

Lexical Errors Made by Instagram Machine Translation in Translating the Account of “CNN Indonesia” News Article

Raihan Nadilla^{1}, and Syahron Lubis,²*

^{1,2} Faculty of Cultural Sciences, Universitas Sumatera Utara, Medan, Indonesia

Abstract. This study aims to depict the types of errors made by Instagram Machine Translation and to find out the most dominant types of lexical errors made by Instagram Machine Translation on ‘CNN Indonesia’ Instagram account. The design of this research was qualitative design. The data were in words, phrases and sentences that contained lexical errors made by Instagram Machine Translation on the “CNNIndonesia” Instagram account. The data were taken by running an Instagram in one account of various captions related to the lexical errors of the study object. The data were collected through stages: finding out and determining, classifying and separating the words, phrases and sentences that contained lexical errors made Instagram Machine Translation on the “CNN Indonesia” Instagram account. The techniques of analysis data researcher translated the captions using Instagram machine translation and then the translation result is compared to the source language. The next step is to examine the lexical errors produced by Instagram machine translation. The research result shows that the types of lexical errors made by Instagram Machine Translation on the “CNN Indonesia” Instagram account based on the error categories theory by Vilar et al founds are: 4 missing errors, 10 incorrect words and 8 unknown words. All errors indicated that Instagram machine translation could not represent the target language in the “CNN Indonesia” Instagram account. The users of Instagram need to filter every translation that is translated by Instagram machine translation before receiving it as information.

Keywords: Instagram, Instagram Machine Translation, Lexical Error, Machine Translation

Received [20 Dec 2022] | Revised [20 Jan 2023] | Accepted [23 Jan 2023]

1 Introduction

The concept of translation introduces to communicate messages from one language to another. Translation is an activity that aims to convey the meaning or meanings of a given linguistic discourse from one language to another. Translation may be described in phrases of sameness of

* Corresponding author at: Faculty of Cultural Sciences, Universitas Sumatera Utara, Medan, Indonesia

E-mail address: raihan.nadilla@gmail.com

meaning across languages. Translation is a challenging task in transferring meaning from a source language (SL) to a target language (TL). It's concluded for this sense because an irresponsible translation system will result in a misunderstanding of the message located inside the supply language within the target language. The equivalence of a translation must be expressed in an appropriate way in SL to TL so that readers can enjoy the translation and forget for a moment that they are reading is only a translation [1].

According to [2] when translating, it is important to understand the meaning of the source text in order to create an accurate translation in the target text so that meaning is translated in terms of grammar, style and sound. It is very important to understand the meaning of the source language when we translating, by understanding the meaning of a source language the translating will reach the target language according to the structure, writing style, and sound of the writing. And we need to be able to understand or be proficient in the two languages we are going to translate so that the translation does not look awkward [2].

Nowadays in the world of translation it is different in translating the text. People are getting used to using machine translation. But still, machine translation still needs to be check by human rectification. This is to avoid the translation error found in the translation result. Machine translation means automatic translation. Machine translation is designed to translate text from one language (source language) to another (target language) without human help. Machine Translation offers a machine that interprets textual content from the source language to the target language. The interpretation expresses the same meaning as inside the source language [2].

As a foreign language, learning English is not as easy as Indonesia language learning because of the different structure and minimal knowledge of vocabulary makes it difficult to understand what words are being conveyed. It is very difficult to understand every language, but language is the only medium of communication in human society. Nowadays technology is very sophisticated, in the past, we often get information from radio, then we start to switch to television which is more colorful and has a visual that we can see through our eyes, and it's more interesting, and now the world begins to discover even more sophisticated technology that is called mobile phones, and from time to time mobile phones are made more varied and have even more functions, there are many applications that we can use on mobile phone, one of which is the Instagram application. Instagram was originally used only to be an album that can be seen by other people, people can see the photos that we upload through our Instagram, and this makes more people start to be interested and use Instagram social media. But

increasingly, Instagram is starting to be used by people to provide information that is around. There is a lot of information that we can get from Instagram.

To give users with a better experience, Instagram put automatic translation. This feature is useful for a small group of users who automatically translate Instagram-caption in a language they may not be able to read. Currently, users whose primary language is set to English on Instagram can tap or click the button to manually translate caption in their feed in another language. This is to make user easier to understand the conversations and news that we follow on Instagram. However, quite often some users have voiced their annoyance with the machine translation feature because automatic translation is enabled by default, and some machine translations are not always accurate. If there are many translation errors within the translation result, it can cause misunderstanding of turning in a message of the source language. This is become new problem that give author to analyzing error found in Instagram Translation Machine (ITM).

In general, translation is changing one language to another language or to explain another word without change the meaning a good translation is not be seen as translation, but mildly should be read as original compositions to deliver all original meaning. It is critical to provide a translation with being portrayed simply put. This means that we must translate lightly and convey the language's original meaning [3].

Machine translation still being debatable for there is no errors or not. And what we know word “error” means something is wrong. According to Nord, translation errors can be categorized into four classes, particularly pragmatic translation errors, cultural translation errors, linguistic translation errors and text-specific translation errors. state that translation errors are frequently linked to one another. It means that one error or problem has an effect on other errors. The solution follows suit. Because the errors are linked, the solutions for one error will influence the solutions for others. It's as if there are networks or hierarchies in which the solution to one problem influences how other problems are addressed [3].

To evaluate the error generated by Instagram Machine Translation, the error must be classified and categorized. Vilar state there are five categorizes of the errors. These are missing word, word order, incorrect words, unknown words and punctuation errors. Missing word mean when some word produced in a sentence is missing. Word order occurred when the word in the target text is in the wrong position. Then, the Incorrect Words errors are the most general type of error. This category occurs when

the system is unable to find the correct translation of a given word. Unknown word error is refers to untranslated word. The last is punctuation error, which refers to mistakes in the use of punctuation marks in writing [4].

The results of machine translation do contain many types of errors. One of the types of error is lexical error. If a text contains a lot of lexical errors, then the translated text will affect the meaning of the translation itself. According to [5] lexical errors are the most common type of mistake in written English. They have a significant impact on the quality of academic writing, and native speakers find them the most vexing. Ellis also state, lexical errors have categories, particularly in writing; if the percentage of errors is low, the quality of academic writing is good; if the percentage of students writing is high, and the quality of academic writing of students is low [4].

The “Incorrect Words” errors are the most general type of error. When the system is unable to find the correct translation of a given the word. We distinguish five subcategories here. The incorrect word in the first example distorts the meaning of the sentence. In this case, we could further distinguish two additional subclasses: when the system selects an incorrect translation and when the system is unable to disambiguate the correct meaning of a source word in a given context, though the distinction is certainly hazy. The following subcategory of “Incorrect Words” errors occurred when the system was unable to produce the correct form of a word, even though the translation of the base form was correct. It is especially important for inflected languages, where the high variability of open word classes makes machine translation difficult. Extra words in the generated sentence cause another type of error. This type of error was introduced when researching the translation of spoken language input, as artifacts of spoken language may produce additional words in the generated sentence.

The last two classes are of lesser importance. The first is style errors refer to a poor choice of words when translating a sentence, but the meaning is preserved even if it is not entirely correct. A common example is the repeated use of a word in a close context. A translator would choose a synonym and avoid word repetition in this case. The second is about idiomatic expressions that the system does not recognize and attempts to translate as normal text. Normally, these expressions cannot be translated in this manner, resulting in additional translation errors.

Unknown words can also cause errors. In this case, we can differentiate between truly unknown words (or stems) and unseen forms of known stems. A variation of this category is particularly important for the Chinese-English language pair. The majority of European languages, or even languages with the same alphabet, can be "translated" simply by copying the input word to the generated sentence, with no further processing.

Lastly, there may be punctuation errors, but for the current machine translation output quality, these are minor annoyances for languages with no fixed punctuation rules and are not taken into account further in this work. Of course, the error types defined in this manner are not mutually exclusive. In fact, it is not uncommon for one type of error to result in another. A bad word translation, for example, can result in a bad ordering of the words in the generated sentence. Then, according error classification based on [4], actually, there are only three categories are related to the type of lexical error. As in [6], many of the errors made were as a result of the student's lack of understanding in differentiating the word class. There are missing words, incorrect words and unknown words. For instance, to reveal the adjective form, the students did not add the suffix to the noun.

Instagram is one application that is booming and is used by many people in the world. Quoting from statista.com in Ideoworks.id article, India is the country with the most Instagram users in the world, reaching 180 million users. Followed by the United States with 170 million users, Brazil with 110 million users, Indonesia with 93 million users, and Russia with 61 million users. Instagram was created by Kevin Systrom and Mike Krieger. Instagram also has many features such as comment, like, and Instagram Machine Translation. Instagram Machine Translation is one of the features found in the Instagram application. It has a function to translate captions on photos posted on Instagram. This feature will make it easier for many people to know the translation of captions in other languages.

Cable News Network (CNN) is an American cable news station founded in 1980 by the American media group Ted Turner. When CNN was first launched, it was the first television station to broadcast 24-hour news coverage and the first television news station in the United States. CNN has many branches, but more often broadcasts from the CNN Center headquarters in Atlanta, the Time Warner Center in New York City, and studios in Washington, DC and Los Angeles. CNN is owned by its parent company, WarnerMedia, and is a division of WarnerMedia News and Sports. The first CNN channel in Asia started in Indonesia on August 17, 2015 and started on December 15, 2015. CNN Indonesia has become Indonesian CNN News.

“An Analysis of Grammatical Errors Of Using Google Translate From Indonesia To English In Writing Undergraduate Thesis Abstract Among The Students’ English Department Of Iain Metro In The Academic Year 2016/2017” by (Kurniasih, 2017) in this thesis she focused on grammatical error in translation. This study applied Miles and Huberman. The research result show that the student using Google translate in translating the abstracts and show results of a finding translate by Google Translate is not accurate in English [7].

The similarity with my research is the topic about machine translation. The differences are the focuses in Kurniasih's research is on grammatical error while this researcher focuses on lexical error and in Kurniasih's used [8] while this research refer use [4]. Previous study contributing to enhances the writer knowledge about Miles and Huberman theory and shows that Google Translate is still not accurate.

“An Error Types Analysis on YouTube Indonesian-English Auto-Translation in Kok Bisa? Channel” by [9] in this article author investigates the error types that commonly occur in the translation produced by YouTube auto-translate. This research uses error classifications from [4]. 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.

The similarities with my research are the topic about machine translation and uses error classifications from [4]. The difference are the focuses in Laksana's research is investigates the error types that commonly occur in the translation produced by YouTube auto-translate while researcher focuses on lexical error produced by Instagram Machine Translation. Previous study contributing to enhances the writer knowledge about Vilar theory and shows that YouTube auto translates is still made error.

“Error Analysis of English Translation of Islamic Texts by Iranian Translators” by [10] in this journal author analysis of the type and frequency of the errors occurring in the English translation of Islamic texts by Iranian translators and analyze the possible cause of the errors. This study used Morgan's sample selection table. The errors were categorized based on the classification of error types developed by [11]. The results of the study revealed that the register category was the most frequent error area.

The similarity with this research is the topic about machine translation. The differences are the focuses in Jahanshahi's analysis of the type and frequency of the errors occurring in the English translation of Islamic texts by Iranian translators and analyze the possible cause of the errors while this researcher focus on lexical error made by Instagram Machine Translation and in Jahanshahi's research Morgan's sample selection table, while researcher refer use [4]. Previous study contributing to enhances the writer knowledge about Morgan's sample selection table and revealed that errors occurring in the English translation of Islamic texts by Iranian translators and the possible cause of the errors revealed that the register category was the most frequent error area.

“Lexical Errors Produced By Instagram Machine Translation” by [12] in this thesis she focuses on lexical error in translation. This study applied theory of [4]. The research result shows that Instagram machine translation produced so many errors and shows the weakness of machine translation to represent the genuine language.

The similarities with my research are the topic about machine translation, and then the thesis focuses on lexical error in translation and use theory of Vilar [4]. The difference between Susanti’s and mine is only Instagram account. But the application we use is the same. Previous study contributing to enhances the writer knowledge more about Vilar theory and shows that Instagram translation is not represent the genuine language.

2 Method

The research design of this study used qualitative method to explain the data. Qualitative data was collected to look for the phenomenon of the Instagram Machine translation output. The data of this study are in the form of text which consists with phrases, sentences and paragraphs. The data source of this study is the caption on “CNNIndonesia” account on Instagram that translated by Instagram Machine Translation. The data for this study were derived from the captions of one Instagram account, CNN Indonesia, which was translated from Indonesia into English using Instagram Machine Translation. The information is presented in the form of phrases, sentences, and paragraphs. Researcher take some steps to obtain the data. After collecting all of the data, researcher read the texts or captions translated using Instagram machine translation on the "CNNIndonesia" Instagram account. The translation result then is compared to the source language by the researcher. The next step is to examine the lexical errors produced by Instagram machine translation

3 Result and Discussion

3.1 Result

IMT translated the sentences “*Komisiaris utama PT Pembangunan Perumahan (PP) Andi Gani Nena Wea mengatakan bahwa proses pengerjaan mengalami cukup kesulitan, karena jalur melintasi rawa-rawa*” into “*Chief Commissioner of PT Housing Development (PP) Andi Gani Nena Wea said that the work process experienced enough difficulties, because the path crossed the swamp*”. Here the word “Pembangunan Perumahan” is the name of a company. This company is one of BUMN in the planning and construction of buildings. This company’s name can’t be translated because this is an

official name that used as a marketing name to be known as the company name. From this data, the target language made by Instagram Machine Translation is categorized as incorrect because an institution name cannot be translated into another language. This translation shows that IMT translated language by word. For the next, Instagram machine translation also translated the sentence "*Namun, Ganjar mengapresiasi pengerjaan proyek tersebut*" into "*Ganjar, however, appreciated the work of project*". Based on [4], it categories as missing word. It happens because when the missing word is only required to form a grammatically correct sentence but, the meaning is preserved. In this case, Instagram machine translation incorrectly missing the use of article "*the*" in the sentence. The article "*the*" is used to refer to something that has already mentioned before. The omission article "*the*" found in sentence above. It can be seen in the sentence, before the word "*project*" should be added the article "*the*", to form "*the project*".

Table 1 Classification of Data 1

Source Language (IDN)	Target Language (ENG)	Errors made by IMT based on Vilar et al theory
Komisaris utama PT <u>Pembangunan Perumahan (PP)</u> Andi Gani Nena Wea mengatakan bahwa proses pengerjaan mengalami cukup kesulitan,karena jalur melintasi rawa-rawa.	Chief Comissioner of PT <u>Housing Development</u> (PP) Andi Gani Nena Wea said that the work process experienced enough difficulties, because the path crossed the swamp.	Incorrect word
Namun, Ganjar mengapresiasi pengerjaan proyek tersebut.	Ganjar, however, appreciated the work of <u>(the)</u> project	Missing word

In this case Instagram machine translation omitted the subject in the target language. This omission is shown in the sentence "*Transportation Minister Budi Karya Sumedi reveals will create a strict health protocol mechanism arranged by his side and the covid-19*" where IMT not translated the word "*gugus tugas*" in the source language as the subject of the sentence. Subject is the most important element that must appear in the sentence. Therefore, it is useful to make the context easy for the reader to understand. Therefore, the correct sentence should be translated into "*Transportation Minister Budi Karya Sumedi reveals will create a strict health protocol mechanism arranged by his side and the task force covid-19*".

The next category of error is unknown word. In this data, researcher also found the unknown word. Unknown words are easily translated by copying the words into the generated sentence without further processing. The unknown word shown in the sentence “*“We have petted some important issues. There will surely be a surge vaccination programs predicted to make people want to travel,” said Menhub Budi*” which Instagram machine translation cannot translate the word “*menhub*” into target language. This word can be translated into “*Minister of Transportation*”. IMT can't translate it may be because the word “*menhub*” is an abbreviation of the “*Menteri Perhubungan*” which in English is the “*Minister of transportation*”.

The next is still same as the previous error. The next error is unknown word. Unknown also indicated as a source of error. In this sentence contains the error in unknown category “*This policy differs from lebaran 2020*”. The error happens because Instagram machine translation not translated the word “*lebaran*” into target language. Meanwhile, the word “*lebaran*” was translated into “*Eid Mubarak*” in the previous sentence that doesn't show in the data table. The sentence of target language is “*The government will not ban people from going home to Eid Mubarak this year, although the coronavirus pandemic still struck*” and the sentence of source language is “*Pemerintah tidak akan melarang masyarakat untuk mudik Lebaran tahun ini, meski pandemi corona masih melanda.*” The sentence shown word “*Lebaran*” translated into “*Eid Mubarak*”. But at the last sentence in the caption shows that IMT did not translate it.

Table 2 Classification of Data 2

Source Language (IDN)	Target Language (ENG)	Errors made by IMT based on Vilar et al theory
Menteri perhubungan Budi Karya Sumadi mengungkapkan akan membuat mekanisme Protokol Kesehatan ketat yang disusun pihaknya dan Gugus Tugas Covid-19	Transportation Minister Budi Karya Sumadi reveals will create a strict health protocol mechanism arranged by his side and the covid-19	Missing word
“Kami sudah petakan beberapa isu penting. Pasti akan terjadi lonjakan, program vaksinasi diprediksi akan membuat masyarakat ingin berpergian,” Ujaran <u>Menhub</u> Budi	“We have petted some important issues. There will surely be a surge vaccination programs predicted to make people want to travel,” Said <u>Menhub</u> Budi	Unknown Word
Kebijakan ini berbeda dari <u>lebaran</u> 2020.	This policy differs from <u>lebaran</u> 2020	Unknown word

The word “*middleman*” is the word made by IMT that does not reach the target language. For this case, an incorrect word was included in the subcategory word error.

As explained by Vilar et al., word error is when the resulting word cannot form a word, even though the basic translation is correct. The word 'warga tengah' referred to "warga sedang," in this sentence, and the word "sedang" is used as an adverb. So, the correct sentence should be "Next there is a photograph of Beijing origin, China depicts the resident spending time in public parks implementing health protocols." Again, this shows that Instagram machine translation provides a translation by word, so the sentence in question is not conveyed correctly.

Table 3 Classification of Data 3

Source Language (IDN)	Target Language (ENG)	Errors made by IMT based on Vilar et al theory
Selanjutnya ada foto asal Beijing, China menggambarkan <u>warga tengah</u> menghabiskan waktu di taman publik dengan menerapkan protokol kesehatan.	Next there is a photograph of Beijing origin, China depicts <u>the middle men</u> spending time in public parks implementing health protocols.	Incorrect word

The word "Mari" translated into "Come". In this case, the incorrect word is in wrong word choice subcategory. Because, the sentence that show in this data is persuasive. According to Oxford, if the context of the sentence is used for making suggestions or as a polite way of telling people what you want them to do, so it more appropriate if using the word "let's". Therefore, the word "come" is more appropriate if changed to be "let's" in the sentence. And also in this sentence IMT translated the platform that shown in the sentence. It is "berbuatbaik.id" into "doinggood.id". In literal translation in Instagram machine translation, it is not wrong to translate the word, but the platform name is not needed to be translated. Based on explanation above the sentence should be changed into "Let's, help our brothers and sisters affected by the eruption of Mount Semeru through the platform of berbuatbaik.id"

Table 4 Classification of Data 4

Source Language (IDN)	Target Language (ENG)	Errors made by IMT based on Vilar et al theory
<u>Mari</u> , bantu saudara-saudara kita yang terdampak erupsi Gunung Semeru melalui platform <u>berbuatbaik.id</u>	<u>Come</u> , help our brothers and sisters affected by the eruption of Mount Semeru through the platform of <u>doinggood.id</u>	Incorrect word

Here, Instagram machine translation used the word “for him”. Actually, the word “Baginya” in the source language is referring to the word “perempuan” (woman) because the previous sentence in the caption on Instagram mention the name of Minister of Employment Ida Fauziah. So, the appropriate objective pronoun in this sentence is “for her”. Furthermore, the correct translation used for this data is “For her, the ideal median wage index is in the range of 0.4 to 0.6 percent”.

Table 5 Classification of Data 5

Source Language (IDN)	Target Language (ENG)	Errors made by IMT based on Vilar et al theory
<u>Baginya</u> , indeks upah rata-rata yang ideal berada di kisaran 0,4 hingga 0,6 percent	<u>For him</u> , the ideal median wage index is in the range of 0.4 to 0.6 percent	Incorrect word

Here, the word “RI” is the abbreviation of “Republik Indonesia”, and if it is translated into English, it will become “Republic of Indonesia”. But, Instagram machine translation translated it into “Republic of Kazakhstan”. However, Instagram machine translation also translated the word “Republic of Indonesia” into incorrect word, it translated become “Republic of Kazakhstan”. This error will interfere the meaning of the sentence so that the sentence's meaning will change.

The word “komnas” is abbreviation from “komisi nasional”. If it translates in target language it will be “The National Commission”. However, in source language, it is “komnas perempuan”, and for the target language itself is “The National Commission on Violence against Women”. Here, IMT showed that IMT translated word by word. The word “permendikbud” becomes an unknown word. This error shown is because IMT can’t translate the word properly. The source language of “permendikbud” if translated to target language is become “Regulation of the Minister of Education and Culture” because, the word “permendikbud” is abbreviation of “Peraturan Menteri Pendidikan dan Kebudayaan” in source language.

Table 6 Classification of Data 6

Source Language (IDN)	Target Language (ENG)	Errors made by IMT based on Vilar et al theory
Sejumlah aktivis berbondong- bondong melakukan aksi di depan gedung <u>DPR RI</u> , KAMIS	A number of activist in droves staged an action in front of the House of <u>Representative of the Republic of Kazakhstan</u> , Thursday	Incorrect word

(25/11)	(25/11)	
Berdasarkan data dari <u>komnas</u> perempuan per 2015-2020, 27 kasus kekerasan seksual terjadi di perguruan tinggi.	According to data from the women's <u>comnas</u> for 2015-2020, 27 cases of sexual violence occurred in collages.	Unknown word
Sementara itu, <u>permendikbud</u> PPKS menuai pro kontra setelah disahkan.	Meanwhile, <u>permendikbud</u> PPKS reaps pros and cons after being confirmed	Unknown word

Here the word “*semrawut*” being translated became “*rusty*”. This is the wrong translated that IMT did. Actually, the word “*rusty*” in Bahasa is “*berkarat*”. The meaning “*berkarat*” and “*semrawut*” it is really different. The word “*semrawut*” in KKBI is kacau balau; acak-acakan; tidak teratur. In English it’s “*Messy*” or “*Chaotic*”. And also in this sentence IMT translated the word “*menjuntai*” into “*cut*”, the correct word should be changed into “*dangled*”. So, the correct sentence must be “*The rusty cables that cut or stacked on the poles in Jakarta become the everyday scene of citizens.*”

The word “*juntaian-juntaian*” means “*bergantung-gantung*” if translated to target language it is “*dangled*”. But, IMT translated it become “*dandruffs*”. The word “*dandruffs*” in Bahasa is “*Ketombe*”. So, IMT have wrong translated the word for this sentence. Furthermore, the correct translation it should be “*In fact, at the end of the year, there are still many shapeless dangled in the streets of the Capital.*”

Table 7 Classification of Data 7

Source Language (IDN)	Target Language (ENG)	Errors made by IMT based on Vilar et al theory
Kabel <u>semrawut</u> yang <u>menjuntai</u> atau bertumbuk di tiang-tiang di Jakarta jadi pemandangan sehari-hari warga.	The <u>rusty</u> cables that <u>cut</u> or stacked on the poles in Jakarta become the everyday scene of citizens.	Incorrect word
yatanya dipenhujung tahun ini masih banyak terlihat <u>juntaian-juntaian</u> <u>tak</u> <u>berbentuk</u> di ruas jalan Ibu kota	In fact, at the end of the year, there are still many <u>shapless</u> <u>dandruffs</u> in the streets of the Capital.	Incorrect word

The word “*Gunung*” becomes an unknown word in this sentence. This error shown is because IMT can’t translate the word properly. The source language of “*Gunung*” if translated to target language is become “*Mount*”. Furthermore, the correct translation for this sentence should be “*Location of disaster eruption of Mount*

Semeru, Lumajang, East Java 'used' by several parties, including politicians, tourists, to film industry”.

Table 8 Classification of Data 8

Source Language (IDN)	Target Language (ENG)	Errors made by IMT based on Vilar et al theory
Lokasi bencana erupsi <u>Gunung</u> Semeru, Lumajang, Jawa Timur 'dimanfaatkan' oleh beberapa pihak, mulai dari politikus, wisatawan, sampai industri perfilman.	Location of disaster eruption of <u>Gunung</u> Semeru, Lumajang, East Java 'used' by several parties, including politicians, tourists, to film industry.	Unknown word

The first error is missing error, missing word happens when the missing word is only required to form a grammatically correct sentence but, the meaning is preserved. In this case, Instagram machine translation incorrectly missing the use of article “*the*” in the sentence. The article “*the*” is used to refer to something that has already mentioned before. The omission article “*the*” found in sentence above. It can be seen in the sentence, before the word “*AFF Cup 2021*” should be added the article “*the*”, to form “*the AFF Cup 2021*”

Table 9 Classification of Data 9

Source Language (IDN)	Target Language (ENG)	Errors made by IMT based on Vilar et al theory
Thailand menang telak 4-0 atas Timnas Indonesia di leg pertama final Piala AFF 2021, Rabu (29/12)	Thailand won 4-0 against Indonesia National Team in the first leg of <u>(the)</u> AFF Cup 2021 final, Wednesday (29/12)	Missing word
<u>Guna</u> mengejar <u>ketinggalan</u> Pelatih Shin Tae Young melakukan pergantian tiga pemain sekaligus sejak awal babak kedua.	<u>Guna</u> pursues the <u>height</u> of coach Shin Tae Young to make three player changes at a time since the start of the second half.	Unknown word Incorrect word
Timnas Indonesia akan menghadapi Thailand di leg kedua pada 1 Januari mendatang.	Indonesia national team will face Thailand in the second leg on January 1st.	Incorrect word

The first error in the data 10 is incorrect word. IMT translated the word “*TNI Angkatan Udara*” into “*TNI Air Force*”. Actually, the word “*TNI*” is abbreviation of “*Tentara Nasional Indonesia*” if translated in English it becomes “*Indonesian National Armed Forces*” and “*TNI Angkatan Udara*” if translated to English it becomes “*Indonesian Air Force*”. And for the next sentence IMT also translated “*TNI AU*”

into “AU army”. Based on explanation above it should be “Indonesian Air Force”. So, the correct translations of those sentences are “*“That Halim airport will be closed from 1 January. This is the order of Mr. President because Mr. President feels that the runway is rough,” said the Chief of Staff of Indonesian Air Force Marshal Indonesian National Armed Forces Fadjar Prasetyo*”, “*That means some Indonesian Air Force units based in Halim will be temporarily relocated*”.

The second error is missing error, missing word happens when the missing word is only required to form a grammatically correct sentence but, the meaning is preserved. In this case, it also has the same mistake line in previous data. Instagram machine translation translated it with missing the use of article “the” in the sentence. The omission article “the” found in sentence above. It can be seen in the sentence, before the word “VVIP building” should be added the article “the”, to form “the VVIP building” so the correct sentence should be “*Revitalization will also be done to expand the military apron, as well as renovate the VVIP building and terminal*”.

For the next error is unknown word. The error in this sentence is translated simply by copying the input word to the generated sentence, with no further processing. Instagram machine translation translated the sentence “*Dalam kajian awal Kemenhub, proses revitalisasi diperkirakan akan memakan waktu kurang lebih satu tahun*” into “*In a preliminary study of Kemenhub, the revitalization process is estimated to take approximately one year.*” the word “Kemenhub” becomes an unknown word. The source language of “Kemenhub” is abbreviation of “Kementerian Perhubungan” if translated to target language is become “Ministry of Transportation”. So the correct sentence for this translation is “*In a preliminary study of Ministry of Transportation, the revitalization process is estimated to take approximately one year*”.

Table 10 Classification of Data 10

Source Language (IDN)	Target Language (ENG)	Errors made by IMT based on Vilar et al theory
"Bahwa Bandara Halim akan ditutup mulai tanggal 1 Januari. Ini adalah perintah Bapak Presiden karena Bapak Presiden merasa kan betul runway-nya sudah kasar," kata Kepala Staf TNI Angkatan Udara Marsekal TNI Fadjar Prasetyo	"That Halim airport will be closed from 1 January. This is the order of Mr. Presiden because Mr.Presiden feels that the runway is rough," said the Chief of Staff of <u>TNI Air Force</u> Marshal TNI Fadjar Prasetyo	Unknown word
Revitalisasi juga akan dilakukan untuk memperlus military appron,	Revitalization will also be done to expand military apron, as well as renovate	Missing word

serta renovasi VVIP gedung dan terminal.	(the) VVIP building and terminal	
Itu membuat beberapa satuan <u>TNI AU</u> yang berada di Halim akan direlokasi sementara waktu	That means <u>some AU army</u> units based in Halim will be temporarily relocated.	Incorrect word
Dalam kajian awal <u>Kemenhub.</u> proses revitalisasi diperkirakan akan memakan waktu kurang lebih satu tahun.	In a preliminary study of <u>Kemenhub.</u> the revitalization process is estimated to take approximately one year.	Unknown word

3.2 Discussion

. **Table 11** Frequency of types of Instagram Machine Translation

NO	Types of IMT Errors	Frequency	Percentage
1.	Missing Words	4	18,18%
2.	Incorrect Words	10	45,45%
3.	Unknown Words	8	36,36%
Total		22	99,99%

From the table above, it can be seen that types of IMT errors based on Vilar et al (2006) related to lexical categories are missing words 18,18%, incorrect words 45,45%, and unknown words 36,36%. The incorrect words shows the highest percentage because incorrect word happened when the system or machine translation unable to find the correct translation in the translation result.

4 Conclusion

According to the analysis and findings in the previous chapter, it can be concluded that the ten data from the captions on "CNNIndonesia" Instagram account contain three types of errors in the lexical category. Those errors are missing words, incorrect words, and unknown words. But, three types of error are not always found in every data. Incorrect word and unknown word become the highest frequent error found in ten data on the captions of "CNNIndonesia" Instagram account. In general, the error that is often encountered in IMT is translating the trade name or the name of the Institute literally. Thus, there is an unnecessary translation process. In addition, all errors indicated that Instagram machine translation cannot represent the target language in the "CNNIndonesia" Instagram account. So, the users of Instagram need to filter every translation that translated by Instagram machine translation before receiving it as the information.

REFERENCES

- [1] M. J. Muchtar, *Translation Theory, Practice and Study*. Medan: Bartong jaya, 2015.
- [2] H. Ghazala, *Translation as problem and solutions*, 4 th ed. Sriya: Dar Elkalem Elrabi, 1995.
- [3] M. A. Balan, "An Analysis of Google Translate Accuracy.," *Transl. J.*, vol. 16, 2011.
- [4] et al Vilar, *Error Analysis of Statistical Machine Output*. Universidad Politecnia de Madrid, 2006.
- [5] R. Ellis, *The Study of Second Language Acquisition*. Oxford University Press, 1994.
- [6] B. Amin, "Lexical Error In Writing English Words Made By Students Of The Junior High School," *Exposure*, vol. 3, no. 1, 2014, [Online]. Available: <https://journal.unismuh.ac.id/index.php/exposure/article/view/831>
- [7] I. Kurniasih, "An Analysis Of Grammatical Errors Of Using Google Translate From Indonesia To English in Writing Undergraduate Thesis Abstract Among The Students," IAIN METRO, 2017.
- [8] M. B. Miles, A. M. Huberman, and J. Saldana, *Qualitative Data Analysis: A Methods Sourcebook*. 1984.
- [9] N. L. Putri, "An Error Types Analysis on YouTube Indonesian-English," *J. Lang. Lit.*, vol. 18, pp. 76–81, 2018.
- [10] M. J. Kafipour, "Error Analysis of English Translation of Islamic Texts byIranian Translators," *J. Appl. Linguist. Lang. Res.*, vol. 2, no. 3, pp. 238–252, 2015.
- [11] J. Liao, "The impact of interactive discussions on L2 Chinese composition writing," University of Iowa, 2010.
- [12] E. Susanti, "Lexical Errors Produced By Instagram Machine Translation," Universitas Islam Negeri Maulana Malik Ibrahim Malang, 2018.