

A Semantic Analysis of the Entailment on the EFL Students' Paraphrased Texts: A Content Analysis

Tika Ruby Indrawati¹, Dian Inayati.², Alies Poetri Lintangari³

{tika.ruby09@gmail.com¹, dianina@ub.ac.id², alieslintang@ub.ac.id³}

Faculty of Cultural Studies, Universitas Brawijaya

Abstract. Analysis on the students' paraphrased text using its semantic relations helps to provide information on their problem or competence in paraphrasing. This study analyzed the entailment on students' paraphrased texts to gain information on their difficulties. Content analysis method was chosen in investigating and interpreting the data based on the context.

This study reveals the 3 types of entailment in students' paraphrased texts. The lack of entailment in the students' paraphrased texts and the absence of negative entailment indicate the students are insufficient in several competences to paraphrase (reading, vocabulary, grammar) and incapable to create the negative sentence. Using Five-Dimensional Paraphrasing Score, the average score are 2.19 in the content aspect, 2.78 in paraphrase quantity, 2.57 in paraphrase quality, 2.31 in language use, and 2.23 in overall quality which implies that they have difficulties in the meaning aspect (content) and the structure of the sentence (language use). This finding shows that the use of semantic or linguistic analysis is going to be beneficial for EFL teaching and learning process in evaluating students' competence or difficulties.

Keywords: paraphrasing skill, paraphrased text, entailment.

1 Introduction

Paraphrasing is a crucial skill for the students in Academic Writing. However, this skill is proven to be challenging for them. The factors affecting students' paraphrasing ability are their limited information about paraphrasing, citing and summarizing [10], low level of reading comprehension, lack of understanding on the subject, and inappropriate chosen references caused by lack of vocabulary [3, 21, 22]. The struggle is particularly increasing in English for Foreign Learners (EFL) Classroom due to low language proficiency of the students [5].

Paraphrasing is mainly focused on the exact meaning of the sentences. However, an exact logical equivalent of meaning is just a narrow interpretation of paraphrasing because there are rarely two similar words that have an exact similar meaning [2]. To understand whether the paraphrased text is linear with the original text, there must be a comparison of meaning and analysis of the relationship between the two propositions. In Semantics, the study of a relation between two propositions falls under the topic of semantic relation. There are several types of semantic relations. However, this study will mainly focus on the relationship of entailment between two propositions. This is on the grounds that the research will cover not only the paraphrase texts which have similar meanings to the original sources but also to analyze the relationship of meaning between the paraphrase text and the original text whether it has similar meaning or not. Thus, this study aims to do a deep analysis of the students' paraphrased text in terms of its semantic relations, mainly entailment as it covers several types of semantic relations, to analyze how similar the meaning between two propositions are. Additionally, by analyzing the similarity of meaning between the two propositions, the researcher seeks to find the quality of students' paraphrasing text or further students' paraphrasing ability.

2 Paraphrased Texts and Paraphrasing Skill

Paraphrased text refers to a version of a written passage from a source text to show that the authors understand the topic given, so they are able to re-state it in their own words to support their argument [6] in which it usually has equivalent textual length as the original using different word choices and structure while also mentioning the original author [19].

In order to identify whether a text is a good paraphrase or not, there are several aspects to assess the paraphrase quality. McCarthy (2009) introduced the term paraphrase quality as a term for the evaluation of 3 aspects which are semantic similarity, syntactic similarity, and lexical similarity. As stated, paraphrasing is the activity to re-state a sentence with a similar meaning, but in different semantic and lexical aspects [13]. In paraphrasing, the focus is on changes in word use and syntactic structure while maintaining the similarity of meaning between the original and paraphrased text [15]. In summary, the paraphrased text serves similar meaning text while produced by different lexical terms and/or variations of grammatical structure.

Meanwhile, Paraphrasing skill is the skill that students need to master in order to provide a good academic writing paper. In paraphrasing, students need to understand what the author stated before stating it in their own words. Paraphrasing skill could be stated as a set of skills which related to one another, leading to the ability to understand a text whether orally or in writing, and showing their comprehension using their own words to create ideas [23]. Paraphrasing skill is important in academic writing as it can demonstrate whether the author can grasp the ideas of what they have read, understand it, and use that information to support their own ideas [17]. Admittedly, paraphrasing skill is a skill that students must have in Academic Writing.

The term of paraphrasing skill mainly focused on the ability to deliver ideas using their own words, but it also covers a skillset that students should have (reading, writing, listening, and speaking). Thus, paraphrasing skill can also be an indication of students' competence because

the students' attempts at paraphrasing can give information on "how well students read as well write" [17]. Additionally, those who had a higher score in paraphrasing tended to use complex paraphrasing techniques, reformulate the paraphrase using the author's terms, and make use of condensation, implication, generalization, and particularization in meaning [14]. On the contrary, inferior paraphrasing often uses simple methods in which some words were replaced by synonyms found in online dictionaries or software to process words [17]. In summary, students' techniques in paraphrasing can illustrate their capabilities and comprehension.

3 Semantic Relations and Entailment

In order to know the meaning of any lexemes in a language, it needs to recognize the lexemes that possess any semantic relation, for example, father and mother, father and son; father and paternal; employer and employee; big and large; big and little; red, yellow and blue. Each of the examples above serves a different relationship [11]. The lexemes father and mother are related generally, but they do not share any relation in genetics. On the other hand, the lexemes father and son share both general and biological relations. The relation between meanings or concepts is under the study of semantics relation [9].

There are several categories: lexical, phrasal, and sentential relations [7]. Entailment is one of the semantic relations of sentential relations. It deals with truth-semantic relations between two propositions. Entailment is the main inference in Semantics because in the process of doing semantic description, entailments between sentences are the evidence for sense relations between words [8].

In entailment, a text can entail another text without any regard for the completeness of meaning [13]. For example, the target sentence "A glacier's own weight plays a critical role in the movement of the glacier" is entailed by "The glacier's weight is an important role in the glacier". The entailment of the target sentence could be done partly, in contrast, or even mutually as long as the sentence delivers a truth conditioning (if the target sentence is true, then the entailed one should also be true).

4 Method

This research used a content analysis method to provide information from students' paraphrased texts. This method was chosen to make effective inferences about the context of the text provided (or other meaningful things) and to gain insight into the students' writing product as the purpose of this study was to identify the pattern frequencies of entailment that exist in students' paraphrases. This approach helped to be able to investigate and interpret within the context of the courses and skills needed in terms of relevance and non-relevant issues. The subjects of this study were 53 of the fourth semester students of English Language Education Study Program who were taking Academic Writing Course in Universitas Brawijaya.

Data Collection process was using Paraphrase Test which was adopted from a paraphrasing exercise in University of Georgia open courses. There were 3 prompts that the students needed to answer. These answers were later evaluated and analyzed in data analysis process. The data

analysis method used 2 instruments: Five-Dimensional Paraphrase Scoring Rubric and the worksheet for analyzing the types of entailment along with the guidelines. In this research, there were four major steps in analyzing the data. The first step is de-contextualization. The next step is recontextualization in which the researcher reread and rechecked whether all the aims had been covered by the coding list. The third step was categorization or coding the data. Another process in the third process of categorization was to score the students' paraphrase text based on the Five-Dimensional Scoring Rubric of Paraphrase. This scoring rubric was used as a tool to assess the originality and the other aspects of students' paraphrases. Lastly, the compilation was the process of analyzing and writing the data.

In analyzing the data, there were other investigator for inter-rater analysis in order to assess the reliability of the study. The result of the inter-rater analysis was compared and calculated using Fleiss Kappa method which resulted in substantial agreement based on the interpretation table by Landis and Koch (1977).

4 Analysis of Students Paraphrased

4.1 Entailment on Students' Paraphrased Texts

The students' paraphrased text was categorized using a data sheet into 5 types of relations to comprehend the sense relation of meaning between the original sentence and the paraphrased text. The 4 types of entailment were referred to as two-way entailment, one-way entailment, negative entailment, and other relation of entailment. If there was no sense of relation to the meaning of the text, then the paraphrased text would be defined as "no relation of entailment".

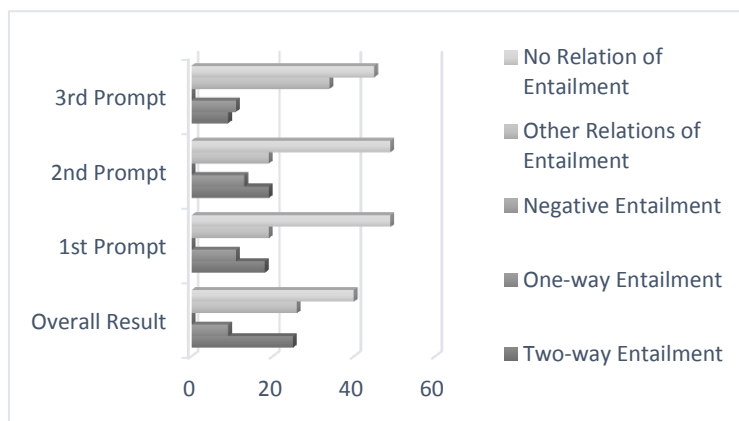


Fig. 1. Types of Entailment in Students' Paraphrased Texts.

The highest percentage of the types of entailment was no relation to entailment. From 159 paraphrased texts, this type of entailment was found in 70 of them (45%) which was mostly caused by syntactical errors that obscured the meaning of the sentences. In either 1st, 2nd or 3rd prompt, the highest percentage was no relation of entailment between propositions. The cause was the misinterpretation by the student which rendered them created similar meanings. This type of text "Invented or Gist" in his Taxonomy of Paraphrase Types [10]. This type of

text referred to a text which did not have any strings of words similar to the original but also could not be called a paraphrase because it only contained the author's ideas. This could be said as a response to the original text but failed to re-state it in a similar meaning. There was no negative entailment (0%) found in students' paraphrased texts which implied that there was no usage of negative synonyms by the students, and that the student did not have the ability to do so. The second highest in every prompt was the ' Other Relation of Entailment ' which accumulated 23% of the data. This type of entailment was mostly caused by the implication of mistranslation of the students' 1st language (Indonesian). The "other relation" in entailment defined sentences, phrases, or words that are somehow related, usually through misalignment or polysemy [16]. It could be also somewhat a response to the original text by adding or omitting words related to the original text [4]. In this type of entailment, the paraphrased text would be considered to be inaccurate because it didn't deliver the main ideas of the original text even though they comprehended the meaning of the original text. Although paraphrased text did not have to be similar word for word, replacing the original words with irrelevant words would create an original sentence and not a paraphrased text.

Then, from 159 entries, two-way entailment was found in 27 of them, which meant that only 17% of the paraphrased texts had similar meanings to the original text semantically and used the correct syntax. On two-way entailment, both sentences entail each other [8]. Therefore, two-way entailment refers to paraphrased text which had similar meanings to the original text while using their own words. Although two-way entailment could be an indication for a good paraphrase, there were also cases where two-way entailment text could be directed as an exact copy of the original text, thus creating plagiarism and failing the purpose of paraphrasing. Lastly, one-way entailment was found in 18 data of texts (11%). One-way entailment referred to the one-way direction of entailment. In this case, the paraphrased text entailed the original text. One way to identify this type of entailment was to find out the relation of hyponymy on both propositions. Another way to know it was to use a scale of words based on its informativeness, for example, excellent > good > OK [8]. Another thing to focus on was that in one-way entailment, the truth semantic relation was in one direction because it did not deliver the whole meaning of the sentence. In this case, the most common occurrence was caused by incompleteness to deliver the whole meaning of the sentence. This resulted in unequal meaning from the target sentence to the paraphrased texts. However, the paraphrased texts are still able to portray several main points of the target sentence

4.2 Students' Problems in Paraphrasing

The students' paraphrased texts were also assessed using Five-Dimensional Scoring Rubric for Paraphrase by Yamanishi, Ono, and Hijikata (2019). This scoring rubric was specifically chosen because it could analyze the paraphrase analytically (for each aspect) and holistically (as an overall assessment). There were 5 dimensions described content (meaning related), the paraphrase quantity (percentage of originality), the paraphrase quality (attempt to use different vocabulary and different structure of the sentence), language use (grammar and structure of the sentences), an overall assessment of the paraphrase. The range of scores was divided into 4 categories: 4 for "Very Good", 3 for "Good", 2 for "Fair", and 1 for "Poor".

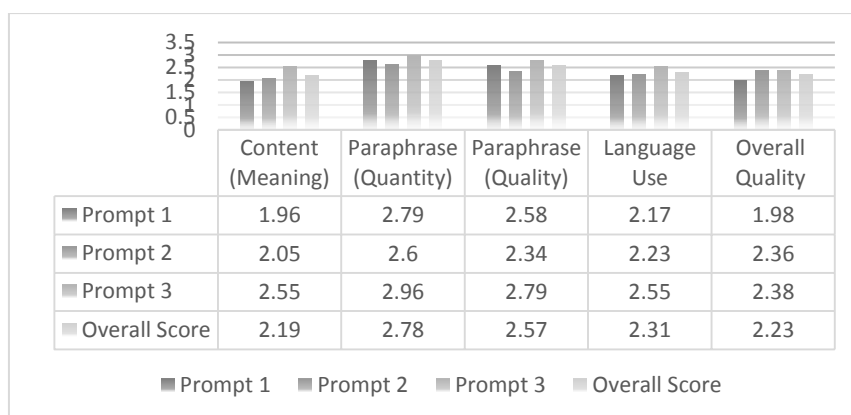


Fig. 2. Students' Paraphrase Score.

This scoring rubric had a different approach as several aspects should be presented separately. Each dimension should be analyzed in its part as the overall assessment of the paraphrase was given as a part of the dimension.

The calculation of the average paraphrased texts of 159 data initiated the final results in which similar to the results of prompt 1-3. The lowest average score was on the content dimension (2.19). Thus, this implied that the students in either prompt 1, 2 or 3 had problem regarding comprehending and delivering the meaning of the original text. The second lowest was on the overall quality of students' paraphrased texts with the average score of 2.23. Based on the Five-Dimensional Paraphrase Scoring Rubric, the students' paraphrase texts were considered poor in quality. Next was the dimension of language use with the average score of 2.31 which was caused by lack of knowledge in vocabulary and grammatical structure. Lastly, the dimension paraphrase quality (2.57) and paraphrase quantity (2.78) had the highest average. Even though it was still considered to be poor, however it was reaching the score of good quality. It could be concluded that the students actively made attempts in paraphrasing using their own words.

5 Discussion

5.1 Students' Competence based on the Types of Entailment

There was a strong relationship between reading comprehension (ability to understand the text) and the ability to paraphrase [23]. Further, based on the list of 20 paraphrase challenges for EFL students, the 3 biggest challenges were: vocabulary skill (did not know the words of the original text and inability to use appropriate word choice), grammar/syntax/structure, and reading comprehension (inability to understand the meaning of the text) [23]. Based on the research findings of this study, it was also found that most students' paraphrased texts did not have any entailment which indicated that most students' paraphrased texts did not have similarities to the original text caused by misunderstanding of the meaning of the text. A lot of the EFL students did text appropriation in which they resorted to cutting and pasting the text in different structures which was caused by the inability to comprehend the meaning of the

original text. It could be interpreted that the students' lack of reading comprehension skills also affected the students' ability in paraphrasing. Moreover, there is a positive correlation between reading comprehension and paraphrasing skills and the increasing ability of reading comprehension will increase the paraphrasing skills [1]. The misunderstanding of the content of the text was often caused by the lack of knowledge of sentence structure, referring to the comprehension of words, phrases, and clauses [1].

Additionally, vocabulary skills and syntactic skills are essential skills in paraphrase writing [23]. While the use of synonyms is the common use for paraphrasing, there is also the need to use appropriate word choices based on the context of the sentence. The incorrect word choices could alter the whole meaning of the sentence. This difficulty in choosing the appropriate words along with the grammatical errors were also the challenges that he mentioned as the challenges in paraphrasing for EFL students. All these skills (reading comprehension, ability to choose a correct word choice, and ability) are important in Academic Writing as it is a final step in creating a writing product.

Several paraphrased texts had two-way entailment, indicating having similar meanings. From the students' paraphrased texts, it could be found that the students were using a lot of paraphrasing techniques, showing that the students had capabilities to understand the texts and used proper word choices and sentence structures that were suitable for the context.). There were also several complex paraphrasing techniques such as condensation, implication, generalization, and particularization of meaning [14]. Other methods of paraphrasing were using synonyms, changing the structure of the sentence (active to passive or vice versa), changing the order of the sentence, changing the word form, combining sentences, changing clause to phrase or vice versa, etc. However, only using synonyms would create an inferior paraphrased text as it often reflected that the authors didn't truly comprehend the original text [17].

Further, there was no found negative entailment in the students' paraphrased texts, indicating that the students could not do so. In psycholinguistics, processing negative sentences was rather difficult compared to a positive sentence [24]. The difficulty originated from the extra step of implanting the 'non' prefix. Thus, creating a negative sentence was considered a later stage in the construction of a sentence. As none of the students were able to create a negative entailment in paraphrasing, it could be derived that the students still did not have the competence to do so.

5.2 Students' Competence based on the Students' Paraphrase Score

The 5 aspects of scoring rubrics were focused on the important aspects of paraphrasing such as semantics, vocabulary mastery, grammatical knowledge, and originality. However, it was found that the students were lacking in the semantic aspects and the language uses. The problems were the students' low reading comprehension along with the insufficient mastery of vocabulary and grammar. The cause for paraphrasing difficulties in EFL students was the low language competence [5]. In contrast, the originality of the students' paraphrased texts showed that rather than plagiarizing, the students rather attempted in paraphrasing even though several errors were causing it to have different meanings. This underlined the purpose of paraphrasing which was to avoid plagiarism [18]. This might be caused by a specific case as this study

focused on the result from the paraphrase test. For more certain details, it needs more analysis on the general students' writing product.

6 Conclusion

This study discussed the mistakes made by the EFL students in paraphrasing and several aspects which caused it. In addition, this study showed that the use of semantic analysis could benefit in EFL teaching and learning process, specifically in analyzing the students' difficulties and their competence. Due to the limitations of the study, it was not found whether the students know the importance of paraphrasing. However, the students were proven to know the basic techniques in paraphrasing.

References

- [1] Arifah, U. P.: A Correlational Study Between Reading Comprehension, Sentence Structure Mastery and Paraphrasing Skill. *English Education Journal* (January 2017) 5 Vol. 2, pp. 61–67 (2017)
- [2] Bhagat, R. and Hovy, E. H.: What is a Paraphrase? *Journal of Computational Linguistics*, 39 (3), pp. 463-37 (2013)
- [3] Bloch, J.: Plagiarism, intellectual property and the teaching of L2 writing. New York, NY: *Multilingual Matters* (2012)
- [4] Boguslavsky, I., Timoshenko, S., & Iomdin, L.: *Semantic Paraphrasing for Information Retrieval and Extraction* (2009)
- [5] Davis, M. & Carroll, J.: Formative feedback within plagiarism education: Is there a role for text-matching software? *International Journal for Educational Integrity*, 5(2), pp. 58-70 (2009)
- [6] Esain, A. E.: *English for Academic Purposes: The challenge of paraphrasing*. (2015)
- [7] Fromkin, Victoria, Rodman, R, and Hyams, N.: *An Introduction to Language* (9th ed.). Boston: Thomson Wadsworth. (2003)
- [8] Griffiths, P.: *An Introduction to English Semantics and Pragmatics* (Edinburg Textbooks on the English Language). Edinburg: Edinburg University (2006)
- [9] Hjørland, B.: Semantics and knowledge organization. *Annual Review of Information Science and Technology*, 41(1), pp. 367-405 (2007)
- [10] Keck, C.: Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practices. *Journal of Second Language Writing*, 25, pp. 4-22 (2014)
- [11] Kreidler, W. C.: *Introducing English Semantics*. London: Routledge (1998)
- [12] Landis, J. R. & Koch, G. G.: An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers. *Biometrics*, 33(2), pp. 363-374 (1977)
- [13] McCarthy, P. M., Rebekah H. G., & McNamara, D. S.: The components of paraphrase evaluations. *Behavior Research Methods*, 41 (3), pp. .682-690 (2009)
- [14] Milićević, J. and Tsedryk, A.: *Assessing and Improving Paraphrasing Competence in FSL* (2011)
- [15] Na, C. D. and Mai, N. X. N. C.: Paraphrasing in Academic Writing: a case study of Vietnamese learners of English. *Language Education in Asia*, 8 (1), pp. 9-25 (2017)
- [16] Pavlick, E., Bos, J., Nissim, M., Beller, C., Durme, B. V., & Callison-Burch, C.: *Adding Semantics to Data-Driven Paraphrasing* (2015)

- [17] Rogerson, A.M., & McCarthy, G: Using Internet based paraphrasing tools: Original work, patchwriting or facilitated plagiarism? *Int Journal for Education Integrity*, Vol. 13, pp. 2 (2017)
- [18] Roig, M.: Plagiarism and Paraphrasing Criteria of College and University Professors. *Ethics & Behavior*, 11 (3), pp. 307-323 (2001)
- [19] Roig, M.: Avoiding plagiarism, self-plagiarism, and other questionable writing practices: A guide to ethical writing. Retrieved from U.S. Department of Health and Human Services (2003)
- [20] Sadrzadeh, M., Kartsaklis, D. & Balkir, E.: Sentence entailment in compositional distributional semantics. *Annals of Mathematics and Artificial Intelligence*, Vol. 82, pp. 189-218 (2018)
- [21] Shi, L.: Rewriting and paraphrasing source texts in second language writing. *Journal of Second Language Writing*, Vol. 21, pp. 134-148 (2012)
- [22] Shi, L. and Dong, Y.: Chinese graduate students paraphrasing in English and Chinese contexts. *Journal of English for Academic Purposes*, Vol. 34, pp. 46-56 (2018)
- [23] Thadphoothon, J.: Students' Perceptions of Paraphrasing Skills and Challenges. *The 4th International Conference on Innovative Education and Technology 2019*, Vol. 4, pp. 274 (2019)
- [24] Tian, Y., Breheny, R.: Dynamic Pragmatic View of Negation Processing. In: Larrivé, P., Lee, C. (eds) *Negation and Polarity: Experimental Perspectives*. Language, Cognition, and Mind, vol 1. Springer, Cham (2016)
- [25] Yamanishi, H., Ono, M. & Hijikata, Y.: Developing a scoring rubric for L2 summary writing: a hybrid approach combining analytic and holistic assessment. *Language Testing in Asia*, Vol. 9, pp. 13 (2019)