

CHAPTER THREE METHODOLOGY

In this Chapter, the methodology of the study such as assumptions, research design and implementation will be discussed. This chapter begins with nature of the study and scope of the research, followed by assumptions, which cover subjects' language proficiency and the disturbances in the study. As for research design, the structure of the case study, the speech text, the pilot study, and the evaluation criteria are introduced. The implementation of the case study includes subject groupings, their backgrounds, sample size, and how the case study is conducted. The end of this chapter is dedicated to the discussion of research limitations.

3.1 Nature of the Study

According to Neuman (21-23), the purposes of research could be classified into for *exploration*, *description*, or *explanation*, depending on the kind of accomplishment that the researcher is after.

In an exploratory research, the researcher tries to learn about a new issue, which others has touched little on and to formulate more precise questions that future research can answer. It addresses the “what” question and is usually the first stage of the following studies. In a descriptive research, the researcher begins with a well-defined subject and then conducts research to describe it in specific details. It

deals with the “how” and “who” questions. An explanatory research usually builds on exploratory and descriptive research and goes on to probe into the reason “why” something occurs.

In this study, the difference *sight translation* (henceforth ST) training makes is the main focus. Because ST is often considered as only a small part of the newly developed field named interpretation, there has been limited research on it. This endows the study with an exploratory nature.

On the other hand, research, depending on its use, can be divided into *basic research* (also called *academic research* or *pure research*) and *applied research*. In simple terms, the former focuses on using research to advance general knowledge, whereas the latter uses it to address specific practical issue. (Neuman, 23-26) This study, which aims to explore more about ST, should belong to the former based on the above definitions for the results can help enhance our general knowledge about ST.

As for the academic research, Gile (*Observational Studies* 70-71) simply divides it into two fundamental categories: theoretical and empirical. The latter can be further classified into two categories: observational research, which consists in studying situations and phenomena as they occur ‘naturally’ in the field, and experimental research, in which situations and phenomena are generated for the specific purpose of studying them.

In this study, a speech text and three groups of subjects were chosen for the purpose of the case study. The subjects performed ST with the same text and under the same circumstances, which fits the description of “experimental research.” (Situations are controlled for the specific purpose of studying them.) However, the study also shares an observational nature since the situations are not really identical due to the limitations that will be described at the end of this chapter.

It should be noted that having the same task performed under identical conditions by more than one person is something rarely happens in the interpreting field, because in the real life situation, one interpreting task is usually done by one or two people only. It is almost impossible to observe and induct a general phenomenon this way. (Gile, *Observational Studies* 77) Therefore, this case study provides a good opportunity to observe more in ST with different groups of people performing the same task. In the future, if there are more follow-up studies with more statistics, a certain pattern might be formulated.

3.2 Scope of the Research

This research is based on a sight translation task performed by subjects from different backgrounds and the objective is to study its implications —especially the influence of ST training and its possible improvement.

The variables of this case study are levels of ST training and language

proficiency. Three groups of people of different ST training and language proficiency perform a same ST task and their different performances are analyzed to demonstrate the influence of the two factors.

The research would focus on the cross-analysis of the following dimensions: fidelity, ST chunking, and delivery, each with two to four subdivisions such as miss-interpretation, omissions (for the first dimension), chunk-moving, chunk-connection (for the second dimension), backtracks, and fillers (for the third dimension).

3.3 Assumptions

Due to the limited time and scale of the case study, it is not practical to try to cover all the variables. Hence, certain assumptions such as subjects' language proficiency for the speech text, the efforts they need, are made to narrow down the scope, reduce the complexities, and keep the case study in a more 'controlled' condition. The researcher tries to draw assumptions as reasonably and objectively as possible so that they can exert their proper functions.

The assumptions are: first, the speech text is effective in terms of examining the subjects' ST performances despite of their different language proficiencies because the words used in the speech are all listed in the high school English vocabulary in Taiwan, (Chang, 張武昌主持 八十四年度基礎科目英文考科試題研發工作計畫研

究報告 - 高中常用字彙表) which means the vocabulary poses no threat to the comprehension of the subjects, who have all completed college education. Besides, the issues covered (US-China dialogue, human right) are frequently discussed on papers or other news media. The possible difficulties lie in some complex sentence structures, which can serve as a good indicator of the subjects' ST skills based on how subjects deal with them.

The second assumption is that the disturbances such as possible differences of the environment and subjects' psychological conditions in this study are negligible because the text, the way the case study was conducted are the same, and the subjects of each group have similar language proficiency or ST training.

3.3.1 Subjects' Language Proficiency

The speech text was originally used as teaching material of general topics in the sight translation class for the first year student at the Graduate Institute of Translation and Interpretation, National Taiwan Normal University (henceforth GITI, NTNU). It concerns US-China dialogue and human right issue and was delivered by former US Secretary of State, Madeleine K. Albright to the Chinese Ambassador and his fellow officials on the National Day banquet at the Chinese Embassy.

Although the language proficiency for each group might be different, the subjects, whichever group they are in, should be able to understand the speech, for the

words used in the speech are all listed in the high school English vocabulary in Taiwan and the issues covered are familiar to the ear, as mentioned earlier. However, all the seemingly comprehensible words put together might generate difficulties for the subjects because comprehension of the sentences involves more than the sum of the individual words, and the relatively more complex syntactic structure thus produced might become the real challenge. The subjects' ways of handling the sentences can also manifest their ST skills, which will be ST chunking and delivery in this case study.

3.3.2 The Disturbances in the Study

To help the study going smoothly and make the results valid, the disturbances in the study must be reduced to the minimum. The possible disturbances include: the differences in the text, the subjects' psychological or physical conditions, the environment under which the ST task is performed, the steps taken to implement the study, etc.

In this case study, the environment (the language laboratory), the text, the steps of implementation, the size of each group are the same. The disturbances are the few exceptions that partook in the case study individually, the differences in their background, language deficiency or ST training within the same group.

3.4 Research Design

In this section, the structure of the case study, the features of the speech text, the definitions of the evaluation criteria: fidelity, ST chunking, and delivery, the pilot study, along with the significance of these evaluation criteria, will be described.

3.4.1 The Structure of the Case Study

There are three different groups to perform the ST task. Each group contains 15 people. Group A is of GITI students, who have received ST training for at least 3 months up to completion, which is 3 semesters. Group B is of graduate students from English department at NTNU, whose majors are linguistics, literature or English teaching, etc. Group C is of students under Government Employees English Program at NTU, level 5 (There are totally 7 levels in that program, so these students should be at upper intermediate level among all the students of the program.) Since they are government employees, their language proficiency is at least of college graduate level. Therefore, the variables are language proficiency and ST training.

The case study has the three groups sight-translate the same speech text and then analyze the results. The analysis adopts fidelity, ST chunking, and delivery as evaluation criteria based on the following viewpoints of scholars: the criteria concern ST chunking and delivery are formed according to the opinions of Weber (*Training*

Translators 33) on speed and clarity & conciseness of the rendition, Tang (155) on time control, speech flow, and backtracking, Gläser (125-127) on phraseological units, Fromkin and Rodman (65-69, 157-158) on meaning units, and Yang (112-13) on segmentation and information supply (the skills of chunking). From the information above, also for the purpose of this study, ST chunking is subdivided into chunk-moving and chunk-connection for observation, and delivery is further broken down to backtracks, fillers, pauses, and time control.

The criteria on Fidelity are derived from the definitions of errors provided by Barik (127-33), Altman (26, 28), Tang (152-54), and Her (115-29). The four scholars all have different categorizations toward errors specific to their researches, which are already mentioned in Chapter 2. However, they all mention errors that distort the original message, omissions or incomplete sentences, awkward renditions. Therefore, the criteria on fidelity for this study are miss-interpretation (error), omissions, unnatural Chinese (the awkward rendition of the target language).

3.4.2 The Speech Text

The choice of text should be general and comprehensible to trainees so that they don't have to worry about the obscurity or technical terms in the text and can concentrate on the mastery of ST skills. (Tang 154) Also, the text should be practical or come from a real source. For instance, opening remarks of a conference,

introduction of an institution, a speech, etc., so that trainees can emulate the real situations. (Yang 111) In this case study, a speech text prepared for the US Secretary of State to the Chinese Ambassador back in 1998, is used based on the logic above.

The text, as mentioned earlier, was chosen from the first year ST class materials. It is an excerpt of a speech delivered by Ms. Madeleine K. Albright to the Chinese Ambassador and his fellow officials on the National Day banquet at the Chinese Embassy. Ms. Albright, other than expressing her congratulation to the Chinese National Day, conveyed at the same time her opinions about China's human rights situation. All the words in the speech are of high school level as mentioned in the previous section. However, some of the sentence structures are rather long and complex, which is common in diplomatic rhetoric, and might pose challenges to the subjects. Generally speaking, the more embedded clauses or phrases a sentence has, the more complex it becomes. In this case study, the phrases are broken into or combined as meaning units for translation purposes, which possess complete semantic meanings even in the target language. For example, in the first sentence, "this dialogue," "is designed," "to help us" are the first three meaning units because they can respectively be a complete meaning unit, and each meaning unit is usually uttered as a whole, not broken down further, in the target language. Following this logic, sentence 8 (I/ could not/ fairly represent them/ if/ I/ did not/ emphasize/ America's

belief/ that/ organized and peaceful/ political expression/ is not/ a crime or a threat;), which consists of 2 embedded clauses or 13 meaning units, is one of the most complicated sentences in the text. Meanwhile, the only proper noun phrase, which is the name of a convention, had been explained to the subjects right before they performed the ST so as to reduce the possible interferences.

The speech text makes a proper ST test because it's not suitable for literal translation, and therefore, interpreters are forced to try other methods such as paraphrasing or rearranging to make the delivery natural in the target language. In so doing, the ST skills such as chunking, smooth delivery that the interpreters possess can be observed (Lambert and Meyer, 277-278, qtd. In Tang 145). The speech excerpt contains 262 words, which, though short, is complete in meaning. Besides, the subjects were briefed before the task to make sure they wouldn't lose the context. For the speech text, please refer to Appendix A.

3.4.3 Subjects' Processing Capacity

It is stated in the previous chapter that ST requires three efforts: reading, the speech production, and coordination; therefore, the results of the case study can be broken down into the three areas. By examining the subjects' fidelity (reading & speech production efforts), chunking skills (reading & speech production efforts), and delivery (speech production & coordination efforts), the different ST skills or styles

employed by subjects with different language proficiencies, and with/without ST training should be revealed. A further analysis of the influence of language proficiency and ST training, as well as what can be done for future improvement can then be discussed.

This study assumes that the processing capacity of the subjects is similar if they are in the same group because of their similar language proficiency and their ST training (all with or all without training in the group). For example, subjects in Group A all passed a stringent exam to enter GITI. All of their TOEFL (Test of English as a Foreign Language, usually used as a requirement for admission of foreign students into U.S. universities) scores are above 600, the full marks being 677. According to the Cultural Division of Taipei Economic and Cultural Office in New York (駐紐約台北經濟文化辦事處文化組), the average score for people in Taiwan taking TOEFL is 510. This indicates that Group A subjects possess rather high language proficiency, and all had received or are receiving intensive ST training. Subjects in Group B all passed the entrance exam of graduate school of English at NTNU, which is a renowned graduate institute of English literature, language, and teaching in Taiwan with a rather low admission rate of around 5%. According to their director Chang in a telephone interview, the students' TOEFL scores range from 580~620. Therefore, these subjects also possess rather high language proficiency,

but they don't receive any ST training.

Subjects in Group C all took an English proficiency test to enter the same class at the same level (5 out of 7 levels, about upper intermediate) of the GEEP (Government Employees English Program), and were receiving ST training at the time of the case study. Though their TOEFL scores are unspecified, they all completed their English language training in college at least. However, English is no longer a requirement after they left school, and based on the evaluation of the English proficiency test conducted by the National Taiwan University language center, their language proficiency is just the average (level 5). Therefore, it can be assumed that their language proficiency is of average college level, and consequently lower than the other two groups. For more information of their background, please refer to 3.5.2 Subjects' Background later in this Chapter.

3.4.4 The Pilot Study

In order to find out what to focus on, the researcher conducts a pilot study by transcribing the ST of one subject from each group. The original text can be found in Appendix A and the transcriptions, Appendix B. After examining the transcriptions, the differences among the subjects from each group emerged. For example, the different ways of comprehending and segmenting sentences, the various numbers of mistakes and redundancies in respective transcriptions, the common

difficult words or structures for all the subjects, etc. The evaluation criteria were later revised based on the findings of this pilot study (See 3.4.7 for a description of evaluation criteria.).

3.4.5 Definition of Fidelity

When it comes to evaluating the performance of an interpreter, most people in this field agree that fidelity should be the first priority. Yang (164) suggests that when conducting a professional exam for potential interpreters, fidelity should take up the most weight (50%), and it should include the examination of comprehension, omissions, and over translation. Among the three items, comprehension is deemed the most important, while the other two have room for ‘negotiation’ or ‘self-explanation.’

The study therefore proposes to observe 1. miss-interpretation (mainly concerns comprehension, although there could be other factors such as slip of the tongue or lack of monitoring, they are the minority in this study, probably due to the lack of incoming sounds and time pressure); 2. omissions; and 3. unnatural Chinese (the literal rendition of the target language), evolved from the three items mentioned in the previous paragraph.

Nevertheless, this is one single observation without follow-up interviews or traces, so there is no opportunity for subjects to explain themselves, which means the

omissions would be judged rather rigidly, while miss-interpretation, or, meaning errors, which indicate the misunderstandings of the speech text would be calculated first of the three subdivisions.

3.4.6 Definition of ST Chunking

The so-called “chunking” means segmenting and arranging sentences into the target language based on proper “meaning units” or “conceptual units.”(Jones 80-82) During SI, the continuous incoming message makes the interpreter cannot but deal with the it by following the sequence closely, so that s/he can release the storage/memory burden and then deal with the new incoming message. With proper chunking, the interpreter can greatly improve the quality and speed of the SI. (Chang 張維為 48) Since ST can be deemed as a basic part of SI and an important skill when doing SI with text, the same principle should apply.

As for the definition of “meaning unit,” linguistically speaking, it is a unit that conveys some agreed-upon meaning that is supposed to be understood (Fromkin and Rodman 157-58). The smallest unit of linguistic meaning is a morpheme, and some small meaningful units can be combined into larger units that also convey meaning. For example, morphemes can be combined into words, and words into phrases and sentences. (Fromkin and Rodman 63, 157, 171) Based on this logic, there could be different combinations of meaning units as long as they convey meanings. Therefore,

the researcher decides to divide the speech text into some proper units that carry meanings, that people can understand even in the target language, based on the theory of “phraseology” (Ed. Cowie) mentioned in 2.4. These meaning units can be referred to as “word-like” and “sentence-like” phraseological units. According to Gläser (125-27), word-like phraseological units are ‘nominations’ and designate a phenomenon, an object, an action, a process or state, a property or a relationship in the outside world. They are manifest in the traditional parts of speech, which are related to these conceptual entities: nouns, verbs, adjectives, and adverbs. They include idioms and non-idioms (technical terms, proper names, clichés, and other set expressions). Sentence-like phraseological units are ‘propositions’ and designate a whole state of affairs in the outside world. Their logical structure consists of a nomination and a predication. They include: proverbs, commonplaces, (e.g. Boys will be boys.) routine formulae, (e.g. Mind the step.) slogans, commandments and maxims, (e.g. Thou shalt not kill. Thou shalt not steal. Do it yourself.) and quotations and winged words, (e.g. A Jekyll and Hyde. Catch 22.) etc.

In the text lack of the sentence-like phraseological units, the sentences of the speech will thus be divided according to the word-like phraseological units. The transitions between two sentences, which should be as natural and logical as possible, will also be observed. The following are some examples of chunking of a sentence

from the speech text done by subjects, and the slashes on the original sentence represent the division of meaning units:

Sentence 5: And/ it has allowed us/ to speak frankly/ to each other,/ in the
knowledge/ that our relations/ will be stronger/ if we/ are honest/
about our disagreements.

Subject A: 並且使得雙方能夠坦承的交流意見，因為我們知道在意見發生衝突時，唯有互相坦承才能讓關係更為密切。

Subject B: 它使我們能夠彼此坦誠相見。基於這樣的認識，我們的關係會更加穩固，如果我們能夠誠實表達我們的歧見。

Subjects A and B basically follow the original chunk order. Subject B follows even more closely, but the chunk connection is not as natural as that of subject A.

For the segmentation of meaning units in the speech text, please refer to Appendix C.

3.4.7 Definition of Delivery

ST is generally considered a warm-up activity for SI or a basic ingredient of interpretation. Therefore, when exercising ST, delivery should be taken into account. Delivery can be further divided into fluency, body language, and intonation, etc (Yang 163-65). The focus in this dimension is put on fluency, which includes backtracking, fillers, unnatural pauses, and time control. The four subdivisions are evolved from

the literatures reviewed on 2.2.3 Delivery, the pilot study of this research, and Yang's (163-65) evaluation on delivery in a professional exam. Backtracking means unnecessary repetitions of words or phrases. Fillers mean any unnecessary words used when interpreting such as 那麼、呢 that are not in the source language and do not fit in to the target language. For unnatural pauses, the research gives it a quantitative measure of 3-second-or-above silence (the reason for this standard has been stated in 2.2.3 Delivery).

3.4.8 Evaluation Criteria

Based on the pilot study, a few phenomena are observed. For example, miss-interpretation (including vocabulary, tense, etc.) or under-interpretation (including omissions), backtracks, shuffles of sentence orders, and fast or slow speed when there's no time limit, etc. can be found in all three transcriptions. It's only a matter of degree. However, the 'different degrees' reveal some information such as the subjects' processing capacity and the role of ST training, etc.

Turning to the theoretical side, a set of guidelines for sight translation is provided by Weber (*Training Translators* 33), stating that the instructor of ST should closely observe the students' skills in public speaking, speed, and clarity & conciseness. In this way, students can improve their delivery—the output. Yang (112-13) also suggests that when teaching ST, the emphasis should be put on segmentation,

filling-in information, rhythm control, natural colloquial expressions, etc. Both of them focus more on the production and coordination sides, with the reading effort hidden as seemingly a prerequisite that is already there.

However, through the pilot study, it can be revealed that to be able to comprehend the material in the case study, one needs certain language proficiency to handle the “reading effort.” Besides, with the aid of ST training, the results in the 3 dimensions—fidelity, ST chunking, and delivery--will all be better. Because the subjects with high language proficiency and ST training have the most satisfactory ST performance, and the subjects with lower language proficiency have more trouble comprehending the text. The details of the results will be described in the next chapter “Research Findings.”

3.5 Implementation

This section discusses subject groupings, their backgrounds, sample size, and how the case study is conducted.

3.5.1 Subjects’ Groupings

There are 3 groups of subjects in the case study. Group A is made up of GITI students, who passed a very stringent exam to study at the institute and have received 3 months to 3 semesters of ST training. Group B is made up of graduate students

from English Department, NTNU, and some of them have completed their studies. Though they never receive any ST training, their language proficiency is widely recognized. Group C is made up of government employees under the GEEP at NTU, level 5 out of the total 7 levels, and they had received over one month of ST training at the time of the observation.

The purpose of the groupings is to examine the influences of different language proficiencies and ST training to ST performance. Group A has both high language proficiency and ST training and expected to perform the best. Group B has high language proficiency but lacks ST training, and probably can't perform as well as Group A because of the latter. Group C has relatively lower language proficiency but has received some ST training. The results of Group B and Group C could probably demonstrate the significance of language proficiency or ST training.

3.5.2 Subjects' Background

The three groups of subjects are categorized and described of their college major, work experience, and ST training in the following three charts:

Table 3.1

Background information of Group A subjects

The subject	College major	ST training
A	English	1 semester
B	Entomology	1 semester
C	English	1 semester
D	English	3 semesters
E	Chinese	1 semester
F	English	1 semester
G	English	1 semester
H	English	2 semesters
I	English	3 semesters
J	English	1 semester
K	English	1 semester
L	English	1 semester
M	English	2 semesters
N	Aeronautical Engineering	2 semesters
O	English	2 semesters

Source: compiled by this study

Table 3.2

Background information of Group B subjects

The subject	College major	ST training
A	English	None
B	English	None
C	English	None
D	English	None
E	English	None
F	English	None
G	English	None
H	English	None
I	English	None
J	English	None
K	English	None
L	English	None
M	English	None
N	English	None
O	English	None

Source: compiled by this study

Table 3.3

Background information of Group C subjects

The subject	College major	ST training
A	Finance	1 month
B	Law	1 month
C	Electronics	1 month
D	Marine	1 month
E	Industrial Management	1 month
F	Law	1 month
G	Law	1 month
H	Water Resource Engineering	1 month
I	Finance	1 month
J	Economics	1 month
K	Accounting	1 month
L	Financial Management	1 month
M	Transportation Management	1 month
N	History	1 month
O	History	1 month

Source: compiled by this study

3.5.3 Sample Size

Each group contains 15 subjects, which is a bit too small for quantitative research. It is due to the fact that GITI has only been established for less than 5 years with a very strict standard of student screening. 15 is about the number of all the students in the interpretation section still can be found in the institute (some students have gone abroad due to the requirement of GITI) at the time of the observation. The other two groups are set with the same number for the convenience of case study and comparison.

3.5.4 Implementation

In order to save time and to reduce the disturbances to the minimum, the case study is done by recording each group's ST in a language laboratory respectively, so that each group can perform ST at the same time under the same environment. Therefore, it took three times to record all of the groups. A few of the subjects who could not join their groups in the lab were recorded separately in the teacher's lounge at GITI.

The subjects were given the speech text and a briefing of the background of the speech (Ms. Albright's speech at the Chinese embassy "in celebration of" Chinese National Day). There was a 5-minute preparation time and no time limit for delivery. During the preparation, the subjects were allowed to make any notes on their sheet of

text or look up the dictionaries, and they could talk to each other or ask questions if they want to. After the case study, their performances were evaluated one by one under the same criteria, and the criteria would be discussed separately in the following chapter.

3.6 Research Limitations

The way this case study was conducted makes it hard for the researcher to observe some parts of the performance such as the body language, the strategies taken, and the interaction with the audience, etc. Furthermore, the researcher's judgment calls of the three dimensions, namely, fidelity, ST chunking, and delivery, though based on some theories and principles, are not objective enough because there is not a set of rules to distinguish all the ST phenomena and the subjects' intentions were not followed to help the judgments.

3.6.1 Gray Areas in Evaluation

As mentioned earlier, there are some gray areas in the evaluation done by the researcher because of insufficient information. For example, when “nuclear proliferation” is interpreted into 核子彈的問題 (nuclear bomb issue), should it be considered miss-interpretation, or an omission? What about “I could not fairly represent them if I did not emphasize America's belief” translated into 「今天我代表美國人民，就必須強調美國的信念」 (Today I represent the American people, so I

have to emphasize America's belief.)?

Therefore, the researcher has to make rather subjective judgments when there are no specific rules but only principles. Consequently, the first example in the previous paragraph is considered an omission rather than miss-interpretation, because it dilutes the original message but still maintain the gist of the message. The second example is considered OK because it only involves phrasing change but not meaning change.

3.6.2 Missing Audience

ST is needed not only for SI exercises but also for conference interpreting situations. It is an important skill when performing SI with text. Therefore, the pressure or disturbances brought by facing the audience should also be considered in ST performances.

Since this ST task was recorded, there would not be any audience, so this stress factor is excluded. The subjects are supposed to perform more naturally and freely during the case study than in the market. Therefore, the results in the case study may be an overestimate of their ability in the real situation.

3.6.3 Limited Sample Size

The limited sample size makes it difficult to draw quantitative conclusions. That is, the conclusions won't be convincing and representative enough since each

group contains only 15 subjects.

However, 3 groups are used to increase the variety for comparison. Group A contains students from first year to graduation, almost everyone available (still in school and still in Taiwan) at that time, about 15 people. For the convenience of observation and calculation, the numbers of subjects in the other two groups are set to be 15 as well. Hopefully, in the future, when GITI is more established and have more students, the choice of subjects will not be as limited and this type of quantitative case study will then become meaningful and solid.

Nevertheless, it should also be noted that Group C is comprised of subjects with very different college majors and ages. Thus, it is hard to prove that their abilities in memory, rhetoric, etc. are similar since their backgrounds are so different. The only thing that can be assumed is that their language proficiency is similar because they all took the same English proficiency test and then were all categorized in level 5.

3.6.4 Limited Scope

After the pilot study, the researcher discovered many different dimensions to analyze such as the types of errors made, different styles of delivery, etc. They all deserve to be further examined. However, for the sake of efficiency and focus, the researcher has to narrow the scope down to 3 dimensions with assumptions to make them easier to measure. After the analysis, some phenomena, such as the intonation

and voice quality of the subjects, or their murmuring the speech text when interpreting, are not observed because they don't belong to any of the criterion.

According to Gile (*Observational Studies* 84), complex human behavior such as interpreting is difficult to measure directly in a precise, thorough and uniform way. This case study also faces the same difficulty. Without a set of comprehensive measures and long-term tracking, the study can only try to quantify the assumed criteria within the 3 categories to compare and analyze the differences generated in the 3 groups through the ST task.

However, through the comparison of the 3 groups, the effect of ST training and language proficiency can still be observed, which can then still serve the purpose of the study.

3.6.5 Other Limitations

The type and length of the speech text are limited because this is a one-time observation. The results of ST under a diplomatic context, which contains 262 words, are observed. However, more researches need to be done if we want to know about how the subjects will perform with different types of speech or with a longer text.

On the other hand, due to the nature and scope of the study, it is hard to make follow-up observations to see the subjects' progress. This along with the limitations

mentioned above reduce the possible achievements of the study and leave considerable room for future research.