Abdulaziz Sanosi
PhD Applied Linguistics, Lecturer, Department of English language & Literature
Prince Sattam Bin Abdulaziz University, Howtat Bani Tamim, Saudi Arabia
ORCID ID 0000-0003-3447-2818
a.assanosi@psau.edu.sa

TO ERR IS HUMAN: COMPARING HUMAN AND AUTOMATED CORRECTIVE FEEDBACK

Abstract. The importance of Corrective Feedback (CF) to language learners has been a controversial topic for a long time. While some studies recognised CF's importance for accurate language use, others considered it deterrent to the meaningful acquisition of a second language. Recently, modern types of corrective feedback that utilise the vast advance in IT and Artificial Intelligence (AI) have emerged. This advancement has opened new investigation areas. Up to now, researchers have acknowledged the role of Automated Written Evaluation (AWE) in enhancing students' writing and motivating them. Other studies have focused on students' and teachers' perceptions of such tools. However, the particular variance between this type of CF and the traditional one is still an area to explore. Accordingly, the present study aimed to compare CF provided by teachers to that offered by a well-known writing assistant Grammarly. The descriptive design was used to analyse the CF instances provided by five college professors to the Grammarly suggestions on a corpus of 115 texts, 23700 words, written by college students. The descriptive statistics method was adopted to summarise the findings. The study's main results indicated no significant difference in the number of errors detected by the two techniques. However, human raters outperformed Grammarly in detecting grammatical errors and were more accurate in identifying structure-related mistakes. On the other hand, Grammarly was found more effective in detecting errors related to spelling and punctuation. These findings imply using focused CF to exploit both methods. Teachers can implement their regular CF approach to develop structural aspects of language. Further, they can encourage students to adopt sophisticated writing assistants to develop their writing mechanics. To account for the potential limitations of the current study, further research that employs a larger sample size and is conducted on longitudinal and experimental bases is required.

Keywords: Grammarly; writing correction; feedback; automated learning, automated written evaluation.

1. INTRODUCTION

Writing is a fundamental skill in any EFL teaching programme. Through writing, students can disseminate knowledge and manage effective communication [1, 2]. However, students face many challenges in mastering this skill as they suffer confusion in selecting the correct vocabulary, using the proper structure, or organising their writing properly [3]. Consequently, they make frequent writing errors in grammar and writing mechanics. It is suggested that "for students who learn how to write, receiving feedback on their writing is an integral part of their learning [4, p. 141]".

Statement of the problem. Corrective Feedback (CF) is widely used by EFL instructors. When applying such a strategy,"students become aware of their inadequacies and are assisted to overcome the problems they face in their language learning experience [5, p. 226]". CF has been classified in various taxonomies, including explicit versus implicit, oral versus written, focused vs comprehensive, and input-providing versus output-prompting [6]. Parallel to the vast advancements in technology and the development of artificial intelligence, a new style of CF has emerged. Writers now can get feedback on their writing on computer applications about linguistic structure, organisation, and diction by Automatic Writing Evaluation (AWE) [7]. One of the

© Abdulaziz Sanosi, 2022.

common applications in this strand is Grammarly, which is an "AI-powered writing assistant that supports clean mistake-free writing while offering suggestions that go way beyond grammar [8]".

The use of both CF and AWE entails some problems. Human CF is deemed not consistently accurate because it is affected by raters' fatigue and inconsistency, especially with the large volume of students' writing, while using AWE is criticised for missing some students' errors and overgeneralising or underestimating others. Consequently, using the proper method for correcting students' mistakes is a problem that requires broader research.

Analysis of recent studies and publications. Researchers investigated CF from several perspectives and generated different results pointing to controversies regarding how it is viewed [6, 9, 10], approving its positive impact on developing learners' writing [11], and addressing its different types and styles [12, 13]. With the emergence of the modern advance in AWE, further research studies were conducted to account for the efficacy of such technology and students' and teachers' perceptions of it [14, 15, 16]. Generally, most researchers admit the usefulness of such a technique, especially when using the Grammarly app.

As Grammarly rapidly gained ground both in the classroom and autonomous learning settings, researchers have started investigating it, especially in the last three years. These studies explored different variables and adopted various methods to scrutinise the software and how teachers and students perceive its use.

Most of the research conducted to study Grammarly was focused on measuring teachers' and students' perceptions of the software [17]. In this regard, Nova [18] reports three Indonesian postgraduate students' experiences. He collected the data through interviews and documentation. He found that Grammarly is a convenient software that provides proper fast free annotated feedback. Nevertheless, it has shortcomings, such as several false alarms relating to the content and the context.

Another study [19] compared 54 students' perception of Grammarly's feedback with 42 students who received traditional grammatical feedback. They used a 15- item survey to collect their data. The results showed that students receiving automated CF were significantly more satisfied with the grammar feedback they received than the other group.

The quasi-experimental method is used to test the role of Grammarly in assessing EFL writing. Ghufron and Rosyida [14] conducted a study involving two groups of 20 participants. The students were tested before the experiment. Then, the experimental group members used Grammarly for the whole semester to correct their writing and rectify it according to their feedback. Meanwhile, the control group members learn to write in the usual way, with CF from their teachers. Finally, the researchers administered a posttest to the two groups. The results showed a significant drop in the level of errors committed by the experimental group members compared to their pretest scores. Further, the findings revealed that the development occurred more in diction, mechanics, and grammar, but a more negligible effect was witnessed in content in the organisation.

Studies on Grammarly were not exclusive to classroom writing to evaluate errors and check them or the software versions on Microsoft word and web browsers. Studies also investigated the potential of the Grammarly virtual keyboard that works on mobile phones. To explore mobile L2 writing, Dizon and Gayed [15] studied further aspects of Grammarly that reduce errors and assist students in composing proper texts, i.e., predictive texts and intelligent writing. The participants (N = 31 university students) engaged in an 8-week free-writing task, where they used their mobile for writing under two modes: Grammarly-assisted and non-Grammarly-assisted. The results showed that students' writing was more error-free when using the predictive text feature powered by Grammarly. It was also found that more lexical variety was witnessed when using Grammarly virtual keyboard. This study is a breakthrough, since there is a paucity of research that addresses predictive texts, real-time feedback, and intelligent writing in general.

Other studies explored Grammarly from different viewpoints. Barrot [20] reviewed the software and reported its efficacy for writing development and its flexible affordances. However, some limitations of the programme should be considered, including inaccuracy in detecting errors and misjudging common phrases as plagiarised. Barrot [20] and most researchers called for more studies to explore different aspects of the software and its implementation. As there is a gap in the research on differences in CF provided by human teachers and Grammarly (in both quantity and quality), this research explores this point.

The research goal. Although a plethora of studies has focused on the use and potential of Grammarly recently, the results are still inconclusive [7], and the actual effect of using Grammarly on developing EFL students' writing is unclear. One of the variables that are not covered thoroughly regarding the new technology assessment is how exactly using Grammarly for CF provision differs from that presented by teachers. Accordingly, the purpose of this article is to explore such an area by comparing a set of CF feedback provided by college professors to those suggested by Grammarly and hence attempting to answer the following research question:

What are the differences between CF provided by human teachers and Grammarly?

2. THE THEORETICAL BACKGROUNDS

Using modern methods of AWE relies on profound literature in SLA, TESOL, and writing, as discussed below.

2.1. Corrective Feedback

Corrective feedback is defined by Chaudron [21] as "any reaction of the teacher which clearly transforms, disapprovingly refers to, or demands improvement of the learner utterance" (p. 131). Implementing (CF) techniques is one aspect of the structure-based instructional approach where "errors are frequently corrected, and accuracy tends to be given priority over meaningful interaction [22, p. 126]". It also represents one way of the *focus on form* method [23]. In this method, students perform different activities while focusing on formal properties, and "teachers can elect to provide corrective feedback on learners' errors during the course of communication activities [24, p. 639]".

CF can be traced back to the old traditional method of teaching writing that continued until the twentieth century. At that time, "writing instruction was based on a somewhat rigid set of assumptions: good writing was done from a set of rules and principles. … A student essay was then graded for its grammatical accuracy and correct organisation as well as its content [25, pp. 88-89]". Notwithstanding, there has been considerable controversy regarding the efficacy and practicality of written CF [6, 9, 12, 26]. It has firstly, and still to a great extent, been seen as a "means of fostering learner motivation [6, p. 3]", and there is sufficient evidence that it "significantly improves accuracy [27, p. 115]". Later, other counter-claims have emerged, disregarding the efficacy of written CF.

The initial view in this strand was raised by Truscott [28], who not only claimed that CF is ineffectual for both logical and functional reasons but also believed that it "has a harmful effect [28, p. 327]". His view was not a one-time opinion; instead, after 25 years, Truscott insists that he is still sceptical about the efficacy of correction. He admitted that if he could turn the clock back, he "might adjust some details of the presentation and change the emphasis in places, but [he does] not see a need for any substantive changes in the arguments [29]". The major concerns behind Truscot's views about the inefficacy of written CF lie in the view that, basically,"grammar correction is a bad idea

[30, p. 111]". Moreover, he finds methodological issues pertaining to the research that conclude its efficacy and misinterpretation of the findings yielded by those studies [31].

The views of Truscott [28] were challenged by many subsequent studies that support using CF. Bitchener and Ferris [32] admitted some constraints regarding using CF, which may lower its efficacy, such as "cognitive load, selective attention, learner readiness, and so on" (p. 27). However, CF would foster language learning when considering these constraints as suggested by different theories and approaches reviewed by the researchers. Kang and Han [33] reviewed 21 studies related to CF practice and results. They concluded that "corrective feedback can lead to greater grammatical accuracy in second language writing, yet its efficacy is mediated by a host of variables, including learners' proficiency, the setting, and the genre of the writing task." (p. 1). Up to now, there is no conclusive answer to the question of the exact effect of written CF in improving EFL learners' writing. However, it is a state of the art, familiar, and indispensable practice in the EFL classroom [2] that needs more critical research to explain its value and impact.

2.2. Automated Corrective Feedback

Using the computer and the associated technologies in the educational setting has become a fundamental practice today. Now, it is hard to imagine executing different activities without utilising the computer as an essential or assisting tool in a language classroom. Furthermore, the new circumstances imposed on the world by the outbreak of the Corona pandemic in 2019 make sceptical families and practitioners rethink their attitudes towards CALL, including online learning, CMC-based and autonomous learning activities. On the other hand, the vast advance in computer science and AI enables highly advanced applications that have transformed computers into smart tutors that can teach, follow up and evaluate learners. Of the common applications in this strand are those that adopt AWE technology.

Automated Writing Evaluation was introduced in the EFL classroom setting to make up for teachers' heavy workload, which prevents them from accurate follow up of their students' writing [34]. Moreover, the unique features of computers outweighed teachers in evaluating students' writing, as "computer scoring can be faster, reduce costs, increase accuracy, and eliminate concerns about rater consistency and fatigue [35, p. 3]". These applications were firstly used for scoring and enabling more practice for students. Although some uncertainty arises about its value for L2 learners and regarding errors and consistency in detecting them [36], AWE's ability to check students' spelling, grammar, wordchoice, tone, and plagiarism is recognised [34, 35]. Accordingly, such applications are used by students in different settings. A well-known example of software with these potentials is Grammarly.

2.3. Grammarly

Grammarly is an AI-Powered writing assistant that can perform much more than its name implies. It does not settle for just grammar correction; instead, it checks and corrects writers' spelling, word choice, punctuation, clarity, engagement, and delivery errors. The software was founded in 2009 and has developed fast over the past few years to reach 30 million daily active users [8]. Since it has widely been used in EFL settings, Grammarly has attracted researchers' attention, and most studies claimed its convenience, helpfulness, accuracy and efficacy [37, 38, 18, 19, 20, 17].

Grammarly is presented to users on different platforms. As of May 2022, it has an application for Microsoft Windows where users can import, paste, or type texts and check them for correctness, clarity, engagement, and delivery. Before using the features, users should adjust their goal to enable

the app to adapt the feedback it presents. The process is done by choosing different variables, including the writer's audience, level of formality, and domain. Further, Grammarly enables users to check for errors while typing in Microsoft Word or outlook. A plugin integrated into the office package allows this potential in real-time. Another plugin is available for the Chrome web browser. The plugin monitors users' typing on different websites and suggests editing them to generate proper texts. Recently, Grammarly introduced its virtual keyboard to work in Android-powered smartphones in all users' favourable apps and an application for iPad.

3. METHODOLOGY

This study adopted a descriptive quantitive method that compared the number and types of corrective feedback suggested by human raters and Grammarly writing assistant.

3.1. Corpus

A total of 115 texts constitute the corpus of the study. The average text length is 206 words, making the total word count of the corpus 23700 words. The texts were written by Arab L1 students studying English Language and Literature at Prince Sattam ibn Abdulaziz, Saudi Arabia. The texts are narrative. Students were asked to write about themselves in three paragraphs to describe their childhood, preferences, and future dreams.

3.2. Procedures

The texts were collected through Blackboard Learning Management System; then, they were rated in two phases. In the first phase, five university professors evaluated the files manually. The raters were asked to determine the errors and mark each error type with a unified code set (see Appendix A). After that, the texts were uploaded individually to Grammarly App for Windows with a premium subscription. The goals of the Grammarly app, which control the style of feedback required, were set, as seen in Figure 1 below.

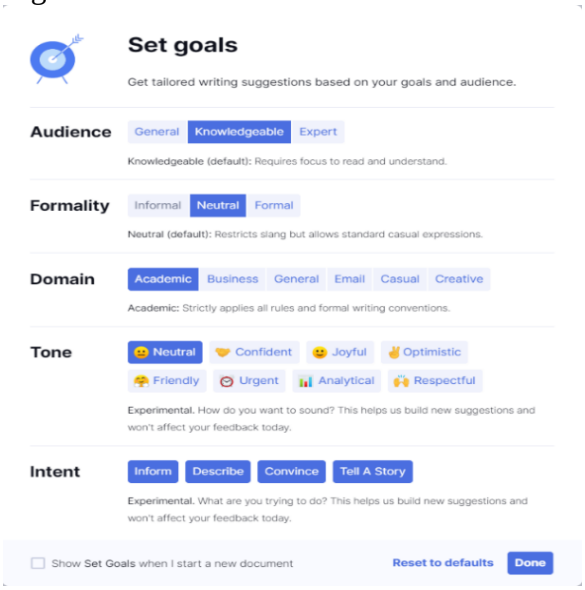

Figure. 1 Grammarly Goals as Set by the Researcher

The goals were set to generate suggestions suitable for informative academic writing.

3.3. Data Analysis:

The corrective feedback suggestions were categorised into three groups: (1) Grammar, (2) Spelling (including capitalisation), and (3) Punctuation (marks and spaces). Grammar errors were categorised into two subgroups: (a) sentence structure (including verb tense; subject-verb agreement; singular/plural agreement; fragment; pronoun reference; possessive; and missing conjunction), and (b) word choice (including wrong, redundant, or missing article; wrong word form; or missing word).

The researcher designed an Excel workbook with columns for human raters and others for Grammarly suggestions for the data analysis. Types and numbers of feedback suggestions presented by human raters and Grammarly were recorded for each text. Finally, each type's total and percentage of suggestions were calculated. The findings compare the quantity and quality of CF provided by human and automated raters, as presented in the following section.

4. THE RESULTS AND DISCUSSION

Table 1 summarises the feedback presented by human raters and the Grammarly suggestions to the corpus:

Table 1

Errors Corrected by Human Raters and Grammarly

Rater	Grammar	Spelling	Pun	Total
Human	879	354	312	1545
Grammarly	624	515	470	1609

The findings show differences in the corrective feedback suggestions provided by human raters and Grammarly in terms of the total mistakes detected and the distribution of those mistakes to different types of errors. Figure 2 shows the percentage of each type of error as seen by human raters.

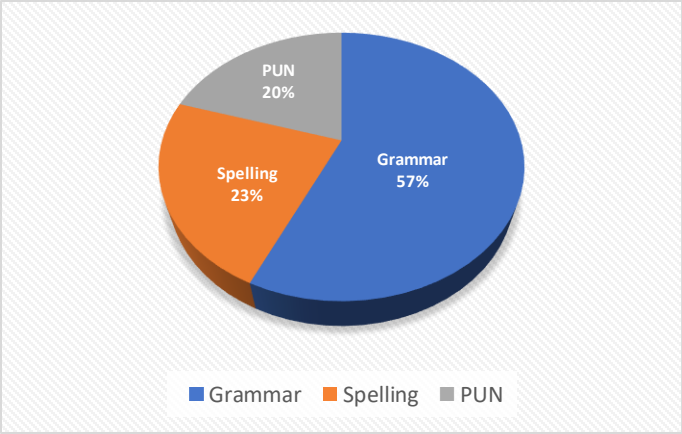

Figure 2. Types of Errors Corrected by Human Raters

Contrary to the human raters, whose feedback concentrates primarily on grammar, Grammarly suggestions addressed different types of mistakes, as can be seen in Figure 3 below:

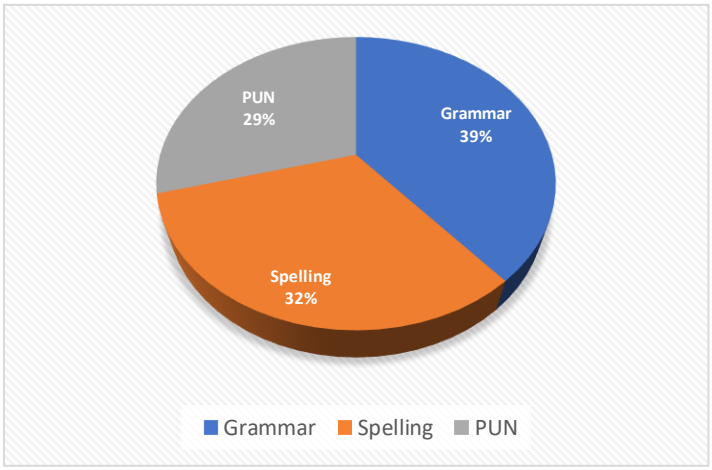

Figure 3.Types of Errors Corrected by Grammarly

A detailed analysis of the data shows a further difference in the types of grammatical mistakes detected by the two raters. Figure 4 details the types of grammatical errors detected by Grammarly and the human raters.

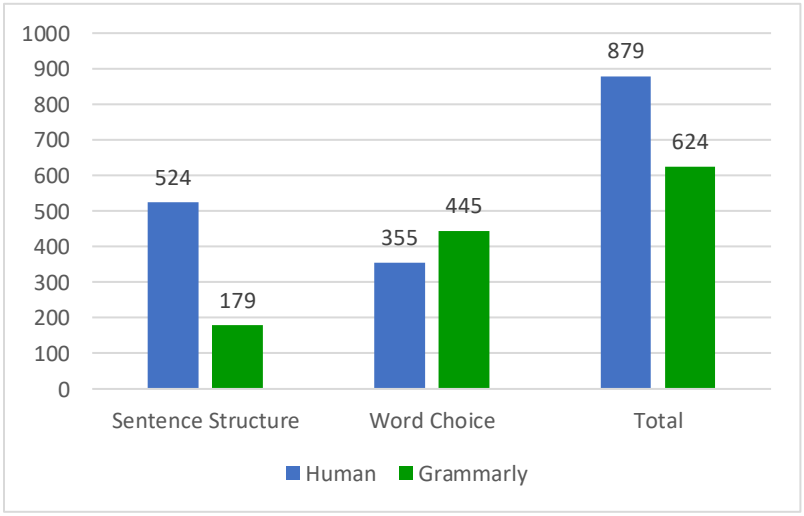

Figure 4.Types of Grammatical Errors Corrected by Human Raters and Grammarly

The result suggested a negligible difference in the volume of the mistakes detected by human and automated raters in general. Grammarly identifies more mistakes than human raters; however, the difference is insignificant. Grammarly surpassed human raters by only 64 errors in the 115 texts (with an average of 0.5 mistakes per text). This difference, though trivial, can be traced back to the nature of AWE in general, which is more sensitive to formatting and punctuation than human raters. For example, many corrective suggestions were labelled (correctness: convention) with the feedback: "*It appears that you have an extra space between the words"*. Teachers cannot easily detect these faults, especially when the extra space is placed at the end of the line. Other mistakes that might be identified by Grammarly only and cause the outnumbering are proper Arabic nouns

used frequently in almost all the texts because of the nature of the assignment (narrative and personal). When considering these factors, it can be stated that human and automated raters addressed the mistakes almost similarly. Primarily, this finding is consistent with previous studies, as it implies the potential of Grammarly in assessing students' writing in general. However, more details on the quality and quantity of mistakes should be discussed to reveal the implications of such findings.

Considering the type of mistakes spotted by the two raters, it was found that human raters focused more on grammatical errors. More than half of the errors detected (57%) were grammatical. The remaining errors were distributed almost evenly between spelling errors (23%) and punctuation (20%). This finding implies an essential difference between human and automatic feedback, given that the grammatical errors represented only 39 % of the mistakes detected by Grammarly. The supremacy of teachers' detection of grammatical errors was not only on the quantity level; it was also found that several feedbacks concerning the detected grammatical errors were inaccurate or not understandable. Figure 5 presents some examples of these false alarms.

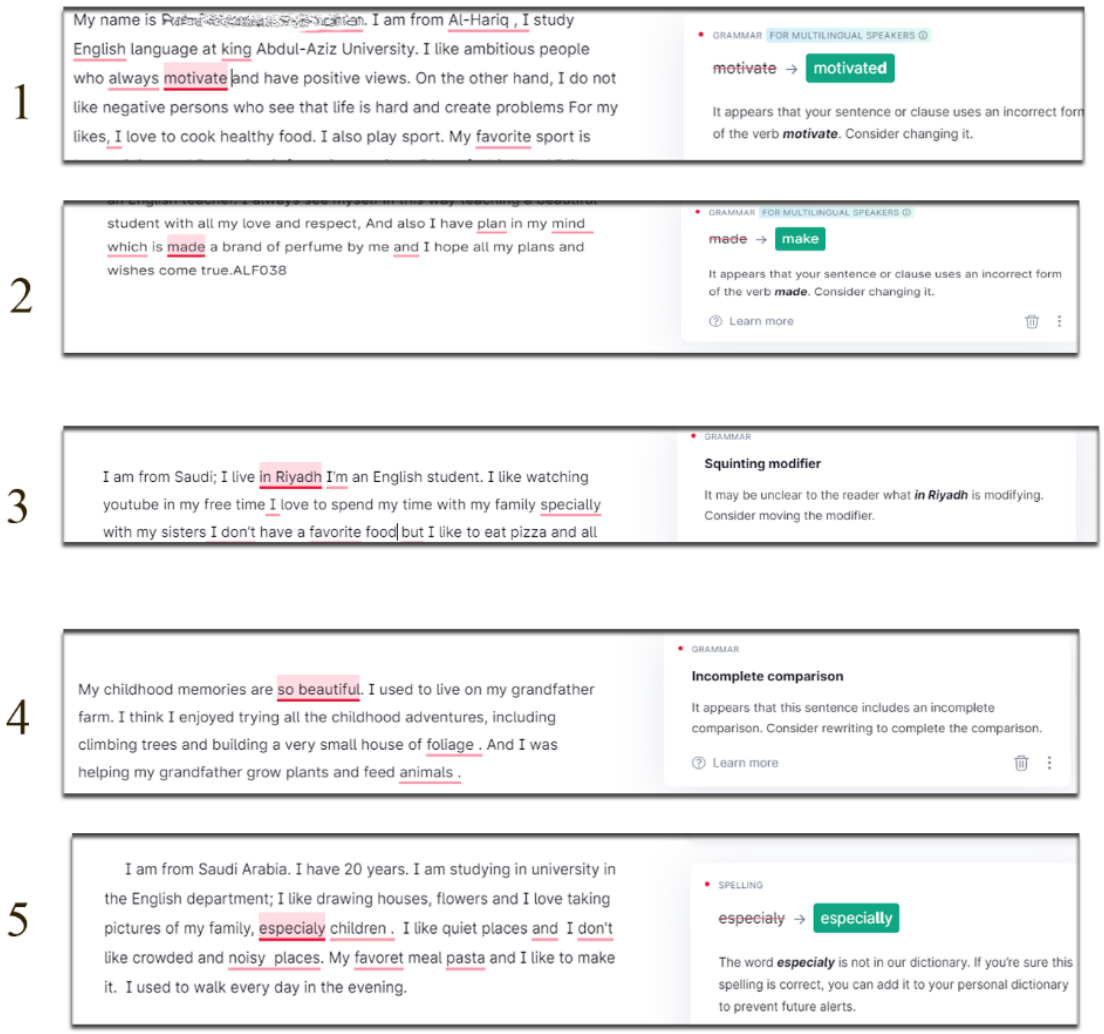

Figure 5.Examples of grammatical errors improperly corrected by Grammarly

Example 1 shows a case where the suggestion was not understandable. The writer used present simple tense to express her preferences. The other verbs in the sentences: *like* and *have*, were not considered wrong; notwithstanding, Grammarly suggested changing the verb *motivate* to the past form for an unclear reason. It is worth mentioning that there were no correction suggestions for the same sentence made by the human raters. For the second example sentence, human raters agree that the verb *make* is in the wrong form, just like what Grammarly suggests. However, the correction offered by Grammarly, i.e. to change the verb to *make*, is also incorrect and did not match any of the teachers' suggestions. It is clear that the sentence is considered passive voice as the auxiliary *is* precedes the verb. Nevertheless, this suggestion would be unhelpful and even confusing for students. In Example 3, the phrase *in Riyadh* is considered a squinting modifier. According to Joki [39], a squinting modifier is "a misplaced modifier that, because of its location in a sentence, could modify either the phrase that precedes it or the one that follows it". This suggestion seems improper as it is clear that the phrase modifies the phrase *I live …*, which precedes it. The mistake is then more likely to be considered a punctuation error rather than a grammatical one.

The inaccuracy of grammar detection is also represented in Example 4. The sentence *My childhood memories are so beautiful* is considered erroneous and labelled with an *incomplete comparison* tag. It is again vague why it is presupposed that the writer is comparing two things here. The adverb *so* is used mainly to mean extent or degree when it is placed before adjectives [40, 41]. Accordingly, the sentence, or at least the use of the adjective, is correct. The feedback presented by two professors for this sentence was to replace the verb 'are' with 'were'. Grammarly does not offer this suggestion, of course, as it is unexpectable from an AI-powered application to understand the context of describing a past event in this case. It is noted that even the human raters did not agree that the writer describes the past events or his recent memories about them.

Example 5 shows a missing, rather than an improper grammatical correction feedback. In this example, the writer used the word *have* to write about his age. The L1 interference is evident in this sentence, as it is common to use the Arabic cognate of the word 'have' to talk about one's age. All human raters marked this as a word-choice mistake and suggested replacing it with *I am 20 years old*, whereas Grammarly did not take action. Again, this may give wrong feedback to students and distort their learning as they may consider their use correct.

From the above discussion, it is implied that human CF is more related to theories on AWE. This implication lies mainly on the ground that "effective L2 pedagogy should involve – at least at times – attention to linguistic form [9, p. 5]". The focus-on-form approach is primarily applied in written CF to draw students' attention to their non-target-like use of language and develop their learning competence. Although Grammarly is a powerful writing assistant tool whose contribution to developing writing can not be neglected, its role in developing language acquisition and students' interlanguage can be questioned according to these theoretical bases. Two remarks can justify this; firstly, some feedback is unclear, which contradicts the concept of comprehensible input [42] demanding that the message presented to students should be understandable to guarantee proper language acquisition. Secondly, the nature of AI prevents AWE applications from detecting the semantic and pragmatic connotations of some expressions such as L1 proper nouns, culture-related expressions, and other L1-transferred units. For this reason, wrong feedback is occasionally presented to students and, hence, negatively affects students' learning and acquisition.

The nature of grammar error detection is also reflected in the types of errors detected by human and automatic raters. It is found that human raters focused on grammar errors related to sentence structure. Most of the errors (60%) marked by teachers were related to sentence structure. These errors were distributed among fragments, especially non-auxiliary sentences, e.g. *my favorite meal pasta* in example 5; verb-tense errors, e.g. *which is made a brand of perfume by me*, example

2; and singular/plural or subject/verb agreement. The remaining errors were on word choice, i.e. wrong-preposition errors and wrong word-class selection.

On the other hand, most of the grammatical errors detected by Grammarly (71%) were related to word-choice type. It was interesting that most of these errors were not even identified by human raters. For example, a missing article, especially the indefinite article before a singular noun phrase of adjective + noun format, was the most frequent grammatical error detected by Grammarly and the least one marked by human raters. This finding indicates that the case of missing indefinite articles was more challenging for teachers, as many cases detected wrong by Grammarly passed the human check successfully. When one compares the two raters' performance, it can be claimed that Grammarly can better mark word-choice errors while human raters outperform AWE in finding structure-related errors.

The above finding implies using focused feedback to evaluate students' writing in classroom or autonomous learning settings. Focused feedback "addresses one or two error types [6, p. 6]" and is believed to "improve learning in terms of task achievement and grammar [16, p. 208]". Accordingly, it can be adopted to enhance students' writing and inform teachers' instruction by selecting the appropriate technique for correcting each type of error. In the case of the present study, activities on using articles can be administered using Grammarly as a writing assistant. As far as the sentence structure is concerned, teachers' rating would be more practical. This implication is also consistent with students' preferences. It was found that students prefer teacher feedback on their grammatical performance as they think grammar mistakes need expert raters to identify and correct [4]. Accordingly, other types of feedback, e.g. peer or automated feedback, may not be considered helpful by them.

Focused CF can also be applied concerning other types of errors detected by the human raters and Grammarly. It was found that 32% of the errors detected by Grammarly were spelling errors, and 29% of them were punctuation errors. When comparing this finding to human raters' output, it is evident that Grammarly outperforms teachers in detecting punctuation and spelling errors. Spelling errors detected by teachers represent 23%, while punctuation errors represent 20% of the overall errors. Again, the nature of the AI systematic analysis can justify this. Proper nouns, extra spaces, and capitalisation caused several errors. However, the difference between the number of errors detected by the two systems is still significant.

5. CONCLUSIONS AND PROSPECTS FOR FURTHER RESEARCH

This study aimed at comparing human to automated CF. It analysed CF provided by university professors and Grammarly on a corpus of 115 texts written by college students. The results indicated that human raters were more accurate in detecting grammatical errors. Moreover, errors related to sentence structure are more likely to be seen accurately by teachers than by Grammarly. On the other hand, Grammarly was more successful in detecting spelling, punctuation, and word-choice errors. We suggest several justifications for these findings, including the nature of AI application and the linguistic characteristics of the texts, i.e. L1-based expressions. The latter can be considered a limitation of the study, which should be looked into in further research. Excluding these items can provide a more transparent image of how Grammarly can be compared to human correction practice.

Moreover, several incomprehensible suggestions by Grammarly concerning grammatical errors were spotted. This finding indicates that Grammarly may not be beneficial in second language acquisition and interlanguage development. Nevertheless, it can not be denied that it

provides accurate feedback concerning spelling, punctuation, and word choice. The above results imply using focused CF. Our suggestion is to adopt human raters for checking grammar and structure and Grammarly for spelling and punctuation. Further research can focus on more extensive data that exclude L1 expressions and include more human raters. This research should adopt a longitudinal and experimental approach that explores the development of students at specific checkpoints in both human and automatic CF settings.

REFERENCES (TRANSLATED AND TRANSLITERATED)

[1] D. Anh, "EFL Student's writing skills: Challenges and remedies," *IOSR J. of Res.& Methods in Educ.*, vol. 9, no. 1, pp. 74-84, 2019.

[2] Y. Luo and Y. Liu, "Comparison between peer feedback and automated feedback in college English writing: A case study," *Open J. of Modern Ling.*, vol. 7, pp. 197-215, 2017.

[3] R. Moses and M. Mohamad, "Challenges Faced by Students and teachers on writing skills in ESL contexts: A literature review," *Creative Educ.*, vol. 10, no. 13, 2019.

[4] S. Thirakunkovit and B. Chamcharatsri, "A Meta-Analysis of the effectiveness of teacher and peer feedback: Implications for writing instructions and research," *Asian EFL J.*, vol. 1, no. 21, pp. 140-170, 2019.

[5] J. Nikaoopour and A. Zoghi, "Analysing EFL learners' errors: The plausibility of teachers' feedbacks and students' uptakes," *J. of Lang.Teaching and Res.*, vol. 5, no. 1, pp. 226-233, 2014.

[6] R. Ellis, "Corrective feedback and teacher Development," L2 J., vol. 1, no. 1, pp. 3-17, 2009.

[7] J. Li, S. Link and V. Hegelheimer, "Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction," *J. of L2Writing*, vol. 27, pp. 1-18, 2015.

[8] "Grammarly," 2021. [Online]. Available: https://www.grammarly.com/about.

[9] C. Beuningen, "Corrective feedback in L2 writing: Theoretical perspectives, empirical insights, and future directions," *Int. J. of English Stud., vol. 10*,no. 2, pp. 1-27, 2010.

[10] R. Al-Jarrah and S. Al-Ahmad, "An Optimality-theoretic account of corrective feedback in process writing," *The Asian EFL J.*, vol. 14, no. 3, pp. 9-39, 2012.

[11] J. Bitchener, "Evidence in support of written corrective feedback," *J. of L2Writing*, vol. 17, pp. 102-118, 2008.

[12] S. Jang, "The efficacy of different types of metalinguistic information in L2 written corrective feedback," *English Teaching*, vol. 75, no. 4, pp. 33-56, 2020.

[13] H. Yaqoob, "Traditional and metalinguistic teachers' corrective feedback to EFL Iraqi university-level students," *Al BasrahStud. J.*, no. 37, pp. 73-106, 2020.

[14] M. Ghufron and F. Rosyida, "The role of Grammarly in assessing English as a foreign language EFL writing," *Lingua Cultura*, vol. 12, no. 4, pp. 395-403, 2018.

[15] G. Dizon and J. Gayed, "Examining the impact of Grammarly on the quality of mobile L2 writing," *The JALT CALL J.*, vol. 17, no. 2, pp. 74-92, 2021.

[16] M. Ibrahim, "The effect of Computer-Mediated Corrective Feedback in improving IELTS students' writing skills," *The Asian ESP J.*, vol. 17, no. 4.2, pp. 205-221, 2021.

[17] M. Nova, "UtilisingGrammarly in evaluating academic writing: A narrative research on EFL students' experience," *J. of English Educ. and Appl. Linguistics*, vol. 7, no. 1, pp. 80-86, 2018.

[18] R. O'Neill and A. Russel, "Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly," *Aust. J. of Educ. Technol.*, vol. 35, no. 1, pp. 42-56, 2019.

[19] J. Barrot, "Integrating Technology into ESL/EFL Writing through Grammarly," *RELC*, pp. 1-5, 2020.

[20] C. Chaudron , "A descriptive model of discourse in the corrective treatment of learners' errors," *Lang. Learn.*, vol. 27, pp. 29-46, 1977.

[21] P. Lightbown and N. Spada, *How languages are learned*, 4th ed., Oxford University Press, 2013.

[22] M. H. Long, "Focus in form in task-based language teaching," in Language policy and pedagogy. *Essays in honor of A. Ronald Walton, R. Lambert and E. Shohamy*, Eds., John Benjamins, 2000, pp. 179-192.

[23] R. Ellis, *The study of second language acquisition*, Oxford University Press, 1994.

[24] M. Sokolik, "Writing," in *Practical English Language Teaching*, D. Nunan, Ed., McGraw-Hill, 2003, pp. 87-108.

[25] A. Maleki and E. Eslami, "The effects of written corrective feedback techniques on EFL students' control over grammatical construction of their written English," *Theory and Pract. in Lang.Stud.*, vol. 3, no. 7, pp. 1250-1257, 2013.

[26] J. Bitchener, "Evidence in Support of written corrective feedback," *J. of L2Learn.*, vol. 17, pp. 102-118, 2008.
[27] J. Truscott, "The case against grammar correction in L2 writing classes," *Lang. learn.*, vol. 46, pp. 327-369, 1996.
[28] H. Mohebbi, "25 years on, the written error correction debate continues: an interview with John Truscott," *Asian-Pacific J of 2nd. and Foreign Lang.Educ.*, vol. 6, no. 3, pp. 1-8, 2021.
[29] J. Truscott, "The Case for "The case against grammar correction in L2 writing classes": A Response to Feris," *Journal of second language learning*, vol. 8, no. 2, pp. 111-122, 1999.
[30] J. Truscott, "The Efficacy of written corrective feedback: a critique of a meta-analysis," May 2020. [Online]. Available: https://www.researchgate.net/publication/341453398_The_Efficacy_of_Written_Corrective_Feedback_A_Critique_of_a_Meta-analysis.
[31] J. Bitchener and D. Ferris, *Written corrective feedback in second language acquisition and writing*, Routledge, 2012.
[32] E. Kang and Z. Han, "The efficacy of written corrective feedback in improving written L2 accuracy: A meta-analysis," *The Modern Lang.J.*, vol. 99, no. 1, pp. 1-18, 2015.
[33] J. Burstein and M. Chodorow, "Automated essay evaluation: The Criterion online writing Service," *AI Mag.*, vol. 25, no. 3, pp. 27-36, 2004.
[34] L. Rudner and T. Liang, "Automated essay scoring using Bayes' theorem," *The J. of Technol., Learn., and Assessment*, vol. 1, pp. 1- 21, 2002.
[35] R. Jim, "Automated written corrective feedback: How well can students make use of it?" *Comput. Assisted Lang. Learn.* , vol. 3, no. 7, pp. 653-674, 2018.
[36] P. Daniels and D. Leslie, "Grammar software ready for EFL writers," *OneCue journal*, vol. 9, no. 4, pp. 391-401, 2013.
[37] A. Qassemzadeh and H. Soleimani, "The impact of feedback provision by Grammarly software and teachers on learning passive structures by Iranian EFL learners," *Theory and Pract. in Lang.Stud.*, vol. 6, no. 9, pp. 1884-1894, 2016.
[38] K. Joki, "The Basics of Ambiguous (Squinting) Modifiers," 14 January 2021. [Online]. Available: https://www.grammarly.com/blog/the-basics-of-ambiguous-squinting-modifiers/.
[39] Merriam-Webster, "So," n.d. [Online]. Available: https://www.merriam-webster.com/dictionary/so . Accessed on: Nov.10 2021.
[40] Cambridge-Dictionary, "So," n.d. [Online]. Available: https://dictionary.cambridge.org/grammar/british-grammar/so. Accessed on: Nov.10 2021.
[41] S. Krashen, *The input hypothesis: Issues and implications*, London: Longman, 1985.

Text of the article was accepted by Editorial Team 23.05.2022

ЛЮДИНІ ВЛАСТИВО ПОМИЛЯТИСЯ: ПОРІВНЯННЯ КОРИГУЮЧОГО ЗВОРОТНОГО ЗВ'ЯЗКУ ЛЮДИНОЮ ТА АВТОМАТИЧНО

Абдулазіз Саносі
PhD прикладної лінгвістики, викладач кафедри англійської мови та літератури
Університет принца Саттама бін Абдулазіза, Хаутат Бані Тамім, Саудівська Аравія
ORCID ID 0000-0003-3447-2818
a.assanosi@psau.edu.sa

Анотація. Важливість коригуючого зворотного зв'язку (КЗЗ) для тих, хто вивчає мову, протягом тривалого часу була суперечливою темою. У той час, як деякі дослідження визнавали важливість КЗЗ для правильного використання мови, інші вважали, що він перешкоджає осмисленому засвоєнню другої мови. Нещодавно з'явилися сучасні типи коригуючого зворотного зв'язку, які використовують значні досягнення ІТ та штучного інтелекту (ШІ). Цей прогрес відкрив нові сфери дослідження. Сьогодні дослідники визнали роль автоматизованого письмового оцінювання (АПО) у поліпшенні грамртності учнів та їх мотивації до цього. Також досліджувалось сприйняття студентами та вчителями таких інструментів. Проте конкретна різниця між цим типом КЗЗ і традиційним все ще залишається предметом дослідження.

Відповідно, дане дослідження мало на меті порівняти КЗЗ, який надають вчителі, з тим, що пропонує відома онлайн-платформа Grammarly. Описовий метод використовувався для аналізу екземплярів КЗЗ, представлених п'ятьма професорами коледжу, як пропозиції до Grammarly корпус, які складаються із 115 текстів, 23700 слів, написаних студентами коледжу. Для узагальнення результатів було використано метод описової статистики. Основні результати дослідження показали відсутність істотної різниці в кількості помилок, виявлених двома методами. Проте оцінювачі-люди перевершили Grammarly у виявленні граматичних помилок і були більш точними у визначенні структурних помилок. З іншого боку, Grammarly було визнано більш ефективним у виявленні помилок, пов'язаних з орфографією та пунктуацією. Ці висновки означають використання сфокусованого КЗЗ для використання обох методів. Учителі можуть застосовувати свій звичайний підхід КЗЗ для розвитку структурних аспектів мови. Крім того, вони можуть заохотити студентів використовувати більш складні сервіси з письма для розвитку механіки письма. Щоб врахувати потенційні обмеження поточного дослідження, необхідні подальші дослідження, які використовують більший розмір вибірки та проводяться на лонгітюдній та експериментальній основі.

Ключові слова: Grammarly; корекція письма; зворотний зв'язок; автоматизоване навчання; автоматизоване письмове оцінювання.

APPENDICES

Appendix A. Error Annotation Codes

Please underline the mistakes and identify the error type above them using the following codes.

Error Type	Description	Code
Sentence Structure	Verb tense	VT
	Singular/plural	S/P
	Subject-verb Agreement	S-V
	Conjunction (correlative, coordinating, subordinating)	FR
	Pronoun (Person, Number, Case)	PRO
	Fragment	FR
Word Choice	Wrong word choice/Missing word (Preposition, adjective, adverb)	WC
	Article (*missing, redundant, wrong)*	ART
Spelling	Spelling mistakes and writing conventions, e.g. (capitalisation)	SP
Punctuation	Marks (commas, periods, question marks)	PUN

(cc) BY-NC-SA

This work is licensed under Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.