

단문형의 영작문 자동 채점 시스템 구축

김 지 은[†] · 이 공 주^{**} · 진 경 애^{***}

요 약

영어 작문 자동 채점 시스템은 수험자가 작성한 영작문을 사람의 개입 없이 시스템이 처리하여 점수나 피드백을 줄 수 있는 시스템이다. 본 연구에서는 영작문 중 여러 문장이나 단락으로 구성된 에세이가 아닌 단문형의 영작문을 채점하는 시스템을 개발하였다. 단일 문장을 채점하기 때문에 정답 문장과 좀 더 자세한 비교를 할 수 있고 수험자들에게 좀 더 상세한 피드백을 제공해 줄 수 있다. 단일 문장을 채점하기 위해서는 크게 두 단계의 처리가 요구된다. 첫 번째 단계는 문장내의 오류를 탐지하는 과정으로, 수험자의 영작문을 분석하여 문장 내에 포함되어 있을 수 있는 철자 및 구문 오류를 검사한다. 둘째 단계는 문장 간 오류를 탐지하는 과정으로 문제 출제자가 제공한 정답문장과 수험자의 영작문을 비교하여 두 문장 사이의 차이를 오류로 인식한다. 실제로 중학교 3학년 학생들을 대상으로 영작문 시험을 수행하였고, 이를 본 연구에서 개발한 영작문 자동 채점 시스템을 이용하여 채점해 보았다. 인간 채점자와의 비교를 통해서 영작문 자동 채점 시스템의 효용성을 살펴보았다.

키워드 : 자동 채점 시스템, 영어 작문, 인간 채점자, 영작문 자동 채점, 오류 인식 규칙, 문장 내 오류, 문장 간 오류

Building an Automated Scoring System for a Single English Sentences

Jee Eun Kim[†] · Kong Joo Lee^{**} · Jin, Kyung-Ae^{***}

ABSTRACT

The purpose of developing an automated scoring system for English composition is to score the tests for writing English sentences and to give feedback on them without human's efforts. This paper presents an automated system to score English composition, whose input is a single sentence, not an essay. Dealing with a single sentence as an input has some advantages on comparing the input with the given answers by human teachers and giving detailed feedback to the test takers. The system has been developed and tested with the real test data collected through English tests given to the third grade students in junior high school. Two steps of the process are required to score a single sentence. The first process is analyzing the input sentence in order to detect possible errors, such as spelling errors, syntactic errors and so on. The second process is comparing the input sentence with the given answer to identify the differences as errors. The results produced by the system were then compared with those provided by human raters.

Key Words : Scoring System, English Composition, Human Rater, Error Production Rules, Intra-Sentential Error, Inter-Sentential Error

1. 서 론

영어 작문 능력을 향상시킬 수 있는 좋은 방법 중의 하나는 되도록 많은 영어 작문 시험을 거치면서 채점 결과를 통해 자신의 영어 작문에 대한 문제점을 파악해 가는 것이다. 이 과정에서 교사는 학생의 영어 작문을 채점하고 동시에 학생이 작성한 영어 문장에 포함된 오류를 찾아 학생에게 피드백을 줄 수 있어야 한다. 이 같은 작업은 교사의 많은 시간과 노력을 요구하며 대량의 채점을 어렵게 한다. 또한, 여러 수험자의 영작문 채점에 대한 일관성 유지가 어렵다는

문제가 있다. 그렇기 때문에 영어 작문을 자동으로 채점할 수 있는 시스템이 존재한다면 학생들의 영어 작문 시험의 향상에 큰 기여를 할 수 있을 것이다.

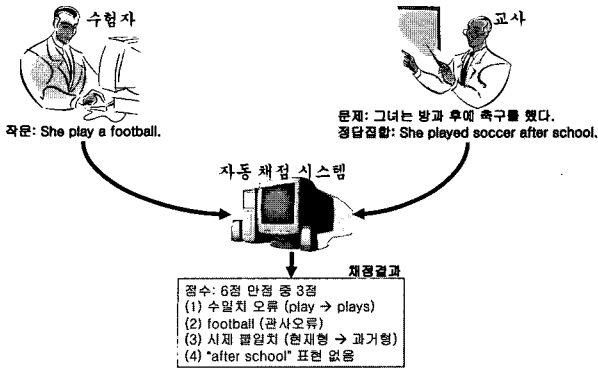
최근 영어 작문 자동 채점 시스템에 대한 많은 연구들이 진행되어 오고 있다[3,4,5,6]. 이와 같은 연구들은 주로 여러 문장이나 여러 단락으로 구성된 에세이에 대한 자동 채점 시스템들이다. 에세이 자동 채점은 개별 문장에 대해 철자 오류, 구문 오류를 검사한 후 에세이의 주제(topic)가 문제에 적합하지, 에세이의 형식이 적절한지에 대한 평가를 하여 채점을 수행한다. 영어를 제 2 외국어로 접한 지 얼마 되지 않은 학생들의 경우에는 에세이에 대한 평가보다는 개별 문장의 기본 문법에 대한 채점을 세밀하게 하고 피드백을 줄 수 있는 시스템이 더 도움이 될 수 있다. 이런 이유로 본

[†] 정 회 원 : 한국의국어대학교 영어학부 강사

^{**} 정 회 원 : 충남대학교 전기정보통신공학부 조교수

^{***} 정 회 원 : 한국교육과정평가원 선임연구원

논문접수 : 2007년 1월 17일, 심사완료 : 2007년 4월 11일



(그림 1) 영작문 자동 채점 시스템의 처리 과정

연구에서는 중학교 3학년 수준의 영어 작문 자동 채점 시스템을 개발하였다. 이 시스템은 단일 문장에 대한 채점을 수행하며 발견된 오류에 대한 피드백을 줄 수 있다.

수험자가 작성한 영어 문장을 자동 채점하는 과정은 (그림 1)과 같다. 우선 교사는 영어 작문 문제와 동시에 영어 작문에 대한 정답집합을 시스템에 제공한다. 수험자는 제시된 문제를 읽고 영어 문장을 작성하여 시스템에 입력한다. 영어 작문 자동 채점 시스템은 수험자의 영어 작문 결과와 교사가 제공한 정답집합을 비교하여 채점을 수행한다. 채점이 완료되면 (그림 1)의 '채점결과'와 같은 정보가 수험자에게 제공된다.

본 연구에서 개발한 영어 작문 자동 채점 시스템은 크게 두 단계로 구성된다. 첫 번째 단계에서는 수동으로 작성된 규칙을 이용하여 수험자의 작문을 분석한다. 수험자가 작성한 영어 문장에는 다양한 형태의 오류가 발생할 수 있기 때문에 정문에 대한 규칙뿐만 아니라 비문(ungrammatical sentence), 즉 오류가 포함된 문장을 처리하는 규칙도 동시에 이용해야 한다. 이와 같은 분석 과정을 통해 수험자의 문장에 포함되어 있는 철자 오류, 구문 오류 등을 검사할 수 있다. 분석된 수험자의 영어 문장은 단순한 의존구조(dependency structure)로 변환시킨다. 동일한 방법으로 교사의 정답집합도 의존구조로 변환시킨다. 두 번째 단계에서는 두 개의 의존구조를 비교해 가면서 수험자의 문장과 정답집합의 문장과의 차이를 계산한다. (그림 1)의 '채점결과' 중 (1), (2)는 첫 번째 단계로부터 도출된 결과이며, (3), (4)는 두 번째 단계로부터 도출된 채점결과이다.

본 논문은 다음과 같이 구성되어 있다. 2장에서 관련연구를 간단히 살펴보고, 3장에서 영작문 자동 채점 시스템의 전체 개요를 살펴본다. 4장에서 자동 채점을 위한 오류 처리 및 점수 계산 모델을 살펴보고 5장에서 실험결과를 제시한다. 마지막으로 6장에서 결론을 맺는다.

2. 관련 연구

참고문헌 [4,5,6]에서는 ETS(Educational Testing Service)에서 개발한 온라인 작문 평가 시스템인 *Criterion*SM을 소개

하고 있다. *Criterion*은 두 개의 요소로 구성되어 있는데, 점수 채점을 담당하는 *e-rater*[®]라는 시스템과 수험자가 범한 작문 오류에 대한 피드백을 줄 수 있는 *Critique*라는 시스템으로 구성되어 있다. *E-rater*[®]는 에세이 작성에 사용된 구문의 다양성(syntactic variety), 에세이의 구성, 주제(topic)의 부합 여부, 어휘사용의 복잡성(lexical complexity) 등을 이용하여 점수를 부여한다. 수험자가 작성한 영어 에세이로부터 채점을 위한 50여 개의 속성정보를 추출해 내고 통계모델(statistical model)을 이용하여 속성정보를 1~6점 범위로 점수화하였다. 이를 위해 실제 수험자가 작성한 270여 개의 에세이를 사람이 채점한 코퍼스가 학습용으로 사용되었다. *Critique*는 문법(grammar), 용법(usage), 구두점 사용, 자주 혼동되는 단어 사용의 정확성, 작문 스타일의 오류 등을 검사하고 오류가 있을 시 피드백을 제시한다. *Critique*는 대량의 코퍼스로부터 추출한 N 그램 정보나 상호정보(mutual information)를 이용하여 문장 내의 수/인칭의 일치(agreement error) 오류, 동사형태(verb formation) 오류, 단어 사용의 오류, 문장부호 오류, 철자 오류 등을 검사하였다. *E-rater*[®]는 실제 영작문 평가 시스템으로 사용되고 있다.

참고문헌 [7]은 자동 채점 시스템의 한 구성요소인 구문 분석 시스템의 구문 오류 처리 방법을 다루었다. 이 연구에서는 두 단계의 구문분석을 수행하여 구문 오류를 검사하였다. 첫 번째 단계로 차트기반의 구문분석을 수행한다. 이 단계에서 사용하는 규칙은 부분구문규칙(partial grammar)이기 때문에 구문 분석의 결과는 명사구(noun phrase), 전치사구(prepositional phrase)와 같은 구절 단위가 된다. 또한 구문 분석은 오류를 허용하면서 구절(phrase)을 완성한다. 두 번째 단계에서는 유한상태기계(finite state machine)를 이용하여 오류를 찾는다. 유한상태기계는 첫 번째 구문분석의 결과로 나온 구절을 입력으로 받아 수동으로 작성된 정규문법(regular grammar)을 이용하여 오류를 검사하게 된다. 이 연구는 일반적인 문장 자체의 오류를 검사하지 못했고 날짜 표현(date expression)에서 발생하는 오류에 대해서만 평가가 이루어졌기 때문에 시스템의 타당성 및 확장성을 정확히 알 수 없다.

참고문헌 [2]에서는 영어를 제 2 외국어로 사용하는 사용자들의 작문을 도와주기 위한 시스템에서의 문법 오류를 찾아내는 시스템을 구현하였다. 이 연구에서는 오류를 포함하고 있는 문장을 분석하기 위한 오류규칙(mal-rules)을 이용하여 오류를 인식한다. 오류규칙은 사용자들이 가장 자주 범하는 오류에 대해 오류를 포함하는 규칙을 직접 작성하되, 하나의 문장에서는 많아야 하나의 요소만이 생략될 수 있다는 가정하에 오류 규칙을 작성하였다.

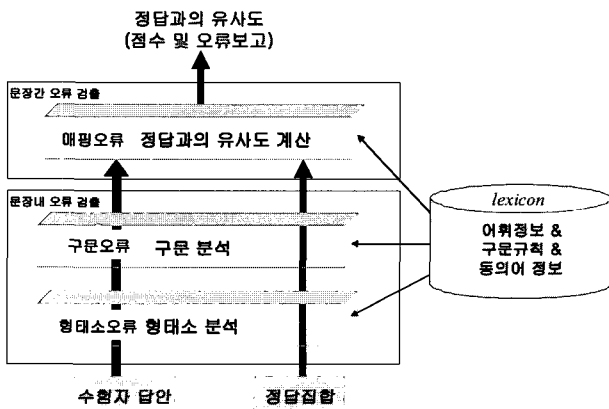
참고문헌 [8]에서는 통계모델을 이용하는 ALEK (Assessing Lexical Knowledge)이라는 시스템을 사용하여 영어 작문에서 오용된 단어들을 찾아낼 수 있다. 이 연구에서는 단어 의미의 모호성 해소 (word sense disambiguation) 방법을 이용하여 쓰임이 적절치 않은 단어의 사용을 찾아낸다. 대량의 코퍼스로부터 바이그램과 트라이그램을

추출하고 바이그램과 트라이그램의 단어들 사이의 상호정보 (mutual information) 값을 저장해 놓는다. 학생이 작성한 작문에 나타난 단어들에 대해 바이그램과 트라이그램을 추출하여 상호정보를 계산해 보았을 때, 그 값이 임계값 (threshold)보다 적으면 단어 사용에 오류가 있다고 간주하였다. 이 방법은 오류를 찾아내는데 필요한 정보를 수동으로 구축하지 않고 코퍼스에서 자동으로 추출한다는 데에 큰 장점이 있는 반면, 바이그램이나 트라이그램처럼 연속되지 않은 단어들 사이에서 발생하는 일치(agreement)와 같은 부류의 문제는 다루지 못한다는 단점이 있다.

3. 영작문 자동 채점 시스템

본 연구에서 개발한 영작문 자동 채점 시스템의 전체 구성은 (그림 2)와 같다. 자동 채점 과정은 우선, 교사가 제공한 정답집합에 대해 형태소분석, 구문분석을 수행하여 단순화된 의존구조를 만들어 놓는다. 수험자가 영어 문장 입력을 종료하면 입력된 문장에 대해 형태소 분석, 구문 분석을 동일하게 수행하는데, 수험자가 작성한 문장에는 오류가 있을 수 있기 때문에 오류 발생을 허용하면서 분석한다. 수험자의 문장도 구문 분석의 결과를 토대로 단순화된 의존구조로 만든다. 두 개의 의존구조를 비교하면서 매핑 오류를 찾아내고 정답과의 유사도를 계산한다.

(그림 2)에서 보듯이 전체 시스템은 크게 두 모듈로 구성되어 있다. 첫째 모듈은 '문장내 오류 검출' 모듈로서, 단일 문장 내에서 검사할 수 있는 오류만을 인식한다. 입력 문장에 대해 형태소 분석과 구문 분석을 수행하면서 문장부호 사용 오류, 철자 오류, 구문 오류 등을 검출한다. 두 번째는 '문장간 오류 검출' 모듈로서 이는 정답문장과의 비교를 통해서만 오류를 인식할 수 있다. 비록 작성한 영어 문장 자체에는 형태 오류나 구문 오류가 존재하지 않지만 그 문장이 교사가 제시한 문제에 부합하지 않을 수도 있다. 교사가 제시한 문제에 부합하는 시제를 사용하지 않았든지, 문제에서 원하는 의미를 다 기술하지 않았든지 또는 불필요한 요소가 포함되어 있다든지 하는 것이 두 번째 단계에



(그림 2) 영작문 자동 채점 시스템의 전체 개요

서 처리할 수 있는 매핑 오류이다. 이런 경우는 반드시 교사가 제공한 정답집합과의 비교를 통해야만 오류를 인식해 낼 수 있다. 또한, 작문 문제에는 일반적으로 하나 이상의 정답이 가능하다. 그렇기 때문에 교사는 정답으로 간주될 수 있는 영어 문장들을 정답집합으로 제공해야 한다. 정답집합의 문장들 중에서 수험자가 작성한 문장과 가장 유사한 문장을 결정하고, 두 문장 사이의 차이점을 비교하여 매핑 오류를 찾아낸다.

(그림 2)에서 보듯이 시스템 처리 과정에서 사용되는 지식베이스(knowledge base)는 사전(lexicon)이며 사전에는 어휘 정보와 구문규칙, 그리고 단어 또는 구(phrase) 간의 동의어 정보가 저장되어 있다. 자동 채점 시스템의 최종 출력은 수험자 답안과 정답 답안과의 비교를 통해서 얻어진 유사도를 환산한 점수이며, 동시에 각 처리 단계에서 얻어진 오류 보고이다.

3.1 자동 채점 시스템의 사전 정보

자동 채점 시스템이 사용하는 사전에는 크게 어휘정보, 구문규칙, 동의어 정보가 포함되어 있다. 자동 채점 시스템의 경우 처리하는 입력 문장에 오류가 포함될 수 있기 때문에 오류 문장도 처리할 수 있도록 정보를 유지해야 한다.

3.1.1 어휘 정보 및 동의어 정보

어휘정보에는 일반적으로 각 단어나 숙어에 대한 품사, 활용정보, 하위범주화 정보 등이 포함된다. 숙어 오류를 효율적으로 처리하기 위해 (그림 3)에서 보는 바와 같이 "go_to_home"이라는 오류가 포함된 숙어 엔트리를 사전에 저장하였다. 예를 들어 수험자의 입력 문장이 "You had better go to home."일 때, 오류 엔트리를 이용하면 'go_to_home'을 하나의 숙어표현으로 분석해 낼 수 있고, 동시에 "IDIOM-MISUSE"의 오류가 발생했고 정확한 표현은 'go home'임을 사전정보를 통해 알 수 있다.

영작문에 대한 정답이 항상 여러 개인 이유 중 하나는 영작문에 사용되는 단어들을 항상 동의어로 대체할 수 있기 때문이다. 즉, 문장 "I can swim"과 "I am able to swim"은 동일한 문장으로 간주된다. 'can'과 'be able to'가 동의어이기 때문이다. 수험자의 답안과 정답을 제대로 비교하기 위해서는 사전에 동의어 정보를 갖고 있어야 한다. 본 논문에서는 Princeton 대학에서 오랫동안 개발해온 WordNet@[1]에서부터 추출한 동의어 집합(synsets) 정보와 집합 간의 계층구조를 이용하여 각 단어나 구의 동의어 정보로 사용한다. 동의어 정보는 사전의 각 엔트리가 갖고 있는 'SYNONYM' 필드에 저장한다 (그림 3)에서 표제어 'give_up' 정보에는

LEX:	"go_to_home"	LEX:	"go_home"	LEX_ENTRY:	"give_up"
POS:	verb	POS:	verb	ROOT:	"give_up"
INF_CODE:	IRREGULAR	INF_CODE:	IRREGULAR	POS:	verb
SUBCAT:	10	SUBCAT:	10	INF_CODE:	IRREGULAR
WEIGHT:	0.5	WEIGHT:	0.5	SUBCAT:	T1 10
CORRECT:	"go_home"			WEIGHT:	0.5
ERROR:	IDIOM MISUSE			SYNONYM:	"abandon"

(그림 3) 사전의 어휘 및 동의어 정보

'SYNONYM' 필드에 'abandon'이라는 값이 저장되어 있는데, 이는 'abandon'이 'give_up'의 동의어임을 알려준다. 본 연구에서 구축한 사전(lexicon)에는 6,285 개의 사전 엔트리와 7,500 여개의 동의어 집합이 포함되어 있다.

3.1.2 구문 규칙

입력 문장의 구문 구조를 파악하는 구문규칙은 구구조 문법(phrase structure grammar)을 기본으로 하여 작성되었다. 규칙들은 영어의 정문(grammatical sentences)을 분석하는 일반 규칙과 구문 오류(syntactic errors)를 포함하고 있는 비문(ungrammatical sentences)을 분석하는 오류 규칙으로 구분된다. 규칙의 기본 형태는 (그림 4)와 같다. 각각의 규칙에는 규칙이 적용되기 위한 조건(CONDITION)을 검사하는 부분이 함수로 구현되어 있다. 조건은 여러 개가 가능하며 모든 조건절이 '참'의 결과를 낼 때에만 규칙이 성공적으로 적용될 수 있다. 규칙이 적용되면 수행해야 하는 부가적인 동작(ACTION)들 역시 함수의 형태로 구현되어 있다.

대부분의 비문을 처리하기 위한 규칙은 정문을 처리하는 규칙과 많은 부분이 공유된다. 가령, 오류인 'a flowers'를 처리하기 위한 규칙과 'a flower'를 처리하기 위한 규칙은 관사와 명사 사이의 수(number) 정보가 일치해야 한다는 조건을 제외하고는 동일하다. 그렇기 때문에 본 논문에서는 비문을 처리하는 규칙을 정문을 처리하는 규칙 내의 조건절(CONDITION)을 완화시킴으로써 구현하였다. (그림 4)의 예제 규칙에서 조건절을 검사하는 함수 '_CHECK_DN_AGR'은 관사와 뒤따르는 명사 사이의 수 일치를 검사하는 함수로써, 함수의 반환값(return)이 '참'일 때는 일반 규칙처럼 작용하나, 반환값이 '참'이 아닌 경우는 오류규칙으로 동작하면서 'DN_AGR_ERR' 오류가 발생했음을 알려준다. 시스템에 구축되어 있는 구문 규칙은 순수 정문을 위한 구문규칙이 235개, 순수 비문만을 위한 구문 규칙이 26개, 그리고 정문과 비문을 모두 처리할 수 있는 구문 규칙이 55개로 모두 316개의 구문규칙이 사용되고 있다.

RULE:	NP:np0 → DET:det1 NP:npl
CONDITION:	_CHECK_DN_AGR (WEAK, [DN_AGR_ERR], det1.agr, npl.agr) _CHECK_CV(WEAK, [CV_ERR], det1.lex, npl.lex)
ACTION:	_ASSIGN(npl, np0)
WEIGHT:	0.5
HEAD:	npl

(그림 4) 예제 구문 규칙

3.2 처리 오류의 범위

자동 채점 시스템이 처리하는 오류의 종류와 예제, 그리고 실제 처리 단계는 <표 1>과 같다.

<표 1> 시스템이 처리하는 오류의 종류와 예제

	오류 종류	예제	처리 단계
형태소 오류	명사 복수형 활용 오류	boxes, childs, toies	형태소 분석
	명사 수정보 오류	sugars	
	동사 활용형 오류	watches, comming	
	비교급/최상급 활용 오류	famouser, prettyer	
	철자 오류	intuton, nife	
	숙어 사용 오류	You had better go to home now.	
	문장부호 오류	He is a student (the period is missing)	
구문 오류	관사 명사 수 일치 오류	a sheep / a bits	구문분석
	관사 형태 오류	a apple / an yacht	
	관사 추가 오류	the Jane swimming	
	한정사 형태 오류	for him wife	
	한정사 명사 일치 오류	much leaves / many furniture	
	전치사 추가 오류	She went to there / She looks like healthy.	
	전치사 타입 오류	It is better to me to go home.	
	비교급 오류	more prettier / more pretty	
	최상급 오류	most successful / the most tallest student	
	형용사 어순 오류	special something	
	형용사 타입 오류	an asleep baby	
	주어 동사 일치 오류	He have a book.	
	동사 형태 오류	He spoken French. / He can speaks English. / He enjoyed to play tennis.	
	동사 수동형 오류	He got pull.	
	동사 완료형 오류	He has went home.	
	동사 진행형 오류	She is knowing the news.	
	be동사 보어 일치 오류	He is the tallest students.	
	조동사 타입 오류	I would better go home.	
	명사 추가 오류	the book that I gave it to him	
	명사 수 오류	He is more intelligent than any other students.	
	부사 추가 오류	He had better than go home now.	
	부사 타입 오류	He is much famous.	
	접속사 추가 오류	Although he is poor, but his is happy.	
	의문대명사 오류	Which she said is true.	
	관계대명사 생략 오류	She is a teacher came to my school last week.	
	관계대명사 추가 오류	She is the teacher who she came last week.	
	관계대명사 타입 오류	He is the man whom gave me the book. He is the man who I have seen before.	
매핑 오류	필수 성분 생략 오류	He e-mail to his wife. → He sent e-mail to his wife.	정답과의 유사도 계산
	속성 불일치 오류	He sends e-mail to his wife. → He sent e-mail to his wife.	
	부가 성분 생략 오류	She is the teacher who came to our school last week. → She is the teacher who came to our school.	
	불필요 성분 삽입 오류	I don't know why she did go there. → I don't know why she did go there then.	

4. 오류 처리 및 점수 계산 모델

4.1 문장내 오류 처리 모듈

단일 문장 내의 오류는 사전의 어휘정보와 구문규칙을 이용하여 검출하고 수정한다. 처리하는 오류의 대상은 <표 1>에 제시하였다. (그림 2)에서 보는 바와 같이 형태소 분석하여 사전의 어휘정보를 적재하고 동시에 사전에 등재되어 있지 않은 단어에 대해 형태소오류를 인식한다. 각 단어에 대한 어휘정보와 구문규칙을 이용하여 구문 분석한다. 구문분석 과정에서 오류규칙에 의해 분석된 구절에서는 구문오류가 인식된다. 구문규칙에는 각 규칙의 중요도(WEIGHT)가 부여되어 있기 때문에 이를 이용하여 여러 개의 구문트리 중 가장 높은 중요도를 갖는 트리를 최종 결과로 얻어낼 수 있다. 중의성이 해소된 하나의 구문트리를 문장간 오류 처리를 위하여 의존구조로 변환한다. 본 모듈은 기존의 철자 검사기와 구문 검사기의 기능들을 다소 확장한 것이므로 여기서는 자세한 설명을 생략한다.

4.2 문장간 오류 처리 모듈

문장간 비교를 통해 발견할 수 있는 오류는 교사가 제시한 문제에 부합하지 않은 요소들을 찾아내는 것으로써, 크게 속성 정보 불일치 오류, 불필요 성분 추가 오류, 부가 성분 생략 오류로 나눌 수 있다. 문장간 오류 처리 모듈의 입력은 (그림 5)에서 보는 바와 같이 정답집합의 문장들을 분석하여 얻은 의존구조들과 수험자 답안을 분석하여 얻은 의존구조이다. 수험자 답안의 의존구조는 항상 1개이고, 정답집합 내의 문장들에 대한 의존구조는 N 개이므로, 수험자답안의 의존구조를 N 개의 의존구조와 비교하여, N 개 중 가장 유사한 의존구조 하나를 선택한다. 자세한 알고리즘은 (그림 6)에 제시하였다.

(그림 6)의 알고리즘에서 β는 10을, α는 2의 값을 사용하였다. (2.3) 단계에서 비교하는 속성값으로는 품사정보, 서법(mood), 양상(modal), 시제(tense), 태(voice), 상(aspect), 부정표현 사용여부, 인칭-수 정보, 한정사 정보 등이 있다.

4.3 점수 계산 모델

본 논문에서는 형태소 단계, 구문 단계, 매핑 단계의 세 단계로 나누어 점수를 계산하였으며 이 세 단계의 점수를

<표 2> 점수 계산을 위한 기준표

점수	형태소오류	구문오류	매핑오류
2	없음	없음	모두 적절
1	1~2개	1개 이하	1~2개 틀림
0	3개 이상	2개 이상	3개 이상 틀림

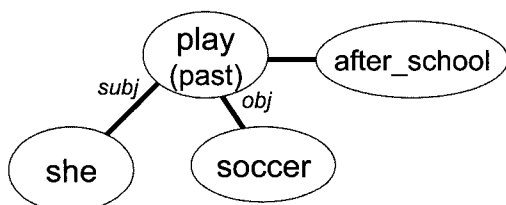
합산하여 최종 점수화하였다. 각 단계마다 0~2점 사이의 점수가 할당되어 있기 때문에 최종 점수의 총점은 6점이다. 점수를 부여하는 기준은 현직 중학교 교사들과의 협의하에 <표 2>와 같은 기준표를 얻어 사용하였다.

(그림 6)과 같은 과정을 거쳐 정답집합 중 수험자의 답안과 가장 유사도가 높은 한 개의 의존구조가 결정되고 나면 (그림 7)과 같은 휴리스틱을 이용하여 유사도를 점수로 환산하였다.

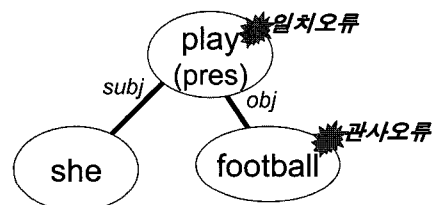
수험자의 영작문이 “He is the tallest his class”이고 정답집합에 “He is the tallest in his class.”가 포함되어 있을 때, 수험자의 영작문은 전치사 ‘in’만이 생략되었을 뿐, 정답집합과 동일하다. 그러나, 구문분석이 실패하면서 6점 중 0점을 받게 되는 상황이 발생할 수 있다. 본 논문에서는 이와 같이 정답문장과 매우 유사함에도 불구하고 구문분석이 실패한다거나 전혀 다른 의존구조가 만들어짐으로써 매핑이 제대로 이루어지지 못하는 수험자 답안을 고려하기 위해 스트링 비교를 이용한 점수 계산 모형을 부가적으로 사용하였다. 즉, 구문분석의 불완전성으로 인해 발생하는 문제점을 보완하기 위해 수험자 답안과 정답 스트링의 스트링 비교만을 통해 보완 점수를 계산해 낸다.

수험자 답안과 정답집합의 정답들을 스트링 대 스트링으로 비교하여 그 차이를 계산한다. 한 스트링을 다른 스트링으로 변환시키기 위한 편집거리(edit distance)를 최소화하면서 두 스트링 사이의 차이를 다이나믹 프로그래밍(dynamic programming) 방식으로 계산해 내는 최소편집거리(minimum edit distance) 방식을 이용한다. 두 스트링 사이의 편집거리는 삽입(insertion) 오류, 삭제(deletion) 오류, 치환(substitution) 오류의 세 종류가 가능하다. 이렇게 계산한 편집거리(d)가 2 이하이고 삭제된 요소가 문장의 필수성분이 아닌 경우에는 (그림 7)에서 계산된 값을 무시하고, 6점 만점 중 고정된 점수 SCORE2를 부여하였다. 최종적인 점수 계산 방법은 (그림 8)과 같으며, 본 논문에서는 SCORE2로 5점을 사용하였다.

정답집합: She played soccer after school.



수험자 답안: She play a football.



(그림 5) 문장간 오류 처리 모듈의 입력: 두 개의 의존구조

```

// 수험자 답안의 의존구조를 Dep_a라고 하고, 정답의 의존구조를 Dep_c라고 하자.
(1) Dep_c와 Dep_a의 모든 노드를 NO_MAPPING으로 표시해 둔다.
(2) Dep_a의 각 노드 X에 대해 (2.1)~(2.3)을 반복적으로 수행한다.
    (2.1) Dep_c에서 노드 X와 동일한 NO_MAPPING node Y를 찾는다.
        성공한 경우 (2.3)를 수행하고 실패한 경우 (2.2)을 수행한다.
        (노드의 원형단어가 같을 때 동일한 노드로 간주한다.)
    (2.2) node X의 동의어인 NO_MAPPING node Y를 Dep_c에서 찾는다.
        성공한 경우 (2.3)를 수행하고 실패한 경우 node X를 NO_MAPPING으로 표시한다.
        (같은 SYNONYM 정보를 갖는 단어를 동의어로 간주한다.)
    (2.3) 노드 X와 노드 Y의 NO_MAPPING을 MAPPING으로 바꾸고,
        노드 X와 노드 Y의 각 속성을 비교한다.
        일치하지 않는 속성에 대해 유사도 값을  $\alpha$ 만큼 감소시킨다.
(3) Dep_a중 NO_MAPPING으로 표시되어 있는 노드 각각에 대해
    유사도 값을  $\beta$ 만큼 감소시킨다.
(4) Dep_c중 NO_MAPPING으로 표시되어 있는 노드 각각에 대해
    유사도 값을  $\beta$ 만큼 감소시킨다.
(5) Dep_a와 Dep_c가 여러 개 있을 때,
    (1) ~ (4)의 과정을 모든 Dep_a와 Dep_c의 조합에 대해 수행하고,
    최대 유사도를 갖는 Dep_a와 Dep_c를 결정한다.
    
```

(그림 6) 문장간 오류 처리 알고리즘

최종점수가 무조건 0점인 경우

```

(1) 구문 오류가 3개 이상이면 최종점수는 0점
(2) 정답 의존구조에서 필수격에 해당하는 노드가 매핑되지 않은 경우 최종점수는 0점
(3) num_node_std_dp = 수험자 답안의 의존구조의 노드 개수
    num_succ_node = 두 의존구조사이에 서로 매핑이 성공한 노드 개수
    num_fail_node = 정답 의존구조의 노드 중 매핑되지 않은 노드 개수
        + 매핑은 되었으나 속성정보가 일치하지 않는 노드 개수
    if ( num_node_std_dp > num_succ_node * 2 AND
        num_node_std_dp < num_fail_node * 2) then
        최종점수는 0점
    
```

최종점수를 형태소/구문/매핑 단계로 나누어서 합산하는 경우

```

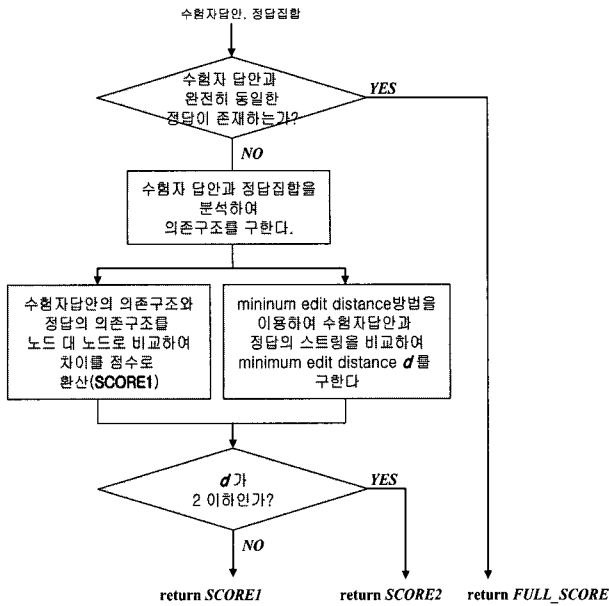
(1) 형태소 단계의 오류 계산 (score_MA):
    if (형태소_단계의_오류_개수 == 0) { score_MA = 2; }
    else if (형태소_단계의_오류_개수 <= 2) { score_MA = 1; }
    else { score_MA = 0; }

(2) 구문 단계의 오류 계산 (score_SA):
    if (구문_단계의_오류_개수 == 0) { score_SA = 2; }
    else if (구문_단계의_오류_개수 <= 1) { score_SA = 1; }
    else { score_SA = 0; }

(3) 매핑 단계의 오류 계산 (score_MAP):
    num_fail_node = 정답 의존구조의 노드 중 매핑되지 않은 노드 개수
        + 매핑은 되었으나 속성정보가 일치하지 않는 노드 개수
    if (num_fail_node == 0) { score_MAP = 2; }
    else if (num_fail_node <= 2) { score_MAP = 1; }
    else { score_MAP = 0; }

(4) 최종점수 계산:
    score = score_MA + score_SA + score_MAP;
    
```

(그림 7) 점수 계산을 위한 휴리스틱



(그림 8) 정답과의 유사도 및 점수 계산 과정

5. 영작문 자동 채점 결과 및 평가

영작문 자동 채점 시스템의 성능을 평가하기 위해 중학교 3학년 수준의 영작문 19문항을 개발하여 실제로 서울의 두 중학교 3학년 학생들(각 학교당 약 5개 학급, 총 282명)을 대상으로 시험을 실시하였다. 문항은 중학교 3학년 현직 영어교사와의 협의하에 개발되었으며, 우리나라 중 3 영어 교과서 공통 문법 영역을 분석한 결과를 바탕으로 개발되었다. 채점은 중학교 영어 교사 2명이 동시에 채점함으로써

<표 3> 자동 채점 시스템의 실험 결과

문항번호	시스템과 인간채점자간의 차이의 평균	표준편차	인간채점자간의 차이의 평균	표준편차
1	0.65	1.19	0.52	1.14
2	0.6	1.17	0.57	1.38
3	0.65	1.28	0.35	0.96
4	0.4	1.03	0.41	1.08
5	0.84	1.42	0.26	0.9
6	1.36	1.83	0.63	1.39
7	0.45	1.13	0.46	1.21
8	0.7	1.37	0.33	0.98
9	0.62	1.07	0.45	1.01
10	0.61	1.17	0.57	1.29
11	0.52	1.09	0.73	1.43
12	0.52	0.95	0.82	1.39
13	0.59	1.04	0.5	1.17
14	0.66	1.19	0.73	1.38
15	0.87	1.17	0.79	1.37
16	0.52	1.09	0.37	0.97
17	0.72	1.14	0.71	1.47
18	0.57	1.13	0.13	0.45
19	0.95	1.29	0.44	0.95
TOTAL	0.67	1.20	0.51	1.15

동일한 답안에 대해 2명의 채점결과를 확보할 수 있었다. 인간채점자도 시스템과 마찬가지로 형태소, 구문, 매핑 부분으로 나누어 채점하였으며, 시스템과 동일하게 <표 2>의 기준을 사용하였다.

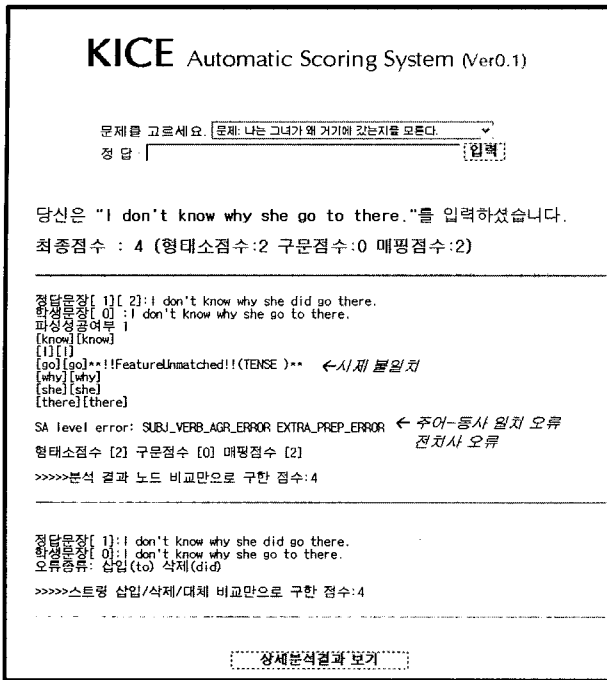
(그림 9)는 본 논문에서 개발한 영작문 자동 채점 시스템의 인터페이스 화면이다. 그림에서 보는 바와 같이 인터페이스는 아직 개발 초기 단계이기 때문에 사용자 측면을 많이 고려하지 못한 상태이다. 사용자가 오류 내용을 쉽게 이해할 수 있도록 사용자 피드백에 대한 고려가 많이 이루어져야 한다.

<표 3>에 실험 결과를 정리하였다. 우선 각 문항별로 시스템이 부여한 점수와 인간채점자가 부여한 점수의 차이에 대한 평균을 계산하였다. 동일한 답안에 대해 2명이 채점하였기 때문에, 2명의 인간채점자의 점수 평균과 시스템의 점수 차이를 계산하였다. 19문항에 대한 점수차이의 평균은 0.67점이었다. 차이의 평균에 대한 표준편차는 1.20이었다. 동일한 문항에 대해 2명의 인간채점자가 채점을 수행했기 때문에 2명의 인간채점자 사이의 점수차이의 평균도 구해보았다. 인간채점자 사이에도 0.51점의 점수차이를 보였다. 인간채점자 간의 차이와 비교해 보았을 때, 시스템의 채점결과가 희망적임을 알 수 있었다. 또한, 시스템 채점의 가장 큰 장점 중의 하나는 동일한 답안에 대해서는 동일한 점수를 항상 부여하는 일관성을 확보할 수 있다는 점이다.

6. 결론

본 연구에서는 단문형 영작문 문제를 자동으로 채점할 수 있는 자동 채점 시스템을 개발하였다. 이와 같은 자동 채점 시스템은 학교 현장뿐만 아니라 컴퓨터와 인터넷 사용이 가능한 곳이라면 어디서든 수시로 사용할 수 있다는 점에서 학생들의 영작문 실력 향상에 지대한 공헌을 할 수 있을 것으로 생각된다. 단문형 영작문 자동 채점은 여러 문장으로 구성되는 에세이 채점과는 달리 문장과 문장을 일대일로 비교해 가면서 세밀한 채점을 수행할 수 있다. 그렇기 때문에 영어 문법이 아직 완벽하게 완성되지 않은 저학년 학생들을 위한 유용한 시스템이 될 수 있다. 또한, 영작문 문항을 자동으로 채점해 줌으로써 학교 현장에서 영어 작문 시험의 출제 부담을 줄여줄 수 있습니다. 더불어 시스템에 의한 채점이기 때문에 인간 채점자가 채점을 하는데 있어서 가장 애로 사항인 채점의 일관성 유지를 보장해 줄 수 있습니다. 그럼으로써 학생들은 영작문 평가에 대한 더 많은 기회를 부여받을 수 있고, 결과적으로 영작문의 실력 향상을 기대해 볼 수 있습니다.

본 연구는 아직 개발 초기 단계로써 다음과 같은 개선이 요구된다. 첫째, WordNet에서 추출한 동의어 정보가 중학교 수준의 동의어 정보에 적합하지 않았다. 동의어 정보를 대상 영역에 맞도록 재조정해야 한다. 둘째, 문장 분석에서 발생하는 중의성을 해결하는데 좀 더 신중을 기해야 한다. 품사 중의성이나 구문구조 중의성을 해결할 수 있는 좀 더 효



(그림 9) 영작문 자동 채점 시스템의 인터페이스 화면

올직한 알고리즘이 사용되어야 할 것이다. 셋째, 오류가 중복되어 계산되는 상황들이 발생하는데 중복 오류를 검사할 수 있는 방법론이 고려되어야 할 것이다.

참 고 문 헌

[1] Christiane Fellbaum, WordNet An Electronic Lexical Database. The MIT Press, 1998.

[2] David A. Schneider and Kathleen F. McCoy. Recognizing Syntactic Errors in the Writing of Second Language Learners. In *Proceedings of the 17th International Conference on Computational Linguistics (COLING-ACL '98)*, 1998.

[3] Higgins, D., Burstein, J., Marcu, D., & Gentile, C. Evaluating multiple aspects of coherence in student essays (PDF). In *Proceedings of the Annual Meeting of HLT/NAACL*, 2004.

[4] Jill Burstein, Daniel Marcu and Kevin Knight, Finding the WRITE Stuff: Automatic Identification of Discourse Structure in Student Essays. *IEEE Intelligent Systems*, Vol 18, Issue 1, 32-39, 2003.

[5] Jill Burstein, Martin Chodorow and Claudia Leacock. Criterion Online Essay Evaluation: An Application for Automated Evaluation of Student Essays. In *Proceedings of the Fifteenth Annual Conference on Innovative Applications of Artificial Intelligence*, 2003.

[6] Jill Burstein and Derrick Higgins, Advanced Capabilities for Evaluation Student Writing: Detecting Off-Topic Essays Without Topic-Specific Training, In *Proceedings*

of the International Conference on Artificial Intelligence in Education, July 2005.

[7] Koldo Gojenola and Maite Oronoz. Corpus-based syntactic error detection using syntactic patterns. In *Proceedings of the workshop on Student research*. 24-29, 2000.

[8] M. Chodorow and C. Leacock. An Unsupervised Method for Detecting Grammatical Errors. In *Proc. First Meeting of the North American Chapter of the Association for Computational Linguistics (ANLPNAACL -2000)*, 140-147, 2000.



김 지 은

e-mail : jeeunk@hufs.ac.kr,
jeeunk@naver.com

1985년 한국외국어대학교 영어과 (학사)
1989년 Georgetown University
언어학과 (석사)
1993년 Georgetown University

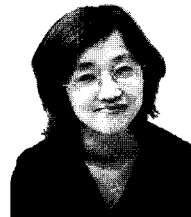
언어학과 (박사)

1993년~1994년 대학 강사
1995년~2002년 한국마이크로소프트(유) 연구원
2003년 한국외국어대학교 영어학부 강사
2006년~현재 한국외국어대학교 영어학부 강사
관심분야: 자연어처리, 전산 언어학, 코퍼스 언어학, 형태론

이 공 주

e-mail : kjoolee@cnu.ac.kr

1992년 서강대학교 전자계산학과(학사)
1994년 한국과학기술원 전산학과
(공학석사)
1998년 한국과학기술원 전산학과
(공학박사)
1998년~2003년 한국마이크로소프트(유)
연구원



2003년 이화여자대학교 컴퓨터학과 대우전임강사
2004년 경인여자대학 전산정보과 전임강사
2005년~현재 충남대학교 전기정보통신공학부 조교수
관심분야: 자연어처리, 자연어인터페이스, 기계번역, 정보검색

진 경 애

e-mail : kajin@kice.re.kr

1985년 한국외국어대학교 영어과 졸업
1990년 University of Pittsburgh 언어학
석사
1993년 University of Pittsburgh
영어교육 박사



1994-1998 LG 인화원 국제화 교육팀 과장
1998-2006 한국교육과정평가원 영어교육정책연구센터장
관심분야: 영어교육, 영어평가