

LKB로 구현한 한국어 단층구조(Flat Structure) 분석¹⁾

신효필, 고성연 (서울대학교 언어학과)

1. 서론

- 가. 전통적으로 주어는 동사구(VP) 밖의 요소이며 기타 다른 문장성분과 차이가 있다고 분석
- 나. 구구조문법을 비롯한 문법기술은 S-> NP, VP 식의 양분적 동사구 계층적 구조에 바탕
- 다. 영어와 같은 언어에서는 이런 특질이 잘 드러나나 문법관계가 문장에서의 위치보다는 접사에 의해서 실현되는 교착어에서 계층적 동사구 설정의 타당성 의심.
- 라. scrambling 과 extraction 관점에서의 비교적 자유로운 어순으로 인한 설명의 타당성
 - a. 주제화는 도출이나 도치에 의해서만 실현되는가?
 - b. 계층적 VP를 설정하지 않을 경우 비교적 자유로운 한국어 어순의 설명
- 마. 한국어의 문장구조를 계층적인 동사구를 구성하지 않고 단층구조(Flat Structure)로 분석
 - a. 장석진(1993)의 기술을 기초
 - b. 이 분석을 LKB(Linguistic Knowledge Building)를 통하여 실제로 구현
- 바. 기술적 타당성과 구현적 타당성 관점에서의 논의
- 사. 이 분석방법의 의의 및 한계 (LKB의 구조적인 제약 포함)
- 자. 분석방법의 다양화 추구 (cf. 김종복, 양재형 2003)

2. 단층적 분석의 동기

- 가. 한국어의 상대적으로 자유로운 어순
 - (1) a. 철수가 미아를 때렸다.
 - b. 미아를 철수가 때렸다.
 - c. ?때렸다 미아를 철수가
- > 적격성 판단 여부 어려움
- (2) 주제화 등의 도출 또는 동사구 내에서의 공백

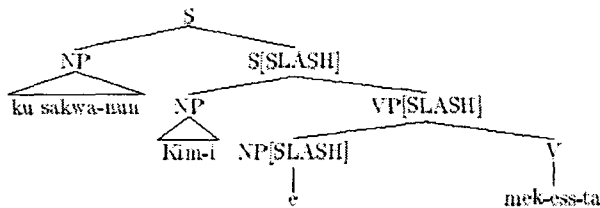


그림 1. 김종복 (1998)의 extraction

1) 이 글은 아직 완전한 논문이 아니며 발표용으로 불완전하게 정리된 글임. 여기서 flat structure를 단층구조로 번역한 것은 정확한 의미를 전달하지 않을 수 있으나 목적어나 간접목적어 등을 VP에 계층적으로 명시하지 않고 전체 문장 S의 성분으로 VP와 평행하게 설정한다는 의미이다. 장석진(1993)에 기초.

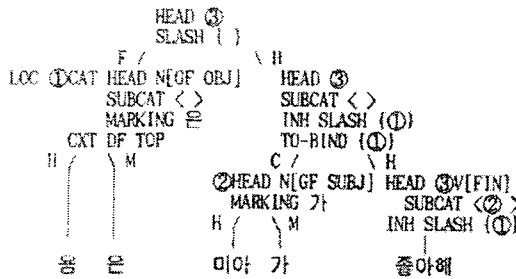


그림 2. 장석진 (1993)에서의 계층적 구조에서의 도출

(3) 순수히 평면적인 관점 2)

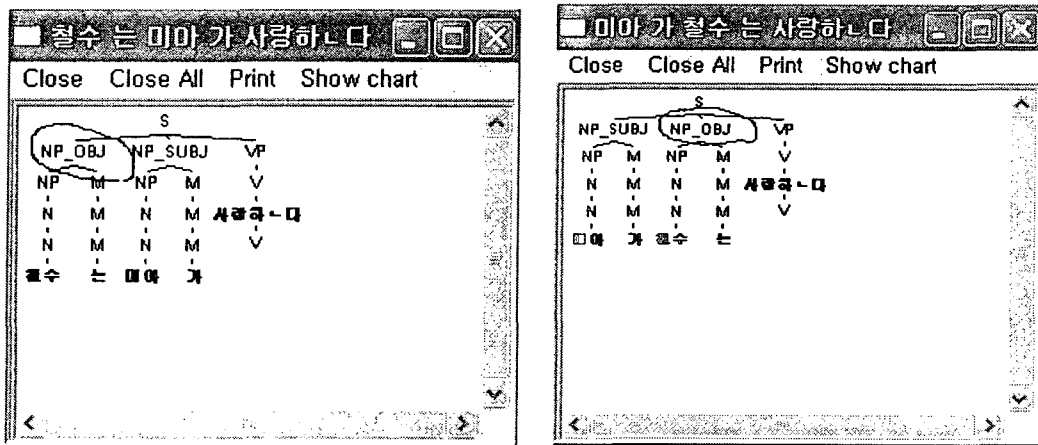


그림 3. LKB에 기초한 단층구조 분석

(4) 계층구조와 양분분석 (binary analysis)

- a. 언어구조의 양분분석은 동사구의 계층구조 분석과 관련
- b. 많은 문법기술 기재들이 양분분석에 기초
- c. LKB도 양분적 구조기술에 기초 단층구조 분석에 적합하지 않음.

3. 한국어 단층구조 분석

가. 논의의 출발인 단층구조 분석은 장석진(1993)에서 기초.

나. 장석진(1993): 한국어를 정보기반법으로 기술하고 한국어 기술에 필요한 자질체계와 성분구조, 원리, 정형도식 등을 설정하고 있다. 이런 원리를 바탕으로 한국어의 여러 문형들이 평면으로 기술되고 있다.

2) 이 분석은 본 논의에서의 LKB로의 구현에 기인한다. 계층적 동사구를 설정하지 않기 때문에 도출된 것이 아니라 scrambling된 구조로 분석이 된다. '-는' 이 붙어도 하위범주와 정보에 의해 목적어로 분석되며, 주제화된 성분의 표시는 현재 이루어지지 않고 있다.

다. 장석진(1993)에서 시도된 정보기반 한국어 문법을 실제로 LKB로 구현해 보는 시도. 이 구현은 장석진(1993)에서 논의된 모든 문법현상을 다 포괄하지 않으나 다음은 실제 그 구현에서 고려된 사항들이다.

(5) 머리자질원리(Head Feature Principle: HFP)

하위범주화원리(Subcategorization Principle: SCP)

비지역적자질원리(Nonlocal Feature Principle: NFP)

보충어추출규칙(Complement Extraction Rule)

정형도식-1: 계층적 vp를 인정할 경우: s NP + VP

정형도식-2: 계층적 vp를 인정할 경우 주어자리 하나만 남아 있는 vp 구조

정형도식-3: 단층구조인 경우

정형도식-4: 두-표(head-marker structure) 구조. 문법기능조사/보문자 ‘-고’

정형도식-5: 두-첨(head-adjunct structure)구조. 수식어와 피수식어와의 관계

정형도식-6: 두-충(head-filler structure)구조. 주제화

4. LKB로의 구현

4.1. LKB

가. 유형화된 자질구조문법을 바탕으로 한 문법과 어휘 개발 틀

나. 주로 HPSG문법기술에 기초

다. Stanford의 CSLI

4.2. 한국어 단층구조 기술을 위한 LKB 변경사항 및 구현내용

가. 양분구조 분석인 LKB에서 단층구조인 다중 분지관계를 구현하기 위해 user-fns 수정 (linear precedence rule)

나. 수식어구에 따른 많은 규칙이 필요 <-> 양분분석에서는 규칙의 간단.

다. 형태소 분석을 가정한 입력 설정 (조사 분리)

라. 형태적/의미적 체계화된 유형설정보다는 구문분석에 초점

마. 특정이론보다는 한국어 실제 문장분석에 초점

바. 분석문장들

a. 기본문장들(±조사): 미아(가) 온다

꽃이 예쁘다

미아(가) 철수(를) 사랑한다

미아(가) 철수(에게) 꽃을 주었다

미아가 꽃이 예쁘다고 생각했다.

미아가 영희에게 그 편지를 읽으라고 요청했다.

b. 기본문장들 (+기본 수식어구): 이 꽃이 예쁘다

철수가 영희에게 저 꽃을 준다

- c. 이중주어 3)/-이다 문장 + (기본 수식어구): 코끼리가 코가 길다
미아가 학생이다
- d. 관계절 문장/관형절 문장: 미아는 아름다운 영희에게 꽃을 주었다.
신사의 입은 옷이 더럽다. (관계절내의 관계화)
나의 살던 고향이 아름답다 (소유격 관계화)
영희가 미아가 먹었다고 말한 사과가 붉다(내포된 절에서의 관계화)
그 학생이 노래를 부른 시간이 길다. (첨가어의 관계화)
미아가 돈을 훔친 사실이 밝혀졌다 (동격)
- e. 명사화구 (Nominalized Phrase): 미아가 돈을 훔쳤음이 밝혀졌다
철수는 집에 돌아가기가 싫다.
- f. 부사적 연결문: 천둥이 치면 비가 온다
바람이 부니까 살구가 떨어진다.
- g. 대등적 연결문: 바람이 불고 살구가 떨어진다.
- h. 복합 문장: 저 붉은 장미가 붉으면 미아가 철수에게 주던 꽃이 예쁘고 나의 살던 고향이 살구로 아름답다

4.3. LKB 기술 세부사항

4.3.1. 유형화된 자질들

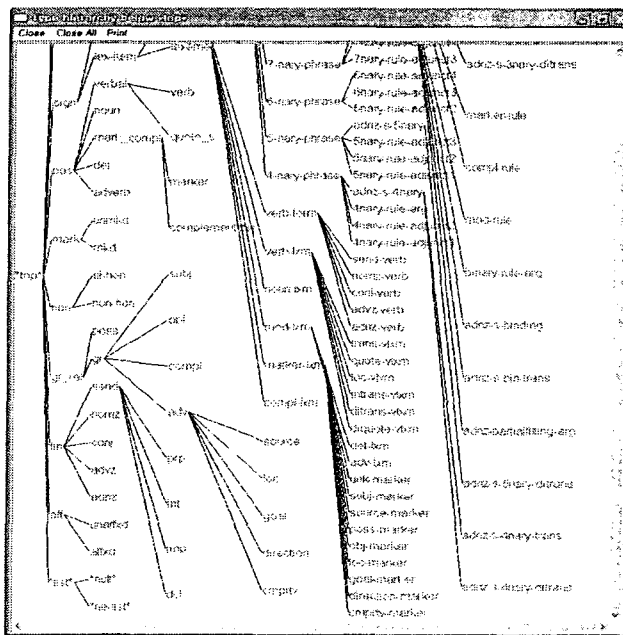


그림 4. 사용된 자질구조

- 3) 소위 중주어 문장과 '이다' 구성은 장석진(1993) 분석과 다르다. Subj2를 설정하고 단층으로 기술하지 않고 일단 서술어와 논항 하나가 결합하여 서술절을 이루고 다시 나머지 논항을 취하는 방식을 따른다.

4.3.1.1 sign

sign := *top* & [HEAD pos, AFF aff, GR gr_rel, COMPS *list*, ARGS *list*].

가. lex-item

- a. lexeme과 word로 구분.
- b. verb-lxm만이 inflection rule을 통해(inflr.tdl) 어미와 결합하여 verb-form을 형성.
- c. lexeme의 구조: verb-lxm은 동사가 취하는 논항의 수와 gr에 따라 구분, marker도 gr에 따라 구분

나. phrase

- a. n-nary phrase의 설정: unary부터 7-nary까지
- b. 동사의 종류와 논항부가어의 결합을 고려하여 문장 형성 규칙의 기본 타입을 정의.
(ex. 5nary-rule-adjunct2: 논항부가어 2 + 논항 2 + 동사 로 이루어진 문장의 기본 타입.)
- c. LKB에서 양분분석이 아닌 단층분석의 한계 더 좋은 방법 모색 필요

4.3.1.2 GR (Grammatical Relations)

가. poss와 gr 구분: 동사의 논항일 수 없는 것과 있는 것의 구분

나. adv: 동사에 따라서 논항일 수도 있고, 논항부가어일 수도 있다.

4.3.1.3 동사 활용과 관련한 자질

가. pos := *top* & [MOD *list*, FIN fin].

- a. MOD: 어떤 구를 피수식으로 요구하는지를 나타내는 자질
- b. FIN: 어미 관련 자질

나. fin := *top* & [HON hon, TNS tns].

- a. fin: 어말어미
- b. send 문장 종결 어미: dcl 평서문, int 의문문, imp 명령문, prp 청유문

다. nomz: 명사화 어미 (-ㅁ, 음, 기)

라. adnz 관형형 어미 (는, 은, 르, 던)

마. advz 부사화어미 (면, 니까, 나)

바. conj 접속어미 (-고, -데, -며, -나)

사. HON, TNS: 존칭/시제 등 선어말어미 관련 자질

4.3.1.4 기타 자질

가. AFF: 조사 결합 여부를 나타내는 자질

나. MARK: gr의 구체적 명사 여부를 나타내는 자질

4.3.2. 규칙들

가. (end) sentence rules: 동사와 동사가 요구하는 논항, 논항외 부가어로 이루어진다. 기본적인 규칙 패턴을 types.tdl에 정의하고, rules.tdl에서는 각 기본 패턴의 규칙 type을 논항과 논항 부가어들 간 순서 변화에 따라 세분하여 기술한다.

나. modifier rules

- a. types.tdl의 기본 유형: 수식어구+피수식어

- b. mod-rule type을 바탕으로 rules.tdl에 정의된 다양한 modifier rule들
 - det + NP : modNP-by-det
 - NPposs + NP : modNP-by-NPmkd / modNP-by-NPunmkd
 - NPposs + VP_nomz : modVPnomz-by-NPmkd (철수의 책을 읽음이... 등등)
 - adv + VP : modVP-by-adv
 - adnominalized sentence + NP : modNP-by-Sadnz

다. marker/complementizer rules: 조사(혹은 보문소) 결합 규칙

- a. marker_attatching_rule
- b. compl_attatching_rule

라. adnominalized sentence rules

- a. 완전한 구조의 문장이 NP를 수식하는 경우는 이미 modifier rule에 정의.
adnominalized sentence + NP : modNP-by-Sadnz

b. 관계화된 논항

- end sentence rule과 같이 논항수와 논항부가어의 수/위치 고려한 다수의 규칙이 필요.
- gap을 이용.
- 예: '입은'+ '옷'의 결합

adnz-partialfilling-subj := adnz-partialfilling-arg &

[HEAD noun, GR.MARK unmkd, COMPS < #1 >, (결합 후에도 주어를 여전히 요구함)

ARGS < phrase & [HEAD verb,

HEAD.MOD < #2 & [HEAD noun, GR.MARK unmkd] >,

HEAD.FIN adnz,

COMPS < #1 & [GR subj], #2 & [GR obj] >],

#2 >].

마. pumping rules

- a. word-to-phrase pumping
- b. 이중 주어 구문을 위한 자동사 문장 펌핑
 - 자동사 문장이 주어를 논항으로 요구하는 VP가 된다.
- c. nomz-pump rules: 명사형 어미가 결합한 문장을 NP로 바꾸어 준다.
 - 완전한 문장 펌핑 (철수가 책을 읽음...)
 - 주어가 없는 불완전한 문장 펌핑 + NP poss와 결합 (철수의 책을 읽음...)
- d. 인용문 펌핑
 - 완전한 구조의 인용문 펌핑.: 영희는 미아가 책을 읽었다고 생각했다.
 - 논항이 부족한 불완전한 구조의 인용문 펌핑 : gap을 이용한 것임. 결합 후에도 여전히 상위 노드의 COMPS가 요구하는 논항이 남아 있다. 예) 철수가 미아에게 집에 돌아가라고 요청했다.
 - 이외에도 타동사, 수어동사 문장에서 주어나 목적어 등이 빠진 인용문을 펌핑하는 rule 들

4.3.3. 사용된 노드 4)

S: 문장

S_nomz1: 주어가 주격으로 표시된 명사절

S_nomz2: 주어가 소유격으로 표시된 명사절

S_nomz-NP: 소유격 주어와 결합하기 전의 불완전 명사절

S_adnz: 관계(형)절

S_advz: 부사연결절

S_conj: 접속절

S_quo: ‘-고’ 보문절

S_quo-NP: 주어가 빠져 있는 (-NP) ‘-고’ 보문절

NP_S: 관계 수식절을 가진 명사구, "내가 살던 고향"과 같은 phrase 이름.

NP+VP-NP: 동사구에서 빠진 성분(VP-NP)을 계속 상위노드로 전달하여 올라갈 때의 명사구
"입은 옷이 벗진 신사"

NP_subj: 주어 명사구 (PP_subj 또는 NP_nom 등에 해당)

NP_obj: 목적어 명사구(PP_obj 또는 NP_acc 등에 해당)

NP_goal: 간접목적어 명사구(PP_goal, 또는 NP_dat 등에 해당)

NP_poss: 소유격 명사구

NP_loc: 처소를 나타내는 조사와 결합된 명사구 (PP_loc)

NP_dir: 방향을 나타내는 ‘-으로’ 와 결합된 명사구. 더 세분될 수 있다. ‘-으로’ 가 도구, 자격 등을 나타내기도 하기 때문에

NP_src: ‘-에서’ 조사와 결합된 명사구

NP_cmp: ‘-처럼’ 조사와 결합된 명사구

VP-NP_ARG5): 자매어 논항이 공백인 동사구 표시

VP-NP_ADJ: 자매어 첨가어가 공백인 동사구 표시

VP-NP_subj: 자매어 주어가 공백인 동사구 표시

VP-NP_obj: 자매어 목적어가 공백인 동사구 표시

VP-NP: 문장성분이 구체적으로 명시되지 않은 명사구가 공백인 동사구

VP: 동사구

NP: 명사구

N 명사

V: 동사

4) 여기서 사용되는 노드들은 장석진(1993)에 완전히 일치하는 것은 아니다. 이론적 관점에서의 문법기술과 대조하기 위해서 그리고 기존 기술과의 관련성을 보이기 위해 잠정적으로 설정된 것으로 앞으로 더 정제될 필요가 있다. 실제로 이런 노드는 LKB에서 parse node의 수정으로 쉽게 변경된다.

5) 여기서 VP-NP 등의 표시는 계층적 동사구를 가정한 경우 NP가 없는 것으로 여겨질 수 있으나 여기서는 자매어 노드로 NP가 생략된 것을 표시한다. 그 이유는 5.3에서 기술되지만 sag(1997) 등에서 제안된 대로 흔적없는 동사구 설정과 관련하여 설명하기 위해 이런 노드로 표시한다.

Det : 지시사

Adv: 부사

M : marker, 조사. 문법조사와 기타 부사어 조사를 구별안하고 모두 marker로 파악 6)

Comp : 보문소 '-고' . 동사의 활용보다는 조사와 같이 취급

5. 공백(Gap)/ 도출(Extraction)

5.1. 한국어에서의 extraction

가. 주제화(Topicalization) vs. Scrambling (김용범, 2003)

-- 주제화를 scrambling과 구별

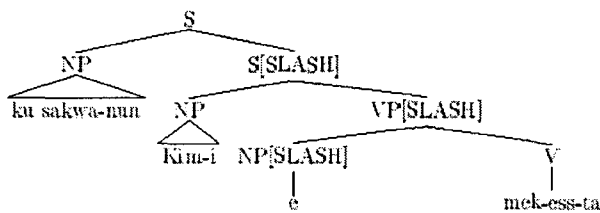


그림 1. 앞의 김종복 (1998)의 extraction

나. 김종복(1998)의 extraction 분석 -> Topicalization (김용범, 2003)

a. island constraints (complex Noun Phrase Constraint)

ㄱ. *[_j 방치가] 이 사태를 초래한] 병;

ㄴ. 우리가 [존이 메리에게 보낸 편지가 분실되었다는 사실]을 발견했다.

b. tough 구문

ㄱ. 우리가 [철수가 영화를 만났다는 사실]을 증명하기가 어렵다.

c. topic/focus construction (김용범,2003)-> relativization from a relative clause (김종복 1988)

ㄱ. 이 신사는 [_i, _j 입은] 옷이 더럽다.

ㄴ. [[_i _j 입고] 있는 옷이] 멋진 신사

d. relativization from an embedded sentential complement clause (김종복(1988)

ㄱ. 존이 [메리가 _i 먹었다고] 말한 사과;

5.2. 실제 구현관점에서의 공백/도출

가. 단층구조를 설정할 경우 공백은 관계절과 보문소 '-고' 에 의해 이루어지는 구성에서

나. 실제로 공백이 된 성분을 어떻게 찾고 결합하는지 구현

다. 이론적 기술과 평행성을 보이기 위해 VP-NP 등의 노드명 사용

6) 장석진 (1993)에서는 주어, 목적어 등을 표시하는 것은 marker로 그 외의 부사어 조사는 후치사구로 파악되나 여기서는 그 방법을 따르지 않는다.

5.3. 한국어 관계절 분석

가. 논항(argument)의 관계화

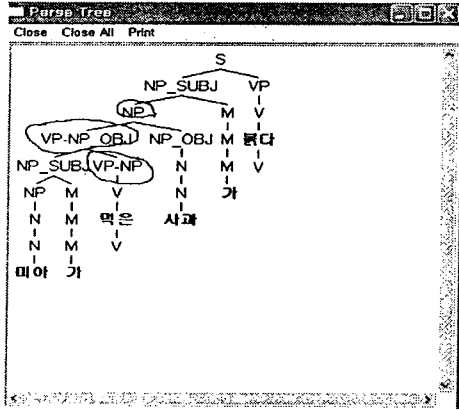


그림 5. 논항의 관계화 LKB 분석

다. 관계절에서의 관계화

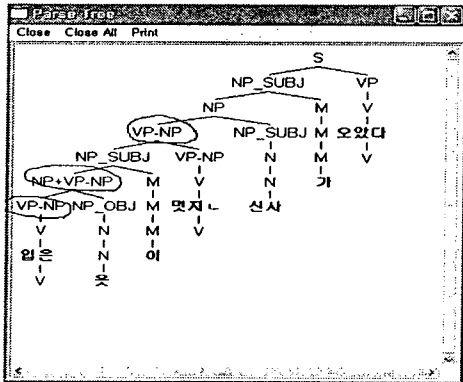


그림 7. 관계절에서의 관계화 LKB에서의 분석

라. 소유격 관계문

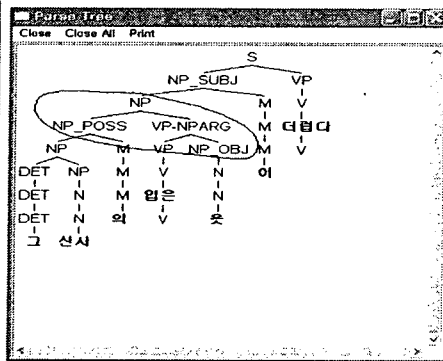
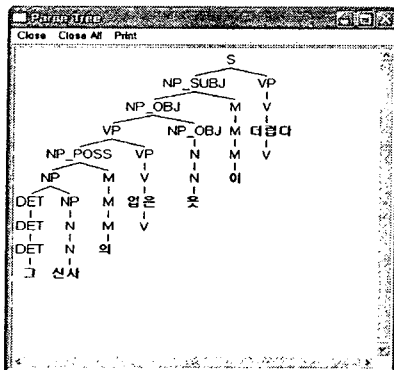


그림 9. 소유격 관계문의 두 가능한 분석

나. 부가어(adjunct)의 관계화

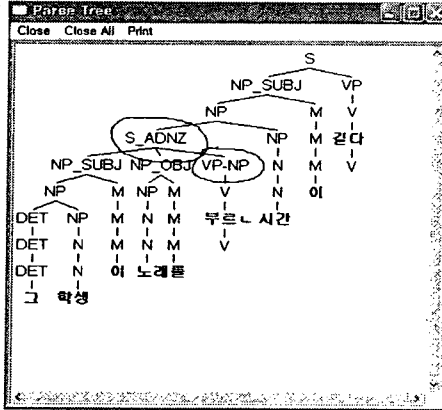


그림 6. 부가어의 관계화 LKB 분석

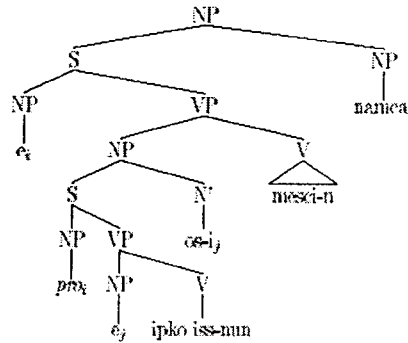


그림 8. 김종복 (1998)

마. 내포된 보문절에서의 관계화

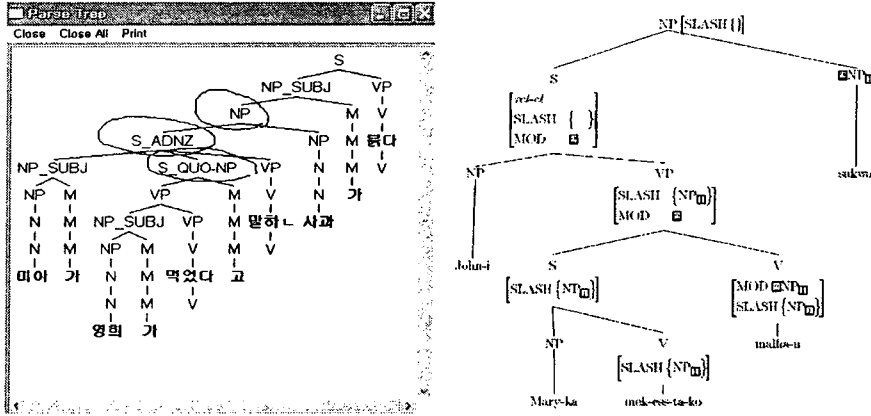


그림 10. 내포된 보문절에서의 관계화 분석

5.4. 명사화 분석

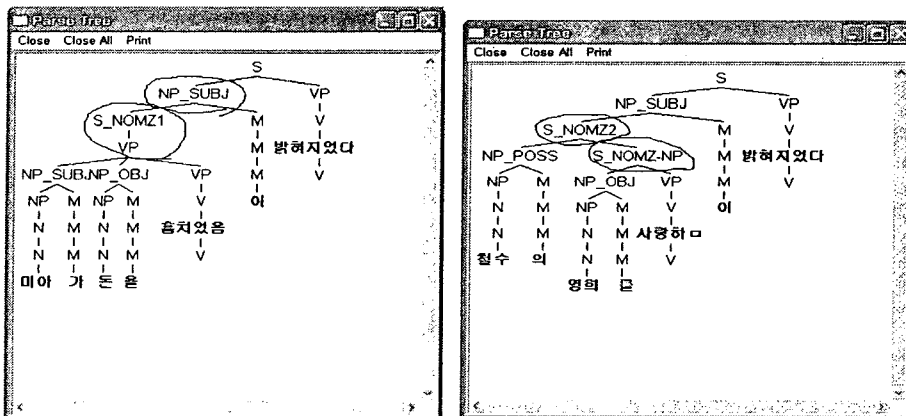


그림 11. 명사화 분석 결과

5.5. 복합문 분석

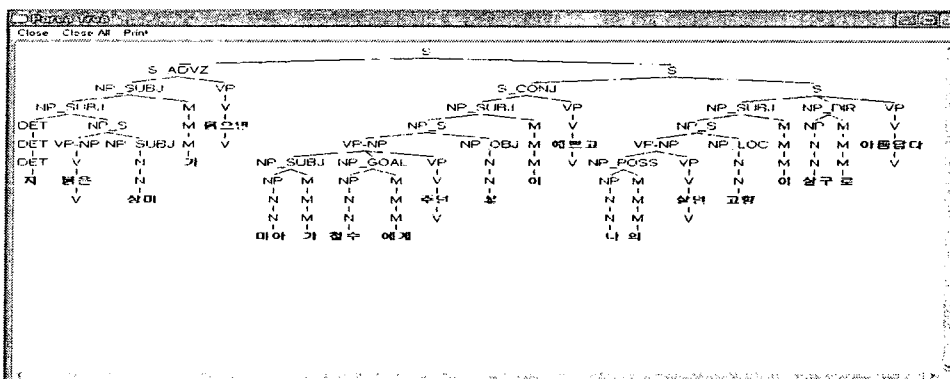


그림 12. 복합문 분석 1

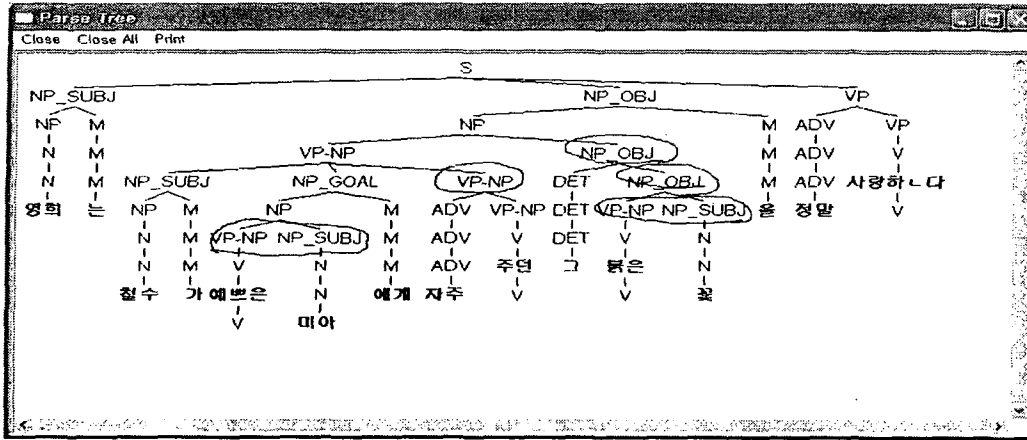


그림 13. 복합문 분석 2

6. 결론

- 가. 한국어 단층구조 분석의 LKB로의 구현 가능성
- 나. 양분 분석 대 단층분석
- 다. LKB의 여러 한계점 및 문제점 극복 필요
- 라. 언어학적 문법기술의 타당성 vs. 실제 구현적 타당성

참고문헌

- 김용범, 2003, *Some Extraction Constructions of Korean and Island Constraints*, 구구조문법연
구회 3월 발표자료.
- 김종복, 1998. *A Head-Driven and Constraint-Based Analysis of Korean Relative Clause
Constructions*, 어학연구 34-4, 서울대학교.
- 김종복, 양재형, 2003. *제약기반이론에서의 한국어구구조문법과 LKB(Linguistic Knowledge Base)
시스템을 이용한 구문분석기 구축*, draft.
- 신효필, 2000. *Toward More Efficient Processing of Typed Feature Structure in Korean*, 언어학
29호.
- 장석진, 1993. *정보기반 한국어 문법, 언어와 정보*.
- 채희락, 2003, *An IPSP Analysis of Gerund Phrases*, 한국언어정보학회 5월 발표자료.
- CopeStake, Ann, 2002. *Implementing Typed Feature Structure Grammars*, CSLI Lecture
Note 110, CSLI publications.
- Sag, Ivan, 1998, *English Relative Clause Constructions*. *Journal of Linguistics*, 33.