

# 언어기반의 인지시스템을 위한 시공간적 기초화

## Spatiotemporal Grounding for a Language Based Cognitive System

안 현 식\*  
(Hyunsik Ahn)

**Abstract** : For daily life interaction with human, robots need the capability of encoding and storing cognitive information and retrieving it contextually. In this paper, spatiotemporal grounding of cognitive information for a language based cognitive system is presented. The cognitive information of the event occurred at a robot is described with a sentence, stored in a memory, and retrieved contextually. Each sentence is parsed, discriminated with the functional type of it, and analyzed with argument structure for connecting to cognitive information. With the proposed grounding, the cognitive information is encoded to sentence form and stored in sentence memory with object descriptor. Sentences are retrieved for answering questions of human by searching temporal information from the sentence memory and doing spatial reasoning in schematic imagery. An experiment shows the feasibility and efficiency of the spatiotemporal grounding for advanced service robot.

**Keywords** : cognitive system, grounding, human-robot-interaction, intelligent robot, sentence processing, service robot

### I 서론

로봇 기술이 발전하면서 인간과 로봇이 생활공간 안에서 일상생활을 함께하는 서비스 로봇에 대한 관심이 늘어나고 있다. 이러한 서비스 로봇은 인간과 경험을 공유하고 시공간적 맥락에 의거한 상호작용을 할 수 있는 능력이 필수적으로 요구된다. 인간과 로봇 사이의 상호작용은 외부로부터의 센싱과 외부로 향한 행동, 인지 정보에 대한 기억, 그리고 맥락에 의거한 재생 등의 기능이 요구된다. 이를 위해서는 인지 정보를 적절한 형태로 기술하고 저장하는 기능이 필요하다. 또한 상호 정보를 주고받기 위해 언어를 활용하기 위해서는 로봇이 인간의 언어를 이해하고 필요에 따라 발화하는 기능이 요구된다. 이러한 기능들은 필수적으로 언어와 인지 정보의 결합 또는 연결이 요구되며 이로부터 맥락적 상호작용을 위한 정보를 이끌어 낼 수 있다. 이와 같이 일상생활을 공유하는 로봇은 감지, 행위, 기억 등과 같은 인지 정보에 대한 언어적 표현과 인간으로부터 발화된 언어를 인지적 정보와 결합시키는 언어의 기초화(grounding) 과정을 필요로 한다[1,2]. 이와 관련된 연구는 인지과학, 인지언어학과 로봇 공학 컴퓨터 과학 등의 학제간 연구 분야로서 아직은 연구의 초기단계로 볼 수 있다.

그 동안의 기초화와 관련된 연구들을 살펴보면, Roy는 인간 로봇 상호작용의 관점에서 기초화에 대한 연구를 진행하여 왔는데, Riprey 로봇에 의해 얻어지는 시각적 물체 정보를 마음영상화(mental imagery)를 통해 모델화 하고 이를 이용하여 초보적인 대화가 가능하도록 하였다[3]. 또한 에이전트가 기준이 되는 물체가 주어진 지도상에서 공간적 언어를 해석하여 명령에 따라 이동하는 실험을 보여주었으

며[4], 그래픽으로 표현한 공간속의 물체를 구(phrase) 단위로 기술할 수 있음을 보여주었다[5]. 물체의 동적인 움직임을 언어적 기호로 표현하고자 하는 연구도 기초화와 관련된다. Siskind는 동영상속에 존재하는 물체와 손의 움직임을 역학적 의미를 가지는 언어적 표현으로 기술하였는데, 손의 동작과 물체의 움직임을 공간 동작 동사를 응용한 사건논리(event logic)로 기술하였다[6]. Mooney는 축구경기 시물레이션 장면을 언어로 표현하여 방송하기 위한 목적으로 그래픽과 언어 사이의 연결에 관한 연구를 진행하고 있다[7]. 이러한 기초화 관련 연구들은 공간적 상황에 대한 구나 명령어 단위의 연구로 이루어지고 있으므로 시간과 공간적 상황을 동시에 처리할 수 있는 문장 단위의 기초화 모델이 제시되고 있지 못하며, 그 기초화된 정보를 기억으로 저장하고 재생하는 것에 대한 연구는 아직 충분히 이루어지고 있지 않다. 또한 다중 모드 대화 시스템(multimodal dialog system)의 경우 언어 외에 손동작 등과 같은 인지 정보를 결합하여 대화를 보다 원활히 하고자 하는 연구가 이루어지고 있으나[8], 로봇 등에 적용하기 위해서는 사물에 대한 인지 정보를 기반으로 한 상호 작용 분야로 확대되어야 할 필요가 있다.

한편 인지과학 분야에서는 인간의 마음의 구조와 연결하여 이를 모델화 한 인지 구조(cognitive architecture)에 대한 연구가 진행되고 있는데 인간의 기억을 분석하고 그 구조를 모방한 모델로 통해 인간의 인지를 이해하고 시물레이션하는 연구들이다[9-11]. 이 연구들은 인간의 인지 기능을 가능한 한 유사하게 모방하는 데에 일차적 목적을 두고 있으며 다양한 인지정보를 통합적으로 저장하는 형식이 존재하지 않는 등 로봇에 적용될 경우에 대한 효율성의 문제는 충분히 고려되고 있지 않다[12]. 인지 로봇(cognitive robot) 연구도 그 연장선상에서 로봇에 인지 모델을 적용한 것으로써 인간 로봇의 맥락적 상호작용보다는 인지 능력 자체에 연구가 집중되고 있다[13].

\* 책임저자(Corresponding Author)

논문접수 : 2008. 10. 9, 채택확정 : 2008. 11. 13.

안현식 : 동명대학교 로봇시스템공학과(hsahn@tu.ac.kr)

※ 이 논문은 2006학년도 동명대학교 연구년 지원에 의하여 이루어진 것임.

본 논문에서는 인간로봇 상호작용을 위한 인지 시스템의 구성을 목적으로 인지정보에 대한 언어적 표현과 언어의 인지정보와의 연결을 위한 시공간적 기초화 방법을 제안한다. 어떤 사건이 로봇의 내부 또는 외부에서 발생했을 때, 문장 형태로 그 사건을 표현하고 파싱을 거쳐 논항구조 형태로 표시하며, 구 단위로 문장을 분석하여 인지정보와 연결한다. 또한 이 문장은 기억 공간에 저장함으로써 외부의 요구에 따라 로봇이 맥락적 정보를 도출하여 재생할 수 있도록 한다. 본 논문은 이러한 일련의 흐름을 통합적으로 실현하기위한 인지 시스템의 모델을 제안하고 영어를 기반으로 하되 단문(simple sentence)으로 이루어진 문장을 이용하여 기초화 과정을 실험한다.

먼저 다음 장에서는 언어를 이용한 인지 정보를 표현하는 방법의 특징과 장점을 언급한다. 그 다음 장에서는 기초화를 위해 필요한 주어진 문장의 의미적 파싱(parsing)과 논항 분석 과정을 설명하고, 이어서 인지 시스템 모델의 구성과 그 내부 모듈에서 얻어진 인지 정보와 문장의 논항(argument)을 연결하는 시공간적 기초화를 하는 방법을 제시한다. 그 다음 장에서는 제안한 방법을 2차원 공간상의 물체를 이용하여 시공간적으로 물체를 기초화 하는 과정을 실험하고 그 결과에 대한 효용성에 대해 논한 후 결론으로 논문을 맺는다.

**II 언어 기반의 인지 시스템 모델**

인간과 로봇이 상호작용하기 위해서는 로봇이 단순히 인지된 정보나 질의에 대한 단순한 대답 뿐 아니라, 인지된 정보를 시간적으로 저장하고 이를 근거로 하여 상황의 맥락에 따라 대응할 수 있어야 한다. 그러면 인지된 정보를 어떠한 방법으로 표현하고 저장하고 재생할 것인가? 인지 정보를 표현하고 저장하고 재생하는데 언어적 표현이 효과적이라고 할 수 있다. 문장을 중심으로 한 인지 정보의 표현과 저장 추론 등은 다음과 같은 장점이 있다.

- 대화 자체가 언어적 표현이므로 인지정보의 언어적 표현은 발화나 청취 정보를 그대로 표현할 수 있으며 다른 형태로 코드화할 필요가 없다.
- 로봇이 감지하는 센싱 정보 행동 정보 뿐 아니라 내부적 추론도 독백 형태의 언어로 표현할 수 있다.
- 문장의 시간 태그를 포함한 순서적인 저장으로 통해 인지 정보를 시간적으로 저장하고 필요시 재생할 수 있다.
- 한 문장은 한 로봇이나 물체의 동작을 표현할 수 있는 최소 단위로서 의미적으로 완결적이며 사건의 기본단위가 될 수 있다.

그림 1은 인간과 로봇의 상호작용을 위해 언어를 기반으로 하여 구성한 인지 시스템 모델의 개념도를 보여주고 있는데 모든 인지 정보를 문장과 연결한 것을 볼 수 있다. 외부로부터 감지한 정보를 저장하고 어떤 행위를 하며 외부의 인간으로부터 음성을 청취하고 필요한 발화를 할 수 있는 로봇의 인지 시스템으로서 기억과 추론을 위해 인지 정

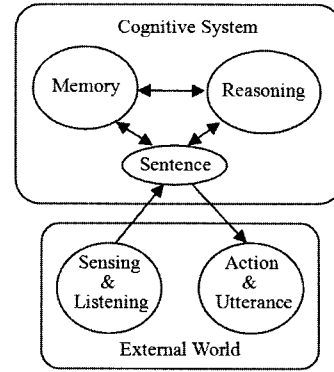


그림 1. 언어 기반의 인지 시스템의 개념도.  
Fig. 1. The concept of language based cognitive system.

보를 모두 문장으로 표현한다. 기존의 연구들이 센싱, 행동 등을 언어와 상호 연결하기 위한 연구임에 비해[3,4,6], 본 논문은 인지 정보와 언어를 연결할 하고 이를 문장을 기본 단위로 한 메모리에 저장하고 필요시 추론하고 재생하는 기능을 포함한다.

그러나 추론하는 단계에서 기억 속에 저장된 문장만으로는 한계가 있을 수 있으므로 문장의 연속으로만 표현하기 어려운 인지 정보에 대한 추론을 별도로 고려할 필요가 있다. 제안한 인지 시스템 모델은 문장을 기반으로 하되 시각 등과 같은 추가적인 인지 정보의 표현을 보조 수단으로 활용함으로써 보다 효과적인 추론이 가능하도록 한다. 따라서 인지 시스템은 인지 정보의 언어적 표현과 보조적 인지 정보를 연결하는 과정을 포함한다.

**III . 문장의 의미적 파싱**

앞 장에서 서술한 바와 같이 로봇의 인지 정보를 문장으로 표현하고 문장을 인지 정보와 연결하기 위해 필요한 의미적 파싱을 설명한다. 의미적 파싱은 먼저 문장을 통사적 파싱(syntactic parsing)을 실시하고, 의도에 따른 문장의 구조를 분류한다. 이어서 통사적 파싱에 의해 구분된 구를 하나의 의미 단위로 두고 논항을 분석한다.

**1. 통사적 파싱**

주어진 문장에 대해 통사적 파싱을 실시한다. 통사적 파서는 Penn Treebank 파서에서 제시하는 규칙에 따른다[14]. Penn Treebank는 언어분석을 위해 Wallstreet Journal의 일부분을 대상으로 하여 자동 또는 부분적 수작업으로 태깅(tagging)한 것이다. 태깅은 크게 Part-of-Speech (POS) 태깅과 통사적 태깅(syntactic parsing)으로 나누어진다. POS 태깅은 단어 품사를 밝히고 품사 태그를 붙이는 과정이며, 통사적 태깅은 문장을 구 단위의 구조로 태깅하는데 괄호를 이용하여 구를 구분하거나 나무(tree) 구조로 표현할 수 있다. 부록의 표 1은 Penn Treebank의 통사적 태그(tag) 집합을 보여주고 있다.

Penn Treebank에 따라 문장을 파싱하여 구분된 구는 몇 가지 종류의 구로 나누어지는데, 형용사구, 부사구, 전치사구, 동사구, wh-부사구, wh-명사구, wh-전치사구 등이 있다. 예를 들어 (1)은 (2)와 같이 Penn Treebank에 따라 구 단위

의 통사적 파싱이 이루어진다.

- (1) Ele went to the back of Raby at 6.
- (2) (S (NP Ele) (VP went (PP to (NP (NP the back) (PP of (NP Raby)))) (PP at (NP 6)))) .)

2. 문장 구조의 의미

문장은 문장이 쓰여지는 의도에 따라 평서문, 명령문, yes-no 의문문, wh-의문문, 감탄문 등으로 나누어진다. 이러한 문장의 구조는 로봇이 외부로부터 인식한 경우와 로봇이 내부로부터 외부로 발화하는 경우 각각 다른 의미를 가진다. 표 1은 문장의 구조와 파싱 규칙을 보여주고 있다.

2.1. 평서문

평서문은 사실에 대한 단순한 진술이다. 여기서 사실이 라는 것은 로봇 스스로 인지한 사실과 대화하고 있는 대상인 인간으로부터 전달 받은 내용도 포함된다. 로봇이 인지한 사실은 인지사실에 대한 단순한 서술과 기억 장치로부터 추론된 결과에 대한 서술로 나누어진다. (1)은 로봇의 기억의 추론에 의한 결과를 서술한 것이며 (3)은 로봇의 시각적 인지정보의 입력에 의해 인식된 물체를 서술한 것이다.

- (3) A new object appeared.

외부 대상으로 부터의 서술은 로봇에게 전달하는 대화속에서 외부의 인간의 발화를 입력한 내용을 의미한다. (4)는 인간이 로봇에게 어떤 물체의 이름을 가르쳐 주는 발화로부터 얻어진 서술이다.

- (4) The name of the object is Ele.

2.2. 의문문

wh-의문문의 경우 로봇 관점에서는 로봇이 가지고 있는 인지 시스템 내부에서 인지 정보가 기억으로 명확히 정의되지 않을 때 인간에게 질문하는 경우이다. 예를 들어 (5)는 물체의 이름이 정의되지 않았을 경우 인간에게 질의하는 경우이다.

- (5) What is the name of the new object?

외부로부터 언어모듈로 입력된 wh-의문문은 로봇에게 질

의한 대상의 정보를 인지 구조로부터 찾아내도록 요구하는 기능을 가진다. (6)은 로봇 내부의 관점에서 보면 로봇이 메모리를 이용하여 질문에 대한 답변을 추론해 낼 것을 요구하는 것으로써 물체의 이동이 언제 일어났는지를 질문하는 것이다.

- (6) When did Ele go to the back of Raby?

한편 yes-no의문문의 경우는 질문의 내용이 로봇의 인지 시스템 내에 들어있는 정보와 동일한지를 질문하는 것이다. (7)은 물체가 새로운 위치로 이동한 것의 사실 여부를 기억으로부터 확인하도록 요구하는 것이다.

- (7) Did Ele go to the back of Raby?

2.3. 명령문

외부로부터 입력된 명령문의 경우는 인간이 로봇에게 어떤 요구를 할 때 나타난다. 로봇으로 하여금 어떤 구체적인 행위를 하도록 요청하는 경우에 해당한다. 로봇이 인간에게 명령을 내리는 경우는 일반적으로 고려의 대상이 되지 않으므로 논외로 한다.

2.4. 기타 문장들

위에서 언급한 문장의 구조 외에도 기타의 예외적인 경우가 있다. 여기에는 감탄문이나 인사문 등이 포함된다. 이와 같은 문장들은 변형이 단순하므로 형식단위의 분석을 하지 않고 문장 단위로 의미를 분석한다.

3. 의미 구조의 분석

통사적 파싱의 결과로서 구분된 각 구에 대해 논항 구조의 분석과 의미역(thematic role)을 설정함으로써 문장의 의미를 분석하고 인지 정보와 연결되도록 한다. 일반적으로 영어 문장은 각 단어의 의미 중에서 동사가 문장의 구조에 가장 큰 영향을 끼친다[15]. 예를 들어 (8)에서 다른 단어들보다 동사인 gave가 가장 중요한 영향을 끼친다.

- (8) He gave an apple to her at 6.

논항구조(argument structure)는 동사의 중요성을 토대로 하여 문장의 통사구조에서 의미구조로 연결하는 기능을 한다. 논항은 한 문장을 구성하기 위해 꼭 필요로 하는 요소이지만, 부가항은 문장의 의미를 드러내기는 하지만 필수적으로 포함하지 않아도 문장이 성립될 수 있는 항을 의미한다. 예를 들어 (8)에서 'He'나 'an apple'은 'gave'가 필수적으로 동반해야 하는 논항이지만, 'at 6'는 문장의 구성에 있어서 필수적이지 않은 부가항(adjunct)이 된다. 이러한 논항들은 논항이 행할 수 있는 대표적 의미역들로 분류해 놓은 의미역(thematic role)을 설정할 수 있다. 부록의 표 2는 논항의 의미역의 종류를 보여주고 있다[16]. (8)에서 동사 'gave'는 문장의 주체에 해당하는 행위주(agent)와 목적어에 해당하는 수혜자(benefactive)인 'her'와 피험체(theme)인 'apple'로 구성된 의미역들을 수반한다.

부가항은 논항으로 포함되지 않지만 문장이 표현하고자

표 1. 문장의 구조.

Table 1. Types of sentences.

Types	Parsing rule	Examples
Declaratives	S→NP VP	The name is Ele.
Imperatives	S→VP	Stop it!
Yes-No Questions	S→Aux NP VP	Does Ele go to the back of Raby?
WH Questions	S→WH Aux NP VP	When does Ele go to the left of Raby?
Others	No rule	Oh!, Hello., Good morning.

하는 사건의 중요한 정보를 제공하여 준다. 본 논문에서는 의미역항과 부가항을 구분하지 않고 분석에 포함시키며 문장이 기술하는 사건이 로봇의 인지 정보와 구체적으로 연결되지 않을 경우 문장의 의미 자체가 완결되지 않은 것으로 간주한다. 이것은 비록 의미역항으로 구성되어 문장으로 완성된 경우라 하더라도 인지 정보와 구체적으로 연결되지 아니하면 추가적으로 필요로 하는 정보가 있어야 함을 의미한다. 만약 인지 정보를 정확히 지시하지 않을 경우 로봇 내부적으로 이를 명확히 하고자 하는 과정을 요구하여야 한다. 예를 들어 (1)의 경우 먼저 'Ele'와 'Raby'가 정의되어야 한다. 이 둘 모두 정의가 되었다고 했을 때 'to the back of Raby'가 어떤 지점 또는 영역을 의미하는지가 드러나야 한다. 다음으로 'went'라는 동사는 물체가 이 전의 위치로부터 다른 위치로 바뀌었다는 것을 의미하는 것으로 정의되어야 한다. 그러나 기억공간이 시간과 공간적 정보를 담아야 하므로 시간에 대한 정의가 이루어 지지 아니하면 그 문장을 기억공간 상에 배치할 수 없다. 따라서 로봇의 인지 시스템의 입장에서는 'at 6'가 있어야 배치가 가능하게 된다. 이를 토대로 'Ele'가 새로운 위치로 이동한 정보가 메모리상에서 정확하게 정의될 수 있다. 그렇지 않다면 대화의 흐름 속에서 시간을 충분히 추정할 수 있어야 한다. 예를 들어 (10)은 시간 정보가 들어있지 않아서 인지정보와 연결할 수 없지만 (9)로부터 그 시점을 을 추정할 수 있다.

- (9) Ele met Raby at 6 PM.
- (10) Ele went to the behind of the Raby.

**IV. 문장의 시공간 기초화**

의미적으로 파싱된 문장은 인지 시스템의 보조적 인지 정보와 연결되어야 한다. 먼저 인지 시스템의 구조를 설명하고 문장의 구 단위의 의미와 인지 정보를 연결하는 시공간 기초화 과정을 설명한다.

**1. 인지 시스템 모델**

제안하는 시공간 기초화 모델은 그림 2에 나타난 바와 같다. 시각모듈(visual module), 청취모듈(listening module), 발화모듈(utterance module)은 외부 세계와 접촉해 있으며 기억모듈(memory module)은 인지된 정보를 저장하는 기능을 한다.

기억모듈은 인간의 기억 역할을 수행하는데 장기기억

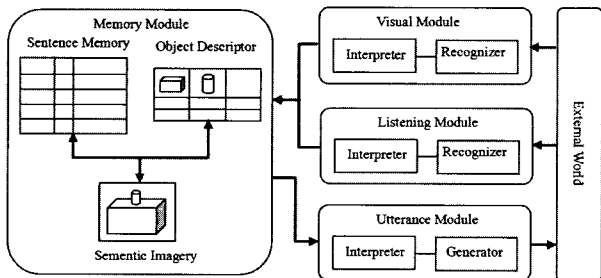


그림 2. 시공간적 기초화를 위한 인지 시스템의 구조.  
Fig. 2. The schematic of cognitive system for spatiotemporal grounding.

표 2. 문장메모리와 문장.

table 2. sentence memory and sentences.

sentence number	time	module	sentence
S <sub>1</sub>	t <sub>1</sub>	v	a new object appeared at x,y,θ.
S <sub>2</sub>	t <sub>2</sub>	u	what is the new object
S <sub>3</sub>	t <sub>3</sub>	l	it is l1

module tag: v: visual, l: listening and u: utterance modules

(long term memory)과 작업기억(working memory) 기능을 수행한다. 인지심리학에 의하면 인간의 장기기억은 의미적 기억(semantic memory), 에피소드적 기억(episodic memory), 과정적 기억(procedural memory)으로 나누어지며 작업 기억은 인지 정보를 처리하는 작업을 위한 잠정적인 기억장치이다 [17,18]. 기존의 인지구조의 연구에서는 대부분 이러한 구분에 의해 기억을 정의하고 있다[19]. 본 논문에서는 이러한 연구가 인간의 기억구조를 모방하는 것에 목표를 두는 것에 비해, 로봇의 실제적인 상호작용의 효율성에 목표를 둔 모델을 제시한다.

문장메모리(sentence memory)는 로봇의 내외부에서 일어나는 모든 사건에 대해 그 사건이 발생한 모듈과 시간 및 사건을 인지하는 모듈들이 만들어 낸 문장을 사건의 발생 순서에 따라 연속적으로 저장하는 기본적 메모리를 구성한다. 표 2는 문장메모리의 예를 보여주고 있다.

문장메모리의 언어적 표현과 함께 보다 효율적인 사건을 표현하기 위한 보조적 기억 매체를 이용한다. 시각모듈의 경우 새로운 물체가 등장했을 경우, 그 물체에 대한 실제 모양과 색깔 크기 등을 별도로 저장하는 물체묘사기(object descriptor)를 이용한다. 물체묘사기는 물체의 시각적 특성 뿐 아니라, 현재의 자세정보를 저장함으로써 문장메모리로부터 물체와 관련된 맥락을 도출해 내는데 이용한다.

도식영상(schematic imagery)은 임식적으로 시각적 기억을 저장하거나 공간적 추론을 위해 물체를 가상으로 배치시키는 역할을 한다. 평상시에는 현재의 상태에 해당하는 장면을 표현하는데, 입력된 영상 자체가 아니라 로봇이 설정한 좌표계의 공간에 물체 단위로 표현된 가상 물체를 직접 배치시킨다. 현재의 기억과 다른 상황이 발생하면 변경된 부분을 반영하여 새롭게 배치시킨다. 만약 과거 사건에 대한 공간적 추론이 필요할 경우는 문장메모리와 물체묘사기를 이용하여 과거 시점에서의 물체들의 배치에 의해 과거 공간을 구성하고 그로부터 시공간적 상황에 대해 추론한다.

시각모듈은 인간의 시각에 해당하는 역할을 하는데 영상의 입력과 물체인식 그리고 해석의 과정을 포함한다. 해석자(interpreter)는 입력된 영상을 문장형태로 표현하는 역할을 한다. 로봇이 경험하는 인지 능력은 로봇이 보유하고 있는 센서들의 차원에 따라 결정된다. 예를 들어 2차원 시각 조건을 가지고 있는 로봇은 그 조건 하에서 인식과 이해 과정이 제한되며 3차원적 정보가 제시된다 하더라도 의미를 가지지 못한다. 따라서 시각 모듈의 차원은 로봇에 따라서 다르게 정의된다.

청취모듈은 로봇이 인간으로부터 언어를 받아들이기 위한 모듈이다. 이 모듈은 음성인식, 발화 언어 이해 등을 포

함한다. 해석자는 음성의 인식을 거친 문장을 의미적 파싱과 문장과 인지 정보를 연결하는 기초화 과정을 거친다.

발화 모듈은 외부의 인간에게 로봇이 언어로 의사를 표시하기 위한 모듈이다. 내부적이거나 외부적 요청에 의해 추론된 문장을 생성하는 경우 생성기(generator)로 통해 외부로 발화된다.

## 2. 논항과 인지 정보

영어의 단문은 Penn Treebank 규칙에 따라 통사적으로 파싱했을 때, 동사를 중심으로 명사구, 전치사구, 형용사구, 부사구 등으로 구성된다. 문장의 동사는 술어로서 어떤 물체의 동작이나 이동을 표현한다. 또한 동사 자체가 내포하는 의미가 구 단위의 문장의 구조를 스스로 드러낸다. 예를 들어 (2)의 경우는 동사 'went' 가 수반하는 것은 이동하는 행위주와 그 행위주가 가는 목표를 수반한다. 이 문장은 동사를 중심으로 논항은 행위주 (11)과 목표 (12)로 구성되며 시간을 나타내는 부가항 (13)이 포함되어 있다.

(11) (NP Ele)

(12) (PP to (NP (NP the back) (PP of (NP Raby))))

(13) (PP at (NP 6))

### 2.1. 명사구

논항 중 주어나 목적어에 해당하는 명사구는 기억 속에 저장된 물체를 의미하는데, 물체묘사기에 저장된 물체들과 연결된다. (2)에서는 (11)이 명사구이며 의미역은 행위주이다.

### 2.2. 전치사구

전치사 구는 시간, 장소, 범위, 원인, 목적, 방법, 방향 등의 의미를 나타낸다. 먼저 장소를 나타내는 경우의 예를 들면 (2)의 전치사구에 해당하는 (12)의 경우 'to'는 뒤에 나오는 명사구가 목적지임을 나타낸다. 명사구 내부는 'of'의 목적어에 해당하는 'Raby'가 있으며 이것은 기억 속의 한 물체와 연결된다. 'the back'은 한 물체를 중심한 임의의 방향이나 위치를 설명한다. 따라서 (12)는 (14)라는 전치사구 패턴의 한 예라고 할 수 있다. 전치사구 (14)는  $X_1$ 이 장소나 방향을 나타내는 명사가 올 수 있으며  $X_2$ 는 기억속의 임의의 물체로 대체될 수 있으므로  $X_2$ 의 위치에 대한 상대적 위치를 나타낸다.

(14) to the  $X_1$  of the  $X_2$

본 논문에서 사용하는 위치 명사들은 기준이 되는 물체를 중심으로 하여 대상이 되는 물체의 중심점의 상호 관계로 정의하는 위치나 방향 명사를 사용한다[5].

시간적 정보는 그 문장의 일부를 구성하거나 문장의 연속적 흐름 속에서 암시되기도 한다. (2)에서 전치사구에 해당하는 (13)은 시간의 'at'가 시간의 어떤 지점을 의미하며, (15)와 같이 시간  $X_3$ 에 의해 다른 수로 대체될 수 있다.

(15) at  $X_3$

여기서  $X_3$ 는 시간을 나타내는 수사이어야 하며 이것은 문장메모리의 임의의 위치를 지시한다. 따라서  $X_3$  시점에 'Ele'가 새로운 곳으로 이동하게 된 것을 나타낸다.

### 2.3. 형용사구

형용사구의 경우는 행위주나 피형체 등에 성질을 부여할 수 있다. 예를 들어 (16)의 경우 'red'는 그 물체의 색깔을 나타내 준다. 여기서 색깔은 물체가 나타내는 중심 컬러 정보 RGB 값으로부터 표현된다.

(16) (S (NP The object) (VP is (ADJP red)) .)

### 2.4. 부사구

부사구의 경우는 동사가 의미하는 행위주의 동작에 대한 부가적인 정보를 제공해 준다. (17)에서 'slowly'는 동작 'went'의 성격을 나타내준다.

(17) (S (NP Ele) (VP went (PP to (NP (NP the back) (PP of (NP Raby)))) (ADVP slowly)) .)

## 3. 모듈 단위의 시공간 기초화

본 논문에서 제안한 언어기반 인지시스템에서는 외부 세계로부터 입력되거나 외부세계로 출력되는 인지 정보는 각 모듈을 단위로 하여 기초화된다. 각 모듈 단위의 시공간적 기초화 과정은 다음과 같이 이루어진다.

### 3.1. 시각 정보의 입력

시각 모듈은 입력 영상 내에 변화가 발생했을 때 활성화된다. 먼저 영상을 물체 단위로 분할하고 그 분할된 물체를 묘사한다. 묘사된 물체의 특징을 물체묘사기에 저장된 물체의 특징과 비교하여 동일 물체인지를 판단한다. 만약 새로운 물체이면 이를 물체의 모양과 색깔 위치 등 물체의 인지적 특징을 물체묘사기에 등록한다. 만일 동일 물체이면서도 위치나 자세가 바뀌었을 경우 이 변경된 정보를 물체묘사기의 현재위치 정보란에 저장한다. 새로운 물체일 경우에는 (18)과 같이 문장으로 표현하고 그 문장을 문장메모리에 저장한다. 위치가 변경된 경우에는 물체묘사기의 현재 위치 정보를 변경시키고 (19)와 같은 문장을 발생시켜서 문장메모리에 저장한다. 시각 모듈에서 발생한 문장은 모두 평서문 형태이며 모든 문장은 문장의 생성 시점과 모듈명이 함께 저장된다.

(18) A new object appeared at X,Y, $\theta$ .

(19) O moved to X,Y, $\theta$ .

### 3.2. 청취 정보의 입력

로봇 외부의 인간으로부터 발화된 음성은 청취모듈로 입력되는데 그 음성은 음성인식 과정을 통하여 문장으로 변환된다. 변환된 문장은 통사적 파싱과 의도에 따른 문장 구조로 분류되고 이어서 구 단위의 의미역을 정의한다. 청취 모듈은 평서문, 명령문, 의문문, 기타 등 모든 종류의 의도에 따른 문장 구조를 가질 수 있다.

평서문은 로봇에게 어떤 객관적 사실을 알려주는 역할을 한다. 새롭게 발견된 물체가 있어서 그 물체의 이름을 질문했을 때 (4)와 같은 대답을 듣고 물체를 명칭화(labeling) 할 수 있다. (20)과 같은 답변의 경우 물체묘사기 내에 들어 있는 새로운 물체의 이름(label)은 X로 설정된다.

(20) The name of the object is X.

명령문은 (21)과 같이 로봇에게 어떤 행동을 요구할 때 사용하게 된다.

(21) Go to the X1 of X2.

wh-의문문의 경우는 메모리 모듈 내에 있는 정보를 이끌어 내어 문장으로 발화하라는 의미를 가진다. (6)과 같은 문장의 경우 'Ele'가 'Raby'를 향한 움직임을 밝히기 위해 문장메모리 상에 기록된 'Ele'의 위치 정보를 현재로부터 과거로 역행하면서 검색한다. 이 문장에 기록된 위치 정보는 각 시간마다 도식영상 위에 표현하면서 'Ele'가 'Raby'의 뒤(back)에 도달한 시점을 찾는다.

yes-no 의문문의 경우 메모리 모듈 내에 있는 정보로부터 질문의 내용의 사실 여부를 찾아서 문장으로 발화하라는 의미를 가진다. (7)의 경우 (6)과 같은 동일한 과정을 거치되 'Ele'가 'Raby'의 뒤로 이동한 것이 사실인지 아닌지 여부를 찾는 과정을 요구한다.

### 3.3. 발화 정보의 출력

발화모듈은 메모리 모듈 내에 들어있는 정보로부터 외부의 인간에게 문장을 출력하는 과정이다. 발화모듈은 로봇과 인간 사이의 대화가 가지는 특성상 평서문, 의문문, 기타 문장 등의 출력이 가능하다.

평서문은 외부의 인간의 요구에 대해 답변하는 경우이거나 로봇 내부적으로 설정한 상태에 도달했을 때 인간에게 설정한 정보를 제공하기위해 발화하는 경우이다. 예를 들어 (6)과 같은 질문에 대한 답을 위해 메모리 모듈을 검색하고 그 결과를 출력하고자 할 때, (1)과 같이 답변할 수 있다. 문장을 만들어 내는 과정은 그 동사가 가지는 논항 구조를 이용하되 청취 모듈로 통해 입력된 문장의 논항들의 의미를 그대로 사용한다.

의문문의 경우는 메모리 모듈 내에 저장된 정보들이 정보의 기술이 불충분 할 경우 그 정보를 외부의 인간으로부터 받아들이기 위해 사용된다. 예를 들어 시각 모듈로부터 새로운 물체가 입력되었을 때 그 물체에 대한 이름은 스스로 결정할 수 있는 것이 아니므로 인간에게 대화로 질문하여야 한다. (5)는 새롭게 등장한 물체의 이름을 묻는 경우이다. 질문을 완료한 후 얼마 동안 그 질문에 대한 답변이 일어날 가능성이 크므로 청취 모듈이 그 질문에 대한 답변을 듣기 위한 대기 상태에 들어간다.

### 3.4. 시공간적 재생

문장메모리에 저장된 문장들은 그 문장을 생성한 시간과 함께 저장된다. 만약 사용자가 특정한 시간에 발생한 사건에 대해 질문하면 인지 시스템은 시간 기반으로 문장메모

리를 검색하여 문장을 찾고 그 문장으로부터 당시 사건에 대한 답변을 한다. 시간적 재생은 (6)이나 (22)의 질문과 같이 인간이 어떤 특정한 시점을 지적하고 그 시점에 어떤 사건이 발생했는지를 질의할 경우, 문장메모리 상에서 그 시점의 문장을 검색하고 그 문장을 과거 시제로 변환하여 재생한다.

공간적 재생은 특정한 시간에 물체나 로봇의 위치를 언어로 재생해 내는 것을 의미한다. (23)과 같이 공간상에 존재하는 여러 개의 물체간의 공간적 상호 관계를 표현하게 된다. 특정 물체의 현재의 위치를 찾고자 할 경우 로봇은 청취 모듈로부터 얻어진 물체의 이름을 이용하여 물체묘사기에 저장된 물체의 정보를 찾는다. 이로부터 공간적 추론이 이루어지며 그 결과를 문장으로 표현하고 발화모듈로 응답이 출력된다.

(22) (S What (S (VP did (VP appear (ADVP last)))) .)

(23) (S What (S (VP is (NP the foremost object))) .)

## V. 실험 및 결과

본 논문에서 제안한 시공간적 기초화 방법을 실험하기 위해 2차원 평면에 존재하는 물체들의 등장, 이동, 퇴장에 대한 시공간적 기초화의 기능과 인간 로봇 상호작용에 대해 실험하였다. 실험을 위한 시나리오는 평면의 실험 테이블 상에 임의 물체 모형을 두고 수작업으로 이를 이동시키거나 조작했을 경우 이에 대한 인지 시스템의 각 모듈의 기능과 대응을 관찰하는 것이었다. 본 실험에서는 영상 내에서 물체의 움직임이 정지되고 일정시간 유지된 영상만을 처리하여 물체가 이동 중이거나 인간의 손동작 영상은 제외하고 정지되어 있는 물체에 대해서만 다루었다. 통사적 파서는 Link Grammar Parser의 Penn Treebank 파서를 응용하여 구성하였으며[20], 입력되는 문장에 대해 자동으로 파싱이 되도록 하였다.

그림 3은 시각모듈로 통해 입력된 영상을 보여주고 있다. 여러 개의 물체가 등장하고 이동하며 퇴장하는 입력 영상이다. 입력된 영상들은 배경과 물체를 구분하는 영역분할 과정을 거친다. 영역 분할된 영상은 각각의 물체의 무게중심을 중심으로 물체의 가장자리까지의 거리를 각도단위로 표현하는 시그니처(signature)에 의해 기술된다[21]. 물체의 정보는 표 3과 같이 물체묘사기에 저장되며, 이때 물체의 모양과 색깔, 현재 중심 위치 중심에서의 거리가 최장인 화소의 각도 등이 저장된다. 동시에 기억메모리는 표 4와 같이 물체가 시각모듈로 입력된 시점과 사건을 기술한 문장을 기록한다. 또한 물체의 이동과 퇴장 및 재등장이 일어났을 경우 이에 대한 표현이 이루어지며, 외부의 사용자로부터의 질문을 의미적 파싱과 인지 정보와의 연결로 통해 해석하고 그에 따른 답변을 문장메모리와 도식영상 정보로부터 만들어 낸다. 한편 도식영상은 그림 4와 같이 묘사된 물체를 인지 시스템의 고유의 좌표계 내에 현재 시점의 물체들을 위치시킨다. 또한 공간적 추론이 필요한 경우 문장메모리 상에 저장된 내용과 도식영상 내에 저장된 물체의 정보를 이용하여 과거에 일어난 사건에 대한 맥락에 의거하

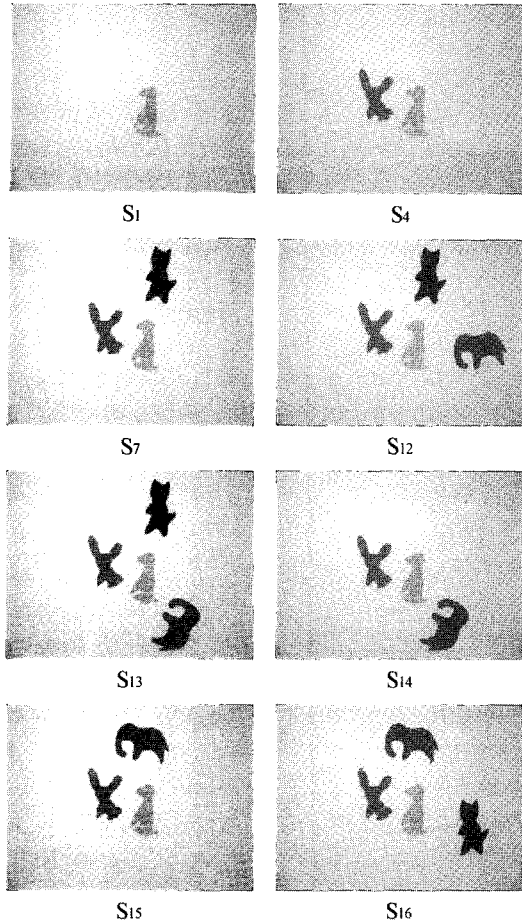






그림 3. 입력 영상.  
Fig. 3. Input images.

표 3. S<sub>12</sub>에서의 물체묘사기.

Table 3. Object descriptor at s<sub>12</sub>.

Object	Label	Shape	Color	Current Pose (x, y, $\theta$ )
O1	Puppy		Yellow	129,51,-8.1
O2	Rabi		Green	77,72,-149.0
O3	Caty		Blue	146,141,-171.0
O4	Ele		Red	210,49,-118.6

여 물체를 배치시키고 미리 정의된 방향과 위치 단어의 2차원 공간상에서의 정의를 활용하여 추론한다.

문장메모리를 중심으로 실험결과를 살펴보면 영상이 입력되면 물체를 표현한 후 이를 물체묘사기에 저장하고 이름을 붙이기 위해 물체의 이름을 질문한다.(S<sub>2</sub>,S<sub>3</sub>,S<sub>8</sub>,S<sub>11</sub>) 외부의 사용자(user)로부터 받은 대답은 문장으로 표현되며 통사적 파싱과 의미구조 분석으로 문장을 분석한 후 물체의 이름을 추출하고 물체묘사기 상에 기록한다.(S<sub>3</sub>,S<sub>6</sub>,S<sub>9</sub>,S<sub>12</sub>) 영상내의 물체를 현재 물체묘사기에 저장된 정보와 비교하

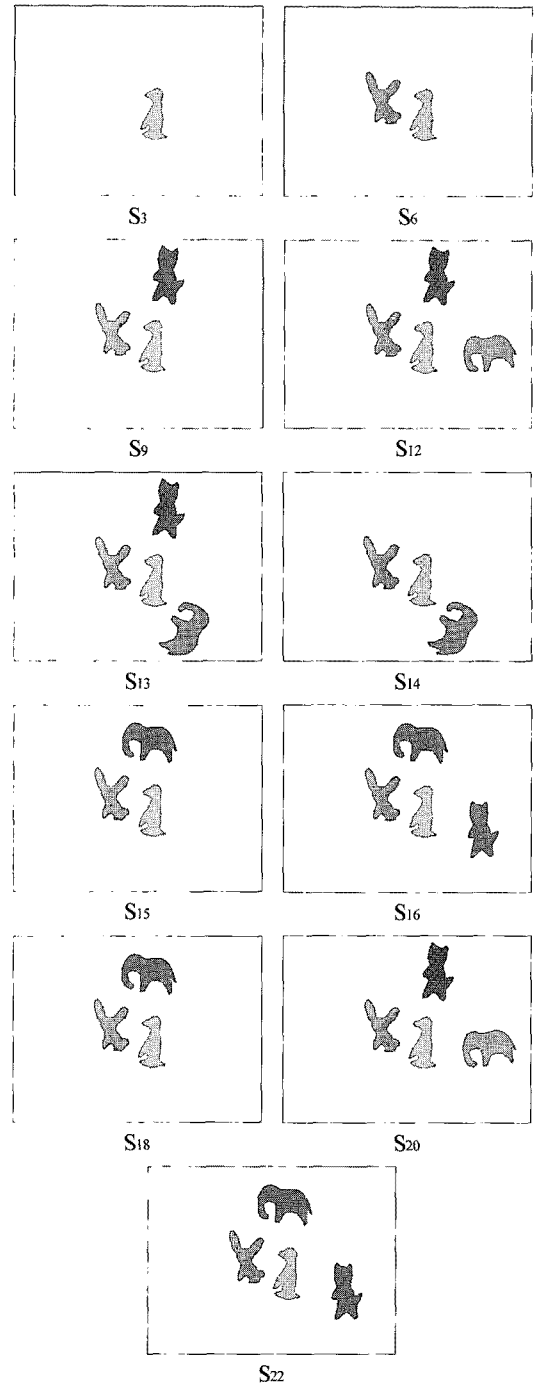


그림 4. 도식영상 내의 물체.  
Fig. 4. The objects of schematic imagery.

여 서로 상이할 경우, 물체가 이동한 것으로 보고 물체묘사기에 저장된 그 물체의 위치 정보를 갱신한다.(S<sub>13</sub>,S<sub>14</sub>,S<sub>15</sub>,S<sub>16</sub>) 시공간적 재생을 실험하기 위해 사용자가 로봇에게 한 질문에 대한 결과를 실험하였다. 시간적인 추론의 경우 문장메모리 상에서, 질의한 시간 내에 존재하는 물체와 관련된 문장을 찾고, 물체의 위치를 찾은 후 이를 기준으로 도식영상 공간상에 표현하여 당시 물체가 이동한 시점을 찾고 그 결과를 발화모듈로 출력하였다.(S<sub>17</sub>,S<sub>19</sub>) 공간적으로는 질문한 시점의 물체의 위치를 답변하거나 물체 간의 상대적인 위치를 도식영상 공간에서 추론하여 질문한 위치에

표 4. 실험의 문장메모리.

Table 4. Sentence Memory of the experiments.

#	Time	Module	Sentence
S <sub>1</sub>	9-10:40:15	V	A new object appeared at 129,51,-8.1
S <sub>2</sub>	9-10:40:18	U	what is the new orange object
S <sub>3</sub>	9-10:40:29	L	It is Pupy
S <sub>4</sub>	9-10:40:49	V	A new object appeared at 77,72,-149.0
S <sub>5</sub>	9-10:40:52	U	What is the new green object.
S <sub>6</sub>	9-10:41:1	L	It is Raby
S <sub>7</sub>	9-10:41:27	V	A new object appeared at 146,141,-171.0
S <sub>8</sub>	9-10:41:30	U	what is the new blue object
S <sub>9</sub>	9-10:41:46	L	It is Caty
S <sub>10</sub>	9-10:41:57	V	A new object appeared at 210,49,-118.6
S <sub>11</sub>	9-10:42:01	U	What is the new red object
S <sub>12</sub>	9-10:42:37	L	It is Ele
S <sub>13</sub>	9-10:43:56	V	Ele went to 170,-8,-63.4
S <sub>14</sub>	9-10:46:29	V	Caty disappeared
S <sub>15</sub>	9-10:51:55	V	Ele went to 123,142,-131.8
S <sub>16</sub>	9-10:52:8	V	Caty reappeared at 201,30,-163.4
S <sub>17</sub>	9-10:54:4	L	When did Ele go to the back of Pupy
S <sub>18</sub>	9-10:54:9	U	Ele went to the to the back of Pupy at 10:51
S <sub>19</sub>	9-10:56:22	L	What did appear last
S <sub>20</sub>	9-10:56:27	U	Ele appeared last
S <sup>21</sup>	9-10:57:47	L	What is foremost object
S <sub>22</sub>	9-10:52:51	U	Caty is foremost object

해당하는 물체의 이름을 출력하였다.(S<sub>20</sub>,S<sub>21</sub>)

**VI. 결론**

본 논문에서는 인간 로봇 상호작용을 위한 언어의 시공간 기초화를 제안하였다. 먼저 로봇의 인지 정보를 언어적으로 표현하고 이를 시간에 따라 저장하는 문장메모리를 이용하는 문장 기반의 인지 시스템의 개념을 새롭게 제시하였다. 문장을 메모리 모듈상의 로봇의 인지 정보와 연결하기 위한 통사적 파싱과 의도에 따른 문장의 구조를 분석

하였으며, 논항과 부가항 모두를 사용하는 로봇의 인지 정보 중심의 논항 분석 방법을 제시하였다. 문장의 시공간 기초화를 위해 인지 시스템을 모델링하고 각 모듈로부터 입력되거나 출력되는 문장의 의미를 분석하였으며 그에 따르는 인지 정보와의 연결 방법을 논하였다. 2차원 평면상의 물체의 등장과 이동 및 퇴장을 언어적으로 기술하고 인지 정보와 연결시키는 시공간적 기초화 방법을 실험하였으며 제안한 기초화 모델이 실제 서비스 로봇에 응용 가능한 모델임을 보였다. 제안한 방법은 인간과 로봇이 일상생활에서 상호작용할 수 있는 시공간적 기초화 방법으로서 기존의 방법에 비해 맥락에 의거한 상호작용이 가능한 방법이며 인지 시스템에 대한 접근 방법은 언어에 기반한 새로운 시도라고 할 수 있다.

본 논문의 인지 시스템은 인지의 차원이 응용 가능하고 표현 가능한 범위로 제한된다는 측면에서 기존의 컴퓨터 비전 및 영상 인식, 센서 기술 등 로봇의 인지 기능 자체에 대한 연구 결과가 뒷받침 되어야 보다 실용적인 인지 시스템으로 활용될 수 있을 것이다. 시간이 지남에 따른 문장 메모리의 증가와 다양한 동사에 대한 인지정보와 표현 문제도 추후 해결과제로 남아 있다. 또한 인지과학이나 인지언어학 등과의 학제간 연구로 통한 인간의 인지 기능과 마음 구조에 대한 이해는 보다 지능화되고 인간과 상호 소통할 수 있는 로봇을 설계하는데 중요한 바탕을 제공할 것이다.

**부록**

표 1. Penn Treebank의 통사적 파싱의 태그집합.

Table 1. Penn Treebank syntactic tagset.

1. ADJP	Adjective phrase
2. ADVP	Adverb phrase
3. NP	Noun phrase
4. PP	Prepositional phrase
5. S	Simple declarative clause
6. SBAR	Clause introduced by subordinating conjunction or 0
7. SBAEQ	Direct question introduced by wh-word or wh-phrase
8. SINV	Declarative sentence with subject-aux inversion
9. SQ	Subconstituent of SBARQ excluding wh-word or wh-phrase
10. VP	Verb phrase
11. WHADVP	wh-adverb phrase
12. WHNP	wh-noun phrase
13. WHPP	wh-propositional phrase
14. X	Constituent of unknown or certain category Null elements
1. *	"Understood" subject of infinitive or imperative
2. 0	Zero variant of that in subordinate clause
3. T	Trace-marks position where moved wh-constituent is interpreted
4. NIL	Marks position where preposition is interpreted in pied-piping contexts.



표 2. 논항의 의미역의 종류.

Table 2. Thematic roles of arguments.

1. 행위주(agent)	술어가 기술하는 행위를 의도적으로 일으키는 것
2. 피험체(Theme)	술어가 기술하는 행위로 인해 변화를 겪는 것
3. 심리경험주(Experiencer)	술어가 기술하는 정신적 심리적 상태를 겪는 것
4. 도구(Instrument)	술어가 기술하는 행위에서 쓰여지는 도구
5. 장소(Locative)	술어가 기술하는 행위가 행해지는 곳
6. 목표(Goal)	술어가 기술하는 행위가 향해지는 곳
7. 근원(Source)	술어가 기술하는 행위에서 개체가 움직여 온 곳
8. 수혜자(Benefactive)	술어가 기술하는 행위로부터 수혜를 받는 것
9. 사역주(Causer)	어떤 사건을 일으키는 것
10. 피사역체(Causee)	사역주에 의해서 일어나는 것

### 참고문헌

- [1] D. Roy, "Semiotic Schemas: A Framework for Grounding Language in Action and Perception," *Artificial Intelligence*, vol. 167, pp. 170-205, 2005.
- [2] S. Coradeschi and A. Saffiotti, "An Introduction to the Anchoring Problem," *Robotics and Autonomous System*, vol. 43, pp. 85-96, 2003.
- [3] D. Roy, K.-Y. Hsiao, and N. Mavridis, "Mental Imagery for a Conversational Robot," *IEEE SMC Part B*, vol. 34, no. 3, pp. 1374-1383, June 2004.
- [4] M. Levit and D. Roy, "Interpretation of Spatial Language in a Map Navigation Task," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 37, no. 3, pp. 667-679, 2007.
- [5] P. Gorniak and D. Roy, "Grounded Semantic Composition for Visual Scenes," *Journal of Artificial Intelligence Research*, vol. 21, pp. 429-470, 2004.
- [6] J. M. Siskind, "Grounding the Lexical Semantics of Verbs in Visual Perception Using Force Dynamics and Event Logic," *Journal of Artificial Intelligence Research*, no. 15, pp. 31-90, 2001.
- [7] R. J. Mooney, "Learning to Connect Language and Perception," *Proceedings of the 23th AAI Conference on Artificial Intelligence*, Chicago, pp. 1598-1601, July 2008.
- [8] F. Huang, J. Yang, and A. Waibel, "Dialogue Management for Multimodal User Registration," *ICSLP-2000*, vol. 3, pp. 37-40, 2000.
- [9] A. Nuxoll and J. Laird, "Extending Cognitive Architecture with Episodic Memory," *Proceedings of the 22nd National Conference on Artificial Intelligence*, 2007.
- [10] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An integrated theory of the

mind," *Psychological Review*, vol. 111, no. 4, pp. 1036-1060, 2004.

- [11] D. E. Kieras, S. D. Wood, and D. E. Meyer, "Predictive Engineering Models Based on the EPIC Architecture for a Multimodal High-performance Human-computer Interaction Task," *ACM Transactions on Computer-Human Interaction*, vol. 4, pp. 230-275, 1997.
- [12] S. D. Lathrop and J. E. Laird, "Towards Incorporating Visual Imagery into a Cognitive Architecture," *Proceedings of the Eighth International Conference on Cognitive Modeling*, Ann Arbor, 2007.
- [13] P. Ratanaswasd, W. Dodd, K. Kawamura, and D. Noelle, "Modular Behavior Control for a Cognitive Robot," *12th International Conference on Advanced Robotics(ICAR 2005)*, pp. 18-20, Seattle, July 2005.
- [14] M. P. Marcus, B. Santorini, and M. A. Marcinkiewicz, "Building a Large Annotated Corpus of English: the Penn Treebank," *Computational Linguistics*, vol. 19, 1993.
- [15] S. M. Garnsey, N. J. Pearlmuttter, E. Myers, and M. A. Lotocky, "The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences," *Journal of Memory and Language*, vol. 37, pp. 58-93, 1997.
- [16] 정태구, *논항구조와 영어 통사론*, 한국문화사, 2002.
- [17] P. N. Johnson-Laird, *Mental Models*, Cambridge, MIT Press, 1983.
- [18] E. Tulving, "Précis of Elements of Episodic Memory," *The Behavioral and Brain Sciences*, vol. 7, pp. 223-268, 1984.
- [19] W. G. Kennedy and J. G. Trafton, "Long-term Learning in Soar and ACT-R," *Proceedings of the Seventh International Conference on Cognitive Modeling*, pp. 166-171, Italy, 2006.
- [20] <http://www.abisource.com/projects/link-grammar/>
- [21] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, 2008.

### 안 현 식

제어 · 자동화 · 시스템공학 논문지 제11권 제8호 참조