English Correction Software

ToBIT Paper

Tobias Koller

University of Applied Science Northwestern Switzerland (FHNW),
4000 Basel, Switzerland

Abstract. This paper forms ....

Keywords: Grammarly · Natural Language Processing · Language Correction Software

1 Introduction

Numerous software solutions on the market promise to help ~~in particular~~ non-native English speakers with grammatical error detection and improving the style and structure of their writing. In the course of the module "Innovative Topics in Business Information Technology" (ToBIT), I am going to evaluate different products regarding their functions, ease of use and effectiveness in supporting the writing process. The main goal is to use literature review to discover how effective the correction software is under real conditions.

One widely known and used writing tool for grammar checking, spell checking, and plagiarism is Grammarly®. Since it is one of the leading products in this field and there are ~~innumerable~~ research papers ~~to support~~, I focus on this particular tool.

many

analyzing its use

After the evaluation of the tools, I will ~~also~~ describe different natural language processing (NLP) techniques that are being applied. In the final part of the document, the findings will be discussed to show ~~clearly~~ the current state of the art in this field. A recommendation to the University of Applied Sciences and Arts Northwestern Switzerland (FHNW) will be given on what products to consider or how to develop a new product internally.

increase the

You need to ~~change~~ spacing (to 1.5 or 2), so that someone who corrects your work has room to write!

This lowers the barrier ~~of~~ entry ~~immensly~~ since all a user needs to have is an e-mail address and a computer or smartphone with an internet connection. However, a premium service is offered against payment. Grammarly's website ("Write your best with Grammarly." n.d.) shows the deviation from the premium plans to the free version. The plans "Premium" and "Business" both comprise the same advanced features including suggestions in the following categories.

- Fluency
- Readability
- Engagement
  - Compelling vocabulary
  - Lively sentence variety
- Delivery
  - Confident language
  - Politeness
  - Formality level
  - Inclusive language
- Plagiarism detection

"Business" plan only differs in the account management tools that are available to the organisation's administrator ~~as well as~~ business-oriented billing. The available software and plugins (browser, MS Word, MS Office, mobile keyboard) are equally available for paid and free users. Since Grammarly has its focus on supporting students, they target educational institutions with their programme "grammerly@edu". The grammar checking functionalities seem to be the same as in the "Premium" and "Business" plans but they offer specialised licenses and 24/7 support. Different versions of the same tool make the evaluation of the tool in the following chapter more difficult since some of the research found is based on the limited functionalities. Furthermore, the software developed greatly in the past years as can bee seen on print screens from Dembsey (2017). This needs to be taken into account when judging the results from this research.

some notes to the author

3 Evaluation

3.1 Grammarly

The Grammarly's web presentation ("Write your best with Grammarly." n.d.) makes strong claims about the effectiveness and usefulness of the system. Over the past years, ~~several~~ researchers attempted to measure the impact of using

his writing. The notion of something being wrong with his writing motivates to comply with general phrases and standard structure in the future. Instead of promoting better writing style, they "fear that the persistent underlying urge towards conformity may stifle individual creativity" (Vojak et al., 2011).

Grammar correction on a sentence level follows rather clear rules whereas connections within a paragraph or even the logical structure of the whole document are much more sophisticated tasks. Unsurprising that also automated correction software like Grammarly have their difficulties. Students experienced the lack of context-aware checking like coherency and cohesiveness within a text. Those who needed the software only to check the grammar did not find this an issue (Nova, 2018). The same results were found by Dembsey (2017) who observes that Grammarly treats each word and sentence individually and not making any connections between them, therefore drastically reducing the learning opportunity, compared to expert feedback.

Amount of errors found On first sight, high quantitative figure of detected errors seems to demonstrate the superiority of a correction algorithm. We will see why this appearance is deceptive.

During the comparison of Grammarly with writing consultants in analysing three student essays, Dembsey (2017) observed a total of 118 issues whereas the cumulative average of the 10 consultants only brought up 51. Repetition of the same issues was the main driver for such a high number of detected issues. A human proofreader could encourage the student to look for additional instances of the same mistake by themselves, leaving more time for different issues. To get a better view of the issues discovered, all issues were categorised, which led to a total of 16 categories to which every issue could be assigned. In all the essays combined, Grammarly's correction cards could be assigned to only six types of issues. This again shows the rather narrow range of recommendations. Cumulated, all 10 consultants addressed 15 issue categories and even on average they addressed more (8) diverse topics than Grammarly.

Despite having found more issues than a human proofreader, Grammarly's issue detection was highly repetitive and only addressed a narrow range of issues. The consultants used less comments but gave a more in-depth explanation and could even connect sentence-level issues to general (thesis) level issues. Furthermore, a high number of issues is often not beneficial for the learning rate of students, as they might become intimidated and demotivated. (Dembsey, 2017)

Accuracy A more crucial measure of value provided by the feedback than number of issues detected is the accuracy of the results. False positives are reported issues that are no problems at all. Dembsey (2017) also considered incorrect use of term or incorrect explanation as inaccuracy. 41% of Grammarly's

start using Grammarly is the registration with an e-mail address and password. Technically, the installation of any software is not required since the system can be used immediately through the browser which serves the main correction features ("Write your best with Grammarly." n.d.). The features are better integrated into the writing process when the provided plugins are used. The interviewed students in Nova's study found no barriers in the download and setup process.

Perceived Usefulness According to the Technology Acceptance Model (TAM), besides ease of use, perceived usefulness is a key factor that influences the people's intention to use computer systems (Davis, Bagozzi, & Warshaw, 1989). In the survey conducted by Cavaleri and Dianati "most students reported that they found the suggestions helpful for improving the particular paper they had submitted to Grammarly and half felt it helped them achieve a better mark" (Cavaleri and Dianati, 2016). Effects were not only short term, students felt the card's feedback helped them in understanding issues better and improve their writing skills also long-term. Therefore, usefulness is not limited to the current piece of writing but rather on the whole learning experience of each user and supports self-directed learning. Besides the direct improvements on the correctness of the grammar, 77.8% of respondents felt an increase in their confidence level after using Grammarly.

These results cover the conclusion by students interviewed by Nova (2018). They mention the positive impact of feedback cards on their self-revision. Increased reflection on the issues found helped them to improve the quality and avoid the repetition of errors. Especially the indication of example sentences helped them to understand the issues better and apply a correction.

However, some students detected also disadvantages that reduced the overall usefulness of Grammarly. Both Ventayen and Orlanda-Ventayen and Nova claimed that some parts of the document should be excluded from grammar checking like the bibliography that follows certain standards. Checking on a reference list does not yield any benefit and only distracts (Nova, 2018; Ventayen & Orlanda-Ventayen, 2018). Another limitation found by Cavaleri and Dianati (2016) was the complex language used in some of the recommendations. Deciding on whether to accept the change or not required some deeper understanding of the problem at hand. When students were not able to understand the issue and the underlying grammar rule, they were not able to make those decisions. Therefore, advanced English writers could benefit more than others. The complex language used in the feedback cards can be seen as a barrier for beginner-level students.

by MA. Teachers indicated that the grade given was influenced by MA by an average of 18%. This number is relatively low since most teachers did not put much confidence in the accuracy and fairness of the automated scoring. On average, they treated the fairness of the system slightly lower than neutral. Knowing the limitations of the automated scoring, it is not surprising that teachers still read the students work very thoroughly. 41% reads them even as thoroughly as when they would not use MA.

Teachers observed different reactions to the automated scoring feature (Grimes & Warschauer, 2010). While some students were increasingly motivated to write a high-quality text for the immediate reward, others were highly distracted by the score and could no longer focus on their task. Some teacher even disabled the automated scoring and only showed their students after submission. Some of the high-performing students that reached a very high score on their first submission were no longer motivated on revisioning whereas if they would not have known, they could have still found parts to improve. Another development observed was students that tried to learn how the scoring algorithm works and then submit text that would simply lead to a higher score but does not make sense in the context of the paper. From those reports, we can conclude that teachers are advised to tightly observe the usage of the AWE by their students. Only if they support their students and prevent misuse, the automated scoring can provide real value by allowing the writers to assess and motivate themselves.

Student's attitude towards AWE Increased motivation towards writing and revising was found by Grimes and Warschauer (2010). Reasons identified were the immediate feedback by the AWE instead of week-long waiting time for a human feedback. For them, the automatic score took the characteristics of gamification and they tried to outperform each other which increased motivation even further. They were also able to use the time after the first submission for further improvements since the feedback is available immediately.

Students also did not rate the fairness of the automatic grades as critical as the teachers. They rated the fairness with 3.4 (on a 5-point scale) whereas the teacher's rating was only 2.8.

Regarding the amount of revisions done by students, the first year did not show any increase and only 12% of essays had more than one revision (Grimes & Warschauer, 2010). In the following year, this changed to 53%. On the one hand, it can be reasoned that teachers allocated more time for the revision process but also the students who learned how to properly use the system and make the best use of its features. Students who revision their writings first focus on low-level issues like spelling and punctuation before moving to feedback about organisation and development. This seems to be a natural behaviour to focus on the low-hanging fruits that can be fixed with lower efforts. Improving on the

Grimes, D., & Warschauer, M. (2010). Utility in a Fallible Tool: A Multi-Site Case Study of Automated Writing Evaluation. *The Journal of Technology, Learning and Assessment*, *8*(6). Retrieved October 14, 2019, from https://ejournals.bc.edu/index.php/jtla/article/view/1625

Nova, M. (2018). UTILIZING GRAMMARLY IN EVALUATING ACADEMIC WRITING: A NARRATIVE RESEARCH ON EFL STUDENTS' EXPERIENCE. *Premise: Journal of English Education*, *7*(1), 80–97. doi:10.24127/pj.v7i1.1332

Qassemzadeh, A., & Soleimani, H. (2016). The Impact of Feedback Provision by Grammarly Software and Teachers on Learning Passive Structures by Iranian EFL Learners. *Theory and Practice in Language Studies*, *6*(9), 1884–1894. doi:10.17507/tpls.0609.23

Ventayen, R. J. M., & Orlanda-Ventayen, C. C. (2018). *Graduate Students' Perspective on the Usability of Grammarly® in One ASEAN State University* (SSRN Scholarly Paper No. ID 3310702). Social Science Research Network. Rochester, NY. Retrieved October 23, 2019, from https://papers.ssrn.com/abstract=3310702

Vojak, C., Kline, S., Cope, B., McCarthey, S., & Kalantzis, M. (2011). New Spaces and Old Places: An Analysis of Writing Assessment Software. *Computers and Composition*, *28*(2), 97–111. doi:10.1016/j.compcom.2011.04.004

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, *10*(2), 157–180. doi:10.1191/1362168806lr190oa

Write your best with Grammarly. (n.d.). Retrieved November 27, 2019, from https://www.grammarly.com/