

## 자연언어의 본질과 기능\*

고 창 수\*\*

### Abstract

**Ko Changsoo. 2003. The Substance and Function of Natural Language.** *Bilingual Research* 23, 1-12. This paper aims to define the Substance of the Natural Language in Natural Language Processing (NLP). Two ideas have developed that the substance of the human language is the symbolic system or language faculty. But NLP regards the language as the corpus of the linguistic element. The purpose of the NLP is the description of the process which parsing the word or sentence of the corpus. As follow, the grammar for the NLP only satisfies the observatory adequacy of the huge corpus. (Hansung University)

**[Key words]** 언어 도구론(language instrumentalism), 언어 능력(language faculty), 언어정보처리(natural language processing), 정보전달기능(communicative function)

### 1. 서론

자연언어는 인공언어에 대비되는 것으로 주로 전산언어학이나 언어정보처리 분야에서 널리 사용되는 개념이다. 전통적인 언어학에서 말하는 언어가 바로 자연언어인 셈이다. 그러나 일반 언어학에서 다루는 언어의 개념과 언어정보처리에서 다루는 자연언어의 개념은 꼭 일치하지 않는다. 이는 연구 목표와 방법론의 차이에서 기인한 것이다. 이 연구에서는 언어가 자연언어로 통칭되는 언어정보처리 분야에서 의미하

---

\* 본 연구는 2003학년도 한성대학교 교내연구비 지원 과제임.

\*\* 한성대학교

는 언어 즉 자연언어의 본질과 기능을 살핌으로써 언어정보처리와 일반언어학의 차이와 언어과학의 발전 방향에 대해 논의하고자 한다.

인간은 오랫동안 언어를 탐구해 왔다. 그 모든 탐구가 과학적 활동인 것은 아니지만, 나름대로 언어의 본질과 기능에 대한 숙고의 역사인 것은 분명하다. 고대 철학에서는 주로 언어의 본질론과 협약론이 대립하였고, 20세기 언어 과학에서는 능력론과 기호론이 대립되었다. 20세기 초반의 구조주의 언어학은 언어 기호론 및 도구론을 중심으로 발전되어 왔지만, 중반기의 생성문법은 언어 능력론을 표방하였다. 양 방향의 연구가 언어에 대한 다른 개념을 가질 수밖에 없는 것은 피할 수 없다. 언어 도구론은 언어를 객관적 자료체의 하나로 간주하는 반면, 언어 능력론에서는 언어를 인간의 생물학적 능력의 일부로 간주하기 때문이다. 이러한 대립에 따라 연구 목표와 방법에 차이를 가지게 된다. 언어 도구론에서는 개별 언어의 문법을 기술을 목표로 하는 반면, 언어 능력론에서는 추상적인 보편 문법을 형식화하는 데에 주력한다.

언어정보처리의 자연언어는 언어 능력론의 입장보다는 언어 도구론에 기대는 개념이다. 정보처리의 대상이 되는 자료가 유한할 뿐 아니라, 컴퓨터가 인간과 거의 같은 능력을 발휘하기도 어렵기 때문이다. 먼저 구조문법이 주장하는 언어 도구론과 생성문법이 주장하는 언어 능력론의 개념을 비교하면서 언어정보처리의 자연언어의 본질적 개념을 궁구해 보기로 한다.

## 2. 언어 도구론 대 언어 능력론

언어 도구론에서 언어는 '의사소통의 도구'이며, '기호의 체계'이다. 즉 언어는 '체계적인 기호를 이용하여 의사를 전달하는 도구'라는 관점이다. 도구라는 말 자체가 의미하는 것은 언어는 인간의 사고 활동 밖에 존재하는 객관물이라는 것이다. 여기에 언어 본질론은 언어와 사고가 구별되지 않는다는 '언어 대 사고의 일원론'이라는 입장을 견지한다.<sup>1)</sup> 도구론이 언어의 시니피앙적 측면을 강조하는 것이라면 본질론은

1) 여기서 언어 본질론은 앞서 언급한 언어 능력론과는 다른 개념이다. 초기의 구조주의에서 언어와 사고의 상호 관계에 대한 여러 언급이 있었는데, 독일식 구조주의에서는 언어가 곧 정신 세계라는 입장을 견지하였다. 일반의미 이론에서도 언어와 사고의 일원적 사고가 유

시니피에적 측면을 강조한다.

그러나 현상을 좀 더 현상적으로 바라보면 도구론이라는 개념에 혼란이 야기된다. 우리는 앞서 언어의 정의에서 '도구'란 특정한 '객관물'이라는 점을 의미한다고 보았다. 이제 이 특정한 객관물이란 말을 '체계적으로 구성된 기호의 집합'이라고 구체적으로 지적해 보자. 이 때 '기호'는 추상적인 상징이다. 그것이 객관화된 실체는 '언어음'과 '문자'이다. '언어음'과 '문자'는 대부분 일대일의 대응관계를 갖는다. 그것은 '언어음'이 분명한 물리량을 가지며, '문자'는 인위적 기호 체계이기 때문이다. 아마 엄밀하게 일대일 대응하지 못한다고 하는 것은 '언어음'이 인위적인 고안물이 아니기 때문이다.<sup>2)</sup>

여기서 문자가 인위적 고안물이라는 것은 중요한 시사점을 갖는다. 모든 언어 기호와 관련된 객관물 중에서 문자만이 폐쇄성과 분명한 체계성을 갖는 것은 '문자'는 확실히 역사적 존재이기 때문이다. 즉 문자는 인공적이라는 점에서 확실히 도구성을 갖지만, '언어음'은 자연적 산물이라는 점에서 문자에 비해 폐쇄성과 체계성이라는 객관물 및 도구적 성질을 덜 갖추게 된다는 점이다. 이러한 점에서 언어가 갖는 시니피에는 언어음보다는 좀 더 객관적이고 도구적 성격에서 멀다고 할 수 있다.

츨스키(1965)는 언어가 가진 이러한 자연물의 속성을 '언어의 창조성'이라고 보았다. 츨스키 입장에서 언어의 창조성은 문장의 무한성이며, 이러한 무한성은 객관물의 존재 영역을 넘어서는 것이다. 인간인 어린 아이는 누구나 언어를 배울 수 있다. 그러나 인간이 아닌 어떤 존재도 언어를 배울 수 없다. 언어가 단순한 객관물이며 도구라면, 인간 언어의 일부와 유사한 어떤 체계가 다른 자연물에 나타날 수 있는 것이다. 이러한 의미에서 인간인 어린 아이는 날 때부터 언어 습득 능력(LAD language aquisition device)을 갖고 태어난다고 보는 것이다. 그가 언어를 단순한 도구로 보지 않고 '생득적 능력'이라고 보는 이유가 여기에 있다. 언어가 능력이라면, 언어는 더

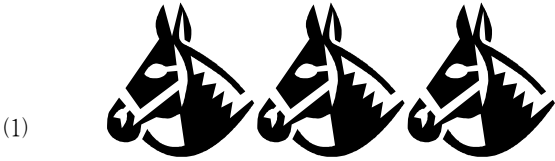
---

행하였으나, 언어 과학의 주류는 역시 언어를 객관적 기호 체계로 다룸으로써 의미를 배제하는 것이었다.

2) 언어음은 언어의 발생과 함께 나타난 자연물이지만, 문자는 인간 역사 발전에서 출현한 인공물이다. 문자는 언어음을 효과적으로 기술한 과학적 고안물이기 때문에 언어음에는 문자가 기술할 수 없는 영역이 존재하게 마련이다.

이상 도구일 수 없지만, 이 능력이 도구로 사용될 수는 있다. 마치 팔이 신체의 일부이지만, 밥을 먹거나 물건을 드는 도구로 보지 않는 것과 같은 것이다.

이러한 관점에서 도구론과 능력론의 대립은 어쩌면 보다 본질적인 문제를 궁구하는 것과 실제적인 분석적 문제에 천착하는 것 사이의 문제일 수 있다. 여기서 우리는 인간 언어의 무한성 혹은 창조성이라는 점에 기대어 언어를 인간 능력의 일부 특히 인간 사고 능력의 일부로 볼 수 있을 것이다. 그 능력은 다름 아닌 객관 세계를 분절하여 기호 체계와 일대일 대응하는 일종의 추상화 능력이다. 우리는 문자 발명의 초기 단계에서 이러한 추상화 능력의 일단을 엿볼 수 있다.



(2) 말 III

(3) 말 3

실제 객관 세계에 ‘세 마리의 말’이 있다고 가정해 보자. (1)은 이를 직접적으로 대응시킨 것이다. 그러나 이러한 대응이 가능하기 위해서는 객관 세계의 ‘말’을 ‘말’이라는 말로 구별하는 것이 필요했다.<sup>3)</sup> 인간은 문자를 이용하여 이러한 대응 관계를 (3)와 같은 단순한 방법으로 또 다른 의미의 개념적 이중 분절을 시도한다. 이는 수학적 의미에서 “III에 3이라는 집합적 이름을 붙이는 이름 짓기”인 셈이다. 기원전 4000년경의 가장 오래된 문자는 바로 이러한 객관 세계의 이름과 그 이름을 가진 것의 개수를 보여주는 것이다.<sup>4)</sup>

---

3) 이것이 바로 이름짓기인데 자연언어는 본질적으로 마르띠네(1977)가 지적하였듯이 ‘음소의 형태소 되기’를 이용하여, 무한한 이름짓기가 가능하다.  
 4) 우르크 대신전에서 발견된 최초의 문자판에는 곡식의 포대수와 가축의 수를 기록하였다.(Jean 1995)

언어 도구론이 주장하는 언어의 개념은 언어의 본질을 전면적으로 드러내는 것은 아니다. 언어는 언어 능력론의 주장처럼 언어가 일반적이며 인위적인 도구들과 본질적으로 다른 성질이 있다. 그 성질은 다름 아닌 '언어의 창조성'이며, 이는 객관 세계의 무한 대응이 가능한 인간의 추상화 능력이라고 할 수 있다.

인간은 주어진 능력을 이용하여 객관 세계의 사실을 추상화하고 이를 사회 안에서 전달 보존한다. 이렇게 추상화된 전달 내용을 정보라는 말로 바꾸어 보자. 정보란 체계화된 사실이다. 어떤 사실이 존재할 때 그 사실을 대신하는 체계화된 기호 내용을 의미한다. 언어가 정보를 전달한다는 점에서 언어는 정보 전달의 도구이다. 그것은 어쨌든 가능성 있는 능력이 아니라, 객관화된 기호들의 실체로 나타나기 때문이다. 말하자면, 정보는 이미 존재하는 내용으로 폐쇄적이다. 이 점에서 언어정보처리의 자연언어의 개념은 도구론이 주장과 일치한다. 자연언어에서 다루는 정보는 이미 현현된 폐쇄성을 담보하고 있기 때문이다.

### 3. 언어정보처리에서 자연언어란 무엇인가?

언어정보처리에서 자연언어는 인간 언어의 본질적 능력을 의미하지 않는다. 이는 말 그대로 이미 전달된 정보를 처리하는 것이지 언어를 처리하는 것이 아니라는 뜻이다. 처리라는 말은 인간이 이해할 수 있는 기호 형태를 기계가 이해할 수 있는 기호 형태로 바꾼다는 뜻이다.<sup>5)</sup> 그렇다면 언어정보처리는 인간만이 이해 가능한 정보 내용을 기계가 이해 가능한 정보로 바꾸는 과정을 형식화하는 것이 된다.

언어의 본질이 인간의 언어 능력이라면 언어정보처리의 대상이 되는 자연언어의 본질은 정보 내용을 담는 기호 체계이다.<sup>6)</sup> 그것은 실제 인간 언어와 상당히 유사하지만, 본질적으로는 다를 수 있다. 자연언어에는 창조성이 없기 때문이다. 정보 내용을 담는 기호 체계라는 개념은 도구적 언어관의 코퍼스(corpus)와 유사한 점이 있다. 그러나 코퍼스가 단순한 폐쇄적 언어를 의미하는 것이라면 자연언어는 코퍼스의 문법이라고 할 수 있다. 촘스키는 문법이 곧 언어 능력이며 언어라고 보았는데, 자연언

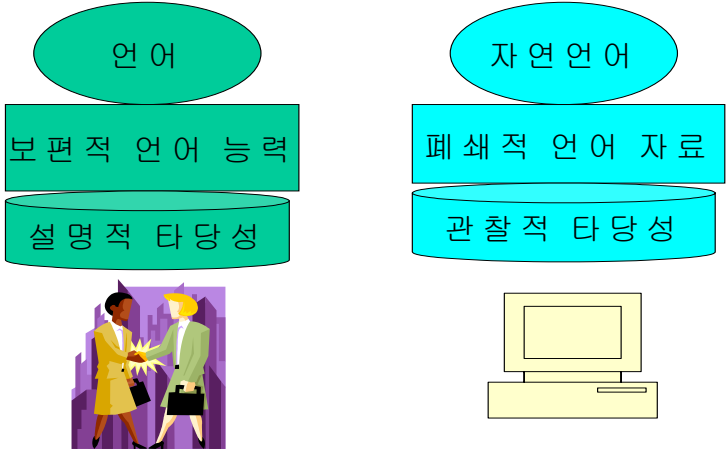
5) 여기서 기계는 튜링이 고안한 생각하는 기계 즉 컴퓨터를 의미한다.

6) 이하 자연언어로 약칭하기로 한다.

어의 문법은 기호 체계이며 폐쇄성을 가진 규칙의 집합이다.

츨스키는 또한 문법을 설명적 타당성을 가진 규칙의 집합이라고 보았는데, 자연언어의 문법은 관찰적 타당성을 가진 규칙의 집합이라는 점에서 큰 차이가 있다.<sup>7)</sup> 자연언어의 문법을 구성하기 위한 코퍼스는 유한할 수밖에 없기 때문이다.<sup>8)</sup>

(4)



언어정보처리의 대상이 되는 자연언어는 결국 기계를 위한 문법을 의미한다. 이 문법은 기계가 자연언어를 받아들여 특정한 자료 구조로 환원하는 절차를 기술한 것이다. 자연언어의 문법은 말하자면 일종의 인공언어인 셈이다. 자연언어가 인공언어

---

7) 어떤 문법이 관찰되는 1차 언어 자료를 올바르게 정확하게 생성할 때 관찰적 타당성을 가진다고 하고, 모국어 화자의 언어 직관을 설명할 수 있을 때 기술적 타당성을 가진다고 한다. 이에 대해 어떤 문법이 설명적 타당성을 갖기 위해서는 1차 언어 자료를 대상으로 기술적 타당성을 가진 문법을 선택하는 일반 원리를 함축하고 있어야 한다는 것이다. 현실적으로 기술적 타당성을 만족하는 문법을 알 수 없기 때문에 설명적 타당성을 가진 문법 이론 또한 허구적 개념에 불과하다. 1차 언어 자료도 대량의 코퍼스를 대상으로 한다면 실제적인 적용에서 관찰적 타당성을 만족하는 문법을 기술하는 문제도 쉽지 않다.

8) 실제의 코퍼스는 수천만 어절을 포함하고 있기 때문에 유한하기는 하더라도 그 전모를 장악하기 어려운 정도의 크기임에는 틀림없다.

에 대해 크게 다른 점은 정보 표현에서 중의성을 허용한다는 데에 있다. 즉 자연언어의 문법은 자연언어를 인공언어의 자료구조로 환원하면서 자연언어의 표현이 함축하고 있는 중의성을 제거하는 것이다.

(5) 들을<sub>i</sub> 걸을 때 들을<sub>j</sub> 수 있는 음악

(5)에서 '들을<sub>i</sub>/j'는 각각 다른 단어가 결합하여 동일한 표현형을 이룬 것이다. 언어의 문법에서는 이러한 현상을 심각하게 다루지 않겠지만, 자연언어의 문법에서는 이러한 어절 중의성의 처리가 중요하다. 일반적인 인공언어에서는 이러한 중의적인 표현형을 허용하지 않지만, 자연언어의 속성상 이러한 표현형이 부지기수로 나타난다는 점이다. 어절 중의성이라는 측면에서 실제 코퍼스 상에 나타난 어절 중의성의 집합조차 아직 파악이 되지 않은 실정이다.

언어정보처리에서 자연언어는 기계가 처리하려는 대상 언어의 표현이며, 이를 분석하는 문법 체계를 의미한다. 다음에서 이 문법 체계의 근간을 살피면서 자연언어의 정보 전달 기능에 대해 살펴기로 하겠다.

#### 4. 자연언어의 정보 전달 기능

자연언어를 기계가 이해하도록 하는 이유는 자연언어의 정보 전달을 특정한 자료구조로 환원하여 다양한 정보 가공이 가능하도록 하는 데에 있다. 예를 들어 우리는 인물에 관한 간단한 데이터베이스 구조를 생각해 보자.

(6)

이름	나이	소속	지위
김아무개	50	삼성전자	상무
이아무개	56	백제대학교	총장

(6)과 유사한 데이터베이스를 일일이 입력하기로 한다면 매우 번거로울 수 있다. 그러나 (7)과 같은 신문 기사를 기계가 자동적으로 (6)과 같은 데이터베이스를 구성한다면 사람이 하는 복잡한 일을 대신하는 기계가 될 것이다.

- (7) ㄱ. 삼성전자 상무 김아무개씨(50)가 오늘 미국 컴텍스 쇼에서 기조연설을 하였습니다.
- ㄴ. 이아무개씨(56)는 오늘 교수협회의 전폭적인 지지로 백제대학교 15대 총장에 취임하였습니다.

기계는 (7ㄱ)의 명사구 표현 '삼성전자 상무 김아무개씨(50)'에서 수식어와 피수식어의 관계 이름과 직위의 상호 관계를 인식하고, (7ㄴ)에서는 주어와 술어의 상호 관계를 인식하여 (6)과 같은 데이터베이스를 자동 생성하게 되는 것이다. 인간은 (7)과 같은 언어 표현에서 많은 정보를 취득할 수 있지만, 기계는 분석 목적이 제한적이기 때문에 그 기능 또한 제한적일 수밖에 없다.

이러한 의미에서 자연언어의 정보 전달 기능은 목적에 따라 각각 다르게 설계될 수 있음을 알 수 있다. 언어의 정보 전달 기능은 주로 정보 내용 자체에 초점을 두고 있는 반면 자연언어의 정보 전달 기능은 문법적 분석 내용이 되는 것이다. 즉 자연언어의 정보 전달 기능은 분석 목적에 맞는 구조화된 분석 결과가 된다. 예를 들어 자동번역을 위한 분석에서는 (7ㄴ)의 명사구 '이아무개'와 '백제대학교 15대 총장'과는 의미적 관련성을 포착할 필요가 없지만, 인물 데이터베이스를 만들기 위해서는 필요하다는 뜻이다. 그러나 무엇보다 필요한 것은 범용적 분석을 위한 분석 결과값의 표준을 정하는 것이다. 제대로 기능하는 구문 분석 알고리즘을 작성하기 위해서는 다수의 인력이 수년간의 세월을 보내야 가능할 수 있기 때문에 하나의 분석 엔진을 가능한 다수의 정보 전달 기능에서 활용하는 바람직하기 때문이다.

이런 의미에서 자연언어 분석을 위한 분석 결과 표준은 언어 분석에서 학교 문법이 차지하고 있는 비중과 같다고 할 것이다.

인간의 언어는 표면적인 형태와 내면적인 구조로 대별된다. 자연언어도 분석대상 문자열과 분석된 자료 구조로 대별된다. 일반 언어학에서는 인간의 사고 체계가 작동하기 위한 심리적 실재로서의 특정한 내면 구조를 파악하는 것이 언어 연구의 주



된 목표가 된다. 즉 심리적 실재에 가장 가까운 언어 구조가 무엇인가를 가정하고 그 구조의 정합성이나 정당성을 확인하는 것이 다양한 연구 주제들을 갖게 된다. 이러한 배경에서 특정 시제 체계나 격 체계가 논의되게 마련이다. 그러나 언어정보처리에서는 앞서 언급한 대로 특정 구조의 정당성을 논의하는 것이 아니라, 특정 구조가 어떤 정보 전달 기능에 도움을 주는가에 따라 구조를 정하고, 핵심적인 논의는 실제의 문자열이 정해진 구조에 얼마나 정확하게 분석되는가를 문제 삼는다.

(8) 한나라당과 민주당, 자민련이 재신임 국면에서 공조를 적극 모색하였다.

(9) <한나라당 & 민주당 & 자민련>Sub <재신임 국면>Time <공조>Obj <모색하다>V

(9)는 (8)에 대한 분석 구조의 한 예를 보인다. 특정 기제가 (8)에 대해 (9)와 같은 분석을 내놓았다고 가정하자. 이 분석은 일단 옳은 분석이지만, 이러한 결과 구조가 정보 분석의 최선의 방안인가에 대해서는 논란의 여지가 있다. 특히 '재신임 국면'이 과연 시간 표현의 하위 분류인지 혹은 '공조'와 같은 단어가 동사의 직접 목적어로 기능하는 것인지에 논의할 수 있을 것이다.

(10) <한나라당 & 민주당 & 자민련>Agent <재신임 국면>Local <공조>Theme <모색하다>V

(8)을 또한 (10)과 같이 분석하는 방법도 있을 것이다. 이러한 분석에서는 좀더 적극적으로 '공조'가 Theme을 가질 수 없다고 주장하고, '재신임 국면'이 단순한 Local이 아니라고 주장하여 (11)과 같은 구조를 제안할 수 있다.

(11) <한나라당 & 민주당 & 자민련>Agent <재신임 국면>Range <공조>Result <모색하다>V

그러나 실제적인 작동이라는 점에서 본다면 (8)의 문장이 과연 (9, 10, 11)과 같이 원하는 대로 정해진 정보 템플릿에 사상될 수 있는지가 더 문제될 것이다. 이러한

점에서 오히려 가장 사상되기 쉬운 구조를 택하는 것도 시스템의 성능을 높이는 일이 될 수 있기 때문이다.

다른 각도에서 생각해 보면 (11)과 같은 구조는 언어 분석에서 바로 처리될 수 있는 것이 아닐 수 있다.<sup>9)</sup> (9)와 같은 통사 구조에 가장 유사한 형태로 분석하고, 이를 (10)이나 (11)과 같은 의미 구조에 가깝게 분석할 수 있기 때문이다. 이러한 분석 결과에 대한 문제는 결국 분석의 정확성과 정보 분석의 목적 다시 말하면 정보 구조의 기능에 달려 있다. 지능형 검색과 같은 분야에서는 (9)나 (10)의 구조로도 충분히 그 기능을 다할 수 있으며, 자동 번역과 같은 분야에서는 (11)에 가깝게 분석될수록 더 정밀한 번역 결과를 기대할 수 있을 것이다.

이와 같이 자연언어의 분석에서는 어떤 구조 분석체가 정당한가를 논의한다기 보다는 분석의 효율성과 목적성에 맞게 특정 구조체를 목표로 하게 된다. 이 점에서 정보처리에서 추구하는 자연언어의 연구 목적과 일반 언어학에서 추구하는 언어 연구의 목적이 분명히 구별됨을 알 수 있다.

## 5. 결론

자연언어는 인공지능에 대비되는 개념으로 주로 언어정보처리에서 널리 사용되는 개념이다. 실제로 이 개념은 인간 언어를 지칭하는 것으로 사용되고 있지만, 인간끼리의 의사소통을 전제하는 언어라기보다는 사용된 언어의 코퍼스와 그 분석 절차를 의미한다.

언어과학은 주로 인간 언어의 도구론에 입각한 기호체계의 분석이나 창조성에 기댄 인간의 생득적 능력을 설명하기 위한 언어 이론을 발전시켜 왔다. 최근 언어정보처리 분야의 발전은 언어를 특정한 자료구조로 환원하는 대상 자료체라는 관점에 눈을 돌리기 시작했다. 언어정보처리에서 자연언어의 연구는 이러한 점에서 생생문법적 견해보다는 구조주의적 관점을 중시하게 된다.

그러나 언어정보처리의 연구 목표와 구조주의적 접근 방법 사이의 분명한 차이는

---

9) 물론 (11)과 같은 구조가 특정 정보 전달에 효율적이기 때문에 곧바로 (11)을 산출하도록 의도할 수는 있다.

구조주의가 코퍼스로부터 기호 체계를 귀납하는 과정이라면, 자연언어의 연구는 귀납된 기호 체계로부터 코퍼스를 해석하는 과정이라는 점이다. 이 점에서 자연언어의 연구는 생성문법이 지향하는 설명적 타당성이 아니라, 대규모 자료로부터 획득되는 관찰적 타당성을 논하게 되는 것이다.

자연언어의 분석은 정보 전달 내용을 특정 자료 구조로 환원하기 위한 것으로 그 전달 기능은 자료 구조 자체이다. 즉 기계는 특정 언어 자료를 해석하여 특정 자료 구조에 대응하는 작업을 반복하는 셈이다. 이 때 자료 구조는 분석 목적에 따라 다르게 설계되는 것이 바람직하나 하나의 분석 엔진을 구성하는 데에 따른 시간과 비용을 감안하면 표준 자료 구조를 지향하는 것이 옳다. 앞서 언급한 바와 같이 언어 분석 구조체가 서로 다른 구조를 가질 수밖에 없다고 하더라도 여러 종류의 분석 결과 혹은 정보 전달 기능을 감안하고 또한 분석 효율성을 염두에 둔다면 가장 기본적이고 효율적인 정보 분석 모형을 가정하고 이를 표준 자료 구조로 삼는 것이 바람직하다는 것이다. 또한 다국어 처리에 있어서 특정 언어를 염두에 두지 않는 표준 분석 모형의 존재는 다양한 용도의 시스템 개발에 도움을 줄 것이다.

그러나 이러한 표준 분석 모형이 꼭 분석 효율성을 높이는 것은 아니다. 언어마다 고유의 언어 특성이 존재하기 때문에 다른 언어의 특성들을 하나의 분석 모형에서 수용하기가 쉽지 않기 때문이다. 또한 이러한 표준 분석 모형을 채택하기 위해서는 다양한 분석 모형을 검증하고 적용해 보아야 하는데, 이에 따른 비용과 시간은 각각의 분석 모형을 염두에 둔 총합보다 많을 수 있기 때문이다. 결국 자연언어의 연구는 이성적 결단이기 보다는 경험적 합리성을 추구하는 것이라고 할 수 있다.

최근 정보 사회의 복잡성이 날로 심화되면서 언어정보처리의 중요성도 날로 증가하고 있다. 그러나 아직도 언어정보처리의 대상이 되는 자연언어의 개념과 일반 언어학에서 다루는 개념을 같은 것으로 이해하는 경우가 종종 있다. 물론 언어의 개념 자체가 근본적으로 다른 것은 아니지만, 연구 목적과 자료를 다루는 태도 그리고 연구 절차에서 중시하는 것들이 다르기 때문에 다른 학적 경향을 갖게 된다. 이에 대한 이해가 깊어질수록 언어정보처리도 한 단계 진전될 수 있을 것이다.

〈참고 문헌〉

- 고창수(1999). 「한국어와 인공지능」. 태학사  
고창수(2002). 「자질연산문법이론」. 월인  
홍윤표 외(2002). 「한국어와 정보화」. 태학사  
김원경(2000). 「한국어 격정보와 자질연산문법」. 고대 박사학위논문  
권중성(1994). 「조선어정보처리」. 한국문화사 영인본(1996)  
송도규(1996). 「인지언어학과 자연언어 자동처리」. 흥릉과학출판사  
양동휘(1996). 「최소이론의 전망」. 한국문화사  
Chomsky(1965). *Aspects of the Theory of Syntax*. Cambridge, Mass: MIT press  
Jean(1995). 「문자의 역사」(이종인 역). 시공사  
Martinet(1977). 「일반언어학 개요」(김방한 역). 일조각  
Somers(1987). *Valency and Case in Computational Grammar*. Edinburgh Univ. press

고창수(Ko, Chang Soo)  
한성대학교 한국어문학부  
서울 성북구 삼선동 2가 389번지  
전화: 02)760-4174  
e-mail: kochs@hansung.ac.kr

접수일자 2003. 9. 30  
게재결정 2003. 11. 11