

한국어 어휘의미망에 기반한 논항 정보를 이용한 의존문법 구문분석기의 구현*

임경업 정영임 권혁철
부산대학교 컴퓨터공학과
{iku88, acorn, hckwon}@pusan.ac.kr

Implementation of Dependency Parser using Argument Information based on Korean WordNet

Gyeong-eop Im Youngim Jung Hyuk-Chul Kwon
Dept. of Computer Science & Engineering, Pusan National University

요 약

한국어는 한 어절이 한 개 이상의 형태소로 이루어졌으며, 이 때문에 지역 중의성이 발생한다. 대부분의 선행 연구에서는 이러한 지역 중의성을 배제하거나, 태거를 사용하여 지역 중의성을 제거해왔다. 본 연구에서는 문장의 모든 형태소 분석에 대해 구문분석을 시도하며, 중의성을 제거하고자 적용된 의존문법 규칙과 구 묶음, 부사 하위범주화, 논항 정보 사전 이용 등의 다양한 기법을 설명하고, 구문분석 성능을 실험으로 나타낸다. 특히, 말뭉치마다 논항 정보 사전을 따로 구축하는 번거로움을 피하고자 한국어 어휘의미망을 사용한다.

1. 서 론

한국어는 문장 요소들의 잦은 생략과 어순이 자유롭다. 이러한 특징 때문에, 한국어는 구구조문법(phrase structure grammar)보다 의존문법(dependency grammar)에 의한 분석이 더 적합하고, 따라서 한국어 구문분석에는 의존문법이 많이 이용되어 왔다[1][5][6].

한 어절이 거의 하나의 형태소로 이루어진 영어와 달리, 한국어는 하나의 어절이 보통 여러 개의 형태소로 이루어져 있으며, 이 때문에 지역 중의성(local ambiguity)이 많이 존재한다. 즉, 하나의 어절에 대한 형태소 분석 후보가 여러 개가 나올 수 있다. 대부분의 기존 한국어 구문분석 연구는 이러한 지역 중의성에 대해 언급 하지 않고 있다[1][4][5][11][12].

[6][7]은 품사 태거(tagger)를 통해 지역 중의성을 제거하고 있는데, 이 경우 태거의 한계가 그대로 구문분석기에 반영된다는 약점이 있다.

본 논문에서 제안하는 구문분석기는 지역 중의성을 인정하고, 그에 따른 모든 가능한 구문분석구조를 결과로 출력하며, 이를 위해 실제 자료와 언어적 특성을 연

구하여 만든 지배-의존 규칙을 적용한다. 다만, 이 과정에서 수많은 중의성이 발생하며, 이를 줄이고자 다양한 방법을 사용한다.

본 논문에서는 2절에서 기존의 구문분석 연구들을 살펴보고, 3절에서 전체적인 시스템과 구문분석 및 중의성을 줄이고자 사용한 방법을 소개한다. 4절에서는 중의성을 줄이는 방법에 대한 실험과 분석을 기술하며, 마지막으로 5절에서 결론 및 향후연구에 대해 논한다.

2. 관련 연구

영어 구문 분석은 대용량의 정제된 말뭉치를 바탕으로 한 확률적 구문 분석이 연구되어 상당한 수준에 이르렀다. 40단어 미만의 문장을 대상으로 E. Charniak의 immediate-head parsing model은 평균 90.1% LP(labeled precision)/LR(labeled recall)를 보이며[16], M.Collins의 dependency-based parsing model의 경우 88.6%의 LP/LR를 보이고 있다[18].

한국어에도 확률적 기법이 적용된 사례가 있다. 은지현[7]의 immediate-head 구문분석 모델이 78.87%의 LP와 79.55%의 LR 성능을 보이며, Hoojung CHUNG[17]의 경우 85.00%의 정확률과 84.51%의 재현율을 보인다.

*이 논문은 2007년도 정부(과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(No. R01-2007-000-20517-0)

하지만, 확률적 시스템은 정제된 말뭉치를 기반으로 구현되는데, 아직 국내에서는 기존에 구축된 말뭉치에 대한 검증이나 보완에 대한 연구가 미비한 실정이다[8].

이와는 달리, 언어적인 특징을 이용하여 구문 분석기의 성능을 향상시키는, 규칙 기반 연구도 진행 중이다. 그 중 하나가 구 묶음(chunking) 기법을 중심으로 한 구문 분석 연구이다. 황이규[14]는 본용언과 보조용언의 묶는 규칙을, 김창제[4]는 서술 어절과 기능 어절을 묶는 규칙을 사용했다. 박익규[6]는 의존명사를 구 묶음으로 처리하여, 시스템의 성능을 43.57% 향상시켰다. 이러한 방법은 매우 효과적이며, 본 연구에서도 구문분석 전에 몇몇 경우에 구 묶음을 수행하여 생성되는 구문분석 결과의 수를 줄였다.

동사의 논항 정보 역시 중의성을 줄이는 유용한 정보이다. 전은희[10]는 논항 사전을 이용하는 구문분석 후처리기를 이용해 구문분석기의 성능을 높였다.

이 외에도, 김광백[2]은 구간 분할을 통해 긴 문장에 대해 구문 분석의 효율을 높였으며, 황선영[12]은 명사 수식 부사의 언어적 특징을 분석하여 구문 분석기의 성능을 향상시켰다.

3. 시스템 전체 개요

본 논문에서 제안하는 구문분석기(이하 KLParse)는 다음 표 1을 전제로 한다.

<p>전제 1. 모든 지배소는 의존소 뒤에 위치한다. 전제 2. 모든 의존소는 하나의 지배소를 가진다. (단, 문장의 최종 지배소는 지배소를 가지지 않는다.)</p>
--

표 1. KLParse의 전제사항

전제 1에 의해, 도치된 문장과 같은 몇몇 특수한 문장에 대해서는 특수한 처리가 필요해진다. 하지만, 이 전제가 없으면 구문분석기의 복잡도(complexity)가 너무 올라가게 된다. 전제 2에 의해, KLParse가 출력하는 트리의 최상위 노드(top node)는 항상 문장의 마지막 요소가 된다.

문장 분석의 기본 단위로 형태소를 선택했다. 이 경우 어절 단위의 분석보다 중의성이 높아지나, 규칙을 만들어 있어서 언어학적으로 훨씬 깔끔하다. 예를 들어, “예쁜 집으로”를 분석할 때, 형태소 단위는 [[예쁜 집]으로]로 분석이 되며, 어절 단위는 [[예쁜][집으로]]로 분석이 되는데, 형태소 단위가 훨씬 자연스럽다. 또한, 어절 단위 분석의 경우, 어절을 표현하는 품사 패턴의 종류가 늘어날 수 밖에 없다. 앞의 예에서, [집으로]의 품사 패턴을 ‘부사’라고 할 경우, 앞 어절인 [예쁜]의 수식을 받을 수 없다. 관형사가 부사를 수식하는 것은 문법적으로 틀리기 때문이다. 그렇다고, [집으로]의 품사 패턴을 ‘명사’라고 할 경우, 이 어절이 나중에 동사의 지배를 받을 때

필요한 정보를 잃어버리게 된다. 따라서, 많은 경우에 어절의 품사 패턴은 그것을 구성하는 형태소의 품사를 2개 이상 반영해야 한다. [7]에서는, NP_OBJ, NP_MOD 등과 같이 어절의 품사 패턴을 설정하여 필요에 따라 어절을 구성하는 형태소의 품사를 2개 이상 조합하고 있다. 이로 인해 어절의 품사 패턴의 수는 형태소의 품사의 수보다 훨씬 커지며, 품사 패턴의 수의 증가는 곧 규칙 수의 증가를 가져온다.

3.1. KLParse의 구문분석과정

KLParse의 구문분석과정은 그림 1과 같다. 입력으로 형태소 분석기의 결과를 그대로 사용하며, 출력으로 모든 가능한 트리를 출력한다.

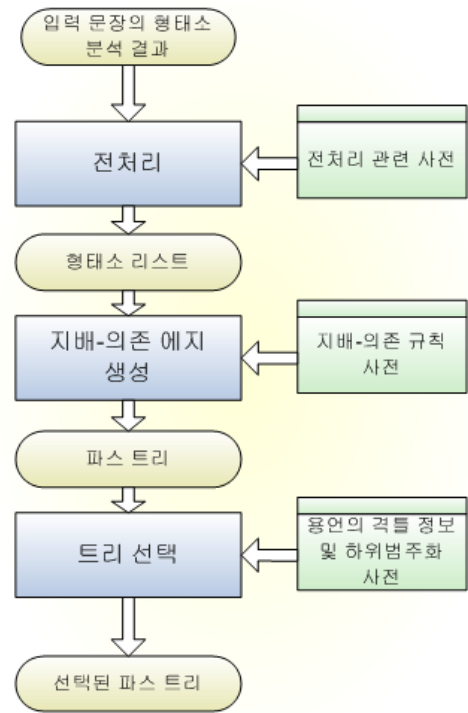


그림 1. KLParse의 구문분석과정

3.2. 전처리

KLParse는 품사 태거를 사용하지 않고 모든 형태소 분석 조합에 대해 분석을 시도한다. 전처리 단계에서는 ①몇몇 형태소들은 미리 합치고, ②형태소 분석 후보 중 일부를 제거하며, ③형태소 분석 결과를 조합하여 형태소 리스트를 만들어낸다.

①에서, 선어말어미 등과 같은, 통사적 의존관계를 가

지지 않는 형태소들은 미리 그전이나 후의 형태소와 합친다. 또한, 일부 복합명사와 관용구에 대해 구 묶음을 수행한다. 구 묶음의 예로는 (관형형 전성어미+수+있다), (관형형 전성어미+도+있다), 등과 같은 것들과 숙어처럼 사용되는 “시도때도없이”, “눈코 뜰 새 없이” 등이 포함된다.

②에서, 보조용언과 그 앞에 올 수 있는 어미를 사전으로 구축하여 분석후보를 제한한다. 또, 일부 의존명사들이 홀로 나타난 분석후보 등, 좌우 형태소를 보고 확실하게 제거할 수 있는 분석후보를 제거한다.

③의 예는 그림 2에 나타나있다.

그림 2에서, 3개의 어절로 된 문장이 각각 3, 2, 2개의 형태소 분석 후보를 가진다면, 조합 가능한 형태소 분석 리스트는 모두 $12(=3 \times 2 \times 2)$ 가 된다. 이 12개의 형태소 분석에 대해 각각 구문분석을 시도하게 된다.

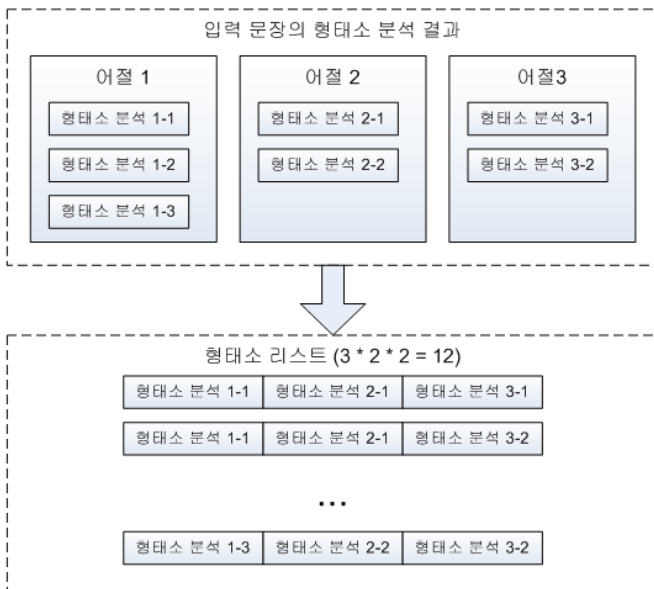


그림 2. 형태소 분석 리스트 생성

3.2. 지배-의존 예지 생성

실제 구문분석이 실행되는 단계이다. 수작업으로 구축된 의존문법 규칙을 바탕으로 지배-의존 예지(edge)를 생성하며 트리를 구축한다. 이 단계의 결과로는 모든 입력 형태소를 끝 노드(terminal node)로 가지는 트리가 출력된다.

3.2.1. 의존 규칙 및 제약 규칙

의존문법 규칙은 입력 형태소의 두 품사의 이진 연산(binary operation)으로 이루어져 있으며, 그 결과가 다시 품사로 출력되는 닫힌 연산(closed operation)이다. KLParse는 명사구, 동사구와 같은 구(phrase)의 개념을 도입하였다. 비록 품사 분류의 수와 규칙의 수가 늘어났다는 단점이 있지만, 하나의 형태소일 때와 그것이 다른 것을 지배한 후의 규칙 적용이 다를 때를 처리하기 쉽다. 예를 들어, ‘이다.’와 같은 지정사는, 단순히 지정사일 때는 바로 앞의 명사 등만 지배할 수 있지만, 그것을 지배하여 지정사구가 되면 논항을 지배할 수 있게 된다. KLParse에는 구를 포함하여 108개의 품사 분류와 이를 바탕으로 한 207개의 의존문법 규칙이 적용되었으며, 이러한 품사 분류와 규칙들을 계속해서 개선해 나가고 있다.

또한, 규칙을 적용할 때, 보다 하위 범주의 규칙을 우선시하여 적용한다. 예를 들어, 형태소 품사 중 ‘체언’이 있고, 그 하위품사로 ‘명사’가 있다. 규칙을 적용할 품사가 ‘명사’일 때, ‘체언’에 대한 규칙과 ‘명사’에 대한 규칙이 충돌한다면 ‘명사’에 대한 규칙을 우선시한다.

의존문법 규칙을 적용함에 2가지 제약요소 - 인접과 건너뛰기를 추가하였다.

인접이란, 어떤 규칙이 두 형태소에 적용될 때, 그 두 형태소가 입력 형태소 리스트에서 반드시 인접한 형태소들이어야 한다는 것이다. 즉, 어떤 규칙이 인접한다면, 바로 그 규칙은 연속해서 나오는 두 형태소일 경우에만 적용할 수 있다. 인접하지 않다면 그에 상관없이 지배할 수 있다. 형태소 단위로 의존문법을 적용할 때는 어절 단위로 의존문법을 적용할 때보다 모호성이 커지는데, 인접은 이를 어느 정도 완화해 준다.

지배소	의존소	결과	인접	건너 띄기
명사	관형사	명사구	FALSE	FALSE
명사	관형사구	명사구	FALSE	FALSE
주격보격조사	명사	격조사구	TRUE	FALSE
주격보격조사	명사구	격조사구	TRUE	FALSE
목적격조사	명사	격조사구	TRUE	FALSE
목적격조사	명사구	격조사구	TRUE	FALSE
동사	명사	동사구	FALSE	FALSE
동사	격조사구	동사구	FALSE	FALSE
동사구	명사	동사구	FALSE	TRUE
동사구	명사구	동사구	FALSE	TRUE
동사구	격조사구	동사구	FALSE	TRUE
종결어미	동사	종결구	TRUE	FALSE
종결어미	동사구	종결구	TRUE	FALSE
온점	종결어미	문장	TRUE	FALSE
온점	종결구	문장	TRUE	FALSE

표 2. 의존문법 규칙의 예

건너뛰기란, 어떤 규칙이 두 형태소에 적용될 때, 입력 형태소 리스트 상에서 두 형태소 사이에 지배소 형태소가 없을 때만 지배 가능하다는 것이다. 예를 들어, [관형

사 명사1 명사2] 와 같은 입력 형태소 리스트가 있을 때, 명사가 관형사를 지배하는 규칙이 건너뛰기가 허용이 안 되었다면, 명사2는 관형사를 지배할 수가 없다. 사이에 명사1이 있기 때문이다. 이는 지배-의존 관계에 지배소와 의존소 사이의 거리를 고려할 수 있게 한다. 이러한 의존문법 규칙의 예가 표 2에 있다.

KLParser는 품사를 필요에 따라 세밀하게 분류하여 중의성을 줄인다.

한국어에서 부사는 문장 안에서의 위치가 자유롭다는 특성을 지닌다[12]. 이러한 부사의 특성 때문에, 문장에 여러 개의 용언이 나타날 때, 부사에 의한 중의성이 발생한다. KLParser에서는 일반부사를 그 수식 범위에 따라 표 3과 같이 명사 수식 부사, 형용사(관형사) 수식 부사, 동사 수식 부사로 하위범주화하였다. 총 3,112개 부사를 하위범주화하여 사전으로 구축하고, 이를 기반으로 의존관계 생성을 제약하였다.

의존소	지배소	예	개수
명사 수식 부사	명사, 수사, 대명사	바로, 단, 마치, 유독	8
형용사(관형사) 수식 부사	형용사, 관형사, 부사	아주, 매우, 덜	74
동사 수식 부사	동사	꼭, 쿡, 그만	2,921
문장 수식 부사, 접속부사	연결구, 종결구	사실상, 예컨대, 그리고	109

표 3 부사의 하위범주화를 통한 의존관계 제약

부사 하위 범주화 중 형용사 수식 부사에 의해 중의성이 제거되는 과정이 그림 3에 있다. 생성되는 파스 트리의 수를 16개에서 8개로 줄였다.

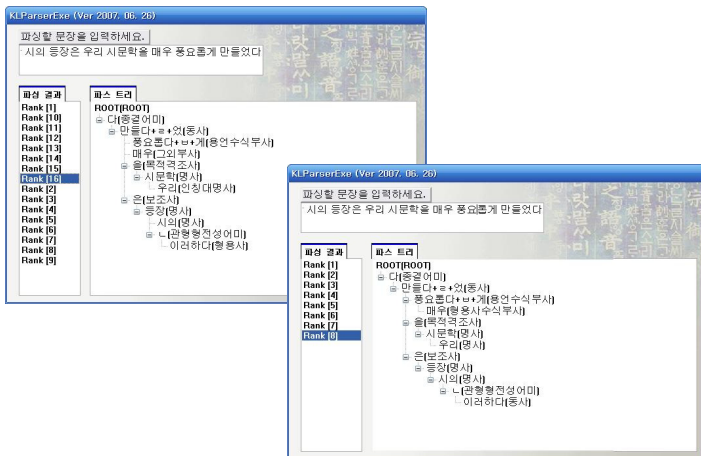


그림 3 형용사 수식 부사와 형용사가 아닌 용언과의 배제관계

하나의 문장에 두 개 이상의 절이 접속부사 혹은 연결어미로 연결되는 경우, 접속부사와 연결어미를 경계로 앞의 절에 있는 의존소는 연결어미나 접속부사 뒤에 있는 지배소의 지배를 받지 않는다. (단, 용언과 보조용언, 용언과 용언을 연결하는 연결어미나 어휘를 연결하는 접속부사는 이 제약에 해당되지 않는다.)

따라서 표 4와 같이 연결어미를 세분화하여 ‘절 간 연결만 가능한 연결어미’를 경계로 해당 연결어미 앞의 의존소가 연결어미 뒤의 지배소에 의해 지배를 받는 의존관계를 배제하였다. 이 제약은 문장의 길이가 길고 두 개 이상의 절이 연결어미로 복잡하게 연결되는 문장에서 생성되는 파스 트리의 수를 크게 줄일 수 있다.

연결어미	예	예문	제약
어휘 연결도 가능한 연결어미	-며, -고, -나, -매, -며, -디, -으나, -자	철수가 맑디 맑은 물을 마셨다.	연결어미를 넘을 수 있음
절 간 연결만 가능한 연결어미 (절 간 연결어미)	-는데, -지만, -느라고, -니까, -면서	생물 선생님은 학교를 나가면 서 또다시 실수를 하였다.	연결어미를 넘는 의존관계를 배제

표 4 연결어미의 세분화를 통한 의존관계 제약

이에 대한 예가 그림 4에 나타나있다. 생성되는 파스 트리의 수를 크게 줄이고 있음을 확인할 수 있다.

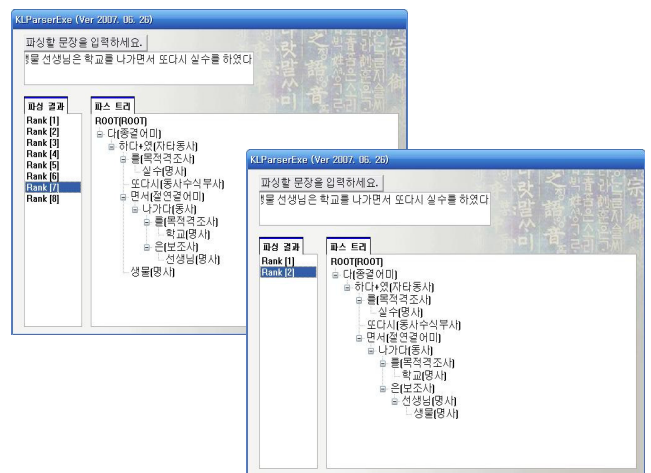


그림 4 절 간 경계를 넘는 의존관계의 배제를 통한 파스 트리 생성의 제약

3.2.2. 구문분석 알고리즘

KLParser의 구문분석 알고리즘은 [1]에 있는 구문분석 알고리즘을 사용한다. 이 알고리즘은 자료구조로 차트를 사용하는 차트 파싱이며, 입력 형태소 리스트를 백트래킹(back-tracking)하지 않는 1-way 알고리즘이다.

3.3 트리 선택

이전 단계에서 생성된 모든 파스 트리에 대해 적용된다. 부적절한 트리를 제거하고자, 용언의 논항 정보를 이용한다. 용언들이 논항으로 취할 수 있는 대상의 의미 범주를 제한하고자, 보통 용언의 논항 정보 사전은 실험 말뭉치를 기반으로 작성된다. 따라서 실험 말뭉치마다 사전을 재구축 해야 하는데, 이를 피하고자 부산대학교에서 구축한 한국어 어휘의미망(이하 KorLex)[9][13]을 사용한다. 논항 정보 사전을 구축할 때 KorLex에 있는 단어를 사용하고, 시스템에 적용할 땐 KorLex에서 논항 정보 사전의 의미범주와 실제 문장에 나타난 의미범주의 상·하위 정보를 받아와 사용한다. 즉, 실제 문장에 나타난 의미범주가 논항 정보 사전의 의미범주의 하위 범주 이면 맞다고 하는 것이다. 이에 대한 예가 그림 5에 나타나 있다. 논항 정보를 적용하기 전에 생성된 파스 트리의 수는 7개이며, 논항 정보 적용 후의 파스 트리 수는 3개로 4개가 감소했음을 알 수 있다.

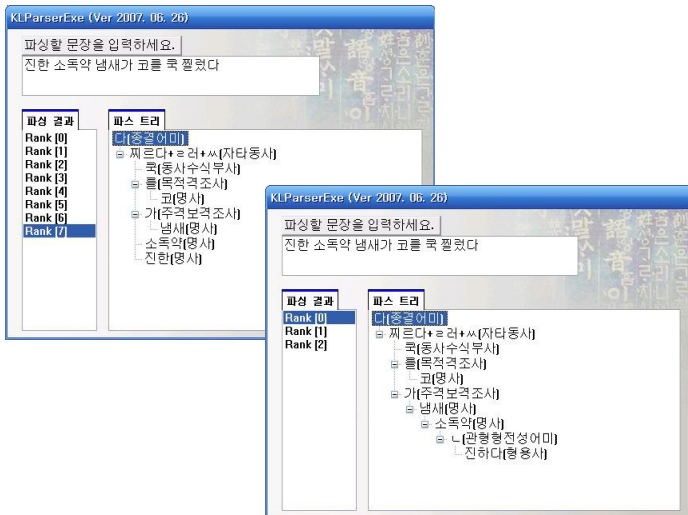


그림 5 논항 정보를 이용한 형태소 중의성 해소

논항 정보 사전을 적용할 때는 문장의 필수 성분이 생략된 경우도 고려한다. 어떤 용언이 있다면, 그가 취한 논항들이 논항 정보 사전을 위배하는지만 확인한다. 즉, 지배할 수 없는 조사나 의미를 지배한 여부만 파악하며, 그러한 논항을 취한 용언이 있는 트리는 결과에서 제외된다.

논항 정보 사전은 표 5와 같은 형식으로 구성되었다.

실험 말뭉치를 바탕으로 의미범주가 정해졌으며, 이 의미범주는 추후 실험을 통해 더 상위 개념으로 바뀌거나, 다른 개념이 추가되는 등 확장될 수 있다.

용언	품사	논항1		논항2	
		조사	의미범주	조사	의미범주
다투다	자동사	이	인간, 집단	에서	장소
매다	타동사	이	인간	을	논, 밭, 풀, 끈, 밧줄
쓰다	타동사	이	인간	을	수단, 행위, 속성, 정신적특징
찌르다	자타동사	가	냄새	를	신체부위

표 5 논항 정보 사전의 예

4. 실험

4.1. 실험환경

국내에서 공식적으로 개발된 한국어 구문트리 부착 말뭉치로 세종 계획에서 구축된 것이 있다[6]. 하지만, 이 말뭉치는 어절을 기본 단위로 삼고 있으며, 지역 중의성을 고려하지 않고 있다. 따라서 중학교 교과서에서 임의로 문장을 선택하여 직접 답안을 작성하였다. 현재까지 10어절 이내(27형태소 이내)의 120개 문장에 대해 의존관계에 대한 답안이 작성되었다.

성능 평가는 PARSEVAL[15]을 사용한다. 이 평가 기준은 구문 트리 전체에 대해서 정확도를 평가하는 것이 아니라, 구문 트리를 구성하는 각각의 구성성분(constituent)에 대해서 그 정확도를 평가하는 것이다[12]. 의존문법에선 의존관계, 즉 에지의 정확도(precision)과 재현율(recall)을 평가한다.

$$\text{정확도} = \frac{\text{결과 트리 내에서 올바른 의존관계의 수}}{\text{결과 트리 내의 의존관계의 수}} \times 100$$

$$\text{재현율} = \frac{\text{결과 트리 내에서 올바른 의존관계의 수}}{\text{정답 트리 내의 의존관계의 수}} \times 100$$

실험에 사용된 시스템의 사양은 Intel CPU 1.87GHz, 2GB RAM이며, 운영체제는 Microsoft Windows XP이다.

4.2. 실험결과

본 논문에서 제시한 기법들을 적용하여 만든

KLParser의 성능은 표 6과 같다. KLParser의 성능은 정확도 74.3%, 재현율 98.5%이다.

현재 KLParser는 하나의 파스 트리만을 선택하지 않고 복수개의 답안을 출력하므로, 타 시스템에 비해 상대적으로 낮은 정확도와 높은 재현율을 보이고 있다. 이 성능을 다른 구문분석기와 직접적으로 비교하는 것은 무리가 있으며, 하나의 파스 트리만을 선택하기 위해서, 단어 자체의 형태소 분석 확률 정보를 이용하는 등의 추가적인 모듈이 필요할 것이다.

PARSEVAL과는 별도로, 중학교 교과서 말뭉치에서 임의로 선택한 512문장에 대해서 출력하는 평균 구문분석 트리의 수는 9.2개이다.

정확도	재현율
74.3%	98.5%

표 6 KLParser의 성능

5. 결론 및 향후과제

본 논문에서는 규칙 기반 의존문법 한국어 구문분석기에 대해 다루었다. 한 문장의 모든 형태소 분석 조합에 대해 구문분석을 시도했으며, 여러 언어적 특성에 기반을 두어 수작업으로 작성된 규칙들이 적용되었다. 10어절 이내의 문장이라는 점을 고려하면 정확률이 많이 높지 않다. 하지만, 아직 적용할 언어적 정보들이 남아 있으므로, 규칙 기반 구문분석 역시 충분히 가능하다는 것을 보여준다 하겠다. 또한, 각 문장의 구문분석 트리 수의 편차도 비교적 심한데, 이 부분은 의존문법 규칙을 좀 더 발전시켜서 해결 가능할 것이다.

KLParser는 품사 태거를 거치지 않고 모든 형태소 분석 조합에 대해 구문분석을 시도하기 때문에, 어절 수가 늘어날수록, 특히 용언의 숫자가 늘어날수록 중의성이 급격하게 커진다. [2]는 문장을 절 단위로 나누는 방법을 제안했는데, 앞으로 이것을 접목시키면 좋을 것 같다.

최종 출력되는 트리를 하나만 선택하려면 확률기법이 반드시 필요하다. 최종 출력된 트리 들 중, 형태소 분석의 확률의 합이 가장 높은 트리를 출력하는 방법 등이 있을 수 있다. 확률 외에도, 지배-의존 관계(수식, 논항 등)와 지배거리의 관계도 단서가 될 것 같다. 예를 들어, 수식 관계는 그 지배거리가 가까운 경향이 있으며, 논항 관계는 가장 첫 주어는 가장 마지막 용언 등과 결합할 확률이, 목적어는 가까운 타동사에 지배될 확률이 높았다. 이러한 정보를 잘 찾아내어 성능향상을 도모할 수 있을 것이다.

참고 문헌

[1] 권혁철, 최준영, “단일화 기반 의존 문법을 이용한 한국어 분석기”, 한국정보과학회 논문지 '92. 9 Vol.19,

No.5, SEPTEMBER.

- [2] 김광백, 박의규, 나동열, 윤준태, “구간 분할 기반 한국어 구문분석”, 제14회 한글 및 한국어 정보처리 학술대회 pp.163-168, 2002.
- [3] “자연언어처리”, 김영택 외 공저, 생능출판사.
- [4] 김창제, 정천영, 김영훈, 서영훈, “부분적인 어절 결합을 이용한 효율적인 한국어 구문 분석기”, 한국정보과학회 가을 학술발표논문집 vol. 22, No.2, 1995.
- [5] 류법모, 이태승, 이종혁, 이근배, “술어 중심 제약 전파를 이용한 2-단계 한국어 의존 파서”, 1996년도 한국정보과학회 봄 학술발표논문집 Vol. 23, No. 1, pp.923-926.
- [6] 박의규, 나동열, “한국어 구문분석을 위한 구문을 기반 의존명사 처리”, 인지과학, 제17권 제2호, pp.119-138, 2006.
- [7] 은지현, 정민우, 이근배, “확률적 차트 파싱에 기반한 한국어 의존 구조 분석기”, 제17회 한글 및 한국어 정보처리 학술대회 논문집, pp.105~111, 2005.
- [8] 이미경, 정한민, 성원경, 박동인, “품사 표지 부착 말뭉치 검증”, 제17회 한글 및 한국어 정보처리 학술대회 논문집, pp.145-150, 2005.
- [9] 이은령, 윤애선, “피동 정보를 통한 한국어 동사 어휘 의미망 정제”, 한국어학 28, pp.139-166, 2005.
- [10] 전은희, 이성욱, 서정연, “한국어 동사의 격틀 정보를 이용한 구문분석 후처리”, 제13회 한글 및 한국어 정보처리 학술대회 논문집, pp.445-449, 2001.
- [11] 정석원, 박의규, 나동열, 윤준태, “격 관계와 상호정보를 이용한 한국어 의존 파서”, 제13회 한글 및 한국어 정보처리 학술대회 논문집, pp.450-456, 2001.
- [12] 황선영, 이공주, “구조적 중의성 해결을 위한. 명사 수식 부사 연구”, 한국인지과학회 논문지. 제 13 권 제 4 호, pp.43-53, 2002.
- [13] 황순희, 윤애선, “의미자질을 고려한 명사어휘의미망 구축(1): 남/여성 자질을 중심으로”, 한국어학 29, pp.309-338, 2005.
- [14] 황이규, 이현영, 이용석, “형태소 및 구문 모소형 축소를 위한 구문단위 형태소의 이용”, 한국정보과학회 논문지 제27권 7호, pp.784-793, 2000.
- [15] E.Black, Meeting of interest group on evaluation of broad-coverage grammars of English. LINGUIST List 3.587, <http://www.linguistlist.org/issues/3/3-587.html>.
- [16] Eugene Charniak, "A Maximum-Entropy-Inspired Parser", Preceedings of North American Chapter of the Association for Computational Linuistics, pp.132-139, 2000.
- [17] Hoojung CHUNG, Hae-Chang RIM, "A New Probabilistic Dependency Parsing Model for Head-Final, Free Word Order Languages", IEICE TRANS. INF. & SYST., VOL.E86-D, No.11 NOVEMBER 2003.

- [18] Michael Collins, "Three generative, lexicalised models for statistical parsing", Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, San Francisco, Morgan Kaufmann.