

한국어 특수구문 처리를 위한 파서의 구현

(An implementation of parser for special syntax processing in Korean)

金載文*, 李相國**, 李相祚*

(Jae Mun Kim, Sang Kuk Lee and Sang Jo Lee)

要約

본 논문에서는 한국어 특수구문의 처리를 위한 국어 구문 분석 시스템을 제안한다. 문법의 기술은 통사와 의미를 통합적으로 처리하는 HPSG를 채택하고, 파싱 기법으로는 한국어에 유리한 단방향 활성 차트 파싱을 사용한다. 본 논문의 파서는 포괄적인 문장 구조(보어-중심어 구조, 수식어-중심어 구조, 중심어-중심어 구조)의 처리뿐만 아니라, 실용적인 문장에서 많이 나타나는 보조용언 구문, 사동문, 피동문, 명사화 어미, 존칭, 화제와 같은 특수구문에 대해서도 파싱을 할 수 있도록 구현되었다.

Abstract

In this paper, we propose a Korean syntax analysis system for special syntax processing. HPSG, which processes syntactic and semantic analysis unificationally, is chosen for grammar description. Head-driven unidirectional active chart parser, which is efficient in Korean processing, is used for parsing mechanism. The parser of this paper can analyze not only general sentence structure which consists of complement-head, adjunct-head and head-head structure but also special syntax which consists of auxiliary verb sentence, causative sentence, passive sentence and so on.

1. 서론

컴퓨터의 사용이 보편화됨에 따라 사용자들이 쉽게 컴퓨터 시스템을 사용할 수 있는 자연어 인터페이스 시스템이나, 기계 번역(machine translation), 정

보 검색(information retrieval)등의 다양한 응용 시스템이 연구 개발되고 있다. 이러한 자연어 응용 시스템에서, 입력 문장의 구조 및 의미를 추출하는 파서는 중요한 역할을 한다. 실용적인 시스템에 이용되는 파서의 효율은 주로 다양한 문장을 처리할 수 있는 파싱의 범위와 모호성의 처리에 달려있다. 한국어의 문법 기술을 위해서는 통합기반문법^{[14][16][17][18]} 중 어휘부의 역할이 특히 강조되어 구절 구조 규칙의 수와 복잡도를 매우 감소시킨 중심어 주도의 통합 문법

*正會員, **準會員, 慶北大學校 컴퓨터 工學科
(Dept. of Computer Eng., Kyungpook Nat'l Univ.)
接受日字: 1994年 3月 22日

(HPSG)^[18]을 이용한다. 파싱 기법으로서는 국어의 중심어 후행 어순 특성과 교착어 특성을 이용한 중심어 주도의 단방향 활성 차트 파싱^[3]을 이용한다. 이 파싱 기법은 일단 중심어를 찾으면 중심어의 왼편으로만 분석하기 때문에, 일반적인 중심어 주도의 파싱 기법인 양방향 차트 파서의 중복 분석 문제를 해결할 수 있다. 또, 입력 문장으로부터 획득한 의미를 표현하기 위해서 HPSG에서 도입한 상황 의미론^[13]을 이용해서 형식화한다.

이러한 분야에 대한 연구는, 먼저 정희성 파서^[10]로서, 국어 분석을 위해 Left-corner 차트 파싱 알고리즘을 이용하여 구현하였고, 통사 분석에 치중하여 의미부를 소홀히 다루었다. 양재형 파서^[5]는 중심어 주도의 양방향 차트 파싱 알고리즘을 사용하여 구현하였고, 이는, 국어와 같은 중심어 후행 언어에 있어서는 중복 분석의 문제가 존재하여 비효율적인 면이 존재한다. 서영훈 파서^[3]는 중심어 주도의 단방향 차트 파싱 알고리즘을 제안 구현하였고, 이는 국어 분석에 효과적이나 통사분석에 치중한 나머지 의미부를 소홀히 한 면이 있다. 또한 김운호 파서^[11]는 기존의 대부분의 파서들이 통사적인 모델만 제시한데 반해 의미부의 의미 표현을 중점으로 다루었으나, 세부사항의 고려없이 포괄적인 처리로 분석 가능한 문장 처리가 다양치 못한 점이 있다.

본 논문에서는 지금까지 HPSG에 근거하여 일반적인 해결책을 제시한, 보어-중심어 구조, 수식어-중심어 구조 및 중심어-중심어 구조등의 포괄적인 문장 구조의 처리뿐만 아니라, 여러가지 실용적인 문장에서 많이 나타나는 보조용언 구문, 사동문, 피동문, 명사화 어미(‘ㄴ, 기’), 존칭, 화제와 같은 세부적인 특성을 지닌 문장에 대해서도 적절한 처리 방법을 제안함으로써, 대부분의 실용적인 국어 문장에 대해서 분석할 수 있는 파서를 구현하였다. 또한, 입력받은 문장에 대해서 통사 및 의미를 통합적으로 처리하고, 입력 문장에서 추출한 정보를 응용 시스템의 생성기 부분에서 사용할 수 있는 내적 표현(internal representation)으로 형식화 한다. 즉, 본 논문에서는 기존의 HPSG파서^{[11][3]} 보다 파싱의 범위가 확장된 국어 문장 분석기를 제안하므로써, 여타의 응용 시스템의 일반적인 한국어 분석 모델로 참조될 수 있을 것이다.

논문 구성은 2장에서는 본 논문에서 채택한 문법 이론인 중심어 주도의 통합 문법에 대해 기술하고, 국어 구문 분석기의 전체적인 구성도를 보인다. 그리고 중심어 주도의 파싱 기법에 대해 기술하고, 특수 구문을 처리할 수 있도록 분석 시스템을 확장한다. 3

장에서는 구현된 국어 분석기에 대한 실험 및 분석 결과에 대해서 검토하고, 4장에서는 결론 및 추후 연구 과제에 대해 기술한다.

II. 파싱 시스템의 설계 및 구현

1. 중심어 주도의 문법

본 논문에서는 국어 문법 기술을 위해 중심어 주도의 통합 기반 문법인 HPSG^[18]형태의 문법체계를 도입한다.^{[1][2][3][4][7]} HPSG는 Pollard에 의해 주도된 이론으로서 기존의 구 구조 규칙에 의해 표현되던 대부분의 통사 정보^{[10][16]}를 어휘부의 어휘 정보로 표현함으로써, 구 구조 규칙의 수를 감소시키고, 비변형적이고 단층적인(unistratal) 문법으로의 기술을 가능하게 하였다. 통사와 의미를 통합적으로 기술하는 HPSG는 보편 문법을 지향하고 있으며 자질 구조에 입각한 정보구조를 바탕으로 어휘부(lexicon)와 소수의 보편적 원리들, 개별 언어의 원리들, 개별 언어의 어휘부의 기호 및 문법 규칙들로 구성되어 있다. 보편 원리에는 중심어 자질 원리(head feature principle), 하위범주화 원리(subcategorization), 의미 원리(semantics principle)등이 있다.

1) 자질 구조

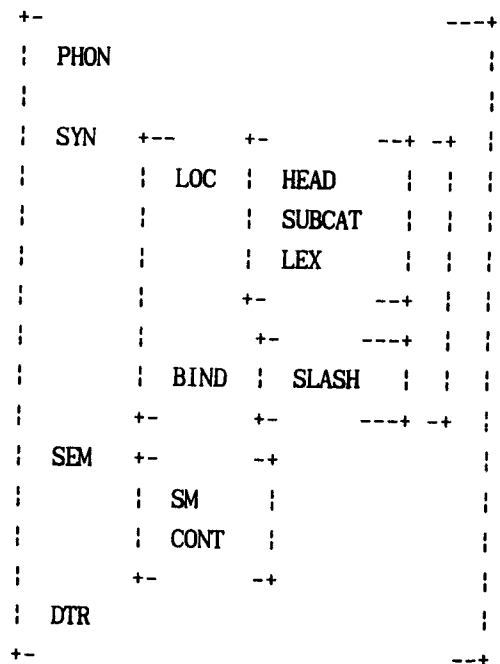


그림 1. 자질 구조

Fig. 1. Feature structure.

본 논문에서 기술하는 국어 문법은 언어적 정보의 형식화를 위해 (그림 1)과 같은 자질 구조(Feature Structure)를 이용한다.

각 자질들에 대한 자세한 설명은 [1][3]에 상세히 기술되어 있다. 여기서는 새롭게 도입한 자질에 대해서 중점적으로 기술한다. 먼저, 어휘의 품사를 나타내는 MAJOR 자질과 중심어 자질에 어간의 형태를 나타내기 위해 STEM 자질을 설정한다. STEM 자질이 가질 수 있는 값은 {root(원형), fcause(사동화), inact(피동화)} 중 하나이다. 그리고 본용언과 보조용언의 구별 자질 AUX{+, -}를 설정하고, 화계를 나타내는 SL(speech level)자질을 설정한다.

2) 의미 표현 및 합성

HPSG 이론에서는 의미론으로서 상황 의미론^[3]을 받아들여 통사와 의미를 통합적으로 기술하고 있다. 즉, 문장 내의 각 어휘 항목 또는 구성 성분이 갖고 있는 개개의 의미 정보를 '통사적 결합과 동시에 의미 정보의 합성'을 행함으로써 문장의 부분 또는 전체 의미(situation-theoretic semantics)를 합성해 낸다.^[1]

▲ 보어-중심어 구조의 의미 합성

```
[ DTRS headed-structure [ ] =>
--
:          +- RELN 13: !SEM !CONT !DST !RELN --+
: SEM !CONT !DST ! ARG-n 12: !PHON
:
:          +-
:          +-
: DTRS ! HEAD-DTR 13: !SYN !LOC !SUBCAT union(!1!, !2!)
: ! COMP-DTRS 12:
:
:          +-
:          +-
--
```

▲ 수식어-중심어 구조의 의미 합성

① 수식적 결합의 경우

```
[ DTRS headed-structure [ ] =>
--
:          +- RELN 11: !SEM !CONT !DST !RELN --+
: SEM !CONT !DST ! ARG1 11: !SEM !CONT !DST !ARG1 <13!>
:
:          +-
:          +-
: DTRS ! HEAD-DTR 11:
: ! ADJT-DTRS 12: !SEM !CONT !DST 13:
:
:          +-
:          +-
--
(<13!>표기는 연결 connection을 나타냄)
```

② 부가적 결합의 경우

```
[ DTRS headed-structure [ ] =>
--
:          +- RELN 11: !SEM !CONT !DST !RELN <13!> --+
: SEM !CONT !DST ! ARG1 11: !SEM !CONT !DST !ARG1
:
:          +-
:          +-
: DTRS ! HEAD-DTR 11:
: ! ADJT-DTRS 12: !SEM !CONT !DST 13:
:
:          +-
:          +-
--
```

▲ 중심어-중심어 구조의 의미 합성

```
[ DTRS headed-structure [ ] =>
--
:          +- RELN 12: ! RELN & 11: ! RELN --+
: SEM !CONT !DST ! ARG1 12: ! ARG & 11: ! ARG
:
:          +-
:          +-
: DTRS ! Righthead ! SEM !CONT !DST !1:
: ! Lefthead ! SEM !CONT !DST !2:
:
:          +-
:          +-
--
```

2. 전체 시스템 구성도

본 논문에서 구현한 한국어 구문 분석기의 전체적인 구조는 (그림 2)와 같다.

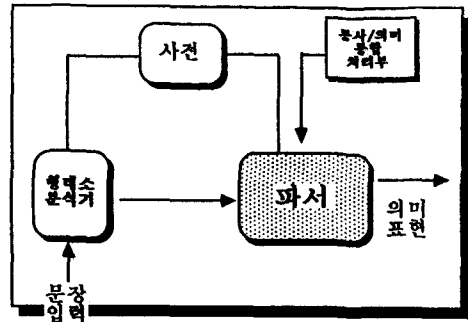


그림 2. 분석 시스템의 구조
Fig. 2. Structure of analysis system.

국어 입력 문장이 들어오면 기존의 형태소 분석기^[6]를 이용하여 형태소 분석을 행하고, 형태소 분석 한 결과가 파서로 입력되어 통사 및 의미 분석을 거친 후 입력 문장에 상응하는 의미 표현을 생성한다.

3. 일반문형의 파싱

본 논문에서 사용하는 상향식 차트 파싱의 흐름도를 간략히 나타내면 (그림 3)과 같다.

중심어 후행 어순이 '전역적' 통사 현상인 국어 분석에서는 통사 분석을 하기 전에 '지역적' 통사 현상인 어순 특성(조사가 채운뒤에 위치, 보조용언이 본용언뒤에 위치에 해당하는 성분들을 미리 처리하고 통사분석으로 들어가게 함으로써 '전역적' 구문 현상인 중심어 후행 어순 특성에 의거한 일관된 통사 분석이 가능하다.

본 논문에서는 국어 통사 분석을 위한 파서로 국어의 중심어 후행 어순 특성을 이용하여 파서의 효율과 융통성을 최대화 하는 "중심어 주도의 상향식 차트 파서"^{[1][3]}를 사용한다. 파서에 이용되는 대표적인 자

료 구조로서 INPUT은 형태소 분석 및 중간 분석을 마친 문장 성분에 대한 자질 구조에 노드 정보를 추가한 에지들을 저장하기 위한 리스트 형태의 자료 구조이다. PENDED는 활성 에지(active edge, 현재 에지들을 포함하여 현재 에지에 의해 결합되어 추가적 분석을 기다리는 에지들)를 저장하기 자료 구조이다. CHART는 분석이 종료된 비활성에지(inactive edge)들을 저장하기 위한 자료 구조이다.

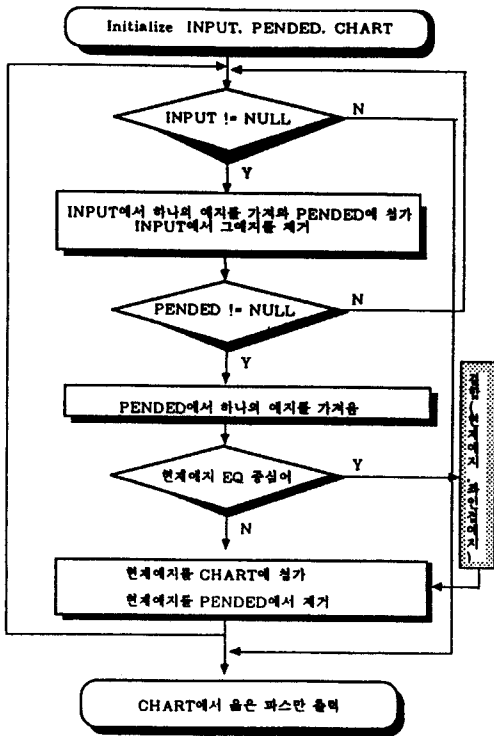


그림 3. 상향식 차트 파싱 흐름도
Fig. 3. Flow chart of bottom-up chart parsing.

4. 특수구문의 파싱

본 논문에서는 기존의 파서들^{[17][6]}이 고려하지 않고 있는 보조용언 구문, 사동문, 피동문, 명사화 어미, 존칭, 화계와 같은 문장 구조에 대한 의미 표현 처리 방안을 제시한다.

1) 보조용언 구문

보조용언의 겉모습은 일반적인 동사, 형용사와 다름이 없지만, 자립성이 약하여 본용언에 기대어 그 말의 뜻을 도와주는 역할을 하며, 자립성을 가지고 실질적인 의미를 나타내며 단독으로 서술 능력을 갖는 본용언에 양태적 의미만 더할 뿐 주어에 대한 서

술어 기능을 하지 못한다. 보조용언은 형식상으로는 본용언의 한정을 받는 위치이지만 의미 및 기능상으로는 본용언을 보조하는 주종(主從)관계이다. 보조용언구문은 보조연결어미가 첨가된 본용언과 보조용언의 복합용언으로 형성된다. 보조용언을 서상, 서법의 의미에 따라 분류하면 (그림 4)와 같다.

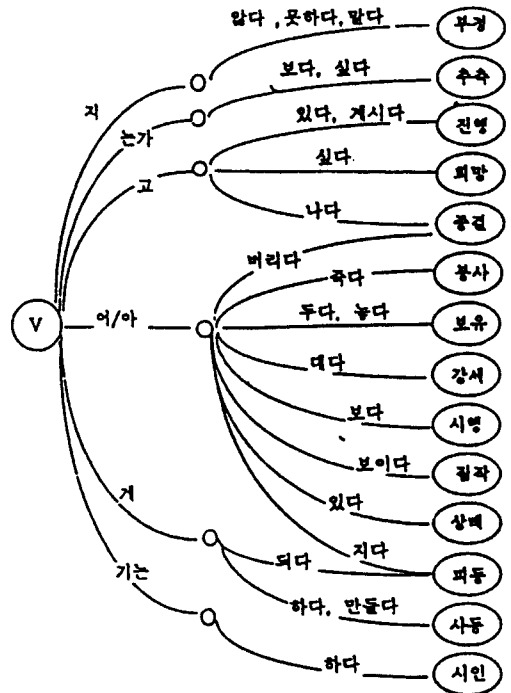


그림 4. 보조용언의 분류
Fig. 4. Specification of auxiliary verb.

본 논문에서는 보조용언 구문의 처리를 위해 동사의 어말어미 정보를 나타내는 VFORM 자질에 보조연결어미(complementizer) -'아/어, 게, 지, 고, 고야, 기는, 는가'- 에 해당하는 COMP 값을 설정, 보조연결어미의 어휘정보에 부여한다. 이 VFORM 자질은 이후 각 보조용언이 하위범주화하는 보어를 제약하는데 사용된다. 또한 보조용언/본용언의 구별을 위해 중심어자질(HEAD)에 AUX[+, -] 자질을 설정한다.

(1) 부정 보조용언 구문

'아니다(않다), 못하다, 말다' 등의 부정사는 보조연결어미 '지'가 첨가된 본용언과 결합하여 부정보조용언 구문을 형성한다. 부정보조용언들이 문과 결합하는 구조는 다음 세가지가 가능하다.^[8]
예) 저 여자는 예쁘지 않다.

㉔ 내부(어휘적) : 저 여자는 [예쁘지 않다]

㉕ 외부(문장적) : [저 여자는 예쁘지] 않다.

㉖ 중간(VP적) : 저 여자는 [_ 예쁘지] 않다.

㉔는 내부부정(internal negation)적이고, 어휘규칙으로 도출할 수 있다. ㉕는 하나의 전체 문장을 요구하는 외부부정(external negation)으로서, 문장을 주어로 하는 지정사 '아니다' 처럼 기술하는 분석으로 무리가 있다. ㉖는 '않다'가 하위범주화 값으로서 SUBCAT { NP(subj), VP() }를 가지며, VP의 주어는 전체 주어와 동일하다는 통사적인 분석 방법이다. 이러한 부정문에 대한 세가지 가능한 분석 중에서 외부부정은 부적당하고, 어휘적 분석과 통사적 분석은 서로 장단점이 있으나 본 논문에서는 파서의 특성상 통사적 분석 방법을 택한다. 전체 문장의 의미 및 구조를 지배하는 것은 본용언이고, 부정보조용언은 단지 그문장에 부정적인 의미만 부여하는 부여할 뿐이다. 즉, 전체 문장의 의미는 본용언의 하위 범주화 값에 의해 결정되는 의미부에 의해 나타내어진다. 그래서 본용언이 요구하는 보어들도 부정보조용언 "않다"가 요구하는 보어들과 중복되어 생략 가능하다고 본다. 그러므로 본 논문에서는 전체 문장의 의미부는 본용언의 의미부를 승계시키며, 통사적 분석 방법의 의해 부정보조용언을 파싱한다.

이의 처리를 위한 부정보조용언의 어휘 정보는 다음과 같다.

```

+-
: PHON 아니하(않)/ 말 / 못하
: SYN:LOC:HEAD:AUX '+'
: SYN:LOC:HEAD:NEG '+'
: SYN:LOC:SUBCAT (NP(subj), VP(SUBCAT(NP(subj))), VFORM comp지) ;
+-

```

부정 보조 구문의 처리 과정은 현재에지(current edge)가 보조용언이고, 좌인접에지(LAE)가 자신의 하위범주화 값을 만족하는 보어일 경우, 보조용언이 중심어가 되면서 보편원리에 입각하여, 보어-중심어 결합을 행한다. 보어-중심어 구조에 적용되는 하위범주화 원리에 의해 본용언이 결합된 VP 성분(보어)과 부정보조용언(중심어)의 통합적 결합이 행해질 때, 통합 성분의 의미 자질은 본용언의 의미 자질로 통합되고, 통합 성분의 의미부 중 극성을 나타내는 POLARITY 자질은 부정보조용언의 자질로 통합된다. 그러므로, 기존의 보어-중심어 의미합성과는 달리 새로운 의미합성 원리가 세워져야만 한다 (그림 5).

다음의 예는 본 논문의 parser에 의해 분석된 결과를 보여준다.

예문) 저 여자는 예쁘지 않다.

```

+-
: PHON 저 여자는 예쁘지 않다
: SYN:LOC:HEAD:AUX '+'
: SYN:LOC:HEAD:NEG '+'
: SYN:LOC:SUBCAT { }
:
+- RELN pretty +-
: SEM:CONT: DST | PRETTIER 여자 |
:
: (MODIFIER) 저 |
:
: POLAR ZERO |
:
+-
+-

```

```

[ DTRS headed-structure [ ] =>
+-
:
+- RELN 12: | SEM | CONT | DST | RELN +- |
: SEM | CONT | DST | ARG-n 12: | SEM | CONT | DST | ARG-1: |
:
+- POLAR 13: | SEM | CONT | DST | POLAR +- |
:
+-
: DTRS | HEAD-DTR 13: | SYN | LOC | SUBCAT union(11, 12): |
:
: COMP-DTRS 12:
:
+-
+-

```

그림 5. 본용언(VP)-부정보조용언(HEAD)의 의미 통합
 Fig. 5. Verb-Negative auxiliary verb semantic unification.

(2) 피동 보조용언 구문

'지다' 보조용언은 일부 한정된 타동사나 사동사에 붙어 보조연결어미 '어/아'를 개재하여 피동의 구문을 형성한다. 본 논문에서는 이런 피동보조구문을 처리하기 위해서, '지다'가 결합된 피동사를 어휘규칙에 의해 사전에 등재한다. 본 논문에서 설정하는 피동화 규칙은 다음과 같다.

▲ 피동화규칙 #1 : 어간이 '쇠되하다, 붉다' 등의 자동사/형용사의 경우, 어휘규칙 전후에 하위범주화 자질 값, SUBCAT {NP(subj)} 값을 그대로 유지한다. 즉, 보어가 하나이므로 의미 역할(semantic role) 에도 아무런 변화도 없다.

```

+-
: PHON '자동사의 어간'
: SYN:LOC:HEAD:STEM root
: SYN:LOC:SUBCAT (NP(subj)<1>) => SYN:LOC:SUBCAT (NP(subj)<1>)
+-RELN +-
: SEM:CONT: DST | ARG1(ER)<1> | : SEM:CONT: DST | ARG1(ER)<1> |
+-
+-

```

▲ 피동화규칙 #2 : 어간이 '풀다, 밝히다' 등의 타동사/타동사화한 사동사의 경우, 하위범주화 자질 값 SUBCAT{NP(subj)<1>, NP(obj)<2>}이 SUBCAT{NP(subj)<2>, NP(adv, pform 에 의해)<1>}로

변화된다. 세자리 보어를 요구하는 수어동사와 같은 경우는 하위범주화 자질 값이 SUBCAT{NP(subj)<1>, NP(obj)<2>, NP(obj2)<3>}에서 SUBCAT{NP(adv. pform에의 해)<1>, NP(subj)<2>, NP(obj2)<3>}으로 변화된다.

PHON '타동사의 어간'	PHON '타동사의 어간-어-지'
SYN:LOC:HEAD:STEM root	SYN:LOC:HEAD:STEM inact
SYN:LOC:SUBCAT {NP(subj)<1>, NP(obj)<2>}	SYN:LOC:SUBCAT {NP(adv. pform에의 해)<1>, NP(subj)<2>}
+RELN _	+RELN _
SEM:CONT:DST ARG1(_ER)<1>	SEM:CONT:DST ARG1(_ER) <1>
ARG2(_EE)<2>	ARG2(_EE) <2>

피동보조구문의 처리는 피동화 규칙이 적용된 복합 용언의 하위범주화 값에 의한 보어-중심어 결합을 한다. 파싱 도중 현재 예지가 복합용언이고 확인접예지가 자신의 보어일 경우 현재 예지가 중심어가 되면서 통합을 시도한다. 의미 합성 과정은 기존의 보어-중심어 합성 과정에 의해 보어 성분을 하위범주화 자질에 명시된 것과 같은 지표를 가진, 중심어의 의미부논항 값에 통합한다.

다음의 예는 본 논문의 파서에 의해 분석된 결과를 보여준다.

예문) 그의 오해가 광수에 의해 풀어졌다.

PHON 그의 오해가 광수에 의해 풀어졌다
SYN:LOC:HEAD:AUX '+'
SYN:LOC:HEAD:STEM inact
SYN:LOC:SUBCAT { }
+RELN solve
SEM:CONT:DST SOLVER 광수
SOLVEE 오해
(MODIFIER) 그의

(3) 사동 보조용언 구문

보용언에 보조연결어미 '게'와 '하다, 만들다' 보조 용언이 결합하여 사동 보조용언 구문을 형성한다. 보조 용언에 의한 사동 구문을 사동접사에 의한 어휘적인 사동과 구별하기 위하여 통사적으로 분석한다. 이런 보조 용언 구문은 '하다, 만들다'와 같은 보조용언이 중심어 성분이 되며, 하위범주화 값으로서 주어, 목적어, VP의 세자리 보충어를 취한다. 전체 문장의 목적이 성분 이 이 VP의 주어이다. 사동 보조용언 구문을 처리하기 위한 '하다, 만들다'의 어휘 정보는 다음과 같다.

PHON 하/만들
SYN:LOC:HEAD:AUX '+'
SYN:LOC:HEAD:STEM fcause
SYN:LOC:SUBCAT {NP(subj), NP(obj) or NP(obj2), VP(SUBCAT{NP(subj)}, VFORM comp)}

사동 보조 구문의 처리는 입력 문장의 분석 도중 사동보조용언 '하다/만들다'가 현재예지일 때, 이것의 하위범주화 값에 의한 보어-중심어 결합을 한다. 보어-중심어 구조에 적용되는 하위범주화 원리에 의해 본용언이 결합된 VP(보어)와 사동보조용언(중심어)의 통합적 결합이 행해질 때, 일반 보편 원리들은 그대로 적용되고, 의미 합성의 경우 통합 성분의 관계 자질은 VP의 관계 자질로 통합된다. VP의 논항 값들은 같은 지표를 갖는 통합 성분의 논항 값에 통합되고, 의미 합성 원리는 부정 보조 구문의 합성원리와 유사하다.

다음의 예는 본논문의 파서에 의해 분석된 결과를 보여준다.

예문) 엄마가 애에게 밥을 먹게 했다.

PHON 엄마가 애에게 밥을 먹게 했다
SYN:LOC:HEAD:AUX '+'
SYN:LOC:HEAD:STEM inact
SYN:LOC:SUBCAT { }
+RELN eat
SEM:CONT:DST EATER 애
EATEE 밥
CAUSER 엄마

(4) 기타 보조용언 구문

그 밖의 다른 여러 형태의 보조용언 구문의 처리는 각 보조용언의 SUBCAT 리스트에 의한 통사적 분석을 행함으로 처리를 할 수 있다. 보조용언들의 하위범주화 값은 SUBCAT{NP(subj), VP(SUBCAT{NP(subj)}, VFORM comp)}으로 VP의 주어는 문장 전체의 주어와 동일하다. 그리고 VP의 VFORM은 여러 형태의 문장에 사용되는 각 보조용언들이 그 문장 특성에 맞게 제약하는 보조연결어미에 따라 적당하게 부여한다. 또한, '고-있다'와 같은 진행, '고-나다'와 같은 종결, '어-있다'와 같은 상태를 나타내는 구문등에는 서상(aspect) 정보를 부여한다.

2) 사동문과 피동문

자동사나 타동사의 기본형 어간이 어휘 규칙에 의해서 사동/피동화될 때, 용언이 요구하는 보어의 수와 전체 문장에서 각 성분들의 의미 역할이 바뀌게

된다. 그러므로 기본형 어간의 하위범주화 값을 그대로 유지하면 접사의 결합에 의한 사동문/피동문을 분석할 수 없거나 의미상으로 오분석을 초래할 수 있다. 본 논문에서는 어간이 사동/피동 접사와 어휘 규칙에 의해 결합할 시에 용언의 하위범주화 값과 의미를 적절하게 조정함으로써 이 문제를 해결하고자 한다. 그리고 이런 형태 문장의 통사적인 정보를 더욱 정확히 표현하기 위해서 'STEM' 자질을 설정한다. STEM 자질이 가질 수 있는 값은 { root, fcause, inact } 중의 하나로서, 기본형일 경우는 root, 사동형일 경우는 fcause, 피동형일 경우는 inact이다.

(1) 사동문

사동문에는 사동사에 의한 것과 '-게 하다'와 같은 보조 구문에 의한 형성법으로 나누어 지는데, 보조구문에 의한 사동은 앞서 해결책을 제시하였으므로 여기서는 사동사에 의한 것만 처리 대상으로 한다. 사동사는 용언의 어근에 파생 접사('이, 히, 기, 리, 우, 구, 추')를 결합시킴으로써 형성된다.

① 자동사에서 유래한 경우

주동문의 주어가 사동문의 목적어가 되며, 자동사인 서술어가 사동화 규칙에 의해 사동사로 바뀌고, 사동문에서는 새로운 행위자인 주어(사동주, causer)가 나타난다. 여기에 적용되는 사동화 규칙은 다음과 같다.

▲ 사동화규칙 #1

+-	+-	+-	+-
: PHON 자동사 어간	:	: PHON 자동사 어간 + 사동접사	:
: SYN:LOC:HEAD:STEM root	:	: SYN:LOC:HEAD:STEM fcause	:
: SYN:LOC:SUBCAT { NP(subj)<1> }	=>	: SYN:LOC:SUBCAT { NP(obj)<1>, NP(subj)<2> }	:
: ++RELN	--	: ++RELN	--
: SEM:CONT:DST	:	: SEM:CONT:DST	:
: ARG1(_ER)<1>	:	: ARG1(_ER)<1>	:
+-	+-	+-	+-

주동문에서, 자동사는 필수적인 보어로서 주어 하나만을 요구하므로 사전에 등재된 기본형 어간의 하위범주화 자질 값은 SUBCAT{ NP(subj) }이다. 하지만 자동사에서 유래된 사동문을 처리하기 위해 설정한 사동화 규칙 #1이 적용된 사동사는 필수 보어로서 주어, 목적어를 요구하므로 용언의 하위범주화 자질 값은 SUBCAT{ NP(subj), NP(obj) }으로 변경되어야 한다. 그리고 문장의 전체 의미를 올바르게 전달할 수 있도록 각 성분의 의미역도 바뀌어야 한다. 이러한 사동문의 처리는 어휘 규칙이 적용된 사동사를 사전에 등재하고, 입력 문장의 분석 도중 사동사가 현재에서로서 중심어 성분이 될 때, 자신의 하위범주화 값에 의한 보어-중심어 결합을 시도한다.

예를들어, 기본형 어간 '눅'의 자질 구조와 사동화된 '눅+이'의 자질 구조를 간략히 보인다.

+-	+-	+-	+-
: PHON 눅	:	: PHON 눅+이	:
: SYN:LOC:HEAD:STEM root	:	: SYN:LOC:HEAD:STEM fcause	:
: SYN:LOC:SUBCAT { NP(subj)<1> }	:	: SYN:LOC:SUBCAT { NP(obj)<1>, NP(obj)<2> }	:
: SEM:CONT:DST	--	: SEM:CONT:DST	--
: MELTER<1>	:	: MELTER<1>	:
+-	+-	+-	+-

다음의 예는 본 논문의 파서에 의해 분석된 결과를 보여준다.

예문) 아이들이 얼음을 녹인다.

+-	+-	+-	+-
: PHON 아이들이 얼음을 녹인다	:		:
: SYN:LOC:HEAD:STEM fcause	:		:
: SYN:LOC:SUBCAT { }	:		:
: SEM:CONT:DST	--	: RELN melt	--
:		: MELTER 얼음	:
:		: CAUSER 아이들	:
+-	+-	+-	+-

② 타동사에서 유래한 경우

목적어는 변화하지 않으나, 서술어가 주동사에서 사동사로 바뀌며, 주동문의 주어가 '에게'라는 부사격 조사를 취하여 부사어가 되고, 사동문에서는 새로운 행위자인 주어(사동주, causer)가 등장하게 된다. 여기에 적용되는 사동화 규칙은 다음과 같다.

▲ 사동화규칙 #2

+-	+-	+-	+-
: PHON 타동사 어간	:	: PHON 타동사 어간 + 사동접사	:
: SYN:LOC:HEAD:STEM root	:	: SYN:LOC:HEAD:STEM fcause	:
: SYN:LOC:SUBCAT { NP(subj)<1> }	=>	: SYN:LOC:SUBCAT { NP(obj2)<1>, NP(obj)<2>, NP(subj)<3> }	:
: ++RELN	--	: ++RELN	--
: SEM:CONT:DST	:	: SEM:CONT:DST	:
: ARG1(_ER)<1>	:	: ARG1(_ER)<1>	:
: ARG2(_EE)<2>	:	: ARG2(_EE)<2>	:
+-	+-	+-	+-

주동문에서, 타동사는 필수적인 보어로서 주어와 목적어를 요구하므로 사전에 등재된 기본형 어간의 하위범주화 자질 값은 SUBCAT{ NP(subj), NP(obj) }이다. 하지만 사동화규칙 #2가 적용된 사동사는 필수 보어로서 주어, 목적어를 요구하므로 용언의 하위범주화 자질 값은 SUBCAT{ NP(subj), NP(obj2), NP(obj) }으로 변경되어야 한다. 그리고 문장 전체의 의미를 표현하는 의미역도 적절하게 바뀌어야 한다. 이러한 사동문의 처리는 어휘 규칙이 적용된 사동사를 사전에 등재하고, 입력문장의 분석 도중 사동사가 현재에서로서 중심어 성분이 될 때, 자신의 하위범주화 값에 의한 보어-중심어 결합을 시도한다. 예를들어 기본형 어간 '먹'의 자질 구조와 사동화된 '먹+이'의 자질 구조를 간략히 보인다.

```

+-          +-          +-          +-
: PHON 먹          : PHON 먹이
:SYN:LOC:HEAD:STEM root      :SYN:LOC:HEAD:STEM root
:SYN:LOC:SUBCAT { NP(subj)<1> :SYN:LOC:SUBCAT { NP(obj2)<1>
:          NP(obj)<2> } :          NP(obj)<2>
:          NP(obj)<3> } :          NP(subj)<3>
:SEM:CONT:DST +- RELN eat +- :SYN:LOC:HEAD:STEM fcause
:          EATER<1> :          :
:          EATEE<2> :SEM:CONT:DST +- RELN eat +-
:          :          : EATER <1> :
:          :          : EATEE <2> :
+-          +-          +-          +-
+-          +-          +-          +-

```

다음의 예는 본 논문의 파서에 의해 분석된 결과를 보여준다.

예문) 어머니가 철수에게 밥을 먹였다.

```

+-          +-
: PHON 어머니가 철수에게 밥을 먹였다
:SYN:LOC:HEAD:STEM fcause
:SYN:LOC:HEAD:SUBCAT { }
:SEM:CONT:DST +- RELN eat +-
:          EATER 철수
:          EATEE 밥
+-          +-
+-CAUSER 어머니+-

```

(2) 피동문

피동문에는 사동 구문에서처럼 접사가 첨가된 피동사에 의한 것과 '-아/-어지다'와 같은 보조용언이 결합된 피동구문이 있다. 이밖에 동사성 명사에 '되다'가 붙어 피동의 뜻을 나타내기도 한다. 보조용언에 의한 피동구문은 앞서 해결책을 제시하였으므로, 여기서는 피동사에 의한 것과 '되다'에 의한 피동 구문을 처리 대상으로 한다. 피동사는 타동사의 어근에 파생 접사('이, 히, 리, 기')를 결합시킴으로써 형성된다. 능동문의 목적어가 피동문의 주어 자리에 오게되며, 능동문의 주어는 부사격 조사 '에게(한테)'가 붙어 부사어가 되는 동시에 타동사였던 서술어는 피동사로 바뀌어 피동태를 이룬다. 여기에 적용되는 피동화규칙은 다음과 같다.

▲ 피동화규칙 #3

```

+-          +-          +-          +-
: PHON 타동사 어간          : PHON 타동사 어간 + 피동결사
:SYN:LOC:HEAD:STEM root      :SYN:LOC:HEAD:STEM root
:SYN:LOC:SUBCAT { NP(subj)<1> :SYN:LOC:SUBCAT { NP(obj2)<1>
:          NP(obj)<2> } :          NP(obj)<2>
:          NP(obj)<3> } :          NP(subj)<3>
:SEM:CONT:DST +- RELN eat +- :SEM:CONT:DST +- RELN
:          ARG1(_ER)<1> :          ARG1(_ER)<1>
:          ARG2(_EE)<2> :          ARG2(_EE)<2>
+-          +-          +-          +-

```

능동문에서, 타동사는 필수적인 보어로서 주어와 목적어를 요구하므로 사전에 등재된 기본형 어간의 하위범주화 자질 값은 SUBCAT{ NP(subj), NP(obj) }이다. 하지만 피동문에서는, 능동문의 목적어가 주어로 되고 주어가 부사어로 되므로, 하위범

주화 자질 값 NP(obj)는 NP(subj)로, NP(subj)는 NP(obj2)로 변경되어야 한다. 이러한 피동문의 처리를 위해서 피동화규칙 #3을 설정하고, 이 어휘 규칙이 적용된 피동사를 사전에 등재한다. 입력 문장의 분석 도중 현재어지가 피동사일 때, 이 피동사가 중심어 성분이 되며 자신의 하위범주화 값에 의해 좌인 접어지와 보어-중심어 결합을 시도한다. 그리고 하위범주화 조건이 만족되는 성분들의 의미 자질은 지표에 의해 피동사의 논항에 통합된다. 예를들어 기본형 어간 '잡'의 자질 구조와 피동화된 '잡+히'의 자질 구조를 간략히 보인다.

```

+-          +-          +-          +-
: PHON 잡          : PHON 잡히
:SYN:LOC:HEAD:STEM root      :SYN:LOC:HEAD:STEM freact
:SYN:LOC:SUBCAT { NP(subj)<1> :SYN:LOC:SUBCAT { NP(obj2)<1>
:          NP(obj)<2> } :          NP(subj)<2>
:SEM:CONT:DST +- RELN hold +- :SEM:CONT:DST +- RELN hold +-
:          HOLDER <1> :          HOLDER <1>
:          HOLDEE<2> :          HOLDEE <2>
+-          +-          +-          +-
+-          +-          +-          +-

```

다음의 예는 본 논문의 파서에 의해 분석된 결과를 보여준다.

예문) 도둑이 순경에게 잡혔다.

```

+-          +-
: PHON 도둑이 순경에게 잡혔다
:SYN:LOC:HEAD:STEM inact
:SYN:LOC:SUBCAT { }
:SEM:CONT:DST +- RELN hold +-
:          HOLDER 순경
:          HOLDEE 도둑
+-          +-
+-          +-          +-          +-

```

또한, 하나의 단어가 자동사, 타동사, 사동사, 피동사등 여러가지 의미로 해석 가능한 경우에는, 각 경우에 해당하는 구문 정보및 의미 정보를 부여해서 각각 사전에 등재한다. 즉, 사전에는 phon 자질이 같고 특성이 다른 여러개의 단어가 등재된다. 그래서, 자동문, 타동문, 사동문, 피동문이 입력되었을 경우, 해당되는 동사 만이 통합에 성공을 하고 나머지는 통합에 실패를 한다. 예를 들어,

- ① 그소리는 너무 자주 들어서 물리다. (자동사, get sick)
SUBCAT {NP(subj), VP(subcat(NP(obj)))}
- ② 의자를 뒤로 물리다. (타동사, take away)
SUBCAT {NP(obj)}
- ③ 그에게 손해를 물리다. (사동사, make compensate)

기반 문법의 대표적인 연산인 통합(Unification)에 의해서 존칭을 해결한다. HONOR 자질이 가질 수 있는 값은 존칭을 나타낼 경우에는 '+', 그렇지 않을 때는 '-'값을 가진다. 그러므로 사전에 등재할 '시, 계서, 계옵서, 님'의 HONOR 자질을 '+'로 한다. 용언의 어간과 선어말 어미 '시'가 결합할 때에, 결합된 용언의 HONOR 자질은 '시'의 HONOR 자질 값으로 통합된다. 마찬가지로 체언과 높임조사, 접미사의 결합시에도 결합된 성분의 HONOR 자질은 높임조사, 접미사의 HONOR 자질로 통합된다. 전체 문장을 분석할 경우, 보어(주어)와 중심어가 하위범주화에 의한 결합시, 주어 성분의 HONOR 자질과 중심어 성분의 HONOR 자질이 일치하면 통합에 성공하고, 그렇지 않으면 실패하게 된다. 즉, 주술간의 일치 현상에 의해 기술한다. 직접 높임의 경우에는 이런 방법으로 해결이 가능하다. 간접 높임은 이중 주어 문제와 연계되어 있으므로 좀 더 많은 연구가 필요하다.

상대 높임은 화자가 청자의 사회적 지위나 나이 등에 따라, 특정한 종결 어미를 씀으로써 상대방을 높이고 낮추는 것이다. 본 논문에서는 화자와 청자의 사회적 관계를 나타내기 위해서 화계(speech level)를 사용한다. 화계는 세가지의 대인 관계 -사회 계층적인 수직 관계, 심리적 유대 관계인 수평 관계, 외형적인 격식 관계-의 복합적인 작용에 의해서 결정된다.^[7] 화계를 나타내는 자질로서 중심어 자질에 SL을 설정하고, 그 값은 어말어미에 따라 분류할 수 있다.

III. 분석 및 검토

본 논문에서는 통사 분석 및 의미를 통합적으로 처리 할 뿐만 아니라, 문장으로부터 추출된 의미를 더욱 적절하고 일반성 있게 표현하기 위해 의미부의 의미 표현도 중요시 다루었다. 그리고 포괄적인 문장 구조의 처리 및 사동/피동문, 명사화 어미, 존칭, 화계, 등의 특수구문에 대해 적절한 처리 방법들을 제시함으로써 분석 가능한 문장의 범위를 확장, 다양한 국어 실용 문장을 처리하도록 파서 시스템을 설계하였다. 보어-중심어 구조의 하위범주화 원리, 수식어-중심어 구조의 관계 관형화 구성, 명사구보문화 구성, 결어 성분 통제, 첨가어에 의한 부가 구성과, 중심어-중심어 구조의 등위 접속 구문, 특수구문 등, 각 문장 구문별로 예문을 선택하여 실험을 한 결과 모든 문장에 대해 옳은 파스가 생성되었다. 그렇지만, 자연어 처리의 가장 중요한 문제 중 하나인 모호

성으로 인해 과도한 파스 트리가 생성되는 경우가 많았다. 특히 수식 범위의 애매성으로 인한 수식 관계 모호성이 많이 발생했다. 의미 표지를 이용한 의미 선택 제약을 가함으로써 많은 모호성이 해결될 수 있었으나, 모든 모호성을 해결하기는 미흡하다.

실험예문의 대상은 고교 문법책^[11]에서 각 분장별로 발제를 하여 실험하였다. 실험결과는 표 1에서 보여 주고있으며, 결과에서 보듯이, 실패한 구문들은 "V-고-싶-어-하고-있다"와 같이 보조용언들이 중첩되는 다중 보조용언구문, 주어 및 목적어가 중첩되는 이중 주어, 이중목적어 구문, "어/아" 로서 피동문을 형성한 '주어지다'와 같은 동사를 띄어쓰기 할 경우, 격 설정이 애매한 경우, 예를 들면 '어머니가 너로 하여금 애를 보게하였다' 등의 경우에는 분석이 실패하였으며, 이에 대해서는 계속 연구 중에 있다. 사용된 예문의 대부분이 몇개의 모호한 파스가 출력 되었는데, 이는 대부분 '수식어-중심어' 구성에서, 수식어가 한정하는 중심어 범위의 애매성에서 오는 모호성과 문장 자체의 모호한 경우가 대부분이다.

표 1. 구문별 파스 출력 결과

Table. 1. The classification results of syntax parse.

구분	부정보조용언	피동문	자동사 유대 사동문	타동사 유대 사동문
일체 문장수	78	72	68	74
실패한 문장수	8	4	5	6
평균 출력 파스	5.7	1.9	2.5	2.2

IV. 결론

본 논문에서는 특수구문의 처리를 위한 국어 파서 시스템을 구현하였다. 국어의 문법 기술을 위해서 국어가 가지는 부분 자유 어순 특성을 고려하여, 구 구조 규칙의 수가 매우 감소된 중심어 주도의 통합 기반 문법을 이용하였다. 그리고 국어와 같은 중심어 후행 언어에 대하여 효율적인 중심어 주도의 단방향 활성 차트 파싱을 사용하여, 보어-중심어, 수식어-중심어, 중심어중심어 구조뿐만 아니라 보조용언구문, 사동문, 피동문, 존칭, 화계, 명사화 어미 등 다양한 문장에 대해서도 의미 표현에 중점을 두어 분석을 할 수 있었다.

본 논문의 연구 성과는 언어학의 전산학적 구현과

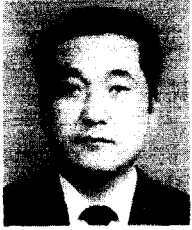
응용 가능성이라는 두가지 측면에서 요약될 수 있다. 첫째, 통사 분석 모델에만 치중된 기존의 틀에서 벗어나, 통사와 의미를 통합적으로 처리하여 의미부의 의미 표현을 중요시 다루었으며, 둘째, 여러가지 특수구문에 대하여 새로운 해결 방안을 제시함으로써 파싱 범위를 확장하였고, 대부분의 실용적인 국어 문장에 대해서 처리를 할 수 있음을 실험을 통하여 입증하였다. 그러므로 본 논문에서 구현된 국어 분석 시스템이 여타의 자연언어 처리 시스템에 활용될 수 있는 응용 가능성을 확인하였다.

앞으로 더욱 실용적인 자연어 처리 시스템을 구축하기 위해서는 모호성 해결을 위한 처리 기법과, 아직까지 일반적인 해결책을 제시하지 못한 여타의 실용적인 구문에 대해 집중적인 연구가 있어야 할 것이다.

參 考 文 獻

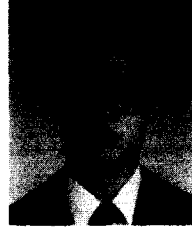
- [1] 김윤호, HPSG에 기반한 국어 분석기의 구현, 경북대 컴퓨터공학과 석사학위 논문, 1992
- [2] 김혜리, 한국어 어휘부-기반 문법 정립을 위한 기초연구, 서울대 언어학과 석사학위 논문, 1988
- [3] 서영훈, 의미 정보를 이용하는 중심어 주도의 한국어 파싱, 서울대 컴퓨터공학과 박사학위 논문, 1991
- [4] 신호필, HPSG를 기초로한 한국어 동사의 하위범주화, 서울대 언어학과 석사학위 논문, 1990
- [5] 양재형, HPSG에 기반한 한국어 분석기의 연구, 서울대 컴퓨터 공학과, 공학 석사 학위 논문, 1990
- [6] 이상조, 한국어 자연어 인터페이스를 위한 사전 구성에 관한 연구(II), 한국전자통신연구소 위탁 과제, 경북대 전자기술연구소, 1992
- [7] 장석진, "HPSG: 통사와 의미의 통합적 기술", 일봉 최준기 박사 회갑기념논문집, pp. 1001-1029, 1987
- [8] 장석진, 정보기반 한국어 문법, 도서출판 언어와정보, 1993
- [9] 정희성, "통합 지식 표현에 의한 자연 언어 처리 기법: 한글 구 구조 문법", 제2회 기계 번역 workshop 발표 논문집, PP.143-158, 1989
- [10] 정희성, "한글 구 구조 문법", 제 1회 자연 언어 처리 워크샵 발표 논문집, pp.3-37, 1987
- [11] 조규빈, 하이라이트 고교 문법 자습서, 지학사, 1993
- [12] Aho, A.V and J.D. Ulman, *The theory of Parsing, Translation, and Compiling*, Prentice-Hall, 1972.
- [13] Barwise and Perry, "Shifting Situations and Shaken Attitudes", *Linguistics and Philosophy* 4, pp.159-219, 1985.
- [14] Bresnan and Kaplan, "Lexical Functional Grammar: A Formal System for Grammatical Representation", *The Mental Representation of Grammatical Relations*, ed. J. Bresnan, 1982.
- [15] Early, "An Efficient Context-Free Parsing Algorithm", *Readings in Natural Language Processing*, ed. B.J. Grosz et al., pp.25-34, 1970.
- [16] Gazdar, et al., *Generalized Phrase Structure Grammar*, 1985.
- [17] Gunji, *Japanese Phrase Structure Grammar, A Unification based Approach*, D.reidel Pub, 1987.
- [18] Pollard and Sag, *Information-Based Syntax and Semantics, Vol.1: Fundamentals*, CSLI Lecture Notes, 1987.
- [19] Satta and Stock, "Head-driven Bidirectional Parsing: A Tabular Method", *International Workshop on Parsing Technologies*, pp.43-51, 1989.
- [20] Stock, et al., "Island Parsing and Bidirectional Charts", *Proc. 12th International Conference on Computational Linguistics*, pp.636-641, 1988.
- [21] Tomita M., *Efficient Parsing for Natural Language*, Kluwer Academic Publishers, 1986.

著者紹介



金 載 文(正會員)

1980年 2月 육군사관학교(전자공학전공) 졸업(이학사). 1988年 2月 경북대학교 컴퓨터 공학과 졸업(공학석사). 1992年 3月 ~ 현재 경북대학교 컴퓨터 공학과 대학원 박사과정 재학중. 주관심 분야는 자연어 처리, 기계 번역, 데이터 베이스 등임.



李 相 國(準會員)

1992年 2月 경북대학교 컴퓨터 공학과 졸업(공학사). 1994年 2月 경북대학교 컴퓨터 공학과 대학원 졸업(공학석사). 1994年 3月 한국 MJL 기술 연구소 연구원. 주관심 분야는 자연어 처리, 기계번역, 데이터 베이스 등임.

李 相 祚(正會員) 第 31卷 第 8號 參照
현재 경북대학교 컴퓨터 공학과 교수