Are They There Yet? Revisiting the Accuracy Level of Computer-Generated Translation

Asep Suparman

*English Education Program, Institut Pendidikan Indonesia Garut, Jl. Terusan Pahlawan No. 32 Garut, 44151, Indonesia
Email: asep.suparman@institutpendidikan.ac.id*

Abstract— Speed, cost, and quality have been the central issues in the translation industry. While computers can offer fast and economical translation, their ability to produce accurate translation at a publishable level like human translators may do is still questionable. In the present study, the author linguistically evaluated the accuracy of English to Indonesian translation generated by Google Translate (GT). To this end, GT-generated translation was compared with human translation as the benchmark for accurate translation. The results of analysis showed that GT failed to accurately reconceptualize the message of the English source texts in the Indonesian target texts. Therefore, although Google's Neural Machine Translation system is quite an improvement to the conventional phrase-based translation system, highly accurate translation does not seem to be achieved by Google Translate in the near future.

Keywords— Machine Translation; Google Translate; Translation Accuracy; English to Indonesian Translation, Computational Linguistics.

I. INTRODUCTION

Since as early as 1951, research on fully automatic machine translation (MT) systems has already commenced, and MT system is believed to be a potential economical substitute for human translators [1], [2]. However, the ALPAC report, published in 1966, showed that MT was not feasible, and since then research on the MT system declined considerably [3]. Furthermore, the significance of MT research is received with great skepticism by scholars particularly due to the fact that language involves human's real-world knowledge, which can never be assimilated by computers [4]. It is believed that computers cannot produce high-quality translation without human post-editing [5]. Nevertheless, users of the MT system are increasing in numbers. In 2016; for example, MT market size is reported worth over USD 400 million [6]. Since the market of MT is there to grab, producers of the MT system seem to strive to improve their approach to win the competition.

There have been three major approaches to MT: rule-based approach, statistical approach, and hybrid approach [7]. However, the MT approach seems to begin to shift by 2016 as major players of MT systems like Google, Microsoft, PROMT, and Yandex started to use a new technology using a deep learning approach called neural machine translation (NMT). Mimicking the human brain's neural networks, NMT has been deemed revolutionary in the sense that it can learn from mistakes to improve translation quality [8].

Google's Neural Machine Translation (GNMT) system has even been reported to deliver a better output from its previous system and reduce translation errors as much as 60% [9]. Due to the potential contribution of MT to the language industry, the author believes that it is time to revisit its accuracy.

To achieve the purpose of this study, the following research question is posed: can GNMT accurately translate English to Indonesian?

II. MACHINE TRANSLATION SYSTEM: NOW AND THEN

Despite the fact that two patents for machine translation (MT) system were independently granted in 1933 to Artsrouni in France and Troyanskii in the Soviet Union [10], the history of machine translation is suggested to date from a period just after the Second World War during which computers had been used for code-breaking [1]. It is noted that the idea of computerized translation was attributed to Warren Weaver of the Rockefeller Foundation. He proposed his idea of designing a computer that would translate to Norbert Wiener of Massachusetts Institute of Technology (MIT) on March 4, 1947, but a reply from Wiener on April 30, 1947 said that the idea was very premature. However, interest in automatic machine translation was growing as newspapers published stories about the use of computers in California for translating word-to-word level [11].

In a very brief description, the history of translation technology could be outlined in the following stages: the

first generation MT system, the second generation of MT systems, and practical and latest MT systems.

There was a growing interest in automatic translation in the period of 1951-1953, during which it became one of the most studied subjects in the field of what has ever since been called computational linguistics [2]. In 1951, MIT finally decided to appoint Yehoshua Bar-Hillel to a full-time research post, and in 1952 they hosted a conference on MT attended by 18 scholars interested in the subject. Over the next 10-15 years, studies on MT were carried out in the likes of the USA, in the USSR, Great Britain, and Canada [1].

In the early 1960s, the so-called first generation of MT program was developed by Georgetown University using enormous dictionaries [2], and in 1964 the US government set up the Automated Language Processing Advisory Committee (ALPAC) tasked with evaluating if its money spent on MT research was worthwhile. In their 1966's report, ALPAC concluded that MT was slower and less accurate than and twice as expensive as human translation, but they suggested that machine-aided translation may be feasible [1], [3], [12].

Studies on MT systems were extensively carried out in Canada, Japan, and Western European countries in the period of 1970-1980s. The design of MT systems developed during this time commonly parsed the SL, transferred SL syntactic structures to the TL, and generated the TL [1], [2].

In 1980s, with the emergence of small-scale computer hardware, which is today known as personal computers, PCs, MT research shifted its focus to computer-based tools to assist human translators called translator's workstation, which is now known as machine-aided translation system, computer-assisted translation tools or Translation Memory (TM) software [1], [13].

In the mid-1990s, MT and TM software products began to be available in the market for professionals and amateurs [14], and conferences on MT has been increasingly held since then; for example, Conference of the Association for Machine Translation in the Americas, AMTA every other year [15], and extensive research papers on this field have been prolifically published.

III. METHOD

The purpose of this study is to evaluate the accuracy level of computer-generated English to Indonesian translation from the perspective of linguistics, not computer science. To achieve this purpose, the study began with a random selection of different types of English texts available for free on the Internet including scientific, news, movie subtitle, and advertisement texts. These selected texts were to heuristically exemplify the units of analysis. The Indonesian translation of the selected texts was then generated by Google Translate (GT). GT was chosen to model the MT systems in the present study due to its massive popularity. It has been reported to translate a staggering 143 words daily [16]. The author then asked three professional Indonesian translators with 9-10 years of experience in the Indonesian translation industry to do volunteer postediting to the GT-produced English to Indonesian translation. The author used their post-edited translation as the benchmark for accurate translation using Vinay and Darbelnet's comparative stylistic approach [17].

IV. RESULTS AND DISCUSSION

The first text is an abstract of a scientific article [18] freely accessible at www.sciencedirect.com. The source text (ST) is in English, and the target text (TT) is in Indonesian. The result of GT-generated translation and professional post-edited (PE) text is presented in Table I.

TABLE I
EXAMPLE OF ENGLISH TO INDONESIAN SCIENTIFIC TRANSLATION

ST (EN)	(1)	At the center of academic writing sits the practice of paraphrasing.
	(2)	Despite this commonplace presence, it continues to be a challenging activity for both students to engage in and for teachers to instruct.
GT (ID)	(1)	<i>Di pusat penulisan akademik duduk praktik parafrase.</i>
	(2)	<i>Terlepas dari keberadaan yang biasa ini, kegiatan ini terus menjadi kegiatan yang menantang bagi siswa untuk terlibat dan mengajar guru.</i>
PE (ID)	(1)	<i>Parafrase merupakan bagian penting dari penulisan akademis.</i>
	(2)	<i>Walaupun sudah lazim, parafrase tetap saja sulit baik bagi para siswa untuk mempelajarinya atau para guru untuk mengajarkannya.</i>

In Example (1), GT failed to interpret the word *sits*. In its Indonesian translation, *duduk*, which means *to be seated*, the word *sits* is conceptualized as physical activity. In addition, GT employs a literal translation by maintaining the ST's syntactic construction in the TT [19], which makes its Indonesian translation logically incomprehensible. In this case, as the PE translation suggests, the translation should not be approached with a source language emphasis, but with a target language emphasis. In its Indonesian translation, GT failed to reconceptualize the message that the paraphrasing practice is a part of academic writing; instead, it viewed the paraphrasing practice as a concrete object being seated in an academic writing center.

At subsentence level, the use of transition phrase *terlepas dari* in Example (2) as a translation for the word *despite* does not sound natural to the TT audience both textually and contextually since it cannot convey a contrast in meaning. The rest of the sentence shows that GT failed to translate the ST into the TT at so many levels. The redundant use of the word *kegiatan*, the use of the word *terlibat*, and the incorrect construction *mengajar guru* show translation failures at syntactic, semantic and pragmatic levels.

The second text is taken from a news article published online by BBC (<https://www.bbc.com/news/world-africa-47843843>) dated April 7, 2019, entitled Rwanda genocide: Nation marks 25 years since mass slaughter. Table II exemplifies its GT-generated English to Indonesian translation.

TABLE II
EXAMPLE OF ENGLISH TO INDONESIAN NEWS TRANSLATION

ST (EN)	(3)	Rwanda genocide: Nation marks 25 years since mass slaughter
	(4)	Rwandans will mourn for 100 days, the time it took in 1994 for about a tenth of the country to be massacred.
GT (ID)	(3)	<i>Genosida Rwanda: Bangsa menandai 25 tahun sejak pembantaian massal</i>

	(4) <i>Rwanda akan berkabung selama 100 hari, waktu yang dibutuhkan pada tahun 1994 untuk sekitar sepersepuluh negara dibantai.</i>
PE (ID)	(3) <i>Parafrase merupakan bagian penting dari penulisan akademis.</i>
	(4) <i>Walaupun sudah lazim, parafrase tetap saja sulit baik bagi para siswa untuk mempelajarinya atau para guru untuk mengajarkannya.</i>

In Example (3), the author found that the use of the word *menandai* as a contextually wrong translation of the word marks. As suggested by PE, the word *memperringati* is the perfect translation. Again, the literal translation employed by the GT system seems to fail to capture the fact that the word *sejak* may imply that the massacre is still happening right now, but it is not the case. Therefore, the word *since* is better not translated, and the word *silam* is used in the TT to compensate for it. Such compensation is required in order to make the translation not only accurate but also natural to the TT audience [20]. The same compensation strategy also applies to *sama dengan lamanya waktu pembantaian sepersepuluh rakyat negara tersebut pada tahun 1994* in Example (4) as literal translation employed GT system results in the incorrect translation.

Translators must be equipped not only with linguistic knowledge but also with encyclopedic knowledge. Computational linguistics may help MT with the first, but not with the latter – at least up to this day. A perfect example where MT is restricted by the lack of real-world knowledge is shown in Table III.

TABLE III
EXAMPLE OF ENGLISH TO INDONESIAN MOVIE SUBTITLE

ST (EN)	(5) <i>Ally! Ally! Ally! Ally!</i>
GT (ID)	(5) <i>Sekutu! Sekutu! Sekutu! Sekutu!</i>
PE (ID)	(5) <i>Ally! Ally! Ally! Ally!</i>

Example (5) shows that *Ally*, a movie character's name from *A Star Is Born*, is translated by GT into *sekutu*, which if back-translated would literally mean *an ally (friend)* in English. Human translators can easily identify that *Ally* is a non-translatable unit, but GT cannot.

The last text is an advertisement text. The author Google-translated the advertisement text available in Figure 1, and the result is presented in Table IV.

Fig. 1 Advertisement text (taken from piximus.net)

Example (6) shows that GT even failed to translate a simple sentence accurately. If back-translated into English, the GT-produced translation *tidak pernah begitu memuaskan* would mean *has never been really satisfying*, which is opposite to the original meaning of the ST.

TABLE II
EXAMPLE OF ENGLISH TO INDONESIAN NEWS TRANSLATION

ST (EN)	(6) Four inches has never been so satisfying.
GT (ID)	(6) <i>Empat inci tidak pernah begitu memuaskan.</i>
PE (ID)	(6) <i>Empat inci tak pernah sememuaskan ini.</i>
	(7) <i>Memperkenalkan menu hemat kami yang baru.</i>

The use of the word *nilai* in Example (6) is also incomprehensible. In the given context, the word value means that the product is offered at an economical price. The perfect Indonesian translation, as suggested by PE, is then *hemat*.

V. CONCLUSION

GT system seems to tend to maintain the ST, in this case, English, in the TT, Indonesian. Due to the different linguistic systems between English and Indonesian, such literal translation resulted in inaccurate translation at the lexical, syntactic, semantic, and pragmatic levels. In addition, GNMT does not seem to be able to solve common MT problems associated with the lack of human encyclopedic knowledge. To conclude, in the case of English to Indonesian translation, GT still cannot produce text at a publishable level without intervention of human translators' post-editing.

ACKNOWLEDGMENT

The author is greatly in debt to the three professional translators, whose names for ethical reasons could not be mentioned, for their voluntary participation in this study.

REFERENCES

- [1] H. L. Somers, "Introduction," in *Computers and translation*, H. L. Somers, Ed. Amsterdam & Philadelphia: John Benjamins, 2003, pp. 1-11.
- [2] B. Vauquois, "Automatic translation—A survey of different approaches," in *Readings in machine translation*, Sergei Nirenburg, H. L. Somers, and Yorick Wilks, Eds. Massachusetts: The MIT Press, 2003, pp. 333-337.
- [3] W. J. Hutchins, "ALPAC: The (in)famous report," in *Readings in machine translation*, Sergei Nirenburg, H. L. Somers, and Yorick Wilks, Eds. Massachusetts: The MIT Press, 2003, pp. 131-135.
- [4] S. Barrachina et al., "Statistical approaches to computer-assisted translation," *Computational Linguistics*, vol. 35, no. 1, pp. 3-28, 2008.
- [5] S. Bennett and L. Gerber, "Inside commercial machine translation," in *Computers and translation*, H. L. Somers, Ed. Amsterdam & Philadelphia: John Benjamins, 2003, pp. 176-190.
- [6] A. Bhutani and P. Bhardwaj. (2016) Global Market Insights. [Online]. <https://www.gminsights.com/industry-analysis/machine-translation-market-size>
- [7] S. Nirenburg and Y. Wilks, "Machine translation," in *Advances in Computers*, Marvin V. Zelkowitz, Ed. San Diego & London: Academic Press, 2000, pp. 159-188.
- [8] B. M. Gupta and S. M. Dhawan, "Machine translation research: A scientometric assessment of global publications output during 2007-

- 16," *DESIDOC Journal of Library & Information Technology*, vol. 39, no. 1, pp. 31-38, 2019.
- [9] Y. Wu et al. (2016, October) Arxiv.org. [Online].
<https://arxiv.org/abs/1609.08144>
- [10] L. Cocci. (2009) Translation Journal. [Online].
<http://translationjournal.net/journal/50caten.htm>
- [11] W. Weaver, "Translation," in *Readings in machine translation*, S. Nirenburg, H. L. Somers, and Y. Wilks, Eds. Massachusetts: The MIT Press, 2003, pp. 13-17.
- [12] J. S. White, "How to evaluate machine translation," in *Computers and translation*, H. L. Somers, Ed. Amsterdam & Philadelphia: John Benjamins, 2003, pp. 211-244.
- [13] H. L. Somers, "Translation memory system," in *Computers and translation*, H. L. Somers, Ed. Amsterdam & Philadelphia: John Benjamins, 2003, pp. 31-47.
- [14] E. Lagoudaki, "Translation memory systems: Enlightening users' perspective," London, 2006.
- [15] R. E. Frederking and K. B. Taylor, "Machine translation: From real users to research," in *6th Conference of the Association for Machine Translation in the America, AMTA 2004*, Washington DC, 2005.
- [16] T. Wolverton, Google CEO Sundar Pichai revealed a jaw-dropping fact about its translation app that shows how much money is still sitting on the table, July 24, 2018.
- [17] J. P. Vinay and J. Darbelnet, *Comparative stylistics of French and English: A methodology for translation*. Amsterdam & Philadelphia: John Benjamins, 1995.
- [18] M. Mori, "Our speech is filled with others' words: Understanding university student and instructor opinions towards paraphrasing through a Bakhtinian lens," *Ampersand*, vol. 5, pp. 45-54, 2018.
- [19] P. Newmark, *A textbook of translation*. Englewoods Cliffs, New Jersey: Prentice Hall, 1988.
- [20] S. Hervey and Ian Higgins, *Thinking translation: A course in translation method: French to English*. London: Routledge, 1992.