View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by E-Journal Universitas Muhammadiyah Semarang

2nd English Language and Literature International Conference (ELLiC) Proceedings – (ELLiC Proceedings Vol. 2, 2018)
Electronic ISSN: 2579-7263
CD-ROM ISSN: 2579-7549

THE DETECTED AND UNDETECTED ERRORS IN AUTOMATED WRITING EVALUATION PROGRAM'S RESULT

Muhamad Nova[1*)], Iwa Lukmana[2)]
*English Language Education Study Program, School of Postgraduate Studies, Universitas Pendidikan Indonesia (UPI)
Indonesia
*mohnova@student.upi.edu

Abstract

The investigation of error is quite challenging to be conducted at school and university. Interestingly, with the technology development, detecting error can be conducted by using an automated writing evaluation program. This study aimed to analyze the errors in writing by applying an automated writing evaluation program. This study applied a mixed methods research with exploratory design. The total of 48 undergraduate students participated in this study and each student submitted one essay which was then re-submitted to the automated writing evaluation program. As the result, there are 483 errors detected by the program which came from 21 types of errors. However, the program still left some errors undetected with the total number of 157 errors which came from 24 types of errors; 12 types of errors have been identified and 12 types of errors have not been identified by the program. From the result of the program, the use of automated writing evaluation program in detecting error seems giving some benefits for the user. However, the application of this program still needs the teacher and lecturer's supervision to reduce the weaknesses of the program in detecting the errors.

Keywords: automated writing evaluation program; detecting error; error; error analysis

Introduction

One of the problems in learning a foreign language is making an error. Previous studies in many EFL learning context have investigated the errors problems made by the students in writings. Various results of their research findings have been revealed in many aspects, including the types of errors (Pouladian, Bagheri, & Sadighi, 2017), the cause of errors (Bosuwon, 2013), and the frequency of error production (Pratiwi, 2015). However, in the reality, the identification of error is still viewed as a challenging demand to be conducted at school and university. The issue of big classroom size makes the error identification seem taking much time and effort (F. Wang & Wang, 2012; Wilson & Czik, 2016). As a solution in addressing this problem, the utilization of technology, such as an automated writing evaluation program, can help the teacher and the lecturer in dealing with error identification.

An automated writing evaluation program, acronymic as AWE program, is a computer software which is utilized to evaluate writing. This writing evaluation program has analytical features which can be used to analyze writing. Several types of automated writing evaluation programs have been improved with artificial intelligence technology which can detect and analyze the sentence on grammar, syntactic, lexical, and discourse levels (Chou, Moslehpour, & Yang, 2016). Then, some versions of these programs have also been featured with diagnostic analysis and feedback to enrich the quality of evaluation given by the program (Chen & Cheng, 2008). Moreover, most of the automated writing programs nowadays are not only giving the result of detection, but also providing the correction and suggestion to improve the quality of writing (P. Wang, 2013; Wilson & Andrada, 2016).

Reflecting from the features built in the program, the automated writing evaluation program seems promising to help the teacher and the lecturer identifying the errors on their students' writings. With the

Electronic ISSN: 2579-7263
CD-ROM ISSN: 2579-7549

diagnostic analysis and feedback had by the program, this program may have a potential to be utilized as error detection. Previous studies have also conducted similar research on automated writing evaluation program utilization (Cotos, 2011; Ebyary & Windeatt, 2010; Scharber, Dexter, & Riedel, 2008; F. Wang & Wang, 2012), but their research only focused on evaluating the writings in general and did not focus on utilizing the program to detect and analyze the errors.

Thus, this study aimed to investigate the new potential of automated writing evaluation program for error detection and to evaluate the result of its error identification result. As the result, the outcomes of the study can be a consideration for the teacher and lecture in applying the automated writing evaluation program for identifying and analyzing the errors on their students' writings.

Methodology

This study applied a mixed methods research with an exploratory design. This research design allowed the researchers to investigate the study deeper by applying two approaches which are qualitative and quantitative approaches. The qualitative approach was applied to identify and to classify the errors detected by the program and the quantitative approach was applied to calculate the frequency of errors production in students' writings.

Related to the process of data collection, the data was gained from students' writings. There were 48 undergraduate students who participated in this study and each of them submitted an essay. These essays were then re-submitted to an automated writing evaluation program, namely *Grammarly* free-version program, to identify and classify the errors made by the students in their writings. The errors detected by the AWE system were underlined with red color and they were recorded in a log. Then, the result of program evaluation was re-analyzed manually to identify the undetected errors in students' writings. Thus, the researchers analyzed the writings and record any errors found in another log. As the result, two logs were produced in this study in which one log of the detected errors and one log of the undetected errors. These logs became the primary data for the data analysis.

Related to the process of data analysis, the data was analyzed through two phases, which were qualitative phase and quantitative phase. In the qualitative phase, the errors recorded in each log were then sorted and classified based on its types. The process of classifying the errors was based on AWE program classification. After being classified, the process of data analyses moved to the second phase which was the quantitative phase. In this phase, the frequency of each type of error was calculated and the percentage of each frequency was identified. As the result, a complete description of types of errors and each frequency was created. Then, a further interpretation was also created in line with the findings of the study.

Findings and Discussion

This findings and discussion section is divided into three sub-sections. The first section discusses the result of error identification done by the AWE program. Then, the second section discusses the findings related to the errors which are found by the researchers but were not detected by the AWE program. Last, the third section discusses the results of error detection with the previous studies.

1. The Detected Errors

The utilization of AWE program as an error analyzer revealed some types of errors in students' writings. With the diagnostic features built in the program's system, this program detected the total number of 483 errors. Each of the detected errors was analyzed and the program also gave the explanation of error occurring in the sentence. This explanation became the consideration to classify the types of errors

Electronic ISSN: 2579-7263
CD-ROM ISSN: 2579-7549

detected by the program. Then, from the classification result, it was revealed that these 483 errors came from 21 types of errors (Table 1). From the result of evaluation, it can be inferred that the system used by the program has been quite successful in detecting the major errors made by students occurring in seven major classifications, i.e. grammatical rule agreements, incorrect forms of the word used, missing item needed in a sentence, additional item which is unnecessary, redundancy item, miswritten words, and rule of word capitalization.

Table 1. The Detected Errors

Types of Errors	Frequency
Missing a determiner	135 (27.95%)
Miswritten	61 (12.63%)
Subject-verb agreement	60 (12.42%)
Incorrect preposition	48 (9.94%)
Missing a comma	42 (8.70%)
Incorrect word class form	40 (8.28%)
Unnecessary comma	26 (5.38%)
Singular-plural agreement	22 (4.55%)
Missing a hyphen	13 (2.69%)
Quantifier-object agreement	13 (2.69%)
Unnecessary preposition	7 (1.45%)
Redundancy	3 (0.62%)
Modal-verb agreement	2 (0.41%)
Uncountable noun	2 (0.41%)
Incorrect word choice	2 (0.41%)
Capitalization	2 (0.41%)
Missing apostrophe	1 (0.21%)
Missing a preposition	1 (0.21%)
Unnecessary punctuation	1 (0.21%)
Missing an auxiliary verb	1 (0.21%)
Incorrect article form	1 (0.21%)
Total	483 00%)

a. The Grammatical Rule Agreements

The result of AWE program evaluation detected the error on grammar rule agreements. The artificial intelligence system can detect the incorrect pair between words, such as subject-verb, quantifier-object, singular-plural, and modal-verb. This incorrectness was then evaluated and corrected by the system by changing the form of the word, such as shown in the following example, an example of quantifier-object agreement error detected in Text 10:

> [1] Each plates has boundaries. (Text 10)
> [*] *Each plate has boundaries.*

The example above shows the recognition of the system in detecting incorrect pair between the quantifier *each* with the noun *plates*. Based on the explanation given, the quantifier *each* is used as singular quantifiers which should be paired with singular countable nouns. Therefore, the system suggests changing the plural noun *plates* into its singular noun *plate*.

b. The Incorrect Word Forms

The AWE program has detected some of the errors which are caused by the incorrect forms of the word used. This type of error includes the incorrectness of preposition, word class form, word choice, and article form. Similar to the previous error, the grammar rule agreement errors, the incorrect forms errors detected by the system were corrected by giving the correct form of the words, whether changing the preposition, word class, or even the word. An example below, which is taken from Text 30, shows the correction given by the system in dealing with incorrect word class form error.

> [2] However, people have to response to this issue wisely… (Text 30)
> [*] *However, people have to respond to this issue wisely…*

The word *response*, which belongs to a noun, has been detected as an error in this sentence. The program corrected the word by changing it into its verb form, i.e. *respond*. The program also gave further explanation on the reason behind of the error which majorly comes from the confusion of the word class had by some words, including *response* and *respond*. Both of these words have similar meaning, but they belong to different word class family which can affect its usage in the sentence. Therefore, the system suggests changing the word into its appropriate form.

Electronic ISSN: 2579-7263
CD-ROM ISSN: 2579-7549

2. The Undetected Errors

After the result of the error detection was revealed, there were found some errors which were not detected by the program. There were 157 errors which were not detected by the program. Then, the researchers classified the errors based on its occurrences and the classification result revealed that these 157 errors came from 24 types of errors (Table 2).

Table 2. The Undetected Errors

Types of Errors	Frequency
Incorrect word class form	35 (22.29%)
Run-on sentence*	30 (19.11%)
Missing an auxiliary verb	16 (10.19%)
Misplacement of word *	14 (8.92%)
Subject-verb agreement	9 (5.73%)
Incorrect tense *	7 (4.46%)
Missing a verb (to-be) *	7 (4.46%)
Unnecessary preposition	6 (3.82%)
Incorrect word choice	5 (3.18%)
Missing a relative pronoun *	5 (3.18%)
Missing a conjunction	4 (2.52%)
Missing a subject *	4 (2.52%)
Singular-plural agreement	2 (1.27%)
Modal-verb agreement	2 (1.27%)
Redundancy *	2 (1.27%)
Multiple verb *	2 (1.27%)
Miswritten	1 (0.64%)
Unnecessary comma	1 (0.64%)
Quantifier-object agreement	1 (0.64%)
Missing a preposition	1 (0.64%)
Unnecessary auxiliary *	1 (0.64%)
Unnecessary article *	1 (0.64%)
Unnecessary relative pronoun *	1 (0.64%)
Multiple determiner *	1 (0.64%)
Total	157 (100%)

Interestingly, from the 24 types of errors, 12 of them have actually been identified by the AWE program and the other 12 types of errors, which are marked with (*) in Table 2, have not been identified by the program. This finding also revealed the weaknesses of the program in evaluating the writings submitted. It was identified that the sensitivity of the AWE program's system in recognizing the structure of the sentence was limited to some cases, including long phrase identification, passive voice recognition, and question structure.

a. Long Phrase Identification

The first weakness identified in applying this AWE program for detecting errors is identifying long phrases failure. The system built in this AWE program failed in recognizing the main focus conveyed in a long phrase. The system calculates the proximity between the words in which the nearest word before the verb is identified as the subject of the sentence. However, the focus discussed in the phrase is located at the beginning of the phrase. As an example, one case was found in Text 46 which is shown below:

> [3] The chemicals in Botox is made to relax the tense muscles. (Text 46)

From the example shown above, it can be seen that the focus discussed is the word *chemicals*. However, the system of the AWE program detects that the focus is the word *Botox*, which is located before the verb (to be) *is*. As the result, the verb (to be) *is* is identified as an appropriate verb form for the singular subject, *Botox*. In contra, the verb (to-be) should be *are* since the focus of the phrase is the word *chemicals*, not the word *Botox*. Thus, the identification of the phrase's focus still needs to be maintained to avoid a misleading context.

b. Passive Voice Recognition

Passive voice recognition becomes the second weakness had by the program. As the AWE program uses algorithm calculation, the artificial intelligence system can detect a sentence by identifying a noun, which becomes the subject of the sentence, and a verb, which becomes the predicate of the sentence. However, this system failed in recognizing the context of the verb used in the sentence, whether it is in an active form or in a passive one. The AWE system cannot differentiate the use of past participle as a verb indicating the past event or as a verb indicating a passive voice form. To make it

Electronic ISSN: 2579-7263
CD-ROM ISSN: 2579-7549

clearer, the example below shows a case of a sentence which is left uncorrected by the system.

> [4] These photoreceptors mostly packed together in the center part of retina. (Text 33)

The sentence above was identified having a correct structure of grammar by the AWE program. The verb *packed* may be inferred as the indicator to tell the reader that the event happened at the past time and thus, the system identifies it as a past participle verb for past tense. However, reuniting to the context of the text whole fully, this sentence seems to miss the object which is being packed by the *photoreceptors* and the meaning of the sentence is more relevant when the verb *packed* is in a passive voice form. Thus, the failure in recognizing the context of the text causes the misleading process in passive voice identification and relevancy of the sentence.

c. Question Structure

This AWE program also has a weakness in identifying the structure of a WH-questions, i.e. *what, who, where, when, why, and how*. In English grammar, the structure of this question should consist of WH question word followed by auxiliary, subject, and verb as a predicate. Indifferent from what has been found in this study, this AWE program failed in recognizing the structure of this question and did not detect the sentence as a question. This recognition failure was found in the following example.

> [5] Why we need them? (Text 7)

The AWE system detects the example of the question above having a correct structure. As what has been mentioned above, the structure of WH-questions should consist of WH question word followed with auxiliary, subject, and verb as a predicate and this sentence has consisted of WH-question word, i.e. *why*, a subject, i.e. *we*, and a predicate, i.e. *need*. However, the system failed to recognize the missing auxiliary for this question. The algorithm of the program recognizes this question has a similar structure with an affirmative reported question in which it is started with the question word *why* which is identified as a relative pronoun, and then, the word *we* is identified as the subject and the verb *need* is identified as the predicate. Then, another possible rationalization for this miscalculation is that the result of this AWE program analysis seems not identifying the question mark put at the end of the sentence. Thus, from the miscalculation result, it may lead the user to keep their structure incorrectly by having a question with an affirmative structure.

3. Discussion

The main focus of this study is to examine the use of an automated writing evaluation program for detecting errors. It carried out the result of the program evaluation on students' writings. The program has detected 483 errors which came from 21 types of errors. This result becomes an evidence for the successful error identification by the program. In this study, the AWE program has detected various types of errors in students' writings, which cover grammar, spelling, and punctuation errors. Not only detecting the incorrect form, but this program also has detected the addition and omission of some syntactical items in the sentence. This result matches the features of AWE program which been identified by previous scholars. Related to the utilization of AWE program, this kind of program can diagnose the problem on grammar, syntactic, lexical, and discourse levels in writing i.e. errors (Chou et al., 2016; F. Wang & Wang, 2012) and it can be a positive input for language learner in understanding sentence structure. Moreover, the feedback and explanation provided by this AWE program after evaluating the writing seem beneficial for both teacher and students. This confirms that this AWE program can also enhance the quality of writing (Chen & Cheng, 2008; P.

Electronic ISSN: 2579-7263
CD-ROM ISSN: 2579-7549

Wang, 2013; Wilson & Andrada, 2016) since the correction and suggestion given may increase the user knowledge in writing, especially in English grammar. As the result, the use of an AWE program seems helpful for both teachers, who can use it for detecting the students' writing errors, and students, who can apply it for improving the writing quality.

Interestingly, it is also critical to note some weaknesses of this AWE program in detecting errors found in this study. It was found that there are 157 errors which were not being detected. The artificial intelligence system still has some misleading recognitions in detecting the errors, especially in dealing with long phrase, passive voice structure, and question structure. The majority of these misleading recognitions came from the less sensitive algorithm calculation in identifying the context of the sentence and the focus of the sentence. As the impact, these cases can influence the students' learning outcomes in which the students still keep their misconception in writing or even the result leads them to their confusion in revising the writing (Scharber et al., 2008). Even though this AWE program has detected various types of errors, the application of this AWE program in detecting errors still needs professional supervision to overcome the misleading result of the evaluation.

Conclusions

This study has investigated the application of an AWE program for detecting writing errors. The result of the program evaluation revealed the total number of 483 errors which came from 21 types of errors. These errors cover three major writing problems, i.e. grammar, spelling, and punctuation. However, the system of this program also still has some weaknesses in recognizing long phrase, passive voice structure, and question structure which result in 157 errors left undetected. From the result of the program evaluation, it indicates that the use of AWE program in detecting error seems giving some benefits for the user. However, the application of this program still needs supervision from language expert to reduce the weaknesses of the program in detecting the errors. Therefore, the assistance from the teacher and lecturer is needed to overcome the misleading result given by the program; The teacher and the lecturer can add further explanation toward the error detected by the program and together with the students, revise their writings. The researchers believe that the result of this study can be a critical consideration for teacher and lecturer who are encouraged to apply an automated writing evaluation program for detecting writing errors.

Finally, a number of potential limitations of this study also need to be considered. First, the process of error detection using the AWE program was conducted only by the researchers without involving the students. The students only submitted their writings to the researchers and the researchers re-submitted these writings to the AWE program. The researchers assume that the participation of the students in using the AWE program may reveal new finding in which the students can give their opinions and views about the results of the evaluation. Therefore, further research is required to investigate the students' attitude toward the use of AWE program in detecting their writing errors. Second, this study only applied the free-version of AWE program, *Grammarly* program. As a free-version, the AWE program's utilized in this study has some limited features and it can influence the result of error detection and identification. Thus, further data collection would be needed to determine exactly the result of full-version of AWE program utilization in detecting errors.

Acknowledgement

This study was fully sponsored by Indonesia Endowment Fund for Education Scholarship (LPDP), Ministry of Finance, Republic of Indonesia (Grant No. 20160611087314).

Electronic ISSN: 2579-7263
CD-ROM ISSN: 2579-7549

Note

Due to the page constraint, the researchers only discussed two types of errors detected by the AWE program in this paper.

References

Bosuwon, T. (2013). Common syntactic errors in interlanguage writing output: A case study of EFL undergraduates. *University of the Thai Chamber of Commerce Journal*, *33*, 229–246.

Chen, C. E., & Cheng, W. E. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, *12*(2), 94–112.

Chou, H. C., Moslehpour, M., & Yang, C. (2016). My access and writing error corrections of EFL college pre-intermediate students. *International Journal of Education*, *8*(1), 144–161.

Cotos, E. (2011). Potential of automated writing evaluation feedback. *CALICO Journal*, *28*(2), 420–459.

Ebyary, K. E., & Windeatt, S. (2010). The impact of computer-based feedback on students' written work. *International Journal of English Studies*, *10*(2), 121–142.

Ferster, B., Hammond, T. C., Alexander, R. C., & Lyman, H. (2012). Automated formative assessment as a tool to scaffold student documentary writing. *Journal of Interactive Learning Research*, *23*(1), 21–39.

Pouladian, N., Bagheri, M. S., & Sadighi, F. (2017). An analysis of errors in writing skill of adult Iranian EFL learners preparing for the IELTS. *International Journal of English Linguistics*, *7*(3), 85–96.

Pratiwi, A. P. (2015). Interlingual and intralingual errors in writing narrative text made by junior high school and senior high school students. *Jurnal Penelitian Humaniora*, *16*(2), 1–11.

Scharber, C., Dexter, S., & Riedel, E. (2008). Students' experiences with an automated essay scorer. *The Journal of Technology, Learning, and Assessment*, *7*(1), 1–45.

Wang, F., & Wang, S. (2012). A comparative study on the influence of automated evaluation system and teacher grading on students' English writing. *Procedia Engineering*, *29*(2012), 993–997.

Wang, P. (2013). Can automated writing evaluation programs help students improve their English writing? *International Journal of Applied Linguistics & English Literature*, *2*(1), 6–12.

Wilson, J., & Andrada, G. N. (2016). Using automated feedback to improve writing quality: opportunities and challenges. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of Research on Technology Tools for Real-World Skill Development* (p. 678). USA: IGI Global.

Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English language art classrooms: Effects on teacher feedback, students motivation, and writing quality. *Computers and Education*, *100*(206), 94–109.