CHAPTER I

INTRODUCTION

1.1 Background of Study

As the world of technology has been growing sophisticatedly, people around the world are able to access information, knowledge, and entertainment easily from a lot of media such as written media, audio-visual media, online media, and vice versa. One of the most accessed media at this moment is online media such as YouTube. YouTube became the second most accessed platform in the world with 2 billion active users' respectively (Oberlo.com).

YouTube provides various kinds of content such as news, beauty, comedy, lifestyle, cooking, and learning. However, the differences in languages of the YouTube content creators and viewers could be an obstacle for people to understand the purpose of the given information. For that matter, aside from the video itself, YouTube announced caption support in 2006 and enabled automated captions since 2009. YouTube provides auto-generated subtitles for over a hundred languages from around the world. Gottlieb (2004, as cited in Farid, G., & Janin, B. 2010) states that the word subtitling is defined as the rendering of the verbal message in filmic media in a different language, in the shape of one or more lines of written text, which are presented on the screen in sync with the original verbal message. The function of the subtitle was for the user to enjoy watching and understanding the content of the video despite the differences of languages.

Nevertheless, since the YouTube Auto-Generated subtitle for most videos using speech recognition technology, generated by machine learning algorithms, it may result in a varied subtitle quality. Consequently, a good translation is needed to make information and knowledge well-delivered to non-native speakers around the world. Newmark (1982) considers "translation as a craft consisting in the attempt to replace a written message and/or statement in one language by the same message and/or statement in another language." (as cited in LONG, 2013). In addition, Rahimi states that:

A translation will be considered inaccurate if it consists of the following cases:
i) Inadvertently omitting some pieces of information

ii) Adding information which is not really in the source text

iii) Committing mistakes during the analysis of the source text resulting in a different meaning (2004, as cited in Fadaee, 2011).

Thus, an inaccuracy that could possibly occur in the YouTube Auto-Generated Subtitle, since it was produced by the automatic machine translation, may result in the misconception or wrong-delivered concept of the source language in the target text by the viewer of the content on YouTube. For that matter, the researcher in this study is interested in analyzing the types of errors that occur in the YouTube English-Indonesian Auto-Generated Subtitle by comparing it with the Official Indonesian Subtitle.

According to TED.com, the subtitles in TED-Ed Channel are translated by TED Translators to help disseminate knowledge, research, and big ideas across languages and borders. TED Translators are volunteers from around the world who create subtitles for published TED, TED-Ed, and TEDx talks. The requirement to volunteer in TED, as a translator is one should be fluent in both the original Talk language and the translated language. Meanwhile, to volunteer as a transcriber; one should be fluent in the original Talk language. Furthermore, new volunteers should work on one assignment at a time, which is given within 30 days to be completed. In addition, before aiming to review, volunteers must have subtitled 90 minutes of talk content. After that, a TED Language Coordinator or the staff member will approve reviewed translations before publishing it. For that matter, the researcher is interested in analyzing the error types that occur in the YouTube English-Indonesian Auto-Generated Subtitle while comparing it with

the Official Indonesian Subtitle made by TED Translators in the content of TED-Ed Channel.

There are other researches discussing error analysis. One of the studies is a journal entitled *An Error Types Analysis on YouTube Indonesian-English Auto-Translation in Kok Bisa? Channelby Naftalia Laksana & Siegfrieda A. S. Mursita Putri* published in the Journal of Language and Literature, *Universitas Kristen KridaWacana*. The study analyzes the error types that frequently occur in the translation made by YouTube Auto-translate. The error classifications from Vilar et al. (2006) are used in this research. The data analyzed are fourteen videos from *Kok Bisa?* Channel. The type of analysis was a mixed quantitative and qualitative research. From the result, the most frequent error types in this research are wrong lexical choice, bad word form, missing auxiliary word, short-range word-level word order, and extra word.

Another study discussing error analysis is an undergraduate thesis entitled *YouTube Auto-Generated Subtitle Performance in Translating Content in Vogue Magazine Channel by Kesia Ruth Peneduh Sukma (2019)* published in the Repository Sanata Dharma University. The focuses of this research are the errors made by YouTube Subtitle and scrutinize the possible errors that occur in the translation produced by YouTube subtitles in the beauty and food content of the *Vogue* Magazine Channel. This research is a descriptive study using the error category based on Koponen's theory (2010). According to this research, the most common error types that occur in the two analyzed videos are a mistranslated concept.

The third study regarding error analysis is an undergraduate thesis entitled Google Translate assessment with error analysis: an attempt to reduce errors by Damianus Deni Kurnianto (2012) published in the Repository Sanata Dharma University. The focuses of this research are the errors made by Google Translate in translating three texts from English to Indonesian and the suggestions to reduce errors in using Google Translate. This research uses the error category based on Koponen's theory (2010). The result of this study shows a total of 206 errors in the three translated texts with the most frequent error type being a mistranslated concept with a total of 136 errors. In addition, Kurnianto suggested three methods to reduce errors which are typing in isolated form, text edition, combined methods.

The differences between this research and the previous studies above are the object and the theory. In this research, the researcher analyzes the error found in the English-Indonesian YouTube Auto-Generated subtitle with the error category by (Vilar, Xu, D'Haro, & Ney, 2006). The first previous study analyzes errors type analysis on YouTube Indonesian-English auto-translation. Furthermore, the third previous study analyzed a Google Translate assessment with error analysis and an attempt to reduce the errors. In addition, the second and third previous studies both use the error category by Koponen (2010). The similarity between this study and the first study is using the theory of error category by Vilar et al. (2006) to analyze the errors. Furthermore, similar to the second study, both are analyzing YouTube content. Besides, the second study analyzed beauty and food content, while this study analyzed educational content on YouTube.

1.2 Statements of Problem

1. What are the types of errors found in the Auto-Generated subtitle in TED Ed Channel?

2. What is the most frequent error found in the Auto-Generated subtitle in TED-Ed Channel?

1.3 Research Objectives

1. To identify the types of errors found in the Auto-Generated subtitle in TED-Ed Channel**.**

2. To identify the most frequent error found in the Auto-Generated subtitle in TED-Ed Channel.

1.4 Scope and Limitation of the Research

 To avoid misunderstanding and to clarify the problem, it is necessary to make a limitation of the problem. The researcher focuses on the errors of language made by YouTube Auto-Generated subtitles in TED-Ed Channel; not the technical issues. The unit of analysis is the English-Indonesian Auto-Generated subtitle of the video *The wicked wit of Jane Austen - Iseult Gillespie* with the length of 05:00 minutes compared to the Official Indonesian subtitle translated by Elda Indria Sari and reviewed by Reno Kanti Riananda as TED Translators in TED-Ed Channel. The Auto-Generated subtitle from the video in TED-Ed Channel was taken from April 17, 2020. The data were taken due to the errors that occur in the Auto-Generated subtitle in some of the videos on the TED-Ed Channel. The theory used to identify the errors found in the English-Indonesian

Auto-Generated subtitle of TED-Ed Channel is the error classifications by Vilar et al. (2006).

1.5 Research Method

The researcher analyzes the data by using a qualitative method. Qualitative research can be defined as an effective model that develops in a natural setting that allows the researcher to establish a level of detail from being highly connected in the actual experiences (Creswell, 2014). Denzin and Lincoln (2000), mention that qualitative research concerns in an interpretive and naturalistic approach: This means that qualitative researchers study things in their natural settings, attempting to make sense of, or to interpret, phenomena in terms of the meanings people bring to them." (as cited in Ospina, 2004)

Qualitative research focuses on reports, description and interpretation, which might lead to the development of new concepts or to an evaluation of an organizational process, on data that cannot be expressed in numbers using statistical techniques (Hancock, Ockleford, & Windridge, 2007). Thus, qualitative research is involved with transferring interpretations of the natural setting, approach, and absolute phenomena in some experiences or data that cannot be expressed numerically.

1.6 Technique of Data Collecting

The researcher analyzes the data by using descriptive analysis techniques. The research is accomplished through the steps as follows:

a.   Watching and reading the English-Indonesian Auto-Generated subtitle of TED-Ed Channel from the video which is *The wicked wit of Jane Austen - Iseult Gillespie*.

b.   Identifying where the errors occur.

c.   Collecting and writing the errors in the data card.

1.7 Data Analysis

There are some steps used in order to analyze the data by the researcher in conducting this research:

a. Applying errors classifications by Vilar et al. (2006).

b. Identifying and classifying the specific types of the errors.

c. Comparing the errors with the official subtitle.

d. Analyzing the types of errors that occur.

e. Summed up the data based on each error subcategory.

1.8 Data source

The data sources are obtained from the YouTube English-Indonesian Auto-Generated subtitle in TED-Ed YouTube Channel. The data are compared to the official Indonesian subtitle translated by TED Translator. TED-Ed YouTube Channel creates educational videos with animations. The subtitle is an Auto-Generated English-Indonesian YouTube subtitle.

1.9 The Organization of Thesis

The thesis organization of this research is consisting of Chapter I, Chapter II, Chapter III and Chapter IV. Chapter I includes an introduction that describes

the background of study; statement of the problem; research objective; scope and limitation of research; research method; the technique of data collecting; data analysis; data source; and the organization of the thesis. After that in Chapter II, the researcher gives further explanation of theoretical framework which consists of the definition of translation; subtitle; a subtitling process; translation issues; problems with automating translation of movie/tv show subtitles; problems related to subtitle creation guidelines; machine translation adaptability problems; and error analysis in machine translation. Furthermore, in Chapter III the data analyses from the data found in this research are explained. The last chapter, Chapter IV, provides the conclusion of the whole research.