

# 서술형 문항 채점을 위한 복합문 구문의미분석 시스템에 대한 연구

강원석<sup>†</sup>

## 요 약

서술형 문항은 수준 높은 사고능력 평가에 적합하나 채점이 쉽지 않다. 동일한 채점기준을 적용하더라도 채점자마다 서로 다른 채점을 할 수 있으므로 객관적인 채점 시스템이 필요하다. 그렇지만 채점 시스템을 구축하기 위해서는 표현 언어인 한국어 분석이 전제되어야 한다. 특히 서술형 문항의 답변은 단 문형태가 아닌 복문이 내포된 복합 문장으로 표현되기 때문에 복합문의 한국어 분석이 이루어져야 한다. 본 연구는 이와 같은 서술형 문항의 자동 채점을 위해 한국어의 복합문을 구문의미 분석하는 시스템을 개발하고 시스템의 성능을 분석하였다. 이 시스템은 의미속성 사전을 활용하여 대상 구문의 의미적 제한 조건을 검사하고 적합한 피수식 대상을 선별하여 구문의미분석을 실행한다. 실험 데이터에 대해 93%의 정확률을 얻어 제안한 시스템이 효과 있음을 알 수 있었다. 본 연구의 시스템은 서술형 문항 채점에 활용 할 수 있고 한국어 처리의 응용영역에 활용할 수 있을 것으로 기대된다.

주제어 : 서술형 채점, 구문의미분석, 복문

## Research on the Syntactic-Semantic Analysis System on Compound Sentence for Descriptive-type Grading

WonSeog Kang<sup>†</sup>

### ABSTRACT

The descriptive-type question is appropriate for deep thinking ability evaluation, but it is not easy to grade. Since, even though same grading criterion, the graders produce different scores, we need the objective evaluation system. However, the system needs the Korean analysis. As the descriptive-type answering is described with the compound sentence, the system has to analyze the compound sentence. This paper develops the Korean syntactic-semantic analysis system for compound sentence and evaluates performance of the system. This system selects the modifier of the word phrase using syntactic-semantic constraint and semantic dictionary. The 93% accurate rate shows that the system is effective. This system will be utilized in descriptive-type grading and Korean processing.

**Keywords** : Descriptive-type Grading, Syntactic-Semantic Analysis, Compound Sentence

<sup>†</sup> 정 회 원: 안동대학교 교수(교신저자)

논문접수: 2018년 8월 30일, 심사완료: 2018년 10월 19일, 게재확정: 2018년 10월 19일

\* 본 논문은 안동대학교 기본연구지원사업에 의하여 연구되었음.

## 1. 서론

서술형 문항은 수준 높은 사고능력 평가에 적합하여 평가 분야에서 그 가치가 높으나 이를 적용하기에는 문제가 많다. 평가자에 따라 평가기준이 다르게 나타날 수 있어 평가의 신뢰성을 얻기 힘들기 때문이다. 또한 객관적인 평가기준이 제시된다고 하더라도 평가의 주관성이 개입되어 같은 답변에 대해 같은 평가점수를 얻기 힘들다. 이를 해결하기 위해서는 사람이 개입되지 않는 객관적인 자동 채점 시스템이 필요하다.

이와 같은 필요성에 근거하여 자동채점 시스템에 대한 연구가 진행되어 왔다. 정상목과 동료들(2005)의 연구는 한국어 문장으로 표현된 답변에서 형태소 해석을 거쳐 단어를 추출하고 이를 벡터로 표현하여 채점을 하였다[1]. 이 연구의 출발은 단어가 서술형 문항의 답변을 대표하고 있다는 점이다. 그렇지만 단어들의 나열이 문항의 답변이 가진 의미를 모두 표현할 수 있다고 보기에 무리가 있다.

조우진과 동료들(2005)는 한글 워드넷을 이용하여 단어에 대한 유의어를 활용하여 채점을 시도하였다[2]. 이 연구는 문항의 답변들이 단어들의 나열이 아닌 개념들의 나열로 표현된 것이라고 보았다. 그렇지만 개념들간의 구조적인 관계는 고려되지 않았다.

서술형 문항은 문항에 따라 여러 가지 유형으로 구분할 수 있지만 그 내용은 유형과 관계없이 사실이나 지식을 묘사하고 있다. 따라서 사실이나 지식을 묘사하는 의미표현을 찾아내고 이를 비교하는 것에 초점을 두어야 한다. 그러한 점에서 강원석(2007)의 연구는 단어를 찾아내는 방법이 아니라 사실이나 지식을 나타내는 의미표현과 유사한 구문구조나 의미구조를 찾아내는 구문의미분석 방법을 적용하였다[3]. 그렇지만 이 연구는 의미 표현에 한계가 있는 단문 표현에 대한 구문의미분석으로 제한하였다.

글을 쓸 때 복잡한 의미와 개념을 표현하기 위해서는 단문이 아닌 복문을 많이 사용한다. 서술형 문항의 답변은 그 목적에 맞게 단순한 사실보다는 사실간의 연결 관계나 인과 관계 등의 차원 높은 표현을 많이 사용하기 때문이다. 따라서 본

연구는 이와 같이 서술형 문항의 답변에 많이 사용되는 복문을 처리하는 구문의미분석 시스템을 설계, 개발하였다.

본 연구의 구문의미분석 시스템은 지식표현을 찾아내고 이를 비교하는 방법에 근거하여 지식표현과 유사한 구문의미트리를 생성한다. 구문의미트리는 구문구조임과 동시에 의미적 역할과 개념을 반영한 의미구조로 정의된다. 생성된 구문의미트리는 정답과의 유사성을 찾아내기 위해 비교를 하게 된다. 본 연구에서는 구문의미트리의 비교가 쉽게 트리의 형식과 구조를 단순하게 정의하였다.

2장은 서술형 문항채점을 위한 구문의미분석의 관련 연구를 기술하고 3장은 복합문 구문의미분석 시스템을 기술한다. 4장은 시스템의 실험과 결과 분석을 기술하고 5장은 결론과 논의사항을 기술한다.

## 2. 관련 연구

서술형 문항의 답변 표현을 처리하기 위한 한국어 분석에 대한 관련 연구를 살펴본다.

정상목과 동료들(2005)은 한국어 형태소 해석을 이용하여 추출한 단어들을 근거로 LSA 방식을 적용하였다[1]. 조우진과 동료들(2005)은 대량의 말뭉치를 이용하여 의미커널을 구축하고 답안을 채점할 때 구축된 의미커널과 한글 워드넷을 이용하여 답안을 채점하였다[2]. 박일남과 동료들(2013)은 단어와 구 수준의 정답 템플릿을 설계하고 이를 토대로 답안과 일치여부를 검사하였다[4]. 이 연구들은 형태소 해석을 기초로 단어를 추출하거나 단어와 단어의 유의어를 활용하여 개념 중심의 일치여부를 검사하였다. 그렇지만 형태소와 의미사전 활용에 그치고 있어 단어나 개념간의 구조적 연결관계인 구문구조 등을 이용하지 못하고 있다.

박소영과 동료들(2004)은 구문분석에서의 중의성을 해결하기 위해 구문태그와 기능태그, 단어 정보 등을 활용한 확률적 방법을 제시하였다[5]. 박의규와 나동열(2006)은 의존명사와 관련된 구뭉음을 통하여 문장의 복잡도를 낮춘 구문분석을 시도하였다[6]. 안광모와 서영훈(2011)은 단일어절 규칙과 구뭉음 규칙, 동사하위범주화의 전처리 과

정을 통해 중의성 해소를 시도하였다[7]. 박용욱과 권혁철(2013)은 구간 분할과 동사와 격정보에 해당하는 논항정보를 이용하여 중의성을 축소하는 방법을 적용하였다[8]. 이 연구들은 구문분석에서 발생하는 구문적 애매성의 문제를 해결하는 연구로 범용의 연구로 서술형 문항 평가에 그대로 적용하기 어렵다. 또한 이 연구들은 부분적으로 활용할 수 있으나 의미적 분석이 이루어지지 않았다. 따라서 사실이나 지식을 찾아 활용해야 하는 서술형 문항 채점에 적용하기 어렵다.

박용욱(2009)은 구간분할 방법과 통합을 이용한 의존문법 구문분석기를 개발하였다[9]. 남웅과 동료들(2014)은 형태소 해석의 오류를 줄이기 위해 문장 전체를 대상으로 하는 의존 구문분석기를 개발하였다[10]. 이 연구들은 의존 구문분석기를 개발한 것으로 응용영역에 적용할 수 있는 시스템이다. 그렇지만 구문구조 분석에 맞추어져 있어 의미적 분석과정과 비교가 필요하다.

이공주와 김재훈(2003)은 규칙에 근거하여 한국어의 부분 구문분석기를 제안하였다[11]. 강원석(2007)은 서술형 평가를 위해 구문의미분석기를 개발하였다[3]. 그렇지만 이 연구들은 복문이 아닌 단문에 초점을 맞춘 분석으로 서술형 평가에 적용하기에 한계가 있다.

강원석과 황도삼(2004)은 질의응답을 위해 질의 유형에 대한 한국어 구문해석을 하였다[12]. 이 연구는 질의유형에 한정되어 있어 한계가 있다. 강원석과 동료들(2014)은 유사문서를 판별하기 위해 구문의미분석을 적용하였다[13]. 이 연구는 유사문서 판별영역이라 서술형 평가영역과 다르고 단문의 구문의미분석을 적용하였다. 박천음과 이창기(2017)는 멀티태스크 기반의 포인터 네트워크를 이용하여 한국어 의존 구문분석을 실시하였다[14]. 이 연구는 학습기능을 이용하였지만 의존구문구조를 판별한 것으로 의미적 분석과 비교가 필요하다.

유혜원(2009)은 복문구조에 대한 한국어 구문분석 방법론을 연구하였다[15]. 이 연구는 서술형 평가에 적합한 복문에 대한 연구를 하였지만 이론적 연구로 시스템이 개발되지 않아 서술형 평가에 적용할 수가 없다.

### 3. 서술형 문항 채점을 위한 복합문 구문의미해석 시스템

#### 3.1 시스템 설계 원칙

서술형 문항 채점을 위한 복합문 구문의미 해석 시스템은 다음과 같은 원칙으로 설계되었다.

원칙 1) 사실이나 지식을 표현하는 재귀적인 의미표현으로 구문의미트리를 정의하고 사용한다.

이것은 서술형 문항 평가의 원래 목적인 수준 높은 사고능력 평가를 달성하기 위해 답변의 내용을 표현하는 의미표현을 찾아내고 비교하기 위한 것이다. 이를 위해 본 시스템은 의미표현과 유사한 구문의미트리를 정의하고 사용한다.

그 형식은 <표 1>과 같다.

<표 1> 트리 형식과 예

트리 형식	( ‘노드의 역할’ ‘노드의 헤드단어’ ‘이 노드를 수식하는 복합명사들’ ‘노드의 자식트리들’ )
예 문장	이 프로그램은 제어 클래스와 자연수 합 클래스로 구성된다.
예 문장의 트리	<p>( FINEN 구성 ( TOP 프로그램 ( BLA 이 ) ) )                  ( COMP_WITH 클래스 합 자연수 ( JUNC 클래스 제어 ) ) )</p>

하나의 트리는 ( )로 표현된다. ‘노드의 역할’은 그 노드의 기능에 해당하는 것으로 수식대상과의 의미적 역할관계를 나타낸다. ‘노드의 헤드단어’는 그 노드의 주 단어를 의미한다. 예를 보면 ‘FINEN’은 노드의 역할을 나타내는 것으로 문장을 종결짓는 서술기능임을 나타내고 있고 ‘구성’은 서술어에 해당하는 단어이다. ( ‘ TOP 프로그램 ( BLA 이 ) ’ )는 노드의 자식트리들 중 하나로서 부모노드와 TOP 의미관계를 맺고 있고, 이 자식은 다시 재귀적인 트리로 정의되어 있다. ( ‘ COMP\_WITH 클래스 합 자연수 ( JUNC 클래스 제어 ) ’ )도 자식트리 중 하나이다. 이 자식트리의 헤드단어는 클래스이고 합과 자연수는 ‘이 노드를 수식하는 복합명사들’에 해당한다.

지식표현의 한 종류인 의미망의 경우 객체의

개념을 노드로 표현하고 개념간의 의미관계를 간선으로 표현한다. 본 연구에서는 어절에 해당하는 주단어의 개념을 노드로 표현하고 개념간의 의미관계를 노드의 역할관계로 표현하였다. 그 의미적 역할은 76개로 정의되었고 그 일부는 <표 2>와 같다.

<표 2> 의미적 역할관계 일부

의미적역할	설명	의미적역할	설명
SUBJ	주격	SUCOCAUSE	인과적
OBJ	목적격	SUCOPURP	의도적
COMP	보격	COMP_WITH	구성적
ADNM	관형격	SOUR_FR	원시적
TOP	화제격	STAT_AS	자격적
COCO	대등의미적	AND	순접
COCOENUM	열거적	BUT	역접
COCOSEL	선택적	...	...
SUCO	종속의미적		

원칙 2) 의미표현 비교에 용이하도록 구문의미트리의 형식을 체계적이고 단순하게 정의한다.

서술형 평가의 영역에서는 의미표현을 찾아내고 이를 비교하는 과정이 필수적이다. 의미관계를 비교할 때 정확히 일치하는 경우 이외에도 유사한 관계를 찾아내어야 할 경우가 있다. 이를 위해 본 연구에서는 상하위관계를 이용할 수 있도록 역할관계를 상하위 체계로 구성하였다. 또한 역할관계를 연관된 것끼리 군집화하여 비교에 용이하도록 정의하였다.

<표 3> 의미적 역할관계의 상하위 관계의 예

상위 역할	하위 역할
COCO	COCOENUM, COCOSEL, COCOCONT, ...
SUCO	SUCOCAUSE, SUCOCOND, SUCOPURP, ...
DEST	DEST_TL_BL, SEDT_TL_TO, DEST_TO
LOC	LOC_AT, LOC_IN, LOC_ON
...	...

<표 4> 의미적 역할관계의 군집

군집명	의미적 역할관계
CASE	SUBJ, OBJ, ...
ENDING	COCO, SUCO, AUCO, QUCO, ...
CONJ	AND, BUT, TRAN, ADD, CAUSE,...

그리고 구문의미트리를 비교할 때 트리의 깊이가 깊게 될 경우 비교에 어려움이 있다. 같은 의미표현을 할 때 가능한한 트리의 깊이를 낮게 하여 비교를 쉽게 하도록 정의하는 것이 필요하다. 본 연구에서는 복합명사의 경우 앞에 있는 명사를 뒤에 있는 명사의 하위노드로 구조화하지 않고 뒤에 있는 명사노드에 앞 명사를 포함시키는 방법을 사용하여 트리의 깊이를 낮게 하였다. 트

리 형식의 세 번째 항목 ‘이 노드를 수식하는 복합명사들’이 포함된 앞 명사들에 해당한다.

원칙 3) 어절을 하나의 노드이자 트리로 정의하고 어절의 순서적인 포리스트에서 시작하여 하나의 트리가 되도록 합병한다. 이때 적용할 원칙은 no crossing, locality principle, 지배소 후위, 지배소 유일이다.

문장을 분석하여 구문의미트리를 생성하기 위해서는 단계적인 방법을 취한다. 다른 트리에 영향을 주지 않는 독립적인 트리부터 합병을 시작하여 마지막으로 서술어에 해당하는 트리를 합병한다. 트리를 구성할 때 no-crossing 원칙을 고려하여 트리간의 교차가 일어나지 않도록 구성한다. 그리고 locality principle를 적용하여 기본적으로 가장 가까이 있는 트리와의 연결관계를 고려한다. 지배소는 한국어 문법상 후위에 나타나므로 이를 고려하고 지배소는 유일하므로 하나의 지배소를 찾아 구조화하도록 한다.

원칙 4) 구문의미분석을 위하여 (의존소, 의존소역할, 지배소)의 트리플과 그에 대한 의미속성의 정보를 이용한다.

구문의미분석할 때 사용하는 원칙과 함께 노드간의 의미적 역할 관계를 식별하거나 구조화 대상을 찾을 때 본 시스템은 트리플 정보와 그에 따른 의미속성을 이용한다. 트리플은 의존소, 의존소역할, 지배소로 구성된다. 의존소는 대상이 되는 어절의 주단어 정보, 그 어절의 어절 유형으로 구성된다. 어절 유형은 어절이 어떠한 종류의 어절인지를 구분하는 것으로 구문의미분석의 첫 번째 단계인 어절처리 단계를 거쳐 결정이 된다. 어절 유형의 종류는 <표 5>와 같다.

<표 5> 어절유형의 종류

어절 유형	설명
HNOUN	명사어절 : 명사(+조사)
HVERB	동사어절 : 동사/형용사+어미
HBEVERB	서술격조사어절 : 명사+서술격조사+어미
HNXS	하다어절 : 명사+파생접미사(하)+어미
HPTN	동명사어절: 동사/형용사+명사형어미(+조사)
HMA	부사어절 : 부사
HMM	관형사어절 : 관형사

의존소 역할은 의존소가 어떠한 의미적 역할을 하는지를 말한다. 역시 어절처리 단계에서 어절의 유형과 함께 의존소의 의미적 역할이 1차로 결정

된다. 역할의 종류는 <표 2>의 의미적 관계와 같다.

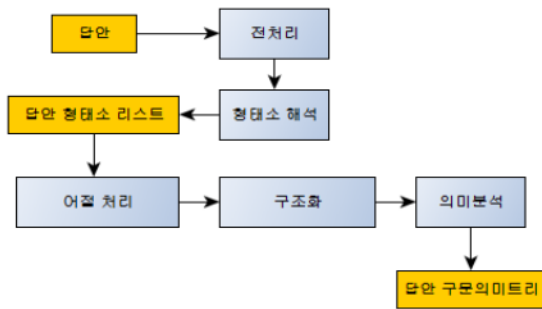
지배소는 지배소에 해당하는 어절의 주 단어 정보와 그 어절의 어절 유형으로 구성된다. 의존소와 지배소의 단어에 대한 의미적 정보는 의미속성 추출기를 이용한다. 본 연구에서 사용한 의미속성과 의미속성 추출기는 [13]을 이용하였다. 그 의미속성은 크게 3그룹으로 구분되고 속성간의 상하위 관계가 정의된다. 일부의 예는 <표 6>과 같다.

<표 6> 의미속성 상하위관계의 예

thing physical-thing animate-thing ... inanimate-thing ... abstract-thing ...	event composition intellectual-behavior ...
feature situation animate-feature location ...	domain philosophy ...

### 3.2 구문의미분석 시스템

구문의미분석 시스템의 흐름은 [그림 1]과 같다.



[그림 1] 구문의미분석 시스템의 흐름

전체 시스템은 전처리, 형태소 해석을 거쳐 형태소 리스트를 생성한다. 이를 입력으로 어절처리, 구조화, 의미분석을 실행하여 구문의미트리를 생성한다.

전처리는 입력문장의 특수기호, 따옴표 등의 처리를 하여 구문의미해석의 처리 대상을 표준화한다.

형태소 해석은 [16]을 이용한다. 예 문장에 대한 처리 결과는 <표 7>과 같이 형태소/형태소태그의

형식으로 구성된다.

<표 7> 형태소 해석의 예

예 문장	단위변환 클래스는 단위변환에 필요한 값을 매개변수로 받아 변환한 후 결과를 반환하는 메소드를 구성한다.
형태소 해석 결과 리스트	단위/ncn+변환/ncn 클래스/ncn+는/jxt 단위/ncn+변환/ncn+에/jca 필요/ncps+하/xsm+ㄴ/etm 값/ncn+을/jco 매개/ncn+변수/ncn+로/jca 받/pvg+아/ecs 변환/ncn+하/xsm+ㄴ/etm 후/nbn 결과/ncn+를/jco 반환/ncpa+하/xsv+는/etm 메소드/ncn+를/jco 구성/ncpa+하/xsv+ㄴ다/ef+./sf

#### 3.2.1 어절처리

다음으로 어절처리, 구조화, 의미분석의 단계를 거친다. 어절처리는 대상이 되는 어절의 유형과 어절의 의미적 역할을 1차로 결정짓는다. 그 알고리즘은 다음과 같다.

```

procedure 어절처리 {
    어절의 포리스트 초기화;
    while(모든 어절트리) {
        어절의 주단어 찾기;
        주단어와 첨가어에 근거한 어절유형 정하기;
    }
    while(모든 어절트리) {
        조사나 어미찾기;
        if(찾은 경우)
            조사나 어미에 근거한 1차 의미적역할 정하기;
        else
            어절유형에 근거한 1차 의미적 역할정하기;
    }
}
    
```

어절처리는 먼저 어절마다 어절의 주 단어와 어절유형을 찾는다. 그 후 어절유형, 어절의 주단어, 어절의 조사나 어미 정보를 근거로 어절의 의미적 역할을 결정짓는다. 예 문장에 대한 처리 결과는 <표 8>과 같다.

<표 8> 어절 처리의 예

예 문장	단위변환 클래스는 단위변환에 필요한 값을 매개변수로 받아 변환한 후 결과를 반환하는 메소드를 구성한다.	
어절처리 결과	( BLA 단위 변환 ) ( TOP 클래스 ) ( ADV 단위 변환 ) ( ADNMEN 필요 ) ( OBJ 값 ) ( ADV 매개 변수 ) ( SUCO 받 )	( ADNMEN 변환 ) ( BLA 후 ) ( OBJ 결과 ) ( ADNMEN 반환 ) ( OBJ 메소드 ) ( FINEN 구성 )

### 3.2.2 구조화

다음은 구조화의 단계이다. 한국어의 특성에 맞추어 원칙 3과 원칙 4를 적용하여 포리스트를 하나의 구문트리로 병합해 간다. 그 순서는 1) 접속사어절 처리, 2) 인용표현 처리, 3) 복합명사 처리, 4) 관형사 처리, 5) 관형격 처리, 6) 부사처리, 7) 격조사어절 처리, 8) 관형절 처리, 9) 시점처리, 10) 기타어절 처리, 11) 용언 처리이다.

#### 1) 접속사 어절 처리

문두의 접속사는 문장간의 연결관계를 명시하는 역할을 하는 것으로 인식하여 문장전체의 서술어에 부착하여 구조화한다. 이 과정의 구조화로 구성된 트리는 SIMSNT로 정의된다. 구문트리의 유형은 <표 9>와 같다.

<표 9> 구문의미트리 유형의 종류

구문의미트리 유형	설명
SIMSNT	단문구조: 하나의 술어를 가진 단문 유형
QUOTATION	인용구조 : 인용 구문유형
COMPNOUN	복합명사구조: 복합명사 유형
DNOUN	명사연결구조: 명사와 명사의 관계로 구성된 유형
CNOUN	절내포명사구조: 하나 이상의 술어를 내포한 명사 유형
COMSNT	복문구조: 둘 이상의 술어를 가진 복문 유형

#### 2) 인용표현 구조화

따옴표의 인용표현을 하나의 구조로 묶는 역할을 한다. 따옴표 사이의 모든 어절을 마지막 어절의 자식트리로 하고 마지막 어절을 루트로 하는 구조화를 진행한다.

#### 3) 복합명사 처리

명사의 앞에서 뒤 명사를 수식하는 명사들은 복합명사라고 할 수 있다. 원칙적인 트리로 표현한다면 앞 명사는 뒤 명사를 수식하므로 앞 명사는 뒤 명사의 자식트리로 복층 트리로 정의된다. 그렇지만 본 연구에서는 원칙 2)에 근거하여 복합명사를 트리 비교하기에 쉬운 단층 구조의 트리로 표현하였다. 그 처리의 결과는 <표 10>과 같다.

<표 10> 복합명사 처리의 예

처리전	( BLA 단위 변환 ) ( TOP 클래스 )
처리후	( TOP 클래스 변환 단위 )

#### 4) 관형사 처리

관형사의 경우 그 관형의 대상이 되는 명사를 찾고 그에 부속되도록 병합한다. 이때 원칙 3의 locality principle에 따라 근접한 명사를 찾아 DNOUN 유형으로 구조화한다.

#### 5) 관형격 처리

관형격 조사 ‘의’에 해당하는 어절의 처리이다. 관형격조사도 바로 근접한 명사를 찾아 DNOUN 유형으로 구조화한다.

#### 6) 부사 처리

부사의 경우 수식대상이 다양하게 나타난다. 그렇지만 본 연구에서는 대부분을 차지하는 용언수식으로 한정하고 이를 구현하였다.

#### 7) 격조사어절 처리

격조사 유형에 따라 제한조건에 맞는 수식대상을 찾고 구조화한다. 그 알고리즘은 다음과 같다.

```

procedure 격조사어절처리 {
  while(모든 격조사어절) {
    switch(격조사어절의미적역할) {
      case SUBJ, VOC, ADV, COMM :
        S_수식동사찾기; S_구조화하기; break;
      case OBJ : O_수식동사찾기; O_구조화하기; break;
      case COMP : C_수식동사찾기; C_구조화하기; break;
      case COMI : CO_수식동사찾기; CO_구조화하기;
        break;
      case TOP : T_수식동사찾기; T_구조화하기; break;
      case JUNC : J_접속명사찾기; J_구조화하기; break;
      case BLA : if(어절내명사있으면) { S_수식동사찾기;
        B_구조화하기;}
      default :
    }
  }
}
    
```

<표 11>는 알고리즘의 수식대상 찾기함수에서 찾을 대상의 제한조건을 표시하였다. 각 함수는 원칙 3의 no-crossing원칙을 준수하도록 하였다. 비관형절인 경우는 관형절이 아닌 조건을 의미한다. 관형절은 대상이 되는 명사를 수식하는 절이다. 문장에서 토픽은 관형절을 설명하기 보다는 문장 전체를 설명하기 위한 것이므로 이를 반영하여 관형절이 아닌 서술어를 찾도록 하였다.

그리고 J\_접속명사찾기의 경우는 대상 어절에 대해 여러 개의 수식대상이 있을 경우 그 중 가장 의미가 유사한 것을 찾는다. 이것은 ‘와’와 같은 접속격 조사의 경우 병렬적으로 나열되는 대상을 찾는 것이다. 예를 든다면 ‘학교의 교훈과 부모의 가훈을 비교한다.’에서 ‘교훈과’ 어절이 구조화될 수 있는 병렬 대상을 찾아야 한다. 그 경

<표 11> 찾기함수의 대상

격조사어절의 지배소 찾기함수	찾고자 하는 어절 유형
S_수식동사찾기	HVERB, HBEVERB, HNXS, HPTN
O_수식동사찾기	HVERB, HNXS, HPTN
C_수식동사찾기	HBEVERB, HVERB
CO_수식동사찾기	HVERB, HNXS
T_수식동사찾기	HVERB, HNXS, HBEVERB, 비관형절
J_접속명사찾기	HNOUN, HBEVERB이면서 의미적 유사성이 가장 높은 것

우 locality principle에 따르면 ‘부모’가 연결될 것이다. 이것은 맞지 않으므로 적합한 것을 찾으려 하기 위하여 교훈, 부모, 가훈의 의미속성을 찾고 이를 비교한다. <표 12>에서 알 수 있듯이 교훈과 가훈의 의미속성 일치도가 더 많으므로 ‘교훈과’ 어절은 ‘가훈은’ 어절과 연결을 한다.

<표 12> 예 단어의 의미속성과 일치여부

항목	내용
예 문장	학교의 교훈과 부모의 가훈을 비교한다.
시스템 결과	( FINEN 비교 ( OBJ 가훈 ( ADNМ 부모 ) ) ( JUNC 교훈 ( ADNМ 학교 ) ) )
교훈의 의미속성	abstract-thing control-behavior domain education event idea intellectual-thing society thing
부모의 의미속성	animal animate-thing feature human physical-thing relation thing
가훈의 의미속성	abstract-thing customs domain idea intellectual-thing society thing
교훈&부모 의미속성	thing
교훈&가훈 의미속성	abstract-thing domain idea intellectual-thing society thing

제시된 예 문장의 격조사 어절 처리 결과는 <표 13>과 같다.

<표 13> 격조사어절 처리의 예

처리후	( ADNМEN 필요 ( ADV 단위 변환 ) ) ( SUCO 받 ( TOP 클래스 변환 단위 ) ( OBJ 값 ) ( ADV 매개 변수 ) ) ( ADNМEN 변환 ) ( BLA 후 ) ( ADNМEN 반환 ( OBJ 결과 ) ) ( FINEN 구성 ( OBJ 메소드 ) )
-----	---

8) 관형절 처리

관형절은 수식대상인 어절인 HNOUN, HBEVERB 유형인 것을 찾아 절내포명사구조인 CNOUN으로 구조화된다. 이 구조화는 근접한 것을 선택한다. 그 결과는 <표 14>와 같다.

<표 14> 관형절 처리의 예

처리후	( SUCO 받 ( TOP 클래스 변환 단위 ) ( OBJ 값 ( ADNМEN 필요 ( ADV 단위 변환 ) ) ) ( ADV 매개 변수 ) ) ( BLA 후 ( ADNМEN 변환 ) ) ( FINEN 구성 ( OBJ 메소드 ( ADNМEN 반환 ( OBJ 결과 ) ) ) )
-----	--

9) 시점 처리

후, 때, 뒤, 경우 등의 시점을 나타내는 어절에 대해 TIME\_AT로 의미적 역할을 변환하고 가장 후위에 있는 HVERB, HNXS인 어절유형을 찾아 구조화한다. 그 결과는 <표 15>와 같다.

<표 15> 시점 처리의 예

처리후	( SUCO 받 ( TOP 클래스 변환 단위 ) ( OBJ 값 ( ADNМEN 필요 ( ADV 단위 변환 ) ) ) ( ADV 매개 변수 ) ) ( FINEN 구성 ( OBJ 메소드 ( ADNМEN 반환 ( OBJ 결과 ) ) ) ( TIME_AT 후 ( ADNМEN 변환 ) ) )
-----	---

10) 기타 어절 처리

나머지 처리되지 않은 어절에 대해 묵시적으로 서술어를 찾고 구조화한다.

11) 용언 처리

이전까지의 처리로 단문 구문의미트리의 포리스트가 된다. 이를 하나의 구문의미트리로 합병한다. 그 처리 알고리즘은 다음과 같다.

```

procedure 용언처리 {
  while(모든 용언어절) {
    switch(용언어절 의미적역할) {
      case COCO :
        COC_수식동사찾기; COC_구조화하기; break;
      case COCOENUM :
        CE_수식동사찾기; CE_구조화하기; break;
      case SUCO :
        COC_수식동사찾기; SU_구조화하기; break;
      case AUCO :
        COC_수식동사찾기; AU_구조화하기; break;
      case QUCO :
        COC_수식동사찾기; QU_구조화하기; break;
      case FINEN : break;
    }
  }
}
    
```

용언은 크게 두 방법으로 구조화 대상을 찾는다. 하나는 대등적 연결의 경우이고 다른 하나는 종속적 연결의 경우이다. 대등적인 연결의 경우는 술어의 의미가 대등한 관계에 놓여있는 것을 찾아야 하고 종속적인 연결은 의존적인 연결에 놓여 있는 것을 찾는다.

<표 16> 찾기함수의 대상

용언어절의 지배소 찾기함수	찾고자 하는 어절 유형
COC_수식동사 찾기	HVERB, HNXS, HBEVERB, 비관형절 (단, 마지막술어이면서 HBEVERB인 것은 이전 술어)
CE_수식동사 찾기	유사한 의미의 HVERB, HNXS(우선적으로 마지막 술어를 선택함, 단, 마지막 술어가 불완전명사와 결합된 것이면 이전 술어)

<표 16>에서 COC\_수식동사찾기는 종속적인 연결을 찾는 함수이고 CE\_수식동사찾기는 대등적인 연결을 찾는 함수이다. 종속적 연결을 찾을 때 비관형절은 관형절이 아닌 것을 의미한다. 관형절은 어떤 명사를 자세히 표현할 때 사용하게 된다. 따라서 용언의 경우 관형절에 연결되기 보다는 일반적 술어에 연결되는 점을 고려하여 관형절을 피하였다. 또 문장의 표현상 일반적으로 두 개 이상 내포된 관형절은 잘 사용하지 않는다. 만약 두 개 이상 사용된다면 그것은 특수한 경우이고 이 경우는 관형절도 수식대상에 포함시키는 규칙을 적용하였다. 대등절의 경우 술어간의 의미적 대등성을 찾기 위해 의미속성 추출기를 이용하여 유사한 술어를 선택하였다. 그렇지만 문장 표현의 특징을 반영하여 우선적으로 마지막에 나오는 술어를 선택하였고 마지막 술어가 불완전 명사와 결합된 경우는 이전 술어를 선택하여 구조화하였다. 용언 처리로 COMSNT유형으로 구조화되고 하나의 트리가 되면 구조화가 종료된다. 예 문장에 대한 처리 결과는 <표 17>과 같다.

<표 17> 용언 처리의 예

처리후	( FINEN 구성 ( OBJ 메소드 ( ADNMEN 반환 ( OBJ 결과 ) ) ) ( TIME_AT 후 ( ADNMEN 변환 ( SUCO 받 ( TOP 클래스 변환 단위 ) ( OBJ 값 ( ADNMEN 필요 ( ADV 단위 변환 ) ) ) ( ADV 매개 변수 ) ) ) ) ) )

3.2.3 의미분석

구조화 다음의 단계는 의미분석의 단계이다. 의미분석은 1차로 분석된 의미적 역할을 보다 세분화된 역할로 분석한다. 이때 사용하는 제한조건은 (의존소, 의존소역할, 지배소)의 트리플과 의미속성으로 [12]의 연구를 토대로 개선한 것이다. 그 내용의 일부는 <표 18>과 같다.

<표 18> 트리플의 제한조건

1차 의미적 역할	의존소	지배소	2차 의미적 역할
ADV(예)	location, structure	positioning, moving	LOC_IN
	time		TIME_IN
	state, organization, situation		STAT_AS
...			

예 문장에 대해 2차 의미적 역할의 분석 결과는 <표 19>와 같다.

<표 19> 의미분석의 처리 결과

처리후	( FINEN 구성 ( OBJ 메소드 ( ADNMEN 반환 ( OBJ 결과 ) ) ) ( TIME_AT 후 ( ADNMEN 변환 ( SUCO 받 ( TOP 클래스 변환 단위 ) ( OBJ 값 ( ADNMEN 필요 ( TARG_TO 단위 변환 ) ) ) ( STAT_AS 매개 변수 ) ) ) ) ) )

3. 실험 및 결과 분석

본 연구에서 제안한 서술형 문항 채점을 위한 복합문 구문의미분석 시스템의 실험을 위해 컴퓨터 분야의 전공과목 시험의 답변 자료를 수집하였다. 시험 문항은 서술형 유형이고 답변의 학생수는 30명이고 답변에 들어있는 복합문장은 175문장이다. 이 문장에 포함된 어절의 총 수는 2487개이고 술어의 수는 862개로 한 복합문장당 평균 4.9개의 술어와 14.2개의 어절로 구성되었다. 실험 데이터의 일부는 <표 20>과 같다.

<표 20> 실험 데이터의 일부

학생id	서술형 답안
01	자바 프로그램을 만들 때는 객체 중심으로 만들어야 한다. 따라서 프로그램에 필요한 자료의 내용과 특징을 기준으로 클래스를 정의하고 자료의 내용과 값을 계산하거나 변경하는 연산의 메소드를 만든다. 예상 키를 구하는 프로그램은 부모의 키와 자신의 성별을 입력하면 자신의 예상 키를 계산하는 프로그램이다. ...
02	3 개의 프로그램의 흐름과 구조를 이해하여 응용하여 프로그램을 작성할 수 있다.
03	예상 키를 구하는 클래스를 만들기 위해서 우선 속성변수에 부모의 키, 성별을 저장한다. 생성자 메소드에서 이 값들을 넘겨준 뒤 키를 구하는 메소드에서 계산 후 값을 돌려준다. n까지 자연수의 합을 구하는 클래스는 n을 속성에 저장하고 계산 후 값을 돌려준다. 단위를 변환하는 클래스는 생성자 메소드에 별다른 내용이 필요없을 수도 있는데 이 때 내용을 기술하지 않아도 생성자 메소드가 묵시적으로 동작한다. ...

본 연구의 실험은 구문의미분석의 대상인 어절



별로 유형을 파악하고 구문의미분석이 제대로 되었는지 검사하였다. 총 2487개의 어절에서 유형별로 조사한 결과 2088개의 유형으로 구분되었다. 어절수와 다른 수가 나온 것은 관형절의 경우 술어에도 포함이 되어 중복되어 표현되었고 복합명사의 경우 연속된 두 개 이상의 어절이 있을 경우 이를 하나의 복합명사로 다루었기 때문이다. 그 유형과 검사결과는 <표 21>과 같다.

<표 21> 시스템 실험 결과

검사유형	수	성공수	성공률(%)
술어	862	810	93.97
복합명사	247	237	95.95
접속조사	54	46	85.19
부사	69	65	94.20
접속사	15	9	60.00
관형사	63	61	96.83
관형격조사	138	137	99.28
격조사	841	776	92.27
관형절	318	317	99.69
시점	21	15	71.43
총 유형	2088	1933	92.58

실험 결과 전체 92.58%의 성공률을 얻을 수 있었다. 성공률은 실험자료에 들어있는 검사유형의 총 수에 대해 시스템이 올바르게 분석한 유형의 수에 대한 비이다. 이 결과로 본 시스템은 서술형 문항 자동채점에 충분히 활용할 수 있을 것으로 기대된다. 시스템의 실험결과 관형사, 관형격조사, 관형절의 경우 시스템에 반영된 locality principle에 의해 가장 근접한 용언과 체언에 부착하여 100%에 근접하는 성공률을 얻을 수 있었다. 이것은 답변의 내용이 컴퓨터 분야의 서술형 답안으로 복잡하지 않은 문장유형이라 이와 같은 결과를 가져온 것으로 파악된다. 그렇지만 나머지 유형에 대한 것은 본 연구에서 적용한 원칙으로 해결할 수 없는 것이 나타났다. 그 원인을 분석한 결과 다음과 같은 사실을 알 수 있었다.

<표 22> 실패원인 유형

실패원인	경우수	비율(%)
중의성	17	10.97
형태소해석	17	10.97
생략, 중복	2	1.29
병렬	12	7.74
사전부재	15	9.68
문맥의미분석	84	54.19
기타	8	5.16
계	155	100.0

1) 애매성 해소의 필요성

실패의 예	부모의 키와 자신의 성별을
시스템 분석 결과	( OBJ 성별 ( ADNM 자신 ( JUNC 키 ( ADNM 부모 ) ) ) )
'키'의 의미속성	feature inanimate-thing man-made-thing measurement physical-thing spatial-measurement thing tool
'자신'의 의미속성	abstract-thing animal animate-thing human physical-thing psychological-thing thing
'성별'의 의미속성	abstract-thing feature intellectual-thing relation thing word

실패 원인 가운데 많은 사례가 중의성의 사례이다. 위 예 어절 '키와'의 경우 접속조사 유형이므로 findsamen()에 의해 연결할 대상을 찾는다. 즉, '키와'의 주 단어가 '키'이므로 '키'의 의미속성과 유사한 명사를 찾는다. '자신'과 '성별' 가운데 '자신'이 일치하는 의미속성이 더 많다. 따라서 '성별'이 아니라 '자신'이 들어있는 어절과 결합한다. '키'는 신장에 해당하는 길이를 나타내는 것, 곡식을 까부는 도구, 배의 진행방향을 조절하는 도구, 열쇠 등의 여러 가지 의미를 나타내므로 의미속성 추출기는 모든 의미를 추출한다. 그렇지만 실제 예 어절은 그 중 신장에 해당하는 길이를 나타내는 의미로 사용되었다. 따라서 '자신'보다는 '성별'을 채택하는 올바른 해석을 하기 위해서는 '키'의 중의성이 해결되어야 할 것이다.

2) 형태소 해석의 개선

실패의 예	속성변수로 정의하고 부모의 키와 자신의 성별을 매개변수로 계산하는
시스템 분석 결과	( STAT_AS 변수 매개 ( JUNC 정의 ) )
형태소해석 결과	속성/ncn+변수/ncn+로/jca 정의/ncn+하고/jcj
형태소 해석 수정후의 결과	( COCOENUM 정의 ( STAT_AS 변수 속성 ) )

위 예 어절 '정의하고'의 경우 명사와 접속조사로 형태소 해석이 되어 구문해석이 올바르게 되지 않았음을 알 수 있었다. 10.97%의 사례가 형태소 해석의 실패로 발생하였고 본 실험에서 형태소 해석 결과를 수정하여 적용한 결과 옳게 됨을 알 수 있었다.

3) 생략 처리

실패의 예	메소드에서 계산 후 값을 돌려준다
시스템 분석 결과	( FINEN 돌리 ( SOUR_FROM 메소드 ) ( OBJ 값 ) ( TIME_AT 후 계산 ) )
수정 후 결과	( FINEN 돌리 ( OBJ 값 ) ( TIME_AT 후 ( ADNMEN 계산 ( SOUR_FROM 메소드 ) ) ) )

생략의 표현은 간략한 표현을 선호하는 특정 학생들의 답변에 집중하여 나타나 많은 사례는 출현되지 않았으나 이의 처리가 필요함을 알 수 있었다.

4) 병렬처리

실패의 예	1부터 n, 1부터 n까지의 짝수 합을 구한다.
시스템 분석 결과	( FINEN 구하 ( SOUR_FROM 1 n ) ( SOUR_FROM 1 ) ( OBJ 합 짝수 ( ADNM n ) ) )

예에서 1부터 n까지의 합과 1부터 n까지의 짝수합을 구하는 표현을 간략히 콤마를 사용하여 병렬적으로 표현하였다. 예에서는 콤마를 병렬접속으로 인식하여 n과 1을 복합명사로 표현하였다. 여러 개의 접속조사와 콤마로 연결된 경우 이에 대한 경계를 짓는 것과 구조화하는 것이 쉽지 않다. 이를 위해서는 정확한 경계와 의미를 파악하는 것이 필요하다.

5) 사전 추가

실패의 예	자동으로 구한다.
시스템 분석 결과	( FINEN 구하 ( TARG_TO 자동 ) )
사전 추가 후 결과	( FINEN 구하 ( MEAS_IN 자동 ) )

실패 중 9.68%의 사례가 의미속성 추출기의 사전에 포함되지 않아 발생하였다. 사전 추가 후 올바른 분석이 될 수 있었다.

6) 문맥의미적 분석

실패의 예	- 생성자 메소드로 아무 내용이 없으면 따로 기술하지 않아도 생성자 메소드가 묵시적으로 동작한다. - 문제점은 난이도 조절과 그로 인한 방관자적 태도를 갖는 것이다. - 세 프로그램에서 각 기능을 하는 클래스를 제어할 클래스가 필요하다.
시스템 분석 결과	( AUCO 기술 ( ADVERB 따로 ) ( SUCOCOND 없 ... ) ) ( TARG_TO 그 ( JUNC 조절 난이도 ) ) ( SOUR_FROM 프로그램 ( BLA 세 ) )

실패의 과반수 이상이 상황에 대한 문맥적 분석이 필요한 것이다. 예에서 '없으면'의 경우 문맥상 '동작한다'와 연결되어야 하는데 '기술하지'와

연결되어 분석이 실패하였다. 이를 개선하기 위해서는 문맥상황에 대한 지식과 상황 정보를 근거로 한 논리적 추론 규칙 정의와 적용이 필요하다.

두 번째 예는 의미적으로 복잡한 것으로 두 가지 문제점을 언급하는데 하나는 난이도 조절이고 다른 하나는 난이도조절로 야기된 방관자적 태도를 취할 수 있는 점이다. 이를 분석하기는 상당히 어렵다. 정확히 해석하려면 언급하는 내용에 대한 정확한 이해를 근거로 해야 한다.

세 번째 예는 '세 프로그램에서'가 주어나 토픽이 되어야 하는데 SOUR\_FR로 분석된 경우이다. 이를 수정하기 위해서는 조사 '에서'에 대해 토픽이나 주어가 되는지를 검사하기 위해 문맥적 상황을 분석해야 한다.

7) 불완전동사의 의미적 통합

개선 예	필요로 한다.
시스템 분석 결과	( FINEN 하 ( STAT_AS 필요 ) )

본 연구의 구문의미분석의 실패에는 해당하지 않지만 예의 경우 '하다'의 경우 '필요로 하다'를 묶어 하나의 의미로 표현한다면 전체 문장의 구문의미분석이 보다 명확해질 것이다.

실패의 주요 원인 6가지 가운데 형태소 해석과 사전 부재의 경우 이를 수정하고 사전에 추가하여 시스템을 수행한 결과 성공률이 94.11%로 개선됨을 알 수 있었다. 나머지 실패의 사례를 개선하기 위해서는 문맥 상황에 대한 분석과 의미적 제한조건을 고려하는 방법의 연구가 필요하다.

4. 결론 및 논의

서술형 문항의 자동 채점을 위해 답변에 대한 구문의미분석이 필요함을 제시하였다. 이에 따라 본 연구는 복잡한 개념과 의미를 표현할 수 있는 복문에 대해 구문의미분석을 할 수 있는 복합문 구문의미분석 시스템을 제안하였다. 본 연구의 실험 결과 92.58%의 성공률을 얻을 수 있었고 복문에 해당하는 술어에 대한 구문의미분석의 성공률도 93.97%를 얻어 시스템이 효과 있음을 알 수 있었다. 또한 시스템의 실패의 원인 가운데 형태소 해석과 사전부재의 경우, 형태소 해석의 수정과 사전에 엔트리를 추가한 후 시스템을 실험한

결과 더 나은 94.11%의 성공률을 얻을 수 있었다.

본 연구의 구문의미분석 시스템은 복잡한 개념과 의미를 표현하는 복문을 분석할 수 있으므로 서술형 문항 채점의 질 향상에 기여할 것으로 기대된다. 또한 이 시스템은 문서 분류나 자연어처리 영역에 활용할 수 있을 것이다. 그렇지만 실용적인 시스템에 활용하기 위해서는 오류평가를 수정할 수 있는 시스템적인 접근의 연구가 필요하다. 한편으로는 문장에서 발생하는 애매성 해소와 문맥의미를 해석하는 방법에 대한 연구가 진행되어야 할 것이다. 또한 문장간의 연결 관계 등을 고려한 분석도 필요하다.

### 참 고 문 헌

[1] 정상목 · 한병래 · 송기상 (2005). LSA를 이용한 서술형 주관식 평가 시스템의 설계 및 구현. **한국정보교육학회**, 9(2), 288-298.

[2] 조우진 · 오정석 · 이재영 · 김유섭 (2005). 의미커널과 한글 워드넷에 기반한 지능형 채점 시스템. **정보처리학회논문지A**, 12-A(6), 539-546.

[3] 강원석 (2007). 구문-격의미 정보를 이용한 주관식 문제 채점 시스템 설계 및 구현. **한국컴퓨터교육학회 논문지**, 10(5), 61-69.

[4] 박일남 · 강승식 · 노은희 · 김명화 · 성태제 (2013). 정답 템플릿 작성 방식에 의한 한국어 서답형 문항 자동채점 시스템. **정보과학회논문지 : 컴퓨팅의 실제 및 레터**, 19(12), 630-636.

[5] 박소영 · 김수홍 · 임해창 (2004). 문장성분의 다양한 자질을 이용한 한국어 구문분석 모델. **정보처리학회논문지B**, 11(6), 743-748.

[6] 박의규 · 나동열 (2006). 한국어 구문분석을 위한 구문음 기반 의존명사 처리. **인지과학**, 17(2), 119-138.

[7] 안광모 · 서영훈 (2011). 구문분석의 중의성 해소를 위한 전처리 규칙의 확장. **한국엔터테인먼트산업학회논문지**, 5(2), 159-165.

[8] 박용욱 · 권혁철 (2013). 구간 분할과 논항정보를 이용한 구문분석시스템 구현에 관한 연구. **멀티미디어학회논문지**, 16(3), 366-374.

[9] 박용욱 (2009). 의존문법 기반의 구간 분할법을 활용한 한국어 구문 분석기. **한국해양정보통신학회논문지**, 13(8), 1705-1712.

[10] 남용 · 윤애선 · 권혁철 (2014). 무제한 한국어 의존 구문분석기 'PNU-KLParser 2.0'의 개발. **정보과학회논문지 : 컴퓨팅의 실제 및 레터**, 20(6), 354-358.

[11] 이공주 · 김재훈 (2003). 규칙에 기반한 한국어 부분 구문분석기의 구현. **정보처리학회논문지B**, 10(4), 389-396.

[12] 강원석 · 황도삼 (2004). 다국어 질의응답을 위한 한국어 해석 시스템 설계 및 구현. **한국컴퓨터교육학회논문지**, 7(4), 43-50.

[13] 강원석 · 황도삼 · Jung H Kim (2014). 구문의미분석을 이용한 유사문서 판별기. **한국콘텐츠학회논문지**, 14(3), 40-51.

[14] 박천음 · 이창기 (2017). 포인터 네트워크를 이용한 한국어 의존 구문 분석. **정보과학회논문지**, 44(8), 822-831.

[15] 유혜원 (2009). 한국어 구문분석 방법론 연구 -복문 구조 분석을 중심으로. **민족문화연구**, 50, 153-183.

[16] 김재훈 · 선충녕 · 홍상욱 · 이성욱 · 서정연 · 조정미 (1999). KTAG99: 새로운 환경에 쉽게 적응하는 한국어 품사 태깅 시스템. **제11회 한글 및 한국어정보처리 학술대회논문집**, 99-105.



### 강 원 석

1985 경북대학교  
전자공학과(공학학사)  
1988 한국과학기술원  
전산학과(공학석사)  
1995 한국과학기술원 전산학과(공학박사)  
1995~현재 안동대학교 컴퓨터교육과 교수  
관심분야: 컴퓨터교육, 자연어처리, 정보검색  
E-Mail: wskang@anu.ac.kr