Errors in Errors:
An Exploration of Grammarly's Corrective Feedback

Joshua Kloppers, National Taiwan Normal University, Taiwan*

https://orcid.org/0000-0001-5608-0933

ABSTRACT

Automated writing evaluation (AWE) software is an increasingly popular tool for English second language learners. However, research on the accuracy of such software has been both scarce and largely limited in its scope. As such, this article broadens the field of research on AWE accuracy by using a mixed design to holistically evaluate the accuracy of the corrective feedback of the leading AWE program Grammarly. 1136 Grammarly-identified errors related to style, lexis, and form were graded and discussed by two native English speakers. An overall accuracy rate of 78.86% and an accuracy rate of 91.60% when excluding style-related errors were found. However, several issues relating to the promotion of a set writing style, variance in feedback quality, and accuracy of style-related corrective feedback were also identified.

KEYWORDS

Automated Writing Evaluation (AWE), Computer Assisted Language Learning (CALL), Second Language Acquisition (SLA)

INTRODUCTION

In 2007, Nadasdi and Sinclair claimed that "a pedagogically-oriented grammar checker… is at least as good as human correction for helping students improve their L2 [second language] writing skills" (p. 12). Though the enthusiasm of this claim definitely does not represent the attitudes of all educators or researchers, there is undoubtedly interest in the use of automatic writing evaluation (AWE) software and how it can be used in the L2 learning process by educators, researchers, and students. This interest can be seen in the variety of published works that have recommended using AWE applications in the L2 classroom, either by themselves or in conjunction with an educator (Bailey & Lee, 2020; Hockly, 2018; Link et al., 2022; Nova, 2018; Wang et al., 2013). Outside of the classroom, this interest can be seen in the success of commercial AWE applications such as *Grammarly*, which has reported over 30 million daily users (Lytvyn, 2022).

This interest in AWE software has resulted in an increasing amount of research aimed at exploring L2 teachers' and learners' perceptions of AWE software as well as its effect on L2 acquisition (Link

*Corresponding Author

et al., 2022). However, despite this increase, there is still a dearth of research on the accuracy of AWE applications (Link et al., 2022; Xi, 2010). Considering that accuracy and consistency are deemed important factors affecting the value of written corrective feedback—here understood as any "response to erroneous output with the aim of improving accuracy" (Karim & Nassaji, 2019, p. 29)—this is a significant gap in the current research (El Tatawy, 2002).

In terms of *Grammarly*, the recent isolated studies that have sought to evaluate its accuracy have largely focused only on specific pieces of language such as lexical (Dodigovic et al., 2016) or form errors (Dodigovic & Tovmasyan, 2021) and have not looked at the accuracy of the software holistically or in regards to other types of errors, such as those related to style. Furthermore, these studies (Dembsey, 2017; Dodigovic et al., 2016; Dodigovic & Tovmasyan, 2021) have evaluated *Grammarly*'s accuracy by comparing it to teacher-provided corrections. However, *Grammarly* is widely used outside of the classroom—"outside of the classroom" is here used to simply refer to situations where instructor feedback is unavailable, even if the work being done is academic in nature—and so needs to also be evaluated as a tool that does not replace instructor feedback but rather functions separately or alongside it (Grammarly Inc., 2022a). To do this, the information that *Grammarly* provides to the learners (i.e., the errors that it identifies and corrects) needs to be evaluated in detail by itself.

Another issue that AWE research faces is the speed at which the software and applications that it aims to research change (Dodigovic & Tovmasyan, 2021). *Grammarly*, for example, averages over three updates a month and recently released a large-scale update that it named *Grammarly* 2.0 (Grammarly Inc., 2022b). It is unclear how generalizable the results of research based on earlier forms of AWE applications are to current AWE applications.

This paper aims to partly address these issues, at least for a time, by thoroughly evaluating the accuracy of *Grammarly*'s identified form, lexical, and style errors as well as its recommended corrections for said errors. It will do so by having two separate native English speakers rate and then comment on 1136 *Grammarly*-identified and corrected errors in forty separate texts produced by university-level L2 English speakers.

LITERATURE REVIEW

Automated Writing Evaluation (AWE) software utilises the work in Natural Language Processing, a subfield of linguistics and computer science interested in how computer programs can process and analyse large amounts of language data, and specifically Latent Semantic Analysis, a further specialisation which works by analysing the relationship between two documents or groups of documents, to assess a text and provide a score, which can be supported by either specific or general corrective feedback (Hockly, 2018). What this means is that AWE engines use algorithms to identify morphological, syntactical, semantic, and pragmatic rules and patterns in research corpora (Khurana et al., 2023). These rules and patterns are utilized to compare a text that is input by the AWE user to the research corpora (Grammarly Inc., 2023; Khurana et al., 2023). This comparison allows for the identification of parts of the input text that diverge from the identified rules or differ from the expected patterns (Grammarly Inc., 2023; Khurana et al., 2023). These parts are then flagged as errors and, where possible, corrections based on the research corpora are generated and provided (Grammarly Inc., 2023). This process involves both Natural Language Understanding, in the identification of the divergent pieces of text, and Natural Language Generation, in the generation of corrections (Grammarly Inc., 2023; Khurana et al., 2023). *Grammarly* specifically also uses machine learning to adjust its corrections based on user interactions with provided corrective feedback (Grammarly Inc., 2023).

The subject of this paper, *Grammarly*, is the "world's leading automated proofreader" with over 30 million daily users (Grammarly, 2014; Lytvyn, 2022). It is both a plagiarism and grammar checker that can be accessed through an application, online, or via an extension as either a commercial product with unlimited features or a free product with limited features (Fitria, 2021). It identifies

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/errors-in-errors/325792

Related Content

The Effect of Serious Mobile Games on Student English Vocabulary Acquisition and Attitude Toward English
Nuri Kara (2022). *International Journal of Computer-Assisted Language Learning and Teaching (pp. 1-16).*
www.irma-international.org/article/the-effect-of-serious-mobile-games-on-student-english-vocabulary-acquisition-and-attitude-toward-english/297203

Towards a More Naturalistic CALL: Video Gaming and Language Learning
Phil Bensonand Alice Chik (2011). *International Journal of Computer-Assisted Language Learning and Teaching (pp. 1-13).*
www.irma-international.org/article/towards-more-naturalistic-call/56324

Researching Mobile-Assisted English Language Learning Among Adult Distance Learners in China: Emerging Practices and Learner Perception of Teacher Role
Qian Kanand Jinlan Tang (2018). *International Journal of Computer-Assisted Language Learning and Teaching (pp. 1-28).*
www.irma-international.org/article/researching-mobile-assisted-english-language-learning-among-adult-distance-learners-in-china/215414

The Use of the CMC Tool AMANDA in the Teaching of English
Esrom Adriano Iralaand Patrica Lupion Torres (2009). *Handbook of Research on E-Learning Methodologies for Language Acquisition (pp. 307-321).*
www.irma-international.org/chapter/use-cmc-tool-amanda-teaching/20036

Materials Design and Pedagogy for Technology-Enhanced Language Learning
Hazel L.W. Chiu (2015). *International Journal of Computer-Assisted Language Learning and Teaching (pp. 22-34).*
www.irma-international.org/article/materials-design-and-pedagogy-for-technology-enhanced-language-learning/128254