The Impact of Crowdsourcing Post-editing with the Collaborative Translation Framework

Takako Aikawa[1], Kentaro Yamamoto[2], and Hitoshi Isahara[2]

[1] Microsoft Research, Machine Translation Team
takakoa@microsoft.com
[2] Toyohashi University of Technology
yamamoto@lang.cs.tut.ac.jp, isahara@tut.jp

Abstract. This paper presents a preliminary report on the impact of crowdsourcing post-editing through the so-called "Collaborative Translation Framework" (CTF) developed by the Machine Translation team at Microsoft Research. We first provide a high-level overview of CTF and explain the basic functionalities available from CTF. Next, we provide the motivation and design of our crowdsourcing post-editing project using CTF. Last, we present the results from the project and our observations. Crowdsourcing translation is an increasingly popular-trend in the MT community, and we hope that our paper can shed new light on the research into crowdsourcing translation.

Keywords: Crowdsourcing post-editing, Collaborative Translation Framework.

1 Introduction

The output of machine translation (MT) can be used either as-is (i.e., raw-MT) or for post-editing (i.e., MT for post-editing). Although the advancement of MT technology is making raw-MT use more pervasive, reservations about raw-MT still persist; especially among users who need to worry about the accuracy of the translated contents (e.g., government organizations, education institutes, NPO/NGO, enterprises, etc.). Professional human translation from scratch, however, is just too expensive. To reduce the cost of translation while achieving high translation quality, many places use MT for post-editing; that is, use MT output as an initial draft of translation and let human translators post-edit it. Many researchers (both from academia and industry) have been investigating how to optimize the post-editing process and developing tools that can achieve high productivity gains via MT for post-editing.[1]

Recently, another type of approach to reduce the cost of translation has surfaced; namely, crowdsourcing translation. Crowdsourcing translation started as a method to create training/evaluation data for statistical machine translation (SMT). For instance, with Amazon's Mechanical Turk, one can create a huge amount of bilingual corpus

[1] See Allen (2003, 2005)[1][2], O'Brien (2005)[3], Guerberof (2009a/b)[4][5], Koehn and Haddow (2009)[6], for instance.

H. Isahara and K. Kanzaki (Eds.): JapTAL 2012, LNAI 7614, pp. 1–10, 2012.

data to build a new SMT system in a relatively inexpensive and quick way (Ambati et al. (2010)[7], Zaidan and Callison-Burch (2011)[8], Ambati and Vogel (2011)[9]).[2] This paper introduces a new way of crowdsourcing translation. Our approach is unique in that it focuses on post-editing and uses a different platform; namely, the Collaborative Translation Framework (CTF) developed by the Machine Translation team at Microsoft Research. For our project, we used foreign students at Toyohashi University of Technology as editors and asked them to post-edit the MT output of the university's English websites (http://www.tut.ac.jp/english/introduction/) via Microsoft Translator (http://www.microsofttranslator.com) into their own languages using the CTF functionalities. This paper is a preliminary report on the results from this project. The organization of the paper is as follows: Section 2 provides a high level overview of CTF while describing various functionalities associated with CTF. Section 3 presents the design of our crowdsourcing project using Toyohashi University of Technology websites. Section 4 presents a preliminary report on the results from the project and Section 5 provides our concluding remarks.

2 Collaborative Translation Framework (CTF)

As mentioned at the outset of the paper, CTF has been developed by the Machine Translation team at Microsoft Research. CTF aims to create an environment where MT and humans can help each other to improve translation quality in an effective way. One of the prominent functionalities of CTF is to allow users to modify or edit the MT output from Microsoft Translator. Thus, with CTF, we can utilize the power of crowdsourcing to post-edit MT output. There are other types of functionalities associated with CTF, and in the following subsections, we describe these in more detail.

2.1 Basic Functionalities of CTF

CTF functionalities have been fully integrated into Microsoft Translator's Widget (http://www.microsofttranslator.com/widget), and one can experience how CTF works by visiting any website(s) with this Widget.[3] For instance, let us look at Figure 1, which is the snapshot of the Widget on the English homepage at Toyohashi University of Technology (http://www.tut.ac.jp/english/introduction/). With this Widget, users (or visitors of this website) can translate the entire web site automaticallyinto their own languages; select their target languages (in Figure 1, Japanese is being

[2] Amazon's Mechanical Turk has been also used for creating different types of data as well. For instance, see Callison-Burch (2009) [10] and Higgins et.al (2010)[11].

[3] CTF functionalities can also be called via Microsoft Translator's public API's. For more details on Microsoft Translator's API's, visit `http://msdn.microsoft.com/en-us/library/dd576287`

selected) and click the translate button (red-circled), so that the entire page can be translated into that selected target language instantly as shown in Figure 2.[4]

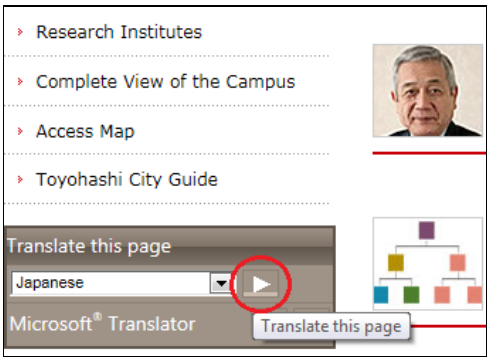

Fig. 1. Widget on the Toyohashi University of Technology website (English)

Fig. 2. Translated into Japanese

Besides the webpage translation described above, CTF functionalities integrated into the Widget can offer users various types of controls. First, by hovering over an individual sentence, users can evoke the CTF user interface (UI) (see Figure 3), and inside the CTF UI, users can see the original sentence and the editorial window where the MT output can be modified as shown in Figure 3.

Second, CTF allows users to see edits from other users. For instance, in Figure 4, the first string next to the icon is the MT output and the two translations below are the alternative translations provided by other users.

[4] Microsoft Translator currently supports 38 languages.
See `http://www.microsofttranslator.com/help/` for the list of languages supported by Microsoft Translator.

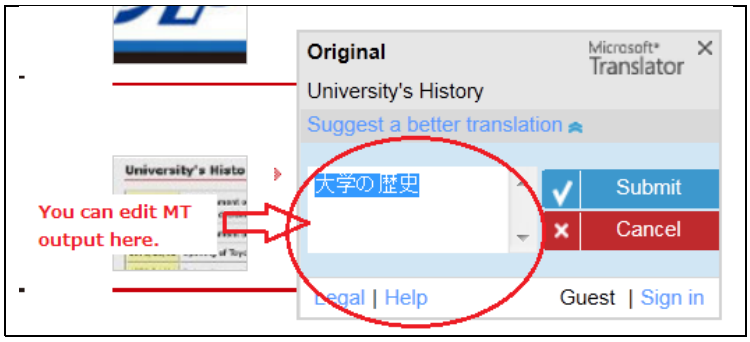

Fig. 3. The Edit control inside the CTF UI

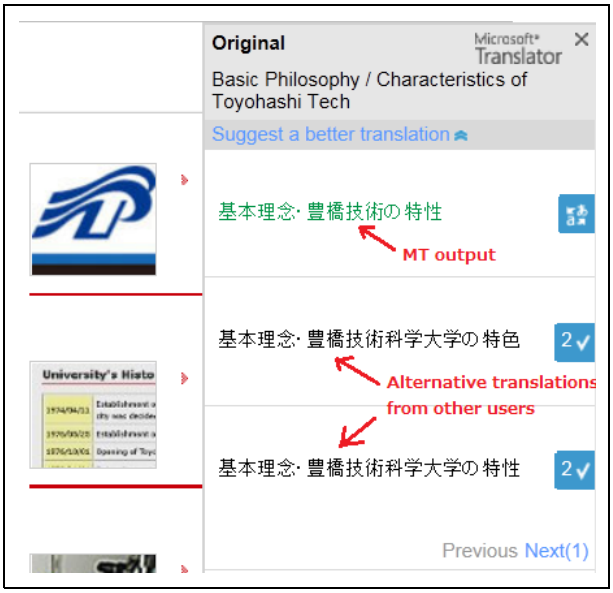

Fig. 4. Alternative translations from other users

Third, users can modify the alternative translations from other users (as well as MT output) as shown in Figure 5.

Note that users can report spam or bad translations by clicking the [Report] button. Furthermore, they can vote for alternative translations if they wish. For instance, the numbers next to the alternative translations in Figure 4/Figure5 above (i.e., “2”) indicate that these translations have already received 2 votes from users.

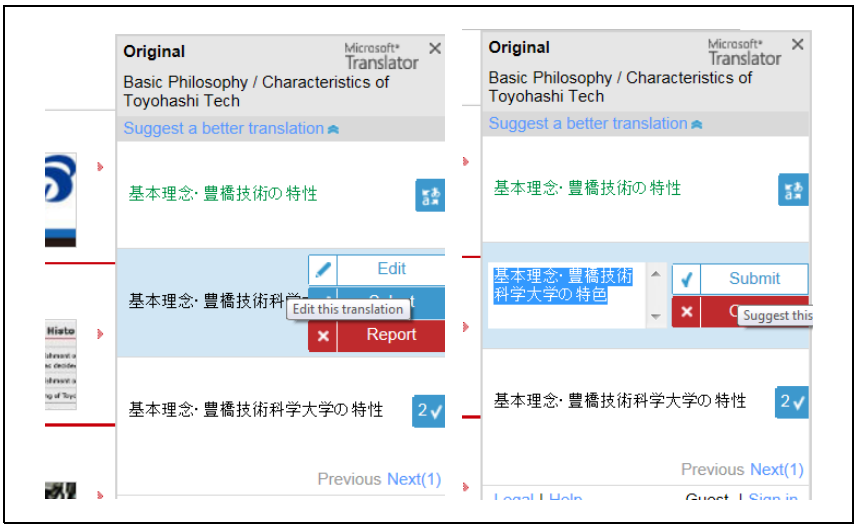

Fig. 5. Post-edit alternative translations

Last but not least, CTF allows a web owner to delegate a set of people as moderators and/or translators. This functionality is critical to our crowdsourcing project and is described in detail in 2.2 below.

2.2 Role-Assignment

One of the biggest concerns for crowdsourcing translation is the quality assurance of crowd-sourced translation. That is, how can we verify the quality of the edits or the translations coming from anonymous users? To address this concern, CTF allows a web owner to assign "trusted" human translators the role of moderator and/or that of translator.

The moderator role can be assigned to someone who can oversee and moderate the quality of the translations coming from translators, and the translator role can be assigned to individuals who can provide their translations. The edits/translations doneby these "trusted" users can overwrite MT output or the edits from anonymous users.[5] This way, the web owner can have more control over the quality assurance of the crowd-sourced post-edits or translations, and she or he can do the assignment of these moderators and translations easily on the CTF dashboard, which is illustrated in Figure 6.

[5] The hieratical order of CTF users is: web owner -> moderator -> translator -> anonymous users. The translations from the web owner, the moderator, or the translator can overwrite MT output but those from anonymous users cannot.

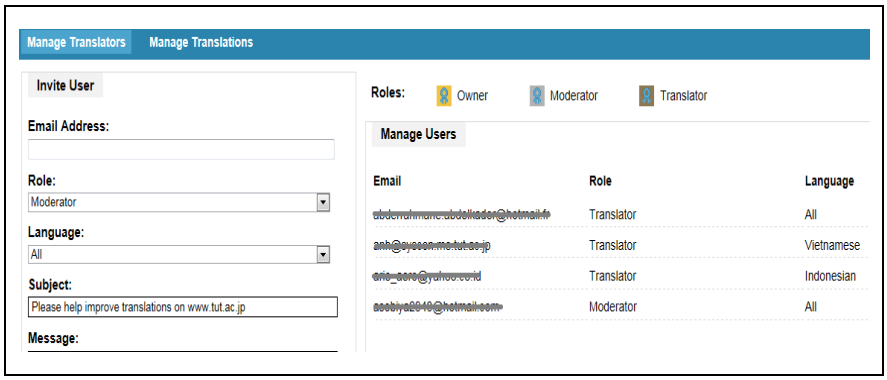

Fig. 6. Role Assignment

3 Crowdsourcing Post-editing: A Case Study at Toyohashi University of Technology

3.1 Background

Toyohashi University of Technology (TUT) has more than 200 foreign students from various countries, and the demand to localize the information on their websites into various languages has always been strong. Yet, localizing the websites using professional translators is just too expensive to do. To make the university information more accessible to these foreign students or to upcoming new students, they created English translations of their websites. However, still lots of foreign students had problems in understanding TUT's website information because of their language barrier. To overcome this issue, the university decided to conduct this crowdsourcing post-editing project using Microsoft Translator's CTF.[6]

3.2 Design

For crowdsourcing post-editors, we hired foreign students with different language background at TUT and assigned each of them the role of translator. These students are familiar with the contents of TUT's websites, which we thought would be a great advantage in post-editing the MT output of TUT's websites. The total number of student participants was 22, and their language background is provided in Table 1.[7]

---

[6] This project is a collaboration between Toyohashi University of Technology and the Machine Translation team at Microsoft Research. See Yamamoto et al. (2011) for our initial report [12].

[7] Strictly speaking, the total number of student participants was 21 as one of the students edited both Arabic and French MT output.

Table 1. The language background of the student participants

Language	Number of the participants
Arabic	2
Indonesian	2
Portuguese	1
Spanish	4
Chinese (simplified)	6
Vietnamese	2
French	2
German	1
Korean	2

Prior to starting the project, we gave these students a brief introduction on how to use the CTF UI and explained the background of the project. We also provided some specific instructions on how to post-edit MT output. The instructions we provided are summarized below:

- Avoid over-editing: don't try to over-edit if the existing translation(s) (whether they are MT output or other human edits) are grammatical and readable.
- Ignore stylistic differences: don't try to modify stylistic differences unless they are critical or matter for readability.
- Start from scratch: if the quality of MT output is too low, provide your translation from scratch (as opposed to modifying MT output).

Fig. 7. Editing alternative translations from other users

It is important to note here that we did not prevent the students from modifying already existing translations provided by other students. For instance, in Figure 7, there are two already existing alternative translations in addition to the MT output. The students are allowed to modify not only the MT output but also any one of these alternative translations if they think it is necessary to modify.

Time-wise, we assigned each student 30 hours of post-editing work on TUT's websites. We conducted this project in November-December, 2011.

4 Results

We have gathered a decent amount of edits from the students as shown in Table 2.

Table 2. Results

(A) Language	(B) Number of sentences edited	(C)Number of edits	(D) Ratio	(E) Average number of edits
Arabic	397	723	45%	361.5
Indonesian	1285	1559	18%	779.5
Portuguese	204	308	34%	308
Spanish	1841	3643	50%	910.75
Chinese (simplified)	1637	2269	28%	378.1
Vietnamese	1341	1929	31%	964.5
French	512	647	21%	323.5
German	147	192	24%	192
Korean	598	707	16%	353.5

Column B refers to the number of original sentences that have been modified. Column C, on the other hand, refers to the total number of edits we got from the student translators. As just mentioned, we did not prevent the students from modifying the edits from others. So some sentences ended up having multiple (alternative) translations, resulting in the gap in number between Column B and Column C. Column D indicates the overall percentage of sentences that have more than one edit(s). For instance, for the case of Arabic, 45% of the original sentences that have more than one edit. Column E indicates how many edits are provided on average by one student (i.e., the numbers of edits divided by the number of students in Table 2).

A couple of observations can be made here. First, the average number of edits varies quite radically depending on the target language. If we simply assume that this average number is pseudo-equivalent to the productivity of post-editing, Vietnamese and Spanish students are expected to be the most productive post-editors, and the German editor the least productive one.[8] Another interesting phenomenon observed

[8] It is true that the quality of the MT system also varies depending on the target language, and we can't or shouldn't come up with any conclusion based solely on these numbers. But, let's assume that we can.

here is the high ratio of multiple edits for Spanish. The English->Spanish system is allegedly the best system in terms of the quality of Microsoft Translator, yet this language pair has the highest ratio of multiple post-edits. It is unclear why this is the case, and we would like to investigate this as a future research topic.

5 Concluding Remarks

In this paper, we provided a preliminary report on the results from our crowdsourcing post-editing project using CTF. Using the crowdsourcing power of the foreign students at TUT, we could localize the majority of TUT's English websites into 9 languages within 2 months with inexpensive cost, and we are very happy about this outcome. We also asked these foreign students to give us their verbatim feedback, some of which are provided in Table 3 below. As seen there, the overall feedback from the students is very positive and it is great to see that the students felt the "sense of community" by participating in this project.

Table 3. Verbatim feedback from the participants

Indonesian student	Working as Microsoft translator give me great benefit especially knowing in detail about the content of the TUT website which I previously didn't know.
Spanish student	….I have to say that I am very glad that I could be a part of it, and about the project itself, I think it's a great way to Internationalize and attract more overseas students …….Since Machine translation in websites is probably the easiest way of helping readers from different countries communicate, it is often used as first choice, but most of the time it translates incorrectly or some sentences do not make any sense, that's why I believe this was a great opportunity to help improve the webpage by the use of humans.
Spanish student	While doing this job, I was able to realize the complexity of translating without changing the original meaning. Sometimes I had to check and correct my own translations at least once to make them sound coherent in Spanish.
Vietnamese student	I could look at most of the sites and provided my changes. But, translation of technical sentences was tough. I think that the Vietnamese translations become much better and more natural now. [9]
French student	This work session for the enhancement of TUT's website was a good idea and I am sure it will permit the University to be well known abroad and reveal its potentialities to students and partners who plan to come in Japan for studies or partnership.

[9] This Vietnamese student provided his/her feedback originally in Japanese. This is the translation of the original feedback.

Our next step is to examine the accuracy and the quality of these crowd-sourced translations from the students. We are currently working on this together with professional translators, and would like to make a report on this investigation in the future.

Acknowledgments. This project was funded by International Affairs Division at Toyohashi University of Technology, and we would like to give special thanks to all the members of International Affairs Division for their support during the project. We are also thankful to Midori Tatsumi for her feedback on our paper.

References

1. Allen, J.: Post-editing. In: Somers, H. (ed.) Computers and Translation: A Translator's Guide, pp. 297–317. John Benjamins Publishing Company, Amsterdam (2003)
2. Allen, J.: What is post-editing? Translation Automation 4, 1–5 (2005)
3. O'Brien, S.: Methodologies for measuring the correlations between post-editing effortand machine translatability. Machine Translation 19(1), 37–58 (2005)
4. Guerberof, A.: Productivity and quality in MT post-editing. In: MT Summit XII – Workshop: Beyond Translation Memories: New Tools for Translators MT, Ottawa, Ontario, Canada, p. 8 (2009a)
5. Guerberof, A.: Productivity and quality in the post-editing of outputs from translation memories and machine translation. Localisation Focus. The International Journal of Localisation 7(1) (2009b)
6. Koehn, P., Haddow, B.: Interactive Assistance to Human Translators using Statistical Machine Translation Methods. In: MT Summit XII (2009)
7. Ambati, V., Vogel, S.: Can crowds build parallel corpora for machine translation systems? In: Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, Los Angeles, CA, pp. 62–65 (2010)
8. Zaidan, O.F., Callison-Burch, C.: Crowdsourcing translation: professional quality from non-professionals. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pp. 1220–1229 (2011)
9. Ambati, V., Vogel, S., Carbonell, J.: Active learning and crowd-sourcing for machine translation. Language Resources and Evaluation (LREC)
10. Callison-Burch, C.: Fast, cheap, and creative: Evaluating translation quality using Amazon's Mechanical Turk. In: Proceeds of the 2009 Conference on Empirical Methods in Natural Language Processing, Singapore, pp. 286–295 (2009)
11. Higgins, C., McGrath, E., Moretto, L.: MTurk crowdsourcing: A viable method for rapid discovery of Arabic nicknames? In: NAACL Workshop on Creating Speech and Language Data With Amazon's Mechanical Turk, Los Angeles, CA, pp. 89–92 (2010)
12. Yamamoto, K., Aikawa, T., Isahara, H.: Impact of collective intelligence on the post-editing machine translation output (機械翻訳出力の後編集の集合知による省力化). In: Proceedings of NLP 2012, Hiroshima, Japan (2012)