Improving Mutual Understanding in Machine Translation Mediated Communication

MONDHEERA PITUXCOOSUVARN

Improving Mutual Understanding in Machine Translation Mediated Communication

Mondheera Pituxcoosuvarn
Department of Social Informatics
Kyoto University, Japan

Improving Mutual Understanding in Machine Translation Mediated Communication

Doctoral Thesis Series of
Global Information Network Laboratory
Department of Social Informatics
Kyoto University

Abstract

Diversity provides a good amount of benefits for society and organizations, as it has a considerable impact on innovativeness. Therefore, diversity of cultural background, i.e., cultural difference, could lead to a wider variety of ideas, which is crucial in design.

Cultural diversity has significant advantages; however, it comes with difficulties in communication. With different cultural and language backgrounds, collaboration across cultures becomes troublesome. It is difficult but crucial to establish a common ground to have a fruitful collaboration. The lack of common ground could cause misunderstanding, and it is essential to maximize the opportunity to understand communication, regardless of their cultural or language background, to maximize the benefit of diversity.

Since establishing the common ground is essential, my research goal is to design a support system that promotes mutual understanding in intercultural collaboration and multilingual communication. This thesis aims to increase mutual understanding by focusing on preventing miscommunication and misunderstanding.

The most significant burden in intercultural collaboration is the language difference. It is almost impossible to communicate without an interpreter when people do not have a shared language. Fortunately, machine translation(MT) has been available for many decades, and it has been used in intercultural collaboration. However, because of its quality, mistranslation often happens, and it could cause misunderstanding. Besides mistranslation, the cultural background also plays a significant role in MT-mediate communication. Even when MT translation is correct, different cultural backgrounds made people think differently, and it obstructs them from having a mutual understanding.

In this thesis, we present three contributions toward improving mutual understanding in MT-mediated communication. Because it is necessary to prevent miscommunication and misunderstanding to have mutual understanding, first, we proposed BBMT to improve communication quality by decreasing miscommunication due to the unequal language burden and the unequal quality MT. Second, since BBMT computation needs private user data, i.e., language score, to be disclosed, we proposed a privacy-aware version of BBMT that protects user private data while still allow the system to compute BBMT. Lastly, we proposed a method to automatically detect words that could cause misunderstanding based on cultural difference, so it can warn the user and prevent misunderstanding when the user tries to use those words.

1. Best-balanced machine translation(BBMT)

   In MT-mediated communication, it is important that the users understand the translated message. If the user cannot understand the translated message, they cannot establish mutual un-

derstanding and might fail to collaborate. However, a lack of shared language and gaps in the language backgrounds of group members could bring about imbalanced participation, which is likely to hinder problem-solving, idea generation, and collaborative learning, among participants using various languages.

Our contribution here is to propose a model of best-balanced communication based on Pareto optimality of the Quality of Messages. Our model helps users to select languages and MT services that will maximize the quality of communication. We conducted a laboratory experiment to compare the communication channels using BBMT, entirely using MT, and using English as a shared language. We found that, with the BBMT channel, there was the least number of conversation breakdowns, usually cased by translation mistakes, comparing to English and full MT channels, which reflect effective communication. Both BBMT channel and full MT channel created a better balance in participation comparing to the English channel. From the experiment result, our proposed method successfully enables users to interact and communicate with better equality while minimizing the problems that can arise from machine translation usage.

2. A privacy-aware computation of BBMT
   The best-balanced machine translation mechanism (BBMT) can help us to assign proper languages to be used. However, the calculation process needs the information of participants' language skills, typically participants' language test scores. Since it is essential to keep test score confidential, as well as other private

information, we introduce agents, which exchange encrypted information and secure computation to ensure that agents can select appropriate languages and translation services without disclosing participants' scores. Our contribution is to introduce a multi-agent system and secured computing together into protecting the privacy of users in multilingual communication. To our best knowledge, it is the first attempt to do so. The central concept is to model interactions among agents who deal with user's private data and to distribute calculation tasks to three different types of agents, together with data encryption. We have made the security argument to confirm that none of the other agents can access or recover participants' scores except for the natural leakage of information, which is unavoidable even if the system is perfectly secured.

3. An automated detection of cultural difference
   To establish mutual understanding in intercultural collaboration, the team member must realize the misunderstanding. However, it is difficult for people to identify words that might cause misunderstanding. It is more difficult for monolingual and monocultural people since they have never experienced the language and culture of the other party. Many researchers have been trying to identify cultural differences using survey studies, but the resulting coverage is limited and requires excessive effort. Here, we propose a novel method that applies an image comparison technique to an image database to automatically detect words that might cause misunderstanding between two languages. We exam-

ine our method by on 2,500 sets of words in a Japanese-English concept dictionary called Japanese WordNet. Our result of low and high similarity sets reflects the high and low possibility of misunderstanding concepts. We provide explanations of the results gained. We also found that the low similarity sets are not only caused by cultural differences but also caused by inequality of meaning, homonymous, and usage as a specific noun in each language.

These contributions are beneficial for any party that wants to enhance MT-mediated communication and intercultural collaboration.

Acknowledgements

This thesis and my work during the Ph.D. study would not be possible without the help of many people. First and foremost, I would like to express my sincere gratitude to my former supervisor, Professor Toru Ishida, for the continuous support from my master's to my Ph.D. study and research. I greatly appreciate his encouragement, his patient, and immense knowledge, which always help me out whenever I encounter difficulties. I want to thank my current supervisor, Associate Professor Donghui Lin, who has been my supervisor for my last year of Ph.D. study, for encouraging my research and give me those precious comments since I have entered this lab. From both of my supervisors, I earned priceless advice for both research and my future career. I could not have imagined having a better supervisor and mentor for my Ph.D. study than both of you.

I also owe my sincere gratitude Professor Tatsuya Kawahara, Professor Hiroaki Ogata, Professor Takashi Kusimi, and Associate Professor Adam Jatowt, for serving as advisers to keep monitoring my research progress and providing useful comments and suggestions.

I would especially like to thank Associate Professor Takao Nakaguchi

for his close support to my research and for his help setting up software for many experiments. I have been very appreciated by his so much effort. I also thank Associate Professor Yohei Murakami for helping with the Language Grid server preparation.

I would like to show my gratitude to Associate Professor Shigeo Matsubara for his kind support. I also greatly appreciate the laboratory and Design School coordinators, Ms. Terumi Kosugi, Ms. Hiroko Yamaguchi, and Ms. Yoko Kubota, for their help in administrative affairs.

I sincerely express my gratitude to Professor Masayuki Abe for a very important lecture and discussion on security and Professor Masaaki Iiyama for his excellent lecture on image comparison.

I sincerely thanks all of my co-author, Dr. Naomi Yamashita, Ms. Yumiko Mori, and Dr. Toshiyuki Takasaki, for their commitment to work with me on my research papers..

Special thanks also go to all my Ph.D. friends and previous lab mates: Kyosuke Yamashita, Huichen Chou, Daichi Hamada, Shiyao Ding, Andrew W. Vargo, Mairidan Wushouer, Amit Pariyar, Kemas Muslim Lhaksmana, Xin Zhou, Shinsuke Goto, Hiroaki Kingetsu, Xun Cao, Nguyen Cao Hong Ngoc, Arbi Haza Nasution, Victoria Abou Khalil, and many others. I am happy to be a part of this beautiful lab with wonderful people.

This thesis is presented to my beloved husband Kai Ihara, who spent a lot of effort on being a good partner of me in my daily life, and my two lovely sons, Lennon and Sage, for their charming smiles and refreshing hugs.

This research was partially supported by a Grant-in-Aid for Scientific Research (A) (17H00759, 2017-2020) 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. My third year of stay and tuition in the Ph.D. program was supported by the Japanese Government (Monbukagakusho) Scholarship.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter gives an overview explanation of how mutual understanding is important in machine translation(MT) mediated communication. Later in this chapter, we also describe our research objectives, issues, and approaches to solve those issues. Lastly, we provide an outline of this thesis.

1.1 Overview

The world is becoming borderless. With faster and cheaper transportations, going abroad becomes easier. Moreover, with the Internet, It transforms the complex relationships between local activities and interaction across distance. With new opportunities and dilemmas[Slevin, 2007], there are more opportunities to collaborate.

Intercultural collaboration creates greater diversity in organizations and

design team. For organizations, diversity provides vast benefits. Diversity of top management can strongly improve innovativeness and firm performance [Talke et al., 2010]. Besides top management diversity, racial diversity in the general workforce is also connected to increased sales revenue, more customers, more significant market share, and greater relative profits [Herring, 2009]. From these studies, we can infer that cultural difference, which is diversity in the cultural backgrounds, is important and beneficial for organizations and teams. It could lead to a broader range of ideas, and some of those ideas could lead to promising innovations.

However, language differences and cultural differences are significant issues in intercultural collaboration. Even though diversity has many advantages, it also raises difficulties in communication. With different cultural and language backgrounds, communication and collaboration become challenging. Nowadays, support tools and services are available for multilingual communication [Ishida et al., 2018]. MT is now available and helps to offset translation problems when there is no bilingual human or translator around. Various MT embedded chat systems have also been developed to support communication across languages. Unfortunately, MT is not perfect and can cause difficulties in communication, including misunderstanding due to mistranslation, cause conversation breakdown, and difficulties in establishing mutual understanding [Yamashita et al., 2009], etc. An existing study suggested that mutual understanding is a state that arises from effective communication processes [Margaret, 1994], so it is essential to establish a mutual understanding to achieve effective communication.

Because of the importance and difficulties of creating a mutual understanding in communication, this research aim at improving mutual understanding in MT-mediated intercultural collaboration.

1.2 Objectives

From our motivation stated earlier, this thesis aims to create tools to support intercultural collaboration by improving mutual understanding, given that MT is available. We focus on designing a collaboration environment where the users can effectively understand and share their opinions mutually. Toward the design of environment for intercultural collaboration based on the two major problems that prevent mutual understanding, the objectives of this thesis are as follows:

1. To provide a method to optimize the quality of communication in MT-mediated communication

   Since MT services are available, but their abilities to translate in every language are not yet perfect. An effective design of how these tools are used could improve the quality of overall communication.

2. To provide a method to create a common ground for multilingual communication

   Because of cultural difference, grounding in intercultural collaboration is difficult. Creating a common ground is an important approach to have a fruitful conversation that is understandable for every party.

1.3 Issues and Approaches

To design our collaborative environment for intercultural collaboration, we listed three issues and approaches to solve those issues:

1. **How to select MT services and languages to be used, to maximize the quality of communication?** Even though MT is available, using MT can also cause conversation breakdown when translation result goes wrong [Yamashita and Ishida, 2006]. Many researchers attempt to improve communication by improving the quality of machine translation [King and Redpath, 2001, Wang et al., 2007, Mikolov et al., 2013, Drewes, 2009], while some researchers use human intelligence to improve the quality of communication [Morita and Ishida, 2009]. However, none of the existing work has ever considered adjusting language channels using the foreign language ability of the users themselves with the quality of MT. When users speak many languages, choosing to choose one language over another could be more beneficial to the overall communication. Especially when MT is available, there are even more choices of language medium to be used. Sometimes, MT quality is low, but a user foreign language could be good enough to have even better communication than using MT. And when there are more than two users, and each one could speak more than one language, it is even more complicated to choose the languages and MT services that yield the best communication. Our approach to this issue is, first, list all the possible communication channels. Then, measure the values called quality

of messages in each combination of languages and services. Quality of message is calculated using the message sender writing skill, MT quality, and message reader reading skill. And finally, select the best combination of languages and services by selecting the Pareto optimal set of quality of messages. This solution is called best-balanced machine translation(BBMT)

2. **How to computing BBMT without disclosing user language score?** Our first version of BBMT can solve the language selection issue; however, the computation process requires the user language skill. We need user foreign language scores of standard examinations, and some users do not wish to disclose their scores. There exists a various method to protect user privacy[Jain and Bhandare, 2011, Sweeney, 2002, Yi et al., 2015]; however, none of them is suitable for BBMT and enable the BBMT calculation. Hence, we tried to alter the calculation process to ensure that the user language score will not be sent away for the calculation. Language scores will be calculated and encrypted at each user's personal computer instead of being sent to the central computer as in the original BBMT. However, encryption and self calculation are not enough for BBMT computation since the process of selecting a Pareto optimal language combination is more complex. We use two additional agents to deal with the processes. The computation tasks are distributed to at least four agents, so none of them will be able to see or guess the user language score except its own user's score.

3. **How to detect the cultural difference and possible mis-**

understanding? For human to detect and identify a difference, experiences related to both cultures are needed. Most researchers use human experiences and effort to study cultural difference [Hofstede, 1984, Yoshino and Hayashi, 2002, Cho et al., 2007]. For machines, it could be possible to process existing data to look for the difference. Our proposal deals with this issue by using an existing image database and a corpus. Images are linked to keywords in different languages. But the same word in two different languages can be linked to a different type of images if the speaker of those languages thinks about the words differently. Hence, it is possible to compare the images that are linked to a word and its translation in another language by looking for the difference that could cause misunderstanding between the speakers of two languages.

1.4 Thesis Outline

This thesis is organized into seven chapters, including Chapter 1. The content of each of the remaining chapters is summarized below.

Chapter 2 summarizes the problems and solutions in intercultural collaboration from the existing studies. We studied the difficulties in Intercultural collaboration and methods that people are using to cope with the intercultural and language difference issues. We also report our ethnographic study on a cross-cultural children workshop and discussed the problems we found. We also give design implications to solve that problem.

Based on the problems we studied, Chapter 3 proposes a solution to help to balance the language burden by providing a method called best-balanced machine translation(BBMT). BBMT can recommend suitable languages combination and services to be used in multilingual communication. Later in this chapter, we report the experiment result we conducted to study the effect of BBMT, comparing to full usage of MT, and using English as a shared language.

Chapter 4 presents a privacy-aware version of BBMT, so private user information, i.e., test score, will not be disclosed during BBMT calculation. We introduced the methods of calculation that combine secure computation technics and multi-agent systems to intercultural collaboration. We also made a security argument to prove the privacy protection effect.

Chapter 5 shifts the focus on creating mutual understanding from the language aspect to involve more cultural elements. We propose a method to automatically detect words that could cause misunderstanding when it is used in multilingual communication. We also discuss the result, the application, and its limitations.

In Chapter 6, we applied the two aspects of supporting collaboration with mutual understanding together and implemented them in a workshop. We describe our experimental design and discuss the result we found.

Finally, Chapter 7 concludes this thesis by discussing the summary of contributions made for improving mutual understanding in MT-mediated communication and suggesting possible future directions.

Chapter 2

Problems and Solutions in Machine Translation Mediated Communication

In the design process, the first process is to sense the intent to define the problems, following by information collection about context and user [Kumar, 2012]. Since we are designing a support system for intercultural collaboration, it is crucial to study the usage and problems of MT-mediated communication both on the surface and in-depth to develop an effective support system. This chapter discusses how MT is used to enable intercultural collaboration and the difficulties of using MT in multilingual communication. To study how MT is used and its problems in real-world collaboration, we conducted an ethnographic study of an intercultural children workshop. In this chapter, we report the ethnographic study results and design implications based on the

results.

2.1 Introduction

To create new ideas, gathering ideas from people with different backgrounds, including cultural and language backgrounds, bring about different ideas that lead to creativity. Unfortunately, those background diversity comes with difficulties in communication, especially when people speak a different language.

Most of the time, people learn English to collaborate [Ishida, 2016], because English is now the global language. We can often hear people around the world talk in English [Crystal, 2012]. However, in international discussions, native speakers might have advantages over the non-native speaker and do not lead to high-quality results. The big gap in language skills can damage opportunities for non-native speakers to participate in intercultural communication. When English is used in a group with language diversity, it can affect socialization and interpretation since it can become a hidden barrier. Non-native speakers sometimes receive negative assessments because of their low language skills. Moreover, their intelligence be underestimated because they speak slowly [Henderson, 2005].

There are various methods developed to help non-native speakers to join in the conversation with native English speakers. One approach is creating artificial delays to help the non-native speaker understand the conversation before continuing the conversation [Morita and Ishida,

2009]. Other methods include signaling the native speaker about the status of non-native speakers [Yamashita et al., 2013], helping non-native writers by providing vocabulary navigation [Gao et al., 2015], and providing real-time translation using eye-gaze input of the non-native reader [Toyama et al., 2014]. Even though these methods can reduce the burden of non-native speakers, they cannot give a wholly balanced communication environment. Since different languages are the main barrier to the collaboration of multilingual groups. MT services are now available and have been used as support systems [Ishida, 2016]. They allow a multilingual team to work together without having a shared language.

2.2 Machine Translation and Its Problems in Intercultural Collaboration

Although MT is a useful tool for multilingual communication, it can still create difficulties due to its unreliable quality. Yamashita et al. [Yamashita and Ishida, 2006] studied how MT affects human communication. They gave pairs of users two sets of ten tangram figures, which were placed in different sequences. The users were instructed to match the arrangements using an MT tool. They found that using MT lead to asymmetries in the machine translation process, which yielded trouble in identifying tangrams and sequences through the expressions used and accepted. The asymmetric quality of each MT service can also cause difficulties, especially for low language resource users who

cannot converse very well, because they cannot understand messages and communicate correctly. Hida [Hida, 2016] studied the children workshop called KISSY in 2014 and 2015. He suggested that issues were present in the children's communication and collaboration. One such problem is some of the messages were incomprehensible when mistranslation made my MT occurred too often. However, still, more MT problems remain to be studied.

With nowadays technologies, the quality of MT is not yet perfect as many researchers tried to improve MT-mediated communication by evaluating [Del Gaudio et al., 2016, Scarton and Specia, 2016] and improving MT quality [Yamashita and Ishida, 2006, Shi et al., 2013, Avramidis et al., 2012].

2.3 Field Observation

Some researchers studied how MT is used in general [Shigenobu, 2007, Hara and Iqbal, 2015] , but how MT supports users in real-world face-to-face communication remains an unstudied area. To fill this gap, we conducted a field study at an event called Kyoto Intercultural Summer School of Youth or KISSY.

Pangea, a non-profit organization (NPO), organizes this event once a year. Its goal is to encourage children to develop social bonds across boundaries and motivate them to communicate with children from different countries with different languages. KISSY is an event that encourages children from different countries to collaborate by working on

a shared project using KISSY tool, which is a machine translation tool. It augments the face-to-face communication established among children and staffs with different language backgrounds.

Children aging 8 years old to 14 years old from different countries gathered together at a university to participate in a workshop and collaborate with each other with no foreign language skills being required.

The main task for the workshop was to create a short clay animation using clay figures. The participants were asked to create a story from one or two given objects: one brown rectangular block, and one white clay piece shaped like a bottle gourd. Each team had to create a scenario, model the clay, take photos, draw backgrounds, record sound effects, assemble the results, and edit the videos.

The workspace for each team was separated, but within the same hall, partitions were not used. Each team had a main table for discussions around a laptop PCs; the children sat in a U-shape facing the middle of the table, see Figure 2.1. There was a shared screen linked to the team leader's PC, who sat next to the screen. There was a table for clay sculpting next to the photo booth.

Each group had its own editing table with two PCs for sound and video editing. The positions of the photo booth, clay work table, and editor PC table were similar but slightly different for each team. The order of participant's seating within the group was changed a few times during the four-day workshop. In addition to the team workspace, there was also a space for administrative use, for example, to distribute equipment in different parts of the activity hall. In this area, also had a small tent

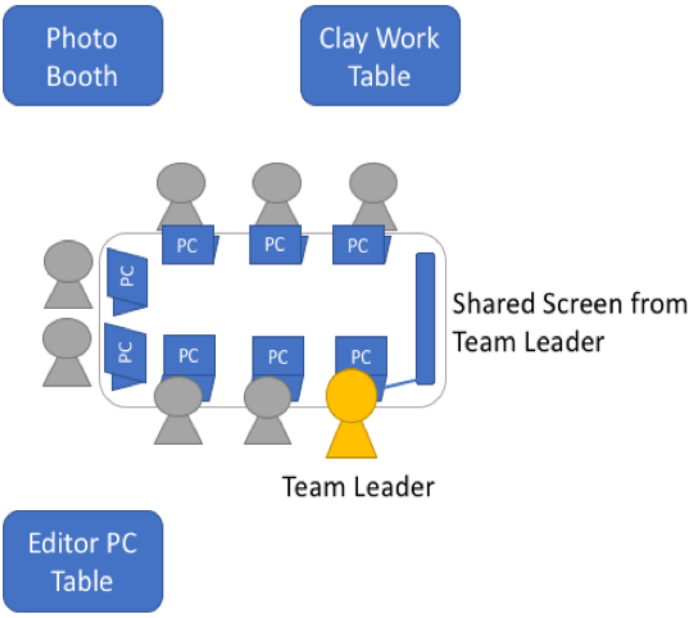


Figure 2.1: Sketch of team work space.

some distance from the team workspace for making sound and voice recordings.

2.3.1 KISSY Tool

KISSY tool is a web application with various functions to support multilingual collaboration; it was created specifically for KISSY. In order to use the system, each participant was provided with a laptop PC and internet connection. Each user was given a username and a password to access the system. They could choose the interface language that they were comfortable with, from English, Khmer (Cambodian language), Korean, and Japanese.

The main function was the multilingual chatroom, where the user could input text in her/his language and see everything in her/his language while the other users saw the text in their own languages. It has two interfaces for sending messages and reading messages in team chatroom. One of them is a general chat interface, the arrival of a new message pushes out the oldest message being shown, see Figure 2.2 . Another chat interface gave everyone their own space in different boxes. Messages from each participant appeared only in each person' s box, as in Figure 2.3. This function makes it easy to see the messages of all users at a time. And Figure 2.4 shows ideaboard, the interface of another function of KISSY tool that was used often. The team leader could use ideaboard to pose a question and the children could express their ideas by typing in virtual responses in their own language. Each member could vote (click) for one favorite idea per question.

MT services used by KISSY tool were provided by LanguageGrid [Ishida, 2006, Ishida, 2011]. Services selected for each pair of languages were provided by GoogleTranslate.

2.3.2 Study Method

We conducted an ethnographic study by observing the participants and staff at KISSY. Ethnography is the most basic form of social research [Hammersley and Atkinson, 2007]. It is a method, most often used in anthropology, that involves encounters, respecting, recording, understanding, and representing human experience [Willis and Trondman, 2000]. This method is now practiced in various discipline.

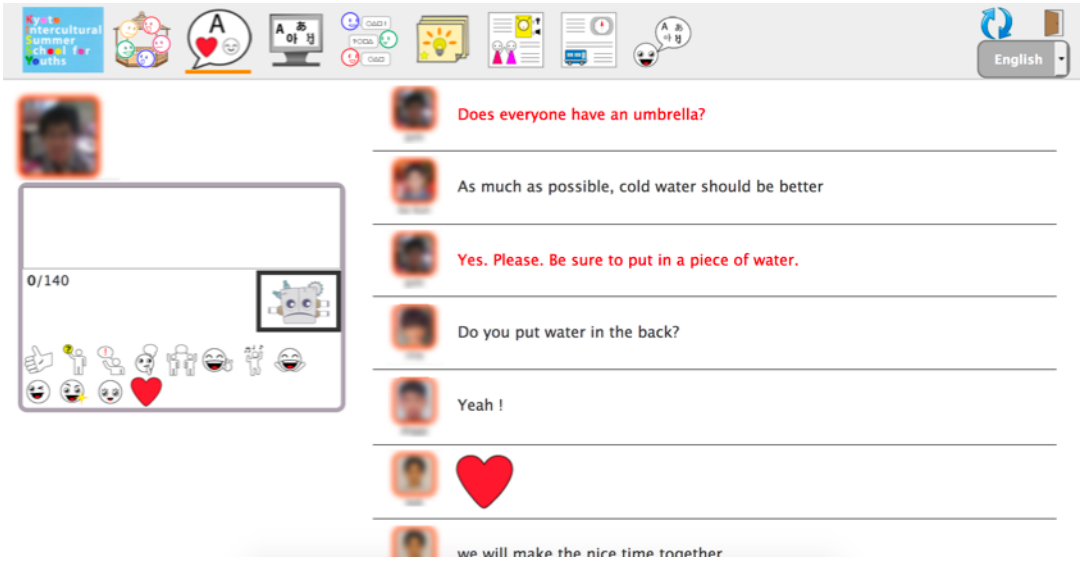


Figure 2.2: A screen shot of KISSY tool team chatroom.

Of the five teams participating in KISSY, two teams were observed for this study. Each team consisted of one adult facilitator, called team leader, and seven participants. Team Red had a Japanese team leader two Korean children, four Japanese children, and one Cambodian child. Team Green had a Japanese team leader, three Korean children, three Japanese children, and a Cambodian child. Korean children could communicate in simple English, while most Japanese children could not communicate in English. Cambodian children also could speak very little English.

Videos were recorded from afar to minimize interference with the activities and to encourage the participants to relax and react in a natural manner. The video were analyzed in our laboratory.

After the event, we conducted face-to-face and online interviews. For

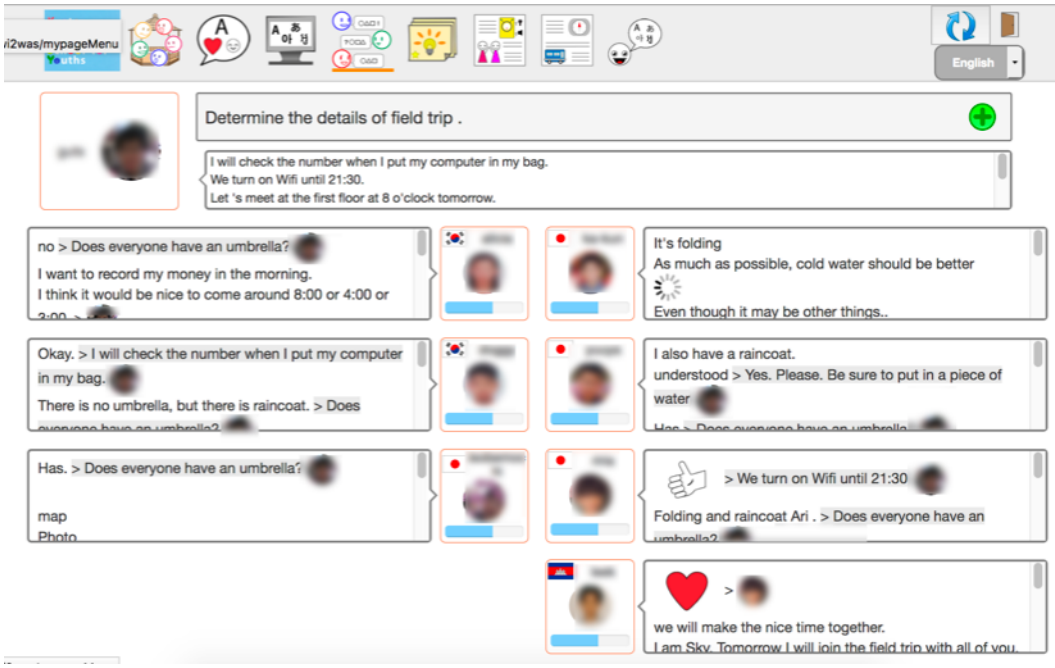


Figure 2.3: A screen shot of discussion screen.

each team, we interviewed two children and the team leader. We also interviewed bilingual and trilingual staff who were not part of any team but helped with the translation

2.3.3 Coping with Mistranslation

During the workshop, when a mistranslation occurred, users tried to solve the problem both by themselves and by asking a human interpreter for help, as shown in Figure 2.5. The chart is discussed in detail in the following subsections.

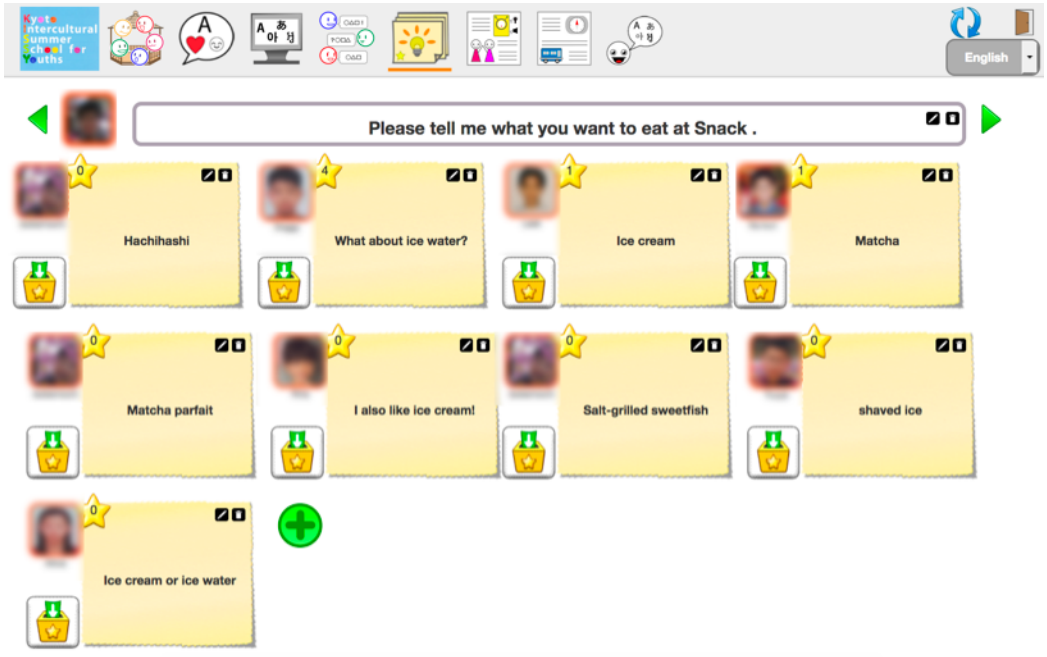


Figure 2.4: A screen shot of ideaboard.

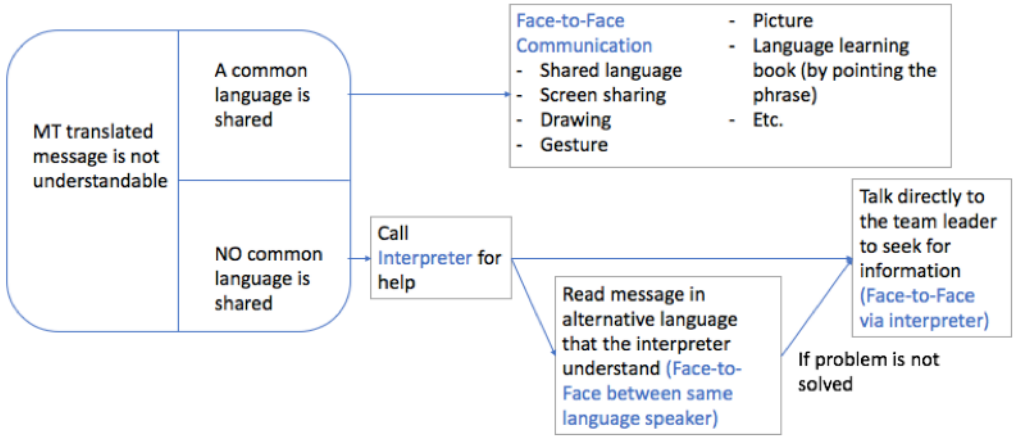


Figure 2.5: Behavior of users when MT translated message was not understandable.

Alternative Channels of Communication

When an MT problem occurred and users shared a language, even if the language skill was low, they turned into face-to-face communication. As shown in Fig. 3, the alternative methods included using shared language, other media including screen, drawing, gesture, picture, and language learning books. Many times, they used more than one communication channel at a time, usually including trying to use their shared language even though their shared language skill was low. Here is an example of MT failure and the user's response. The paragraphs in italic are from the ethnography transcript.

During role decision of Team Green, team leader(TL2) asked everybody a question via the ideaboard function. He typed this question in Japanese "どの役割がいいですか？" which basically means "what role do you want to take?". It appeared on the Cambodian child(C2)'s screen as "A position that is this good?" in Khmer. C2 did not understand. Since she sat next to TL2, she suddenly turned to him and TL2 noticed that she could not understand the question. He switched his tool into English language mode and tried to speak with her in English. She understood. The translation in English was shown as "Which role should I take?". TL2 noticed the mistranslation so he suddenly fixed it by saying "Not me, but you". In this situation, they both tried to communicate, using gesture when C2 turned to TL2, alternative translation on TL2's screen, and then English.

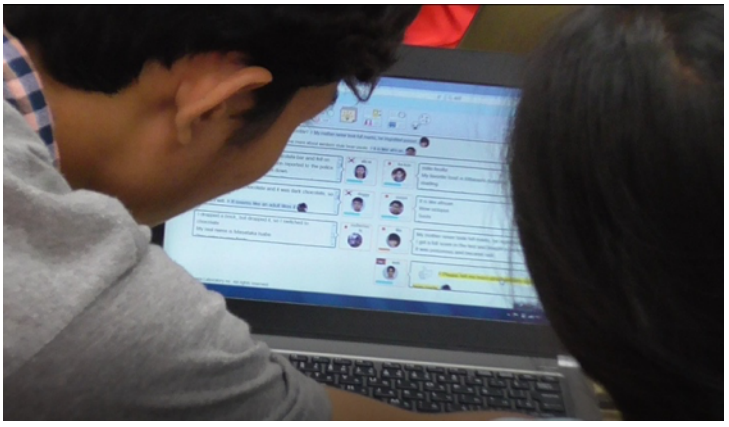


Figure 2.6: Cambodian staff helped a Cambodian child by reading English translation instead of Khmer translation.

Asking Others

Many times, the MT tool failed to enable users to communicate, especially for the user of low resource language (Cambodian children). There was only one Cambodian staff to support three Cambodian children in three different groups.

In the interview, the Cambodian staff stated that "They only understood Khmer language and the translations on the screen in Khmer weren't correct. It was almost correct if English was translated into Khmer. But they typed in Korean or Japanese so this is the problem for translation, I think. The kids need more help from me in this activity". As shown in Figure 2.6 and the interview, Cambodian staff helped the children to understand messages by switching the interface to English and reading English texts instead of Khmer. When the translated messages were incomprehensible, the communication changed from MT among participants and team leader to face-to-face with a human in-

terpreter. When the bilingual staff helped, he normally used one of two options.

We asked how he helped them when the Khmer message was difficult to understand. He answered, "Sometimes I checked in English, if I still did not understand because of translation from Japanese or Korean to English, I asked the team leader to explain it to me in English" .

First, the bilingual staff read the messages and tried to understand the message by reading the message translated into other language. For example, the Cambodian staff, who could speak both Khmer and English, helped the Cambodian kid by reading messages in English and translated for her, since English translation had higher quality and thus easier to understand.

If the staff still could not understand the message, he asked the team leader directly in their shared language and translated his understanding to the kid.

Understand Culturally Dependent Context

Sometimes, the children sometimes could not understand each other when the messages or words depended on the culture of the originator. At the beginning of the the workshop, our staff showed the object in Figure 2.7, a brown rectangular block, to the children and asked them what did they think this block looked like.

One of the Japanese used KISSY tool to say that it looked like 'Anko' or Japanese style red bean paste. People from a different culture could not

Figure 2.7: A clay block shown to the children.

understand the comment, because red bean paste in the other countries does not look like a block. Moreover, the Khmer MT translation was "something made by Japanese", which was incorrect. Khmer speaker could not understand the whole translated phrase.

Since the MT translated message was incomprehensible, face-to-face communication was needed and a human interpreter was also needed to help the children create mutual understanding of the object. As shown in Figure 2.8, the Cambodian staff helped the Cambodian participant to search for photos of 'Anko'. It helped them understand why the Japanese participant referred to the object as red bean paste.

Another similar situation arose when the name of thing that exists in one language meant something different in another language. During the self-introduction period, when the children were talking about movies, even though MT could translate the movie names correctly, many movies have specific names in different languages. As a consequence, the users had trouble understanding which movies was being referred to. In this case, as well, they solved the problem by searching

Figure 2.8: Cambodian staff helped the Cambodian child to search for 'Anko' picture.

for images or posters of the movie to show the others.

Tool Navigation and Instruction

During the workshop, the children could use different software or different windows in the KISSY tool. The tool includes many functions for example, ideaboard, chatroom, etc. Working as a team requires that everyone be on the same page for collaboration to succeed.

In the workshop, TL2 had just finished creating a new question on ideaboard but one Korean child(K1) was using a webpage outside the KISSY tool. Since K1 was not on the chatroom page, she did not know what was going on and which page should be viewed at that moment. TL2 told an English-Korean interpreter(I3) to tell K1 "Go to the next question" .

As shown in Figure 2.9, the team leader wanted everybody to look at the ideaboard after he finished entering a question. Because K1 was

Figure 2.9: Interpreter (I3) , on the left-hand side, helped team leader (TL2), the man on the right, to talk to (K1), the girl who is using a PC on the left, about what to do next.

not viewing the chatroom screen, it was impossible for her to read messages even if TL2 sent a message to the chatroom for everybody to navigate to ideaboard. In this case, direct face-to-face communication or communication via human interpreter was needed. Because MT cannot cope with this kind of situation, an alternative communication channel was used, and an interpreter was needed.

Sometimes, the team leader communicated non-verbally by showing the page on the shared screen and pointing to the page so the children could follow him on the same page. In another example, when the users were not familiar with the tool or when the users faced difficulties with the tool, they shifted to face-to-face communication, as in Figure 2.10.

‘At the very beginning of the workshop, while the other children are voting for the idea they like. C2 was still not sure what to do. TL2

Figure 2.10: Team leader (TL2), on the right sitting next to the shared screen, communicated directly to the Cambodian Child (C2), the girl next to him.

had to point at her screen and ask her "Can you choose one?" verbally in English.'

In this case, users could not communicate using MT in the KISSY tool. Hence, conversation in shared language was needed when there was no human interpreter.

Substituting Machine Translation

In many situations, MT was not used. This section describes situations wherein other communication methods were preferred over MT.

1. Using Common Words and Signs

   In the workshop, when there was a common word among all lan-

guages or there was a simple word that could be understood by everyone. For some words, i.e. "Okay" or an object that everybody understood such as "Soba", "Sushi", face-to-face communication was often used. Non-verbal communication was also used. For example, pointing index finger down at the keyboard meant "Vote!", and clapping hands expressed that something has been done. This kind of gesture can be easily understood by every member.

2. Involving Physical Objects in Communication
   When the communication involved physical objects, MT was used less often. The following situation is from the video taken in the morning of the third day of the workshop.

   Team Red had a short meeting before working separately. In the meeting, before team leader (TL1) explained the work plan of that day using a physical board with written papers on it, he called a Korean-Japanese interpreter(I1) and a Khmer-English interpreter(I2) to help him. Then he explained the work, mostly in Japanese and sometimes in English, while pointing on the board from time to time. I1 translated what TL1 said from Japanese to Korean for Korean Children in parallel. TL1 spoke English later but not for all messages in Japanese.

   Instead of using MT, TL1 decided to ask for interpreter's help and speak directly in this mother language. In this case, using machine translation would make the use of gesture and the physical board difficult, especially when the MT tool requires typing.

However, not using MT can cause the inequality in successful receipt of messages. In this situation, I2 could not speak Japanese so he had to wait until TL1 spoke English to him. TL1 spoke much shorter sentences in English due to his limited ability to communicate in English and this prevented C1 from understanding the whole meeting while the Japanese and Korean children could understand more quickly about what was going on.

2.3.4 Discussion

From our investigation of KISSY 2017, there are problems that are deeper than mistranslation.

1. Low Language Resource User Support
   Good collaboration should have team members equally and actively participate in the conversation and activity. In KISSY, the low language resource users, the Cambodians, faced the biggest barrier to participation. One problem was the quality of Khmer MT. Since the language is low resource for translation, the messages translated from and to Khmer are difficult to understand and sometimes incomprehensible. When the messages on the screen do not make sense, it is difficult to know what are people talking about and it is almost impossible to talk or express one's idea in this situation.

2. Problem Detection
   Another problem as a consequence of low language resource user support is, it is difficult to detect the problem when a user needs

help. If the children know they need help and ask for help, their problem can be solved easier, but many times they did not recognize they needed help. During the workshop, many times the team leaders had to identify who needed helps and then ask the interpreters to help the children. The interpreters also looked around and checked if any of the children needed help. However, if help is not promptly available, it will raise difficulties with their participation.

3. Human Interpreter Task Overload

   When the children had to solve their understanding problems whether by using alternative communication methods or asking the interpreter for translation help, the communications were usually one-to-one. Unfortunately, KISSY multilingual chat is inherently not suitable if the goal of communication is one-to-one. At the workshop, when the team leader and the Cambodian child wanted to communicate one-to-one, the Cambodian interpreter was called to help with the translation. The number of human resource or staffs is limited, especially for the minority languages. One-to-one communication without multilingual tool support increased the need for human interpreters. Because of the problems caused by MT and others, the Cambodian children needed a lot more help than the other children. The Cambodian interpreter also reported that he could not manage to help all the children at the same time. Many times, the children need his help but saw him busy with others; the children did not want to disturb him and so waited until he was free. The time spent waiting delayed

their participation.

2.3.5 Design Implication

Design implication to support communication with low MT quality

1. Image Browser in Multilingual Chatroom
   As mentioned with regard to culturally-dependent context, the participants tried to search for images and show them to the other participants to help their understanding. However, it was not convenient to search and share images since no shared display was provided, other than that controlled by the team leader. Adding a shared image browsing function to the tool could save time and raise user effectiveness. It would also encourage the users to use more photos or figures to express their ideas and to understand the others. This design guideline can also help to solve the problem of understanding the culturally dependent context that cannot be explained easily by words, for example, travel attractions, and ethnic foods.

2. Interpreter Calling Function with Prioritization.
   One of the main problems for minority users was the paucity of MT and human resources, since MT quality is low for low resource-languages and it is difficult to find speakers of the minority language to support the children. Predicting the help needed would be a useful function as children often failed to notice that

they needed help or were shy in asking for help. A prediction model could be made by timing the periods of inactivity of the participant and altering the interpreter would help too. Raising the priority of users who have been idle or who need more help might be useful. Developing a language profile of each member of the team is also possible. If there are two children waiting for help but one of them can speak better English as a shared language with the team leader, that person might need less help. The flag for help can also be sent to the team leader since sometimes the team leader checked the progress of minority users and tried to communicate directly or call for the interpreter to help them. Showing Translated Result in Known Foreign Language in Parallel.

In the workshop, non-native English speaker staff helped the children to understand written messages by reading messages in English, instead of reading message in his own language with mistranslated messages, and translating it to the children. Even though a user might have limited skill of second language, it is still possible that messages translated in second language could be more readable than low quality messages in the main language. Hence, showing both results, those in the user's language and those in the user's foreign language, could increase the probability of understanding messages.

Design implication for more convenient communication

1. MT for 1:1

   Even though collaboration tools should focus on group communication, sometimes one-to-one communication is also needed to run the team activity, as mentioned already with regard to human interpreter task overload. Providing 1:1 translations by stand-along portable devices would be extremely useful if permitted by the group activities. Such a function would allow members to share messages directly without having to involve everybody. This will help to reduce the costs created by the human interpreters.

2. Graphic Signs and Keywords for Changing Method of Communication

   We already noted that sometimes users used common words and signs to substitute for MT. Better communication requires the greater use of a common language. Images yield better and easier understanding even if the users do not speak the same language. Thus, ideograms might be useful for multilingual collaboration without a shared language. For example, showing the chatroom logo to the children would ensure that they turned to the chatroom.

   Having some basic shared keywords is another way to create easier communication. We can give a list of keywords with their translations and pronunciations in their languages to the participants before the workshop starts. The keywords can be those that are often used in the event, for instance, for KISSY, "Let's

vote!" as "Vote" .

2.4 Conclusion

We reviewed existing problems and some solution and reported on a field study of machine translation (MT) usage in a social collaboration event for children. The children were asked to conduct a project using KISSY tool, an MT embedded system for multilingual communication.

In the workshop, participants and staffs faced various types of problems due to and related to the use of MT. They chose alternative communication methods when they could not understand the translated messages. The alternative methods involve using a shared language, screen sharing, drawing, gesture, picture, etc. When problems arose due to cultural differences or culturally-dependent words, they turned to an interpreter for help and/or used web browsers to search for related photos to increase understanding or to confirm the understanding of the others. They also needed to communicate via face-to-face methods, when, for example, one or more users were not on the chatroom page, because they could not read the intstruction messages.

Some problems have yet to be solved. Better support for low language resource users is still needed. The interpreters can be become overloaded, particularly for low resource languages.

Finally, we drew a few design implications based on our study. We suggest that the future designs should consider the inclusion of image browsers to assist user understanding, a 1:1 translation function in

addition to the group chat, an interpreter calling function with priority, and the use of common keywords or images, to be used together with MT. Showing the translation results in the user's second language in parallel with her/his mother language could also be effective if the user's first language is a low resource language or machine translation quality is low.

Chapter 3

Best-balanced Machine Translation

To establish mutual understanding, it is important for everybody to understand the messages and be able to communicate effectively, however, the unequal quality of MT prevents the user to achieve that. This chapter discusses a balancing problem in MT-mediated communication regarding to MT quality. We provide a scenario when it is difficult to choose MT services and how it might cause imbalance in understanding and participation. To cope with this problem, we provides a solution called best-balanced machine translation(BBMT). Here, we also report an experiment we conducted to evaluate our BBMT proposal.

3.1 Introduction

To collaborate across culture, people usually study a shared language to communicate. Most of the time, people use English to collaborate [Ishida, 2016], because English is now the global language [Crystal, 2012]. However, the difference of language skills can reduce the opportunities for non-native speakers to participate in the communication. Moreover, sometimes non-native speakers receive negative assessments because their foreign language skills are low. Their intelligence be underestimated because they speak slowly [Henderson, 2005].

There are many researchers have been working on supporting the multilingual communication. One method is to create artificial delays to help the non-native speaker understand the conversation before continuing the conversation [Morita and Ishida, 2009]. Some other researchers try signaling the native speaker about the status of non-native speakers [Yamashita et al., 2013], helping non-native writers by providing vocabulary navigation [Gao et al., 2015], and providing real-time translation using eye-gaze input of the non-native reader [Toyama et al., 2014]. Even though these methods can reduce the burden of non-native speakers, they cannot give a wholly balanced communication environment.

It is also significant to learn other languages and respect different culture. Because it is impossible for one person to learn every language, machine translation(MT) and other technologies on the internet can be a solution [Ishida, 2016]. Using MT can improve the efficiency and effectiveness of discussions [Aiken, 2009]. Yet, MT can also cause many communication problems during collaboration. Because people

has varied level 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 [Fellbaum, 2010], common problems with machine translations, can also cause conversation breakdown when translation result go wrong [Yamashita and Ishida, 2006]. On the other hands, if a foreign language is used chosen, the user with lower skill of that language can have less chance to communicate, or feel left out of the conversation.

Some researchers attempt to improve communication by improving the quality of machine translation as well as using human intelligence. For example Morita D. [Morita and Ishida, 2009] introduced a method to use monolinguals to help with the fluency and adequacy of both sides of two language translations. Taking a direction form the outsourcing of human intelligence, we realized that the ability of the users themselves can also be used. 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 presented various methods to help non-native speakers and support multilingual collaboration. Our research is novel and orthogonal to existing research. Our model aims at supporting non-native speakers with different proficiencies of shared language skill. Our method create the best balance in terms of opportunity to participate in communication. To obtain the best balanced communi-

cation environment, we start with a previous study called user-centered QoS [Bramantoro and Ishida, 2009]. Generally, services are evaluated by users based on the Quality of Service (QoS). However, information of users is important in selecting the best machine translation service. Thus, they introduced a new function to calculate the Quality of Message (QoM) by using the users' skills in writing and reading messages when machine translators are used. In this work, we extend QoM to define a model of the best balanced channel using the parameters of user language skills and machine translation accuracies. Then, we investigate our model in a real world experiment to study the effectiveness of our approach in creating effective environments for multilingual collaboration.

3.2 Scenario

Machine translation can cause communication balance problems. For example, Figure 3.1 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. It is easier to use translator especially for Japanese user. Later, a Korean user with good English skill joins the conversation, then, it becomes more complicated to choose what languages or what services should be used.

One possibility is to use a shared foreign language, English. Another possibility is to use machine translation. It is also possible to combine both options together. If only English is used for this conversation, it

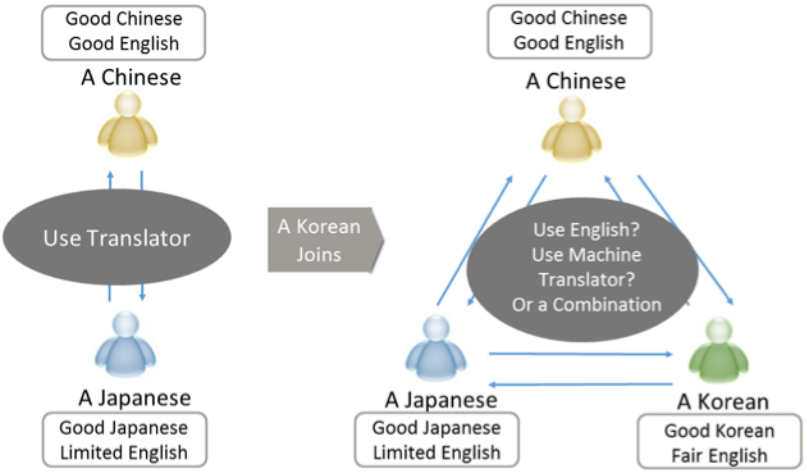


Figure 3.1: Situation where the multilingual communication problem exists.

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 of machine translation imperfection. The situation is a problem of asymmetry in collaboration caused by language proficiency difference. In that group, people have asymmetric opportunity to participate in the conversation. We believe that to have the best communication is to have mutual understanding and equal chance to take part in the conversation. As a consequence, we would like to propose a method to solve this problem.

3.3 Modeling Multilingual Communication

3.3.1 Best Balanced Channel

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

To select languages for the best-balanced communication channel, the algorithm as follows should be carried out.

Based on existing work [Bramantoro and Ishida, 2009] related to user-centered QoS, we model the Quality of Message (QoM) that user P_i uses language L_i to send message to user P_j, who uses language L_j via a 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 QoMi,j, can be represented as follows:

$$QoM(P_i, MT_{i,j}, P_j) = writing\_skill(P_i, L_i) \times accuracy(MT_{i,j}) \times reading\_skill(P_j, L_j) \tag{3.1}$$

From this model, writing skill of the sender, accuracy of machine translation, and reading skill of the receiver affect the quality of message.

ALGORITHM 1: Best-balance machine translation

Input : P_i: User i (1 <= i <= N, n is the number of users)
L_i: Language i used by user P_i
$WS(P_i, L_i)$: Writing skill of P_i when using language L_i
$RS(P_i, L_i)$: Reading skill of P_i when using language L_i
$MTQ(MT_{ij})$: MT accuracy to translate from L_i to L_j

Output: Best-balanced language combination BBC

Generate the set of all possible language combinations LC
forall *language combination* C_k *in* LC **do**
  forall *language pair* $< L_i\ L_j >$ *in* C_k **do**
    $QoM(P_i, MT_{ij}, P_j) \leftarrow WS(P_i, L_i) MTQ(MT_{i,j}) RS(P_j, L_j)$;
    $QoM(P_j, MT_{ji}, P_i) \leftarrow WS(P_j, L_j) MTQ(MT_{j,i}) RS(P_i, L_i)$;
    $AvgQoM_{ij} \leftarrow (QoM(P_i, MT_{ji}, P_j) + QoM(P_j, MT_{ji}, P_i))/2$;
  end
end
Acquire the set of Pareto optimal language combination PLC;
$m \leftarrow$ number of acquired language combinations in PLC;
if $m = 1$ **then**
  $BBC \leftarrow$ the only language combination in PLC;
else
  forall *Pareto optimal language combination* C_k *in* PLC **do**
    Compute the variance between each $AvgQoM$ in C_k;
  end
  $BBC \leftarrow$ language combination with minimum variance;
end

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.

QoM pair between user P_i and user P_j is written as $(QoM_{i,j}, QoM_{j,i})$

and an MT pair between language L_i and language L_j can be 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 is needed to be extended. QoMi,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 of a QoM pair 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 optima, it does not need to calculate variance.

3.3.2 Example

Assume that languages L1, L2, L3 are used by three users P1, P2, P3 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. Possible combinations of languages Cx = (L1,L2,L3) for the communication among three users are listed 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 (QoM1,2, QoM2,1), (QoM2,3, QoM3,2), and (QoM3,1, QoM1,3). C1 utilizes three pairs or six of machine translation services, including (MTja,ko, MTko,ja), (MTko,zh, MTzh,ko), and (MTzh,ja, MTja,zh).

With user language profiles andmachine translator qualities in the example situation, the only Pareto optimal combination is C4, which means the conversation will be best balanced when the Japanese user uses Japanese while Korean and Chinese user use English, and the machine translation service needed is (MTjp,en, MTen,jp), since (MTen,en, MTen,en) represents using English with no translation .

In cases, there exist more than one Pareto optimal combinations. The best balanced combination can be chosen by calculating the differences among the QoMs using variance, because the lower differences indicates the higher equality of the conversation.

3.4 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.

3.4.1 Task

In the experiment, the participants were asked to play three collaborative games together. Three survival problems were used: desert survival problem (DSP) [Lafferty et al., 1974], winter survival problem (WSP) from the project ARISE [Project ARISE Fermilab, 2019], and lunar survival problem (LSP) from NASA [Nasa, 2019]. 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 safety. WSP is similar to DSP, but with the extremely cold weather in the wood. The task for the participants is the same as DSP but the item list is different from the first game. LSP is slightly a unique situation, landing 80 kilometers from the target place on the moon. Yet, LSP task is also the same as the first two games with different set of items.

In the original problems, they describe the situation using a number of paragraphs in English. But our game is narrated the situation using easy short sentences in English and figures. Simplification is needed to cover the various English proficiencies of the players. Each story explains the different situation with time, location, and events that happened while the participants played survivor roles in the story. Our version of the game also simplified the choice of items. While the original games provide more items to be ranked, our participants are asked to rank a set of only 6 items by their importance for each situation.

The participants were asked to collaborate and finish the game within time limit but we did not mention the game score or the right answer of the ranking. Because getting right answer or high score is depended on specific knowledge of the team members, for example, science, geography, survival skill and camping skill, it does not show how well the participants collaborate.

3.4.2 Experiment Design

At first, game instructions were introduced to the participants. Afterward, 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 [[Ishida, 2011].

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 selected language into the message box then clicks 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 three survival games, we played an example game for twenty minutes. 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 balance machine translation (BB). The strategy was chosen randomly. The second game was played using one of the strategies which was not selected for the first game. The last game was played with the remaining strategy.

First, we gave them the 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 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 influences 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.

3.4.3 Participant

We divided our nine participants into three groups. In each group, there were a Chinese participant, a Japanese participant, and a Korean participant. All the them were either undergraduate, graduate, or research students from varied fields.

Table 3.1: Profile of participants

Group/	Chinese		Japanese		Korean	
Participant	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)

English skill profiles of the participants, displayed in Table 3.1, consisted of (writing_skill, reading_skill) normalized with range of 0 to 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) [Cambridge Assessment, 2019] 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 [ETS, 2019] for TOEIC and TOEFL and data from Cambridge assessment [Cambridge Assessment, 2019] for IELTS. With the conversion matrix, we gives 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 described as M, for male, and F, for female.

3.4.4 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 chose twenty 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 in to the other three languages. To illustrate, sentences in Japanese were translated into Chinese, English, and Korean.

Even though quantitative metrics are useful for evaluation, they cannot completely replace human assessment
citecallison2007meta. The translated sentences were rated by educated native speakers also holding at least a bachelor's degree. This methodology of rating adequacy and fluency is widely used to measure machine translation as proposed by LDC [Linguistic Data Consortium, 2005]. Fluency of the translated sentences 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. After that, the evaluation results were confirmed by another native speaker of the same language if the evaluation were acceptable or not.

The evaluation for each sentence was averaged to decide the quality of machine translation service from one language to another. Adequacy and fluency rating evaluated by the human were added up and normalized to the scale of 0 to 1 as displayed in Table 3.2.

Table 3.2: Quality of translation services

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

Table 3.3: QoM values of all possible combination

Combination/QOM Pair	Pair1	Pair2	Pair3
C1	0.771875	0.540625	0.7125
C2	0.771875	0.625	0.790625
C3	0.790625	0.570313	0.7125
C4	0.790625	0.875	0.790625
C5	0.46875	0.540625	0.404688
C6	0.46875	0.625	0.75
C7	0.5	0.570313	0.404688
C8	0.5	0.875	0.75

3.4.5 Communication Channel

The value of QoM pairs for each combination from C1 to C8 can be calculated, as shown in Table reftab:QoMCombination, using the participant profile from Table 1, and quality of translation services from Table 3.2. In this case, the only row containing Pareto optimal sets of QoM pairs is C4, so variance does not need to be calculated and the best balanced machine translation channel is C4.

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

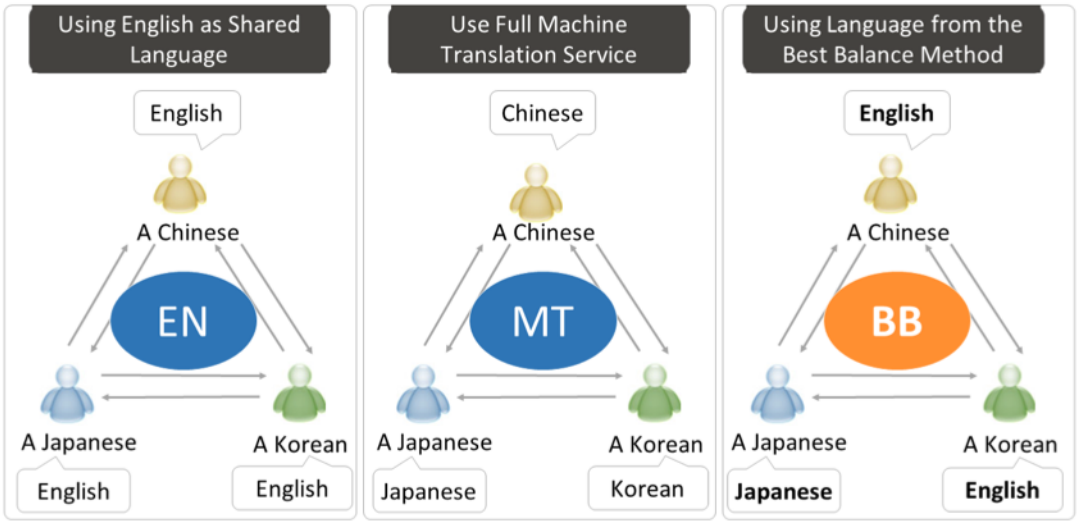


Figure 3.2: Strategies of communication in this experiment.

To investigate validity of our result, we chose the other two communication channels that are normally used at the moment of this thesis 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 Figure 3.2. Hence, the communication channels used in this experiment includes BB, EN, and MT, while EN represents C8 (en, en, en) , and MT represents C1 (ja, ko, zh) in the calculation.

3.5 Behaviors of Participants with Low Shared Language Skill

3.5.1 Simpler sentences used by Japanese when using English

With EN channel, the sentences typed by Japanese users were simpler and shorter because of their limited language skill. Having simple sentences does not mean the conversation is bad but with complex sentences, it might be more natural and it can easily lead to interesting discussions or new ideas.

3.5.2 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 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 very limited. Sentences which are not understandable are normally ignored by other parties [10]. When a low-English

skill participant input 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

3.5.3 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 conversa-

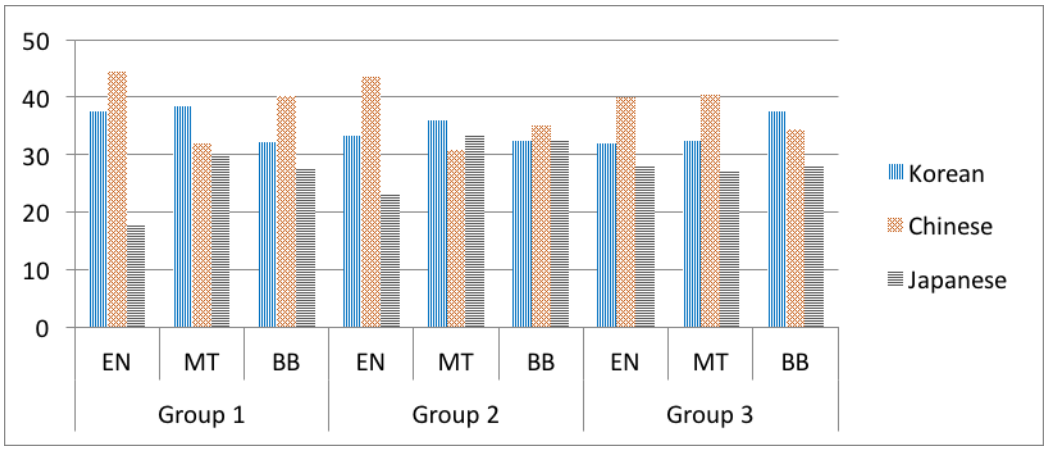


Figure 3.3: Talkativeness of each participant in each group measured by percentage of utterances each participant made.

tion by comparing the talkativeness, here measured by the number of utterances.

The percentage of utterances made by each participant is shown in Figure 3.3 , and the average percentage of utterances created by each nationality with similar English-skill level is shown in Figure 3.4 From both figures, the EN channel shows the most unequal participation in the conversation. The Japanese tended to talk much less when using English, while the balance became better when they used MT in MT channel and BB channel.

Conversation Urging

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 parts of a conversation

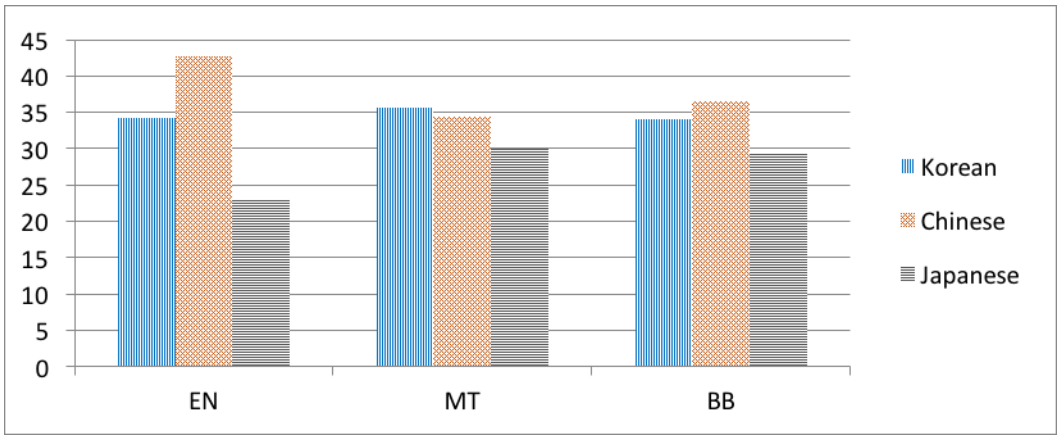


Figure 3.4: Average talkativeness grouped by country of origin measured by percentage of utterances.

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?

After analyzing their chat logs of each team to look for the reasons 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. With MT, it can help the participant dealing with difficulties in languages, in terms of expression and understanding. Using MT can also increase confidence when using her/his mother language. Asking for a specific participant' s opinion appears much less when MT or BB channel is used.

3.6 Machine Translation Problem

It is normal that the machine translation output many mistakes, although most of the time people can still understand what the other tried to say. Often, however, using the MT channel mistakes more likely, which lead to more serious problems, such as miscomprehension or misunderstanding. Some translation errors are severe enough to interrupt the flow of conversation.

The conversation in Figure 3.5 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 grey. The serious translation mistake is printed in bold font.

The participants had already discussed about the 3rd 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 "Is not 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 in the first Chinese sentence already converts the sentence into a yes/no question. Unfortunately, the translation result, "Number 3 is not a mirror", had completely opposite meaning. This time, the translation mistake is serious enough, so the users could notice. The Korean user also asked the Chinese user to explain it again. This kind of situation costs us more time and disrupt the smoothness

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. 翻译会异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?

Figure 3.5: A part of English translation, original messages, and translated messages shown for each participant when MT was used.

of conversation.

3.7 Conversation Breakdowns

Breakdowns are issues in communication as they interrupt the flow of conversation. Conversations breakdowns were common in the MT channels and EN channels. The number of breakdowns in English conversation made by group 1, group 2, and group 3, were 20, 9, and 7, respectively, for a total of 36 breakdowns. With machine translation, the number of breakdowns were 17, 14, and 7 by group 1 to 3, respectively, totally 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, there are more issues when only machine translation is used and when only English is used.

Our investigation showed that when MT is used, machine translation errors triggered by 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 translator problem, conversation can also fail.

Using only the EN channel can also cause breakdowns. The causes include misunderstanding due to lack of language skill. User 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 is normally raised address the participant' s behavior and the original topic was interrupted.

3.7.1 Relationship between breakdowns and QoM

A smaller number of breakdowns indicates 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 concept of the QoM or the better message quality when both user skills and machine translation qualities are taken into consideration.

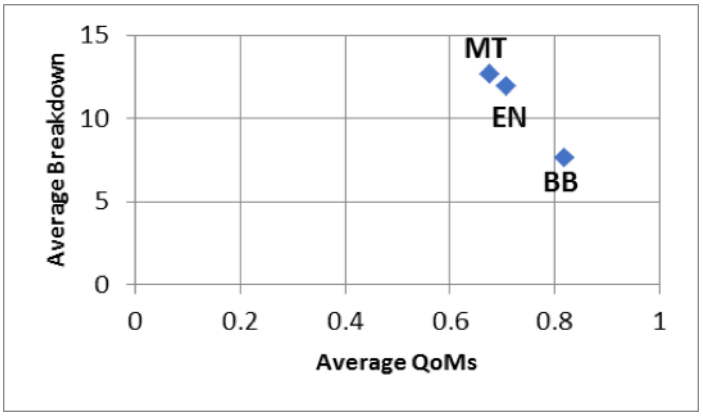


Figure 3.6: Comparison between the average breakdowns and average QoM.

As shown in Figure 3.6, 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 lowest number of breakdowns on average. Using full translation causes the highest average breakdown rate and the lowest average value of QoMs. EN channel lay between BB and MT. These results show the relationship between the number of breakdowns and the QoM value. Higher QoM, indicative of better message quality, is associated with fewer breakdowns, and confirms better communication quality.

3.8 Discussion

3.8.1 Talkativeness and The QoMs

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 3.3, 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 trend in Figure 3.7. The lower the CV value of QoM, the more equitable quality of message each user can have.

We also calculated the CV of average talkativeness, to see how well distributed talkativeness was. The lower is the CV value of average talkativeness, the more equally the participants engage in the conversation. We can see the similar trend that when the users employ the languages that they have medium to high skill in, the variation in QoM and Talkativeness is much lower. From Figure 3.7, using EN has much higher CV of 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 or 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 do not count the data associated with machine translation problems. If

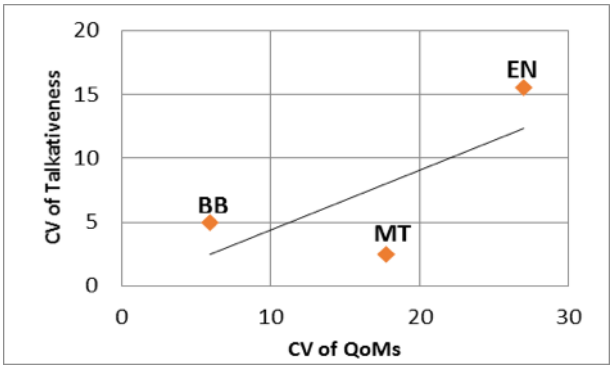


Figure 3.7: Coefficient of Variation of Average Talkativeness vs. Coefficient of Variation of QoMs.

more accurate MT was used, the number of utterances related to problems with MT and overall talkativeness might be smaller, since some utterances are about complaining and trying to fix understanding problem cause by MT mistranslation. These utterances can be counted and included in talkativeness but it does not mean that the users are 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 result of CV.

Combining the results in Figure 3.6 and Figure 3.7, 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 balance machine translation model creates minimizes conversation breakdowns since it avoids the use of low quality machine translation services, which tend to create translation mistakes.

3.9 Possibility of Using Different Language to Read and Write

BBMT is proposed based on the assumption that a user read and write the same language at a time, for example, the Japanese user read messages in Japanese and write messages in Japanese. However, in real world, it is possible that a Japanese user can read in English and write in Japanese.

BBMT can be extended to consider this case by adding the detail of language combinations. For instance, current BBMT combination for two users when a user speak Japanese and English and another user speak English could be (ja,en) which means the first user uses Japanese and another user uses English to communicate. To cope with the situation above, the language combination could be ((en,ja),(en,en)), meaning the first user read in English and write in Japanese, while the second user reads and writes in English. In this case, the same equation of QoM can still be used and the process of selecting Pareto optimal can still be performed. The difference is, there will be more combinations than the original BBMT. With the previous example, Table 3.3 contains eight combinations. If the users can read in one language and write in another language, using the same language profile, there could be 64 combination of languages.

3.9.1 Possibility for Gaining Higher Overall Talkativeness

Even though BBMT can improve the balance of talkativeness, however; assigning languages to be used could make users feel uncomfortable to communicate. On the contrary, allowing users to use their preferred languages might make them feel more comfortable to share their ideas.

BBMT is focusing on having equal opportunity to participate but it might not lead to the highest number of utterances and ideas.

3.10 Conclusion

Our main contribution is proposing best balance machine translation model that 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 when the quality of machine translation is too low.

To confirm the validity of our result, we set up an experiment comparing our method to widely used methods of communication including using English as a shared foreign language and fully use machine translation. In the experiment, our best balance 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 can indicate the better flow of conversation when our model is used. 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 original model promotes creativity by enhancing communication quality by allowing 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.

Chapter 4

Privacy-Aware Best-balanced Machine Translation

Because the BBMT, we proposed in the third chapter, may invade user's privacy of having test score sent to the central system. This chapter propose a method protect user privacy while allowing the data to be used for computation. Later in this chapter, we made a security argument to prove the privacy protection effect of our protocol.

4.1 Introduction

In multilingual communication, machine translation(MT) is a useful tool to overcome the language barrier. There are many services available with varied quality, but sometimes users' foreign language skill can yield better for the communication if the MT quality is too low.

Previously, we have proposed a method called best-balanced machine translation method. Since it is difficult for a human to decide what languages should be used when there is a group of people who speaks different languages and with different level of proficiency, their system suggested what languages to be used in multilingual communication when machine translation (MT) services exist. In order to calculate and suggest the best languages to be used, quality of MT and users' language test score, i.e. TOEIC, TOEFL, are needed to be shared between personal agents and optimizing agent for the calculation.

As well as other private personal information, it is important to protect the confidentiality of test score. Education Testing Service (ETS), a non-profit organization who administers international tests, including, TOEFL, TOEIC, etc., also pays attention to confidentiality. Their policy inquires that private information, including score data, must be kept confidential unless there is an informed consent of the individual. In the U.S., the Family Educational Rights and Privacy Act (FERPA) also required to have written permission from students' parent or eligible student to release any data from a student record, except for some special conditions [US Department of Education, 2018]. We can infer that, it is crucial to treat language test score confidential. The issue with the previous version of best-balanced machine translation model is, distributing the test score for calculation violates confidentiality for users' private data. Calculating best-balanced language combination has specific procedures and its characteristic creates challenges to protect user privacy. Our research problem is how to calculate best-balanced language set without disclosing nor distributing user's language score from

each user's personal agent(PA), the agent who deals with each user's personal data and activity.

Multi-agent system and decentralization have a great number of applications. They have been used not only for general problem solving [Anders et al., 2012, Matsui et al., 2012], but they can also be used to protect private information [Greenstadt et al., 2006, Yokoo et al., 2002]. With multi-agent system and data encryption, we would like to introduce secured computing into human agent interaction. As far as we can tell, this is the first trial on applying secured computation on multilingual communication support system. We propose a solution to protect test score data privacy from being disclosed by user's PA while keeping the language optimization based on best-balanced machine translation possible. We distribute optimization works to different types agents and use data encryption to support human-human communication, while the original method uses one agent for optimization in order to keep user private data invisible for any other agent except user's own PA.

4.2 Related Work

Several groups of researcher use agents to support human-human communication in various way. Agents are used to support remote collaboration, for example Vartiainen et al.[Vartiainen et al., 2015] proposed a concept of a mobile tele-presence robot to support collaboration. There existed research related to agents for group discussion and meeting support. A group of researchers [Ishii et al., 2017] tried to create a conver-

sational interface that allows agent and people communicate smoothly by studying head movements in multi-party meetings. Another group of researcher [Kimura et al., 2017] aimed to build an agent that can participate group discussion to improve communication skill of users, so they proposed a model to determine attention target for the agent in group discussion.

Each existing works focuses on different parts of supporting collaboration, including user experiences, conversation training, etc. On the contrary, our work focuses on using multi-agent system to support intercultural collaboration and multilingual communication and focuses on user data privacy. In our study, we apply multi-agents to create a privacy-aware system for selecting languages to be used by each user by combining existing concept of best-balanced machine translation and secure computation technics together. Currently, there are many methods to protect private data.

Since private data is not only needed to be protected but also needed to be processed with some other data, some researchers proposed various method to make use of the data without violating the data privacy. Jian and Bhandare [Jain and Bhandare, 2011] proposed a method to preserve privacy based on min max normalization transformation for the data to be used for data mining.

K-anonymity [Sweeney, 2002] has been widely used to protect data privacy, especially for data publication, usually done by data suppression and generalization. Since different kinds of data have different characteristics, calculation used are varied. For example, test scores of students also have its own characteristic, Yi [Yi et al., 2015] proposed a

method to publish test scores while having students' privacy protected by using K-anonymity into location based service. Some researchers use cryptography techniques to help protecting user's privacy. When the data is online, private data can be leaked. There exist studies using encryption to protect the confidentiality. Popa et al. [Popa et al., 2011] proposed CryptDB, a system to protect sensitive data with encrypted database query. They executed SQL queries over encrypted data and also link encryption keys to user passwords so even the database admin cannot access the data.

Besides protecting user's data, Yokoo et al. [Yokoo et al., 2002] proposed a method to protect agent's private information using multi-party techniques. Their method utilizes public key encryption scheme allowing the information to be computed cooperatively but cannot be linked back the agents.

There exist various methods to protect data privacy while making use of the data. However, each method works for specific types of data and specific situations, for instance data suppression is not suitable for the process that need real or specific data for computation, but this method is useful for some data publications. None of the stated methods is suitable to keep language score data privacy while enabling best-balanced calculation. The challenge is using only data encryption and decryption or purely secure computation is not enough to make user score invisible from the other agents, since the processes of calculation are fairly complex and there are chances that some agent can guess users' language level.

4.3 Privacy-aware Best-balanced Machine Translation

4.3.1 Agents and Their Roles

Here, we would like to propose a multi-agent system to optimize the languages that could be yield the best balance in MT mediated communication. Our method includes three different types of agent Personal agent, Process managing agent, and Selection agent, instead of using only personal agents and optimizing agent in the original method. Works that used to be done by one optimizing agent in the original method are distributed to all types of agents. The major functions of this system are secured AvgQoM calculation done by Personal agents, key handling and secured language combination selection by the Selection agent. Process managing agent does the rest of the works that cannot be done by the other two agents due to security reason, including creating language combination and managing the data. In this section, we describe the agents we introduced and their roles using role schema by following Gaia methodology [Wooldridge et al., 2000]. The role schemas include protocols, which are linked to the activities and interactions. Activities are works done by the agent itself, without interaction, written with underlined text. The interactions among agents are shown in Figure 4.1. The arrows on the diagram represents data transmission among three different types of agents, including function name, linked to the agent role schemas, and data to be sent in the parentheses.

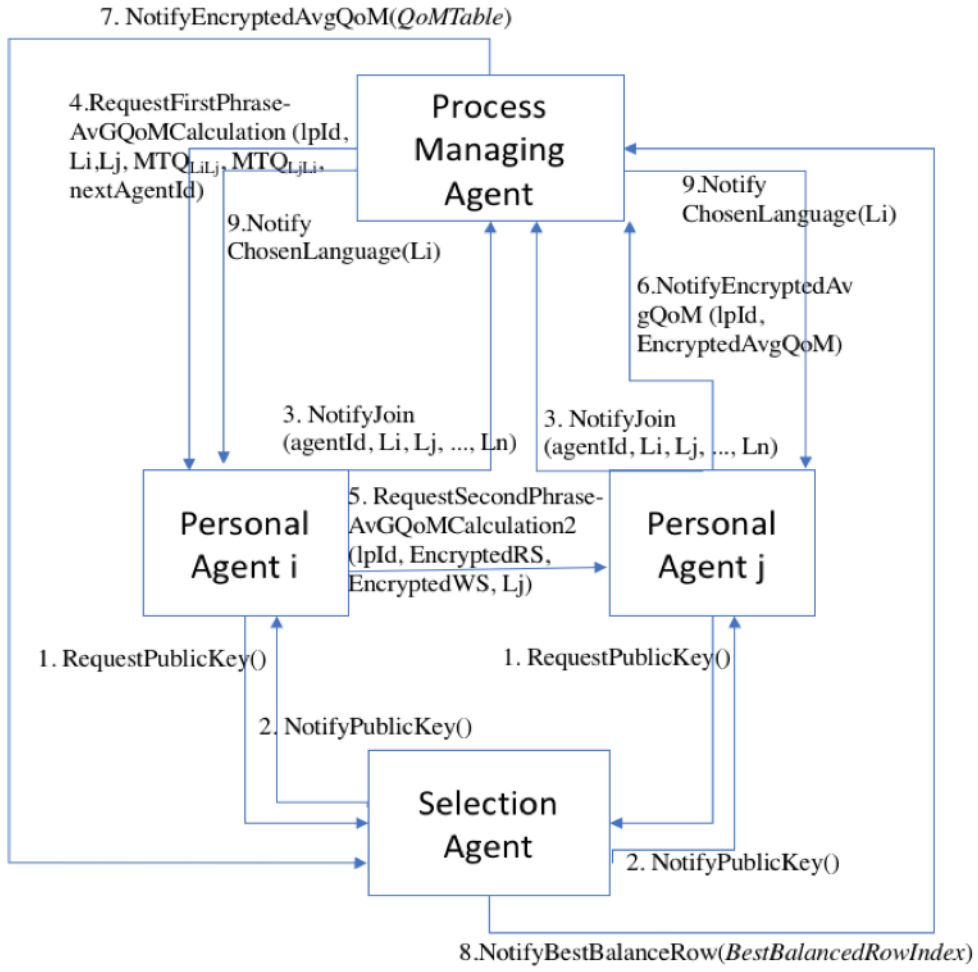


Figure 4.1: Interactions and data sharing among agents in privacy-aware BBMT calculation.

Personal Agent

In best-balance calculation, language score data is needed only to calculate QoM. Later, QoMs of each language pair are averaged. Averaged

ALGORITHM 2: Firstphrase AvgQoMCalculation

Input : $IpId, Li, Lj, MTQ_{LjLi}, nextAgentId$

Output: $EncryptedWS, EncryptedRS$

1 Calculate $EncryptedWS = MTQ_{LiLj} * Encrypt(WS_{UiLi})$;

2 Calculate $EncryptedRS = MTQ_{LjLi} * Encrypt(RS_{UiLi})$;

3 Request $SecondPhrase - AvGQoMCalculation(lpId, EncryptedRS, EncryptedWS, Lj)$ to Personal Agent of the language pair whose $id = nextAgentId$;

QoM (AvgQoM) is an important value to select the best-balanced communication channel

In the original calculation, AvgQoM is calculated by one single agent so this agent needs to see every user's score. In our proposal, a Personal agent is not only responsible for a user's data and take part in the computation of AvgQoM so user's score is not sent to any other agent.

To calculate AvgQoM without disclosing user language score, encryption is used and the calculation is done by two Personal agents, the score owners. Each AvgQoM is consisted of two parts for calculation FirstPhraseAvgQoMCalculation done by the personal agent of the first score owner and SecondPhraseAvgQoMCalculation done by the personal agent of the second score owner, using the following procedure. The number in parenthesis is linked to the process in Figure 4.1.

The first calculation is decided and requested by Process managing agent(PMA). The first agent encrypts its user writing skill(WS) and reading skill(RS) of the request language L_i and multiply by given MT quality score (MTQ). After the first calculation is done, the first agent requests the second calculation to the next agent identified earlier by

ALGORITHM 3: Second-phrase AvgQoMCalculation

Input : $IpId, EncryptedWS, EncryptedRS$

Output: $EncryptedAvgQoM$

1 Calculate $EncryptedAvgQoM = (EncryptedWS * Encrypt(RS_{UjLj}) + EncryptedRS * Encrypt(WS_{UjLj}))/2$;

2 Send $NotifyEncryptedAvgQoM(lpId, EncryptedAvgQoM)$ to Process Managing Agent;

Role Schema: PERSONAL AGENT(PA)
Description: This role involves handling user's data for privacy-aware BBMT computation, requesting for a public key, notifying user usable language to Process managing agent, and calculating Encrypted AvgQoM with its own user's data and forward the calculation result as requested by Process managing agent.
Protocol and Activity: RequestPublicKey, NotifyJoin, First-Phrase AvgQoMCalculation, RequestSecondPhrase-AvgQoMCalculation, Second-Phrase AvgQoMCalculation, NotifyEncryptedAvgQoM
Permission: accesses PublicKey
Responsibility: REGISTER = (RequestPublicKey, NotifyJoin) FIRST_PHRASE_CALCULATION = (First-Phrase AvgQoMCalculation, RequestSecondPhrase-AvgQoMCalculation) SECOND_PHRASE_CALCULATION = (Second-Phrase AvgQoMCalculation, NotifyEncryptedAvgQoM)

Figure 4.2: Role schema of Personal agent.

PMA with EncryptedRS and EncryptedWS.

The second agent also encrypts its own RS and multiply by first agent's EncryptedWS and encrypts its WS then multiply by the first agent's EncryptedRS. Afterward, it averages those two values by sum two values together and divide it by two. After both phrases of calculation are

executed, the second agent send the encrypted AvgQoM back to the Process managing agent with the reference, IpId.

Selection Agent

Selection agent creates a public key for encryption and a private key for decryption. When a Personal agent joins the system and request for the public key, selection agent send a public key to the Personal agent. In this system, Personal agents have public key for secured AvgQoM calculation and Selection agent has both public key and private key, while Process managing agent does not have any key.

The reason why keys are handled by Selection agent is when an agent knows the pure value of AvgQoM and know that whose AvgQoM belongs to, it is possible to guess the value of user scores. For example, if an AvgQoM between user i and user j is very low, the agent who knows the AvgQoM can imply that machine translation quality is low and both user i's skill and user j's skill of particular languages are also low. Selection agent can see both encrypted AvgQoMs and AvgQoMs, but this agent does not know who

owns each AvgQoM since all the encrypted AvgQoMs are forwarded from Process managing agent in QoM Table without any information about user, Personal agent, nor combination.

Role Schema: SELECTION AGENT(SA)
Description: This role involves handling public and private keys, decryption, and selecting the best balance AvgQoM. It sends public keys to Personal agents as requested. After receiving QoMTable from Process managing agent, it selects the best-balanced and notify the selected row index back to the Process managing agent.
Protocol and Activity: NotifyPublicKey, SelectBestBalancedRow, NofifyBestBalancedRow
Permission: accesses PublicKey, PrivateKey
Responsibility: KEY_REQUEST_RECEIVE = (NotifyPublicKey) QOM_TABLE_RECEIVE = (SelectBestBalancedRow, NofifyBestBalancedRow)

Figure 4.3: Role schema of Selection agent.

Process Managing Agent

This agent generates tables and creates language combinations based on user languages as Personal agent notifies. It can see the encrypted AvgQoMs data from Personal agents, but without decryption key, it is impossible to see the real AvgQoM values. It also handles index, as reference codes, to distribute calculation and selection task and to keep Selection agent from linking its data back to any user nor combination.

Role Schema: PROCESS MANAGING AGENT(PMA)
Description: This role creates necessary tables for all the calculation when any new Personal agent joins and requests AvgQoM first calculation to the Personal agent. It also forwards encrypted AvgQoMs collected from Personal agents to Selection agent and informs Personal agents selected language after receiving the best balance row index from Selection agent.
Protocol and Activity: GenerateCombinationTable, GenerateQoMTable, GenerateLanguagePairListAndIpId, RequestFirst-Phrase AvgQoMCalculation, FillAvgQoMs, NotifyAvgQoMTable, FindBestBalancedLanguages, NotifyChosenLanguages
Permission: -
Responsibility: REGISTRATION_RECEIVE = (GenerateCombinationTable, GenerateQoMTable, GenerateLanguagePairListAndIpId, RequestFirst-Phrase AvgQoMCalculation) FILL_QOM_TABLE = (FillAvgQoM) SEND_QOM_TABLE = (NotifyAvgQoM) BEST-BALANCE_ROW_RECEIEVE = (FindBestBalancedLanguages, NotifyChosenLanguages)

Figure 4.4: Role schema of Process managing agent.

4.4 Computation Flow

When a user joins this system, her/his Personal agent sends a request for public key to selection agent(SA). After receiving the request, SA sends the public key to the Personal agent. Each Personal agent also notifies Process managing agent when it joins the system with its id called agentId, and languages which its user can use, displayed as language L_i and language L_j.

Upon receiving the notification, Process Managing Agent receives the notification, the agent creates a table called Combination Table which contains all possible language combinations based on users' usable lan-

Table 4.1: An example of Combination Table.

Combination ID	Language and User
C1	((i, L1), (j, L1))
C2	((i, L1), (j, L2))
C3	((i, L1), (j, L3))
C4	((i, L2), (j, L1))
C5	((i, L2), (j, L2))
C6	((i, L2), (j, L3))
C7	((i, L3), (j, L1))
C8	((i, L3), (j, L2))
C9	((i, L3), (j, L3))

guages. Each row contains an ID called combination ID, username and language of each user. For example, $(i,Li),(j,Lj)$ is a language combination for user i using language L_i and user j using language L_j to communicate. In case, the third user, user k, who speaks language L_k joins, one of the possible combinations would be $(i,L_i),(j,L_j),(k,L_k)$.

When there are 2 users and both user i and user j can use languages L_1, L_2, and L_3, the combination table will contain information as shown in Table 4.1. The total number of combinations equal the number of language(s) used by user i multiplied by the number of language(s) used by user j. If there are more than two users, the multiplication continues to the number of languages used by the third user, fourth user, until the last user language(s) is(are) considered.

Afterward, Process managing agent also create a blank QoM Table to store encrypted AvgQoM in the future. The table height equals the heights of Combination Table and the table width is equal $n*(n-1)/2$ when n is the number of users. This table will store encrypted AvgQoM

for each link between two agents. If there are three users, there are three links among the users, including links between user i and user j, between user i and user k, and between user j and user k. So, there will be three encrypted AvgQoMs for each combination.

Besides Combination Table and QoM Table, the Process managing agent also creates Language Pair List, including Language pairs whose AvgQoM needed to be calculated. A language pair is a pair of language that allow two users to communicate. For example$(i,L_1),(j,L_1)$ is a language pair when bout user i and j use language L_1 to communicate. When there are two users, this Language Pair List contains all the language pair in the Combination Table, but when there are three or more users, some language pairs in the Combination Table are redundant, so only one of the redundant language pairs appear in the Language Pair List. For three-user communication, each combination contains 3 language pairs including language pair of user i and user j, user i and user k, user j and user k. Each language pair in this list also has an ID called IpId generated by the Process managing agent, in order to link the language pair with its location in QoM Table.

Every language pair in the Language Pair List needs to have AvgQoM calculated, so the Process managing agent, send one message per on language pair in the list to the first Personal agents of the pair. The Process managing agent also have a list of machine translation quality or MTQ. MTQ(L_i,L_j) represents machine translation quality when translate from language L_i to language L_j. For instance, for the language pair $(i,L_i),(j,L_j)$, the Process managing agent sends a message to Personal agent i , requesting for the first phrase of AvGQoM calcu-

lation. This message includes IpId of the pair, language L_i, language L_j, MTQ(L_i,L_j), MTQ(L_j,L_i) and the nextAgentId which is the Id of the second agent in the language pair.

When the Personal Agent of user P_i receives the request for AvgQoM calculation, it calculates Encrypted writing skill (Encrypted WS) by using the public key to encrypt the value of user i's writing skill in language L_i or $WS(P_iL_i)$), then multiplies the value by $MTQ(L_iL_j)$. Encrypted reading skill (Encrypted RS) is calculated by encrypting user i's reading skill of language L_i or $RS(P_iL_i)$ multiplied by $MTQ(L_iL_j)$. Both Encrypted WS and Encrypted RS are encrypted values and need the private key, held only by the Selection Agent, to read the real values. Hence, there is no agent can read the pure data of this value, since this data is not sent to the Selection Agent either.

After the first part of calculation is done by the first agent i, the first agent sends a request for the second calculation to the second agent j of the language pair. The message includes IpId, Encrypted WS, Encrypted RS, and Language L_j. When the second agent receives the request, it calculate Encrypted AvgQoM by using the public key to encrypt and multiply own reading skill of language L_j (RS_{UjLj}) to Encrypted WS and own writing skill of language L_j (WS_{UjLj}) to Encrypted RS, then average these values. This whole process of calculation also needs the public key and the result value is an encrypted value of AvgQoM called Encrypted AvgQoM. When the calculation is done, the second agent of the language pair sends Encrypted AvgQoM together with the reference IpId back to the Process managing agent.

Process managing agent collects Encrypted AvgQoM from Personal

agents and put them on previously created QoM Table based on IpId. After QoM Table is filled with Encrypted AvgQoM received from Personal agent(s), this QoM Table has no index to refer to any combination nor user. Then, Process managing agent send this QoM Table to Selection agent.

Selection agent selects the Best balance row by, first, decrypt the Encrypted AvgQoMs to AvgQoMs with its private key, then select the row that has Pareto optimal value. After decryption, Selection agent can only see numbers on the table, since there is no reference to any combination nor user. If there are more than one Pareto optimal roles, this agent calculated variance among the value in Pareto optimal roles, and selects the row with the least variance as Best balance row. After that, Selection agent send the index of the Best balance row back to the Process managing agent.

After Process managing agent receives the row index, it searches for the information from the row index to see which language combination is linked to the Best balance row and what language each user should use in that combination. Later, it notifies each personal agent the recommended language for each user based on the best-balanced combination selection.

4.5 Discussion

4.5.1 Security Argument

In this section, we discuss privacy protection effect. The information that we are trying to protect here is user language score(s) that each Personal agent owns, including reading skill and writing skill of its user. This information should not be seen by any other agent except its owner. Beside user reading and writing skill, AvgQoM should also be treated as a fairy sensitive information. When AvgQoM is very low, the agent who knows the AvgQoM owners can guess that both user scores are low. So, another security requirement is, any agent who knows pure AvgQoM value must not know which agents' skills are used for the calculation.

We investigate situations when each agent, one by one, acts as an adversary, trying to violate user privacy, given the information known to that agent, called view. A view of an agent is all the information that is visible to that agent, including the information it owns, and information from the other agents or the other source. View of each agent is shown in Fig 6. Information that each agent owns or created by itself is underlined. Information without underline is information received from the other agent. We assume that each agent is Honest-but-Curious(HbC). HbC agents follow the steps of the protocol but try to learn as much information as possible [Wang and Kissel, 2015].

When we consider the Personal agent of user i, or PA_i, as an HbC adversary, PA_i should get no information about another Personal agent's

Table 4.2: View of each agent

Personal Agent (PA) View	Selection Agent (SA) View	Process Managing Agent (PMA) View
- Own language data - Public key for encryption	- Public and private key - Decrypted AvgQoMs	- Combination Table - QoM Table
- Some MTQ	- Best balanced row index in QoM Table	- All MTQ
- Encrypted RS and encrypted WS	- User(s) joined the system	- User(s) joined the system
- Suggested language to be used	- All encrypted AvgQoM in the QoM table	- User usable language(s) - All encrypted AvgQoM - Best balanced row index in QoM Table

private information than what is trivially compatible from its own input and the final language outcome. From Personal agent's view in Table 4.2, data related to another agent, PA_j, is encrypted, including Encrypted RS and Encrypted WS, hence there is no private information leak to PA_j.

Considering Selection agent as an HbC adversary, we would like to prove that, given the view of the protocol, there is no way for this agent to recover the inputs from Personal agent(s). In the worst case for information security, when there are only two users and each user speaks only one language. Selection agent knows the pure or decrypted

AvgQoM value which is calculated from the following equation.

$$AvgQoM(Pi, MT_{i,j}, Pj) = \{[(writing\_skill(Pi, Li) \times accuracy(MT_{i,j}) \times reading\_skill(Pj, Lj)] + [writing\_skill(Pj, Lj) \times accuracy(MT_{j,i}) \times reading\_skill(Pi, Li)]\}/2 \tag{4.1}$$

In this equation, there are 6 variables, unknown to Selection agent, so we can make an information theoretic argument that, Selection agent cannot recover the scores from each Personal agent. However, in reality when each user only speaks one language, there is no need to look for the best-balanced language combination from the beginning. AvgQoM and all language score should be naturally and undoubtedly high since everybody use her/his native language. But when user(s) speak more than one language, the complexity is even higher for Selection agent to know or guess user language score, since it does not know whose data and what language each AvgQoM linked to, as QoM Table, sent from Process managing agent, only contains encrypted AvgQoM value. There is no user data, Personal agent's data, nor language combination data included in the QoM Table.

For Process managing agent, if this agent gets plain AvgQoM, then it implies violation of semantic security. However, AvgQoM seen by Process managing agent are all encrypted and this agent does not have the private key for decryption. Hence, Process managing agent cannot violate the semantic security. Moreover, if there are two encrypted AvgQoMs that link the best-balanced row indices, and Process manag-

ing agent can identify which value leads which index, there is a violation in distinguish ability of encryption. However, different encrypted AvgQoMs, such as higher AvgQoM in the best-balanced row, can also give the same best-balanced index result. As a result, distinguish ability of encryption is not violated.

Using view of the protocol for each agent, we can conclude that, this protocol satisfies our privacy-protection objective. However, some actions that lead to information leakage cannot be prevented. Natural leakage of information could happen, regardless of how secure the system is. Even though the system is ideally trusted and secured, the leakage is going to happen unavoidably and naturally. In our case, when a Personal agent is malicious and joins the system numerous times, it could change its data input little by little, and look for the change in recommended language result. With this method, the Personal agent could guess the level of language ability of another user especially when there are only two users in the system.

4.5.2 Implementation

Calculation time can be tremendous when there are many users and each user speaks many languages, since it is related to 2 factors, including the number of users and the number of languages. However, in the real setting, best-balanced machine translation is created to be used in multi-lingual chat and usually used by 3-4 users for discussion, while each user speaks 2-3 languages. So, the computation time is still acceptable. As in real implementation, we have tried to measure time

used and message count on a personal computer, with 4 users, each speaks 2-3 languages, it took around 1.8 seconds to finish the plain calculation using out proposed method.

Due to the different implementation environment, the implement adjustment might be needed. Even though some encryption algorithms, which are fully homomorphic encryption, allows multiplication of two encrypted value. However, this is relatively new and only available when implemented on C++. Many encryption algorithms, such as Paillier, do not support multiplication of encrypted value but supports multiplication of one encrypted value and a pure value. In this case, instead of encrypting reading skill and writing skill, encrypting machine translation quality and multiply by own reading skill and own writing skill could be another option.

4.5.3 Privacy-aware BBMT as A Framework

This method is proposed based on the privacy problem existed in the original BBMT, however; this method can work as a framework for other computations. It is possible to apply our method of combining secure computation and multi-agent system to different situations where the data is needed to be protected but also needed to be used as a part of computation.

4.6 Conclusion

The original method to compute best-balanced machine translation combination requires users to disclose their language test scores for the calculation. Since there are some users who do not wish to share their language scores, language scores should also be treated as private information. We have proposed a protocol to compute best-balanced language combination without disclosing user's language scores. We combined cryptography technic and multi-agent system together and propose a 3-agent-type system to create a privacy-aware protocol for this calculation, since the existing privacy protection methods are not suitable for best-balanced machine translation. Our method enables the calculation for the best-balanced language recommendation while having language score kept with the user's Personal agent. We also proved the privacy protection effect using security arguments.

We introduce secured computation to protect user's private information for multilingual collaboration, as we would like to emphasize the importance of user's data privacy in human-computer interaction and computer-mediated intercultural collaboration. Because it is significant to treat test scores confidential, as they are user private information, our proposal ensures that user language score will not be disclosed except the natural leakage which is unavoidable.

Chapter 5

Automated Cultural Difference Detection

To successfully establish mutual understanding, is it essential for people to realize if they understand or misunderstand each other. Since it is difficult for human to identify the misunderstanding, this chapter provides a method to automatically detect the possible misunderstanding based on words. We discuss the relationship between culture difference and images, then propose our processes and discuss the results of our experiment.

5.1 Introduction

In intercultural collaboration, the absence of a common ground can result in misunderstanding, and it might take time for the participants to

realize that they are misunderstanding or failing to establish common ground. To identify the ongoing misunderstanding in an intercultural team, the team member need to have enough knowledge of both cultures, or knows the team members very well. An attractive solution is extend MT systems such they can identify likely causes of misunderstanding.

Therefore, the purpose of this work is to propose a baseline method to automatically detect words that could cause misunderstanding when the team members speak different languages. We utilize an image comparison technique to detect these words. Because people with different language background sometimes see things differently [Deutscher, 2010], images from existing databases can be linked back to the languages and keywords which would allow us to identify the cultural differences. As far as we know, this is the first study to base the automatic detection of cultural differences on image comparison.

The outputs of this method can be used to support intercultural design workshops and MT-mediated communication. Potential cultural differences can then be visualized or the MT system can warn the user that the word used has a high probability of triggering a misunderstanding.

5.2 Motivation

5.2.1 Motivating Scenario

Obviously, multilingual communication is more difficult than monolingual communication. The differences in cultural backgrounds can yield even more problems if no common ground exists between the parties. Our previous work [Pituxcoosuvarn et al., 2018] reported collaboration difficulties among children during a workshop. Briefly a block of clay was shown to the children and they were asked "what does this looks like". A Japanese participant said it looked like 'red bean paste' but the children from other countries did not understand, since red bean paste in their culture looks different. The problem was solved by finding images of Japanese red bean paste, and showing those images to the other children. In this case, even if MT output was correct, this problem still exists because the team members have different backgrounds. Fortunately, they realized that they did not fully understand each other after a moment, so they could resolve the misunderstanding. Failure to identify misunderstanding rapidly delays the collaboration significantly and might lead to failure. In many cases, people do not realize that there is misunderstanding happening during the conversation, instead they think they understand but, they do not. Our solution is to create a tool that can detect cultural differences and thus possible misunderstandings.

The prior study found that the children had different mental images due to their diverse cultural backgrounds, the problem was solved by

searching for images in their language and share them with the other children, so they finally understood each other. From this case, we would like to investigate if the difference or the similarity of images can link us back to the cultural difference or give us information about cultural background.

5.2.2 Cultural Difference and Images

Cho and Ishida [Cho and Ishida, 2011] referred to the detection of cultural difference as detecting semantic difference based on the culture definition of Geertz [Geertz, 1973], which defines culture as "a historical transmitted pattern of meanings embodied in symbols". They said that it can be viewed as 'a pattern of interpretation' or 'a pattern of semantic'. Our method is designed to detect semantic difference in two languages and thus could be viewed as cultural difference. We expect it to be used as a tool to predict misunderstanding that could happen. The children' s workshop indicated that the world could look different through the language glass. Besides the event, Deutscher [Deutscher, 2010] gave a lot of interesting examples of how different language speakers see the world differently. One example provided is that the historical evidence of ancient people shows no reference/usage of blue color, not because they saw less color than we do, but because blue is extremely rare in nature, so there was no need to find a name for this color. He also mentioned that some languages have no word for 'time', and some languages use four cardinal directions instead of 'left' and 'right'. It is reasonable then to assume that people with different language

Figure 5.1: Samples from images of '団子' (dan-go, Japanese sweet) (left) and 'dumpling' (right)

backgrounds might have different images in their mind and might have different thoughts when presented with the same word or the translation of the same word. Nowadays, there are databases of images with annotations in various languages and many image search tools are available. If these images and their keywords satisfy users by providing good images for the keyword input in different languages, we could use the image database as a tool to identify cultural difference. For instance, the word '団子' (dan-go), which means Japanese sweet dumpling made of rice flour, can be translated into 'dumpling', however, when we look for images of '団子' and 'dumpling' the results are totally different. Even though the translation of this word is not wrong, Japanese speakers and English speakers have different mental images when presented with these two words, as shown in Fig. 5.1. From this example, we can infer that an annotated image database covering the different languages can help us identify cultural differences between or among the speakers of different languages.

To detect cultural difference, there are various resources that can be used, such as dictionaries, thesauruses, etc. However, those resources

are created focusing on the words' meaning, not their shapes and textures. In design, how object looks like is also substantial. From the problem in KISSY workshop about the read bean paste, if we look at text-based resource, the description of this word in Japanese and the other languages might not be different, but when it comes to image, we can easily see the different. Because of this, images could be linked to cultures based on how they look like in different languages.

5.3 Related Work

5.3.1 Cultural Difference Detection

Cultural difference detection Groups of researchers have been studying and identifying cultural difference. Most of them collect data using cross-national surveys and then analyze the survey results. For example, Hofstede's cultural dimension [Hofstede, 1984] is a well-known model of culture. Yoshino [Yoshino and Hayashi, 2002] also conducted a cross-national survey and compared the social values, ways of thinking, and other attributes. The results from these studies are interesting, but it seems impossible to apply the results to achieve real-time detection.

Cho et al. [Cho et al., 2007] studied the cultural difference in pictogram interpretations and its pattern. They did a web survey to understand the difference in pictogram interpretation in the U.S. and Japan. They found that, 19 of 120 pictograms were judged to have culturally different interpretations. They also found three patterns of the interpretation difference including, "two cultures could share the same concept but

with different perspectives", "two cultures partially share the concept", and "two cultures do not share any concept".

Yoshino et al. [Yoshino et al., 2015] proposed a method for cultural difference detection in Wikipedia. The initial dataset is created by examining words and phrases with different meaning and usage in Japanese and Chinese by 18 Japanese students and five Chinese students. They proposed a flow of judgements based on the initial dataset. They evaluated four judgements as being successful in assessing cultural difference, including, The article is not explained from a global viewpoint", "While the Japanese Wikipedia version of the article mentions Japan, the Chinese version does not.", "Existence of a defining statement, categorization by country name, and reference to origin or target country.", and "Neither country name is mentioned in either language version of Wikipedia."

Existing works on cultural difference identification and detection have been conducted within specific areas of use. Moreover, the need for human judgement is inevitable. Accordingly, our work focuses on developing a method that can automatically detect cultural differences present in a broad range of domains.

5.3.2 Support Tools for Heterogeneous and Intercultural Team

In interdisciplinary and cross-cultural collaboration, the team members have various backgrounds and if they are not aware of it, difficulties in

collaboration can arise. A group of researchers [Mattioli et al., 2018] proposed a support tool to create awareness of the bias in design teams. Their goal was to make each member be aware of her/his own interpretation of a topic while understanding and respecting the other team members' viewpoints. Their process includes asking each member to make a bias card by choosing three pictures for a topic, together with text (up to 140 characters), and sharing the cards among team members.

Their proposed tool can help create mutual understanding among the members. However, topics must be selected and the same process must be performed repeatedly. By comparison, our proposal can create mutual understanding without the need for topic assignment and eliminates the need for repetitive checks.

5.4 Method

To resolve the problem mentioned in Section 1, we introduce a method that can detect possible misunderstanding when communicating across languages by comparing images that are associated with keywords in different languages. We chose to use image comparison for several reasons. First, images are well linked to language and culture, as explained in Section 2. Second, it is often said that "A picture is worth a thousand words". Images contain information that might not be present in a dictionary. Finally, since image databases and image search engines are available, it is more convenient and less time consuming to use them together with an image comparison technique to automate

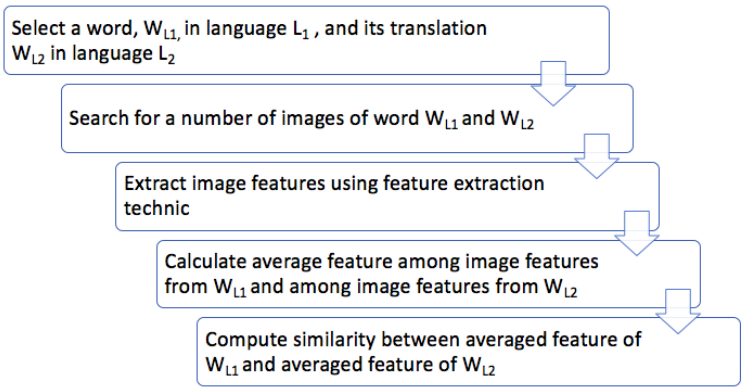


Figure 5.2: Similarity calculation procedure.

the detection of possible misunderstanding. Fig. 5.2. shows the overall procedure used to compute the similarity of a word and its translation in another language. If the similarity is low, there is more possibility of misunderstanding when those words are used in cross-culture and cross-language communication.

First, a word, W_{L1} is selected in language, L_1. Its translation in language L_2 is W_{L2}. We look for a certain number of images in an image database or in the put of an image search engine for both words. The number of images should be sufficiently large, since many images may not well represent the keywords. Ideally, all the images linked to a word should be similar, but in many cases the images are rather diverse. For example, when the keyword is 'lion', the image results are mostly pictures of lions. But, when the keyword is 'zoo', the image results could include pictures containing different kinds of animals. If we only chose one image for the word 'zoo', we might get only the image of a lion, which does not well represent 'zoo' but has high similarity to the word

'lion' instead. To calculate similarity, after randomly selecting images, an image processing technique is used to extract features of all images for word W_{L1} and word W_{L2}. Feature extraction is a dimensional reduction process so the original image data can be simplified and processed. The features extracted from the images for the same word are averaged and compared with processing results for the other word. It is also possible to compare every extracted feature but this would take too long time and consume large amounts of computation resources. Lastly, the similarity between the two averaged features is computed. This similarity usually ranges from 0 to 1. A lower similarity indicates a bigger difference between the images and thus a higher chance of misunderstanding.

5.5 Experiment

We conducted an experiment to examine the proposed detection method. The selection of the data source and tools used were based on simplicity, and they could be adjusted or replaced with other resources or services for better accuracy. The software for this experiment, was written in Python.

The system computed the similarity between words in English and Japanese using words from Japanese WordNet [Isahara et al., 2008], a Japanese-English lexical database, created from the original English WordNet of Princeton University [Miller, 1998]. In WordNet, lemmas, the dictionary form of words, are linked to sets of synonyms called synsets. For Japanese WordNet, Japanese lemmas and English lemmas

are linked to the same synset. We conducted this experiment on 2,500 randomly-selected noun synsets. Around half of the synsets Japanese lemma linked to the synset, so it is impossible to calculate the similarity for those synsets. Based on the detection method proposed in Section 4, first, we randomly selected a synset from the Japanese WordNet database and selected one or two lemmas in each language, based on its availability. If there were more than two lemmas, we run the same similarity calculation program with ten images for each lemma and choose the most similar two lemmas. Since some synsets have excessive number of lemmas making the calculation infeasible, if there were more than five lemmas for a language in one synset, we randomly selected just five lemmas and calculated the similarity among them to find two lemmas for the next step of calculation. The reason behind this is when there are too many lemmas, some lemmas have slightly different, boarder, or more specific meaning than the others, so we attempt to find the two most similar lemmas that can strongly represent the synset in each language.

For this experiment, we downloaded 30 images for each word if there was one lemma per language and 15 images for each word if there were two lemmas, with Google Image Download API [1]. The images were resized to 224 pixels x 224 pixels for feature extraction. To extract image features, we used one of the most popular tools, VGG16 [2] from Keras. The software transformed visual information of the image into a vector space. The result of feature extraction was a vector that contains 4,096 feature values for each image. An example of feature extraction

[1]https://github.com/hardikvasa/google-images-download
[2]https://keras.io/applications/#vgg16

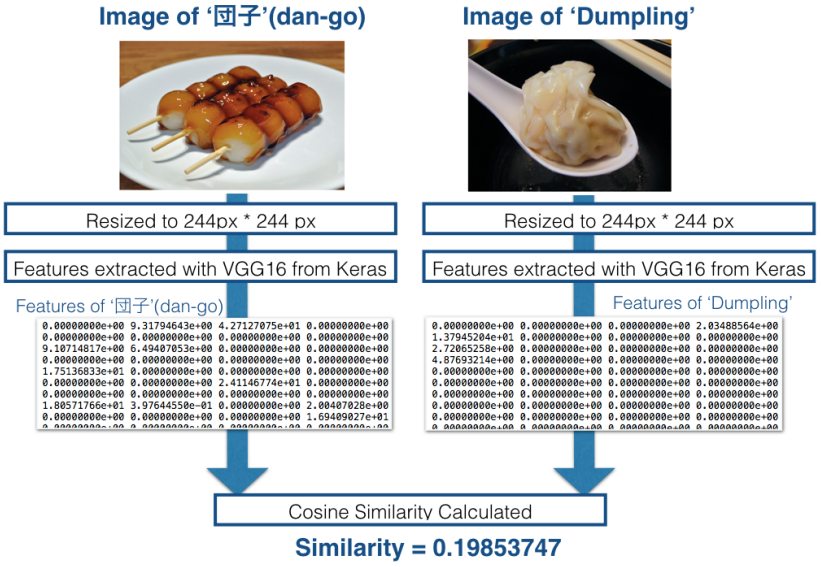


Figure 5.3: An example of feature extraction and image comparison process

and comparison is shown in Fig. 5.3. After that, averaged features were calculated. The synsets with two lemmas had their averaged features calculated from both lemmas.

Several similarity measures can be used, but for simplicity, we used cosine similarity, one of the most common methods of comparing vectors,. Given A is the feature vector for one word and B is the feature vector for its translation, cosine similarity was calculated as follows:

$$similarity = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} \boldsymbol{A}_i \boldsymbol{B}_i}{\sqrt{\sum_{i=1}^{n} \boldsymbol{A}_i^2} \sqrt{\sum_{i=1}^{n} \boldsymbol{B}_i^2}} \tag{5.1}$$

We iterated the program 2,500 times on Windows servers. The calculation took several days since it involved around 150,000 images and heavy computation loads.

5.6 Result

We calculated the similarity of words in 2,500 synsets. Fig. 5.4 displays the numbers of synsets grouped by the calculated similarity result. Most synsets, almost 60% of the synsets, contain lemmas with similarity values between 0.7 and 0.9. Synsets with similarity lower than 0.6 can be considered as low similarity synsets and the possible cause of misunderstanding, misperception, or different interpretation of words in those synsets. There are more synsets with high similarity than synsets with low similarity, because in the real world, most words do not cause misunderstanding or misperception.

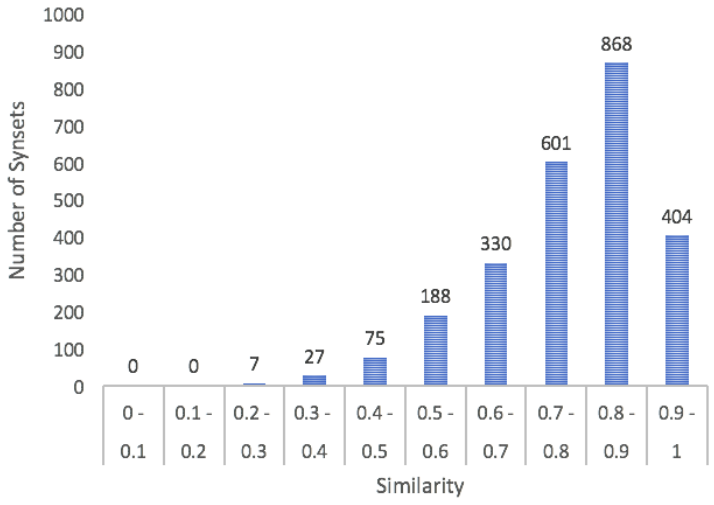


Figure 5.4: The Histogram of word similarity in synsets

Fig. 5.5 displays typical examples of synsets with similarity values lower than 0.6 and higher than 0.6. The similarity values are rounded to four decimal places. Each synset used for the calculation contained one to two lemma(s) in English and one to two lemma(s) in Japanese.

The images under each lemma are chosen just to demonstrate the meaning of the lemma and represent overall images to the reader. Some images in the table are not used in the real calculation, but are similar to those images that were used in the calculation. These images are displayed for the reader understanding, because most images downloaded in our experiment are not permitted to be publicly reused.

From Fig 5.5, it is obvious that synsets with high similarity contain more similar images from its lemmas. For synset 04317175, the word

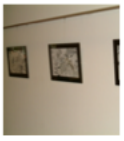
Similarity		Synset Data	English lemma 1	English lemma 2	Japanese lemma 1	Japanese lemma 2
Low	0.2731	10913871 {william_cowper, cowper, クーパー}	william cowper	cowper	クーパー (Mini Cooper)	-
						
	0.3660	00269674 {makeover, リフォーム}	makeover	-	リフォーム (reform)	-
						
	0.5024	07491476 {amusement, 御楽しみ, 娯しみ, 楽しみ, 御楽, お慰み, 慰み, 御楽み, 御慰み, 楽しび, 面白さ, お楽しみ, 愉しみ, 興}	amusement	-	娯しみ * (recreation, playfulness)	お慰み * (recreation)
						
High	0.7632	07673397 {oil, vegetable_oil, 植物油}	oil	vegetable oil	植物油 (vegetable oil)	-
						
	0.8396	13490343 {growth, 成長, 進化, 発展, 進歩, 発達, プログレス, 発育}	growth	-	成長 (growing)	進歩 (progress, improvement)
						
	0.9677	04317175 {stethoscope, 聴診器}	stethoscope	-	聴診器 (stethoscope)	-
						

Figure 5.5: Example of synsets with high and low similarity from the result

‘stethoscope’ and ‘**聴診器**’(stethoscope) refer to the same object and so give the same interpretation to the reader. Synset 13490343 also has high similarity but slightly lower than the synset of word ‘stethoscope’. Images of all the words in this synset, including ‘growth’ ‘**成長**’(growing) and ‘**進歩**’(progress, improvement), yield similar images and they have similar meaning. But, since these words are abstract nouns, the images are slightly different. For example, images for all those 3 words contain images of small trees, in addition to graphs with upward arrow and figures representing evolution and development. Synset 07673397 contains the words ‘oil’, ‘vegetable oil’, and ‘**植物油**’(vegetable oil); it has slightly lower similarity than the first two synsets. However, the word ’oil’ has broader meaning than the other two words so it yielded images of drilling rigs and oil in containers.

To detect the cultural difference and misunderstanding due to different language and backgrounds, we focus on synsets with low similarity. Synset 10913871 has remarkably low similarity. The reason for the similarity result between ‘Cowper’ , ‘ William Cowper ’ and ‘**クーパー**’(Mini Cooper) could be the nature of language. Because Japanese language has relatively few phonemes, many words derived from foreign language can be written in Japanese character but they are pronounced differently from the original words. Both ‘Cowper’ and ‘William Cowper’ are person’s names but ‘**クーパー**’ is pronounced ‘Kuu-Paa’. It is the homophone of ‘cooper’ in Japanese which makes most people think of Mini Cooper, i.e. the car made by the automobile marque called Mini. When an English speaker person uses the word ‘Copwer’, without any further explanation, a Japanese speaker might misunderstand

that the speaker is talking about cars.

Synset 00269674 involves two words: 'makeover' and '**リフォーム**'(reform). 'Makeover' is defined as "the process of improving the appearance of a person or a place, or of changing the impression that something gives" [Hornby and Cowie, 1995]. But when it is used in English, people will think about a personal makeover, usually involving make up and change in appearance. Image results from this word mostly include picture of women before and after a beauty makeover. Whereas '**リフォーム**', in a Japanese dictionary, is defined as: 1) Revise, improvement 2) To remake, to resize or redesign clothes, to renovate building(s). The main image created by a Japanese is usually related to building renovation. Images of this word are usually images of a renovated room.

Synset 07491476 links 'amusement', '**娯しみ**'(recreation,playfullness), to '**お慰み**'(recreation). The Oxford Advanced Learner's Dictionary defines 'amusement' as " the feeling that you have when something is funny or amusing, or it entertains you" and "a game, an activity, etc. that provides entertainment and pleasure". The images of this word show mostly pictures of amusement parks and rides. The word '**娯しみ**', and '**お慰み**' are rarely used in daily life. They have similar meaning to 'amusement' in terms of fun and pleasure, but their use is different and more complex. '**娯しみ**' is not included in standard dictionaries but are used by only some groups of people. However, its adjective, '**娯しい**', exists in dictionaries describing the feeling of fun and enjoyment. The kanji, Chinese character used in Japanese language, '**娯**' means recreation and pleasure. The images of '**お慰み**' are

related to ink painting magazines and painting exhibitions since they are used as magazine names and exhibitions. The image results are much different from each othes, since this word is ambiguous and intangible. A few images from '**お慰み**' are related to handmade objects and craftwork due to the usage of this word. Even though the meaning of this word is mainly about fun and recreation, this word can be explained as "To create enjoyment, give people pleasure". Unlike the other words in this synset, this word involves persosnal pleasure. It is also often used conditional, for example "it will be fun, if I succeed".

5.6.1 Pattern of low similarity synsets

From the results of our experiment, low similarity has a few reasons that cause. Synsets with low similarity could lead to misunderstanding and different interpretations between English and Japanese speakers. When we look at the synsets with similarity lower than 0.6, we find two interesting causes related to cultural difference.

First, a word and its translation have the same meaning but represent different images in the languages. Because of the different backgrounds, language and culture, people look at a word and its translation differently.

Second, words in one language can have more specific or broader meaning than in the other. The cases in this category includes:

1. Broader meaning in one language - When the word in one language has broader meaning, it might yield a greater variety of

images.

2. Several homonyms in one language - When one word has several meanings it could confuse people even for native speakers. Some of the meanings might include slang or a negative interpretation compared to its translation.

3. Specific noun in one language, i.e. name of famous person, brand name, product name, company name - When a word is used as a specific noun, such as a product name, in one area, many people will think about the product instead of the original meaning of that word.

5.7 Discussion

5.7.1 Visualization of the cultural difference

Existing work [Mattioli et al., 2018] raise the possibility of sharing images and short texts made by each team member in design workshops to help mutual understanding and own bias. Here, we would like to present an alternative tool to be used in the same situation. The key advantage of our tool is that the information to be shared can be created automatically using data output by the proposed cultural difference detection method. Using visualization can help team members realize the difference and understand each other better; Fig. 5.6 is an example. This graph can be shown to the team member together with the images already downloaded for the calculation. When the

word ‘makeover’ is mentioned, we grow the tree graph from the root nodes for a few levels and investigate the neighbor nodes using data from Japanese WordNet. In this graph, we use 0.6 and 0.75 as thresholds to determine node color. We use green for similarity values of 0.75 and above (high similarity nodes), yellow for values between 0.6 and 0.75 (medium similarity). Yellow nodes might contain different interpretations of words or concepts where the difference is not obvious enough to cause misunderstanding but caution in advised. Red nodes are for words with similarity under 0.6; they have high possibility of causing misunderstanding. The hierarchy of this tree is based on the relationships in WordNet. The upper nodes are words with broader meaning than its lower originating node, or a hypernym of the lower node. Lower nodes are words with more specific meaning than their upper nodes, i.e. a hyponym of the upper node. In a workshop, such as a brainstorming workshop for an advertising idea, when the English speaker mentions ‘makeover’, the Japanese member might look at the translation and think about it as ‘reform’ and assume that the discussion is about ‘reform’ not ‘makeover’. The automated system can warn and display this tree to the participants. Since, this tree seems to fit the images from Japanese speaker’ s side, the Japanese can use it to confirm their meaning and the English speaker can see what Japanese speaker thinks about this word since this word is about reconstruction, fixing, and improvement, given that the upper nodes have fair or high similarity.

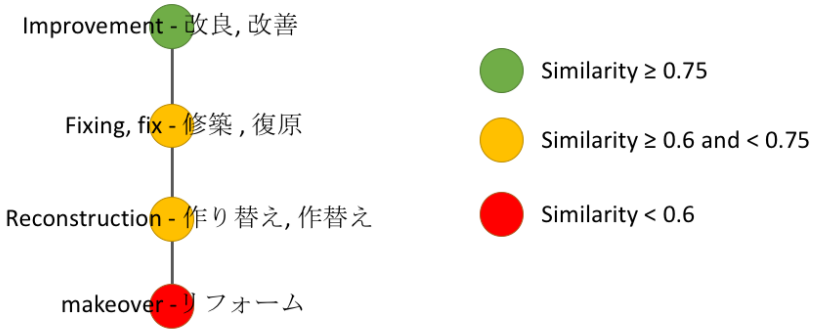


Figure 5.6: An example of visualization of cultural difference

5.7.2 Application of the Result

The results from our detection method can be used in MT-mediated communication. When a word used in the translation process, i.e. in a chat system, is identified as a low similarity word in our database, we can warn the team members and offer the word tree and related images to them. Even for intercultural workshops or intercultural design workshops that are conducted in one language, it is still useful to warn the team about possible misunderstandings when there is non-native speaker in the team. We can also apply this warning implementation in computer mediated communication between native and non-native speakers. The result can be used as reference for culture and language studies as well. It can be modified and used as an additional tool in existing language learning research, for example to help identify false friends, words in two languages that look or sound similar but differ significantly in meaning [Abou-Khalil et al., 2018] in language studies .

5.7.3 Limitation

Besides cultural differences low similarity is sometimes the consequence of imperfect translation of the language resource. Low similarity can also be caused by technical errors, for instance, few images are available for download. We note that abstract nouns are one of the problems our method has difficulty with, since they are not linked to unambiguous images.

5.8 Conclusion

Our main contribution of this chapter is to present a novel method to automatically detect cultural difference, and possible misunderstanding, or possible misinterpretation of words. We aim to use it for MT-mediated communication and intercultural workshops. We believe that, this is the first work to apply image comparison to the identification of cultural differences. Since existing works [Hofstede, 1984] [Yoshino and Hayashi, 2002] [Cho et al., 2007] [Yoshino et al., 2015] mostly studied and identified cultural difference using the survey method, they can only study cultural differences across specific culture pairs. It also is too expensive and too slow. The proposed method achieves our goal of detecting possible cultural differences, misunderstandings, and misinterpretations in intercultural collaboration automatically. As such, it can be used in broader areas of study and does not need human effort.

We investigated our method by applying it to Japanese WordNet. We looked for dissimilarities between sets of images that represent words in

English and Japanese using an existing image database and automated image comparison. We conducted an experiment on 2,500 synsets and presented some of our results. Low similarity can be due to different images being linked to a word and its translation, resulting from the different backgrounds. The second cause is the unequal meaning assigned to words in each language. For example, when words in one language have several meanings and when one meaning in one language is used as a specific noun, including name of commercial brands, companies, or people.

Chapter 6

Implementation

Since we have proposed the methods to improve mutual understanding in the previous chapter. This chapter proposes an implementation. We also report an experiment to see how our idea is applicable by implementing those proposed ideas in it.

6.1 Introduction

Here, we report an experiment that applied BBMT and the cultural difference detection system, in order to study the effect when both methods when they are implemented in a workshop. To evaluate our two main solutions, we organized an experimental workshop and a controlled workshop. The task of both workshops is the same. The participants in each workshop are different but have similar profiles. We asked the participants to solve a problem using MT for communication

collaboratively. They had to share their information, brainstorm, and negotiate to have a team answer to the task. Lastly, we evaluate our proposals by comparing the qualitative result between the controlled group and the experimental group. Since it is difficult to control participant conversation and there is no guarantee that every conversation will raise a problem of cultural difference, we guided the conversation by giving them a task that has its detail relating to the cultural difference issue.

6.2 System Proposal

We would like to propose a mutual understanding enhancing system in this section, which works as an extension to the multi-language chat system. The multi-language chat system is developed by Connections Co., Ltd. [1]. This chat system allows users to chat using their preferred languages. MT services can be selected from the Language Grid; for example, we used a combination of Toshiba English-Chinese Machine Translation and J-server in our previous experiments and this implementation experiment.

Our implementation proposal is based on the usage shown in the use case diagram, displayed in Figure 6.1. The system requirements are as follow:

1. The system can compute BBMT combination without disclosing language scores except to each user's Personal agent and success-

[1] https://cnnc.jp/

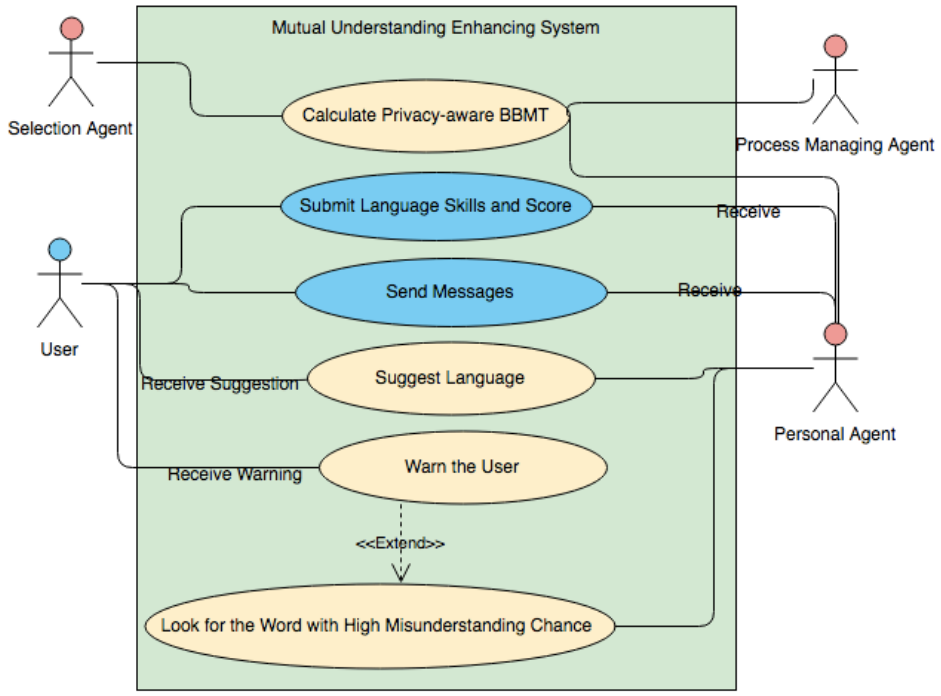


Figure 6.1: Use case diagram of the mutual understanding enhancing system

fully suggest the most suitable language for each user.

2. The system can warn the user of the misunderstanding that might happen when to detect the typing input of the user.

Combining privacy-aware BBMT and the automated detection system, there are four actors in the system, including all agents and the user.

To have BBMT combination computed, the user needs to submit the language skill data to its Personal agent, and the calculation process will be done by the personal agent, process managing agent, and the se-

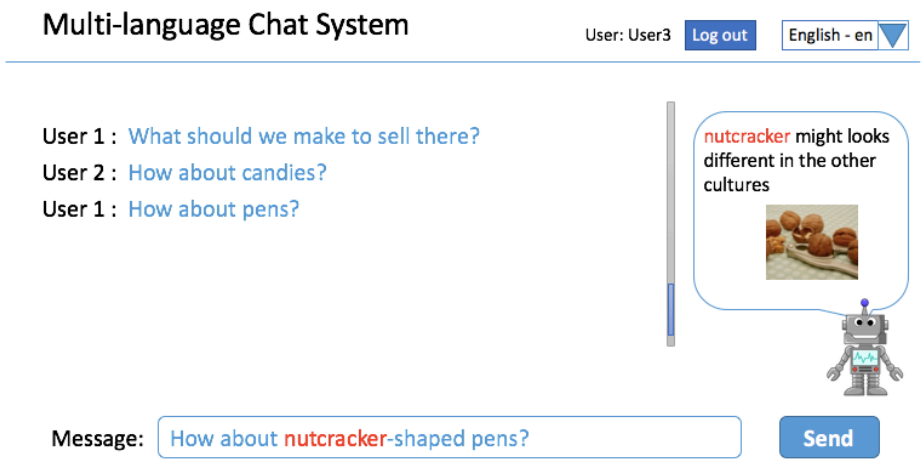


Figure 6.2: An example of the proposed system's screenshot

lection agent together. After the calculation is completed, the personal agent will suggest the language to the user.

During the conversation, when the user sends messages to the system, the personal agent will look for the words that match the words in its list of low similarity. If the word is found, the personal agent will warn the user of the possible misunderstanding, as shown in the screenshot in 6.2

6.3 Experiment Design

6.3.1 Participant

Both the experimental group and the controlled group, each has three participants. First, a Japanese participant with low English language skills. Second, a Chinese participant with medium English skills. Third, a native-level English speaker. All the participants are either students or researchers in their 20s to 30s. The native-level English speaker is educated in the English language environment for at least five years and has a clear understanding of English speaking cultures. All of the participants are well computer literate.

6.3.2 Language Channel

The control group and the experimental group were asked to solve one problem, together with their teammates. The controlled group was assigned to use full MT to translate among Japanese, Chinese, and English, while the experimental group was asked to communicate via the BBMT channel. Using the participants' profile, the BBMT channel recommends the Japanese speaker to use Japanese and the Chinese speaker and the English speaker to use English.

6.3.3 Wizard-of-Oz Cultural Difference Detection Help

Wizard-of-Oz [Maulsby et al., 1993] method is used for prototyping the intelligence system that warns the users of cultural difference. So we could study the effects and avoid the cost of real software implementation in this stage. The wizard used the list of words that might cause misunderstanding due to cultural differences. Most of the words are automatically found by our program described in the previous chapter. The wizard warned the users of each language of the misunderstanding that might occur when using words in the list.

Since there is no guarantee that the participants will use the words in the cultural difference list, we guide the participants with the task and instruction that shape the conversation, which needs the words in that list. The wizard warned the participants of the experimental group before the discussion started.

From the use case diagram in 6.1, the use cases in yellow or light color are actually executed by a wizard.

6.3.4 Task

Modified from the survival games used in BBMT experiments, a situation, as follows, is given to the participants in their native language except for English for a native-level English speaker.

While you and your two new colleagues taking a train across the country

to start your new job abroad, your train has been crashed with a big unidentified object in the middle of a drought area in the picture. Most of the passengers are dead. You have to try to get to the nearest town in the north which is about 50 km away, and each of you has some snacks and drink in your bag, enough for one day only.

Each participant has a list of three in their languages. They need to share the list with the team members. The team member has to choose only one out of three from each list collaboratively. Totally, there were nine on the list given, and only three can be chosen. The collaboration was ended when all members agree on the chosen list.

The original lists given to each participant are shown in Figure 6.3. The words that could cause misunderstanding are underlined to emphasize. For the control group, the lists given to them did not contain any emphasized text.

The list in Japanese are translated into English as
You found a train container that has not been destroyed yet. You can take one of the listed items.

1. *Two bicycles (**ママチャリ**)*
2. *Two corollas (**花冠**)*
3. *A sword*

The list in Chinese can be translated into English as
You found the other three people who took the same train. They are alive but wounded by the wreck. You can choose to help one person and take that person with you. The rest will be helped and carried by the

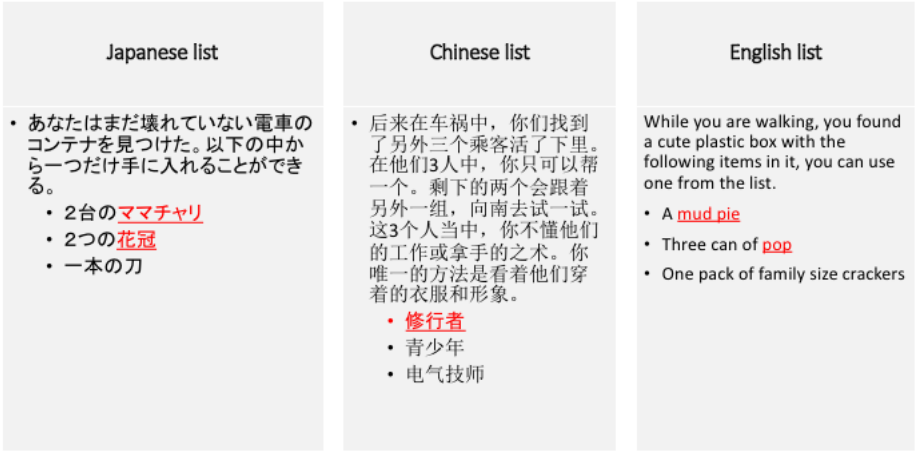


Figure 6.3: Lists given to participants in each language.

other group who are heading to the south. You don't know them, but you can guess who they are from how they look and dress.

1. *A religious practitioner (**修行者**)*
2. *A young teenager*
3. *An electrician*

6.4 Result

This section reports the qualitative result of the experiment.

6.4.1 Quality of Translation

In the control group, where full MT is used, Chinese participant reports that the translation was not correct. However, he could guess the meaning in its context. However, it is difficult to send an understandable message when MT is not good enough to translate precisely. Chinese participant told the other that he has '**修行者**' in his list. This word is translated into 'Practitioner' in English and '**開業医** (Medical practitioner)' in Japanese. But the term '**修行者**' actually means a practitioner of a religion, usually Buddhism. The translation in English was too vague because the practitioner has several meanings, and the translation in Japanese was wrong, as shown in Figure 6.4.

User	English	Japanese	Chinese
CaUser	I have a traveler, a teenager, an electrical technician.	旅行者、ティーンエイジャー、電気技術者がいます。	我这里有一位旅行者 一位青少年 一位电气技师
…	…	…	…
CaUser	Oh, not a traveler is a practitioner.	ああ、旅行者は開業医ではありません。	哦不是旅行者 是修行者
JaUser	practitioner?	開業医？	从业者？

Figure 6.4: Chat log of the control group when the word 'pactitioner' can is mentioned.

User	English	Japanese	Chinese
EnUser	I have a mud pie, three cans of pop and family size crackers	マッドパイ、ポップ缶とファミリーサイズのクラッカー3缶があります	我有一个泥饼，三罐汽水和全家大小的饼干
…	…	…	…
JaUser	Popular music?	ポップミュージック？	流行音乐？

Figure 6.5: Chat log of the control group when the 'pop(soda)' can is mentioned.

The Figure 6.5 shows the chat messages of the controlled group. The

Japanese could not understand the word 'cans of pop'. The word 'Pop can' is translated into 'ポップ缶'. He could not understand, so he asked if it means 'pop music' or not.

In the experimental group, where only English and Japanese are used, there is no obviously serious translation mistake in the chat log. From the interview of the experimental group the participants shared that they have no problem at all with the translation, as in the interview script as follow: *We asked the Chinese user who used English as the conversation medium "Do you have some difficulty in reading", the Chinese user simply answered "No".* The English speaker reported that he has no difficulties except the unexpected choices in a survival game, including, the flower crown(corolla) and human as a choice. The Japanese user was asked if there are any difficulties, but he did not report any except that he did not actually agree on the answer with his teammate.

6.4.2 Skipping the Discussion of Incomprehensible Topic

In the control group, the English speaker was not sure if the *corolla* actually means the headwear or the very popular car. But she did not start the conversation about that and followed the flow of the conversation, especially when the other two persons agree to take *the bicycles* without the discussion of the selection of the other items.

6.4.3 Behaviour of User When the Cultural Different is Warned

Every user of the experimental group tried to explain the items farther when they are warned that those words might cause misunderstanding, as displayed in Figure 6.6, Figure 6.7, and Figure 6.8. While the control group members who were not warned hardly explain the detail of the items except when he/she is asked by their teammates, the experimental group members were more careful to explain when they introduce the word or explain in a few utterances later.

User	English	Japanese	Chinese
EnUser	for me i have a mud pie, a pie made from mud. three can of soda and a 1 pack of family size cracker	私にとっては、泥のパイ、泥から作られたパイがあります。 3缶のソーダと1パックのファミリーサイズのクラッカー	对我来说，我有一个泥饼，是用泥做的。三罐苏打水和一包家庭装饼干

Figure 6.6: A chat message from English speaking user when is warned of cultural difference.

User	English	Japanese	Chinese
JaUser	I have two bicycles, two flower crowns and a sword	私は2台の自転車と，花でできた冠を2つ，そして一本の刀を持っています	我有两辆自行车，两个花冠和一把剑
JaUser	Two people can ride on one bicycle	1台の自転車には，2人が乗ることができます	两个人可以骑一辆自行车

Figure 6.7: Chat messages from Japanese speaking user when is warned of cultural difference.

The English speaker of the controlled group explained the word ‘mud pie’ and change the word ‘pop’ to ‘soda’. The Japanese user uses the word **自転車** (Bicycle) and explains that two people can ride on one

User	English	Japanese	Chinese
CaUser	The two left are monk and eletrician	左の2つは僧kとエレトリシャ	剩下的两个是和尚和贵族
...	...	...	...
CaUser	sometimes monk travel a lot	時々僧kはよく旅行します	有时和尚旅行很多

Figure 6.8: Chat messages from Chinese speaking user when is warned of cultural difference.

bicycle. And the Chinese user who uses English used the word 'monk' that closely describes the word '**修行者**' (practitioner of religion), added the detail: "sometimes monk travel a lot".

6.5 Discussion

6.5.1 Underlying Miscommunication

Sometimes, the discussion seems to be smooth, without any problem; however, after the users actually did not understand the conversation correctly. We experienced this situation from the interview when we asked if the members of the control group understand the word '**修行者**' translated as a practitioner. They thought they understand, but they actually misunderstood. In that situation, there was no question raised, and the conversation was too short for the team members to realize that there is a misunderstanding. This problem is deeper than the other problems on the surface, but it can lead to ineffective collaboration.

6.5.2 Altering Words

Both in the controlled group and experimental group, sometimes the participant alter the word they used from the word given in their documents. In the real world, especially when writing, we usually use synonyms to provide variety and interest for the reader [Bailey, 2003]. People also choose different synonyms that they think it is the most suitable to describe their thought[Gookin, 2013]. In our experiment, we found that even though the participants in the controlled group were not warned of the cultural difference, they sometimes choose different synonyms to communicate, but less than the experimental group. For instance, the Japanese participant in the control group used the word '**自転車**' (bicycle) instead of the word '**ママチャリ**' (A specific type of bicycle often used by mother and often seen with a child seat attached). He explained in the interview that he thinks to think the word '**自転車**' (bicycle) is more suitable and more formal. However, a piece of information was missing with this situation since '**自転車**' (bicycle) can give a different impression than '**ママチャリ**' and it is easier to imagine that two-person can ride on a '**ママチャリ**' (mother bicycle).

In the future, it is also interesting to study how people adjust their words when they realize or not realize that they are communicating with foreigners with or without MT. This study could help with the design of multilingual communication in the future.

6.6 Conclusion

We proposed a system that implements both BBMT and automated cultural difference detection. We conduct a small experiment based on the proposal to demonstrate that applying the result of BBMT and automated cultural difference detection to workshops are beneficial. We used the result from automated cultural difference detection and used the Wizard-of-Oz method to warn the experimental group members of possible misunderstandings that might occur when using some particular words. From the experiment result, the team members of the experimental group are more careful and tried to explain those words clearly, so the other team member would understand and successfully establish mutual understanding. The result also shows that the experimental group who follow BBMT language recommendation had less mistranslation, which allows them to have a mutual understanding.

Chapter 7

Conclusion

This chapter concludes the contributions of this thesis and future direction based on our study.

7.1 Contributions

This thesis presents three contributions toward improving mutual understanding in MT-mediated communication. To establish mutual understanding, it is crucial to prevent miscommunication and misunderstanding. First, we proposed BBMT to improve help in selecting proper languages to be used based on MT quality and users' language skills to decrease miscommunication. Second, we proposed a privacy-aware version of BBMT, so it does not disclose private data, i.e., language score, to be disclosed for the computation. Third, we proposed a method to automatically detect words that could cause cultural background

related misunderstanding between two languages, so we can use it to prevent misunderstanding and increase mutual understanding. We, in this section, review these contributions.

1. Best-balanced machine translation(BBMT)
   We proposed a model that helps to select the best language combination using the concept of balancing the quality of communication. We adopted the existing concept of quality of message and combined it with Pareto optimality. We set up an experiment comparing our method to widely used methods of communication, including using English as a shared foreign language and fully use of MT. In the experiment, our BBMT method produced fewer communication breakdowns comparing to using full MT, which indicates a better quality of communication. It helps to deal with the unequal translation that might cause misunderstanding for low-language resource users and might prevent users from establishing mutual understanding. Moreover, the experiment showed that utterances of participants who have limited skills in a shared foreign language increased when using MT services both in with MT and BBMT. Our model can help to deal both with imbalanced participation in conversations and minimizing the errors caused by MT as the users' language skills are taken into consideration when the usage foreign language of the user is better than the machine translation in some situations.

2. A privacy-aware computation of BBMT
   Because it is essential to keep private data, such as test scores, protected, but the data is needed to compute BBMT combina-

tion, we have proposed a protocol for the computation without disclosing the user's language scores. We applied a cryptography technic and multi-agent system together and proposed a system with three types of agents for the privacy-aware protocol. Since none of the existing privacy protection methods is suitable for BBMT calculation. Our approach enables the calculation while having a language score kept only with the user's Personal agent. To prove the privacy protection effect, we made security arguments. In this study, our contribution includes introducing secure computation into intercultural collaboration. We aimed to emphasize the importance of user's data privacy in human-computer interaction and computer-mediated intercultural collaboration, considering that it is significant to treat test scores confidential, as they are private user information. Except for the natural leakage, which is unavoidable even if the system is entirely secured, our proposal can ensure that the user language score will not be disclosed.

3. An automated detection of cultural difference
   We presented a novel method that is able to detect words that might cause misunderstanding automatically. We examined our approach on 2,500 sets of words in the Japanese WordNet, a Japanese-English concept dictionary. We looked for dissimilarities between sets of images that represent words in English and Japanese. Using an image database and automated image comparison technic. The result of the low and high similarity words set to reflect the similar and dissimilar concepts in English and

Japanese, which shows that our proposed method achieves our goal of detecting possible cultural differences, misunderstandings, and misinterpretations automatically.

To show the applicability of the whole framework, we proposed implementation. We conducted an experiment based on the Wizard-of-Oz version of the application with a control group who use full MT for communication and an experimental group with the support of BBMT language recommendation, and cultural difference warning from the detection. We found that, in the experimental groups, a misunderstanding occurred much less than in the control group. The result shows that our proposals to improve mutual understanding are applicable and successfully decrease misunderstanding and increase mutual understanding.

7.2 Future Direction

In this section, we will describe a few methods to improve the current research and a few areas for future research using the current research result. The proposed solutions open the possibility to pursue future direction as follows:

- *Implementing the support system in a real-world situation.* We have conducted laboratory experiments for both BBMT and cultural difference detection system. However, in the real-world setting, the result and problems in collaboration could be different from the laboratory setting. It would be beneficial to study the

effect of the proposal in real usage by implementing a tool that applied both functions, so we could farther improve the system.

- *Near real-time calculation of BBMT and quality of translation message.* Because the quality of translation also depends on the context of the message. It might be able to translate simple conversation correctly while fail to translate academic discussion. It is necessary to measure the quality of MT for each different conversation in near real-time. After having a real-time evaluation of MT, we can actively compute BBMT combinations and suggest the user the best combination in each stage of communication.

- *Improving accuracy of cultural difference detection by considering its confidence.* The resulting quality of the cultural difference detection system can be improved by considering its confidence value. The current result includes the 2,500 sets of words, but some sets of words have too various image data. For example, the keyword 'Zoo' will be linked not only to the image of the zoo but also to images of lions, monkeys, and other kinds of animals. This variation of image data seriously interferes with the similarity value when calculated by the machine. When this kind of word is compared with its translation, the similarities of both languages are usually lower than they suppose to be. My hypothesis for this case is, if we give a lower confidence value for the concept with various styles of image and involve this value in the similarity calculation, the warning system would warn the user more accurately. The system can distinguish the low similarity concept with low confidence as the less priority to-warn word.

- *Extending cultural difference detection to cover different language resources.* We used only WordNet as in the current cultural difference detection system as our language resource. The system quality can be improved by integrating different language resources, such as dictionaries and thesaurus, in the computation. My research challenge in this topic is to automatically compare dictionaries in two or more different languages to look for the cultural difference in the word usages and its synonym(s). My hypothesis is the meaning and synonym can be linked to its concept and originality of the word. So, the same word in different languages might include synonyms that have a different concept from the synonym of another word. If my hypothesis is correct and the automation is successful, we will have another list of words that might trigger misunderstanding and culturally related problems. This list can enhance the prediction of the current research, and it can correctly warn the users of cultural differences more correctly.

- *Evaluation and comparison of the node graph of the cultural difference visualization tool.* In Chapter 5 discussion, we proposed a visualization of the cultural difference based on the result of the cultural difference detection system. In the future, we could implement and evaluate this visualization tool. Moreover, comparing word trees that include similarity data is also interesting and could bring about the relationships between the low similarity nodes that are different or similar to their roots.

Publications

Journal

Mondheera Pituxcoosuvarn, Takao Nakakuchi, Donghui Lin, and Toru Ishida. Privacy-aware Best-balanced Multilingual Communication. *IEICE Special Issue of Knowledge-Based Software Engineering.* 2020. (*conditionally accepted*)

Mondheera Pituxcoosuvarn and Toru Ishida. Multilingual Communication via Best-Balanced Machine Translation. *New Generation Computing.* vol. 26, no. 4, pp 349-364, 2018.

Conferences

Mondheera Pituxcoosuvarn, Takao Nakakuchi, Donghui Lin, and Toru Ishida. Secure Agents for Supporting Best-balanced Multilingual Communication. *HCI International 2020*, Copenhagen, Denmark, 2020. (*accepted*)

Mondheera Pituxcoosuvarn, Donghui Lin, and Toru Ishida. A

Method for Automated Detection of Cultural Difference Based on Image Similarity. In *Proceedings of the International Conference on Collaboration and Technology 2019*, Kyoto, Japan, September, 2019.

Mondheera Pituxcoosuvarn, Toru Ishida, Naomi Yamashita, Toshiyuki Takasaki, and Yumiko Mori. Machine Translation Usage in a Children's Workshop. In *Proceedings of the International Conference on Collaboration and Technology 2018*, Costa de Caparica, Portugal, September, 2018.

Mondheera Pituxcoosuvarn, Toru Ishida, Naomi Yamashita, Toshiyuki Takasaki, and Yumiko Mori. Supporting a Children's Workshop with Machine Translation. In *Proceedings of the 23rd International Conference on Intelligent User Interface Companion (IUI)*, Tokyo, Japan, March, 2018. (*short paper*)

Mondheera Pituxcoosuvarn and Toru Ishida. Enhancing participation balance in intercultural collaboration. In *Proceedings of the International Conference on Collaboration and Technology 2017*, Saskatoon, SK, Canada, August, 2017. (*best student paper award*)

Bibliography

[Abou-Khalil et al., 2018] Abou-Khalil, V., Ogata, H., et al. (2018). Learning false friends across contexts. In *LAK'18: 8th International Learning Analytics and Knowledge (LAK) Conference*. Association for Computing Machinery (ACM).

[Aiken, 2009] Aiken, M. (2009). Transterpreting multilingual electronic meetings. *Int. J. Manag. Inf. Syst*, 13(1):35–46.

[Anders et al., 2012] Anders, G., Siefert, F., Steghöfer, J.-P., and Reif, W. (2012). A decentralized multi-agent algorithm for the set partitioning problem. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 107–121. Springer.

[Avramidis et al., 2012] Avramidis, E., Burchardt, A., Federmann, C., Popovic, M., Tscherwinka, C., and Vilar, D. (2012). Involving language professionals in the evaluation of machine translation. In *LREC*, pages 1127–1130.

[Bailey, 2003] Bailey, S. (2003). *Academic Writing: A Handbook for International Students*. Taylor & Francis.

[Bramantoro and Ishida, 2009] Bramantoro, A. and Ishida, T. (2009). User-centered qos in combining web services for interactive domain. In *2009 Fifth International Conference on Semantics, Knowledge and Grid*, pages 41–48. IEEE.

[Cambridge Assessment, 2019] Cambridge Assessment (2019). International language standards. Available at http://www.cambridgeenglish.org/exams-and-tests/cefr/_.

[Cho and Ishida, 2011] Cho, H. and Ishida, T. (2011). Exploring cultural differences in pictogram interpretations. In *The Language Grid*, pages 133–148. Springer.

[Cho et al., 2007] Cho, H., Ishida, T., Yamashita, N., Inaba, R., Mori, Y., and Koda, T. (2007). Culturally-situated pictogram retrieval. In *International Workshop on Intercultural Collaboration*, pages 221–235. Springer.

[Crystal, 2012] Crystal, D. (2012). *English as a global language*. Cambridge university press.

[Del Gaudio et al., 2016] Del Gaudio, R., Burchardt, A., and Branco, A. (2016). Evaluating machine translation in a usage scenario. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1–8.

[Deutscher, 2010] Deutscher, G. (2010). *Through the language glass: Why the world looks different in other languages*. Metropolitan Books.

[Drewes, 2009] Drewes, W. (2009). Method for increasing the accuracy of statistical machine translation (smt). US Patent App. 12/321,436.

[ETS, 2019] ETS (2019). Compare toefl scores. Available at https://www.ets.org/toefl/institutions/scores/compare/.

[Fellbaum, 2010] Fellbaum, C. (2010). Wordnet. In *Theory and applications of ontology: computer applications*, pages 231–243. Springer.

[Gao et al., 2015] Gao, G., Yamashita, N., Hautasaari, A. M., and Fussell, S. R. (2015). Improving multilingual collaboration by displaying how non-native speakers use automated transcripts and bilingual dictionaries. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3463–3472. ACM.

[Geertz, 1973] Geertz, C. (1973). *The interpretation of cultures*, volume 5019. Basic books.

[Gookin, 2013] Gookin, D. (2013). *Word 2013 For Dummies*. –For dummies. Wiley.

[Greenstadt et al., 2006] Greenstadt, R., Pearce, J. P., and Tambe, M. (2006). Analysis of privacy loss in distributed constraint optimization. In *AAAI*, volume 6, pages 647–653.

[Hammersley and Atkinson, 2007] Hammersley, M. and Atkinson, P. (2007). *Ethnography: Principles in practice*. Routledge.

[Hara and Iqbal, 2015] Hara, K. and Iqbal, S. T. (2015). Effect of machine translation in interlingual conversation: Lessons from a formative study. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3473–3482. ACM.

[Henderson, 2005] Henderson, J. K. (2005). Language diversity in international management teams. *International Studies of Management & Organization*, 35(1):66–82.

[Herring, 2009] Herring, C. (2009). Does diversity pay?: Race, gender, and the business case for diversity. *American Sociological Review*, 74(2):208–224.

[Hida, 2016] Hida, S. (2016). Supporting multi-language communication in children's workshop. *Master thesis*.

[Hofstede, 1984] Hofstede, G. (1984). Cultural dimensions in management and planning. *Asia Pacific journal of management*, 1(2):81–99.

[Hornby and Cowie, 1995] Hornby, A. S. and Cowie, A. P. (1995). *Oxford advanced learner's dictionary*, volume 1430. Oxford university press Oxford.

[Isahara et al., 2008] Isahara, H., Bond, F., Uchimoto, K., Utiyama, M., and Kanzaki, K. (2008). Development of the japanese wordnet. In *Sixth InternationalConference on Language Resources and Evaluation*.

[Ishida, 2006] Ishida, T. (2006). Language grid: An infrastructure for intercultural collaboration. In *International Symposium on Applications and the Internet (SAINT'06)*, pages 5–pp. IEEE.

[Ishida, 2011] Ishida, T. (2011). *The language grid: Service-oriented collective intelligence for language resource interoperability*. Springer Science & Business Media.

[Ishida, 2016] Ishida, T. (2016). Intercultural collaboration and support systems: a brief history. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 3–19. Springer.

[Ishida et al., 2018] Ishida, T., Murakami, Y., Lin, D., Nakaguchi, T., and Otani, M. (2018). Language service infrastructure on the web: the language grid. *Computer*, 51(6):72–81.

[Ishii et al., 2017] Ishii, R., Kumano, S., and Otsuka, K. (2017). Prediction of next-utterance timing using head movement in multi-party meetings. In *Proceedings of the 5th International Conference on Human Agent Interaction*, pages 181–187. ACM.

[Jain and Bhandare, 2011] Jain, Y. K. and Bhandare, S. K. (2011). Min max normalization based data perturbation method for privacy protection. *International Journal of Computer & Communication Technology*, 2(8):45–50.

[Kimura et al., 2017] Kimura, S., Huang, H.-H., Zhang, Q., Okada, S., Ohta, N., and Kuwabara, K. (2017). Proposal of a model to determine the attention target for an agent in group discussion with non-verbal features. In *Proceedings of the 5th International Conference on Human Agent Interaction*, pages 195–202. ACM.

[King and Redpath, 2001] King, D. D. and Redpath, R. J. (2001). Method and system for improving machine translation accuracy using translation memory. US Patent 6,278,969.

[Kumar, 2012] Kumar, V. (2012). *101 Design Methods: A Structured Approach for Driving Innovation in Your Organization*. Wiley.

[Lafferty et al., 1974] Lafferty, J., Eady, P., Pond, A., and Synergistics, H. (1974). *The Desert Survival Problem: A Group Decision Making Experience for Examining and Increasing Individual and Team Effectiveness: Manual*. Experimental Learning Methods.

[Linguistic Data Consortium, 2005] Linguistic Data Consortium (2005). Linguistic data annotation specification: Assessment of fluency and adequacy in translations(technical report). Technical report, Linguistic Data Consortium.

[Margaret, 1994] Margaret, T. (1994). Establishing mutual understanding in systems design: An empirical study. *Journal of Management Information Systems*, 10(4):159–182.

[Matsui et al., 2012] Matsui, T., Silaghi, M., Hirayama, K., Yokoo, M., and Matsuo, H. (2012). Distributed search method with bounded cost vectors on multiple objective dcops. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 137–152. Springer.

[Mattioli et al., 2018] Mattioli, R., Ferraris, S. D., Ferraro, V., et al. (2018). Mybias: A web-based tool to overcome communication issues and foster creativity in heterogeneous design teams. In *DS 93: Proceedings of the 20th International Conference on Engineering and Product Design Education (E&PDE 2018), Dyson School of Engineering, Imperial College, London. 6th-7th September 2018*, pages 271–276.

[Maulsby et al., 1993] Maulsby, D., Greenberg, S., and Mander, R. (1993). Prototyping an intelligent agent through wizard of oz. In

Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems, pages 277–284. ACM.

[Mikolov et al., 2013] Mikolov, T., Le, Q. V., and Sutskever, I. (2013). Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*.

[Miller, 1998] Miller, G. (1998). *WordNet: An electronic lexical database*. MIT press.

[Morita and Ishida, 2009] Morita, D. and Ishida, T. (2009). Collaborative translation by monolinguals with machine translators. In *Proceedings of the 14th international conference on Intelligent user interfaces*, pages 361–366. ACM.

[Nasa, 2019] Nasa (2019). Survival! exploration then and now. Available at https://www.nasa.gov/pdf/166504main_Survival.pdf.

[Pituxcoosuvarn et al., 2018] Pituxcoosuvarn, M., Ishida, T., Yamashita, N., Takasaki, T., and Mori, Y. (2018). Machine translation usage in a children's workshop. In *International Conference on Collaboration Technologies*, pages 59–73. Springer.

[Popa et al., 2011] Popa, R. A., Redfield, C., Zeldovich, N., and Balakrishnan, H. (2011). Cryptdb: protecting confidentiality with encrypted query processing. In *Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles*, pages 85–100. ACM.

[Project ARISE Fermilab, 2019] Project ARISE Fermilab (2019). Instructional materials guides. Available at http://ed.fnal.gov/arise/guide.html.

[Scarton and Specia, 2016] Scarton, C. and Specia, L. (2016). A reading comprehension corpus for machine translation evaluation. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3652–3658.

[Shi et al., 2013] Shi, C., Lin, D., and Ishida, T. (2013). Agent metaphor for machine translation mediated communication. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 67–74. ACM.

[Shigenobu, 2007] Shigenobu, T. (2007). Evaluation and usability of back translation for intercultural communication. In *International Conference on Usability and Internationalization*, pages 259–265. Springer.

[Slevin, 2007] Slevin, J. (2007). Internet. *The Blackwell encyclopedia of sociology*.

[Sweeney, 2002] Sweeney, L. (2002). Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):571–588.

[Talke et al., 2010] Talke, K., Salomo, S., and Rost, K. (2010). How top management team diversity affects innovativeness and performance via the strategic choice to focus on innovation fields. *Research Policy*, 39(7):907–918.

[Toyama et al., 2014] Toyama, T., Sonntag, D., Dengel, A., Matsuda, T., Iwamura, M., and Kise, K. (2014). A mixed reality head-mounted

text translation system using eye gaze input. In *Proceedings of the 19th international conference on Intelligent User Interfaces*, pages 329–334. ACM.

[US Department of Education, 2018] US Department of Education (2018). Family educational right and privacy act.

[Vartiainen et al., 2015] Vartiainen, E., Domova, V., and Englund, M. (2015). Expert on wheels: an approach to remote collaboration. In *Proceedings of the 3rd International Conference on Human-Agent Interaction*, pages 49–54. ACM.

[Wang and Kissel, 2015] Wang, J. and Kissel, Z. A. (2015). *Introduction to network security: theory and practice*. John Wiley & Sons.

[Wang et al., 2007] Wang, W., Knight, K., and Marcu, D. (2007). Binarizing syntax trees to improve syntax-based machine translation accuracy. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 746–754.

[Willis and Trondman, 2000] Willis, P. and Trondman, M. (2000). Manifesto for ethnography. *Ethnography*, 1(1):5–16.

[Wooldridge et al., 2000] Wooldridge, M., Jennings, N. R., and Kinny, D. (2000). The gaia methodology for agent-oriented analysis and design. *Autonomous Agents and multi-agent systems*, 3(3):285–312.

[Yamashita et al., 2013] Yamashita, N., Echenique, A., Ishida, T., and Hautasaari, A. (2013). Lost in transmittance: how transmission lag enhances and deteriorates multilingual collaboration. In *Proceedings*

of the 2013 conference on Computer supported cooperative work, pages 923–934. ACM.

[Yamashita et al., 2009] Yamashita, N., Inaba, R., Kuzuoka, H., and Ishida, T. (2009). Difficulties in establishing common ground in multiparty groups using machine translation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 679–688. ACM.

[Yamashita and Ishida, 2006] Yamashita, N. and Ishida, T. (2006). Effects of machine translation on collaborative work. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 515–524. ACM.

[Yi et al., 2015] Yi, T., Shi, M., and Hong, Z. (2015). Privacy protection method for test score publishing. In *2015 7th International Conference on Information Technology in Medicine and Education (ITME)*, pages 516–520. IEEE.

[Yokoo et al., 2002] Yokoo, M., Suzuki, K., and Hirayama, K. (2002). Secure distributed constraint satisfaction: Reaching agreement without revealing private information. In *International Conference on Principles and Practice of Constraint Programming*, pages 387–401. Springer.

[Yoshino and Hayashi, 2002] Yoshino, R. and Hayashi, C. (2002). An overview of cultural link analysis of national character. *Behaviormetrika*, 29(2):125–141.

[Yoshino et al., 2015] Yoshino, T., Miyabe, M., and Suwa, T. (2015).

A proposed cultural difference detection method using data from japanese and chinese wikipedia. In *2015 International Conference on Culture and Computing (Culture Computing)*, pages 159–166. IEEE.