



Multilingual Communication via Best-Balanced Machine Translation

Mondheera Pituxcoosuvorn¹  · Toru Ishida¹

Received: 1 August 2017 / Accepted: 26 July 2018 / Published online: 4 August 2018
© Ohmsha, Ltd. and Springer Japan KK, part of Springer Nature 2018

Abstract

It is known that creative ideas are often generated by correspondents from different cultures, but it remains true that distance still matters due to the language barrier. To enhance multilingual communication, this paper proposes the model of best-balanced machine translation. Our model is based on the quality of messages among participants (assumed to have different levels of language skill) and takes not only machine translation quality but also users' language skill into account. We provide a method to select languages to be used with machine translators, and a way of creating the best-balanced communication environment. Many studies have addressed machine translation with the goal of helping the non-native speaker to understand what was said. Our approach is totally differently, since we focus on helping the non-native speaker by enhancing the opportunity to join in the conversation. We conduct an experiment and find that this model allows machine translation technologies to benefit multilingual communication while making best use of the participants' different language skills. The proposed model addresses the talkativeness of the participants. It also improves communication by reducing serious machine translation errors and the number of conversation breakdowns.

Keywords Communication support environment · Intercultural collaboration · Multilingual communication · Usability of machine translation

Introduction

To create new ideas, gathering ideas from people with different backgrounds including cultures and languages is expected to enhance creativity. Most of the time people learn English to collaborate [1], because English is now the global language. We can often hear people around the world talk in English [2]. However, in international

✉ Mondheera Pituxcoosuvorn
mondheera@ai.soc.i.kyoto-u.ac.jp

¹ Graduate School of Informatics, Kyoto University, Yoshidahonmachi, Sakyo Ward, Kyoto, Kyoto Prefecture, Japan

discussions, native speakers might have advantages over non-native speakers which is likely to yield less-than-optimal results. In fact, big gaps in language skill can reduce the opportunities of non-native speakers to participate in intercultural communication. When English is used in a group with language diversity, socialization and interpretation and will be impacted since it can become a hidden barrier. Non-native speakers sometimes receive negative assessments because of their low language skill. Moreover, their intelligence tends to be underestimated because they speak slowly [3].

Various methods have been developed to help non-native speakers participate in conversations with native English speakers, for example, creating artificial delays to help the non-native speaker understand the conversation before continuing the conversation [4], signaling the native speaker about the status of non-native speakers [5], helping non-native writers by providing vocabulary navigation [6], and providing real-time translation using eye gaze input of the non-native reader [7]. Even though these methods can reduce the burden of non-native speakers, they cannot provide a completely balanced communication environment.

While sharing a language can smooth the communication process, innovation is greatly enhanced by learning other languages and respecting different cultures. Because it is impossible for one person to learn every language, machine translation (MT) and other technologies on the internet are attractive solutions [1]. Using MT can improve the efficiency and effectiveness of discussions [8]. Yet, MT can also cause many communication problems during collaboration. Because people have various levels of language proficiency and machine translation accuracy is uncertain, it is a difficult task to decide which languages or which translation services should be used. If MT services are used but some of the users have better common language skill than the translation quality of MT, the conversation will not be as fruitful as it should be. Polysemy and synonymy [9], common problems with machine translations, can also cause conversation breakdown [10]. On the other hand, if a foreign language is used chosen, the user with lower skill in that language will have less chance to communicate, or come to feel left out of the conversation.

Some researchers have attempted to improve communication by improving the quality of machine translation as well as using human intelligence. For example, Morita [11] introduced a method to use monolinguals to help with the fluency and adequacy of both sides of two language translations. Taking a direction from the outsourcing of human intelligence, we realized that the ability of the users themselves is also a resource that should be better utilized. Many people know more than one language and to communicate in a group, we can combine the ability of those users and machine translation services to realize best quality communication.

The studies mentioned earlier introduced various methods to help non-native speakers and support multilingual collaboration. Our research is novel and orthogonal to existing research. Our model aims to support non-native speakers with different proficiencies in a shared language. Our method creates the best balance in terms of opportunity to participate in communication. To obtain the best-balanced communication environment, we start with a previous study called user-centered QoS [12]. Services are generally evaluated by users in terms of the quality of service (QoS). However, information of users is important in selecting the best machine

translation service. Thus, they introduced a new function that calculates the quality of message (QoM) using the users’ skills in writing and reading messages when machine translators are used. In this paper, we extend QoM to define a model of the best-balanced channel using the parameters of user language skills and machine translation accuracy. Then, we investigate our model in a real-world experiment to confirm the effectiveness of our approach in creating effective environments for multilingual collaboration.

Scenario

Machine translation can cause communication balance problems. For example, Fig. 1 shows a situation in multilingual communication. It is not complicated to choose the best communication method for a conversation between a Chinese user with fair English skill and a Japanese user with limited English skill. Selecting the appropriate translator is straightforward, especially for the Japanese user. Later, a Korean user with good English skill joins the conversation; this makes it more complicated to choose what languages or what services should be used.

One possibility is to use the shared foreign language, English. Another possibility is to use machine translation. It is also possible to combine both options. If only English is used for this conversation, it might cause difficulties for the Japanese whose English skill is limited. Machine translation could be useful; however, the other two participants have good enough English skill to communicate which might be better than using machine translation because machine translation is still imperfect. The situation is a problem of asymmetry in collaboration caused by differences in language proficiency. In such groups, people have asymmetric opportunity to participate in the conversation. We believe

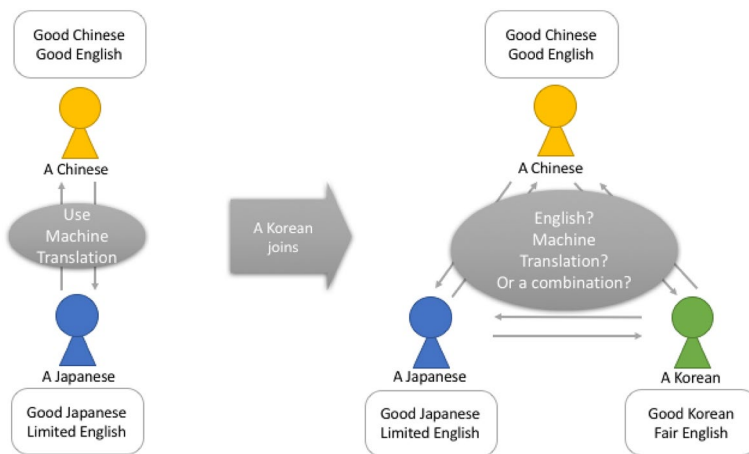


Fig. 1 Situation where the multilingual communication problem exists

that to have the best communication is to have mutual understanding and equal chance of participating in the conversation. Our proposal tackles this problem.

Modeling Multilingual Communication

Best-Balanced Channel

In this paper, we propose a model to cope with asymmetrical participation in terms of unequal opportunity to take part in a conversation and the asymmetrical nature of machine translation. Our model is called best-balanced machine translation.

Based on the existing work related to user-centered QoS [12], we model the quality of message (QoM) where user P_i uses language L_i to send a message to user P_j , who uses language L_j via machine translation service $MT_{i,j}$. Service $MT_{i,j}$ translates messages from language L_i to language L_j . To calculate QoM, we consider the input language writing skill of the message sender, $MT_{i,j}$ machine translation accuracy, and output language reading skill of the message receiver. The quality of message from user P_i to P_j via machine translation service $MT_{i,j}$. QoM ($P_i, MT_{i,j}, P_j$), or simply $QoM_{i,j}$, can be represented as follows:

$$QoM(P_i, MT_{i,j}, P_j) = \text{writing_skill}(P_i, L_i) \times \text{accuracy}(MT_{i,j}) \times \text{reading_skill}(P_j, L_j) \quad (1)$$

In this model, writing skill of the sender, accuracy of machine translation, and reading skill of the receiver affect QoM. As a result, choosing the most appropriate language pairs is crucial.

Because messages are major parts of conversations, to increase the overall quality of multilingual communication, the quality of message should be maximized. Our model also provides a method of selecting the language pairs that will maximize the quality of message.

The QoM pair between user P_i and user P_j is written as ($QoM_{i,j}, QoM_{j,i}$) and the MT pair between language L_i and language L_j is written as ($MT_{i,j}, MT_{j,i}$). A QoM pair is Pareto optimal when we cannot make a QoM better, without making another QoM worse. A QoM pair is selected as best balanced when it is Pareto optimal and has the least variance. If there are more than two users, Pareto optimality must be extended. $QoM_{i,j}$ can be maximized by selecting appropriate language pair (L_i, L_j), under the constraint that each user can speak one language. The average QoM is defined as the average of $QoM_{i,j}$ and $QoM_{j,i}$.

A set of QoM pairs is Pareto optimal when it is impossible to make a better average QoM, without making any of the other average QoMs worse off. A set of QoM pairs will be the best-balanced set when it is Pareto optimal and the variance of average QoMs is minimum among all Pareto optimal sets of QoM pairs. However, if there is only one Pareto optimum, it is not necessary to calculate the variance.

Example

Assume that languages L_1, L_2, L_3 are used by users P_1, P_2, P_3 , respectively. From the situation in Fig. 1, let ja, ko, and zh represent Japanese Korean, and Chinese language. Under the assumption that everyone has various level of English skill, the possible combinations of languages $Cx=\{L_1, L_2, L_3\}$ for the communication among the three users are as follows:

$$C1 = \{ja, ko, zh\}, C2 = \{ja, ko, en\}, C3 = \{ja, en, zh\}, C4 = \{ja, en, en\}, \\ C5 = \{en, ko, zh\}, C6 = \{en, ko, en\}, C7 = \{en, en, zh\}, C8 = \{en, en, en\}$$

For n number of users, each combination consists of $n(n-1)/2$ QoM pairs. For instance, combination $C1$ is composed of three QoM pairs including $(QoM_{1,2}, QoM_{2,1})$, $(QoM_{2,3}, QoM_{3,2})$, and $(QoM_{3,1}, QoM_{1,3})$. $C1$ utilizes three pairs or six of machine translation services, including $(MT_{ja,ko}, MT_{ko,ja})$, $(MT_{ko,zh}, MT_{zh,ko})$, and $(MT_{zh,ja}, MT_{ja,zh})$.

Given the user language profiles and machine translator qualities set in the example, the only Pareto optimal combination is $C4$, which means that the conversation will be best balanced when the Japanese user uses Japanese while Korean and Chinese users use English; note that the best machine translation service is $(MT_{jp,en}, MT_{en,jp})$, since $(MT_{en,en}, MT_{en,en})$ represents using English with no translation.

In some cases, more than one Pareto optimal combination will exist. The best-balanced combination can be chosen by calculating the differences among the QoMs using variance, because a lower difference indicates a higher quality of conversation.

Experiment

To investigate our model, we designed and conducted an experiment. This experiment is designed to compare our best-balanced machine translation channel (BB) with other channels including using English as common foreign language and using a full translation service. The calculation and the selection of language were done once before each game started.

Task

In the experiment, the participants were asked to play three collaborative games. Three survival problems were used: desert survival problem (DSP) [13], winter survival problem (WSP) from the project ARISE [14], and lunar survival problem (LSP) from NASA [15]. DSP is a popular collaborative task. The participants have to arrange items in a list by their importance given a situation of a crash landing in a desert, in order to survive and reach the destination safely. WSP is similar to DSP, but the environment is a forest and the weather is extremely cold. The task for the

participants is the same as DSP but the items different from those in the first game. LSP is slightly a unique situation, landing 80 km from the target place on the moon. Yet, LSP task is similar to the first two games but again with a different set of items.

The original problems describe the situation using a number of paragraphs in English. In the experiment, the game descriptions are narrated using easy short sentences in English and figures. Simplification is needed to cover the various English proficiencies of the players. Each story explains its situation using time, location, and events that happened; the participants then played survivor roles in the story. Our version of the game also simplified the choice of items. While the original games provide many items to be ranked, our participants were asked to rank a set of only six items in decreasing order of their importance for each situation.

The participants were asked to collaborate and finish the game within a time limit but we did not mention the game score or the right ranking responses. Because the getting right answer or high score depends on the specific knowledge of the team members, for example, science, geography, survival skill and camping skill, the responses do not show the effectiveness of participant collaboration.

Experiment Design

At first, game instructions were introduced to the participants. Next, we demonstrated how to use Online Multilingual Discussion Tool (OMDT), which is a web application created for multilingual symposia. OMDT enables multilingual chat and it uses translation services from The Language Grid which is a collective intelligent system that allows users to combine existing language services for their own usage [16].

With OMDT, the user can choose a language to be shown by selecting from a drop-down on the right-top of the screen. The user can type her/his message in the language selected into the message box then click to send the message. The message will appear below in her/his selected language, but on the screen of the other users, the same message appears in the language selected by that user. In this way, users can chat using their mother language or their foreign language.

Before playing the three survival games, we played an example game for 20 min. During this example game, participants can still ask questions and talk. Later, after the participants understood how to play and how to use OMDT, we asked the participants to move and sit separately so they could not see each other. The games were played using three strategies of communication. The participants played the first game using their shared language, English (EN), fully using machine translator (MT), or using best-balanced machine translation (BB). The strategy was chosen randomly. The second game was played using one of the strategies not used in the first game. The last game was played using the remaining strategy.

First, we gave them an explanation of the situation and asked them to try to understand the given problem, then write down their personal answers before discussing the selections with the other participants by chatting or using machine translation online. After that, they created the team answer by discussing the options with the

other team members. At the end of the game, the participants could choose to give a new personal answer set if the discussion influenced their thought.

After playing those three games with different communication channels, the participants were interviewed about how they felt when they play each game with a different channel.

Participant

We divided our nine participants into three groups. Each group consisted of a Chinese participant, a Japanese participant, and a Korean participant. All of them were either undergraduate, graduate, or research students from various fields.

English skill profiles of the participants, displayed in Table 1, consisted of (writing_skill, reading_skill) normalized to the range of 0–1. English skills were measured using normalized standard test score from TOEIC, TOEFL, or IELTS. Test scores were converted to Common European Framework of Reference for Languages (CEFR) [17] which is an international standard for English language ability. CEFR has six points including A1 (basic), A2 (basic), B1 (independent), B2 (independent), C1 (proficient), C2 (proficient). Score conversion is done with data from ETS [18] for TOEIC and TOEFL and data from Cambridge assessment [17] for IELTS. With the conversion matrix, we give the language score for QoM calculation as follows: 1 for C1 and above, 0.75 for B2, 0.5 for B1, 0.25 for A2, and 0 for A1 and lower. Gender is written as M, for male, and F, for female.

Machine Translation

At the moment, there are several machine translation services available. The services we used in this experiment are from J-Server and Toshiba English-Chinese Machine Translation. J-Server was used for all translations except between English and Chinese.

We randomly chose 20 sentences from a corpus provided by Japan Electronics and Information Technology Industries Association (JEITA) in English to be translated. Each sentence was translated into Chinese, Japanese, and Korean by human, native speakers holding at least bachelor's degree. The translations were approved by another native speaker of the same language. Later, each sentence in each language was translated by machine into the other three languages. To illustrate, sentences in Japanese were translated into Chinese, English, and Korean.

Table 1 Profile of participants

Group/participant	Chinese		Japanese		Korean	
	Gender	English skill	Gender	English skill	Gender	English skill
Group 1	M	(1, 1)	F	(0.75, 0.5)	M	(1, 0.75)
Group 2	F	(1, 1)	M	(0.75, 0.5)	F	(1, 0.75)
Group 3	M	(1, 1)	M	(0.75, 0.5)	F	(1, 0.75)

Table 2 Quality of translation services

From > to	English	Chinese	Japanese	Korean
English	1	0.5875	0.7	0.7375
Chinese	0.6625	1	0.75625	0.6625
Japanese	0.88125	0.7875	1	0.75
Korean	0.5875	0.41875	0.675	1

Table 3 QoM values of all possible combination

Combination/QoM pair	Pair 1	Pair 2	Pair 3
C1	0.771875	0.540625	0.712500
C2	0.771875	0.625000	0.790625
C3	0.790625	0.570313	0.712500
C4	0.790625	0.875000	0.790625
C5	0.468750	0.540625	0.404688
C6	0.468750	0.625000	0.750000
C7	0.500000	0.570313	0.404688
C8	0.500000	0.875000	0.750000

Even though quantitative metrics are useful for evaluation, they cannot completely replace human assessment [19]. The translated sentences were rated by educated native speakers also holding at least a bachelor's degree. This methodology of rating adequacy and fluency adopted is widely used to measure machine translation as proposed by LDC [20]. Translated sentence fluency was scored from 0 to 5. Adequacy was also rated from 0 to 5 rated as how much meaning of the sentence was expressed by the translated sentence. Finally, the evaluation results were confirmed by another native speaker of the same language.

The evaluation for each sentence was averaged to decide the quality of machine translation service from one language to another. Adequacy and fluency ratings assigned by e humans were added up and normalized to the scale of 0–1 as displayed in Table 2.

Communication Channel

The value of QoM pairs for each combination from C1 to C8 can be calculated, as shown in Table 3, using the participant profile from Table 1, and quality of translation services from Table 2. In this case, the only row containing Pareto optimal sets of QoM pairs is C4, so variance was not calculated and the best-balanced machine translation channel was C4.

From Sect. 3.2, C4 contains {ja, en, en}, which represents the best-balanced machine translation channel or BB. With this channel, Chinese and Korean participants use English while Japanese participant uses Japanese.

To investigate the validity of our result, we chose the other two communication channels that are often used when this paper was written: using their foreign language, English as conversation medium (EN) and fully using MT(MT); all the members use their mother language and communicate via MT, as shown in Fig. 2. Hence, the communication channels used in this experiment include BB, EN, and MT, where EN represents C8 {en, en, en}, and MT represents C1 {ja, ko, zh} in the calculation.

Behaviors of Participants with Low Shared Language Skill

Simpler Sentences Used by Japanese When Using English

With the EN channel, the sentences typed by Japanese users were simpler and shorter because of their limited language skill. Simple sentences do not inherently create poor communication, but longer, more complex sentences are more likely to establish natural communication and trigger interesting discussions or new ideas.

Ignorance of Incomprehensible English Sentence

Low language proficiency can lead to incomprehensible sentences. The conversation below shows a part of conversation when all participants used English (EN channel) for the WSP game. From our given choice of items, Ko thought that item *lighter* was useful for making fires while Zh wondered if this were really possible. Zh thought that *chocolate* was the useless choice and asked if the others agreed or not. Then, the Japanese participant asked something about *shortening* in English, but the other could not understand the word “solve” in that sentence, since her English skill is

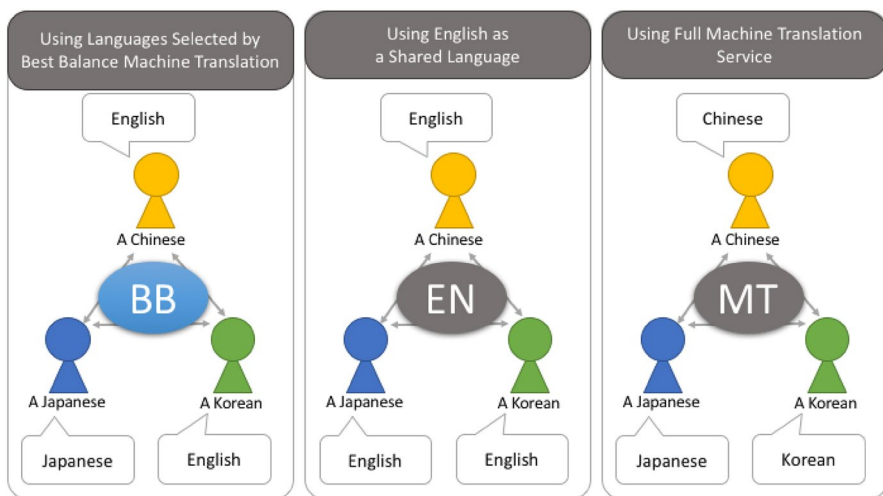


Fig. 2 Strategies of communication in this experiment

very limited. Sentences not understandable are normally ignored by other parties [10]. When a low-English skill participant inputs an incomprehensible sentence, sometimes the other participants simply ignore that sentence. Instead of asking about *shortening*, they continued the conversation without referring to what Ja said.
(Using EN Channel)

Ko We can make fire with *lighter* and *tree*
Zh But it is so cold and wet, I wonder if we can make it.
Zh Do you agree that the *chocolate* is the most useless one?
Ja can we solve *shortening*...?
Ja *chocolate* is most useful
Ko Wait a minute, we can get fire from crash

Less Engagement in Conversation of Japanese Users When Using English

When using English, Japanese users tend to be less active in the conversation. The same Japanese participant can be more active in the conversation when he/she communicated via MT. MT helps people with lower shared language skill worry less about what to say. It is easier for them to think in their own language and simply type in their mother language. Communication via MT can be more comfortable for the participants since it can provide more confidence in joining the conversation. With machine translation, low language skill participants could engage in the conversation more often and took less time to come up with a sentence. We can see the less engagement in the conversation by comparing the talkativeness, here measured by the number of utterances.

The percentage of utterances made by each participant is shown in Fig. 3, and the average percentage of utterances created by each nationality with similar English skill level is shown in Fig. 4. From both figures, the EN channel shows the most unequal participation in the conversation. The Japanese tended to talk much less when using

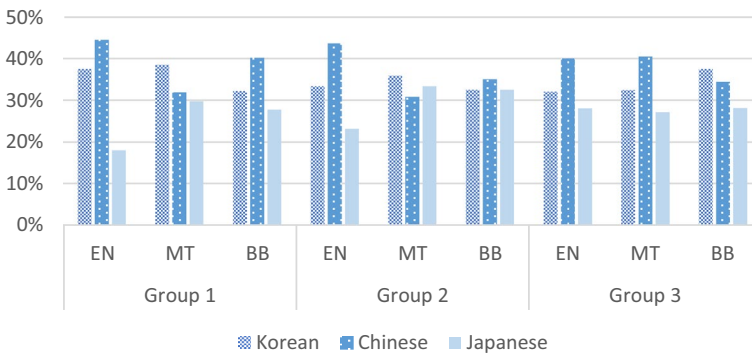


Fig. 3 Talkativeness of each participant in each group measured by percentage of utterances each participant made

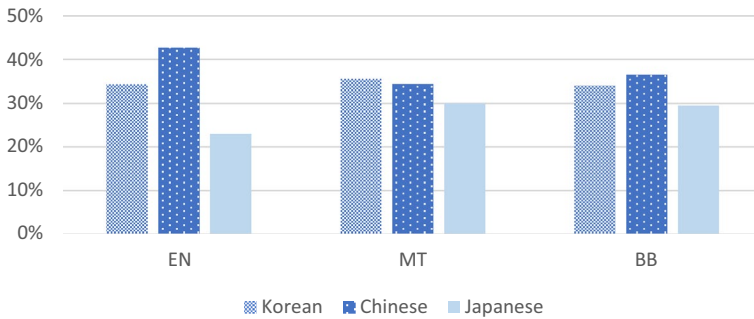


Fig. 4 Average talkativeness grouped by country of origin measured by percentage of utterances

English, while better balance was achieved when they used MT in MT channel and BB channel.

Conversation Encouragement

Sometimes, a participant had not written any reply for a long time in games using English Channel (EN). For example, the Japanese user was asked for her opinion many times at different times by the other users.

(Using EN channel)

15:37:39 **Zh** Ja, what do you think?

15:48:19 **Ko** How to you think about Ja?

15:57:42 **Zh** How about Ja?

The chat logs of each team were analyzed to discover why a participant stopped talking. The possible reasons are not understanding the current conversation, taking long time to express her opinion due to the language difficulties, having no opinion, or the participant's personality. MT use can help the participant to deal with the language difficulties, in terms of expression and understanding. MT use can also increase confidence when using her/his mother language. Asking for a specific participant's opinion appears much less often when the MT or BB channel is used.

Machine Translation Problem

Noticeable Machine Translation Problem When Using MT Channel

It is normal for machine translation output to contain many mistakes, although most of the time people can still understand what the other tried to say. However, using

the MT channel made mistakes more likely, which lead to more serious problems, such as miscomprehension or misunderstanding. Some translation errors can be severe enough to interrupt the flow of conversation.

The conversation in Fig. 5 is an example of an English translation of a conversation held in the MT channel. The original messages typed by the participant are highlighted in gray. The serious translation mistake is printed in bold font.

The participants had already discussed about the third item. But with the translation mistake, there was a misunderstanding. After discussing item number 3, the Korean participant thought that number 3 should be a radio transmitter–receiver, but the other two participants agreed on a mirror. The Chinese participant wanted to say, “Isn’t number 3 a mirror?”, without the question mark. Although question mark should be used in standard written modern, in conversation, it is totally understandable when reading the message even without question mark since the last letter “吗” already converts the sentence into a yes/no question. Unfortunately, the translation result, “Number 3 is not a mirror”, had the completely opposite meaning. This time, the translation mistake was so serious that the users noticed it. The Korean user also asked the Chinese user to explain it again. This kind of situation wastes time and disrupts conversation fluidity.

Conversation Breakdowns

Breakdowns are serious issues in communication as they interrupt the flow of conversation. Conversation breakdowns were common in the MT channels and EN channels. The number of breakdowns in English conversation made by group 1,

User	Message on Japanese Screen	Message on Chinese Screen	Message on Korean Screen
Zh	3番は鏡ではありません	3号不是镜子吗	3번은 거울이 아닙니다
	No. 3 is not a mirror	Isn't No. 3 a mirror?	No.3 is not a mirror.
	翻訳が誤った	翻译错误	번역이 틀렸다
	The translation was wrong.	The translation was wrong.	The translation was wrong.
Ko	申し訳ありません。翻訳が変でもう一度具体的に説明お願いしてもかまわないでしょうか？	对不起。翻译会奇怪可以再一次具体地说明拜托吗？	죄송합니다. 번역이 이상해서 다시 한 번 구체적으로 설명부탁드려도 될까요?
	Excuse me. The translation is strange. Would you please explain again specifically?	Excuse me. The translation will be strange. Would you please explain again in details?	Excuse me. Since the translation is strange, would you please explain it again in details?

Fig. 5 Part of English translation, original messages, and translated messages shown to each participant when MT was used

group 2, and group 3 was 20, 9, and 7, respectively, for a total of 36 breakdowns. With machine translation, the number of breakdowns was 17, 14, and 7 by groups 1 to 3, respectively, for a total of 38 breakdowns, while with BB, only 23 breakdowns occurred, 12 by group 1, 6 by group 2, and 5 by group 3. This indicates that issues were more frequent and serious when only machine translation was used and when only English was used.

Our investigation showed that when MT is used, machine translation errors triggered breakdowns. Wrong translation might not lead to breakdown but if the mistake is big enough to confuse the reader or the reader cannot understand what the writer said, breakdown is likely. When the topic being discussed was suddenly changed because of a machine translation problem, conversation can also fail.

Using only the EN channel can also cause breakdowns. The causes include misunderstanding due to a lack of language skill. Users with low English skill can make more language mistakes and if those mistakes are severe enough, breakdown is likely. The behavior of the participants with limited English skill can also trigger breakdowns, most often when the participant was too quiet for too long; another conversation topic was normally raised to address the participant's behavior which interrupted the original topic.

Relationship Between Breakdowns and QoM

A smaller number of breakdowns indicate a better flow of conversation. Fewer breakdowns can be linked to fewer problems in translation in the case of using MT and less misunderstanding due to the language proficiency problem when using EN. This reflects the vision of better QoM when both user skill and machine translation quality are taken into consideration.

As shown in Fig. 6, which shows the average number of breakdowns in each game, our model, using BB, has the highest average QoM from our previous calculation and yielded the fewest breakdowns on average. Using full translation causes the highest average breakdown rate and the lowest average value of QoM. The EN channel lay between BB and MT. These results show the relationship between the

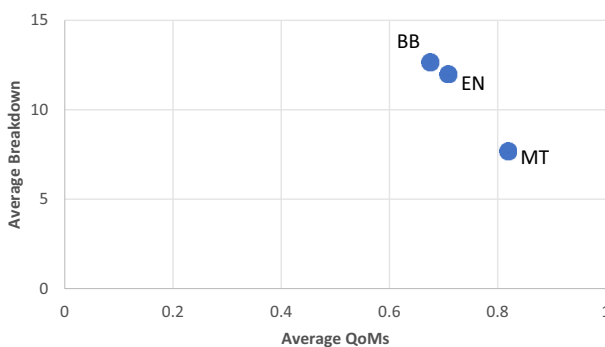


Fig. 6 Comparison between the average breakdowns and average QoM

number of breakdowns and the QoM value. Higher QoM, indicative of better message quality, is associated with fewer breakdowns and confirms better communication quality.

Discussion of the Model

To investigate the effectiveness of our proposed model, we compare talkativeness to the Quality of Message (QoM).

First, we calculate the coefficient of variation (CV) of QoM for each channel. From Table 3, C8 {en, en, en}, C1 {ja, ko, zh}, and C4 {ja, en, en} which were used in the experiment, we can see the CV of QoM in Fig. 7. The lower the CV value of QoM, the more equitable is the QoM of each user.

We also calculated the CV of average talkativeness, to see how evenly talkativeness was distributed. The lower the CV value of average talkativeness, the more equally the participants engage in the conversation. We can see a similar trend in that the variation in QoM and Talkativeness is much lower when the users employ the languages that they have medium to high skill in. From Fig. 7, using EN yields much higher CV values for QoM and Average Talkativeness than MT or BB.

The line in the graph is the trend that our model reflects. Unfortunately, due to the paucity of data, MT creates less imbalance, i.e., CV, of talkativeness than expected.

However, upon investigating the causes, we found that sometimes machine translation error caused small conversations that were not related to the game being played. CV of talkativeness might be different if we discounted the data associated with machine translation problems. If more accurate MT was used, the number of utterances related to problems with MT and overall talkativeness might be smaller, since some utterances are complaints about and attempts to fix the understanding problems caused by MT mistranslation. These utterances can be counted and included in talkativeness but it does not mean that the users are

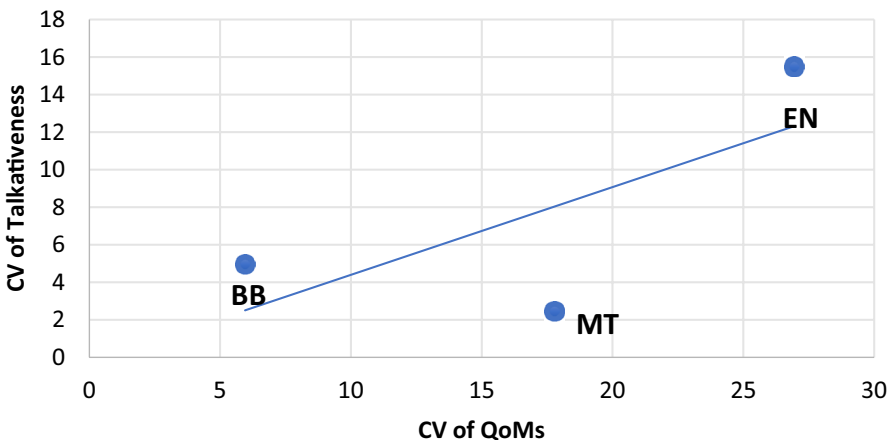


Fig. 7 Coefficient of variation of average talkativeness vs. coefficient of variation of QoMs

more talkative. However, in this experiment the number of utterances related to MT is very small. The change of average talkativeness before and after omitting the utterances related to MT is less than 0.8%, so it does not have a big effect on the CV values.

Combining the results in Figs. 6 and 7, the joint use of MT and BB creates good equality, allowing people to join the conversation more equally. Nevertheless, our proposed model, BB, has a big advantage over full machine translation. The best-balanced machine translation model minimizes conversation breakdowns since it avoids the use of low quality machine translation services, which tend to create translation mistakes.

Conclusion

Our main contribution is proposing the best-balanced machine translation model; it enhances multilingual communication via the selective use of machine translation. Using our model can help dealing both with imbalanced participation in conversations and machine translation problems created by low machine translation quality.

To confirm the validity of our proposal, we set up an experiment that compared our method to widely used methods of communication including using English as a shared foreign language and the full use of machine translation. In the experiment, our best-balanced machine translation method demonstrated better performance. Observations made during the experiment showed that utterances of participants who have limited skill in a shared foreign language increased when using machine translation services. Although the average percentage of utterances made by each participant when using machine translation and best balance is not significantly different, the fewer breakdowns recorded when our model is used indicate a better flow of conversation. When machine translation is simply used without considering the language ability of participants, more translation errors will occur, which will lead to interruption and misunderstanding.

Our model promotes creativity by enhancing communication quality as it allows people with different backgrounds to participate in conversations equally while minimizing the errors caused by machine translation as the users' language skills are taken into consideration. Our concept is to harness the intelligence of both machines and people to enhance multilingual communication.

Acknowledgements This research was partially supported by a Grant-in-Aid for Scientific Research (S) (24220002, 2012–2016) from Japan Society for the Promotion of Science (JSPS), and the Leading Graduates Schools Program, “Collaborative Graduate Program in Design” by the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

1. Ishida, T.: Intercultural collaboration and support systems: a brief history. In: International conference on principles and practice of multi-agent systems. Springer International Publishing, Phuket. pp. 3–19, (2016) https://doi.org/10.1007/978-3-319-44832-9_1
2. David, C.: English as a global language. Cambridge University Press, Cambridge (1997)
3. Henderson, J.K.: Language diversity in international management teams. *Int. Stud. Manag. Organ.* **35**, 66–82 (2005)
4. Morita, D., Ishida, T.: Collaborative translation by monolinguals with machine translators. In: 14th international Conference on Intelligent User Interfaces (IUI '09), pp. 361–366. ACM, New York (2009) <https://doi.org/10.1145/1502650.1502701>
5. Yamashita, N., Echenique, A., Ishida, T., Hautasaari, A.: Lost in transmittance: how transmission lag enhances and deteriorates multilingual collaboration. In: 16th ACM Conference on computer supported cooperative work and social computing (CSCW 2013), pp. 923–934. ACM, Texas (2013) <https://doi.org/10.1145/2441776.2441881>
6. Gao, G., Yamashita, N., Hautasaari A.M., Fussell, S.R.: Improving multilingual collaboration by displaying how non-native speakers use automated transcripts and bilingual dictionaries. In: 33rd Annual ACM Conf. on human factors in computing systems, pp. 3463–3472. ACM, Seoul (2015) <https://doi.org/10.1145/2702123.2702498>
7. Toyama, T., Sonntag, D., Dengel, A., Matsuda, T., Iwamura, M., Kise, K.: A mixed reality head-mounted text translation system using eye gaze input. In: 15th International Conference on intelligent user interfaces (IUI '10), pp. 329–334, ACM, Haifa (2014) <https://doi.org/10.1145/2557500.2557528>
8. Aiken, M.: Transterpreting multilingual electronic meetings. *Int. J. Manag. Inf. Syst.* **13**(1), 35–46 (2009). <https://doi.org/10.19030/ijmis.v13i1.4940>
9. Fellbaum, C.: WordNet. Theory and applications of ontology: computer applications, pp. 231–234. Springer, Dordrecht (2010). https://doi.org/10.1007/978-90-481-8847-5_10
10. Yamashita, N., Ishida, T.: Effects of machine translation on collaborative work. In: computer supported cooperative work and social computing 2006 (CSCW 2006), pp. 515–524. ACM, Alberta (2006) <https://doi.org/10.1145/1180875.1180955>
11. Morita, D., Ishida, T.: Collaborative translation by monolinguals with machine translators. *Procs. 14th IUI*, pp. 361–366 (2009)
12. Bramantoro, A., Ishida, T.: User-centered QoS in combining web services for interactive domain. In: 5th International Conference on semantics, knowledge and grid, pp. 41–48. IEEE Press, Guangdong (2009) <https://doi.org/10.1109/skg.2009.106>
13. Lafferty, J.C, Eady, P.M., Elmers, J.: The desert survival problem. *Exp. Learn. Methods Michigan* (1974)
14. Fermilab Project ARISE. <http://ed.fnal.gov/arise/guide.html>
15. NASA. http://www.nasa.gov/audience/foreducators/5-8/features/F_Exploration_Then_and_Now.html
16. Ishida, T.: The language grid: service-oriented collective intelligence for language resource interoperability. Springer Science & Business Media, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21178-2_1
17. Cambridge Assessment. <http://www.cambridgeenglish.org/exams-and-tests/cefr/>
18. ETS. <https://www.ets.org/toefl/institutions/scores/compare/>
19. Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C., and Schroeder, J.: (Meta-)evaluation of machine translation. In: 2nd Workshop on statistical machine translation, pp. 136–158. ACM, Prague (2007) <https://doi.org/10.3115/1626394.1626403>
20. Linguistic Data Consortium: Linguistic data annotation specification: assessment of fluency and adequacy in translations. Tech. Rep. (2005)