# An Error Types Analysis on YouTube Indonesian-English Auto-Translation in *Kok Bisa?* Channel

**Naftalia Laksana & Siegfrieda A. S. Mursita Putri**
*Naftalaksana@gmail.com & Siegfrieda@ukrida.ac.id*
Faculty of Humanities and Social Sciences, Universitas Kristen Krida Wacana

## Abstract

*This study investigates the error types that commonly occur in the translation produced by YouTube auto-translate. This research uses error classifications from Vilar et al. (2006). The data were fourteen videos from "Kok Bisa?" channel. The source text and target text from YouTube auto-translate were aligned and analyzed in terms of the error types. This was a mixed quantitative and qualitative study and a primary research. The result shows that the most frequent error types are wrong lexical choice, bad word form, missing auxiliary word, short range word level word order and extra word. The other error types rarely occur in the translation.*

*Keywords: translation error, error types, YouTube auto-translate*

## Introduction

Language is fundamental to the communication in this world. There are many different languages in this world. Therefore, translation is needed to connect people with different language. One of the forms of translation is subtitle. Subtitle is to help the audience understand the content of the video or movie. YouTube, as the second most popular website (Gray, 2017), provides translation in a form of subtitle, it is named YouTube auto-translate. This study aims to investigate the error types commonly found in the YouTube auto-translation. This study analyzes the translation from Indonesian to English. This study uses the error classifications from Vilar et al. (2006).

The error classifications from Vilar et al. (2006) consists of four main classifications which are punctuation, missing word, word order and incorrect words. Missing word is divided into missing content word and missing auxiliary word. Word order is divided into word level and phrase level. Each level is divided again into short range and long range.

Incorrect word is divided into extra word, bad word form, untranslated and bad word sense. Bad word sense is divided into wrong lexical choice and bad disambiguation. In this study, punctuation is not analyzed. Therefore, there are eleven error types used to analyze the error types. This study can show the weakness of machine translation and help to improve the quality of machine translator.

## Method

The study was qualitative and quantitative study. This is also a primary research. The data were fourteen YouTube videos from *Kok Bisa?* channel. *Kok Bisa?* channel is an Indonesian educational channel which discusses many topics starting from technology, biology, history, physics and many more. This study chose fourteen videos because it was 10% of all videos published in the channel until September 2017. The source text and the target text were aligned in a table and analyzed in terms of the error types. The result of the analysis was presented in percentages.

## Error Types Analysis

This study analyzed the error types commonly found in YouTube auto-translation from Indonesian to English. This study also analyzes the emphasis of the translation, whether the translation is source language emphasis or target language emphasis according to Newmark's V diagram (1988). This study also analyzes whether the error is single error or multiple error. Single error is the error that happens individually. It is not caused by another error and does not cause other error. Meanwhile, multiple error is the error that causes another error. It means the errors are related to each other. Here is the result of the error types analysis.

| Error Type | Percentage |
|---|---|
| Missing Content Word | 3.99% |
| Missing Auxiliary Word | 19.93% |
| Short Range (word order in word level) | 9.52% |
| Long Range (word order in word level) | 2.10% |
| Short Range (word order in phrase level) | 1.88% |
| Long Range (word order in phrase level) | 0.44% |
| Extra Word | 8.97% |
| Bad Word Form | 21.93% |
| Untranslated | 4.43% |
| Wrong Lexical Choice | 24.14% |
| Bad Disambiguation | 2.66% |
| Total Error | 100.00% |

**Table 1. Error Types Analysis Result**

From the result, it is seen that the most common error types are wrong lexical choice, bad word form, missing auxiliary word, short range word level word order and extra word. The other error types rarely occur in the translation from YouTube auto-translate. Here are the examples of the most common error types.

### Wrong Lexical Choice

This is the most commonly found error type in the translation produced by YouTube auto-translate. This error happens when the machine translator chooses the wrong word choice for the target text.

(1) ST: Padahal, saat Cornelis de Houtman pertama kali berlayar ke Nusantara tahun 1596, Belanda hanya bertujuan untuk berdagang rempah-rempah.
    TT: In fact, when Cornelis de Houtman first sailed to the archipelago in 1596, Dutch only intended to trade spices.
(TG12)

In this sentence, the word "Nusantara" which refers to Indonesia was translated into "archipelago". Archipelago means an extensive group of island. The word archipelago in the target text is not explaining enough the intention of the word "Nusantara". Therefore, it is considered as wrong lexical choice. It is also a single error, it does not influence other error and is not influenced by the other error. The error happens only to the word "archipelago". This sentence is translated in source language emphasis, because the target language is following the lexical meaning of the source language. In this case, the machine translator does not deliver the context of the sentence. It makes the translation of "nusantara" becomes "archipelago" which means a which has lexical meaning "a group of islands".

### Bad Word Form

This is the second most common error found in YouTube auto-translation. This error happens when the word has incorrect morphological form.

(2) ST: "meskipun belum ada angka pasti…"
    TT: "although there are no exact figures…"
(TN4)

This sentence has two errors. Both errors are classified as bad morphological form. The first error is in the auxiliary verb that should be singular verb. Since the sentence says "no exact figure", the verb should be singular verb. It is the same as "no one" and "nothing" which counts as singular. This error possibly occurs because of the error in the word "figures" which should be singular too. It should be singular because the sentence says "no exact figure" which means "there is no figure".

Therefore, there errors are multiple error because it is related to each other. This is also a source language emphasis because the lexical choice in the target text is following the lexical choice in source text. It means it does not change any lexical choice or the form of the sentence to convey the context of the sentence. The target text only follows the source text.

## Missing Auxiliary Word

This is the third most common error found in the translation produced by YouTube auto-translate. This error happens when the sentence is missing an auxiliary word.

(3) ST: Mungkin kita benar-benar sendirian di galaksi ini.
    TT: Maybe we really alone in this galaxy.
(TE33)

In this sentence, the target text is stating a subject "we" and adjective "alone" without the auxiliary verb. This sentence has to be completed with auxiliary verb "are" since the subject is "we". Therefore, this error is classified as missing auxiliary word. This error occurs because the target text is only translating the word in the source text. It does not focus on the target text grammatical structure, but it focuses more to the form of source text. Due to that reason, the target text does not have the auxiliary verb which is actually important in English.

## Short Range Word Level Word Order

This is the fourth most common error found in the translation produced by YouTube auto-translate. This error happens when the sentence has wrong word order but only in word level and short range. Short range means that the word just should be moved not far away from the original place or within the same chunk.

(4) ST: Fakta uniknya, Belanda sendiri sekarang menggunakan jalur kanan, bukan jalur kiri.
    TT: The fact unique, Netherlands itself is now using the right lane, not the left lane.
(TI21)

The error in this sentence is in the word "fact" and "unique". The order of those word should be switched. It should be the adjective first, then the noun. Therefore, this error is classified as short range word level word order error. This is also an error that is caused by following the same word order as the source text. This is source language emphasis. It focuses on the source text. The error in this sentence is single error. It does not cause other error in the sentence.

## Extra Word

This is the fifth most common error found in YouTube auto-translation. This error happens when there is an extra word which is actually not needed in the sentence.

(5) ST: Dalam 500 tahun tempat-tempat yang kita kenal sekarang akan kembali ke kondisi semula sebelum dibangun oleh manusia
    TT: In 500 years places we know it today will return to its former condition before it was built by humans
(TD25)

In this sentence, there is a word that is actually not needed. The word "it" is not needed in this sentence. The word "it" here refers to the subject which is "places". Meanwhile, the subject is already stated in the sentence. It means the pronoun "it" is not needed. This error is classified as extra word. This translation is source language emphasis because the lexical choice is following the source text. It does not change the form to deliver the context, but it just follows the source text's form. The error in this sentence is a single error. It does not make other error and is not affected by other error.

In the translation of the data taken, the translation is only source language emphasis. It possibly happens because it is a translation done by machine that only identifies the lexical meaning of the word. YouTube auto-translate is translating the text in source language emphasis. Target language emphasis occurs when a translation is focusing on delivering the context without following the form or the lexical choices of the source text. In target text emphasis, the target text often uses different

lexical choice in order to convey the context of the sentence. In the translation of the subtitle in the videos, there is no translation with target language emphasis.

There are some error types that occur in YouTube auto-translate. However, the percentage is not high. Here are the examples of those error types.

(6) TT: Tapi kok internetnya bisa cepet, ngga lemot?
    ST: But how can internet cepet, guns slow?
(TH27)

The word "cepet" in the source text remains the same in the target text. It should be translated into "fast". However, the word "cepet" in this sentence is not translated. Therefore, it is classified as untranslated error type.

(7) ST: Atau mungkin kita belum bertemu kehidupan lain karena kita adalah kehidupan yang pertama.
    TT: Or maybe we have not met life because we are the first life.
(TE32)

This sentence is missing one word. In the source text, there is the word "lain" which describes the noun "kehidupan". But, the target text only has the word "life". The target text is missing the word "other". Therefore, this error is classified as missing content word.

(8) ST: Oleh karena itu seperti layaknya pohon bangunan super tinggi dapat berayun ketika diterjang angin kencang
    TT: Therefore like a tree super tall buildings can swing when buffeted by strong winds
(TN10)

This sentence has an error that classified as bad disambiguation error. The bad disambiguation error in this sentence is in the word "tall". The word "tinggi" can be translated into "tall" and "high". However, for this context, the word "high" is more suitable to use because it is talking about building. It shows that the translation for this sentence did

not produce a correct translation for a source word that has multiple meaning.

(9) ST: Dan kemudian, terbentuklah Black Hole
    TT: And then, formed Black Hole
(TB24)

In this sentence, there is a problem related to word order. The word formed should be in the end of the sentence. Besides, it also needs the auxiliary verb. The word order error occurs in a long range because the word "formed" has to be moved to the end of the sentence. Therefore, this error is classified as long range word level word order error.

(10) ST: Kemudian, secara tidak sengaja peneliti di Sheffield and Warwick University di Inggris berhasil memecahkan pertanyaan ini.
     TT: Then, inadvertently researchers at Sheffield and Warwick university in the UK this question successfully solved.
(TK12)

In this sentence, there is a problem with the word order but in the phrase level. The order of phrase "this question" has to be switched with the phrase "successfully solved". Therefore, this error is classified as short range phrase level word order.

The least frequently occurred error type is long range phrase level word order. There is only 0,44% of all errors classified as this error type. Here is the example of this error type.

(11) ST: Ada objek misterius dibalik langit tersebut
     TT: There the sky behind mysterious object
(TB3)

This sentence has a problem with word order in phrase level. The phrase "mysterious object" has to be moved before the word "behind". Meanwhile, the phrase "the sky" has to be moved after the word "behind". This is classified as a long range phrase level word order error. Besides, this sentence also needs auxiliary verb.
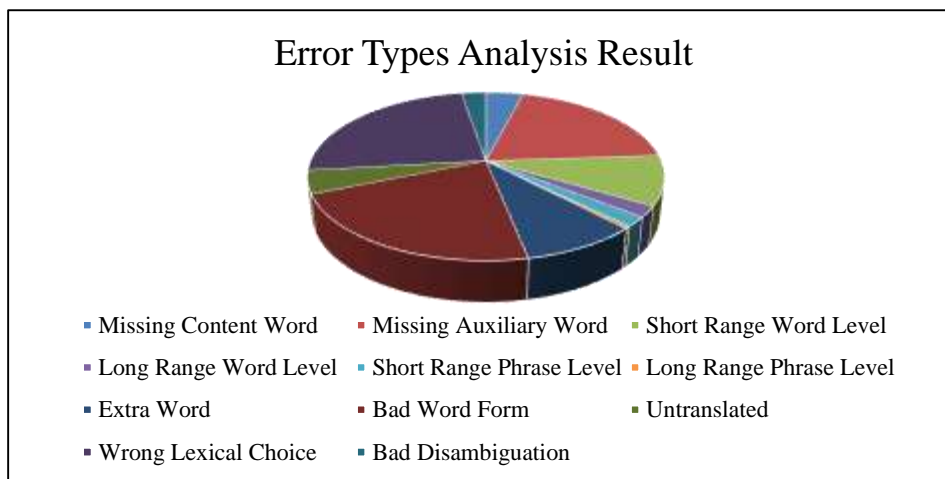
Below is the chart of the error types analysis result.



**Chart 1. Error Types Analysis result**

From the analysis, it is found that the most frequent error is wrong lexical choice and bad word form with the percentage more than 20%. Then, missing auxiliary verb error also frequently occurs with the percentage more than 15%. Next, short range word level and extra word are in the middle with the percentage less than 10% but more than 5%. The error types that rarely occur in YouTube auto-translate translation of "Kok Bisa?" channel's videos are untranslated, missing content word, bad disambiguation, long range word level word order, short range word level word order and the least frequently occurred error type is long range phrase level word order.

**The Implication of the Error Types Analysis**

The result of the error analysis is in line with the result from the previous study. The previous study by Ghasemi and Hashemian (2016) was analyzing the error types found on the translation from Google Translate. It used an error analysis method by Keshavarz (1999). There were six error types used in the previous research to classify the error types. The result showed that the most common error is lexicosemantic error type. In this study, it was also found that the most common error is in lexical choice. It shows that machine translator has a weakness in the lexical level. The previous study also found that machine translation also has problems in the tenses. It confirmed the findings of this research that shows that bad word form is also an error that frequently occur in the YouTube auto-translation. Bad word form is an error of morphological form which can be caused by the wrong tense used. The previous study also confirmed that machine translation has problem in word order.

From the previous study by Koponen (2010), it was found that the most typical errors are mistranslating an individual concept and omitted relation. This study also shows the similar error types. The most frequent error type in this research is wrong lexical choice. It is included in the category of incorrect words. Incorrect word is an error when a lexicon is translated incorrectly. It is the same as the mistranslating an individual concept in the study from Koponen (2010). It proves that the most common error type in machine translation is when a single lexicon is not translated correctly. Meanwhile, omitted relation is when the target does not convey the source text because of morpho-syntactic errors. It is in line with the result of this study which shows that bad word form error is also commonly found in the translation result from YouTube auto-translate. Bad word form is an error type that is caused by a bad

morphological form. It shows that morphological error is also commonly found in machine translation.

The most common error found in this study is similar to the most common error in the study conducted to Google translate. It shows that the problem occurred in machine translation is similar although the machine translator is different. Besides the commonly found error types, this study also found that the errors found are mostly single error. The errors are not related each other. It means the error in the machine translation mostly does not happen because of another error. It occurs individually. This study also found that the translation is source language emphasis. It can also be the reason behind the errors. There is some word for word and literal translation found in the translation. These methods are included as source language emphasis. This can cause an error in the translation because the word for word and literal methods are usually used to translate a difficult text. Then, the result of the word for word and literal translation should be reorganized and rewritten in accordance with the context. However, the word for word and literal translation in the translation produced by YouTube auto-translate is not revised again to meet the context and grammatical structure. This causes errors in the translation result. The result implies that machine translators have similar common errors. The result also proves that the weaknesses of machine translators are in lexical choice and grammar. It also implies that machine translation is source language emphasis because the common error found in YouTube auto-translate and Google Translate is similar. It means the factor causing the error also can be similar.

## Conclusion

There are some error types that are commonly found in YouTube auto-translation. The first is wrong lexical choice. The second one is bad word form. The next is missing auxiliary word. It is followed by short range word level word order and extra word. The other error types are also found in the translation from YouTube auto-translate, but they rarely occur. The translation from YouTube auto-translate is source language emphasis. The errors found in the translation from YouTube auto-translate are mostly single errors. Machine translators have similar common errors. The weakness of machine translation is in lexical level and grammar. The factor behind the error can also be caused by the source language emphasis. For further research, it is suggested to analyze the source of the errors found in the translation produced by YouTube auto-translate.

## References

Bojar, Ondřej. "Analyzing Error Types in English-Czech Machine Translation". *The Prague Bulletin of Mathematical Linguistics.* April 2011: 63-76.

Ghasemi, Hadis & Hashemian, Mahmood. "A Comparative Study of Google Translate Translations: An Error Analysis of English-to-Persian and Persian-to-English Translations". *English Language Teaching.* February 2016: 13-17.

Gray, Alex. "These are the world's most popular websites". *Weforum.org*. World Economic Forum, 10 April 2017. Web. 5 January 2018.

Koponen, M. *Assesing Machine Translation Quality with Error Analysis.* Thesis. Helsinki: University of Helsinki, 2010.

Newmark, Peter. *A Textbook of Translation*. New York: Prentice-Hall International, 1988.