



Evaluating performance of different grammar checking tools

Subham Sahu¹, Yogesh Kumar Vishwakarma², Jeevanlal Kori³, Jitendra Singh Thakur⁴

Department of Computer Science and Engineering, Jabalpur Engineering College, India,

¹subhamsahu1120@gmail.com

²ykvishwakarma304@gmail.com

³jeevankori280@gmail.com

⁴jsthakur@jecjabalpur.ac.in

ABSTRACT

Grammar checking is the process of detecting and correcting Grammatical errors present in text. A Lot of online apps are available for grammar checking. How efficient these apps are? To answer this question in this paper we have presented a systematic evaluation of five such most popular apps viz. Grammarly, Ginger, ProWritingAid, LanguageTool and After the Deadline. To evaluate performance of these apps we have created a dataset of 500 sentences with the grammar error types and subtypes. These erroneous sentences were then fed to these apps, and the ability of these apps to identify each type of the grammar errors were recorded. Grammarly has achieved highest overall accuracy of 44.4%. After The Deadline has achieved lowest overall accuracy of 28.74%. None of the apps were able to achieve accuracy more than 11% in case of Sentence Structure errors. Similarly, in Semantic error no app has achieved accuracy more than 25%. The results of the study suggest that there is a lot of scope for the improvements to be done in automated grammar checking apps.

Key words: grammar checking app, erroneous sentence.ss

1. INTRODUCTION

Nowadays every person prefers English as their secondary language, due to its global acceptance. And to get proficiency in English, first we need to go through its syntax and semantics (Grammar). Writing is one of the most challenging area in English learning [10]. A very small mistake like a missing punctuation, misplace punctuation, a missing word changes the meaning of the sentence. So it becomes essential to identify and correct various grammar errors in English writing. A lot of free, open source and proprietary grammar checking tool are available such as SpellCheckPlus [6], Ginger [2], ProWritingAid [7], Reverso [5], After the Deadline [13], LanguageTool [4] etc. Each of these tools claims to give good accuracy, but lacks empirical evaluations. Although, a few efforts have been reported in the literature for the evaluation of some grammar checking apps. The literature lacks a systematic evaluation of the available grammar checking apps.

Caroline Haist [12] evaluated the grammar checker of Microsoft Word 97 in 2000. Hao-Jan Howard Chen [9] compared performance of two web-based grammar checkers (Microsoft ESL Assistant and NTNU statistical grammar checker). He performed the experiment by taking few example sentences. These two grammar checkers are not active at present. In 2016, Michelle Rose Cavaleri and Saib Dianati [8] did the evaluation of Grammarly [3] by analyzing the feedback received from users (language experts or non experts). The feedback was based on the acceptance and use of Grammarly among higher education students. Randy Joy Magno Ventayen and Caren C Orlanda-Ventayen [16], in 2018, used System Usability Scale (SUS)¹ below to measure the usability of the Grammarly based on responses received from graduate students (app users). They reported that Grammarly is useful in improving writing and understanding of grammar rules. In 2018, Muhammad Ali Ghufroon and Fathia Rosyida [11] reported that Grammarly software is more effective in reducing students errors in EFL writing compared to teacher corrective feedback (indirect corrective feedback) to teach EFL writing.

In this paper, we present the evaluation of the five most popular grammar checking apps that we have done with the help of an experiment. A flurry of grammar checking apps are available on the web, but the aim of our study is to evaluate the performance of the popular free and open source grammar checking apps because most of the users prefer to use freely available apps rather than the paid one.

As a dataset annotated with the grammar error types was not available, we created a dataset of 500 English sentences based on the scheme of classification of grammar error. We have made our dataset publicly available on [26]. These erroneous sentences were then used to evaluate the error detection abilities of the five popular grammar checking applications viz. Grammarly [3], Ginger [2], ProWritingAid [7], LanguageTool [4], and After the Deadline [13] (see Table 1). The rest of the paper is organized as follows; section 2 mentions previous efforts made in this area, section 3 discussed how we performed the experiment, section 4 shows

¹ Simple tool to measures usability of any application based upon responses of 10 item questionnaires each having 5 response options.

Table 1: Grammar Checking Apps

Grammar Checking app	Tested on	Version
Grammarly	Web Interface	Free Version
Ginger	API	Full Version
ProWritingAid	Web Interface	Free Version
LanguageTool	Web Interface	Free Version
After the Deadline	Web Interface	Free Version

the results of the experiment and answers these research questions.

1. Which Grammar checking app has outperformed?
2. Which app is specifically good for each type of error?
3. What are the areas for improvement in these grammar apps?

2. RELATED WORK

2.1 Corpus Based Evaluation

Caroline Haist [12] analyzed the performance of Microsoft Word 97 grammar checker on common errors frequently made by college students. He ran thousands of sentences on grammar checker and closely analyzed the output against each error type (fragment, run-on, punctuation, word-usage error etc.). The result showed that MicrosoftWord 97 performed reliably on very few error types, like subject-verb agreement with 72% accuracy, fragment error 60% accuracy. So Word 97 preferably not used by students, until their writing is relatively free of errors.

Hao-Jan Howard Chen [9] evaluated performance of Microsoft ESL Assistant and NTNU statistical grammar checker. He randomly selected five sentences from a large set of incorrect sentences for each error type. He ensured that errors in these sentences are correctly identified by ETS Criteria [1] (base tool for comparison). Then these sentences were given as input to grammar checkers, and output is compared with the output of ETS Criteria. The result showed that Microsoft ESL Assistant has achieved overall precision of 50%, recall 30% and NTNU has achieved precision of 61%, recall 72%.

2.2 Feedback Based Evaluation

Michelle Rose Cavaleri and Saib Dianati [8] analyzed performance of Grammarly grammar checker via conducting a survey among higher education students of two colleges, the Australian College of Applied Psychology (ACAP) and Navitas College of Public Safety (NCPS). The Survey consists of three parts 1) asked students about their qualification and language abilities, 2) students asked to give rating to statements about usability and usefulness, of grammar checker and 3) students asked about the how effectively grammar checker helped on their writing quality, confidence and assignment marks. The aim of this survey is to get feedback on usefulness and ease of use of grammar checker,

which are the two main principals of the Technology Acceptance Model (TAM)². Total of 37 students registered for this survey, out of which 18 has successfully taken part in this survey. In terms of usefulness, 15 out of 18 students strongly agreed. For ease of use, 17 out of 18 students strongly agreed.

Muhammad Ali Ghufron¹ and Fathia Rosyida [11] evaluated performance of Grammarly based on improvements seen in EFL Writings of the students during their semester. In experiment total 40 English Students of a private university in Indonesia were selected using cluster random sampling technique, and divided into two groups of 20-20 students, one is experimental group and another is control group. The experiment is conducted in two phases, pre test and post test. Pre test is conducted to know the background of students in ESL writing without any help of grammar checker or teacher. In post test, experimental group advised to use grammar checker to improve their EFL writing, and control group asked to improve their EFL Writings by getting corrective feedback from teachers. And finally at end of semester they were asked to submit their writings to teacher. Then teacher evaluates their writing on the basis of content, organization, diction, language use (grammar), and mechanics (spelling and punctuation). Reid's [14] modified scoring rubric is used to score students' writings. From results it is concluded that Grammarly is more effective than teacher corrective feedback to teach EFL writing in reducing errors in dictation, grammar, spelling and punctuation.

Hyejin Yang [17] performed study on SpellCheckPlus [6] grammar checker using mixed approach. He used corpus based approach to measure efficiency of grammar checker, and used feedback based approach to measure acceptance and usefulness of grammar checker. Total 16 students involved in this experiment. Two sets of questionnaire prepared, one is pre-questionnaire consists of 11 questions, aim is to gather students personal information and their prior knowledge about grammar checkers before participating in this study. Another is a post questionnaire consists of 12 questions; aim is to know students perspective on easiness, helpfulness, trustworthiness toward grammar checker. And To evaluate efficiency of grammar checker students asked to submit three writing assignments (corpus). Efficiency is measured in terms of the mean of Normalized error rates per 100 words. Mean error rates of assignment 2 and assignment 3 significantly reduced from 6.21 to 4.60 and 6.75 to 4.17 respectively, after using SpellCheckPlus [6].

3. METHODOLOGY

This section presents the methodology used for the evaluation of the grammar checking apps. The following are the steps involved in the evaluation process:

² Offers a Conceptual framework for predicting the acceptability and use of a technology.

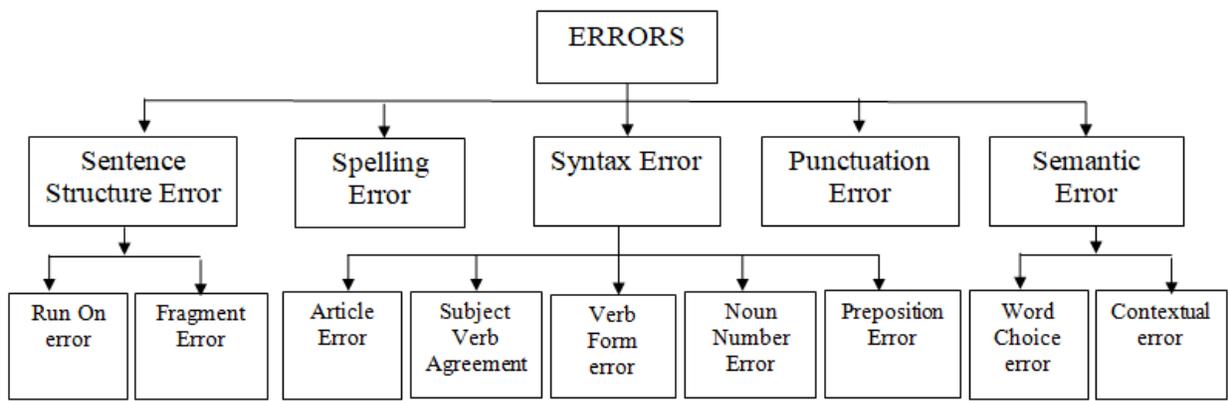


Figure 1 : Scheme for Classification of error

3.1 Selection of the grammar checking tools for evaluation

There are many grammar checking apps available in the web; so, we decided to evaluate the five most popular grammar checking apps. To find the popular grammar checking apps, we searched the web using keywords such as “Top grammar checkers apps”, “Best apps for grammar checking”, “popular grammar checker tools”, “free grammar checker applications”, “top free editors for grammar checking” and many more. From the search results of each keyword, top 10 articles were selected. After reading these articles, rank is assigned to every app based on number of articles claimed that a particular app is useful and popular. From the results of ranking, we have selected the top 5 apps (see Table 1) for evaluation.

3.2 Grammar Error types and metrics used for evaluation

The grammar error types and the classification scheme for the error types used in our study is obtained from a systematic review that we had conducted in our previous work [15]. We have classified the grammar errors into five main types as shown in Figure 1. These errors are: 1) Sentence Structure error, 2) Punctuation error, 3) spelling error, 4) Syntax and 5) Semantic error. Sentence Structure error, Syntax error and Semantic errors are further classified into subtypes. Each error type with examples taken from [18] is described below.

1) Sentence Structure error: Hornsby [19] has formulated 25 English patterns (arrangement of POS components in a sentence), if none of these patterns found in a sentence, then the sentence is said to have Sentence Structure error. The Sentence Structure error is further classified into subtypes a) Fragment error and b) Run-On error. A fragment is an incomplete sentence in which either subject or verb is missing or it may be a sentence having dependent clause without the main clause [18]. A run-on sentence is two independent clauses missing a punctuation or necessary conjunction between them, which affects the readability of text [15]. Example 1 and 2 are correct; Example 3 contains fragment error and Example 4 contains Run-on error.

Example 1: She began singing. (S-V-Gerund)

Example 2: She wants to go. (S-V-to-infinitive)

Example 3: A fair little girl under a tree. (Verb is missing)

Example 4: I ran fast missed the train. (Conjunction ‘but’ is

missing)

2) Punctuation error: Punctuation is used to enhance sentence readability and expressivity. A missing punctuation or unnecessary punctuation changes the meaning of sentence. *Example 5:* He lost lands money reputation and friends. (lands, money, reputation and friends)

Example 6: Alas she is dead ! (Alas ! She is dead.)

Example 7: How are you ? Mohan? (How are you, Mohan?)

3) Spelling error: A meaningless word in a sentence makes the whole sentence meaningless. This is the most common type of error found in English text.

Example 8: Death lays his iccy hand on kings. (icy)

Example 9: Many are called, but few are choosen. (chosen)

4) Syntax error: If any sentence breaks a grammar rule, then it has syntax error. Based on type of grammar rule it breaks, syntax error is further classified into following subtypes:

a) Preposition error: Wrong usage of prepositions in a sentence leads to preposition error. Example 10 missing prepositions ‘on’; Example 11 misused preposition ‘of’.

b) Subject verb agreement: Mismatch between form or number of a verb and subject in a sentence leads to Subject verb agreement error. Example 12 and 13 shows an valid agreement between subject and verb.

c) Article error: Wrong usage of articles (a, an, the) in a sentence leads to article error. Example 14 missing article ‘The’; Example 15 misused article ‘a’.

d) Noun number error: Plural form of uncountable or mass noun is wrong in English grammar. It is wrong because uncountable or mass noun does not have any plural form. Example 16 and 17 contains Noun number error.

e) Verb form error: Wrong usage of verb form in a sentence leads to Verb form error. Example 18 and 19 contains Verb form error.

Example 10: He sat a stool. (He sat on a stool.)

Example 11: He has recovered of his illness. (from his illness)

Example 12: He is not to blame. (Subject-‘he’ and verb-‘is’ both are third person singular)

Example 13: They are not on good terms. (Subject-‘they’ and verb-‘are’ both are third person plural)

Table 2: Sample dataset of four sentences

Correct Sentence	Errneous Sentence	Error Type	Error Subtype	Error Description
The tiger of Serghei took a lot of space.	The tiger of Serghei took a lot of place.	Semantic Error	Word Choice Error	used 'place' instead of 'space'
Eliza will bring a small gift to Sophie's party.	Eliza will bring an small gift to Sophie's party.	Syntax Error	Article error	used 'an' instead of 'a'
John is sleeping at the moment.	John am sleeping at the moment.	Syntax Error	Subject verb agreement	used 'am' instead of 'is '(Third person singular)
I read the license plate.	I read the licsence plate.	Spelling Error		used 'licsence' instead of 'license'

Example 14: Book you want is out of print. (The book)

Example 15: He returned after a hour. (an hour)

Example 16: He paid a sum of money for the informations. (information)

Example 17: The sceneries here are very good. (The scenery here is very good.)

Example 18: She leaves school last year. (left) ('last year' indicates a finished event of the past)

Example 19: The boys are play hockey. (playing) (the event is currently happening, so -ing form of verb is required)

5) Semantic error: The errors that do not violate English grammar rules, but make the sentence senseless or absurd, are called as semantic errors [15]. Semantic error is further classified into subtypes a) context error and b) wrong word choice error. A correctly spelled word but wrong in context of sentence, is said to have context error. Examples 20, 21 are contextual errors. Wrong word choice error is due to limited knowledge of vocabulary. Example 22, 23 are word choice errors.

Example 20: Our team is better then theirs. ('then' is not a spelling mistake, but the context gives an idea of comparison, indicating correct word as 'than')

Example 21: The jury were divided in there opinions. (their opinions)

Example 22: A group of cattle is passing. (A herd of cattle)

Example 23: I am going to the library to buy a book. (use 'bookstore' instead of 'library').

To measure performance of grammar apps we have calculated two types of accuracy, error wise and overall accuracy. Error wise accuracy reflects the performance of a particular grammar app on each error type (also on subtype). And overall accuracy reflects average performance considering all error types. Equation (1) is used to calculate both type of accuracy.

$$accuracy = \frac{\text{No. of incoorrect sentences corrected by app}}{\text{Total no. of Sentences}} \quad (1)$$

For overall accuracy, Total no. of Sentences = 500.

For main error type , Total no. of Sentences =20% of 500.

3.3 Collection of Dataset

To conduct the evaluation of the grammar checking tools, we need a dataset of erroneous sentences annotated with the grammar error types and subtypes along with the corresponding set of correct sentences. We searched various sources on the web for the dataset required in our study, but unable to find any such dataset. Therefore, we created a dataset of 500 sentences collected from various grammar books [25], articles [20]-[22] and grammar learning websites [23]-[24] and created the required dataset by inserting the errors of each type in the sentences. The distribution of the error types in the dataset is shown in Figure 2(a). Figure 2(b) shows the distribution error subtypes under Sentence Structure error, Semantic error and Syntax error. We got the dataset checked by two experts (professors having PhD degree in English Literature). The sample dataset of four sentences is shown in Table 2. We have made our dataset publicly available on [26].

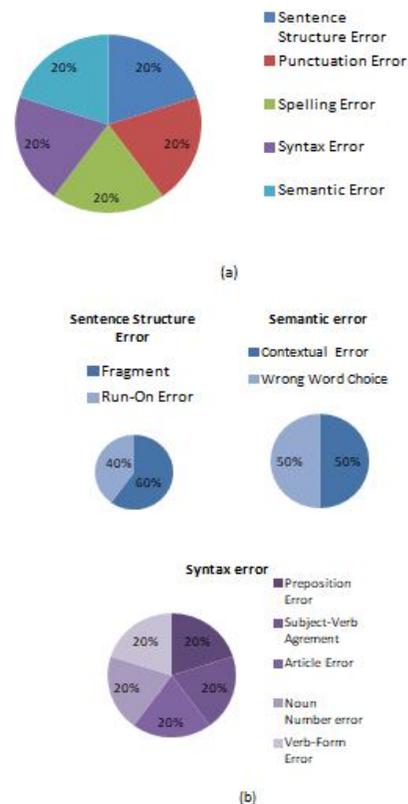


Figure 2: Distribution of total 500 sentences (a) among main error categories (b) further distribution of Sentence Structure, Syntax, and Semantic error among subtype error categories

3.4 Testing of Five Grammar apps

In this step, we tested the five grammar apps (viz. Grammarly, Ginger, ProWritingAid, LanguageTool and After the Deadline) via feeding erroneous sentence (Table 2 col. 2) to these apps. For each app, this is done in following three sub steps:

- i) Each erroneous sentence (Table 2 col.2) from the dataset is given as input to the app via an interface (GUI or API) provided by the app.
- ii) The App may suggest some tip (see Figure 3) to correct the error present in the sentence. We accept the underlying suggestion (if suggestions are more than one, than we choose one with the highest priority).

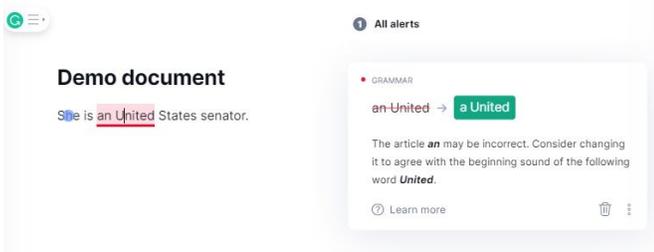


Figure 3: Grammarly suggestion on sentence “She is an United Seates senator”

- iii) The sentence is obtained after accepting suggestion in previous sub step, is called output sentence (see Figure 4). This output sentence is matched against every possible correct sentences (Table 2 col.1), if the output sentence matches with any of possible correct sentences (Table 2 col.1), then we conclude that error is correctly identified and corrected by app, otherwise not.

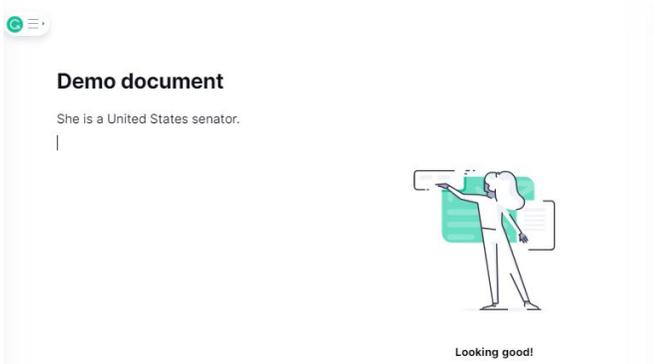


Figure 4: Output sentence after accepting suggestion.

4. RESULT AND DISCUSSION

This step analyzes the results obtained after testing of the five apps using the dataset. Figure 5 shows the comparison between overall performances of the five grammar checking apps viz. Grammarly, Ginger, ProWritingAid, LanguageTool and After the Deadline in identifying all the types of grammar errors. Grammarly has achieved highest overall accuracy of

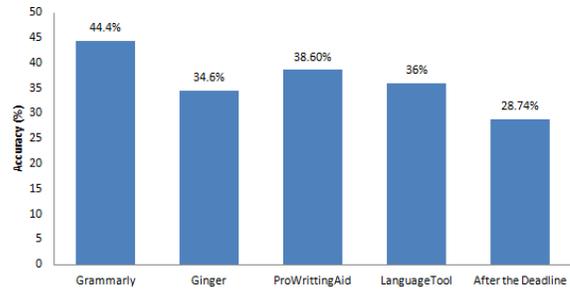


Figure 5: Overall accuracy comparisons of apps.

44.4%. Grammarly outperformed specifically in spelling and syntax error with highest accuracy of 99% and 69% respectively (see Figure 6 and Figure 7).

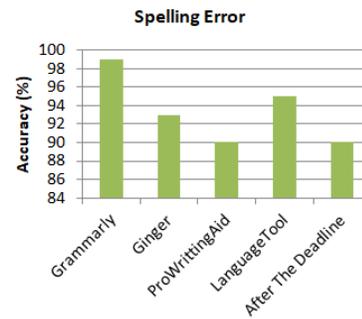


Figure 6: Comparison of accuracy in spelling error

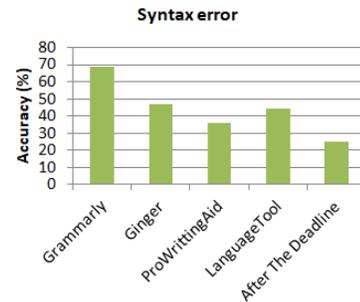


Figure 7: Comparison of accuracy in Syntax error

No app has not performed well on semantic error. All apps has comparable accuracy of 18% to 25%. (see Figure 8). In semantic error, apps are mostly failed to identify word choice error. (see Figure 13).

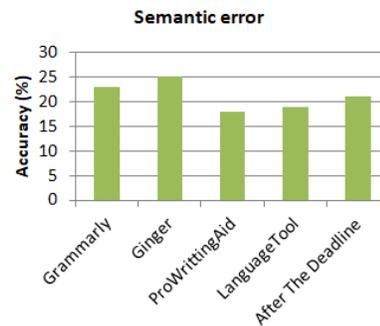


Figure 8: Comparison of accuracy in Semantic error

Similarly, no app has performed well on Sentence Structure error. Ginger and After the Deadline did not correct, even a single Sentence Structure error. LanguageTool is one with the highest accuracy of only 11% (see Figure 9 and Figure 12). In punctuation error, ProWritingAid has achieved the highest accuracy of 42%. And all other apps have accuracy below 27% (see Figure 10).

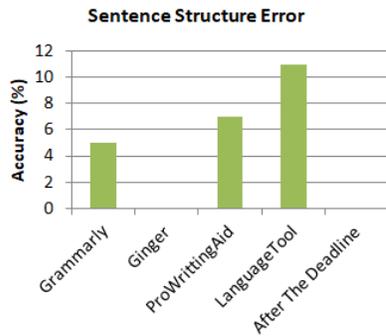


Figure 9: Comparison of accuracy in Sentence Structure error

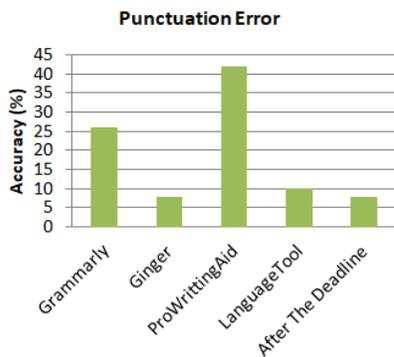


Figure 10: Comparison of accuracy in Punctuation error

Figure 11, Figure 12 and Figure 13 shows error wise comparison among Syntax error subtypes, Sentence Structure error subtypes and Semantic error subtypes respectively.

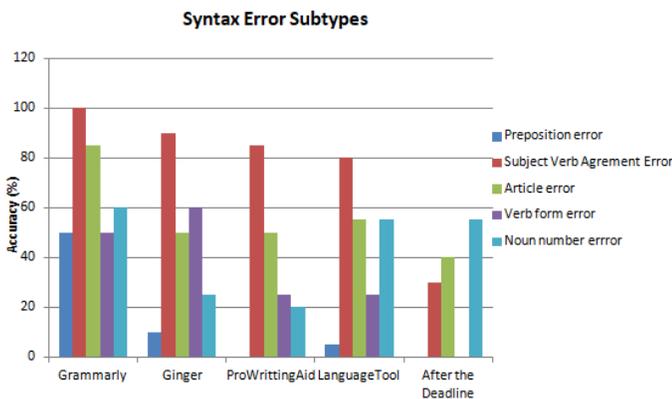


Figure 11: Error wise accuracy of Syntax error Subtypes

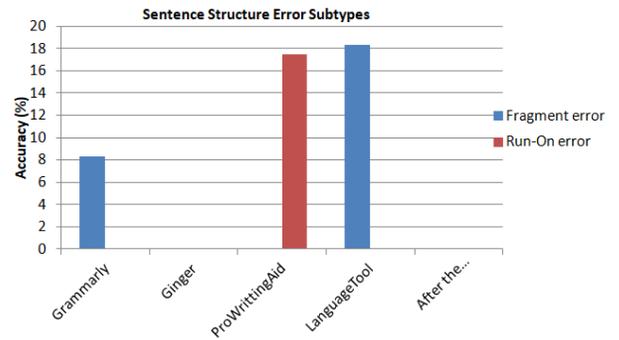


Figure 12: Error wise accuracy of Sentence Structure error Subtypes

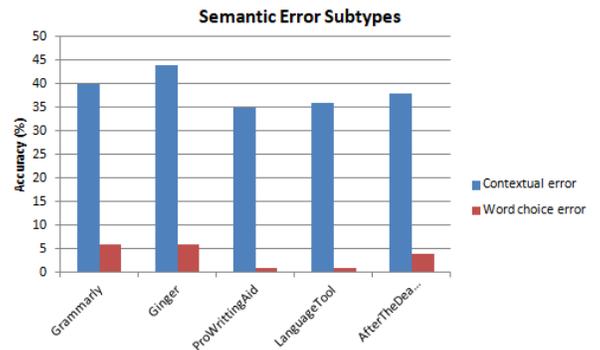


Figure 13: Error wise accuracy of Semantic error Subtypes

5. LIMITATIONS AND FUTURE WORK

In this work, we have performed the experiment for evaluating the performance of the five most popular grammar checking apps available on the web. In future, the same dataset can be used for evaluating the other grammar checking apps also. As the aim of our study was to evaluate the free and open source grammar checking apps, we have evaluated the performance of the free versions of these apps. The future works can evaluate the paid (or premium) versions of these apps. Our dataset consists of 500 sentences and contains equal distribution of each type of errors. The experiment can be repeated for a larger dataset of such kind.

6. CONCLUSION

In this paper, we presented the evaluation of the performance of the five most popular grammar checking apps. The results of the study suggest that there is a lot of scope for the improvement to be done in these apps. Specifically, all the apps failed to identify the sentence structure errors. From the users' perspective, this experiment shows that one cannot completely trust on these apps for the identification and correction of the grammar errors. And from the developer's perspective, this experiment is found to be very useful in a way that they can easily identify the areas for improvement in these apps. Another main contribution of this research is that we have made our dataset publicly available on [26].

REFERENCES

1. **Writing Evaluation Service.** ETS Criterion, available at criterion.ets.org/criterion/default.aspx.
2. **Grammar and Spell Checker.** Ginger Software, available at www.gingersoftware.com/grammarcheck
3. **Write Your Best with Grammarly.** Grammarly, available at www.grammarly.com/.
4. **Spell and Grammar Checker.** LanguageTool, available at languagetool.org/.
5. **Check Grammar and Spelling for English Texts.** Reverso, www.reverso.net/spell-checker/english-spelling-grammar/
6. **Online Spelling and Grammar Checker for English** SpellCheckPlus, spellcheckplus.com/en/.
7. **For the Smarter Writer A Grammar Checker, Style Editor, and Writing Mentor in One Package.** ProWritingAid, prowritingaid.com/Free.
8. Michelle Rose Cavaleri and Saib Dianati. **You want me to check your grammar again? The usefulness of an online grammar checker as perceived by students**, *Journal of Academic Language and Learning*, Vol. 10, No. 1, pp. 223-236, January 2016.
9. Hao-Jan Howard Chen. **Evaluating Two Web-based Grammar Checkers-Microsoft ESL Assistant and NTNU Statistical Grammar**, *The Association for Computational Linguistics and Chinese Language Processing*, Vol. 14, No. 2, pp. 161-180, June 2009.
10. Muhammad Fareed, Almas Ashraf and Muhammad Bilal. **ESL Learners Writing Skills: Problems, Factors and Suggestions**, *Journal of Education & Social Sciences*, Vol. 4, No. 2, pp. 83-94, October 2016. <https://doi.org/10.20547/jess0421604201>
11. Muhammad Ali Ghufuron and Fathia Rosyida. **The Role of Grammarly in Assessing English as a Foreign Language (EFL) Writing**, *Lingua Cultura*, Vol. 12, No. 4, pp. 395-403, November 2018. <https://doi.org/10.21512/lc.v12i4.4582>
12. Haist, Caroline. **An Evaluation of Microsoft Word 97's Grammar Checker**, January 2000.
13. **After the Deadline.** *After the Deadline* - Spell, Style, and Grammar Checker for WordPress, Firefox, TinyMCE, JQuery, and CKEditor, www.afterthedeathline.com/.
14. Reid, Joy M. **Teaching ESL Writing.** Englewood Cliffs, N.J.: Prentice Hall Regents and London: Prentice-Hall International, 1993.
15. Madhvi Soni and Jitendra Singh Thakur. **A systematic review of automated grammar checking in English language**, *arXiv preprint arXiv:1804.00540*, 2018.
16. Randy Joy Magno Ventayen and Caren C Orlanda-Ventayen. **Graduate students' perspective on the usability of grammarly® in one ASEAN state university**, *Asian ESP Journal*, Vol. 14, pp. 9-30, January 2018.
17. Jin, Yang Hye. **Efficiency of Online Grammar Checker in English Writing Performance and Students' Perceptions**, *Korean Journal of English Language and Linguistics*, Vol. 18, No. 3, pp. 328-348, 2018. <https://doi.org/10.15738/kjell.18.3.201809.328>
18. Yeung, C. Y., and Lee, J. **Automatic detection of sentence fragments**, in *Proc. Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Vol. 2, pp. 599-603, July 2015 <https://doi.org/10.3115/v1/P15-2099>
19. AS Hornby. **Guide to Patterns and Usage in English.** Oxford University Press, 1995.
20. **STUDY.COM**, study.com/academy/lesson/spelling-mistakes-common-examples-quiz.html.
21. Walker, Nicholas. **Word Choice Errors.** Virtual Writing Tutor Blog, 14 Oct. 2018, blog.virtualwritingtutor.com/word-choice-errors/.
22. **Incorrect Verb Forms.** English Grammar, 26 Dec. 2013, www.englishgrammar.org/incorrect-verb-forms/.
23. Shrives, Craig. **What Are Articles? (with Examples).** Articles What Are Articles? www.grammar-monster.com/glossary/articles.htm.
24. **Indefinite and Definite Articles.** Articles - Indefinite and Definite Articles - Basic English Grammar, www.talkenglish.com/grammar/articles.aspx.
25. Peter Herring, **Complete English Grammar Rules.** The Farlex Grammar book, July 2016.
26. **Dataset:** <https://sites.google.com/view/idoresearch/home>