

형태소 및 구문 모호성 축소를 위한 구문단위 형태소의 이용

(Using Syntactic Unit of Morpheme for Reducing
Morphological and Syntactic Ambiguity)

황 이 규[†] 이 현 영^{**} 이 용 석^{***}

(Yi-Gyu Hwang) (Hyun-Young Lee) (Yong-Seok Lee)

요약 기존의 한국어 형태소 해석은 한국어의 교착어 특성상 다양한 형태론적 모호성을 가진다. 이러한 형태론적 모호성은 많은 구문 모호성을 만들어 내어 올바른 의미를 가지는 파스트리의 선택을 어렵게 한다. 한국어에서 이런 형태론적 모호성의 대부분이 보조용언이나 의존 명사와 관련이 있다. 보조용언이나 의존명사는 주위의 형태소들과 강한 결합 관계를 가지고 있으며, 대부분 자립성이 없는 기능형태소이다. 결합된 형태소들은 문장 내에서 하나의 기능적 역할이나 구문적 역할을 수행한다. 우리는 품사 태깅된 20만 어절 크기의 말뭉치로부터 이 형태소열을 찾아 이를 3가지 유형으로 분류하였다. 그리고, 이를 구문 형태소로 정의하고 구문 형태소를 구문 분석의 기본 입력 단위로 간주하였다. 본 논문에서는 구문 형태소가 아래와 같은 문제의 해결에 있어서 효율적인 방법임을 제안한다: 1) 형태소 해석의 축소, 2) 구문 해석 도중 불필요한 부분 파스트리의 배제, 3) 구문 모호성의 축소. 마지막으로, 실험 결과를 통해 구문단위 형태소가 형태소 및 구문 모호성을 축소하기 위해 반드시 필요함을 보인다.

Abstract The conventional morphological analysis of Korean language presents various morphological ambiguities because of its agglutinative nature. These ambiguities cause syntactic ambiguities and they make it difficult to select the correct parse tree. This problem is mainly related to the auxiliary predicate or bound noun in Korean. They have a strong relationship with the surrounding morphemes which are mostly functional morphemes that cannot stand alone. The combined morphemes have a syntactic or semantic role in the sentence. We extracted these morphemes from 0.2 million tagged words and classified these morphemes into three types. We call these morphemes a syntactic morpheme and regard them as an input unit of the syntactic analysis. This paper presents the syntactic morpheme is an efficient method for solving the following problems: 1) reduction of morphological ambiguities, 2) elimination of unnecessary partial parse trees during the parsing, and 3) reduction of syntactic ambiguity. Finally, the experimental results show that the syntactic morpheme is an essential unit for reducing morphological and syntactic ambiguity.

1. 서론

한국어 형태소 해석의 결과는 많은 형태론적 모호성

을 가지고 있기 때문에 형태소 해석의 결과를 바로 구문 해석의 입력으로 하여 구문 해석을 수행하기에는 여러 가지 문제가 있다. 형태소 해석의 과생성 문제는 구문 해석 과정에서 불필요한 연산을 수행하게 하며, 불필요한 구문 트리를 만들어 내기도 한다. 이러한 문제를 해결하려고 형태소 해석을 하면서 간단한 형태소 배열 규칙을 이용하여 제약하거나 과생성된 형태소 해석 결과에서 하나의 품사열로 추정하기 위해 태깅을 도입하였다[1].

태깅은 형태소 해석 결과를 품사열로 가정하고, 품사열에 대한 바이그림이나 트라이그림 등의 통계정보를

† 학생회원 : 전북대학교 컴퓨터과학과
yghwang@konji.chonbuk.ac.kr

** 비 회원 : 전북대학교 컴퓨터과학과
hylee@konji.chonbuk.ac.kr

*** 종신회원 : 전북대학교 컴퓨터과학과 교수
yslee@moak.chonbuk.ac.kr

논문접수 : 1999년 7월 6일
심사완료 : 2000년 5월 24일

축적한 후, 이러한 통계정보를 바탕으로 주어진 여러 품사열에서 하나의 품사열을 선택함으로써 구문 해석단계의 부담을 줄인다. 물론 이러한 방법을 통해 구문 해석의 입력을 줄임으로써 구문 해석 단계의 과부하를 줄일 수 있지만, 태깅 단계에서 100% 올바른 결과를 보장하기 어렵고 또한 태깅 시스템의 학습을 위한 자료 등을 축적하기 어려운 문제점이 있다[2].

예를 들어 “먹어 보았다”라는 문장에 대한 해석 과정을 살펴보자. 형태소 분석을 “여러 형태소들의 묶음이 표층형태로 나타나는 하나의 단어로부터 의미를 갖는 최소 단위인 각 형태소를 분석해 내는 것[3]”으로 정의하였을 때, “먹어 보았다”란 문장을 의미의 최소단위인 형태소들로 분해하면 [그림 1]과 같다¹⁾. [그림 1]은 가능한 형태소 해석 후보를 모두 생성한 후 형태소 배열 규칙을 적용하여 불필요한 형태소들을 제거한 것이다. 예를 들어, “다”라는 형태소는 의존명사나 일반명사로 간주할 수 있으나 “았”이라는 선어말 어미 뒤에는 명사가 나타날 수 없다는 배열 규칙을 이용한 것이다. 만일 형태소 배열 규칙을 적용하지 않는다면 [그림 1]보다 많은 형태론적 모호성이 발생한다. 아래의 그림에서 형태소 배열 규칙을 이용해 많은 형태론적 모호성이 아직 해결되지 않은 상태로 남아 있다.

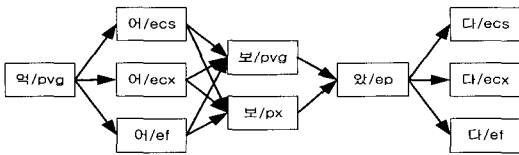
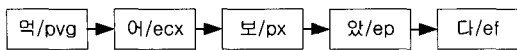


그림 1 “먹어 보았다”의 형태소 해석 결과

이를 해소하기 위해 일반적으로 태거를 이용한다. 태거를 이용할 경우, 아래와 같은 결과를 얻을 수 있다.



문장 해석의 최종 목표가 형태소 해석이 아니면, 입력 문장으로부터 구문 해석이나 의미 해석 단계를 거쳐 주어진 문장에 대한 의미를 얻거나 형태소 해석 결과로부터 명사를 추출하는 정보검색의 인덱싱 과정을 거칠 수 있다. “먹어 보았다”란 문장의 최종적인 해석 결과를 의미 해석의 관점에서 볼 때, “먹다”라는 어휘의 의미에 “시도”, “과거”, “평서”라는 의미가 추가된 형태로 볼 수 있

다. “어/ecx 보/px”로부터 “시도”, “았/ep”으로부터 “과거”, “다/ef”로부터 “평서”의 정보를 유추할 수 있다.

먹/pvg(시도, 과거, 평서)

그런데, “시도”와 같은 양상 자질을 가지는 형태소열은 특징이 있다. 이들은 특정한 기능형태소들이 결합하여 문장에서 하나의 기능을 수행한다. 또한 이외에도 많은 형태소가 서로 결합하여 하나의 기능적, 구문적 단위를 형성할 수 있다. 이들은 특정 품사 배열과 특정 어휘 배열을 가지므로 쉽게 하나의 단위로 묶을 수 있으며, 형태소끼리 서로 강한 결합 관계를 형성하고 있다. 따라서 이들을 구문 해석 전에 효과적으로 묶을 수 있다면 형태소 모호성을 크게 축소할 수 있으며, 구문 해석의 부담을 크게 줄일 수 있다.

이 논문에서는 기능형태소 해석 결과에 있는 형태소 모호성을 줄여 구문 해석의 부담을 줄이는 방안으로 여러 형태소가 결합하여 하나의 구문적 단위나 의미적 단위를 가질 수 있는 형태소들을 구문 형태소로 정의하고, 한국어에 어떠한 형태소열들이 구문적 형태소 단위를 이루고 있는지 살펴본다.

속어나 관용어는 특정 내용어와 특정 기능어가 결합하여 하나의 의미적 단위를 이루므로 그 수가 매우 한정적이다. 그러나 구문 형태소는 개개의 기능 형태소들이 결합하여 하나의 의미적 형태소 단위를 이루며, 개개 형태소들의 의미가 결합하여 형태소열의 의미를 한정하는 것이 아니라 형태소들이 결합하여 새로운 의미를 나타내는 것을 말한다. 또한 구문 형태소는 문장에서 내용어의 역할을 수행하지 않고 기능어의 역할을 수행하므로 속어나 관용어에 비해 문장에서 자주 발생하며, 다양한 체언이나 용언에 이 구문 형태소가 결합할 수 있기 때문에 언어 생성력이 매우 크다. 따라서 이러한 구문 형태소에 나타나는 형태론적 모호성의 해소는 구문 해석의 부담을 크게 줄일 수 있다. 이 논문에서는 코퍼스로부터 일정 횟수 이상의 빈도를 가지며, 한 어절 이상에 걸친 기능형태소들을 대상으로 체언이나 용언에 구조적/의미적 영향을 미치는 빈도 상위 어절을 대상으로 이러한 기능형태소의 특징을 조사하고 이를 구문형태소로 정의하였다. 구문 형태소는 선행하는 형태소에 정보를 더하거나 그 자체로 하나의 문법 범주가 될 수 있다. 이것을 관용구의 관점에서 본 연구로 [5]가 있었다.

이 연구는 복수어 단위 분석 방법[6]과 구문 해석전에 여러 어절을 하나의 단위로 결합하는 방법[7, 8]을 일반화하였다. 문장 성분사이의 통합관계를 중심으로 구

1) 본 연구에서 사용하는 품사 태그는 통합 국어정보베이스[4]를 위한 한국어 형태·통사 태그 설정 기준을 따른다.

성되는 복수어 단위 정보를 이용하는 [6]의 경우, 말뭉치를 기반으로 모든 결합유형을 사전에 수록하고 숙어나 연어 구조도 복수어 단위 사전을 기반으로 분석한다. 이는 단순히 형태론적 모호성의 해소를 위해 이러한 정보를 이용한 것으로 모든 결합유형을 사전에서 수록하기 어렵다. 부분적인 어절 결합 관계를 이용한 방법[7]이나 최장 묶음을 이용한 구문 해석[8]의 경우, 중심어에 수식하는 첨가어나 보충어를 구문 해석전에 결합함으로써 파서의 부담을 줄이려는 노력이다. 본 연구를 [6]와 비교하면, 단순히 형태론적 모호성 해소에 그치지 않고 구문 형태소를 구문 해석을 위한 하나의 단위로 묶어 파서의 입력 단위로 가정하기 때문에 구문 해석의 효율성을 높일 수 있다. 또한, [7]과 [8]에서는 구문해석 전에 간단한 구구조 규칙을 이용하는 부분 파싱과정이 필요하지만, 본 연구는 기능적 단위의 형태소만을 결합 단위로 간주하여 일관성을 유지하였다. 또한, [9]와 [10]의 경우, 본용언과 보조용언의 기능형태소에 대한 결합의 필요성을 주장하였는데, 본 논문에서는 이를 일반화하여 한국어 문장에서 나타날 수 있는 여러 형태의 기능 형태소 결합을 살펴보고 이들을 하나의 단위로 처리하는 방법에 대해 기술하였다.

[11]에서는 품사 태깅을 위해서 묶인말이라는 형태소 패턴을 사전에 저장하고 이를 이용하여 태깅을 시도하였으며, [12]에서는 문장상에서 나타나는 단어열 중 언어적 속성을 가지는 단어들을 다중 단어로 정의하고 이를 품사 태깅 모델에 적용하였다. 또한, [2]에서는 관용 어구를 정의하고 이를 이용하여 태깅시 형태론적 모호성의 축소에 이용하였다. 이러한 연구의 특징은 단순히 형태론적 모호성의 축소에 초점을 맞추었다. 본 연구에서는 구문 해석의 관점에서 어떠한 구문형태소들이 반드시 기능적 단위로 결합되어야 하며, 이 구문 형태소가 문장에서 어떠한 기능을 하고, 이들을 어떠한 구문적 단위로 보고 문장을 해석해야 하는지에 초점을 맞추었다.

2장에서는 구문 형태소의 종류를 살펴보고, 문장에서 어떠한 문법적 역할이나 의미적 표현으로 나타나는지 살펴본다. 3장에서 이러한 구문 형태소들을 결합하는 방법과 이를 이용한 구문해석 과정에서 검토되어야 할 점들을 설명한다. 4장에서는 구문 형태소를 이용한 구문 해석과 일반 형태소를 이용한 방법을 비교해 보고 5장에서 결론과 향후 연구 방향에 대해 토의한다.

2. 구문 형태소

2.1 구문 형태소의 정의

구문형태소란 여러 기능형태소들이 결합하여 하나의 구문/의미적 단위를 형성하는 형태소 열을 말한다. 한국어에서 구문형태소로 정의할 만한 기능 형태소의 결합으로 크게 두 가지가 있다. 기능형태소들이 결합하여 하나의 양상 자질을 나타내는 경우와 기능형태소와 용언이 결합하여 하나의 심층격 조사역할을 하는 형태소열이 그것이다. 양상이란 어떤 사건이나 행동, 상태에 대한 화자의 태도를 표현한다. 예를 들면 겸양, 추측, 피동, 소망, 가능, 부정, 진행, 시도, 완료 등이다. 이 양상은 구문 해석 단계에서 구구조 규칙에 의해 용언에 대한 부가 자질의 형태로 표현될 수 있고, 의미 해석 단계에서 여러 지식을 이용하여 인식된 구구조를 하나의 의미 자질로 나타낼 수 있다. 또한 심층격 조사상당 형태소 열은 구구조 관점에서 볼 때, 주어진 문장의 정확한 문법적 구조를 파악하는데 도움이 되지만 의미 해석을 어렵게 한다. 이런 양상 자질이나 조사 상당 어구를 구문 형태소 단위로 구문 해석 전에 인식하면 형태소의 모호성 축소와 구문 해석의 과정을 단순화시킬 수 있다.

2.2 구문 형태소의 종류

2.2.1 보조 용언을 매개로 한 구문 형태소

한국어에서 용언의 뒤에 보조적 연결어미를 매개로 선행 용언에 속성을 첨가하는 또 다른 용언이 나타날 수 있는데 이를 보조용언이라고 한다. 보조용언은 구문적 역할을 수행하지 않고 선행하는 용언에 단순히 양상 정보만 부가한다. 따라서 이러한 보조용언을 본용언의 한 부분으로 간주함으로써 형태론적 모호성을 줄일 수 있다. 예를 들어, “먹어 보았다”에서 “보았다”가 선행하는 “먹다”에 ‘시도’라는 양상 정보만 추가함을 알 수 있다. 따라서 “먹[시도]+았+다”라는 하나의 어절로 해석할 수 있다[9, 10]. 또한 선어말 어미가 나타내는 시제, 높임, 공손 등도 양상의 의미를 나타내며, 어말어미가 나타내는 평서, 감탄, 의문, 명령, 청유, 연결, 전성 등도 용언 “먹”에 양상 의미를 첨가하므로 어절 분석 결과는 “먹[시도, 과거, 평서]”과 같이 나타낼 수 있다.

아래에 보조용언의 형태 및 보조용언이 선행 용언에 어떠한 양상 자질을 추가해 주어야 하는지 나타나 있다. Type 1 형태는 구문 구조에 어떠한 영향도 미치지 않고 용언에 양상 의미만 추가한다.

Type 1 : <v> {<ecx> [jxc] <px>}+ {<e>}+

예) “먹/pvg+고/ecx 싹/px+다/ef”

“먹/pvg+어/ecx+는/jxc 보/px+다/ef”

“먹/pvg+어/ecx 보/px+고/ecx+도/jxc 싹/px+다/ef”

[표 1]은 Type 1형 구문 형태소가 결합하여 선행 용언에 어떠한 양상 자질을 포함하여야 하는지를 나타내

며, [그림 2]에서는 Type 1형 구문 형태소 인식하기 위한 오토마타를 표현했다.

표 1 Type 1 구문 형태소의 양상 자질

구문 형태소	양상 자질
-어 보	시도
-고 싶	희망
-어야 하	당위
-려고 하	의도
-어 버리	완료

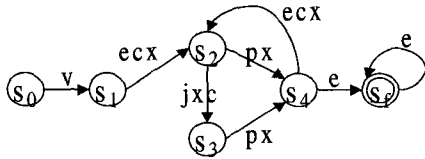


그림 2 Type 1형 구문형태소 인식을 위한 오토마타

[표 2]에서는 실제 코퍼스에서 추출된 Type 1형 구문 형태소의 일부를 정리했다.

표 2 Type 1형 구문 형태소의 예

(-아/-어) 지다, (-게) 되다, (-게) 하다, (-게) 만들다, (-아/-어) 가다, (-아/-어) 오다, (-고) 있다, (-고) 계시다, (-아/-어) 내다, (-아/-어) 버리다, (-고) 나다, (-고) 말다, (-아/-어) 주다, (-아/-어) 드리다, (-아/-어) 두다, (-아/-어) 놓다, (-아/-어) 가지다, (-아/-어) 대다, (-지) 말다, (-지) 못하다, (-지) 않다, (-아/-어) 보다, (-아/-어) 보이다, (-어야) 하다, (-고) 싶다, (-아/-어) 있다, (-아/-어) 계시다, (-는가/-나 가/-나) 보다, ...

Type 1형 구문 형태소는 선행하는 용언의 어간과 통합되어 하나의 단위로 인식된다. 이러한 양상 정보는 자질의 형태로 표현된다. 예를 들어 “먹고 있다”의 구문 해석 전의 결과는 “떡/pvg[진행, 평서]”이다. 즉 ‘진행’의 양상 자질과 평서형 종결어미가 용언 “떡”과 결합하여 하나의 구문 형태소를 이루었다.

2.2.2 의존명사를 매개로 한 구문 형태소

의존 명사가 용언과 결합하여 선행하는 용언에 양상 정보를 첨가하는 기능을 수행할 수 있다. 이때 분절적인 해석은 여러 가지 형태론적 모호성을 야기한다. 한국어에서는 이런 현상이 자주 발생하는데 이러한 형태소열은 구문 해석 트리의 문장 구조와 관련이 없고, 선행하는 용언에 양상 정보만 더한다. 그러한 형태소들을 보면

아래와 같다.

Type 2: <v1> <etm> {<nbm>}+ [<j>]* <v> {<e>}+
 예) “떡/pvg+을/etm 수/nbn 있/paa+다/ef”

“떡/pvg+을/etm 리/nbn+도/jxc 없/paa+다/ef”

Type 2': <v1> <etm> <nbm> <jp> {<e>}+

예) “떡/pvg+을/etm 터/nbn 이/jp+다/ef”

“떡/pvg+을/etm 모양/nbn+이/jp+다/ef”

[표 3]는 Type 2형과 2'형 구문 형태소가 결합하여 선행 용언 “<v1>”에 어떠한 양상 자질을 포함하여야 하는지를 나타내고 [그림 3]은 Type 2, 2'형 구문 형태소 인식하기 위한 오토마타를 표현했다.

표 3 의존명사 포함열의 양상 자질

구문 형태소	양상 자질
ㄹ 수(도) 있	가능
ㄹ 것이	희망
ㄹ 수밖에 없	필연
ㄹ 뿐이	결론
ㄹ 리 없	부정

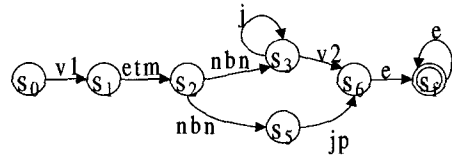


그림 3 Type 2, 2'형 구문형태소 인식을 위한 오토마타

[표 4]에서는 실제 코퍼스에서 추출된 Type 2, 2'형 구문 형태소의 일부를 정리했다.

Type 2, 2'형 구문 형태소는 선행하는 용언의 어간과 통합되어 하나의 단위로 인식된다. 이러한 양상 정보도 Type 1형과 마찬가지로 양상 자질의 형태로 표현된다. 예를 들어 구문 해석을 하기 전의 “먹을 수 있다”에 대

표 4 Type 2, 2'형 구문 형태소의 예

(-ㄹ/-는) 경우가 많다/있다/흔하다, (-ㄹ/-는) 바(가/도) 있다/없다, (-ㄹ/-는) 바에 따르다, (-ㄹ/-는) 셈이다, (-ㄹ/-는) 수가 많다/있다, (-ㄹ/-는) 적(도/은/이) 있다/없다, (-ㄹ/-는) 줄(도/은) 물랐다/알았다, (-ㄹ/-은) 편이다, (-ㄹ/-을) 리(가/는) 없다, (-ㄹ/-을) 만(은) 하다, (-ㄹ/-을) 모양이다, (-ㄹ/-을) 바(는/를) 모릅니다/없습니다, (-ㄹ/-을) 뿐(만) 아니다, (-ㄹ/-을) 뿐이다, (-ㄹ/-을) 줄 모르다/알다, (-ㄹ/-을) 지 모르다/알다, (-ㄹ/-을) 지경에 이르다, (-ㄹ/-을) 지경이다, (-ㄹ/-을) 터이다, ...

한 형태소 해석 결과는 “먹/pvg[가능, 평서]”이다. 즉 ‘가능’의 양상 자질과 평서형 종결어미가 용언 “먹”과 결합하여 하나의 구문 형태소를 이룬다.

2.2.3 의사 조사

조사가 “대하다”, “인하다”, “비하다”와 같은 의존명사에서 파생한 용언이나 “비롯하다”와 같은 일반 용언과 결합하여 문장 내에서 조사 상당 어구의 역할을 수행하는 경우가 많이 발견된다. 이는 의미 해석의 관점에서 형태소들을 관찰하였을 때 명백해진다. 예를 들어 “철수로 인하여 싸움에 졌다”라는 문장에서 “-로 인해”는 원인격 조사 상당어구로 간주할 수 있다. 이러한 단어열을 하나의 구문적 단위로 처리할 경우, 구문 해석과 의미 해석에서 많은 모호성을 제거할 수 있다. 이런 경우를 영어 표현으로 변환하면 단 하나의 전치사로 대응된다. 이 연구에서는 이러한 형태소열을 **의사 조사**로 정의하고 이를 구문 형태소의 범주에 포함시켰다. 아래는 그러한 구문 형태소들의 예이다. Type 1과 2, 2'형이 양상 자질의 형태로 표현되는데 반해 Type 3, 3'형은 문법 범주인 조사의 형태로 변환되어야 한다.

Type 3 : <n> <j> <v> <ecs>

예) “컴퓨터/ncn+에/jca 대하/pvg+어/ecs”

Type 3' : <n> <j> <n> <j>

예) “한국/ncn+과/jcj 마찬가지/ncn+로/jca”

[표 5]은 Type 3, 3'형 구문 형태소가 결합하여 선행체인 “<n>”에 어떠한 양상 자질을 포함해야 하는지를 나타내며, [그림 4]에서는 Type 3, 3'형 구문 형태소 인식하기 위한 오토마타를 표현했다²⁾.

표 5 의사 조사와 의사격

구문 형태소	의사격
-에 대해	to
-에 관해	about
-를 위해	for
-로 인해	cause

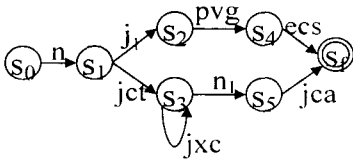


그림 4 Type 3, 3'형 구문 형태소 인식을 위한 오토마타

2) j1은 “와”, “에”, “로”, “를”등을 포함하는 몇 종류의 조사를 나타내며, pvg는 “달하-”, “따르-”, “비하-”, “의하-”등의 일부 용언을, n1은 “마찬가지”, “반대”, “별도”등의 일부 명사를 나타낸다.

[표 6]에서는 실제 코퍼스에서 추출된 Type 3, 3'형 구문 형태소의 일부를 정리했다.

표 6 Type 3, 3'형 구문 형태소의 예

(-에) 달해, (-에) 따라, (-에) 비해, (-에) 의해, (-에) 처해, (-에) 한해, (-에) 반해, (-와) 같이, (-와) 견주어, (-와) 관련하여, (-와) 달리, (-와) 함께, (-와) 더불어, (-를) 비롯해, (-를) 통해, (-를) 향해, (-를) 두고, (-를) 맞아, (-을) 가지고 (-로) 말미암아, (-로) 미루어, (-와) 마찬가지로, (-와) 반대로, (-와) 별도로, (-에) 있어.

Type 3, 3'형 구문 형태소는 선행하는 체인과 통합되어 체인구로 인식될 수 있다. 이러한 체인구에서 Type 3, 3'형은 의사 조사의 역할을 수행하고 있다. 예를 들어 “철수로 인해”의 구문 해석 전 결과는 “철수/nq+로 인해/j[cause]”가 될 것이다. 즉 “로 인해”가 하나의 문법 범주인 조사로 결합되어 체인인 “철수”와 함께 체인구를 이룬다.

2.2.4 기타 강한 어휘적 공기 관계를 가지는 형태소

코퍼스 분석의 결과, 이외에도 몇 가지의 기능형태소들의 강한 결합이 나타나고 있는데, 그중 명사형 전성어미인 ‘기’를 매개로 하는 형태소열도 빈번하게 발생하며 이들 또한 양상 정보를 가진다. 이들은 실제 어휘와 품사가 결합하여 구문형태소를 이룬다. 이를 정리해 보면 [그림5]와 같다³⁾.

예) “먹/pvg+기/etn+도/jxc 하/pvg+었/ep+다/ef”
 “먹/pvg+기/etn 때문/nbn+이/jp+다/ef”

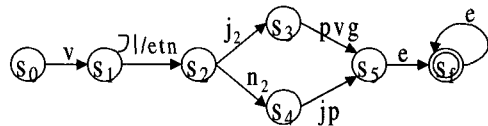


그림 5 다른 구문 형태소의 예

3. 구문 형태소의 생성과 적용

3.1 구문 형태소 생성

본 연구에서는 개개의 형태소들이 의미적으로 강한 결합 관계를 형성하는 형태소 열을 구문 형태소로 정의하였다. 구문 형태소 단위로 형태소들을 결합함으로써 불필요한 형태론적 모호성을 제거하기 위해 형태소 해석과 구문 해석 사이에 구문 형태소 생성기를 두었다.

3) j2는 “는”, “도”, “로”, “만”등을 포함하는 몇 종류의 조사를 나타내며, n2는 “때문”, “마련”, “심상”등의 일부 명사를 나타낸다.

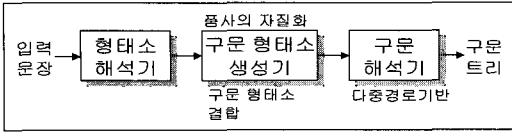


그림 6 형태소 및 구문 모호성 축소를 위한 자연어 처리 시스템의 구조

이 연구에서 제안한 구문 형태소 생성기는 형태소 해석 결과를 입력으로 하여 내부의 오토마타를 통해 구문 형태소를 인식하고 이를 양상자질 또는 품사의 형태로 변환한다. 이러한 방법은 기존의 형태소 해석기를 그대로 이용할 수 있으며, 새로 발견되는 구문 형태소들을 쉽게 추가할 수 있는 장점을 가진다.

본 논문에서는 구문 형태소 생성기의 기능을 크게 두 가지 관점에서 정의하였다. 첫째, 형태소 후처리의 역할이다. 형태소 해석기가 생성해 내는 많은 형태론적 모호성의 축소나 제거를 위해 다양한 형태소 제약 정보를 이용하는 방법[6, 11]이나 태거를 이용하는 방법[2, 12]와 마찬가지로 구문 형태소 단위의 인식을 통해 불필요한 형태론적 모호성을 제거한다. 둘째, 구문 해석기 전처리 과정의 역할이다. 구문 해석의 효율을 위해 부분적인 어절 결합을 이용하는 방법[7]과 유사하게 기능적 단위의 형태소열을 구문 해석전에 결합하여 불필요한 구문 해석 과정을 줄인다. 또한 구문 형태소 생성기가 형태소 해석기와 독립적인 모듈로 설계되었기 때문에 인터페이스 부분만 수정한다면 다양한 형태소 해석기에 쉽게 결합할 수 있다. 구문형태소 생성기의 결과를 바탕으로 구문 해석기는 구문 형태소 단위의 입력을 파싱함으로써 불필요하게 발생하는 구문 트리를 미리 제거할 수 있다. 실제 시스템에서 나타나는 구문 형태소 자질 생성의 결과는 아래와 같은 형태를 가진다⁴⁾.

표 7 시스템의 자질 생성 결과

“잡고 싶다” (@VP (form 잡고_싶다) (root 잡) (cat VP) (subcat pvgt) (modal hope) (eform dec)) “먹을 수 있다” (@VP (form 먹을_수_있다) (root 먹) (cat VV) (subcat pvgt) (modal able) (eform dec)) “컴퓨터에 대하여” (@NP (form 컴퓨터에_대하여) (root 컴퓨터) (cat NP) (subcat ncn) (iform to))
--

4) [표 7]에서 나타나는 자질값인 pvgt는 transitive verb를 나타내며, dec는 declarative를 나타낸다. 또한 eform은 ending form,iform은 particle form을 나타낸다.

3.2 구문 분석을 위한 품사의 계층적 표현

품사 태그의 수가 많을수록 문법적 행동에 대한 구분이 명료해지고 좀더 자세하게 단어의 의미를 파악할 수 있는 반면, 한 단어에 대해 많은 품사를 가지게 되어 다음 단계인 태깅을 필요로 한다. 또한 이렇게 많은 품사 태그를 이용하여 구문 해석을 하면 구구조를 명확히 기술하기 위한 문법의 크기가 크게 증가한다. [13, 14]에서는 2654개의 구구조 규칙을 이용하여 한국어 문장을 해석했다.

부분 자유 어순을 가지는 한국어를 구구조 문법으로 기술할 경우, 많은 구구조 규칙이 필요하다. 따라서 국어 정보베이스[4]의 대분류 품사를 이용하여 구구조 제약을 하고, 자세한 품사정보인 소분류 품사는 제약 자질로 표현하여 문법내의 단일화 과정에서 제약하는 방법을 이용하면 형태소 해석의 결과도 줄일 수 있고, 소분류가 가지는 장점을 수용할 수 있다. 어절 w 가 최대 두 개의 형태소 결합으로 이루어졌다고 가정할 때, 어절 w 는 최대 $\sum_x \sum_y w_x w_y$ 의 분석 결과를 가진다(x 와 y 는 각각 어절의 전반부와 후반부를 이루는 형태소가 불규칙 현상이나 음운 현상에 의해 하나 이상의 형태를 가질 때 그 수를 말한다). 각 w_x 와 w_y 가 여러 품사를 가진다면 이의 형태소 분석 결과는 [그림 7]의 (1)과 같다. 소분류 품사를 구문 분석을 위한 구구조 규칙의 기본 단위로 하는 방법(1)과 본 논문에서 사용하는 표현 방법 (2)은 [그림 7]과 같다.

$$\begin{aligned} \text{소분류 품사태그 이용: } & \sum_x \sum_i w_x / \hat{p}_i + \sum_y \sum_j w_y / \hat{p}_j & (1) \\ \text{대분류 품사태그 이용: } & \sum_x w_x / P. \hat{p}_1 \dots \hat{p}_l + \sum_y w_y / P. \hat{p}_1 \dots \hat{p}_m & (2) \end{aligned}$$

($w_x, w_y, \hat{p}_i, \hat{p}_j$: 형태소
 i, j : w_x, w_y 가 가질 수 있는 소분류 품사수
 $\hat{p}_i, \hat{p}_j, \hat{p}_l, \hat{p}_m$: 소분류 품사태그, P : 대분류 품사태그)

그림 7 형태소 해석 결과의 표현

표 8 어절 “이는”에 대한 형태소 해석 결과

일/pvg+는/etm 일/paa+는/etm 이/pvg+는/etm 이/ncn+는/jxc 이/npd+는/jxc 이/nnc+는/jxc 이/nbn+는/jxc 이/ncn+는/jxt 이/npd+는/jxt 이/nnc+는/jxt 이/nbn+는/jxt

일/v.pvg.paa+는/e.etm
 이/v.pvg+는/e.etm
 이/n.ncn.npd.nnc.nbn+는/j.jxc.jxt

예를 들면, 어절 “이는”에 대해 소분류 품사 태그를 적용하여 형태소 해석하면 [표 8]과 같다.

이를 대분류 품사를 기준으로 하고, 소분류 품사를 자질의 형태로 표현하면 아래와 같다. 이는 [표 8]에서 나타나는 모든 정보를 포함하면서 형태소 결과를 줄인 것이다.

3.3 구문 형태소를 이용한 형태소 모호성 축소

아래는 가) 문장에 대해 품사를 계층적으로 표현하고 구문 형태소 단위의 결합을 적용하지 않은 형태소 해석 결과이다. [그림 8]에서 볼 수 있는 것처럼 품사를 계층적으로 표현하였음에도 4가지 모호성이 발생하였다.

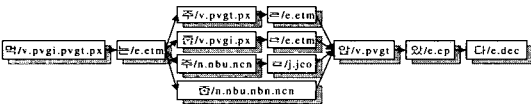


그림 8 가)에 대한 형태소 해석

가) 먹는 줄 알았다

그런데, “줄”이라는 의존 명사는 Type 2형의 구조에 의하면 앞에는 관형형 어미가, 뒤에는 “알다/모르다”와 같은 용언이 반드시 나타나며, 구문적인 역할보다는 양상 자질의 역할을 하는 것을 알 수 있다. 그러므로 이러한 유형의 의존 명사들은 양상 자질로 보는 것이 타당하다. 따라서 위의 문장은 “먹[시도, 과거, 평서]”라는 한가지 유형으로만 분석하는 것이 타당하다. 또한 자질 생성과정에서 세 어절이 하나의 문법 자질인 동사구로 인식되고 구구조 규칙에서 하나의 문법 단위로 인식되기 때문에 구문 해석 과정이 간단해 질 수 있다.

구문형태소 단위의 인식으로 모든 형태소 모호성이 해결되는 것은 아니다. 몇몇 문장의 경우, 구문 해석 전 단계에서 품사 모호성이 해소된 형태소열을 정확히 선택하는 것이 어려운 경우가 있다. 그러므로 허용될 수 있는 형태소 열들은 구문 해석기가 받아들여야만 한다[15]. 즉, 태깅이나 구문형태소 생성 단계에서 해결될 수 없지만 구문 해석이나 의미 해석에 의해 해결될 수 있는 형태론적 모호성은 구문 해석을 위해 허용되어야 한다. 구문 형태소 생성기를 통과한 형태소 해석 결과도 일부분의 형태론적 모호성을 포함하며, 이 중 대부분은 구문 해석 과정에서 모호성이 축소될 수 있다.

예를 들어, “나는” 이라는 어절에 대한 형태소 해석 결과중 “나/pvg+는/etm”과 “날/pvg+는/etm”과 같은 이형 동품사의 경우 태깅 단계에서 적절한 후보를 선택하는 것은 주변 어절의 문맥 정보를 필요로 한다[2]. 그러나 이러한 주변 어절 문맥 정보를 구축하기 위해서는 대량의 품사 태그된 코퍼스가 필요하며, 문맥 정보를 유지하기도 쉽지 않다. 이러한 후보 중 올바른 결과는 구문 해석 과정에서 쉽게 찾아질 수 있다. “나다”는 자동사이며, “날다”는 타동사이기 때문에 주어진 문장을 구문 해석하는 도중에 올바른 선택을 할 수 있다.

또한, “정치가 세상을 어지럽게 하다”를 형태소 해석하면 [그림 9]와 같다. 이때, “정치가”라는 어절에서 “정치+가”와 “정치가”라는 형태소 모호성이 발생하였다. 이중 어떠한 해석 결과를 선택하는가를 형태소 해석이나 태깅을 통해서 정확하게 결정할 수 없다. 그러나 구문 해석 과정에서는 이러한 선택을 구문 정보를 통해 선택할 수 있다.

나) 정치가 세상을 어지럽게 하다.



그림 9 나) 문장의 형태소 해석 결과

이 연구에서는 구문 해석 과정에서 각 규칙에 가중치를 부여함으로써 이러한 문제를 부분적으로 해결할 수 있다. 나) 문장은 두 가지 해석이 모두 가능한데 문장의 완전성(completeness)[16] 관점에서 볼 때 주어를 포함한 문장이 그렇지 않은 문장보다 가중치가 높기 때문에 “정치+가”를 좀 더 가중치를 두어 분석할 수 있다. 따라서 이러한 종류의 형태소 모호성에 대한 해소는 구문 해석까지 연기되어야 한다.

3.4 다중 경로를 허용하는 조건단일화 제약

본 연구에서는 여러 세분류 품사로 나누어지지만 하나의 대분류 품사 범주에 속하는 경우, 단어를 하나의 대분류 품사로 하고 각 세분류 정보를 제약 자질의 형태로 표현하였다. 이에 따른 문제점으로 문장에서 불필요한 제약 자질 정보가 계속 유지된다. [그림 10]에서 “먹”의 보조용언(px) 정보나 “이”의 주격조사(jcs) 정보, “줄”의 보조용언(px) 정보는 불필요한 정보이다.

다) 먹는 것이 좋다



그림 10 문장 다)의 형태소 해석

이러한 정보는 구문 해석 단계에서 구구조의 단일화 도중 구구조의 결합 제약으로 쉽게 제거할 수 있다. 즉, 체언과 용언이 결합할 때 용언의 제약 자질에 보조용언(px)이 나타날 수 없다는 단일화 제약을 표현하고, 조사“이” 등은 “좋다” “아니다” 등과 결합할 때는 자질 제약 정보로 보격조사(jcc)이어야 한다는 것을 조건 단일화 제약[17]으로 표현하면 된다. 이를 통해 불필요한 정보를 구문 해석 단계에서 제약할 수 있다.

4. 실험 및 평가

4.1 구문 형태소의 추출

구문 형태소들을 찾아내기 위해 본 연구에서는 국어 정보베이스[4]의 품사태깅된 말뭉치에서 추출한 어휘화된 바이그램, 트라이그램, 4-gram을 대상으로 상위 빈도를 나타내는 형태소열 중 하나 이상의 어절에 걸쳐 있으며, 내용이 포함되지 않고 기능어만으로 구성된 형태소열을 추출하였다. 아래의 예 중 “때문”, “예정”과 같은 단어는 내용어이지만 출현 빈도가 높고 자립성이 약하기 때문에 포함시켰다. 이들을 대상으로 강한 결합성 여부와 내용에 어떠한 기능적, 의미적 정보를 부가하고 있는지를 기준으로 분류하였다. [표 9]는 약 196,350여 어절에서 얻어진 바이그램, 트라이그램, 4-gram 중 상위 20개를 보여주고 있다.

이중 어휘를 제외하고 인접한 품사들을 다시 출현 빈도수로 정렬하여 인접한 품사들 사이의 강한 관계를 바

표 9 상위 빈도수의 인접한 형태소 열

bigram	trigram	4-gram
고/ecx 일/px	고/ecx 일/px 다/ef	고/ecx 일/px 습니다/ef /sf
에/jca 대하/pvg	고/ecx 일/px 습니다/ef	고/ecx 일/px 다/ef /sf
지/ecx 않/px	하/xsv 고/ecx 일/px	하/xsv 고/ecx 일/px 다/ef
ㄴ/etm 것/nbn	적/xsn 이/jp ㄴ/etm	것/nbn 이/jp 다/ef /sf
ㄹ/etm 것/nbn	에/jca 대하/pvg ㄴ/etm	하/xsv 기/etn 로/jca 하/pvg
ㄷ/etm 것/nbn	고/ecx 일/px ㄷ/etm	기/etn 로/jca 하/pvg 일/ep
에/jca 따르/pvg	기/etn 로/jca 하/pvg	하/xsv 고/ecx 일/px 습니다/ef
기/etm 위하/pvg	ㄹ/etm 것/nbn 으로/jca	ㄹ/etm 것/nbn 으로/jca 보이/pvg
어/ecx 주/px	ㄴ/etm 것/nbn 으로/jca	하/xsv 고/ecx 일/px ㄷ/etm
계/ecx 되/px	에/jca 따르/pvg 아/ecs	하/xsv ㄹ/etm 예정/ncn 이/jp
어/ecx 일/px	에/jca 대하/pvg 어/ecs	이/npd 에/jca 따르/pvg 아/ecs
어/ecx 지/px	ㄹ/etm 것/nbn 이/jp	ㄹ/etm 것/nbn 이/jp 라/ef
기/etm 때문/nbn	기/etn 위하/pvg 아/ecs	되/xsv 고/ecx 일/px 습니다/ef
어/아/ecx 하/px	것/nbn 이/jp 다/ef	고/ecx 일/px ㄷ/etm 것/nbn
어/ecx 오/px	ㄹ/etm 예정/ncn 이/jp	하/xsv 기/etn 위하/pvg 어/ecs
지/ecx 못하/px	하/xsv 기/etn 위하/pvg	ㄷ/etm 것/nbn 이/jp 다/ef
을/jco 위하/pvg	하/xsv ㄹ/etm 것/nbn	기/etn 때문/nbn 이/jp 다/ef
ㄹ/etm 예정/ncn	것/nbn 으로/jca 보이/pvg	기/etn 로/jca 하/pvg 일/ep
을/jco 통하/pvg	ㄴ/etm 것/nbn 이/jp	어/ecx 일/px 습니다/ef /sf
어/ecx 보/px	기/etn 때문/nbn 에/jca	에/jca 대하/pvg 어서/ecs ㄷ/ep

탕으로 분류하여 구문 형태소들을 정리하였다. 이들 중 강한 결합을 나타내는 품사열은 “<ecx> px>”, “<etm> <nbn>”, “<jca> <pvg>”, “<jco> <pvg>” 등이 있었다.

4.2 구문 형태소를 이용한 형태소 모호성 축소 실험

국어정보베이스[4]에서 어절 “수”를 포함하는 어절은 8224번 나타났는데, 이중 8177개가 “nbn”으로 태깅되었으며, 8번은 태깅 오류였다. 다른 39개는 고유명사(nq)나 보통명사, 수사였다. 그러나 만일 “수”가 의존명사로 태깅되었다면 어절 “수”에 후행하는 단어는 반드시 “일/pvg” 또는 “없/pvg”이었다. 따라서 “ㄹ 수 있/없”은 항상 하나의 구문 형태소로 간주해도 된다. 만일 이를 하나의 형태소로 간주하지 않는다면 이 어절들에 대해 태깅을 하지 않을 경우, 12개의 형태론적 모호성이 발생한다. 따라서 이러한 어절들을 구문 형태소 단위로 묶으면 구문 해석 단계에서의 부담은 크게 줄어든다.

또한, “대해”를 포함한 어절은 743번 나타났는데, 이중 45개의 태깅된 결과는 오류 또는 선행 어절에 조사 “에”가 포함되지 않았으며, 698개의 결과가 “-에 대해”의 제대로 된 결과를 가지고 있었다. 즉, “-에 대해”가 문장에 나타난다면 반드시 하나의 품사 태그 결과인 “-에/jca + 대하/pvg + 어/ecs”를 나타내고 있다. 따라서 이러한 어절은 굳이 분절적으로 나누어 분석할 필요가 없으며, 구문적으로 조사로 간주하여도 문장의 분석이나 의미구조의 생성에서도 영향을 미치지 않는다.

일반 형태소 해석 방법과 비교할 때, 본 연구에서는 [표 10]과 같은 결과를 얻었다. 여기에서 A1은 비교 어절에 대한 일반 형태소 해석기를 적용한 결과이다. A2는 소분류 품사 태그를 제약 자질로 표현한 것이며, A3은 A2에 구문 형태소 단위를 적용한 것이다. 구문 형태소 단위를 사용하지 않고 형태소 해석 후 태깅을 하는 방법과 비교하여 보면, 태깅의 경우 보통 95-97%의 정확률을 보이기 때문에 앞에서 의존 명사 “수”나 용언 “대해”의 경우처럼 잘못된 태깅 결과를 보일 수 있다. 그러나 구문 형태소를 이용할 경우, 결과가 어절에 대해

표 10 형태소 해석 결과의 비교

어절	일반 형태소 분석기 이용 (A1)	대분류 품사 태그 이용 (A2)	구문 형태소와 대분류 품사태그 이용(A3)
서울에 대해	36	6	1
서울에 관해	14	3	1
서울을 위해	8	4	1
서울로 인해	10	3	1

반드시 결정적이지는 않지만 형태론적 모호성의 수가 크게 감소함을 알 수 있었다.

구문 형태소열에 대한 형태소 해석 결과는 [표 11]과 같다. 구문 형태소에 내용 형태소를 결합하여 부분적인 어절을 형성한 후 형태소 해석한 결과이다. 각 어절별 평균 형태소 모호성 수는 2어절일 경우에 개개 어절의 형태소 모호성의 수를 나타낸 것이다.

표 11 구문 형태소를 개별적 형태소로 간주할 경우 형태소 수

	Type 1 형 (2어절)	Type 2 형 (2어절)	Type 2 형 (3어절)	Type 3 형 (2어절)
실험 문장 수	22	11	29	24
각 어절별 평균 모호성 수	7.45, 14.60	2.72, 24.36	3.20, 4.90, 7.45	3.96, 10.04

이러한 형태소열을 구문 형태소로 간주하면, 구문 해석의 입력은 하나의 문법 형태소만 가지므로 구문 해석의 효율성을 기대할 수 있다.

초등학교 6학년 사회교과서를 바탕으로 200문장을 구문 분석한 결과, Type 1이나 Type 2, 2'형 구문 형태소는 182번이 나타났고 Type 3, 3'형 구문 형태소는 24번 발생하였다. 실험 문장은 문장 당 평균 10.91개의 어절을 가지고 있는데, 양상 자질을 나타내는 구문형태소가 예상보다 많이 나타나는 이유는 사회 교과서의 성격상 “-해 보자”와 같은 청유형 등이 많이 사용되었기 때문이다. 이 문장에 대해 일반 형태소를 이용할 경우, 한 문장에 평균 41.30개의 형태소가, 구문 형태소 단위의 결합을 이용할 경우 문장 당 20.38개의 형태소를 얻어 약 51%의 형태소 모호성 축소 결과를 얻었다. 또한 단순히 품사 태그셋을 제약 자질구조로 표현한 방법으로 구문 해석 할 경우, 한 문장당 3.83개의 구문 트리가 생성되었고, 구문 형태소 자질 생성기를 이용한 경우, 2.16개의 구문트리가 생성되었다. 구문형태소 단위의 구문 분석이 아닌 경우, 보조용언의 처리나 의존 명사를 위한 구구조 규칙이 39개 사용되었지만 구문 형태소를 이용할 경우, 이러한 규칙은 필요가 없었다. 이를 정리하면

표 12 초등학교 사회교과서 분석 결과

	구문형태소 단위 적용 안함	구문형태소 단위 적용
한 문장당 평균 형태소 수	41.30	20.38
한 문장당 평균 파스 트리 수	3.83	2.16

[표 12]와 같다.

4.3 구문 형태소 단위의 인식의 고려 사항

구문 형태소 단위의 형태소 인식으로 발생하는 문제점으로 용언과 이에 부가되는 양상 자질의 주체가 서로 다른 경우가 있다. 예를 들어,

“나는 영화가 밥을 먹는 줄 몰랐다”

라는 문장에서 ‘추측’의 양상을 나타내는 “-ㄴ 줄 모르”와 “먹다”의 주체가 다르다. 따라서 이들을 하나의 단위로 결합할 경우, 의미 해석에서 전혀 다른 결과를 만들어 낸다. 이러한 구문 형태소로 “-게 하”, “-게 만들-”, “-지 말-”, “-는 줄(은/도/만) 모르-/알-”, “-ㄴ 리(가/도) 없-”, “-을 모양이-” 등이 있다. 이런 경우는 전체 구문 형태소에서 차지하는 비중이 작기 때문에 구문 형태소 단위의 인식으로 모호성 제거 후 의미 해석 단계에서 이를 해결하거나 구문 형태소 인식 단계에서 이런 형태소를 구문 형태소로 묶지 않도록 해야 한다.

5. 결 론

구문 해석을 어렵게 하는 대표적인 원인은 형태소 해석의 파생성에 있다. 특히 한국어나 일본어와 같은 교착어의 특성을 가지는 언어일수록 이러한 현상이 빈번하게 발생한다. 본 연구에서는 한국어에서 형태소 파생성을 유발하는 복합 동사구와 의존 명사 등을 포함하는 어절에 대해 구문 형태소 단위의 처리를 제안하였다. 기존 연구에서는 태깅 시스템을 이용하여 이러한 형태소 모호성을 해결하고자 하였지만 이 방법이 반드시 100% 올바른 결과를 찾아내는 것은 아니다. 본 논문에서는 구문 형태소를 구문 해석의 기본 단위로 묶어주는 구문 형태소 생성기를 이용하여 구문 해석에 도움을 주는 방법을 보였다. 구문 형태소 생성기는 강한 결합을 가지는 기능 형태소들을 하나의 형태소로 간주하며, 이들 사이의 결합 관계를 이용하여 형태소 모호성을 축소할 수 있다. 또한 구문 해석의 입력 단위로 구문 형태소 단위를 사용함으로써 구문 해석 과정을 간략화시킬 수 있었다.

본 논문에서 제안한 방법은 기본적인 태거의 역할을 상당부분 수행하기 때문에 태깅 시스템이 없어도 만족할 만한 구문 해석 성능을 기대할 수 있으며, 구문 해석 단계에서 자질 정보와 구문 정보를 이용하여 가능한 올바른 구문 트리를 생성할 수 있었다. 그러나 몇몇 구문 형태소의 경우, 항상 양상 자질로 쓰이지는 않는 것을 발견하였다. 이는 발생 빈도가 낮지만 이에 대한 연구가

진행되어야 하며, 구문 형태소들을 좀 더 확장할 수 있도록 다양한 한국어 기능 어절에 대한 분석이 진행되어야 한다.

참 고 문 헌

[1] Eugene Charniak, Curtis Hendrickson, Neil Jacobson, Mike Perkowitz, "Equations for Part-Of-Speech Tagging," Proc. of the 11th National Conference on Artificial Intelligence (AAAI), pp.784-789, 1993

[2] 임희석, 언어 지식과 통계 정보를 이용한 한국어 품사 태깅 모델, 고려대학교 대학원 컴퓨터학과 박사학위 논문, 1997

[3] 고영근, 국어 형태론 연구, 서울대학교 출판부, 1989

[4] KIBS : Korean Information Base System, <http://kibs.kaist.ac.kr/>

[5] 이희자, "현대 국어 관용구의 결합 관계 고찰", 제6회 한글 및 한국어 정보처리 학술대회, p.333-352, 1994

[6] 강승식, 음절 정보와 복수어 단위 정보를 이용한 한국어 형태소 분석, 서울대학교 대학원 컴퓨터공학과 박사학위 논문, 1993

[7] 김창제, 정친영, 김영훈, 서영훈, "부분적인 어절 결합을 이용한 효율적인 한국어 구문 분석기", 제22회 정보과학회 가을 학술발표 논문집, pp. 597-600, 1995

[8] 박상규, 정창민, 조준모, 이상조, "최장 묶음을 이용한 효과적인 한국어 구문분석기", 제22회 정보과학회 춘계 학술발표 논문집, Vol. 21, No. 1, pp. 961-964, 1995

[9] 안미정, 옥철영, "한국어 구문 구조 분석을 위한 복수 동사 처리", 제21회 정보과학회 추계 학술발표 논문집, Vol. 21, No. 2, pp. 625-628, 1994

[10] 송연정, 배우정, 이기오, 이용석, "형태소 분석의 자질 구조 생성에 관한 연구", 제21회 정보과학회 춘계 학술발표 논문집, pp. 817-820, 1994

[11] 박혜준, 윤준태, 송만석, "말뭉치 품사 꼬리달기 시스템", 제21회 정보과학회 춘계 학술발표 논문집, pp. 829-832, 1994

[12] 김재훈, 오류-보정 기법을 이용한 어휘 모호성 해소, 한국과학기술원 전산학과 대학원 박사학위 논문, 1996

[13] K. J. Lee, J. H. Kim, G. Ch. Kim, "Probabilistic Language Model for Analyzing Korean Sentences," Proceedings of the 17th International Conference on Computer Processing of Oriental Languages (ICCPOL'97), HongKong, pp. 392-395, 1997

[14] 이공주, 언어 특성에 기반한 한국어의 확률적 구문분석, 한국과학기술원 전산학과 대학원 박사학위 논문, 1998

[15] G. O. Lee, et al., "Multi-Path LR Parsing for Nonsegmental Words Using Interactive Strategy," The 3rd Pacific Rim International Conference on AI, Vol. 2, pp. 668-672, 1994

[16] Bresnan, J.(Ed.), The Mental Representation of Grammatical Relations, Cambridge Mass.: MIT Press, 1982

[17] 양승원, 박영진, 이용석, "조건 단일화 기반 PATTERN을 이용한 한국어 구문분석", 한국정보과학회 논문지, 제 22권, 4호, pp. 653-662, 1995



황 이 규

1993년 전북대학교 전자계산학과 졸업(이학사). 1995년 전북대학교 전산통계학과 졸업(이학석사). 1995년 ~ 현재 전북대학교 전산통계학과 박사과정. 관심분야는 한국어 정보 처리, 인공지능



이 현 영

1991년 전북대학교 전산통계학과 졸업(이학사). 1996년 전북대학교 전산통계학과 졸업(이학석사). 1996년 ~ 현재 전북대학교 전산통계학과 박사과정. 관심분야는 한국어 정보처리, 정보 검색



이 용 석

1977년 서울대학교 공학사(전자공학). 1979년 한국과학기술원 공학석사(전산학). 1995년 일본 국립도쿠시마대학교 공학박사(지능정보). 1979년 한국표준연구소 선임연구원. 1983년 ~ 현재 전북대학교 컴퓨터공학과 교수. 관심분야는 기계번역, 한국어 정보처리, 정보 검색, 음성 처리