

学校代号 10532

学 号 S2012W1032

分 类 号

密 级 公开



硕士学位论文

基于文学文本的翻译修订与神经机器 翻译译后编辑对比研究

学位申请人姓名	李小叶
培 养 单 位	外国语学院
导师姓名及职称	王湘玲 教授
学 科 专 业	翻译硕士
研 究 方 向	英语笔译
论文提交日期	2022 年 5 月 13 日

学校代号：10532

学 号：S2012W1032

密 级：公开

湖南大学硕士学位论文

基于文学文本的翻译修订与神经机器翻译
译后编辑对比研究

学位申请人姓名	李小叶
导师姓名及职称	王湘玲 教授
培 养 单 位	外国语学院
专 业 名 称	英语笔译
论文提交日期	2022 年 5 月 13 日
论文答辩日期	2022 年 5 月 17 日
答辩委员会主席	余承法 教授

**Translation Revision and Neural Machine Translation
Post-editing of Literary Texts: A Comparative Study**

by

LI Xiaoye

**B.A. (Southwestern University of Finance and
Economics) 2020**

A thesis submitted in partial satisfaction of

the Requirements for the degree of

Master of Arts

in

English Translation in the Graduate school

of

Hunan University

Supervisor

Professor WANG Xiangling

May, 2022

摘 要

近些年来,翻译技术的进步促使译者的身份开始发生转变,“翻译服务提供者”这一称呼更能体现当前译者的工作内容。这意味着译者不再是单纯地翻译文本,而是会花更多的时间从事修订、译后编辑等工作。市场的迫切需求催促我们开展有关修订和译后编辑能力的研究。然而,与二者能力相关的实证研究数量较少,多以产品数据为主,且研究结果并不一致。因此,有必要进一步开展实证研究对比译者的翻译修订和译后编辑能力。除此之外,神经机器翻译的出现为进一步拓展所研究的文本类型提供了可能。以往机器翻译译后编辑与人工翻译的对比研究多以非文学文本为主,基于文学文本的机器翻译译后编辑研究数量较少,机器翻译的文学译文在译后编辑之后能否在质量方面超过“人工翻译+修订”这一模式,这一问题有待回答。

因此,本文拟从文学文本着手,综合运用调查问卷、击键记录和回溯性报告,旨在探究学生译者在修订和译后编辑当中的行为差异,并尝试探索神经机器翻译译后编辑在文学翻译当中的应用前景。本研究拟回答以下问题:(1)被试在修订和译后编辑文学文本时如何修改译文?(2)与修订相比,被试在译后编辑文学文本时会投入多少译后编辑努力?具体而言,被试在译后编辑文学文本时是否会投入更多的时间努力、技术努力和认知努力?(3)与修订的文本相比,译后编辑的文学文本的翻译质量如何?具体而言,两类译文的充分性和流利度如何?

本研究共收集到48名受访者的问卷反馈,并从中随机抽取21名学生译者参加修订和译后编辑实验,经分析发现:(1)从修改行为来看,学生译者在译后编辑中的修改次数将近修订中的1.5倍。学生译者在译后编辑中的修改质量显著优于其在修订中的修改质量。(2)从译后编辑努力来看,学生译者在译后编辑中投入了更多的时间努力、技术努力和认知努力。(3)从翻译质量来看,修订和译后编辑之后的文学译文在准确度和流利度方面都比较相近,并不存在显著差异。最后,本文将为相关领域实证研究提供方法论参考,本研究发现将为翻译教学及译者修订和译后编辑能力的培养提供有益启示。

关键词: 翻译修订; 神经机器翻译译后编辑; 修改行为; 译后编辑努力; 翻译质量; 文学文本

Abstract

With the advancement of translation technology, translators have transformed into Translation Service Providers (TSPs) who are probably spending more time on revision and revision-related activities such as post-editing. The urgent market demand has encouraged more research on translation revision and post-editing competences. However, few empirical studies of those two competences have been conducted. Most of them rely on product data and the results of previous studies are mixed. Therefore, it is necessary to conduct an empirical research to further investigate the difference between translation revision and post-editing competences. Moreover, neural machine translation (NMT) might make it possible to apply MTPE into literary translation. Previous comparative studies between MTPE and human translation mainly focus on non-literary translation. But there is a concern whether the quality of post-edited literary NMT will exceed that of revised literary HT.

Therefore, the present study has adopted research methods like questionnaire, key logging and retrospective protocol so as to investigate the differences of translation trainees' behaviors in revision and post-editing of literary translations and explore the application prospect of MTPE in literary translation. The research aims to answer the following questions: (1) How do participants rectify literary texts in revision and post-editing? (2) Compared with revision, what is post-editing effort spent by participants in post-editing of literary texts? Specifically, how much temporal, technical and cognitive effort do translation trainees spend in translation revision and NMT post-editing respectively? (3) Compared with revised text, what is final translation quality of post-edited literary texts? Specifically, what are the adequacy and fluency of revised and post-edited literary texts?

The present study has collected questionnaire data from 48 respondents and invited 21 among them to involve in translation revision and post-editing tasks. The results showed that: (1) From the perspective of intervention behaviors, the number of changes made by translation trainees in NMT post-editing is one and half of that in translation revision. Translation trainees obtained higher intervention quality in

post-editing than in translation revision. (2) As for post-editing effort, translation trainees indeed spent much more temporal, technical and cognitive effort in post-editing than in translation revision. (3) When it comes to translation quality, the average adequacy and fluency of revised HT and post-edited NMT are close to each other and there is no significant difference. To the end, the present study will provide some suggestions for related empirical research in terms of methodology. The findings of the present study will benefit for the translation training.

Key Words: translation revision; neural machine translation post-editing; intervention behaviors; post-editing effort; translation quality; literary texts

Contents

Chapter 1 Introduction	1
1.1 Research Background and Significance	1
1.2 Research Purpose and Research Questions	4
1.3 Research Methodology	6
1.4 Organization of the Thesis	7
Chapter 2 Literature Review	9
2.1 Translation Revision	9
2.1.1 Revision of translation	9
2.1.2 Intervention behaviors.....	11
2.2 Machine Translation Post-editing	14
2.2.1 Post-editing of neural machine translation	14
2.2.2 Post-editing effort.....	16
2.2.3 Translation quality.....	17
2.3 Comparative Studies of Translation Revision and Post-editing	18
2.4 Research on Revision and Post-editing of Literary Texts	20
2.5 Summary	21
Chapter 3 The Empirical Study	23
3.1 Research Design.....	23
3.1.1 Participants.....	23
3.1.2 Materials.....	24
3.1.3 Instruments.....	33
3.1.4 Experimental site.....	36
3.1.5 Pilot study.....	36
3.2 Data Collection	37
3.2.1 Experiment procedures	37
3.2.2 Data resources	39
3.3 Data Exclusion	39
3.4 Data Analysis	40
Chapter 4 Results and Discussion.....	41
4.1 Intervention Behaviors in Revision and Post-editing of Literary Texts	41
4.1.1 Intervention behaviors.....	41
4.1.2 Intervention quality	45
4.2 Post-editing Effort in Revision and Post-editing of Literary Texts	47

4.2.1 Temporal effort.....	47
4.2.2 Technical effort	49
4.2.3 Cognitive effort	50
4.3 Translation Quality in Revision and Post-editing of Literary Texts	54
4.3.1 Adequacy	54
4.3.2 Fluency	56
4.4 Summary	57
Conclusion	59
Notes	63
References.....	64
Appendix B 调查问卷	75
Appendix C 实验文本	78
Appendix D 任务纲要.....	81
Appendix E 翻译修订指南	82
Appendix F 译后编辑指南	83
Appendix G 实验知情同意书.....	84
Appendix H 实验操作步骤.....	86
Appendix I 任务难度及认知负荷主观评分	87
Appendix J 详细中文摘要	88

List of Figures

Figure 1.1	Framework of Research Purpose and Research Questions	5
Figure 1.2	Independent and dependent variables	6
Figure 1.3	Framework of research methodology	7
Figure 3.1	Number of error annotations in HT and NMT of ST1 and ST2	29
Figure 3.2	Error annotation and error score per text in HT and NMT	30
Figure 3.3	Number of error annotations in specific types in NMT and HT	30
Figure 3.4	Proportion of error types in NMT of ST1 and HT of ST2	33
Figure 3.5	Screenshot of User interface of Translog-II	35
Figure 3.6	Screenshot of EVCapture	36
Figure 3.7	Screenshot of Supervisor interface of Translog-II	38
Figure 3.8	Experiment procedures	38
Figure 4.1	Distribution of intervention behaviors per task	42
Figure 4.2	Intervention quality per participant	46
Figure 4.3	Temporal effort of post-editing (P) and translation revision (R)	48
Figure 4.4	Technical effort of post-editing (P) and translation revision (R)	49
Figure 4.5	Cognitive effort of post-editing (P) and translation revision (R)	51
Figure 4.6	AVEAdequacy score per participant in both tasks	54
Figure 4.7	AVEFluency score per participant in both tasks	56

List of Tables

Table 3.1	Basic information of ST1 and ST2	25
Table 3.2	Error types	28
Table 4.1	Descriptive Statistics and Normality Tests of intervention quality	47
Table 4.2	Significance Test of intervention quality	47
Table 4.3	Descriptive Statistics and Normality Tests of temporal effort	48
Table 4.4	Descriptive Statistics and Normality Tests of technical effort	49
Table 4.5	Significance Tests of technical effort indicators	50
Table 4.6	Descriptive Statistics and Normality Tests of cognitive effort	51
Table 4.7	Significance Tests of cognitive effort indicators	52
Table 4.8	Descriptive Statistics of subjective assessment on task difficulty	53
Table 4.9	Significance Tests of cognitive effort indicators	53
Table 4.10	Descriptive statistics and Normality Tests of AVEAdequacy	55
Table 4.11	Significance Test of AVEAdequacy	55
Table 4.12	Descriptive Statistics and Normality Tests of AVEFluency	56
Table 4.13	Significance Test of AVEFluency	57

List of Abbreviation

ST	Source Text
TT	Target Text
HT	Human Translation
MT	Machine Translation
NMT	Neural Machine Translation
MTPE	Machine Translation Post-editing
TRC	Translation Revision Competence
PEC	Post-editing Competence
IV	Independent Variable
DV	Dependent Variable
TSPs	Translation Services Providers
SMT	Statistical Machine Translation
SPSS	Statistical Package for Social Science

Chapter 1 Introduction

Translation revision and post-editing are two primary ways to safeguard translation quality. In digital translation environments, translation revision and post-editing are parts of the same workflow as translation and depend on similar skill sets. In this chapter, the background and significance of the present study will be introduced firstly, which is followed by the research purpose and questions. Moreover, the research methodology, including research methods and instruments, will be stated since the present study is an empirical research. Finally, the organization of the thesis will be introduced.

1.1 Research Background and Significance

The ongoing technologization of translation services has transformed translators into Translation Service Providers (TSPs) who are probably involving revision and post-editing tasks more frequently than before (see Mossop 2007; Martin 2007; O'Brien 2010). The increasing applications of translation revision and post-editing demonstrate the pivotal role of these two tasks in recent translation industry.

According to ISO 17100, translation revision is a necessary activity that must be carried out by someone other than the translator, which is composed of “a bilingual review of the content of the target language and the source language to determine its suitability for the agreed purpose” (ISO 2015:2). Therefore, translation revision in this study mainly refers to the act of checking and amending others' translation. Post-editing is defined as editing and correcting the output of machine translation (ibid). Both revision and post-editing are common methods adopted by people to increase the final translation quality. The intuitive difference between revision and post-editing lies on the fact that the former is aimed at correcting errors in human translation (HT) and the latter is aimed at correcting errors in machine translation (MT). While correcting errors in texts, translators may encounter different translation errors and have different performance in two scenarios. Given the same profiles among translators, their different performances can be attributed to the original

translation quality and their perceptions towards translation revision and post-editing (see Daems & Macken 2020).

When reviewing comparative studies on revision and post-editing, we find that most of them have been written from a theoretical perspective (see Jakobsen 2018; Mossop *et al.* 2020; do Carmo & Moorkens 2020; Konttinen *et al.* 2020). Some researchers propose competencies or competence models with regard to revision based on their respective studies (see Robert *et al.* 2017a; Scocchera 2019; Künzli 2006; Mossop *et al.* 2020). Rico & Torrejón (2012) and Pym (2013) list the necessary competencies for post-editing respectively. Konttinen *et al.* (2020) look at commonalities and differences in revision and post-editing competences, like error detection in strategic subcompetence, in order to formulate an efficient model for revision and post-editing training. However, these research results lack the support from empirical experiments. Relevant empirical studies are inclined to focus on a specific behavior of translators such as over-editing (see Nitzke & Gros 2020). Previous studies have left a lot of questions that remain unanswered. Therefore, it is urgent to explore the differences of translation trainees' behaviors between translation revision and NMT post-editing from an empirical perspective.

Besides, the rapid development of translation technology has encouraged more empirical studies on MT and post-editing in recent years. Researchers always compare them with translation from scratch when exploring the potential functions of MT. In some comparative cases of MTPE and translation from scratch, efficiency and quality of MTPE outperform that of translation from scratch (see Plitt & Masselot 2010; Koponen 2016; Daems *et al.* 2017; Wang *et al.* 2021a). However, those previous studies have been conducted with nonliterary texts as the experimental texts. MT might have an advantage to deal with this text type. Therefore, some researchers expand their studies to post-editing of literary translation. Toral Ruiz *et al.* (2018) test the usability of MT in literary translation and found that MTPE will improve the productivity of literary translators. But there is still a concern whether the quality of MTPE will exceed that of translation revision when it comes to literary translation. Therefore, we would like to investigate the final translation quality of revised literary HT and post-edited literary MT.

There are two reasons why we would like to investigate the differences of translation trainees' behaviors between translation revision and NMT post-editing of literary texts. Firstly, we would like to have a better understanding of their strategic subcompetence under different scenarios and draw the attention of translation trainers to the post-editing and revision competences of translation trainees. And the improvement of translator training system will enable translation trainees to quickly adapt to the translation industry after graduation. The second reason lies on exploring the potential of NMT post-editing on literary texts, compared with revised literary HT. And the significance of the current study is shown as follows.

First, the present study investigates the differences between translation revision and NMT post-editing of literary texts in a more comprehensive way. More specifically, when it comes to the comparative studies of revision and post-editing, researchers tend to focus on specific behaviors of translators, such as over-editing (Nitzke & Gros 2020) and necessary changes (Daems & Macken 2020). However, we use the Translog II program to record all intervention behaviors of translation trainees during the experiments and analyze their differences in two scenarios. The survey of questionnaire and retrospective reports will help explore the motivations behind the behaviors. In addition, this study also investigated post-editing efforts and translation quality in two tasks.

Second, this study provides empirical supports for theoretical research on the competences and teaching of revision and post-editing. In recent years, a raft of studies on revision and post-editing competences are emerging. Robert *et al.* (2017a) and Scocchera (2019) propose revision competence models respectively. Rico and Torrejón (2012) and Pym (2013) enumerate the competences required for post-editing. The theoretical studies have compared the differences between revision and post-editing competences, but few studies have empirically differentiated the concrete performance of translation trainees between revision and post-editing. Although revision and post-editing share some common subcompetences such as strategic subcompetence, they are different in certain aspects (see Robert *et al.* 2022). Revisers and post-editors will face different translation errors and use different tools in two scenarios (Konttinen *et al.* 2020). Therefore, this study hopes to explore the

translation errors that translation trainees might encounter in two tasks, triggering a discussion about translator competence development and providing an empirical support for theoretical studies.

Third, the present study also explores the potential of NMT post-editing on literary translation, compared with translation revision of literary texts. Recently, NMT has emerged as a new paradigm in MT and has drew a lot of attention. Scholars have compared them with translation from scratch in terms of productivity and quality. Whereas, most research results are acceptable for nonliterary texts such as technical documents since the main purpose of nonliterary translations is to preserve the meaning of the original text. It might not be the case for literary texts because literary translation requires translators to preserve not only the meaning of the source texts but also its reading experience (Jones & Irvine 2013). Though some researchers have applied MTPE into literary texts in their experiments, they mainly focus on the issue of productivity (see Toral Ruiz *et al.* 2018). There remains a concern whether the translation quality of post-edited literary MT exceed that of revised literary HT. Therefore, we conduct this experiment and try to answer the questions left by the previous studies.

1.2 Research Purpose and Research Questions

The present study aims to explore the differences between translation revision and NMT post-editing of literary texts from the perspectives of intervention behaviors, post-editing effort and translation quality. Firstly, we will exploit keystroke logging to record translator trainees' behaviors so as to compare the differences in two modes. The assessment of intervention behaviors is based on the translation errors annotated. Secondly, we will calculate post-editing effort translator trainees spent in two tasks and an adapted scale will be used to collect participants' subjective assessment of task difficulty and cognitive load. Thirdly, the present study will also assess and compared the final translation quality in both scenarios based on fluency and adequacy (TAUS 2013). Besides that, translation trainees' attitude towards translation revision and NMT post-editing will be presented in the discussion for a better explanation of those objective data. By doing this, the thesis intends to address the following questions:

- RQ1. How do participants rectify literary texts in revision and post-editing?
- RQ2. Compared with revision, what is post-editing effort spent by participants in post-editing of literary texts?
- 2.1 Do participants spend more temporal effort in post-editing of literary texts?
- 2.2 Do participants make more technical effort in post-editing of literary texts?
- 2.3 Do participants invest more cognitive effort in post-editing literary texts?
- RQ3. Compared with revised text, what is final translation quality of post-edited literary texts?
- 3.1 What is the adequacy of revised and post-edited literary texts?
- 3.2 What is the fluency of revised and post-edited literary texts?

The research questions are put forward to investigate the differences between translation revision and NMT post-editing of literary texts. Figure 1.1 below will help visualize the research purpose and research questions of the present study.

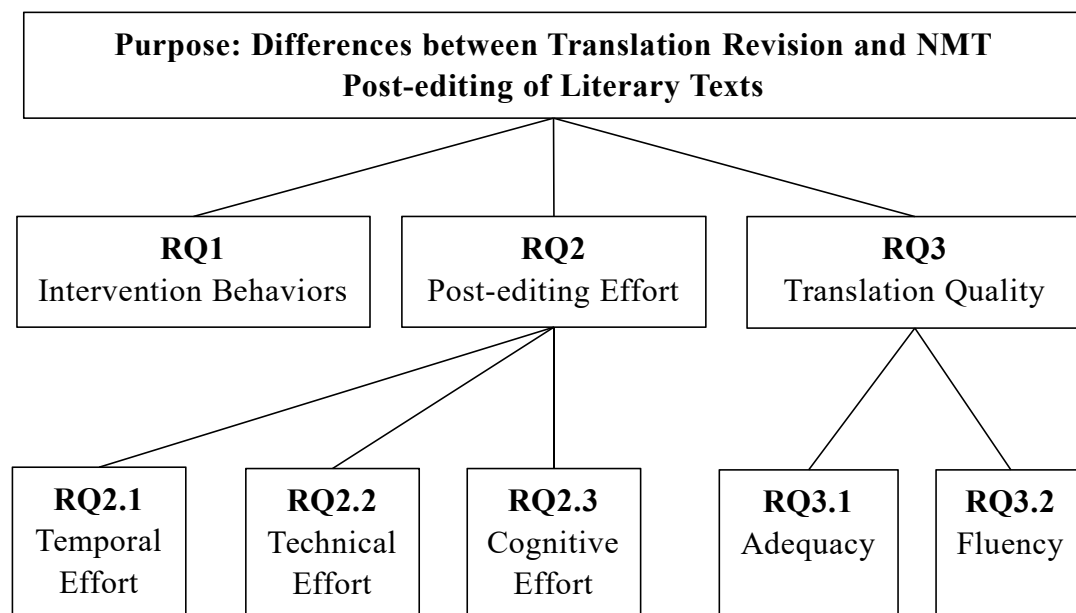


Figure 1.1 Framework of Research Purpose and Research Questions

In the present study, the independent variable is task type (translation revision and NMT post-editing). The dependent variables to probe into the differences of translation revision and NMT post-editing are intervention behaviors, post-editing effort and translation quality. To be specific, intervention behaviors will be assessed in accordance with the intervention typology proposed by Robert *et al.* (2017b).

Post-editing effort will be presented on the basis of Krings' model (2001) and the detailed measurement methods will be introduced in section 1.3. Translation quality will be evaluated based on adequacy and fluency (TAUS 2013). Figure 1.2 displays all variables in a much clearer way.

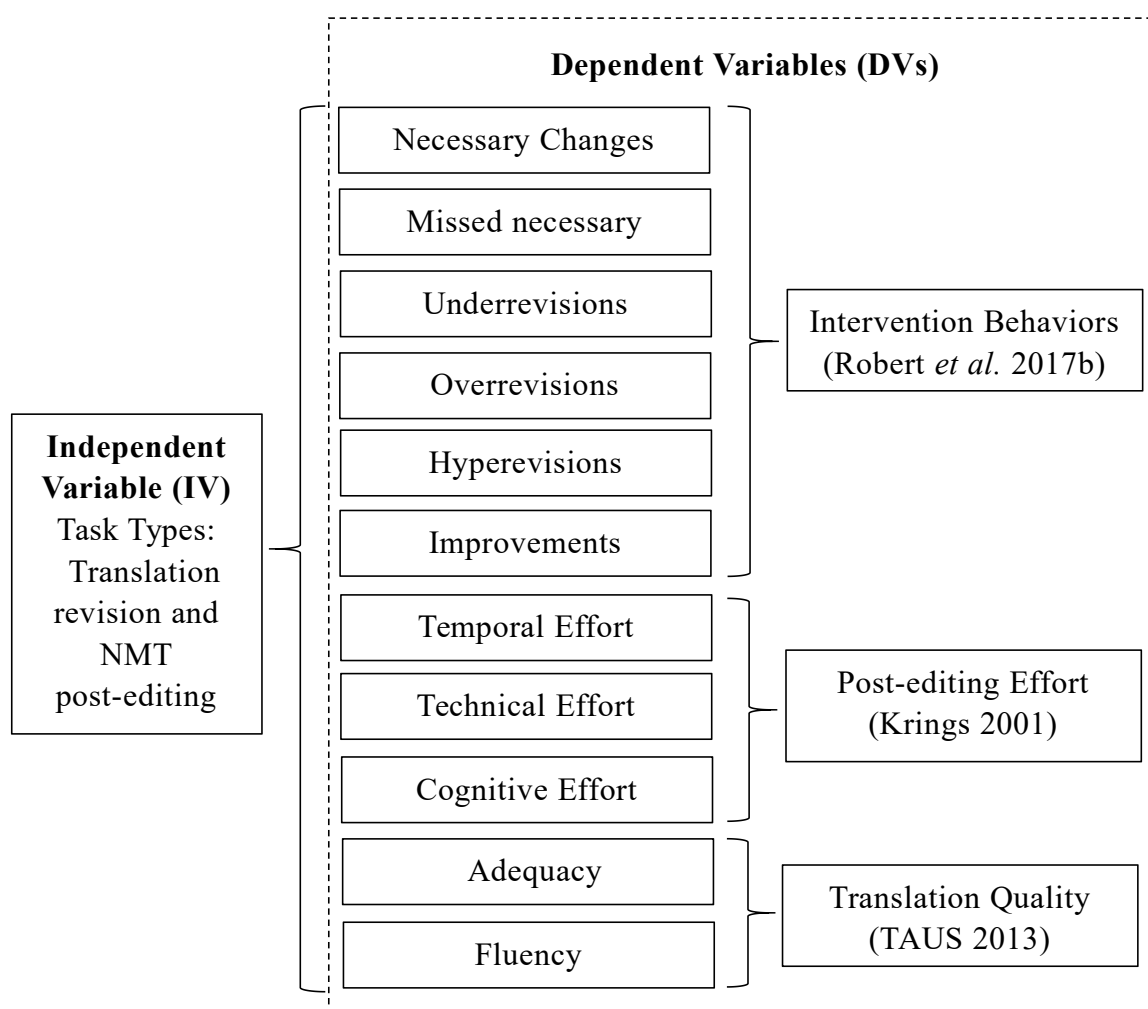


Figure 1.2 Independent and dependent variables

1.3 Research Methodology

The research questions will be addressed with the aid of quantitative and qualitative data. Quantitative data in the present study mainly comes from the questionnaire and key logging and qualitative data, a supplement of the results, will be collected from retrospective protocol (see Figure 1.3). The questionnaire aims to collect basic information of respondents and their attitudes towards translation revision and NMT post-editing. Items with regard to attitudes are designed as 5-point

Likert scale. The key logging instrument will record translator trainees' typing and pausing activities in two tasks, which will offer data with regard to post-editing effort. To be specific, temporal effort will be measured by the drafting time and the total task time. Technical effort includes the total number of keystrokes, insertions and deletions. Cognitive effort, the most difficult indicator to capture, will be measured by the number of pause and pause duration. Furthermore, participants are required to rate the task difficulty and cognitive load after each task with the aid of an adapted NASA Task Load Index (Sun & Shreve 2014) (see Appendix I). Retrospective report, as the main source of qualitative data, will be exploited to explain the results arising in the experiment.

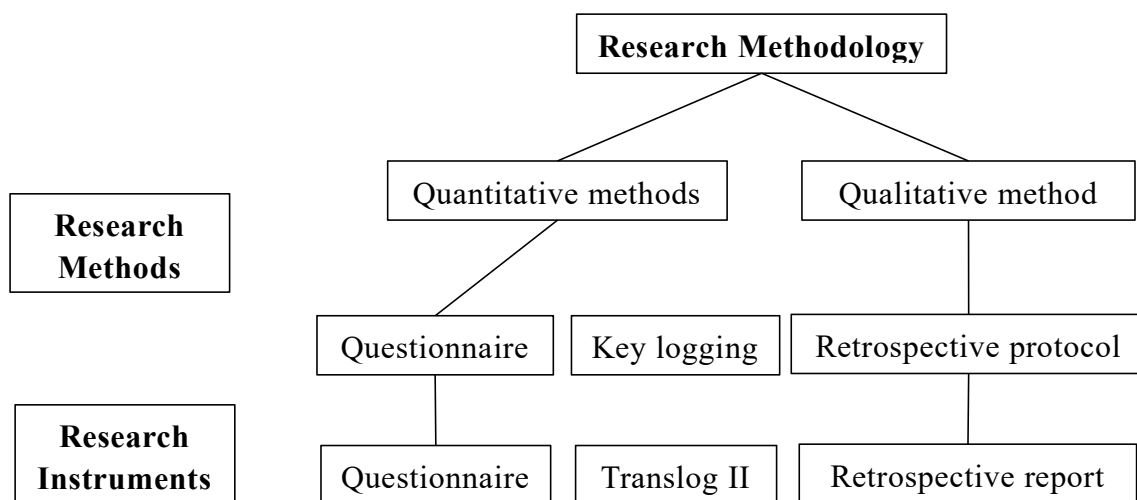


Figure 1.3 Framework of research methodology

1.4 Organization of the Thesis

This thesis is structured as follows. Chapter one is a general but essential introduction of the present study, which is composed of research background and significance, research purpose and research questions, research methodology and the structure of the thesis.

Chapter two reviews relevant previous studies of in translation revision, MTPE, comparative studies of translation revision and NMT post-editing and research on revision and post-editing of literary texts. Research on intervention behaviors, post-editing effort and translation quality also has been introduced respectively in this

chapter. It provides a theoretical and methodological foundation for the present study.

Chapter three introduces the research design firstly, including participants, materials, instruments, experimental site and pilot study. Text selection is of importance in the present study. We need to manipulate the quality of draft translations since it is a comparative study and the sources of translations are different. So, we highlight how to annotate translation errors and the distribution of errors in each text. After the preparation of experimental texts, we present a clear introduction of data collection, exclusion and analysis orderly. Experiment procedures and data resources are described in section 3.2.

Chapter four is a presentation and discussion of the results obtained in the present study. It could be divided into three sections with regard to translator trainees' intervention behaviors, post-editing effort spent and translation quality respectively. Six types of behaviors are assessed in section 4.1. Intervention quality is measured as well for a more intuitive understanding of translator trainees' performance in two scenarios. Afterwards, results of temporal, technical and cognitive effort are presented in section 4.2. A subjective assessment of cognitive load is added as a supplementation of cognitive effort since we measure this indicator by pauses not fixation. Section 4.3 analyzes the results of translation quality, which might shed light on the potential of NMT post-editing on literary translation. All data is presented in tables and figures. This chapter also attempts establish a link between the findings in the present study and those in previous relevant studies.

Chapter five is a conclusion of the present study. It generalizes the results of each research question and puts forward implications of the present study. Limitations are described in this chapter and some suggestions are provided for future studies.

References and other relevant materials such as experimental texts and questionnaire could be found at the end of the thesis.

Chapter 2 Literature Review

This chapter firstly introduces research on translation revision and MTPE respectively. To be specific, it includes the definition and current studies of these two key words. Research with regard to important variables like interventions behaviors, post-editing effort and translation quality are also introduced in the sections. Then, we climb comparative studies of translation revision and NMT post-editing from theoretical and empirical perspectives, which is followed by research on revision and post-editing of literary texts. Section 2.3 and 2.4 mainly focus on questions left by the previous studies. Given that it is an empirical study, we also review relevant research instruments so as to collect quantitative and qualitative data needed in the present study.

2.1 Translation Revision

In many translation agencies, translators who have transformed into translation service providers no longer simply translate texts. Among series of translation services, the proportion of revision is getting larger, which indicates that the demand for revision service is gradually increasing. Moreover, translation revision, as a key stage of the translation process, has gained an increasing attention from scholars in the translation studies over the last fifteen years or so (Scocchera 2018).

2.1.1 Revision of translation

Revision of translations is an ancient European activity dating back to the end of the 4th century in which Saint Jerome¹ is appointed by Pope Damasus I² to revise the existing Bible. However, it is not until the 20th century that revision has received the attention of major publishers, translation academics and institutions. Large institutions like the UN and the EU have attached importance to the function of reviser for a few decades, but it is not the case in translation agencies until the publication of the

European standard for Translation Services EN 15038 (2006) and its successor, ISO 17100 (2015).

As defined in these standards, revision is a necessary activity that must be carried out by someone other than the translator. In fact, it takes an arduous process for scholars to refine the concept of revision (see Robert 2008). Künzli (2005) points out that there were some “terminological confusion” in some studies of translation revision. To avoid terminological fuzziness, Robert *et al* (2017a) proposes four factors which need to be taken into consideration in research on translation revision: the text to be revised, the person who revises, the moment at which the text is revised and the way in which the text is revised.

According to the person who revises, the term “revision” could be categorized into self-revision and other-revision. Self-revision is an activity in which a translator revises his or her own translation (see Mossop 2007; Shih 2006). Other-revision is to revise a translation of another translator (see Mossop *et al.* 2020). And “translation revision” investigated in the present study refers to the act of checking and amending others’ translation. When revising a translation, translators are required to read a draft translation and then detect and correct errors according to the translation revision brief. They need to make appropriate corrections and improvements before the translation is delivered to the client (Rensburg 2017).

Empirical studies with regard to revision have continued to emerge in recent years. Most of previous studies focus on revisers’ revisions by establishing or using existing parallel corpora (see Popič 2014; Robin 2014; McDonough Dolmaya 2015; Bisiada 2018). Intervention behaviors are also an object of interest for scholars (see Brunette *et al.* 2005; Künzli 2007; Robert 2013; Robert & Van Waes 2014). Other objects of research on revision include background of translators (see Marashi & Okhowat 2013), revision style (Huang Xiaohan 2018), revision principles (Künzli 2005), the relationship between revision quality and profiles of revisers (Rensburg 2017).

Domestic research on translation revision is still incipient in comparison. Only two articles with regard to revision have been published in domestic journals to date. For instance, Huang & He (2017) explore the patterns of translation revision

behaviors of professional and translation trainees in terms of the distribution characteristics of those behaviors and the linguistic levels at which revision occurs, but this study did not subdivide those revision behaviors according to the practical impacts of behaviors on translation quality.

In general, studies of translation revision at home and abroad reflect that domestic translation studies do not pay enough attention to revision even though revision plays an increasingly pivotal role in translation. It might affect the development of revision competence among translator trainees. Therefore, the present study hopes to serve as an evidence for the asymmetric phenomenon.

2.1.2 Intervention behaviors

Revision in essence is an intervention to the draft translation made by revisers, which could be categorized into different groups. “Intervention behaviors” is a term adapted from “revision interventions” for avoiding puzzling readers.

The phrase “unnecessary changes” makes its first appearance in the early 1980s, when Arthern (1983) adopts the phrase to evaluate the performance of revisers in his research. Unnecessary changes are an indicator not only used to quantify the quality of revision or of the revisers’ performance, but also used to reflect the principle of revision. The principle of revision, as noted by Robert *et al.* (2017b) and Mossop (2020), is that revision is neither a matter of retranslating nor one of rewriting. Therefore, unnecessary changes in revision need to be avoided. Arthern’s typology was subsequently used by Künzli (2007), Robert (2013) and Robert & Van Waes (2014) to measure the quality of revision in their studies.

In order to describe a revision product appropriately, Rensburg (2017) categorizes revision interventions, which we called intervention behaviors, into four types including necessary changes, unnecessary changes and error overlooked and error introduced. Necessary change means that a reviser has corrected an error appropriately. Unnecessary change is used to describe the action in which a reviser has made a change that could not be justified. An error is overlooked if the reviser has not corrected it in the draft translation. An error introduced refers to the action in which the reviser has introduced a new error that was not present in the draft

translation. It is indeed clearer to describe revisers' behaviors from these four types, but what should be noted is that this typology doesn't include behaviors such as detecting but failing to correct an error.

Robert *et al.* (2017b) proposes an intervention typology including necessary changes, missed necessary changes, underrevisions, overrevisions, hyperrevisions and improvements. Necessary changes refer to rectify the translation errors in the translation. Missed necessary changes are used to describe the action that translators fail to detect the translation errors in the original translation. Underrevisions mean that translators detect an error but fail to correct it. Overrevisions refer to introducing a new error in the draft translation. Hyperrevisions are unnecessary changes made by translators, which have no effect to the final translation quality. Improvements means that translators change some segments without errors but the final translation quality has been enhanced. Compared with previous intervention typologies, this one is more systematic. Though research on intervention typology have continuously developed, researchers remain inclined to concentrated on one or two specific types of intervention (see Nitzke & Gros 2020; Daems & Macken 2020).

In addition to the description of revisers' intervention behaviors, it is also necessary to evaluate their work from a quantitative perspective. It enables revisers to get realize what is insufficient in their revision and then improve the quality of revision, which we called intervention quality.

When quantifying the quality of revision product, Arthern (1983) includes major errors, minor errors and unnecessary changes in his formula and assigned different weights to each error and every change. The quality was measured by adding major errors to one-half of the minor errors and one third of unnecessary changes. It meant that a product of revision with few errors and unnecessary changes would obtain a higher quality. Though Arthern (1991) simplifies his formula by not considering the number of unnecessary changes and by eliminating the relative weighting, his formula in 1983 still has a far-reaching impact on subsequent research. On the one hand, any wasting of time will affect the cost-effectiveness in the revision process, and if feedback needs to be given to the translator, unnecessary changes could negatively influence the relationship between the reviser and the translator. Furthermore,

different weights in the formula were of importance since the consequences of a major translation error, for instance, do have a more serious impact on the final translation quality than that of a minor error or an unnecessary change.

Künzli (2007) compares the number of necessary changes performed with the total of unnecessary changes, new errors introduced and errors overlooked. The study adds the concepts of necessary changes and errors overlooked but failed to take into consideration the impacts of different errors on the final translation quality.

Robert (2012) takes the quality of draft translation before revision into account and proposed two formulas to evaluate the quality of a revision product. In the first formula, the quality is measured by dividing the number of necessary changes by the total number of errors in the draft translation. In the second formula, the quality is measured by firstly subtracting the number of new errors introduced from the number of necessary changes and then dividing the total by the total number of errors in the draft translation.

Rensburg (2017) develops a new formula to evaluate revision quality on the basis of work done by Robert (2012). Rensburg adds the number of errors overlooked and unnecessary changes into the formula and weighted different errors and changes based on their effects. Besides that, the total number of errors in the draft translation is replaced by the number of words in the text. Moreover, this number is multiplied by one thousand for readability purposes.

Terry *et al.* (2017) insert 25 intentional errors in the experimental texts and only took necessary changes into account when calculating the intervention quality. Each necessary change is assigned a value of one point. However, other types of interventions like overrevisions (introduce of an error) were discussed briefly. The method is quite convenient but failed to record the impact of other behaviors on the final translation quality. It is certain that other interventions like overrevisions would increase the cognitive effort and even deteriorate the quality of translation.

Daems & Macken (2020) also have been inspired by the second formula suggested by Robert (2012). They adapt Robert's formula by weighting errors and changes according to their corresponding severity. A severity weight ranging from 0 to 3 has been introduced to the formula.

2.2 Machine Translation Post-editing

Machine translation, as defined by Arnold *et al.* (1994:1), is “an attempt to automate all, or part of the process of translating from one human language to another”. According to its definition, automation is an obvious characteristic that distinguishes MT from other task types such as translation from scratch, translation memory and computer-aided translation. Research on machine translation (MT) starts from 1950s (Somers 2003), which might be inspired by Warren Weaver³, the director of natural science. Since then MT system has evolved in past decades.

2.2.1 Post-editing of neural machine translation

Neural machine translation (NMT) has emerged as a new paradigm in MT since 2014. Compared to statistical machine translation (SMT), NMT has been shown to considerably improve the translation quality achieved, regardless of the language pair (see Toral Ruiz & Sánchez-Cartagena 2017). SMT, such as Google Translate before 2017, is essentially corpus-based. It relies on complex algorithms which learn how to transfer strings from one language to another using the probabilities derived from bilingual parallel corpus. Unlike SMT, NMT learns to translate sentences by using large artificial neural networks, and trains itself by deep learning and representation learning. Therefore, NMT has become the mainstream MT technology. Google and DeepL translate are popular online NMT today.

The first study focusing on post-editing can be dated to 1960s (García 2012). Orr & Small (1967) conduct a study to compare reading comprehension by subjects who read a raw machine translation, a post-edited machine translation, or a manual translation of scientific texts. In accordance with the required translation quality, we could categorize post-editing activities into two different degrees: light post-editing and full post-editing. In light post-editing, post-editors would make the minimum number of changes and essential accuracy corrections only, while in full post-editing, post-editors are required to produce publishable and similar-to-human translation (Koglin & Cunha 2019).

In recent years, empirical studies of NMT post-editing have yielded a great progress with the popular use of NMT. Research has been inclined to compare NMT post-editing with other task types, like human translation (see Guerbero Arenas 2017; Koponen *et al.* 2019; Yamada 2019) and translation with the aid of a translation memory (see Sánchez-Gijón *et al.* 2019). Moreover, usability of NMT post-editing from ergonomics has drawn a lot of interest. Screen (2019) and Wang *et al.* (2021a) have explored usability of NMT post-editing respectively. The topics of research on NMT post-editing are similar to that of MTPE, including translation quality, translation directionality, task difficulty, different language pairs, post-editing performance and the like. For instance, Daems & Machen (2019) and Jia *et al.* (2019a, 2019b) investigate the effect of NMT on translators' post-editing performance in her studies. Jia & Zheng (2022) explore the interaction effect between source text complexity and machine translation quality on the task difficulty of NMT post-editing from English to Chinese. Researchers also show a great interest on post-editing competence and some of them have conducted pilot study to reveal the competence by evaluating translators' behaviors in NMT post-editing (see Nitzke & Gros 2020).

In domestic translation studies, there is no literature review with NMT post-editing as its core since NMT model is a new paradigm emerging in 2014. However, most researchers have mentioned NMT or NMT post-editing in their studies on MTPE (see Yang & Wang 2019; Cui 2021) and post-editing competence and course model (Zhong & Shu 2020). Among domestic empirical studies of NMT post-editing, Li (2021) explores the impact of source texts on translators in time and quality. Wang *et al.* (2021b) compare the differences between cognitive effort and translation quality in NMT post-editing and human translation in order to explore the potential of NMT post-editing in metaphor translation. Yang & Fan (2021) investigate the semantic, pragmatic and syntax features of NMT so as to improve the translation quality of NMT post-editing.

It suggests that previous studies have neglected to examine and analyze post-editing behaviors. Research on post-editing behaviors might shed some light on relevant competence and pedagogical research. Therefore, the present study aims to answer the question left by previous studies.

2.2.2 Post-editing effort

In the translation industry, there might be an assumption that MTPE requires less effort than translation from scratch. For this reason, post-editing effort has received much interest. The most influential model of post-editing effort is proposed by Krings (2001), including temporal effort, technical effort and cognitive effort. Temporal effort is the amount of time needed for post-editing. Technical effort refers to the changes made by post-editors when correcting errors. Cognitive effort is one of pivotal indicators to reflect how post-editors detect errors and make corrections in MTPE. When attempting to measure the effort involved in post-editing, scholars have used different approaches.

Temporal effort could be considered the most visible part of post-editing effort (Krings 2001). It is easy to capture this indicator with keylogging and eye tracking tools (see Aziz *et al.* 2012; Elming *et al.* 2014).

Though compared with temporal effort, technical effort is more difficult to measure, some computerized metrics could be used to reflect the technical effort to some extent. Snover *et al.* (2006) and Daems *et al.* (2017) adopt Human-targeted Translation Edit Rate (HTER) to compare how many words have changes between the initial MT and the post-edited MT. However, Elming *et al.* (2014) point out that keylogging is a better measure to track the actual technical effort. Indeed, scholars tend to use keylogging instruments such as Translog II (Carl 2012) to capture this indicator (see Feng 2017; Jia *et al.* 2019a, 2019b).

Cognitive effort is the most difficult of the three aspects to capture and measure, although considerable research interest has been focused on this topic. Some research investigates the use of think-aloud protocols (O'Brien 2005; Vieira 2017; Koglin & Cunha 2019). One of the approaches to identifying cognitive effort is based on detecting pauses in keylogging data. It has been suggested that extended pauses are points of increased effort (O'Brien 2005; Screen 2017), as are clusters of short pauses (Lacruz & Shreve 2014). Eye tracking technology has also been applied to investigating cognitive effort, on the assumption that more frequent and longer gaze fixations indicate increased cognitive effort (Moorkens 2018). Vieira (2017) investigates and compares different indicators of cognitive effort using eye tracking,

think-aloud protocols and subjective ratings. Herbig *et al.* (2019) conducts an experiment with physiological measures such as skin- and heart-based indicators.

Beyond that, another possible way of investigating post-editing effort is subjective assessments. In other words, we could require participants to assess how much effort they perceive during the tasks. Subjective assessments have been very frequently used to evaluate mental workload mainly because of their high face validity and easy implementation (Vidulich 1988). And the most used subjective measure is the rating scale. Specia *et al.* (2010) propose a four-point scale for professional translators to rate sentences of MT. Callison-Burch *et al.* (2012) use a five-point scale to record how of the the sentence needed to be edited. Sun & Shreve (2014) use NASA Task Load Index, a multidimensional scale for measuring mental workload, to assess the level of translation difficulty. These scales are designed to indicate the post-editing effort perceived by participants, which can be used as a supplementation for objective data like pause duration and fixation duration.

2.2.3 Translation quality

The quality of translation is not only an indicator to assess a translation work, but also an important feedback to translators how they did in the task. Translation quality, as suggested by Van Slype (1979: 12), is not an absolute concept and thus should be assessed “relatively, applying several distinct criteria illuminating each special aspect of the quality of the translation”. Though the focus of his report was on the evaluation of machine translation, the definition also holds true for human translation. Since then, a large number of methods to assess translation quality have been proposed.

Translation quality assessment (TQA) is an essential part in translation studies. According to Moorkens *et al.* (2018), there are two main ways to assess the quality of translated texts: manually and automatically. The manual assessments could be done in several ways and the following three manual methods are relatively popular: Dynamic Quality Framework (DQF) (O’Brien *et al.* 2011), Localization Industry Standard Association Quality Assessment Model (LISA) (2011) and Multidimensional Quality Metric (MQM) (Lommel *et al.* 2014). These approaches evaluate the final quality of a translation and also allow for the integration of severity scores. As for

automatic assessments, there are also a variety of approaches, for instance, Bilingual Evaluation Understudy (BLEU) (Papineni *et al.* 2002), Metric for Evaluation of Translation with Explicit Ordering (METEOR) (Banerjee & Lavie 2005), and Translation Error Rate (TER) (Snover *et al.* 2006). One or more reference translation will be taken into consideration in these ways of measuring the translation quality. In order to apply into more scenarios and text types, these approaches, manual or automatic, are allowed to be adjusted in light of concrete conditions.

2.3 Comparative Studies of Translation Revision and Post-editing

Unlike comparisons of revision and translation or comparisons of MTPE and from scratch translation, comparisons of translation revision and Post-editing has been researched empirically by very few scholars. The relationship between revision and post-editing might be one of the main reasons why the relevant research is in its initial stage. Therefore, we attempt to figure out those different voices before reviewing the empirical comparative studies.

In the translation studies, some held a view that post-editing is a form of revision. Krings (2001) and Schwartz (2014) support this opinion for post-editing, like revision, can be performed by monolinguals without access to the source text. Another argument in favor of this view is that the translator spends most of the time pausing rather than tapping in the post-editing tasks (Koehn 2009; Ortiz-Martinez *et al.* 2016). However, this argument has also faced challenges from other translators and scholars. For instance, Jakobsen (2019) theoretically suggests that the application of technologies in professional translation makes the distinction between three activities increasingly blurred. It is worth noting that Jakobsen's conclusion is based on a precondition that revision and post-editing are different activities. do Carmo & Moorkens (2020) also notice that and investigated the relationship between editing, post-editing and revision. They discussed how the view of PE as a form of revision may have contributed to devaluing this process and affecting pedagogical approaches. Meanwhile, scholars have proposed different revision and post-editing competences

(see Robert *et al.* 2017a; Scocchera 2019; Mossop *et al.* 2020; Rico & Torrejón 2012; Pym 2013). Konttinen *et al.* (2020) generalize revision competence models and post-editing competences, presenting the differences between two competence models, and advocated a holistic and integrated approach to teaching a strategic subcompetence that applies to correcting both human and machine translations. All these suggest that the latter argument is more popular in the translation studies.

Few comparative studies of translation revision and post-editing have been conducted from the empirical perspective to date. Nitzke & Gros (2020) point out that translators usually generate their own translations in their minds before revision and post-editing by comparing their over-editing behaviors. “Over-editing” is synonymous with what we refer to as “unnecessary changes” and “hyperrevision” in Section 2.1.2. The experiment suggested that personal preferences, especially in terms of word choice and lexical style, should be suppressed in post-editing and translation revision competence training. Daems and Macken (2020) propose a method to quantify revision quality and revision optimization based on the product data. The results showed that participants performed better when they had the wrong assumption about the provenance of a text. Their results are in contrast with Guerberof Arenas (2008) and Teixeira (2014). After an in-depth analysis of inconsistent results, they questioned whether it is necessary to distinguish revision and post-editing. Apparently, this question is up for debate since they didn’t use other methods like keystroke logging and eye tracking to support their argument. Robert *et al.* (2022) exploit Inputlog to discuss four perspectives on how to measure translation revision competence (TRC) and post-editing competence (PEC) and called for more empirical research into TRC and PEC.

Though a raft of theoretical research supports that revision and post-editing are different activities, empirical evidence remains insufficient to intensify this argument. In terms of content, previous empirical studies mainly focus on a specific behavior in revision or post-editing, which might not fully reflect all the intervention behaviors of translators in two scenarios. Moreover, the product-oriented approaches without process data as supplementation might lead to an inaccurate result. Therefore, in order to narrow the gap between theoretical research and empirical research, the present

study will explore the differences between translation revision and MTPE in a more comprehensive way by integrating process data with product data. The results in the present study will also be interpreted by the information from the questionnaire and written retrospective protocol.

2.4 Research on Revision and Post-editing of Literary Texts

As noted by Mossop (2007), research on the revision of translation is relatively scarce and has mainly focused on nonliterary texts. Literary translation has rarely been the object of research on revision. It might be partly attributed to the market demand, but it is still of importance to translate literary texts. Unlike nonliterary translation, literary translation is a more difficult and creative activity. Translators are required to preserve the original meaning of the source text and its reading experiences for targeted readers (Toral Ruiz *et al.* 2018). It is a great challenge for translators. Therefore, revision becomes the linchpin before obtaining a quality literary translation. But previous research on revision of literary translation mainly focuses on general topics such as revision processes (see Jones 2006; Kolb 2013), loyalties of literary revisers (Feinauer & Lourens 2017) and interactions between actors involved in revision process (Marin-Lacarta & Vargas-Urpi 2019). These studies might benefit to improve the efficiency of the workflow, but a more detailed investigation is needed to reveal how translators revise literary translation and what might affect their performance in revision tasks, which might shed light on the reviser training.

Furthermore, researchers have made a lot of attempts to explore the possibility of post-editing of literary translations. Jones & Irvine (2013) apply MT into literary texts and attempt to address what makes literary translation particularly hard and the potential role of the machine in literary translation. Toral Ruiz *et al.* (2018) conduct an experiment and found that MT would increase the productivity of literary translators in literary translation. There is another experiment conducted by Toral Ruiz & Way (2018) and the results show that 17-34% of MT of novels could match that of professional literary translators. Creativity is a vital indicator that distinguishes HT and MT. Arenas & Toral Ruiz (2020) further compare the creativity and readers'

experiences between HT and MT and found that there was no significant difference in these two indicators. Those attempts serve as an evidence for the surprising development of MT. Indeed, MT enables translators to increase their productivity. And with the involvement of post-editing activities, it is possible to obtain a quality literary translation.

After the review, we found that there are a few empirical studies of revision and post-editing of literary translation. The exploration of translation quality of post-edited literary texts might put forward some implications for further study. Therefore, we would compare the quality of revised and post-edited literary translation.

2.5 Summary

This chapter has reviewed the existing literature with regard to translation revision, MTPE, comparative studies of translation revision and MTPE and research on revision and post-editing of literary translation. It provides a solid theoretical foundation as well as a helpful methodology instruction for the present study.

Firstly, there is a gap between theoretical and empirical research on differences between translation revision and NMT post-editing. For instance, Konttinen *et al.* (2020) generalized revision competence models and post-editing competences, presenting the commonalities and differences between two competence models. But it lacks empirical evidence to support this argument. The existing empirical studies of revision and post-editing mainly focus on one or two specific behaviors of translators or analyze two tasks based on product data, which might lead to an inadequate comparison. Therefore, it needs an empirical research to explore the differences between translation revision and NMT post-editing.

Secondly, with the rapid development of translation technology, NMT has emerged as a new paradigm in MT. It has been shown to considerably improve the translation quality. However, most findings in previous comparative studies are only applicable to nonliterary texts. Although some attempts have been made to investigate the efforts spent in MTPE and translation from scratch of literary texts, it is still insufficient to reveal the potential of NMT post-editing in literary translation,

especially compared with translation revision.

Therefore, we conduct an empirical experiment to investigate the differences between translation revision and NMT post-editing based on literary texts, which might shed light on relevant research and translator training in practice.

Chapter 3 The Empirical Study

After reviewing research on translation revision and post-editing, we found that there remains a gap between theoretical and empirical studies concerned these two task types, though some attempts have been made to address this issue. Moreover, it is also suggested to extend the text types from nonliterary one to literary one when comparing translation revision and NMT post-editing. In order to narrow the research gap, we hereby design a specific experiment to address the research questions in section 1.2.

3.1 Research Design

3.1.1 Participants

The present study aims to investigate the differences of translation trainees' behaviors in revision and post-editing of literary translations from English to Chinese. Thus, participants are required to satisfy the following criteria: 1) limited translation experience; 2) a basic knowledge of literary translation, revision and post-editing; and 3) native speakers of Chinese with English as their second language.

We distributed the questionnaire online and recruited participants for two tasks based on the criteria above. A total of 48 postgraduate students, specializing in translation studies at a Chinese university, volunteered to fill the questionnaire. It includes 45 females and 3 males with an average age of 22.83 years (range 21-30, $SD=1.585$). The data collected indicates that all of them have no professional translation experience in translation companies. They have acquired the basic knowledge about translation revision and MTPE in "Computer-Aided Translation" course. They also have attended a course of literary translation. The survey also shows that they all are native speakers of Chinese with English as their second language. They have learned English for 12.42 years on average (range 5-17, $SD=3.121$) and all of them have passed the Test for English Majors-Band 8⁴ (TEM-8).

21 (P01-P21) among those respondents are randomly selected to participate in the following tasks which are designed to investigate the differences between translation revision and NMT post-editing. They were 19 females and 2 males with an average age of 22.38 years (range 21-23, $SD=0.898$). Though the proportion between male and female in the present experiment is remarkable different, it could reflect and comply to the reality. There are 51 females and 4 males studying translation and interpreting in this university this year. And all participants are translation trainees without professional translation experience. We have collected their scores in TEM-8 for having a better understanding of their language proficiency. The data shows that they all passed TEM-8. Specifically, 12 (54.14%) of them have got a score between 60 and 69, seven (33.33%) of them have gained a grade between 70 and 79, and two (9.52%) have achieved a grade above 79.

3.1.2 Materials

3.1.2.1 Text selection

The experimental texts are two excerpts from a real literary translation task of *Walden* (Thoreau 1999) from English to Chinese with similar text complexity and readability. Participants in the experiment are allowed to use online and offline resources but not allowed to refer to the existing translation version. In order to ensure the validity of outcomes, we have recorded the behaviors of all participants by EVcapture, a screen-recording software, throughout the experiment. This experimental source texts are composed of approximately 188 and 192 English words respectively, for the consideration of the display of the text in Translog-II window without scrolling. In addition to the text length, we also take other criteria such as number of sentences, text readability and complexity into consideration for the comparability of the two texts (see Table 3.1). The results from ReadabilityFormulas.com⁵ and Lexile Analyzer⁶ indicate that ST1 and ST2 share a similar text readability and complexity.

Table 3.1 Basic information of ST1 and ST2

Indicators	ST1	ST2
Length (words)	188	192
Number of Sentences	8	8
Flesch-Kincaid Grade Level	9.8	10
Flesch-Kincaid Reading Ease	66.9 standard / average	66.2 standard / average
Gunning Fog Score	12.8 hard to read	13 hard to read
SMOG Index	8.5	8.5
Coleman Liau Index	7	8
Automated Readability Index	10.5	11.5
Linsear Write Formula	13.9 college	14.2 college
Lexile Measures	1210L - 1400L	1210L - 1400L

Jensen (2009) points out that readability tests provide indication for assessing the amount of production effort and comprehension effort needed in the translation. ReadabilityFormulas.com is a free website to score texts including documents, books, policies, technical materials and to find the reading level and grade level that readers need to read and comprehend the text. Its free readability formulas include Flesch Reading Ease formula, Flesch-Kincaid Grade Level, Fog Scale (Gunning FOG Formula), SMOG Index, Coleman-Liau Index, Automated Readability Index and Linsear Write Formula. It might be not friendly for readers to catch up with the real meaning of those figures, so we have added the relevant explanations behind each item. According to the results from ReadabilityFormulas.com, ST1 and ST2 got similar scores for all seven indicators.

Text complexity is a critical consideration in matching translation trainees to appropriately challenging materials. Simply stated, text complexity is how easy or hard a text is to read, based on quantitative and qualitative text features. So, we adopted Lexile Analyzer® to measure the complexity of the text by breaking down the entire piece and studying its characteristics, such as sentence length and word frequency, which represent the syntactic and semantic challenges that the text presents to a reader. The outcome is the text complexity, expressed as a Lexile® measure,

along with information on the word count, mean sentence length and mean log frequency. The range of text complexity in Lexile Analyzer is from 200L to 1600L. ST1 and ST2 both scored 1210L-1400L, meaning that they were both complex text to some extent.

Based on the above indicators, the two texts are similar in readability and complexity, which can be expected to bring similar objective task difficulty to translation trainees. The experimental HT of ST1 and ST2 were draft translations of translation trainees in the real project with 329 and 338 words respectively (see Appendix C). With the aim of mimicking a real scenario where translation trainees resort to the tools they know, we chose to use the online NMT systems Google Translate, DeepL, Baidu Translate and Youdao Translate to translate the two source texts due to their free accessibility and reliability. The quality of NMT was evaluated by two professional translators with more than 5 years of experience in translation. The two source texts with better translation quality were chosen for the experiment. The better translation quality refers to the NMT with lower number of errors and lower number of critical errors. Therefore, DeepL was chosen for the final NMT version, as its quality was slightly better than that of other three NNMT systems for these texts. The MT versions were 326 and 321 words long respectively (see Appendix C). After comparisons between HT and NMT of ST1 and ST2 in the section 3.1.2.2, we decided to adopt NMT of ST1 and HT of ST2 for revision and post-editing tasks respectively.

Translation briefs (see Appendix D) which introduced the target audience and the quality expectation also were provided for participants. Moreover, translation revision and TAUS MT Post-Editing Guidelines (2016) (see Appendix E and F) have been provided for instruction in the revision and post-editing task respectively.

3.1.2.2 Translation Errors in HT and NMT of Literary Texts

This section will introduce the error taxonomy used for annotation tasks and the annotation procedure. The HT and NMT output of two texts were annotated by the three professional translators with more than five years of translation experience respectively. All of them have experience in error annotations in this language pair.

(1) Error taxonomy

Error annotation in the present study is based on an error taxonomy derived from the TAUS Dynamic Quality Framework subset of MQM⁷ (Lommel *et al.* 2014). MQM is a flexible hierarchical translation error taxonomy that can be tailored for different languages, texts and purposes.

There is a total of eight error categories in MQM. We adopt accuracy, fluency and terminology error categories to annotate errors arising in our experimental texts since we would investigate the differences in content. The other four categories are related to format, design and presentation of translation, which doesn't confirm to our research purpose. In the error taxonomy, accuracy errors are occurring when the target text does not accurately correspond to the propositional content of the source text, which includes mistranslation, over-translation, under-translation, addition and omission, do not translate (DNT) and untranslated. Fluency errors are related to the linguistic well-formedness of the text, including problems with grammar, punctuation, spelling, unintelligible and character encoding. And terminology errors are arising if a term does not conform to normative domain or organizational terminology standards or when a term in the target text is not the correct, normative equivalent of the corresponding term in the source text. It includes inconsistent use of terminology resource, inconsistent use of terminology and wrong term. The MQM taxonomy suggests how these error categories can be divided into more detailed and specific items and allows the use of only a subset of the entire error taxonomy.

Based on the previous research (Carl & Baez 2019), those errors could also be marked as “critical” and “minor”. “Critical” would be appropriate for errors with a severe impact on understanding or fluent reading. For instance, errors that cause a shift in meaning between the source text and the translation or that affect the intelligibility of the translation belong to this category. “Minor” errors have a slight impact on understanding or fluent reading. For example, a sentence with a minor error can still be understood without effort and the information contained in the translation is equal to that of the source text. For the convenience of quantifying participants' intervention behaviors, we assigned critical errors in the present study a value of “2

points” and minor errors a value of “1 point”.

However, Carl & Baez (2019) hold a view that only four error types in their taxonomy could be marked as “critical” and they did not take terminology errors into consideration as well. We updated the error taxonomy in accordance with the latest DQF-MQM and this experimental task. Annotators in the present study could mark errors as “critical” if they felt it is necessary and reasonable. The adapted error taxonomy in Table 3.2 was discussed and developed by annotators. And then they were assigned to annotate the HT and NMT of ST1 and S2.

Table 3.2 Error types

Accuracy	Fluency	Terminology
Mistranslation	Grammar	Wrong term
Addition	Punctuation	
Omission	Spelling	
	Unintelligible	

(2) Error annotation

The error annotation schema in our experiment was implemented in Word office not YAWAT for two reasons. YAWAT (German 2008) is a tool used to manually align translation equivalences. On the one hand, it was complicated to use YAWAT in this study for we invited three annotators to mark translation errors arising in the HT and NMT. We were intended to compare annotations of several annotators in order to reduce individual preference. But YAWAT could only save the latest one alignment and annotation. On the other hand, it was also not accessible to use YAWAT without the foreign official permission. Therefore, we decided to use Word office for a better comparison.

We aligned the ST1 and ST2 with their respective HT and NMT in sentence level. Annotators were asked to fragment the translations into alignment groups, and then assign an error category to that alignment groups. In case words in the ST or in the TT could not be aligned, they could be marked as “Omission” or “Addition” if the

content from the ST was missing or added in the TT. The annotation schema allows for 16 annotation categories: 8 critical errors and 8 minor errors. Annotators had time to be familiar with the MQM and a discussion on how to mark errors was conducted before the formal error annotation.

Overall, three annotators made 157 times of annotation. Figure 3.1 showed the distribution of error annotations across HT and NMT of ST1 and ST2 respectively.

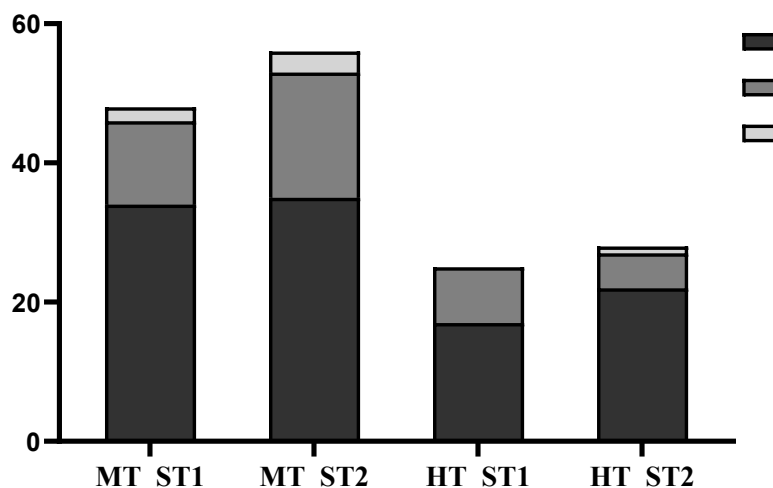


Figure 3.1 Number of error annotations in HT and NMT of ST1 and ST2

From the perspective of text origin, 104 total annotations were produced by three annotators for two NMT outputs, 48 for ST1 and 56 for ST2. In reverse, a relatively low number of annotations were produced by annotators in HT with 25 for ST1 and 28 for ST2. To be more specific, the distribution into accuracy, fluency and terminology errors was slightly different across both data sets. A total of 66% of the NMT error annotations (N=69) were labeled as accuracy errors, while this was the case for 74% of the HT error annotations (N=39). 29% of the NMT error annotations (N=30) were labeled as fluency errors, while this was the case for only 21% of the HT error annotations (N=11). The terminology errors accounted for 5% of the total NMT error annotations (N=5) and 5% of the HT error annotations (N=3) respectively.

According to Figure 3.2 below, there were slightly more critical error annotations than minor ones for both NMT and HT outputs. Specifically, a total of 67.3% of NMT error annotations (N=70) were labeled as critical while the figure was about 56.6% (N=30) in HT. 32.7% (N=34) of NMT error annotations was marked as minor, while

the minor error annotations accounted for 43.4% (N=23) in HT. Furthermore, we calculated and presented the error score per text in Figure 3.2 base on its severity weight. The error scores were significantly different between HTs and NMTs. The highest score was given to NMT of ST2 and the lowest score went to HT of ST1. We took number of error types and error scores into consideration when choosing experimental texts.

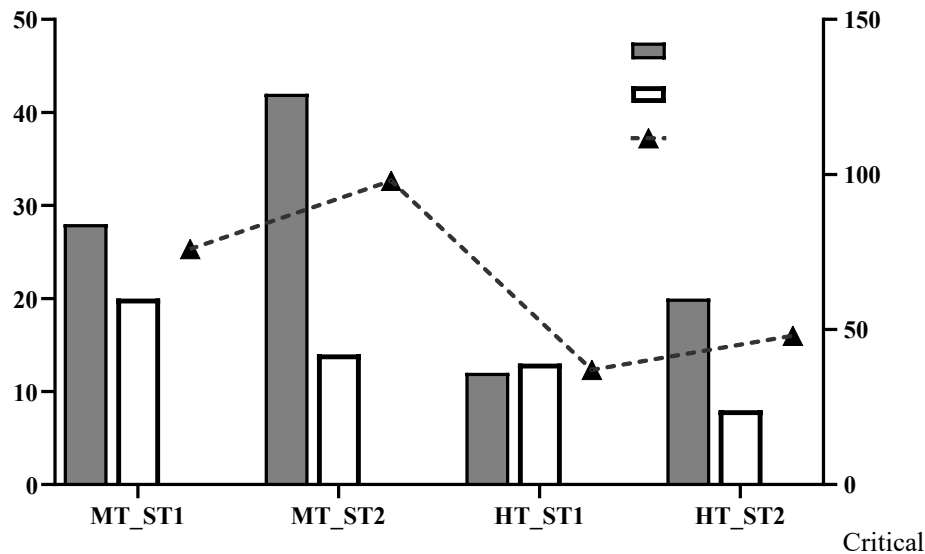


Figure 3.2 Error annotation and error score per text in HT and NMT

A more detailed view of HT and NMT was presented in Figure 3.3. We used “Termi” to replace “wrong term” in the Figure below for a clear and comfortable view. The choice of the error annotations was quite different in two texts.

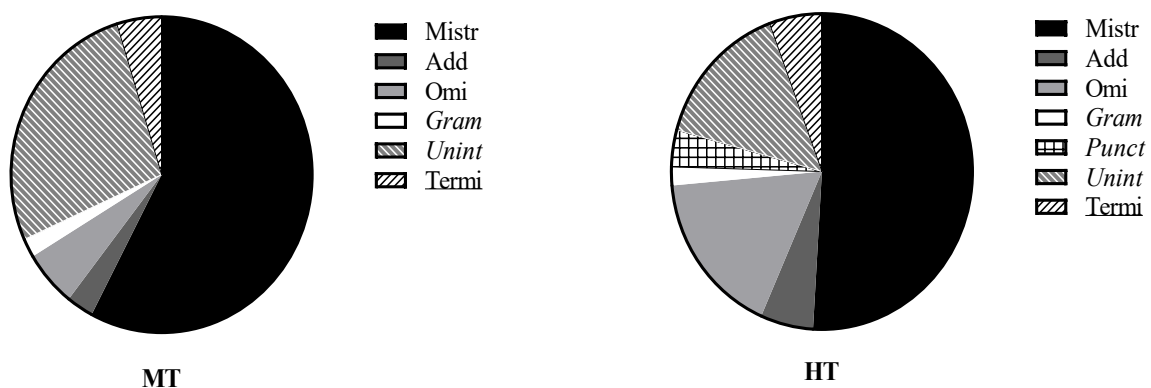


Figure 3.3 Number of error annotations in specific types in NMT and HT

Mistranslation in accuracy errors accounted for the largest proportion in both NMT and HT. There were relatively more mistranslations (mistr) annotated by three translators in the NMT outputs (N=60) than that in the HT output (N=27). It indicated that understanding the original meaning was critical for literary translation. NMT relies on algorithms and translation corpus to translates texts, which might be seriously affected by the expression way of ST. HT is more flexible in this term for translators could exert their creativity. However, the translation quality of HT always rests with individual translation proficiency. In the present study, translation trainees who translated ST1 and ST2 initially performed better than NMT in terms of understanding the original meaning. That is the reason why the number of annotated mistranslation errors was quite different between HT and NMT. Error with the second largest number in NMT was unintelligible error in fluency category (N=30). 18 unintelligible errors were annotated as minor and 12 critical in the NMT output. In contrast, only 8 errors in HT were labeled as unintelligible. It noted that the expression was more fluent and native in HT. Translation trainees in the present study tended to present their understanding in a reader-friendly way. Expressing the idea in a native way was a great challenge for NMT when it processed complex texts such as literary texts in the present study. The number of omission errors in two translation tasks was slightly different from other error types since there were more omission errors in the HT output than that in the NMT output. Different from HT, NMT would process all contents in the text so that omission errors were relatively rare in NMT. Omission in accuracy was the second largest error in HT with 3 minor errors and 6 critical errors. Translation trainees would adopt this method to deal with the segment that they didn't understand, which was a reasonable and intelligent approach in some conditions. Remarkably, NMT in the present study also omitted an important segment when translating ST1. It failed to translate “rapt in a reverry” in the second sentence with a total of 76 words into Chinese. It might be a fault of NMT and we need to further improve the function of MT to deal with long and complicated sentences. It was also noted that punctuation errors arose only in the HT in the present study. No spelling error was in the NMT and HT. The reason might be that Chinese language is an alphabetic language, so errors like spelling are rare in Chinese.

(3) Inter annotator agreement

To determine the validity of the approach, we also computed inter annotator agreement using the `kappam.fleiss` function in R which is provided with the `Orr`⁸ package.

$$K = \frac{\text{Pr(a)} - \text{Pr(e)}}{1 - \text{Pr(e)}}$$

Equation 3.1 Kappa fleiss function

The `kappam.fleiss` implementation in R allows the comparison of several annotators and the assessment of their agreement above chance. The Kappa score is shown in Equation 3.1. According to the research of Carl and Baez (2019), the factor Pr(e) amounts to the agreement by chance. The factor Pr(a) refers to the observed agreement which can range between 0 and 1 theoretically. “ $\text{Pr(a)} - \text{Pr(e)}$ ” is then the degree of agreement achieved above chance and “ $1 - \text{Pr(e)}$ ” the degree of agreement that can be maximally achieved above chance. Landis and Koch (1977) labeled kappa scores as “poor” (<0), “slight” (0.0 - 0.2), “fair” (0.2-0.4), “moderate” (0.4-0.6), “substantial” (0.6-0.8) and “perfect” (0.8-1).

After the analysis, only 27 of the 54 accuracy error annotations were highlighted by three annotators. This led to an agreement of 50% with $\kappa=0.714$. For fluency errors, only 11 of the 19 cases were highlighted by three annotators, equal to an agreement of 57.9% with $\kappa=0.734$. As for four terminology errors, only 2 among them were marked by three annotators. It was equal to an agreement of 50.0% with $\kappa=0.557$.

According to this classification, accuracy and fluency error categories in this study showed a “substantial” agreement. The terminology category obtained a “moderate” agreement. According to the kappa value of accuracy, fluency and terminology errors, a few things must be taken into account. First, certain errors were highlighted by only two annotators simply because the third annotator hadn’t observed the error, not because the third annotator did not agree with the judgement. Secondly, in some cases, annotators might detect the errors but group them differently. Therefore, we consulted with annotators and obtained the final errors arising in four

texts. Based on the analysis and discussion among the annotators, we chose HT of ST2 for translation revision task and NMT of ST1 for post-editing task due to their similar number of errors and error score. NMT of ST1 involves 18 errors with 11 accuracy errors, 6 fluency errors and one terminology error. HT of ST2 includes 14 errors with 12 accuracy errors, one fluency error and 1 terminology error. And NMT of ST1 had an error score of 28 and HT of ST2 had a score of 24.

The proportion of each error type could be found in Figure 3.4. Error types in bold stand for critical, the rest are minor. For a better distinction of three error categories in the Figure, accuracy errors were presented in the normal type, fluency errors in italic type, and terminology errors were highlighted by an underline. The distribution of error could represent the characteristic of HT and NMT.

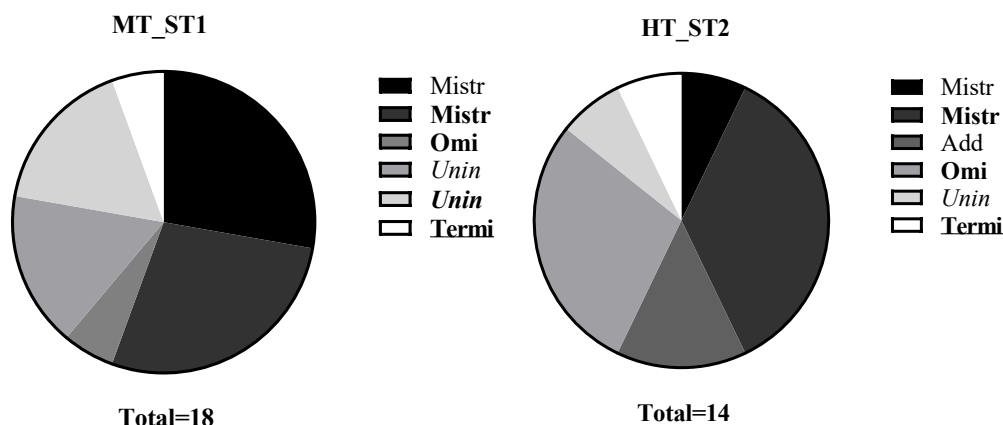


Figure 3.4 Proportion of error types in NMT of ST1 and HT of ST2

This method of errors design differed from previous studies since we kept the ecological validity of the experimental texts by using HT from authentic projects instead of inserting some translation errors into HT. It could reflect the real performance of translation trainees when revising HT and post-editing NMT. Therefore, it was reasonable that the distribution of error types and draft translation quality were slightly different after the variable control.

3.1.3 Instruments

The questionnaire (see Appendix B) aims to collect the basic information of translation trainees and survey their attitudes towards translation revision and NMT post-editing. It contained three parts including Part I: Demographic information,

which collected information such as gender, age, translation experience and English proficiency, Part II: Experience of translation revision and MTPE (6 items) and Part III: Attitudes towards translation revision and MTPE (11 items). We adopted a 5-point Likert scale, in which “1” meant “strongly disagree” and “5” meant “strongly agree”.

Translog-II (Carl 2012) is a program which can record human writing and translating process on a computer without interfering it. It is an instrument specially designed for translation process research to collect objective data of translation process. It consists of two main components, i.e. Translog-II Supervisor and Translog-II User. The former is designed for researchers to create project files and replay recorded sessions, and the latter is for users to complete their tasks such as writing and translating a text. Translog-II can record keystroke activities including insertion, deletion, navigation, cut/copy-and-paste, return key and mouse operation. Translog-II produces a log file in XML format for every session which records writing and translating activities.

In this study, Translog-II is used to collect participants’ translation process data such as keystrokes and pauses. The projects were created first. The participants were shown the Graphical User Interface (GUI) (Figure 3.5). In accordance with the computer interface, the source texts were displayed on the left part of the screen with the typeface Times New Roman and double-spaced at 14-point size. Participants could click the “Stop logging” button and save xml. files after finishing the tasks.

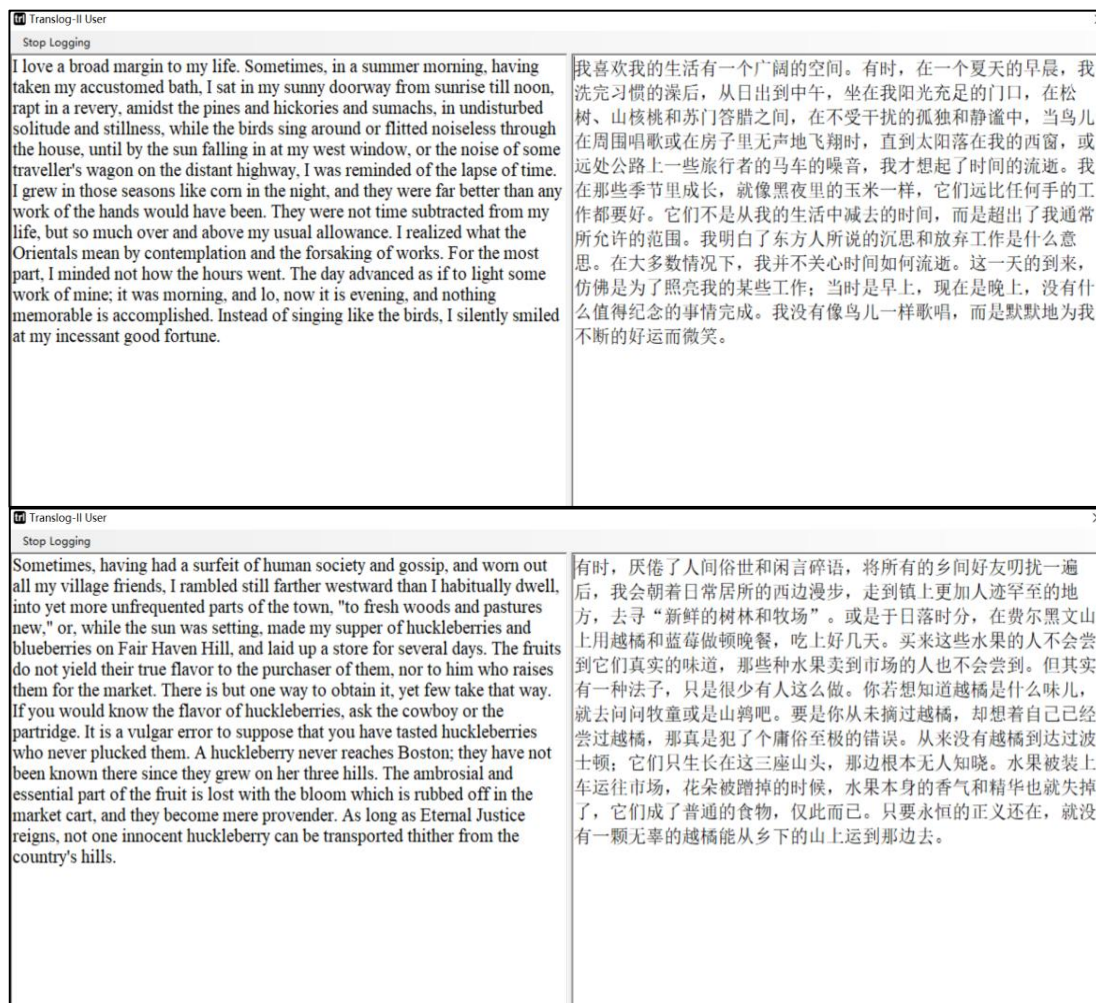


Figure 3.5 Screenshot of User interface of Translog-II

(the upper user interface is for post-editing task, the lower for revision task)

EVCapture⁹ is a free screen recording tool developed by a Chinese company. It not only enables the researchers to have a careful observation of translation trainees' behaviors in two scenarios, but also helps participants recall what they have done in two experiments when writing the retrospective papers. The interface of the software ware shown in Figure 3.6. In the formal study, participants were required to turn on EVCapture software before the tasks and save relevant files after two tasks. Therefore, we received a total of 21 screen recording files in this study. Moreover, EVCapture is easy to operate and will not disturb participants during the tasks.



Figure 3.6 Screenshot of EVCapture

3.1.4 Experimental site

The questionnaire was distributed on the “Computer-Aided Translation” course. Respondents could answer the questionnaire on their smartphones or laptops conveniently. They could take their time to answer those questions.

The translation revision and NMT post-editing tasks were completed in the office rooms in the library. Those rooms were soundproof to prevent participants from being disturbed. Moreover, the quiet settings in the rooms also provided the participant with a relaxing and comfortable environment and made them fully engrossed in the text.

3.1.5 Pilot study

A pilot study was conducted to ensure that the questionnaire could collect the needed information and Translog II and EVCapture programs could operate well. Translation trainees who participated in the pilot study were excluded from the formal experiment. In the pilot study, all the participants needed to complete tasks the same as in the formal one. Factors that might affect the experiment were modified in later formal experiment. For instance, we added one item to collect the contact information of respondents as well as explanation of translation revision and MTPE in the questionnaire. Furthermore, we used experiment ID rather than participants' real names in the formal experiments since some participants felt stressed if we connected

the poor performance to their real name. It was also suggested to slightly adjust the layouts for translation revision and post-editing tasks.

3.2 Data Collection

3.2.1 Experiment procedures

The experiment was conducted in October and November, 2021. All participants signed an Informed Consent Form (see Appendix G) before the experiment. They were informed that their anonymity and confidentiality would be ensured. They were given a ¥40 reward for their work. The experiment was approved by the Ethics Committee of the College of Foreign Languages at Hunan University.

The questionnaire were distributed to respondents through Kanchenjunga¹⁰ online and they were given adequate time to complete the whole questionnaire. Data were collected without missing cases.

In the revision and post-editing tasks, 21 participants were tested individually in the study learning office in the school library. They needed to read the experimental procedures (see Appendix H). According to translation briefs, translation revision and post-editing guidelines, they were asked to produce publishable translations for general readers in China and to use as much original translation as possible.

Participants downloaded the Translog II program on their laptops and then completed a warming-up task to get used to the translation interface. They would revise HT of ST2 and post-edit NMT output of ST1. The two tasks were conducted in a random order to avoid the sequence effect. They had access to Internet or other resources in the experiment for having a real translation environment. 30 minutes were suggested for each task. Right after participants completed every task, they were asked to subjectively evaluate the task difficulty based on the adapted NASA Task Load Index (Sun and Shreveport 2014) (see Appendix I). Participants were given time to relax themselves between two tasks. When participants finished two tasks, their translation process data recorded by Translog II and EVCapture would be replayed. A screenshot of replaying the post-editing session was presented in Figure 3.7.



Figure 3.7 Screenshot of Supervisor interface of Translog-II

At the end of the experiment, participants will be asked to write a retrospective report about their feeling and differences between translation revision and MTPE. The qualitative data will serve as an evidence to interpret the analysis of quantitative data. And the experiment procedures could be found in Figure 3.8.

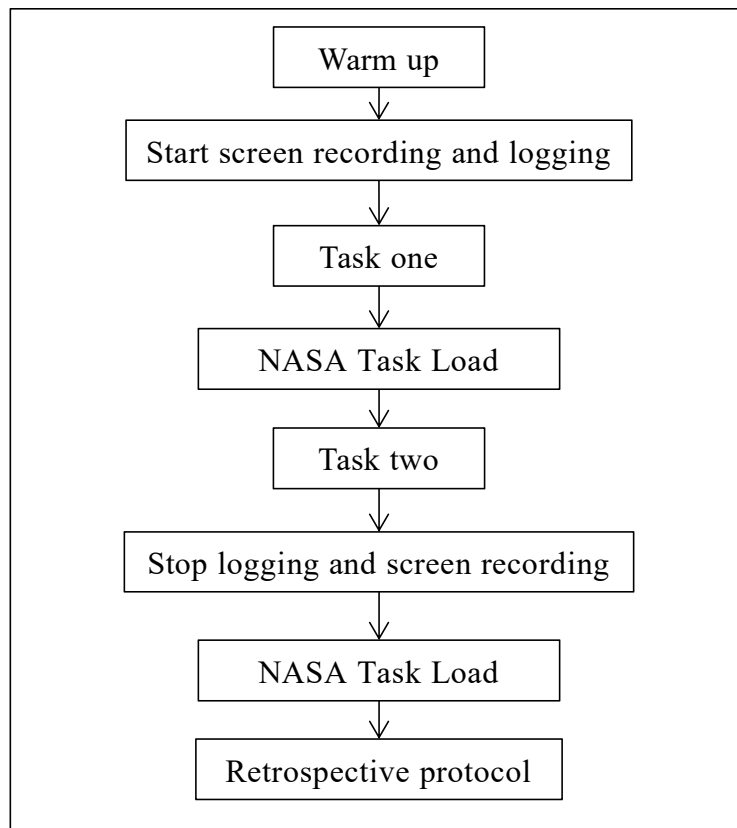


Figure 3.8 Experiment procedures

3.2.2 Data resources

The present study drew on the following six data resources during tasks:

- (1) participants' final translations after translation revision and NMT post-editing;
- (2) xml* files saved from Translog-II which contained process and product data;
- (3) mp4* files saved from EVCapture;
- (4) participants' subjective assessment of task difficulty and cognitive effort;
- (5) respondents' attitudes towards translation revision and MTPE;
- (6) participants' written retrospective report about their feelings of differences between translation revision and MTPE.

These data resources include quantitative data that could be further statistically analyzed and qualitative data that could help explain results. Data from these resources were examined in the next part.

3.3 Data Exclusion

To ensure reliability and validity of data analysis, we examined data quality and excluded those didn't meet the standards.

We collected a total of 49 questionnaires. No one returned unfinished questionnaire. One respondent filled the questionnaire twice, therefore, all 48 valid questionnaires were included in the further analysis.

In the following translation revision and post-editing tasks, 42 logged files from 21 participants were correctly saved. However, we found the time spent in one logged file (ST2_P06) different from that in screen recording. After checking, we found that P06 failed to save her logged file in revision task but she conducted this task again in a short time. The data of post-editing effort in her second attempt was invalid, so we decided to analysis 40 logged files correctly saved from other 20 participants.

In total, 48 questionnaires, 40 final translations, 40 logged process data from 20 participants, 40 task difficulty subjective assessment data and 40 written retrospective reports were included for further analysis.

3.4 Data Analysis

Data analysis was carried out in Statistical Package for Social Science 26.0 (SPSS 26.0) and RStudio. We performed descriptive analysis, normality test, paired t-test and non-parametric test in SPSS 26.0 and used the `kappam.fleiss` function in R which is provided with the `Orr` package respectively. The significant level was set as $p < 0.05$. GraphPad Prism 8 was applied to make high-quality graphs so as to visualize the results.

Chapter 4 Results and Discussion

This chapter provides statistical analysis of the data from resources previously mentioned and detailed discussion of the results. The structure of this chapter follows our research questions. First of all, translation trainees' intervention behaviors were investigated. Secondly, we compared the post-editing effort spent by participants in two tasks from the perspectives of temporal, technical and cognitive effort. Thirdly, we presented the final translation quality of revised and post-edited texts for it was also a vital indicator to reflect translation trainees' behaviors in two scenarios. Descriptive statistics of those data were presented with the help of SPSS 26.0 and GraphPad Prism 9.3. Besides, possible explanations of the results were provided with reference to the questionnaires and retrospective reports. The results were mainly produced in table or figure format for a better visualization. With these analyses and discussion, we tried to address the research questions.

4.1 Intervention Behaviors in Revision and Post-editing of Literary Texts

4.1.1 Intervention behaviors

Revision intervention is a concept adapted from translation revision studies. It refers to the behaviors of translators when they detect and correct errors in the translation. There are various types of behaviors occurring the translation. We adopted the systematic intervention typology generalized by (Robert *et al.* 2017b). We also invited two translators with more than 5 years professional translation experience to assess the intervention behaviors arising in two tasks.

The result of 20 participant's intervention behaviors was shown in Figure 4.1. Participants made more changes in post-editing than in revision, which was consistent with the number of errors in NMT and HT. When post-editing NMT, participants made a total of 476 changes with an average of about 24 changes. Compared with NMT post-editing, a significantly fewer changes were made by participants in

translation revision with a total of 317 and an average of around 16 changes. It could be interpreted by the result of item 6 in our questionnaire, 65% of participants strongly agreed that they would made more changes in NMT post-editing than in translation revision (N=13). It indicates that translation trainees keep a cautious attitude towards the quality of NMT. The results were in line with Briggs (2018) and Wang *et al.* (2021a) who report that most students use MT tools but had limited trust in the accuracy of their output.

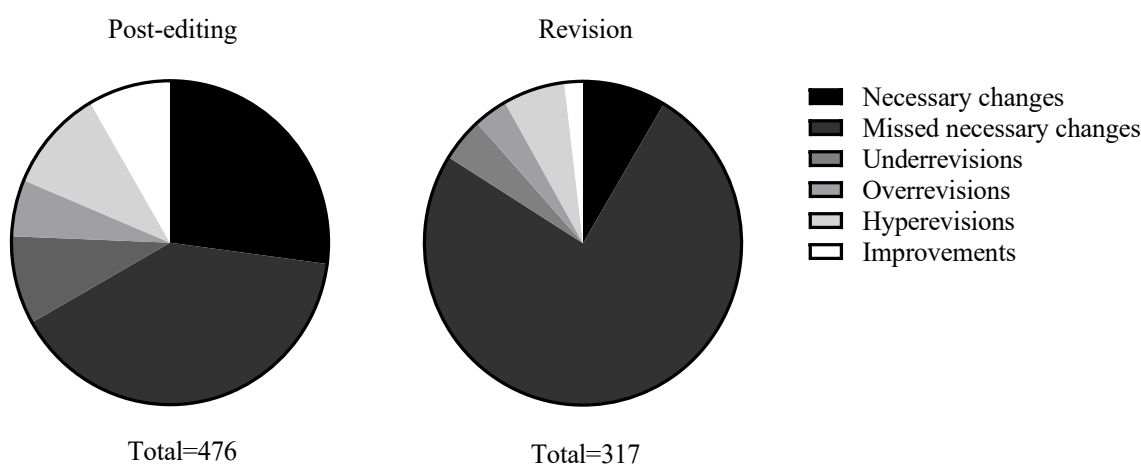


Figure 4.1 Distribution of intervention behaviors per task

According to Figure 4.1, 27.10% of intervention behaviors were necessary in NMT post-editing (N=129), while only 8.52% of changes were necessary in translation revision (N=27). Taking the total number of intervention behaviors per task into consideration, we found that participants corrected more errors in NMT post-editing than in translation revision, while they missed more necessary changes in translation revision. Participants missed 189 necessary changes in NMT post-editing, accounting for nearly 38.77% of total changes. The figure was much higher in translation revision, more than three quarter of total changes (N=239). As for underrevisions, participants detected but failed to correct 42 errors in NMT post-editing, while this figure decreased to 14 in translation revision. The results of these three intervention types revealed that it was more difficult for translation trainees to rectify errors in the HT. Those results could also be explained by the results of questionnaires. 55% of participants strongly agreed and they feel it easier to detect errors in NMT than in HT (N=11) and 40% agreed that (N=8). We further explored the reasons why translation trainees have such an idea. We hypothesized that

translation with less fluency errors might arouse less intervention behaviors. A translation with less fluency errors means that it reads more smoothly, which might leave a false impression that the translation was in a good quality and impair translators' alertness. Errors in HT are trickier since HT reads fluently and smoothly. If translation machine couldn't figure out some sentences, its translation might expose those problems. However, translators could adopt some methods to hide those issues.

Apart from the overall data, the proportion of overrevisions was also vital to reflect translation trainees' understanding of ST. As defined in the first paragraph, overrevisions are new errors introduced by translation trainees. Participants added 27 and 11 new errors into NMT post-editing and translation revision respectively. Those errors were a presentation of their misunderstanding to ST.

(1) ST1: I grew in those seasons like corn in the night...

NMT: 我在那些季节里成长, 就像黑夜里的玉米一样...

P02: 在四季更迭中, 我长大了, 就像黑夜中的玉米一样...

In this example, P02 changed “那些季节” in NMT of ST1 into “在四季更迭中” when translating “I grew in those seasons like corn in the night”. “Seasons” in this context refers to the time periods that the author spent alone, rather than four seasons in a year in common sense.

(2) ST2: or, while the sun was setting, made my supper of huckleberries and blueberries on Fair Haven Hill, and laid up a store for several days.

HT: 或是于日落时分, 在费尔黑文山上用越橘和蓝莓做顿晚餐, 吃上好几天。

P14: 或在日落时分, 在费尔黑文山上用越橘和蓝莓做顿晚餐, 攒点果子叫卖, 待上好几天。

In this example, P14 added “攒点果子叫卖” at the end of first sentence in HT of ST2 when translating “laid up a store for several days”. “叫卖” in Chinese is to sell or peddle products, which is an addition for there was no such an idea conveyed by the ST.

The term “hyperrevisions” has drawn more attention from scholars in the

previous studies. Hyperrevisions have been used as an indicator of one of the components of Translation Revision Competence (TRC). Robert *et al.* (2017b) conduct a research to investigate the fairness and tolerance of translation trainees and revision trainees. The results show that revision trainees show more fairness and tolerance, in that they make fewer hyperrevisions. On the basis of one of the major principles in revision, hyperrevisions, that is unnecessary changes, must be avoided. Revision is not about asking oneself whether a sentence can be improved, but whether it needs to be improved (Mossop 2007). More recently, Mossop *et al.* (2020: 226) list six bad attitudes of revisers, among which two are related to unnecessary changes: (1) “I wonder if this passage can be improved. (Of course it can, but does it need to be?)” and (2) “I’m revising, so I have to make some changes. (No, you don’t)”. To examine such mind set in the experiment, the following would probe deeply into the related data. In Figure 4.1, participants have made hyperrevisions in NMT post-editing with a total of 49. In translation revision, participants made a total of 20 unnecessary changes, in which P20 contributed 4 of them.

(3) ST1: I realized what the Orientals mean by contemplation and the forsaking of works.

NMT: 我明白了东方人所说的沉思和放弃工作是什么意思。

P20: 在这样的状态中，我终于明白了东方人所说的沉思和放弃工作是什么意思。

P20 made 10 unnecessary changes when post-editing NMT, such as adding some nonsense words like “在这样的状态中”. Those unnecessary changes might not affect the quality of translation but they made the translation wordy.

(4) ST2: The fruits do not yield their true flavor to the purchaser of them, nor to him who raises them for the market.

HM: 买来这些水果的人不会尝到它们真实的味道，那些种水果卖到市场的人也不会尝到。

P20: 在商店里买来这些水果的人是品尝不到它们真实的味道的，那些专门

种这些水果，然后卖到市场的人也尝不到这些滋味。

In the fourth example, P20 tended to add some inappropriate punctuation to cut a coherent sentence into two short sentences. When translating “the fruits do not yield their true flavor to the purchaser of them, nor to him who raises them for the market”, P20 broke the initial HT into several short sentences. The corresponding initial HT was shorter and much clearer than the revised one. Therefore, such kind of changes were unnecessary.

Compared with other five types, improvements were relatively rare, especially in translation revision. Participants made 40 improvements in NMT post-editing and only six in translation revision. It indicated that the draft translation quality of HT might be slightly better than NMT. Therefore, we couldn't just compare the final translation quality of post-edited NMT and revised HT. We also needed to take into account the comparison between the draft and final translation quality.

4.1.2 Intervention quality

Based on the distribution of intervention behaviors, we have assessed the actual intervention quality. A few scholars have created formulas for quantifying the quality of a revision product. We decided to adopt a latest formula (Daems & Macken 2020) to calculate intervention quality (see Equation 4.1). The slight difference was that we assigned critical errors a value of 2 and minor errors a value of 1, while the researchers made a severity weight ranging from 0 to 3. The reason why we assigned two degrees was to be consistent with the error taxonomy in the present study.

$$\text{Intervention quality} = \frac{(\text{necessary changes} * \text{severity weight}) - (\text{overrevisions} * \text{severity weight})}{\text{total number of errors} * \text{severity weight}}$$

Equation 4.1 Intervention quality

After collecting the value of intervention behaviors, the overall intervention quality was presented below (see Figure 4.2). Most participants scored a higher intervention quality in NMT post-editing than in translation revision, except P08.

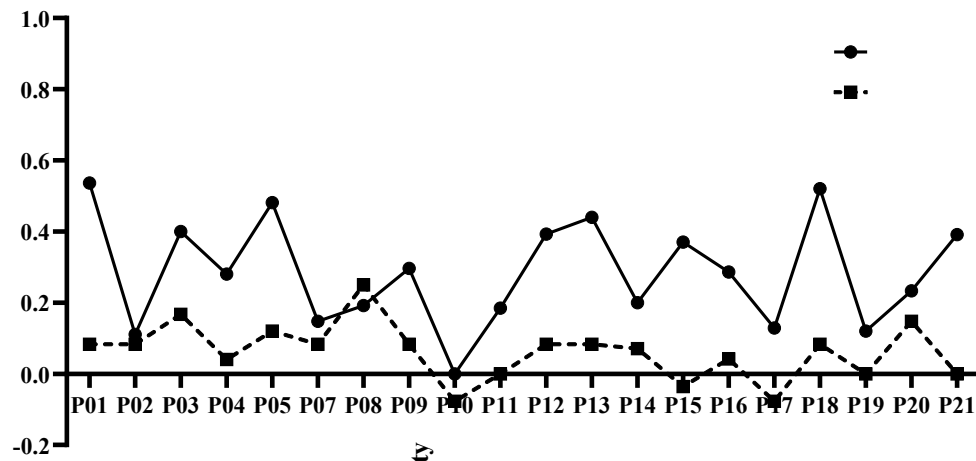


Figure 4.2 Intervention quality per participant

The line chart was supported by the descriptive statistics in Table 4.1. The mean score of intervention quality in NMT post-editing was about 0.286, higher than 0.061 in translation revision. There was no negative score in NMT post-editing, while three participants obtained negative scores in translation revision. For instance, P10 made no necessary changes but introduced a critical error in translation revision. The highest score occurred in NMT post-editing. P01 corrected 11 translation errors but introduced one critical error in NMT. The performance of translation trainees could be interpreted by the result of item 8 in our questionnaire. 35% of participants admitted that they were not sure whether their changes in revision and post-editing are right ($N=7$). And most of participants held a neutral attitude towards this item ($N=9$). Such an idea might influence their efficiency of error detection in two scenarios. It indicates that they need a systematic training of revision and post-editing in which they might acquire the knowledge of how to detect and correct errors in different tasks accurately.

Moreover, the test of normality indicated that intervention quality in NMT post-editing and translation revision were identical to normal distribution ($p>0.05$). And then we further tested the significance of difference of intervention quality in two scenarios (see Table 4.2). The results showed that the intervention quality in NMT post-editing was significantly higher than that in translation revision ($p<0.05$).

Table 4.1 Descriptive Statistics and Normality Tests of intervention quality

	Descriptive statistics					Shapiro-Bilk		
	N	Min	Max	Mean	Std. Deviation	Statistic	d	Sig.
P_Intervention quality	20	.000	.536	.28555	.151235	.965	20	.658
R_Intervention quality	20	-.077	.250	.06145	.079787	.946	20	.307

Table 4.2 Significance Test of intervention quality

	Paired Samples Test		
	t	d	Sig. (2-tailed)
P_Intervention quality - R_Intervention quality	6.828	19	.000

4.2 Post-editing Effort in Revision and Post-editing of Literary Texts

4.2.1 Temporal effort

Differences of temporal effort in two tasks were discussed firstly. Drafting time (Durd) and total task time (Dur) were included as dependent variables in the study. Drafting time is “the time offset from the beginning of the session until the first keystroke, which coincides with the end of the orientation phase” (Carl *et al.* 2016: 20-21). Total task time, in the present study, refers to the time participants spent in each task from clicking the button of “start logging” to that of “stop logging”. The descriptive statistics were displayed in Table 4.3 and Figure 4.3.

In this experiment, translation trainees spent an average of 71124.10ms on drafting during post-editing, while doubled the drafting time during transaction revision (M=142299.75ms). By contrast, the whole NMT post-editing task took 2016574.05ms on average, while translation revision task only took half of the time needed in NMT post-editing (M=1066190.65ms).

Table 4.3 Descriptive Statistics and Normality Tests of temporal effort

Descriptive statistics						Shapiro-Bilk		
	N	Min	Max	Mean	Std. Deviation	Statistic	d	Sig.
P_Durd	20	11046	295594	71134.10	72553.601	.664	20	.000
P_Dur	20	785297	14439016	2016574.05	2940409.161	.327	20	.000
R_Durd	20	16400	320407	142299.75	100283.093	.916	20	.083
R_Dur	20	393265	1960531	1066190.65	465857.052	.957	20	.482

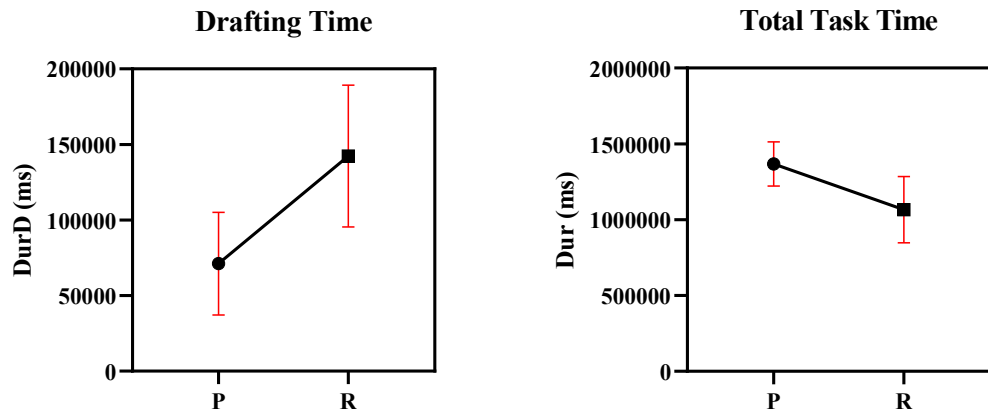


Figure 4.3 Temporal effort of post-editing (P) and translation revision (R)

The Normality test (see Table 4.3) suggested that the drafting time data and the total task time data in NMT post-editing were not normally distributed. Given this fact, Wilcoxon Signed Ranks Test was applied to Durd and Dur. The difference of drafting time in two scenarios were significant (Sig.=0.006, $p<0.05$). The slightly longer drafting time during translation revision might be explained by the fact that there were less errors in HT than in NMT, particularly accuracy errors. Therefore, it was more difficult for translation trainees to detect the first error before making the first keystroke. For total task time, translation trainees significantly spent more time to finish NMT post-editing than translation revision (Sig.=0.003, $p<0.05$). After comparing translation trainees' performance in two scenarios, we found that the reduction of temporal effort in total task time supported the finding in intervention behaviors (Sec. 4.1.1).

4.2.2 Technical effort

Technical effort was measured by the number of total keystrokes (Key), insertions (Ins) and deletion (Del). We collected the keystrokes data through the Translog II. The results could be detected in Table 4.4 and Figure 4.4.

Participants performed 721.75 keystrokes on average when post-editing NMT, while performed much less in revision (M=415.1). Translation revision could reduce participants' total keystrokes by 42.49% compared with NMT post-editing. For insertion, the result was similar to total keystrokes, where participants inputted 53.85 insertions on average in post-editing and 29.9 in revision. As for deletion, participants deleted more in post-editing (M=66.45) than in revision (M=31.45).

Table 4.4 Descriptive Statistics and Normality Tests of technical effort

Descriptive statistics						Shapiro-Wilk		
	N	Min	Max	Mean	Std. Deviation	Statistic	df	Sig.
P_Key	20	330.00	1740.00	721.7500	364.20655	.838	20	.003
P_Ins	20	19.00	130.00	53.8500	31.09963	.867	20	.010
P_Del	20	8.00	188.00	66.4500	46.85417	.877	20	.015
R_Key	20	56.00	2021.00	415.1000	502.23101	.691	20	.000
R_Ins	20	.00	172.00	29.9000	43.02985	.632	20	.000
R_Del	20	.00	197.00	31.4500	45.96850	.651	20	.000

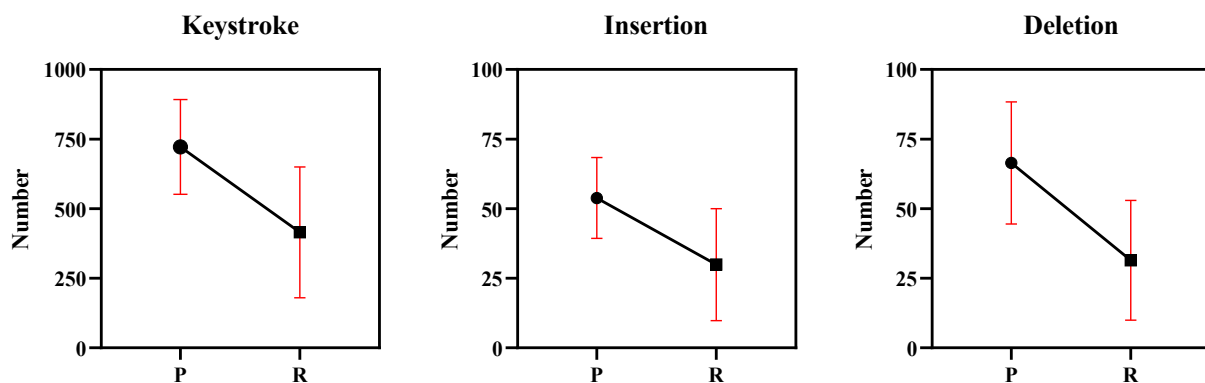


Figure 4.4 Technical effort of post-editing (P) and translation revision (R)

The normality test (see Table 4.4) revealed that the comparative analysis of the number of total keystrokes, insertions and deletions should use nonparametric test because the principle that three data sets must be normally distributed couldn't be satisfied. In accordance with Wilcoxon Signed Ranks Test (see Table 4.5), we concluded that HT could significantly reduce the number of total keystrokes, insertions and deletions for participants for revision tasks ($p < 0.05$).

Table 4.5 Significance Tests of technical effort indicators

Wilcoxon Signed Ranks Test			
	R_Key-P_Key	R_Ins-P_Ins	R_Del-P_Del
Z	-2.950 ^b	-3.140 ^b	-3.305 ^b
Asymp. Sig. (2-tailed)	.003	.002	.001

b. Based on positive ranks.

4.2.3 Cognitive effort

Different from temporal and technical effort, the measurement of cognitive effort is never straightforward. Whilst accepting that supplementary methods should ideally be taken advantage of when using pause analysis to measure cognitive effort (O'Brien 2006:1), pauses are used here as the use of eye-tracking equipment to collect gaze data was not available. We also involve subjective assessment of cognitive load as a supplementation. When collecting pause data, a threshold of 1.5 seconds was used and pauses below were not analyzed. It was decided to use this threshold as it has been linked to the theoretical capacity of the Phonological Loop (Dragsted 2006), a component of working memory which, as noted above, is linked to written language production.

In this study, the cognitive effort was measured by number of pause (PauN) and pause duration (PauD, in ms) recorded by Translog II. Number of pauses were the total number of pauses in two tasks. Pause duration was the sum of time of all pauses for the duration of the entire tasks. According to Table 4.6, when completing revision task, translation trainees had less average number of pauses ($M=48.65$) than

completing post-editing task ($M=79.35$). The result was also true for pause duration, where participants spent an average of 951101.20ms in translation revision and 1177948.35ms in NMT post-editing. The increase was about 226.8s on average. It was also noted that the standard deviations of both indicators in NMT post-editing were smaller than those in translation revision, which meant the degree of dispersion of data sets in NMT post-editing were lower. The differences were more clearly shown in Figure 4.5 below.

Table 4.6 Descriptive Statistics and Normality Tests of cognitive effort

Descriptive statistics						Shapiro-Wilk		
	N	Min	Max	Mean	Std. Deviation	Statistic	df	Sig.
P_PauN	20	42	164	79.35	32.380	.882	20	.019
P_PauD	20	630451	1653750	1177948.35	309326.631	.954	20	.431
R_PauN	20	11	217	48.65	52.230	.711	20	.000
R_PauD	20	335812	1560874	951101.20	393508.834	.941	20	.247

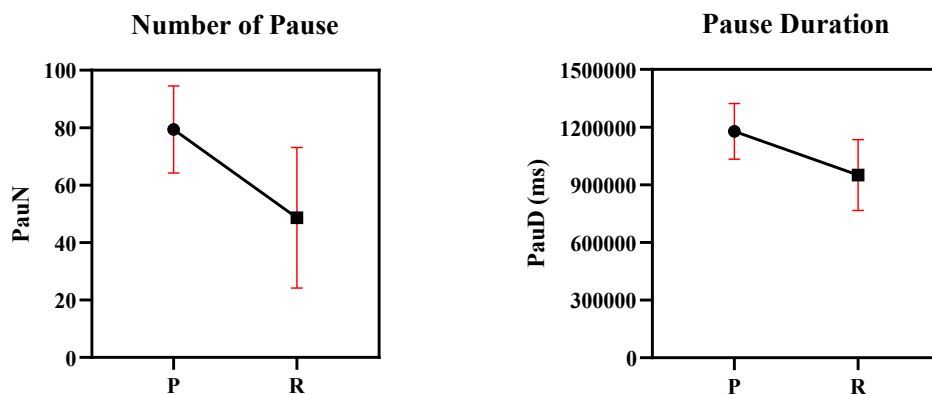


Figure 4.5 Cognitive effort of post-editing (P) and translation revision (R)

Normality test suggested in Table 4.6 that the data set of PauD was normally distributed while the data set of PauN was not. Given this fact, Paired Sample text and Wilcoxon Signed Ranks Test were adopted respectively. In Table 4.7, the number of pauses was significantly lower in translation revision ($\text{Sig.}=0.007$, $p<0.05$). The pause duration was also significantly lower in translation revision ($\text{Sig.}=0.016$ $p<0.05$).

Table 4.7 Significance Tests of cognitive effort indicators

	Paired Samples Test			Wilcoxon Signed Ranks Test	
	t	df	Sig. (2-tailed)	Z	Asymp. Sig. (2-tailed)
R_PauN-P_PauN				-2.688 ^b	.007
P_PauD-R_PauD	2.630	19	.016		
b. Based on positive ranks					

We further investigated translation trainees' subjective assessment on the task difficulty and cognitive load. Translation trainees were asked to scale the task difficulty from four subdimensions, namely mental demand, effort, frustration and performance. Mental demand is an indicator to measure how mentally demanding the task is. Effort is aimed at reflecting how hard the participants have to work to accomplish the task. Frustration reflects how insecure, discouraged, irritated, stressed and annoyed the participants feel during the task. Finally, performance refers to the perceived assessment of their own performance by participants.

The statistics were displayed in Table 4.8 The average mental demand ($M=6.25$) in NMT post-editing was higher than that in translation revision ($M=4.85$). The average effort in NMT post-editing was rated as 6.10 by participants, while the average effort in translation revision was about 4.85. The average frustration in NMT post-editing ($M=4.65$) was slightly higher than that in translation revision ($M=3.55$). Therefore, under the condition that the complexity of source texts was controlled to be similar, translation trainees felt more frustrated during post-editing task. However, according to Table 4.8, participants believed that they might produce a similar performance in two scenarios for the average score of perceived performance were 6.5 in post-editing and 6.4 in translation revision. According to the table above, we conducted tests of normality among those indicators in both scenarios and founded that mental demand and effort were distributed normally, while frustration and performance didn't satisfy the requirements. Thus, we would further conduct paired samples tests for mental demand and effort and nonparametric tests for frustration and performance (see Table 4.9).

Table 4.8 Descriptive Statistics of subjective assessment on task difficulty

	Descriptive statistics					Shapiro-Wilk		
	N	Min	Max	Mean	Std. Deviation	Statistic	df	Sig.
P_Mental Demand	20	2.00	10.00	6.25	1.77334	.931	20	.161
P_Effort	20	3.00	9.00	6.10	1.71372	.937	20	.210
P_Frustration	20	1.00	8.00	4.65	1.75544	.959	20	.533
P_Performance	20	5.00	9.00	6.50	1.27733	.890	20	.027
R_Mental Demand	20	2.00	10.00	4.85	2.43386	.921	20	.102
R_Effort	20	1.00	10.00	4.85	2.71981	.929	20	.145
R_Frustration	20	1.00	9.00	3.55	2.28208	.881	20	.019
R_Performance	20	3.00	9.00	6.40	1.63514	.936	20	.204

Table 4.9 Significance Tests of cognitive effort indicators

	Paired Samples Test		Wilcoxon Signed Ranks Test	
	t	df	Sig. (2-tailed)	Asymp. Sig. (2-tailed)
P_MentalD- R_MentalD	2.208	19	.040	
P_Effort - R_Effort	1.965	19	.064	
R_Frustration-P_Frustration				-1.793 ^b .073
R_Performance-P_Performance				-.177 ^b .859

b. Based on positive ranks.

According to significance tests, we found that translation trainees believed that translation revision task takes significantly less mental demanding (Sig.=0.04, $p<0.05$). However, there was no significant difference in the perceived effort, frustration and performance. The results of perceived performance could be interpreted that translation trainees were required to finish two tasks and their performance in two scenarios might mainly depends on personal translation proficiency.

4.3 Translation Quality in Revision and Post-editing of Literary Texts

Translation quality is measured by adequacy and fluency. And the adequacy and fluency guidelines developed by TAUS (2013). Adequacy is related to how much of the meaning is represented in the translation. Adequacy of translations could be rated from 1 to 4 points with 1 for none and 4 for everything. Fluency refers to to what extent the translation is intuitively acceptable and can be sensibly interpreted by a native speaker. Fluency of translations could be rated from 1 to 4 points with 1 for incomprehensible and 4 for flawless.

Two evaluators with more than 5 years of professional translation experiences were invited to rate the 40 final translations of 20 participants from adequacy and fluency dimensions respectively. In order to avoid personal preference, evaluators in the present study were not informed of the source of translations. And a reference translation produced by a famous Chinese literary translator was provided.

4.3.1 Adequacy

AVEAdequacy of final translations in two tasks were shown in Figure 4.6. Among 20 participants, nine of them (P03, P05, P07, P09, P12, P13, P16, P17, P21) got a higher AVEAdequacy scores in NMT post-editing than in translation revision.

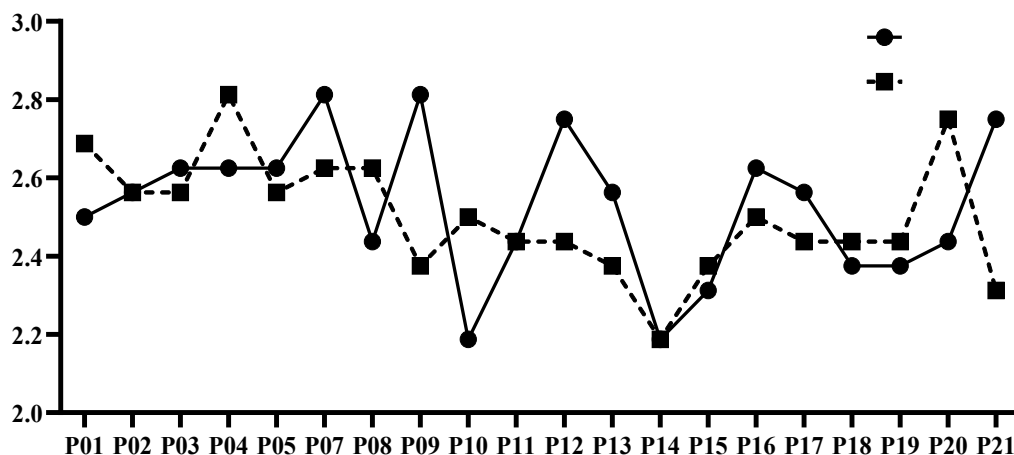


Figure 4.6 AVEAdequacy score per participant in both tasks

We presented descriptive data and the results of normality in Table 4.10. According to the Table below, we could find that the AVEAdequacy of post-edited NMT and revised HT were close to each other. There was a little difference in terms of mean of AVEAdequacy, in which the AVEAdequacy of post-editing was slightly higher than that of translation revision. The results of Shapiro-Wilk tests showed that the AVEAdequacy in post-editing and translation revision was identical to normal distribution ($p>0.05$). We then conducted Paired Samples Test but found that there was no significant difference between two indicators ($p>0.05$) (Table 4.11).

Table 4.10 Descriptive statistics and Normality Tests of AVEAdequacy

	Descriptive Statistics					Shapiro-Wilk		
	N	Min	Max	Mean	Std. Deviation	Statistic	df	Sig.
P_AVEAdequacy	20	2.1875	2.8125	2.528125	.1852663	.955	20	.448
R_AVEAdequacy	20	2.1875	2.8125	2.500000	.1517442	.973	20	.823

Table 4.11 Significance Test of AVEAdequacy

	Paired Samples Test		
	t	df	Sig. (2-tailed)
P_AVEAdequacy - R_AVEAdequacy	.582	19	.567

Furthermore, we took the draft translation quality into consideration. Though the AVEAdequacy of original NMT was rated as 2.25, participants have significantly improved the adequacy of NMT after post-editing tasks ($M=2.52$). As for translation revision tasks, we found that it was a little bit hard for participants to rectify the errors in HT in Section 4.1.1. The AVEAdequacy of original HT was 2.325, higher than that of original NMT, but the AVEAdequacy of final translation in translation revision was 2.5. We thought participants would be more cautious when post-editing NMT, which would help them rectify more errors in the NMT and then improve the adequacy. When participants revising HT, they showed more trust on the HT which might lead to more missed necessary changes and the introduction of new errors might deteriorate the initial adequacy of HT.

4.3.2 Fluency

The results of AVEFluency were presented in Figure 4.7. There were nine participants (P01, P05, P09, P10, P12, P14, P16, P17, P21) whose translation quality in NMT post-editing was rated higher than that in translation revision.

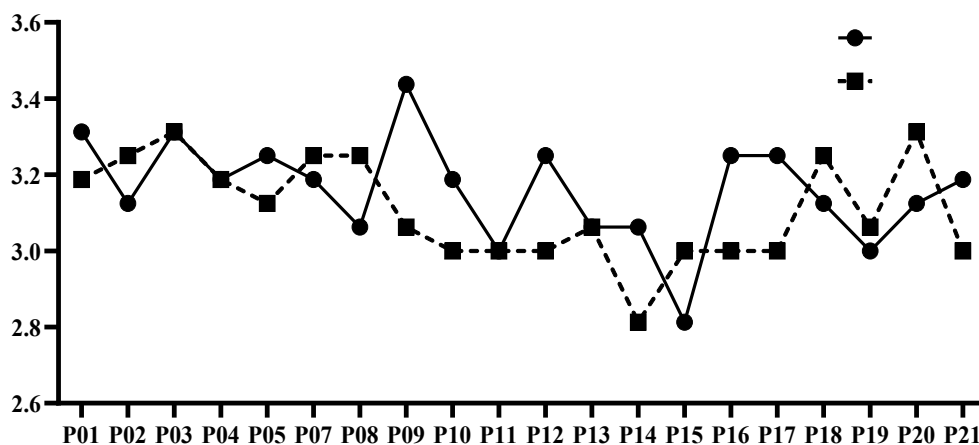


Figure 4.7 AVEFluency score per participant in both tasks

We further examined the descriptive statistics of AVEFluency in two tasks (Table 4.12). The result was similar to that of AVEAdequacy since the mean of AVEFluency of post-edited NMT ($M=3.159$) was slightly higher than that of revised HT ($M=3.106$). And then we conducted tests of normality on both groups of data and found that the AVEFluency scores of NMT post-editing was identical to normal distribution ($\text{Sig.}=0.639, p>0.05$). However, the AVEFluency score of translation revision didn't satisfy the normality distribution ($p<0.05$). The results of Wilcoxon Signed Ranks Test (see Table 4.13) indicated that the AVEFluency between NMT post-editing and translation revision was not statistically different ($\text{Sig.}=0.131, p>0.05$).

Table 4.12 Descriptive Statistics and Normality Tests of AVEFluency

	Descriptive Statistics					Shapiro-Wilk		
	N	Min	Max	Mean	Std. Deviation	Statistic	df	Sig.
P_AVEFluency	20	2.8125	3.4375	3.159375	.1382382	.965	20	.639
R_AVEFluency	20	2.8125	3.3125	3.106250	.1361779	.902	20	.045

Table 4.13 Significance Test of AVEFluency

	Wilcoxon Signed Ranks Test	
	Z	Asymp. Sig. (2-tailed)
R_AVEFluency - P_AVEFluency	-1.510 ^b	.131
b. Based on positive ranks.		

However, we conducted the same comparison in terms of AVEFluency. Compared with the initial AVEFluency, the figure was increased from 3.0 to 3.159 in NMT post-editing, while the score was decreased from 3.1 to 3.106 in translation revision. It was a great challenge for translation trainees to improve the fluency of translations, especially when the initial fluency was good. The results indicated that translation trainees need to intensify their translation revision competence.

In the present study, every participant had completed both tasks in the same site, and the complexity of the source texts for the two tasks had been controlled to be similar. However, we had found the translation quality of final products in two task types had no significant difference in terms of adequacy and fluency. Synthesizing results of adequacy and fluency, the results confirmed with Screen (2017) and Jia *et al.* (2019a) who find no significant difference in quality between NMT post-editing and from-scratch translation. The practical translation quality was not in conformity with what respondents expected. Item 8 and 9 in our questionnaires were about text types. 25% of participants strongly agreed (or agreed) that, compared post-editing, revised nonliterary texts has a higher quality (N=5). Most participants held a neutral attitude (N=9). All participants strongly agreed (or agreed) that the quality of revised literary texts is higher than that of post-edited one (N=20). However, the final translation quality in two scenarios was close to each other in the present study.

4.4 Summary

In short, literary NMT has more translation errors than literary HT, especially mistranslation in accuracy error type and intelligible in fluency error type. In translation revision and NMT post-editing tasks, translation trainees could detect and correct more errors in NMT but miss more errors in HT. On the basis of intervention

quality, translation trainees showed a more effective post-editing competence and they needed more practices to enhance their translation revision competence, in particular error detection and correction in the strategic sub-competence. The results from post-editing effort also confirmed this idea. The final translation quality was a pivotal indicator for this comparative study of translation revision and NMT post-editing. The data showed that the translation quality of revised literary HT was similar to that of post-edited literary NMT, which might encourage the further studies on literary NMT. In accordance with item 10 and 11 of our questionnaires, we found that translation trainees show a positive attitude towards NMT post-editing of literary texts, which might also encourage the further development and research of NMT post-editing.

Conclusion

This study has investigated the differences between translation revision and NMT post-editing of literary texts from the perspectives of intervention behaviors, post-editing effort and translation quality. Translation revision and NMT post-editing require different competences, which has been demonstrated in the previous theoretical studies. But it needs empirical support. Thus, we conducted an experiment to explore and verify the differences. Moreover, it has been proved that NMT will help literary translators improve their productivity in literary translation. Those studies have laid a foundation for the present study which would like to reveal whether the quality of NMT post-editing will exceed that of translation revision in literary translation. In line with previous research in translation revision and NMT post-editing, the present study collected data through methods including questionnaire, key logging and retrospective report, trying to address those issues.

(1) Revisiting the Research Questions

RQ1. How do participants rectify literary texts in revision and post-editing?

The results showed that the number of changes made in NMT post-editing was one and a half times that in translation revision, which could be interpreted by the slightly different translation quality and translation trainees' attitudes toward HT and NMT. According to our survey, most respondents held a view that there were more errors in NMT, compared with HT. However, translation trainees in fact corrected more errors in NMT post-editing but missed more errors in translation revision. It seemed that it was easier for participants to detect errors in NMT than in HT. As for the large number of missed necessary errors in HT, we made an analysis based on the retrospective reports of participants. Most of them believed that the quality of HT was trustworthy for it was expressed in a fluent and native way. Their trust towards HT might lead to an insufficiently rigorous detection.

Intervention quality in the present study were calculated for a better observation of translation trainees' behaviors. The results showed that participants obtained a

higher intervention quality score in post-editing than in revision, which was consistent with the results of intervention behaviors. It also supported the study of Daems and Macken (2020). The result suggested that translation trainees in the present study performed better in terms of error detection and correction in NMT post-editing. It also suggested that translation trainees need to improve their strategic subcompetence in translation revision competence, especially error detection and correction.

RQ2. Compared with revision, what is post-editing effort spent by participants in post-editing of literary texts?

2.1 Do participants spend more temporal effort in post-editing of literary texts?

The results show that participants significantly spent more time on drafting time in revision than in post-editing, while they significantly spent more time to finish the post-editing task compared with the total task time of revision one.

2.2 Do participants make more technical effort in post-editing of literary texts?

Data of technical effort was collected via Translog II program, including total number of keystrokes, insertions and deletions. The results showed that participants made more keystrokes in NMT post-editing than in translation revision from three key logging indicators. Those figures demonstrated the assessment in intervention behaviors were reliable. The results of significance tests suggested that participants spent significantly less technical effort in translation revision than in NMT post-editing.

2.3 Do participants invest more cognitive effort in post-editing of literary texts?

We used number of pause and pause duration to calculate translation trainees' cognitive effort. The results show that translation trainees paused more times in post-editing than in revision. And the total pause duration in post-editing task is significantly longer than that in revision task. We also adopted a scale for collecting translation trainees' perception of task difficulty and cognitive load. They assumed that NMT post-editing might take more mental demand. However, participants believed that they would obtain a similar performance in NMT post-editing and translation revision. The average performance score in NMT post-editing was even slightly higher than in translation revision. This could be interpreted by the translation

proficiency of translator trainees for they engaged in both tasks. The significance tests suggested that mental demand was significantly reduced in translation revision, compared with NMT post-editing. However, the differences of effort, frustration and performance between two tasks were not significant.

RQ3. Compared with revised text, what is final translation quality of post-edited literary texts?

3.1 What is the adequacy of revised and post-edited literary texts?

We evaluated the translation quality based on TAUS (2013) which was measured by adequacy and fluency. The results showed that the average adequacy score in NMT post-editing was slightly higher than that in translation revision. But the Paired Sample Tests showed that there was no significant difference of average adequacy scores between translation revision and NMT post-editing. In addition, we also compared the average adequacy score of final translation quality with the original one. The score has been increased significantly in NMT post-editing, which indicated that participants did a good job in post-editing of literary translation. By contrast, there was a slight decline of the average adequacy score in translation revision. It provided an evidence for the suggestion that translation trainees in the present study need to improve their translation revision competence.

3.2 What is the fluency of revised and post-edited literary texts?

The results showed that the difference of the average fluency between translation revision and MEPT was not significant. Though the average fluency scores were not significantly different, we found something interesting when comparing the average fluency scores. The final average fluency score of HT was lower than the initial average fluency score in translation revision and even the final average fluency score in NMT post-editing. It also served as an evidence for the urgent need to improve translation trainees' revision competence.

(2) Implications of this Study

Based on the above results, the present study may provide some implications for further research.

To begin with, revision of translation is an increasingly important topic today due to its significance to translation quality and the urgent market demand. However, the attention paid on translation revision remain far behind its pivotal role in the translation. This could be supported by the results of our survey. 85.4% of respondents (N=41) said they have not attended a class specialized for revision of translation but 95.8% of respondents (N=46) wished have such a course in post-graduate. One of the major reasons why translation revision failed to draw the public attention in the past decades lies on inappropriate perceptions. Some translators often believe that revision is only a part in the translation. Revision of translation could be a separate activity in the translation market. We conducted the experiment to investigate participants' revision competence and found that translation trainees indeed need to improve their translation revision competences, especially error detection and correction.

Moreover, the development of MT will help translators involve in a better translation experience, which has been verified by some researchers. Toral Ruiz *et al.* (2018) expand his research into literary texts and find that MT could improve literary translators' productivity. When it comes to literary translations, it is reasonable that some translators might refuse to use MT since MT of literary texts lacks of creativity. But we couldn't deny the advantages of MT for this reason. NMT post-editing is an activity full of creativity due to the participation of translators or post-editors. The potential of NMT post-editing of literary texts also has been demonstrated in the present study, in which the final translation quality of post-edited literary texts is close to that of revised literary texts.

(3) Research Limitations and Future Directions

Apart from the findings and implications, there are also some limitations in the present study. In the light of these limitations, we come up with some suggestions for further research. Firstly, the comparative study between translation revision and NMT post-editing has been conducted with a literary text as its experimental text. Though the results show that the final translation quality of post-edited NMT is close to that of revised HT, it indeed takes more cognitive effort in post-editing tasks. It will cost

more cognitive effort in translation revision tasks if we take the translation from scratch into account. Furthermore, it is suggested that a similar investigation should be undertaken for non-literary translation projects since the text type might affect the universality of these findings in the present study. It will provide a more powerful evidence for the differences between translation revision and NMT post-editing. Thirdly, we found that the size of quantitative data was relatively small. It is suggested to expand the size of respondents and participants and analyze the data from an overall view. Finally, we could further investigate whether there is correlation between specific translation error types and the error detection and correction of translation trainees.

Notes

1. Saint Jerome (347-420) was a biblical translator and monastic leader, traditionally regarded as the most learned of the Latin Fathers.
2. Pope Damasus I (305-384) was the bishop of Rome from October 366 to his death. He presided over the Council of Rome of 382 that determined the canon or official list of sacred scripture.
3. Warren Weaver (1894-1978) was an American mathematician famous as one of the pioneers of machine translation.
4. Test for English Majors-Band 8 (TEM-8) is based on the highest level of standard for English major students in China. It is taken at the eighth term. TEM-8 evaluates students' English ability in listening, reading, writing and translating.
5. <https://readabilityformulas.com/> (Nov. 30th, 2021)
6. Lexile Analyzer® <https://hub.lexile.com/analyzer> (Nov. 30th, 2021)
7. MQM <https://themqm.info/typology/> (Jan. 6th, 2022)
8. <https://cran.r-project.org/web/packages/irr/irr.pdf> (Dec. 24th, 2021)
9. EVCapture <https://www.ieway.cn/evcapture.html> (Dec. 21th, 2021)
10. <https://www.wjx.cn/vj/he3LwAn.aspx> (Dec. 12th, 2021)

References

- [1] Arenas, A. G., & A. Toral Ruiz. 2020. The impact of post-editing and machine translation on creativity and reading experience [J]. *Translation Spaces* 9(2):1-28.
- [2] Arnold, D., *et al.* 1994. *Machine Translation: An Introductory Guide* [M]. London: Blackwell.
- [3] Arthern, P. J. 1983. Judging the quality of revision [J]. *Lebende Sprachen* 28(2): 53-57.
- [4] Arthern, P. J. 1991. Quality by numbers: assessing revision and translation [A]. In Picken (eds.). *Proceedings of the Fifth Conference of the Institute of Translation and Interpreting* [C]. 85-94.
- [5] Aziz, W., C. M. S. Sheila and L. Specia. 2012. PET: A Tool for Postediting and Assessing Machine Translation [A]. In Nicoletta Calzolari *et al.* (eds.). *Proceedings of the 8th International Conference on Language Resources and Evaluation* [C]. European Language Resources Association (ELRA). 3982-3987.
- [6] Banerjee, S., & A. Lavie. 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments [A]. In *Proceedings of the ACL workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization* [C]. Michigan: Association for Computational Linguistics. 65-72.
- [7] Bisiada, M. 2018. The editor's invisibility: Analysing editorial intervention in translation [J]. *Target: International Journal of Translation Studies* 30(2): 288-309.
- [8] Briggs, N. 2018. Neural machine translation tools in the language learning classroom: Students' use, perceptions, and analyses [J]. *The JALT CALL Journal* 14(1): 3-24.
- [9] Brunette, L., C. Gagnon & J. Hine. 2005. The GREVIS project: Revise or court calamity [J]. *Across Languages and Cultures* 6(1): 29-45.
- [10] Callison-Burch, C., P. Koehn, C. Monz, M. Post, L. Specia. 2012. Findings of the 2012 workshop on statistical machine translation [A]. In Callison-Burch *et al.* (eds.). *7th Workshop on Statistical Machine Translation. Proceedings of the workshop* [C].

- Stroudsburg, PA: Association for Computational Linguistics. 10-51.
- [11]Carl, M. 2012. Translog-II: A program for recording user activity data for empirical reading and writing research [A]. In *the Eight International Conference on Language Resources and Evaluation* [C]. Istanbul, Turkey. 153-162.
- [12]Carl, M., & M. C. T. Báez. 2019. Machine translation errors and the translation process: a study across different languages [J]. *The Journal of Specialised Translation* 31: 107-132.
- [13]Carl, M., S. Bangalore & M. Schaeffer. 2016. *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB* [C]. New York: Springer.
- [14]Castilho, S., & S. O'Brien. 2016. Evaluating the impact of light post-editing on usability [A]. In *Proceedings of the 10th International Conference on Language Resources and Evaluation* [C]. LREC 2016. 310-16.
- [15]Daems, J., *et al.* 2017. Translation methods and experience: A comparative analysis of human translation and post-editing with students and professional translators [J]. *Meta* 62: 245-270.
- [16]Daems, J. & L. Machen. 2019. Interactive adaptive SMT versus interactive adaptive NMT: a user experience evaluation[J]. *Machine Translation* 33:117-34.
- [17]Daems, J. & L. Machen. 2020. Post-editing human translations and revising machine translations [A]. *Translation Revision and Post-editing* [C]. London/New York: Routledge. 50-70.
- [18]do Carmo, F. & J. Moorkens. 2020. Differentiating editing, post-editing and revision [A]. *Translation Revision and Post-Editing* [C]. London/New York: Routledge. 35-49.
- [19]Dragsted, B. 2006. Computer aided translation as a distributed cognitive task [J]. *Pragmatics and Cognition* 14(2):443-463.
- [20]Elming, J., Balling, L. W., & Carl, M. 2014. Investigating user behaviour in post-editing and translation using the CASMACAT workbench [A]. In O'Brien *et al.* (red.). *Post-editing of Machine Translation: Processes and Applications* [C]. Cambridge Scholars Publishing. 147-169.
- [21]EN. 2006. *EN 15038: Translation Services—Service Requirements*. European

Committee for Standardization.

[22]Feinauer, I., & A. Lourens. 2017. The loyalty of the literary reviser: author, source text, target text or reader? [J]. *Stellenbosch Papers in Linguistics Plus* 53: 97-118.

[23]García, I. 2012. A brief history of postediting and of research on postediting [J]. In Anthony Pym and Alexandra Assis Rosa (eds.). *New Directions in Translation Studies. Special Issue of Anglo Saxonica* 3(3): 292-310.

[24]Gaspari, F., H. Almaghout and S. Doherty. 2015. A survey of machine translation competences: insights for translation technology educators and practitioners [J]. *Perspectives: Studies in Translatology* x:1-26.

[25]Germann, U. 2008. Yawat: Yet Another Word Alignment Tool [A]. In *Proceedings of the ACL-08: HLT Demo Session (Companion Volume)* [C]. 20-23.

[26]Guerberof Arenas, A. 2008. Productivity and quality in the post-editing of outputs from translation memories and machine translation [J]. *Localisation Focus* 7(1): 11-21.

[27]Guerberof Arenas, A. 2017. Quality is in the eyes of the reviewer: a report on postediting quality evaluation [A]. In Jakobsen and Mesa-Lao (eds.). *Translation in Transition: Between Cognition, Computing and Technology* [C]. 188-206.

[28]Herbig, N., S. Pal, M. Vela, *et al.* 2019. Multi-modal indicators for estimating perceived cognitive load in post-editing of machine translation [J]. *Machine Translation* 33(1): 91-115.

[29]Huang, J. 2018. Working styles of translation trainees in self-revision, other-revision and post-editing [A]. *Eye Tracking and Multidisciplinary Studies on Translation* [C]. Amsterdam/Philadelphia: John Benjamins Publishing Company. 145-184.

[30]ISO. 2015. *ISO 17100: Translation Services—Requirements for Translation Services*. Geneva: International Organization for Standardization.

[31]ISO. 2017. *ISO 18587: Translation services--Post-editing of machine translation output—Requirements*. Geneva: International Organization for Standardization.

[32]Jakobsen, A. L. 2019. Moving translation, revision, and post-editing boundaries [A]. In Helle V. Dam *et al.* (eds.). *Moving Boundaries in Translation Studies* [C].

ondon/New York: Routledge. 64-80.

[33]Jensen, K. T. H. 2009. Indicators of text complexity [J]. *Copenhagen Studies in Language* (37): 61-80.

[34]Jia, Y., M. Carl & X. Wang. 2019a. How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study [J]. *The Journal of Specialised Translation* 31: 60-86.

[35]Jia, Y., M. Carl & X. Wang. 2019b. Post-editing neural machine translation versus phrase-based machine translation for English-Chinese [J]. *Machine Translation* 33:1-2, 9-29.

[36]Jia, Y. & B. Zheng. 2022. The interaction effect between source text complexity and machine translation quality on the task difficulty of NMT post-editing from English to Chinese: A multi-method study[J]. *Across Languages and Cultures* 23(1):36-55.

[37]Jones, F. R. 2006. Unlocking the black box: researching poetry translation processes [A]. In Pertecheggla and Lofredo (eds.). *Translation and Creativity: Perspective on Creative Writing and Translation Studies* [C]. 59-74.

[38]Jones, R., & A. Irvine. 2013. The (un)faithful machine translator [A]. In *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (Sofia)* [C]. 96-101.

[39]Koehn, P. 2009. A process study of computer-aided translation [J]. *Machine Translation* 23(4): 241-263.

[40]Koglin, A., & R. Cunha. 2019. Investigating the post-editing effort associated with machine-translated metaphors: a process-driven analysis [J]. *The Journal of Specialised Translation* 31(1): 38-59.

[41]Kolb, W. 2013. Who are they? Decision-making in literary translation [A]. In Way et al. (eds.). *Tracks and Treks in Translation Studies* [C]. 207-21.

[42]Konttinen, K., L. Salmi & M. Koponen. 2020. Revision and post-editing competences in translator education [A]. *Translation Revision and Post-Editing* [C]. London/New York: Routledge. 187-202.

[43]Koponen, M. 2016. Is machine translation post-editing worth the effort? A survey of research into post-editing and effort [J]. *The Journal of Specialised Translation* 25:

131-148.

[44]Koponen, M., L. Salmi, M. Nikulin. 2019. A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output [J]. *Machine Translation* 33(1): 61-90.

[45]Klings, H. P. 2001. *Repairing Texts: Empirical Investigations of Machine Translation Post-editing Processes* [M]. Kent, Ohio & London: The Kent State University Press.

[46]Künzli, A. 2005. What principles guide translation revision? A combined product and process study [A]. In Kemble (ed.). *Translation Norms: What Is 'Normal' in the Translation Profession? Proceedings of the Conference Held on 13th November 2004 in Portsmouth* [C]. Portsmouth: University of Portsmouth. 31-43.

[47]Künzli, A. 2006. Teaching and learning translation revision: some suggestions based on evidence from a think-aloud protocol study [A]. In Garant (ed.). *Current Trends in Translation Teaching and Learning* [C]. Helsinki, Finland: University of Helsinki. 9-23.

[48]Künzli, A. 2007. Translation revision: A study of the performance of ten professional translators revising a legal text [A]. In Gambier. *et al.* (eds.). In *Doubts and Directions in Translation Studies, Selected Contributions from the EST Congress, Lisbon 2004* [C]. Amsterdam/Philadelphia: John Benjamins. 115-126.

[49]Lacruz, I., & G. M. Shreve. 2014. Pauses and cognitive effort in post-editing [A]. In O'Brien *et al.* (eds.). *Post-Editing of Machine Translation: Processes and Applications* [C]. 246-272.

[50]Landis, R. J., & G. G. Koch. 1977. The measurement of observer agreement for categorical data [J]. *Biometrics* 33(1):159-174.

[51]Linguistic Data Consortium. 2002. Linguistic Data Annotation Specification: Assessment of Fluency and Adequacy in Arabic-English and Chinese-English Translations [OL]. <https://catalog.ldc.upenn.edu/docs/LDC2003T17/TransAssess02.pdf>[Consulted on Jan. 8, 2022].

[52]Lommel, A., *et al.* 2014. Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics [J]. *Tradumàtica* 12: 455-463.

- [53]Marashi, H., & M. Okhowat. 2013. The comparative impact of editing texts translated into Farsi with and without the original English texts [J]. *Perspectives* 21(3): 299-310.
- [54]Marin-Lacarta, M., & M. Vargas-Urpi. 2019. Translators revising translators: a fruitful alliance [J]. *Perspectives: Studies in Translation Theory and Practice*. 27(3): 404-18.
- [55]Martin, T. 2007. Managing risks and resources: a down-to-earth view of revision [J]. *The Journal of Specialised Translation* (8):57-63.
- [56]McDonough Dolmaya, J. 2015. Revision history: Translation trends in Wikipedia [J]. *Translation Studies* 8(1): 16-34.
- [57]Moorkens, J. 2018. Eye tracking as a measure of cognitive effort for post-editing of machine translation [A]. In Walker and Federici (eds.). *Eye Tracking and Multidisciplinary Studies on Translation* [C]. 55-69.
- [58]Moorkens, J., S. Castilho, F. Gaspari and S. Doherty. 2018. *Translation Quality Assessment: From Principles to Practice* [M]. Springer.
- [59]Moorkens, J., & S. O'Brien. 2017. Assessing user interface needs of post-editors of machine translation [A]. *Human Issues in Translation Technology* [C]. London/New York: Routledge. 127-148.
- [60]Mossop, B. 2007. Empirical studies of revision: what we know and need to know [J]. *The Journal of Specialised Translation* (8):5-20.
- [61]Mossop, B., J. M. Hong & C. Teixeira. 2020. *Revising and Editing for Translators* [M]. London/New York: Routledge.
- [62]Nitzke, J., S. Hansen-Schirra and C. Canfora. 2019. Risk management and post-editing competence [J]. *The Journal of Specialised Translation* 31: 39-59.
- [63]Nitzke, J., & K. A. Gros. 2020. Preferential changes in revision and post-editing [A]. *Translation Revision and Post-Editing* [C]. London/New York: Routledge. 21-34.
- [64]O'Brien, S. 2002. Teaching post-editing: a proposal for course content [A]. In *Proceedings of the 6th EAMT Workshop Teaching Machine Translation* [C]. 99-106.
- [65]O'Brien, S. 2005. Methodologies for measuring the correlations between post-editing effort and machine translatability [J]. *Machine Translation* 19(1): 37-58.
- [66]O'Brien, S. 2006. *Pauses as indicators of cognitive effort in post-editing machine*

- translation output [J]. *Across Languages and Cultures* 7(1):1-21.
- [67]O'Brien, S. 2010. Introduction to post-editing: Who, what, how and where to next? [A]. In *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas (AMTA)* [C]. Denver, Colorado, USA: AMTA.
- [68]O'Brien, S. *et al.* 2011. Translation quality evaluation framework. [www.taus.net/component/rsfiles/download?path=Reports%252FFree%2BReports%252Ftausdynamic quality.pdf](http://www.taus.net/component/rsfiles/download?path=Reports%252FFree%2BReports%252Ftausdynamic%20quality.pdf).
- [69]Orr, D. B., & V. H. Small. 1967. Comprehensibility of Machine-aided Translations of Russian Scientific Documents [J]. *Mechanical Translation and Computational Linguistics* 10(1-2): 1-10.
- [70]Ortiz-Martinez, D., J. González-Rubio, V. Alabau, G. Sanchis-Trilles and F. Casacuberta. 2016. Integrating online and active learning in a computer-assisted translation workbench [A]. In Carl *et al.* (eds.). *New Directions in Empirical Translation Process Research* [C]. 57-76.
- [71]Papineni, K., R. Salim, W. Todd & Z. Wei-Jing. 2002. BLEU: a method for automatic evaluation of machine translation [A]. In *Proceedings of the 40th annual meeting on association for computational linguistics* [C]. Philadelphia: Association for Computational Linguistics. 311-318.
- [72]Plitt, M., & F. Masselot. 2010. A productivity test of statistical machine translation post-editing in a typical localisation context [J]. *The Prague Bulletin of Mathematical Linguistics* 93:7-16.
- [73]Popič, D. 2014. Revising translation revision in Slovenia [A]. *New Horizons in Translation Research and Education 2* [C]. University of Eastern Finland. 72.
- [74]Pym, A. 2013. Translation skill-sets in a machine-translation age [J]. *Meta: Translators' Journal* 58(3): 487-503.
- [75]R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna, Austria: R foundation for statistical computing. Retrieved from <https://www.r-project.org/>
- [76]Rensburg, A. 2017. Developing assessment instruments: The effect of a reviser's profile on the quality of the revision product [J]. *Linguistica Antverpiensia, New Series: Themes in Translation Studies* 16: 71-88.

- [77]Rico Pérez, C., & E. Torrejón. 2012. Skills and profile of the new role of the translator as MT post-editor [J]. *Tradumàtica* (10): 0166-178.
- [78]Robert, I. S. 2008. Translation revision procedures: An explorative study [A]. In Pieter (eds.). In *Translation and Its Others. Selected Papers of the CETRA Research Seminar in Translation Studies* [C]. Boulogne. 91.
- [79]Robert, I. S. 2012. *Translation revision: Revision procedures and their impact on the revision process and product (Doctoral dissertation)* [M]. University of Antwerp, Antwerp, Belgium.
- [80]Robert, I. S. 2013. Translation revision: does the revision procedure matter? [A]. *Tracks and Treks in Translation Studies* [C]. Amsterdam/Philadelphia: John Benjamins. 87-102.
- [81]Robert, I. S., A. Remael, J. J. J. Ureel. 2017a. Towards a model of translation revision competence [J]. *The Interpreter and Translator Trainer*. 11(1): 1-19.
- [82]Robert, I. S., J. J. J. Ureel, A. Remael and A. R. Terryn. 2017b. Conceptualizing translation revision competence: a pilot study on the ‘fairness and tolerance’ attitudinal component [J]. *Perspectives: Studies in Translation Theory and Practice* 26(1):2-23.
- [83]Robert, I. S., R. Terryn, J. J. J. Ureel & A. Remael. 2017c. Conceptualizing translation revision competence: A pilot study on the ‘tools and research’ subcompetence [J]. *The Journal of Specialised Translation* (28): 293-316
- [84]Robert, I. S., & L. Van Waes. 2014. Selecting a translation revision procedure: do common sense and statistics agree? [J]. *Perspectives* 22(3): 304-320.
- [85]Robert, I. S., S. Iris & J. J. J. Ureel. 2022. Measuring translation revision competence and post-editing competence in translation trainees: methodological issues [J]. *Perspectives*.1-15.
- [86]Robin, E. 2014. Explicitation and implicitation in revised translations [A]. Complex Visibles Out There: *Proceedings of the Olomouc Linguistics Colloquium* [C]. Olomouc, Czech Republic: Palacký University. 559-574.
- [87]Sánchez-Gijón, P., Moorkens, J. & Way, A. 2019. Post-editing neural machine translation versus translation memory segments[J]. *Machine Translation* 33:31-59.
- [88]Schwartz, L. 2014. Monolingual post-editing by a domain expert is highly

- effective for translation triage [A]. *Proceedings of the Third Workshop on Post-Editing Technology and Practice* [C]. Nice, France: AMTA. 34-44.
- [89]Screen, B. 2017. Machine translation and Welsh: Analysing free Statistical Machine Translation for the professional translation of an under-researched language pair [J]. *The Journal of Specialised Translation* 28: 317-344.
- [90]Screen, B. 2019. What effect does post-editing have on the translation product from an end-user's perspective[J]. *The Journal of Specialised Translation* 31: 133-157.
- [91]Scocchera, G. 2018. Collaborative revision in editorial/literary translation: some thoughts, facts and recommendations [A]. In Monti & Schnyder (eds.). In *Traduire à plusieurs* [C]. 281-94.
- [92]Scocchera, G. 2019. The competent reviser: a short-term empirical study on revision teaching and revision competence acquisition [J]. *The Interpreter and Translator Trainer* 14(1):19-37.
- [93]Shih, C Y. 2006. Revision from translators' point of view. An interview study [J]. *Target* 2: 295-312.
- [94]Snover, M., B. Dorr, S. Richard, L. Micciulla, & J. Makhoul. 2006. A study of translation edit rate with targeted human annotation [A]. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas* [C]. Cambridge: The Association for Machine Translation in the Americas. 223-231.
- [95]Somers, H. 2003. *Computers and Translation: A Translator's Guide* [M]. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- [96]Specia, L., D. Raj and M. Turchi. 2010. Machine translation evaluation versus quality estimation [J]. *Machine Translation* 24(1), 39-50.
- [97]Sun, S., & G. M. Shreve. 2014. Measuring translation difficulty: An empirical study [J]. *Target* 26 (1): 98-127.
- [98]TAUS. 2013. Adequacy/Fluency Guidelines. <https://taus.net/academy/best-practices/evaluate-best-practices/adequacy-fluency-guidelines>.
- [99]TAUS. 2016. TAUS Post-editing Guidelines [OL]. [https://www.taus.net/thinktank/articles/postedit-articles/taus-post-editing-guidelines\(a](https://www.taus.net/thinktank/articles/postedit-articles/taus-post-editing-guidelines(a)

ccessed 02/10/2021)

- [100] Teixeira, C. S. C. 2014. *The handling of translation metadata in translation tools* [A]. In O'Brien *et al.* (eds.). *Post-Editing of Machine Translation: Processes and Applications* [C]. 109-25.
- [101] Terryn, A. R., I. S. Robert, J. J. J. Ureel, *et al.* 2017. Conceptualizing translation revision competence: a pilot study on the acquisition of the knowledge about revision and strategic subcompetence [J]. *Across Languages and Cultures* 18(1): 1-27.
- [102] Thoreau, H. D. 1999. *Walden* [M]. Oxford University Press.
- [103] Toral Ruiz, A., & A. Way. 2015. Translating literary text between related languages using SMT [A]. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature* (Denver, CO) [C]. 123-132.
- [104] Toral Ruiz, A., & A. Way. 2018. What level of quality can neural machine translation attain on literary text? [A]. *Translation Quality Assessment: From Principles to Practice* [C]. Springer.
- [105] Toral Ruiz, A., *et al.* 2018. Post-editing effort of a novel with statistical and neural machine translation [J]. *Frontiers in Digital Humanities* 5(9):1-11.
- [106] Toral Ruiz, A., & V. M. Sánchez-Cartagena. 2017. A multifaceted evaluation of neural versus phrasebased machine translation for 9 language directions [A]. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics* [C].
- [107] Van Slype, G. 1979. Critical Methods for Evaluating the Quality of Machine Translation [A]. Prepared for the *European Commission Directorate General Scientific and Technical Information and Information Management, Report BR-19142* [C]. Bureau Marcel van Dijk.
- [108] Veale, T. & A. Way. 1997. Gaijin: A bootstrapping approach to example-based machine translation [A]. In *International Conference, Recent Advances in Natural Language Processing* [C]. Tsigov Chark, Bulgaria. 239-244.
- [109] Vidulich, M. A. 1988. The Cognitive Psychology of Subjective Mental Workload [A]. In Peter A. H. and Najmedin M. (eds.). *Human Mental Workload* [C]. Amsterdam: North-Holland. 219-229.

- [110] Vieira, L. N. 2017. From process to product: links between post-editing effort and post-edited quality [A]. *Translation in Transition: between Cognition, Computing and Technology* [C]. Amsterdam: John Benjamins Publishing. 162-186.
- [111] Wang, X., T. Wang, R. M. Martín & Y. Jia.. 2021a. Investigating usability in postediting neural machine translation: Evidence from translation trainees' self-perception and performance [J]. *Across Languages and Cultures* 22(1): 100-123.
- [112] Yamada, M. 2019. The impact of Google neural machine translation on post-editing by student translators[J]. *The Journal of Specialised Translation* 31:87-106.
- [113] 崔启亮, 2021, 人工智能在语言服务企业的应用研究[J], 《外国语文》 37(1):26-32, 73.
- [114] 冯佳, 2017, 译入/译出认知负荷比较研究——来自眼动追踪的证据[J], 《中国外语》 (4): 79-91.
- [115] 黄笑菡、何培芬, 2017, 基于 Translog 程序的不同水平译者翻译修改行为模式研究[J], 《外语电化教学》 (178): 32-39.
- [116] 李梅, 2021, 机器翻译译后编辑过程中原文对译员影响研究[J], 《外语教学》 42(4):93-99.
- [117] 杨文地、范梓锐, 2021, 科技语篇机器翻译的译后编辑例析[J], 《上海翻译》 (6):54-59.
- [118] 杨艳霞、王湘玲, 2019, 中外机译应用研究的可视化分析(1998-2018)[J], 《上海翻译》 (5):33-39, 95.
- [119] 王湘玲、赖思、贾艳芳, 2021b, 人工翻译与神经网络机器翻译译后编辑比较研究——基于对隐喻翻译的眼动追踪与键盘记录数据[J], 《外语教学理论与实践》 (4):115-126.
- [120] 仲文明、舒超, 2020, 译后编辑的能力结构与课程设置——基于国外译后编辑课程的前沿分析[J], 《外语电化教学》 (6):86-91,7.

Appendix B 调查问卷

亲爱的受访者:

您好!

我们是湖南大学外国语学院机器翻译译后编辑研究团队。感谢您积极参与本次有关学生译者对翻译修订和机器翻译译后编辑的态度调查。问卷试题无对错之分。请按照您的真实体验回答以下问题。本次调查建议用时 4-5 分钟。我们将对您提交的个人信息和回答进行严格保密,并承诺该问卷内容仅供研究使用。再次感谢您的参与!

一、 个人信息

您的姓名

您的性别

•男 •女

您的年龄

所在专业

所在年级

QQ 联系方式

学习英语 _____ (年)

已获得证书

•TEM-4_____ •TEM-8_____

我做过职业或兼职译者(通过翻译获得报酬)

•是 •否

我常用的在线机器翻译软件是

•谷歌翻译

•百度翻译

•有道翻译

•DeepL

•其他_____

二、基本情况

2.1 翻译修订（指修改他人译文中的错误）

在日常翻译中，我经常从事翻译修订活动。

•是 •否

我接受过专业的翻译修订培训。

•是 •否

我希望学校开设有关翻译修订的课程。

•是 •否

2.2 机器翻译译后编辑（指修改机器译文中的错误）

在日常翻译中，我经常从事机器翻译译后编辑活动。

•是 •否

我接受过专业的机器翻译译后编辑培训。

•是 •否

我希望学校开设有关机器翻译译后编辑的课程。

•是 •否

三、态度对比

1. 我认为翻译修订和机器翻译译后编辑对译者的能力要求是不一样的。

1=非常不同意，2=不同意，3=中立，4=同意，5=非常同意

2. 我认为自己在修订和译后编辑同一篇文章时会有不同的表现。

1=非常不同意，2=不同意，3=中立，4=同意，5=非常同意

3. 我认为在翻译教学中有必要区分翻译修订和机器翻译译后编辑。

1=非常不同意，2=不同意，3=中立，4=同意，5=非常同意

4. 我认为机器译文中的错误比人工译文的多。

1=非常不同意，2=不同意，3=中立，4=同意，5=非常同意

5. 我认为机器译文中的错误比人工译文的更好找。

1=非常不同意，2=不同意，3=中立，4=同意，5=非常同意

6. 与人工译文相比，我会对机器产出译文做出更多的修改。

1=非常不同意，2=不同意，3=中立，4=同意，5=非常同意

7. 多数情况下，我不确定自己对译文的修改是否正确。

1=非常不同意, 2=不同意, 3=中立, 4=同意, 5=非常同意

8. 翻译非文学文本时, 我认为“人工翻译+修订”的译文质量比“机器翻译+译后编辑”的更高。

1=非常不同意, 2=不同意, 3=中立, 4=同意, 5=非常同意

9. 翻译文学文本时, 我认为“人工翻译+修订”的译文质量比“机器翻译+译后编辑”的更高。

1=非常不同意, 2=不同意, 3=中立, 4=同意, 5=非常同意

10. 我愿意尝试用“机器翻译+译后编辑”的方式翻译文学文本。

1=非常不同意, 2=不同意, 3=中立, 4=同意, 5=非常同意

11. 我可以接受由“机器翻译+译后编辑”产出的高质量文学类译文。

1=非常不同意, 2=不同意, 3=中立, 4=同意, 5=非常同意

您的问卷已提交。感谢您的参与。祝您学习进步!

Appendix C 实验文本

(ST1) I love a broad margin to my life. Sometimes, in a summer morning, having taken my accustomed bath, I sat in my sunny doorway from sunrise till noon, rapt in a revery, amidst the pines and hickories and sumachs, in undisturbed solitude and stillness, while the birds sing around or flitted noiseless through the house, until by the sun falling in at my west window, or the noise of some traveller's wagon on the distant highway, I was reminded of the lapse of time. I grew in those seasons like corn in the night, and they were far better than any work of the hands would have been. They were not time subtracted from my life, but so much over and above my usual allowance. I realized what the Orientals mean by contemplation and the forsaking of works. For the most part, I minded not how the hours went. The day advanced as if to light some work of mine; it was morning, and lo, now it is evening, and nothing memorable is accomplished. Instead of singing like the birds, I silently smiled at my incessant good fortune. (Source: *Walden*; Length: 188 words)

(ST1-HT) 我喜欢让生活富有余地。有时，在某个夏日清晨洗漱完后，我会坐在门前，从日出到晌午，在阳光的沐浴下沉思遐想，周边是松木，山胡桃树和漆树，我完全沉浸在这份寂静里，唯有鸟儿放声歌唱，轻快地在屋里飞来飞去，直到夕阳落在了西边的窗子上，抑或从远处传来了游人的马车声，我才恍然惊觉于时光的流逝。那些日子里，我就像夜里的玉米一样生长，而那些玉米远比任何经手工劳作出来的要好得多。这并不是在浪费时间，相反，它们延长甚至超出了我的生命。我懂得了东方人所说的沉思和放弃的意义。大多数情况下，我并不在乎时间是如何流逝的。白昼在前行，仿佛只是为了照亮我的工作；那时是清晨，可你瞧，转眼间就变成了黑夜，但一件有意义的事都没做成。我没有像鸟儿一样歌唱，我只是静静地微笑，笑我自己好运不断。

(ST1-DeepL NMT) 我喜欢我的生活有一个广阔的空间。有时，在一个夏天的早晨，我洗完习惯的澡后，从日出到中午，坐在我阳光充足的门口，在松树、山核桃和苏门答腊之间，在不受干扰的孤独和静谧中，当鸟儿在周围唱歌或在房子里

无声地飞翔时，直到太阳落在我的西窗，或远处公路上一些旅行者的马车的噪音，我才想起了时间的流逝。我在那些季节里成长，就像黑夜里的玉米一样，它们远比任何手的工作都要好。它们不是从我的生活中减去的时间，而是超出了我通常所允许的范围。我明白了东方人所说的沉思和放弃工作是什么意思。在大多数情况下，我并不关心时间如何流逝。这一天的到来，仿佛是为了照亮我的某些工作；当时是早上，现在是晚上，没有什么值得纪念的事情完成。我没有像鸟儿一样歌唱，而是默默地为我不断的好运而微笑。(obtained in October 29th, 2021)

(ST2) Sometimes, having had a surfeit of human society and gossip, and worn out all my village friends, I rambled still farther westward than I habitually dwell, into yet more unfrequented parts of the town, “to fresh woods and pastures new,” or, while the sun was setting, made my supper of huckleberries and blueberries on Fair Haven Hill, and laid up a store for several days. The fruits do not yield their true flavor to the purchaser of them, nor to him who raises them for the market. There is but one way to obtain it, yet few take that way. If you would know the flavor of huckleberries, ask the cowboy or the partridge. It is a vulgar error to suppose that you have tasted huckleberries who never plucked them. A huckleberry never reaches Boston; they have not been known there since they grew on her three hills. The ambrosial and essential part of the fruit is lost with the bloom which is rubbed off in the market cart, and they become mere provender. As long as Eternal Justice reigns, not one innocent huckleberry can be transported thither from the country's hills. (Source: *Walden*; Length: 192 words)

(ST2-HM) 有时，厌倦了人间俗世和闲言碎语，将所有的乡间好友叨扰一遍后，我会朝着日常居所的西边漫步，走到镇上更加人迹罕至的地方，去寻“新鲜的树林和牧场”。或是于日落时分，在费尔黑文山上用越橘和蓝莓做顿晚餐，吃上好几天。买来这些水果的人不会尝到它们真实的味道，那些种水果卖到市场的人也不会尝到。但其实有一种法子，只是很少有人这么做。你若想知道越橘是什么味儿，就去问问牧童或是山鹑吧。要是你从未摘过越橘，却想着自己已经尝过越橘，那真是犯了个庸俗至极的错误。从来没有越橘到达过波士顿；它们只生长在这三座山头，那边根本无人知晓。水果被装上车运往市场，花朵被蹭掉的时候，水果

本身的香气和精华也就失掉了，它们成了普通的食物，仅此而已。只要永恒的正义还在，就没有一颗无辜的越橘能从乡下的山上运到那边去。

(ST2-DeepL NMT) 有时，我受够了人际交往和闲言碎语，把村里的朋友都打发了，就向西走得比我平时住的地方还远，到镇上更不常去的地方，“到新鲜的树林和新的牧场”，或者在太阳落山时，在费尔黑文山上用胡瓜子和蓝莓做晚餐，积攒几天的食物。这些果实并不向购买者提供其真正的味道，也不向为市场饲养它们的人提供。只有一种方法可以获得它，但很少有人采取这种方法。如果你想知道胡瓜子的味道，请问牛仔或鹧鸪。假设你品尝过从未采摘过的胡克里果，那是一个粗俗的错误。哈克贝利从未到达过波士顿；自从它们在她的三座山上生长以来，那里就不为人所知。果实的芳香和基本部分随着在市场上被擦掉的花朵而消失，它们变成了单纯的食物。只要永恒的正义还在，就没有一个无辜的胡瓜果能从这个国家的山上被运到那里。(obtained October 29th, 2021)

Appendix D 任务纲要

Tasks Brief

Source language: English

Target language: Chinese

Text type: Literary text

Text source: *Walden* (Thoreau, 1999)

Target Audience: General readers in China

Publishable quality is expected.

Where and how will the translated text be used?

This text will be read by the Chinese audiences who want to know the content in *Walden* (ibid).

What is the desired response from readers/listeners?

Learn more about *Walden* (ibid).

Tone of Voice / Image / Style of Address to the Reader: Close to original text.

Note: You have access to external resources but existing translations during the task.

30 minutes are imposed for each task.

Appendix E 翻译修订指南

General principles on revision

(European Commission, 2010; Mossop, 2007)

1. If you cannot understand the translation without reading it twice, or without consulting the source text, then a correction is necessary.
2. Do not ask whether a sentence can be improved but whether it needs to be improved.
3. Do not retranslate. Do not retranslate. Do not retranslate.
4. Minimize [the] introduction of error by not making changes if in doubt about whether to do so.
5. When you make a linguistic correction or stylistic improvement, make sure you have not introduced a mistranslation.
6. When you make a change, check whether this necessitates a change elsewhere in the sentence or a neighboring sentence.
7. Check numbers as well as words – they are part of the message.
8. Do not make changes you cannot justify.
9. Do not impose your own approach to translating on others.
10. Do not impose your linguistic idiosyncrasies on others.
11. Remember the purpose of revision: correct any errors and improve the readability so that the reader of the translation will receive the intended message.

Appendix F 译后编辑指南

TAUS MT Post-Editing Guidelines (2016)

1. Aim for grammatically, syntactically and semantically correct translation.
2. Ensure that key terminology is correctly translated and that untranslated terms belong to the client's list of "Do Not Translate" terms.
3. Ensure that no information has been accidentally added or omitted.
4. Edit any offensive, inappropriate or culturally unacceptable content.
5. Use as much of the raw MT output as possible.
6. Basic rules regarding spelling, punctuation and hyphenation apply.
7. Ensure that formatting is correct.

Appendix G 实验知情同意书

一、研究背景介绍

同学，您好！我们系湖南大学外国语学院机器翻译译后编辑研究团队，现诚挚地邀请您参与一项翻译修订和机器翻译译后编辑对比研究。该研究旨在探究学生译者在翻译修订和机器翻译译后编辑过程中的差异。如果您同意参与，请仔细阅读下列说明。如有任何疑问，请随时提出。

二、研究目的与内容

本研究通过探索学生译者在翻译修订和机器翻译译后编辑过程中的差异，需要您配合完成以下任务：

- 1、填写调查问卷；
- 2、完成翻译修订和机器翻译译后编辑实验；
- 3、填写回溯性报告。

三、研究意义

本研究旨在对比学生译者在翻译修订和机器翻译译后编辑过程中的差异，为翻译修订和机器翻译译后编辑对比研究提供一定的启示，有助于学生译者的能力培养，帮助学生译者快速适应翻译市场的需求。此外，该实验将有助您更加全面地认识翻译修订和机器翻译译后编辑这两种模式。

四、研究风险与不适

本实验将在学校图书馆的研讨空间进行，所选地点安静舒适，尽可能避免让被试产生任何生理心理上的风险和不适。

五、隐私问题

本研究承诺将对所有实验被试的个人信息进行保密。发表该项研究结果时会采用编号的方式对被试进行标识，不透露被试的具体身份。

六、报酬

按照要求顺利完成该项实验的被试将会获得一份价值¥40的礼品。

七、自由退出

参与实验的被试可以随时了解与研究有关的信息和进展，自行决定参与（或不参与）本实验。被试有权在实验过程中的任一阶段向研究者退出本实验，中途退出实验的被试信息及数据将被立即销毁，不再纳入本研究的后续数据分析之中。

八、联系方式

如您对本研究感兴趣，请随时与研究者 XXX 联系。联系方式如下：

电话：199XXXXXXXX 邮箱：xyz115XXXXXXXX@163.com

九、知情同意及签字

我已经阅读该实验知情同意书，清楚该实验的目的、内容、风险和受益等基本情况，并自愿参与本项实验。

被试签名：

手机号码：

签字日期：

Appendix H 实验操作步骤

感谢您参与本次实验！您的个人实验数据（包括译文、录屏和键盘记录数据等）都将在研究中以匿名的形式呈现。本实验旨在体现真实的翻译修订和机器翻译译后编辑过程，对实验结果不做任何价值判断。在熟悉操作步骤后，请按平时的状态开展修订和译后编辑任务。

一、实验内容

- 1、填写调查问卷；
- 2、完成翻译修订和机器翻译译后编辑任务；
- 3、填写回溯性报告。

二、实验工具

- 1、Translog II 键盘记录软件
- 2、EVCapture 录屏软件

三、实验步骤

- 1、阅读任务纲要、翻译修订和译后编辑指南；
- 2、下载并安装 Translog II 软件，利用热身文本熟悉该键盘记录软件，具体流程包括导入实验文本、开始记录、切屏进行网络检索、终止记录、保存文件等；
- 3、向工作人员示意可以开始实验，打开 EVCapture 录屏软件，导入相应文件，开始翻译修订和机器翻译译后编辑任务，允许查找网络资源，但不允许参考现有译文。各个任务分别建议用时为 30 分钟。完成单个任务后请点击终止记录并保存键盘记录数据；
- 4、完成两个任务后请终止录屏，保存文件并根据各自体验撰写回溯性报告。

四、数据收集

2 份译文、2 份键盘记录数据、1 份录屏软件和 1 份回溯性报告。

五、注意事项

- 1、实验开始前，请将手机调至静音；
- 2、实验过程中，请保持安静并按照规定要求完成实验操作；
- 3、实验结束后，如有任何问题，请及时联系工作人员。

感谢您的支持与配合！

Appendix I 任务难度及认知负荷主观评分

Please scale the overall difficulty you felt when completing the tasks (0 for the least difficult, and 10 for the most difficult) and tick translation revision and NMT post-editing according to the actual task condition. Then scale the four dimensions in Adapted NASA Task Load Index.

Part I Basic information

Experimental ID _____

Which text

•Text 1 •Text 2

Which task

•Translation revision •NMT post-editing

Part II Subjective scoring of task difficulty and cognitive effort

1. Mental demand: How mentally demanding was the task?										
0 Low	1	2	3	4	5	6	7	8	9	10 High
2. Effort: How hard did you have to work to accomplish your level of performance?										
0 Low	1	2	3	4	5	6	7	8	9	10 High
3. Frustration: How insecure, discouraged, imitated, stressed, and annoyed were you?										
0 Low	1	2	3	4	5	6	7	8	9	10 High
4. Performance: How successful were you in accomplishing what you were asked to do?										
0 Poor	1	2	3	4	5	6	7	8	9	10 Good

Appendix J 详细中文摘要

翻译修订和译后编辑是提升翻译质量常见的两种手段,二者在当前翻译市场中的比重愈来愈大,这要求译者必须具备一定的翻译修订和译后编辑能力。但是从整体来看,学生译者和职业译者的修订和译后编辑能力都有提升的空间。过度编辑的现象在这两种任务类型下屡见不鲜,过度编辑便是译者无法很好驾驭修订和译后编辑能力的表现之一。

迫切的市场需求催促我们开展有关修订和译后编辑的研究。但目前有关二者的对比研究数量较少,尤其是相关的实证研究。一些学者在前人的基础上分别提出了有关翻译修订和译后编辑的能力与模型,但是相关的实证研究略显单薄,这些理论层面的论点亟需实证数据的论证。

与此同时,翻译机器的快速发展也为进一步拓展所研究的文本类型奠定了基础。以往的对比研究以非文学文本为主,基于文学文本的对比研究较少,其结果如何仍有待探究。

因此,本文拟从文学文本着手,综合运用调查问卷、击键记录和回溯性报告等研究方法,一方面探究学生译者在修订和译后编辑当中的行为差异,另一方面尝试探索机器翻译译后编辑在文学翻译当中的潜力。本研究拟回答以下问题:(1)被试在修订和译后编辑文学文本时如何修改译文?(2)与修订相比,被试在译后编辑文学文本时会投入多少译后编辑努力?具体而言,被试在译后编辑文学文本时是否会投入更多的时间努力、技术努力和认知努力?(3)与修订的文本相比,译后编辑的文学文本的翻译质量如何?具体而言,两类译文的充分性和流利度如何?

本研究首先开展了背景调查,收集了48名翻译专业研究生的问卷反馈。受访者均通过英语专业八级考试,具备一定的翻译能力。受访者们均参加过“翻译技术实务”和“文学翻译”两门课程,因此掌握了有关翻译修订、译后编辑和文学翻译的基本知识。基于受访者的反馈和预实验,研究者完善和改进了实验内容。随后,研究者从48名受访者中随机要求21名参与正式的翻译修订和机器翻译译后编辑实验。

本研究的实验材料包括调查问卷、实验文本、翻译纲要、翻译修订和译后编辑指南。调查问卷包含四个部分,旨在了解被试的基本情况以及被试对翻译修订

和机器翻译译后编辑的态度,为后续的结果分析与讨论提供相应的解释。本实验的源语文本均取自于梭罗的《瓦尔登湖》,两篇文本字数相当(分别是182和190个英文单词),翻译方向为英译汉,文本类型为文学文本。选择《瓦尔登湖》的缘由如下:首先,这是一个真实的翻译项目,能体现一定的市场需求。其次,《瓦尔登湖》是一本文学价值很高的作品,目前有很多中文译本。多个译本和翻译前辈的支持意味着本次实验能够找到很好的译文作为翻译质量评估的参考。再次,对比学生译者在修订和译后编辑文学文本时的行为差异可以为相关的理论研究提供实证数据支撑,并有助于学生译者能力和批判精神的培养。为了保证研究结果的有效性,本文利用多个指征对源语文本进行检验,以保证两个文本具有相似的复杂度从而控制翻译难度。同时,本研究对人工译文和机器译文的错误进行了标注。从译文错误来看,机器翻译的译文中的错误数量将近人工译文中的一倍,主要体现在误译和译文晦涩两个方面。这说明人工翻译的文学译文比机器翻译更为准确、更为流利,但人工译文中的漏译和增译问题却比机器译文多,这符合人工译文的特点。译者在不理解原文内容时通常采用这类讨巧的办法规避问题,以保证译文的流利度。而机器译文中不存在增译现象,但本研究发现翻译机器在处理个别英文长难句时会出现漏译的问题。这说明当前的神经翻译系统仍有不足。本研究最终选取神经机器翻译系统 DeepL 的译文作为译后编辑任务中的源语文本。经过对比,最终选取文本一的机器译文和文本二的人工译文分别作为译后编辑和修订的实验文本。本研究所使用的工具以击键记录软件 Translog II 和录屏软件 EVCapture 为主。

正式实验开始之前,所有被试签署了知情同意书。完成热身任务之后,被试开始翻译修订和机器翻译后编辑实验。任务顺序完全随机。翻译过程中,被试可以查阅外部资源,并不可以查看网上已有的译文。每项任务建议用时30分钟。被试每完成一项任务都需要对该项任务的难度及认知负荷进行主观打分。两项任务完成之后,被试需要保存译文、键盘记录和录屏文件,并撰写回溯性报告。实验结束后,每位参与实验的被试都会获得一份价值40元的礼物作为报酬。

在数据检查过程中,研究者发现有一名学生译者重复填写了调查问卷。经核实和沟通,研究者剔除了其中一份问卷。另外,在对比录屏数据和键盘记录数据时,研究者发现被试 P06 在译后编辑中的击键用时明显少于录屏所记录的时间。经核实,被试 P06 在实验中无法正常保存键盘记录数据,为弥补实验数据,自行开展了二次实验。剔除上述数据之后,本研究共收集了48份问卷、40份译文、

40 份键盘记录文件、40 份录屏文件、40 份任务难度及认知负荷主观评分及 40 份回溯性报告。

通过 SPSS 26.0 和 RStudio 的分析,研究者凭借 Prism 9.3 软件绘制了较为清晰直观的图表。研究结果显示:(1)从修改行为来看,学生译者的译后编辑次数将近修订次数的 1.5 倍。具体来看,学生译者在译后编辑机器译文时的错误修改正确率为 27.10%。这一数值虽然不高,却远超翻译修订中的情况。学生译者在修订人工译文时的错误修改正确率只有 8.52%,遗漏了 75.39%的译文错误。这一方面说明人工译文较高的流利度会使其中错误更为隐秘,另一方面也反映了学生译者不擅长发现人工译文中的错误,学生译者相关的修订能力有待提升。(2)从译后编辑努力来看,学生译者在译后编辑机器译文时投入了更多的时间、技术和认知努力。具体来看,学生译者在修订中的起草用时明显多于译后编辑中的起草用时,但修订任务的总用时却显著少于译后编辑任务。其次,学生译者在修订人工译文时的击键总次数、增加和删除总次数均显著少于译后编辑机器译文中的情况。再次,停顿数据显示学生译者在译后编辑中投入的认知努力显著多于修订。基于学生译者对任务难度和认知负荷的主观评估,本研究发现学生译者认为自己在翻译修订中投入的努力显著低于机器翻译译后编辑,但是学生译者对自己在两项任务中的表现却给出了相近的预期。上述结果表明,学生译者更信任人工译文的质量,特别是当人工译文具备不错的流利度时。这会降低学生译者在修订过程中的警觉,导致其遗漏了很多严重的译文错误。相反,学生译者对机器译文的不信任有助于他们发现更多的机器译文错误。这也间接表明学生译者在修订和译后编辑能力上存在差异,翻译教学应当重视对这两种能力的培养。(3)从翻译质量来看,修订和译后编辑之后的文学译文在准确度和流利度上不存在显著差异。同一批被试在不同的任务类型得到了相近的翻译质量,这反映了文本的翻译质量很大程度上受制于学生译者自身的翻译能力与水平。在本研究中,机器翻译的使用对最终翻译质量影响的不够明显。换言之,机器翻译只是给译者提供参考和辅助,并不会主导整个翻译过程,而译者的译后编辑活动才是决定翻译质量的首要因素。此外,虽然修订和译后编辑之后的翻译质量比较接近,但是这并不意味着最终的译文质量达到了“可以出版”的水平。两种任务类型下的译文都必须进行反复的修改,才能真正满足广大读者的需求。这与翻译文学作品的市场要求和流程相吻合,一部好的译著往往凝结了译者、审校者和出版社的良苦用心,经年累月才得以现世。本研究为进一步将机器翻译译后编辑与文学翻译相结合提供了

实证支持。虽然机器译文缺乏创造性，但是融入了译者的译后编辑活动之后，这样的文学译文在质量上并不会比翻译修订的结果逊色多少。同时，值得注意的是质量好的译文总是离不开反复打磨与修改。

本文将为翻译修订和机器翻译译后编辑领域的实证研究提供方法论参考。本次研究发现将为翻译教学及译者修订和译后编辑能力的培养提供有益启示。

关键词：翻译修订；神经机器翻译译后编辑；修改行为；译后编辑努力；翻译质量；文学文本