*영어어문교육 13*권 1호 2007년 봄

영작문 자동 채점 시스템 개발 연구

진 경 애
(한국교육과정평가원)

Jin, Kyung-Ae. (2007). Development of automated scoring system for English writing. *English Language & Literature Teaching, 13*(1), 235-259.

The purpose of the present study is to develop a prototype automated scoring system for English writing. The system was developed for scoring writings of Korean middle school students. In order to develop the automated scoring system, following procedures have been applied. First, review and analysis of established automated essay scoring systems in other countries have been accomplished. By doing so, we could get the guidance for development of a new sentence-level automated scoring system for Korean EFL students. Second, knowledge base such as lexicon, grammar and WordNet for natural language processing and error corpus of English writing of Korean middle school students were established. Error corpus was established through the paper and pencil test with 589 third year middle school students. This study provided suggestions for the successful introduction of an automated scoring system in Korea. The automated scoring system developed in this study should be continuously upgraded to improve the accuracy of the scoring system. Also, it is suggested to develop an automated scoring system being able to carry out evaluation of English essay, not only sentence-level evaluation. The system needs to be upgraded for the improved precision, but, it was a successful introduction of an sentence-level automated scoring system for English writing in Korea.

[automated scoring system/English writing, 자동 채점 시스템/영작문]

Ⅰ. 배경

대학수학능력시험과 같은 대규모의 시험은 시행과 채점의 효율성 때문에 대부분 선택형 문항으로 구성되어 있으나 국가 수준의 학업성취도 평가, 중등임용고사와 같은 국가 고사와 각급 학교 수준의 시험에서는 서답형[1] 문항을 포함하고 있다. 서답형 문항은 문항이 지시하는 바에 따라 수험자의 생각이나 의견을 직접 서술 하도록 하기 때문에 창의성, 표현력, 문제해결력, 비판력, 논리적 일관성 등의 고등정신 능력을 평가하는 데 효과적이고 피험자의 학습 태도와 학습 방법을 개선하는 데 효과가 있다.

해외의 평가 전문 기관에서는 서답형 문항 채점에 드는 시간과 비용을 절감하기 위하여 서답형 문항 자동 채점 프로그램을 개발하여 활용하고 있는 데, 인간 채점자(human rater)와 자동 채점(automated scoring)과의 채점자 간 신뢰도가 인간 채점자들의 채점자 간 신뢰도 지수만큼 확보되고 있다는 연구 결과가 이미 많이 발표되었다(Khaliz, 2004; Shermis & Burstein. 2003). 따라서 우리나라에서도 이제 서답형 문항 자동 채점 시스템(automated scoring system)을 개발하여 활용하게 되면 사람이 채점할 때 드는 비용과 시간을 절약할 수 있으면서도 평가의 신뢰도나 타당도를 확보할 수 있어 매우 유용할 것이다. 본 연구에서는 영어 과목의 대표적인 서답형 문항인 영작문 문항 자동 채점 프로그램의 자동 채점 시스템 프로토타입 개발 과정 및 결과를 소개하고 향후 개발 방향에 대하여 논의하고자 한다. 해외에서는 이미 영어 에세이 자동 채점 프로그램이 상용화 되어 있으나, 우리나라 학생들의 경우에는 에세이 이전에 문장 단위의 영작문을 수행 할 수 있어야 하기 때문에 문장 단위의 영작문 자동 채점 프로그램을 개발하게 되었다.

Ⅱ. 문헌 연구

1. 자연어 처리 (Natural Language Processing)

영작문 자동 채점 시스템은 자연어 처리를 기반으로 하는 응용 시스템이다. 자연어 처리란 인간의 언어(natural language)를 의미 없는 단순 문자열이 아닌 “언어”로서 컴퓨터 시스템을 이용하여 처리하는 것이다. 즉, 영어와 같은 자연어 문장을

1) 서답형 문항은 답이 문항 내에 주어지는 것이 아니라 써 넣는 형태의 문항을 말하며, 논술형, 단답형, 괄호형, 완성형 등이 있다. (성태제, 2002)

컴퓨터 시스템에 입력하면 결과물로서 입력 문장이 전달하고자 하는 의미가 구현된 문장을 출력하게 된다. 이러한 출력 문장을 얻기 위해서는 대상 언어에 대한 전반적인 문법 지식, 즉, 어휘, 구문, 의미에 관한 지식(lexical, syntactic and semantic knowledge)이 포괄적이고 체계적으로 기술되어 있는 시스템을 구축해야 한다. 시스템에 구현된 어휘에 관한 지식은 형태소 분석 과정을 통한 사전 정보 검색으로 입력 문장 분석 시 활용된다. 구문 지식은 규칙으로 구현되어 구문규칙 사전에 저장되며, 구문 분석 과정(syntactic processing /analysis, 혹은 parsing) 시 입력 문장에 적용된다. 구문 분석의 결과물로 트리(tree)가 생성되며, 자연어가 갖고 있는 특성 중의 하나인 구문의 중의성(ambiguity)으로 인해 한 개의 구문 구조에 여러 개의 트리가 생성되는 경우가 대부분이다. 의미에 관한 지식 중 의미 제약 정보는 이러한 구문 트리의 중의성을 해소하는 데 적용되며, 그 결과물은 logical form으로 표현된다.

2. 해외의 자동 채점 프로그램

1) E-Rater

E-rater(Version 1.3) (Powers, Burstein, Chodorow, Fowles & Kukich, 2001)는 ETS에서 개발한 영어 논술 문항의 자동채점 도구로 6 단계로 점수를 부여한다. E-rater는 통계적인 접근을 하는데 인간채점자에 의하여 이미 채점된 논술의 샘플들을 사용하며, 답지를 읽는 방법은 자연언어 처리 방법을 사용한다. 인간의 채점을 따라하기 위하여 E-rater는 답지 내에서 다양한 변인들을 추출한 후 최소자승법에 의한 회기 계수 추정법(least squares regression weights)를 사용하여 점수를 산출하여 인간 채점의 결과와 마찬가지로 0-6단계의 점수를 예측하는 것이다. E-rater는 에세이 주제(prompt)에 따라 그 채점을 위한 변인들이 다르게 산출되지만 인간 채점자들이 사용하는 채점 기준표(rubric)의 내용과 논리와 어느 정도 부합한다. 그러나 초기 버전인 V 1.3의 경우 어떤 변인은 암묵적으로 다른 변인을 측정하거나 인간 채점과 일대일 대응이 되지 않은 경우가 있어, 새로운 모델을 제시하게 되었다. 새로운 E-rater 버전(V 2.0)은 V 1.3이 사용한 50개 정도의 총 변인 수 중 8-12개의 변인들만을 산출하여 회귀 분석을 하게 된다. 점수 산출을 위한 변인으로는 문법, 용법, 구두점, 스타일의 오류, 조직/전개, 단어의 복잡성, 주제에 적합한 단어 사용, 작문 길이 등이 있다. E-rater 채점 시 작문의 전체 길이는 인간 채점의 결과를 예측하는 가장 중요한 변수이다. 따라서 2.0 에서는 전체 단어수로 계산되는 작문의 길이를 변인 set에 포함하되 다른 변인들에게 미치는 영향은 최소화 하도록

하였다.

2) Criterion

Criterion(Burstein, Chodorow & Leacock, 2003)은 상용화된 온라인 영어 작문 교수 학습 프로그램이다. 2002년 9월 ETS는 E-rater를 웹 기반 실시간 버전인 Criterion의 채점 엔진으로 사용했다. 두 개의 컴퓨터 응용 프로그램: (1) E-rater (자동 채점 시스템) 와 (2) Critique(작문 분석 도구) 으로 구성되어 있으며, 에세이 자동 채점과 진단 피드백 제공의 두 가지 기능을 수행한다. E-rater는 자연언어처리를 기본으로 에세이의 전반적인 질을 평가하며, 이를 위해 관련된 언어학적인 요소들을 도출하여 통계적인 방법으로 점수를 산출한다.

Criterion은 기본적으로 교실에서 교사의 작문 교정에 대한 시간을 덜어 주는 목적으로 개발되었기 때문에 학생들이 작문 실력을 향상할 수 있는 기회를 무제한 적으로 제공한다. 교실이나 시험 상황보다 덜 위협적인 사이버 환경에서 작문을 연습하고, 수정하며, 작문 실력 향상 정도에 관하여 교사와 편안하게 의견을 교환할 수 있다. 교사들은 작문 주제(prompt)를 메뉴에서 선택할 수 있고, 직접 만든 주제를 추가할 수도 있다. Criterion은 초등학교 학생부터 고등학교 학생 및 대학원생에 이르기까지 다양한 수준의 작문 주제를 포함하고 있다. Criterion은 교사에게 점수 및 오류, 포트폴리오 등 다양한 정보를 제공하기 때문에 개별학생과 반 전체 구성원들의 진도 및 작문 실력 향상을 쉽게 파악할 수 있다.

2002년 출시된 이후로, Criterion 개별 학생에서부터 학교, 학군, 교육 기관, 온라인 학습 및 평가 제공업체와 관련 서비스 기관들을 포함한 200여개 기관에서 구입하여 사용하고 있다. 2002년 12월 말 50,000명의 사용자가 집계되었다. 사용자는 초, 중, 고 공립학교와 해외에서는 중국, 일본, 대만 등에서 사용된다. 초, 중고등학교 중에는 중학교에서 가장 많이 사용된다.

3) ALEK(Assessment of Lexical Knowledge)

ALEK(Leacock & Chodorow, 2003)은 ETS에서 학생들의 에세이에서 문법 오류를 감지해서, 진단 피이드백을 제공하여 학생들이 문법 오류를 수정할 수 있도록 하기 위해 개발 중인 자동 문법 감지 시스템이다. 시스템 개발은 코퍼스 및 통계 기법에 기초를 두고 있다. 특히, 두 가지 종류의 문법 오류를 감지하는 데 주력하고 있다. 첫째 오류는 일반적인 구문 규칙과 관련된 오류(Syntactic error, e.g. 관사-명사 일치, 서술어 오류 등)와 두 번째는 특정 단어의 용법과 관련된 오류(word-specific usage error,

e.g. 불가산 명사에 복수형 어미 -s를 붙인다거나 , 전치사 선택 오류 등)으로 크게 분류할 수 있다.

ALEK는 코퍼스(신문사용, 머리기사와 표등은 제외)로부터 문장을 추출하여 단어와 기능을 분리하여 각각의 기호(token)로 표기한다. 또한, 품사 태거를 사용하여 각각의 기호(token)에 정보를 부여하고, 바이그램(bigram)을 작성한다. 점수 산출은 전체 코퍼스에서 바이그램의 상대적 빈도수 산출하여 계산한다. ALEK에서는 오류 코퍼스(negative evidence)는 시간 낭비이고 효율적이지 못하며, 정문 코퍼스를 사용하는 것이 적절하다고 판단하여 정문 코퍼스만을 사용하였다.

지금까지 살펴본 해외의 상용화된 자동 채점 프로그램은 모두 에세이 수준의 프로그램이고 문장 단위의 채점 프로그램은 상용화 되어 있지 않다. ALEK은 문법 수정 및 진단용 프로그램으로 현재 개발 중이다. 따라서 본 연구에서는 해외의 프로그램을 참조하여 국내에서 문장 단위 자동 채점 개발 기술을 축적하고 동시에 우리나라 학생들의 수준에 맞는 시스템을 개발하고자 하였다.

III. 시스템 개발을 위한 데이터 구축

1. 오류 코퍼스 구축

본 연구에서 개발하는 영작문 자동 채점 시스템은 인간 채점자가 채점 기준에 의해 채점하는 논리로 시스템이 채점을 하는 방식을 취하기로 하였다. 따라서 시스템의 오류 처리 과정 및 채점 기준을 인간 채점자와 최대한 동일하게 하기 위해서, 인간 채점자가 채점한 다양한 문장들의 오류를 분석해서 오류의 경중을 나누고 그에 따른 가중치를 결정하는 오류 분석이 우선이다. 오류 분석을 위해서는 우선 지필 검사를 실시한 후 그 결과를 분석하는 절차가 필요하다. 지필 검사는 중학교 3학년 학생들을 대상으로 실시하였으며, 시스템에서 다양한 오류 규칙을 적용하기 위하여 중학교 3학년 수준의 주요 영작문 구문을 도출하여 문항을 만들기로 하였다. 중학교 3학년을 대상으로 한 이유는 국가 수준의 학업성취도 평가에서 중학교 3학년 수준에서 영작문 문항이 출제되기 때문이다.

주요 구문 도출을 위해서, 제7차 교육과정의 중학교 3학년 영어 교과서 중 여섯 가지 종류의 교과서(천재, 중앙, 지학사, 대한교과서, 교학사(심), 교학사(배))를 선택하여 공통 문법 영역을 분석하였다. 중 3 영어 교과서의 영어 문법 공통 영역은 ① 문장의 구성과 품사, ②문장의 종류, ③문장의 형식, ④동사의 시제, ⑤부정사, ⑥동명사, ⑦분사, ⑧조동사, ⑨수동태, ⑩관사, ⑪명사, ⑫대명사, ⑬형용사와 부사, ⑭비

교급, ⑮관계사, ⑯전치사, ⑰접속사, ⑱가정법, ⑲일치와 화법, ⑳특수 구문(도치, 강조, 부정, 생략, 삽입, 동격 등) 총 20개이다.

위의 20 가지 기본 문법 구조와 교과서의 예문들을 중심으로, 기본적인 구문형태로 비교적 평이한 문장으로 문항을 구성하였다. 총 20개의 문항이 개발되었는데, 각 문항의 내용은 우리말로 문장을 제시하면 영어로 작문 하는 형태이다. 문항들은 주요 문장 구조가 겹치지 않게 각 10 문항씩 Form A와 Form B에 수록되어, 지필검사에 사용되었으며 전체 학생들의 반은 Form A의 10 문항에 대하여 시험을 치르고 나머지 학생들은 Form B의 10 문항에 대하여 시험을 봤다. Form A와 Form B로 나눈 이유는 채점 및 오류 분석의 부담을 줄이면서도 총 20 문항에 대한 오류 정보를 얻기 위해서였다.

2. 지필 검사 시행

영작문 시험 Form A는 서울 A 중학교와 B 중학교의 중학교 3학년 학생들 (각 학교당 약 5개 학급, 총 282명)을 대상으로 실시되었고, Form B는 서울 C 중학교, 성남 E 중학교 (각 학교당 약 5개 학급, 총 306명)에서 실시되었다. 시험을 집행한 교사들은 모두 경력이 10년이 넘는 현직 중 3 영어 교사이며, A, C, E 중학교 교사의 경우 2004년도 학업성취도 문항 출제자이기도 하다. 시험은 정규 수업 시간에 선생님의 감독 하에 40분간 실시되었으며, 답안은 채점을 하지 않은 채로 오류분석을 위해 회수하였다.

3. 오류 유형화

오류의 유형은 형태소 오류, 구문 오류, 어휘 오류로 구분하였다. 형태소 오류는 문장 구조나 의미에 영향을 미치지 않고, 형태소 범위 내에서 발생하는 오류이다. 구문 오류는 문장 구조에 영향을 미치는 오류로서, 영문법 규칙의 적용 과정에서 잘못된 판단 및 모국어의 전이, 과일반화(over-generalization) 등의 원인에 의해 발생되는 오류이다. 어휘 오류는 문장 구조에 영향을 미치지 않는 범위 내에서, 의미에 영향을 미치는 단어 선택, 생략, 추가와 관련된 오류이다. 본 연구에서는 문장의 필수격 (주요 성분)에서 오류가 발생하면 구문 오류로 간주했다. 이 때, 주어, 서술어, 목적어, 보어 즉 문장의 필수 성분의 중심어(head)에 해당하는 단어가 생략되어 있으면 구문오류로 본다. 반면에, 명백한 내용(content)을 갖고 있음에도 불구하고, 문장 안에서 부가적인 요소의 생략일 경우 이를 내용(어휘) 오류로 본다. 다음의 예를 한 번 살펴보자.

Ex) He sent (his) (younger) (brother) an e-mail.

위의 4형식 예문에서, 간접 목적어의 Head 인 brother의 생략은 목적어 생략으로 보고, 간접 목적어의 부가성분이나 명백한 내용을 포함하고 있는 his와 younger의 생략은 어휘 생략으로 분류한다. 또한, 각 오류 유형은 중복되어 기술될 수 없다.

1) 형태소 오류

앞서 총 306명을 대상으로 한 총 20 문항(Form A와 Form B)영작문 시험 답안의 오류 분석 결과 우리나라 중학생들이 흔히 범하는 형태소 오류의 유형과 예문은 표 1과 같다.

표 1
형태소 오류

오류의 종류	예
명사복수형활용오류	toies, boxs, childs
명사수정보오류	sugars
동사활용오류	watchs, turnning, comeing
비교급/최상급 활용오류	bigest, tinyest
철자 오류	teh, translaion
confusable 오류	affect effect
idiom_usage오류	in the front of in front of
fmw (frequently misused word)오류	ninetyth

2) 구문 오류

예비 검사 결과로 도출한 오류 분석 결과 일반적으로 우리나라 중 3 학생들이 자주 범하는 구문 오류는 표 2와 같다.

표 2
구문 오류

SA level ERROR	예
관사오류	Playing with **the** fire is dangerous.
일치	I don't know why she **goes (->went)** there
동사오류	If I had money, I would **bought (->buy)**that computer
태	I was seen **(->saw)** much leaves yesterday.
조동사	You have better **(-> had better)** go home now
가정법	If I **would have** (->had) money, I could buy the computer.
주요성분 생략	He (sent) e-mail to his wife
전치사	Seoul is the city in which I was born **from.**
접속사	Although he is poor, **but** he is happy.
관계사	**Which (->What)** she said is true. She is teacher **who she** comes to my school last week.
어순	He is **famous more** than his father
비교급/최상급	He is **much (->more)** famous than his father. He is the **most** tallest boy.
준동사	I saw a lot of **falling (->fallen)** leaves yesterday. **Play (->Playing)** with fire is dangerous. It is dangerous **to playing (-> to play)**with fire.

3) 어휘 오류

어휘 오류는 대다수 단어 선택, 필요한 단어 생략, 또는 불필요한 단어 추가 등으로 분류할 수 있으며, 어휘 오류의 예를 살펴보면 아래 표 3과 같다.

표 3
어휘 오류

LA level ERROR	examples
단어선택	Sujin seems to be very ***healthful*** **(->healthy)** She is a teacher who ***went*** **(->came)** to our school last week.
단어생략	You had better go home **(now)** He sent **(his)** wife an e-mail I saw many **(fallen)** leaves in yesterday.
단어추가	The thing that she said **yesterday** is true. She was **new** teacher that she came last week.

4. 채점 기준

주요 구문을 중심으로 개발된 영작문 시험을 시행 한 후, 오류분석 자료를 토대로 시스템에 적용할 분석적 채점 기준을 수립하였다. 표 5의 채점 기준은 1 문장 기준이며, 학생이 기술한 문장에 전혀 오류가 없는 경우는 6점이 되고 3개 영역(어휘, 구문, 형태소)의 만점은 2점이 된다.

표 5
채점기준

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

Ⅳ. 영작문 자동 채점 시스템 개요

본 연구에서 개발한 영작문 자동 채점 시스템은 수험자의 작문 답안과 정답 문장을 비교함으로써 채점을 수행한다. 이러한 과정은 시스템 처리의 측면에서 보았을 때, 크게 두 단계를 거쳐 수행된다. 첫째, 수험자가 작성한 답안 자체에 문법적 오류가 존재하는지의 여부를 검사하고, 둘째, 수험자의 답안과 문제출제자가 제공한 정답집합과의 유사도(similarity)를 측정하는 것이다. 각각의 과정에서는 오류 검사 모델, 유사도 계산 모델 등이 필요하게 된다. 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. 시스템 처리 모듈은 크게 형태소 분석, 구문 분석, 정답과의 유사도 계산 모듈로 나눌 수 있다.

형태소 분석은 입력 문장의 각 단어에 해당하는 사전 정보를 적재하는 과정이다. 형태소 분석 과정을 통해 수험자 답안에 있는 단어 오류를 검사할 수 있다. 구문 분석 과정은 구문 규칙을 이용하여 입력 문장의 구문 구조를 파악하는 과정이다. 역시 이 과정을 통해 수험자 답안에 있는 구문 오류를 검사할 수 있다. 구문 분석의 결과는 구문 트리를 단순화시킨 의존구조(dependency structure)로 얻어진다. 일반적으로 영작문 문제에서의 정답은 하나 이상인 경우가 많다. 제시된 정답집합도 수험자 답안과 동일한 과정을 거쳐 의존구조로 변환된다. 수험자 답안의 의존구조와 정답집합의 의존구조들을 서로 비교하여 수험자 답안이 정답 답안과 얼마나 유

사한지를 계산해 냄으로써 영작문 자동 채점이 완료된다. 자동 채점 시스템의 최종 출력은 수험자 답안과 정답 답안과의 비교를 통해서 얻어진 유사도를 환산한 점수이며, 동시에 각 처리 단계에서 얻어진 오류의 보고이다.

1. 지식베이스(knowledge base)

사용하는 지식베이스는 어휘에 관련된 정보와 어휘 간의 동의어 정보, 그리고 구문 규칙으로 나눌 수 있다. 영어 문장을 분석하기 위해서는 영어 단어의 어휘 정보와 영어 구문 규칙이 필수 요소이다. 정답 문장과 수험자의 답안에 사용된 표현은 동일한 의미를 가지면서 서로 다른 표현으로 나타나는 경우가 빈번하다. 이와 같은 바꿔쓰기(paraphrasing) 현상을 다루기 위해서는 단어나 구(phrase) 간의 동의어 정보가 필요하다. 본 연구에서는 Princeton 대학에서 오랫동안 개발해 온 WordNet으로부터 동의어 정보를 추출하여 사용한다.

1) 어휘 사전

어휘사전에는 단어와 숙어에 해당하는 구(phrase)에 관한 모든 정보들이 저장되어 있다.

2) 구문 규칙

구문 규칙은 구구조 문법(phrase structure grammar)을 기본으로 하여 작성되었다. 규칙들은 영어의 정문(grammatical sentences)을 분석하는 일반 구문 처리 규칙과 구문 오류(syntactic errors)를 포함하고 있는 비문(ungrammatical sentences)을 분석하는 오류문 처리 규칙으로 구분된다. 각각의 규칙은 독립적 단위로 기술됨을 원칙으로 하였으나, 비문을 처리하는 규칙은 정문 처리 규칙 내에 조건(condition)으로 기술되거나 독립된 규칙으로 기술되었다.

3) 오류 규칙

오류 규칙은 입력문의 오류를 체크하여 오류 유형을 인식하는 역할을 하며, 두 가지 형태로 기술된다. 첫 번째는 일반 구문 규칙 내에서 조건문으로 기술되며 일반 규칙의 조건문(CONDITION)과 구분지어 조건문 내에 'WEAK'라는 레이블을 갖는다. 두 번째는 일반 구문 규칙과 같은 독립된 규칙으로 기술하되, 오류 규칙임을

명시하기 위해 조건문 내에 'ERROR'라는 레이블을 갖는다. 오류 구문 규칙은 오류가 발생할 수 있는 조건을 입력 문장에 대해 검사하고 오류 발생 조건이 만족된 경우, 발생한 오류에 대한 정보를 유지해야 한다. 이와 같은 작업을 수행하기 위하여 오류 규칙의 조건문에는 반드시 함수가 사용되어야 하는데, 이 함수의 역할은 오류 조건을 검사하고 조건을 만족할 때 오류 아이디 값을 부여하게 된다. 오류 아이디는 차후 사용자에 대한 오류 설명 등으로 활용될 수 있다.

4) 구축된 구문 사전의 크기

구문 규칙 사전의 전체 규칙 개수는 316개이다. 이를 규칙의 종류에 따라 나누어 보면 표 6과 같다. 레이블 'WEAK'를 갖는 규칙들은 일반 구문 규칙으로도 사용될 수 있고 오류 구문 규칙으로도 사용될 수 있기 때문에 전체 구문 규칙을 두 집합(일반 구문 규칙과 오류 구문 규칙)으로 나눌 때는 양쪽에 모두 포함된다.

표 6
구문 사전의 크기

순수 일반 구문 규칙 (NORMAL)	일반 구문 규칙 + 오류 구문 규칙 (WEAK)	순수 오류 구문 규칙 (ERROR)
235개	55개 (오류 종류: 76개)	26개 (오류 종류: 29개)
일반 구문 규칙: 290개	오류 구문 규칙: 81개 (오류 종류: 105개)	
전체 규칙 개수: 316개		

5) 동의어 정보

영작문에 대한 정답이 항상 여러 개인 이유 중 하나는 영작문에 사용되는 단어들을 항상 동의어로 대치할 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 시스템의 지식베이스에 동의어 정보를 갖고 있어야 한다. 본 연구에서는 Princeton 대학에서 오랫동안 개발해온 WordNet®에서부터 추출한 synsets 정보와 synsets 간의 계층구조를 이용하여 각 단어나 구의 동의어 정보로 사용한다.

WordNet은 단어 간의 관계를 표현하고 있는 일종의 어휘 데이터베이스로 네트

워크 형태로 구성되어 있다. 각 노드는 synset이라고 불리는 동일한 의미를 가지는 단어들의 집합이며, 링크는 synset들 간의 의미적 관계를 표현하고 있다. 사용되는 의미적 관계로는 반의어 관계, 의미간의 상하위 관계, 부분-전체 관계 등이 있다. WordNet에는 약 15만 개의 단어가 등재되어 있으며, 약 11만 개의 synset으로 구성되어 있다.

2. 자동 채점 처리 단계

시스템의 처리 단계는 크게 형태소 분석, 구문 분석, 정답과의 유사도 계산 과정으로 나눌 수 있다. 형태소 분석은 입력 문장의 각 단어에 대한 사전 정보를 적재하는 과정이며 구문 분석은 입력 문장의 구문구조를 파악하는 과정이다. 구문분석의 결과는 단어와 단어 간의 관계를 나타내고 있는 의존구조(dependency structure)이다. 정답의 의존구조와 수험자 답안의 의존구조를 비교함으로써 두 답안간의 유사도 및 점수를 계산해 낼 수 있다.

제시된 정답 문장들은 우선 분석되어 그 결과가 시스템 내부 파일에 저장되어 있어야 한다. 수험생이 답안을 작성하면 수험생의 답안만을 분석하여 시스템 내부 파일에 저장되어 있는 정답의 분석결과와 비교한다. 정답 문장에는 오류가 없다는 가정을 제외하고는 정답 문장들을 처리하는 과정과 수험자 답안을 처리하는 과정은 동일하다.

1) 형태소 분석과 형태소 단계 오류 처리

(1) 형태소 분석의 입력과 출력

형태소 분석의 목적은 입력 문장에서 발생한 각 단어의 사전 정보를 적재하는 것이다. 그렇게 함으로써 각 단어의 품사나 의미, 활용형 등과 같은 기본적인 정보를 얻을 수 있다. 형태소 분석 과정의 입력은 입력된 문장 스트링이 되며 출력은 각 단어에 어휘 사전 정보가 적재된 형태이다. 어휘 사전 정보에는 단어의 원형 정보만이 저장되기 때문에 ‘succeeded’와 같이 동사 활용을 한 경우에는 ‘-ed’를 떼어내고 사전 정보를 참조해야한다. 또한 ‘flowers’와 같이 명사가 복수형으로 사용된 경우에는 ‘-s’를 떼어내고 사전 정보를 참조해야 한다. 이러한 과정을 통해 입력 단어가 복수형으로 사용되었는지 과거형으로 사용되었는지와 같은 부가적인 정보를 획득할 수 있다.

(2) 형태소 분석

형태소 분석의 기본 목적은 어휘 사전 정보를 적재하는 것이다. 그러나 각 단어에 대한 사전 참조를 통해 사전 정보를 적재하기에 앞서 입력문장으로부터 정규화된 토큰(token)을 분리해내는 tokenization 작업을 우선 수행한다. 정규화된 각 토큰에 대해 사전참조를 통해 원형과 사전 정보를 적재한다.

(3) 형태소 오류 처리

형태소 분석 과정을 거친 후 사전정보를 적재하지 못한 입력 단어가 존재할 경우, 형태소 단계의 오류가 발생했음을 알 수 있다. 본 연구에서는 수험자의 입력 문장에는 언제든지 오류가 발생할 수 있다는 가정을 하기 때문에 입력 문장에서 형태소 오류가 발생했다 하더라도 그 오류를 인식하고 오류를 수정한 후 구문 분석 단계로 넘겨야 한다.

2) 구문 분석과 구문 단계 오류 처리

(1) 구문 분석의 입력과 출력

구문 분석은 입력 문장의 구문 구조를 파악하는 과정이다. 즉 입력 문장에서 주어부, 서술부와 같은 문장의 필수 구성요소들과 부사구 등의 부가 구성요소들을 인식하는 작업이다. 구문 분석의 입력은 형태소 분석의 결과이다. 즉, 어휘 사전 정보와 형태소 분석을 거쳐 얻어진 부가정보가 포함된 단어열이 구문 분석의 입력이다. 구문 분석 단계의 최종 출력 형태는 중심어(head)와 의존어(dependent) 사이의 관계로 표현된 의존구조이다. 더불어 구문 분석의 결과에는 입력 문장에 구문 오류가 있을 경우 구문 오류 정보가 적재된다.

(2) 구문 분석

구문 분석은 차트 기반의 상향식(bottom-up) 차트 파싱 알고리즘을 사용한다. 사용하는 구문 규칙이 구구조(phrasal structure)를 기반으로 하고 있기 때문에 구문 분석 단계의 출력은 기본적으로 구구조 형태의 구문 트리이다. 구조적 중의성으로 인하여 결과 구문 트리는 항상 여러 개가 발생 가능하다. 본 연구에서는 구문 규칙에 부여된 우선 순위 정보를 이용하여 여러 개의 구문 트리 중 가장 높은 점수를 부여받은 최대 N개의 구문 트리만을 최종 출력으로 간주한다.

본 연구에서 사용하는 구문 규칙은 정규 구문 규칙이나 오류 문장을 위한 오류 구문 규칙이 동일한 형태를 취하고 있기 때문에 이들을 구문 분석 단계에서 처리하는 방법은 크게 다르지 않다. 오류 구문 규칙이 사용된 구(phrase)에는 해당 구문

오류가 발생했기 때문에 구문 오류 정보를 담고 있는 'SA_ERR' 정보를 추가시킨다. 다음은 예제 문장 "He made her happy."에 대해 차트를 기반으로 구문 분석을 수행해 가는 과정을 보여준다. 전체 문장을 모두 포함하는 차트가 만들어지고 나면 구문 분석이 성공한 것이다.

<table>
<tr><td colspan="4">S (S NP VP)</td></tr>
<tr><td>NP</td><td colspan="3">VP (VP VP NP ADJP)</td></tr>
<tr><td></td><td>VP</td><td>NP</td><td>ADJP</td></tr>
<tr><td>pron</td><td>v</td><td>pron</td><td>adj</td></tr>
<tr><td>He</td><td>made</td><td>her</td><td>happy.</td></tr>
</table>

구문 분석의 입력은 형태소 분석 과정을 거쳐서 사전 및 형태소 정보를 갖고 있는 단어들로 구성된 문장이다. 그렇기 때문에 형태소 분석 과정에 발견된 오류 정보를 구문 분석 과정에서도 계속 유지하고 있어야 한다.

모든 구문 규칙에는 조건문(Condition)과 실행문(Action) 부분이 함께 저장되어 있다. 한 규칙에는 여러 개의 조건문과 실행문이 가능하다. 하나의 구문 규칙에 해당하는 구가 인식되면 (즉, 점표시가 규칙의 제일 오른쪽 끝에 위치할 때) 해당 구문 규칙을 활성화하기 전에 조건문들을 모두 평가한다. 조건문에서는 구에 대한 속성(feature) 정보들을 검사하고, 실행문에서는 구에 대한 속성 정보들을 다시 할당한다. 모든 조건문의 평가 결과가 참(true)일 때에만 해당 구문 규칙이 활성화되며, 이 때 실행문들을 수행한다.

(3) 구문 분석 단계의 오류 처리

입력 문장이 "A dogs are black."일 경우, 'A dogs'는 관사와 명사 사이의 수일치 오류를 포함하고 있다.

4) 의존 구조 형태로의 변환

구문 분석 과정에서 출력으로 나온 구구조의 구문 트리는 의존구조 형태로 단순화된다. 의존구조는 중심어(head)와 의존어(dependent), 그리고 그 두 단어 사이의 관계로 구문구조를 표현한다. 의존구조는 일종의 유향 비사이클 그래프(directed acyclic graph)이다. 노드는 문장에서의 중심어와 의존어가 되며, 관계는 노드 사이의 유향 에지(directed edge)로 표현되며 의존어에서 중심어로 방향이 설정된다. 의

존구조를 사용하면 구문 트리에서 나타나는 중간 구들을 제거함으로써 단순화된 구문 분석 결과를 만들 수 있고, 이렇게 단순화된 의존구조를 이용하면 정답 문장과의 비교가 좀 더 용이해질 수 있다.

5) 정답과의 유사도 계산 및 점수 계산

(1) 정답과의 유사도 계산 방법

정답과의 유사도를 계산하는 과정은 크게 세 가지로 나누어 볼 수 있다. 첫째, 수험자 답안이 정답집합 중 하나와 완전히 동일한 경우에는 수험자 답안을 분석하는 일이 무의미해 진다. 그렇기 때문에 우선 수험자 답안이 정답집합 중 하나의 정답과 완전히 동일한지 검사한다. 만약 완전히 동일한 정답이 발견될 경우 FULL_SCORE를 반환하면서 더 이상의 처리가 불필요하다. 동일한 정답이 발견되지 않은 경우에는 수험자 답안을 분석하여 수험자 답안의 의존구조와 정답집합의 의존구조들을 노드 대 노드로 비교해서 유사도를 계산한다. 간혹 구문 분석기나 구문 규칙의 불완전함으로 인하여 수험자 답안이 정답과 매우 유사함에도 불구하고 노드 비교를 통한 유사도 계산의 결과가 불완전한 경우가 발생한다. 이와 같은 문제에 대한 보완책으로 정답 문장과의 스트링 최소편집거리(minimum edit distance) 방식을 이용하여 유사도를 계산한다.

(2) 노드 비교를 통한 정답과의 유사도 계산 모듈의 입력 및 출력

유사도 계산 모듈의 입력은 구문 분석 결과인 의존구조인데 수험자 답안을 분석한 의존구조와 정답집합을 분석한 의존구조를 함께 입력으로 받아들인다. 수험자 답안의 문장이 항상 한 문장인데 비해, 정답집합의 문장은 한 개 이상이 가능하다. 수험자 답안의 문장은 항상 하나이지만 이를 구문 분석한 결과인 의존구조는 여러 개가 가능하다. 그렇기 때문에 유사도 계산은 N개의 수험자 답안의 의존구조와 M개의 정답집합의 의존구조의 조합으로 이루어진 N*M 개의 의존구조 쌍 가운데서 가장 일치도가 높은 두 개의 구조를 찾아내어 그 두 구조 사이의 유사도를 계산한다.

두 구조 사이의 유사도를 점수로 환산한 결과가 최종 출력이 되며, 동시에 수험자 답안의 구조에서 발생한 형태소 단계, 구문 단계의 오류와 두 구조 사이의 매핑 차이점이 부가적인 오류 정보로 출력된다.

(3) 점수 계산

본 연구에서는 형태소 단계, 구문 단계, 매핑 단계의 세 단계로 나누어 점수를

계산하였으며 이 세 단계의 점수를 합산하여 최종 점수화하였다. 각 단계마다 0~2점 사이의 점수가 할당되어 있기 때문에 최종 점수의 만점은 6점이다. 가장 유사도가 높은 두 개의 의존구조가 결정되고 나면 다음과 같은 휴리스틱을 이용하여 유사도를 점수로 환산하였다.

(4) 최소편집거리 기법을 이용한 정답과의 유사도 계산 모듈

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) 오류 세 종류가 가능하다.

정답 스트링이 “I sing the song”이고 수험자 답안이 “I sing song”일 때의 최소편집거리는 1(‘the’ 삭제)이다. 수험자 답안이 “I the song”일 때의 최소편집거리 역시 1(‘sing’ 삭제)이다. 두 경우 모두 최소편집거리가 1이지만 첫 번째 예제의 경우는 문장의 부수적인 성분(관사 ‘the’)이 삭제된 반면, 두 번째 예제의 경우는 문장의 필수성분인 주동사가 생략되었다. 그렇기 때문에 최소편집거리 알고리즘을 이용해 삭제된 단어가 문장의 필수성분(주동사, 주어, 직접목적어, 간접목적어)인지 아닌지를 검사한다. 필수성분인지 아닌지의 검사 방법은 삭제된 단어를 정답문장의 의존구조에서 찾아, 그 단어의 격정보를 검사함으로써 가능하다. 최소편집거리가 2이하이면서 삭제된 단어가 필수성분이 아닐 때, 최소편집거리에 따른 점수를 계산한다. 최소편집거리가 1인 경우, 최종점수는 FULL_SCORE-1을 최소편집거리가 2인 경우 최종점수는 FULL_SCORE-2로 계산한다. 다음은 최소편집거리를 통해 얻은 최종점수 예제들이다.

정답문장 : French is spoken in France.
수험자 답안: French is talked France.
오류종류 : 삭제(in) 대치(talked ↔ spoken)
최종점수 : 4점

정답문장 : He is more famous than his father.
수험자 답안: He is much famous than his father.
오류종류 : 대치(much ↔ more)
최종점수 : 5점

정답문장 : I don't know why she went there.
수험자 답안: I don't know why does she went there.
오류종류 : 삽입(does)
최종점수 : 5점

3. 자동 채점의 실제

다음은 영작문 자동 채점 결과 화면이다. 인터페이스는 개선의 여지가 많이 남아 있는 상태이다.

그림 1
자동채점의 실제-1

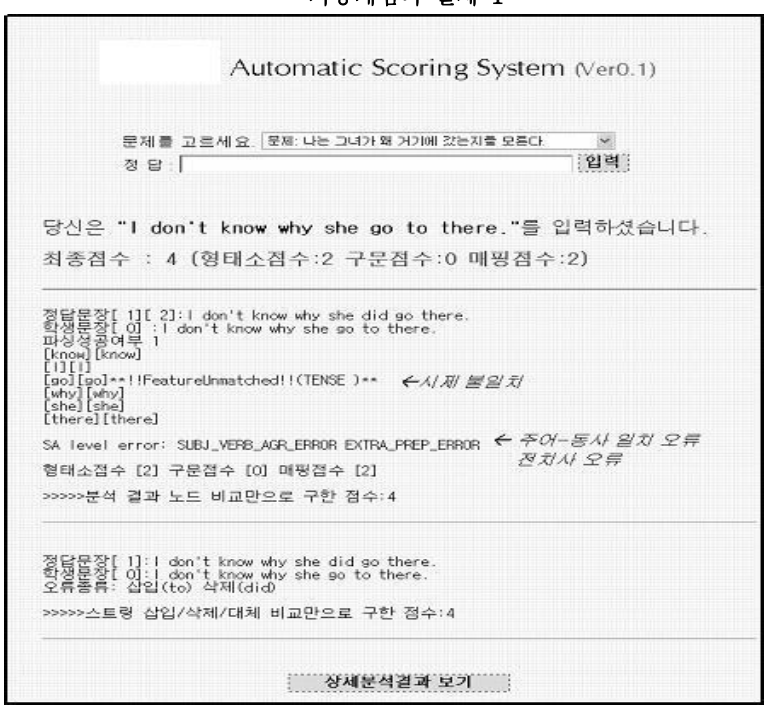
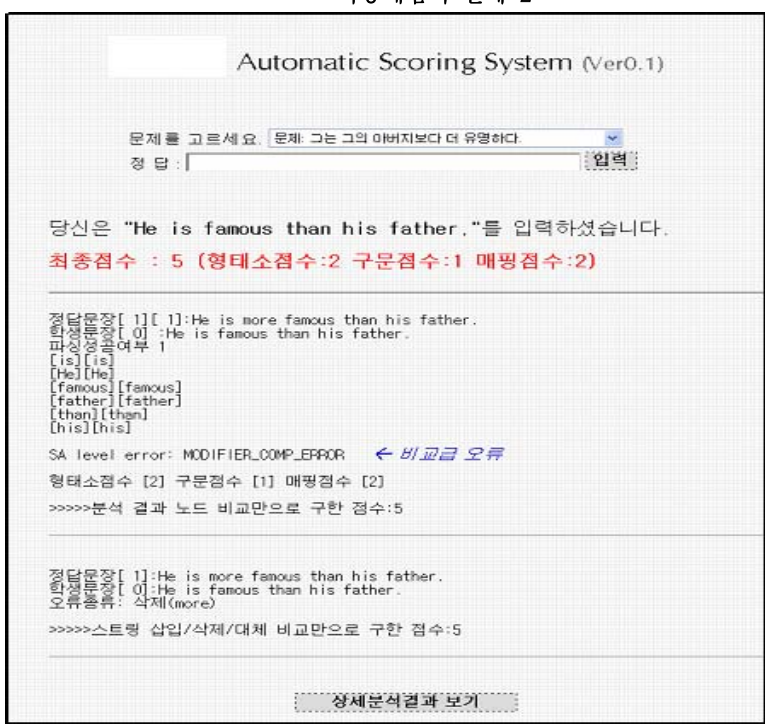


그림 2는 문장 “Seoul is I was burn”에 대한 분석 및 자동 채점 결과 화면이다. ‘burn’은 동사 ‘born’을 혼동하여 잘못 쓴 경우이다. 그러나 ‘burn’이 동사원형으로

존재하는 단어이기 때문에 be동사와 동사원형이 함께 사용된 'was burn'은 잘못 사용된 동사 형태이다. 그러므로 '동사 형태오류'가 발생하였고 관계대명사생략오류가 발생하였다. 또한 비교가 이루어진 정답 문장이 "I was born in Seoul"로 사용된 문형이 일치하지 않았기 때문에 문장패턴불일치오류가 발생하였다.

그림 2
자동채점의 실제-2



Ⅴ. 결론

서답형 문항 자동 채점 프로그램은 자연어 처리를 기반으로 서답형 문항의 채점 알고리즘을 적용하여 컴퓨터가 채점하는 시스템이다. 외국의 자동 채점에 대한 연구는 1960년대에 최초로 PEG가 개발된 이래 지금까지 지속적으로 연구 개발되어

지고 있다. 많이 알려진 프로그램으로는 인공 지능을 기반으로 에세이 채점을 하는 밴티지러닝사의 IntelliMetric 엔진과 자연어 처리 및 통계적 방법으로 에세이를 채점하는 ETS의 E-rater 등이 있는데 이미 많이 활용되어질 뿐 아니라 끊임없는 연구 개발로 정확성을 계속 높이고 있다. 서답형 문항은 피험자가 답안을 직접 구성함으로서 피험자가 가지고 있는 지식을 직접적으로 측정할 수 있다. 그러나 채점에 드는 비용과 시간으로 대규모 시험에 적용하는 데에 시행의 문제점을 안고 있다. 만약, 국내에서도 자연어처리 및 인공 지능에 대한 이론의 발전과 더불어 자동 채점 프로그램이 활성화된다면, 학생들의 종합적 사고력 및 수행 능력을 측정하는 데에 획기적인 역할을 수행하고, 학교 현장을 변화시키는 데에 큰 역할을 할 것이다.

영작문 자동 채점 프로그램의 경우 해외에서 개발된 프로그램을 살펴보면 영어를 모국어로 하는 피험자들의 논술 시험 (My Access, Vantage Learning 사) 과 영어를 외국어로 하는 학습자들의 교수 학습용 프로그램인 Criterion 등이 있다. 그러나 이러한 프로그램들은 영어 능력이 상당 수준 이상이 되어서 에세이를 쓰는 수준의 사용자들의 능력을 측정하기 때문에 영어를 학습하는 기초 단계의 영어 작문 능력 측정에는 적절하지 않을 수 있다. 본 연구에서는 중학교 학생들을 대상으로 하여 학업성취도 평가 등에 활용하거나, 교수 학습에 적용할 수 있는 영작문 문항 자동 채점 프로그램을 개발하는 것으로 목표를 수립하였기 때문에 이를 위한 프로그램을 국내 자연어 처리 전문가의 기술로 직접 개발하였다.

자동 채점 프로그램을 개발하는 절차에 대한 경험이 없는 상태에서 연구를 진행하며 그 개발 절차를 확립하여 갔다. 영작문 자동 채점 프로그램을 개발하기 위하서는 우선 문항 개발, 채점 기준 수립, 영작문 시험 시행을 통한 오류 코퍼스 구축을 필요로 한다. 중학교 3학년 검인정 교과서를 분석하여 주요 영어 구문을 추출하고 이를 바탕으로 총 20개의 영작문 문항을 개발하여 시험을 시행하고 그 결과를 분석하여 오류 코퍼스를 구축하였다.

자동 채점 시스템에서 구현될 점수 산출 프로그램은 회귀 분석이나 베이지안 방법으로 점수를 예측하는 방법이 있고, 전문가들에 의하여 시스템에 점수 산출 공식을 직접 구축하는 방안이 있는데, 본 연구에서는 영작문 문법과 오류 규칙을 토대로 공식을 직접 입력하는 방식을 취하였다. 영작문 문항의 단위가 문장 단위여서 피험자가 구성한 각 문장에 어떻게 점수를 부여하는 지가 관건이었다. 문장 단위의 작문이라도 단순히 맞고 틀리고의 기준으로 점수를 주지 않고, 오류의 종류와 횟수에 따라 0점에서 6점으로 점수를 부여하는 것으로 하였다. 채점 영역은 구문, 형태소 및 어휘로 하여 각각 2점 만점으로 하였으며, 이 채점 기준에 의하여 인간 채점자들이 채점을 하였으며, 동시에 시스템의 점수 산출 알고리듬 정답과의 유사도 요소와 함께 적용되었다.

본 연구에서 개발한영작문 채점 시스템은 수험자의 작문 답안과 정답 문장을 비교함으로써 채점을 수행한다. 이러한 과정은 시스템 처리의 측면에서 보았을 때, 크게 두 단계를 거쳐 수행된다. 첫째, 수험자가 작성한 답안 자체에 문법적 오류가 존재하는지의 여부를 검사하고, 둘째, 수험자의 답안과 문제출제자가 제공한 정답집합과의 유사도(similarity)를 측정하는 것이다. 시스템 처리 과정에서 사용되는 지식베이스(knowledge base)는 사전(lexicon)이며 사전에는 어휘 정보와 구문 규칙, 그리고 단어 또는 구(phrase) 간의 동의어 정보가 저장되어 있다.

본 연구에서 개발한 자동 채점에 의한 내적일관성 계수는 A형의 경우 0.92, B형의 경우 0.87 정도로 나와 신뢰도가 높은 것으로 나타났다.

1차년도에 개발한 영작문 자동 채점 시스템의 한계 및 향후 연구 방향은 다음과 같다. 현재 버전의 시스템은 아직 개발 초기 단계이기 때문에 어휘 사전, 구문 규칙 사전 및 시스템 전반에 걸쳐 개선할 부분들이 많이 남아 있다. 향후 시스템 보완의 방향은 다음과 같다. 첫째, 좀 더 다양한 문형을 다룰 수 있도록 구문 분석 규칙을 보강해야 한다. 둘째, 채점 시스템의 성능은 정답 문장과 수험자 답안 문장을 얼마나 잘 매핑 시키느냐에 달려있다고도 해도 과언이 아니다. 그러나, 매핑에는 다양한 예외 상황이 발생할 수 있기 때문에 이러한 예외 상황을 체계적으로 처리할 수 있는 방법에 대한 연구가 필요하다. 셋째, 스펠링 오류가 발생한 경우에는 오류를 반드시 복원하여 그 다음 처리 단계에 복원된 형태를 넘겨주어야 하는데 정답 문장을 이용하여 스펠링 오류를 복원할 수 있는 효율적인 방법에 대한 연구가 필요하다. 넷째, 입력 문장에 오류가 있을 경우에도 유효하게 적용할 수 있는 품사 중의성 해소 방법에 대한 연구가 필요하다. 다섯째, 구문 오류가 발생한 구문 트리와 구문 오류가 발생하지 않은 구문 트리가 동일한 문장에 대해 함께 발생한 경우, 오류를 고려하면서 정답과 더 유사한 구문 트리를 선택하는 방법에 대한 연구가 필요하다. 여섯째, 자동 채점 결과로써 나오는 오류 메시지를 사용자에게 유용한 정보로 바꾸어서 해 줄 수 있도록 인터페이스를 보강하여야 한다.

아직 프로토타입의 형태이나 국내에서, 규칙 적용 방법을 통하여 영작문 자동 채점 프로그램을 개발한 것은 그 시사점이 매우 크다고 할 수 있다. 본 연구에서 개발된 프로그램은 문장 단위의 영작문 채점으로 해외에서 기 개발된 에세이 자동 채점과는 차별성을 갖고 있으며, 우리나라 학생들을 대상으로 하는 영작문의 기본 문형을 익히고 평가 하는데 더 적절할 수 있다. 해외에서 수십년에 걸친 연구 개발로 그 정확도를 높여 상용화 한 것 같이 우리의 기술로 개발된 영작문 채점 프로그램의 보완 및 발전을 위한 지속적인 연구 개발 및 지원이 필요하다.

참 고 문 헌

김부자. (2005). 형태에 초점을 맞춘 교수가 영어 쓰기 능력에 미치는 영향. *영어어문교육, 11*(5), 77-98.

성태제. (2002). *현대교육평가*. 서울: 학지사.

민찬규, 김보경. (2006). 협동학습을 활용한 고등학교 영어 쓰기 지도 효과. *영어어문교육, 12*(4), 185-210.

Burstein, J., Chodorow, M., & Leacock, C. (2003) An Application for Automated Evaluation of Students Essays. *Proceedings of the Fifth Annual Conference on Innovative Application of Artificial Intelligence* (pp. 3-10). Mexico: Acapulco.

Leacock, C. & Chodorow, M. (2003) Automated Grammatical Error Detection. In M. D. Shermis & J. C. Burstein (Eds.), *Automated Essay Scoring: A Cross-Disciplinary Perspective* (pp. 195-207). NJ: Lawrence Erlbaum Associates.

Powers, D. E., Burstein, J. C., Chodorow, Fowles, M. E., & Kukic, K. (2001). *Stumping E-rater; challenging the validity of automated essay scoring* GRE No. 98-08bP, ETS, NJ 08541.

Khaliz, S. N. (2004). *An evaluation of automated scoring programs designed to score essays.* Unpublished doctoral dissertation, University of Massachusetts.

Shermis. M. D., & Burstein. J. (2003). *Automated essay scoring: A cross-disciplinary perspective.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

부 록

우리 나라 중학생들이 범하는 영어 오류 빈도 및 종류

순서	오류제목	예
1	관사오류	The **French** is spoken in France.
		He saw that **the Jane** swimming in the river
		Playing with **the fire** is dangerous.
		I learn English for **a** seven **years**.
		I saw many leaves **a yesterday**
		He is (the) tallest in his class
		He sent **(an) e-mail** to his wife.
		He sent ***a*** **(->an) e-mail** to his wife.
2	시제 불일치	**I *have seen* (->saw)** many leaves yesterday.
		I ***learn*** **(->have learned)** English for a seven years.
		She is a teacher who ***come*** **(->came)** to our slchod last week
		I don't know why she ***goes*** **(->went)** there
		What she ***say*** **(->said)** is true
		Su-jin ***is looking*** **(->looks)** so healthy
		We ***enjoy*** **(->enjoyed)** listening to music
		He ***was*** **(-> is)** famous than his father.
3	술어동사 오류	**He *can speaks* (-> can speak)** not only English but also Franch.
		I don't no why she ***going*** **(->went)** there
		I don't know why she ***is go*** **(->went)** there
		French **is speak (->is spoken)** in French
		He **is speak** of not only English but also French
		He see Jain **is swam (-> swam)** in river.
		If I had money, I **would bought (->buy)** that computer
		He ***can speaking*** **(-> can speak)** both English and French.
		You'd better **to go (->go)** home now.
		I ***had studing*** **(->have studied)** English for 7 years.
		I ***have learn*** **(->have learned)** English for 7 years.
		I ***learning*** **(->have learned)** English for 7 years

4	수 일치/ 격 일치	There ***are*** (->is) no water to drink.
		He sent e-mail for ***him*** (->his) wife
		French ***are*** (->is) spoken in France.
		*비교급+ than any other +단수명사 He is taller than any other ***students*** (->student) in his class.
		If I have money, I buy ***these*** (->this) computer.
5	태	I ***was seen*** (->saw) much leaves yesterday.
		Sujin ***is looked*** (->looks) very healthy.
		He ***was sent*** (->sent) E-mail to wife
		French ***is speaking*** (-> is spoken) in France
6	조동사	You ***have better*** (-> had better) go home now
		He ***does*** (->can) speak English well and French well too.
		If I **would** have money, I could buy the computer.
7	서술어 생략	**He (sent)** e-mail to his wife
		Playing with fire (is) dangerous.
		Su-jin seems to (be) very healthy.
		I don't (know) why she go there.
		Seoul is the city where I **(was)** born
8	주어 생략	**I don't know the reason why (she)** went there
		If I have money, **(I)** buy the computer.
9	전치사	**Seoul is the city in which I was born from.**
		He is the tallest **(in)** his classroom
		He saw the Jane swim **(in)** the river
		We enjoyed listening **(to)** music.
		She is a teacher who came **(to)** our school last week
		I learned English **(for)** seven years.
		He saw Jane swimming **at (->in)** the river.
		He sent E-mail ***for*** (->to) his wife
10	접속사	Mabye (->If) I have money, I can buy the computer.
		He saw ***what*** (->that) Jane swim in reaver.
		Although he is poor, ***but*** he is happy.
11	관계사	(What) She say is fact.
		Seoul is the city **(which)** I was born in.

		She is a teacher **(who)** come to my school last week
		I don't know **(why)** she went there.
		Which **(->What)** she said is true.
		He saw Jane swim **where** in the river.
		She is teacher who ***she*** comes to my school last week.
12	어순	**I saw many leaves falling** yesterday.
		He sent **to his wife** an E-mail.
		we enjoyed **to listening** music.
		He is **famous more** than his father
13	비교급	**He is much (->more)** famous than his father.
		He is more **famously** than his father.
		No one is ***tall*** **(->taller)** than him in his class.
		He is **famouser (-> more famous)** than his father.
		He is famous.**(than)** hisfatcher.
		He is **(more)** famous then his father
14	최상급	**He is the most** tallest boy.
		He is ***the most tall*** **(->the tallest)** His classroom
		He is taller than **(any other)** his classrom students.
15	분사	**I was saw many *fall* (->fallen)** leaves yesterday.
		I saw a lot of ***falling*** **(->fallen)** leaves yesterday.
16	동명사	***Play*** **(->Playing)** with fire is dangerous.
		enjoy +동명사 대신-> to 부정사 or 동사원형: we enjoy ***to listen*** **(->listening)** to music
		we enjoyed **listen (-> listening)**to music
17	to 부정사	**It is dangerous *that (->to)*** play with fire
		It is Dangerous ***playing*** **(-> to play)** with fire.
		It is dangerous ***to playing*** **(-> to play)**with fire.
		He saw that Jane ***to swim*** **(->swim)** in the river.
		I don't have watter ***for*** **(->to)** drink
18	형용사/부사	**There is *not*** **(->no)** water to drink
		There are ***not (->no)*** boys in the class taller than he
		He is ***happly*** **(->happy)** though he is Poor
		it is ***danger*** **(->dangerous)** to play with fire

		I saw lots of ***dying*** **(->dead)** leaves yesterday.
		2형식동사+형용사보어앞 불필요한 전치사: She looks ***like*** very healthy.
		I saw ***much*** **(->many)** leaves yesterday.
19	**목적어 생략**	**He sent e-mail to (his wife)**
		I saw the many (fallen leaves) yesterday.
		I've learned **(English)** for seven years
20	**주어 동사 반복**	**He watched Jane that she** swimed in river
		Seoul **I** was born in city.
		If **you I** have a money, I bought it computer
		Not only **does** he **can** speak English, but also French,
		I don't know why ***does*** she went there.
		Su-jin looks feel great.

예시 언어(Examples in): English
적용 가능 언어 (Applicable Languages): English
적용 가능 수준 (Applicable Levels): Secondary

진경애
한국교육과정평가원
110-230 서울 종로구 삼청동 25-1
Tel: 02-3704-3653
Fax: 02-3704-3558
Email: kajin@kice.re.kr

Received in Jan. 2007
Reviewed by Feb. 2007
Revised version received in Mar. 2007