentence : 의과대학에서 수술이나 해부학교육을 할 때 실감통신기술을 이용하여 인체를 직접 만지는 느낌이 나게하는 의료시뮬레이터의 개발이 실용화되면 의학용시체를 구할 필요가 없어진다.

```

---DATA BASE---
(ISA (X1 의과대학(MLC)))
(ISA (X2 수술(ACT)))
(ISA (X3 해부학교육(ACT)))
(ISA (X4 때(TIM)))
(ISA (X5 실감통신기술(NCC)))
(ISA (X6 인체(CON)))
(ISA (X7 느낌(FEL)))
(ISA (X8 개발(ACT (의료시뮬레이터(CON) 의))))
(ISA (X9 의학용시체(CON)))
(ISA (X10 필요(NCC)))
((C1) (하다 (2 (PRES, F3)
           (X1(LOC 에서) (X2 OR X3)(OBJ 을))))
((C2) (이용하다 (어 (PRES, F17)
           (X5 (OBJ 을) X4(C1))))
((C3) (만지다 (는 (PRES, F3)
           (X6 (OBJ 를) 직접(ADV))))
((C4) (나게하다 (는 (PRES, F3)
           (X7 (C3 AGT 이))))
((C5) (실용화되다(면 (FUTU, F9)
           (X8 (C4 AGT 이))))
((C6) (구하다 (2 (FUTU, F3)
           (X9 (OBJ 를))))
((C7) (있다 (다 (PRES, F1, NEG)
           (X10 (C6 AGT 가))))
    
```

QUESTION->의과대학에서 무엇을 하나?
 (하다 (2 (PRES, F3)
 (의과대학(MLC LOC 에서) ANS(ACT OBJ 을)))
 ANS = X2 OR X3 = 수술이나 해부학교육
 ANSWER->의과대학에서 수술이나 해부학교육을 한다.

QUESTION->의과대학에서 수술이나 해부학교육을 할 때 무엇을 이용하는가?
 (이용하다 (어 (PRES, F17)
 (ANS (NCC OBJ 을) (때(C1))))
 ANS = X5 = 실감통신기술
 ANSWER->의과대학에서 수술이나 해부학교육을 할 때 실감통신기술을 이용한다.

(그림 5) 실험 예
 (Fig. 5) An example of experiment

4. 결 론

본 논문은 체언과 용언으로 이루어진 한국어의

문장을 분석하고 이해하기 위하여 가변패턴네트를 구성함으로써 문장으로부터 절형식의 개념단위를 도출하여 지식베이스를 구성하고 이것을 이용하여 간단한 질문응답 시스템으로 적용가능성을 시도하였다.

용언을 문장의 중심으로 생각하고 용언과 결합할 수 있는 조사를 수반한 명사구와 그 의미소성의 짝으로 구성된 필수기본패턴과 임의패턴 그리고 격결합패턴으로 이루어진 가변패턴네트를 구성하여 구문 및 의미해석을 한다. 또한 여러 개의 문장으로 구성된 복합문에서 용언의 지배를 받는 개념단위들의 논리적 결합관계를 해석하였으며, 추출한 개념단위로부터 절 형식의 논리적 지식베이스를 구성하였다. 본 논문의 연구는 한국어 이해처리를 위한 기초단계로서 실험대상이 된 문장은 국민학교 교과서의 정형화된 문장을 중심으로 패턴을 추출하여 교과서 및 간단한 신문기사를 중심으로 실험하였으며, 비교적 문장이 간단하고 규칙에 맞는 표현에 대해서는 실패없이 분석하여 지식베이스를 구성할 수 있었지만, 복잡한 수식관계나 여러개의 문이 결합되어 내포된 복문을 처리하는 데는 아직 정확도가 떨어졌다. 따라서 앞으로 개선해야 할 과제는 현재 수집되어 있는 480만 어절의 말뭉치를 기반으로하여 보다 정형화된 가변패턴네트를 구성하고, 복잡한 수식관계로 이루어진 복문의 정확한 분리처리와 문맥정보와 화용론의 접목을 시도하고, 보다 많은 예문을 처리하여 미비점을 개선해야 할 것으로 생각된다. 그리고 현재 실험진행 중인 480만 어절의 말뭉치는 한국전자통신연구소의 데이터를 이용하고 있다.

참 고 문 헌

- [1] T.Moto-oka, Fifth Generation Computer Systems, North Holland Publishing Company Amsterdam, New York, 1982.
- [2] B. Bruce, "Case System for Natural Language", Artificial Intelligence, Vol. 6, pp. 327-360, 1975.
- [3] N. Chomsky, Aspect of the Theory of Syntax, MIT Press, Cambridge, Mass., 1965.
- [4] C.J. Fillore, "The Case for Case", In E. Bach & R.T. Harms(ed.) : Universals in Linguistics Theory, NewYork, Holt. Rinehart & Winston, 1968.
- [5] Winograd, T, Understanding Natual Language, Academic Press, 1971.
- [6] Harry Tennant, Natual Language Processing, PBI a petrocelli book, New York/Prin-ceton, 1981.
- [7] E.V. Poopv, Talking with Computers in Natural Language, Springer-Verlag, New York, 1986.
- [8] E.A. NIDA, Computatinal Analysis of Meaning, Mouton Publisher, New York, 1979.
- [9] S.H. Lee, K.R. Han. S. Ozawa, J.K. Lee, "A Study on extraction of aspect and modality information in Korean", Korean Journal of Cognitive Science, Vol. 1, No. 2, pp. 255-277, 1989.
- [10] 井佐原均, 石崎俊, "日本語新聞記事解析にをける構文情報および意味情報の抽出法", 人工知能學會誌, Vol. 3, No. 5, pp. 607-616, Sept. 1988.
- [11] 遠藤勉, "日本語ニュース文の意味解析について", NLC88-5, pp. 33-40, 1988.
- [12] 黃道三, 長尾眞, "分類語彙表からの韓國語シソ-ラスの作成", 情報處理學會論文誌, Vol. 35, No. 2, pp. 210-221, 1994.
- [13] 中井浩, 佐藤雅元, "語の収集と體言を中心とする辭書について", 自然言語處理38-7, pp. 1-6, 1983.
- [14] 石綿敏雄 et al., 文法と意味 I, 朝倉書店, 1983.
- [15] 이정현, "한국어처리를 위한 구.절 문법과 질문응답 시스템", 인하대학교박사학위청구논문, 1988.
- [16] 최병욱 외, "의미네트워크를 이용한 언어 이해과정의 구현에 관한 연구", 한국전자통신연구소 최종 연구보고서, 1989.
- [17] 이기식 외, "한·일, 일·한 자동번역시스템

개발에 관한 연구(III), 한국과학기술원 부설 시스템공학센터 연구보고서, 1987.

[18] 박창호 외, 자동번역시스템 개발에 관한 연구(I), 한국과학기술원 부설 시스템공학센터 연구보고서, 1987.

[19] 박병수, "영한 기계번역을 위한 일반 구구조 규칙의 수립", 언어 13권, 1호, 1988.

[20] 우요섭, 최병욱, "자연언어처리 및 인터페이스를 위한 기술", 한국통신학회지, Vol. 11, No. 5, pp. 46-61, 1994.



한 광 복

1984년 인하대학교 전자공학과 졸업(공학사)

1986년 인하대학교 대학원 전자공학과(공학석사)

1989년 인하대학교 대학원 전자공학과(공학박사)

1989년~91년 한국체육과학연구원 선임연구원

1991년~현재 호서대학교 컴퓨터공학과 조교수

관심분야 : 자연언어처리(한국어 정보처리), 기계번역, 전문가 시스템