

Google, Translate This Website Page

Flipping through Google Translate's Ability

Bayu Budiharjo

English Department

Universitas Sebelas Maret

Surakarta, Indonesia

budiharjo_b@staff.uns.ac.id

Abstract— Google Translate is one of the most universally used Machine Translations in the world. This Machine Translation is claimed to be equipped with Artificial Intelligence technology in performing translation between a range of language pairs, enabling it to produce “translations that are usually more accurate and sound closer to the way people speak the language”. One of the services this product of advanced technology offers is website translation. This paper aims to investigate both the strong points and the infirmities of Google Translate in relation to transferring the message of and naturalness in the reconstruction of website page content, which may take different forms ranging from linguistic content, such as single words, terms and sentences, to other types of content: images and hyperlinks. This research was undertaken by translating the homepage of the official website of English Premier League from English into Indonesian. The data consist of linguistic units ranging from word to sentence and the visual elements on the website page. The content of the original site was compared to that of its translation. Analysis was done on certain elemental areas covering the verbal and visual elements of the content in connection with its nature as content placed on website. This research gains the following findings. Some terminologies and expressions are flawlessly rendered despite the false translations and inconsistencies in the translation product. Interesting phenomena are identified: spatial shift and glitches in terms of web layout. Apparently, Google Translate and the technology it has still require prolific advancement to be able to function as a reliable website translator.

Keywords—Google Translate; website translation; strengths and weaknesses

I. INTRODUCTION

Technology has made people's life a lot easier in almost every aspect, including translation. The activity of rendering message across languages can today be done with the help of products of technology ranging from CAT Tools to online translation services available free of charge. The latter seems to be promising, especially for laypeople, to get instant and trouble-free translation service. One popular online translation service is Google Translate, which has the ability to translate a single word or term to a whole text. The automated translator powered by Google also offers website translation service. This feature of Google Translate can help web users access the content of websites from which they want to get information in the language they understand.

Despite its ability to produce quick translation, until recently, there are still gaps between translations resulting from Google Translate and those human translators would produce. Researches carried out in the last few years support this statement. Errors are still produced by Google Translate (Vidhayasai, Keyuravong and Bunsom, 2015; Chen, Acosta and Barry, 2016; Afshin and Alaeddini, 2016; Nadhianti , 2016; Darancik, 2016; Sanchez Martin, 2017; Napitupulu, 2017; Yusran, 2017; Allué, 2017; Amanah, 2017) and the errors range from those dealing with linguistic issues to problems with style and mistranslation of technical terms. The aforementioned researches concentrate the investigation on texts of diverse types. The ability of Google Translate to translate website has not seemed to have equal attention and this is a focal area of scientific investigation as this Google Translate's service has been campaigned worldwide.

Exploring separate area from the researches' areas of study, this research focuses on a different type of object, namely website content. Text of this type involves not only linguistic elements but also visual and spatial ones. The distinct characteristic initiates different challenge for the online Machine Translation. This research directs its main focus to website translation, which is set apart from the discussion of localization. It is due to the distinct characteristics of the translation of website page investigated in this research. As its name suggests, Google Translate translates websites and does not perform website localization, which involves putting cultural differences into consideration. Google Translate simply converts units of language in one language into those in another language. The Machine Translation does task other than “producing a new (translated) target website based on the source website that meets the specified purpose for the new cultural and linguistic community” (Sandrini, 2006).

The focus of attention of this research is Google Translate, an online Machine Translation powered by Google which has the ability to work on translation involving more than 100 languages (<https://translate.google.co.id>). This Machine Translation is supported by the high-tech Neural Machine Translation (NMT) for some language pairs, including English- Indonesian (Tempo, 2017). The production NMT at Google (GNMT) is able to eliminate slower training and

inference speed, ineffectiveness in dealing with rare words, and occasional failure to translate all words in the source language sentence, the three inherent weaknesses of Neural Machine Translation (Wu, et.al., 2016). GNMT works by considering entire input sentence as a unit for translation. It works in a different way from the previous generation Phrase-Based Machine Translation (PBMT) which breaks an input sentence into words and phrases to be translated largely independently (<https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html>

II. METHODOLOGY

This research aims to identify the strengths and weaknesses of Google Translate in translating the homepage of the official website of Premier League (<https://www.premierleague.com/>) from English into Indonesian. The homepage analyzed in this research is dated May 12, 2018, which has undergone content changes because of updates on regular basis. This research is an attempt to assess whether the translation is an ideal translation serving as bridge facilitating Indonesian users who do not have access to English. The data analyzed in this research take form of elements composing website content on the homepage of the official website of Premier League, which are linguistic units, images and hyperlinks. The website is chosen based on the following considerations (1) English Premier League is the most well-liked football league in Indonesia, (2) the Indonesian version of the website has currently not been available and (3) the homepage of the website and its Google-Translate-generated translation provides numerous phenomena to explore, i.e. technical terms, images and other elements which are potential to complicate the work of the Machine Translation. The data are analyzed by means of comparison between the source text and the target text to evaluate the equivalence of the original content and that of the translation, covering the semantic and semiotic dimensions referring to the assessment kit proposed by Hariyanto (2015). The assessment kit comprises a set of assessment parameters for the four dimensions from which website translation can be analyzed: pragmatic dimension, semantic dimension, stylistic dimension and semiotic dimension. As the translation is done by machine, the points of assessment exclude the equivalence of intention and speech act as well as style in expressing message. The evaluation of the semantic dimension involves the use of the online version of dictionaries (Cambridge, Merriam-Webster and Oxford). The aim of evaluating the two dimensions is examining to what extent the translation generated by Google Translate fits the boundaries of how a translated website page should be. The comparison, along with deeper study, is also done to discover strengths and weak spots tracked from the translation product.

III. RESULT AND DISCUSSION

A. Result

The first and the main strength of Google Translate is its ability to produce instant translation. The online Machine

Translation translates the homepage in a matter of seconds. In addition to the speed, based on the analysis, some other strengths of Google Translate are identified.

Several club names are recognized and translated as technical terms instead of being translated word for word, for example “Man City” (the short form of “Manchester City”) and “West Ham”. The identical forms “Man City” and “West Ham” are displayed on the Indonesian version of the website. The Machine Translation does not work by translating every word in isolation despite the fact that the words “Man”, “City”, “West” and “Ham” have their own out-of-context meaning when each of them occurs as individual word.

The other terms faultlessly translated are the football terms “matchweek 38” and “fixtures”. The term “matchweek 38” is detected by Google Translate as a technical term and translated into “pekan 38”. The equivalent of the term “matchweek” used by Google Translate here is the one commonly used in the target language. Further, the use of “pekan” is a correct “decision” despite other options of equivalent, for example “matchweek” when the term is translated in isolation by the Machine Translation. Similarly, the latter is detected as a term in the subject of football, i.e. the upcoming dates on which matches are held. The term is impeccably translated into the target language.

A more complex phrase is also successfully translated. The phrase “Rooney’s long range volley” is translated with an additional detail. The expression used as the equivalent in the target language is “*tendangan voli jarak jauh Rooney*”, in which the word “*tendangan*” is added. It successfully delivers the message of the original phrase. In football, “volley” refers to “kicking a moving ball in the air before it touches the ground”. The Machine Translation accurately identifies the context and picks the suitable equivalent.

The Machine Translation’s strengths in translating come with its weak points. These fragilities deal not only with linguistic components in the website content but they are also concerned with images and spatial configuration. The fragilities make the homepage of the website send different message; they cause the translation to look different as well.

The typical problems of the Machine Translation generated by Google subsist in the result of the website translation. Some source language words appear on the translated web page as their mismatched equivalents, among the possible alternatives of equivalent. Some instances to give details about the case are:

| | |
|--------|--------------------|
| ST | TT |
| Youth | <i>pemuda</i> |
| Broken | <i>rusak</i> |
| Spurs | <i>Kemasyhuran</i> |

The three translation cases have one thing in common and each case comes with different problem. The word and terminologies are translated on the basis of one-to-one correspondence in the target language, with no identification of their function in the text. This makes the translation fail to

convey the message as the original text does. In the first example, “youth” functions as a hyperlink navigating users to a different page displaying content about competitions and other features concerning youth teams. Therefore, the proper translation of “youth” is not “*pemuda*”, which refers to young person or people. The term is part of the navigation bar and thus it stands on its own (does not occur together with other word). In the second example, the word “broken” is present under different environment from the one surrounding the occurrence of “youth”. The word “broken” on the website page is part of an interrogative “Which records have Man City broken?” Despite its occurrence in a more well-defined environment, the word is translated into a word carrying the most basic sense, “*rusak*”. The collocation between “break” and “records” is excluded from the analysis and transfer of message from the source language into the target language. This results a poor, out-of-context translation. Meanwhile, the third example further illustrates the problem in the product resulting from Google Translate. The term “Spurs”, the short form of a football club name “Tottenham Hotspurs”, is translated into “*Kemasyhuran*”. The name is translated neither as a name nor as a word carrying its most basic sense. The equivalent chosen as the equivalent contains a “new” meaning, not listed in the dictionaries.

Another case in relation to the linguistic elements is the inconsistency in translating some of the same terms on the website page. The inconsistency is observable from the translation of identical terms occurring under similar environments which are translated into different forms of equivalent. The following examples illustrate the irregularity in the transition of some terms:

| ST | TT |
|------------------------|---|
| Fantasy Premier League | Fantasy Premier League <i>Liga Premier Fantasi</i> |
| ad | ad <i>iklan</i> |
| more | lainnya lebih banyak lebih |

The term “Fantasy Premier League” is the name of an online fantasy football game which can be played via the official website <http://www.premierleague.com>. The same terminology occurs, at different sections, on the website page as hyperlink to link the homepage of the website to the page on which the users can access the game. The term is translated differently, in which one of the two undergoes no change while the other is translated word for word. Similar case is observable in the translation of “ad”. It occurs at different sections and is also translated in different ways, resulting “ad” and “*iklan*” as the equivalents. The more intriguing case concerns the translation of the term “more”, used for referring to “further related content”. The term “more” on the navigation bar is translated into “*lebih*”. Meanwhile, “more” as part of phrases is inconsistently translated into “*lebih*” in “*Lebih Sosial*” (ST: More Social), “*lebih banyak*” in “*Lebih Banyak*” (ST: More News).

“Stats” (ST: More Stats) and “*lainnya*” in “*Berita Lainnya*” (ST: More News).

The other problems identified in the translation of the homepage from English into Indonesian are outside Linguistics concern. These problems deal with visual and spatial aspects. Shifts and glitches occur, causing the website page in the target language to have different appearance from the original.

The shifts include the reversed position of “search” and “sign in” buttons as can be seen below:

ST:



TT:

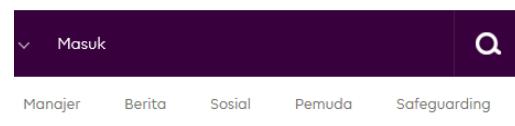
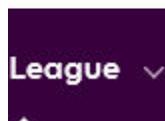


Figure 1. Reversed Position

In the original website page, the position of the “search” button is on the left and the “sign in” button is on the right. The position changes on the translated website page. The “*masuk*” button moves to the left and leaves wider blank space within the menu bar. From both images taken from the original and translated homepages, it can be observed another shift. Buttons located on the right of “social” in the original website page are grouped into drop down menu “more” while those displayed on the translated page are lined up so that no drop down menu can be found.

The shifts also involve the layout in relation to club logos. The images below provide a clearer picture about this spatial shift. The images on the left side are the beginning sections of the horizontal series of club logos. Meanwhile, images on the right show the end of the space displaying the club logos.

ST:



TT



Figure 2. Club Name Bar

On the display of the English website page, the twenty club logos occupy the same row. The Indonesian website page displays different images. The image of the last club (West Ham) logo moves and causes the second row to appear.

Shifts happen to another part of the translation of the homepage: the section displaying fixtures, as seen in the following images.

ST:

| | | |
|-----|-------|-----|
| HUD | 21:00 | ARS |
| LIV | 21:00 | BHA |

TT:

| |
|-----|
| HUD |
| |
| |
| ARS |
| |
| |
| LIV |
| |
| |
| BHA |

Fig. 3. Fixture Layout

In the original text, the abbreviated names of clubs and local time are displayed in single rows. In the translated text, the section looks very much different. The names occupy different rows and the time indicating the kick off is nowhere to be found. This problem in the translation of the website page involves not only changes in relation to space but also loss of information.

In addition to changes in relation to position of images and texts, blank space becomes an issue in the translation performed by Google Translate. Blank spaces occur in some parts of the Indonesian version of the homepage. These blank spaces set off problems concerning both the layout and message of the website page.

A blank space creates layout change, more precisely; the translation of the website page displays additional space which is nonexistent in the source text. The following extract from the original homepage and its translation provides clearer notion.

ST:



TT:



Fig 4. Additional Space

In the source text, the text is right below the image but when the website page is translated into the target language, the result displays additional blank space between the image and the text (the image in the TT is not presented in the extract because of the limited available space in the paper). The blank space forces the text to appear on the area outside the box in which it should appear.

A different problem exists caused by another blank space in the different section on the translation result. The position of the blank space is next to one image which the Machine Translation can render in the Indonesian version of the homepage. The different features in both the source and target languages can be obviously seen in the following screenshots:

ST:



TT:



Fig. 5. Blank Space and Missing Text

In the English version, the website page shows an image with a text. In the translated version, the text is not present. This blank space makes users unable to access the information contained in the original text.

The points described in the previous sections provide more or less a certain amount of picture about the performance of website translation service powered by Google. The quick translation of the website pages has many things to take into account to be able to come with the quality as expected.

In relation to verbal expressions, the translation generated by the Neural Machine Translation still contains considerable amount of “funny target text” (Hariyanto, 2015), as the result of literal translation of technical terms, the presence of distorted equivalences and the use of equivalent outside those listed in dictionaries. These observable facts cause more than translation which sound funny. In conjunction with some missing parts of the website homepage, the funny translations convey different message from the original. Difference in a website and its translation into other language is to some extent justified (Chuang, 2011), but it is a different kind of difference.

B. Discussion

The discussion section starts with the ability of Google’s Neural Machine Translation to maintain speed as Google’s scientists improve neural machine translation over the past few years while finding a way to make it work on large data sets (<https://www.zdnet.com/article/google-announces-neural-machine-translation-to-improve-google-translate/>). The Machine Translation is claimed to have the abilities which “tend to be better than other models both in terms of speed and accuracy” (Wu et.al, 2016) The statement about how speed is maintained is proven in this research as Google Translate took only several seconds to complete the translation of the homepage.

The wordpiece models also demonstrate brilliance in handling several cases in relation to rare words. It is in line with the statement that Google’s Neural Machine Translation system attempts to address many issues concerning Neural Machine Translation (*ibid*), one of which is translation of rare words. The Machine Translation recognizes “West Ham” and also “Aston Villa” as single terms instead of common phrases and this also applies to some other cases, in which recognition of the terms as common phrases would produce incorrect translation. This reveals that the machine no longer processes data word-for-word. The NMT system can correctly process data on sentential base based on a range of context in

determining the most relevant translation (Tempo, 2017). In relation to context, Google Translate also shows its quality in translating the terms “Gameweek 38” and “fixtures”. In translating the two terms, Google Translate places the two in the right context, the sphere of football. The quality is also reflected from the action done on the term “volley” and the choice of correct equivalent. The Machine Translation “adds” detail in the translation of the term.

Apparently, the word “quality” used earlier not only refers to the positives but also negative qualities. The translation of the homepage shows different characteristics from the criteria of good translation, viewed through Larson’s theory (1984: 485), namely accuracy, clarity and naturalness, which coincides with the parameters of translation quality proposed by Nababan, Nuraeni and Sumardiono (2012): accuracy, acceptability and readability. The discussion about translation quality here is limited to analyzing the accuracy and naturalness of the translation and it uncovers the fact that the translation of the homepage of the official website of English Premier League is still away from the criteria of good (high-quality) translation.

At this point, the discussion continues with comparison between the observable facts, presented previously, and the theory about good translation proposed by Larson which is in line with the one proposed by Nababan, Nuraeni and Sumardiono. In terms of verbal element, at the micro level (sentence and the units below), the quality of the translation of micro units within the website page vary. Some units of translation successfully convey the message of the original text and at the same time sound natural in the target language, in other words they fulfill the criteria of accuracy and naturalness. Some others, however, contain different information.

A contrast is in evidence from the previously-presented translation of “Rooney’s long range volley” and the translation of “youth” into “*pemuda*”. The latter exemplifies translation using the most basic sense, otherwise speaking, out of context. The result shows inadequacy in complying with the criteria of accuracy as the term used as the (mistaken) equivalent points to different referent despite the lack of problem with naturalness. To resolve the problem, the translated term which should be present there is “*tim junior*”.

Another identified item exemplifying problem with selection of equivalent is the translation of “Which records have Man City broken?” into “*Rekaman mana yang membuat Man City rusak?*” The translation evokes the need of further improvement of the Machine Translation despite the result of a series of experiments which suggests that “the quality of the resulting translation system gets closer to that of average human translators.” (Wu et.al., 2016). The example shows some uses of equivalents which are out of context, such as in the selection of the equivalent terms for “record” and “broken”. The collocation formed by “record” and “broken”, which can serve as context within the sentence, is left unrecognized, resulting translation into most basic senses: “*rekaman*” and “*rusak*” instead of “*rekor*” and “*pecah*”, which are the correct equivalents. The translation does not fulfill the

criteria of accuracy. In addition, in terms of naturalness, the translation is grammatically correct but sounds funny.

In relation to the other element, the visual element, the discussion is more or less the same. Some images appear exactly identical and on the same position, both on the original website page and on the translated page, some of which are the followings:



Fig. 6. Identical Images

On the contrary, some images are not rendered in the same way. The visual glitch found on the translation generated by Google Translate causes various impacts on the accuracy and naturalness. The switch between “sign in” and “search” (magnifying glass) buttons (Fig. 1) is found not to lessen any of the accuracy or the naturalness. The term “masuk” and the symbol of magnifying glass in the target language carry the original message and are the natural forms to express the message. While the switch of position is still “safe”, some other glitches taking form of spatial shifts and blank spaces decrease the degree of naturalness and even accuracy of the translation.

The section presenting an image and verbal content “Final-day relegation battles: 2010/11” (Fig. 4) looks unusual because of extra blank space lying between. The extra space intensifies the unusual look by forcing the text downward, making the text break through the space in which it should occur. Also, the shift happening to the section displaying fixtures of English Premier League matches in the target language (Fig. 3) causes the section to look odd with the blanks between club names, causing it to appear as a peculiar box containing rows with club names. Further, the information suggested by the source text (the teams playing in the league matches and the kick off time) is not conveyed in the target language. The last two instances of spatial glitch show problems: the first with naturalness and the second with naturalness and accuracy.

Further analysis was done by applying the assessment kit designed by Hariyanto (2005) and the result indicates that the translation of the homepage of English Premier League official website generated by Google Translate is not an ideal one. The statement is constructed based on the answers to some relevant questions in assessing the semantic and semiotic dimensions (*ibid*). The relevant questions related to semantic dimensions deal with (1) whether there is any change of meaning (referent) in TT as compared to ST and whether the change (if any) is justifiable and (2) whether there is any omission or addition of information in TT as compared to ST and whether it happens systematically. Meanwhile, the questions related to semiotic dimensions deal with whether there is any change in non-verbal elements (color, illustration, pictures, etc.) of the TT as compared to that of ST and whether the change (if any) is justifiable.

The answers to the questions can be derived from the details specified in the previous sections. Unjustified changes of referent occur in the translation, which result from the use

of mistaken equivalents. Unjustified omission (missing information) occurs at some sections of the website page, which is caused by spatial shifts. All of these give negative impact on the overall semantic dimensions of the translation. Unjustified changes also happen to the non-verbal elements, which take forms of spatial shifts and additional spaces. The alterations result change of web layout and make the translation of the homepage still away from the criteria of naturalness.

IV. CONCLUSION

Based on the trial, analysis and discussion, in a relatively short amount of time, some elements of the website page (verbal and visual) are successfully rendered in the target language despite the presence of false translations and inconsistencies in the translation of the verbal elements of the website page. The false translations and inconsistencies altogether lower the level of accuracy and naturalness of the translation. In addition, spatial shifts and glitches are identified in relation to the transfer of the visual elements from the source language to the target language, causing changes in terms of web layout. These changes give similar effect on the text as a whole: cutback in terms of accuracy and naturalness. Apparently, Google Translate with its improved Neural Machine Translation still requires prolific advancement to be able to produce reliable translation of website. As this research is a basic research, involving a very limited quantity of data, this research can only describe a small piece of the phenomenon of website translation produced by Google Translate. Further studies investigating deeper topics about website translation performed by Machine Translation are encouraged to get a more complete understanding about the issue. Particularly for developers of Machine Translation technology, attention can be drawn on upgrade, one of which is the ability of Machine Translation to recognize context.

References

- Afshin, H. and Alaeddini, M. A. (2016). A contrastive analysis of machine translation (Google Translate) and human translation: Efficacy in translating verb tense from English to Persian. *Mediterranean Journal of Social Sciences*, 7. Retrieved from www.mcses.org/journal/index.php/mjss/article/viewFile/9502/9172
- Allué, B. R. (2017). The reliability and limitations of Google Translate: A bilingual, bidirectional and genre-based evaluation”, *Entreculturas*, 9. Retrieved from <http://entreculturasmusica.comimagine.es/wp-content/uploads/2017/05/articulo05.BlancaRoigAllué.Entreculturas9.pdf>
- Amanah, F. H. (2017). Errors made by Google Translate and its rectification by human translators (Unpublished thesis). University of Malaya, Malaysia.
- Chen, X. Acosta, S. and A. E. Barry, A. E. (2016). Evaluating the accuracy of Google Translate for diabetes education material. *JMIR Diabetes*, 1, 1-11. <https://doi.org/10.2196/DIABETES.5848>.
- Chuang, Y. T. (2011). The challenges of new media translation: A multimodal approach to website translation. *Linguistics Applied*, 4. Retrieved from <https://repozytorium.ukw.edu.pl/bitstream/handle/item/4182/The%20challenges%20of%20New%20Media%20Translation%20a%20multimodal%20approach%20to%20website%20translation.pdf?sequence=1&isAllowed=true>
- Daranick, Y. (2016). The effect of data-based translation program used in foreign language education on the correct use of language. *The Turkish Online Journal of Educational Technology*, 15. Retrieved from <https://eric.ed.gov/?id=EJ1117647>

- Hariyanto, S. (2016). *Website translation: With special reference to English–Indonesian language pair*, Malang: Transkomunika Kencana.,
- Larson, M. (1984). *Meaning-based translation*, Lanham, Md: University Press of America.
- Martin, B. S. (2017). *Translation quality assessment of Google Translate and Microsoft Bing Translator* (Unpublished thesis). Universidad de Valldolid, Spain.
- Nababan, M. R. Nuraeni, A. & Sumardiono. (2012). Pengembangan model penilaian kualitas terjemahan. *Kajian Linguistik dan Sastra*, 24, Retreived from <https://publikasiilmiah.ums.ac.id/handle/11617/2220>.
- Nadhianti, N. (2017). *An analysis of accuracy level of Google Translate in English-bahasa Indonesia and bahasa Indonesia-English translations* (Unpublished thesis). Universitas Negeri Yogyakarta, Indonesia.
- 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. Retreived from <http://www.eajournals.org/wp-content/uploads/Analyzing-Indonesian-English-Abstracts-Translation-in-view-of-Translation-Errors-by-Google-Translate.pdf>
- Sandrina, P. (2006). *Website translation: A new training challenge*. Retreived from https://www.researchgate.net/publication/258106669_Website_Translation_a_New_Training_Challenge
- Tempo (2017). *Google Translate Indonesia adopts NMT technology*. Retreived from <https://en.tempo.co/read/news/2017/04/28/310870526/Google-Translate-Indonesia-Adopts-NMT-Technology>
- Vidhayasai, T., Keyuravong, S. and Bunsom, T. (2015). Investigating the use of Google Translate in “terms and conditions” in an airline’s official website: errors and implications. *PASAA*, 49. Retreived from <https://files.eric.ed.gov/fulltext/EJ1077919.pdf>
- Wu, Y. et.al. (2016). Google’s Neural Machine Translation System: Bridging the gap between human and Machine Translation. *arXiv:1609.08144v2*. Retreived from <https://arxiv.org/abs/1609.08144>
- Yusran, N. (2017). *An error analysis of legal terminology translation using Google Translate from English to Indonesian* (Unpublished thesis). Universitas Islam Negeri Syarif Hidayatullah, Indonesia