Evaluation and Usability of Back Translation for Intercultural Communication

Tomohiro Shigenobu

Language Grid Project, National Institute of Information and Communications Technology, 3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0289, Japan
shigenobu@nict.go.jp

Abstract. When users communicate with each other via machine translation, it is important to improve the quality of the translations. The "Back Translation" technique can improve the translation accuracy. A back translation, first, translates the input language into the target language (outward), and then translates the target language into the input language (homeward). This allows the users to confirm the accuracy of the machine translation by themselves. If the user finds that his input sentence is unsuitable for machine translator, he can rewrite the input sentence. For effective multilingual communication, it is important that the back translation offer good accuracy and good usability. This paper focuses on these two points; we evaluated the accuracy of back translation, and developed a user interface that improves the usability of back translation. The outward and homeward translations show a correlation. Back translation can improve the accuracy of outward translation for users.

Keywords: Machine translation, Back translation, Intercultural communication, Usability.

1 Introduction

Several communities are now communicating via machine translation[1], and such communities are expected to increase [1]. While machine translation is useful, it comes with some risk. As users depend on the accuracy of the translation, the lack of any necessary information could lead to misunderstanding and confusion [2]. This problem was elucidated in the Intercultural Collaboration Experiments in 2002 (ICE2002[2]) with Chinese, Korean and Malaysian colleagues [3]. More than forty students and faculty members from five universities joined this experiment. The goal was to develop open source software using the participants' first language: Japanese participants use Japanese; Chinese participants use Chinese, and so on. The experiment used the multilingual communication tool TransBBS (bulletin board systems) to translate the message of one user into the other languages. We found that the accuracy of current machine translators was inadequate for intercultural communications. Even if the message was written as usual, other users did not

[1] Enjoy Korea: http://www.enjoykorea.jp/
[2] Intercultural Collaboration Experiment: http://www.ai.soc.i.kyoto-u.ac.jp/ice/

N. Aykin (Ed.): Usability and Internationalization, Part II, HCII 2007, LNCS 4560, pp. 259–265, 2007.

understand its translation. We took a new approach to remove this problem. Participants check the result of translation and rewrite their message before posting. Most participants could understand English to some extent. They translated their messages into English and confirmed the English translation. If the translation appeared to be incorrect, they altered the original message and retranslated it. When the English version was adequate, or no further improvement could be discerned, the message was posted. We found that this incremental improvement did raise the quality of the message significantly. However, most people who doesn't understand the reference language cannot execute this method. One method for solving this problem is "Back translation."

Users need to know what information is being sent to their partners. Back translation allows a user to write a sentence that is machine-translatable using only the user's mother language. Back translation is therefore an important technique as it provides confirmation of translation accuracy and highlights significant problem areas. Back translation is defined as the original language obtained by translating input into a target language and then retranslating the resulting text back into the original language. The effectiveness of back translation is based on the assumption that when back translation (homeward) is correct, the target language translation (outward) must also be correct. This method enables the translation accuracy to be confirmed in the input language. When the outward and homeward translations differ greatly from each other estimating the accuracy of the target language translation is difficult. Correcting text using back translation makes no sense without reliability of accuracy. We need to evaluate the accuracy of outward and homeward translations.

Many communication tools have been developed to date, but most of them only display the translated text [4]. Furthermore, existing back translation tools only show the translation results, and a user must decide whether the input text needs to be corrected [5]. Therefore, by examining the experiences from past applications, the following problems became apparent. 1) Users require a long time to complete the cycle of correcting input and back translated text because users usually prepare an entire message and then translate it. 2) As users do not know what part of the input should be corrected, they have to repeat the process of correcting the input and executing back translation many times. 3) The accuracy of translated sentences does not improve even if users correct the input sentence due to operating in a multilingual environment. These problems have reduced a user's willingness to initiate corrections. From a practical viewpoint, merely using back translation will not provide effective support. Developing a more effective user interface for back translation is necessary.

This paper describes our evaluations of the relation between the outward and homeward translations offered for Japanese-English machine translation as an example. Furthermore, the development of a user interface that makes back translation more effective is described.

2 Accuracy of Back Translation

When both quality of an outward translation result and that of back translation result are far from each other in quality, it is difficult to estimate the quality of target

language. Therefore, outward and homeward quality evaluations of back translation are necessary. This section describes evaluations of the relation between the quality of outward and homeward translation using Japanese-English machine translation as an example.

2.1 Method of Evaluation

Hereafter, target language translation is referred to as "outward translation" while back translation is called "homeward translation." The instructions given in the subjective human evaluations followed those used for TIDE [6]. 5-step evaluations were performed. The evaluation method compares translated sentences and reference sentences to judge how much information was transferred correctly. Evaluation criteria were "How much of the meaning expressed in the gold-standard translation was also expressed in the target translation?" the evaluation scores were "5: All, 4: Most, 3: Much, 2: Little and 1: None". We used J-Server by KODENSHA[3] as the machine translation system. As the evaluation texts, we used 186 sentences extracted at random from material provided by NTT[4]. The material examined consisted of a set of original Japanese sentences and paired original English sentences. The evaluation contents are described below:

- Evaluation A: We compared outward translations (English) to the reference sentences (original English). This represents the outward evaluation.
- Evaluation B: We compared the homeward translations (Japanese) to the input Japanese sentences (original Japanese). This represents the homeward (back translation) evaluation.

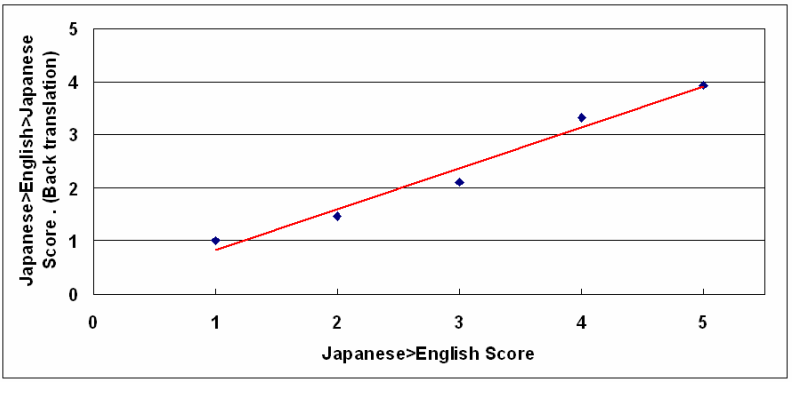

Fig. 1. Correlation between outward and homeward

2.2 Results

Figure 1 shows the relation of evaluation values between outward (Evaluation A) and homeward (Evaluation B). We found that the two translation directions were

[3] KODENSHA: http://www.kodensha.jp/
[4] NTT Natural Language Research Group: http://www.kecl.ntt.co.jp/mtg/resources/index-j.php

correlated. The homeward evaluation value was about 20% less than the outward evaluation value. Figure 2 shows the relation of evaluation values between outward and homeward translations based on the number of words. The point size represents the number of samples. As an inevitable consequence, values of both evaluations tend to fall as the sentences become long. Additionally, the homeward evaluation becomes lower as the outward evaluation become low. The average value of outward translation was 3.1 and that of homeward translation was 2.4. Given that the quality of evaluation value 5 is 100% and that of evaluation value 1 is 0%, the quality of outward decreases to 77% of its original value and that of homeward quality decreases to 60%. However, given that evaluation value 3 is the "threshold limit value for understanding contents", understandable outward translated sentences consist of about 8 words on average and homeward translated sentences consist of about 5 words on average. The evaluation texts is used for testing of machine translator, the evaluation value is low as there are several sentences that the machine translator finds difficult to translate. We consider that the evaluation value will improve as users tend to write suitable sentences for machine translation. The evaluation value decreased if the machine translation result was used as the input sentence since these sentences were generally incomplete; however, there is a high possibility that when back translation (homeward) is correct, the target language translation (outward) also is correct.

The quality of back translation in which outward translation is used as input sentence decreases significantly as sentence becomes long. However, the quality between outward and homeward translated sentences has correlation; thus, it is highly possible that if quality of back translation is high, quality of target translation becomes high. Therefore, users may be able to estimate the quality of outward translation by confirming homeward translation sentences.

We will evaluate various viewpoints, and increase the number of subjects and translators to be evaluated.

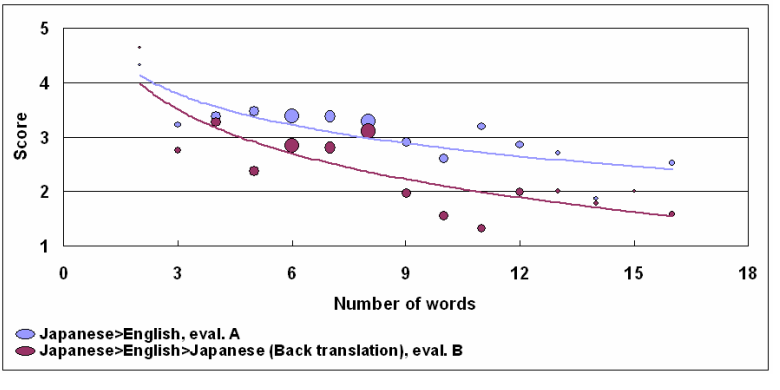

Fig. 2. Relation of outward and homeward translation quality

3 Usability

Our case studies showed that back translation suffers from several problems. Back translation needs some support functions if it is to become truly practical. Simply showing the result of back translation to the user is not good enough. This section describes a user interfaces to improve the usability of back translation.

3.1 Real-Time Back Translation

When users used the back translation provided by our first tool, they had to push the back translation button by themselves. The back translation result was shown only after they input their entire message. Since the system took several seconds to display the back translation result, the user quite often gave up on using the result. Our solution is real-time incremental back translation. It shortens the response time and improves usability so that users can obtain the back translation results while inputting their message. An incomplete sentence is likely to yield errors in the back translation. However, users can find their typographic and grammatical errors. We abut the input area and the back translation area to grab the user's attention. Figure 3 shows the display position of the back translation result.

3.2 Highlighting Translation Unsuitable Parts

Users who tried to correct sentences made many unsuccessful attempts because they did not know what part of the input caused problems for the machine translator. To help the user in locating problem areas in the input sentence, we have developed a

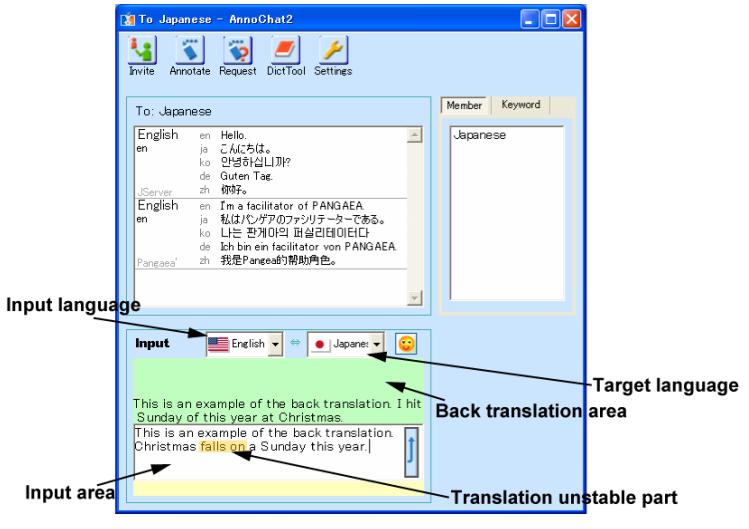

Fig. 3. Display of back translation result

function that highlights the translation unsuitable parts. This function is based on natural language processing technologies. Based on morphological analysis and syntactic parsing, it uses the degree of similarity as employed for the automatic evaluation of translation accuracy. This method can calculate the degree of similarity between an input message and the back translation result automatically [7]. When the degree of similarity is low, the method simplifies the original text gradually. The processes of simplification yield several texts. The method can estimate unsuitable parts by the differences of these texts [8]. The user is shown which parts of the message are unsuitable for translation. This method has been shown to improve message quality for machine translation [9]. Figure 3 shows the user interface with modification candidate words highlighted.

3.3 Multilingual Environment

Corrections performed by Japanese users when referring to the back translations based on English were very effective in improving the accuracy of the English translations. However, the corrections were not so effective for Chinese and Korean translations [10]. When a user corrects input using back translation, the accuracy of the translations is not improved for all languages simultaneously. Therefore, we extended the back translation function described above to suit multilingual environments. Figure 4 shows the user interface of this extended function. Users can write a suitable message for each language. When they decide that the translation quality is suitable (via back translation), they select the "fixed" check box to fix the translation or that language.

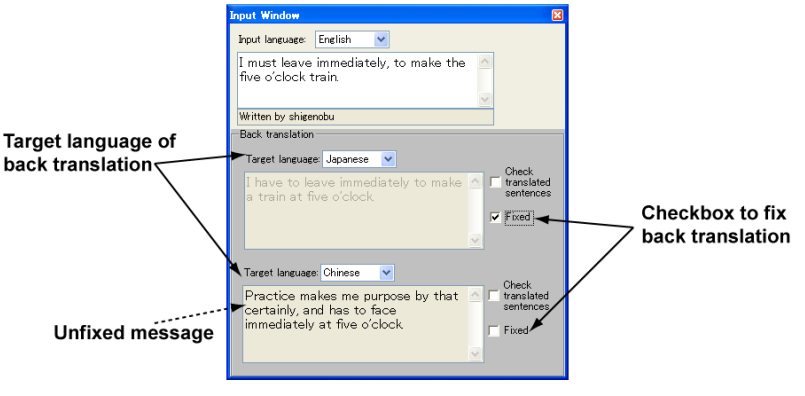

Fig. 4. Fixing function for back translation in multilingual environment

4 Conclusion

Current machine translation systems have insufficient quality to support intercultural communication. Therefore, repairing input messages by back translation is essential to improve translation quality. We verified the effectiveness of back translation by

evaluating the relation between outward and homeward translations, and developed am effective user interface for back translation.

The quality of back translation in which outward translation is used as input sentence decreases significantly as sentence becomes long. However, the quality between outward and homeward translated sentences has correlation. Therefore, users may be able to estimate the quality of outward translation by back translation. Moreover, to improve the usability of back translation, we developed a user interface that offers real-time back translation, back translation proceeds automatically as the message is being input. We arranged the input area close to the back translation result area so that the user can notice typographic and grammatical errors immediately. The interface also highlights message passages that unsuitable for translation. For multilingual environments we developed an interface in which users can repair and write messages for each language.

References

1. Climent, S., More, J., Oliver, A., Salvatierra, M., Sanchez, I., Taule, M., Vallmanya, L.: Bilingual Newsgroups in Catalonia: A Challenge for Machine Translation. In: Journal of Computer Mediated Communication, 9(1), (2003)
2. Yamashita, N., Ishida, T.: Automatic Prediction of Misconceptions in Multilingual Computer-Mediated Communication. In: International Conference on Intelligent User Interfaces (IUI-06), pp.62–69 (2006)
3. Nomura, S., Ishida, T., Yamashita, N., Yasuoka, M., Funakoshi, K.: Open Source Software Development with Your Mother Language: Intercultural Collaboration Experiment 2002. In: International Conference on Human-Computer Interaction (HCI-03) vol. 4, pp. 1163–1167 (2003)
4. Raymond, S.F., Chris, C.B.: Secondary Benefits of Feedback and User Interaction in Machine Translation Tools. In: Workshop paper for “MT2010: Towards a Roadmap for MT” of the MT Summit VIII (2001)
5. Yoshino, T., Shigenobu, T., Maruno, S., Ozaki, H., Ohno, S., Munemori, J.: Development and Application of an Intercultural Synchronous Collaboration System. In: Proceedings of Eighth International Conference on Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies (KES 2004), pp.869–882 (2004)
6. Linguistic Data Annotation Specification: Assessment of Fluency and Adequacy in Arabic-English and Chinese-English Translations. (2002), http://www.ldc.upenn.edu/Projects/TIDES/Transl ation/TransAssess02.pdf
7. Uchimoto, K., Hayashida, N., Ishida, T., Isahara, H.: Automatic Rating of Machine Translatability, 10th Machine Translation Summit (MT Summit X), pp. 235–242 (2005)
8. Uchimoto, K., Hayashida, N., Ishida, T., Isahara, H.: Automatic Detection and Semi-Automatic Revision of Non-Machine-Translatable Parts of a Sentence. In: International Conference on Language Resources and Evaluation (LREC-06) (2006)
9. Hayashida, N., Ishida, T.: Performance Prediction of Supporting Self-Initiated Repair by Translation Agents, IEICE, vol. J88-D-I(9), pp, 1459–1466, (Japanese 2005)
10. Ogura, K., Hayashi, Y., Nomura, S., Ishida, T.: User Adaptation in MT-mediated Communication. In: Su, K.-Y., Tsujii, J., Lee, J.-H., Kwong, O.Y. (eds.) IJCNLP 2004. LNCS (LNAI), vol. 3248, pp. 596–601. Springer, Heidelberg (2005)