

# 한국 대학생이 보이는 영어작문 실수 유형\*

임 희석<sup>1)</sup>, 박 종원<sup>2)</sup>, 남 기춘<sup>3)</sup>

1)천안대학교 정보통신학부, 2)천안대학교 어문학부 3)고려대학교 심리학과

E-mail: limhs@cheonan.ac.kr, chongwon@mail.chonan.ac.kr, kichun@korea.ac.kr

## The Type of English Writing Error of Korean Undergraduate Students

Heesuck Lim, Chongwon Park, Kichun Nam

Dept. of Information & Communications, Cheonan Univ., Division of Language and Literature, Cheonan Univ. Dept. of Psychology, Korea. Univ.

E-mail: limhs@cheonan.ac.kr, chongwon@mail.chonan.ac.kr, kichun@korea.ac.kr

본 연구는 한국과학재단의 특정목적기초과제(R01-2000-000-00407-0)의 지원으로 수행되었음.

### Abstract

This study was conducted to extract the feature set of English writing error for suggesting adequate English writing program and making automated scoring system. The frequent committed error and the error across the level of writing proficiency were reported.

Also, It is reported that the correlation between type of error and native speaker's rating score.

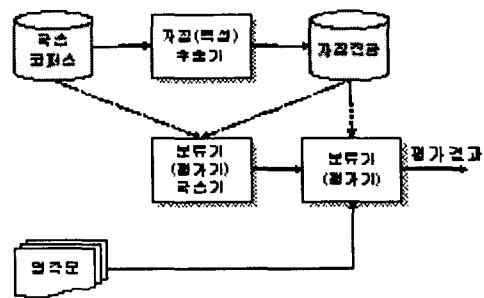
과 투자에 비추어 볼 때 국내에서도 영작문 자동 평가 시스템에 관한 연구가 이루어져야 한다고 사료되며 이러한 필요성에 의하여 본 연구팀의 연구가 시작되었다.

기존의 많은 연구에서와 같이 본 연구팀은 영작문 자동 평가 시스템의 기능을 학생들의 영작문을 입력받아 그 문서를 이미 몇 개의 부류로 나뉘어져 있는 영작문 수준에 분류하는 문서 분류 문제로 정의하고 연구를 진행하고 있다. 예를 들어, 작문 수준이 0~5까지 6개의 수준으로 분류되어 있다면 영작문 자동 평가 시스템은 입력된 영작문이 6개 수준의 어떤 수준에 해당하는 문서인가를 결정하는 작업으로 생각할 수 있다는 것이다.

### 1. 서론

효과적인 언어 학습을 위하여 학습자 능력의 평가와 진단은 매우 중요한 요소이다. 영어 학습에 있어서 필수적인 요소인 작문에 대한 학습도 효과적인 평가와 진단이 매우 중요하다. 하지만 영어 작문에 대한 평가와 진단에 필요한 원어민을 학습자가 주위에서 찾아 활용하기에는 현실적으로 많은 어려움이 따른다. 최근에 인터넷을 이용한 언어 학습 발달로 영작문 수정 및 교정을 위한 사이트들이 있지만 실시간으로 영작문에 대한 평가 결과와 진단을 받을 수 없는 실정이며, 이에 자동 영작문 평가 시스템에 대한 수요가 매우 크다 할 수 있다.

최근 언어처리 기술의 발달로 영작문 자동 평가 시스템 구축에 관한 시도가 국외의 언어 관련 업체와 대학을 중심으로 이루어지고 있으며, 몇몇 데모 프로그램도 시험 중에 있다[4, 5, 6, 8, 9]. 영작문 자동 평가 시스템 개발에 대한 연구는 대부분 해외의 연구이며, 국내 연구진에 의한 연구는 미흡한 실정이다. 우리나라에서의 영어를 비롯한 외국어 학습에 대한 지대한 관심



[그림 1] 문서분류기 기반의 영작문 자동평가기 구조

문서 분류기는 일반적으로 [그림 1]과 같이 크게 분류기 학습기와 분류기로 구성되며 언어 자원으로는 분류기 학습 코퍼스와 자질 집합이 사용되며 문서 분류기를 기반으로 한 영작문 자동 평가기 개발을 위한 일련의 과정은 [표 1]과 같이 설명할 수 있다.

자질 집합은 문서의 논리적인 표현을 위하여 문서를 대표할

[표 1] 영작문 자동 평가기 개발 과정

1. 학습 코퍼스 구축
  - 1.1 영작문 원시 코퍼스 수집
  - 1.2 원어민에 의한 영작문 점수 할당
2. 자질 집합 구축
  - 2.1 학습 코퍼스를 이용한 자질 후보 집합 구축
  - 2.2 원어민에 의한 자질 후보 집합 구축
  - 2.3 자질 후보와 영작문 점수와의 상관 관계를 이용한 자질 집합 선정
3. 영어 자동 평가기 학습
  - 3.1 자질 추출 도구 개발
  - 3.1 문서 분류기(평가기) 학습 방법 개발
  - 3.3 영어 자동 평가기 학습
4. 영작문 자동 평가기 평가 및 수정

수 있는 특징의 집합으로 자질 집합의 선정이 문서 분류기 성능에 매우 중요한 영향을 미친다. 이상적으로 자질은 영작문의 점수를 결정하는데 많은 영향을 미치는 요인들로 구성되어야 하나 실제 평가를 수행하는 원어민들조차 평가에 사용되는 자질을 명확하게 기술하기 어려운 실정이다. 본 연구팀은 원어민으로부터 영작문 평가에 사용될 수 있는 자질의 후보를 수동으로 획득하는 한편 영작문에서 보이는 오류와 실제 영작문간의 상관 분석을 통하여 자질 집합을 자동으로 획득하는 방안을 사용한다. 본 논문은 영작문 자동 평가기 구축을 위한 단계로서 자질 집합 구축을 위해 한국 대학생들의 영작문을 분석한 결과를 오류 유형과 오류 유형과 영작문의 점수와의 상관 관계 분석 결과를 중심으로 기술한다.

## 2. 학습 코퍼스 구축 및 분석 자료 구축

### 2.1. 원시 및 학습 코퍼스 구축

영작문 원시 코퍼스 구축을 위하여 서울 지역 대학교 3곳과 지방 대학 1곳의 500명을 대상으로 TOEFL의 TWE 주제 중 비교적 학생들에게 친근한 주제라고 생각되는 “시골 생활과 도시 생활에 대한 자신의 견해를 적으시오”라는 주제로 영작문을 수집하였다. 서울 지역 대학교와 지방대로 구성된 4곳의 대학교에서 영작문을 수집한 이유는 원시 코퍼스의 영작문 점수에 따른 균형을 유지하기 위한 목적이었다. 성의 있고 개인의 영작문

실력이 반영될 수 있도록 하기 위하여 영작문 시간은 30분으로 제한하였고, 강의실에서 작문이 수행되었다. 500개의 영작문 중 10개의 작문은 주제와 상관없는 잘못된 작문 내용을 가진 결과였으며 최종적으로 490개의 영작문 학습 코퍼스를 구축하였다.

### 2.2. 영작문 코퍼스 오류 태깅 및 분석

490개의 원시 코퍼스를 두명의 원어민이 각각 0~6점 사이의 작문 점수를 부여하게 하였으며 두 평가자의 의견이 불일치하는 경우 서로의 의견을 교환하여 서로 합의한 점수로 영작문 점수를 할당하는 작업을 수행하였다. 오류 태깅은 질적 연구(qualitative research)를 위하여 최근 많이 사용되는 QSR NVIVO 2.0을 이용하여 상향식 방법에 따라 임의의 164개의 영작문을 대상으로 오류를 태깅하였다[10]. 상향식 방법에 따른 오류 태깅이란 태깅할 오류 유형을 미리 설정하지 않고 태깅을 수행하는 평가자가 글을 읽으며 영작문에 포함되는 오류 유형을 만들어 나가는 방식을 의미한다. 이는 평가자가 미리 정해진 부류에 해당하는 오류만을 분석 대상으로 하지 않고 작문자 입장에서의 오류를 좀더 찾아내기 위함이었다. [표 2]는 164개의 영작문에 대한 통계를 자동 분석한 결과를 보이고 있다. [표 2]에서 1번 열은 영작문 점수 부류, 2~5번 열은 각 부류에 속하는 영작문의 평균 단어수, 평균 토큰수, 평균 내용어의 수, 내용어의 비율을 나타내며 6번째 열은 각 부류에 속하는 영작문의 개수를 나타낸다.

[표 2] 영작문 코퍼스의 어휘에 관련된 통계

Level	Avg Num. of Words	Avg Tok. of Words	Avg. Num. of Content Word	Cont. Word Ratio	The Num. of Subj.
0	138.625000	158.875000	87.687500	0.632552	16
1	169.050003	187.300003	100.300003	0.593316	20
2	239.441177	264.441162	137.941177	0.576096	34
3	278.274506	303.098053	154.666672	0.555806	51
4	328.962952	346.962952	175.555557	0.533664	27
5	333.933319	350.000000	177.866669	0.532641	15
6	357.000000	368.000000	198.000000	0.554622	1

### 3. 영작문 오류 분석 및 상관 분석

QSR NVIVO 2.0을 이용한 오류 태깅에서 태깅된 오류의 유형은 총 60가지였으며, 이들을 어휘 오류, 구문 오류, 의미 및 담화 오류로 병합할 수 있었다. 60가지의 오류 중 오류 유형은 영작문에 태깅된 오류의 개수를 이용하여 오류 유형과 영작문 점수간의 상관 계수를 구하고자 하였다. 그러나 [표 2]에서와 같이 각 그룹에 속하는 피험자의 수와 평균 작문의 길이가 달라서 오류의 절대 개수로 상관을 계산할 수 없으며, 작문의 길이와 피험자의 수에 의해서 태깅된 오류의 수를 정규화하였다. 아래의 [표 3], [표 4], 그리고 [표 5]는 각 수준별 오류 유형 중 상위 10개씩을 나타낸 것이다.

[표 3] ~ [표 5]에서 많은 오류들이 공통적으로 나타남을 볼 수 있다. 흥미로운 사실은 3~5수준과 같이 비교적 작문 실력이 있는 그룹에서의 제일 순위의 오류가 의미에 관련된 오류이고 비교적 용이한 시제 및 수 일치에 관련된 오류는 2수위에 있고 저수준 그룹은 그 반대라는 것이다. [표 6]은 [표 2]의 통계값을 포함한 60가지

의 발견된 오류 유형과 원어민이 평가한 영작문 점수와의 상관값을 계산하여 상위 10개를 나열한 것이다.

### 4. 결론 및 향후 연구

본 연구에서는 영작문 자동 평가 시스템 개발을 위한 사전 연구로서 영작문의 논리적인 표현을 위하여 사용될 자질(특징) 집합 선정을 위한 사전 연구로 한국 대학생 영작문 실수 코퍼스의 구축과 영작문 오류 유형 및

오류 유형과 영작문간의 상관 관계에 대한 분석 결과를 제시하였다. 본 연구 결과는 164개의 비교적 소량의 영작문 자료를 이용하였으며 분석 대상의 분포가 영작문 점수대별로 균형 있지 못한 편향된 자료를 사용하였기 때문에 의미있는 결과로 해석하기에는 다소 무리가 따를 것으로 생각된다. 향후 영작문 자동 평가기를 위한 자질 집합의 구축을 위하여 좀더 많고 균형 있는 영작문에서의 오류 분석과 오류 유형과 영작문 점수와의 연관이 높은 자질을 추출할 수 있는 방법에 대한 연구를 수행하

[표 3] 오류 유형과 점수와의 상관

/Verbals/to Root /Content word Ratio /Sentence types/complex sentences /article /usage/There be structure Avg. Num. Of Content Word Avg. Num. Of Token /run on sentences /article /usage/There be structure
--

[표 4] 4수준과 5수준의 오류 유형 상위 10개

4수준의 오류의 상위 10개	5수준의 오류의 상위 10개
1. /meaning/structure	1. /meaning/structure
2. /agreement/agreement-noun	2. /agreement/agreement-noun
3. /meaning/word selection	3. /Main verbs issues/main verbs/transitive
4. /parallel structure	4. /connectives
5. /Sentence types/complex sentences	5. /meaning/word selection
6. /Main verbs issues/main verbs/transitive	6. /parallel structure
7. /connectives	7. /agreement/agreement-Sand V
8. /preposition	8. /Sentence types/complex sentences
9. /article	9. /adjective
10./pronoun	10./run on sentences

[표 5] 0수준과 1수준의 오류 유형 상위 10개

0수준의 오류의 상위 10개	1수준의 오류의 상위 10개
1. /article	1. /agreement/agreement-noun
2. /mechanism/capitalization	2. /parallel structure
3. /agreement/agreement-Sand V	3. /article
4. /agreement/agreement-noun	4. /meaning/word selection
5. /usage/There be structure	5. /Main verbs issues/main verbs/transitive
6. /Sentence types/complex sentences	6. /Main verbs issues/main verbs/intransitive
7. /parallel structure	7. /Sentence types/complex sentences
8. /preposition	8. /mechanism/capitalization
9. /meaning/word selection	9. /adjective
10. /connectives	10. /relative pronouns

[표 6] 2수준과 3수준의 오류 유형 상위 10개

2수준의 오류의 상위 10개	3수준의 오류의 상위 10개
1. /agreement/agreement-noun	1. /agreement/agreement-noun
2. /meaning/word selection	2. /meaning/structure
3. /mechanism/capitalization	3. /meaning/word selection
4. /Main verbs issues/main verbs/transitive	4. /mechanism/capitalization
5. /connectives	5. /article
6. /agreement/agreement-Sand V	6. /Main verbs issues/main verbs/transitive
7. /parallel structure	7. /parallel structure
8. /adjective	8. /preposition
9. /Sentence types/complex sentences	9. /agreement/agreement-Sand V
10. /preposition	10. /Sentence types/complex sentences

여야 할 것이며, 본 연구는 이러한 연구 수행에 있어서의 연구 방향과 문제점을 사전에 점검하는 기회로서의 의미를 찾을 수 있을 것이다.

### 참고문헌

- Scoring sYstem, Univ. of Maryland, <http://ericae.net/betsy/>
- [7] M. R. Lawrence, T. Liang, "Automated Essay Scoring Using Bayes' Theorem", *The Journal of Technology, Learning, and Assessment*, Vol. 1. No. 2, June 2002.
- [8] PEG(Project Essay Grade), Duke Univ., <http://134.68.49.185/pegdemo/ref.asp>
- [9] Vantage Learning, Inc., IntelliMetric, <http://www.intellimetric.com/>
- [10] <http://www.qsrinternational.com/products/nvivo.html>
- [1] J. C. Burstein, & M. Chodorow, "Automated Essay Scoring for Nonnative English Speakers," Joint Symposium of the Association of Computational Linguistics and the International Association of Language Learning Technologies, Workshop on Computer-Mediated Language Assessment and Evaluation of Natural Language Processing, College Park, Maryland, 1999.
- [2] J. C. Burstein, S. Wolff, and C. Lu., *Using Lexical Semantic Techniques to Classify Free-Responses. The Depth and Breadth of Semantic Lexicons*, Edited by Nancy Ide and Jean Veronis. Kluwer Academic Press, 1999.
- [3] J. C. Burstein, K. Kukich, S. Wolff, C. Lu, M. Chodorow, L. Braden-Harder, & M. Dee Harris, "Automated Scoring Using A Hybrid Feature Identification Technique", *Proceedings of the Annual Meeting of the Association of Computational Linguistics*, August 1998. Montreal, Canada.
- [4] J. Burstein, "The E-rater Scoring Engine: Automated Essay Scoring With Natural Language Processing", In *Automated essay scoring: A cross-disciplinary perspective*. Edited by Mark D. Shermis and Jill Burstein, Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 2003.
- [5] Knowledge Analysis Technologies, Intelligent Essay Assessor, <http://www.knowledge-technologies.com/>
- [6] M. R. Lawrence, BESTY(Bayesian Essay Test