ISSN Online : 2709-4162
ISSN Print : 2709-4154
Volume.6, Issue.4
(2023)
(October- December)

Quality Assessment of Automatic Paraphrasing Tool for English: An Analysis at Syntactic Level

Zala Wahab
Irfan Ullah
Liaqat Iqbal*
MPhil Scholar, Department of English, AWKUM
Assistant Professor, Department of English, AWKUM
Associate Professor, Department of English, AWKUM
Corresponding Author Email: liaqat@awkum.edu.pk

Abstract

Tools for paraphrasing are regarded as significant educational resources that support academia. Both professionals and students can use these technologies to make their jobs easier. But the effectiveness of these paraphrasing techniques needs to be evaluated and assessed. This study examines the syntactic similarities and differences between the original and the paraphrased text to evaluate the quality of automated paraphrasing performed by such tools. The data used in this analysis comes from QuillBot's paraphrasing of both literary and non-literary texts. Through the corpus tool AntConc, syntactic features were studied. The HSO measure in WordNet was also utilized to measure the relatedness between sentences at the aforementioned level. There were many variations between the original and the paraphrased text. The automated paraphrase of non-literary text by QuillBot is closer to the original text than that of literary text. Syntactic modifications were discovered, including changes to word order, tense, voice number, and grammatical category. These modifications occasionally skewed the message while other times they elaborated it. Therefore, manual revision and rechecking of automatic paraphrases should be done rather than taking it for granted. While conventional technologies, like QuillBot, might be depended on for paraphrasing of non-literary text, it needs to be manually verified and updated in addition to the automated paraphrase of literary content carried out by such programs.

Keywords: Automatic paraphrasing, Syntactic analysis, QuillBot, HSO measure, WordNet.

1. Introduction

Paraphrasing is an important activity encountered by students as well as professionals on a daily basis. Advancement in every phase of life has also made this task easy for them. Many paraphrasing tools have been designed and are available online. These tools facilitated students and professionals but, at the same time, the quality of tool paraphrasing lags behind manual paraphrasing. There is a need to assess and evaluate tool paraphrasing. Therefore, this study aimed at exploring the quality of an online paraphrasing tool, QuillBot. QuillBot is selected for the study because it is considered a standard tool and mistake expectancy is less in it. It gives grammatically correct paraphrasing up to a certain extent. It has optimistic reviews. It is a "fantastic grammar and paraphrasing writer and can produce good and presentable writing in seconds" (Shahidul., write in draft, build a formal letter, February 2, 2022). According to Fitria (2021), it is a convenient paraphrasing tool that rewrites text material by modifying the structure of sentences and replacing words with synonyms while maintaining the meaning of the original content.

For the syntactic analysis, the grammatical status and parts of speech of the paraphrased texts were analyzed and then compared to that of the original text by using close observation and AntConc. AntConc was used because it provided an easy analysis of parts of speech and comparison and allowed numerically calculating the grammatical status of the words of sentences. The word order modification, changes in narration, sentence repetition, wrongly paraphrased sentences, and un-paraphrased sentences were highlighted

ISSN Online : 2709-4162
ISSN Print : 2709-4154
Volume.6, Issue.4
(2023)
(October- December)

in the tool's paraphrasing. The data used is selected randomly. A literary text (a few chapters of a novel) and a non-literary text i.e., newspaper articles are considered. Therefore, the first two chapters of *Emma* (Jane Austen's novel) were taken and for the non-literary text, an article from *DAWN* newspaper was used. These texts were paraphrased and the product was analyzed. The syntactic relevance of original and paraphrased text is explored in this study. The syntactic structure of paraphrasing obtained from Quillbot is analyzed. Various modifications as well as relevancies were found in the original and its alternate paraphrased text at syntactic level.

The relatedness and resemblance between the original and the paraphrased text are examined using an exploratory qualitative method. As more interpretation was needed to analyze both the meaning and the structure of the chosen texts, the qualitative technique was adopted. Both qualitative and computational observations and analyses were made of the text's substance. We looked at changes in grammatical structure and word order. For various POS comparisons, AntConc was employed. It was determined how closely connected different phrases and words from the literary and non-literary texts were. Changes in the syntactic structure of the literary text include those in the word order, speech parts, additions and deletions of words, etc. Many sentences were paraphrased correctly but often the little modifications introduced affect the sense of complete text. Grammatical modifications were observed along with incorrect substitution of words. Certain sentence narration has also changed that affected syntax and sense of sentences. Moreover, repetitive sentences have been discovered in automatic paraphrasing, indicating that the quality of tool paraphrasing may suffer if specific lines are repeatedly used. In certain sentences, words that were absent from the original text were added. When paraphrasing, QuillBot omits several terms from the original text. Several words and phrases from the original text were missing from the computerized paraphrase of the literary text. Therefore, the sentence fragments were also found not paraphrased. Furthermore, grammatical modifications were also observed in the automated paraphrasing of the non-literary text. The non-literal paraphrased text contained no words or phrases that were omitted. Although the word order and some of the words were changed, the overall concept remained conserved. The literary text was found to have fewer ambiguities and differences than the non-literary material when tool paraphrasing of the two types of texts was compared. The occurrence of tense changes was uncommon. Most of the synonyms used were real and appropriate for the situation.

2. Literature Review

The basic kind of syntactic information obtainable from a text is the word order and grammatical construction of a sentence. This type of information is easily accessible because no language-specific tools are needed for the utilization of word data. Different researchers worked on finding syntactic similarities between texts. They utilized different tools and techniques, for example, Manning (2011) claims that modern POS taggers could be used, which, have high accuracy (around 97%), but they still generate more errors than POS taggers. The sentences are analyzed at word as well as phrase level. "Syntactic similarity of sentences is based on measuring the similarity of the given words. If two sentences are similar then structural relations between words will be similar and vice versa" (Kaur, 2015, p. 216). The similarity at the phrase level is also worked upon. A researcher, Burch (2008), improves the quality of paraphrases taken from parallel corpora by considering the phrases and then their paraphrases of the same type. He parsed the parallel corpus English side, changing the algorithm of phrase extraction to get phrase labels with bilingual phrase pairs. He introduced complex syntactic labels for retaining high coverage of the non-constituent

ISSN Online : 2709-4162
ISSN Print : 2709-4154
Volume.6, Issue.4
(2023)
(October- December)

phrases. He showed manual evaluation revealed a 19% improvement over the baseline method in paraphrase quality.

The sentence similarity methods are used for measuring the degree and extent of similarity between sentences. Ferreira, Cavalcant, Freitas, Lins, Simske, and Riss, (2018), measured the similarity between words in sentences. The lexical, syntactic, and semantic components were considered. They used Li-McLean, which gave them results according to their expectations (that could be compared to state of art systems and evaluations made by humans. They used CNN-corpus for evaluating the degree of similarity between summaries. Moreover, Kaur (2015), says that syntactic similarity is a key activity in high-level text documents, natural language processing, data mining, and information retrieval fields. NLP (Natural language processing) is a type of intelligent machine that can translate text between natural languages like English and computer languages like C++. He asserts that Web mining is a technique for performing tasks on the internet, such as document clustering and community mining. Finding similarities between the two manuscripts, on the other hand, is a challenging endeavor. As the breadth of NLP expands, techniques for carrying out many aspects of language, including semantics, syntax, and paradigms, are required. He says that syntactic similarity measuring depends on that of the words. If the structural relation between words is similar then the two sentences will be considered similar. It is not more difficult for measuring syntactic similarity between two documents, but there is no more work on the syntactic similarity as part of the deep research. As a result, Kaur (2015) opted to improve the syntactic similarity between the two publications. There are many algorithms that may be used to discover word similarity, including longest common substring, Edit distance, Soundex algorithm, and bi-gram algorithm. But in such algorithms, certain issues are finding syntactic similarities between the words. He says that these approaches do not work in some conditions.

The Soundex Algorithm is a similarity algorithm determining whether or not certain rings are similar. It would not describe, however, any similarity between 'French' and the 'Republic of Franch,' because they do not begin with the same letters rather they start with different letters. Edit Distance algorithm, on the other hand, would differentiate some better results between the two strings than the Soundex algorithm, but would rate 'France' and 'French' (having 6 distance) as being highly similar to 'French' and the 'Republic of French'. Finally, he says, that the Longest Common Substring gives 'French' and the 'Republic of French' a high similarity rating (a common substring length of 6). However, the 'French Republic' string is similar to the two strings equally, the 'Republic of France' and the 'Republic of Cuba', according to the new technique. He proposed as a solution a new string similarity measure that is independent of the ordering method. He also provided a novel method that considers not only one longest common substring but other common substrings also. Kaur (2015) further says that, if two strings are uttered the same then their similarity is usually more; yet, there is a difference in both strings, therefore this does not rule out the possibility of similarity between the words. Therefore, he first verified how many neighbouring characters are present in both strings. He showed that the given two strings should be regarded as similar because they have the same words in the given documents but in a different order. On the other hand, if one of the strings is simply a character's anagram in the other text, the two documents should be considered dissimilar. Moreover, the algorithm they used also works in many different languages and gives better results for finding similarities between the two documents. (Kaur, 2015).

ISSN Online : 2709-4162
ISSN Print : 2709-4154
Volume.6, Issue.4
(2023)
(October- December)

3. Methodology

Although automated paraphrasing programs are frequently used, there are numerous inconsistencies and ambiguities in them. The automatic paraphrase differs from the original text in terms of lexicon, syntax, and frequently even semantics. The alterations in the original and paraphrased texts were the focus of this investigation. For this, exploratory qualitative methodology was used. Typically, the primary goals of qualitative research are to describe, comprehend, and explain (Lichtman, 2010). When there is no prior research to draw upon or rely upon to forecast an outcome, exploratory design is used. Swedberg claims that the goal of exploratory research is to find something novel and significant (Stebbins, 2001). This study employed both a literary piece (Emma) and a non-literary text (an article from DAWN) to to be used as data. Given (2015) asserts that the sole emphasis of data collection and analysis can be documented. The source texts and the automatic translations were used in the document. Lexical, semantic, and syntactic analysis is done on the automatic paraphrase. The morphological construction of the source text's terms and their alternatives in the paraphrased text are examined for lexical analysis. Additionally, the words' similarity is taken into account. In syntactical analysis, the word order and grammatical changes are analyzed. The relatedness and similarity of the sentences in the source text and the paraphrased version were noted for the semantic analysis.

For analyzing relatedness and similarity between words and sentences in texts, many measures are available, primarily divided into corpus-based measures and knowledge-based measures. These include Wu & Palmer (1994); Hirst and St-Onge Measure (HSO), (1998) and Resnik (1999). For the analysis of text relatedness and similarity, this study used the HSO measure. HSO is used because it produces accurate results. The knowledge-based, HSO measure is used since it produced an accurate estimate of four parts of speech, the maximum number of parts of speech that can be analysed in WS4J. We don't use Wup, Res, etc. because they can't supply adjective values. They emphasize words and verbs. Although there is no upper limit to the value that LESK can produce, this might cause confusion for the reader. The maximum value is significant since it helps determine the cutoff point. The cutting point reveals if there is a tiny difference in semantic measure between the two ideas above or below a particular point. In light of all benefits and drawbacks of various measures, we chose the HSO measure and gathered and examined the relatedness values it produced.

HSO is a path-based measure that determines the relationship between concepts (specific meaning of a word). The HSO measure formalises the relationship between semantic relatedness. It determines the amount of path alterations and the greatest separation between the words, for which a relatedness rating system was established. Using the specifics of the path that connects two concepts, the following formula can be used to give the semantic similarity between them a numerical score or weight.

Weight= C – path length – k. # changes in direction

Where C and K have constant values and are determined by trials, and the path denotes a permissible connection between concepts.

To find the HSO relatedness values, a WordNet Similarity for Java (WS4J) web application was employed. The WS4J web tool uses established measures of semantic similarity to compare ideas using the Java API. It employs the measurements of semantic similarity and depends on relationships between ideas in WordNet. It illustrates how similar the notions are. The online demo gives users two choices: one is to compare the similarity and relatedness of two words, and the other is to gauge the compatibility of several words spoken at once.

The data sets employed in this investigation were made up of a limited sample of automated Quillbot paraphrases of English literary and non-literary texts (standardized

ISSN Online : 2709-4162
ISSN Print : 2709-4154
Volume.6, Issue.4
(2023)
(October- December)

paraphrasing tool). Using the random sample technique, the first two chapters of Jane Austen's novel Emma were utilized for the literary text, and an article from the DAWN newspaper was used for the non-literary material. These texts were rephrased, and the resulting text was examined. The focus was on the parallelism between the original text and the paraphrased version in lexical, semantic, and grammatical terms. Also investigated was the automated paraphrase's incongruity.

4. Analysis and Findings

Tool modifies the text syntactically. Modifications in word order, narration, and grammar are evident from QuillBot's paraphrasing of the literary and the non-literary text. The details are given as follow:

4.1 Syntactic Modifications in the Literary Text

Various syntactic changes have been observed between the paraphrased text and the original text. In certain sentences, just the word order has been changed, while in other sentences the parts of speech have been modified by the tool in the practice of paraphrasing. Sometimes a single or few words of the original text is modified in form or order by the tool. for example, *She had enough resolve to pursue her own desire* is used as a substitute for *She had resolution enough to pursue her own will.* Here, in this sentence, the word *resolve* is substituted for *resolution* and *desire* for *will* while the position of *enough* was also changed.

4.1.1 Word order modification

In paraphrasing, the word order of the literary text sentences has rarely been modified by the tool. The word order was kept almost the same in the paraphrasing but the word order of certain parts of long sentences could be found changed. For example, the sentence "*Sixteen years had Miss Taylor been in Mr. Woodhouse's family, less as a governess than a friend*" has been changed into, "*Miss Taylor had been in Mr. Woodhouse's family for sixteen years, more as a friend than a governess*,".

4.1.2 Grammatical changes and ambiguities

Grammatical ambiguities are found in automated paraphrasing. In the original text, it was said about Mr. Knightley that "after some days' absence" he revisited Hertfield at dinner timing. But, in the automated paraphrasing it was paraphrased grammatically wrong as, "a few days away". Here the word *away* does not fit with the word *days*. Similarly, at another place in the automatic paraphrasing, "she had then only to sit and think" is rewritten as "all she had to do was sit and reflect". In this paraphrasing grammatical modification is evident as the paraphrase lacks proper prepositions. Furthermore, the sentence, "the wedding completed and the bridesmaids gone" is substituted for "The wedding over, and the bride-people gone", instead of bride-people the term bridesmaid is used which carries a different sense the what is used in the original.

4.1.3 Narration changes

In certain places of the automated paraphrase of the literary text, it was observed that the narration of sentences has also been modified. For example, the sentence "Mr. Woodhouse saw the letter, and he says he never saw such a handsome letter in his life." has been changed into "When Mr. Woodhouse saw the letter, he exclaimed, "I've never seen such a beautiful letter in my life". Therefore, the syntax as well as the sense of the sentence has been affected by this change.

4.1.4 Sentence repetition

Repeated sentences were also found in the automated paraphrasing, which indicates that in tool paraphrasing certain sentences could be repeated which affects the quality of tool paraphrasing. An example from automated paraphrasing carried out at QuillBot can be:

ISSN Online : 2709-4162
ISSN Print : 2709-4154
Volume.6, Issue.4
(2023)
(October- December)

"Emma was of no feeble character..." This sentence was not paraphrased by QuillBot and was repeated in QuillBot while the actual first sentence was completely ignored and missed in the paraphrase. The second sentence was repeated in its place, without being paraphrased.

4.1.5 Addition of words and phrases

It was found from the analysis of paraphrased text that various words and phrases were added in various sentences which were not found in the original text. Sentences and the words added into it in the paraphrased text are given in the following table:

Table 1. Added Words in Paraphrased Text

Sentences taken from Paraphrased Text	Added Words	Changes Observed
His father was completely unconcerned about it.	completely	Emphasis
"Look at the carriage!	Look at	Sense
and think of what she had lost.	and reflect on everything she had lost.	Sense

In the original text, there is no emphasis on the point of indifference of Mr Woodhouse but here the element of emphasis was added by adding the word *completely*. In the second sentence, in the original text, it was said with the wonder that 'a carriage' but it was completely wrongly paraphrased as 'look at the carriage'. Similarly, it is said in the original text that 'and think of what she had lost' paraphrased as 'and reflect on everything she had lost', using the word *everything* in place of *what*. By the term *what* is meant in the original text only Miss. Taylor, not everything? Similarly, at another place, Mr. Woodhouse says to Mr. Knightly that dear Emma bears everything nicely, but it was wrongly paraphrased by adding the word *she* after the phrase, dear Emma which changes the entire direction of the utterance: from being said to Mr. Knightly, it seems that he is addressing his daughter Emma directly as it as paraphrased as 'Dear Emma, she bears everything so beautifully'.

4.1.6 Phrase and words omitted

QuillBot removes certain words from the original text in paraphrasing. In the automated paraphrasing of the literary text, several words and particularly phrases of the original text were found missing. The following table summarizes these changes:

Table 2: Phrases and Words Omitted from Paraphrased Text

Original Text	Paraphrased Text	Omission
And *of* moments only of regret.	And moments only of regret.	The preposition has been dropped out which affected the meaning.
Having rather too much her way.	Have *a little* too much of her way.	The word little has been added.
Attended by her pleasant husband *to a carriage of her own.*	Attended by her pleasant husband.	Phrase omitted.
But a few weeks brought some alleviation *to Mr. Woodhouse.*	—but a few weeks brought some relief.	A noun or indirect object dropped out.
What was unwholesome to him *he regarded* as unfit for anybody.	What was unwholesome to him was unfit for anyone.	The narrative phrase *he regarded* been omitted.

From the table, it is evident that the omission of certain words and phrases did not affect the flow, sense and function of the complete sentence, while that of others did changed it.

ISSN Online : 2709-4162
ISSN Print : 2709-4154
Volume.6, Issue.4
(2023)
(October- December)

4.1.7 Sentences left unparaphrased

Many sentences of the literary text are left unparaphrased by the tool, which affects the quality the tool parapharsing. While paraphrasing chapters from Emma through QuillBot it was observed that certain sentences remained as it were in the original text. Moreover, the parts of sentences are also left unparaphrased many times. Example could be given from the text, the paragraph part: *and the shadow of authority…esteeming Miss Taylor's judgement*, is left unparaphrased. Similarly, the sentence "*and many a long October…once more*" is left unparaphrased with only one word was *society* was replaced with *company*.

4.1.8 Syntactically wrong paraphrased sentences

The word order of certain sentences was changed in such a way that it seems not correct though syntactically they are allowable. One such sentence is, *It became desirable to make a complete life change*, which is tool paraphrasing of the sentence, *A complete change of life became desirable.* In the literary text, it was found that certain sentences were completely wrongly paraphrased. For example, it was said in the original text that Miss Taylor loss brought grief to Emma, but the tool paraphrased the sentence as: *Miss Taylor's death was the first to cause anguish.* Emma got separated from Miss Taylor because Miss Taylor got married. But the tool paraphrased it wrongly by taking the cause of separation as the death of Miss Taylor.

4.2 Syntactic Modifications in the Non-Literary Text

The tool also syntactically modified the non-literary text. Certain sentences lacks the words expressed in the original text, while other sentences were added extra words by modifying the original text. The grammatical categories of certain words were found changed. The word order has also been modified.

4.2.1 Grammatical/ Tense modifications in non- literary texts

In automated paraphrasing, the grammar of the original text is often changed by the tool. Sometimes the tense has been changed while at other times just the lexicon is modified.

Table 3. Grammatical/Tense Modifications in Non- literary Texts

Original Text	Paraphrased Text	Tense Changes
Melting of glaciers and the *warming* of seawater	Glaciers *melt* and seawater *warms*.	Present participle form of verb changed into base form
Crop yields are facing growing threats from floods.	Agricultural harvests are increasingly threatened by floods.	Present participle to Past Participle
Pakistan *has* been hit particularly hard by these factors.	These issues have *had* an especially negative impact on Pakistan.	Present perfect to Past perfect
Government has *recognised* the seriousness of this issue.	Administration *recognises* the gravity of the situation.	The present perfect tense has been changed into present simple.
Had been *rising* into gentility.	That had *risen* through the ranks of gentility.	Present participle is changed into past participle.

It is observed in the above table that in the first sentence, present participle form of the verbs *melting* and *warming* has been changed into simple form of verb *melt* and *warms.* In the second sentence, the present participle form of the verbs, *are facing*, has been substituted to Past Participle, *threatened*. Similarly, in the next example it was noted that, '*…has been* hit hard by these factors' was changed into '*had negative impact* on Pakistan'. Thus, changing the Present perfect to Past perfect.

ISSN Online : 2709-4162
ISSN Print : 2709-4154
Volume.6, Issue.4
(2023)
(October- December)

4.2.2 Word order modification

The word order of sentences or clauses have often modified by the tool. For example, the word order of '*Nearly 70 per cent of ...the impact of anthropomorphic activities*' is changed as, '*the impact of human activities can be seen in the fact that about 70% of this rise occurred in the previous three decades*.' Moreover, at another place, the sentence *A large chunk of ... in greenhouse gas emissions* was changed into *Because of increased greenhouse gas emissions, a huge portion of the heat radiated from our planet is being forced back into the ocean rather than rising into space.* The word order of all sentences were not changed, as is evident from the following sentences of automated paraphrase:
Climate change is not only affecting animal life but also human prosperity changed into *Climate change is having an impact not only on animal existence, but also on human prosperity*.

Therefore, the word order of certain non-literary text sentences has also been found changed. In certain places it does not, affect the sense while at other times it does has an impact on the sense of the text.

4.2.3 Insertions

The quality of tool is also affected by its inserting additional elements in paraphrased text. As is evident from the first sentence of paragraph of article's paraphrase that an element of modality has been inserted in the following sentence,
'...its effects *are* all around us', changed into, '...its consequences *can be* found all around us'.

Considering the syntactic relevance and modifications of both the literary and non-literary texts and their alternate tool paraphrasing, it was noted that, the word order was modified and words were substituted but the sense and function of sentence was kept mostly the same. Moreover, comparing tool paraphrasing of literary and non-literary text, it was found that there were less ambiguities and differences in the non-literary text as compared to the literary text. Tense changes were there it was rare. No words or phrases were found omitted in the non-literary paraphrased text. The synonyms use used were mostly authentic and according to the context in the literary text.

Changes in the literary text's syntactic structure include word order, speech parts, additions and deletions of words, etc. Although the paraphrased phrases largely maintained the same word order, some lengthy statements had their word order altered. The computerized paraphrasing also had grammatical problems. The purpose of the sentence was altered and the sense as a whole suffered since some words were replaced with the incorrect words. The way that sentences are narrated has also changed, and this change has an impact on both the sentence's syntax and sense. Additionally, repetitive sentences have been discovered in automatic paraphrase, indicating that the quality of tool paraphrasing may suffer if specific lines are repeatedly used. In certain sentences, words that were absent from the original text were added. When paraphrasing, QuillBot omits several terms from the original text. Several words and phrases from the original text were missing from the computerized paraphrase of the literary text. Several lexicons, in particular original text terms, were missing from the automatic paraphrase of the literary text. While leaving out some words and phrases did not modify the overall structure, meaning, or purpose of the statement, leaving out others did. The technology missed many sentences in the literary text when paraphrasing them. It was discovered that some sentences stayed unchanged when paraphrasing chapters from Emma using QuillBot. Additionally, it happens frequently that sentence fragments are not paraphrased. Though they are acceptable syntactically, certain statements have had their word order altered in a way that makes them appear improper. Some sentences were completely mis-paraphrased by substituting inappropriate words for certain words, such as the word loss being substituted for the word death when the cause of loss is not always death. Additionally, grammatical modifications to the non-literary content

were discovered. The non-literal paraphrased text contained no words or phrases that were omitted. Although the word order and some of the words were changed, the overall concept remained maintained. The literary text was found to have fewer ambiguities and differences than the non-literary material when tool paraphrasing of the two types of texts was compared. The occurrence of tense changes was uncommon. Most of the synonyms used were real and appropriate for the situation.

5. Conclusion

In this study, the original text and QuillBot paraphrased text were analysed and their syntactic differences were compared. The tool paraphrased text's syntactic similarity to and divergence from the source texts are investigated. The literary and non-literary texts is used as sample of data so that the findings could be generalised to tool paraphrasing of all other kinds of literary and non-literary materials. Therefore, tool paraphrasing cannot be assumed to be reliable. Although standard tools like QuillBot's paraphrasing could be depended upon to some extent and should not be followed blindly. Moreover, this study is helpful for paraphrasing tools users as the pros and cons of automatic paraphrasing at syntactic level has been explored in this study for them. Further, this work would be expanded to evaluation of paraphrased text at semantic level too.

References

Callison-Burch, C. (2008, October). Syntactic constraints on paraphrases extracted from parallel corpora. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (pp. 196-205).

Ferreira, R., Cavalcanti, G. D., Freitas, F., Lins, R. D., Simske, S. J., & Riss, M. (2018). Combining sentence similarities measures to identify paraphrases. *Computer Speech & Language*, *47*, 59-73.https://doi.org/10.1016/j.csl.2017.07.002.

Fitria, T. N. (2021). QuillBot as an online tool: Students' alternative in paraphrasing and rewriting of English writing. *Englisia: Journal of Language, Education, and Humanities*, 9(*1*), 183-196. https://doi.org/10.22373/ej.v9i1.10233.

Hirst, G., and St-Onge, D. (1998). Lexical chains as representations of context for the detection and correction of malapropisms. In Fellbaum, C., ed., *WordNet: An Electronic Lexical Database. MIT Press.* Pp.305–332.

Kaur A., (2015). A novel Approach for syntactic similarity between two short text. *International Journal of Scientific & Technology Research 4*(6),

Lesk, M. (1986, June). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings of the 5th annual international conference on Systems documentation* (pp. 24-26).

Litchman, M. (2010). *Understanding and evaluating qualitative educational research*.US: Sage.

Manning, C. D. (2011, February). Part-of-speech tagging from 97% to 100%: is it time for some linguistics?. In *International conference on Intelligent Text Processing and Computational Linguistics* (pp. 171-189). Springer, Berlin, Heidelberg.

Resnik, P. (1999). Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. *Journal of Artificial Intelligence Research*, 11, 95-130.

Stebbins, R. A. (2001) *Exploratory research in the Social Sciences*. Sage University Papers series on Qualitative Research methods, 48. Thousand Oaks, CA: Sage.

Wu, Z., and Palmer, M. 1994. Verb semantics and lexical selection. In 32nd Annual Meeting of the Association for Computational Linguistics, 133–138.