

Google vs. Instagram Machine Translation: Multilingual Application Program Interface Errors in Translating Procedure Text Genre

Anisa Larassati, Nina Setyaningsih, Raden Arief
Nugroho, Valentina Widya Suryaningtyas, Setyo
Prasiyanto Cahyono

English Study Program, Faculty of Humanities
Universitas Dian Nuswantoro, Semarang, Indonesia
anisa.larassati@dsn.dinus.ac.id

Stephani Diah Pamelasari
Faculty of Mathematics and Science
Universitas Negeri Semarang, Semarang, Indonesia

Abstract—The field of translation has benefited from the advance of technology and the Internet nowadays. Hence, it is common to find online machine translation tools in the cyber world. Google Translate and Instagram translation feature, a multilingual application program interface (API), are cases in point. This research focuses on the errors in how procedural texts in Instagram posts are translated into Indonesian language by using Google Translate and API. The translation error typology presented by the American Translators Association (ATA) was used as a framework of analysis. The result shows that Instagram higher errors were made by Instagram than Google Translate, with Terminology error (T), Syntax error (SYN), and Literalness (L) error being the most notable error categories in both translation machines. There is an interrelation among these three. Terminology errors were caused by the Literalness of the text elements, creating the translation of terms not suitable for the context. This research suggests that with improvements, Google Translate has a possibility to be one of the most reliable free MTs in the future. Also, the field of linguistics may become a fruitful contribution in the betterment of machine translation results.

Keywords—error; Google Translate, Instagram, translation

I. INTRODUCTION

Translation concerns with transforming one language or source language into another (target language). It can also be regarded as a form of communication since there are certain messages conveyed by the sender to the receiver, especially when the receiver does not speak the same language as the sender. Therefore, it can be said that translation facilitates message transfer process. However, in communication process, it is common that noises sometimes interfere with the process of conveying messages. In the case of translation, such noises may occur when there is inappropriate translation that makes the text hard to understand. Therefore, some of the messages may not be well grasped by the receiver.

The advance of technology has made translating tasks easier. Nowadays, translation machines, electronic dictionaries, and discussion forums of translators have emerged as auxiliary tools that enable translators to produce better translation. Google Translate is one of the most widely

used online translation tools. It is available for different interfaces such as website, mobile applications for smartphones. It also supports over 100 languages and various forms of text and media such as speech, images, and videos. Google Translate serves translation functions among others written words, website, documents, and speech.

Furthermore, people who speak different languages can also be facilitated by translation features integrated in many social media platforms; Instagram is a case in point. Instagram is a social networking application launched in 2010. It is a photo and video sharing application that are available in Android, Apple iOS, and Windows Phone and can be downloaded for free. This application can be easily accessed in internet connected smartphones.

In 2016, Instagram launched a translation feature. The feature translates captions and comments automatically based on the language they are written in and the language setting of the person who is viewing it. It can also be seen by clicking **See Translation** button below the text.

Posts on various Instagram accounts comprise a variety of text types or genres. One of the genres found in Instagram posts is procedure. It is commonly found in accounts that share cooking recipes. Procedural texts instruct how to do a certain activity, such as science experiments, following an itinerary, and recipes. One of the Instagram accounts that post recipes is @howtofoodprep. This account shares the how-to's of food preparations. The language used in this account is English.

This research analyzes the translations of procedural texts in @howtofoodprep posts generated by translation tools. It looks at how the posts, written in English (the source language/SL), are translated into Indonesian language (the target language/TL). Procedural texts are different from any other genres such as narrative, review, news item, etc, in that the readers need to follow certain instructions to achieve certain goals. For Indonesian readers who do not speak English, translation can help them follow the instructions.

Considering the rationale described above, this research will focus on examining the errors that are found in the translations of @howtofoodprep posts resulted from the

Instagram translation feature and Google Translate respectively.

II. THEORETICAL REVIEW

A. Machine Translation (MT)

The discussion over technology involvement in translation field of study always covers two kinds of technology; they are Computer Assisted Translation Tools or CAT Tools and Machine Translation or MT [1], [2]. The difference lies within the actor who or that conducts the translation process. The former puts human translator as the one who translates a text, while the latter places machine as the one that is responsible in translating a text [3], [4], [5]. In the field of information technology, engineers have been developing MT since the 1940s, but the technology usually referred to as multilingual application program interface is continuously undergoing software update due to frequent technical errors up to now [6]. Many believe that the errors occur due to the linguistic architecture approach used by MT. The approach itself is also famously known as The Vauquois Triangle[7]. From it, we can identify that MT has three approaches in translating a source language (SL) text, they are direct approach, transfer based approach, and *interlingua* approach [7]. The figure below illustrates the architecture of the approach.

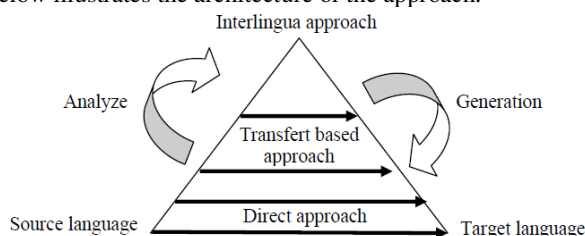


Fig. 1. The Linguistic Architecture of Machine Translation

Direct approach is an approach that enables the analysis of the SL and resolves the SL ambiguity. Aside from the analysis of the SL, the approach allows the program to find the target language (TL) equivalence and analyze the syntactical arrangement of the TL. The approach activates the function of machine or program to imitate the syntactical arrangement of the SL, separate the morphological inflections to get the base form, and match the arrangement in the SL and TL dictionaries. The program component involved in this approach is bilingual dictionary database and computer programs that can analyze and generate a text lexically and morphologically.

Transfer based approach is the stage where the program converts the SL to the TL. The stage has three levels; the first level is converting the SL text to its intermediate representation molding as parse trees. The second level is the conversion of the representation to its equivalent in the TL. In this stage, the program analyzes the abstract representation of the SL. The analysis itself varies from syntax transfer to semantic one. During the syntax transfer, parse trees that are formed to imitate the TL structure appear and convert the SL input. The transfer process uses map structures equivalents

from the SL to the TL whose results are modified in accordance with the syntax and semantic boundaries of the TL.

The last stage is the generation of the TL. This approach is commonly known as the *interlingua* approach. The approach is considered as the most appropriate approach in multilingual system. The approach has two stages, they are the analysis stage (from the SL to its *interlingua*) and the generation stage (from the *interlingua* to the TL). During the process in this approach, the content of an SL sentence is analyzed syntactically and semantically. The result is an *interlingua* content that is free from the language structure of both SL and TL. The *interlingua* content itself is being referred to as the intermediary internal representation of the SL. In order to create the TL, a computer program performing the language analysis generates the *interlingua* to the TL. Many scholars believe that these three approaches are still expected to make translation errors because the incorrect *interlingua* inputs [6]. That is why the phenomenon of translation errors made by two of the biggest multilingual translation programs, i.e. Google Translate and Instagram Translation, becomes our focus of the study in this article.

B. Machine Translation (MT) Errors

The advancement of translation technology such as online dictionary, translation website, and others may, in one hand, make the translation process faster and easier. However, translating one language to another is not as simple as changing the vocabulary of the source language into the target language. Translation process often involved other aspects to take into consideration such as cultural, contextual, and grammatical aspects of language which unfortunately may not be thoroughly understood by MT.

One of the most well-known MT is Google Translate. The convenience provided by Google Translate is the key aspect which makes it one of the most popular translation machines nowadays. It can translate word, texts, and even the whole website by just a simple click. Aside from its popularity; however, researchers have found some errors made by Google Translate. Study of [8] on Indonesian-English translations of news item texts in websites translated by Google Translate has highlighted 13 categories of errors. The errors that are ranged from the highest occurrences to the lowest are: grammatical errors, terminology errors, omission errors, syntax errors, mistranslation/misunderstanding errors, literalness/faithfulness errors, usage errors, punctuation errors, addition errors, ambiguity errors, word form errors, capitalization errors, and spelling errors.

Another similar research has also been conducted by [9] who used different type of text with different language pair, English-Persian. Their research on the accuracy and appropriateness of Google Translate in translating Oliver Twist text has outlined the inability of Google Translate to translate the English verb tense to Persian, which in turn, made the translated passages became unacceptable. In addition to that, grammatical errors in translation of aspects, passives, and compounds were also occurred in the translation output.

In the community of practice of academic language teaching, the widespread use of web-based MT such as Google Translate is inevitable. Non-native students tend to rely on Google Translate to overcome their language barrier or to complete their assignments. The study of [10] had tried to look into the implication of the use of Google Translate in the academic setting. They asked their students to submit an essay in their first language which was then translated into English by using Google Translate. The results of the translations were analyzed for their errors by using [11] error coding. They argue that the results of the translations were far from being perfect or error-free. However, they believe that MT, just like any other technologies, will be develop and getting more sophisticated over time. The development of MT will give a significant influence on the teaching of Languages for Academic purposes and help both the students and the instructors during the process.

The increasing use of social media also contributes to the development of MT. Some social media such as Instagram have been equipped with instant translation button. It provides the “see translation” button/link to translate the caption from English to target language and vice-versa. However, the resulting translations are still far from error-free. Research on Instagram translation error is still limited compared to Google Translate. One current research on Instagram error translation was conducted by [12]. Her study compares the translation made by Instagram MT and university students. The translation results from both sources were then rated for their quality by a professional translator and researcher. The results show that human translator produce better translation results than Instagram MT in both formal and non-formal language.

C. Procedural Texts

Procedural texts aim to provide information and directions in a sequenced way so that the readers can perform the activities appropriately and successfully. The structure of a procedural text generally consists of 1) goal, which is an introductory statement that presents the purpose of the activity, 2) the list of materials or equipments used in the process, and 3) the sequence of steps that are presented in order in which they should be completed.

Procedural texts also have characteristics in terms of the language, commonly referred to as the language features. These may include the use of temporal conjunction (*first, second, before, after*), the use of imperative sentences (*Put the pasta, Wash the potatoes, Pick it up.*), and the use of action verbs (*fold, measure, cut*) [13].

III. METHOD

A. Analytical Framework

This research used translation error typology presented by the American Translators Association (ATA) [14] as the analytical framework since it is considered to be more applicable than [11] error coding. The coding of [11] focuses more on the academic English writing errors. On the other hand, ATA provided more comprehensive categorization of translation errors. They used this list of error categories as the

framework for standardized error marking for professional translator certification. By using this analytical framework, we coded the data as shown in the following table:

TABLE I. FORMAT OF TRANSLATION ERRORS CODING

No.	Word/Phrases/Sentences	Categories of Errors	Comments
1	Source Language		
	Instagram		
	Google Translate		

B. Study Procedure

The data of this research were taken from 3 Instagram posts of @howtofoodprep account. The samples were chosen purposively, that is, the ones that conform to the generic structure and language features of a procedural text. The text in each post was copied and pasted into Google Translate and then translated into Indonesian. As for the Instagram translation features, the **See Translation** button was simply tapped to generate the translation. The translation results were then coded and further processed for analysis.

IV. RESULT AND DISCUSSION

The data investigation process that has been done throughout three Instagram posts revealed a total of 95 and 37 errors made by Instagram Translation and Google Translation respectively. The general findings of each error typology can be seen in the following table below.

TABLE II. GENERAL FINDINGS OF ERRORS MADE BY INSTAGRAM (IG) AND GOOGLE TRANSLATE (GT)

No.	Error Categories	Freq. of IG Errors	% IG	Freq. of GT Errors	% GT	Total of Errors (IG & GT)	% (IG & GT)
1	Terminology (T)	46	48%	23	62%	69	52%
2	Syntax (SYN)	17	18%	6	16%	23	17%
3	Literalness (L)	16	17%	3	8%	19	14%
4	Usage (U)	3	3%	2	5%	5	4%
5	Word Form/Part of Speech (WF/PS)	4	4%	1	3%	5	4%
6	Addition (A)	3	3%	1	3%	4	3%
7	Text Type (TT)	4	4%	0	0%	4	3%
8	Omission (O)	1	1%	1	3%	2	2%
9	Other errors – Meaning Transfer (OTH-MTH)	1	1%	0	0%	1	1%
TOTAL		95	100%	37	100%	132	100%

In general, Instagram made higher errors than Google Translate. The table also shows that the three most notable error categories found in both IG and GT are Terminology error (T), Syntax error (SYN), and Literalness (L) error. These three categories are highly interrelated. The high number of Terminology (T) errors was caused by word-for-word translation (Literalness) of the phrases and sentences. Thus, most words were translated into terms that were not suitable for the context. As a result, most parts of the text, especially those produced by Instagram translation, were less acceptable and less natural. In some instances, Terminology (T) errors were also caused by words or abbreviations such as “tbsp” (tablespoon), “tsp” (teaspoon), and “lb” (libra or pound) in the source language that were not translated in to the target language. One of the examples of Terminology error is as follows:

Excerpt 1. Terminology error for technical term and content words

SL : 1 lb chicken breast
 IG : 1 lb ayam payudara
 GT: 1 lb dada ayam

In this example, the term “lb” (abbreviation for Roman term *libra* or pound in English) is not translated into the appropriate Indonesian term “pon” by both Instagram and Google Translate. “Lb” is not applicable in Bahasa Indonesia. Another terminology error made by Instagram in the example above is the noun “chicken breast” which translated into “ayam payudara”. This translation result is highly unnatural and unacceptable since “payudara” in Bahasa Indonesian is used to refer to female breast or bust. Moreover, the word order does not follow the grammatical rules of Bahasa Indonesia (Syntax Error). The main noun of a compound word in Bahasa Indonesia is placed after the modifying noun. Thus, “dada (breast) ayam (chicken)” made by Google Translate is considered more natural.

Sometimes, the errors found in a phrase or sentence can be further breakdown into several categories which are also interrelated to each other. As seen in Excerpt 1, one of the causes of Terminology errors found in the data was the Literalness (L) of the translation result. In some cases, the Literalness (L) also caused the Syntax (SYN) error as seen in the following example:

Excerpt 2. The Literalness (L) and Syntax (SYN) errors

SL : if you're preparing eating
 IG : jika anda makan persiapan
 GT: jika anda menyiapkan makanan

Excerpt 2 shows that the IG's translation follows the source language word for word and gives incorrect rendition of the intended meaning. It is because the word order of the translation does not follow the grammatical rules of Bahasa

Indonesia. The verb *makan* ‘eat’ and noun *persiapan* ‘preparation’ which together became ‘eating preparation’ is not the intended meaning of the source language. On the other hand, GT gives more acceptable translation by correctly translated the part of speech and word order into the target language. The verb *menyiapkan* ‘preparing’ and the noun *makanan* ‘food’ comes in the correct word order and gives the correct rendition of the intended meaning. It can also be easily understood by Indonesian native speaker.

Another example where three categories of error found in one phrase can be seen in the following excerpt:

Excerpt 3. The Literalness (L), Syntax (SYN), and Terminology (T) errors

SL : Sea salt and fresh ground black pepper to taste
 IG : laut garam dan tanah segar lada hitam untuk mencicipi
 GT: garam laut dan lada hitam segar secukupnya

In the above example, the word for word translation of the phrase made by IG caused Syntax and Terminology error since the arrangement of words does not conform to the syntactic rules of Bahasa Indonesia. The incorrect order of the compound noun *laut garam* makes the translation result highly diverge from the intended meaning. The noun *laut* ‘sea’ which occurs before *garam* ‘salt’ means ‘the sea of salt’ in Bahasa Indonesia.

IG also uses the incorrect term *tanah* ‘soil’ for the word ‘ground’. Thus, makes the translation *tanah segar lada hitam* ‘fresh soil black paper’ sounds very unnatural for Indonesian readers. On the other hand, GT gives more acceptable and natural translation.

The examples above clearly describe that Instagram made a lot more errors than Google Translate. Instagram fails to recognize words with multiple meanings such as ‘ground’ and ‘breast’, and fails to use the correct terminology for the appropriate context. Instagram also tends to made Syntax error as a result of their Literalness in translating the source language.

The frequency of other error categories appears only in less than 10 occurrences. Usage (U) occurs when the collocation of the words is not correct. Word Forms / Parts of Speech occur when the word form is different from the source language, such as verb which translated into noun. Addition (A) occurs when the translation contains word that does not occur in the original source language. Text Type (TT) error occurs when the register or the style of the translation is inappropriate. For example, the procedure text which requires step-by-step instruction that used imperative *membagi* ‘dividing’ instead of infinitives form *bagi* ‘divide’. Omission occurs when there are words being deleted in the target language. Other errors – Meaning Transfer (OTH-MTH) is used to code the translation error that does not fit the descriptions of other categories.

V. CONCLUSION

This case study is limited in terms of language pair, MT engine, data size, and text type. However, a conclusion can be drawn based on the analysis. We can sum up that machine translation does not always provide low quality translation. Google Translate has improved so much as it can recognize technical terms and abbreviations, grammatical rules such as word order, and more appropriate word choice. It also maintains the writing format of the original text including the paragraph spacing and symbols. In translating procedure text, Google Translate successfully follows the style of the genre as there was zero error for Text Type (TT) category. Even though it is not as perfect as the professional translation, Google Translate may become one of the most reliable free MTs in the future.

Instagram, on the other hand, needs to work more on its translation quality. Its error frequency is almost three times higher than Google Translate. It still fails to recognize the correct word choice as the translation results are mostly literal. It also made a lot of errors regarding the grammatical rules of the target language, especially the word order. In addition, Instagram does not keep the format of the original text. The entire paragraph formatting such as spacing from the original text is missing. As a result, it affects the readability of the translated text. Translation service is essential for Instagram as one of the most popular social media platforms. It is hoped that Instagram can improve its translation service in the near future.

The findings of this research are expected to be able to contribute in the improvement of translation generated by machines. The linguistic analysis of translation errors made by MT may give the developers some insights on what to improve for the better translation results.

REFERENCES

- [1] A. Basari and R.A. Nugroho. "Subtitles for movie commercial trailers: a technology-based translation," in the Proceedings of International

Seminar on Application for Technology of Information and Communication (iSemantic), Universitas Dian Nuswantoro: Semarang, 2017, pp. 104-108.

- [2] A. Basari and R.A. Nugroho. "The use of Aegisub in teaching audiovisual translation classes: a review on IT-based subtitling course," in the Proceedings of Education and Language International Conference (ELIC), vol. 1, no. 1. Universitas Islam Sultan Agung: Semarang, 2017, pp. 49-56.
- [3] A.H. Homiedan. "Machine translation," in *Journal of King Saud University, Language & Translation*, vol. 10, 1998, pp. 1-21.
- [4] W.J. Hutchins and H.L. Sommers. "An introduction to machine translation," Academic Press: London, 1992.
- [5] R. Zantout and A. Guessoum. "Arabic machine translation: a strategic choice for the Arab world," in *Journal of King Saud University, Computer and Information Sciences*, vol. 12, 2000, pp. 299-335.
- [6] M. Popovic and H. Ney. "Word error rates: decomposition over POS classes and applications for error analysis," in the Proceedings of ACL Workshop on Machine Translation, 2007, pp. 48-55.
- [7] M.A. Chéragai. "Theoretical review of machine translation," in the Proceedings of International Conference on Web and Information Technologies, 2012, pp. 160-169.
- [8] A. Ismail and R. Hartono. "Errors made in Google Translate in the Indonesian English translations of news item texts," in *Journal of English Language Teaching*, vol. 5, no. 2, 2016, pp.1-6.
- [9] H. Afshin and M.A. Alaeddini. "A contrastive analysis of machine translation (Google Translate) and human translation: Efficacy in translating verb tense from English to Persian," in *Mediterranean Journal of Social Sciences*, vol. 7, no. 4, 2016, p.40-50.
- [10] M. Groves and K. Mundt. "Friend or foe? Google Translate in language for academic purposes," in *English of Specific Purposes*, vol. 37, 2015, pp. 112-121.
- [11] D.R. Ferris, H. Liu, A. Sinha, and M. Senna. "Written corrective feedback for individual L2 writers", in *Journal of Second Language Writing*, vol. 22, no. 3, 2013, pp. 307-329.
- [12] G.F. Arvianti. "Human translation versus machine translation of Instagram's captions: Who is the best?," in the Proceedings of Education and Language International Conference (ELLiC), vol. 2, 2018, pp. 531-536.
- [13] http://wpc.2ee1.edgecastcdn.net/002EE1/teacher_toolkit/rex/writing/pdfs/upper/procedure_text_information_and_overview.pdf.
- [14] American Translator Association. 2017. Framework for standard error marking, an explanation. Web, July 2018. Retrieved from https://www.atanet.org/certification/aboutexams_error.php