

Chapter I: Introduction

This chapter describes several things related to this research which are distributed into several points such as basic consideration, research question, research objective, research delimitation, and research significance.

Basic consideration

In this era, translation activity is not limited only to human translators, but it also expands for machine translation (MT). According to Temizöz (2013), machine translation is an automatic process of the translation system that is run by computer software from a source language to the target language. This tool has become a popular method for the translator, students, etc.

As computational activities become more mainstream, and the internet opens up the wider multilingual and global community, the research and the development of Machine Translation continues to grow, and the rapid rises of social media (networks), which put people speak differently, increases the necessity of the use of machine translation (MT), and the impact is machine translation has become a popular media to translate text. According to Kemp, (2018) total of the population in the world is 7.593 billion while the numbers of internet users are 4. 021 billion, and the number of users of social media in 2018 is 3.196 billion. It is clear that the users of the internet are more than half of the population and they are from different countries.

As the user of the internet is from various of countries, many of website developer that provides online machine translation such as SmartMate, TransPerfect, Google, Systran, etc, have tried to develop their online translation

machine to make their users get better services, unfortunately, those online machines translation, have several issues regarding the accuracy and the inability to understand the text as humans do. According to Hutchins (1994), the main problems in MT systems are divided into four main headings: structural, lexical, contextual, and pragmatic or situational. A lexical error occurs when the machine translates a word that has an independent meaning such as noun, verb, adjective, adverb, and preposition. Structural error in MT occurs when machine translations fail to translate sentences with correct grammar in the target language. The contextual error occurs when machine translation fails to translate the meaning of a word to the target language. From the statement, it can be said that online machine translation has lots of deficiencies. Although online translation machine has lots of deficiency in the translation process, this tool has become a widely popular tool to be used and an interesting topic to be discussed.

Furthermore, this study particularly focuses on the automatic YouTube subtitle translation. The researcher chooses this topic because the user of YouTube is from many countries. According to Merchdope, the number of users of YouTube is about 1,300,000,000. Female users are 38% and male users are 62% from around the world. You can navigate YouTube in a total of 76 different languages (covering 95% of the Internet population). From the statements above, it is a fact that the users of YouTube are from different countries which put the viewers speak in different languages, that is why YouTube create their own online automatic machine translation.

Unfortunately, this automatic YouTube subtitle translation still has a lot of issues in its translation output. This machine is using a speech recognition machine that allows people (viewers) to transcribe the English speakers' sound in real-time, into a subtitle automatically, and also automatically translates the subtitle into specific language that has been provided for example English into Indonesian. According to (Google support) YouTube can use speech recognition technology to automatically create captions (subtitle) for videos in real-time. These automatic subtitles are generated by a machine that is processed the algorithms, so the quality of the captions may vary. This translation process can be run by clicking a specific button in the video for example, while the video in English is playing, the viewer can put the subtitle by clicking the subtitle button (CC icon) and it also can be translated by clicking the auto-generated button using the machine translation to translate the subtitle. This automatic captioning and subtitling can only proceed from English into as the source language to another language because the auto speech recognition technology can only recognize English speakers' video by this time.

This process of automatic speech recognition and automatic translation can proceed because of YouTube have combined their auto speech recognition machine and machine translation, as cited in Google official blog published in (2009) that states Google official have combined their Automatic Speech Recognition Machine with YouTube caption system to make auto-caps or automatic captions that are using the same voice recognition algorithms system with the Google Voice, to automatically generate the captions for the video. This

captioning machine is intended for those who are deaf as stated in Google Official Blog (2009) the captioning machine helped the deaf and hearing impaired.

Since the launch of this machine in 2006 as stated in Google video blog (2006), the use of the caption is not only limited to the deaf but it is expanded for the people (viewers) who want to watch the video with muted sound or for those who can understand the video through written subtitles better than they hear speech in the video. YouTube machine translation is intended for YouTube creators who want to grow their international audience by making their videos more accessible in other languages. According to (YouTube Support) the translation is intended for viewers who speak other languages and it also makes the video available for deaf viewers and non-native speakers of the language being used.

By using this machine, people (viewers) no longer need the content creator or the owner of the video to create captions for their video, it is automatically generated by the machine (ASR) and automatically translated into 51 languages, as stated in (Google Official Blog, 2006) this means that the YouTube machine translation enable people (viewers) to translate the subtitle in any of 51 languages.

Unfortunately, the translation generated from this machine is not perfect; it has some errors in its translation output. Based on the researcher's experience in using YouTube, some errors occur in this machine, such as the differences between the subtitle and what the speakers have said, the errors of the target language, words did not successfully translate, and words did not appear on the caption. Based on all the statements above, the researcher is interested in finding

out what kinds of translation errors especially in English to Indonesian translation that is occurred in YouTube subtitle translation. The objectives of this investigation were to examine: Kinds of translation errors made by automatic YouTube subtitle translation.

One example of these translation problems can be observed in a video with automatic translation into the Indonesian language from the English language posted by user: *Domics* in the title "*Break Ups: Part 4*". In duration time 2:50, the speaker said that "*had the same sense of humor.*" but the machine translated it as "*memiliki arti yang sama humor*" in this text the translation should be "*memiliki selera humor yang sama*". From this text, the machine fails to translate the text into the correct sentence in the target language.

The researcher has observed some previous researches related to the use of online translation machines. The first research is *An Error Analysis of Translated Result of Google Translation-Translate in English Text* this research conduct by Nurmala and Ratna Sari Devi (2018). The investigation noted that Google translates made several grammatical errors to translate three texts from Liputan6.com from Indonesian to English, there is 1 error in omission, 1 error, in addition, 2 errors in misformation, other errors that have not been in the taxonomic category, 4 errors in diction, 8 errors in misordering, 11 errors in tenses, 7 errors in pronoun, and 4 errors in semantics. There are several differences between this research and my research. This previous research is using text only as of the source of the data, while this recent research is using video, and subtitle text as the source of the data. Moreover the previous research is only only

focuses on the error of one machine, while my research is using two different machines, is automatic speech recognition machine and machine translation. The similarity between this research and my research is both of my research and this research are uses the same type of machine translation that is a neural machine translation.

The second research that has related to this research is *Semantic Error Analysis of Instagram Machine Translation from Indonesian to English* conducted by Fadila (2017). This research is the error translation of HTI's Instagram account post, in semantics aspects of Instagram machine-translation output. The result of this research is the error of Instagram machine translation shows that Instagram machine translation cannot translate in good quality the postings from HTI's Instagram account, especially in grammatical and contextual meaning. Moreover, these researches are using different types of data. Fadila uses an Instagram caption that is conducted by a human, while this research is using subtitle that is generated automatically by speech recognition.

In conclusion, all of these problems proceed by YouTube MT may cause the viewers, who are using the output (subtitle) of the machine, cannot understand the meaning of the subtitle in the target language (TL), unfortunately, there has been very limited studies investigating this issue, as this may be a new issue in the machine translation field, especially for the case of automatic YouTube subtitle translation machine. Therefore, this study wants to examine to what extent YouTube automated translation produces a good translation.

Research question

As regards this research related to error analysis in YouTube, so the researcher formulates a research question as follows:

What are the translation errors in Automatic YouTube subtitle translation?

Research objective

From the statement of the research question, the researcher has one objective that is to examine the errors of YouTube translation machine output in translating the subtitle from English to Indonesian.

Research delimitation

The main point of this research is the errors of the subtitles translation results from English to Indonesian, by YouTube's machine translation.

Research significance

The result of the research, theoretically, could give detailed information about kinds of translation errors made by automatic YouTube subtitle translation, and practically, it might be useful informations especially viewers who using this automatic YouTube subtitle translation in their video, to be aware of what kinds of error could occur when they are watching video from YouTube, and the most important is the result of this research hopefully can give useful informations for the further research of machine translation error, especially in YouTube machine translation.