A Design for Korean Phrase Structure Grammar (KPSG) in ALE

Choi, Woon-Ho  Chang, Suk-Jin
Department of Linguistics
Seoul National University

요약
본 논문에서는 한국어의 전산처리를 위한 문법 모형 개발의 일부로서 HPSG에 기반한 문법 모형의 개발을 시도한다. 문법 모형의 개발에는 ALE(Attribute Logic Engine)를 이용하며, 보문 구조와 보조 음역 구문을 분석하기 위한 사전 구조 및 문법 규칙을 제시한다. 그리고 문의 종류 (Sentence Type: ST)와 문법(Sentence Level: SL), 시제, 존대 등을 분석해서 표상하기 위한 유형 개층 및 어휘부, 문법 규칙, 문법 원리 등을 제시한다.

1 들어가기
한국어의 전산처리를 위한 문법 모형에 대한 연구의 일부로, 본 논문은 ALE(Attribute Logic Engine)를 이용하여 한국어 문법을 설계하고, 한국어 문법과 사전 구조의 설계 및 응용 가능성을 모색하는 것을 목적으로 한다.

ALE는 Bob Carpenter와 Gerald Penn에 의해서 자재 구조(Typed Feature Structure) 논리를 바탕으로 Prolog를 이용해 구현한 논리 프로그램 엔진으로, 문법의 설계가 용이하고, 문법의 적합성 및 계산가능성 검사를 문법 개발과 동시에 진행할 수 있는 장점이 있다.

ALE를 이용하여 설계된 한국어 문법(이하 KPSG)은 유형 개층 정의(Type Hierarchy Definition), 어휘 항목, 동사 규칙, 문법 원리, 형태-음운 규칙으로 구성되어 있다. 현재 KPSG는 몇 가지 실험용 문장의 시점으로 응용 가능한 문법 구조의 원형을 설계하는 것을 목적으로 하기 때문에, 분석 가능한 문장에 제한을 두고 있다. 이 부분은 문법 및 어휘부의 확장을 통해서 차차 늘려가야 한다.

본 논문에서는, 보조 음역 구문, 음성 네트워크 구성의 처리를 중심으로 ALE를 이용해 설계된 한국어 문법 구조를 제시한다.

2 KPSG의 구조
KPSG는 기호(Sign)에 대한 유형 개층 정의를 바탕으로 어휘 항목 및 동사 구조, 문법 원리를 구성하고, 일부 형태-음운 예약을 사용한다. KPSG의 기본 구조는 다음과 같다.

(1) KPSG의 구조

<table>
<thead>
<tr>
<th>Lexical Entries</th>
<th>Syntactic Rules</th>
<th>Grammar Principles</th>
<th>Phonological Constraints</th>
<th>Lexical Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type Hierarchy Definition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.1 유형 개층(Type Hierarchy) 정의

(2) KPSG의 유형 개층 정의

<table>
<thead>
<tr>
<th>SYNSEM</th>
<th>CAT</th>
<th>HEAD</th>
<th>COMPS</th>
<th>SPN</th>
<th>MARKING</th>
<th>PN</th>
<th>MCONSTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CAT 자절에 의해 도입되는 HEAD 자절은 
HEAD 유형을 받으므로 가지며, HEAD 유형은 문 
법 기능을 담당하는 여의(FUNC)와 실질( 
SUBST)로 갈라진다. 현재까지 SUBST는 하위 
유형으로 VERB(용인)과 NOUN(세언)을 가지며, 
VERB는 동사, 형용사, 지정사로 세분된다. 
FUNC는 현재 MARK 값만이 담당되어 있으며, 
MARK는 문법 기능을 담당하는 표지로 조사(격 
조사), 부문소(고), 보조용언 연결 여미(고) 등 
이 담당되어 있다. (2)의 계층 구조는 [1]에서 제 
시된 환경(Context) 등에 대한 구조가 아직 도입 
되지 않았다. 

KPSG에서 CAT의 유형 계층 정리는 다음과 
같다.

(3) CAT 유형 정의

cat sub [ ]
intro [head:head, 
comps:list, 
spr:list, 
subj:list, 
marking:marking, 
phonconstr:phonesyntype].

2.2 동사 규칙

KPSG에서 현재 정의되어 있는 동사 규칙은 
Complement-Head Rule, Head-Marker Rule 2가 
자이다.

Complement-Head Rule은 머리어와 머리어의 보 
중어를 결합시켜 주는 규칙이다.

(4) Complement-Head Rule 
complement_head_rule
(Mother, synsem:cat:comps[])

==> goal > comp_num_rstr(Comps),
cats > (Comps, ne_list),
cat > (HeadDtr, synsem:cat:comps:Comps),
goal > head_feature_principle(Mother, HeadDtr),
goal > marking_principle(Mother, HeadDtr),
goal > semantics_principle(Mother, HeadDtr).

(5) Head-Marker Rule

head_marker_rule 
(Mother, synsem:cat:[spr[], comps[]])

==> cat > (HeadDtr,synsem:cat:[spr[MarkerDtr], 
comps[]]),
cat > (MarkerDtr, synsem:cat:[spr[], comps[]]),
goal > phon_constr(HeadDtr, MarkerDtr),
goal > spec_principle(HeadDtr, MarkerDtr),
goal > head_feature_principle(Mother,HeadDtr),
goal > marking_principle(Mother, MarkerDtr),
goal > semantics_principle(Mother, HeadDtr).

(6) Head Feature Principle

% head_feature_principle(Mother, HeadDtr)

head_feature_principle(synsem:cat:head:Head, 
synsem:cat:head:Head)

if true.

위 원칙이 달하는 것은 Mother Node의 Head 
자질의 값은 Head Daughter의 Head 자질의 값과 
일치해야 한다는 것이다.

(7) Marking Principle

%marking_principle(Mother, MarkerDaughter) 
marking_principle(synsem:cat:marking:Mark, 
synsem:cat:marking:Mark)

if true.

Marking Principle은 Mother Node의 marking 
값은 Marker Daughter의 marking 값이 일치해야
(제 10회 한국어 정보처리 학술대회)

(8) SPEC Principle
% spec_principle(HeadDtr, SprDtr)
   spec_principle(X, synsem=cat:head=\{mark, spec:X\})
   if true.

동일한 Mother Node와 Daughter Node 사이에서 Marker Daughter는 그것이 결합하는 Head Daughter를 명시하고 있고, 명시된 Head Daughter와만 결합할 수 있다고 말한다. 따라서, 조사 〜아/거가 동은 명사구와만 결합할 수 있고, 보문소 〜고는 완형 보문만을 이루게 된다. 그리고 〜고가 있다와 같은 보조 용언 구문을 구성하는 〜고는 용언과만 결합하도록 제약한다.

(9) Semantics Principle
% semantics_principle(Mother, SemHead)
   semantics_principle(synsem=cont:Cont,
      synsem=cont:Cont)
   if true.

Semantics Principle은 Mother Node와 Semantic Head Daughter의 CONT(ent)값이 동일해야 한다는 것을 제약한다.

2.4 어휘 항목

어휘 항목 중 ‘있다’와 보조용언 ‘있다’는 다음과 같이 구성되어 있다.

(10) ‘있다’에 대한 KPSG 사전 항목
mit -->
  spr[], subj:\{[@np_gf(gf_subj), @np_ind(Subj)]\},
  comps:[synsem:\{cat:head=\{verbal, vform:base, morphword:bfalse, subj:\{[@np_gf(gf_subj), @np_ind(Subj)]\},
    comps[], marking:unmarked, broncs:closedysl}),
  cont:\{nucleus:\{qfpsoa_state_progress, (soa_arg:Prop)\}\}]).

보조용언 ‘있다’는 사전 항목에서 보조용언이며 ‘〜고’가 첨가된 절을 보충어로 취하며, CONT(ent)는 보충어로 취한 절의 상태 진행을 나타낸다.

2.5 음운 제약 및 어휘 규칙
음운 제약은 실사부와 허사부의 결합에서 매개 모음을 삽입 등과 같은 처리를 하기 위한 제약 규칙이다. 어휘 규칙은 음언의 활용과 관련된 치를 담당하며, 어간을 바탕으로 어휘 생성을 제약한다.

문의 종류(Sentence Type: ST) 및 문계(Sentence Level: SL)의 결정은 음언 생성 어휘 규칙부에서 생성과 함께 결정된다. 즉, 어간에 결합되는 어미의 종류 및 유형에 따라서 문장의 종류 및 문장의 계층이 결정된다. 따라서, '있다'라는 어휘를 바탕으로 생성된 어휘 '있었다'의 HEAD 차질-값을 보면 다음과 같다.

(12) 문의 종류 및 문계
HEAD verbal
HON bfalse
INFL btrue
REGTYPEx irt_reg
MORPHWORD btrue
SL sl_plain
ST st_dcl
TNSF btrue
VTYPE vb_adj
VFORM past

(12)에서 보듯이 '있었다'는 문의 종류로 평서문(st_dcl), 문계로 평제(sl_plain)를 갖는다.

3. 분석 예
문장 "김이 운호가 밥을 먹었다-고 믿고 있었다"에 대한 최종 분석결과는 (12)와 같다.

(13) "김이 운호가 밥을 먹었다-고 믿고 있었다"의 분석 결과

STRING:
0 kim 1 i 2 winho 3 ka 4 pop 5 ul 6 mekessta
7 ko 8 mit 9 ko 10 issoessta 11
CATEGORY:
sign
SYNSEM synsem
CAT cat
COMPS e_list
HEAD verbal
HON bfalse
INFL btrue
REGTYPEx irt_reg
MORPHWORD btrue
SL sl_plain
ST st_dcl
TNSF btrue
VTYPE vb_adj
VFORM past
MARKING marking
PHONCNSTR phsyltype
SPR list
SUBJ list
CONT psoa
NUCLEUS qfpsoa_state_progress
SOA_ARG psoa
NUCLEUS qfpsoa_believe
QFP_BELIEVER ref
SOA_ARG psoa
NUCLEUS brel_mekta
EATEN ref
EATER ref

네포된 문장 "운호가 밥을 먹었다"까지의 중간 분석 결과는 (13)과 같다.

(14) "운호가 밥을 먹었다"의 분석 결과
sign
SYNSEM synsem
CAT cat
COMPS e_list
HEAD verbal
HON bfalse
INFL btrue
REGTYPEx irt_reg
MORPHWORD btrue
SL sl_plain
ST st_dcl
TNSF btrue
VTYPE vb verb
VFORM past
MARKING marking
PHONCNSTR phsyltype
SPR list
SUBJ list
CONT psoa
NUCLEUS brel_mekta
EATEN ref
EATER ref

Edge created for category above:
from: 2 to: 7
string: winho ka pap ul mekessta
rule: complement_head
# of dtrs: 1
(15) "은호-가 밥-을 먹었다-고"의 문석 결과는 다음과 같다.

sign
SYNSEM synsem
  CAT cat
    COMPS e_list
    HEAD verbal
      HON bfalse
      INF btrue
      REGTYPE irt_reg
      MORPHWORD btrue
      SL sl_plain
      ST st_dcl
      TNSF btrue
      VTYPE vb_verb
      VFORM past
    MARKING comp_ko
    PHONCNSTR phsyltype
    SPR e_list
    SUBJ list
  CONT psao
    NUCLEUS brel_mekta
    EATEN ref
    EATER ref

Edge created for category above:
  from: 2 to: 8
    string: wunho ka pap ul mekses ko
    rule: head_marker
# of dtrs: 2

4. 마무리

ALE를 이용한 한국어 문법 설계의 가능성에 대해서 살펴 보았다. 언어처리를 위한 한국어 문법은 사전 구조, 문법 규칙 등이 일치가 되어서 개선 가능해야 한다. KPSG는 현재 용언의 형태- 음운부와 동사부가 어휘 생성 규칙에 의해서 결합되어 있다. 따라서, 선어말 어미나, 어말 어미에 대해서는 통사적 자립성을 허용하지 않는다. 앞으 로의 문법 개발에서는 이 부분을 수정 보완해 나가야 할 필요가 있다. KPSG는 앞으로 통합 문법 에 바탕을 두고 의미-용용 부분까지 포함해서 하나의 둘 안에 설계할 계획이며, 전산 처리 가능한 문법의 설계를 목적으로 확장해 나갈 것이다.