

Gender Bias In Machine Translation (Google Translate) From Indonesian To English

Nurtamin

Faculty of Cultural Sciences, Universitas Hasanuddin
nurtamin20f@student.unhas.ac.id

Herawaty Abbas

Faculty of Cultural Sciences, Universitas Hasanuddin
herawaty@unhas.ac.id

Ery Iswary

Faculty of Cultural Sciences, Universitas Hasanuddin
ery.iswari@gmail.com

Muhammad Hasyim

Faculty of Cultural Sciences, Universitas Hasanuddin
hasyimfrance@unhas.ac.id

Abstract

This study aimed to explain the gender translation pattern from Indonesian to English in machine translation (Google Translate) and the machine translation bias form in gender translation. The method used is descriptive qualitative with Sangkuriang folklore as the object. The primary data were Sangkuriang folklore translation result from Indonesian to English using Google Translate and the secondary data were books or references related to machine translation, language, and gender. Data were collected by documentation technique, document analysis, and note-taking technique. The results showed that Google Translate had a tendency to identify the word "*dia*" in Indonesian for the male gender "he" in English. Although the source language has provided context for the clause before the keywords 'he' and 'she', the machine translator still translated the word '*dia*' in the source language into the word 'he' in the target language. Besides, Google Translate still embedded a positive gender depiction towards the male, namely by depicting brave, optimistic, tough, able to take risks, and likes challenges. Meanwhile, the depiction of negative traits was addressed to the female. In addition, it mapped the roles of men and women where men were as leaders because they are brave, tough, and like challenges; while women were as subordinate figures who were only able to choose, but did not have full power to determine and lead.

Keywords: gender bias, machine translation, Google Translate, gender and language

1. INTRODUCTION

The development of globalization era also provides an influence on patterns of interaction and communication between humans. One of the influences brought about by globalization is the influence of language. This can be seen in the increasing use of foreign languages in aspects of life. For example, in an academic context, students and lecturers have an obligation to upload journals in two language versions, namely Indonesian and English. For those who have minimal English language skills, they often use translation tools, such as

Google Translate, Bing Translate, U-Dictionary, and others to help prepare their journals so that they are immediately accepted and published.

One of the most popular translation machines that attracts people's attention is Google Translate. Based on information compiled from the official Android Police website, www.androidpolice.com, a technology information provider site, data is obtained that Google Translate users globally have reached 1 billion as evidenced by the number of downloads of the Google Translate

application on the Play Store in 2021. Furthermore, the Google Translate users are spread into various professions such as academics, professionals, entrepreneurs, foreign tourists, to football fans.

Google Translate was chosen as the object of this research because this translation tool is quite widely known and used by people from various sectors. Ayuwuragil (2017) cited from the official online news website CNN Indonesia stated that in its journey, Google Translate continues to experience development and refinement in order to provide maximum and accurate translation results. It is known that previously Google Translate used PBMT (Phrase-Based Machine Translation) technology which works by translating word by word or phrase by phrase. Over time, Google Translate changed PBMT technology into NMT (Neural Machine Translation). Damar (2017) on the official online news website Liputan6.com stated that Macduff Hughes claimed that the use of Google Translate for Indonesian was supported by NMT technology. This technology can make machine translation results close to human translation results.

The translation results produced by machine translation tend to favor certain parties who are dominant (male) and marginalize other parties (female). This reinforces the fact that translation tools, such as Google Translate, produce sexist language. Thomas and Wareing (2007:106) state, "sexist language is a language that represents men and women unequally so that members of one sex group are considered lower in humanity, simpler, and have fewer rights than members of another sex group." Sexist language usually presents stereotypes about men and women that are sometimes detrimental to both, but more often to women.

The object of this research is a folklore text from West Java, namely Sangkuriang. This story is about a boy who falls in love with his own mother, Dayang Sumbi. In order to prove his love and fulfill his wish for Dayang Sumbi, Sangkuriang accepted the challenge to build a big boat and the boat had to be finished before dawn. This Sangkuriang folk tale was chosen as the object of research because this story has a gender theme that seeks to explore the depiction of male and female figures in the story. Therefore, this gender-based folklore is considered very relevant to the research topic to be studied.

2. LITERATURE REVIEW

2.1 Concept of Gender and Sex

In general, the majority of people have the notion that gender and sex are similar. However, in principle, these two concepts have different terminology. Fakihi (2013: 7) distinguishes between gender and sex. The definition of sex means the characterization or division of two human sexes based on inherent, unchanging and non-interchangeable biological characteristics. In this case, it is often said to be God's provision or 'nature'. While gender is an inherent trait of men or women that is socially and culturally constructed and can be exchanged.

On the other hand, Wiasti (2017) stated that sex is defined as the biological sex that distinguishes between men and women, while gender is the difference in roles, rights and obligations, power and opportunities between men and women in social life. For more details, Wiasti (2017: 32) classified the differences between gender and sex in the table below.

Table 2.1 The Differences between Gender and Sex

Sex	Gender
It is a biological difference between men and women	It is the difference in roles, rights and obligations, power and opportunities between men and women in community life
Sex differences are the same all over the world that women can get pregnant, while men cannot; universal in nature.	Gender is not the same all over the world, depending on the culture and development of people in one region, it is local.
Sex differences do not change over time. From the past until now and in the future, men do not experience menstruation and cannot get pregnant	Gender changes from time to time. Every event can change the relationship between men and women in society

Source: Wiasti, Ni Made. 2017. *Mencermati Permasalahan Gender dan Pengarusutamaan Gender (PUG). Sunari Penjor 1 (2017), p. 32.*

2.2 Gender Bias

Gender bias can be interpreted as unequal positions and roles assigned to men and women. In social life, it is not uncommon to find perceptions or assumptions that certain genders play a more important role than other genders. With this condition of inequality, this is what causes the emergence of gender bias.

Gender bias arises because of the social and cultural construction adopted by the

community. However, unfortunately, this social construction is one-sided, so that it makes other genders cornered. Regarding this social construction, Wiasti (2017: 34) presented examples of social construction related to the characteristics, nature of work, norms, and values of each gender and the prevailing reality.

Table 2.2 Gender Characteristics and Traits

Social construction of	Male	Female	Reality
Characteristics	Short hair, wearing long pants	Long hair, wearing a skirt and earrings	There are men with long hair and wearing earrings, there are also women wearing long pants and short hair
Traits	Firm, tough, strong, rational	Gentle, sensitive, whiny, shy	There are men who are whiny, gentle, sensitive, shy; there are also women who are firm and strong
Occupations/roles	Making a living (public)	Doing household chores (domestic)	There are men who are unemployed, doing housework; and there are also women who work for a living or both earn a living
Norms	Heir	Not an heir	In Minang, women are heirs, but in Bali, women have no right to inherit; whereas in Java, both men and women inherit.
Status	As the leader, the family head	Husband's companion, housewife	There are women who are leaders and there are men who are led.
Value	First-class creature	Second class creature	Equal

Source: Wiasti, Ni Made. 2017. *Mencermati Permasalahan Gender dan Pengarusutamaan Gender (PUG). Sunari Penjor 1 (2017), p. 34.*

2.3 Nature of Translation

Regarding translation, several experts have presented their perspectives regarding the definition of translation. According to Catford (1965: 20) translation is an activity of transferring a text that has similarities between the source language and the target language, where the important thing in this activity is the similarity or equivalent. In addition, Munday (2006: 5) argued that translation is a change from written text using verbal language in the source language to written text using verbal

language in the target language. From these two definitions, it can be concluded that translation is the process of transferring or changing a text from the source language to the target language by taking into account the aspects of equivalence or similarity.

Basically, translation is not an activity of translating word for word from the source language to the target language. Ideally, translation is the activity of translating the meanings and messages contained in the source language and then expressing it by

looking for its equivalent in the target language. On the other hand, translation is not about simply transferring or reproducing words from the source language to the target language, but about transferring words from the source language to the target language by taking into account the context outside the language. The context referred to here is the context of gender and culture.

2.4 Translation Equivalence Theory of Baker (2011)

According to Baker (2011: 6), one of the most important components related to equivalence is the translator's knowledge in the translation process. Translators are required to see each unit of words and morphemes from the source language to be equated with the target language. There are three aspects that need to be considered by translators in translating, namely 1) internal aspects consisting of aspects of words and phrases; 2) context aspect which consists of semantic and pragmatic aspects; and 3) non-language aspects that include cultural aspects of the source language adapted into the target language.

Baker (2011: 4-5) divides translation equivalence into five types, namely as follows.

A. Equivalence at word level

As a translator, in conveying the meaning as a whole, the translator is asked to carry out the decoding process of the unit or structure that has meaning to the recipient of the target language (Baker, 2011: 9). At this level, Baker provides a distinction between the definition of a word and a morpheme. Baker (2011: 9-10) defines the word as the smallest unit that has its own meaning; while the morpheme as a minimal formal element of the meaning of a language that may or may not have several elements of meaning. In other words, the significant difference between morpheme and word is that morpheme cannot have more than one meaning element and cannot be analyzed. In English, morphemes have several grammatical functions. Baker (2011: 10) suggests that morphemes can function as a plural word pointer (funds), a specific gender indicator (manageress), and an adverb of time (considered).

B. Equivalence at the level above the word

Equivalence at the level above the word corresponds to the equivalence at more than one word. Baker (2011: 51) suggested that a language unit in the form of more than one word or a collection of words does not occur by itself, but is influenced by other words. Furthermore, these words are not formed randomly, but are formed based on rules or boundaries with the aim of conveying meaning to the reader. This one applies to English. As is well known that English has a grammatical rule that determiners should not be placed after nouns. For example in the phrase "beautiful girl the". This phrase is clearly wrong because it does not comply with the rules or restrictions that apply in the English language. In particular, the difference in lexical patterns from the source language to the target language makes it difficult for translators to translate.

C. Grammatical equivalence

Lexical sources are not the only influencing factor in analyzing language equivalence. Another strongest factor that determines the types of language equivalence differences is the grammatical system of a language. Each language has different language equivalence criteria such as time, number, gender, form, visibility, persona, proximity, and so on.

D. Equivalence at text level

In translation, it is not only the meaning and syntactic structure in the text that gets attention. However, translators also need to pay attention to the themes and information contained in the text (Baker, 2011: 131). Equivalence at the text level consists of two types, namely: 1) thematic structure and information structure, and 2) cohesion. Thematic structure relates to how the message from the author or speaker is conveyed, while the information structure relates to how the point of view of the reader or listener. Furthermore, the thematic structure is further divided into two types of sub-categories, namely themes and rhemes: themes are topics and rhemes are messages (Baker, 2011: 131).

E. Pragmatic equivalence

Pragmatics is the study of the use of language. In other words, pragmatics is the study of meaning, which is not obtained from the linguistic system but is obtained and

manipulated by participants in a communication (Baker, 2011: 230). By focusing on the difficulty areas of translators in cross-cultural communication, pragmatics is divided into two, namely coherence and implicature.

2.5 Gender and Language Theory of Lakoff (1973)

One of the most influential figures in the field of language and gender is Robin Lakoff. Robin Lakoff is a professor of Linguistics at the University of California, Berkeley. Through his book entitled *Language and Woman's Place*, Lakoff is considered a pioneer in the birth of the field of language and gender in linguistics.

Lakoff (2003: 45) stated that women are part of the marginalized even down to the smallest part, namely language. The marginality and powerlessness of women is reflected in how women are expected to speak and how women are spoken of. These aspects of marginality can be explored and analyzed from the lexical and syntactic aspects.

According to Lakoff, women acquire unequal behavior from several aspects of life; One of them is related to work. Let us consider the following example:

He's a professional (1)

She's a professional (2)

In the example of clause 1, it can be concluded that the word "he" in the example is assumed to be a man who has a professional job, such as a doctor or lawyer. This is different from the example of clause 2. In the example of clause 2, it can be concluded that the word professional associated with women in the word "she" may have assumptions that have negative connotations. The most likely assumption is that women have jobs as prostitutes. In this context, women are re-identified with their sexuality which is not equal to men.

3. RESEARCH METHODOLOGY

This research is a qualitative descriptive study (Hasyim et al., 2021). Qualitative research is research that seeks to explore and understand the meaning of the behavior of certain elements. According to Sugiyono (2010: 9), qualitative research is research

where the researcher is placed as a key instrument, data collection techniques are carried out in combination and data analysis is inductive. Meanwhile, descriptive research is research that describes a phenomenon, event, and symptom (Sugiyono, 2016: 336).

The data source in this study is the result of translating the Sangkuriang folklore text from Indonesian to English using the Google Translate machine. The primary data source in this study is the translation of the Sangkuriang folklore in the form of clauses from Indonesian to English using the Google Translate machine. Meanwhile, the secondary data in this study are books or references related to Google Translate, translation, language and gender to determine the existence of gender bias in the translation of folklore texts from Indonesian to English using the Google Translate machine.

The sample of this study is the result of translating the Sangkuriang folklore text using Google Translate in the form of clause verbs, adjective verbs, noun verbs, and personal pronouns. This study uses a purposive sampling technique, namely data sampling based on certain considerations.

The data collection techniques used in this study to collect the required data consist of documentation technique, document analysis, and note-taking technique.

4. RESULTS AND DISCUSSION

Patterns of Gender Translation by Google Translate

In the Antconc application, the frequency feature is provided to find out how often a word appears and how dominant it is when compared to other words that you want to research. Because this research is related to gender translation, the two keywords that are the focus of this research are on the singular third-person pronouns, namely the words 'he' and 'she'. After processing the data using the Antconc application, it was found that the frequency of occurrence of the word 'he' was 51 times and the frequency of occurrence for the word 'she' was 9 times. This shows that the translation of Google Translate still has a tendency to translate the word '*dia*' from Indonesian into the word 'he' into English..

Clause and Language and Gender Analysis

As previously explained, the type of clause used in this study is a clause based on

the word or phrase that is the predicate. Based on the classification using the Antconc application on the results of the translation of the Sangkuriang folklore text using Google Translate, 34 verb clauses, 1 noun clause, and 2 adjective clauses were obtained with the type of noun being the focus is the single third-person pronoun.

It is generally known that Indonesian does not have a clear gender designation for the word 'dia'. However, there are several clauses or sentences in Indonesian that have a gender indication by looking at the context of the previous clause or sentence. However, Google Translate has not been able to consider and see the context. One of the proofs can be seen in the translation of the clause 'Dia menikah dengan Si Tumang yang ternyata titisan dewa' from Indonesian which translates to 'He married si Tumang who turned out to be an incarnation'. If it is seen from the clause or sentence before the clause above, it is known that there is a context that has been embedded in the Indonesian language, namely in the snippet of the clause 'Tidak ada yang dapat diperbuat Dayang Sumbi selain memenuhi ucapannya (There is nothing Dayang Sumbi can do other than fulfill her words)'. Here it is clear that there has been a clear gender indication for the word 'he', namely Dayang Sumbi. However, Google Translate does not translate it to 'she' but translates it to 'he'.

The second feature in the Antconc application is the concordance feature. In this feature, users can see what words or sentences are embedded before and after the searched keyword. Two keywords in this research to be included in the Antconc application are the words 'he' and 'she'. After being analyzed, there are two data that show the confusion of the Google Translate machine translator in doing translations related to gender. First, it can be seen in the translation of the clause 'Untuk mengisi waktu luangnya selama dalam pengasingan, dia pun menenun' which is translated into 'To fill his spare time in exile, he also weaves'. In this clause, context has been provided before the clause related to gender designation. In Indonesian, it is clear that the gender indicator is female. However, after being translated, the gender indicator shows the opposite, namely being translated into male gender which is indicated by the word he. Here, it can be seen that in terms of concordance, although the source language has

provided context before the keywords in the form of gender indicators, Google Translate is still not able to make equivalence from the source text to the target text. Therefore, it can be concluded that Google Translate has a tendency to prioritize the translation of the word "dia" in Indonesian by identifying it with the word "he" in English.

Forms of Gender Bias by Google Translate

Based on the results of the translation of the Sangkuriang folklore text using Google Translate, it can be seen that Google Translate provides more descriptions of male gender characteristics. This is evidenced by the more use of the word 'he' in the target language for the word 'dia' in the source language. In terms of depicting gender characteristics, it can be seen that the male gender is described as a person who is brave, tough, full of confidence and optimism, and able to take risks. This depiction of the male gender trait is found in some of the analyzed data.

Based on the aspect of depicting gender traits obtained from the translation of the Sangkuriang story text using Google Translate, it can be concluded that the depiction of positive traits is still aimed at the male gender such as optimistic, brave, tough, able to take risks, and likes challenges. Meanwhile, the depiction of women's gender characteristics is negative, namely women are depicted as figures who are the source of the occurrence of a problem. Therefore, it can be said that gender translation carried out using Google Translate is still biased towards the female gender.

Based on the division of male and female gender roles from the Sangkuriang folklore text, it can be revealed that there is still an ideal role mapping that can only be occupied by men and women. The role mapping lies in the leader's position where Google Translate has the perception that the leader position is identical to the male gender and the subordinate position is identical to the female position. In fact, the reality shows the opposite. Therefore, it can be concluded that Google Translate still shows biased translation results in clauses related to gender.

Based on the aspect of expectation of the gender nature, it can be said that the results of the Google Translate translation indirectly indicate that the female gender is expected to be a person who has a gentle, calm, and patient

nature in dealing with a problem. In fact, it would be unethical if a gender is expected or expected to have socially constructed characteristics. This trait should return to each individual without having to look at the gender of the person.

Translation Equivalence

After analyzing the translation quality of Google Translate and presenting it in the research results section, it is found that Google Translate still does not meet the equivalence or equivalence between the source language and the target language. The equivalence theory used here is the translation equivalence theory of Baker (2011) which divides equivalence into five types, namely equivalence at the word level, equivalence at the level above the word, grammatical equivalence, equivalence at the text level, and pragmatic equivalence.

After examining the data, it was found that the types of equivalences that have not been met by the Google Translate machine translator are equivalence at the word level and grammatical equivalence. Based on the results of the study in the form of clauses that indicate the inability of the Google Translate machine translator in creating equivalence or equivalence, it can be concluded that Google Translate's claim that it has been able to resemble the results of human translation is still not acceptable and still needs to be refined. This is because there are still many results of the translation of Google Translate clauses that have not shown equivalence at the word level and grammatical equivalence, especially in the translation of clauses related to gender.

5. CONCLUSION

From the aspect of the translation pattern formed, it can be concluded that Google Translate has a tendency to identify the word 'dia' in Indonesian for the male gender 'he' in English compared to the female gender 'she'. On the other hand, it was found that Google Translate also tends to translate the word 'dia' into 'he'. Even though the source language has provided context for the clause or sentence before the keywords 'he' and 'she', the machine translator still translates the word 'dia' in the source language into the word 'he' in the target language.

Meanwhile, from the aspect of word class or clause, it can be seen that Google Translate

still embeds a positive gender depiction of the male gender, namely by depicting brave, optimistic, tough, able to take risks, and likes challenges. Meanwhile, the depiction of negative traits is aimed at the female gender. In addition, the results obtained that Google Translate maps the roles of men and women where men are considered as leaders because they are brave, tough, and like challenges; while women are considered as subordinate figures who are only able to choose, but do not have full power to determine and lead.

ACKNOWLEDGMENT

The authors would like to express gratitude for the full support and motivation of the Indonesia Endowment Funds for Education (LPDP) – the Ministry of Finance of the Republic of Indonesia as the main funding sponsor of this paper publication.

REFERENCES

- Ayuwuragil, Kustin. 2017. Teknologi NMT Bikin Google Translate Makin Akurat. Retrieved from <https://www.cnnindonesia.com/technology/20170427170031-185-210704/technology-nmt-bikin-google-translate-makin-akurat>. [Accessed: April 8, 2022].
- Baker, Mona. 2011. In *Other Words: A Coursebook on Translation* Second Edition. New York: Routledge
- Catford, J. 1965. *A Linguistic Theory of Translation*. London: Oxford Univeristy Press
- Damar, Augustine Mario. 2017. Makin Akurat, Google Translate Didukung Kecerdasan Buatan. Retrieved from <https://www.liputan6.com/tekno/read/2933941/makin-akurat-google-translate-disupport-kecerdasan-buatan>. [Accessed: May 4, 2022].
- Fakih, Mansour. 2013. *Analisis Gender & Transformasi Sosial*. Yogyakarta: Pustaka Pelajar
- Lakoff, Robin. 1973. *Language and Woman's Place*. United Kingdom: Cambridge University Press.
- Hasyim, M., Saleh, F., Yusuf, R., Abbas, A. 2021. Artificial Intelligence: Machine Translation Accuracy in Translating French-Indonesian Culinary Texts. *International Journal of Advanced Computer*

- Science and Applications*, 2021, 12(3), 186-191
- Munday, J. 2006. *Introducing Translation Studies: Theory and Applications*. Oxon: Routledge
- Sugiyono. 2010. *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif dan R & D*. Bandung: Alfabeta
- Sugiyono. 2016. *Cara Mudah Menyusun: Skripsi, Tesis, dan Disertasi*. Bandung: Alfabeta
- Thomas, Linda and Shan Wareing. 2007. *Bahasa, Masyarakat, dan Kekuasaan*. (Translated by Ed. Abdul Syukur Ibrahim). Yogyakarta: Pustaka Pelajar
- Wiasti, Ni Made. 2017. *Mencermati Permasalahan Gender dan Pengarusutamaan Gender (PUG)*. *Sunari Penjor*, Vol. 1 No. 1, 29-42