Translation Error Types Analysis on *TikTok* Indonesian-English Auto-Translation Content

Dirahima Berida Shafira
English Study Program, Faculty of Humanities, Universitas Dian Nuswantoro
Jl. Imam Bonjol No.207, Pendrikan Kidul, Kec. Semarang Tengah, Kota Semarang, Jawa Tengah 50131
311202002199@mhs.dinus.ac.id

Raden Arief Nugroho
English Study Program, Faculty of Humanities, Universitas Dian Nuswantoro
Jl. Imam Bonjol No.207, Pendrikan Kidul, Kec. Semarang Tengah, Kota Semarang, Jawa Tengah 50131
2arief.nugroho@dsn.dinus.ac.id

Abstract—**This study investigates the types of errors that are common in translations generated by automatic translation software like TikTok. The researcher used a classification of types of translation errors (D.Vilar, 2006). There are several translation errors, including missing words, incorrect word order, unknown words, and incorrect punctuation. This research data is taken from TikTok which has high viewers or is commonly called for your page (fyp). The source text and target text from TikTok auto-translate were aligned and analyzed in terms of error types. The method of this study is descriptive-qualitative. The researcher analyzed the translation results from the auto-translation of 10 data and classified errors based on the accuracy of transferring the message. This error cannot be ignored because translation deals with conveying the same messages from one language and making them have the same meaning in the target language. Methods of data collection using the methods of observation and note-taking techniques The result is that there are translation errors, including missing words, incorrect word orders, unknown words, and punctuation that are not found in the results of the analysis that has been carried out.**

Keywords—Translation Error, TikTok Auto-translate Error, Machine Translation Error

I. INTRODUCTION

Translation in today's era can be done not only by humans but also by machine translators, commonly known as "auto-translate." (Nida and Taber, 2004) define translation as "reproducing in the receptor language the closest natural equivalent of the source language message, first in terms of meaning and secondly in terms of style." From this definition, it can be concluded that there are at least two factors that play a role in translation, namely the sender or source language to be translated and the recipient or target language to be translated by the translator. The translation is the same as how we communicate to provide information to recipients in text form. That's why they both need someone who can help them translate the language so that the message will be understood by the recipient.

But in modern times, there is already machine translation, commonly called MT. Machine translation is to provide a system that translates text from a source language into a target language such that the translation expresses the same meaning as the text in the source language (Irfan, 2017). Machine translation has many benefits, including the ability to save time, money, and many language choices. However, this machine translation may produce poor results when the source text contains errors or uses words that are not in the built-in dictionary. The only way to fix this is to manually update the dictionary regularly. Many social media applications, such as Facebook, Instagram, Youtube, and Netflix, support automatic translation features through machine translation to make it easier for users to communicate in different languages. In addition to the applications mentioned, the TikTok application already has an automatic translation feature that is carried out via machine translation (MT). TikTok is a social networking application and music video platform where users can create, edit, and share short video clips complete with filters and accompanied by music as support. TikTok calls these automatic subtitles "auto-captions." This aims to be able to quickly translate one language into another. Apart from that, this feature can also translate the words spoken by the creator in the TikTok content, which will then be automatically transcribed by the application.

This is very closely related to language; according to (Wibowo, 2001), language is a system of symbols that are meaningful and articulate sounds (generated by the said tool), which are arbitrary and conventional, and

which are used as a means of communicating by a group of human beings to give birth to feelings and thoughts. That is why translation is needed to connect people of different languages. There are many languages in this world; there are even 36 languages that can be translated into TikTok subtitles. Of course, with real-time translation, which happens at that moment, the results of the translation of TikTok subtitles are not necessarily accurate or there are errors. Translation errors are when the translation appears to differ from the original text in terms of its linguistic meaning, either at the word or sentence level. Translations can be labeled as errors due to imprecision or inaccuracy of the source text, misuse of phrases, misuse of official or official lists, or because they lack coherence or cohesion (Newmark, 1998). The limitations of machine translation include the limited time to re-examine. This limitation makes it interesting and challenging to examine how TikTok subtitles work in translating conversational messages into written subtitles. The main focus of this study is to investigate the types of errors commonly found in the automatic translations of TikTok by analyzing translations from Indonesian to English. Practically, this research is expected to be used as a reference for further researchers who are interested in the same field, namely translation errors in translation studies.

II.   LITERATURE REVIEW

The following are some previous studies on machine translation errors, which found many translation errors in machine translation. There have been several studies of translation errors on YouTube, Google Translate, and other machine translations. (Ary, Donald, Lucy Cheser Jacobs, and Asghar Razavieh, 1979) said that "the search for related literature should be completed before the actual conduct of the study begins." This part discusses some of the previous studies that concern the analysis of translation errors in machine translation. Therefore, they can be used as a reference or guide in this research.

A.   Review of Previous Studies 1

The topic related to the error analysis approach in machine translation was carried out by previous researchers, including Kezia Ruth Peneduh Sukma (2019), who examined errors that often occur on YouTube subtitles by comparing two different videos, namely Vogue Beauty Secret and Vogue Perfect Breakfast, using error analysis. M Koponen (2010), shows the results of research showing that the type of error that most often occurs in the two videos is a concept translation error. In the translation process, YouTube subtitles make mistakes in various aspects, such as inaccuracies in understanding words that are pronounced incorrectly, expressions (idioms), and words that have double meanings. It can be concluded that YouTube subtitles have a better performance at translating food videos than beauty videos.

Almost the same research was also conducted by Sependi Napitupulu (2017). This study seeks to investigate the frequency of errors in the translation of abstracts produced by Google Translate concerning the Keshavarz (1999) error model. However, the researcher focuses on exploring Google Translate errors in translating undergraduate course summaries from Indonesian to English.

Previous studies indicated that tenses are another area where machine translation struggles. This supports research that demonstrates that YouTube's machine translation suffers from misunderstandings of words that are pronounced improperly. Using the incorrect tense results in morphological faults known as bad tenses. Previous studies have also demonstrated that machine translation has issues with target text accuracy and word order. Researchers studied TikTok subtitles and employed theory from (D. Vilar, 2006) to analyse them, which differs from earlier research that was previously acknowledged by researchers and produced different findings.

III.   METHOD

In conducting this research, the authors used a qualitative descriptive method using a language error analysis approach to machine translation. A qualitative descriptive design may be considered most appropriate because it recognizes the subjective nature of the problem and the different experiences participants have, and it will present the findings in a way that directly reflects or closely resembles the terminology used in the original research question (Bradshaw C, 2017). The data collected in this study is objective data, which means it was taken from the source text and target text. In this case, the focus of the error analysis is on the translation errors made by TikTok subtitles from Indonesian to English. The field of translation research used in this research is translation and technology. The research object is the subtitles in the TikTok videos made by the TikTok subtitles. The video that is taken to serve as data has a different type, depending on whether the translation error appears in different variations. In analyzing the data, the researcher uses the classification of types of translation errors (D.Vilar, 2006) to identify and classify errors that occur in TikTok subtitles. Translation errors are classified into five categories proposed by (D.Vilar, 2006), namely "missing words," "word order," "incorrect words," "unknown words," and "punctuation." Data is collected by researchers to provide transcripts or original data; the source text is taken from

TikTok videos. The target text and the target text are arranged in tabular form and analyzed according to the type of error. The results of the analysis are shown as proportions.

IV.   RESULT AND DISCUSSION

A. Error Types Analisis

This study analyzes the types of errors found in the translation by the TikTok machine or auto-translation from Indonesian to English. This study also analyzes whether the error is a single error or multiple errors. Single errors are errors that occur individually. not that it was caused by another error and did not cause any other error. Meanwhile, some errors are errors that cause other errors. That means the errors are interconnected. The data was then analyzed and categorized according to Vilar et al. (2006). What often happens is that wrong words take the form of wrong lexical choices and wrong forms. The following is a table of results from the analysis of 15 sets of translation data.

TABLE I.   RESEARCH TABLE RESULT

No.	Error Analysis		
	Types of Error	*Error Frequency*	*Percentage*
1.	Missing Word	5	33,4%
2.	Word Order	1	6,6%
3.	Incorrect Word	4	26,7%
4.	Unknown Word	4	26,7%
5.	Punctuation	1	6,6%
Total		**15**	**100%**

From these results, it can be seen that the most common types of errors are missing words, incorrect words, unknown words, and word order. However, this type of punctuation error was not found in this analysis.

a)   Missing Words

A missing word occurs because an error is generated when some words in the resulting sentence are missing. Then this error has two types of errors, namely when the missing words are very important to express the meaning of the sentence or content, and when the missing word is only needed to form grammatically correct sentences, but the meaning is retained, or filler. Important words that help explain the significance of a phrase are known as content words. Filler words are those that are required for the sentence to be grammatically correct.

(1) **ST** : Cipung udah bisa bantu mama Gigi endorse.
   TT : **Can you help mama teeth** endorse?
   (@anakdesa at 14/1/2023 on TikTok)

The translation here comes out of the message that must be conveyed. If the target text removes the word "Cipung," which is the name of someone who is the subject in this sentence, then "can you help" should be "has been able to help" which is more appropriate than "can you help" which is more towards offering to do the favors. "Mama teeth" does not refer to the mother's teeth, but the mother of cipung is "Gigi" which is a person's name. This sentence is an example of an error in content words that are translated into nouns. So far, TikTok has not been able to translate someone's name that has an element of object context.

(2) **ST** : Cipung bangun tidur masih ngelag ya, ga nangis ditinggal papa sama mama pergi!
   TT : Cipung woke up still lagging, **ga crying.**
   (@anakdesa at 14/1/2023 on TikTok)

An error was found in the translation results; this time the resulting target text was missing, namely "ditinggal papa sama mama pergi!" which is the message that must be conveyed. Another error found is "ga crying". What is meant is not crying. The correct translation is "does not say". Two errors are found at once in this text, namely that content that should be important to translate is not in the target text sentence.

(3) **ST** : Mille crepes viral belum buka udah antri sejam lebih.
   TT : Viral mille crepes have not been opened for over an hour.
   (@pogocake at 2/3/2023 on TikTok)

The word "antri" is not translated into the target text, so the word "antri" should be translated as "waiting." This includes translation errors because the word "waiting" is a very important word to express the meaning of a

sentence. This is included in the missing words because the content words should be translated into the target language.

(4) ST : Mbak Pungky, disidang kan di pengadilan. Mbak Pungky, adalah salah satu korban penipuan dari sikembar Rihana Rihani.

TT : Ms. Pungky, tried in court. Ms. Pungky, is one of the victims of fraud from **Sikembar** Rihana Rihani.

(@info_updated at 18/7/2023 on TikTok)

A translation error was found in the "sikembar" words which are not translated, should be in the target text the translation is "the twins". And then grammatical errors were also found, after the first word, namely Ms. Punky followed by "was". Additionally, it makes clear the sentence's paradoxical meaning. As a result, the source language's intended meaning is not communicated.

(5) ST : Kalo kita putus aku langsung ada yang baru, bukan berarti aku selingkuh, **kan aku udah bilang yang nungguin kita putus banyak!**

TT : If we break up I immediately a new one. It doesn't mean I'm cheating.

(@faizalramaa at 13/7/2023 on TikTok)

The mistake this time lies in the last sentence which is not translated "kan aku udah bilang yang nungguin kita putus banyak!". Missing words this time occurs not in words but in sentences.

b)   Word Order

This type of error represents the incorrect structure of the sentence. The error occurs when the word order that occurs in the target text is in the wrong position.

(6) ST : Konten Agung Hapsah langganan jadi trending dan kita selalu penasaran resepnya

TT : Agung Hapsah **subscription** content is trending and we are always curious about the recipe.

(@narasi at 20/1/2023 on TikTok)

There are also translation errors in the target text, the source text for the word "subscription" here means that the content that is trending is not the subject that subscribes to the content. This error is the word order does not match the position it should be. That it was a noun phrase, but failed to realise in the text.

c)   Incorrect Word

It is found when the system cannot find the correct translation of the given word. In classified incorrect words according to (D.Vilar, 2006) it is divided into six categories namely wrong lexical choice, incorrect disambiguation, incorrect form, extra words, style, and idiom.

(7) ST : Happy birthday, Xabiru! Jagain Chava sama buna ya… Biru keren banget pokonya.

TT : Happy birthday, Xabiru! Take care of Chava and Buna **ya**… **Blue** is really cool.

(@cindysoebagio_ at 15/3/2023 on TikTok)

In the source text sentence, there is an error in the word "Biru". "Biru" here shows someone's name, namely Xabiru, but the target text translates the name "Biru" to "Blue" which shouldn't need to be translated. In addition, another error word was found, namely "ya..." was not translated into the target text and remained "ya...", the word "ya..." can be translated "okay".

(8) ST : Kemajuan ekonomi Indonesia membuat gue "kesel".

TT   : Indonesia economic progress makes me **"kesel".**

(@satriong at 16/5/2023 on TikTok)

The translation error is found in the word "kesel", which is the word "disappointment" which should be the target text, namely "annoyed". The causes in the text are wrong disambiguation

(9) ST : Perdana nih a day in my life aku yang agak gabut ini.

TT : A day in my life, I'm a **little excited.**

(@itsokbitj at 8/3/2023 on TikTok)

Based on the example above, the translation of "perdana nih" is omitted. Therefore, the message from the source language that this is the first time is not conveyed. Then the word "agak gabut" is not right in the target text which is translated to "little excited", "little exited" can be translated to "bit confused".

(10) ST : Ketua komnas HAM Ahmad Taufan Damanik membantah ketika disebut seolah sebagai pengacara Ferdy Sambo. Menurutnya, saat ini yang paling adil adalah menguji soal dugaan pelecehan seksusal terhadap Putri Candrawati. Apakah rekomendasi Komnas Ham sudah tepat dan adil bagi almarhum Yosua Hutabarat?

TT : Komnas HAM Chairman Ahmad Taufan Damanik denied when referred to as lawyer Ferdy Sambo. According to him, at this time the fairest is to test the alleged sexual harassment of **princess candrawathi**. Is the recommendation of Komnas Ham appropriate and fair the late Joshua Hutabarat. (@KompasTV at 24/8/2022 on TikTok)

In the data source text, there is the name "Putri Candrawati" which is translated into the target text as "princess candrawathi". This is a translation error because someone's name should not need to be translated into the target text.

d) Unknown Word

Are also a source of errors. Unknown word error referring to untranslated words. These words should not be in the target text. Unknown words are those that have been found but are regarded as unknown since they are not listed in dictionaries.

(11) ST : Seorang pemilik kendaraan melapor ke Polres Metro Depok setelah kendaraannya mengalami kerusakan, baret di bagian sebelah kiri.
TT : A vehicle owner reported to Depok Metro Police after his vehicle was damaged, **baret** on the left.
(@depokupdate at 13/08/2023 on TikTok)

The word "baret" in the source text remains the same in the target text it should be translated as "scratched" which means "baret, lecet". but in the target text "beret" is not translated. It is classified as a type of error of unknown word.

(12) ST : Masih inget banget dulu awal main tik tok dikata katain karena bikin telur geprek ini, ya tapi gimana ya, tetep recook karena memang seenak itu. Kalau gasuka tinggal skip aja sih.
TT : I still remember the beginning of the game tick tok said because it made this geprek egg, but what you do, it still recook because it is delicious. If **gasuka** just skip it.
(@hi_erisa at 15/07/2023)

The word "gasuka" means "dislike" but TikTok's auto-translation cannot translate the word. That word should be translated as "dislike this video". Because the meanings are unknown to the readers, it is categorised as an unknown word.

(13) ST : Ya aku emng cowo yang mudah baper, aku sering posesif sering cemburu karna akuu sayang sama kamu, kalo kamuu risih maaf ya.
TT : Yes I am a guy who is easily **baper,** I am often possessive often jealous because I love you, if you are uncomfortable sorry yes.
(@waktuluang at 20/6/2023 on TikTok)

"Baper" in the sentence is not translated into the target language, "baper" should be translated into "bring a feeling". This refers to an unknown word because the reader in the target text does not know the term idiom.

(14) ST : Ngerasain dinginnya kaki gunung sumbing, pas banget dapet sunrise bagus.
TT : Feel the cold **cleft mountain** foot, just right can sunrise nice.
(@jogjaexplore at 20/2/2023 on TikTok)

The word "cleft mountain" in the source text above is a translation of "Gunung Sumbing", this is a translation error because "Gunung Sumbing" should be translated into "Sumbing Mountain". Mount Sumbing is one of the mountains in Indonesia, so the resulting translation does not match the message to be conveyed.

e) Punctuation

While there are punctuation errors that refer to these errors in using signs in writing, bad word translations can also cause bad wording in the resulting sentences.

(15) ST : Total timer buat masak telor 9 mnit yak arena pas awal kurang mateng jd +4mnit lagi lupa ga kevideo.
TT : Total timer for cooking eggs **9mnit ya** because at the beginning less **mateng jd +4mnit** again forgot not **kevideo.**
(@tatadev01 at 26/1/2023 on TikTok)

The caption sentence above proves that there is a punctuation type error, that is, if there is an error in the spelling of the reading, TikTok auto-translate cannot translate into the target language. Auto translate only translates the correct writing word.

V. CONCLUSION

It can be concluded that there are four types of errors in the translation of machine translation results in TikTok media, including missing words, word orders, incorrect words, punctuation, and unknown words. The most common error that occurs is when translating someone's name. TikTok's auto-translation is lacking in translating this, so an error occurs. Then there are words originating from the source text that are not translated into the target language. Then you have to add the vocabulary that needs to be entered into the machine. Another thing that happened quite a bit was the position of the word, which didn't match the structure it should have. Errors in eliminating words that should be messages that must be conveyed also occur in TikTok media. The translation results from the machine still have quite a lot of errors, and there is no accurate translation, so the translation results are better used as pre-translations that the user still has to pay close attention to. However, auto-translation on TikTok is enough to help users understand the contents of the videos they are watching.

VI. REFERENCES

[1] Ary, Donald, Lucy Cheser Jacobs, and Asghar Razavieh. (1979). *Introduction to Research in Education.* United States of America: Holt, Rinehart and Winstons

[2] Basari, A. & Nugroho, R.A. (2017). The Use of *Aegisub* in Teaching Audiovisual Translation Classes: A Review On IT-Based Subtitling Course. *The 1st Education and Language International Conference Proceedings Center for International Language Development of Unissula.*

[3] Bradshaw C, A. S. (2017). Employing a qualitative description approach in health care research. *Global Qualitative Nursing Research*, 1–8.

[4] D.Vilar, J. X. (2006). *Error analysis and statistical machine translation output.* Universidad Politecnica de Madrid.

[5] Irfan, M. (2017). Machine Translation. *Departement of Computer Science*, 1-2.

[6] Larassati, A., Setyaningsih, N., Nugroho, R.A., Suryaningtyas, V.W., Cahyono, S.P. & Pamelasari, S.D. (2019) Google vs. Instagram Machine Translation: Multilingual Application Program Interface Errors in Translating Procedure Text Genre. *Proceedings - 2019 International Seminar on Application for Technology of Information and Communication: Industry 4.0: Retrospect, Prospect, and Challenges, iSemantic* 2019. DOI: 10.1109/ISEMANTIC.2019.8884334

[7] Nababan, M. N., Nuraeni, A., & Sumardiono. (2012). Pengembangan Model Penilaian Kualitas

[8] Newmark, P. (1998). *A textbook of Translation.* New York : Prentice Hall Inc.Terjemahan. *Kajian Linguistik dan Sastra*, 24(1)..

[9] Nida and Taber. (2004). *The Theory and Practice of Translation.* Shanghai: Shanghai Foreign Language Education Press

[10] Nugroho, R. A., Basari, A., Suryaningtyas, V. W., Setyaningsih, N., Cahyono, S. P., & Larassati, A. (2019). Translation as an Alternative to a Language-Based Vocational Course at the Undergraduate Level. 1st Vocational Education International Conference (VEIC 2019). Atlantis Press.

[11] Nugroho, R. A., Nababan, M. R., & Subroto, E. D. (2016). Translation Microstrategies Used by Visually Impaired Translators. *International Journal of Linguistics*, 6(5).

[12] Nugroho, R.A. (2010). Rapport and Address Terms in Family Guy Cartoon: Can (Targeted) Audience Identify a Social Dimension of Relationship? *Journal LITE* 6(2). Retrieved from: https://publikasi.dinus.ac.id/index.php/lite/article/download/490/993

[13] Pratama, A.A., Ramadhan, T.B.L, Elawati, F.N., & Nugroho, R.A. 2021. Translation Quality Analysis of Cultural Words in Translated Tourism Promotional Text of Central Java. *JELTL (Journal of English Language Teaching and Linguistics)*, Vol. 6(1), 2021.

[14] Shalekhah, R.A., Estayani, S.A., Sari, M., Nugroho, R.A. (2020). Linguistic Politeness Analysis of Indonesia's Prominent YouTube Influencers. *JELTL (Journal of English Language Teaching and Linguistics)*, Vol. 5(3), 2020.

[15] Suryaningtyas, V. W., Nugroho, R. A., Cahyono, S. P., Nababan, M. R., & Santosa, R. (2019). Translation Learning Enrichment Using Smart Application Creator 3.0: An Attempt to Design a Mobile Application in Translation for Tourism Purpose Course. *Proceedings - 2019 International Seminar on Application for Technology of Information and Communication: Industry 4.0: Retrospect, Prospect, and Challenges, iSemantic* 2019, 542–547.

[16] Wibowo, W. (2001). *Manajemen bahasa, pengorganisasian karangan pragmatic dalam bahasa indonesia untuk mahasiswa dan praktisi bisnis.* Jakarta: PT. Gramedia Pustaka Utama