Bulletin of Aichi Univ. of Education, 72（Educational Sciences）, pp. 81 − 87, March, 2023

# Automated Editing of EFL Student Writing: A Comparison of Microsoft Word and Google Docs

James VENEMA

*Department of Foreign Languages, Aichi University of Education, Kariya 448-8542, Japan*

## Abstract

This paper uses a sample of EFL student writing to evaluate the effectiveness of the imbedded grammar checkers in Microsoft Word and Google Docs. A sample of EFL writing of first-year university students of just over 12,000 words was analyzed using the imbedded grammar checkers of the two most common digital writing platforms. The analysis found that Google Docs both found more errors and more complex errors while also being slightly better at suggesting corrections that resulted in grammatically correct sentences and clauses. Google Docs would appear be a legitimate option to Microsoft Word for students and educators interested in taking advantage of free grammar checkers.

Keywords: *grammar checkers, Microsoft Word, Google Docs, error correction*

## Introduction

Most student writing today will be done digitally, typically taking advantage of the ubiquitous grammar checkers bundled into common writing software such as Microsoft Word and Google Docs. As the default application in Windows computers, Microsoft Word is undisputedly the market leader in writing software. However, Google Docs has made major inroads, particularly with the growing popularity of Chromebooks in education where Google Docs is the default writing application. Both writing platforms come bundled with their own grammar checkers that students and writers, including the author of this paper, regularly rely on as writing aids. In response to classes being moved online in 2020, the author of this paper shifted from Microsoft Word to Google Drive, and Google Docs, for submitting writing homework in a university EFL class focusing on debate. Over one semester in 2020, students worked in teams to write a series of speeches on six different issues for a total sample of just over 12,000 words. As a teacher transitioning away from Microsoft Word to collect student writing, my first impression of Google Docs was that it was better at identifying and correctly suggesting corrections for a greater number of student errors, including more complex errors. This study was conducted to test that impression.

## Grammar Checkers

Dale (2016) writes that grammar checkers have gone through three paradigms. The first phase, in the late 1970s and 1980s relied on a relatively simple "pattern matching and string replacement" (p. 492) such as replacing *isnt* with *isn't*. Second generation technology relied on larger "rule-based descriptions of permissible syntax" (p. 492) to find ungrammatical elements and suggest corrections. The third generation of grammar checking, including the context-aware checker bundled into Google Docs (Whiteclaw et al, 2009), incorporated statistical models of language use. These systems use statistical analysis of large internet-based corpora to look for potential deviations from common, ostensibly correct, uses to offer suggestions more closely matching corpora examples.

Google docs first became available in 2009 and quickly became the most popular free writing software. In February of 2019 a grammar check was added, but there has been very little published research on its efficacy. Wolber (2019), using just six fabricated sentences with seven errors to compare different grammar checkers, wrote that the Google Docs grammar check only identified two errors while options such as Grammarly and Sapling (<*https:/sapling.ai*>) identified six. Wolber did a follow up study in 2021 (Wolber, 2021) and noted some significant improvements, with Google Docs correctly identifying five of the seven errors. Still, it is hard to draw any conclusions from such a small sample of deliberately manufactured errors.

Despite its ubiquity, there is surprisingly little research into the effectiveness of the grammar checker in Microsoft Word. McGee and Ericsson (2002) were critical of imbedded grammar checker in Word.

> Although the programming might be complicated and the interface slick, users get recycled, often confusing advice about grammar and mixed bag of style suggestions that don't take into account current thinking on the grammar itself, good rhetorical theory, or pedagogical implications. (p. 459)

Twenty years is, of course, a long time in the technology industry, and it would probably be safe to assume improvements, particularly for a corporate giant such as Microsoft with abundant financial resources. However, as ubiquitous as the imbedded grammar checker in Word might be, the author could find very little research published on its relative effectiveness in analysing EFL student writing. Specifically, the author has found no published research on the relative effectiveness of the imbedded grammar checkers of the two most common digital writing platforms, Microsoft Word and Google Docs, incorporating real samples of EFL writing. Effectiveness, of course, is a potentially ambiguous term that requires clarification. This article will focus on three relatively objective questions.

1. Which grammar checker correctly identifies more errors in student writing?
2. Which grammar checker is better able to identify more complex errors?
3. Which grammar checker is better able to provide suggestions that result in correct samples of writing?

## The Language Sample

The 18 first-year students in this course worked in teams to research, write, and present debate speeches on the issues below.

1. Immigration to Japan
2. Part-time jobs for high school students
3. Living alone as a university student
4. Elementary school children being left alone at home
5. Assisted suicide
6. Online vs. face-to-face classes

The simplified form of debate introduced in the course included three speeches. In the 'opening speech' students gave structured reasons for their position, from either the affirmative or negative position. In the 'refutation speech', students attacked the opposing team's arguments given in the opening speech. Finally, in the 'rebuttal speech' students reviewed the arguments and refutations given, outlining why they think they won each point. The debate speeches were submitted as written homework in advance of class and were proofread by the teacher to enhance coherency. Only one of the three speeches was presented in a single class to allow for this process. In this sense, the course was more a writing class where students read from prepared speeches than a free-flowing oral debate class.

The combined sample of student writing came to over 12,000 words. No objective English test results are available for the students, although they were willing and capable of taking on relatively complex writing and speaking tasks on challenging topics. With those capabilities, students' levels could be categorized as B1 or B2 in the CEFR levels (Common European Framework of Reference for Languages). Students were responsible for working in teams to develop arguments but were individually responsible for researching and writing points of argument. For this reason, the writing samples are evenly distributed among the students.

## Method

The speeches were submitted as Google Docs files. Despite being verbally encouraged by the author to use the imbedded grammar check, the submitted assignments included numerous highlighted errors identified by Google Docs. For purpose of this research, the version history option was used to go back to the originally submitted file, prior to any teacher edits. This file was copied and pasted into Microsoft Word using the 'unformatted text' option of 'paste special' to avoid formatting issues that affected the Word grammar and spell check. To activate the Microsoft Word grammar checker, all the text needed to be highlighted and a language option selected (American English). All the work was done on an Apple computer, with the most current version of Microsoft Word through an Office 365 subscription (2021).

While the author found no way to edit the grammar check preferences of Google Docs, Microsoft Word included the

option to adjust the grammar check settings, with almost 150 options to choose from. (To find these setting click the following: Word-Preferences-Spelling & Grammar-Settings). For the purposes of this study the 'reset all' button was used to revert to the default settings. While this left many boxes unchecked (such as passive voice, cultural bias, vague adjectives) clicking all the boxes and rerunning the error check resulted in no further items identified for possible errors in the one document that this was tried on.

In both Google Docs and Microsoft Word all items identified as possible errors were counted and categorized. The categories arrived at are not intended to be a complete categorization of all the error types that the grammar checkers might identify. Instead, they were the minimum number of categories that the author felt could explain the error sample. Seventeen categories were identified based on the errors highlighted as well as the suggested corrections. The error categories, along with examples, can be found in Table 1 below.

Table 1.

*Error Categories, Examples, and Correction*

| Error Categories | Example Error | Example Correction |
|---|---|---|
| article | 30% of population | 30% of the population |
| brevity | is not a must thing for students | is not a must for students |
| capitalization | Firstly, It | Firstly, it |
| conjugation | watching a screen for long hours cause | watching a screen for long hours causes |
| conjunction | They insisted if Japan | They insisted that if Japan |
| hyphenation | nonpayment of tax problem | non-payment of tax problem |
| name | Keizo Yamawaki | Keizo Yamasaki |
| preposition | 70% students | 70% of students |
| punctuation | costs a lot and it | costs a lot, and it |
| sentence construction | it is like putting the cart before the horse that high school students | Second is that it is like putting the cart before the horse. High school students… |
| spacing | social demand . | social demand. |
| spelling | Besedes, | Besides, |
| verb form | even now it is existed | even now it exists |
| word choice | without telling about it to her father | without talking about it to her father |
| word family | hiding something danger | hiding something dangerous |
| word order | you can imagine easily | you can easily imagine |

Finally, a judgement was made on whether the suggestions resolved the error and resulted in a correct sentence. This involves not only correctly identifying and resolving a given error but also correctly identifying other errors in the same clause or sentence. The analysis also required a decision on whether the grammar check resulted in a grammatically correct clause or sentence. In many cases it was quite clear whether the error was correctly identified and resolved as in the following examples.

*It cannot say that - It cannot be said that* (Google Docs, verb form, error resolved)

*the all success countries – all success countries* (Google Docs, article, error unresolved)

*Therefore we – Therefore, we* (Microsoft Word, punctuation, error resolved)

*had not capacity to admit immigrations.This – has not capacity to admit immigrations. This* (Microsoft Word, spacing, error unresolved)

Of course, the argument can be made that the grammar checkers were successful in identifying and correcting the article and spacing errors above. However, correcting one error while leaving another unresolved in the very same clause could result in a false level of confidence regarding a sample of writing that is still objectively incorrect or problematic.

As is often the case in editing writing, particularly writing in a second language, error identification and correction sometimes requires speculation regarding intended meanings.

*is high level in the world – is the highest in the world* (Google Docs, article & word choice, error unresolved)

In the above example, it was assumed that the student's intended meaning would have been best expressed as *is one of the highest levels in the world* and the error was seen as unresolved since the correction resulted in an unintended meaning albeit a grammatically correct phrase. Admittedly, judgements on whether an error was resolved or not could at times be subjective. For example, despite the awkward phrasing of 'recent 9 years' the following error was seen as resolved.

> *the population is decreasing in recent 9 years – the population has been decreasing in recent 9 years* (Google Docs, verb form)

An absolute claim of objectivity is clearly impossible. A different researcher and writing teacher would no doubt arrive at some different decisions on whether a clause or sentence has been sufficiently corrected. Still, this author assumes enough of a consensus to make some conclusions regarding the effectiveness of the grammar checkers.

## Results

There was a striking difference in the number of errors identified. Google Docs identified a total of 248 errors while Microsoft Word identified 130. The categories and number of occurrences for each category can be seen in figures 1 and 2 below.

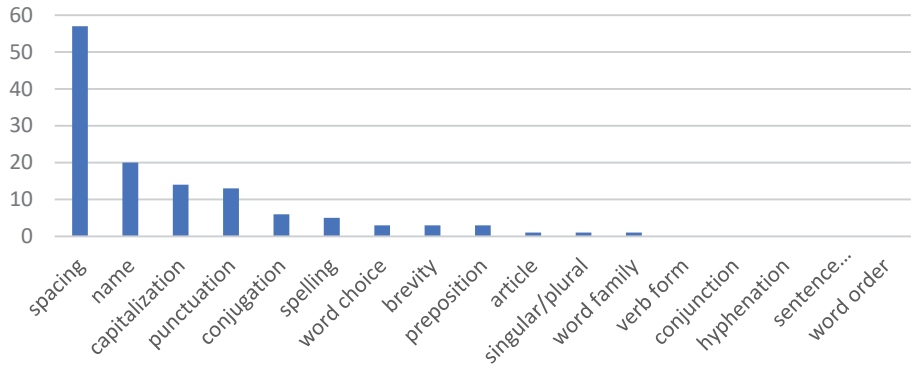


Figure 1.
*Error Categories Identified in Microsoft Word*

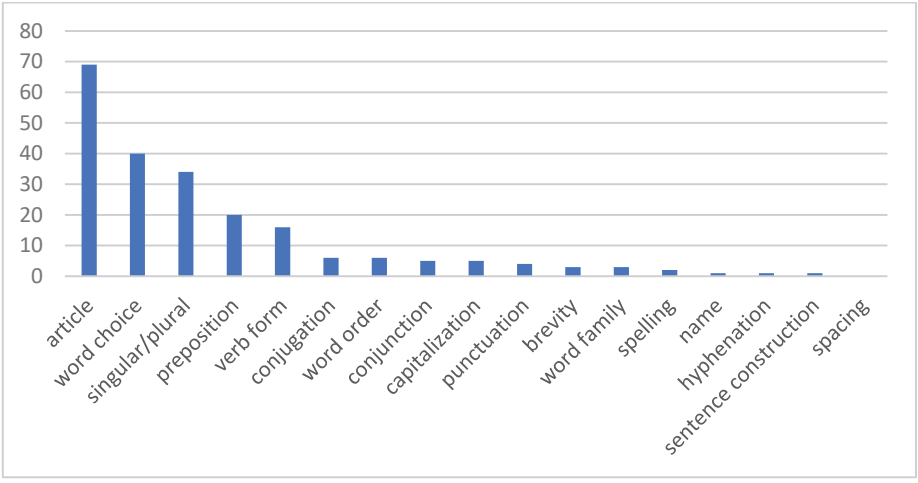


Figure 2.
*Error Categories identified in Google Docs*

The difference is even more pronounced upon looking at the categories. Google Docs failed to identify any 'spacing' problems. The resulting errors in student writing when submitting homework in Google Docs were subsequently identified by Microsoft Word and counted for the largest category of errors at 57. The second highest error category identified in Microsoft Word was 'name', mostly a result of the failure of Microsoft Word to identify Japanese names. In fact, only two out of the 20 'name' errors identified by Microsoft Word resulted in a resolved error. This category was a primary reason why the resolved error rate for Microsoft Word was at 71% while Google Docs was at 82%. (Eliminating this category would bring the resolved error rate to 79% for Microsoft Word.)

Beyond the number of errors identified, there were differences in the complexity of errors identified. The top four error categories identified by Microsoft Word, accounting for 104 of the 130 errors identified, were 'spacing', 'name', 'capitalization', and 'punctuation'. The top four categories for Google Docs were 'article', 'word choice', 'singular/plural', and 'preposition'. Word choice and preposition are typically more complex errors involving examples such as the following.

*when you school student – when you were a school student*
*a service connects the sitter the family – a service connects the sitter to the family*

Articles might seem to be an easier error to target and correct. However, this overlooks the various, often idiosyncratic, usages in English that defy concise explanations. For even advanced EFL students, articles can remain a problematic area. I would expect many EFL students to struggle with the following two corrections.

*As I said in opening speech - As I said in the opening speech (Google Docs)*
*save the money. – save money. (Google Docs)*

For Google Docs, the next most common category (16 errors identified) was 'verb form'. The relatively broad terms such as 'word choice' and 'verb form' were often a result of an attempt to describe the errors identified in the most reasonably small number of categories. For example, under the category word choice we have the following error types:

*policy has changed by the situation - policy has changed due to the situation*
*in their high school ages – in their high school years*
*learn the way to save – learn how to save*

The following errors and suggested corrections were grouped under verb form.

*even now it is existed – even now it exists*
*get used to do house-work – get used to doing housework*
*the negative team miss some points - the negative team missed some points*

In general, the key difference in the top errors lies in the scope at which the error is identified. The Microsoft Word grammar checker was more often identifying errors limited to a single word, letter or space within a clause or sentence. Google Docs was identifying errors that resulted as a function of the relationship between words in a clause or sentence. Google Docs corrections also included many more sentences where multiple errors, in different error categories, were identified.
The total number of multiple category errors identified by Google Docs was 31, of which 21 were resolved. Some examples were as follows.

*increasing the wage rate lead increasing - increasing the wage rate leads to increasing* (conjugation, preposition)

*it should not be illegal in Japan elementary school age children at home alone - it should not be illegal in Japan for elementary school age children to be at home alone* (preposition, word choice)

Microsoft Word, on the other hand, identified just three (capitalization & name, spacing & name, word choice and word form)

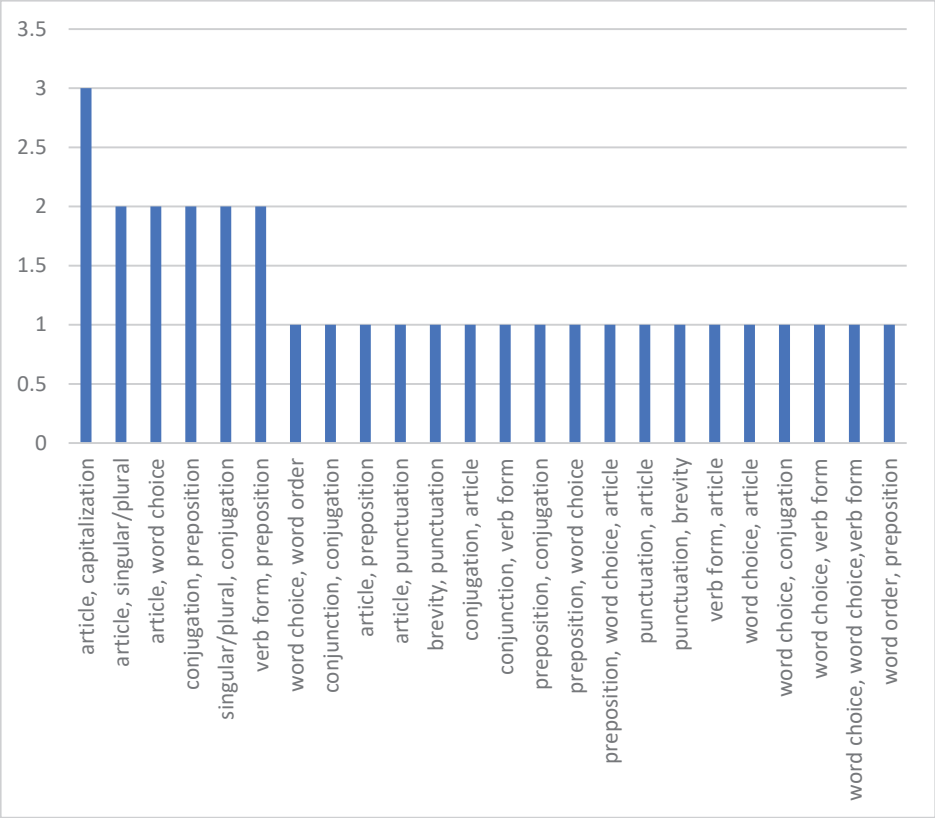


Figure 3.
*Frequency of multiple error corrections in Google Docs*

of which only one (word choice and word form) was resolved, albeit a relatively impressive correction involving reported results of a survey.

*asked 188 doctors how did they feel when - asked 188 doctors how they felt when* (word choice, verb form)

One category, 'sentence construction', (for which only one example was found in Google Docs and none in Microsoft Word) stands out for the complexity of the error identified and in the suggested correction. This correction was unique in that a single sentence was both rephrased and broken into two distinct sentences.

*it is like putting the cart before the horse that high school students - Second is that it is like putting the cart before the horse. High school students…*

However, the rarity of this kind of editing suggestion indicates that automated grammar checkers will not be replacing writing teachers any time soon. Teachers will remain the better source of more holistic corrections and input.

## Discussion

These results provide some evidence of what I had intuitively sensed when using Google Docs to correct student writing: that more complex errors at the clausal and sentence level were being correctly identified and corrected. Not only did Google Docs identify more errors; it also tended to identify more complex errors and offer more complex correction suggestions. At the same time, Google Docs maintained a higher error resolved rate of 84 to 71 percent. (However, a significant part of the difference was due to the incorrect identification of Japanese names as errors by Microsoft Word.) The one error that Google Docs overlooked is somewhat mystifying: spacing. Spacing issues would seem to be easily identifiable, yet Google Docs managed to overlook 57 examples of spacing errors. I do not believe this is an insignificant issue for EFL teachers, as I have

found spacing to be a relatively common problem, albeit one that, like spelling, should be easily identified and corrected with modern grammar checkers.

One possible qualification of the results is that, while students submitted their writing in Google Docs, there is no way to verify what software was used to compose their speeches. While the author was often able to view student writing progress in Google Docs real time, it is possible that some students were composing speeches in Word before copying and pasting the writing into Google Docs. Most students would then have taken advantage of the embedded grammar checker in Word. The numerical discrepancy in the results between the two grammar checkers could then be at least partially explained by differences in the kinds of errors identified. However, the fact that spacing errors, unidentified in Google Docs, were the most common error identified in Microsoft Word suggests that students were not transferring files between writing platforms.

Of course, any student writing submitted digitally has likely been edited to some extent by students using the embedded grammar checkers. It would be interesting to compare the effectiveness of the grammar checkers from an unfiltered sample of student writing, a sample 'untainted' by any digital grammar check. While labor intensive, this could be achieved by inputting samples of student writing done by hand. Admittedly, this would be less effective in identifying spacing errors. In fact, spacing errors in first-year students might be a result of the tendency of Japanese high schools to require all homework to be submitted in handwritten form.

Finally, from an instructional perspective, the fact that 248 errors were identified but not corrected in Google Docs implies the need to provide instruction to students on how to effectively use embedded grammar checkers. While by no means perfect, an 82% success rate in providing error corrections, if constant across other samples of student writing, should result in significantly improved student writing when they are consistently applied. Still, an 18% unresolved error rate means student judgements will remain an important part of the writing process.

## References

Dale, R. (2016). Checking in on grammar checking. *Natural Language Engineering, 22*, 491-495. DOI: *https://doi.org/10.1017/S1351324916000061*

McGee, T. & Ericsson, P. (2002). The politics of the program: MS Word as the invisible grammarian. *Computers and Composition, 19*, 453-470. DOI: *https://doi.org/10.1016/S8755-4615(02)00142-1*

Whitelaw, C., Hutchinson, B., Chung, G. Y., and Ellis, G. (2009). Using the web for language independent spellchecking and autocorrection. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, vol. 2, Stroudsburg, PA: Association for Computational Linguistics*, 890–899. DOI: *https://aclanthology.org/D09-1093.pdf*

Wolber, A. (2019, April 4). How Google Docs grammar check compares to it alternatives. *TechRepublic. https://www.techrepublic.com/article/how-google-docs-grammar-check-compares-to-its-alternatives/*

Wolber, A. (2021, June 10). How well does Google Docs spelling and grammar check work? *TechRepublic. https://www.techrepublic.com/article/how-well-does-google-docs-spelling-and-grammar-check-work/*

(Received September 16, 2022)