

WritingAssistant™
Comprehensive automated feedback

Erica Schramma and Venkat Srinivasan
EnglishHelper, Inc., Dedham, MA, USA

Abstract: *Learning English is becoming the goal of increasing numbers of people around the world. The resources provided in the classroom, however, are not always sufficient for adequate practice and reinforcement for English language learners. This is especially true when learning to write in English, where detailed feedback is a crucial part of the learning process. There is, therefore, need for a tool that will comprehensively detect issues in English as a Second Language (ESL) writing – one that is non-intrusive, contextually relevant, and tailored to the specific needs of users.*

WritingAssistant™ is an automated, tractable, and traceable software solution that enables writers of all levels to improve their writing skills. It provides easy access to a proofreading and self-learning tool that can be used anywhere and in any context, with a flexible architecture that makes it highly configurable for users. Its rule-based approach to error detection and correction produces high rates of precision and recall: on a corpus of texts written by non-native speakers, WritingAssistant™ returned 67% recall and 94% precision, suggesting that the tool can be very beneficial for English language learners.

Keywords: automated error detection, writing tool, ESL technology, feedback, editing

Introduction

English is well known as world's lingua franca – it is estimated that about one-fourth of the world's population can communicate in English with varying fluency. While enabling people separated by great distances and cultures to communicate easily, English language skills also provide global opportunities and a competitive edge in the job market. The opportunity to build English skills is in high demand, but not everyone has equal access to resources to strengthen these skills.

People who learn English in a classroom but do not hear it in a real-world context cannot reinforce the skills they have learned in class until they are faced, often unprepared, with a practical situation that demands it. Oftentimes, a dearth of proficient English as a model – whether it is an English teacher trying to see to the needs of many students or the lack of any proficient speakers in a learner's vicinity – can inhibit the progress of English writing skills, where consistent interaction and corrective feedback are crucial to improvement.

Foreign language classrooms have seen a huge increase in the number of language-learning technologies in recent years as computers and other interactive media have become widely available and less expensive. Most of the interactive tools that are available focus on language acquisition, while writing skills are expected to be learned passively. Based on the current situation, however, where people are heavily reliant on written communication in English through e-mail, news, blogs, SMS, etc., there is need for a tool that will comprehensively detect issues in ESL writing.

The WritingAssistant™ (WA) is an automated writing analysis product that looks for errors in all areas of grammar, mechanics, and style. WA points out the errors to the user and corrects them where it can, giving an explanation about why the construction is incorrect and how the user might fix it for each error. Its architecture is designed to be tractable, traceable, and flexible, performing consistently for all levels of English proficiency and mimicking the mental process of a human editor.

Feedback and Automatic Error Detection and Correction

It is well-documented in educational research that giving learners writing feedback helps reduce the number of mistakes they make (e.g. Leacock et al. 2014; Myles 2002). As ESL learners practice writing as a method to improve fluency in their target language, if they are not supplied with corrective feedback, they run the risk learning and fossilizing incorrect constructions. Feedback on writing provides reinforcement of learned concepts and negative evidence to help learners refine their understanding.

Leacock et al. (2014) summarize the conclusions that can be drawn from existing research:

1. Even feedback on one or a few error types can be of value to the language learner.
2. The benefits are not limited to revisions of the particular essay in which the feedback was given, but instead can generalize across topics and persist over months.
3. Not all types of errors have shown these benefits, so it may be inappropriate to generalize, for example, from article errors to preposition errors or even, within article errors, from mistakes involving *a/the* referential usage to those that involve omission of an article in an obligatory context (e.g., *the students enjoyed course*). (110, emphasis in original)

The feedback from these studies is all provided by human editors, which raises the question if feedback provided by automated writing checkers can have the same benefit to learners. There has not been much research done in this area, and almost no studies examine long-term effects of automated feedback. The effect of the Criterion system (Burnstein et al. 2004) on student writing as compared to human editing and feedback has been examined in several studies, whose results indicated that automated feedback produces positive results, with non-native speakers making statistically significant improvements in their error rate (e.g. Choi 2010; Chodorow et al. 2010). Because automated systems are at risk for pointing out errors that are not incorrect (i.e. false positives), there is the risk that ESL learners will absorb incorrect information given by the automated system; research indicates, however, that learners who use automated feedback “do not blindly accept the system’s output but instead [are] discerning in their choices [for accepting good suggestions from the automated system]. The data seem to reflect their greater ability to recognize correct usage, even when they [are] unable to reproduce it” (Leacock et al. 2014, 111).

The type, quantity, and quality of feedback that writers receive also bear on its overall effectiveness. Individual, formative feedback is regarded as most effective, since it is personalized to the learner’s writing and provides suggestion for improvement. Summative feedback, in contrast to formative feedback, evaluates overall performance. Shute (2008) explains that formative feedback consists of verification and elaboration; verification most basically is identifying a response as ‘correct’ or ‘incorrect’, while elaboration can more specifically address the response or topic, or give examples or guidance for the correct answer. Formative feedback is intended to “[shape] the perception, cognition, or action of the learner” (Shute 2008, 175). Thus, giving a learner only an overall score on their writing with no indication how to improve is not ideal in an ESL context. In addition, writers should not be overwhelmed with corrective feedback, since that can cause cognitive overload and prevent retention of any of the information presented in the feedback. Ideally, a balance should be maintained in giving enough feedback and elaboration that improvement and learning can occur, and not so much that the user becomes overwhelmed and possibly discouraged.

As mentioned above, the quality of feedback given by an automated system also has an impact on user performance. Leacock et al. (2014) classify suggestions given by automated systems as ‘good’ when the feedback provided by the system fixed the problem in the learner’s input (i.e. true positive), ‘neutral’ when the suggestion was either unnecessary or the suggestion was correct but the learner did not fix the problem, or ‘bad’ when a suggestion was generated but there was no error (i.e. false positive). The goal of any automated writing correcting system is to maximize the number of true positives and minimize the false positives for both high accuracy and recall, but they make the assumption that the user will always successfully incorporate the feedback into their writing. Roscoe et al. (2010) highlight this discrepancy that developers of automated writing evaluation tools tend to measure success in terms of accuracy scores rather than efficacy for the learner, even if the ultimate goal of the tool is to improve the user’s writing.

Context and User Interaction with Language Learning Technology

Language learning in the classroom has many advantages with relevance to the learner’s context, as they can be of various ages, proficiencies, backgrounds, abilities, etc., and are continually influenced by their cultural, social, and political contexts. All of these attributes cannot be separated from the individual and therefore reflect in his or her writing (Kelly et al. 2004).

The advantage to an automated writing editing system is that it can itself be context-neutral, leaving the user to apply and work within his or her context. It eliminates any bias that might occur in a classroom setting, because it is completely impartial. While the disadvantage to such a system is that problems that might be common to a specific population, such as certain language patterns from a particular mother tongue, are not often accounted for, these can be addressed by keeping the system flexible so it can be changed to incorporate such patterns in the rules.

Another advantage to using an automated editing system is that it can provide feedback immediately, where a classroom instructor cannot give comprehensive feedback rapidly, especially if they have a large group of students. Shute (2008) explains that immediate feedback is beneficial for difficult tasks and promotes retention of conceptual knowledge, which is relevant to writing and language learning. Over time, the feedback given by automated systems is consistent in both emphasis and application, as every time the system identifies the same type of error, it will deliver the same message, even over long periods of time. In this way, it is not vulnerable to human inconsistency, thereby enabling a reinforcement of the concepts the writer needs to improve.

Considering the factors of context and feedback together, long-term usage of an automated error correction system may have the ability improve the user’s writing skills, even if their original goal was to immediately correct a piece of their writing. If the user finds the feedback informative and useful, this can “influence a learner’s goal orientation (e.g. to shift from a focus on performing to an emphasis on *learning*) [...] via formative feedback” (Shute 2008, 162, emphasis in original).

WritingAssistant™

With so many factors to consider when evaluating ESL writing, it is a daunting task to create an automated solution that adequately meets the needs of all possible users. The product that we have developed, WritingAssistant, attempts to address many of the challenges faced by automated writing evaluation systems. The errors that we assess are those that cover all areas in writing, and for every error we detect, we offer an explanation of how to correct the error. The architecture of the system is designed to be flexible and provide transparency between the user-submitted text and the reason for the error being generated.

The WA Taxonomy of Errors

The WritingAssistant checks text for errors in three broad categories: Mechanics, Grammar, and Style (divided into Vocabulary and Content Quality). The mechanics category includes spelling, punctuation and case; it looks for words that are misspelled or incorrectly capitalized and checks for clear and correct usage of punctuation. The grammar rules find errors in syntax and tense, as well as pointing out words that might be inappropriately chosen. Style covers the overall flow of the sentences, checking if they adhere to conventions of good writing and word choice.

If using WA in a classroom context, the numbers and types of errors can be tallied to create an overall assessment score of the text. For individual usage, an overall assessment score is not calculated; rather, the error count per category is displayed without a calculated score. The categories that are included in this score and whose errors are shown can be customized depending on user requirements; if the user desires that only grammar and not style should be examined, or even that finer subcategories such as article usage within grammar should be omitted, the options can be so configured. If using the overall scoring feature, the weight of each subcategory in the final score can be customized so that an overall assessment rating reflects features that the user wishes to emphasize.

Table 1: Categories of Errors in WritingAssistant™

Mechanics	Grammar	Vocabulary	Content Quality
<ul style="list-style-type: none"> • Punctuation • Spelling • Incorrect Case • Spacing 	<ul style="list-style-type: none"> • Subject-Verb Agreement • Verb Transitivity Problems • Incorrect Prepositions • Article Usage • Singular/Plural Issues • Sentence Fragments • Run-on Sentences • Pronoun-Antecedent Agreement • Mixed Construction • Needed Words Missing • Verb Tense 	<ul style="list-style-type: none"> • SMS Lingo • Indian English • Incorrect Word Usage 	<ul style="list-style-type: none"> • Appropriateness of Tone • Wordy Sentences • Repetitive Word Usage

Architecture

WA is designed to be highly scalable and flexible. It benefits from a unique model-driven architecture that allows it to learn continuously. Experts can enhance many aspects of the WA engine without needing to resort to programming. The engine's architecture also allows it to be configured for different contexts rather easily.

The WA engine is comprised of three logical components: a Part-of-Speech (POS) tagger, an Analyzer, and an Editor. The POS tagger determines the parts of speech of the words in the text. The Analyzer analyses the text and the Editor determines the corrections for any errors that are found.

Accurate POS tagging, assessment of grammar, and determining relevance are all highly dependent on understanding the context of a written piece of work (e.g. Kukich 1992). In order to help determine the context of the text, WA has functionality that allows a 'guided form' in which an administrator can set a writing prompt. The administrator can then also add multiple 'gold standard responses' to the project, which function as ideal responses to the given prompt. The engine then uses the responses and machine learning techniques to generate a list of key concepts in the project. This list can assess the relevance of a writing sample submitted by a user to the prompt. The words in the list can also be added to the custom dictionary, allowing them to take on a different meaning or new properties from the dictionary meaning of the terms: e.g., to denote proper nouns, definiteness for article usage, etc. This prompt-based function is well suited for classroom and corporate training environments. The custom dictionary feature itself is available for all users, which allows them to add relatively uncommon names or concepts that they frequently use. Using this feature improves the accuracy of the analysis for their submissions.

WA can be deployed through several platforms – on the web, as an MS Office plug-in, and, in the future, a native app on tablet devices. When deployed as a plug-in, a synchronization process ensures that the rules in the local version are updated with any changes on the WA server.

POS Tagger

The Part of Speech (POS) tagger is the first major step to parsing and correcting the submitted text. Each word is assigned a part of speech based on its function in the sentence (noun, verb, etc.). The POS tagger has been created using corpora-independent methods, so that there is no rule bias created by the confines of a text corpus.

The rules of the tagger encapsulate the patterns of normal English syntax, but are flexible enough to allow for variation from imperfect user input. They determine the word's POS based on specific context that speakers would use to disambiguate the word's function. The POS tagger also uses *n*-grams, which are common constructions that follow a fixed pattern and might have a different function as a group than in their individual components, possibly affecting the parsing of the sentence.

Because the tagger is comprised of sets of externalized rules, it allows us to achieve higher accuracy in tagging across heterogeneous user bases than a system that uses probabilistic determination would (e.g. Rozovskaya and Roth 2014). The rules allow the POS tagger to be sensitive to recurring error patterns in the writing of English language learners. The WA POS tagger's accuracy competes with well-regarded taggers on tagging accuracy of the Brown

corpus, and outperforms other taggers on learner data since those taggers have been trained on corpora of well-formed text.

Table 2: Part-of-speech tagging accuracy on learner data: WritingAssistant™ versus Stanford POS tagger (Toutanova et al. 2003)

	% Accuracy	N
WritingAssistant™ Tagger	97.52	49,350
Stanford Tagger	94.65	49,324

Based on the POS tagging output, the sentences are parsed into phrases and clauses so that they are understandable for the Analyzer and Editor.

Analyzer and Editor

The Analyzer and Editor perform the type of analysis a human editor would upon reading the text. The Analyzer scans for grammatical errors in the text, focusing on the types of errors encountered in the writing of ESL learners (e.g. Leacock et al. 2014). The Analyzer and Editor, like the POS tagger, have a structure that allows for great flexibility: there is an encoded underlying grammar of correct English, on top of which we have a set of built-in abstract rules that detect deviations from the structure, as well as an externalized rule dataset that looks for certain error patterns. These external rules can be customized for ESL learners coming from various language backgrounds.

The Analyzer searches for these errors while the Editor attempts to correct the errors, especially those that can alter meaning and hamper communication. Sometimes an obvious correction suggests itself, as in the case of the indefinite article usage error *an car*, where the article *a* should be used instead of *an*. In other, more complicated cases such as Mixed Construction, the Analyzer may detect that the grammatical structure is inappropriate, but there is no obvious single way to correct the error.

For every error detected by the Analyzer, whether it is corrected by the Editor or not, a description of the error is given. This description gives a grammatical or conventional explanation as to why the error was detected and how to fix it. The explanation gives guidance to the user for understanding and correcting the error. Should further grammatical explanation be required, the WA web portal includes a library of lessons on English grammar and writing conventions. The formative feedback given by the Analyzer and Editor provides learning opportunity for the user.

The option to customize the categories that the Analyzer checks helps minimize the chance that the user would be overwhelmed by the feedback. Though WA does not do this automatically, the user still has the option to focus on particular areas by using the category checking options rather than working through everything at once.

Our description of all of the uses of our automated error detection system points to optimized use being determined by the learner themselves. This supports the point that a product such as ours is a tool for learning writing skills, not a substitute for a human instructor.

Feedback Analysis

Using a sample from our data corpus, we will give a few examples of the errors the Analyzer detects, and the feedback it gives for each of these errors.

Source Text:

Dear Deepak,
Sincere apologies for late mail.
Would really like to apologies on my teams part that both my nominated riders dint attended this session due to their personal issues & dint even informed us regarding their absence from session.
I had coached them about the same.
I completely take the accountability for this & it's my commitment that will send maximum number of riders in next riders training session.
Regards
Govind

Edited Text:

Dear Deepak,
Sincere apologies for **the**¹ late mail.²
Would really like to apologies on my teams part that both my nominated riders **did not**³ **attend**⁴ this session due to their personal issues **and**⁵ **dint**⁶ even informed us regarding their absence from **session**⁷.
I had coached them about the **same**⁸.
I completely take the accountability for **this**⁹ **and**¹⁰ **it is**¹¹ my commitment that will send maximum **number**¹² of riders in next riders training session.
Regards¹³,
Govind

The errors presented below are in the order they are found in the text. Not all of the errors found by the Analyzer are examined here. The columns contain the index number of the error in the order it is found in the text, the broad category designation, the subcategory, the erroneous text, the error explanation, and the sentence in which the error is found, corrected if possible.

2	Grammar	Sentence Fragment	Sincere apologies for the late mail.	This is a sentence fragment or poorly constructed sentence; it appears to be missing a verb.	Sincere apologies for the late mail.
---	---------	-------------------	--------------------------------------	--	--------------------------------------

The explanation for this error points out that it is a sentence fragment, both in the subcategory designation and in the explanation itself. It also states that it is a sentence fragment because it is missing a verb; this indicates that if the user revises the sentence to include a verb, the error will be corrected. One possible correction is *I sincerely apologize for the late mail*. The construction of the original sentence is common in spoken English, but it is not appropriate for written English.

4	Grammar	Other Verb Problems	did attended	After a helping verb like 'do', 'does' or 'did', use the base form of the verb, 'attend'.	Would really like to apologies on my teams part that both my nominated riders did not attend this session due to their personal issues and dint even informed us regarding their absence from session.
---	---------	---------------------	--------------	---	--

This error corrects *did not attended* to *did not attend*. *Did* is a helping verb, and in compound verb constructions, helping verbs must be followed by the infinitive or base form of the verb (we distinguish between 'base form' and 'infinitive' to minimize confusion between infinitive verb forms and infinitive constructions, e.g. *attend* and *to attend*). The explanation

states that *do* is a verb that requires special attention since it is a helping verb, and so the verb following it must be in base form to create a correct grammatical construction.

6	Grammar	Incorrect Word Usage	dint	This looks like incorrect use of a word similar to the one intended.	Would really like to apologies on my teams part that both my nominated riders didn't attended this session due to their personal issues and dint even informed us regarding their absence from session.
---	---------	----------------------	------	--	---

In this error, the user wrote the word *dint* where he or she intended *didn't*. The Analyzer can determine that the word *dint* within this sentence is probably not correct, so it points out that there is an error even though the Editor cannot provide a correction that it is confident is correct. This encourages the user to review the sentence and try to find the word they intended.

8	Vocabulary	Indian English	same	Use of 'the same' instead of the pronoun 'it' is colloquial. If the audience is not familiar with Indian English, consider revising.	I had coached them about the same.
---	------------	----------------	------	--	------------------------------------

This is an example of a language-specific construction that the Analyzer can be trained to detect. Many of the pilot studies for WA were conducted with English language learners from India, so we were able to create Indian English-specific rules. This construction is very common in Indian English, where *the same* is used where a speaker of a standard English dialect might use *it* or *this*. In our system, this is flagged as a suggestion rather than an error, since Indian English is often considered its own dialect (e.g. Sailaja 2009) and therefore is not required to hold American or British English as the writing standard. Therefore the writer can choose to revise the sentence or not, depending on his or her purpose.

11	Vocabulary	Appropriateness of Tone	it's	Use contractions only for informal situations.	I completely take the accountability for this and it is my commitment that will send maximum number of riders in next riders training session.
----	------------	-------------------------	------	--	--

WA allows the user to indicate the level of formality in their text by toggling the Appropriateness of Tone subcategory. When the Analyzer is looking for these errors, it assumes that the text is written formally, as in a business context. In formal writing, contractions such as *it's* should be written out as *it is*. This error points out this informal construction and indicates to the user that it should be revised to *it is* in order to be appropriate for a formal context.

13	Mechanics	Punctuation	Regards	A comma should follow the closing of an email.	Regards
----	-----------	-------------	---------	--	---------

Letters and e-mails have many structural conventions which can be difficult to master, so WA includes an option to check for letter header and footer errors. In this case, the Analyzer detects that there is a closing word that is missing a comma. This error is fixed in the corrected text that the Analyzer returns to the user, shown above.

In order to be concise on the error report that WA presents to the user after the submitted text has been analyzed, the explanations given in the report are necessarily brief. They are,

however, supported by in-depth Language Lessons that explain all of the relevant aspects of grammar and writing conventions so that the user can understand why the Analyzer is detecting the error and how he or she can correct it.

The calculated accuracy of our system is currently 67% on non-native English speaker text, with a 6% false positive rate, making our results highly competitive in the market, especially on learner text. These results are, of course, only indicative of how effective WA can be. High accuracy and recall rates are important to the overall user experience, but a more formal long-term study needs to be undertaken in order to definitively test WA's effectiveness with improving writing for English language learners. One informal pilot study we have done with a group of adult ESL learners has indicated improvement in English writing over a period of several months, but a controlled study will substantiate those reported findings.

Conclusion

The WritingAssistant™ is a complete tool for analyzing writing and providing corrective feedback for English language learners. Its architecture allows for robust detection of errors, and is highly adaptable to different contexts and language backgrounds. Because of this adaptability, the WritingAssistant™ is also stronger than currently available online grammar checkers for English language learners. Furthermore, the engine performs particularly well in the areas in which English language learners struggle most, such as article usage, verb transitivity, and incorrect preposition usage. While the engine may not detect as many errors overall as human graders, it does avoid the problem of inter-rater reliability – it will detect the same number of errors in a specific example when run multiple times and the same types of errors across samples. The engine will also continue improving as we continue learning from new data submitted. We believe that the WritingAssistant™ solves the pressing need for a tool that assesses writing samples for mechanics, grammar, and style for English language learners.

REFERENCES

- Burnstein, J., Chodorow, M., & Leacock, C. (2004). Automated Essay Evaluation: The *Criterion* Online Writing Service. *AI Magazine*, 25(3), 27-36.
- Chodorow, M., Gamon, M., & Tetreault, J. (2010). The utility of grammatical error detection systems for English language learners: Feedback and assessment. *Language Testing*, 27(3), 335–353
- Choi, J. (2010). The use of feedback in the ESL writing class integrating automated essay scoring. In *Proceedings of the Society for Information Technology and Teacher Education International Conference (SITE)*, 3008–3012. San Diego.
- Kelly, S., Soundranayagam, L., & Grief, S. (2004). Teaching and learning writing: a review of research and practice. National Research and Development Centre for Adult Literacy and Numeracy.
- Kukich, K. (1992). Techniques for automatically correcting words in text. *ACM Computing Surveys (CSUR)*, 24(4), 377-439.
- Leacock, C., Chodorow, M., Gamon, M., & Tetreault, J. (2010). Automated grammatical error detection for language learners. *Synthesis lectures on human language technologies*, 3(1), 1-134.
- Myles, J. (2002). Second language writing and research: The writing process and error analysis in student texts. *TESL-EJ*, 6(2), 1-20.
- Roscoe, R. D., Kugler, D., Crossley, S. A., Weston, J. L., & McNamara, D. S. (2012). Developing Pedagogically-Guided Threshold Algorithms for Intelligent Automated Essay Feedback. In *FLAIRS Conference*.
- Rozovskaya, A., & Roth, D. (2014). Building a State-of-the-Art Grammatical Error Correction System. *Transactions of the Association for Computational Linguistics*, 2, 419-434.
- Sailaja, P. (2009). *Indian English*. Edinburgh University Press.
- Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, 78(1), 153-189.
- Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1* (pp. 173-180). Association for Computational Linguistics.