
중학생 영작문 실력 향상을 위한 자동 문법 채점 시스템 구축

Implementing Automated English Error Detecting and Scoring System for Junior High School Students

김지은*, 이공주**

한국의국어대학교 영어대학 영어학과*, 충남대학교 공과대학 전기정보통신공학부**

Jee-Eun Kim(jeeeunk@naver.com)*, Kong-Joo Lee(kjoolee@cnu.ac.kr)**

요약

본 논문에서는 중학생이 작성한 영작문에서 자동으로 문법오류를 검색하고 채점하는 시스템을 소개한다. 학생의 문장이 입력되면 형태소 및 구문 분석을 하고 오류를 검색한다. 문장 분석이 완료되면 교사들이 제공한 채점기준에 의해 자동으로 채점한다. 문법오류를 탐지하기 위해서 정문을 처리하는 규칙은 물론 오류를 포함하고 있는 문장도 처리하는 규칙을 구현하였다. 본 시스템에서는 영어를 제2외국어로 사용하는 학생들이 한국어의 영향으로 인해 발생시키는 영어 구문오류를 집중적으로 처리하고자 하였다. 이와 같이 영작문에 대한 자동 채점은 학생들에게 오류에 대한 즉각적인 피드백을 제공하여 스스로 자신의 영어실력을 모니터링할 수 있게 해 준다. 이러한 시스템을 활용함으로써 학생들은 언제 어디서나 혼자서 영작문 공부를 할 수 있으며, 이는 학생들의 실력 향상에 많은 도움이 될 것이다.

■ 중심어 : | 문법오류 | 오류 검색 | 오류 탐지 규칙 | 자동 채점 | 영어 실력 향상 |

Abstract

This paper presents an automated English scoring system designed to help non-native speakers of English, Korean-speaking learners in particular. The system is developed to help the 3rd grade students in junior high school improve their English grammar skills. Without human's efforts, the system identifies grammar errors in English sentences, provides feedback on the detected errors, and scores the sentences. Detecting grammar errors in the system requires implementing a special type of rules in addition to the rules to parse grammatical sentences. Error production rules are implemented to analyze ungrammatical sentences and recognize syntactic errors. The rules are collected from the junior high school textbooks and real student test data. By firing those rules, the errors are detected followed by setting corresponding error flags, and the system continues the parsing process without a failure. As the final step of the process, the system scores the student sentences based on the errors detected. The system is evaluated with real English test data produced by the students and the answers provided by human teachers.

■ keywords : | Automated English Scoring System | Korean-Speaking English Learners | Improving English Grammar Skill | Detecting Grammar Errors | Error Production Rules |

I. 서론

영어를 학습하고 있는 학생들의 관심사 혹은 목표는 오류가 없는 완벽한 영어를 구사하는 데 있을 것이다. 한 언어를 완벽히 구사한다는 것은 발음 혹은 철자부터 시작하여 문장의 완성을 통해 에세이 혹은 회화로 자신의 생각을 자유로이 전달하는 것이라 할 수 있다. 이러한 목표는 여러 가지 방법을 통해 달성될 수 있었으나 우선적으로 기본적인 문법지식을 터득하는 것이 선행되어야 할 것이다. 또한 오랜 시간을 요구하는 언어습득의 특성상, 지속적이고 반복적이라는 학습방법도 병행되어야 한다. 이러한 이유로 정확한 문법지식의 습득을 위해서 학생들이 언제 어디서든 자신의 문법오류에 대한 즉각적인 교정과 피드백 받을 수 있는 환경이 제공될 수 있다면 효과적인 것이다. 그러나 교실이나 학원과 같은 교육 현장에서 벗어나면 이러한 도움은 기대하기 어려운 것이 현실이다. 교육 현장이라 할지라도 개개인에게 즉각적인 피드백을 필요할 때마다 줄 수 없는 것도 또한 우리의 현실이다. 이런 현실에 대한 대안으로서 문법검사기와 같은 시스템을 이용한다면 학생 스스로 자신의 문법실력을 테스트해 보고 즉각적인 오류 교정과 피드백을 자신이 원하는 시간에 받을 수 있다. 또한 학생 스스로가 자신의 문법실력을 테스트 해 보고 지속적인 학습을 통해 자신의 실력 향상을 모니터링 할 수 있으므로 더욱 효과적이다[12].

영어문법을 터득한다는 것은 영어의 구문구조를 이해하고 이에 익숙해지며 풍부한 어휘를 사용함에 있어 적절하게 사용하는 것이라 할 수 있겠다. 구문구조의 이해는 구문규칙, 단어의 철자 및 형태, 문장부호 사용의 이해로 세분할 수 있다[9]. 기존의 문법검사기들은 이러한 문법요소들을 대체적으로 탐지하여 어느 정도의 피드백을 주는 기능을 갖추고 있다. 하지만 대부분의 기존 문법검사기들은 영어를 모국어로 하는 사용자를 대상으로 구축되었다. 영어를 제2외국어로 공부하는 학생들은 영어를 모국어로 하는 사람들과는 다른 종류의 오류를 만든다. Ellis(1985)는 언어습득 과정에서 오류가 발생하는 것은 필연적이며 모국어의 영향에 따라 오류형태도 달라질 수 있다고 하였다[5]. 이처럼 한국

학생들은 한국어의 영향을 받은 오류를 발생시키거나 기존의 문법 검사기로는 이러한 오류가 탐지되지 못하는 것이 문제점으로 지적될 수 있다.

본 논문은 중학교 3학년 학생들이 작성한 문장 단위의 영작문에서 자동으로 오류를 탐지하고 채점하는 시스템을 소개한다. 자동채점 시스템은 위에서 언급한 효과적인 학습 환경의 제공 이외에도 영작문을 평가함에 있어 인간 채점자의 주관적이며 일관성(Consistency)이 결여된 채점방식을 객관적이며 일관된 방법으로 대체할 수 있다는 장점을 갖고 있다. 또한 인간채점자가 반복적으로 계속되는 다량의 작문을 채점하는 데 드는 시간과 노력을 최소화 시켜 준다. 이러한 시스템은 적절한 채점방법 기준의 책정(과1) 인간채점자 훈련이 선행되어 구현된다면 더욱 효율적이 될 것이다.

본 시스템에는 한국어를 모국어로 하는 학생들이 작성한 영어문장의 문법을 검사하는 규칙을 구현하였다. 이러한 구문규칙을 구현함에 있어 정문(Grammatical Sentences)을 처리하는 규칙은 물론, 문법오류를 포함하고 있는 비문(Ungrammatical Sentences)을 처리하는 비문규칙 혹은 오류규칙을 기술하였다. 또한 구현된 구문규칙과 시스템을 통해 탐지된 학생들의 오류를 분석하였다. 2장에서는 간략하게 관련연구를 둘러보고 3장에서는 전체시스템의 개요를 설명한다. 4장에서는 구문규칙의 종류 및 구현 방법에 대해 기술한다. 5장에서는 오류의 범위와 언어습득 과정에서 바라본 오류 분석 및 오류 구현에 대해 논하고 시스템에 관한 평가를 한다. 6장에서는 시스템의 향후 개선책과 연구 방향을 제시하고 결론을 맺는다.

II. 관련 연구

최근 영어를 대상으로 문법을 검사하거나 채점을 하는 시스템에 관한 연구, 혹은 이를 상용화 한 시스템들을 어렵지 않게 접할 수 있다. GRADE(GRAMmar Diagnostic Expert System, 2002)라는 시스템은[1] 일

1) 본 시스템에 구현된 채점 기준은 현직 중학교 교사들에 의해 책정되었다.

본어를 모국어로 하는 영어 학습자를 위해 개발된 문법 오류 탐지 시스템이다. 이 시스템은 문장 전체를 분석하기보다는 17가지 종류의 오류를 기술하고 이 오류만을 검색한다. 또한 시스템의 사전 올림말 개수는 219개이며 종속절이나 내포문 등을 포함한 복문 혹은 등위 접속사로 연결된 중문, WH-의문문, 감탄문 등은 처리 범위에서 제외된다. 이 밖에도 명사구의 길이에 제한이 있고 입력문에 'I've'와 같은 축약형도 역시 처리 범위에 포함되지 않은 제한적인 시스템이다.

Schneider & MaCoy(1998)는 미국의 수화(ASL, American Sign Language)를 모국어로 하는 농아들을 위한 영어 문법오류 검색기를 개발하였다[10]. 오류를 탐지하기 위해 오류규칙(Mal-rules)을 구현하였으나 역시 복문, 중문은 제외한 단문만을 처리한다. 또한 한 개의 문장에 한 개의 오류만이 발생한다는 가정 하에 제한적 문법을 구현하였다. 시스템의 자체 평가에 사용된 문장도 180개로 범용 시스템으로 가기에는 보완이 필요해 보인다.

Gojenola & Oronoz(2000)는 자동채점 시스템의 일부인 구문분석 및 구문오류 탐지 기능에 관하여 기술하였다[7]. 이 시스템은 구문분석에 부분 구문규칙(Partial Grammar)을 사용하였으므로 결과는 문장이 아닌 구절 단위(Phrase Level)이다. 오류 검색에 있어서도 문장의 구문오류가 아닌 낱자 표현 방식에 관한 오류만을 다루고 있다.

ETS(Educational Testing Service)는 온라인 영작문 채점 시스템인 CriterionSM이란 시스템을 상용화 하고 있다[2]. CriterionSM은 영작문 중 에세이를 대상으로 다양한 구문 요소(Syntactic Variety)의 사용, 에세이의 구성, 내용의 주제와의 부합 여부, 고난위도의 어휘 사용 여부 등을 채점 기준으로 사용하였다. 이 외에 문법, 용법(Usage), 구두점, 자주 혼동되는 단어의 정확한 사용, 작문 스타일을 검사하고 오류가 발견 되면 피드백을 준다. 문법오류에 대한 피드백은 대량의 코퍼스에서 추출한 N-그램 정보를 이용한다.

III. 전체 시스템 개요

본 연구를 통해 개발된 영작문 자동채점 시스템은 문장에 대해 교사가 제공한 한 개 혹은 여러 개의 정답문과 학생이 작성한 문장을 필요로 한다. 우선, 교사가 제공한 정답문에 대해 형태소 분석 및 구문 분석을 수행하고 그 결과를 단순화된 의존구조로 변환하여 저장해 놓는다. 본 시스템을 이용하는 학생이 작성한 영어 문장이 입력되면 우선 입력문이 교사에 의해 주어진 정답문 중 하나와 정확하게 일치하는지를 확인한다. 일치할 경우 규칙을 적용할 필요가 없다. 정답문의 어느 문장 과도 일치하지 않을 경우 입력문에 대한 분석이 정답문과 동일한 방식으로 진행되어 역시 의존구조로 변환된다. 분석 과정 중 오류가 탐지되면 오류의 종류를 확인하고 표지(Error flag)를 한 후 문장 분석을 계속한다. 입력문의 분석이 종료되면 두 의존구조를 비교하여 두 구조 간의 차이를 찾아내고 유사도를 계산하여 점수를 매긴다. 시스템의 결과물은 이와 같은 처리과정을 통해 얻은 학생의 답안에 대한 점수이며 탐지된 오류에 대한 설명이다.

[그림 1]은 본 연구에서 개발한 시스템의 전체 처리 과정을 보여 준다. 본 시스템은 Linux Fedora Core 5에서 C언어를 사용하여 개발되었다.

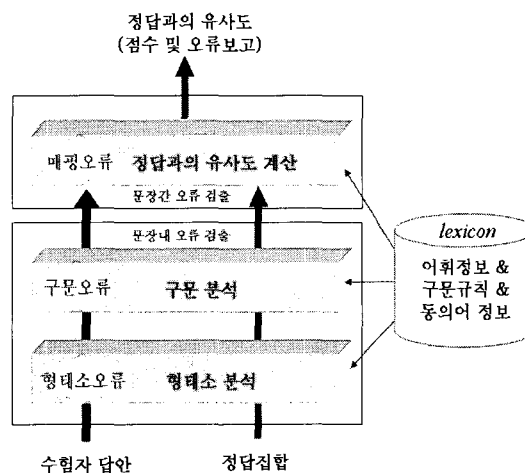


그림 1. 시스템 개요

1. 오류 탐지 방법

본 시스템의 오류 탐지는 두 가지 방법으로 수행된다. 첫 번째는 시스템에 기술된 규칙에 의해 탐지될 수 있는 오류를 찾아낸다. 입력문에 대해 분석을 수행하면서 구문오류, 철자 및 형태오류, 구두점 사용오류 등을 걸러낸다. 두 번째 방법은 입력문을 정답문과의 비교를 통해서 오류를 탐지한다. 이는 문장 차체에는 문법오류가 없다 하더라도 정답문이라 간주될 수 없는 문장도 있기 때문이다.

표 1. 정답문과 학생 답안의 비교

정답문	학생 답안
She sang a song with her younger sister.	She sang a song with a sister. She sings a song with her younger sister.
You had better sleep now.	You should sleep now.
He is the man who came to our house yesterday.	He is the man who came.
He speaks English fluently.	He says English fluently.

[표 1]에서처럼 학생 답안에는 문법적 오류는 없으나 정답과 다른 시제의 사용, 정답으로 요구되는 내용어의 미사용, 혹은 부적절한 동의어의 사용, 불충분한 내용 기술 등이 포함되어 있다. 이와 같은 문장을 처리하기 위해서는 정답과 비교하는 것이 필요하며 비교를 통해 입력문에 대한 정확한 채점이 이루어질 수 있다.

표 2. 복수 정답문

She is the tallest in her school.
She is taller than any other student in her school.
She is the tallest one in her school.
No one is taller than her in her school.
There is no one taller than her in her school.

시험문제의 문장은 대개 [표 2]의 예문들처럼 다수의 문장이 정답문으로 간주되므로 가능한 모든 종류의 정답문을 제공해야 한다. 정답문이 여러 개이면 이들 중 학생의 입력문과 가장 유사한 문장을 선택하고 두 문장을 비교하여 오류를 탐지한다.

2. 사전 구축

문장 분석 및 오류 탐지 과정에 사용되는 지식기반(Knowledge Base)은 사전(Lexicon)이다. 사전은 어휘 정보 및 동의어 정보, 그리고 구문규칙이 저장되어 있다. 사전의 어휘정보는 일반적인 사전 정보, 즉 올림말의 품사, 활용정보, 하위범주화 정보를 포함한다. 본 시스템에서는 효율적인 숙어처리를 하기 위해 자주 틀리는 숙어를 오류가 포함된 형태로 사전에 수록하였다. 이러한 올림말과 숙어 이외에 구문규칙들도 별도의 사전형식으로 저장되어 사용된다. 동의어 사전은 WordNet®으로부터 추출되어 구축되었다. WordNet®에는 올림말의 각 의미(sense)마다 동의어가 세분화 되어 구축되어 있다. 본 시스템에서는 중학교 3학년 학생 수준의 동의어를 추출하여 구축하였다.

표 3. 사전의 어휘 및 동의어 정보

LEX	"put_together"	LEX	"castle"
POS	verb	POS	noun
INF_CODE	IRREGULAR	INF_CODE	REGULAR
SUBCAT	10	SUBCAT	COMMON
SYNONYM	(assemble)	SYNONYM	(palace)

IV. 구문규칙의 구현

본 시스템에 구현된 구문규칙들은 구구조문법(Phrase Structure Grammar)을 기반으로 기술되었다. 규칙은 정문을 처리하는 정문규칙²⁾과 오류가 포함된 비문을 처리하는 비문규칙³⁾으로 나뉜다. 규칙의 기본 형식은 아래 [표 4]와 같다.

표 4. 구문규칙 예제

RULE	NP:np0 → DET:det1 NP:np1
CONDITION	CHECK_DN_AGR(WEAK, [DN_AGR_ERR], det1.agr, np1.agr) CHECK_CV(WEAK, [CV_ERR], det1.lex, np1.lex)
ACTION	ASSIGN(np1, np0)
WEIGHT	0.5
HEAD	np1

- 정문 규칙이란 문장 전체를 처리하는 것뿐만 아니라 구(phrase)와 절(clause)을 처리하는 모든 규칙을 일컫는다.
- 비문 규칙이란 오류를 포함하고 있는 구, 절, 그리고 문장을 처리하는 모든 규칙을 일컫는다.

규칙이 적용되기 위해서는 조건(CONDITION)을 만족시켜야 하며 이를 검사하는 부분은 함수로 구현되었다. 규칙 내에 여러 개의 조건이 기술될 수 있으며 모든 조건이 만족되어 '참'의 값을 내어야 규칙이 성공적으로 적용된다. 결과가 '참'일 경우 동작(ACTION)이 수행되며 이 또한 함수로 구현되었다. 규칙을 구현함에 있어 정문규칙이 우선적으로 기술되었으며 위의 [표 4]의 형식을 갖는다. 비문규칙은 정문규칙의 조건을 완화하거나 오류만을 처리하는 독립적인 규칙으로 나뉘어 기술되었다. 시스템에 구현된 구문규칙은 정문규칙이 235개, 비문규칙 26개, 그리고 오류 탐지 조건을 포함한 정문규칙이 55개로 모두 316개이다.

본 시스템은 이러한 규칙들을 구현함에 있어 Burnstein and et. al (2005)와 Chodorow and et. al (2000)등이 사용한 통계처리 방법과는[2][4] 달리 코퍼스에 나타난 문법현상을 하나하나 기술한 규칙 기반(Rule-based) 시스템이다. 정문규칙은 중학교 영어 교과서와 선행되었던 시험에 대한 선생님의 정답문을 중심으로 기술되었으며 비문규칙은 우선적으로 학생들의 시험결과를 통해 얻어진 오류를 분석하여 기술하였다. 이렇게 문법현상을 일일이 기술하는 방법은 통계처리 방법과 비교해서 시간과 노력이 더 많이 든다고 할 수도 있겠으나, 실제 현상을 확인하고 기술함으로써 규칙이 보다 정확하고 예외적인 현상까지도 처리할 수 있는 장점을 갖고 있다 하겠다.

1. 정문규칙 기술

정문규칙은 중학교 3학년 영어 교과서에 실린 문장을 처리 대상으로 하여 기술되었다. 영어의 문법현상을 올바르게 표현하는 문장들은 정문규칙에 의해 처리된다. 규칙을 구현함에 있어 우선적으로 Greenberg(1961)의 유형학(typology)에 의거하여[6] 영어가 속한 SVO 타입의 언어속성을 기술하였다. 유형학의 분류에 따라 어순은 가지치기(branching), 동사와 조동사 간의 어순, 수식어와 피수식어의 어순, 수식어들 간의 어순, 주절과

종속절의 순서 등을 기술함으로써 처리하였다. 문장의 서법(mood)은 평서문, 의문문, 명령문 및 감탄문이 기술되었다. 문장형식은 기본 5형식의 문장들이 분석되며, 문장형태는 단문은 물론 종속절을 포함한 복문과 등위접속사를 포함하는 중문도 처리하도록 하였다. 기술된 종속절의 종류에는 관계절과 같은 형용사절, 조건문 등을 포함하는 부사절, 동격절 같은 명사절이 있다. 문장의 시제는 단순 및 완료 시제를 분석할 수 있으며 수동태의 문장도 역시 처리하도록 기술되었다.

2. 비문규칙 기술

비문규칙은 문장에 포함되어 있는 오류를 탐지할 뿐만 아니라 기존의 자연어 분석 시스템들이 입력문에 오류가 있는 경우 분석이 실패되는 것과는 달리 오류를 표지하고 계속해서 프로세스를 수행하게 하는 역할을 한다. 그러므로 다양한 오류를 처리하는 규칙들은 유연하고 강력한 시스템 구축에 중요한 역할을 한다 할 수 있겠다.

비문규칙은 두 가지 방법으로 기술되었다. 첫 번째 기술방법은 정문규칙 내에 조건을 추가하되 조건을 완화하여 기술함으로써 구현되었다. 이런 경우, 다른 조건과 동작은 정문규칙과 동일하나 오류가 탐지되면 표지를 하고 프로세스를 계속한다. 예를 들어 'He have a book.'이라는 오류가 있는 문장을 처리할 경우, 명사구 [He]와 동사구 [have a book]을 문장으로 묶는 규칙이 적용될 때 '주어와 동사의 수일치 오류'가 발견된다.

표 5. 주어 동사 수 일치 오류 탐지 규칙

RULE	S:s0 -> NP:np1 VP:vp1
CONDITION	...
	CHECK_SP_AGREEMENT(WEAK,[SUBJ_VERB_AGR_ERROR], np1.agr, vp1.agr)
	...
ACTION	ASSIGN(vp1, s0)
	...
WEIGHT	0.5
HEAD	vp1

오류를 찾아내는 규칙은 [표 5]에서 보듯이, 주어와

4) 본 연구에 사용된 코퍼스는 중학교 3학년 영어 교과서 중 여섯 가지 종류의 교과서 (천재, 중앙, 지학사, 대한교과서, 교학사 (심), 교학사 (배))의 공통 문법 영역에 대한 문장, 교사들의 정답문, 그리고 학생들이 작성한 문장들의 모음이다.

동사의 수 일치 조건을 체크하는 함수 'CHECK_SP_AGREEMENT'의 반환값이 '참'인 경우 정확한 문법으로 간주하여 정문규칙처럼 적용된다. 반면, 반환값이 '참'이 아닌 경우 오류로 간주하여 'SUBJ_VERB_AGR_ERROR'라는 오류⁵⁾를 표시한다.

두 번째 비문규칙 기술방법은 형식에 있어서 첫 번째 방법과 동일하며 입력문에 오류가 있는 경우에만 적용된다. 첫 번째 방법은 정문규칙의 조건을 완화하여 오류를 탐지하고 표시하였으나, 이번에는 조건의 반환값이 정문규칙에서처럼 '참'이어야만 규칙이 적용된다. 이 방법은 'Although he is smart, but he does not study hard.'와 같이 불필요한 접속사의 사용으로 인해 정문규칙으로는 처리할 수 없는 문장을 처리하기 위한 것이다.

표 6. 접속사 오류탐지 규칙

RULE	S:s0 -> SBCL:sbcl1 conj:conj1 S:s1
CONDITION	...
	CHECK_EXTRA_CONJ(ERROR,[EXTRA_CONJ_ERROR], sbcl1, conj1, s1)
	...
ACTION	ASSIGN(vp1, s0)
	...
WEIGHT	0.5
HEAD	vp1

[표 6]의 'CHECK_EXTRA_CONJ'라는 함수가 적용되기 위해서는 반드시 오류가 있어야 하며 오류가 탐지되면 'EXTRA_CONJ_ERROR'와 같은 오류값을 반환한다.

비문규칙은 학생들의 문장⁶⁾을 분석하여 빈도수⁷⁾가 높은 오류부터 기술하였다. 또한 사용된 코퍼스 상에서의 빈도수는 높지 않지만 다른 코퍼스에서는 흔히 발견될 수 있다고 예측되는 오류와 학생 문장에서는 발견되

5) 'SUBJ_VERB_AGR_ERROR'와 같은 모든 오류값은 레이블이며, 실제 시스템 안에서는 ID가 주어졌다.
 6) 규칙을 기술하기 위해 오류유형을 분석한 데이터는 중학교 3학년 을 상대로 치러진 실제 시험문제에 대해 학생들이 작성한 문장들이다.
 7) 본 연구에서 제시하는 빈도수는 중학교 3학년 영작문에서 나타나는 오류를 일반화 할 수 없음을 명시하고자 한다.

지 않았지만 정문규칙에서 파생될 수 있으리라 생각되는 오류⁸⁾들도 기술하였다.

표 7. 오류 종류 및 빈도수

오류 종류	빈도수
문장부호 사용 오류	18.1 %
대/소문자 사용 오류	15.6 %
수일치 오류	13.6 %
철자 오류	5.7 %
시제 일치 오류	5.3 %
관사 오류	3.8 %

[표 7]의 빈도수는 학생들이 작성한 영어문장 4,161개를 분석하여 얻어진 결과이다. 가장 빈번하게 나타나는 오류는 mechanics라고 분류되는 대문자/소문자 구분 오류와 문장부호 사용오류였다. 같은 유형으로 분류되는 철자오류도 빈도의 상위를 차지한다. 대문자/소문자 오류 및 철자오류는 형태소 단계⁹⁾에서 다루지며 문장부호는 구문 단계에서 규칙으로 다루졌다. 시스템의 현재 단계에는 기본적인 구두점오류¹⁰⁾를 탐지하는 규칙이 기술되어 있다.

학생들의 문법오류 유형을 살펴보면 한국어에 없는 문법현상들이 오류빈도가 높은 것을 알 수 있다. 일치 오류 (주어 동사 수 일치, 관사와 명사의 수 일치 및 시제 일치 등의 오류¹¹⁾), 관사 사용 오류 (생략 및 추가)¹²⁾ 등이 상위를 차지한다. 이외에 빈번하게 발견되는 오류로는 동사 형태 오류가 있다. 동사 형태 오류란

8) 예를 들면 'alive'나 'asleep'과 같은 서술형 형용사가 수식어로 사용되는 오류는 실제 코퍼스에서는 발견되지 않았으나 규칙으로 기술되었다.
 9) 형태소 오류는 명사의 복수형 활용 오류(e.g., childs), 명사의 수 정보 오류 (e.g., sugars), 동사의 활용 오류(e.g., watches), 형용사 및 부사의 활용 오류(e.g., sader) 등이 탐지되며 이러한 오류 및 형태소 단계에서 발견되는 오류 처리 규칙은 기존에 알려진 맞춤법 검사 기능들과 많은 차이가 없으므로 이에 대한 설명을 생략한다.
 10) 기본적인 구두점오류 탐지 규칙은 문장의 마침표(. ! ?)와 주절과 종속절의 순서에 따라 사용되는 반점(.)의 사용 오류를 찾아낸다.
 11) 본 논문에서 언급되는 시제 일치란 학생 문장에 정답이 요구하는 시제를 사용하는지를 체크하는 것을 말한다. 학생 문장 분석에는 문제가 없으나 정답이 요구하는 시제와 다른 시제를 사용하였을 경우 오류로 간주된다. 이러한 시제 일치 오류는 규칙의 적용이 아닌 두 문장 간의 비교에 의해 탐지된다.
 12) 관사오류의 대부분이 생략 오류임을 볼 때, 한국어에는 없는 관사 사용이 학생들에게는 어려운 부분임을 보여준다.

동사구를 구성함에 있어 시제, 상(Aspect), 태 (Voice), 혹은 선행 동사의 속성에 따라 적절한 동사 형태를 사용하지 못한 경우를 이른다. 아래 예문은 언급된 오류를 포함하고 있는 실제 학생들의 문장들이다.

표 8. 오류 유형

오류 형태	예문
주어 동사 수 일치 오류	The movie were very funny.
관사 명사 수 일치 오류	These movie was very funny.
동사 형태 오류	I was saw movie.
	I will watching the movie.
	The movie was very interested.

빈도수와 상관없이 정문규칙을 기반으로 기술된 문법오류 유형은 관계대명사 및 부사의 오용, 대명사의 격 사용 오류, 형용사 및 부사의 비교급/최상급 오류 등이 있다. 이러한 규칙들은 독립된 규칙이 아닌 정문처리 규칙에 조건을 추가 하는 형태로 기술되었다. 빈도수가 높은 오류는 하나의 규칙으로 많은 양의 오류를 처리하므로 독립된 규칙으로 기술되어도 효율적이다. 반면에 아주 예외적이거나 빈도수가 극히 낮은 오류는 규칙으로 기술되어야 하는지, 기술된다면 그 기준은 무엇인지에 대한 결정이 필요하다.

V. 오류 처리 및 평가

1. 오류종류

본 시스템에는 28가지의 구문오류가 81개의 규칙으로 구현되었다. 아래 [표 9]는 구문오류의 종류와 예제뿐만 아니라 형태소오류 및 정답문과의 비교 시 발생하는 오류의 종류와 예제를 보여준다.

2. 오류유형 및 구현

시스템이 채점한 학생들의 오류를 분석해 보면 대부분 언어습득 과정에서 발생하는 영어 문법지식의 결여에서 오는 오류(Developmental Errors)로 볼 수 있다. 대표적인 예로 주어 동사 수 일치 오류, 비교급/최상급 오류, 전치사의 오용 등 다양한 오류가 발생하였음을

확인할 수 있었다. 수 일치 오류나 비교급/최상급 오류는 사전에 적재된 정보들을 이용하여 오류를 탐지하도록 하였다. 또한 영어문법 지식이 부족할 때 기존의 문법지식을 과일반화(Overgeneralization)한 오류도 많이 발견되었다. 이러한 유형으로는 관사 첨가 오류, 전치사 첨가 오류 등이 탐지되었다. 전치사 첨가 오류는 규칙 하나로 일반화하기 어려웠고 예외적인 것들이었으나 수가 많지 않아 대부분 규칙으로 기술되었다.

정확한 표현을 모르는 문장을 작성함에 있어서는 한국어의 문법지식을 전이시킨 모국어 전이 오류(Transition Errors)도 눈에 띄었다. 특히 문장 구성 성분이 누락된 생략(Omission) 오류가 발생하는 경우, 이유는 여러 가지가 있겠으나 모국어의 영향¹³⁾을 받아 발생되었다고 보이는 중간언어 오류(Interlingua Errors)를 많이 발견할 수 있다[11]. Ellis(1985)는 모국어에서 비롯되는 오류의 비율을 분석해 본 결과 3%에서 51%의 분포도를 보였으며 대부분의 연구에서도 30%정도의 오류가 같은 원인에서 기인한 것으로 밝히고 있다 [5]. 본 연구에서 가장 많이 발견된 오류는 'be'동사를 생략하는 오류¹⁴⁾였다. 'be'동사 생략 오류는 크게 두 가지¹⁵⁾로 분류된다. 첫째는 형용사가 서술어로 사용되는 경우이다. 한국어에서는 형용사 단독으로 서술어 역할이 가능하나 영어에서는 'be'동사와 같은 연결동사(linking verb)가 필요하다. 이런 경우 아래 예제에서처럼 'be'동사는 생략하고 형용사로 서술어를 대체하였다. 둘째는 명사구가 보어로 사용되는 경우이다. 한국어에는 영어의 'be'동사에 해당되는 서술격조사인 '이다'가 독립어가 아닌 명사의 뒤에 교착되어 사용된다. 예제는 아래 [표 10]에서 찾아 볼 수 있다.

13) 한국어 문장을 영작하는 방식의 시험을 통해 얻어진 답안들이었으므로 다른 경우 보다 이러한 현상이 좀 더 두드러졌을 수도 있다. 또한 자신의 생각을 표현하는 예제이 작문이 아니었으므로 한국어 문장을 번역하려고 하는 데서 발생하는 오류도 찾아 볼 수 있었다.

14) 'be'동사 생략 오류는 3.22%로 빈도수에서 상위권으로 분류된다.

15) 모두 영어 2형식 문장에 해당되는 구조에서 발생된 오류이다.

표 9. 구현된 오류의 종류와 예제

오류 구분	오류 종류	예제	처리 단계		
형태소 오류	명사 복수형 활용 오류	boxs, 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.			
구문 오류	문장부호 오류	He is a student (the period is missing)	구문 분석		
	관사 명사 수 일치 오류	a sheep / a bits			
	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(비교) 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.			

표 10. 'be'동사 생략 오류 예제

정답문	학생 답안
The book was very interesting.	The book very interesting.
The book was very fun.	The book very funny.
His favorite numbers are seven and ten.	His favorite numbers seven and ten.

관사오류¹⁶⁾도 빈도수의 상위권에 포함된 오류이며 관사가 한국어에는 없는 문법적 요소라는 것과 관련지어 볼 수 있겠다. 관사의 오류 탐지는 관사가 수식하는 명사의 속성에 의해 오류가 발견될 수 있는 경우이면 규칙을 적용하였으나 그 외의 경우에는 정답문과의 비교를 통해 오류를 탐지하였다. 관사오류는 생략(Omission), 추가(Addition) 및 대치(Substitution)로 분류되며 이 중 생략 오류¹⁷⁾가 가장 많았다. 생략 오류 중 단수형 명사 앞에서 부정관사의 생략이 가장 많았다. 부정관사의 생략 오류는 수식하는 명사의 가산성(Countable)과 수 정보(Number)에 의해 파악된다. 정관사의 생략 오류는 오류 구성 조건이 규칙으로 기술되기에는 명확하지 않아 현 단계에서는 구현되지 않았다. 추가 오류는 오용, 즉 적절치 않은 곳에 관사를 사용하거나 적소에 사용했다 할지라도 적절치 않은 관사가 사용된 경우이다. 추가오류의 경우도 아래 [표 11]의 예문에서 보듯이 정관사가 수식하는 명사가 '고유 명사'라는 속성을 찾아 오류 탐지 조건을 만족시키는 몇 가지 경우만 기술하였다.

표 11. 관사 오류 유형

정답	학생 답안
The book was very interesting.	Book was very interesting.
I read Harry Potter.	I read the Harry Potter.
This book was fun.	These book was fun.

관사의 대치 오류는 부정관사가 수식하는 명사의 첫 글자가 모음인지 자음인지에 따라 'a'을 'an'을 선택해야 하나 그렇지 못한 경우에 발생된다. 이 오류는 학생들의 답안에서는 발견되지 않았으나¹⁸⁾ 가장 보편적인

16) 앞에서 언급했던 것처럼 3.79%의 빈도를 보였다.
17) 생략 오류는 2.63%로 관사 오류 중 70% 정도를 차지한다.

관사 오류이므로 규칙으로 구현되었다. 또한 관사의 대치 오류로 한정사¹⁹⁾와 명사의 수 일치 오류를 들 수 있다. 위의 [표 11]의 예에서 보듯이 'these'가 'this'를 대치하여 사용되어 오류를 발생시켰다.

이 밖에도 어순오류와 주어 생략의 오류가 모국어의 영향에서 비롯된 오류유형에 속한다고 볼 수 있다. 어순은 한국어의 SOV 특성을 전이시켜 동사를 문장의 마지막에 사용한 경우가 있었다. 주어 생략 또한 문단 안에서 인지된 주어는 생략할 수 있는 (Zero Anaphora) 한국어 특성의 전이로 인해 생긴 오류로 분류할 수 있겠다. 이 두 경우는 규칙 적용이 아닌 정답문과의 비교를 통해 오류를 찾아낸다.

3. 시스템 결과물 및 평가

시스템에 구현된 비문규칙을 구현하기 위해 총 4,161개의 학생들이 작성한 문장을 분석하였다. 영문법의 관점에서 중요하다고 간주되는 문법에서 발생하는 오류 및 학생들의 작문에서 발견된 오류 중 빈도수가 높은 오류를 탐지하기 위한 규칙들을 중심으로 구현되었다.

아래 [그림 2]는 시스템을 이용하여 학생들이 작성한 영어 문장을 채점한 결과의 한 예를 보여준다.

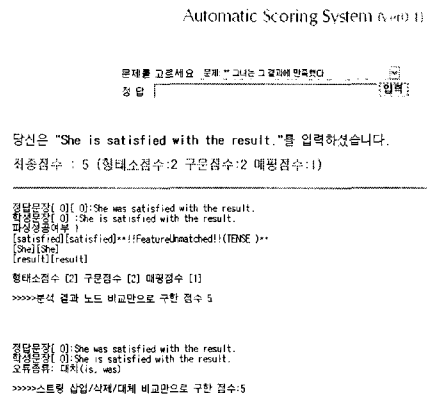


그림 2. 영어 문장 채점 결과

- 18) 본 시스템에 사용된 시험 문제(코퍼스)가 제한적이었으므로 향후 코퍼스에 이런 보편적인 오류가 모두 포함될 수 있도록 계획해야 할 것이다.
- 19) 본 시스템의 사전에는 관사, 지시형용사, 수사, 양화사 등을 한정사(Determiner)라는 그룹으로 묶고 각각의 속성은 사전 정보로 저장하였다.

오류 탐지 및 채점 성능을 평가하기 위해 서울의 두 개 중학교 3학년 학생들이 작성한 11,172개의 문장을 돌려 중학교 교사 2명이 동시에 채점한 결과와 비교하였다. 채점의 만점을 6점으로 하였을 때 교사들 간의 최종 채점 점수 차이는 0.51점이었으며 시스템과 교사들의 차이는 0.67점이었다. 이러한 차이는 지속적으로 학생들의 시스템을 이용한 결과물을 분석하여 수정, 보완한다면 좀 더 좁힐 수 있을 것으로 기대된다.

VI. 결론 및 향후 연구방향

본 연구에서는 중학교 3학년이 작성한 영어문장에서 자동으로 오류를 탐지하고 채점하는 시스템을 개발하였다. 학생들은 인터넷이 가능한 곳이면 언제 어디서든 이 시스템을 이용하여 혼자서 영작문 연습을 할 수 있다. 또한 기존의 문법검사기와는 달리 한국어의 영향을 받아 발생된 오류까지도 검색하고 오류에 대한 피드백을 줄 수 있도록 시스템을 구축하였다. 이처럼 학생들이 장소와 시간에 구애 받지 않고 지속적으로 반복적인 학습을 하며 자신의 문법오류에 대한 피드백을 받으면 영작문 실력 향상에 도움이 되리라 생각한다.

본 연구는 개발 초기 단계에 있으므로 다음과 같은 개선점이 요구된다. 첫째, 좀 더 많은 양의 코퍼스를 구축하는 일이다. 코퍼스는 정문과 비문 코퍼스로 나누어 구축하여 시스템을 보완한다. 정문 코퍼스는 정문규칙을 보완하여 교사의 정답문을 포함한 다양한 종류의 문장을 분석할 수 있는 시스템의 트레이닝 데이터로 사용한다. 비문코퍼스의 보충 및 확장은 시스템을 유연하고 강력하게 만든다. 보다 다양하고 한국어를 모국어로 사용하는 학생들이 만들어 내는 특유한 오류를 찾아내어 규칙으로 기술하여 오류 탐지의 범위를 넓히고 정확성을 높이는 데 사용한다. 둘째, 비문규칙은 때때로 구문의 중의성을 유발하여 파싱트리 개수를 증가시켜 시스템에 부담을 주기도 한다. 이런 규칙에 적절한 제약을 추가하고 트리의 양산을 방지하여 시스템의 효율성을 높여야 할 것이다. 셋째, 동의어사전의 보완이 필요하다. 현재 구축된 동의어사전은 WordNet@에서 필요한

부분을 추출하여 구축되었다. WordNet@의 동의어는 고도로 세분화되어 있어 중학교 3학년들 수준의 동의어로 사용하기에는 부적절하였으므로 동의어를 단순화시키는 작업이 추가되어야 할 것이다. 넷째, 오류가 중복되어 채점되는 문제를 해결하여 채점결과의 정확성을 높여야 할 것이다. 마지막으로 채점기준과 점수매점 방식을 보완하여 수험자간의 변별력을 높여야 할 것이다.

참고 문헌

- [1] M. Bowden and R. A. Fox, "Diagnostic Approach to the Detection of Syntactic Errors in English for Non Native Speakers," Technical Report, 2002.
- [2] J. Burstein and D. Higgins, "Advanced Capabilities for Evaluation Student Writing: Detecting Off-Topic Essays Without Topic-Specific Trainin," Proceedings of the International Conference on Artificial Intelligence in Education, July 2005.
- [3] M. Canale and M. Swain, "Theoretical bases of communicative approaches to second language teaching and testing," *Applied Linguistics*, Vol.1, pp.1-47, 1980.
- [4] M. Chodorow and C. Leacock, "An Unsupervised method for Detecting Grammatical Errors," Proceedings of First Meeting of the North American Chapter of the Association for Computational Linguistics (ANLPNAACL -2000), pp.140-147, 2000.
- [5] R. Ellis, *The Study of Second Language Acquisition*. Oxford University Press, 1994.
- [6] J. H. Greenberg, "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements," *Universals of Languages*. Cambridge, MIT Press, 1961.

[7] K. Gonjenola. and M. Oronoz. "Corpus-based Syntactic Error Detection Using Syntactic Patterns," Proceedings of the Workshop on Student Research, pp.24-29, 2000.

[8] R. Quirk and S. Greenbaum, *A Concise Grammar of Contemporary English*. Harcourt Brace Jovanovich, Publishers, 1973.

[9] R. Scarcella and R. L. Oxford, *The tapestry of language learning*, Heinle and Heinle, 1992.

[10] D. Schneider and K. F. McCoy, "Recognizing Syntactic Errors in the Writing of Second Language Learners," Proceedings of the 17th International Conference on Computational Linguistics (COLING-ACL '98), 1998.

[10] D. Schneider, D. Kathleen, and F. McCoy, Proceedings of the 17th International Conference on Computational Linguistics (COLING-ACL '98), 1998.

[11] L. Selinker, "Interlanguage," International Review of Applied Linguistics, Vol.10, pp.209-230, 1972.

[12] C. J. Wong, "Computer Grammar Checker and Teaching ESL Writing," Paper presented at the 9th Annual Midlands Conference on Language and Literature, Omaha, NE, 1996.

• 2006년 ~ 현재 : 한국외국어대학교 영어학과 강사
 <관심분야> : 자연언어처리, 전산 언어학, 코퍼스 언어학, 형태론

이 공 주(Kong-Joo Lee)

정회원



• 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> : 자연언어처리, 자연어인터페이스, 기계번역, 정보검색

저 자 소 개

김 지 은(Jee-Eun Kim)

정회원



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

• 1995년 ~ 2002년 : 한국마이크로소프트(유) 연구원
 • 2003년 : 한국외국어대학교 영어학부 강사