

The Effect of Automated Corrective Feedback On L2 Writing in POS Categories

Yingxue Wang

Jiamusi University
Busan University of Foreign Studies
wonderfulcristina@163.com

ABSTRACT

Automated corrective feedback is the processing of the Computer-Assisted Language Learning used in L2 English writing assessment that is ubiquitous in current L2 practice and research (e.g., Chen, 2016; Chukharev & Saricaoglu, 2016; Gao et al., 2020). This research examined the effect of automated corrective feedback of Pigainet as a kind of Computer-Assisted language learning instruments in English writing revision. Data were collected from 591 drafts of 31 participants who submitted their drafts on Pigainet and coded as errors frequency ratios according to POS category. Findings suggested that Pigainet could help participants revise their writing errors of Article, Verb, Preposition, and Noun.

Keywords: Automated Corrective Feedback, Automated Writing Evaluation, L2 Writing, POS

1. INTRODUCTION

Automated corrective feedback systems may facilitate the iterative writing process by alleviating teachers of the time-intensive practice of providing formative individual feedback to students (Burstein et al., 2003) [1]. Students can receive specialized diagnostic feedback from an automated system that covers critical aspects of writing such as sentence form, word usage, and organizational structure (Burstein, *ibid*). Students can utilize this information to rewrite their essays independently, allowing them to participate in the write, feedback, and revision cycle independently.

Recently, automated CF as a popular L2 learning software has been utilized in L2 writing evaluation, such as ETS Criterion, MY Access!, Write to Learn, and Pigainet, which is used in all the universities in China. With the combination of automated scoring and feedback, these instructional systems are now referred to as automated writing evaluation (AWE) systems (McNamara et al., 2015)[7]. Particularly in the last decade, these AWE systems have been gradually marketed in classroom settings with English language learners (ELLs), although they were initially designed for native speaker-writers of English in the US (Warschauer & Ware, 2006).

Modern automated writing evaluation (AWE) systems can now provide comments on language and

content in addition to an automatic score, thanks to advances in language processing technologies and statistical approaches. Therefore, it is significant to do some research on the effect of automated corrective feedback provided by AWE.

2. Literature Review

Automatic Writing Evaluation (AWE) tools are based on artificial intelligence, natural language processing, and statistical techniques, enabling them to complete the evaluation of written text in a shorter time than manual evaluation (Grimes & Warschauer, 2010)[4]. As AWE tools reduce workforce essentially, they play an essential role in formative assessment "Assessment of Learning" (Bennett, 2011, p. 8), which helps learners improve writing skills by providing repeated formative feedback.

The AWE tool Project Essay Grade (PEG) was developed in the 1960s to give a solution for teachers. Typical current AWE software, such as Criterion, MyAccess, and Writing Roadmap 2.0, provide a numerical score to reflect the overall quality of writing and diagnostic feedback to help the student improve writing in several areas, including grammar, word choice, sentence structure, organization, and idea development, based on a sample of student writing (Gao et al., 2020; Li et al., 2017; McNamara et al., 2015; Rich, 2012)[3][6][7][8].

While this study focuses on POS categories to investigate the effect of automated corrective feedback after answering the following research questions: what is the EFR of each type in POS categories? And what is the EFR change of each type in POS categories?

3. RESEARCH METHOD

Fifty freshmen in Jiamusi University as participants took part in this study voluntarily. The pre-test and post-test were designed to compare the EFR change and investigate the improvement of accuracy in terms of POS types. Both pre-test and post-test were taken in the classroom for 30 minutes. Furthermore, ten tasks were assigned to finish in 10 weeks, namely once a week. And the data could be collected to explore the characteristics of POS types in L2 learners' writing.

From the participants, thirty-one students' 591 drafts were selected from Pigai.org. In order to make the error coding category reliable, the triangle research methods are used in the research. This study combines two automated writing evaluation tools, Pigainet and Grammarly, and human rater, to code and analyze

sampling drafts errors so that the data coding could be more comprehensive and multidimensional.

4. DATA ANALYSIS AND RESULTS

We calculated the total number of errors identified by triangle method on each draft, standardized the raw number of errors, and then obtained descriptive statistics about the EFR of each type totally from the first draft to the final draft for all papers including pre-and post-tests, and the mean number of EFR changing from pretest to posttest.

RQ1: What is EFR of each type in POS categories?

As the Figure1 shows that in the whole pie, EFR of the article(mean=.0342,SD=.011) is 30%, EFR of the verb (mean=.015,SD=.0067) is 22%, EFR of preposition (mean=.0127,SD=.012) is 13%, EFR of the noun (mean=.0096,SD=.0032) is 10%, EFR of pronoun (mean=.0062,SD=.0055) is 9%, EFR of conjunction (mean=.009,SD=.019) is 8%, EFR of the adjective (mean=.002,SD=.0021) is 4%, EFR of punctuation (mean=.0014,SD=.0008) is 2% as well as adverb's portion.

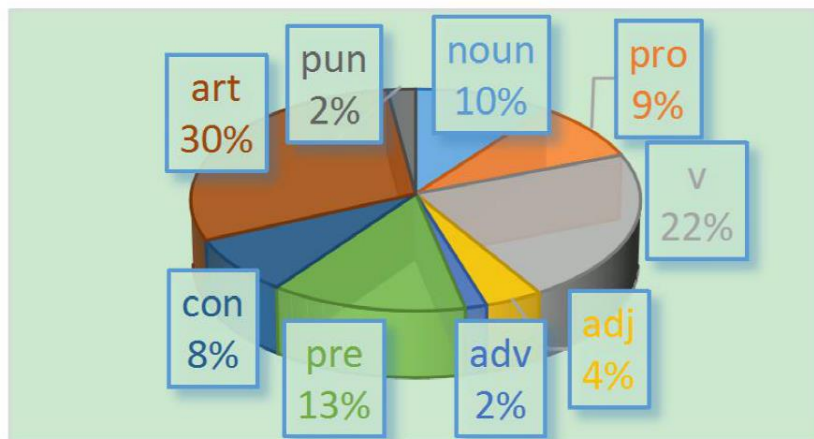


Figure 1 The Portion of Error Types

Note: adj-adjective,adv-adverb,art-article, con-conjunction,noun-noun,pre- preposition, pro- pronoun,pun-punctuation,v-verb

RQ2: What is EFR change of each type in POS categories?

The EFR change of each type is not a normal distribution. Therefore, Wilcoxon pair test should be conducted to check the EFR change of error types between pre and post-test. As figure 2 shows, the lines of EFR of error types are down. And the results of Wilcoxon test shows there are significant differences between pre-

post test in terms of article EFR change ($z=4.103$, $p[2tailed]<.05$),verb EFR change ($z=4.077$, $p[2tailed]<.05$),preposition EFR change($z=3.163$, $p[2tailed]=.002$),noun EFR change ($z=2.175$, $p[2tailed]=.03$). It means the improvement in accuracy of article, verb, preposition and noun between pre-and posttest.

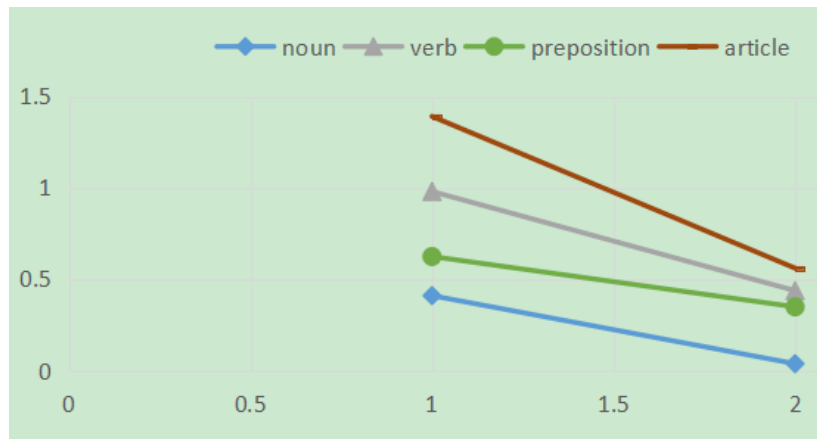


Figure2 The EFR of Change in pre-posttests

5. DISCUSSION

The results show that L2 learners usually confusing in using articles in their target language, such as omitting articles and overusing "the", which may be similar to the previous studies. Ionin(2003) states that the first main cause is that learners have incorrect or incomplete semantic representations[5]. And the second cause is that learners have complete, correct expressions for articles but uneasy choosing the lexical form during production thanks to stress on mental processing or phonological limitations. Tabatabai(1985) argues that the errors of articles have been found to occur due to omission of articles and unnecessary or wrong use of articles. These errors seem to occur due to ignorance of rules restrictions. In addition, Liu (2016) investigates that Chinese students always omit "the" or overuse it in specific contexts. Chinese writers were transferring structures from L1, and they may be likely to omit a/an. Overgeneralization may be linked to the learners who make the most errors due to overuse of "the".

And the results also show that the automated corrective feedback of Pigai.org could help L2 learners correct the POS errors in the revised drafts and the post-test, so that this study provides evidence that automated corrective feedback of Pigainet has been vastly influential in helping L2 learners improve their accurate production immediately and over time in terms of POS types, especially in article, verb, preposition, and noun. The results confirm the previous studies (Gao et al, 2020; Li et al., 2017) that have been investigated that automated corrective feedback, such as Criterion and Bingo, have a positive effect on the learners' writing accuracy in terms of grammar.

6. CONCLUSION

The study outcomes provide helpful and valuable information for future research on AWE integration in the classroom. According to Rich (2012), integrating teaching and research with online technology-based

formative assessment could become regular[8]. The research-based method may not only empower instructors and students in their local classroom teaching and learning, but it also has the potential to influence teacher practice in global English language classrooms. The ubiquity and immediacy of AWE feedback enable students to practice writing when no human assistance is available since They can access the AWE system anytime and anywhere and submit their work for evaluation (Wang and Goodman 2012) [9]. In addition, L2 writers who use an online AWE system can quick and easy access to material about language usage and idea creation via the internet.

Although the results of this study show that automated written corrective feedback effectively ensures the precise creation of specific linguistic forms, such as POS kinds in L2 writing, several limitations should be recognized and considered in future research. The POS categories examined in this study were carefully chosen to provide insight into the efficacy of feedback. But the study depends solely on the Pigainet program's error counts in terms of POS, so that it may not provide a complete picture of writing accuracy. Future research should look into other specific rule-based and item-based forms to better understand not only how but also under what conditions and automated corrective feedback treatment may (or may not) work, as well as use multiple accuracy measures, and investigate how AWE corrective feedback can help students' cognitive and metalinguistic growth and its relationship to autonomous learning.

REFERENCES

- [1] Burstein, J., Chodorow, M., & Leacock, C. (2003). Criterion SM Online Essay Evaluation: An Application for Automated Evaluation of Student Essays. In IAAI (pp. 3-10).
- [2] hukharev-Hudilainen E & Saricaoglu A (2016) Causal Discourse Analyzer: Improving Automated Feedback on Academic ESL Writing. Computer Assisted Language Learning, 29(3): 494–

516.

<https://doi.org/10.1080/09588221.2014.991795>

- [3] Gao, J., Li, X., Gu, P., & Liu, Z. (2020). An Evaluation of China's Automated Scoring System Bingo English. *International Journal of English Linguistics*, 10(6), 30.
- [4] Grimes, D., & Warschauer, M. (2010). Utility in a Fallible Tool: A Multi-site Case Study of Automated Writing Evaluation. *Journal of Technology, Learning, and Assessment*, 8(6), 4–44.13781.
- [5] Ionin, T. (2003). Article semantics in second language acquisition (Doctoral dissertation, Massachusetts Institute of Technology). MIT Libraries: <http://dspace.mit.edu/handle/1721.1/7963>
<http://dspace.mit.edu/handle/1721.1/7582>
- [6] Li, Z., Feng, H. H., & Saricaoglu, A. (2017). The Short-term and Long-term Effects of AWE Feedback on ESL Students' Development of Grammatical accuracy. *Calico Journal*, 34(3), 355–375.
- [7] McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., & Dai, J. (2015). A Hierarchical Classification Approach to Automated Essay Scoring. *Assessing Writing*, 23, 35–59.
- [8] Rich, C. S. (2012). The Impact of Online Automated Writing Evaluation: A Case Study from Dalian. *Chinese Journal of Applied Linguistics*, 35(1), 63–79. <https://doi.org/10.1515/cjal-2012-0006>
- [9] Wang, M-J. & D. Goodman (2012). Automated Writing Evaluation: Students' Perceptions and Emotional Involvement. *English Teaching & Learning* 36/3: 1–3