International Journal of Internet, Broadcasting and Communication Vol.13 No.4 22-29 (2021)
http://dx.doi.org/10.7236/IJIBC.2021.13.4.22

IJIBC 21-4-3

Evaluating Corrective Feedback Generated by an AI-Powered Online Grammar Checker

Dosik Moon

Associate Prof, Dept of English, Hanyang Cyber University, Korea
dmoon@hycu.ac.kr

Abstract

This study evaluates the accuracy of corrective feedback from Grammarly, an online grammar checker, on essays written by cyber university learners in terms of detected errors, suggested replacement forms, and false alarms.The results indicate that Grammarly has a high overall error detection rate of over 65%, being particularly strong at catching errors related to articles and prepositions. In addition, on the detected errors, Grammarly mostly provide accurate replacement forms and very rarely make false alarms. These findings suggest that Grammarly has high potential as a useful educational tool to complement the drawbacks of teacher feedback and to help learners improve grammatical accuracy in their written work. However, it is still premature to conclude that Grammarly can completely replace teacher feedback because it has the possibility (approximately 35%) of failing to detect errors and the limitations in detecting errors in certain categories. Since the feedback from Grammarly is not entirely reliable, caution should be taken for successful integration of Grammarly in English writing classes. Teachers should make judicious decisions on when and how to use Grammarly, based on a keen awareness of Grammarly's strengths and limitations.

Key words: *Online Grammar Checker, Corrective Feedback, Grammarly, Grammatical Accuracy*

1. Introduction

Continuous writing practice and feedback are widely considered essential to achieve a high proficiency in English writing. Particularly, teacher corrective feedback plays a significant role in improving English as a foreign language (EFL) learners' grammatical accuracy at the sentence level. Thus, the learners usually expect to receive frequent and immediate corrective feedback from their writing instructors on their written work [1]. However, providing feedback is such a time-consuming and burdensome task that it is not easy for the instructors to sufficiently meet learners' expectations for feedback. Worse still, the classroom settings with large class sizes and limited class time do not allow an instructor to provide learners with effective feedback which will help improve their writing skills.

Online grammar checkers (GCs) seems to have a considerable potential to be a useful language learning tool with a capability of providing quick, convenient, and easy-to-use feedback [2]. Feedback generated by

Manuscript Received: August. 17, 2021 / Revised: August. 25, 2021 / Accepted: August. 27, 2021
Corresponding Author: dmoon@hycu.ac.kr
Tel: +82-2-2990-0334, Fax: +82-2-2290-0600
Associate Professor, Department of English, Hanyang Cyber University, Korea

GCs is known to have several benefits when used in English writing classes: to lessen teachers' workload for giving feedback, to improve learners' grammatical accuracy, and to promote autonomous learning in writing [3]. Before integrating GCs in English writing classes, one problem remains to be addressed first regarding the accuracy of feedback provided by the GC. Currently, numerous state-of-art online GCs such as SpellCheckPlus, Ginger, Grammarly, Reverso, claim that they have achieved high accuracy by applying artificial intelligence, algorithm application, and natural language processing technology. Conversely, several studies have reported that GCs are weak at identifying certain types of grammatical errors or sometimes provide incorrect suggestions for alternative forms [4].

The accuracy of feedback is especially important for EFL learners, who lack the ability to accurately determine whether their own sentences or texts are grammatically correct. Inaccurate feedback can confuse these learners and lead them to make more errors. Although the feedback accuracy of recently developed online GCs has considerably improved compared with that of the earlier ones, it is still pointed out as a weakness of the GC to provide inaccurate or decontextualized feedback. These mixed results about the accuracy of corrective feedback from GCs suggest that more research is needed to evaluate the strengths and weaknesses of feedback generated by GCs.

This study evaluates the accuracy of the corrective feedback provided by Grammarly, an online grammar checker regarded as achieving the highest level of accuracy among GCs currently available. To assess the accuracy of the feedback, the rate of errors detected, replacement suggestions made, false alarms marked by Grammarly on a broad range of grammatical error types on narrative essays written by Korean EFL learners at a Cyber University. The purpose of this study is to examine the educational value of Grammarly to complement the drawbacks of teacher feedback and meet learners' needs for improving grammatical accuracy in their written work.

2. Prior Research

Early research on the accuracy of corrective feedback provided by GCs focused mainly on a limited set of error types such as articles, determinants, and prepositions, measuring the accuracy rate and recall rate of feedback. The results were somewhat different depending on the type of GC used, but most research showed an accuracy rate of over 60% though the recall rate was less than 50%. For example, a program called web-frequency algorithm was found to have an accuracy of 62% and a recall rate of 41% [5]. On the other hand, a GC called A maximum entropy classifier, which focused only on prepositions, was found to have a higher accuracy of 80% while the recall rate was significantly lower at 30% [6].

Meanwhile, a comparative research evaluating the feedback capability of two types of GCs, NTNU and Microsoft ESL Assistant, produced quite different results. Dealing with a wider scope of error types including articles, verbs, SVA, run-ons/fragments, spelling, and compounds, NTNU achieved an accuracy of 61% and a recall rate of 72%, which was significantly better than the Microsoft ESL Assistant [7]. Despite these overall positive results, these earlier GCs were not considered reliable by English educators and few were widely adopted as educational tools in English writing classes.

Recently, a growing number of studies on the AI-based Grammarly GC have obtained positive results, suggesting that GCs possess high potential for facilitating effective language learning. A study on learner satisfaction with Grammarly reported that learners using Grammarly were more satisfied with the quality of feedback on language use than those who received teacher feedback [8]. Grammarly was also found to be highly effective in helping learners identify and correct grammatical errors to improve the quality of English writing [9]. Furthermore, feedback from Grammarly can lead learners to notice contrary evidence in their writing that may help them to apply their cognitive and metacognitive operations [10]. Grammarly also has

the capability of providing effective grammar support regardless of international or domestic education contexts, either online or face-to-face mode [11].

Meanwhile, research on the accuracy of Grammarly has reported its limitations as well as strengths. According to a study comparing Grammarly with three other GCs, Grammarly achieved the highest rate of overall accuracy, but this rate was only 44.4%, with an exceedingly low accuracy rate (5%) in detecting sentence structure errors [12]. Similarly, in a more recent study, Grammarly revealed some limitations in handling tense-aspect, word order, and pronoun errors [13]. Worse still, it failed to detect any tense shift errors although there were 46 committed errors.

3. Research Methods

The current study partially replicates an earlier study [13] in terms of research design. Like [13], this study used Grammarly free version to evaluate the accuracy of its corrective feedback in terms of the rate of successfully detected errors, of inappropriate replacement forms suggested, and of false alarms, which are the correct forms that CGs misidentify as errors. A main difference between the two studies is that while [13] compares the feedback of Grammarly on authentic essays and fabricated simple sentences with that of other two GCs in an English as a second language (ESL) context, this study examines the feedback of Grammarly on authentic essays in an EFL context.

3.1 Merits of Grammarly as an English Learning Tool

Grammarly is an AI-powered English GC that generates automatic feedback by identifying errors in grammar, vocabulary, and language style. Known as the most accurate GC among various GCs available on the market, Grammarly has several merits as an English learning tool. First of all, it provides real-time corrective feedback on roughly 500 kinds of grammatical errors so that learners can immediately use it to improve their written work. Besides corrective feedback on grammatical errors, it also provides metalinguistic explanations of grammar errors, which help learners resolve errors. Grammarly is also convenient to use to the extent that users need only copy or paste their text into the input box or upload it in order to receive feedback. Lastly, users can choose from three versions of Grammarly available on both mobile and PC platforms: free, premium, or business version [14]. Although the two paid versions have advanced features, Grammarly free version provide immediate corrective feedback on grammar, spelling and punctuation at no cost once the users access it. Figure 1 shows how Grammarly free version provides corrective feedback. The errors in the text are underlined in red, and replacement forms and metalinguistic explanations are provided to the right.

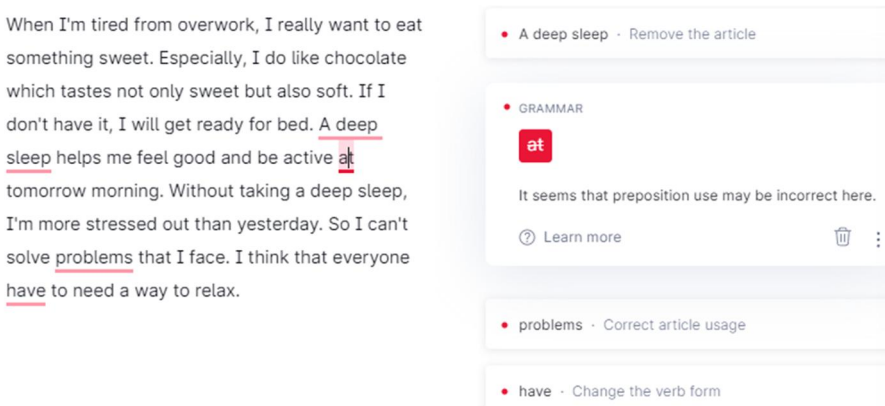

Figure 1. Screenshot of Grammarly

3.2 Data Collection

To evaluate the accuracy of corrective feedback generated by Grammarly, essays were collected from 29 cyber university students enrolled in a "graduation project" course. The participants were all enrolled in the English department and in their final year at a cyber university in Seoul; their English proficiency ranged from intermediate to high-intermediate although there was some variation among individuals. Each participant wrote a 300-400-word narrative essay on the topic of their choice regarding their personal experiences [15]. The primary concern of this study is to investigate how accurate the feedback of Grammarly is, not how many grammatical errors the participants make. Therefore, the participants, as usually done when doing an authentic take-home writing assignment, were allowed to use dictionaries, grammar books, and other references while composing the essays. However, it was prohibited to use any kind of grammar checking software to avoid the potential influence of corrective feedback from CGs.

3.3 Data Analysis Procedure

To classify grammatical error types, this study uses the categories adapted from [13]. As shown in Figure 2, grammatical errors are classified into six main types: nouns, articles, verbs, word forms, sentence structures, and prepositions. Noun, verb, and preposition errors have subcategories. The subcategories of verb errors include tense-aspect, tense shift, subject-verb agreement, and verb forms and those of noun errors are singular/plural and pronoun and incorrect noun selection. Preposition errors are subcategorized into wrong, missing, and unnecessary use of prepositions.

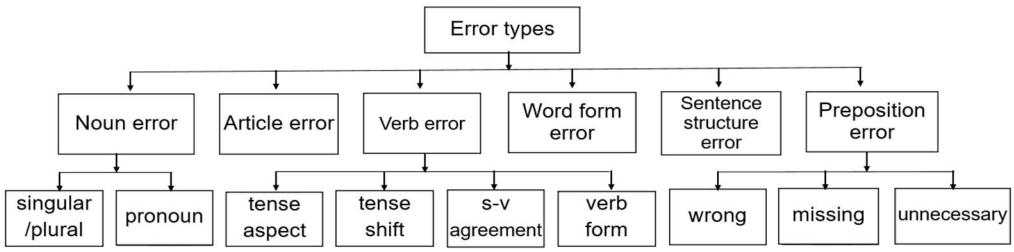

Figure 2. Scheme for Classification of Grammatical Errors

Once all the essays were submitted, two human raters, the researcher and a native English-speaking teacher with 20 years of teaching experience, respectively identified and coded grammatical errors in the essays. We independently analyzed the first three essays and then compared our results to calibrate our classification of errors. The rest of the essays were divided up between the raters, 13 essays respectively, and analyzed individually. For ambiguous errors that were difficult to classify, only those that the two of us agreed upon were included in the analysis. To focus solely on grammatical errors, errors related to word choice or mechanics such as spelling and punctuation were excluded from the analysis. In addition, we excluded cases in which the nature of an error could not be clearly defined due to the unclear meaning of the sentence. Finally, to ensure consistency of classification, the researcher reviewed all identified and classified errors and made corrections where necessary.

Upon completion of the error classification process, the students' essays were uploaded onto Grammarly. To calculate the accuracy of feedback from Grammarly, the frequency of detected errors, suggested replacement forms, and false alarms for each grammatical category was counted first and the percentage of detected errors per each category was calculated.

4. RESULTS

A total of 497 errors were identified by human raters from the 29 essays with the 23,108 words. Table 1 presents the number of errors found in each category along with typical examples. Of the given error categories, article errors (126) had the highest frequency, followed by preposition-related errors (112) and verb-related errors (109). In noun (50), word form (49), and sentence structure (51), a relatively fewer number of errors were detected compared with in article, preposition, and verb errors.

Table 1. Grammatical Errors Found in the Essays

Error Category	Subcategory	Error Examples	Correct Forms	# of Errors
Verb	tense-aspect	Last year I have visited Germany.	**visited**	23
	tense shift	I told my teacher that I can't do it.	**couldn'**t	37
	s-V agreement	She make my study habit great.	**makes**	28
	verb form	I recommend to eat this type of food.	**eating**	21
Subtotal				**109**
Noun	singular/plural	When I was a middle school students	**students**	31
	pronoun	… my parents. My parents gave me courage …	**They**	19
Subtotal				**50**
Preposition	Wrong	The main reason of my stress…	**for**	63
	missing	compare their spouse _ others	**with**	32
	unnecessary	We spent too much time on chatting…	**__**	17
Subtotal		.		**112**
Article		I met the native professor.	**a**	126
Word form		When I am under stressed	**stress**	49
Sentence structure		My girlfriend and I go shopping,___ take a trip to somewhere.	**and**	51
Total				**497**

Grammarly detects grammatical errors in the uploaded essays, and Table 2 presents the number and percentage of errors detected by Grammarly. Grammarly detects a total of 324 (65.2 %) out of 497 errors identified by the human raters, which is more than twice the rate of 29.6% obtained in [12]. Additionally, Grammarly is found to perform better in some categories than others—it is particularly strong on article (77.8%) and preposition (75.0%) errors, and performed well on noun (62.0%), verb (60.6%), and word form errors (57.1%). However, Grammarly shows a considerably low detection rate (35.3%) in dealing with sentence structure errors when compared with other categories. This finding confirms the result of [12] that Grammarly has a very low coverage rate (5%) on sentence structure errors. As for sentence structure errors, the Grammarly free version shows only the locations of the errors by underlining them without suggesting any replacement forms, commenting that this type of error falls under "advanced issues" for which more specific feedback is available only for Premium users. Since this study, unlike [12], treats the underlined feedback as successful identification of structure shift errors, the detection rate of sentence structure errors is about 7 times higher than that of in [12]. If the underlined feedback were treated as a failure to detect errors, the rate would drop to approximately 10%.

Table 2. The Number and Percentage of Errors Detected by Grammarly

Error Category	Subcategory	Committed errors	Detected errors
Verb	tense-aspect	23	13 (56.5%)
	tense shift	37	18 (48.6%)
	sv agreement	28	21 (67.8%)
	verb form	21	14 (66.7%)
Subtotal		**109**	**66 (60.6%)**
Noun	singular/plural	31	21 (67.7%)
	pronoun	19	10 (52.6%)
Subtotal		**50**	**31 (62.0%)**
Preposition	wrong	63	51 (80.9%)
	missing	32	21 (65.6%)
	unnecessary	17	12 (70.6%)
Subtotal		**112**	**84 (75.0%)**
Article		126	98 (77.8%)
Word form		49	28 (57.1%)
Sentence structure		51	17 (35.3%)
Total		**497**	**324(65.2%)**

In respect to the subcategories, Grammarly shows a detection rate of over 50% in all categories except tense shift errors (48.6%). Although the detection rate of tense shift errors in this study is relatively low, it is incomparably higher than the rate (0%) presented in [13]. This huge discrepancy in the detection rate seems to have occurred due to the different error classification criteria of the two studies in terms of tense shift errors. This study classifies inappropriate tense shifts as errors that occur within the sentence boundary, but [13] classifies as errors those that occur at a discourse level beyond the sentence boundary. Moreover, Grammarly is not designed to detect errors beyond the sentence boundary, so such a low detection rate of tense shift errors was obtained in [13].

On the other hand, Grammarly is found to provide very few inaccurate replacement forms and false alarms. As shown in Table 3, Grammarly provides total 6 inaccurate replacement forms on the identified errors and 3 false alarms. Inaccurate replacement forms are suggested on verb, word form, and sentence structure errors. For example, Grammarly inaccurately suggested “under-stressed” rather than “under stress” as a replacement for “under stressed” in “When I am under stressed, I clean my house.” Although the suggested replacement form is not ungrammatical, it changes the intended meaning in the essay. As for false alarms, Grammarly inaccurately marks "TV” as an error in “when I come back home, I turn on TV.” although both "TV” and "the TV” are grammatically right.

Table 3. Inaccurate Replacement Forms and False Alarm

	Verb	Noun	Preposition	Article	Word form	Sentence structure	Total
Inaccurate replacement Forms	1	0	0	0	4	1	6
False alarms	0	0	0	1	1	1	3

According to the results above, on the detected errors, Grammarly mostly provide accurate feedback and appropriate replacement forms, very rarely making false alarms. Grammarly is also found to be strong at catching errors related to articles and prepositions. These results suggest that Grammarly can be used as a useful instructional tool to complement teacher feedback in writing classes. However, it seems to be still too early for Grammarly to completely replace teacher feedback because there is a non-negligible possibility (approximately 35%) of failing to detect errors and a possibility, although very low, of making either false alarms or inaccurate replacement suggestions. Moreover, Grammarly reveals weaknesses in catching errors in certain grammatical categories such as tense shift and sentence structure. These findings suggest that the accuracy of Grammarly needs to be further improved to offer more reliable support to its users.

These mixed findings imply that teachers' active involvement is important for successful integration of Grammarly in English writing classes. Teachers should make judicious decisions on how and when to use Grammarly, being fully informed of both its strengths and limitations. For example, when dealing with prepositions and articles on which Grammarly performs well, learners may be given more autonomy, and with sentence structure and tense shift at which Grammarly is weak, teachers need to increase teacher control. On the other hand, according to some research, teacher feedback can be inconsistent, inaccurate, and/or delayed whereas the feedback generated by Grammarly is highly consistent and immediate. Combined feedback from both the teacher and Grammarly can not only provide learners with more reliable and timely feedback, but also help teachers save time and effort in giving feedback.

5. CONCLUSION

This study evaluates the accuracy of corrective feedback provided by Grammarly free version on essays written by cyber university learners in terms of the rate of detected errors, suggested replacement forms, and false alarms. The results indicate that Grammarly has a high overall error detection rate of over 65%, being particularly strong at catching errors related to articles and prepositions. In addition, on the detected errors, Grammarly mostly provide accurate feedback and appropriate replacement forms, very rarely making false alarms. These findings suggest that Grammarly has high potential as a useful educational tool to complement the drawbacks of teacher feedback and to help learners improve grammatical accuracy in their written work.

However, it is still premature to conclude that Grammarly can completely replace teacher feedback because there still remains the possibility of (approximately 35%) of failing to detect errors; the limitations in detecting errors in certain categories; and the possibility, although very low, of providing inaccurate feedback. Since the feedback provided by Grammarly is not 100% reliable, caution should be taken for successful integration of Grammarly in English writing classes. Teachers should make judicious decisions on when and how to use Grammarly, based on a keen awareness of Grammarly's strengths and limitations. This way Grammarly can be used to effectively meet learners' needs for grammatical accuracy.

Finally, this study has limitations in several respects. First, the data analyzed in this study was collected from a limited number of intermediate level learners situated in the specific context of a cyber university. To

obtain more generalizable, further research needs to be conducted on the texts written by learners with a wide range of proficiency levels in various educational contexts. Second, this study evaluates the accuracy of Grammarly free version only. Given that Grammarly Premium offers far more advanced features, a comparative study on the functions of the Grammarly free version and Grammarly Premium seems to contribute to the more effective use of Grammarly in writing education.

References

[1] I. Lee, "Revisiting teacher feedback in EFL writing from sociocultural perspectives," TESOL Quarterly, Vol. 48, no. 1 pp. 201–213, 2014. DOI: https://doi.org/10.1002/tesq.153
[2] J. Woodworth and K. Barkaoui, "Perspectives on Using Automated Writing Evaluation Systems to Provide Written Corrective Feedback in the ESL Classroom," TESL Canada Journal, Vol. 37, No. 2, pp. 234-247, 2020. DOI: https://doi.org/10.18806/tesl.v37i2.1340
[3] J. Zhang, H. Ozer, and R. Bayazeed, "Grammarly vs. Face-to-face Tutoring at the Writing Center: ESL Student Writers' Perceptions," Praxis: A Writing Center Journal, Vol. 17, No. 20, pp. 33-47, 2020. DOI: https://doi.org/10.26153/tsw/8523
[4] M. Nova, "Utilizing Grammarly in evaluating academic writing: A narrative research on EFL students' experience," Premise: Journal of English Education, Vol. 7, No. 1, pp. 80-97, 2018. DOI: https://doi.org/10.24127/pj.v7i1.1300
[5] R. De Felice and S. G. Pulman, "A classifier-based approach to preposition and determiner error correction in L2 English," in Proc. the 22nd International Conference on Computational Linguistics, Aug. 18 – 22, pp.169–176, 2008. DOI: https://doi.org/10.3115/1599081.1599103
[6] X. Yi, J. Gao, and W. B. Dolan, (2008). "A web-based English proofing system for English as a second language users," in Proc. the Third International Joint Conference on Natural Language Processing, Jan. 8-10, pp. 619–624, 2008.
[7] H.-J. H. Chen, "Evaluating two web-based grammar checkers Microsoft ESL Assistant and NTNU statistical grammar checker," Computational Linguistics and Chinese Language Processing, Vol. 14, No.2, pp. 161–180, 2009.
[8] R. O'Neill and A. Russell, "Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly", Australasian Journal of Educational Technology, Vol. 35, No. 1, pp. 42–56, 2019. DOI: https://doi.org/10.14742/ajet.3795
[9] L. Karyuatry, D. Muhammad, R. Arif, and A. D. Nisrin, "Grammarly as a tool to improve students' writing quality: Free online proofreader across the boundaries," Edulitics Journal, Vol. 3, No. 1, pp. 36-42, 2018. DOI: https://doi.org/10.30595/jssh.v2i1.2297
[10] S. Koltovskaia, "Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study," Assessing Writing, Vol. 44, No. 2, pp. 1-12, 2020. DOI: https://doi.org/10.1016/j.jslw.2021.100816
[11] R. O'Neill and A. M. Russell, "Grammarly: Help or hindrance? Academic learning advisors' perceptions of an online grammar checker." Journal of Academic Language and Learning, Vol. 13, No. 1, pp. 88-107, 2019. DOI: https://journal.aall.org.au/index.php/jall/article/view/591
[12] S. Sahu, Y. K. Vishwakarma, J. Kori, and J. S. Thakur, "Evaluating performance of different grammar checking tools. International Journal," Vol. 9, No. 2, 2020. DOI: https://10.30534/ijatcse/2020/201922020
[13] P. John and N. Woll, "Using Grammar Checkers in an ESL Context: An Investigation of Automatic Corrective Feedback," Calico Journal, Vol. 37, No. 2, pp. 169-192, 2020. DOI: https://doi.org/10.1558/cj.36523
[14] J. S. Barrot, "Integrating Technology into ESL/EFL Writing through Grammarly," RELC Journal, 0033688220966632, 2020. DOI: https://doi.org/10.1177/0033688220966632
[15] Dosik, Moon, "Learner-Generated Digital Listening Materials Using Text-to-Speech for Self-Directed Listening Practice," International Journal of Internet, Broadcasting and Communication, Vol.12, No.4, pp. 148-155, 2020. DOI: https://doi.org/10.7236/IJIBC.2020.12.4.148