

# 구문-격의미 정보를 이용한 주관식 문제 채점 시스템 설계 및 구현

강 원 석<sup>†</sup>

## 요 약

주관식 평가는 객관식보다 높은 수준의 인지 능력을 측정할 수 있으나 채점의 객관성과 신뢰성의 문제, 그리고 한국어 처리의 어려움으로 기피하여 왔다. 본 논문은 이와 같은 문제를 해결하기 위해 한국어의 구문정보와 격의미 정보를 이용하여 주관식 문제의 답안을 채점하는 시스템을 설계, 구현하였다. 이 시스템은 채점의 시간과 노력을 덜고 채점의 객관성을 획득할 수 있다. 시스템의 실험 결과 75%에 달하는 성공률을 얻을 수 있었고 본 시스템을 이용하여 학생의 답지 가운데 정답에 해당하는 답지를 찾는 정보검색의 관점에서 시스템을 평가한 결과 단어추출방법을 이용한 시스템보다 재현율과 정확률의 개선을 얻을 수 있었다. 이 시스템은 앞으로 주관식 채점 방법의 기초가 될 것으로 기대된다.

키워드 : 주관식채점, 구문, 격의미

## Design and Implementation of a Subjective-type Evaluation System Using Syntactic and Case-Role Information

Won-Seog Kang<sup>†</sup>

### ABSTRACT

The subjective-type evaluation can estimate the high-recognition ability, but has the problem of the objectivity and reliability of the evaluation, and the difficulty of Korean language processing. To solve the problem, this paper designs and implements a subjective-type evaluation system using syntactic and case-role information. This system can reduce the time and endeavor for evaluation and provide the objectivity of the evaluation. The system results the 75% success rate to the instructor evaluation and gets the better precision and recall than the word extraction evaluation system. We expect that this system will become a basis of the research on the subjective-type evaluation.

Keywords : subjective-type evaluation, syntactic, case-role

### 1. 서 론

교육 분야에서는 여러 분야의 연구가 진행되고 있다. 그 가운데 평가는 교육의 성취여부, 문제점 파악, 그리고 개선 방향을 알려주기 때문에

이에 대한 연구는 매우 중요하다[1][2]. 일반적으로 평가는 객관식과 주관식으로 나눌 수 있다. 객관식 평가는 채점의 객관성과 신뢰성이 높지만 평가의 질이 떨어진다. 반면에 주관식 평가는 높은 수준의 인지 능력을 측정할 수 있지만 객관성과 신뢰성이 결여되고 자연어 처리의 문제점을 안고 있다. 이와 같은 문제를 해결하기 위하여 주관식 평가에 대한 관련 연구[3][4][5][6][7][8][9]

<sup>†</sup> 정 회 원: 안동대학교 컴퓨터교육과 교수(교신저자)  
논문접수: 2007년 6월 12일, 심사완료: 2007년 7월 19일  
\* 이 논문은 2004년 안동대학교 특별학술연구지원사업에 의하여 연구되었음

가 진행되고 있다.

[1][2]의 연구는 원격교육에서의 상호작용적 평가를 위하여 문제를 출제하고 답안을 피드백하여 채점자가 채점한 후 그 결과를 보여주는 평가관리 시스템을 설계, 구현하여 관리의 효율성을 도모하였다. 이 시스템은 객관식과 주관식 문항이 함께 제시되나 채점자가 채점하고 시스템은 단지 관리하는 관리도구이다. [8]의 연구는 주관식 유형의 채점시스템에 고려할 요소와 인자 등을 제시한 것으로 주관식 자동 채점 시스템은 아니다. [3]의 연구는 주관식 자동 채점을 위하여 단어의 빈도수 등의 정보를 벡터로 표현하고 시소러스를 이용하여 단어의 의미를 추출하는 방법으로 답안과 정답지의 유사도를 계산하는 구현된 채점 시스템이다. 그렇지만 이 시스템은 단어를 추출하는 방법과 의미를 추출하는 시소러스의 한계를 가지고 있어 실용적 시스템이기보다는 간단한 프로토타입 시스템이다. 그리고 [4]의 연구는 채점 보조 시스템으로 자동채점하는 시스템이 아니라 주관식 문제를 채점하는 과정에서의 고려사항을 관리하는 관리시스템이다.

[5]의 연구에서는 주관식 채점의 질을 높이고자 단어와 함께 의미를 활용하였다. 그러나 이 논문에서는 답안에 출현하는 단어에 대해서만 의미를 구할 수 있는 유의어 사전을 구축하였기 때문에 일반성과 범용성이 결여된다. [6]의 연구에서도 단답형이 아닌 주관식 유형에 대해 평가를 하는 시스템을 구현하였다. 역시 이 시스템도 답안 유형이 간단한 문장으로 제한되고 단어를 추출하여 그 단어의 AND 연산과 OR 연산을 통해 채점을 하는 방법을 제시하여 제한이 있다. 또한 단어 중심이므로 단어가 다르면 같은 의미일지라도 다른 것으로 인식하여 채점의 질이 낮다. 이와 같은 문제를 해결하기 위하여 [7]에서는 대량의 사전을 활용하여 일반적인 단어 추출과 함께 의미도 추출하였고 이를 벡터로 표현하여 유사도를 비교하는 방식의 채점을 시도하였다. 그러나 이 시스템도 답안이 하나의 문장으로 표현된 것의 처리에 초점을 맞추어서 답안이 논술형인 복수의 문장이나 복문의 경우에 대한 처리에는 문제가 있고 또한 구구조나 문구조의 처리도 하지 않았다.

본 논문은 논술형의 복수 문장을 처리 대상으로 하고 단어 이용의 한계를 극복하기 위하여 구문정보와 격의미 정보를 이용하는 주관식 채점 시스템을 제안한다. 본 시스템은 답안문서를 구문해석과 격의미해석을 하여 단위구구조와 단위문구조를 추출하고 이를 벡터화하여 유사도 계산의 과정을 거쳐 최종 채점을 하는 시스템으로 구문 정보와 격의미 정보를 이용하는 방법을 제안한다. 시스템을 복수 문장의 논술 유형의 답안에 실험한 결과 75%의 성공률을 얻었다. 학생의 답안 가운데 맞는 답안을 찾는 검색의 관점에서 시스템을 평가한 결과 단어추출만의 시스템보다 구문-격의미 정보를 이용한 본 논문의 시스템이 정확률과 재현율이 더 나아짐을 볼 수 있었다. 본 논문의 시스템에서 사용된 구문-격의미 활용 기법은 지능적인 자동 채점 시스템을 향한 발판이 될 것이다.

## 2. 구문해석과 격의미 해석

논술 형태의 한국어 답안을 정확히 채점하기 위해서는 한국어 답안의 문장을 해석하여 지식구조로 변경하고 정답의 지식구조와 비교를 해야 한다. 문서의 지식구조를 획득하는 과정은 일반적으로 형태소해석, 구문해석, 의미해석, 문맥해석, 그리고 지식구조화 단계를 거쳐야 한다. 최근 연구되고 있는 주관식 채점 방법은 형태소 해석 단계를 거친 다음, 단어를 추출하고 단어를 이용하는 방식을 취하고 있다[3][5][7]. 이와 같은 경우 지식구조와는 차이가 많다. 그러나 현실적으로 구문해석, 의미해석, 문맥해석, 그리고 지식구조화 단계의 어려움이 있기 때문에 단어를 이용한 채점의 방법을 사용하고 있다.

본 논문은 채점의 질을 향상하기 위하여 형태소 해석에서 한걸음 더 나아가 지식구조화의 단계를 위한 구문해석과 격의미해석의 단계를 거친다. 그리고 생성된 단위 구구조와 단위 문구조를 이용하여 답안을 채점한다. 여기에서 이용하는 단위 구구조와 단위 문구조는 다음과 같이 정의한다. 단위 문구조는 하나의 술어동사와 이에 부속되는 문장성분으로 구성되는 트리 형태를 말하는 것으로 단문에 해당하는 것이다. 예를 들면 "나는 학교에 간

다.”의 문장을 트리로 나타낸 것 “(가 verb (나 subj) (학교 targ))”을 단위 문구조 유형으로 정의한다. 단위 구구조는 하나의 독립성분과 이에 부속되는 요소로 구성되는 트리 형태를 말하는 것으로 하나의 구구조이다. “시스템의 실험”의 명사구를 트리로 나타낸 것 “(실험 noun (시스템 adnm))”을 단위 구구조 유형으로 정의한다.

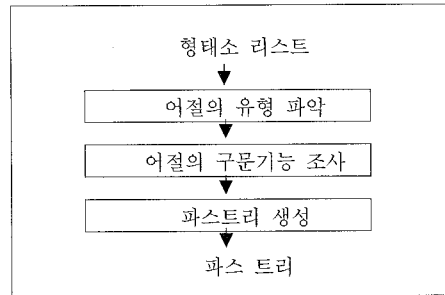
단위 문구조와 단위 구구조의 표현은 단어의 표현보다 더 정확한 채점을 가능하게 한다. 다음 문장의 예를 들자.

“원격 교육이란 의사소통 매체를 통합적으로 활용한다.”

위 문장에서 단어를 추출하면 “원격, 교육, 의사소통, 매체, 통합, 활용”의 단어가 나온다. 단어를 이용하는 채점 시스템은 단어들의 출현 순서와 위치를 고려하지 않는다. 따라서 정답의 내용 속에 들어 있는 단어가 전혀 관계없는 문장 속에 출현해도 이 단어가 출현한 것으로 인정되므로 채점의 질이 떨어진다. 만약 비교 대상을 단어로 하지 않고 구구조나 문구조로 한다면 단지 단어가 출현하는 것보다는 더 정확한 결과를 얻을 수 있을 것이다. 위 문장의 경우 문구조를 추출한다면 “( HNXS 활용 FINEN 다 하 ( HNOUN 교육 SUBJ 이란 원격 ) ( HNOUN 통합적 TARG\_TO 으로 ) ( HNOUN 매체 OBJ 를 ( HNOUN 의사소통 BLA ) ) )”의 문구조를 얻을 수 있고 답지의 문구조가 정답의 문구조와 같아야 답이 맞다고 채점할 수 있다. 즉, 단어만 일치하는 것이 아니라 문장의 유형, 단어와 단어의 유형, 그리고 격정보 등이 총체적으로 일치해야 한다. 이것은 부분적인 지식구조의 일치를 보여 주고 있어 보다 정확한 채점의 결과를 예상할 수 있다.

## 2.1 구문해석

한국어는 첨가어이기 때문에 구성요소의 구조적인 순서와 위치에 민감하지 않다[13][14][15]. 따라서 한국어 구문해석은 단어나 구의 구조적인 분석보다 기능적인 분석에 집중해야 한다. 즉, 단어나 구의 기능을 나타내는 조사나 어미의 분석에 초점



<그림 1> 구문 해석

을 맞추어야 한다. 그렇지만 조사나 어미의 세부 분석은 격해석 단계에서 처리하므로 구문 분석 단계에서는 일반적이고 구문적인 기능을 표현하는 파스 트리를 만드는 것에 목표를 둔다.

구문 해석은 형태소 해석의 결과인 형태소 리스트를 입력으로 받는다. 형태소 해석은 [11]의 시스템을 사용하였다. 구문 해석은 <그림 1>과 같은 세 단계의 처리를 거친다.

첫째 단계는 어절의 유형을 파악하는 것이다. 유형 결정은 형태소 해석의 결과를 토대로 어절의 유형을 결정한다. 그 과정은 다음과 같다.

1. find\_headinwp();
2. determ\_headtype();

먼저 find\_headinwp()를 통해 어절에서 중요한 의미를 가지는 헤드를 찾는다. 헤드는 어절의 주 의미를 담고 있는 명사나 동사가 된다. 그리고 두 번째로 determ\_headtype()를 통해 헤드의 품사 정보와 헤드에 첨가되는 첨가어의 종류에 따라 어절의 유형을 결정한다. <표 1>은 어절의 가능한 유형을 기술한 것이다. HNOUN은 명사와 조사가 결합하는 어절이고 HNXS는 명사와 파생접미사 “하”, 그리고 어미가 결합한 어절이다. HVERB는 일반 동사와 어미가 결합한 어절이고 HBEVERB는 명사와 서술격조사 “이”, 그리고 어미가 결합한 어절이다. HPTN은 동사와 명사형어미, 그리고 조사가 결합한 어절이고 HMA와 HMM은 어절이 부사나 관형사인 경우이다.

구문해석의 두 번째 단계는 어절의 구문적인 기능을 조사한다. 그 과정은 다음과 같다.

1. find\_casepart();
2. determ\_wprole();

<표 1> 어절의 유형

유형	헤드	첨가어	예
HNOUN	명사	(조사들)	매체+를
HNXS	명사	동사파생접미사+어미	활용+하+다
HVERB	동사	어미	달리+다
HBEVERB	명사	서술격조사+어미	교육+이+다
HPTN	동사	명사형어미+조사	살+ㅁ+은
HMA	부사		너무
HMM	관형사		그

먼저 find\_casepart() 함수에서 어절의 헤드와 연관되는 조사나 어미를 찾아낸다. 그리고 determ\_wprole()을 통해 그 조사와 어미의 유형에 따른 구문 기능을 결정한다. <표 2>는 어절의 가

<표 2> 구문기능

기능이름	조사나어미	예
주격(SUBJ)	가,이,에서,서	교사+가
목적격(OBJ)	을,를	매체+를
보격(COMP)	이,가	사용자+가
관형격(ADNM)	의	요소+의
호격(VOC)	야,아,여,이야,이여	철수+야
결합격(JUNC)	와,과,며,랑	교사+와
서술격(PRED)	이다	교육+이+다
부사격(ADV)	에,에서,으로	도구+로
종결격(FIN)	니다,다.	주+다
기타(BLA)	blank	교육

능한 구문 기능을 기술하고 있다. 이 절에서 언급한 구문 기능은 채점 시스템의 비교 정보로 이용하게 된다. 구문 기능의 결정은 어절의 헤드, 헤드와 연관된 조사나 어미, 그리고 피수식어의 세 요소의 정보를 이용한다. 피수식어는 어절이 수식하는 대상으로 서술어이거나 다른 체언 등이 될 수 있다. COMP 기능격의 경우 피수식어는 “이다”나 “되다”와 같은 서술어가 될 것이다. 관형격의 경우 어절은 복문을 만들거나 복합명사를 구성하게 된다. 결합격의 경우는 두 개의 요소가 동등 구조로 병합되는 경우이고 서술격은 영어의 be동사와 같이 “이다”로 맺어지는 문장에 해당한다. 부사격은 필수 기능 이외의 부사적 역할을 하는 조사나 어미로 “에”, “으로”, “에서” 등이 이에 해당한다. 종결격은 동사에 해당하는 것으로 문장의 종료를 나타내는 기능으로 “니다”, “다” 등이 이에 해당한다. 조사가 부착되지 않는 경우거나 <표 2>에 정의되지 않는 격조사가 사용되는 경우 기타 BLA로 정의한다. 이 격은 다음 격의미해석 단계에서 세부 격의미를 다시 검사하게 된다.

세 번째 단계는 파스 트리를 구성하는 단계이다. 그 과정은 다음과 같다.

1. construct\_quot();
2. construct\_compnoun();
3. construct\_mm();
4. construct\_ofnoun();
5. construct\_ma();
6. construct\_simsnt();

파스 트리 생성은 각 구성요소를 하나의 트리 로 보고 연관된 구성요소를 결합하여 여러 개의 트리를 하나의 트리 로 단 들어 가는 것이다. 즉, 포리스트에서 한 개의 트리 로 변환한다고 볼 수 있다. 트리 생성의 첫 단계는 construct\_quot()로 서 괄호와 특수기호 처리이다. 괄호의 내용이 하나의 문장인 경우는 이 내용을 해석하여 파스 트리를 구성한다. 만약 그렇지 않으면 괄호의 내용을 하나로 묶어 인용을 의미하는 하나의 요소로 정의한다. 연관된 요소가 묶여서 하나의 트리 로 되는 경우 각 요소의 피수식어와 트리의 루트어 부 등의 속성을 변경한다. 이 정보는 나중에 파 스트리에서 구성요소의 조상과 후손을 찾을 수 있도록 해준다.

트리 생성의 두 번째 단계는 복합명사를 만드는 과정으로 construct\_compnoun() 함수를 통해 진행된다. 이 단계에서는 연속적으로 나오는 명사들을 하나로 묶는다. BLA 기능이면서 첨부 조사나 어미가 없는 경우의 명사를 연관지어 결합한다. 여러 개가 연속으로 나올 경우 이를 하위 구조로 계속 포함시키고 마지막 단어를 트리의 루트로 설정한다.

트리 생성의 세 번째 단계는 construct\_mm()으로 관형사의 처리이다. 관형사인 경우 어절 유형이 HMM이다. 처리의 원칙은 관형사 고유의 특징 반영과 locality principle, no crossing constraint이다. locality principle은 부착될 가능성이 여러 개 있는 경우 근접하는 것을 선택하는 중의성 해결원리이고 no crossing constraint는 수식 관계가 겹쳐지지 않아야 하는 제한규칙이다. 트리 생성의 네 번째 단계는 construct\_ofnoun() 함수로 조사 “의”에 대한 처리를 한다. 이 경우 후위에 붙는 어절의 유형이 HNOUN이거나 HBEVERB인 어절을 찾아 연결

시킨다. 어절 유형이 HNOUN인 경우는 명사를 의미하고 HBEVERB인 경우는 명사 + “이다”를 의미한다. 이 과정에도 “의” 조사의 특징과 locality principle, no crossing constraint를 적용한다. 트리 생성의 다섯 번째 단계는 construct\_ma() 함수로 부사를 처리한다. 부사의 경우는 단어 유형에 따라 처리하게 된다. 부사는 동사, 파생동사 또는 서술격 동사에 부착한다.

트리 생성의 여섯 번째 단계는 construct\_simsnt() 함수로 문단위구조를 완성하는 과정이다. 이 단계에서 조사 붙은 어절의 피수식어를 찾게 된다. 이것은 그 어절이 어떠한 동사에 대해 구문격기능을 하는지 찾는 것이다. 어절의 부착 대상을 알아보기 위해 세 개의 구성요소를 검사한다. 그 구성 요소는 어절의 헤드 정보, 조사나 어미 정보, 그리고 피수식어의 정보이다. 앞서 언급한 바와 같이 세 개의 구성요소를 검사함과 동시에 no crossing, locality principle을 만족하는지를 검사하여 트리를 구성해나간다.

## 2.2 격의미 해석

주관식 답안의 정확한 채점을 위해서는 단어 수준의 정보 보다는 지식구조 수준의 정보를 이용하는 것이 낫다고 언급하였다. 지식구조는 각 구성요소들간의 의미 관계를 정의하여 설계된다[15]. 결국 지식구조에 근접하기 위해서는 구성요소들간의 의미관계를 파악해야 하고 이를 위해서는 어절의 세부 격의미기능을 밝혀야 한다. 즉, 구문해석에서 다룬 구문격기능보다 더 세분된 격의미기능의 파악이 필요하다.

한국어의 조사는 구문격기능 이외에 다양한 격기능을 가진다. 구문해석 과정을 거치면 입력 문장의 각 구성요소의 구문격기능은 밝혀진다. 그러나 아직 많은 어절들의 세부격기능이 밝혀지지 않은 상태이다. 본 논문의 격의미 해석 과정에서 이 어절들의 세부격의미 기능을 파악한다.

격의미 해석은 어절의 헤드 정보와 조사 정보, 그리고 피수식어의 의미 정보의 세 구성요소의 조건과 기타 사항을 검사해야 한다. <표 3>에 각 세부격의미들과 그 제한 조건을 기술하였다.

<표 3>의 첫 번째, 두 번째, 세 번째 열의 항목은 각 격의미에 대한 제한조건을 표현하고 있다. 첫 번째 열은 어절의 격조사를 나타내고 두 번째 열은 어절의 헤드 정보를 나타낸다. 이 정보는 헤드의 단어가 가지고 있어야 할 의미속성을 나타낸다. 세 번째 열은 어절과 관계를 맺는 피수식어 동사에 대한 정보이다. 역시 이 정보도 피수식어 동사가 가지고 있어야 할 의미속성을 나타낸다. 네 번째 열은 각 세부격의미의 이름이

<표 3> 세부 격의미와 제한조건

격조사	헤드	피수식어	세부격의미	기타
에	location,structure time time-point duration animate-thing	positioning  의(하다)	LOC-IN TIM-IN TIM-AT PERI SUBJ-BY TARG-TO	head+동안+에
로 / 으 로	organization,state, situation tool measurement		STAT-AS  INST-WITH MEAS-IN TARG-TO ADVERB	head+적+으로
에게	animate-thing	수여동사	TARG-FOR	
부터, 으로부터 부터, 로 부터	location time		SRPT SRTI SOUR	
서, 예 서	location situation location,structure		LOC-IN SITU-IN SRPT	head+에서 ... 까지
와, 과	animate-thing animate-thing	competition agreement	COMP-WITH ACCM-WITH	
까지	time location		DEST-TI DEST	에서/부터 ... head+까지 에서/부터 ... head+까지
마다			ADVERB	

고 다섯 번째 열은 기타 정보로서 세 구성요소 이외에 고려해야 할 제한조건을 뜻한다. "head+동안+에"의 정보는 격조사가 "에"와 헤드사이에 "동안"이라는 단어가 추가적으로 사용되는 경우를 말한다. "head+적+으로"의 경우도 마찬가지로 헤드와 격조사 사이에 "적"이라는 단어가 추가적으로 사용되는 경우를 말한다. "head+에서...까지"는 어절의 뒤부분에 "까지"라는 조사가 붙은 어절이 나오는 경우를 말한다. "에서/부터 ... head+까지"의 경우는 해당 어절의 앞 부분에 "에서"나 "부터"라는 조사가 붙은 어절이 나오는 경우를 말한다.

격의미 결정에서 여러 가지 가능성이 있는 경우

(중의성이 있는 경우)는 priority 순서로 격의미를 결정한다. 그 우선권은 각 격조사별로 표에 제시한 순서대로인데 제약조건이 강한 것을 높게 부여하였다.

격의미 해석 과정에서 헤드 단어 또는 피수식어 단어의 의미정보를 검사한다. 일반적으로 해석 과정의 성능은 해석 방법에 달려있다. 그렇지만 좋은 해석 방법을 사용한다고 할지라도 객관성있고 질 좋은 의미속성 획득방법이 없다면 해석과정의 성능이 보장되지 못한다. 본 논문에서는 격의미 해석에 필요한 의미속성을 획득할 수 있는 시소러스 도구를 사용한다[12]. 이 도구는 문서에서 빈번히 출현하는 단어 8631단어에 대해 의미를 획득할 수 있다. 예를 들면 단어 “교육”에 대해 의미속성을 구하면 “domain education event intellectual-behavior society utterance 교육”이 나온다.

아래 예 문장에 대한 구문해석과 격의미해석의 결과는 다음과 같다.

입력 문장 :

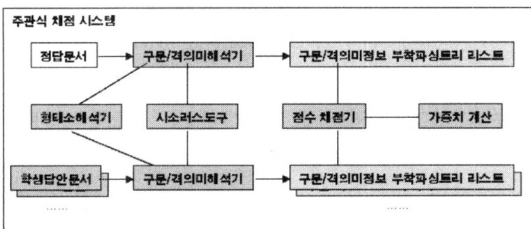
원격 교육이란 의사소통 매체를 통합적으로 활용한다.

구문해석&격의미해석 결과 :

( HNXS 활용 FINEN 다 하 ( HNOUN 교육 SUBJ 이란 원격 ) ( HNOUN 통합적 TARG\_TO 으로 ) ( HNOUN 매체 OBJ 를 ( HNOUN 의사소통 BLA ) ) )

### 3. 주관식 문제 채점 시스템

본 논문의 구문/격의미 정보를 이용한 주관식 문제 채점 시스템의 구조는 <그림 2>와 같다.



<그림 2> 주관식 문제 채점 시스템 구조도

시스템은 먼저 주관식 문제의 정답지에 대해 구문해석과 격의미해석을 실행한다. 그 결과로 구

문/격의미 정보가 부착된 파싱트리 리스트가 생성된다. 이 과정에 구문/격의미해석기는 형태소해석기와 시소러스도구의 도움을 받는다. 다음으로 시스템은 학생들의 답안 문서를 입력받아 구문해석과 격의미해석을 거쳐 구문/격의미 정보가 부착된 파싱트리 리스트들을 출력으로 낸다. 여기에서 시스템이 다루는 답안지는 서술형의 답지이므로 복수의 문장이 서술되고 이를 처리하면 복수의 파싱트리들이 생성된다. 파싱트리 리스트는 파싱트리들을 모두 모은 것을 말한다. 학생마다의 답지에 대해 처리를 하면 여러 개의 파싱트리 리스트들이 생성된다. 다음으로 채점기가 채점을 한다. 그 과정은 네 단계로 진행된다.

1. 가중치 계산
2. 단어와 트리의 벡터화
3. 유사도 비교
4. 최종 점수 계산

먼저 점수 채점기는 가중치를 계산한다. 가중치 부여 방법에 따라 다른 유형의 시스템을 구축하므로 다음과 같이 구분하여 정의한다.

- 비교시스템 : [16]의 단어추출 프로그램 이 용한 시스템. 본 논문의 시스템과의 비교를 위하여 타 단어추출 프로그램을 이용한 시스템을 구축하여 실험한다.
- 시스템 1 : 구문/격의미해석 없는 시스템. 단어에만 기본 가중치를 부여한다. 구문/격의미 해석의 필요성을 검증하기 위해 비교하기 위한 시스템이다.
- 시스템 2 : 구문해석 적용한 시스템. 추출된 단어에 기본 가중치 부여하고 구문해석의 결과인 (단어, 구문정보)의 구조에 기본 가중치를 더 부여한다. 구문해석의 결과인 파싱트리 정보가 매칭된 경우는 단어 매칭의 경우보다 정답에 근접함을 반영하므로 가중치를 더 부여하였다.
- 시스템 3 : 구문/격의미 해석 적용한 시스템. 추출 단어에 기본 가중치를 부여하고 격의미 해석의 결과인 (단어, 격의미정보)의 구조에 두 배의 기본 가중치 부여한다. 격의미 해석의 결과인 구조는 단어 매칭보다 정답에 더 의미적으로 근접하므로 가중치를 더 부여하였다.

- 시스템 4 : 구문/격의미 해석 적용한 시스템. 추출된 단어에 기본 가중치를 부여하고, 구문해석의 결과인 (단어, 구문정보)의 구조에 기본 가중치를 더 부여하고, 구문/격의미 해석의 결과인 (단어, 구문정보, 격정보)의 구조에 두 배의 기본 가중치를 더 부여한다. 이것은 구문해석과 격의미 해석의 결과인 구조가 단어 매칭의 경우보다 의미적으로 정답에 더 근접함을 반영한 것이다.

두 번째로 채점기는 추출된 단어와 생성된 트리구조들의 가중치를 벡터로 표현한다. 가중치는 출현 빈도수를 총 빈도수로 나눈값으로 하여 표현하였고 트리 구조들에 대해서는 빈도수를 더 부가한 것으로 가중치를 표현하였다.

세 번째로 채점기는 정답지의 벡터와 학생 답지의 벡터를 코사인 함수로 비교하여 유사도를 계산한다. 유사도 함수는 다음과 같다.

$$sim(Di, A) = \frac{Di \cdot A}{|Di| \times |A|} = \frac{\sum_{j=1}^m Di_j * A_j}{\sqrt{\sum_{j=1}^m Di_j^2} * \sqrt{\sum_{j=1}^m A_j^2}}$$

Di 는 학생 답안에 대해 단어 또는 트리구조와 가중치를 벡터로 표현한 것이고 A는 정답지에 대해 단어 또는 트리구조와 가중치를 벡터로 표현한 것이다. 수식의 m은 정답지와 학생답지에 나타난 단어 또는 트리구조의 총 수로서 벡터의 차수이다.

네 번째로 채점기는 유사도 함수값을 이용하여 평가 점수를 계산한다. 점수 계산식은 다음과 같다.

$$cal(Di) = a * (\frac{sim(Di)}{b})^2 + c$$

이 식에서 a, b, c는 상수이다. 이와 같은 계산식을 적용한 결과 시스템 4는 세 종류의 학생답지에 대해 평균한 결과 75%의 성공률을 얻었다.

#### 4. 주관식 채점 시스템 실험 결과

시스템을 실험하기 위하여 세 종류의 주관식

문제를 정의하고 출제자의 채점 결과를 정답으로 정의하였다. 다음은 세 종류의 주관식 문제이다.

문제 1. 채널을 이용하여 입출력할 경우 CPU, 채널, 입출력장치 등의 장치를 들어 그 과정을 설명하라.

문제 2. 원격 교육을 100자 내외로 간단히 설명하라.

문제 3. 교육용 소프트웨어를 100자 내외로 간단히 설명하라.

각 문제에 대한 답안의 수는 37, 36, 36이다. 각 답안의 점수는 100점을 만점으로 환산하였다. 세 유형의 답안에 대해 채점 시스템을 실험한 결과 평균 75%의 성공률을 얻을 수 있었다. 이 결과는 기존 관련연구보다 성공률이 낮다. 그러나 기존 관련연구들은 답안이 하나의 문장으로 된 경우로 제한하여 이루어진 것이고 본 연구의 성공률은 답안이 복수의 문장으로 된 경우로 확장한 것이어서 의미가 있다. 본 연구의 성공률은 다음과 같이 정의하였다.

$$성공률 = \frac{cal(Di) - bas(Di)}{총답지수} < 10인 답지수$$

이 식에서 cal(Di)는 답안 Di에 대해 본 논문의 채점 시스템의 채점 결과이고 bas(Di)는 출제자의 채점 결과이다. <표 4>는 실험데이터의 일부

<표 4> 문제 1에 대한 채점시스템의 일부결과

답안 ID	출제자 채점	비교 시스템	시스템4 유사도	시스템4 채점점수
20041962	90	0.86	0.69	98.32
20011299	70	0.80	0.66	91.03
20031311	70	0.85	0.65	89.05
20051252	80	0.79	0.65	87.69
20051250	80	0.87	0.63	84.76
20031326	80	0.83	0.63	84.07
20021236	90	0.88	0.63	83.19
20051255	80	0.84	0.62	81.87
20031334	80	0.87	0.61	78.62
20051256	70	0.90	0.61	78.40
.....				

예이다. 지면상 일부만 기술하였다.

<표 4>의 첫 번째 열은 답안ID이고 두 번째 열은 출제자의 채점 결과이고 세 번째 열은 비교 시스템에서 계산한 유사도 값이고 네 번째 열은 본 시스템 4(구문정보와 격의미정보에 가중치 부여한 시스템)에서 계산한 유사도 값이고 다섯 번

제 열은 본 시스템 4의 채점 점수이다.

주관식 문제 채점 시스템의 성능을 평가하는 것은 쉬운 일이 아니다. 출제자의 채점 점수와 시스템의 채점 점수의 상관계수를 이용하여 평가하기도 한다. 이와 같은 방법은 연속적인 점수인 경우는 가능하나 이산적인 점수의 경우 상관계수의 계산도 여의치가 않다[8]. 일반적으로 채점을 할 때 엄격히 채점한다면 이산치로 점수를 부여한다. 반면에 부분 점수를 부여하여 연속적인 값으로 점수를 부여할 수도 있다. 본 논문에서는 이산적인 값 부여방법이 정보검색 분야에서 신뢰성 있는 평가인자인 정확률과 재현율을 얻을 수 있게 하므로 이 방법을 채택하였다. 문제의 채점 결과가 한계치 이상인 경우 정답으로 하고 한계치 미만인 경우 오답으로 부여하여 정확률과 재현율을 계산한다. 본 연구에서의 한계치는 60점으로 하였다. 본 시스템에 적용한 정확률과 재현율은 다음과 같이 정의된다.

$$\text{정확률} = \frac{\text{시스템 결과가운데 맞는 문서수}}{\text{시스템의 결과 답문서의 수}}$$

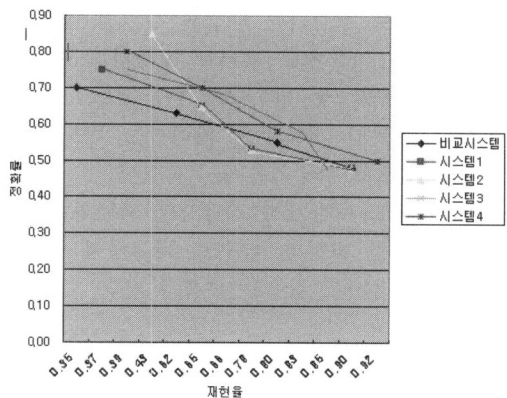
$$\text{재현율} = \frac{\text{시스템의 결과가운데 맞는 문서수}}{\text{채점자가 맞다고 판정한 총 답문서수}}$$

<표 5>가 정확률과 재현율의 척도로 비교시스템과 각 시스템을 평가한 결과이다. 표의 첫 번째 열은 재현율을 나타내고 두 번째부터 마지막 열까지는 각 시스템의 정확률을 나타낸다. 각 시스템마다 결과 답 문서수를 변화시키면서 그때의 재현율과 정확률의 변화를 표로 기술하였다.

<표 5> 각 시스템의 재현율과 정확률

재현율	비교 시스템	시스템1	시스템2	시스템3	시스템4
0.35	0.70				
0.37		0.75			
0.39				0.75	0.80
0.43			0.85		
0.62	0.63				
0.65		0.65	0.65		0.70
0.66				0.68	
0.78		0.53	0.53		
0.80	0.55				0.58
0.83				0.58	
0.85				0.48	
0.90	0.48	0.48	0.48		
0.92					0.50

<표 5>를 도표로 그린 것이 <그림 3>이다. 그림3에서 시스템 1과 시스템 2는 비교 시스템보다 나아짐을 볼 수 있으나 답 문서수를 늘려가면서 좋지 않은 결과가 나오기도 볼 수 있다. 그러나 시스템 3과 시스템 4는 비교 시스템보다 모든 경우에서 재현율과 정확률이 좋은 것을 확인할 수가 있다. 따라서 본 논문에서 구문해석과 격의미해석 정보를 이용한 주관식 채점 시스템이 효용 가치가 있음을 보였다.



<그림 3> 각 시스템의 재현율과 정확률

본 논문에서 적용한 단위구구조와 단위문구조

는 단어 매칭의 경우보다는 정확한 채점이 가능하나 아직도 수정해야 할 점이 많다. 본 시스템의 단위구구조 해석의 성공률은 83%이고 단위문구조 해석의 성공률은 72%여서 개선의 여지가 많다. 또한 구구조나 단위문구조간의 관계, 또는 순서 정보, 그리고 지식 구조의 존재, 수, 부정 등의 정보도 반영할 필요가 있다.

## 5. 결 론

주관식 평가는 그 가치가 높이 인정하나 시간과 경비, 그리고 객관성의 문제로 기피하여 왔다. 본 논문은 이와 같은 문제점을 해결하고자 구문/격의미 정보를 이용한 주관식 문제 채점 시스템을 제안하였다. 본 논문의 시스템은 구문/격의미 정보를 획득하기 위하여 실용적인 구문해석과 격의미해석기를 설계, 구현하여 이를 이용하였다. 본 논문의 시스템을 실험한 결과 단어 매칭을 하



는 기존의 주관식 채점 시스템보다 성능이 개선했음을 알 수 있었다. 본 시스템은 앞으로 주관식 평가 분야의 근간 기술로 활용될 것으로 기대하고 또한 정보검색이나 자연어 처리분야의 기초 기술로 활용될 것으로 기대한다.

본 시스템의 실험으로 앞으로 연구방향을 알 수 있다. 앞으로 다양한 서술유형의 답안의 채점 방법과 정확한 채점의 방법의 개선이 필요하다. 정확한 채점을 위해서는 답문서의 내용에 대한 지식구조화가 필요하다. 따라서 지식구조를 만들기 위해 구조나 단위구조의 연관관계를 정의할 필요가 있고 이를 분석하는 과정도 필요하다. 이와 같은 연구는 주관식 채점 분야 뿐 아니라 질의/응답 시스템, 지식기반추론의 분야에도 파급효과가 클 것으로 기대된다.

### 참고문헌

- [1] 한병래, 홍지영, 송가상, “원격 컴퓨터 교육을 위한 형성평가 프로그램의 설계 및 구현,” 한국컴퓨터교육학회 논문지, 3권 1호, 2000.6.
- [2] 남윤희, 김병기, 이종희, “원격 교육을 위한 평가 시스템 설계 및 구현,” 2001년 정보교육학회 하계 학술발표논문집, 2001.
- [3] 정동경, 벡터유사도와 시소러스를 이용한 주관식 답안의 채점방법, 동국대학교 교육대학원 석사논문, 2001.
- [4] 방훈, 허성구, 김원진, 이재영, “웹상에서 주관식 문제 채점보조 시스템,” 2001년 한국정보과학회 봄학술발표논문집, 28권 1호, 2001.
- [5] 박희정, 강원석, “유의어 사전을 이용한 주관식 문제 채점 시스템 설계 및 구현,” 한국컴퓨터교육학회 논문지, 6권 3호, 2003.7.
- [6] 김홍기, 홍동권, “주관식 평가를 위한 웹기반 온라인 평가 시스템의 구현 및 적용,” 한국정보교육학회 논문지, 8권 2호, 2004.
- [7] 조우진, 오정석, 이재영, 김유섭, “의미커널과 한글 워드넷에 기반한 지능형 채점 시스템,” 한국정보처리학회 논문지 A, 12권 6호, 2005.12.
- [8] 조지민, 김경훈, “서답형 문항의 인터넷 기반 채점 시스템 설계 연구,” 한국컴퓨터교육학회 논문지, 10권 2호, 2007.3.
- [9] G. Salton, Automatic Text Processing, Addison-Wesley, 1989.
- [10] 김명철, 김덕봉, 김유성, 김재훈, 박혁로, 이하규 공역, 최신정보검색론, 홍릉출판사, 1999.
- [11] 김재훈외 5인, “새 환경에 적응가능한 한국어 품사 태깅 시스템 KTAG99,” 11회 한글 및 한국어정보처리 학술대회논문집, 1999.
- [12] W. S. Kang and H. K. Kang, “An Effective Concept-based Text Categorization System Using the Thesaurus Tool,” Proc. of 18th Int. Conf. of Computer Processing of Oriental Languages, March, 1999.
- [13] 서광준, 어절간의 의존관계를 이용한 한국어 파서, KAIST 석사논문, 1993.
- [14] 이공주, 김재훈, “규칙에 기반한 한국어 부분구문분석기의 구현,” 정보처리학회논문지 B 제10-B권 4호, 2003.8.
- [15] 황도삼, 김태석, 최기선 공역, 자연언어처리, 홍릉과학출판사, 2007.
- [16] 강원석, 노주환, 제환주, 조대흠, 황세연, 정부천, “검색엔진을 위한 키워드 관련어 추출기의 설계 및 구현,” 한국컴퓨터교육학회 2007년도 동계학술대회 논문집, 2007.1.

### 강원석



1985 경북대학교 전자공학과 (공학학사)

1988 한국과학기술원 전산학과(공학석사)

1995 한국과학기술원 전산학과(공학박사)  
 1995~현재 안동대학교 컴퓨터교육과 교수  
 관심분야: 자연어처리, 정보검색  
 E-Mail: wskang@andong.ac.kr