

구문 분석 말뭉치 구축을 위한 분석의 원칙, 방법, 문제*

김의수^o 강범모

고려대학교 민족문화연구원 전자텍스트연구소
usk2000@orgio.net / bmkang@korea.ac.kr

Principles, methods, and some problems in compiling a Korean treebank

Ui-su Kim^o Beom-mo Kang
Center for Electronic Texts, Korea University

요 약

본고는 21세기 세종계획의 일환으로 현재 구축 중인 구문 분석 말뭉치의 분석 방법에 대한 연구이다. 논의하고자 하는 내용은 첫째, 구문 분석에서의 기본 원칙과 표지의 종류 및 세부 원칙에 대한 것이다. 둘째, 이러한 구문 분석 방안을 마련하는 데 있어 심각하게 고려된 몇 가지 유형의 문제들에 관한 논의이다. 특히 이 문제들은 자연언어처리에서만 아니라 이론적인 국어학의 연구에서도 매우 중요하다. 화자의 직관에 의해서라기보다는 실제 말뭉치 구축 작업을 통해서 그 실체가 확연하게 드러나는 문제들이라는 점에서 이들은 우리의 관심을 끌기에 충분하다. 본고에서는 이러한 문제들이 실제 구문 분석에서 어떻게 발생하고 어떻게 해결될 수 있는지를 보일 것이다.

0. 도입

본고는 21세기 세종계획의 일환으로 현재 고려대학교 민족문화연구원 전자텍스트연구소에서 구축 중인 구문 분석 말뭉치의 분석 방법과 관련하여 그 개요를 소개하고 그와 관련된 몇 가지 중요한 문제들을 논의하는 데 그 목적이 있다.

본문의 내용은 크게 두 부분으로 나뉘는데, 1장에서는 구문 분석 방법의 기본 원칙과 구문 분석 표지들의 종류, 또한 이를 바탕으로 한 세부 원칙의 목록 등이 논의될 것이다. 이어 2장에서는 그러한 원칙과 표지 및 방법론을 세우는 가운데 불거져 나온 중요한 문제들 일부와 그에 관한 본고의 처리 방안을 다룬다.

이러한 문제들은 자연언어처리에서뿐만 아니라 순수 이론언어학적인 견지에서도 매우 중요하게 다루어져야만 하는 것들이라는 점에서, 그리고 이들은 직관에 의존한 언어 연구만으로는 쉽게 포착되지 않는 것들이라는 점에서 매우 흥미롭다.

1. 구문 분석 개요

구문 분석의 기본 원칙과 구문 분석 표지의 종류는 다음과 같다.

(1) 구문 분석의 기본 원칙

- 자연언어처리에서 일반적으로 고려되는 일관성 유지와 효율성 제고에 초점을 두되, 일반언어학적 관점에서도 크게 벗어나지 않도록 한다.
- 표층 구조를 중시하여 분석한다.
- 이분지 가설을 취하며 다분지를 허용하지 않는다.
- 공범주를 인정하지 않는다.
- 어절을 분석의 기본 단위로 한다.
- 보어와 부가어를 구분하되 보어의 범위를 엄격히 제한한다.
- 원칙적으로 접속과 내포를 구별하지 않으며 접속절은 모두 부사절로 분석한다(다만 명사구 접속만 인정한다).
- 하나의 주어가 모문과 내포문 모두에 관련되어 있어 구문 분석의 중의성이 발생할 경우, 모문의 주어로 우선 분석한다.

* 이 논문은 문화관광부 지원의 21세기 세종계획 기초자료 구축 과제 일부로 연구되었음.

(2) 구문 분석 표지의 종류 및 내역

a. 구문 표지

범주	사례
S	문장
NP	체언구 (명사, 대명사, 수사)
VP	용언구 (동사, 형용사, 보조용언)
AP	부사구
DP	관형사구
IP	감탄사구

b. 기능 표지

범주	사례
SBJ	주어 (NP_SBJ, VP_SBJ, S_SBJ)
OBJ	목적어 (NP_OBJ, VP_OBJ, S_OBJ)
CMP	보어 (NP_CMP, VP_CMP, S_CMP)
MOD	체언 수식어 (NP_MOD, VP_MOD, S_MOD)
AJT	용언 수식어 (NP_AJT, VP_AJT, S_AJT)
CNJ	접속어 (NP_CNJ)
INT	독립어 (NP_INT)

c. 기타 표지

범주	사례
PRN	삽입어구 (예: [NP_PRN ...])

실제 구문 분석 말뭉치를 구축하기 위해 참조해야 할 구문 분석 지침은 다음과 같이 구성되어 있다.

(3) 구문 분석 지침의 기본 목차

- I. 기본 원칙
- II. 구문 분석 표지
- III. 표지별 분류 기준 및 세부 지침
- 1. 절의 기본 구조
 - 1.1 절의 기본 구성 성분
 - 1.1.1 서술어
 - 1.1.2 논항
 - 1.1.3 부가어
 - 1.1.4 독립어
 - 1.1.5 공범주
 - 1.2 절의 기본 유형
 - 1.2.1 단순절
 - 1.2.2 비단순절

2. 구의 세부 구조

- 2.1 명사구
 - 2.1.1 명사 + 명사 구성
 - 2.1.2 명사 + {과/나/도/그리고/혹은/및/또는} + 명사구성
 - 2.1.3 관형절 수식 구성
 - 2.1.4 명사 전성 구문
 - 2.1.5 (명사+명사 구성) + (명사 전성 구문)
 - 2.1.6 동일 기능 명사구 반복 구성
 - 2.1.7 동격 명사구 구성
- 2.2 동사구
 - 2.2.1 본용언 + 보조용언 구성
 - 2.2.2 본용언 + 본용언 구성
 - 2.2.3 의사 보조용언 구성
 - 2.2.4 부사절 수식 구성
 - 2.2.5 경동사 구성
- 2.3 부사구
 - 2.3.1 부사구 연결 구성
 - 2.3.2 논항을 취한 부사구
- 2.4 유동 양화사구
- 3. 절의 세부 구조
 - 3.1 논항절의 기능표지 부착 시점
 - 3.2 접속의 층위 문제
 - 3.3 계사 구문
 - 3.4 주격 중출 구문
 - 3.5 목적격 중출 구문
 - 3.6 주어 인상 구문
 - 3.7 목적어 인상 구문
 - 3.8 장형 사동 구문
 - 3.9 도치 구문(어순재배치 구문)
 - 3.10 삽입 구문
 - 3.11 직접인용문으로서의 단순문
 - 3.12 VP가 아닌 서술어를 가진 문장 구성
- 4. 공범주 문제
 - 4.1 기본 입장
 - 4.2 관련 문제와 분석 방안
 - 4.2.1 성분 이동에 따른 흔적
 - 4.2.2 탈락된 주어나 목적어
 - 4.2.3 공운용소
 - 4.2.4 계사 구문에서의 '이다' 생략
 - 4.2.5 V 혹은 VP 생략
- 5. 기타 구문

그러면 이제부터 실제 구문 분석에서 부딪히는 중요한 문제들 가운데 몇 가지를 살펴보도록 한다.²⁾

2. 문장 성분의 구별 문제

2) 한국어를 대상으로 한 기존의 구문 분석 연구로는 펜실 베니아 대학의 Upenn TreeBank(2001)와 Kaist의 STEP 2000이 대표적이다. 본고에서는 주로 전자와의 비교를 통해 구문 분석의 문제를 논의하고자 한다.

2.1 보어/부가어의 구별 문제

본 연구에서 보어는, 보격조사 '이/가'를 가진 명사구 혹은 그것이 생략되었거나 보조사로 대체된 명사구 및 용언구, 그리고 인용절만을 가리키는 것으로 한정된다.

- (4) a. 철수가 [NP_CMP 의사]가 되었다.
 b. 철수가 [NP_AJT 의사]로 되었다.
 c. 철수가 [NP_CMP 의사]만 되었다도 좋았을텐데.
 d. 철수는 [S_CMP 자신이 잘못을 하지 않았다고] 말했다.

이때 '의사로'와 같은 소위 필수 부사어들은 어느 면에서 보면 필수 성분으로서의 보어라 할 만하지만 개별 문장의 상황에 따라 그 판별이 어려운 경우가 대부분이다. 따라서 보어를 엄격히 제한하고 그 밖에 다양한 스펙트럼을 갖는 (필수) 부사어 성분들을 '_AJT'라는 기능표지로 포착해 주고자 한다. 즉, 기능 표지 '_AJT'는 일반 부사어뿐만 아니라, 비록 보격조사 '이/가'를 동반하지는 않았지만 필수적인 문장 성분으로 간주될 법한 것들도 포함한다. 그 용례를 보이면 다음과 같다.³⁾

- (5) a. 철수가 [NP_AJT 의사]로 되었다.
 b. 우리 반은 [VP [S_AJT 학생들이 매우 열심히 공부했기에] 이번에 1등을 차지하였다].
 c. 우리 반은 [VP [S 학생들이 매우 열심히 공부해서] 이번에 1등을 차지하였다].

2.2 목적어/부가어의 구별 문제

목적어와 부가어를 구별하기 힘든 경우도 있다. 특히 시간이나 거리의 명사구가 그러하다. 이러한 경우에는 그러한 명사구가 조사 '을/를'을 취하거나 그것을 후행하는 동사가 명백히 타동사인 경우에만 '_OBJ'라는 목적어 기능 표지를 부여한다.

- (6) a. 현우는 학교를 하루 쉬었다.
 ⇒ 현우는 학교를 [NP_AJT 하루] 쉬었다.
 b. 현우는 학교를 하루를 쉬었다.
 ⇒ 현우는 학교를 [NP_OBJ 하루를] 쉬었다.
 (7) a. 현우는 아침 일찍 학교를 갔다.
 ⇒ 현우는 아침 일찍 [NP_OBJ 학교를] 갔다.
 b. 현우는 아침 일찍 학교에 갔다.
 ⇒ 현우는 아침 일찍 [NP_AJT 학교에] 갔다.
 c. 현우는 아침 일찍 학교 갔다.
 ⇒ 현우는 아침 일찍 [NP_AJT 학교] 갔다.

3) 반면에, Upenn TreeBank에서는 보어를 목적어나 주어라 아니지만 필수적인 성분인 것으로 간주한다. 그러나 이런 입장에서 항상 문제되는 것은 보어 식별의 명확한 잣대가 없으며 그 때문에 일관성 있는 구문 분석이 어려워지게 된다는 점이다.

3. 문장 성분 생략의 문제

3.1 공범주와 명사구 생략의 문제

성분 생략에서 고려되는 공범주에는 흔적(trace), pro, PRO, 공운용소(null operator) 등이 있으나 본 구문 분석에서는 고려하지 않는다.⁴⁾

- (8) a. 어서 가거라!
 ⇒ [S [VP 어서 가거라!]]
 b. 그는 먹던 사과를 나에게 던졌다.
 ⇒ 그는 [NP_OBJ [VP_MOD 먹던] 사과를] 나에게 내던졌다.
 c. 그가 바보라고 말했다.
 ⇒ 그가 [VP [VP_CMP 바보라고] 말했다].
 d. 오니까 갔다.
 ⇒ [S [VP [VP 오니까] [VP 갔다]].]

3.2 서술어 생략 구문

i) 단순문에서의 서술어 생략

동사 없이 나머지 문장 성분들만이 존재할 경우, 그 전체 범주를 무엇으로 분석해 줄 지가 문제가 되는 예들이 많다. 이러한 사례는 특히 신문 기사에서 많이 발견되는데, 아래와 같은 분석 원칙에 의거하여 처리하는 것이 바람직해 보인다.

- (9) 분석 원칙:
 a. 주어(_SBJ)가 있으면 그것을 포함한 전체 범주를 S로 분석한다.
 예) [S [NP_SBJ 최수 3명이] 탈옥]
 b. 주어가 없고 목적어(_OBJ)가 있으면 그것을 포함한 전체 범주를 VP로 분석한다.
 예) [VP [NP_OBJ 최다 국내외 관광지]와 [NP_AJT 한 자리에서]]
 c. 주어나 목적어가 모두 없고 부사어(_AJT)가 있으면 그것을 포함한 전체 범주는 NP로 분석한다.
 예) [NP [NP_AJT 매일] 민족 공연 등 직접 홍보]

ii) 복합문에서의 V나 VP 생략

단순문이 아닌 복합문, 특히 접속의 경우 선행절의 동사나 동사구가 생략될 수 있다. 이 경우는 다음과 같이 모두 절로 분석할 수 있다.

- (10) a. 철수는 밥을, 나는 떡을 먹었다.

4) Upenn TreeBank에서는 모든 공범주를 다 인정하려 하였다. 그러나 이럴 경우 기계가 아닌 사람만이 공범주를 적절히 분석해 줄 수 있게 되어 결과적으로 구문 분석의 자동화와 효율성을 크게 떨어뜨릴 수밖에 없게 된다.

⇒ [S [S [NP_SBJ 철수는] [NP_OBJ 밥을,]]
[S 나는 떡을 먹었다].]

b. 철수는 병수에게, 영희는 순이에게 책을 주었다.

⇒ [S [S [NP_SBJ 철수는] [NP_AJT 병수에게,]] [S 영희는 순이에게 책을 주었다].]

4. 제시어 구성의 처리

제시어 구성은 동격 명사구 구성으로 처리하는 것이 타당해 보이며, 그것은 후행하는 동격의 명사구와 우선적으로 묶어 준다. 이때 첫 번째 명사구, 즉 제시어는 아무런 기능표지를 갖지 않으며 후행 명사구와의 결합으로 생기는 상위 명사구는 후행 명사구의 기능 표지를 계승하도록 한다.

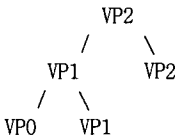
- (11) a. 그런데 현우 넌 어디루 이사 갔니?
⇒ 그런데 [NP_SBJ [NP 현우] [NP_SBJ 넌]] 어디루 이사 갔니?
b. 그동안 나는 철수 그 사람을 죽 지켜보았습니다.
⇒ 그동안 나는 [NP_OBJ [NP 철수] [NP_OBJ 그 사람들]] 죽 지켜보았습니다.
c. 마을에서 약간 떨어진 외딴집, 그것이 현우 집 식구가 새로 든 집이었다.
⇒ [NP_SBJ [NP 마을에서 약간 떨어진 외딴집,] [NP_SBJ 그것이]] 현우 집 식구가 새로 든 집이었다.

5. 의사 보조용언 구성

주서술어 뒤에, 보조용언은 아니지만 서법(mood, modal)을 나타내는 언어 단위들이 올 경우가 있다. 자연 언어처리에서 이들을 의사 보조용언(pseudo-auxiliary verb) 구성으로 처리하는 것이 바람직하다. 이들은 주서술어와 한데 묶여 VP로 투사되는데, 그것은 아래와 같은 일반적인 '본용언'과 '보조용언'의 구성 방식과 동일하게 처리된다.

(12) 참고: 본용언 + 보조용언 분석 방법

a. [VP2 [VP 본용언1 [VP 보조용언2]]



b. 철수가 집에 [VP3 [VP2 [VP1 가지] [VP2 앓고]] [VP3 있다].

그런데, 이와 같은 의존 구성을 처리하는 데 있어서는

다음과 같이 대략 세 가지 방법이 있을 수 있다.

(13) 예문: 그는 밥을 곧 먹을 것이다.

a. <1안>: [VP [S 그는 밥을 곧 먹을] 것이다].

b. <2안>: [S 그는 밥을 곧 [VP [VP_MOD 먹을] 것이다]].

c. <3안>: [S 그는 [VP [VP 밥을 곧 먹을] 것이다].

<1안>은 문면에 나타난 그대로 '관형절+것이다'로 분석하는 것이고, <2안>은 서술어를 먼저 묶고 나서 명제를 구성하는 논항들이나 부사어들을 그 다음에 묶어 나가는 방식이다. 반면 <3안>은 주어를 제외한 나머지 서술부에서 다시 '것이다'를 제외한 나머지 성분을 관형구로 묶고 그것이 뒤의 '것이다'를 수식해 준다는 분석으로서 Upenn TreeBank에서 취하고 있는 입장이다. 본 분석에서는 <2안>을 따르기로 한다. 그 이유는, 이 방식이 앞서 '본용언'과 '보조용언'을 먼저 묶는 것과 동격의 처리 방식일 뿐만 아니라, 문장을 구성하는 서술어가 본질적으로 '술어+서법/양태'으로 구성된다는 점에서, 그리고 구문 분석에서의 비중의적인(따라서 효율적인) 처리안이라는 점에서 합리적인 것으로 판단되기 때문이다.

이와 같은 의사 보조용언 구성에 속하는 사례들은 대략 다음과 같으며, 그 목록의 수는 작업량의 증가에 따라 비례하여 앞으로 더욱 늘어날 것으로 보인다.⁵⁾

- 1) '-는 {모양/노릇}이다'의 구성
예) 철수가 [S 고향이 [VP [VP_MOD 그리운] [VP 모양이다]].
예) 우리가 매일 [VP [NP_SBJ [VP_MOD 매일 놀] 수 만든] [VP [VP_MOD 없는] [VP 노릇이다]].
- 2) '-ㄹ + 의존명사 + 이다' 구성
예) 철수가 밥을 곧 [VP [VP_MOD 먹을] [VP 것이다]].
예) [S [VP_MOD 한국이면 어디서라도 있음직한] [VP [VP_MOD 고향일] 뿐이다].
- 3) '-ㄴ 것 같다' 구성
예) 비가 곧 [VP [NP [VP_MOD 올] 것] 같다].
- 4) '-기로 하다'의 구성
예) 그가 나와 같이 [VP [VP_AJT 가기로] 하였다].
- 5) '-기도 하다'의 구성
예) 그들은 서로 [VP [VP 다르기도] 하다].
- 6) '-든지 하다' 구성
예) 철수는 [VP [VP 밥을 먹든지] [VP 빵을 [VP_MOD [VP 먹든지] 할] 것이다]].
- 7) '-어서는 안 된다' 구성
예) 우리는 경계를 [VP [VP 늦추어서는] 안 된다].
- 8) '-ㄹ {만/법}하다' 구성

5) 이 문제에 관해서는 임해창 외(2001a)와 김의수(2002)를 참고할 수 있다. 한편, Upenn TreeBank에서도 '필요로 하다, -뿐만 아니라, -밖에 안 되다, -밖에 모른다, -것만 하다' 등 몇몇 사례들을 소개하고 있다.

예) 그 영화는 [S 청소년들도 [VP [VP_MOD 볼] 만하다]].

9) '-고 해서' 구성

예) [S 시간도 [VP [VP 없고] 해서]] 그는 친적질에 들르지 않았다.

10) '-ㄴ/ㄹ지도 모르다' 구성

예) 그가 집에 [VP [VP 같지도] 모른다].

11) '-ㄴ/ㄹ지도 모르는 일이다' 구성

예) 앞으로 어떤 별이 [VP [VP_MOD [VP_OBJ 내릴는지] 모를] 일이다].

12) '-ㄹ지경이다' 구성

예) 모두가 [VP [VP_MOD 죽을] 지경이다].

13) '-ㄹ/ㄴ 것이다/냐' 구성

예) 너는 장차 무엇을 [VP [VP_MOD 할] 것이냐]?

14) '-ㄹ/ㄴ 듯(도) 하다' 구성

예) 따스한 손길로 소리 없이 머리를 [VP [NP [VP_MOD 어루만져 주시는] 듯도] 했다].

6. 접속과 내포의 문제

대등 접속절과 종속 접속절은 상호간에 구별이 어려울 뿐만 아니라, 그들은 공히 부사절과 같은 내포문과도 변별이 잘 되지 않는다. 이 점을 의식하여 최근 순수 국어학의 연구에서도 이들을 구별하지 않으려고 하는 경향이 강하며,⁶⁾ 자연언어처리의 입장에서는 그러한 필요성이 더욱더 크게 대두된다. 따라서 본 연구에서는 접속절의 두 유형을 구별하지 않을 뿐만 아니라, 기존에 접속절로 다루었던 모든 문장들을 부사절의 한 유형으로 파악하고자 한다.

7. 구문 분석 도구의 특성과 얽힌 문제

7.1 구문 표지 없이 기능 표지만 무확되는 경우

21세기 세종계획의 구문 분석 말뭉치 구축은 고려대학교 자연언어처리실(NLP lab)에서 제작한 작업 도구 프로그램을 바탕으로 이루어지고 있다. 따라서 실제로 하나의 문장에 대해 구문 분석을 시도할 때에는 자연언어의 특성뿐만 아니라 그것의 분석 도구인 구문 분석 프로그램의 자체 속성 때문에 발생하는 특징적인 상황도 유의미한 고려 대상이 된다. 예컨대, 인용 부호 다음에 조사 가 오는 경우가 그 대표적 사례 가운데 하나이다.

- (17) '나도 아버지가 있다'로서만이 아니라 '내게는 이러한 훌륭한 아버지가 계시다' 하는 벽차오르는 자랑이었다
- (18) 위 문장의 형태소 태깅 결과

6) 이관규(1992)는 종속접속문을 접속문이 아닌 부사절로 파악하였고, 더 나아가 고평주(1999)는 대등접속문도 통사적 대등 구성이 아닌 의존 구성임을 논증하였다. 한편, Upenn TreeBank에서는 종속접속문은 부가어절로 처리하고 대등접속만을 인정하였다.

'나도 '/SS + 나/NP + 도/JX
 아버지가 아버지/NNG + 가/JKS
 있다'로서만이 있/VV + 다/EC + '/SS + 로서/JKB + 만/JX
 + 이/JKC
 아니라 아니/VCN + 라/EC

(19) 구문 분석 도구에 얽힌 모습:

☞ 별첨 [1]

(20) 방안: 구문 표지 없이 기능 표지만을 가지는 경우 구문 표지의 빈 자리에 'X'를 넣어준다.

(21) a. 실제 예문: '나도 아버지가 있다'로서만이 아니라 '내게는 이러한 훌륭한 아버지가 계시다' 하는 벽차오르는 자랑이었다.

b. 구문 분석 결과:

☞ 별첨 [2]

한편, 조사가 떨어져 있을 뿐만 아니라 선행 명사구가 다시 수식을 받는 경우에도 이 문제와 발생한다. 그리고 이때에는 특히 구문 분석의 순서가 관건이 되는데, 본 연구에서는 선행 명사구와 조사의 결합체가 함께 묶이고 난 후 그 전체가 선행하는 수식어에 의해 수식을 받는 것으로 분석하고자 한다.

(22) a. 세계적 조각가인 알렉산더 칼더 조각전행사의 하나로 미술관측이 마련한 '어린이를 위한 모빌실습'의 현장이다.

b. ... [X_MOD [S_MOD 미술관측이 마련한] [X_MOD [S '어린이를 위한 모빌실습'] 의 /JKG]] ...

c. 구문 분석 결과:

☞ 별첨 [3]

7.2 '명사 + 조사' 구성이 논항처럼 기능하는 경우

이러한 경우는 그동안 순수 국어학적인 입장에서조차 별로 주목받지 못했는데,⁷⁾ 특히 본 연구와 같은 자연언어처리의 입장에서는 매우 중요한 문제이다. 그 이유는 형태소 분석 단계에서는 분명히 VP로 태깅되었으나, 구문 분석 단계에서는 그것이 NP로 재분석되어야 마땅해 보이기 때문이다. 이는 결국 특히 자연언어처리에서 형태소 분석 단계의 결과와 통사부 분석 단계의 결과가 불일치할 수도 있음을 단적으로 보여주는 사례인 것이다. 그 사례는 다음과 같다.

7) 이에 관해 최근 이선웅(2000)의 논의가 주목할 만한데, 그는 내포된 '의문사(이)ㄴ가'를 CP가 아닌 NP로 보아야 한다고 주장하였다.

(23) 형제간이라도 있었으면 좀 나왔으려면, 하필 남
들 같잖게 단 혼자였다.

(24) 형태 분석 결과:

형제간이라도 형제/NNG + 간/NNB + 이/VCP + 라도/EC
있었으면 있/VV + 었/EP + 으면/EC

좀
나왔으려면, 낮/VA + 았/EP + 으려면/EC + ,/SP

...

(25) 구문 분석 결과:

☞ 별첨 [4]

7.3 '우측원리(오른쪽 마디 투사)'가 위반되는 경우

본 구문 분석에 도입된 고려대학교 자연언어처리실의
구문 분석 도구는 두 교점이 상위 교점으로 병합될 때
우측 마디의 구문 분석 정보(구문 표지, 기능 표지, 기
타 표지)가 자동으로 상위 마디에 투사되도록 설계되어
있다. 한국어의 실제 구문 분석의 대부분의 경우에는
이와 같은 투사 방식이 유효하다. 그러나 그렇지 않은
경우가 종종 발생한다. 예컨대, 다음과 같은 오른쪽(우
측) 부가의 경우가 그러하다. 이때에는 부득이 왼쪽 마
디 투사를 허용할 수밖에 없어 보인다.

(26) a. 그가 죽었어, 어제.

⇒ [S [S 그가 죽었어,] [AP 어제]].

b. 장사꾼들이 다 그렇지 뭐.

⇒ [S [S 장사꾼들이 다 그렇지] [IP 뭐]].

이와 같은 점은 구문 분석에서의 예외적인 항목 설정 정
도로 끝나는 것이 아니라, 구문 분석의 자동 처리에서
일률적인 우측 마디 투사가 일단 불가능함을 의미한다는
데 더 큰 의미가 있다. 그러나 이 문제로 인해 자동 구
문 분석이 큰 난관에 봉착하지는 않을 것으로 생각된다.
하나로 묶이는 하위 마디들의 표지 유형들을 적절히 고
려한다면 상위 마디의 투사 표지의 결정 또한 규칙화할
수 있어 보이기 때문이다.

참고문헌

강범모. 2001. [범주문법: 한국어의 형태론, 통사론, 타
입논리적 의미론], 고려대학교 민족문화연구원.
고광주. 대등 접속문에 대한 재검토, [한국어학] 9.
김의수. 2002. 언어단위로서의 상대어 설정 시고, [형태
론] 4-1, 박이정 출판사.
김홍규 외. 2001. 형태소 분석 말뭉치 구축 지침, [21세
기 세종계획 국어 기초자료 구축 연구보고서], 문화관
광부.
서정수. 1996. [국어문법], 한양대 출판원.
이관규. 1992. [국어 대등구성 연구], 서광학술자료사.

이기용 외. 1997. 구문분석기 데이터 정보개발, ms. 한
국과학기술원.

이선웅. 2000. '의문사+(이)-+ㄴ가' 구성의 부정(不定)
표현에 대하여, [국어학] 36.

임해창 외. 2001a. 구문 태그 부착 말뭉치 구축 도구 개
발, 21세기 세종계획 국어 기초자료 구축 분과 제2 위
크숍 발표문.

임해창 외. 2001b. 구문 태그 부착 말뭉치 구축 도구 개
발, [21세기 세종계획 국어 기초자료 구축 연구보고
서], 문화관광부.

홍종선·강범모·최호철. 2001. 한국어 언어 관계 연구,
월인.

Han, Chung-hye, Na-Rae Han and Eon-Suk Ko. 2001.
Bracketing Guidelines for Penn Korean TreeBank,
IRCS Report 01-09, Institute for Research in
Cognitive Science University of Pennsylvania.

·김의수(Ui-su Kim)

고려대 민족문화연구원 전자텍스트연구소

02-3290-2493, 017-296-3795

usk2000@orgio.net

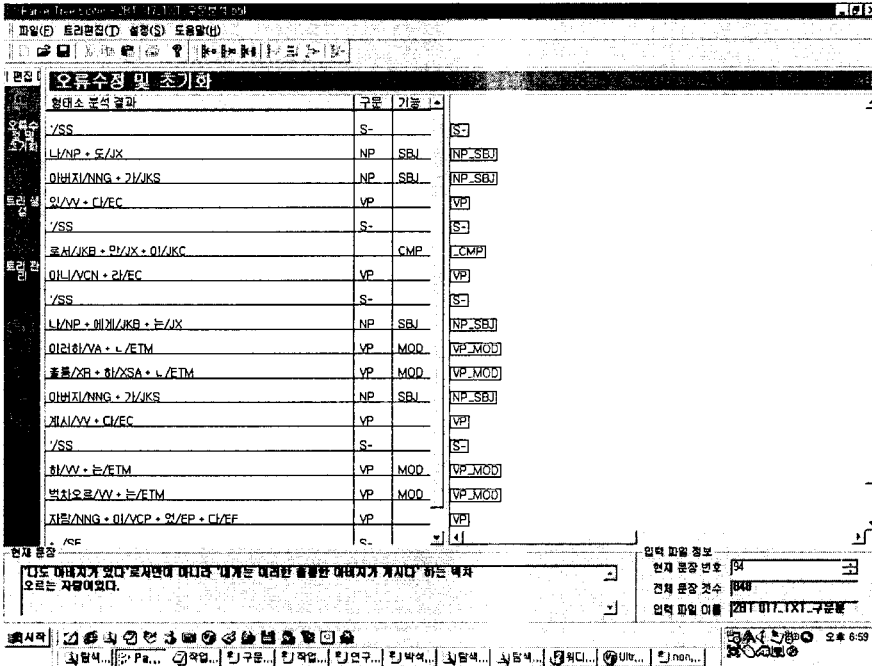
·강범모(Beom-mo Kang)

고려대학교 문과대학 언어과학과

02-3290-2173

bmkang@korea.ac.kr

*별첨[1] : (19) 구문 분석 도구에 의한 모습:



*별첨[2] : (21) b. 구문 분석의 결과:

```

(VP (X_CMP (S- (S- +'/SS)
               (S- (S (NP SBJ 나/NP + 도/JX)
                    (S (NP SBJ 아버지/NNG + 가/JKS)
                       (VP 있/VV + 다/EC))))
              (S- +'/SS)))
 (X_CMP 로서/JKB + 만/JX + 이/JKC)
 (VP 아니/VCN + 라/EC))

```

*별첨[3] : (22) c. 구문 분석의 결과:

```

(S (VP (X_MOD (S_MOD (NP_AJT (NP_MOD (VP_MOD (NP 세계/NNG + 적/XSN)
                                               (VP MOD 조각가/NNG + 이/VCP + ㄴ/ETM))
                                               (NP_MOD (NP (NP 알렉산더/NNP)
                                                         (NP 칸더/NNP)
                                                         (NP MOD 조각전/NNG + 행사/NNG + 의/JKG)))
                                               (NP AJT 하나/NR + 로/JKB)
                                               (S_MOD (NP SBJ 미술관/NNG + 측/NNB + 이/JKS)
                                                       (VP MOD 마련/NNG + 하/XSV + ㄴ/ETM)))
                                               (X_MOD (S (S- +'/SS)
                                                       (S (NP (VP_MOD (NP OBJ 어린이/NNG + 를/JKO)
                                                           (VP MOD 위하/VV + ㄴ/ETM))
                                                           (NP 모빌/NNG + 실습/NNG)
                                                           (S- +'/SS)))
                                                       (X_MOD 의/JKG)))
                                               (VP 현장/NNG + 이/VCP + 다/EF))
      (S- + ./SF))

```

*별첨 [4] : (25) 구문 분석의 결과:

(S (VP (VP (S (NP_SBJ 형제/NNG + 간/NNB + 이/VCP + 라도/EC)
 (VP 있/VV + 었/EP + 으면/EC))
 (VP (AP 줌/MAG)
 (VP 낫/VA + 았/EP + 으련만/EC)))
 (VP (VP (AP 하필/MAG)
 (VP (NP_AJT 남/NNG + 들/XSN)
 (VP 같잖/VA + 게/EC)))
 (VP (DP 단/MM)
 (VP 혼자/NNG + 이/VCP + 었/EP + 다/EF))))
 (S- + ./SF))