

# 문장성분의 다양한 자질을 이용한 한국어 구문분석 모델

박 소 영<sup>†</sup> · 김 수 홍<sup>††</sup> · 임 해 창<sup>†††</sup>

## 요 약

본 논문에서는 효과적인 구문 중의성 해결을 위해 문장성분의 구문자질, 기능자질, 내용자질, 크기자질을 활용하는 확률적 한국어 구문분석 모델을 제안한다. 그리고, 제안하는 구문분석 모델은 한국어의 부분자유어순과 생략현상을 잘 처리할 수 있도록 문법규칙을 이진형식으로 제한한다. 실험을 통해 제안하는 구문분석 모델의 성능을 각 자질조합별로 분석한다. 분석결과를 서로 다른 특징을 갖는 자질의 조합이 서로 유사한 특징을 갖는 자질의 조합보다 구문중의성 해결에 더 유용하다는 것을 보여준다. 또한, 단일자질인 기능자질이 내용자질과 크기자질의 조합보다 성능이 더 우수함을 알 수 있다.

## Korean Parsing Model using Various Features of a Syntactic Object

So-Young Park<sup>†</sup> · Soo-Hong Kim<sup>††</sup> · Hae-Chang Rim<sup>†††</sup>

## ABSTRACT

In this paper, we propose a probabilistic Korean parsing model using a syntactic feature, a functional feature, a content feature, and a size feature of a syntactic object for effective syntactic disambiguation. It restricts grammar rules to binary-oriented form to deal with Korean properties such as variable word order and constituent ellipsis. In experiments, we analyze the parsing performance of each feature combination. Experimental results show that the combination of different features is preferred to the combination of similar features. Besides, it is remarkable that the function feature is more useful than the combination of the content feature and the size feature.

**키워드 :** 확률적 구문분석 모델(Probabilistic Parsing Model), 한국어 문법(Korean Grammar), 자질조합(Feature Combination)

### 1. 서 론

구문분석은 문법을 바탕으로 문장의 구문구조를 분석하는 문제이다. 이를 위해 그동안 다양한 구문분석모델이 제안되었는데, 그 대표적인 예가 확률적 문맥 자유 문법이다. 확률적 문맥 자유 문법은 문법 규칙에 따라 구문구조를 생성하고, 구문구조의 생성확률을 추정하고, 확률이 최대인 구문구조를 문장의 구문구조로 선택한다. 그러나, 문법규칙이 동일한 경우 적절한 구문구조를 선택할 수 없다는 문제가 있다.

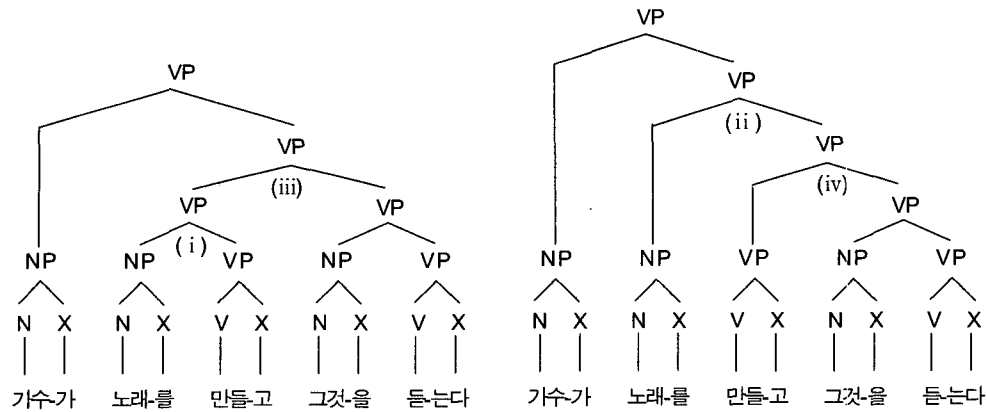
예를 들어, 문장 “가수가 노래를 만들고 그것을 듣는다”의 구문구조를 분석해 보자. 확률적 문맥 자유 문법은 (그림 1)의 두 구문구조를 동일한 확률로 생성하므로, 최적의 구문구조를 선택할 수 없다. 즉, (그림 1)의 두 구문구조에서 서로 다른 부분인 (i) “VP → NP VP”와 (ii) “VP →

NP VP”의 문맥 자유 문법 규칙이 같고, (iii) “VP → VP VP”와 (iv) “VP → VP VP”의 문맥 자유 문법 규칙이 같으므로, 구문구조의 생성확률은 동일하게 된다.

이러한 구문 중의성 해결 능력을 개선하기 위해서 어휘화된 구문분석 모델이 등장하였다[6, 7, 14, 16, 18, 23]. 이러한 모델에서는 (그림 1)과 같이 어휘화 규칙 (i) “VP/만드 → NP/노래 VP/만드”와 어휘화 규칙 (ii) “VP/듣 → NP/노래 VP/듣”이 다르므로, 두 구문구조의 생성확률이 달라진다. 이를 바탕으로 최대의 확률을 가지는 구문구조를 선택할 수 있다. 그러나 어휘화 규칙 (iii) “VP/듣 → VP/만드 VP/듣”과 어휘화 규칙 (iv) “VP/듣 → VP/만드 VP/듣”은 여전히 같다. 따라서, 어휘화된 구문분석 모델에 내부 문맥[20, 23]이나, 외부문맥[6]이나, 생성된 과정[10, 14, 19]을 함께 고려하여 구문분석 성능을 향상시키는 연구가 진행되고 있다.

본 논문에서는 (그림 1)의 구문구조에서 규칙 (i), (ii), (iii), (iv)를 구분할 수 있도록 문장성분의 다양한 자질을

<sup>†</sup> 준 회원 : 고려대학교 대학원 컴퓨터학과  
<sup>††</sup> 종신회원 : 상명대학교 컴퓨터정보통신공학부 교수  
<sup>†††</sup> 종신회원 : 고려대학교 컴퓨터학과 교수  
 논문접수 : 2004년 5월 12일, 심사완료 : 2004년 7월 26일



(그림 1) 문맥 자유 문법에서 생성한 구문구조 후보

사용하는 구문분석 방법을 제안한다. 문장성분을 구문 레이블과 단일 중심어 어휘로 표현하는 기존의 어휘화된 구문분석 모델과 달리, 제안하는 방법은 구문분석에 유용한 정보를 더 많이 활용하기 위해서 문장성분을 구문 레이블, 내용 중심어, 기능 중심어, 크기로 표현한다. 앞으로, 제2장에서는 한국어 구문분석을 위한 기존 문법 모델에 대해 살펴본다. 그리고, 제3장에서는 제안하는 확률적 자기기반 한국어 문법 모델에 대해 설명하고, 제4장에서는 제안하는 모델의 성능을 평가한다. 마지막으로 제5장에서는 결론을 맺는다.

## 2. 기존 연구

문법은 각 문장의 의미 구조와 기능 구조를 명시하는 중요한 역할을 한다. 최근에 한국어 문장을 분석하기 위해서 몇몇 문법이 제안되고 있는데, 단일화 기반문법, 구구조문법, 의존문법, 범주문법, 어휘화된 트리인접 문법으로 구분할 수 있다. 하지만, 한국어의 특성을 제대로 처리하지 못해서 몇 가지 문제점이 제기 되고 있으며, 이를 좀 더 자세히 살펴보면 다음과 같다.

HPSG와 같은 단일화 기반 문법은 복잡한 언어 현상을 정확하게 포착할 수 있도록 도와준다[4, 26]. 그러나, 이러한 문법은 너무 정교하고 복잡해서 다양한 분야에 적용하기에는 개발에 어려움이 따른다[3, 23]. 게다가, 단일화 연산은 심각한 중의성을 야기하며[11], 문장을 분석하는 CPU시간의 대략 90% 정도가 단일화 연산에 소요된다[24]. 그러므로, 최근 경향은 중간 구조 공유나 여과와 같은 기법을 도입하여 효율성을 개선하는 쪽으로 연구되고 있다[11, 24].

구구조문법은 주어진 문장에 구구조를 할당하는 문법이다. 구구조문법을 이용한 구문분석의 최근 경향은 구문분석 말뭉치로부터 문법규칙을 학습하여, 말뭉치 형식과 비슷하게 분석결과를 생성하는 것이다[13]. 그러나, 구조중의성 해결에 중요한 역할을 하는 어휘정보를 제대로 반영하지 못한

다는 문제가 있다[21]. 이러한 문제를 보완하기 위해서, 최근 구문분석 모델은 어휘정보를 함께 포함하는 경향이 있다[7, 16, 23]. 게다가, [6]에서는 한국어가 어순이 비교적 자유롭고 일부 문장성분이 생략될 수 있으므로, 문법규칙의 형식을 제한해야 한다고 기술하고 있다.

의존문법은 문장에서 두 단어 사이의 의존관계를 분석한다. 의존관계는 구구조 규칙보다 훨씬 간단하므로 문법 학습 시 자료 부족 문제에 영향을 덜 받는다[21]. 또한, 이러한 의존관계는 자유로운 어순과 생략현상을 잘 다룰 수 있으므로, 한국어 구문분석에 많이 적용되어 연구되고 있다[1, 2]. 그러나, 구조정보를 이용할 수 없으므로 둘 이상의 서술어가 주어를 공유하는 문장을 표현할 수 없다는 한계가 있다[12]. 게다가, 일부 휴리스틱 규칙에도 불구하고 의존관계를 분석할 때 의미 없는 관계가 지나치게 많이 생성될 수 있다[3, 6].

범주문법은 범주에 순행연산과 역행연산을 적용하여 파스 트리를 유도한다. [22]에서는 한국어의 자유어순을 처리할 수 있도록 집합 개념을 도입하지만, 구조 중의성과 범주 중의성을 여전히 포함하고 있다. 반면에, [8]는 대등구조에서 의미식정보를 사용하여 중의성을 완화하는 방법을 제안한다. [12]는 어휘정보를 활용하지 못하는 범주문법을 보완하여, 중의성 해결시 구문분석모델에 어휘정보를 추가하는 방법을 제안한다. 그럼에도 불구하고, 범주문법은 다양한 범위에 적용되기 위한 어휘사전이 요구된다.

어휘화된 트리인접 문법은 기본트리와 보조트리에 대치연산과 부가연산을 적용하여 문장을 분석한다[5]. 이 문법은 각 트리의 구조정보와 트리 끝에 있는 어휘 정보를 모두 이용할 수 있다. 반면에 각 어휘에 대한 설명을 제공하는 복잡한 기본트리가 필수적으로 요구된다. XTAG 프로젝트에서는 광범위한 적용을 갖는 영어문법을 한국어에 확장하는 연구를 진행하고 있는데, 이는 한국어의 기능어에 대한 통사적 특징을 잘 표현하지 못한다는 문제가 있다[27]. 또한, 각 어휘에 대한 기본트리 중의성은 구조중의성과 함께 과생성 문제를 악화시킬 수 있다[15].

이러한 점을 반영하여 본 논문에서는 말뭉치에서 쉽게 학습할 수 있고, 효과적인 구문 중의성 해결을 위해 문장성분의 다양한 자질을 활용하는 확률적 자질기반 한국어 문법을 제안한다.

### 3. 확률적 자질기반 한국어 문법

확률적 구문분석 모델은 식 (1)과 같이, 품사 부착된 문장  $w_1/p_1 w_2/p_2 \dots w_n/p_n$ 이 입력되면, 문법에 따라 구문구조를 생성하고, 구문구조에 포함된 모든 규칙의 생성확률을 모두 곱해서 구문구조의 생성확률을 추정하고, 추정된 구문구조의 생성확률중 최고의 확률을 갖는 구문구조를 문장의 구문구조로 선택한다.

$$Tree_{best} = \underset{Tree}{argmax} P(Tree) = \underset{Tree}{argmax} \prod_{rule_i \in Tree} P(rule_i) \quad (1)$$

$$= \underset{Tree}{argmax} \prod_{i=1}^n P(N^p \rightarrow w_i/p_i) \times \prod_{i=n+1}^m P(N^p \rightarrow N^l N^r) \quad (2)$$

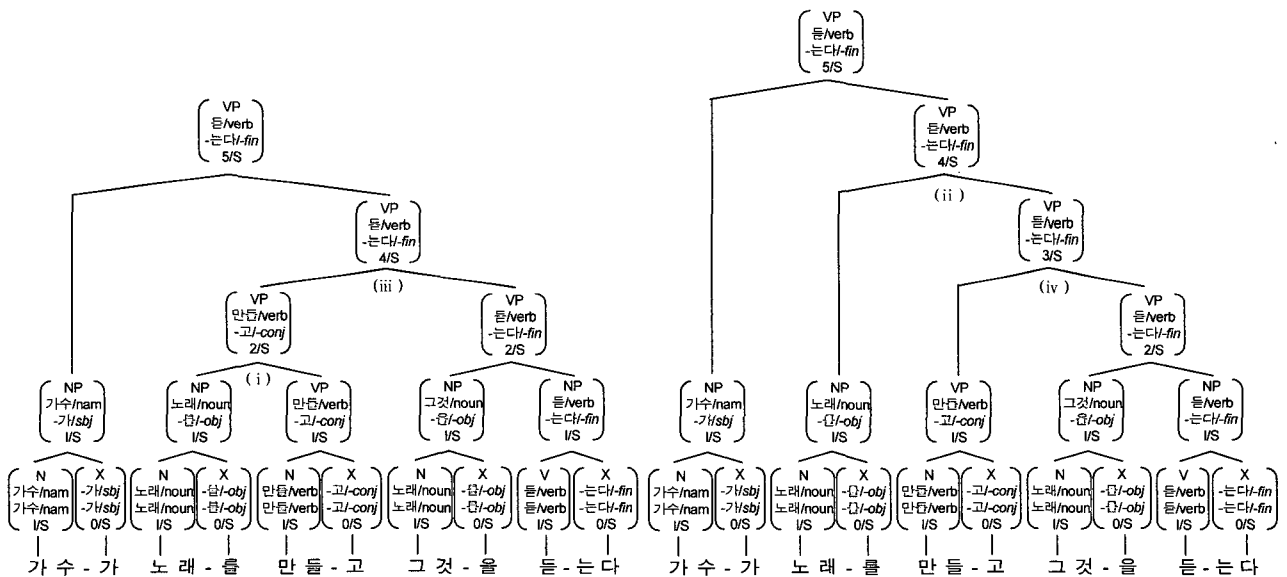
식 (1)에서  $Tree$ 는 문장의 구문구조를 나타내고,  $rule_i$ 는 구문구조에서  $i$ 번째 문법규칙을 나타내며,  $n$ 은 단일형식의 규칙수를 나타내고,  $m$ 은 이진형식의 규칙수를 나타낸다. 그리고, 문법규칙에서  $N^p$ 는 부모 비단말노드를 나타내고,  $N^l$ 는 왼쪽 자식 비단말노드를 나타내며,  $N^r$ 는 오른쪽 자식 비단말노드를 나타내고,  $w_i/p_i$ 는  $i$ 번째 단어와 품사 태그를 나타낸다.

제안하는 확률적 자질기반 한국어 문법 모델은 규칙을 단일형식과 이진형식으로 제한하므로 식 (2)와 같이 유도될 수 있다. 여러 개의 자식 비단말노드를 포함하는 문법규칙은 여러개의 이진형식의 문법규칙으로 변환가능하므로, 문법의

적용율을 줄이지 않고 문법크기를 줄일 수 있다는 장점이 있다[25]. 게다가, 이러한 이진형식의 제약은 비중심어 어순은 비교적 자유롭고 일부 문장성분은 생략되는 한국어의 구문특징을 잘 처리할 수 있다.

한국어의 구문구조 중의성을 해결에 필요한 다양한 정보를 활용할 수 있도록, 문장성분  $N$ 은  $[l, c, f, s]$ 와 같이 구문 자질  $l$ , 내용 자질  $c$ , 기능 자질  $f$ , 크기 자질  $s$ 로 구성된다. 이러한 문장성분은 좀더 구체적으로 나타내면  $[l, w_c/p_c, w_f/p_f, a_s/b_s]$ 와 같다. 구문 자질  $l$ 은  $NP$ 나  $VP$ 와 같은 구문 레이블이 나타낸다. 내용 자질  $c$ 는 문장성분의 내용 중심어를 표현하며, 이 때 내용자질  $w_c/p_c$ 는 의미 중심어의 어휘와 품사를 나타낸다. 기능 자질  $f$ 는 문장성분의 기능 중심어를 표현하며, 이 때 기능자질  $w_f/p_f$ 는 기능중심어의 어휘와 품사를 나타낸다. 크기 자질  $s$ 는 문장성분의 크기를 표현하며,  $a_s/b_s$ 는 문장성분이 포함하는 어절수와 그에 대한 크기 태그  $S$ (small),  $M$ (medium),  $L$ (large)로 구성된다.

예를 들어, (그림 2)에서 문장성분 “가수가”는  $[NP, 가수/noun, -가/sbj, 1/S]$ 와 같이 표현될 수 있다. 이는 명사 “가수”가 주격조사 “-가”에 의해서 문장에서 주어 기능을 수행하며, 문장성분이 1개의 어절로 간단히 구성되어 있다는 것을 나타낸다. 이와 같이, 일반적으로 어절에서 내용중심어는 어절의 마지막 실질 형태소이고, 기능중심어는 마지막 형식형태소가 된다. (그림 1)에서 문장성분 “그것을 듣는다”와 문장성분 “노래를 만들고 그것을 듣는다”는 모두  $VP$ 로 동일하게 표시되므로 구분할 수 없지만, (그림 2)에서는  $[VP, 듣/verb, 는다/fin, 2/S]$ 와  $[VP, 듣/verb, 는다/fin, 4/M]$ 과 같이 표시되므로, 크기자질  $a_s/b_s$ 을 바탕으로 구분된다.



(그림 2) 자질기반 한국어 문법에서 생성한 구문구조 후보

이처럼 제안하는 확률적 자질기반 구문분석 모델은 (그림 2)의 구문구조에서 서로 다른 부분을 나타내는 이진구적 (i) [VP, 만들/verb, 고/conj, 2/S] → [NP, 노래/noun, 를/obj, 1/S] [VP, 만들/verb, 고/conj, 2/S], (ii) [VP, 듣/verb, 는다/fin, 4/S] → [VP, 노래/noun, 를/obj, 1/S] [VP, 듣/verb, 는다/fin, 3/S] (iii) [VP, 듣/verb, 는다/fin, 4/S] → [VP, 만들/verb, 고/conj, 2/S] [VP, 듣/verb, 는다/fin, 2/S], (iv) [VP, 듣/verb, 는다/fin, 3/S] → [VP, 듣/verb, 는다/fin, 2/S]를 구분할 수 있다.

게다가, 제안하는 문장성분의 표현 방법은 단순히 구문 레이 블만 표현하는 방법에 비해 구문 중의성 해결에 필요한 더 많은 자질정보를 제공하고 있다. 그럼에도 불구하고, 문장성분의

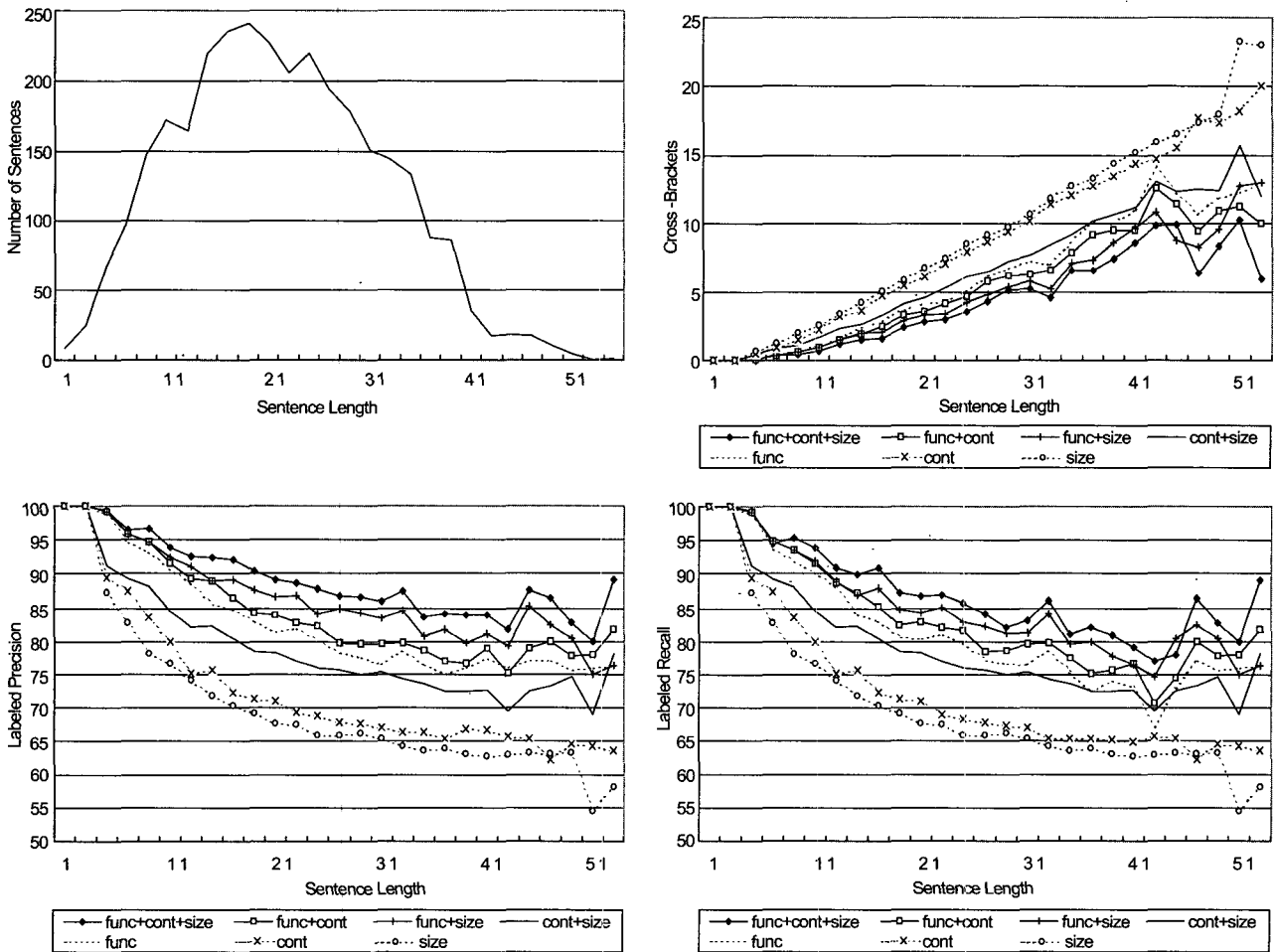
자질정보는 문맥자유문법을 바탕으로 한 구문구조만 있으면 부가적인 언어 자원 없이 자질기반 구문구조를 생성할 수 있다.

#### 4. 실험 및 평가

각 자질조합의 구문중의성 해결능력을 평가하기 위해서, <표 1>과 (그림 3)에 보여지는 바와 같이 일부 선택된 자질을 바탕으로 문장성분을 구성하여 실험하였다. 이때, 구문자질은 기본적으로 모두 포함된다. (그림 3)에서 *func*는 기능자질을 나타내고, *cont*는 내용자질을 나타내고, *size*는 크기자질을 나타낸다. 그리고 *cross brackets*, 정확률, 재현율을 기준으로 평가하였다[17]. 평가대상 말뭉치는 한국어의 다양한 언어현

<표 1> 자질조합별 구문분석 성능

	func+cont+size	func+cont	func+size	cont+size	func	cont	size
CBs	2.79	3.59	3.29	4.55	4.11	6.10	6.78
정확률	89.06%	84.00%	86.59%	78.32%	81.52%	70.95%	67.73%
재현율	86.72%	82.89%	84.31%	78.32%	80.45%	70.95%	67.73%



(그림 3) 자질조합별 구문분석의 성능

상을 포함하고 있는 31,080문장의 KAIST 트리뱅크[9]이며, 어절단위의 이진형식으로 변환된 말뭉치이다. 이 말뭉치를 90%의 학습집합과 10%의 실험집합으로 구분하여 실험하였다.

<표 1>은 전체 실험집합에 대한 자질조합별 구문분석 성능을 나타낸다. 그리고, (그림 3)은 문장 길이에 따른 구문분석 성능을 보여주기 위해 문장당 형태소 수에 따라 실험집합의 문장을 정렬하였다. <표 1>과 (그림 3)의 실험결과는 내용자질, 기능자질, 크기자질의 자질조합의 성능이 가장 높다는 것을 보여준다. 단일 자질조합에서는 기능자질의 성능이 가장 우수하고, 크기자질의 성능이 가장 떨어진다. 특이한 점은 단일자질인 기능자질이 내용자질과 크기자질의 조합보다 성능이 더 우수하다는 것이다. 한편, 기능자질과 크기자질의 자질조합성능이 기능자질과 내용자질의 자질조합 성능보다 우수했다. 이는 어휘와 품사로 구성된 기능자질과 내용자질이 서로 비슷한 특성을 가지므로, 자질의 효과가 서로 겹쳐서 성능 향상에 큰 영향을 끼치지 못한 것으로 보여진다. 반면에, 어휘와 품사로 구성된 기능자질과 숫자와 크기태그로 구성된 크기자질이 서로 다른 특징을 가지므로, 자질의 효과가 서로 완해 주어서 성능향상을 증가시킨 것으로 보여진다. (그림 3)의 좌측 상단 그림에서 보는 바와 같이 실험 집합에서 40 형태소 이상으로 구성된 문장의 수가 많지 않으므로, 긴 문장에 대한 구문분석 성능은 소수 문장의 특징에 크게 영향을 받아서 다소 고르지 않게 나타난다.

## 5. 결 론

본 논문에서는 효과적인 구문분석을 위해 문장성분을 구문자질, 내용자질, 기능자질, 크기자질로 표현한 확률적 자질기반 한국어 문법모델을 제안하였다. 그리고, 제안하는 구문분석 모델은 한국어의 부분자유어순과 생략현상을 잘 처리할 수 있도록 문법규칙을 이진형식으로 제한하였다. 실험에서 각 자질조합별로 구문 중의성 해결 능력을 분석하였는데, 서로 다른 특징을 갖는 자질의 조합이 서로 유사한 특징을 갖는 자질의 조합보다 구문 중의성 해결에 더 많은 도움을 줄 수 있다는 것을 알 수 있었다. 그러나, 제안하는 방법은 어휘정보를 지나치게 많이 요구하는 경향이 있어서, 효율성이 떨어지고 자료부족문제가 심각하게 나타날 수 있다는 문제점이 있다. 따라서, 앞으로 구문분석의 효율성을 개선시키고, 자료부족문제를 완화하기 위한 연구도 함께 진행되어야 한다.

## 참 고 문 헌

- [1] 강호관, 이종혁, 이근배, “새로운 어절 해석에 기반한 한국어 의존관계 파서”, 제9회 한글 및 한국어 정보처리 학술발표논문집, pp.327-331, 1997.
- [2] 김학수, 서정연, “어휘 의존 정보에 기반한 한국어 통계적 구문 분석기”, 한국정보과학회 인공지능 연구회 봄 학술발표논문집, pp.61-65, 1997.
- [3] 나동렬, 한국어 파싱에 대한 고찰 정보과학지, 제12권 제8호, pp.33-46, 1994.
- [4] 박성숙, 한성국, “자질 집합 표현에 의한 자연언어 문법 규칙 기술”, 제6회 한글 및 한국어 정보처리 학술발표논문집, pp.419-425, 1994.
- [5] 이공주, 김길창, “TAG을 기반으로 한 한국어 구문분석기에서 트리 변형 규칙”, 제1회 지능기술 공동학술회의, pp.100-105, 1995.
- [6] 이공주, 김재훈, 김길창. “제한된 형태의 구구조 문법에 기반한 한국어 구문분석”, 정보과학회논문지(B), 제25권 제4호, pp.722-732, 1998.
- [7] 정윤철, 문법 제약규칙을 이용한 중심어 주도의 한국어 구문분석기, 서강대학교 석사학위논문, 2002.
- [8] 조형준, 박종철, 한국어 병렬문의 통사, 의미, 문맥 분석을 위한 결합범주문법 한국정보과학회 논문지, 제27호 제4호, pp.448-462. 2000.
- [9] 최기선, KAIST 언어 자원 2001년도판, 과학기술부 핵심 소프트웨어 과제 결과물, 2001, <http://kibs.kaist.ac.kr>.
- [10] Black, Ezra, Fred Jelinek, John Lafferty, David M. Magerman, Robert Mercer and Salim Roukos. “Towards History-based Grammars : Using Richer Models for Probabilistic Parsing,” In Proceedings of the Annual Meeting of the Association for Computational Linguistics, pp.31-37, 1993.
- [11] Cancedda, Nicola and Christer Samuelsson, Experiments with Corpus-based LFG Specialization. In proceedings of the Sixth Applied Natural Language Processing Conference. pp.204-209, 2000.
- [12] Cha, Jeongwon, Geunbae Lee and Jong-Hyeok Lee, “Korean Combinatory Categorical Grammar and Statistical Parsing,” Computers and the Humanities, Vol.36, No.4, pp.431-453, 2002.
- [13] Charniak, Eugene, Statistical Techniques for Natural Language Parsing AI magazine, Vol.18, No.4, pp.33-43, 1997.
- [14] Charniak, Eugene, “Immediate-Head Parsing for Language Models,” In Proceedings of the Annual Meeting of the Association for Computational Linguistics, pp.116-123. 2001.
- [15] Chung, Euisok, Soojong Lim, JiHyun Wang, Myung-Eun Lim and Bo-Hyun Yun, Korean Syntactic Analyzer with Tree-based Grammar, In Proceedings of the International Conference on Computer Processing of Oriental Languages, pp.132-137, 2001.
- [16] Collins, Michael, Head-Driven Statistical Models for natural Language Parsing, Ph. K. Thesis, University of Pennsylvania, 1999.
- [17] Goodman, Joshua. “Parsing Algorithms and Metrics,” In

[1] 강호관, 이종혁, 이근배, “새로운 어절 해석에 기반한 한국어 의존관계 파서”, 제9회 한글 및 한국어 정보처리 학술발표논문

Proceedings of the Annual Meeting of the Association for Computational Linguistics, pp.177-183. 1996

- [18] Goodman, Joshua. "Probabilistic Feature Grammars," In Proceedings of the International Workshop on Parsing Technologies, pp.89-100. 1997.
- [19] Klein, Dan and Christopher D. Manning. "Accurate Unlexicalized Parsing," In Proceedings of the Annual Meeting of the Association for Computational Linguistics, pp.423-430. 2003.
- [20] Kwak, Yong-Jae, Young-Sook Hwang, Hoo-Jung Chung, So-Young Park, Sang-Zoo Lee and Hae-Chang Rim, GLR Parser with Conditional Action Model(CAM), In Proceedings of the 6th Natural Language Processing Pacific Rim Symposium, pp.359-366, 2001.
- [21] Lee, Seungmi and Key-Sun Choi, Reestimation and Best-First Parsing Algorithm for Probabilistic Dependency Grammars, In Proceedings of the 5th Workshop on Very Large Corpora, pp.41-55, 1997.
- [22] Lee, Wonll, Geunbae Lee and Jong-Hyeok Lee, Chart-Driven Connectionist Cateforial Parsing of Spoken Korean Computer Processing of Oriental Language, Vol.10, No.2, pp. 147-159, 1996.
- [23] Magerman, David M., Statistical Decision-Tree Model for Parsing, In Proceedings of the Annual Meeting of the Association for Computational Linguistics, pp.176-183, 1995.
- [24] Malouf, Robert, John Carroll, Ann Copestake. Efficient Feature Structure operations without compilation. Natural Language Engineering, Vol.6, No.1, pp.29-46, March, 2000.
- [25] Manning, Chirstopher D. and Hinrich Schutze, Foundations of Statistical natural Language Processing, MIT Press, 1999.
- [26] Pollard, Carl and Ivan A. Sag, Head-Driven Phrase Structure Grammar, CSLI and University of Chicago Press, Stanford, Ca. and Chicago, III, 1994.
- [27] Yoon, Juntae, Chung-hye Han, Nari Kim and Mee-sook Kim,

Customizing the XTAG system for efficient grammar development for Korean Proceedings of the International Workshop on Tree Adjoining Grammars and Related Formalisms, 2000.



### 박 소 영

e-mail : ssoya@nlp.korea.ac.kr

1997년 상명대학교 전자계산학과 학사

1999년 고려대학교 컴퓨터학과 석사

1999년~현재 고려대학교 컴퓨터학과 박사과정

관심분야 : 자연어처리, 기계번역, 기계학습



### 김 수 홍

e-mail : soohkim@smu.ac.kr

1992년~현재 상명대학교 컴퓨터정보통신공학부 교수

1998년~현재 한국정보처리학회 학회지편집위원

2002년~현재 한국정보처리학회 e-Business 연구회 위원장

관심분야 : 정보검색, e-learning, 병렬처리시스템, EC/ERP



### 임 해 창

e-mail : rim@nlp.korea.ac.kr

1993년~현재 인지 과학회 이사

1994년~1998년 한국정보과학회 편집위원

1998년~2000년 한국정보과학회 한국어정보처리연구회 운영위원장

1999년~2000년 고려대학교 컴퓨터과학기술연구소 연구소장

1991년~현재 고려대학교 컴퓨터학과 교수

관심분야 : 자연어처리, 구문분석, 정보검색, 기계학습