Study on the Impact Factors of the Translators' Post-editing Efficiency in a Collaborative Translation Environment

Na Ye
Knowledge Engineering Research Center
Shenyang Aerospace University
Shenyang 110136, China
yn.yena@gmail.com

Guiping Zhang
Knowledge Engineering Research Center
Shenyang Aerospace University
Shenyang 110136, China
zgp@ge-soft.com

Abstract

This paper investigates the factors that influence the translators' post-editing efficiency in a collaborative translation environment. We developed GE-CCT, a platform that enables hundreds of translators to be organized to complete one large-scale translation task together. On the basis of the platform, this paper made thorough analysis on the post-editing logs of 325 translators within 8 weeks. Results show that some exterior environmental factors and interior personal factors obviously affect the translators' post-editing speed and quality. Accordingly, we proposed solutions that improved the collaborative translation efficiency.

1 Introduction

Since the translation results of fully-automatic machine translation still cannot meet the requirements of practical applications, recently computer-aided translation (CAT) becomes a hot research topic. In CAT systems, the human translator is provided with an aiding translation of the input sentence and performs post-editing on it under the assistance of plenty of tools (such as dictionary lookup, translation memory retrieval, term recognition, document format processing, etc) until the correct translation is acquired.

With the rapid enlargement of the scale of CAT, multiple spatially distributed translators be organized to complete one translation task together becomes a common phenomenon. This new translation mode can be called collaborative translation (CT). Under the CT environment, each translator works on a client and the CT system is located on a central server to provide assistance following the translators' request commands. The CT system can not only offer all the functions of the CAT system but also other functions such as the management of the translation tasks and the consistency checking of the translation results in order to help the translators cooperate efficiently.

Current research on collaborative translation mainly focuses on how to learn translation knowledge automatically from the post-editing feedbacks to improve the quality of the aiding translation (Llitjós and Carbonell, 2006; Simard et al., 2007; Groves and Schmidtke, 2009; Ortiz-Martinez et al., 2010) and how to design better collaboration strategy (Murata et al., 2003; Bey et al., 2006; Morita and Ishida, 2009; Bederson et al., 2010). However, other than the technical improvements on the system, we find in practice that there are many factors which have great impacts on the translators' post-editing psychology and thus influence their efficiency.

In this paper we developed the GE-CCT collaborative translation platform, on the basis of which we collected the post-editing logs of 325 translators working collaboratively within 8 weeks in a large-scale Chinese-English patent translation project. Then we made thorough investigation on the logs to discover the factors that affect the translators' post-editing efficiency (both speed and

quality). Results show that the exterior environmental impact factors include festival, weekday and the use of automatic proof-reading tool, and the interior personal impact factors include the translator's platform usage time, major and English level. This paper also analyzed how some of the factors play the role on the translators' psychology and proposed solutions that increased the collaborative translation efficiency.

2 Related Work

Martínez (2003) pointed out some impact factors of the translators' post-editing efficiency, including whether the assistant tools of the CAT system are effective, whether the interface is friendly, whether the translator follows reasonable process for post-editing, whether the data are well analyzed before translation, whether a user dictionary is constructed, whether the aiding translation is credible, etc. Aikawa et al. (2007) validated the assumption that using controlled language can improve post-editing efficiency. Doyon et al. (2008) found some other factors including the quality of the aiding translation, the translator's experience and the translator's native language. Koehn (2009a) proved that certain types of assistance including suggestions of sentence completion, word and phrase translation options and machine translation output help to improve the translators' post-editing speed and quality. The above studies are mainly from the view of improving the CAT system and paid little attention to the translator himself. In fact, translators play a central role in post-editing, and their translation psychologies and processes need to be better understood before their capabilities can be brought into full play.

Some researchers investigated the psychological process of translators during post-editing. Because the data at cognitive level cannot be observed directly, think-aloud protocol (Kussmaul and Tirkkonen-Condit, 1995) or choice network analysis (Campbell, 2000) is applied (Krings, 2001; O'Brien, 2005). However, quantified results still can hardly be obtained. Koehn (2009a) believes that the pauses of keystrokes and mouse clicks are reflections of cognitive process, and used the data recorded by the Caitra system (Koehn 2009b) to discover the cognitive meaning of pauses with different lengths. Carl et al. (2010) analyzed the translators' operating data and found that the translation process can be clearly divided into three phases, namely gisting, drafting and post-editing. Student translators have longer gisting phases whereas professional translators have longer post-editing phases. These research fruits help people to better recognize the translator's psychological post-editing process, but they didn't investigate the relationships between the translators' psychology and other factors, which is the focus of this paper and will provide valuable clues for improving the translators' post-editing efficiency. Besides, we conducted analysis in a collaborative translation environment and revealed some psychological characteristics different from those in the CAT environment.

3 User Study

3.1 GE-CCT Collaborative Translation Platform

The GE-CCT platform is implemented as a web-based client-server architecture. It consists of four subsystems as follows:

The collaborative translation subsystem offers an interactive Chinese-English post-editing environment (see Figure 1) to the translators and it contains many aiding tools such as automatic aiding translation generation, assistant input, assistant typesetting, translation memory (TM) retrieval, etc.

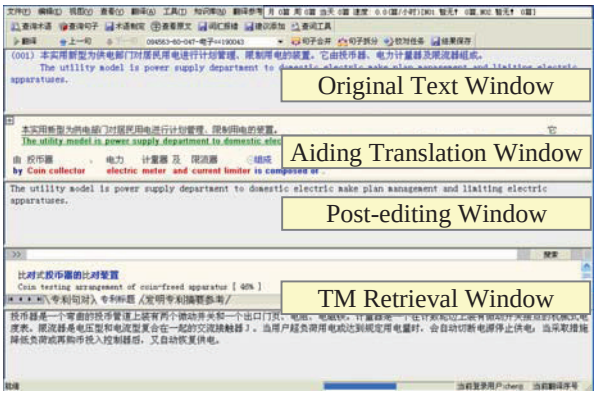


Figure 1. Post-editing Interface of the GE-CCT Platform.

The quality checking subsystem is developed for translation quality controlling and contains five modules including automatic translation sampling, error labeling, consistency checking, quality score computing and comment generating.

The task management subsystem performs the importing, exporting, preprocessing, clustering, assigning and monitoring of the translation tasks.

The translation knowledge management subsystem realizes the automatic learning, updating, accumulating and sharing of translation knowledge in the whole translation process.

The overall framework of the platform is illustrated in Figure 2.

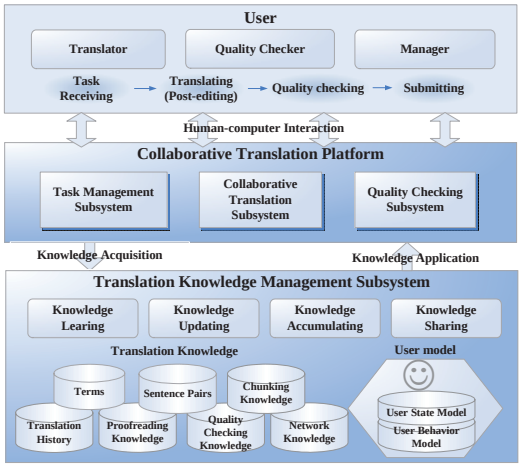


Figure 2. Overall Framework of the GE-CCT Platform.

With the platform, we tracked the daily post-editing amount and quality of each translator, which then allows for a detailed analysis of the translators' efficiency. The personal information of each translator and the activities of the whole group are also recorded in detail.

3.2 Background and Settings

In our study, all of the 325 recruited translators are native speakers of Chinese with English skills at university-level or higher level. 160 of them work full time every day, and the others work half of the day. They were rewarded according to their post-editing amount and quality so as to give them an incentive to be productive.

The translators are required to work on a large-scale technical document translation project. The total amount of translation task is 200 million Chinese characters. These documents come from the Chinese patents of recent years provided by the State Intellectual Property Office (SIPO) of the PRC. It took about 11 months to accomplish the whole task (error rate lower than 1.5‰) and the user data within 8 weeks are recorded for our study.

In order to assign the translation tasks justly, each document is broken up to several blocks with a fixed length of 220 Chinese characters. Punctuations, spaces, digitals and English characters are excluded while computing the lengths. Such block is called the standard work-piece (SWP), and it was taken as the basic unit in the assignment of the translation tasks and the evaluation of the translators' post-editing efficiency.

3.3 Evaluation

In this paper we use the number of SWPs translated by each translator per day to evaluate the translator's daily post-editing speed.

Speed is not the only criterion of success. The translations also have to be correct. We rely on human judges to check the quality of translation results. Due to the large translation amount, 10 percent of examples are randomly selected from the translation result of each SWP and transferred to the human judges to check the errors. Then the system computes the quality score *s* as follows:

$$s = MAX - \sum_{i}^{N} w(t_i) \cdot n(t_i) \qquad (1)$$

where $w(t_i)$ refers to the weight (from 0 to 1) of error type t_i, $n(t_i)$ refers to the error number with the same error type t_i, and *MAX* refers to the full mark which is set to 5. We defined 7 types of error including missing translation, redundant translation, grammatical error, prohibited usage, formatting error, logical error and terminology error.

If a translated SWP gets a quality score lower than 4.5, then it will be returned to the corresponding translator for revising, otherwise it will be accepted. Only those SWPs that have passed the quality checking process will be included in the evaluation of post-editing speed.

4 Results and Analysis

4.1 Speed and Quality

As mentioned above, the daily post-editing speed of each single translator is evaluated by the number of SWPs that he/she successfully accomplishes every day, and his daily post-editing quality is

evaluated by the average quality score of all the translated SWPs every day. Figure 3 and Figure 4 show the average daily speed (number of SWPs per translator per day) and average daily quality (quality score per translator per day) of all the translators within the 8 weeks.

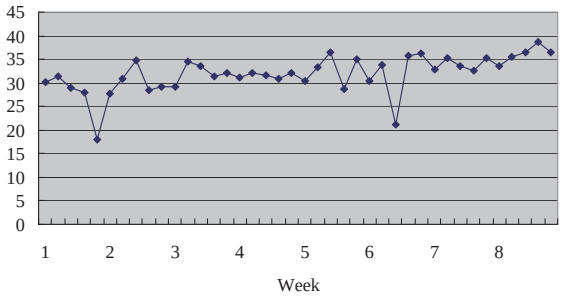


Figure 3. Average Daily Post-editing Speed.

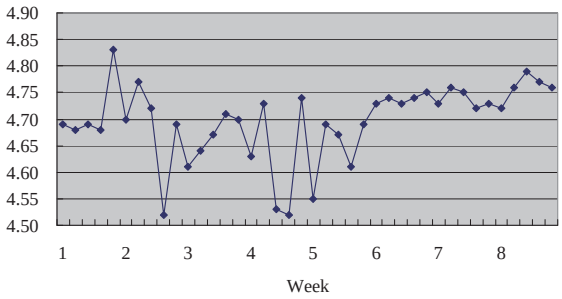


Figure 4. Average Daily Post-editing Quality.

In our study, only working days (from Monday to Friday) are included for the statistics because on weekends the translators are not required to work. It can be seen that the translators' average post-editing speed ranges from 17.9 SWPs to 38.8 SWPs per day (the two great drops correspond to two festivals), and their average daily quality score ranges from 4.52 to 4.83. In fact, the daily speed and quality of different translators vary greatly, and we only focus on the average values in the above graphs. Table 1 shows the average post-editing speed and quality per week.

Week	1	2	3	4	5	6	7	8
Speed	27.3	30.2	32.2	31.6	32.8	31.5	33.9	36.1
Quality	4.71	4.68	4.67	4.63	4.64	4.74	4.74	4.76

Table 1. Average Post-editing Efficiency per Week.

On the whole, the translators work faster with the time, and their post-editing quality is also better.

4.2 Environmental Impact Factors

4.2.1 Festival

There were 4 festivals (including 2 Chinese traditional festivals and 2 Western festivals) during the 8 weeks. On these days, translators are still required to work but may be influenced in the psychology. We compared the average daily speed and quality during the festivals and non-festivals. Table 2 gives the results.

Date	Festivals	Non-festivals
Speed	27.5	32.5
Quality	4.66	4.70

Table 2. Comparative Results of the Post-editing Efficiency on Festivals and Non-festivals.

From Table 2 we can see that the translators' post-editing speed greatly decreases by 18.2% on festivals while the quality score doesn't change obviously. To validate the generality of the observation, we compared the efficiency of each festival with its immediately previous day. Results show that the daily speeds on all the four festivals and the quality scores on two festivals are lower.

We computed the CHI-square statistic (Yang et al., 1997) between the festivals and the change of speed/quality for checking the correlation hypothesis. The significance level is set to 0.05. Results show that there is strong relationship between festivals and the speed variation, indicating that translators tend to become slower on festivals.

4.2.2 Weekday

To answer whether there is any relationship with the day of week and the post-editing efficiency, the average speed and quality score of each weekday are computed, as shown in Table 3 (the 4 festivals are excluded from the statistics).

WeekDay	Mon	Tue	Wed	Thu	Fri
Speed	30.7	33.4	34.4	32.5	31.8
Quality	4.67	4.72	4.69	4.67	4.74

Table 3. Average Post-editing Efficiency per Weekday.

It can be seen that the translators' speed and post-editing quality both tend to increase on Tuesdays and decrease on Thursdays.

We also compared the efficiency of each weekday with its immediately previous day (the 4 festivals and their next days are excluded from the statistics). Table 4 gives the comparative results.

WeekDay	Mon	Tue	Wed	Thu	Fri
Speed (+/-)	-	8/0	3/3	1/3	4/2
Quality (+/-)	-	7/1	2/4	0/4	5/1

Table 4. Comparative Results of the Post-editing Efficiency between each Weekday and its Previous Day.

In Table 4, the two numbers in the cells are the number of days with higher/lower efficiency than its previous day. We computed the CHI-square statistic between the weekdays and the change of speed/quality at a significance level of 0.05. Results show that Tuesday and Thursday have strong relationships with the post-editing efficiency variation. On Tuesdays translators tend to be faster in speed and better in quality, while on Thursdays they tend to be worse in quality.

4.2.3 Use of Automatic Proof-reading Tool

In this section we discuss the effect of an automatic proof-reading tool.

The proof-reading tool works after the translator accomplishes the current SWP. It automatically checks the problems that usually occur in the translation results, including word capitalizations, articles, punctuations and some usages prohibited by the post-editing rules. Whenever an error is detected, the tool labels the corresponding part with different colors to warn the translators. The errors are detected through a rule matching strategy in which the rules are expressed with regular expressions and written manually.

This tool was put into use since Monday of the 3rd week. We had expected the translators' post-editing quality to be improved with the tool. However, from Table 1 we can see that the average quality score didn't change obviously. Interestingly, the average speed increased.

Through discussing with the translators and observing their behaviors, the reason was found. In fact, translators are usually nervous before they submit the translation results. They will check the results over and over again to avoid getting a low score in the following quality checking process. The utility of the proof-reading tool made them more confident with their translations, especially when the translations have passed the examination of the tool. Then the translators will submit the results without checking it too many times and thus became faster.

We also evaluated the average ratio of the number of SWPs that needed revising to the total number of SWPs received in every week. The results are shown in Table 5.

Week	1	2	3	4	5	6	7	8
Ratio	0.67	0.70	0.49	0.50	0.48	0.46	0.38	0.43

Table 5. Average Revising Ratio of SWPs per Week.

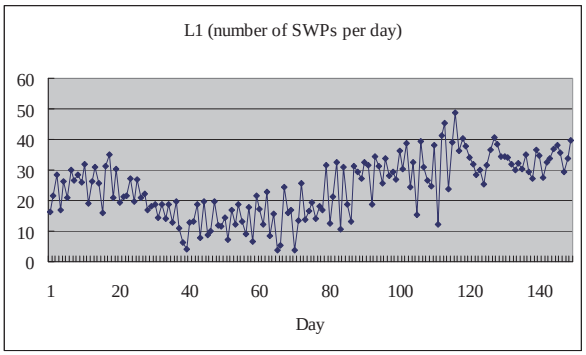
It can be seen that the revising ratio greatly decreases by 30% since the 3rd week, which had also led to the improvement of post-editing speed. All these validate the observation that the proof-reading tool helps the translators to be faster.

4.3 Personal Impact Factors

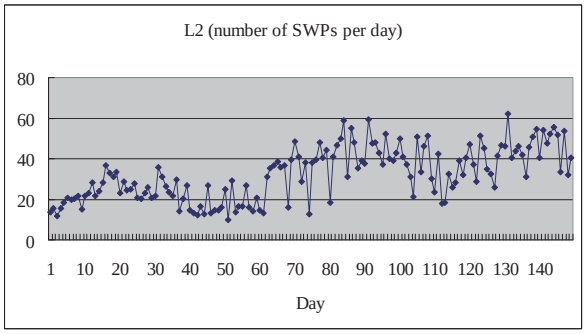
Other than the exterior environmental factors, the post-editing efficiency is also related with some personal factors of the translator himself.

4.3.1 Platform Usage Time

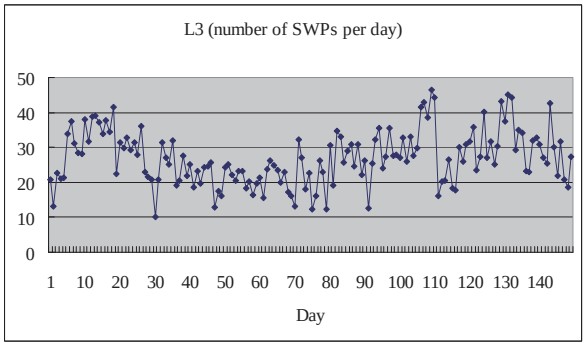
Another important question that we are trying to answer is: What kind of variation in efficiency will the translators produce after they start to work in the collaborative translation environment. This has consequences for the design of the translation aids, since we want to alleviate the most troubling aspects of the post-editing process to increase the translators' productivity. We randomly selected 3 full-time translators and recorded their daily speeds within 150 days since their enrollment dates (see Figure 5).



(Figure 5a) Translator L1



(Figure 5b) Translator L2



(Figure 5c) Translator L3

Figure 5. Daily Post-editing Speed of 3 Translators.

According to the graph, all translators have a similar tendency in the variation of daily speed. In the first a few days (about 10 to 20 days) after enrollment date, the speed increases continuously. Then the translators come into a relatively long period in which they become slower before the speed recovers in the end. The length of the special period ranges from about 60 days to 90 days on different translators.

To explain the reason for the speed recession period that the translators encounter, we made widespread discussions and investigations with them. Results show that the translators are excited with the post-editing job at the beginning because this is a new translation style for them and they find the tools are effective. After these days, the translators' interests weaken and meanwhile they fall into the trouble of quality. Everyday they receive a lot of SWPs returned for revising from the quality checkers. They also need to learn more about the advanced functions of the platform so as to keep up their productivity. All these have led to the inevitably troubling period. In fact, during this time many translators will lose confidence and choose to leave the job.

4.3.2 Major

In our study the translators' majors are diversified. Since the project aims at translating patents, most of the documents are related with technologies of different domains. Therefore, a translator's knowledge of a specialized field will naturally promote his post-editing efficiency.

Table 6 shows the distribution of majors among all the translators. We also selected 54 best translators and computed the distribution among them.

Major	English	Others
All	0.60	0.40
Best	0.64	0.36

Table 6. Distribution of Translators with English Major.

From Table 6 we can see that there is a large portion of translators whose majors are not English. This has partly validated that the specialty knowledge is quite helpful for post-editing. However the portion is relatively lower among the best translators, indicating that the linguistic knowledge plays a more important role if one hopes to become excellent in post-editing.

4.3.3 English Level

In this study we classify the translators' English skills into 4 levels: Level 1, 2, 3 and 4. The classification is based on the level of the National English qualification tests that the translators have passed. Level 1 means the translator has passed the highest qualification test and level 2 means the second highest qualification test, etc. Table 7 gives the distribution of English levels among all the translators and the 54 best translators.

English Level	Level 1	Level 2	Level 3	Level 4
All	0.15	0.55	0.22	0.08
Best	0.33	0.31	0.31	0.05

Table 7. Distribution of Translators' English Levels.

Among all the translators, those at Level 2 occupy the majority; while among the best translators, those at Level 1 are the most. And in both groups, the ratios of translators at Level 4 are very low. The data shows that one may have to be profound in English skills in order to be qualified in post-editing, and he needs to be more skillful in English in order to be excellent.

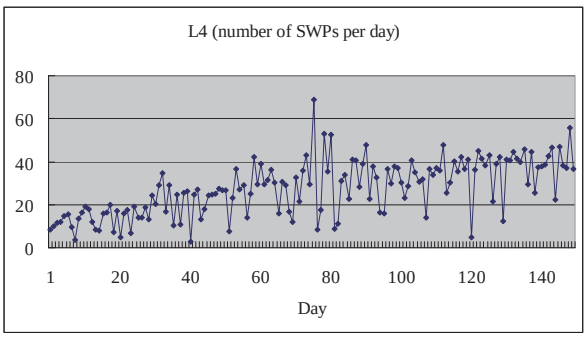
4.4 Solutions

Regarding the aforementioned impact factors of the translators' post-editing efficiency, we proposed some strategies to help translators avoid or reduce the extent of being influenced.

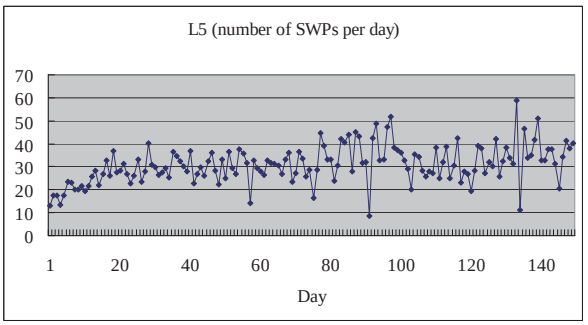
Firstly, focusing on the impact of festivals and weekdays, we adopted an alerting strategy. On the days when the translators are predicted to be less efficient, the CT platform will automatically alert the translators in the morning. In addition, we arrange regular meetings on those days and encourage the translators to be productive. The strategy was proved to be effective in practice.

Secondly, to solve the speed recession problem in section 4.3.1, we proposed an enrollment training strategy. For each new recruit, two quality checkers with rich post-editing experiences will be assigned to give him/her advices on both the post-editing techniques and psychology. The advices are offered in multiple styles such as lectures, panel discussions, question-answering, etc. The training is classified into three levels. The junior-level training focuses on teaching the basic post-editing techniques, including instructions on the usage of the platform's basic functions and post-editing requirements. The intermediate-level training focuses on enhancing the translators' post-editing quality, including post-editing experience exchanging and problem discussions. The senior-level training focuses on helping the translators to maintain the high quality and increase the speed, including instructions on the advanced post-editing techniques and complaint handling.

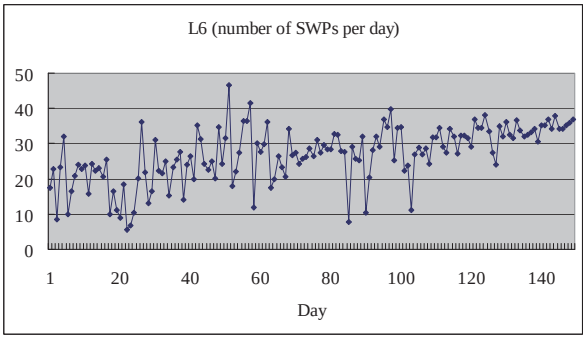
To prove the effectiveness of the training strategy, we recorded the daily speeds of 3 full-time translators who had participated in the training since their enrollment dates (see Figure 6).



(Figure 6a) Translator L4



(Figure 6b) Translator L5



(Figure 6c) Translator L6

Figure 6. Daily Post-editing Speed of 3 Translators with Enrollment Training.

In Figure 6a and 6b, the translator's post-editing speed continuously increased or remained stable with the time. This shows that the training has helped them avoid or reduce the influences of the difficulties in the speed recession period. In Figure 6c, the translator's recession period shortened greatly. At about 30 days after his enrollment date, the translator has grasped the required knowledge points and adapted to the stress from the quality checking process.

Questionnaires among translators also show that the training strategy is useful and helps to keep the people. In the collaborative translation environment, it is desirable for translators to communicate frequently about their common problems and more desirable if an experienced translator gives them suggestions before they encounter the problems.

5 Conclusion and Future Work

In this paper we conducted study on the post-editing efficiency of translators in a collaborative translation environment. Some factors are found to have significant impact on the post-editing efficiency. The main conclusions are:

(1) Festivals have impact on the post-editing speed. The translators are obviously slower on festivals.

(2) There are some relationships between weekday and the efficiency. On Tuesdays the translators tend to be higher in both speed and quality, while on Thursdays their post-editing qualities tend to be lower.

(3) The usage of automatic proof-reading tool helps to increase the post-editing speed by enhancing the translators' confidence on their translation results.

(4) The translators will come into a long period of speed recession since a few days after they start to use the collaborative translation platform. The period can be eliminated or shortened through the adoption of an enrollment training strategy.

(5) Although translators may benefit from the knowledge of a specialized field, the best translators generally need better English skills.

Further study of the post-editing efficiency are needed both to gain insight into what the most time-consuming post-editing activities are and how they can be alleviated. We are also interested in the study of providing personalized aiding to different translators according to their individual post-editing characteristics.

References

Aikawa T, Schwartz L, King R, et al. 2007. Impact of Controlled Language on Translation Quality and Post-editing in a Statistical Machine Translation Environment. *Proceedings of MT Summit XI*: 10-14.

Bederson BB, Hu C, and Resnik P. 2010. Translation by Iterative Collaboration between Monolingual Users. *Proceedings of Graphics Interface (GI) Conference.*

Bey Y, Boitet C and Kageura K. 2006. The TRANSBey Prototype: An Online Collaborative Wiki-Based CAT Environment for Volunteer Translators. *Proceedings of the 3rd International Workshop on Language Resources for Translation Work, Research & Training*: 49-54.

Campbell S. 2000. Choice Network Analysis in Translation Research. *Intercultural Faultlines: Research Models in Translation Studies: Textual and Cognitive Aspects*: 29-42.

Carl M, Kay M, Jensen K. 2010. Long-distance Revisions in Drafting and Post-editing. *Proceedings of CiCling 2010*: 193-204.

Doyon J, Doran C, Means D, et al. 2008. Automated Machine Translation Improvement Through Post-Editing Techniques: Analyst and Translator Experiments. *Proceedings of the 8th AMTA*: 346-353.

Groves D, Schmidtke D. 2009. Identification and Analysis of Post-Editing Patterns for MT. *Proceedings of MT Summit XII.*

Llitjós AF, Carbonell JG. 2006. Automating Post-Editing to Improve MT Systems. *Proceedings of the AMTA Automated Post-Editing Techniques and Applications Workshop.*

Koehn P, Haddow B. 2009a. Interactive Assistance to Human Translators using Statistical Machine Translation Methods. *Proceedings of MT Summit XII.*

Koehn P. 2009b. A Web-based Interactive Computer Aided Translation Tool. *Proceedings of the ACL-IJCNLP 2009 Software Demonstrations*: 17-20.

Krings HP. 2001. Repairing Texts: Empirical Investigations of Machine Translation Post-editing Processes. *Kent State University Press*: Koby GS.

Kussmaul P, Tirkkonen-Condit S. 1995. Think-Aloud Protocol Analysis in Translation Studies. *TTR*, 8(1): 177-199.

Martínez LG. 2003. Human Translation versus Machine Translation and Full Post-Editing of Raw Machine Translation Output. *Dublin City University*.

Morita D, Ishida T. 2009. Collaborative Translation by Monolinguals with Machine Translators. *Proceedings of the 13th International Conference on Intelligent User Interfaces (IUI)*: 361-366.

Murata T, Kitamura M, Fukui T, et al. 2003. Implementation of Collaborative Translation Environment 'Yakushite Net'. *Proceedings of MT Summit IX*.

O'Brien S. 2005. Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Text Translatability. *Machine Translation*, 19(1): 37-58.

Ortiz-Martinez D, Garcia-Varea I, Casacuberta F. 2010. Online Learning for Interactive Statistical Machine Translation. *Proceedings of NAACL 2010*: 546-554.

Simard M, Goutte C, and Isabelle P. 2007. Statistical Phrase-Based Post-Editing. *Proceedings of NAACL 2007*: 508-515.

Yang YM, Pedersen JO. 1997. A Comparative Study on Feature Selection in Text Categorization. *Proceedings of the 14th International Conference on Machine Learning (ICML)*: 412-420.