ERROR ANALYSIS IN ENGLISH-INDONESIAN MACHINE TRANSLATE

Halimah
Universitas Suryakancana
halimah@unsur.ac.id

Abstract

Since the advent of the 21st century, there have been a lot of developments and new technologies have been introduced which have made life more convenient and simple. Although not appropriate for all situations, machine translation (MT) is now being used by many translators to ease their work. Many others use MT to get a quick grasp of foreign text that they would not understand. The quality of Google Translate depends on the number of human translated texts searched by Google Translate. Therefore, the quality of the translation has been considered far from perfection. Thus in order to evaluate the quality of machine translation, error analysis has been suggested to be conducted. This paper presents the results of a research study focusing on the types of Google translation errors found in the English translation of procedural text. The purposes of this paper are (i) the results of English-Indonesian machine translation, categorizing errors in machine translation into 3 types: semantic errors, syntax errors and morphology errors, and (ii) to describe the dominant kind of translation error produced by Google Translate. This study revealed that The most frequently occurring errors were form category of sematic (i.e., 44 errors out of 97 or 45.36%). Syntax errors ranked second (i.e. 34 errors out of 97 or 35.05%) and morphology errors ranked third (i.e., 19 errors out of 97 or 19.59%).

Keywords: machine-translate, error analysis, English, Indonesian

1. Introduction

Carrying out translation is not an easy task; it is a complicated skill (Wongranu, 2017); it is a combination of art and skill (Yousofi, 2014). There are many aspects should be considered in doing translation. One of them is finding the equivalency of word from the source language toward the target language (Halimah, 2018). A successful translator should enjoy a good amount of knowledge in linguistics, sociolinguistics, and other fields which relate to our human life (Yousofi, 2014).

Information technology is developing very rapidly in this century, one of them is internet. Initially the Internet is an information technology that is only utilized by the military in America and newly used for public interest in the 80s. Since then, the internet has penetrated all over the world and into many aspects of human life. By using the internet everyone can search and get the information he needs quickly without having to leave his seat. Formerly more communication is done directly (face to face), whereas now personal contacts face to face slowly has been replaced with the ease of communicating via the internet and cell phones. The Internet is a collection or network of computers that exist around the world. In this case the formerly independent computer can deal directly with the host or other computers. With its ability to connect one computer to another, making the Internet can be used to access or transfer data / information from one computer to another computer.

With the advent of the Internet in the 1990s, and the commensurate rapid growth of information and communication technology, translation has taken a further step towards providing more informed and reliable products for the client. Translator education and training which, not long ago, were to a large extent predicated upon teacher resourcefulness and the (un)availability of parallel and similar

texts, are today almost impossible without resort to information and communication technology, with freely available online web tools and services becoming an ever more significant element in contemporary classroom resources (Korošec, 2013).

Ghasemi & Hashemian (2016) stated that "Google Translate is a provided service to translate different written texts from one language to another and it provides translating 90 languages". It can translate not only a word, but also a phrase, a section of a text, or a Web page. To translate a text, Google Translate different documentaries to find the best appropriate translation pattern between translated texts by human. This pattern machine is called MT. Machine translation (MT) whose aim is to use software in order to translate texts is a subgroup of computational linguistics (Sapar, Ridhuan, & Abdullah, 2018). Although not appropriate for all situations, machine translation (MT) is now being used by many translators to aid their work. Many others use MT to get a quick grasp of foreign text from email, Web pages, or other computer-based material which they would not otherwise understand (Aiken & Balan, 2011). Consequently, the quality of Google Translate depends on the number of human translated texts searched by Google Translate. Translating results from translator machines need to be studied further, especially to see the error of the existing language, because many parties who doubt the quality of its translation.

The use of Google Translate has been increasing either in the academic discipline or in the non-academic discipline. Despite the fast-turnaround time produced by Machine Translation such as Google Translate, the quality of the translation has been considered far from perfection (Putri & Ardi, 2015). Further, they sated that regarding to translation process, Google Translate does not apply grammatical rules because its algorithms are based on statistical analysis rather than traditional rule-based analysis. Thus in order to evaluate the quality of machine translation, error analysis has been suggested to be conducted (Napitupulu, 2017; Fang et al., 2011).

This paper presents the results of a research study focusing on the types of Google translation errors found in the English translation of procedural text. The purposes of this paper are (i) to report the results of English-Indonesian machine translation, categorizing errors in machine translation into 3 types: semantic, syntax, and morphology, and (ii) to describe the dominant kind of translation error produced by Google Translate.

1.1 Error Analysis

Errors are considered by many educators to be an integral part of the teaching – learning process (Aqel & Mohammed, 2017). Errors in simple words are the problematic aspects of learners (Kafipour & Jahanshahi, 2015). Error analysis is a field of study that enters the umbrella of applied linguistics. This review is not new for language teachers, since the results of error analysis are used to improve the language learning process, either to correct mistakes made by learners or to help teachers develop appropriate learning strategies. Through analysis one can detect the problems of a translator in broader sense. They also can reveal the degrees of error and the nature of errors. Another benefit is that the patterns of error can be cleared (Kafipour & Jahanshahi, 2015). Error analysis to identify the common errors and focusing the correction on those errors (Hamzah, 2012). In translation, identifying error during the process of translation is very crucial to do (van der Wees, Bisazza, & Monz, 2015), since it can improve the quality of translation resul

1.2 Translation error

Numbers of researches have been conducted on errors in translation. Among of them are Napitupulu (2017) found that there are five types of error classification is used as the parameters, namely lexico-semantic error, tense error, preposition error, word order error, distribution and use of verb group error, and active and passive voice error. Elmahdi (2015) discovered the types of errors in use of articles (omission of articles, redundant, or wrong use of articles). Uba (2015) found two kinds of translation error; interlingua and intralingua. Wongranu (2017) revealed three kinds of translation errors; semantic errors, syntactic errors, and miscellaneous errors. Yousofi (2014) found that translators' had committed errors in linguistic, cultural and stylistic areas. Kafipour & Jahanshahi (2015) found the most error made in translation was register. Ardeshiri & Zarafshan (2014) found pragmatic error in translation. Aqel & Mohammed (2017) discovered that spelling error occurred in translation. Faisyal (2015) reported that two types of errors; morphological and syntactic errors. Fang, Ge, & Song (2011) found tree types of error made by machine translation: incorrect lexical choices, structural errors and component omissions.

From the finding above it can be simplified that there are 23 types of translation errors that can be used as the parameters in analyzing error in translation; lexico-semantic error, tense error, preposition error, word order error, distribution and use of verb group error, active and passive voice error, omission of articles, redundant, wrong use of articles, interlingua and intralingua, semantic errors, syntactic errors, and miscellaneous errors, linguistic, cultural, stylistic, register, pragmatic error, and morphological error .

Yet, in this study, the researcher limited the types of error occurred in translation of Google Translate into three types only; sematic, syntactic, and morphological errors.

1.3 Machine Translate

Google at the beginning of its discovery is a tool or machine that helps internet explorers to quickly find the information or websites they are looking for. This search engine is very useful considering the number of web pages in the virtual world can be millions in number, while the human brain's ability to remember the address of a web page is very limited. In 2007, *Google* introduced *Google Translate* (*GT*), a statistical machine translation (MT) platform that currently provides automated translations, directly or via a pivot language, between over 50 languages. Slovene was added to the list of Google-supported languages in September 2008. GT's success is to a large extent predicated on its statistical approach, which has proven to produce better results than the previously supported rule-based linguistic systems, most known among which is *Systran*, that for the most part retrieve data from bilingual dictionaries and grammars which are then supplemented by linguistic and other rules (Korošec, 2013).

In its development, Google is not only a search engine that helps internet users find links to a web page, but also provides a translator engine. Being an important member of the "Google family", Google Translate is probably one of the easiest and most accessible tools to help users meet their translation needs (Medvedev, 2016). The translator engine attached to Google will automatically help translate a text or web page from one language to another, so that the reader helps when trying to understand the contents of a web page.

2. Method

2.1 Material

This research was conducted by using descriptive design with qualitative approach- it was done with consideration that the purpose of this research to find out the lexical errors in translating English text into Indonesian using Google Translate.

The data for the study was Text taken from an online site on May 21, 2016 (https://www.huffingtonpost.com/caitlin-barry/teaching-film-in-a-high-_b_1307408. Then the text was translated by an existing translator engine on Google into Indonesian. Sample analysis in this research is ten sentences of procedural contained in procedure text. Furthermore, each of these procedural sentence is decapitated into the phrase or word specified by the researcher.

2.2 Procedure

In determining and analyzing students' errors, the researcher followed Sridhar (1975) who proposed a methodology of EA consisting of the following steps 1) collection of data, 2) identification of errors, 3) classification into error type, and 4) statement of relative frequency of error types.

2.3 Data Analysis

To conduct the analysis, the researcher classifies the error according to language components: morphology, semantic, and syntax. Morphology is scientific study about word structure and formation rules. Morphology, the study of forms, is the branch of linguistics that deals with the internal structure of complex words (Aronoff, 2013). morphology is the study of morphemes and their arrangements in the word formation (Jatnika, Suganda, & Sobarna, 2014). Morphology is understood to involve generalizations about form and meaning that relate words to one another within a language (Inkelas, 2008).

Semantics is the study of the meaning of words, phrases and sentences (Wongranu, 2017; Kreidler, 2010). In semantic analysis, there is always an attempt to focus on what the words conventionally mean, rather than on what a speaker might want the words to mean on a particular occasion. This technical approach to meaning emphasizes the objective and the general. It avoids the subjective and the local. Semantic deals with the conventional meaning conveyed by the use of words and sentences of a language. Syntax is the part of linguistics that studies sentence structure (Hana, 2011)

3. Results and Discussion

Text taken from an online site on May 21, 2016. Then the text is translated by an existing translator engine on Google into Indonesian. Sample analysis in this research is ten sentences of procedural contained in procedure text. Furthermore, each of these procedural sentences is decapitated into the phrase or word specified by the researcher. The findings can be presented in the following table.

Table 1. Recapitulation of Translation Text Translation Data of Google Translator Machine Translation.

Data	Types of Errors			Total
	Morphology	Semantic	Syntax	
1	2	2	4	
2	3	5	3	
3	2	8	3	
4	2	0	0	
5	4	6	4	
6	4	7	4	
7	1	4	7	
8	0	4	4	

9	1	7	4	
10	0	1	1	
	19	44	34	97

Based on the table above, there are 19 errors in the morphological aspects. Such errors are commonly found in the

plural-s that denotes plural. In English there are affixes -s, but they are not in Indonesian. For example "students" has the meaning of "pupils", whereas Google Translate is translated as "pupil".

Semantic error appears in the text 44 times. The ten phrases are: "To prepare", where the Google translator translates the phrase "To prepare" to "prepare". "Preparing" the word is more correctly translated into "preparation". Similarly, the word "Take special note" in the third sentence is more accurately interpreted as "note the parts", so the more suitable match is "note the parts" instead of "Take special note". Examples of other errors can be seen in the paragraph step to the first 2 sentences. Here we find phrases translated as "where necessary", a more precise translation is "needed". . Of the three examples it appears that Google's translation engine difficulties in translating words that refer to the content.

A very prominent error is a syntax error, because the Google translator engine may not be able to produce the correct sentence according to the rules of Indonesian grammar. In the paragraph the third step "Prior to starting the film, plan an" empathy-building "or" connection-making "activity for the students that can start them thinking about the larger issues or themes in the film, and how they relate to their own lives "and translates to" Before starting the movie, plan an "empathy-building" or "connection-making" activity for your students that can start them thinking about bigger issues or themes in the movie, and how they relate to life themselves". A more accurate form of sentence translation is: "Initial activities to begin the film, prepare" build a sense of spirit ", or" relationship making ", for your students who can start them to think about wider issues or themes in film ... . "

Having found out the frequency of each category and subcategory, in the next step, the most frequently occurred errors are identified.

Table 2: Total frequency Errors

Content of Translation Errors	Numbers	Percentage
Morphology	19	19.59%
Semantic	44	45.36%
Syntax	34	35.05%

As shown in table 2, the most frequently occurring errors were form category of sematic (i.e., 44 errors out of 97 or 45.36%). Syntax errors ranked second (i.e. 34 errors out of 97 or 35.05%) and morphology errors ranked third (i.e., 19 errors out of 97 or 19.59%). To understand the results better, see figure 1.

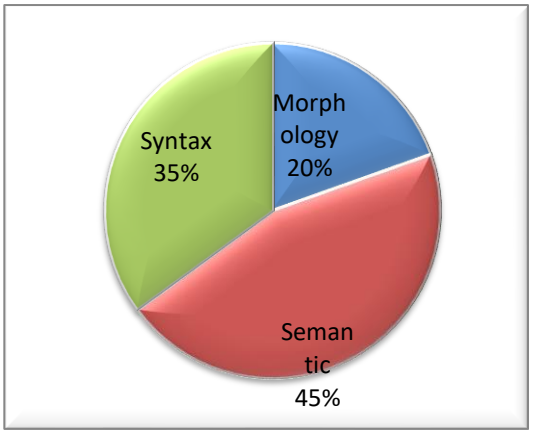

Figure 1. The Percentage of Translation Errors

Based on the studies that have been done, it is clear that the output of Google's translation engine contains many errors. Putri & Ardi (2015) said that “Despite the

fast-turnaround time produced by Machine Translation such as Google Translate, the quality of the translation has been considered far from perfection". In line with, (Napitupulu, 2017) sated that regarding to translation process, Google Translate does not apply grammatical rules because its algorithms are based on statistical analysis rather than traditional rule-based analysis.

4. Closing

From a quick overview of English translations to Indonesian translations by a Google translator engine it appears that this machine translates words by word, the context of a sentence is often overlooked. This aspect is a major drawback of the translation obtained through Google's translator engine. Therefore users who want to translate a text must make improvements to the text of the translation.

In today's globalized world the Internet plays a very important role, in helping humans to find and provide the necessary information without being constrained by time and place. Web pages that are available until now the number could be millions, so that internet users need a tool in the form of a search engine that can find an information along with its web address quickly. One of the most popular search engines is Google.

From the analysis of translation of Google translate from English into Indonesian language found that most errors occurred at the level of semantic, then followed by errors in semantic syntax, and morphology. Translations from a machine of this kind translator still has many shortcomings and does not produce translations which is accurate, so that the translation is more worthy to be referred to as pre-translation that still needs to be perfected by its users. Nevertheless, the translation of Google Translate can be used to understand a text globally.

5. References

Aiken, M., & Balan, S. (2011). An analysis of Google Translate accuracy. *Translation Journal*, *16*(2), 1–3. https://doi.org/1536-7207

Aqel, K. A. J. D., & Mohammed, F. (2017). A LONGITUDINAL ANALYSIS STUDY OF WRITING ERRORS MADE BY EFL. *British Journal of Education Vol.5*, *5*(13), 127–145. Retrieved from http://www.eajournals.org/journals/british-journal-of-education-bje/vol-5-issue-13-december-2017/longitudinal-analysis-study-writing-errors-made-efl-students-al-quds-open-university-qou-case-language-use-course/

Ardeshiri, M., & Zarafshan, M. (2014). Students' Causes of Errors in Translating Pragmatic Senses. *International Journal of English and Education*, *3*(4), 238–253. Retrieved from www.ijee.org

Aronoff, M. (2013). Morphology - Linguistics - Oxford Bibliographies. *Linguistic*. Retrieved from http://www.oxfordbibliographies.com/view/document/obo-9780199772810/obo-9780199772810-0001.xml#firstMatch

Elmahdi, O. E. H. (2015). Sudanese EFL Learners' Sources of Errors In The Production of Articles. *British Journal of English Linguistics*, *3*(4), 25–32.

Faisyal, R. (2015). *Morphological and syntactic errors found in english composition written by the students of daarut taqwa islamic boarding school klaten*. Universitas Muhammadiyah Surakarta. Retrieved from https://www.google.com/search?q=morphologycal+error+in+translation%2C+pdf&ie=utf-8&oe=utf-8&client=firefox-b

Fang, F., Ge, S., & Song, R. (2011). Error Analysis of English-Chinese Machine Translation Categorization of Errors in Machine Translation. *Springer-Verlag Berlin Heidelberg 2011*. Retrieved from https://pdfs.semanticscholar.org/0dcb/236b168398785123fe44cc36ab7df6327419.pdf%0A%0A

Ghasemi, H., & Hashemian, M. (2016). A Comparative Study of Google Translate Translations: An Error Analysis of English-to-Persian and Persian-to-English Translations.

English Language Teaching, *9*(3), 13. https://doi.org/10.5539/elt.v9n3p13

Halimah. (2018). Comparison of Human Translation with Google Translation of Imperative Sentences In Procedures Text. *BAHTERA: Jurnal Pendidikan Bahasa Dan Sastra*, *17*(1), 11–29. Retrieved from http://journal.unj.ac.id/unj/index.php/bahtera/

Hamzah. (2012). An Analysis of The Written Grammatical Errors Produced by Freshment Students in English Writing, *Volume 6 N*, 17–25. Retrieved from http://ejournal.unp.ac.id/index.php/linguadidaktika/article/viewFile/3127/2627

Hana, J. (2011). Syntax 1. In *Intro to Linguistics* (pp. 1–14). Retrieved from https://www.google.com/search?client=firefox-b&biw=1366&bih=654&ei=uBMqW-HYCYr7vgTw4rWoBA&q=syntax+is%2C+pdf&oq=syntax+is%2C+pdf&gs_l=psy-ab.3..0i8i7i30k1l3j0i30k1j0i8i30k1l5j0i8i7i30k1.39971.45328.0.46484.11.8.0.0.0.0.187.973.3j5.8.0....0...1c.1.64.psy-ab..6.5.706...35i39k1j0i7i30k1j0i67k1.0.Ft6EeYAxRN8

Inkelas, S. (2008). The Morphology-Phonology Connection SHARON INKELAS. *Linguistics Society and*, *3564*(January), 1–18. Retrieved from https://journals.linguisticsociety.org/proceedings/index.php/BLS/article/

Jatnika, A. W., Suganda, D., & Sobarna, C. (2014). â€™ Typical â€™ Morphology System of Language Advertising of Cellular Services. *International Journal of Linguistics*, *6*(3), 181–197. https://doi.org/10.5296/ijl.v6i3.5838

Kafipour, R., & Jahanshahi, M. (2015). Error Analysis of English Translation of Islamic Texts by Iranian Translators. *Journal of Applied*

Linguistics and Language Research, *2*(3), 238–252.

Korošec, M. K. (2013). The Internet, Google Translate and Google Translator Toolkit. *Tralogy*. Retrieved from http://lodel.irevues.inist.fr/tralogy/index.php?id=113

Kreidler, C. W. (2010). *Introducing English Semantics*. *English*. London and New York: Routledge. https://doi.org/10.4324/9780203265574

Medvedev, G. (2016). Google Translate in Teaching English. *Journal of Teaching English for Specific and Academic Purposes*, *4*(1), 181–193. Retrieved from http://espeap.junis.ni.ac.rs/index.php/espeap/article/view/318/221

Napitupulu, S. (2017). ANALYZING INDONESIAN-ENGLISH ABSTRACTS TRANSLATION IN VIEW OF TRANSLATION ERRORS BY GOOGLE TRANSLATE. *International Journal of English Language and Linguistics Research*, *5*(2), 15–23. Retrieved from

http://www.eajournals.org/journals/international-journal-of-english-language-and-linguistics-research-ijellr/vol-5-issue-2-april-2017/analyzing-indonesian-english-abstracts-translation-view-translation-errors-google-translate/

Putri, G. D., & Ardi, H. (2015). TYPES OF ERRORS FOUND IN GOOGLE TRANSLATION : A MODEL OF MT EVALUATION. In *Third International Seminar on English Language and Teaching (ISELT-3)* (Vol. 3, pp. 183–188). Padang. Retrieved from http://ejournal.unp.ac.id/index.php/selt/article/view/6306/4910

Sapar, A. A., Ridhuan, M. M., & Abdullah, T. L. (2018). Free Online Translators: A Comparative Assessment in Terms

of Idioms and Phrasal Verbs. *International Journal of English Language & Translation Studies*, *6*(1), 15–19. Retrieved from http://www.eltsjournal.org/index.html

Sridhar, S. . (1975). Contrastive Analysis, Error Analysis and Interlanguage: Three Phases of One Goal? *Studies in Language Learning*, *1*, 400. Retrieved from https://files.eric.ed.gov/fulltext/ED123888.pdf

Uba, S. Y. (2015). An Error Analysis of A Nigerian Postgraduate Student in A United Kingdom University. *British Journal of English Linguistics*, *3*(4), 1–11.

van der Wees, M., Bisazza, A., & Monz, C. (2015). Five Shades of Noise: Analyzing Machine Translation

Errors in User-Generated Text. In *Proceedings of the Workshop on Noisy User-generated Text* (pp. 28–37). Retrieved from http://www.aclweb.org/anthology/W15-4304

Wongranu, P. (2017). Errors in translation made by English major students: A study on types and causes. *Kasetsart Journal of Social Sciences*, *38*(2), 117–122. https://doi.org/10.1016/j.kjss.2016.11.003

Yousofi, N. (2014). Describing the Errors in the Translations of Iranian Novice English Translators. *Procedia - Social and Behavioral Sciences*, *98*, 1952–1958. https://doi.org/10.1016/j.sbspro.2014.03.628