

상황정보에 기반한 한국어대화의 전산적 처리와 표상구조의 구축

이 동 영[†]

요 약

한국어대화에서는 존대현상(honorification phenomenon)이 일어나기도 하고, 존대대명사(honorific pronoun)가 사용되기도 하며, 맥락상 되찾을 수 있으면 주어나 목적어가 완전히 생략되기도 한다. 이러한 특징적인 언어현상이 일어나는 한국어대화를 처리하고 그것의 표상구조를 만들기 위해서 대화참석자에 관한 정보, 발화문의 화행에 관한 정보, 대화에 관련된 사람들의 사회적 지위에 있어서의 상대적 순위에 관한 정보, 대화에 나타나는 발화문 사이의 정보흐름 등을 명시적으로가 아니라 명시적으로 표시하고 이용할 것을 본 논문은 제안한다. 또한, 본 논문은 이러한 상황정보(situational information)를 표시하고 이용하는 방법과 한국어대화의 적절한 표상구조를 제시한다. 본 논문에서 한국어대화표상구조의 설정은 담화표상이론(Discourse Representation Theory)과 분할담화표상이론(Segmented Discourse Representation Theory)을 수정·확대하여 이루어진다. 나아가서, 본 논문은 한국어대화를 전산적으로 어떻게 처리하고 그것의 표상구조를 구축하는지를 프롤로그 프로그래밍 언어를 사용하여 보여 주고나서, 그러한 표상구조의 타당성을 알아보기 위해 자연발생적 한국어대화에도 적용시켜 본다.

Computational Processing of Korean Dialogue and the Construction of Its Representation Structure Based on Situational Information

Dong-Young Lee[†]

ABSTRACT

In Korean dialogue honorification phenomenon may occur, an honorific pronoun may be used, and a subject or an object may be completely omitted when it can be recovered based on context. This paper proposes that in order to process Korean dialogue in which such distinct linguistic phenomena occur and to construct its representation structure we mark and use the following information explicitly, not implicitly: information about dialogue participants, information about the speech act of an utterance, information about the relative order of social status for the people involved in dialogue, and information flow among utterances of dialogue. In addition, this paper presents a method of marking and using such situational information and an appropriate representation structure of Korean dialogue. In this paper we set up Korean dialogue representation structure by modifying and extending DRT (Discourse Representation Theory) and SDRT (Segmented Discourse Representation Theory). Furthermore, this paper shows how to process Korean dialogue computationally and construct its representation structure by using Prolog programming language, and then applies such representation structure to spontaneous Korean dialogue to know its validity.

키워드: 한국어대화(Korean Dialogue), 상황정보(Situational Information), 담화표상이론(Discourse Representation Theory), 분할담화표상이론(Segmented Discourse Representation Theory), 대화의 표상구조(Representation Structure of Dialogue), 프롤로그 프로그래밍 언어(Prolog Programming Language), 자연발생적 대화(Naturally-occurring Dialogue)

1. 서 론

담화(discourse)에서는 나타난 문장에 상관없이 서술자와 청중이 고정되어 있지만, 대화(dialogue)에서는 발화문(utterance)마다 화자(speaker)와 청자(addressee)가 바뀌는 것이 일반적이다. 즉, 대화가 담화보다 더 역동적(dynamic)이다. 또한, 한국어대화에는 문맥에 따라 주어나 목적어가 빈번히

생략되고 존대현상이 항상 관련되어 있고 존대대명사가 쓰이기도 한다. 이러한 특징을 지닌 한국어대화를 처리하고 그것의 표상구조를 적절하게 구축하기 위해서 대화의 발화문에 나타난 표면적인 정보뿐만 아니라 대화에 관련된 상황정보를 명시적으로 이용할 것을 본 논문은 제안한다.

담화의 표상구조와 관련된 이론으로 담화표상이론(Discourse Representation Theory)[1, 2]과 분할담화표상이론(Segmented Discourse Representation Theory)[3]이 있지만, 이들 이론은 원래 담화를 다루기 위해 주창되었기 때문에, 화

[†] 종신회원: 세종대학교 영어영문학과 교수
논문접수: 2002년 7월 2일, 심사완료: 2002년 10월 1일

자와 청자가 빈번히 바뀌는 대화, 더군다나 앞에서 언급된 특징들을 지닌 한국어대화에 그대로 적용시킬 수 없는 한계성이 있다. 본 논문은 한국어대화를 적절히 표상하기 위해서 구체적으로 어떠한 정보를 어떠한 구조 속에 포함시켜야 하는지를 논의하고 한국어대화 표상구조의 뼈대와 예를 보여준다.

또한, 본 논문은 여러 가지 상황정보를 컴퓨터 상에서 나타내고 대화의 처리에 이용하는 방법과 한국어대화의 전산적 처리와 표상구조의 전산적 구축에 대해 입력과 출력의 예를 들면서 자세히 설명한다.

나아가서, 본 논문에서 제시된 표상구조가 실제 한국어대화를 나타내는 데에도 적절하고 타당성이 있는지 알아본다.

2. 담화표상에 관련된 이론

대화를 살펴보기 전에 담화를 다루고 그것의 표상구조를 구축하는 이론들을 알아보기로 한다.

2.1 담화표상이론

담화표상이론[1, 2]은 문장의 연속인 담화를 해석하여 그 결과를 담화표상구조(Discourse Representation Structure, DRS)로 나타낸다¹⁾. 이 담화표상구조는 두 가지 구성요소(즉, 담화지시체(discourse referent)와 담화표상구조조건(DRS-condition))로 이루어져 있다. 예로, <표 1>에 보여진 영어담화를 살펴보기로 한다.

<표 1> 영어담화의 예

a. A soprano sang the aria. b. The audience applauded her. c. She bowed.
--

위에 나타난 영어담화는 담화표상이론에 의해 (그림 1)에 보여진 구조로 표상된다.

p a d soprano(p) aria(a) sang(p,a) audience(d) applauded(d,p) bowed(p)
--

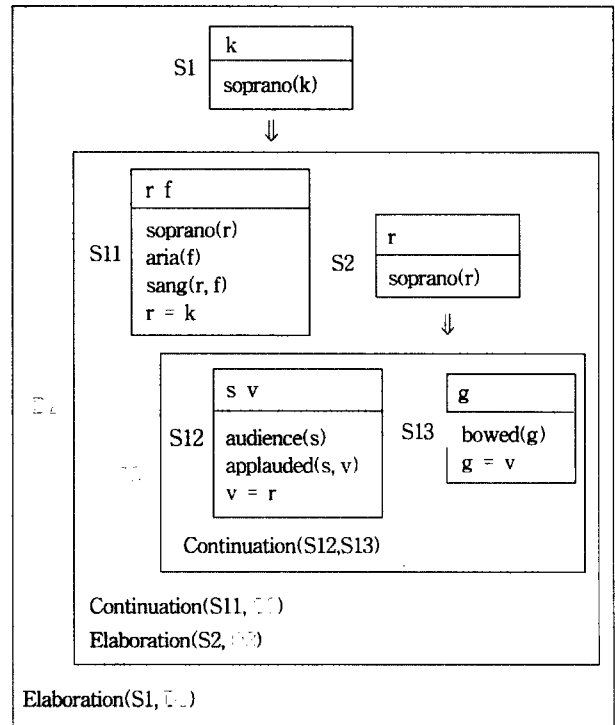
(그림 1) 담화표상구조의 예

1) Lee[4]는 주제, 초점과 같은 기능적 개념을 담화표상이론 내에 수용시키는 방법을 모색하고 영어에서의 질의-응답 쌍들을 분석하는 방법을 보여 주었다. 정소우[5]는 담화표상이론에 선형분석과 상승규칙을 추가하여 영어 조건문이나 양화사가 있는 영어 문장에서의 대명사의 지시체를 찾는 방법을 제시하였다. 그렇지만, 담화표상이론을 한국어담화나 대화에 적용시켜 본 연구는 지금까지 없었다.

그 그림에 나타난 구조는 3개의 담화지시체(즉, p, a, d)와 6개의 담화표상구조조건(한 예는 $sang(p, a)$ 임)으로 이루어져 있다. 담화표상이론에서 대명사의 해결(resolution)은 대명사에 관련된 담화표상구조조건에서 대명사의 지시체가 접근가능(accessible)하면 성공적으로 이루어진다. 예를 들면, <표 1>의 문장 b에 나타난 대명사 her과 관련된 담화표상구조조건은 $applauded(d, p)$ 인데, 지시체 p가 이 조건으로부터 접근가능하기 때문에(즉, 여기서는 동일한 담화표상구조 내에 있기 때문에) 이 대명사는 해결될 수 있다.

2.2 분할담화표상이론

분할담화표상이론[3]은 담화를 분할담화표상구조(Segmented Discourse Representation Structure, SDRS)로 나타냄으로써, 담화표상이론에서와는 달리 담화에 일어나는 각 문장을 독립된 표상구조로 나타낸다. 분할담화표상구조는 크게 두 가지 구성요소로 이루어져 있다. 하나는 담화표상구조나 분할담화표상구조이고 다른 하나는 이들 구조에 대한 분할담화표상구조조건(SDRS-condition)이다. 이 조건의 형태는 $\gamma(T1, T2)$ (여기서 γ 는 문장사이의 담화관계(discourse relation)[6, 7]를 가리키고, $T1$ 이나 $T2$ 는 담화표상구조나 분할담화표상구조를 가리킴)이다. 예를 들면, <표 1>에 이미 나타난 영어담화는 다음의 (그림 2)에 나타난 분할담화표상구조로 표상된다.



(그림 2) 분할담화표상구조의 예

(그림 2)에서 담화표상구조인 $S11, S12, S13$ 은 <표 1>에

나타난 문장 a, b, c를 각각 표상한다(represent). 계속(Continuation)이나 상세설명(Elaboration)과 같은 담화관계는 화제기반의 갱신(그림 2)에서 기호 '↓'로 표시됨을 유발하기 때문에, <표 1>의 담화의 문장과는 직접 대응하지 않는 담화표상구조 S1과 S2가 생기게 된다. 그리고 ①이나 ②는 분할담화표상구조를 나타낸다. 따라서, 담화표상구조와 분할담화표상구조를 사용하여 세분화된 구조를 설정하는 분할담화표상이론에서는 담화 속의 문장들이 각각 별개의 표상구조로 나누어져서 나타내어진다.

이 절에서 담화표상에 관한 두 가지 이론을 살펴보았다. 이 이론들이 한국어대화에 그대로 적용되어 한국어대화표상이 적절하게 이루어질 수 있는지 논의하기 전에 한국어대화의 특징을 다음절에서 알아보기로 한다.

3. 한국어대화의 특징

한국어대화의 특징을 <표 2>에 나타난 대화를 바탕으로 하여 고찰하기로 한다.

<표 2> 한국어대화의 예

a. 민 대리가 일본어를 배우니? (화자 : 박 부장, 청자 : 송 과장)
b. ㄹ 중국어를 배우요. (화자 : 송 과장, 청자 : 박 부장)
c. 민 대리, 한 부장도 중국어를 배우니? (화자 : 박 부장, 청자 : 민 대리)
d. 그분은 아랍어를 배우시어요. (화자 : 민 대리, 청자 : 박 부장)

첫째로, 한국어대화에는 존대현상이 항상 관련되어 있다. 존대현상의 발생은 대화에 관련된 사람들의 사회적 지위에 있어서의 상대적 순위에 의해 제약을 받는다. 존대는 대화에서 존대형태소(즉, 존대접미사 '님', '시'와 존대격조사 '께서', '께', 그리고 존대동사어미 '(어)요', '습니(다/까)' 등)를 통하여 실현된다. 대화에서 화자가 존대형태소를 써서 존대하는 사람의 사회적 지위는 화자보다 높다. 예를 들면, <표 2>에 나타난 발화문 d에서 존대동사어미 '어요'가 쓰였기 때문에 청자의 사회적 지위는 화자보다 높다는 정보의 추론이 가능하다. 이와 반대로, 발화문 c에서는 존대동사어미가 쓰이지 않았기 때문에 그 발화문의 화자의 사회적 지위는 청자와 같거나 더 높다는 정보를 추론할 수 있다.

둘째로, 한국어대화에서는 대화맥락으로부터 파악이 가능하면 주어나 목적어가 완전히 생략된다²⁾. 예를 들면, <표 2>에 나타난 발화문 b에서 주어가 생략되어 있다. 이 생략된 주어가 무엇인지 알기 위해서는 발화문 b보다 앞선 발

화문에서 얻은 정보를 이용해야만 한다. 심지어 어떤 경우에는 대화의 첫 번째 발화문에서도 주어나 목적어가 생략되기도 한다. 이런 경우에는 그 발화문의 화행(speech act) [8-11]에 의존하여 생략된 부분을 되찾을 수 있다. 예를 들면, 대화 첫 번째 발화문의 화행이 *허락요청*(ask-permission)이고 주어가 빠져 있으면, 이 빠진 주어는 화자를 가리킨다.

마지막으로, 존대현상과 관련하여 대화에서 존대대명사(발화문 d에 나타난 '그분'이 그 예임)가 쓰이기도 한다. 이 존대대명사의 지시체는 존대대명사가 쓰인 발화문의 화자보다 사회적 지위가 높아야 한다. 따라서, 이러한 조건을 만족시키는 어떤 지시체를 앞선 발화문들에서 얻은 사회적 지위에 관한 정보를 이용하여 찾아내는 것이 곧 존대대명사를 해결하는 것이 된다.

4. 담화표상에 관련된 이론의 한국어대화 적용에 있어서의 한계성

이 절에서는 2절에서 논의한 이론들이 한국어대화에 일어나는 특징적인 현상들을 다루고 대화표상구조를 구축하는 데 적용될 때 어떠한 한계성이 있는지 알아보기로 한다.

4.1 담화표상이론의 한계성

첫째로, 화자와 청자 같은 대화참석자에 대한 정보는 한국어대화에 일어나는 존대현상을 설명하고 생략된 부분을 찾는 데 필요하나, 담화표상이론에서는 이러한 정보를 전혀 고려하지 않고 있고 담화표상구조에도 포함시키지 않고 있다.

둘째로, 한국어대화에 나타나는 특징적인 현상을 설명하기 위해서는 어느 단일 구조가 어느 발화문을 표상하는지 명확히 구분되어 있어야 하는데, 담화표상구조에서는 그러한 구분이 되어 있지 않다.

마지막으로, 대화의 첫 번째 발화문에서 생략된 주어나 목적어는 그 발화문의 화행에 의존하여야 되찾을 수 있는데, 화행에 관한 정보가 담화표상이론에서는 고려되지 않고 있다.

4.2 분할담화표상이론의 한계성

분할담화표상이론에서는 (그림 2)에 나타난 바와 같이 어느 구조가 어느 문장을 표상하는지 구분이 되어 있는 점에서는 담화표상이론보다는 한국어대화표상에 유리하나 여전히 아래와 같은 한계성을 지니고 있다.

첫째로, 분할담화표상이론이 담화표상이론에서와 같이 대화참석자에 대한 정보를 고려하고 있지 않아서 한국어대화표상에 그대로 적용시킬 수 없다.

2) 생략된 요소는 기호 ㄹ로 표시되어 있다.

둘째로, 발화문의 화행에 관한 정보를 고려하지 않아 한국어대화의 첫 번째 발화문에서 생략된 주어나 목적어를 찾아낼 수 없다.

마지막으로, 분할담화표상이론에서 쓰이는 담화관계의 개념은, 3절에서 논의된 한국어대화에 나타나는 특징적인 현상들을 설명하는 데 아무런 역할을 하지 못한다.

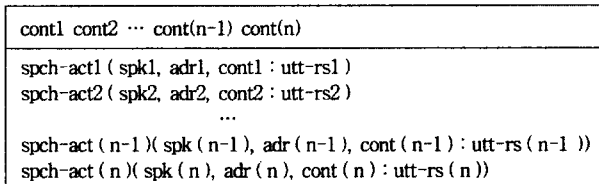
5. 한국어대화의 표상구조의 제안

앞 절에서 살펴본 바와 같이, 담화표상이론이나 분할담화표상이론은 그대로 한국어대화표상에 적용될 수 없는 한계성을 지니고 있다³⁾. 이 절에서는 담화표상이론의 담화표상구조와 분할담화표상이론의 분할 개념을 이용하여 한국어대화의 특징적인 현상을 설명할 수 있는 한국어대화표상구조를 제안하고자 한다.

한국어대화에 일어나는 존재현상을 설명하기 위해서, 그리고 존재대명사를 해결하기 위해서는 대화참석자에 대한 정보와 대화에 관련된 사람들의 사회적 지위에 있어서의 상대적 순위에 관한 정보가 대화표상구조에 명시되어야 한다⁴⁾.

한국어대화에서 생략된 주어나 목적어를 되찾기 위해서는 발화문의 화행, 앞선 발화문들에 관한 정보의 후속 발화문으로의 흐름 등이 대화표상구조에 명시되어야 한다.

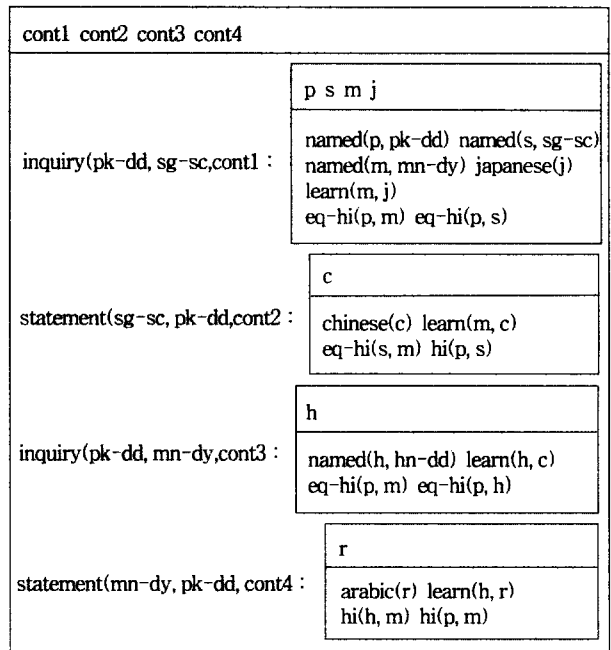
본 논문에서는 이러한 상황정보가 명시되고 정보의 흐름이 발생하도록 하기 위해 (그림 3)에 나타난 구조를 한국어대화 표상구조의 골격으로 제안한다.



(그림 3) n개의 발화문으로 구성된 한국어대화의 표상구조의 골격

(그림 3)에 나타난 바와 같이, 한국어대화에 나타나는 각 발화문은 컴퓨터 프로그래밍 언어의 하나인 프롤로그(Prolog)[20-22]의 술어(predicate)의 형태(기능항<논항1, 논항2, ..., 논항k>의 모습을 가짐)로 표상된다. 즉, 발화문의 화행은 프롤로그 술어의 기능항(functor)((그림 3)에서 spch-act2가 한 예임)으로 나타내고, 발화문의 화자와 청자는 프롤로그

술어의 첫 번째 논항(argument)(spk2가 한 예임)과 두 번째 논항(adr2가 한 예임)으로 각각 나타낸다. 발화문의 내용은 담화표상구조로 나타내고 cont라는 표지를 붙인다. 이 표지가 대화표상구조에서 대화지시체로 사용됨으로써 앞선 발화문의 내용이 뒤에 따르는 발화문의 처리에 활용될 수 있게 된다. 즉, 정보의 흐름이 일어나게 된다. 대화표상구조의 예를 들면, <표 2>에 나타난 한국어대화의 표상구조는 (그림 4)와 같다.



(그림 4) 한국어대화의 표상구조의 예

(그림 4)에 보여진 바와 같이, 대화의 표상구조는 그 대화에 나타난 발화문들에 대한 표상구조를 바탕으로 이루어지고, 각 발화문의 내용을 대화지시체로 사용하여 선행 발화문의 내용이 후행 발화문의 표상구조 구축에 활용되도록 하고 있다.

예를 들면, (그림 4)에 나타난 첫 번째 프롤로그 술어는 <표 2>에 나타난 발화문 a를 표상한다. 그 프롤로그 술어는 '발화문 a의 화행은 질의이고 화자는 pk-dd(박 부장을 가리킴)이며 청자는 sg-sc(송 과장을 가리킴)이다'라는 상황 정보를 전달하고 있다. 또한, cont1으로 표시된, 발화문 a의 내용에 대한 표상은 '발화문에 관련된 세 사람은 p, s, m으로 표시되고 p의 사회적 지위는 m과 s보다 같거나 높다'는 정보를 포함하고 있다(즉, eq-hi(p, s)는 p의 사회적 지위는 s보다 같거나 높다는 것을 의미함).

앞선 발화문에 관련된 정보가 뒤따르는 발화문의 표상구조를 구축할 때 사용되기 때문에, 어떤 발화문에서 처음으로 관련된 개체가 아니면 그 발화문의 표상구조에 지시체로 따로 표시되지 않는다. 예를 들면, (그림 4)에 나타나 있

3) 영어대화에 관한 연구[12, 13]가 있으나, 그 연구들은 영어대화에 일어나는 언어학적 현상보다는 '대화사건'이라는 개념을 사용하여 대화참석자의 정신상태(예를 들면, 의도, 목적 등)를 추론하는 데 주안점을 두고 있다.
 4) 한국어대화에 나타나는 존재현상에 대해 많은 연구(특히, [14-19])가 이루어져 왔다. 하지만, 대화 속에서 누가 존재를 받고 있는지 서술적으로만 설명이 되었고, 존재현상이 대화에 관련된 사람들의 사회적 지위의 상대적 순위의 변화에 따라 어떻게 변하는지에 대한 논의와, 상대적 순위를 형식적, 수학적으로 표시하여 존재현상과 체계적으로 접목시키려는 시도는 없었다.

는, <표 2>의 발화문 *c*에 대한 표상구조에서 *h*(한 부장을 가리킴)만 지시체로 표시되어 있는데, 선행 발화문인 발화문 *a*에 관련된 모든 지시체(즉, *p, s, m, j*)와 발화문 *b*에 새로이 도입된 지시체(즉, *c*)가 정보의 흐름에 의해 자동적으로 활용 가능하기 때문에, 이러한 것들은 다시 따로 표시되지 않는 것이다.

여기에 제시된 한국어대화 표상구조가 어떠한 측면에서 적절하고 유용한지 알아보기 위해, <표 3>에 나타난 대화를 바탕으로 생각해 보기로 한다.

<표 3> 한국어대화의 또 다른 예

a. 주말에 야구 결승전을 보았니?	
(화자: 상철, 청자: 용수)	
b. 예, 보았어요.	
(화자: 용수, 청자: 상철)	
c. 야구경기를 자주 보니?	
(화자: 상철, 청자: 용수)	
d. 아뇨, 평소에는 거의 안 봐요.	
(화자: 용수, 청자: 상철)	

위 대화를 표상하기 위해서는 생략된 구성요소가 누구를 혹은 무엇을 가리키는지, 그리고 누가 누구를 존대하는지를 알아야 한다. <표 3>의 첫 번째 발화문 *a*에서 생략된 주어를 찾기 위해서는 발화문 자체만 보아서는 불가능하고 발화문의 참석자와 화행을 알아야만 가능하다. 이러한 정보가 본 논문에서 제시된 발화문 표상구조에 포함되어 있기 때문에, 이 정보를 이용하여 생략된 구성요소를 찾는 다음, 발화문 *a*의 표상구조를 완성할 수 있게 된다.

발화문 *b*에서는 주어와 목적어 모두가 생략되어 있고 존대현상이 일어나고 있는데, 이러한 현상을 다루기 위해서는 그 발화문 자체에 관련된 상황정보뿐만 아니라 앞선 발화문에 관한 정보와 사회적 지위의 상대적 순위에 관한 정보도 이용해야 한다. 이러한 경우도 발화문의 내용을 가리키는 표지를 대화지시체로 사용함으로써 선행 발화문에서 얻은 정보가 후행 발화문에 활용되도록 하는 장치가 마련되어 있고 대화관련자들에 대한 사회적 지위에 있어서의 상대적 순위에 관한 정보가 명시되어 있는, 한국어대화 표상구조에 의해 해결이 된다.

발화문 *c*와 *d*에 일어나고 있는 주어나 목적어의 생략현상과 존대현상도 마찬가지로 원리에 의해, 제시된 대화표상구조에 나타난 여러 가지 정보를 이용하여 쉽게 다루어지고 그 발화문들의 표상구조가 완성된다.

그러므로, 제시된 한국어대화 표상구조는 한국어 대화에 포함된 발화문에서 일어나는 여러 언어적 현상을 적절하게 다룰 수 있게 하고 각 발화문의 표상구조를 구축할 수 있게 함으로, 한국어대화의 내용 파악에 유용하다.

다음절에서는 상황정보를 전산적으로 어떻게 이용하여

한국어대화를 처리하고 그것의 표상구조를 컴퓨터 상에서 구축하는지를 논의한다.

6. 한국어대화의 전산적 처리와 표상구조의 구축

6.1 상황정보를 이용한 한국어대화의 전산적 처리

한국어대화는 프롤로그 프로그래밍 언어를 사용하는 ALE (Attribute Logic Engine)[23]를 수정하고 확장하여 처리한다.

ALE의 어휘처리 부분을 확장하여 각 어휘에 관련된 상황정보를 추가로 입력한다⁵⁾. 먼저, 각 어휘에 관련된 화자와 청자에 관한 정보를 변수(variable)를 사용하여 임시로 설정하고, 실제로 그 어휘가 어떤 발화문에 쓰이면 이 변수는 상수(constant)로 실현(instantiation)되게 한다. 어휘와 관련된 화자와 청자에 관한 정보는 (그림 5)에 나타난 바와 같이 속성값 행렬(Attribute Value Matrix)[24, 25]로 표시한다.

c_idx : (speaker : Spk, addressee : ADR)

(그림 5) 화자와 청자에 관한 정보의 예

그래서 대화에 나타나는 모든 발화문에 대해서 화자와 청자가 누구인지를 수동으로 입력해주면 (그림 5)에 보여진 변수 *Spk*과 *ADR*는 그 화자와 청자에 해당되는 값(value)을 각각 갖게 된다.

또한, 존대형태소를 포함한 모든 어휘에 대해서 존대현상과 관련이 있으면 사회적 지위의 상대적 순위에 관한 정보를 변수(즉, 어휘의 지시체를 가리키는 변수, 화자를 가리키는 변수, 청자를 가리키는 변수 등)를 써서 입력하고, 그렇지 않으면 입력하지 않는다. 예를 들면, 존대동사어미 '습니까'는 화자가 청자에게 존대를 표시할 때 쓰이므로 이 형태소와 관련된 사회적 지위의 순위에 관한 정보는 (그림 6)과

5) 대화 처리를 위한 시스템의 사전에서는 각 어휘에 대해서 품사, 지표, 상황에 관한 정보 등을 포함시켜 놓고 있다. 예를 들면, 사람 이름인 '상철'에 관한 프로그램은 아래에 나타난 바와 같다.

```
sangchul --->
word,
synsem: loc : ( cat : ( head : ( noun,
                        case : na ),
                    subcat : [ ] ),
  cont : ( index : sc,
    restr : e_set ),
  conx : ( c_idx : ( speaker : SP,
                  addressee : _ ),
    soc_status :
      ( elt : ( nucleus : ( elt : ( equal_higher_stat,
                                equal_higher : SP,
                                equal_lower : sc ),
                              elts : e_set ),
                quants : [ ] ),
          elts : e_set ) ),
  ( @ empty_non_loc ).
```

각 어휘에 관련된 이와 같은 상황정보(conx라는 속성의 값으로 표시되어 있음)는 보다 상위의 구조(즉, 구(phrase), 문장(sentence), 발화문, 대화 등)로 전파되어서, 대화에 관련된 상황정보는 대화 속에 나타난 모든 어휘에서 얻을 수 있는 상황정보를 기반으로 해서 얻어진다.

같이 나타낸다.

(그림 6) 사회적 지위의 순위에 관한 정보의 예

대화에 나타나는 각 발화문의 화행에 관한 정보는 발화문에 쓰인 동사어미의 형태에 기반을 두고 파악한다. 예를 들면, 어떤 발화문이 ‘-(해)도 될까요’로 끝나면 그 발화문의 화행은 *허락요청*이 된다⁶⁾.

<표 2>의 *a*에 나타난 발화문을 처리하기 위한 입력의 형태는 (그림 7)에 나타난 바와 같다.

(그림 7) 발화문 입력의 예

위의 입력에서 볼 수 있는 바와 같이, 발화문 자체뿐만 아니라 발화문의 화자와 청자에 관한 정보도 프롤로그의 리스트(‘[]’ 형태로 표시됨)속에 함께 들어가 있다.

이러한 입력에 대해서 컴퓨터는 발화문을 처리한 결과를 그 발화문에 관련된 상황정보와 함께 (그림 7)의 입력예에 나타난 프롤로그 변수인 *AVM*의 값으로 출력해준다. 예를 들면, <표 2>에 나타난 발화문 *a*에 관련된 상황정보는 (그림 8)과 같이 출력된다. 그 그림에 나타난 바와 같이, 발화문의 처리 결과로서 발화문의 화자와 청자에 관한 정보는 *c_idx*라는 속성의 값으로, 발화문에 관련된 사람들의 사회적 지위에 있어서의 상대적 순위에 관한 정보는 *soc_status*라는 속성의 값으로, 발화문의 화행에 관한 정보는 *spch_act*

6) 표상구조 구축 시스템에서 화행은 대화의 발화문에서 생략된 구성요소가 화자를 가리키는지 아니면 청자를 가리키는지를 파악하는 데 사용된다. 이러한 목적을 위해 분류된 화행의 종류와 각 화행에 관련된 동사어미, 그리고 이를 바탕으로 파악된, 생략된 요소가 가리키는 대상의 예는 아래에 나타난 표와 같다.

화 행	동 사 어 미	생략된 요소가 가리키는 대상
요 구 (demand)	‘-해 주시겠어요’	청자
허락요청 (request-permission)	‘-(해)도 될까요’, ‘-할까요’	화자
질 의 (inquiry)	위에 나타난 어미 이외 의 의문형 동사어미	청자
진 술 (statement)	평서문을 만드는 진술형 동사어미	화자

예를 들면, 발화문 ‘기억에 남는 여행을 말씀해 주시겠어요?’에서 화행은 요구이고 생략된 주어는 청자를 가리킨다.

라는 속성의 값으로 각각 출력된다.

(그림 8) 발화문에 관련된 상황정보 출력의 예

한국어대화의 처리는 대화에 나타난 발화문을 하나씩 차례대로 처리함으로써 이루어지는데, <표 2>에 보여진 대화에 대한 입력형태는 (그림 9)에 나타난 바와 같다⁷⁾.

(그림 9) 대화처리를 위한 입력의 예

위 그림에 나타난 바와 같이, 각 발화문은 프롤로그 리스트의 형태로 입력이 되며, 전체 대화도 리스트의 형태로 입력이 된다. 표상구조 구축 시스템은 이러한 입력대화를 처리한 후, 대화에 관련된 상황정보를 비롯한 여러 가지 정보를 속성값 행렬의 형태로 자동적으로 출력해 준다.

6.2 한국어대화 표상구조의 전산적 구축

표상구조 구축 시스템은 대화에 나타나는 발화문의 처리 결과로부터 자동적으로 얻은 속성값 행렬에서 대화표상구조에 포함되어야 할 정보(예를 들면, 대화참석자, 대화에서 언급된 지시체, 대화내용 등)를 자동적으로 추출한다. 이러한 과정은 (그림 10)에 나타난 프롤로그 프로그램에 의해 이루어진다.

(그림 10) 발화문 처리와 정보추출을 위한 프로그램

이렇게 추출된 정보에서 충돌이 일어나지 않으면 발화문

7) 입력형태에서 화자와 청자에 관한 정보는 실제 이름이나 명칭이 아니라, 어휘부에서 각 개체에 미리 할당된 지표(index)(예를 들면, *pk_did*는 박 부장을 가리키는 지표임)를 사용하여 입력한다.

에서 생략된 구성요소를 복구하거나 존재대명사의 지시체를 파악한다. 그 다음에, 발화문에 관련된 대화지시체와 발화문 내용에 바탕을 둔 대화조건을 결합하여 발화문의 표상구조를 구축한다. 이러한 과정을 각 발화문을 처리할 때마다 적용하여 각 발화문에 해당되는 표상구조를 점진적으로 하나씩 만들어 나간다. 대화의 마지막 발화문에 대한 표상구조가 만들어지면, 그 때까지 구축한 발화문의 표상구조들을 연결시켜서 대화의 표상구조를 완성한다. 이러한 과정을 수행하는 표상구조 구축 시스템은 (그림 11)에 나타난 프롤로그 프로그램에 의해 실현된다.

```
dlog ( DiagLs ) : -
dlog_proc ( DiagLs, 1, [], [], DgRS, _ ),
pretty_print ( DgRS ), nl, usr_respond.

dlog_proc ( [], _ , _ , [], [], [] ) : - !.
dlog_proc ( [ Utt | RestUtts ], UttNum, AccUttCon, AccStat,
[ [ UttDomName | RestDomName ], [ UttCon | RestCon ] ],
[ UttStat | RestUttStat ] ) : -
ale_to_info ( Utt, UttInfoLs ),
recover ( UttInfoLs, UttNum, AccUttCon, AccStat, RecInfoLs ),
resolve ( RecInfoLs, UttNum, AccUttCon, AccStat, UttStat,
CollStat, ResolvInfoLs ),
create_domname ( UttNum, UttDomName ),
cons_uttcon ( ResolvInfoLs, AccUttCon, UttDomName, UttCon ),
NewUttNum is UttNum + 1,
dlog_proc ( RestUtts, NewUttNum, [ UttCon | AccUttCon ],
CollStat, [ RestDomName, RestCon ], RestUttStat ).
```

(그림 11) 표상구조의 자동적 구축을 위한 프로그램

이렇게 대화의 표상구조를 구축함으로써, 대화표상구조의 어느 부분이 어느 발화문을 나타내는지 분명하게 구획되어져서 구성요소성(compositionality)이 있고, 또한 앞선 발화문의 표상구조에 나타난 정보가 뒤따르는 발화문의 표상구조의 구축에 활용될 수도 있다.

표상구조 구축 시스템은 <표 2>에 보여진 한국어대화에 대해 (그림 12)에 나타난 표상구조를 구축해서 출력해 준다⁸⁾.

```
*** Dialogue Representation Structure ***
[ [ cont1, cont2, cont3, cont4 ],
[ inquiry ( pk_dd, sg_sc, cont1 ([ A, B, C, D ],
[ named ( A, pk_dd ), named ( B, sg_sc ),
named ( C, mn_dy ), japanese ( D ), learn ( C, D ),
eq_hi ( A, C ), eq_hi ( A, B ) ] ) ],
statement ( sg_sc, pk_dd, cont2 ([ E ],
[ chinese ( E ), learn ( C, E ),
```

8) 여기서 발화문의 내용을 가리키는 표지인 'cont1 ... cont4'는 표상구조 구축 시스템이 각 발화문을 처리한 후 대화에서 일어난 발화문의 순서에 따라 자동적으로 부여해 준다. 예를 들면, k번째 발화문의 내용에 대해서는 contk라는 표지를 부여한다. 대화에 나타난 각 개체(entity)에 대해서는 대화에 처음으로 나타난 순서에 따라 'A, B, ... Z, A1, B1, ... Z1, An, Bn, ... Zn'의 순으로 표상구조 구축 시스템이 자동적으로 변수를 할당해 준다. 예를 들면, 대화에서 다섯 번째로 나타난 개체에 대해서는 E라는 변수를 할당한다.

```
eq_hi ( B, C ), hi ( A, B ) ] ) ],
inquiry ( pk_dd, mn_dy, cont3 ([ F ],
[ named ( F, hn_dd ), learn ( F, E ),
eq_hi ( A, C ), eq_hi ( A, F ) ] ) ],
statement ( mn_dy, pk_dd, cont4 ([ G ],
[ arabic ( G ), learn ( F, G ),
hi ( F, C ), hi ( A, C ) ] ) ] ) ]

Further Representation Structure? y.

no
```

(그림 12) 한국어대화 표상구조 출력의 예

출력된 표상구조에 보여진 바와 같이, 한국어대화에 일어나는 특이한 현상을 다루는 데 필요한 정보가 각 발화문의 표상구조에 포함되어 있고, 이러한 정보를 바탕으로 여러 언어현상들을 처리한 결과도 나타나 있다.

예를 들면, <표 2>의 발화문 b에서 생략된 주어가 *민 대리*라는 것이 *learn(C, E)*에 드러나 있고(*named(C, mn_dy)*에 의해 프롤로그 변수 C는 *민 대리*를 가리키고, *chinese(E)*에 의해 변수 E는 중국어를 가리킨다는 것을 이미 알 수 있음), 그 발화문에 관련된 사람들의 사회적 지위에 있어서의 상대적 순위에 관한 정보는 *eq_hi(B, C)*와 *hi(A, B)*에 나타나 있다. 이 경우에 생략된 구성요소의 파악은 발화문 a의 표상구조에 나타난 정보를 바탕으로 이루어지는데, 그 발화문의 화행이 *질의*이고 화자와 청자는 각각 *pk_dd*와 *sg_sc*이며, 발화내용에 *learn(C, D)*가 있다는 것이 그 표상구조에 나타나 있다. 발화문 b를 처리하면서 표상구조 구축 시스템은 그 발화문의 화행이 *진술*이고 화자와 청자는 각각 *sg_sc*와 *pk_dd*이며 발화내용으로 *learn(URI, E)*가 있다는 정보를 자동적으로 추출해낸다⁹⁾. 이 정보와 발화문 a의 표상구조에 나타난 정보를 기반으로 표상구조 구축 시스템은 발화문 b는 발화문 a의 질의 내용에 대한 응답이라는 것을 파악하고 발화문 b의 내용인 *learn(URI, E)*를 발화문 a의 내용인 *learn(C, D)*에 견주어서 URI이 C와 같다는 것을 자동적으로 알아낸다. 그리하여 최종적으로 *learn(C, E)*가 발화문 b의 표상구조에서 발화내용으로 출력되어진다. 따라서, 발화문 b에서 생략된 주어는 C(즉, *민 대리*)임을 보여 준다.

발화문 d에 쓰인 존재대명사는 한 부장을 가리킨다는 사실이 *learn(F, G)*에 드러나 있으며, 각 발화문의 표상구조 사이에 정보의 흐름이 일어나기 때문에 선행하는 발화문에 나타났던 지시체는 표기되지 않고, 해당 발화문에서 처음으로 언급된 지시체만 그 발화문의 표상구조에 표기되어 있다.

따라서, 표상구조 구축 시스템에 의해 자동으로 구축된

9) 여기서 URI은 주어진 입력 대화에서 첫 번째로 생략된 구성요소에 대해 컴퓨터가 감정적으로 할당한 변수이다.

한국어대화 표상구조는 (그림 3)에 제시된 골격을 잘 반영하고 있고, 대화에 일어나는 현상들을 다루기 위해 필요한 정보를 모두 포함하고 있다.

7. 실제 한국어대화에의 적용과 유용성

본 논문에서 논의된 한국어대화 표상구조와 그 구조의 전산적 구축방법이 타당한지를 알아보기 위해, 라디오 스튜디오에서 이루어진 자연발생적 대화에 적용시켜 보았다. 실제 대화의 발화문에 나타나는 비문법적인 부분, 더듬거리는 부분, 불필요하게 반복적인 부분에 대해서는 삭제하거나 수정하였고¹⁰⁾, 그렇게 해서 얻은 발화문을 6절에서 논의된 표상구조 구축 시스템에 입력시켜 보았다. 그 시스템에 의해서 자동적으로 구축된 발화문 표상구조에 대한 결과는 <표 4>와 같다¹¹⁾.

<표 4> 자연발생적 대화에의 적용결과

	자연발생적 라디오 대화
총 발화문 수	163
자동적으로 구축된 발화문 표상구조의 수	149
성공율	91.4%

<표 4>에 나타난 바와 같이, 자연스러운 면대면 대화의 발화문 표상구조 구축에 있어서 91%의 비교적 높은 성공률을 보였는데, <표 5>에 있는 실제 대화의 일부분에 대한 실행의 예는 (그림 13)에 나타나 있다.

<표 5> 실제 대화의 일부분

a. 남자친구들하고 갔었나요? (화자 : 사회자, 청자 : 유익종)
b. 아니지요. (화자 : 유익종, 청자 : 사회자)
c. 여자친구가 있었어요? (화자 : 사회자, 청자 : 유익종)
d. 그렇지요, 그건 작게 얘기하지요. (화자 : 유익종, 청자 : 사회자)

표상구조의 자동적 구축에 실패한 경우들은 발화문에서 생략된 구성요소를 제대로 파악하지 못해서 생겼다. 예를 들면, 생략된 구성요소가 대화의 참석자들이 미리 서로 알고 있어서 대화에서 전혀 언급이 되지 않은 개체를 가리키는

경우가 있었고, 대화에서 언급된 내용을 바탕으로 추론을 해야 나오는 추상적인 개체를 가리키는 경우도 있었다. 이러한 경우에 있어서 생략된 구성요소를 찾아내기 위해서는 대화참석자의 믿음이나 배경지식에 관한 정보, 실제 세상에 관한 지식과 논리적 사고가 필요한데, 이런 요소를 수용하여 보다 더 정교한 대화표상구조를 구축하면 이러한 경우도 다룰 수 있게 될 것이다.

```

| ?- dlog ([ prsd, iij,
            namcachinkwu_tul_hako, ka_ss_ess_nayo ],
            [ iij, prsd,
              ani_ciyo ],
            [ prsd, iij,
              yecachinkwu_ka, iss_ess_eyo ],
            [ iij, prsd,
              kuleh_ciyo, kuke_n, cakkey, yaykiha_ciyo ]]).

*** Dialogue Representation Structure ***
[[ cont1, cont2, cont3, cont4 ],
 [ inquiry ( prsd, iij, cont1 ([ A, B, C ],
                               [ named ( A, prsd ), named ( B, iij ),
                                 boyfriend ( C, bf ), went ( B, C ),
                                 polite ( A, B ) ])),
   statement ( iij, prsd, cont2 ([,
                               [ not ( went ( B, C ) ),
                                 polite ( B, A ) ])),
   inquiry ( prsd, iij, cont3 ([ D ],
                               [ girlfriend ( D, gf ), existed ( D ),
                                 polite ( A, B ) ])),
   statement ( iij, prsd, cont4 ([,
                               [ existed ( D ), tell ( B, existed ( D ) ),
                                 polite ( B, A ) ]]) ] ]

Further Representation Structure? y.

no
    
```

(그림 13) 시스템에 의한 실제대화 실행의 예

본 논문이 표상구조에 포함시키고 있는 상황정보와 언어학적 정보를 바탕으로 한국어 발화문에 나타나는 언어학적 현상들을 완벽하지는 않지만 매우 성공적으로 다룰 수 있다는 점(즉, 생략된 구성요소의 회복, 존대현상의 포착, 존대대명사의 해결 등)에 본 논문의 유용성이 있다. 즉, 한국어 대화에 나타나는 발화문을 처리하고 나서 그 발화문에 관련된 상황정보와 언어학적 정보를 그 발화문의 표상구조에 포함시켜 구축하고, 다시 이러한 정보가 뒤따르는 발화문의 처리와 표상구조 구축에 도움을 주도록 정보의 흐름이 일어나도록 하고, 대화의 최종 발화문에 대한 표상구조가 만들어졌을 때 대화전체의 표상구조가 완성되도록 함으로써, 본 논문은 한국어대화의 적절한 표상구조를 제시하여 대화에 나타나는 언어현상을 매끄럽게 다루게 하고 일목요연하게 대화내용의 파악을 가능하게 하는 실제적 효과를 가져다준다.

10) 이러한 삭제나 수정은 수동적으로 이루어졌는데, 현재의 표상구조 구축 시스템에 발화문의 비문법적 부분을 문법적인 부분으로 고쳐주고 반복적인 부분을 제거해 주는 프로그램과 간투어가 들어간 발화문 등 보다 다양한 형태의 발화문을 다룰 수 있도록 하는 프로그램이 포함되면, 표상구조의 구축이 더 한층 자동적으로 이루어질 것으로 기대된다.
 11) 표상구조 구축 시스템에 의해서 발화문의 표상구조가 자동적으로 구축되지 못하는 경우에는 필요한 정보를 시스템에 추가적으로 제공하여 주었다.

또한, 본 논문은 발화문의 화자와 청자, 발화문의 화행, 발화문에 관련된 사람들의 사회적 지위에 있어서의 상대적 순위 등에 관한 정보가 존대현상과 존대 대명사를 설명하고 생략된 구성요소를 되찾는데 필수 불가결하다고 지적하고 그러한 정보를 어떠한 형태로 표상구조에 반영시키고 활용하는지를 보여 주고 있어서, 한국어대화 처리에 대한 시야를 넓혀 준다.

8. 결 론

본 논문은 한국어대화에 일어나는 특이한 현상들(주어나 목적어의 생략, 존대현상의 발생, 존대대명사의 사용 등)을 설명하고 한국어대화를 적절하게 표상하기 위해서는 대화에 관련된 상황정보들(대화참석자에 관한 정보, 화행에 관한 정보, 사회적 지위의 상대적 순위에 관한 정보)을 이용하고 이러한 정보를 표상구조에 명확하게 표시하여야 하며 정보의 흐름이 대화의 발화문 표상구조 사이에 이루어지도록 해야 한다는 것을 주장하였다. 또한, 본 논문은 이러한 주장이 수용된 한국어대화 표상구조의 기본 골격을 제시하였고, 프롤로그 프로그램을 통하여 구현된 표상구조 구축 시스템에 의해서 자동적으로 구축된 한국어대화 표상구조의 구체적인 예를 들었다. 나아가서, 본 논문은 제시된 표상구조와 그것의 구축방법을 실제 자연발생적 한국어대화에 적용했을 때에도 표상구조 구축에 있어서 91%의 성공률을 나타냈음을 보여 주었다.

본 논문이 제시한 한국어대화 표상구조는 한국어대화에서와 비슷한 언어현상이 일어나는 일본어대화(발화문의 구성요소가 생략되고 존대현상이 일어남)의 표상에도 큰 수정을 가하지 않고 적용될 수 있으리라 생각된다. 더욱이, 다른 자연언어의 대화를 표상할 때에도 그 대화에서 일어나는 언어현상에 따라 표상구조에 포함시킬 정보를 달리하면 되므로, 본 논문에 보여진 한국어대화 표상구조는 다른 자연언어의 대화표상구조에 대한 하나의 모형이 될 수 있을 것이다.

참 고 문 헌

[1] Hans Kamp, "A Theory of Truth and Semantic Representation," In Jeroen Groenendijk, Theo Janssen, and Martin Stokhof, editors, *Formal Methods in the Study of Language*, Mathematical Centre, Amsterdam, pp.277-322, 1981.
 [2] Hans Kamp and Uwe Reyle, *From Discourse to Logic : Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Kluwer Academic Publishers, Dordrecht, 1993.
 [3] Nicholas Asher, *Reference to Abstract Objects in Discourse*, Kluwer Academic Publishers, Dordrecht, 1993.

[4] Hae-Yun Lee, "Fokus-Hintergrund-Gliederung in der Diskursrepräsentationstheorie : Analysen der Frage-Antwort-Paare," *Koreanische Zeitschrift für Germanistik*, Vol.40, No.4, pp.160-179, 1999.
 [5] 정소우, "담화표상이론에서의 접근가능성에 대하여", *담화와 인지*, 제2권, 1996.
 [6] Alex Lascarides and Nicholas Asher, "Discourse Relations and Defeasible Knowledge," In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (ACL-91)*, Berkeley, pp.55-63, 1991.
 [7] Alex Lascarides, Nicholas Asher, and Jon Oberlander, "Inferring Discourse Relations in Context," In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics (ACL-92)*, Newark, pp.1-8, 1992.
 [8] Edward Finegan, *Language : Its Structure and Use*, Third Edition, Harcourt Brace College Publishers, New York, 1999.
 [9] Stephen C. Levinson, *Pragmatics*, Cambridge University Press, Cambridge, 1983.
 [10] John R. Searle, *Speech Acts : an Essay in the Philosophy of Language*, Cambridge University Press, Cambridge, 1969.
 [11] John R. Searle, *Expression and Meaning : Studies in the Theory of Speech Acts*, Cambridge University Press, Cambridge, 1979.
 [12] Massimo Poesio, "A Model of Conversation Processing Based on Micro Conversational Events," In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Pittsburgh, 1995.
 [13] Massimo Poesio and David Traum, "A Multi-Purpose Model of Conversational Context," In *Proceedings of the IJCAI Workshop on Context and Knowledge Representation*, Montreal, 1995.
 [14] 김석득, "한국어 존대형의 확대구조", *인문과학*, 제20권, 1968.
 [15] 고영근, "현대 국어의 존대법에 대한 연구", *어학연구*, 제10권 제2호, 1974.
 [16] 김형규, "국어 경어법 연구", *동양학*, 제5권, 1975.
 [17] 황적륜, "한국어 대우법의 사회 언어학적 기술", *언어와 언어학*, 제4권, 1976.
 [18] 서정수, *존대법의 연구*, 한신문화사, 서울, 1984.
 [19] Susumo Kuno and Young-Joo Kim, "The Honorific Forms of Compound Verbals in Korean," In Susumo Kuno, John Whitman, Ik-Hwan Lee, and Young-Se Kang, editors, *Harvard Studies in Korean Linguistics*, pp.178-189, Hanshin Publishing Company, Seoul, 1985.
 [20] Ivan Bratko, *Prolog Programming for Artificial Intelligence*, Third Edition, Addison-Wesley Publishing Company, Reading, 2001.
 [21] William F. Clocksin and Christopher S. Mellish, *Programming in Prolog*, Third, Revised and Extended Edition, Springer-Verlag, New York, 1987.

- [22] Leon Sterling and Ehud Shapiro, *The Art of Prolog*, Second Edition, The MIT Press, Cambridge, 1994.
- [23] Bob Carpenter and Gerald Penn, *ALE(Attribute Logic Engine)*, Version 2.01, <http://macduff.andrew.cmu.edu/ale>, 1995.
- [24] Mark Johnson, *Attribute-Value Logic and the Theory of Grammar*, CSLI Lecture Notes Number 16, CSLI, Stanford, 1988.
- [25] Stuart M. Shieber, *An Introduction to Unification-Based Approaches to Grammar*, CSLI Lecture Notes Number 4, CSLI, Stanford, 1986.

이 동 영

e-mail : dylee@sejong.ac.kr

서울대학교 영어영문학과 졸업(학사)

예일대학교 언어학과 졸업(석사)

에딘버러대학교 인지과학센터 졸업(공학 박사)

서울대학교 어학연구소 연수연구원 역임

현재 세종대학교 영어영문학과 재직

관심분야 : 대화처리, 파싱기술, 기계번역, 전산언어학, 화용론, 의미론