

# 엔트로피 지도 CRF를 이용한 한국어 어절 구문태그 예측 (Eojeol Syntactic Tag Prediction of Korean Text using Entropy Guided CRF)

오진영<sup>\*</sup> 차정원<sup>\*\*</sup>  
(Jin-Young Oh) (Jeong-Won Cha)

**요약** 본 연구에서는 의사결정나무와 Conditional Random Fields(CRFs)를 이용하여 한국어 어절 구문태그를 예측하는 시스템에 대해서 설명한다. 기계학습에서 자질의 선택은 작성자의 직관에 의해서 주로 이루어지는데 이는 작성자의 지식에 의존한다. 본 연구에서는 의사결정나무를 사용하여 보다 체계적으로 조합이 이루어지도록 하였다. 또한 오류 분석을 통하여 최적의 자질이 무엇인지를 파악하여 최고의 성능을 보이도록 하였다. 실험을 통하여 본 논문에서 제안한 방법이 성능향상에 도움이 된다는 것을 확인할 수 있어 앞으로 구문 분석에 많은 도움이 될 것이라고 확신한다.

**키워드** : 어절 구문태그, 의사결정나무, CRFs

**Abstract** In this work, we describe the syntactic tag prediction system for Korean using the decision tree and CRFs. Generally they select features by their intuition. It depends on their prior knowledge. In this works, we combine features systematically using the decision tree. We also analyze errors and optimize features for the best performance. From the result of experiments, we can see that the proposed method is effective for the syntactic tag estimation and will be helpful for the syntactic analysis.

**Key words** : Eojeol Syntactic Tag, Decision Tree, CRFs

\* 이 논문은 제34회 추계학술대회에서 '엔트로피 지도 CRF를 이용한 어절 구문태그 예측'의 제목으로 발표된 논문을 확장한 것임

† 학생회원 : 창원대학교 컴퓨터공학과  
shanghwoe963@changwon.ac.kr

\*\* 종신회원 : 창원대학교 컴퓨터공학과 교수  
jcha@changwon.ac.kr

논문접수 : 2008년 12월 19일

심사완료 : 2009년 3월 1일

Copyright©2009 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 컴퓨팅의 실제 및 레터 제15권 제5호(2009.5)

## 1. 서론

구문분석은 문장을 분석하여 문장을 구성하는 요소들 간의 관계를 정의하는 작업이라 할 수 있다. 하지만, 문장이 길어질수록 문장을 구성하는 요소들의 수도 많아지므로 관계를 정의하기란 어려워진다. 또한 교차어인 한국어는 부분 어순 자유 언어에 속하므로 구문분석에 대한 처리효율이 떨어진다. 이에 Abeny는 구(phrase) 묶음을 함으로써 문장의 원래 단어 수가 감소하여 복잡성이 줄어들 것이고, 대표적인 구의 특징이 하나로 표현된다면 구문분석 과정의 부담감을 줄일 수 있을 것이다[1].

한국어 구문 분석에서 분석 단위를 어절로 할 것인지 형태소로 할 것인지에 대한 논의는 많이 있었다. 차정원은 [2]에서 의미어와 기능어의 구별을 통하여 통사 기능을 명확히 할 수 있어 형태소 단위가 타당하다고 하였지만 다른 많은 연구에서는 형태소 단위는 후보가 너무 많이 생성되어 현실적으로 구현이 어렵고 문장 성분이 어절 단위로 결정되므로 어절 단위가 타당하다고 하였다.

본 연구에서는 형태소 분석 결과로부터 구문분석의 최소 단위인 어절 구문태그를 예측한다. 어절 태그는 문장에서 주어, 목적어, 부사어 등 문장 성분을 나타낸다. 따라서 어절 태그가 결정되면 그 어절이 문장에서 어떤 역할을 수행하는지가 결정되기 때문에 중요하다. 실제 세종계획 21[3]의 구문 코퍼스에서는 어절 구문태그를 정의하고 이를 바탕으로 구문분석을 하고 있다. 본 연구에서는 세종계획 21의 구문 코퍼스를 이용한다.

본 연구에서는 의사결정나무(Decision Tree)를 통하여 자질집합을 자동 조합 및 생성하는 것으로, 사전지식 없이도 의사결정나무를 이용하여 최적의 자질집합을 작성 가능하도록 하는 것이다. 이것을 CRFs로 학습하여 구문태그를 예측한다. 이 결과로 기계학습의 성능에 영향을 미치는 것은 학습 알고리즘뿐만 아니라 어떤 속성을 선택하여 학습할 것인가 중요하다는 것을 보이고자 한다.

또한 자질집합을 자동으로 생성한 것에 대한 성능 영향을 분석하고 결과를 제시하며, 구문분석의 오류에 영향을 주는 문장의 정확도, n-best 어절 태그들에 대한 성능평가 한다.

본 논문의 구성은 다음과 같다. 2장은 관련 연구에 대해서 설명하고, 3장에서는 본 논문에서 제안하는 기계학습을 통한 한국어 어절 구문태그에 대한 설명, 4장에서는 실험 및 분석결과, 마지막으로 5장에서 결론을 맺는다.

## 2. 관련연구

대부분 구문분석 이전에 기본적인 처리단계로 구 인식을 하므로 많은 연구가 이루어져 왔다. 먼저 횡영속은 한국어 문장을 구성하는 기본구를 인식하는 연구를 하

였다. 기계학습 알고리즘에 최적인 자질집합선택을 위해 '점증적 유용성'이란 관점에서 자질의 적합성을 정의, 선택하였다. 기계학습을 위해서 알고리즘이 다른 결정트리 학습방법과 메모리기반 학습 방법을 사용하여 각각 93.40%, 91.75%의 성능을 나타내고 있다[4].

박성배는 규칙기반방법이 구 단위화에 유용한 방법이지만, 예외적인 처리 방법이 없다는 단점을 보안하기 위해 규칙기반방법과 기억기반 학습을 결합하는 방법을 제시하였다. 규칙으로 처리한 단위화의 결과가 임계치를 기준으로 예외를 처리하여 기억기반 학습으로 검증하는 것으로 F-Score가 94.19%를 보여주고 있다[5]. 또한 k-NN으로 확장된 한국어 단위화에 대한 제안을 하였다. [5]와 비슷하며, 규칙기반 단위화에 기초하여 결정이 모호한 경우 k-NN학습방법으로 관련 정보에서 추출한 정보를 이용하여 단위화를 처리하는 것으로 규칙기반만을 사용했을 때보다 제안 시스템은 정확도가 2.3% 증가하였음을 보여준다[6].

김미영은 단위(chunk)분석과 의존문법에 기반을 둔 한국어 구문분석 시스템을 제안하였는데 하나의 노드로 간주될 수 있는 명사구, 동사구를 하나의 단일화된 차트로 묶어준다면 의존문법 등과 같은 단계에서 빠르고 간단한 분석을 할 수 있다는 전제로, 명사구 단위화를 위해서는 6개 규칙을 사용하며, 동사구 단위화는 오토마타를 사용한다. 명사구는 95.68%, 동사구는 84.19%의 성능을 보이고 있다[7].

박의규는 문장의 분석 복잡도를 낮추기 위해 의존명사에 대한 구 묶음 처리를 하였다. 제안된 시스템이 단순한 구묶음 기법을 사용하였을 때보다 에러를 43.57% 감소시킴으로써 의존명사 처리는 구문분석 시스템 성능 향상에 영향을 미치는 것을 확인하였다[8].

신효필은 서술어가 문말에 위치하는 언어의 특성을 바탕으로 한 문장 안에서 다양한 동사의 활용으로 형성되는 파생구조 및 내포된 동사구에 의해 형성되는 지역적 동사구내에서 그 성분들의 논리적인 분할을 구성하고 다시 결합하는 형태로 언어의 의존적인 발견적 규칙(휴리스틱스)들을 점수화하여 적용하였다. 제안된 시스템은 빠른 구문분석이 요구되는 시스템을 위해 고안되었고, 최소의 자원으로도 최대효과를 낼 수 있음을 보여

준다[9].

[10]에서 Bangalore는 TAG(Tree-Adjoining Grammar)에서 단어(word)에 붙은 기본 구문 정보를 결정하는 시스템을 제안하였다. 이것은 기본 구문 정보가 결정되면 문장 전체 구문 분석을 효율적으로 할 수 있기 때문이다.

[11]에서 Ruy Luiz Milidiú는 Transformation based learning에서 의사결정나무를 사용하여 자질 조합을 효과적으로 구성하는 방법을 제시하였다.

본 연구에서는 사전 지식이 필요한 자질 결정을 의사결정나무를 사용하여 자동으로 결정하고 이를 CRFs로 학습하여 구문태그를 결정한다. 따라서 그 언어에 상당한 지식이 필요한 자질 조합을 최적화하여 성능 향상을 이를 수 있다.

### 3. 제안 시스템

본 연구는 어절 구문태그를 결정하기 위해서 이 문제를 분류문제로 취급한다. 따라서 형태소 분석 결과가 주어졌을 때, 이를 바탕으로 구문태그를 생성한다.

#### 3.1 코퍼스 생성

세종계획 21 구문 코퍼스를 변형하여 어절 태그를 예측하는 코퍼스를 생성한다.

그림 1은 어절 구문태그 학습파일에서 한 어절마다 3개의 자질을 가지도록 한 예이며, 이후 4장에서는 다른 형태, 다른 크기 자질을 가진 코퍼스 결과를 확인할 수 있다.

어절 단위로 분리된 문장은 각각의 어절이 하나 이상의 형태소로 구성되어 있다. 그러므로 구문 태그가 어떤 한 형태소의 영향을 많이 받을 것인가에 따라 자질을 추가한다.

그림 1에서는 한 어절에 대하여 3개의 품사 자질을 가지고 있는데, 첫 번째 형태소(자질1, H\_POS), 마지막 형태소(자질3, T\_POS), 마지막 형태소의 앞 형태소의 품사(자질2, t\_POS)로 구성된다. 어절을 이루는 형태소 중 어절 구문 태그에 영향을 주지 않을 형태소(기호 등)는 자질에서 삭제하였다. 4번째는 어절 구문 태그이다. 그림에서 '-'는 해당 품사가 없을 경우에 표시한다.

형태소 분석 결과	자질1	자질2	자질3	구문태그
실로/MAG	MAG	-	-	AP
어쳐구니/NNG+ 없/VA+ 는/ETM	NNG	VA	ETM	VP_MOD
일/NNG+ 이/VCP+ 지만/EC	NNG	VCP	EC	VNP
그것/NP+ 이/JKS	NP	-	JKS	NP_SBJ
현실/NNG+ 이/VCP+ 다/EF+ ./SF	NN	VCP	EF	VNP

그림 1 코퍼스 생성 예

### 3.2 자질집합 조합

표 1은 본 논문에서 사용한 기본 자질들을 나열한 것이다.

표 1 사용된 기본 자질들

자질이름	자질내용
H_POS±2	어절의 첫 형태소의 품사 앞, 뒤 2개씩
T_POS±2	어절의 끝 형태소의 품사 앞, 뒤 2개씩
t_POS±2	어절의 끝에서 두 번째 형태소의 품사 앞, 뒤 2개씩
H_POS	자신의 첫 형태소의 품사
T_POS	자신의 끝 형태소의 품사
t_POS	자신의 끝에서 두 번째 형태소의 품사

기본 자질들을 다양한 조합을 통해서 새로운 자질로 만들 수 있다. 이 경우, 대부분의 연구에서는 연구자의 사전지식이 중요한 요소가 된다. 본 연구에서는 의사결정나무를 사용하여 새로운 자질 조합을 생성하는 방법을 설명한다.

그림 2는 어절 구문태그 예측을 위해 만들어진 의사결정나무의 한 부분을 보여준다. 의사결정나무에서 뿌리(root)에서 가까울수록 정보량이 많은 자질이 된다. 이 의사결정나무에서 깊이우선탐색(depth-first search)으로 조합된 자질을 생성한다. 표 2는 조합된 자질들을 보여준다. 표에서 위의 4개(1,2,3,4)는 조합된 자질들을 보여주고 아래의 3개(5,6,7)는 이 자질들에서 최상위 자질을 제거한 자질들을 보여준다.

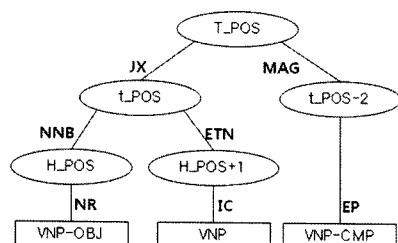


그림 2 의사결정나무. JX는 보조사, MAG는 부사, NNB는 의존명사, ETN은 명사전성 어미, NR은 수사, IC는 독립어, EP는 선어말 어미를 나타낸다.

표 2 조합된 자질들

번호	조합된 자질
1	T.POS
2	T.POS t.POS-2
3	T.POS t.POS H.POS
4	T.POS t.POS H.POS+1
5	t.POS-2
6	t.POS H.POS
7	t.POS H.POS+1

### 3.3 CRFs 학습, 테스트 및 평가

CRFs는 입력 데이터 열과, 클래스로 분류된 열의 조건부 확률을 계산하여 기계학습 방법이다[12].

입력 데이터  $\mathbf{x}$ , 입력된 열에 대응되는  $\mathbf{y}$ 열을 각각  $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$ ,  $\mathbf{y} = \langle y_1, y_2, \dots, y_n \rangle$  라 하고  $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 가 주어졌을 때 식 (1)과 같이 정의가 가능하다.

$$p(\mathbf{y} | \mathbf{x}, \Lambda) = \frac{1}{Z(X)} \exp \left( \sum_{i=1}^n \sum_j \lambda_j f_j(y_{i-1}, y_i, x, i) \right) \quad (1)$$

식 (1)에서  $f_j$ 는  $j$  번째 자질함수이고,  $\lambda_j$ 는  $j$  번째 자질함수의 학습된 가중치이다.  $\lambda$ 값이 클수록 그에 따른 많은 자질 가중치를 갖게 되는 것이다.

본 연구에서는 그림 1과 같은 코퍼스의 입력에 대해 의사결정나무로 조합, 생성된 자질집합을 바탕으로 기계학습한다. 또한 학습 데이터의 과적합(overfitting) 문제를 해결하기 위해서 가우스 사전 평활(Gaussian Prior Smoothing)[13]을 적용한다.

## 4. 실험 및 분석

문장 성분을 결정한 이전 연구결과가 없어, 본 논문에서 제안한 의사결정나무를 이용한 자질 자동 조합 방법의 타당성을 보이기 위해서 다양한 실험을 하였다.

### 4.1 실험환경

실험에서 사용된 세종계획 21 구문 코퍼스는 8개의 구문표지와, 7개 기능표지, 10개의 기타표지가 있다.

학습문서는 90112어절(9899문장)이며, 평가 문서는 학습 문서와 별개의 20685어절, 1430문장이다.

### 4.2 의사결정나무를 통한 자질조합 생성

의사결정나무를 통하여 자질조합을 생성함에 있어 여러 가지 경우의 실험을 통해 타당성을 확인해 보았다.

표 1의 기본 자질을 사용하여 기본 실험을 수행했다. 기본 자질을 그대로 사용했을 때와 이를 실험자의 사전지식을 이용해 임의로 조합한 후의 결과를 표 3에서 볼 수 있다.

표 3 임의 작성된 자질 집합 결과(자질 3개)

문맥	기본자질	조합자질	
		-2~2	-1~1
어절	92.60	92.75	<b>93.01</b>
문장	44.61	41.54	<b>46.50</b>

표 3에서 보듯이 기본 자질의 나열보다는 자질들의 조합이 더 좋은 성능을 보이고 있다. 또한 문맥은 앞뒤 2개씩을 보는 것이 좋은 성능을 보인다.

이제는 표 1의 기본 자질을 의사결정나무를 이용하여 조합하였다. 표 4에서 보는 바와 같이 사전 지식을 가진

사람이 작성한 자질보다 성능이 떨어진다는 것이 알 수 있다.

오류를 분석해보면 어절의 가운데 형태소인 궁정지정사 '이(VCP)'의 존재여부에 따라서 어절 구문태그가 결정되는 예가 많았다.

표 4 의사결정나무에 의해 조합된 자질(자질 3개)

문맥	조합자질	
	-2~2	
깊이	3	4(최상위 노드 제외)
어절	92.73	92.73
문장	45.94	45.59

조용사 'XSV<sup>1)</sup>', 'XSA<sup>2)</sup>'의 존재여부도 많은 영향을 주었다. 또한 뒤에 오는 어절의 처음 형태소의 품사에 따라서 영향을 미치는 경우가 많이 발생하였다.

이런 분석의 결과로 본 연구에서는 표 5와 같이 새로운 자질을 추가하였다.

표 5 추가된 자질들

자질이름	자질내용	
h_POS	자신의 어절에서 XSA, XSV, VCP가 있으면 표시	
n_POS	다음 어절의 첫 형태소의 품사	

이렇게 변형된 자질을 가지는 코퍼스의 예를 그림 3에서 보였다. 그림 3에서 자질 2, 3번이 추가된 자질이다.

변형된 학습 문서로 앞에서 수행했던 평가를 다시 하였다.

표 6 임의 작성된 자질 집합 결과(자질 5개)

문맥	기본자질	조합자질	
	없음	-1~1	-2~2
어절	91.71	92.92	93.31
문장	39.44	45.66	46.85

표 7 의사결정나무에 의해 조합된 자질(자질 5개)

문맥	조합자질	
	-2~2	
깊이	3	4(최상위 노드 제외)
어절	93.37	93.19
문장	48.11	48.53

형태소 분석 결과	자질1	자질2	자질3	자질4	자질5	구문 태그
실로/MAG	MAG	-	-	-	NNG	AP
어처구니/NNG+ 얹/VA+ 는/ETM	NNG	-	VA	ETM	NNG	VP_MOD
일/NNG+ 이/VCP+ 지만/EC	NNG	-	VCP	EC	NP	VNP
그것/NP+ 이/JKS	NP	-	-	JKS	NNG	NP_SBJ
현실/NNG+ 이/VCP+ 다/EF+ ./SF	NNG	-	VCP	EF	-	VNP

그림 3 코퍼스 생성 예

실험을 분석해보면 조사나 어미가 있는 것이 그렇지 않은 어절에 비해서 성능이 높았다. 즉 조사나 어미가 없는 어절이 오류가 많이 발생하였는데, 그 빈도는 오류 어절의 절반이상을 차지하였다.

의사결정나무를 이용해 자질들을 분석해보니 5번짜리 어절이 영향을 많이 주었다. 그래서 5번짜리 어절을 변경하여 모든 품사를 추가하는 것이 아니라 동사류에 대해서만 추가하고 형태소 정보를 병기하였다. 또한 H\_POS ±2와 5번짜리 어절이 겹칠 경우에는 H\_POS±2를 제거하였다. 이렇게 변경된 자질을 사용하는 코퍼스로 실험한 결과가 표 8에 있다.

표 8 의사결정나무에 의해 조합된 자질

문맥	조합자질	
	-2~2	
깊이	3	4(최상위 노드 제외)
어절	91.26	93.90
문장	36.85	49.65

이 실험에서 조사나 어미에 대한 오류를 약 14% 감소 시켰지만, 여전히 용언수식어에 대한 오류가 가장 많다.

표 9는 오류가 가장 많이 발생하는 조사나 어미가 없는 어절을 '미지(未知, unknown)'로 처리하여 모두 맞다고 가정할 경우의 성능을 나타내고 있다.

표 9 조사, 어미가 없는 어절을 '미지'로 할 경우

문맥	조합자질	
	-2~2	
깊이	3	4(최상위 노드 제외)
어절	94.12	96.08
문장	56.19	68.79

이렇게 조사나 어미를 모두 맞다고 가정한 이유는 어절 구문태그를 결정하지 못할 경우, 구문분석에서 결정하도록 넘겨주어야 되기 때문이다.

본 논문에서 제안한 구문태그 예측 모델은 형태소 자질을 사용하지 않았기 때문에 다른 기계학습 모델에 비해 현저히 작은 크기로 위와 같은 성능을 보이고 있다. 실제로 형태소에 대한 자질을 추가하면 CRF학습을 위

한 코퍼스의 양도 작아지고, 모델의 사이즈도 커지는 것을 확인하였다. 그림 3의 코퍼스 환경에 모든 형태소를 부착하여 같은 과정으로 실험을 하였을 때 87%의 성능을 보였으며, 학습코퍼스에서 학습되지 못한 형태소에 대한 오류가 대부분이다.

### 4.3 n-Best 성능

위의 실험결과에 하나를 선택하여 n-best 성능을 구하였다. n-best는 CRFs의 어절태그 후보에서 확률이 높은 n개를 결과로 출력한 것으로, 이 결과 속에 정답이 있으면 올바른 것으로 간주한다. 이것은 구문분석의 입력으로 n-best를 주면 성능 향상을 꾀할 수 있기 때문이다. 즉, 1-best에서는 없는 정답이 n-best에서는 있을 가능성이 커지므로, 이후 단계인 구문분석 성능이 향상될 수 있다.

먼저 표 8에서 2번째 실험인 의사결정나무 사용, 문맥범위 -2~2, 깊이는 4이고 최상위 자질을 삭제한 실험결과에 대한 n-best결과는 표 10과 같다.

표 10 n-best[5개의 자질]

	1-best	2-best	3-best	4-best	5-best
어절성능	93.90	95.36	96.18	95.57	97.05
문장성능	49.65	61.12	67.41	70.77	74.34

그리고 표 9에서는 의사결정나무를 통해 얻어진 깊이가 4이고, 최상위 노드를 제외한 자질집합 실험 결과에 대한 n-best 결과는 표 11과 같다.

표 11 조사, 어미가 없는 어절을 ‘미지’ n-best

	1-best	2-best	3-best	4-best	5-best
어절성능	96.08	97.47	97.86	98.04	98.28
문장성능	68.79	74.20	77.90	79.93	82.10

각 표에서 나타난 것과 같이 n-best의 성능은 후보가 증가할수록 높아지는 것을 확인할 수 있다. 특히 문장 성능확률이 높아지는 것으로 보아 이것을 구문분석에 적용한다면 문장에 대하여 정확도 성능을 높일 수 있을 것이다.

## 5. 결 론

본 논문에서는 한국어 구 인식에 대한 유용한 자질을 생성하기 위해 의사결정나무를 이용하여 자동으로 자질집합을 조합하도록 하였다. 즉, 기계학습에 맞는 자질집합을 적극적으로 생성하는 것이 아니라, 실질적인 유용함을 보이고자 하였다.

실험결과, 구 인식에 대해서는 현재 어절을 구성하는 형태소 집합, 다음 어절의 첫 형태소와 음절 자질을 사

용하였고, 문맥의 범위는 좌우 문맥의 2개를 보았을 때 가장 성능이 좋았다. 결과로 각 93.90%, 49.65%의 어절과 문장 성능을 얻었다.

향후 연구로는 구문분석 방법에 이 방법을 적용하여 보다 향상된 구문분석 방법을 개발할 예정이다.

## 참 고 문 헌

- [1] Abney, S. and S. P. Abney. Parsing by Chunks. Principle-Based Parsing. R. C. Berwick, S. P. Abney and C. Tenny, Kluwer Academic Publishers: pp. 257-278, 1991.
- [2] 차정원, “한국어 결합범주문법을 위한 통계적 구문분석”, 포항공대 박사학위 논문, 2002.
- [3] 세종계획 21, <http://www.sejong.or.kr/>
- [4] 황영숙, 정후중, 박소영, 곽용재, 임해창, “자질집합선택 기반의 기계학습을 통한 한국어 기본구 인식의 성능향상”, 정보과학회논문지, 소프트웨어 및 응용 제 29권, 제9호, pp. 654-668, 2002.
- [5] 박성배, 장병탁, “한국어 구 단위화를 위한 규칙 기반 방법과 기억 기반 학습의 결합”, 정보과학회논문지, 소프트웨어 및 응용 제 31권, 제3호, pp. 369-378, 2004.
- [6] 박성배, 장병탁, 김영탁, “k-NN으로 확장된 한국어 단위화”, 한국정보과학회 가을 학술발표논문집, Vol.27, No.2, pp. 182-184, 2000.
- [7] 김미영, 강신재, 이종혁, “단위(chunks)분석과 의존문법에 기반한 한국어 구문분석”, 한국정보과학회 봄 학술발표논문집, Vol.27, No.1, pp. 327-329, 2000.
- [8] 박의규, 나동열, “한국어 구문분석을 위한 구묶음 기반 의존명사 처리”, 한국인지과학회 논문지 제 17권, 제2호, pp. 119-138, 2006.
- [9] 신효필, “최소자원 최대효과 구문분석”, 한국 정보 과학회 언어 공학연구회, 학술대회지(한글 및 한국어정보처리), pp. 242-248, 1999.
- [10] Bangalore, S. and A. K. Joshi. “Supertagging: An Approach to Almost Parsing,” Computational Linguistics 25: pp. 237-265, 1999.
- [11] Milidiú, R. L., C. N. d. Santos, et al.. Phrase Chunking Using Entropy Guided Transformation Learning. Proceedings of ACL-08: HLT. Columbus, Ohio, Association for Computational Linguistics: pp. 647-655, 2008.
- [12] J. Lafferty, A. McCallum, F. Pereira. “Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data,” Proceedings of International Conference on Machine Learning, ICML-01, pp. 282-289, 2001.
- [13] A. L. Berger, V. J. Della Pietra, S. A. Della Pietra, “A maximum entropy approach to natural language processing,” Computational Linguistics, Vol.22, No.1, pp. 39-71, 1996.