Difficulties in Establishing Common Ground in Multiparty Groups using Machine Translation

Naomi Yamashita[1], Rieko Inaba[2], Hideaki Kuzuoka[3], Toru Ishida[2,4]

[1] NTT Communication Science Labs.
Kyoto, Japan
naomi@cslab.kecl.ntt.co.jp

[2] National Institute of Information and Communications Technology
Kyoto, Japan
rieko.inaba@nict.go.jp

ABSTRACT

When people communicate in their native languages using machine translation, they face various problems in constructing common ground. This study investigates the difficulties of constructing common ground when multiparty groups (consisting of more than two language communities) communicate using machine translation. We compose triads whose members come from three different language communities—China, Korea, and Japan—and compare their referential communication under two conditions: in their shared second language (English) and in their native languages using machine translation. Consequently, our study suggests the importance of not only grounding between speaker and addressee but also grounding between addressees in constructing effective machine-translation-mediated communication. Furthermore, to successfully build common ground between addressees, it seems important for them to be able to monitor what is going on between a speaker and other addressees.

Author Keywords

Machine translation, Referential communication, Grounding, Computer-mediated communication.

ACM Classification Keywords

H.5.3 [**Group and Organization Interfaces**]: Computer-supported cooperative work, Synchronous interaction.

INTRODUCTION

Although communication technology has increased collaboration across international borders, language remains the biggest barrier to intercultural collaboration. In fact, most people have difficulty thinking and communicating in their non-native languages [20, 1].

For such people, machine translation appears to be an attractive technology, since it allows them to speak (write) and listen (read) in their native language. Indeed, an increasing number of multilingual organizations and Internet communities are proposing machine translation for communication support [8, 13]. One project that provides various language supports for such organizations is the "Language Grid Project [13]", which also served as a basis of this study.

Although machine translation liberates people from language barriers, it also poses hurdles to establishing mutual understanding. As one might expect, translation errors are the main source of inaccuracies that complicate mutual understanding [18]. In addition to translation errors, people have trouble constructing mutual understanding because they are not aware how each message is translated into other languages [19]. Furthermore, pairs have trouble grounding references because echoing and shortening of referring expressions are disrupted by asymmetries and inconsistencies in machine translation [22].

Although some novel solutions have been proposed [19, 13], machine translation still imposes excessive burdens on establishing mutual understanding. As a preliminary investigation, we interviewed members of an NPO [17] that has been using a machine-translation-embedded chat system to manage its overseas offices for almost two years. From these interviews, we found that they were facing particular difficulties when conducting multiparty group meetings. All of the interviewees mentioned that it was virtually impossible to conduct a group meeting when the total number of languages within the group was larger than two. For example, it seemed that members were easily left behind in the conversations of such meetings.

This study, inspired by these interviews, aims to clarify the reasons why machine-translation-mediated conversation is so difficult when the number of group members is larger than two. Research has demonstrated the difficulties of grounding references between pairs using machine translation [22]. Building on this previous work by expanding the experiment on referential communication from pairs to triads, we consider ways of supporting machine-translation-mediated collaboration for group work.

CHI 2009, April 3–9, 2009, Boston, MA, USA.

[3] University of Tsukuba, kuzuoka@iit.tsukuba.ac.jp
[4] Kyoto University, ishida@i.kyoto-u.ac.jp

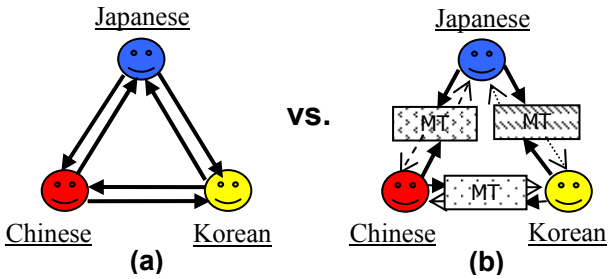


Figure 1 Three members communicating: (a) in their shared second language (English) or (b) in their native languages using machine translation software.

In the remainder of this paper, we first draw on prior research and predict how machine translation might affect referential communication within triads. Next, we describe a study that compares referential communication within triads in English (their shared second language) (Figure 1(a)) and referential communication within triads in their native languages using a machine-translation-embedded chat system (Figure 1(b)). We conclude with a discussion and issues raised by our study.

DIFFICULTIES IN ESTABLISHING COMMON GROUND IN MACHINE-TRANSLATION-MEDIATED COMMUNICATION

Common Ground

Regular Communication

Establishing *common ground* [4, 7, 6]—mutual knowledge, beliefs, assumptions, etc.—is important because communication is more efficient when participants share a greater amount of common ground [4, 9]. According to Clark and Marshall [6], people construct their common ground based on information they share by belonging to the same community, a shared physical setting (i.e., *physical co-presence*) or shared conversational content (i.e., *linguistic co-presence*). In each case, to successfully establish common ground, people not only must share the same information but also be aware that they are sharing this information with others [4, 15].

Grounding [4], then, refers to a process by which "common ground is updated in an orderly way, by each participant trying to establish that the others have understood their utterances well enough for the current purpose." During the grounding process, people become aware of what others do and do not know [5]. Such information helps them to formulate appropriate utterances, which leads to effective communication [5, 12].

In sum, for communicators to efficiently ground their utterances (particularly when members do not share the same physical space), the following three conditions must hold:

(1) they must share the same conversational content with others [4, 15]; (2) they must be aware that they are sharing the conversational content with others [4, 15]; and (3) they must be able to distinguish between information they do and do not share with others [5, 12].

Machine-Translation-Mediated Communication

It is important to satisfy the above three conditions in constructing common ground [4], but these conditions are not satisfied in machine-translation-mediated communication: As for condition (1), members cannot share the same conversational content because machine translation often mistranslates some parts of their utterances. As for condition (2), members cannot be aware whether they have the same conversational content, since they have no idea whether machine translation translated each utterance correctly into every language. Finally, as for condition (3), members cannot assess which parts of the utterance others do or do not understand because they have no idea where translation errors exist in other languages.

To improve machine-translation-mediated communication, researchers have proposed a novel solution called *back translation* [19]. Back translation offers speakers the awareness of how their utterances are translated into other languages by retranslating the translated utterances back to the speaker's language. Studies have demonstrated that the technique improves translation quality in machine-translation-mediated communication [19].

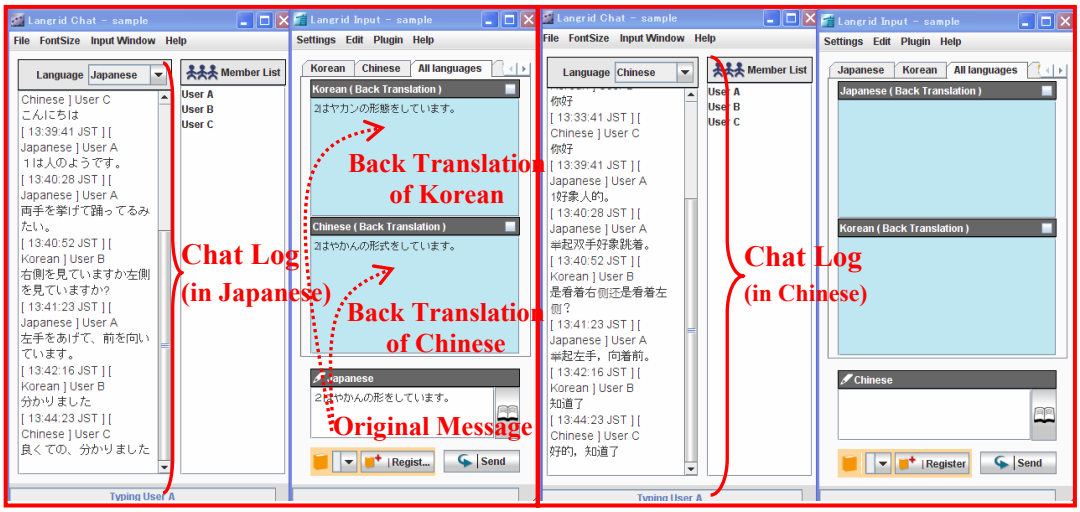
Despite this breakthrough, some problems remain unresolved in multiparty machine-translation-mediated communication. Even with the use of back translation, an addressee in a three-way machine-translation-mediated communication cannot monitor how the speaker's utterance is translated to the other addressee. For example, speaker A's message is translated into B's and C's languages simultaneously and back translations from both languages are shown to A. However, B (C) cannot monitor the translation between A and C (B). Consequently, conditions (2) and (3) do not hold between the two addressees: As for condition (2), the two addressees (B and C) cannot be aware whether they share the same information (i.e.. A's utterance); as for condition (3), addressee B (C) cannot be aware what addressee C (B) did and did not understand of A's utterance.

Since conditions (2) and (3), which are important in establishing common ground, do not hold in three-way machine-translation-mediated communication, it would clearly be difficult to build common ground, even with the use of back translation.

Referential Communication

Regular Communication

One type of communication that has been extensively studied to examine people's grounding process is "referential communication [7, 10, 14]." In referential communication, speakers and addressees work together to build common ground on a referent by adopting the same perspective [7]. Once speakers and addressees have enough evidence to believe that they are talking about the same thing, mapping is grounded between the referent and the perspective [3].



Japanese Interface **Chinese Interface**

Figure 2 Langrid Chat Interface (Japanese Director and Chinese Matcher)

The most basic task for examining referential communication is called the "referential communication task." Research applying this task typically studies how pairs arrange an identical set of figures into matching orders [7, 10, 14]. In each trial, one partner (the Director) is given a set of figures in a predetermined order. The other partner (the Matcher) is given the same figures in a random order. The Director must explain to the Matcher how to arrange the figures in the predetermined order. Typically, this matching task is repeated for several trials, each using the same figures but in different orders.

The process of agreeing on a perspective on a referent is known as *lexical entrainment* [3, 11]. Studies have shown that people make references based on historical factors such as recency, frequency of past references, and partner-specific conceptualization of the referent [2]. Studies have also shown that once communicators have entrained on a particular referring expression for a referent, they tend to abbreviate this expression in subsequent trials [2, 14].

Machine-Translation-Mediated Communication

Research on machine-translation-mediated communication has also studied referential communication between members of pairs. Yamashita [22] compared referential communication within pairs in English (their shared second language) and that within pairs in their native languages using machine translation software. Their results showed that lexical entrainment was disrupted in machine-translation-mediated communication because echoing was disrupted by asymmetries in machine translations. In addition, the process of shortening referring expressions was also disrupted because the translations did not produce the same terms consistently throughout the conversation.

Back translation can be used to alleviate the asymmetry issues because it offers speakers the awareness whether their utterances are symmetrically translated; when back translation does not yield the original expression, it implies that they cannot share the expression with others. While back translation may help communication within pairs, it is still unclear whether it improves communication within triads. Indeed, the NPO we interviewed had been using a machine-translation-embedded chat system with a back translation function, and they managed to conduct communication within language pairs; however, they said this was not possible within language triads.

As mentioned, we assume that problems peculiar to multiparty group communication arise when participants try to build common ground using machine translation; establishing common ground among multiple addressees would be difficult because addressees cannot monitor how the speaker's utterance is translated to the other addressees. To examine how this issue actually leads to real problems in the grounding process, we conducted an experiment using a machine-translation-embedded chat system with a back-translation function.

CURRENT STUDY

The present study builds on Yamashita's research [22] by expanding the experiment of referential communication from pairs to triads. We attempt to reveal how machine translation complicates referential communication within triads by comparing such communication in English (members' shared second language) and that in their native languages through machine translation software (Figure 1).

In the present task, three participants from three different language communities—China, Korea, and Japan—work together in a referential communication task in English or in their native languages. In the task, they must arrange an identical set of tangram figures into matching orders. In each trial, one participant (Director) is given a set of figures in a predetermined order, and the other two participants (Matchers) are given the same figures in different random orders. Using a multilingual chat system embedded with a back-translation function, the Director must explain to the Matchers how to arrange the figures in the predetermined order. Rotating the role of Director for each trial, this matching task is repeated for six trials (i.e., two cycles) using the same figures but in different orders.

Multilingual Chat System: Langrid Chat

For the experiment, we used a machine-translation-

embedded chat system called "Langrid Chat [16]" (Figure 2). Langrid Chat translates each message into other languages while providing awareness information on the typing of other users. The machine-translation software embedded in Langrid Chat is a commercially available product that is rated as one of the very best translation programs on the market, in terms of translation quality. Langrid Chat is also equipped with a back-translation function: when a user types a sentence into the typing area, the system automatically translates the sentence into other languages, retranslates them back to the original language, and shows them to the user (Figure 2 (left)). Back translation is provided in real time so that users can edit their messages before sending them to others.

The chat interface allows each user to select his/her browsing and typing language from Chinese, English, Korean, and Japanese. For example, a Japanese participant who selects Japanese for his browsing and typing language can read and write in Japanese. Similarly, when a triad selects English as their browsing and typing language, they can both read and write in English.[1]

Hypotheses

We use quantitative and qualitative data analyses to examine three hypotheses:

In three-way machine-translation-mediated communication, machine translation translates each message into two other languages. Since translation from language A to B and translation from language A to C are carried out independently of each other, the original utterance in language A is often translated differently in language B than in C. In such conversations, two Matchers will not be able to share the same Director's utterance (i.e. condition (1) does not hold). Furthermore, they will not be aware whether they share the same Director's utterance (i.e., condition (2) does not hold). Under such conditions, we assume that participants will have trouble in identifying referents, leading them to low efficiency in their mutual acceptance process:

H1 (Efficiency of Mutual Acceptance Process): Participants will more efficiently identify a referent when using English rather than machine translation.

In the second cycle, each participant becomes the Director once again. When comparing referring expressions of the same participant between the first and second cycles, we expect that referring expressions will be shorter in the second cycle when using English because people often abbreviate referring expressions over time [2, 14]. However, we expect that abbreviation of referring expressions is at times very difficult when using machine translation for the following reason: Even when a Director A's referring expression is translated correctly to both Matchers (B and C), this does not ensure that the same referring expression will be correctly translated between B and C (i.e., condition (2) does not hold between the three participants); when B (or C) becomes the next Director, he or she might realize that the referring expression does not work between B and C, and thus change the referring expression to something else or add some details so that C (or B) understands it. Such changes in referring expression may complicate their mutual acceptance process, making it difficult to abbreviate their referring expressions:

H2 (Abbreviation of Referring Expressions over Trials): Participants will abbreviate their referring expressions more when using English than when using machine translation.

Not only is abbreviation difficult, but we also expect that making an appropriate reference (that would be smoothly identified by the Matchers) is also difficult when participants rotate their Director roles. When participants rotate their Director roles, the new Director (previous Matcher) typically explains each referent based on what he believes he shares with others [4]. However, in machine-translation-mediated communication, participants are less able to distinguish between information that they do and do not share with others (i.e., condition (3) does not hold). Therefore, we expect that the new Director will not be able to formulate appropriate references that would be smoothly identified by the Matchers:

H3 (Improvements in Making Appropriate References): Participants are less able to improve their efficiency of formulating appropriate references when using machine translation than when using English.

METHOD

Design

Thirteen triads (total of thirty-nine participants) from different language communities—China, Korea, and Japan—participated in the experiment. Nine triads participated in a referential communication task using their native languages through machine translation; four triads participated in the same referential communication task using a common language (English, which is not their native language). The experimental design was a between-subjects design for comparing referential communications carried out using the above two language methods.

Participants

Participants consisted of thirteen Chinese, thirteen Korean, and thirteen Japanese living in Japan. None of the participants knew each other before the experiment. Their English proficiency levels varied, but all of the participants had studied English for more than six years, and they were able to read and write basic English. They frequently used e-mail and instant messaging, but only a couple of them had used machine translation before the experiment. Participants were paid for their participation.

[1] Since machine translation automatically translates all messages, there is no difference in delay between conversation in English and using native languages.

Procedure

Step(1): On arrival, participants were taken to a room and asked to complete experimental consent forms. Next, participants were taken to a room partitioned into three compartments with a computer in each, and asked to sit in front of one of the computers. Participants were then given explanations of how to use Langrid Chat and an overview of the experiment. Participants were told that a) each person has the same set of figures in different orders; b) there are three roles: one Director and two Matchers; c) the Director must explain each figure one by one until both Matchers arrange their figures in the Director's order; d) the matching task is repeated six times using the same figures but in different orders, and each time the role of Director is rotated.

Step(2): As a pre-study, the participants engaged in a short-term referential communication task using three tangram figures (different from those used in Step(3)). The pre-study was conducted to let participants familiarize themselves with Langrid Chat.

Step(3): Triads were presented with eight tangram figures (Figure 3) arranged in different sequences, and they were instructed to match the arrangements of figures using Langrid Chat.

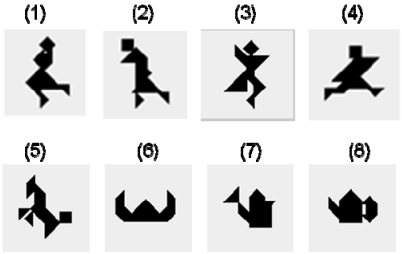


Figure 3. Eight tangram figures used in the experiment.

Rotating the role of Director for each trial, this matching task was repeated for six trials (i.e., two cycles) using the same figures but in different orders.

Step(4): Following the four matching tasks, participants were interviewed, as described below.

Please note that the experimental design was incomplete in that Director role was not counterbalanced for order; Japanese participants played the Director role for the first and fourth trial, Korean participants in the second and fifth trial, Chinese participants in the third and sixth trial.

Measures

Efficiency of Referential Communication. The triads were instructed to complete the task as efficiently as possible. We used the number of utterances (messages) per figure made by Directors to measure the efficiency of referential communication.

Abbreviation of Referring Expressions. We compared the length of referring expressions of the same Director between the first and second cycles and calculated the frequency of the Directors abbreviating their referring expressions. We did not compare the length of referring expressions between different Directors because the number of words differs among different languages even when they use the same expressions.

Improvements in Making Appropriate References. When Directors make appropriate references based on prior mutually accepted descriptions, Matchers should be able to identify the referents through the "*basic exchange* [7]" more frequently, where basic exchange is the most efficient way to identify a referent consisting of two steps: (a) the presentation of a referring expression and (b) its acceptance. To measure the appropriateness of each Director's reference, we calculated the proportion of basic exchange.

Interview. At the end of the experiment, we interviewed each participant separately using Japanese or English. When the participants had trouble understanding or speaking, bilingual translators translated our questions. There were no predetermined questions, but the topics covered the usefulness of the multilingual chat system (Langrid Chat), the ease of constructing and understanding utterances, and the strategies they used for effectively completing the task. The interview also helped to explain some specific incidents observed during the task.

RESULTS

Three groups were excluded from quantitative analysis since the members ran out of time and could not repeat the tasks for six trials using machine translation.

Efficiency of Referential Communication

Number of Utterances

Our first hypothesis *H1* stated that participants would more efficiently identify a referent when using English rather than machine translation. To test this hypothesis, the numbers of Director's utterances per figure were analyzed in a repeated measures ANOVA with Language Condition as a between-subjects factor[2]. Results indicated a significant main effect for Trial ($F[5, 40]=8.95$, $p<.001$) and a significant main effect for Language Condition ($F[1,8]=15.68$, $p=.001$) but no interactions.

As shown in Figure 4, the number of Director's utterances decreased over trials for both Language Conditions. As predicted by *H1*, however, it was proved that the Machine Translation condition yielded more utterances of a director compared to the English condition.

In forming our first hypothesis, we anticipated that participants would have trouble identifying referents through machine-translation-mediated communication due to the following two factors:

[2] Where ANOVA is carried out, the test for homogeneity of variance (Levene test) was also carried out. Unless reported, variances were equal between conditions ($p>.05$).

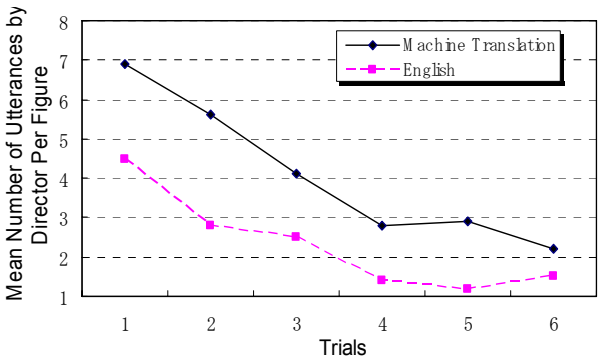


Figure 4. Mean number of utterances by a Director per figure.

- Two Matchers B and C will not be able to share the same Director A's utterance (i.e., condition (1) does not hold) because of the discrepancy in translation between A to B and A to C.
- Two Matchers B and C will not be aware of whether they share the same utterance of Director A (i.e., condition (2) does not hold).

To see how these factors actually affected referential communication, we examined the conversations in our experiment in further detail. In the following, we examine the impact of these factors one by one.

Places of Identifying Referents

When two Matchers do not share the same utterance of a Director (i.e., when condition (1) does not hold), Matchers may not be able to identify the referents based on the same Director's utterances. As expected, we found many cases in which Matchers identified the referents at different places in the conversation; specifically, one Matcher required more information and/or clarification than the other when using machine translation (Excerpt 1).

Excerpt 1. Matchers accepting Director's Proposal at Different Points of the Conversation (translated into English). Underline&Boldface indicates the originator of each message.

	Japanese Screen	Korean Screen	Chinese Screen
<3rd trial> Director: Chinese			
1	C: A head is a square one.	C: The head is square.	**C: Its head is square.**
2	C: The edge run toward the right.	C: The vicinity is attached to the right.	**C: It runs toward its right.**
3	K: Is it the design to which you run?	**K: Does it looks like running?**	K: Is it after we assume that I compare and run?
4	**J: I got it.**	J: I got it.	J: I got it.
5	C: A lower back is the parallelogram.	C: A lower back is the parallelogram.	**C: The lower back is the parallelogram.**
6	K: I got it.	**K: I got it.**	K: I got it.

To understand what the participants were trying to communicate, we translated all messages into English. In addition, to share the automatically translated messages in this paper, we further translated the translated messages into English.

In the excerpt above, a Japanese Matcher and a Korean Matcher identified one of the Tangram figures based on a Chinese Director's explanation. In this trial, the Japanese Matcher identifies the figure in the 4th line, while the Korean Matcher identifies it in the 6th line. Although this was their third time to match the same figures, the Korean Matcher was late in identifying the figure, presumably because the Chinese Director's 2nd utterance made no sense to the Korean Matcher.

To see whether such a case (i.e., Matchers identifying a referent at different places in the conversation) occurred more frequently in machine-translation-mediated communication than in English, we counted the number of such cases for each trial and then performed a repeated measure ANOVA on those numbers.

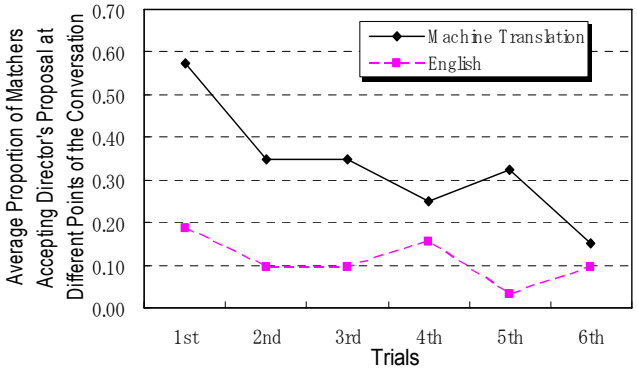


Figure 5. Average proportion of Matchers identifying a figure at different points in the conversation.

As shown in Figure 5, Matchers identified the referents at different points in the conversation more frequently in machine-translation-mediated communication than in English ($F[1,8]=15.99$, $p<.01$). We also found a significant main effect for Trial ($F[5, 40]=3.44$, $p<.05$) but no interactions.

Although back translation offered Directors the awareness of how their messages were translated into the other languages, it appeared from the interviews that rewriting their messages until the back translations of the two different languages reflected the meaning of the original message was difficult and time consuming. As a result, a Director's utterance was often translated differently to the two Matchers, leading them to identify the figures at different points in the conversations (i.e., based on different information). We speculate that such a tendency will increase as the number of languages increases in multiparty machine-translation-mediated communication.

Adaptation of References toward Others

From further observation, we found that referential communication using machine translation was even more inefficient because Matchers were not aware whether they shared the same Director's utterance (i.e., condition (2) did not hold).

Excerpt 2. Director not being able to coordinate his utterance toward the slow Matcher (translated into English). Underline&Boldface indicates the originator of each message.

	Japanese Screen	Korean Screen	Chinese Screen
<2nd trial> Director: Korean			
1	K: Looks like a pitcher.	**K: The shape of a pitcher.**	K: It's a financial aid person electron, an arm is done.
2	C: Sorry, not well understood.	C: Sorry. Not well understood.	**C: Sorry, I don't understand.**
3	K: The third one is swept when watering flowers.	**K: The third one is used when watering flowers.**	K: When giving water to a flower, the third is used.
4	**J: A sprinkler?**	J: A sprinkler?	J: Is this a sprinkler?
5	K: Yes.	**K: Yes.**	K: Yes.
6	C: The mouth was big.	C: The mouth became big.	**C: Its spout is big.**
7	K: The mouth is big.	**K: The mouth is big.**	K: The mouth is big.
8	**J: Is the mouth triangle?**	J: Is the mouth triangle?	J: Is the mouth triangle?
9	C: Got it, no problem.	C: Got it. No problem.	**C: Got it. No problem.**
10	K: Do you understand?	**K: Do you understand?**	K: Do you understand?
11	K: OK.	**K: OK.**	K: OK.
12	K: The mouth is triangle.	**K: The mouth is triangle.**	K: Mouth is triangle.
13	**J: I got it!**	J: I got it!	J: I got it!
<3rd trial> Director: Chinese			
14	C: A sprinkler.	C: A sprinkler.	**C: A sprinkler.**
15	C: Water was given and it was consumed.	C: Water was given and it was consumed.	**C: We use it for watering flowers.**
16	K: I got it.	**K: I got it.**	K: I got it.
17	C: The mouth is big.	C: The mouth is big.	**C: The spout is big.**
18	K: Yes, yes.	**K: Sure, sure.**	K: Nene.
19	K: It has a right triangle mouth, right?	**K: It has a right triangle mouth.**	K: You had a mouth of a right triangle, right?
20	**J: Sorry,**	J: Sorry.	J: Sorry.
21	**J: I got it.**	J: I got it.	J: I got it.

Matcher C, B often acquires knowledge of why C did not accept A's proposal concurrently with him or her by following the subsequent conversation between A and C. B makes use of such knowledge to coordinate his or her own utterances on the referent upon becoming the next Director [5]. However, such coordination was rarely observed in referential communication using machine translation.

In Excerpt 2, for example, a Japanese Matcher and a Chinese Matcher identify one of the Tangram figures based on a Korean Director's explanation. In this (second) trial, the Chinese Matcher identifies the figure in the 9th line, but the Japanese Matcher cannot identify it at the same timing. He asks the Director a question regarding the shape of the pitcher's spout (whether it is triangular) and manages to identify the figure in the 13th line. Although it is typically the case that the next Director coordinates his utterance (i.e., indicating that the pitcher's spout is triangular) so that the previous slow Matcher (i.e., the Japanese Matcher) can easily identify the referent, the Chinese Director in the consecutive trial did not do so. The Japanese Matcher finally manages to identify the figure with the help of the Korean Matcher.

Interestingly, the Korean Director's utterances were translated similarly to both Matchers in the second trial (from line 2). It is likely that the Chinese and the Japanese Matcher shared similar information regarding the Korean Director's utterance. Thus, if the Chinese Director (in the third trial) had coordinated his utterance indicating that the pitcher's spout was triangular, the Japanese Matcher would have been able to identify the figure more smoothly. We infer that the Chinese participant did not do so because he did not know whether he shared the same information with the Japanese Matcher in the second trial; maybe he could not understand why the Japanese Matcher could not accept the Korean Director's proposal concurrently with him in the second trial (whether because of translation error or other reasons), and thus he did not know what strategy to take. Similar cases were found elsewhere.

To examine whether such cases occurred more frequently in machine-translation-mediated communication than in English, we first extracted the cases in which Matchers differed in their places of accepting the Director's proposal. Then, for each case, two independent coders classified whether the next Director coordinated their utterances toward the previous slow Matcher. Since the coders only understood Japanese and English, they classified the transcripts of which Korean and Chinese utterances were translated into Japanese by bilingual translators. Agreement between the two coders was high (Cohen's Kappa values of the transcripts using English and machine translation were 0.91 and 0.95, respectively). We then calculated the rate of Directors coordinating their utterances toward the previous slow Matcher for each triad.

Overall, Directors coordinated their utterances toward the previous slow Matcher more when using English (Avg: 78.8%) than machine translation (Avg: 48.8%). A T-test showed a significant difference between the two language conditions ($t(8)=2.63$, $p<.05$). Since the previous slow Matchers often required further explanation when Directors did not coordinate their utterances toward them, we infer that such a lack of coordination of utterances was one reason leading them to inefficient communication requiring a large number of utterances to match the figures.

Abbreviation of Referring Expressions

Studies using referential communication tasks have shown that once a pair of communicators has entrained on a particular referring expression for a referent, they tend to abbreviate this expression on subsequent trials [2, 14]. However, we predicted in *H2* that abbreviation of referring

expressions is difficult, particularly for triads using machine translation.

To examine *H2*, we compared the lengths of referring expressions of the same Director between the first and second cycles and classified for each referent whether the referring expression was (i) shortened (i.e., certain adjectives or/and explanations are eliminated), (ii) lengthened (i.e., certain adjectives or/and explanations are added), or (iii) other (identical or totally differentiated). For each participant, we calculated the rates of shortened and lengthened referring expressions.

Although the difference was not significant, participants shortened their referring expressions slightly more when using English (Avg: 45%) than machine translation (Avg: 31%) ($F[1,8]=3.98$, $p=.08$). As a more interesting finding, participants lengthened their referring expressions significantly more when using machine translation (Avg: 19%) than English (Avg: 6%) ($F[1,8]=5.21$, $p<.05$).

It seems that participants had trouble finding referring expressions that could be shared with all three members. Even in a case where a Director's reference was smoothly accepted by the Matchers in the first cycle, the Director sometimes lengthened his or her referent in the second cycle because the reference could not be used between the two Matchers (when one of the Matchers became the Director). The excerpt below captures this tendency.

In Excerpt 3, it appears that the Directors could not determine which terms to omit and which to leave (from 4th to 6th trial). We infer that Directors are reluctant to abbreviate their referring expressions once a new adjective or/and explanation is added during their mutual acceptance process, since they do not know which terms are translated correctly among all language pairs or why a new explanation has been added. To minimize their collaborative effort, it seems that they adopt a strategy of listing several references so that some parts of the list would be correctly translated in the translations of any language pair. We speculate that such difficulties in sharing the same reference will increase as the number of languages increases in multiparty machine-translation-mediated communication.

Excerpt 3. Directors not being able to abbreviate their referring expressions (conversation is translated into English). Underline&Boldface indicates the originator of each message.

Japanese Screen	Korean Screen	Chinese Screen
<1st trial> Director: Japanese		
J: Number 2 is a horse.	J: Number 2 is a horse.	C: Number 2 is a horse.
<2nd trial> Director: Korean		
K: Number 4 is	**K: Number 4 is a person standing upside down.**	K: 4 times
--- (snip) ---		
J: Mr. B. Which number is the animal?	J: Mr. B. Which number is the animal?	J: Mr. B. Which number is the animal?
K: Animal?	**K: Animal?**	K: Animal?
--- (snip) ---		
J: Which number is the creature with a square tail?	J: Which number is the creature by which a tail is a square?	J: A tail, what number is a square creature?
C: An animal will be 8 days.	C: An animal is 8 days.	**C: Animal is number 8.**
K: I wouldn't know what to say, but something like an animal is 4 times most.	**K: I don't know what you are saying but the most animal like thing is number 4.**	K: Something like whatever animal says, is it wasteful, an unclear one is 4 times most.
<3rd trial> Director: Chinese		
C: It seems to be an animal.	C: It seems to be an animal.	**C: It looks like an animal.**
C: Horse	C: Horse	**C: Horse**
<4th trial> Director: Japanese		
J: Horse. Animal.	J: Horse. Animal.	J: Horse. Animal.
J: Tail is square.	J: A tail is square.	J: A tail is square.
<5th trial> Director: Korean		
K: It's an animal	**K: It's an animal.**	K: It's an animal.
K: It seems to be a word which raised its foreleg.	**K: It's a shape of a horse raising its front legs.**	K: A word is the design which entered a foreleg.
<6th trial> Director: Chinese		
C: Animal, it seems to be a horse.	C: Animal, it seems to be a horse.	**C: Animal, seems to be a horse.**
C: There is a square on the right side.	C: There is a square on the right side.	**C: There is a square on the right side.**

Improvements in Making Appropriate References

We hypothesized in *H3* that participants are less able to improve their efficiency in formulating appropriate references when using machine translation than when using English because they are less able to distinguish between information that they do and do not share with others (i.e., condition (3) does not hold).

We have already seen much evidence that making appropriate references is difficult. For example, coordinating their utterances toward the previous slow Matcher was difficult; finding a reference that could be shared between all members was also difficult.

To see how much Directors improved in making appropriate references over trials, we calculated for each trial the rate of participants matching the figures through basic exchange (i.e., the most efficient way to match a figure: a Director proposing a reference and two Matchers accepting the reference immediately). Then, we performed a repeated measure ANOVA on those rates.

As shown in Figure 6, participants were able to match the figures more efficiently in English than in machine translation ($F[1,8])=61.43$, $p<.001$). We also found a significant main effect for Trial ($F[5, 40]=6.40$, $p<.01$) as well as a significant Language by Trial interaction ($F[5,40]=12.0$, $p<.001$). It appeared that Directors using machine translation had difficulty improving their references so that both Matchers could identify them immediately.

If Directors had used back translation more rigorously, the increasing rate of basic exchange could have been steeper. However, the problem does not lie only in the disinclination to use back translation. As previously mentioned, Directors were not aware which terms could be shared and which terms could not be shared with all of the members. Such unawareness impeded them from constructing appropriate references; even when they once

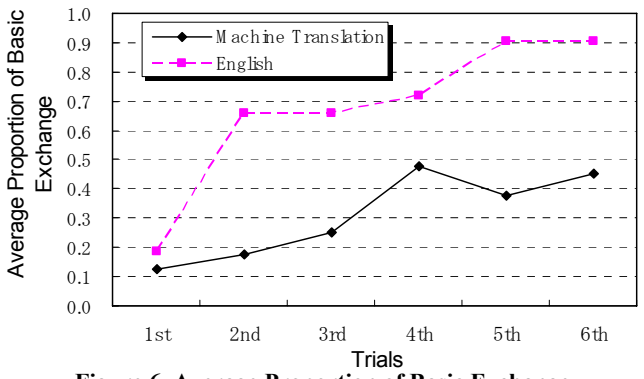


Figure 6. Average Proportion of Basic Exchange.

used a reference that could be shared among all of the members, they added redundant explanations when some problems occurred, and they were reluctant to shorten them because they were not aware which references could be shared among all members.

DISCUSSION

The goal of this study was to clarify why and how grounding conversations is difficult in machine translation-mediated multilingual triads.

Previous studies have documented the importance of satisfying the following conditions for communicators to successively build common ground: (1) they must share the same conversational content with others [4, 15]; (2) they must be aware that they are sharing the conversational content with others [4, 15]; (3) they must be able to distinguish between information they do and do not share with others [5, 12].

However, from our experiments, we found that satisfying these conditions was particularly difficult when the number of languages used in a group was larger than two. First, it appeared that condition (1) was often violated because of the discrepancy between translation from A to B and that from A to C. When condition (1) was violated, Matchers were not able to identify a referent at the same timing; one of the Matchers required more clarification for identifying the referent. Matchers tended to identify the referents based on different information. Furthermore, conditions (2) and (3) were often violated because participants using machine translation could not monitor how each utterance was translated into the other languages. Such a violation seemed to cause many problems in grounding references. In our experiment, we found three issues that seemed to arise from the violation of these conditions.

First, participants were not aware which parts of the conversational content they did and did not share with others. Under such a condition, we infer that Matchers had trouble understanding other Matchers' utterances (e.g., why a Matcher was asking for clarification) because they did not know the basis of their utterances. As a result, Directors were less likely to coordinate their utterances toward the previous slow Matcher. Second, participants were not aware which terms they could and could not share with all of the members. Under such a condition, it seemed that Directors could not determine which terms to omit and which terms to leave. As a result, Directors were less likely to abbreviate their referring expressions over trials. Finally, it appeared that participants using machine-translation-mediated communication had difficulty constructing appropriate (efficient) utterances because they could not distinguish between what they did and did not share with others. As a result, the participants' mutual acceptance process was inefficient and did not improve much compared to using English.

Although participants could always observe conversations between others through machine translation, it seemed that participants could not efficiently achieve mutual knowledge through indirect inferences. We speculate that one reason lies in the participants' behavior that they rarely provided back-channels or their status of understandings; when they had trouble understanding other participants' utterances, they ignored the utterance [22] or asked questions (instead of saying that they do not understand). This made them difficult to distinguish between shared and unshared information.

Theoretical Implications

Our study suggests the importance of not only grounding between speaker and addressee but also grounding between addressees in constructing effective machine-translation-mediated communication. When common ground is not well-established between addressees, communication is likely to become inefficient when they become a speaker. To successfully build common ground between addressees, it seems important for them to be able to monitor what is going on between a speaker and other addressees. By monitoring such conversation, they acquire knowledge of what others do and do not know. However, we speculate that being able to distinguish such knowledge is not sufficient for effective communication. When an addressee has trouble understanding a speaker's utterance, other addressees should be able to assess *why* the addressee fails to understand it by monitoring the conversation between speaker and the addressee (e.g., is it because of mistranslation or another reason?). When they are able to correctly assess the reason, they will be able to construct appropriate utterances that can be smoothly understood by others. We believe that knowledge of others (acquaintance relationships) and communicational context have a strong impact on participants' ability to assess such reasons.

Design Implications

Our findings and the above discussion suggest two recommendations for the design of future machine-translation-embedded communication systems to support group work.

- Provide speakers with an awareness of how their utterances are translated between addressees (i.e.,

whether the terms they are using can also be used between addressees).

- Provide addressees with an awareness of how a speaker's utterance is translated to other addressees using different languages (e.g., whether it is translated correctly or which part of the utterance is mistranslated).

One way of increasing mutual awareness among group members may be to share the video images of each participant's facial expressions. As shown in Veinott et al. study [21], video helps grounding between multilingual participants because it helps them assess other participants' level of understanding by providing their facial expressions.

For our future work, we are interested in investigating machine-translation-mediated communication which actually took place in the NPO that we have interviewed. In the long run, based on the findings from such investigations, we are hoping to contribute to the development of more effective machine-translation-mediated communication systems.

ACKNOWLEDGMENTS

This research was supported by the Kyoto University Global COE Program: Informatics Education and Research Center for Knowledge-Circulating Society. The authors would like to thank the Language Grid Project members, particularly Tomohiro Shigenobu for letting us use the multilingual chat system. We also thank the anonymous reviewers for their constructive comments.

REFERENCES

1. Aiken, M., Hwang, C., Paolillo, J., and Lu, L. A group decision support system for the Asian Pacific rim. *Journal of International Information Management*, 3, 1994, 1-13.
2. Brennan, S. E. Lexical Entrainment in Spontaneous Dialogue. *Proceedings of International Symposium on Spoken Dialogue*, 1996, 41-44.
3. Brennan, S. E. and Clark, H. H. Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 6, 1996, 1482-1493.
4. Clark, H. H. *Using Language*. Cambridge, UK: Cambridge University Press, 1996.
5. Clark, H. H. and Haviland, S. E. Comprehension and the Given-New contract. *Discourse Production and Comprehension*, 1977,1-40.
6. Clark, H. H. and Marshall, C. E. Definite reference and mutual knowledge. *Elements of discourse understanding*, 1981, 10-63.
7. Clark, H. H. and Wilkes-Gibbs, D. Referring as a collaborative process. *Cognition*, 22, 1986, 1-39.
8. Climent, S., More, J., Oliver, A., Salvatierra, M., Sanchez, I., Taule, M., and Vallmanya, L. Bilingual Newsgroups in Catalonia: A Challenge for Machine Translation. *Journal of Computer Mediated Communication*, 9, 1, 2003.
9. Fussell, S., Krauss, R. Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psych*, 62, 3, 1992, 378-391.
10. Fussell, S., Kraut, R., and Siegel, J. Coordination of Communication: Effects of Shared Visual Context on Collaborative Work. *Proceedings of CSCW*, 2000, 21-30.
11. Garrod, S. and Anderson, A. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27, 1987, 181-218.
12. Grice, H. P. Logic and conversation. *Syntax and Semantics, Vol. 3: Speech Acts*, Seminar Press, 1975, 113-127.
13. Ishida, T. Language Grid: An Infrastructure for Intercultural Collaboration. *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, keynote address, 2006, 96-100.
14. Krauss, R. M. and Glucksberg, S. The development of communication: Competence as a function of age. *Child Development*, 40, 1969, 255-256.
15. Krauss, R. P. and Fussell, S. Mutual knowledge and communicative effectiveness. Intellectual Teamwork: Social and Technological Foundations of Cooperative Work, 1990, 111-146.
16. Langrid Chat: http://langrid.nict.go.jp/en/chat.html
17. NPO Pangaea: http://www.pangaean.org/
18. Ogden, B., Warner, J., Jin, W. and Sorge, J. Information Sharing Across Languages Using MITRE's TRiM Instant Messaging. 2003.
19. Shigenobu, T. Evaluation and Usability of Back Translation for Intercultural Communication. *International Conference on Human-Computer Interaction (HCII-07)*, 10, 2007, 259-265.
20. Takano, Y. and Noda, A. A temporary decline of thinking ability during foreign language processing. *Journal of Cross-Cultural Psychology*, 24, 1993, 445-462.
21. Veinott, S., Olson, J, Olson, G. and Fu, X. Video helps remote work: speakers who need to negotiate common ground benefit from seeing each other. *Proceedings of CHI*, 1999.
22. Yamashita, N. and Ishida, T. Effects of Machine Translation on Collaborative Work. *Proceedings of CSCW*, 2006.