Effects of Machine Translation on Collaborative Work

Naomi Yamashita
Social Communication Lab
NTT Communication Science Labs
2-4 Hikaridai Seika-cho
Soraku-gun Kyoto
Japan
naomi@cslab.kecl.ntt.co.jp

Toru Ishida
Department of Social Informatics
Kyoto University
Yoshida-Honmachi
Sakyo-ku Kyoto
Japan
ishida@i.kyoto-u.ac.jp

ABSTRACT
Even though multilingual communities that use machine translation to overcome language barriers are increasing, we still lack a complete understanding of how machine translation affects communication. In this study, eight pairs from three different language communities–China, Korea, and Japan–worked on referential tasks in their shared second language (English) and in their native languages using a machine translation embedded chat system. Drawing upon prior research, we predicted differences in conversational efficiency and content, and in the shortening of referring expressions over trials. Quantitative results combined with interview data show that lexical entrainment was disrupted in machine translation-mediated communication because echoing is disrupted by asymmetries in machine translations. In addition, the process of shortening referring expressions is also disrupted because the translations do not translate the same terms consistently throughout the conversation. To support natural referring behavior in machine translation-mediated communication, we need to resolve asymmetries and inconsistencies caused by machine translations.

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

General Terms
Human Factors, Experimentation, Performance, Languages

Keywords
Multilingual Groups, Machine Translation, Distributed Work, Computer-Mediated Communication, Reference, Lexical Entrainment.

1. INTRODUCTION
As computer-mediated communication increases collaboration across broad distances, collaborations involving people speaking different languages are starting to play a significant role in our lives.

CSCW'06, November 4–8, 2006, Banff, Alberta, Canada.

In multilingual groups, in which members' native languages differ, communication typically takes place in one language, requiring some members to communicate in a non-native language. However, since members who must communicate in their non-native language frequently find communication difficult [27, 1, 17], such collaboration tends to be ineffective [2, 28].

Machine translation is a powerful tool for multilingual groups, because it allows all members to speak (write) and listen (read) in their native language. Indeed, we have already seen several multilingual Internet communities and multilingual projects in which each participant communicates in his or her native languages via machine translation.[1] One such project is the "Intercultural Collaboration Project [24]," which has been conducted annually since 2002 and is now integrated into the "Language Grid Project [15]". The number of such communities and projects are expected to grow in the future [7].

Although machine translation liberates members from language barriers, it also poses hurdles for establishing mutual understanding. As one might expect, translation errors are the main source of inaccuracies that complicate mutual understanding [25]. Climent found that typographical errors are also a big source of translation errors that hinder mutual understanding [7]. Yamashita discovered that members tend to misunderstand translated messages and proposed a method to automatically detect misunderstandings [30].

Despite these breakthroughs, we still lack a complete understanding of how machine translation affects communication. For example, how do speakers and addressees establish common ground when using machine translation? How do they make a reference and identify it without sharing identical referring expressions? Can an addressee smoothly identify a referent after the speaker's referring expressions have been shortened?

Answering such questions will help provide a foundation for designing machine translation-mediated collaboration for computer-supported collaborative work (CSCW). A great deal of research has examined the effects of communication technologies on collaboration among people sharing the same native language. To date, virtually no research examined machine translation-mediated communication.

In this paper, we explore the effects of machine translation on referential communication between speakers of different native

[1] http://bbs.enjoykorea.jp/tbbs/read.php?board_id=tlife&nid=1080185

languages. In particular, we compared participants' referential communication in English (their shared second language) versus referential communication in their native languages when using machine translation software.

2. BACKGROUND

2.1 References and Common Ground

According to Clark, reference is a collaborative process. Speakers and addressees work together to establish shared knowledge or common ground [5, 6, 14, 20]. One way they do so is by adopting the same perspective on a referent [14]. Once speakers and their partners have enough evidence to believe that they are talking about the same thing, mapping is grounded between the referent and the perspective [4].

2.2 Referential Communication

Many social psychological communication studies have employed what has come to be called a "referential communication task." This task allows us to examine the adequacy of communication. Referential communication tasks are not the only way to objectively assess the adequacy of communication, but they have been extensively used [6, 10, 19].

The most notable research applying this task, for example the studies conducted by Clark [6], studied how participants arrange an identical set of figures into matching orders. On 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* [4, 11]. 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.

In the following sections, we review prior research on referential communication and state our hypotheses for how prior findings will apply to machine translation-mediated communication.

2.3 Speakers' Descriptions

When two people in conversation refer repeatedly to the same object, the referring expressions are often simplified and shortened [19], and the expressions converge on the same or similar referring expressions [3, 4]. There are two main strategies for shortening referring expressions: simplification and less often, narrowing [2] [6]. With simplification, certain details, usually adjectives, are omitted while retaining the referent's overall image, as in "a guy running with shiny gold-rimmed glasses" to "a guy running with gold-rimmed glasses." With narrowing, the focus of a perspective is narrowed to just one part of a figure. The perspective typically moves to a peripheral but distinctive part, as in "a guy running with shiny gold-rimmed glasses" to "gold-rimmed glasses."

In machine translation-mediated communication, shortened referring expressions are not necessarily translated correctly; even when referring expressions overlap considerably, machine translation may generate something totally different based on very small changes. For example, a Japanese sentence "Don't worry about such a trivial problem" is translated into "Please be not worried about such a trivial problem" in English, while "There's no need to worry about such a trivial problem" is translated into "Even if we do not care, such a trivial problem is good." Because abbreviation is problematic for machine translation, we expect that participants will identify a figure using identical referring expressions throughout the conversation.

2.4 Addressees' Responses

Each referring expression offered by a speaker constitutes a proposal; the addressee can either ratify it by accepting it immediately, ask questions and/or confirm whether his/her understanding is correct, contribute a counterproposal, or wait for the speaker to propose something else.

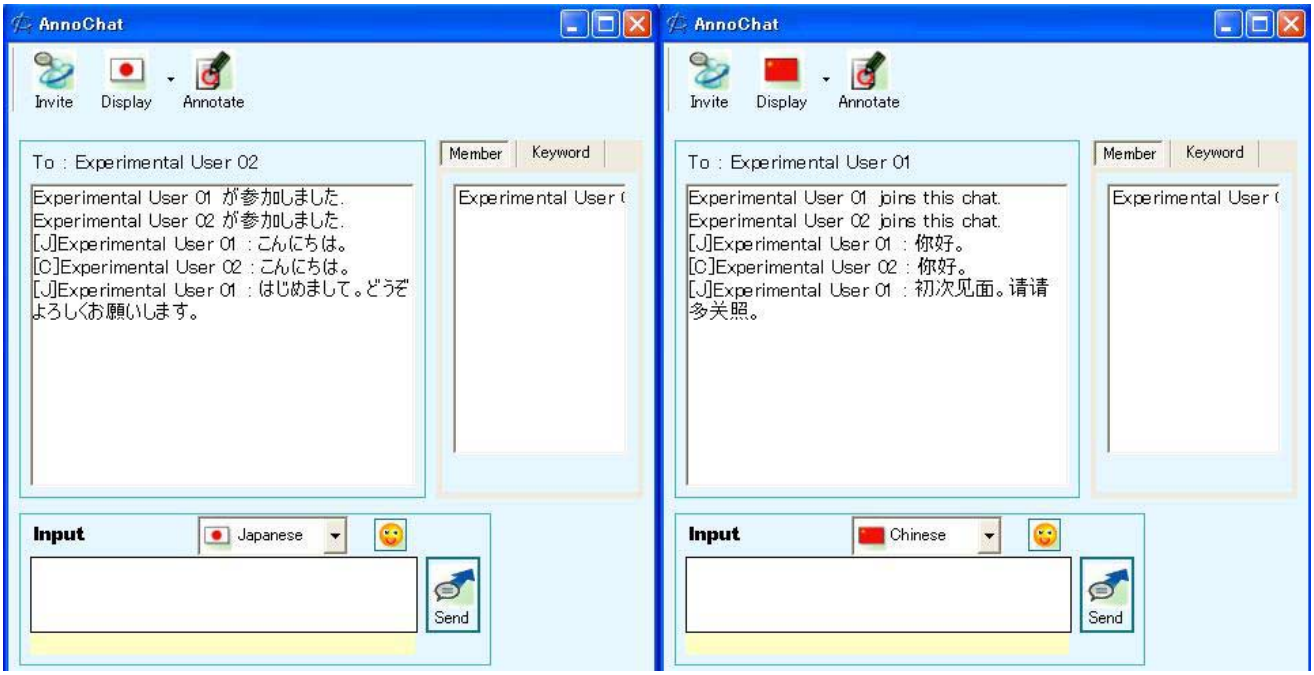
In machine translation-mediated communication, a speaker and an addressee cannot share the same referring expressions; participants not only view messages written in different languages but translations between two different languages are not transitive: translation from language A to B and back to A does not yield the original expression. The intransitive nature of machine translations results from its development process; translation from language A to B is built independently of translation from language B to A. In such conversations, the addressee cannot echo the speaker's expression as a way of accepting it, illustrating that they are referring to the same thing.

Figure-matching is a task that requires two participants to establish that the addressee has understood the speaker's current utterance before continuing [6]. Thus, we expect that addressees will frequently ask questions or confirm their understanding when using machine translations to ensure that they are talking about the same figure as the speaker.

2.5 Efficiency of Mutual Acceptance Process

The most efficient way to identify a referent consists of two steps called a "basic exchange": (a) the presentation of a referring expression and (b) its acceptance [6]. When the referring task is difficult (e.g., when the referent must be selected from among many similar figures or is difficult to describe), basic exchanges seldom occur in the first trial. In later trials, the rate of basic exchanges increases because they can be based on prior mutually accepted descriptions [6]. Indeed, in our preliminary experiment with Japanese participants using Japanese, participants managed to match only 30% of figures using basic exchanges in their first trial, whereas by their second trial, they successfully matched 90% of the figures using basic exchanges.

[2] In preliminary experiments with Japanese participants using Japanese, the participants shortened their referring expressions by simplification.



(a) Japanese Chat Interface **(b) Chinese Chat Interface**

Figure 1 Chat Interface (Japanese and Chinese)

In machine translation-mediated communication, we expect that pairs will have trouble successfully identifying referents, even in their second trials, due to the inconsistent and asymmetric nature of the translations discussed in Sections 2.3 and 2.4.

2.6 Hypotheses

The following hypotheses summarize this discussion:

H1 (how each tangram is described): Speakers will be less likely to abbreviate their referring expressions over trials when using machine translation as opposed to English (shared second language).

H2 (how addressees respond): Addressees will more often ask questions and/or confirm understanding when using machine translation as opposed to English.

H3 (efficiency): Pairs will more efficiently identify a referent when using English rather than machine translation.

3. METHOD

3.1 Design

In this study, we compared the referring and identifying processes of eight pairs using their common language (English, which is not their native language) and using their native languages by machine translation.

The experiment was separated into two phases. The first half was conducted in 2005 as part of the Intercultural Collaboration Experiment (ICE2005), jointly hosted by Chinese, Japanese, Korean, Malay, and Thai universities and research institutes. The other half was conducted in 2005 in Japan. The two phases differed only in the site at which the experiment took place and in whether we conducted detailed interviews, which were performed only in the second phase.

In the experiment, pairs sat in different rooms. Each pair was given the same Tangram figures arranged in different sequences and instructed to match the arrangements of their figures using a multilingual chat system (Figure 1). After matching their arrangements, their figures were placed in two new random orders, and the procedure was repeated. They carried out the task twice in English and twice in their native languages using machine translation[3].

3.2 Participants

Six pairs of university students in China, Korea, and Japan participated in the first phase (six Japanese in Japan, three Chinese in China and three Koreans in Korea). Two pairs of university students joined the second phase (two Chinese in Japan and two Japanese in Japan). For each pair, native languages differed between the two participants. None of the participants knew their partners before the experiments.

None of the participants understood both Chinese and Japanese, or Korean and Japanese. They only understood their native language and English. English proficiency levels varied, but all participants had studied English for more than six years and were

[3] We used a slightly different task from standard referential communication tasks, in that the participants were not explicitly designated as a Director or a Matcher. We were interested in investigating how the participants actually proceed with the matching task; we expected that machine translation would complicate their decisions of how to proceed with the matching. However, we could not find support for our expectation. Participants rarely discussed how to proceed with the task (who will become a Director or a Matcher); instead, one of the participants typically started explaining his/her figure, and the other tried to figure out which figure the speaker was talking about.

exposed to English in their lives; they read and wrote research papers in English.

The participants frequently used emails and instant messaging. However, they rarely used machine translation.

3.3 Apparatus

For the experiment, a multilingual chat system, "AnnoChat[4] [9]" (Figure 1), was prepared that automatically translates each message into the other languages while providing awareness information of the typing of other users. The chat interface allows users to select their browsing and typing languages from Chinese, English, Korean, and Japanese. For example, a Japanese participant who has selected Japanese for his browsing and typing language will be able to read and write in Japanese. Similarly, when a pair selects English as their browsing and typing language, they can both read and write in English.[5]

The machine translation software embedded in AnnoChat is a commercially available product, and rated as one of the very best translation qualities among other translation software. The machine translation quality from Japanese to Chinese, Japanese to Korean, and Korean to Japanese was evaluated as "Good" (within four scales of "Very Good," "Good," "Not Bad," and "Bad"), and Chinese to Japanese as "Not Bad. [26]"[6]

AnnoChat also offers such functions as adding annotations. However, to remove the possibility of influence from the functions, participants were prohibited from using them.

3.4 Procedure

Each pair was presented with ten tangram figures (e.g., Figure 2) arranged in different sequences and instructed to match the arrangements of figures using AnnoChat.

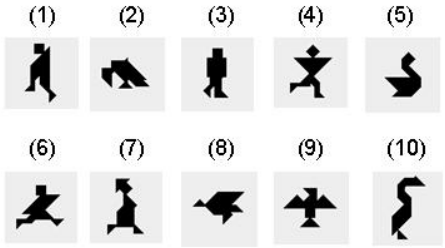


Figure 2. Ten tangram figures used in the experiment.

The experiment's procedure was as follows:

Procedure (1): Participants engaged in a short term free discussion on how to support intercultural collaboration using Annochat to become familiar with it.

Procedure (2): Before matching the figures, participants were told that: a) each person has the same ten figures in different orders; b) their task was to match the arrangements of the figures; and c) they could use any strategy to accomplish the task.

Procedure (3): Each pair worked on four matching tasks:

Procedure (3-1), First trial in English: Each participant accessed a URL individually arranged immediately before the experiment and got a figure set. Participants set their display and descriptive language into English and matched their arrangements in English.

Procedure (3-2), Second trial in English: As in Procedure (3-1), each participant got a figure set in which the same figures were arranged in different orders. Participants repeated the procedure.

Procedure (3-3), First trial in native languages using machine translations: Each participant got a new figure set whose figures differed from those in Procedures (3-1) and (3-2). Participants set their display and descriptive language into their native languages and matched arrangements in their native languages (using machine translation).

Procedure (3-4), Second trial in native language using machine translation: Participants' figures were placed in two new random orders, and then the procedure was repeated.

The two figure sets (used in English and in native languages) were counterbalanced for order.[7] The experimental design was incomplete in that language condition was not counterbalanced for order[8].

Procedure (4): Following the four matching tasks, participants were interviewed, as described in Section 3.6.

3.5 Coding

We developed a coding scheme to capture the primary purpose of each utterance to investigate the relationship between machine translation and dialogue. The categories used for analyses in this paper are presented in Table 1.

Table 1 Utterance Types

Category	*Definition*
Presentation (Description)	A speaker describing a figure: e.g., "Figure 7 looks like a bird flying to the left." "Its neck is long."
Presentation (Noun phrase)	A speaker explaining a figure with a noun phrase: e.g., "Figure 5 is a dancing lady."
Question or Confirmation	An addressee asking the speaker for clarification, more information, or confirming an understanding: e.g., "Is she wearing a long dress?"
Acceptance	An addressee accepting the speaker's presentation: e.g., "Ok," "That's my 5th figure."
Not Understood	An addressee telling the speaker that he/she did not understand the message (e.g., "I don't understand.").
Others	Utterances that don't belong to any of the above categories.

[4] http://yoshino.sys.wakayama-u.ac.jp/spark/?lng=en&page=AnnoChat

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

[6] Translation quality among European language pairs (such as French, Spanish, Italian, etc.) and English were evaluated as "Very Good."

[7] Interview results and average matching times indicate that the two figure sets have similar difficulty.

[8] The flaw in the experimental design actually works against our hypotheses, in that the pairs are more familiar with one another by the time they match their arrangements using machine translation.

Four independent coders classified samples of utterances until they reached 90% agreement. Two coders understood Chinese, Japanese, and English, and the other two understood Korean, Japanese, and English. The first two coded transcripts of the Chinese-Japanese pairs, and the other two coded transcripts of the Korean-Japanese pairs. Agreements between the two coders remained high throughout (Chinese-Japanese pairs: Cohen's Kappa = 0.78, Korean-Japanese pairs: Cohen's Kappa = 0.81).

3.6 Interviews

At the end of the four matching sessions, we interviewed the participants about ease of creating utterances, ease of understanding utterances, how efficiently they conducted the matching tasks, how difficult the matching task was, the usefulness of machine translation, and their English proficiency.

In the first phase, we conducted the interviews in English over AnnoChat. In the second phase, we conducted face-to-face interviews. Here, we used Japanese for Japanese participants and English for Chinese participants.

4. RESULTS

Since no significant differences were found in any of the dependent variables between the first and second phases of data collection, we discuss these results as a whole.

4.1 Descriptions of Tangrams

4.1.1 Description vs. Noun phrase

First, we investigate how the speakers referred to the Tangram figures. We classified each of a speaker's referring expressions into one of two categories: "Description" and "Noun phrase." Typically, speakers use descriptions when first establishing a common perspective on a referent but shorten these descriptions to noun phrases once common ground is established.

Table 2 shows the proportion of referring expressions that were noun phrases for the first and second trials when using English and when using machine translation.

Table 2 Proportion of Referring expressions in "Noun phrase"

	1st trial	*2nd trial*
English	15%	67%
Machine Translation	6%	28%

We performed a repeated measures analysis of variance (ANOVA) on the proportion of noun phase, using condition order and language conditions as repeated factors. The proportion of "Noun Phrase" units increased significantly in second trials ($F[1,7]=152.95$, $p<.001$). The proportion of "Noun Phrase" units differed significantly between the English and machine translation conditions ($F[1,7]=95.53$, $p<.001$). Also, there was a significant language by trial interaction ($F[1,7]=15.90$, $p<.01$).

It appears that participants using English can easily move from descriptions to noun phrases across trials as in general conversations, whereas participants using machine translation have difficulties in moving to noun phrases.

4.1.2 Inconsistencies in Machine Translation

Our first hypothesis stated that speakers would be less likely to abbreviate their referring expressions over trials when using machine translation as opposed to when speaking in English. To test this hypothesis, we classified each of a speaker's referring expressions on the second trial in each condition into one of four categories: identical, narrowed, simplified, and different (Table 3).

Table 3. Ways Participants Explained Each Figure in the Second Trial

	Identical	*Narrowed*	*Simplified*	*Different*
English	33%	6%	50%	11%
Machine Translation	58%	22%	14%	6%

As predicted by *H1*, participants using machine translation rarely shortened referring expressions in a simplified manner ($F[1,7]=25.53$, $p=.001$). Instead, they often identified a figure by using *exactly* the same referring expressions as on the first trial (e.g., Figure 3) or by using a distinctive narrowed term from the referring expression in the first trial (e.g., Figure 4) ($F[1,7]=138.03$, $p<.001$).

To see why the speakers did not shorten their lengthy referring expressions in their second trial, we examined the conversations in our experiment in further detail. As expected, we found many cases in which machine translation translated messages quite differently in the first and second trials, even when the referring expressions overlapped considerably (Figure 3).

Japanese Screen (translated in English) <First Trial>	Chinese Screen (translated in English) <First Trial>
J: My second figure looks like an animal. J: It has four feet and a tail. C: *That's my the 9th.*	J: *My second figure is like an animal.* J: *It has four feet and a tail.* C: That's my 9th.
<Second Trial>	**<Second Trial>**
J: My second figure is an animal with a tail and four feet. C: *What kind of meaning and rice boy?* J: My second figure looks like an animal. J: It has four feet and a tail. C: *Oh, I understand, and am the 8th.*	J: *My role of a young handsome beau is a boy with a tail and 4 feet.* C: What do you mean? A handsome boy? J: *My second figure is like an animal.* J: *It has four feet and a tail.* C: I got it. It's my 8th figure.

Figure 3 Translation Inconsistencies and Participants Using Identical Referring expressions as First Trial. Italicized messages are those outputted from machine translation.

To understand what the participants were trying to communicate, we translated both the Chinese and Japanese messages into English. Also to share the automatically translated messages in this paper, we further translated the Japanese and Chinese

translated messages into English, referring to the automatically translated results of the Chinese and Japanese. (Here, the machine translation quality from Chinese to English was "Not Bad" and Japanese to English was "Good.") The translated output from machine translation is italicized.

In the excerpt above, a Chinese participant and a Japanese participant matched one of the Tangrams in two sentences: "looks like an animal" and "it has four feet and a tail." In their second trial, the Japanese speaker tried to explain the same figure in one sentence: "an animal with a tail and four feet." In ordinary conversation, the addressee would obviously recognize the meaning of the sentence (i.e., recognize the original in the new version). However, machine translation generates something quite different ("My role of a young handsome beau is a boy with a tail and 4 feet.") based on very small changes. As a result, the reference becomes uninterpretable, and the Japanese speaker reuses exactly the same explanation he gave in the first trial.

In the post-experimental interview, the Japanese participant said,

> I got afraid of rephrasing an expression. I thought it was a reliable way to use the same referential expression as in the first trial.

We also found that in their second trial, speakers using machine translation preferred to narrow expressions rather than simplify them. The following excerpt (Figure 4) shows a speaker using narrowed expressions in his second trial.

Japanese Screen (translated in English) <First Trial>	Korean Screen (translated in English) <First Trial>
J: 3 looks like a lady wearing Kimono. J: It looks like she's looking toward right. K: *Twice is wearing a Kimono.*	J: *3 looks like a lady wearing Kimono.* J: *It seems she's looking toward right.* K: Number 2 is wearing a Kimono.
<Second Trial>	<Second Trial>
K: *2 is Kimono.* J: Kimono is 9.	K: 2 is Kimono. J: *Kimono is 9.*

Figure 4 Participants Shortening Referring expressions through "Narrowing"

In the excerpt above, in their first trial a Korean participant and a Japanese participant agree that the figure they are discussing is a lady wearing a Kimono who is looking to her right. In their second trial, they match the same figure by only giving the distinctive term, "kimono."

We infer that "narrowing" is observed more frequently in machine translation-mediated communication because distinctive terms such as "kimono" have few alternatives in translation, and thus, participants feel safe using them to match the figures. Indeed, note the Korean participant's comment in the post-experimental interview:

> In the first trial, it seemed that "Kimono" was translated right. So I thought we can identify the same figure by just saying "Kimono." I thought it's a fast and safe way…"

In sum, it appears that participants using machine translation tend, in their second trial, to use the same referring expressions as they did in the first trial or to select distinctive terms from the first trial to safely identify figures without misunderstandings.

4.2 Addressee's Responses

To investigate how the addressees identified the figures, we classified addressee's each message into four categories: Questions/Confirmation, Acceptance, Not Understood, and Others.

Table 4 shows the proportion of messages in each of the four categories for the first and second trials when using English and when using native languages via machine translation.

Table 4 Proportion of Addressee's Messages Spoken in Each Category

Category	*English 1st trial*	*English 2nd trial*	*MT 1st trial*	*MT 2nd trial*
Questions/ Confirmation	25%	17%	24%	20%
Acceptance	66%	83%	65%	78%
Not Understood	3%	0	5%	2%
Others	6%	0	6%	0

We predicted in *H2* that addressees would often ask questions or confirm their understandings of the translated messages, but we did not find support for our hypothesis; there was no significant difference in the proportion of "Question or Confirmation" messages between conversations in English and in native languages using machine translation.

4.2.1 Asymmetries in Machine Translation

As predicted in Section 2.4, we found many cases in which communication broke down in machine translation-mediated communication due to the asymmetric nature of translation (Figure 5).

Japanese Screen (translated in English)	Chinese Screen	Chinese Screen (translated in English)
J: 1 is a dancing lady. C: *It jumped.* J: 3 is a person with his head down.	J: 1是跳的女性 C: 是跳的 J: 朝向了下的人是3	J: *1 is a dancing lady.* C: Ok, a dancing one. J: *A person looking down is 3.*

Figure 5 Communication Breakdown Due to Asymmetric Nature of Machine Translation

In the above excerpt, the Japanese participant is explaining his first figure, and the Chinese participant shows that she understands the message. From the interview we learned that in her response, she carefully responded, deliberately echoing the same word the Japanese participant had used, to emphasize that she understood the message. However, since the Chinese to Japanese translation translated "dance" into "jump," the Japanese participant got confused. He breaks off matching his first figure and starts explaining his third figure. The Japanese participant said in the interview:

> I couldn't understand what my partner meant, so I decided to proceed with another figure, which looked easier to match.

4.2.2 Failure to Comprehend

As in Figures 3 and 5, we found many cases where addressees had trouble understanding a speaker's messages.

In general conversation, when an addressee does not understand the speaker's message, the addressee tries to pinpoint the problem. The speaker also answers addressee's questions. In other words, speaker and addressee minimize collaborative effort by quickly and informatively indicating what is needed for mutual acceptance [6].

Observing the "Not Understood" statements in our experiment, we found that all such statements in English pinpointed their questions to the incomprehensible part of the message and asked for specific information to match the figures. The participants answered their partner's questions and used these responses to adjust subsequent utterances. In contrast, approximately half of such statements in machine translation-mediated communication did not ask for specific information, as in "Sorry, I don't understand," or "What do you mean?"

Moreover, participants avoided focusing on the incomprehensible part of messages to discover what was wrong. Since translations are not transitive, it appears that they cannot efficiently solve the problem. Speakers have little choice but to offer more information and proceed with the task, as in Figure 3.

In Figure 3, the Japanese participant first describes his second figure. However, the machine translation mistranslates the message, so the Chinese participant asks a question. However, since the translation is not transitive, the question makes no sense to the Japanese participant. So the Japanese participant disregards the question and offers further information about the figure until the Chinese participant confirms her understanding. In the interview, the Japanese participant said,

> I guess my partner was confused by a mistranslation. His question made no sense to me. I didn't say anything about "rice boy"…What's a "rice boy" anyway?

4.2.3 Descriptions from Many Different Perspectives

Instead of addressees asking questions or actively confirming understanding, they tended to wait for the speakers to offer further information until they could confirm their understandings.

Consistent with quantitative results, speakers tended to describe the figures more frequently in machine translation than in English. From further detailed analysis of such utterances, we found that speakers often described the figures from many different perspectives. The excerpt below captures this tendency:

Japanese Screen (translated in English)	Chinese Screen (translated in English)
C: *My 1 wildly galloping horse at four o'dock.* J: Is 4 also a dog? C: *The tail of 4 is square.* C: *4 rise only by a hind leg.* J: It's my 5.	C: My number 4 is a running horse. J: *Is 4 a dog?* C: The tail of 4 is square. C: It's standing on it's hind legs. J: *It's my 5.*

Figure 6 Participant Describing a Figure from Many Different Perspectives

In the above excerpt, the Chinese participant is describing her fourth figure, and the Japanese participant is asking a question or seeking confirmation. However, instead of answering the question, the Chinese participant offers additional information about the figure. In the interview, the Chinese participant said:

> I thought we could finish the task faster if I keep on giving more and more information [instead of responding to the question].

Another participant said:

> In English, I cannot instantly think of many expressions, but in my native language, I can easily think of many different expressions.

It seems that participants can minimize mutual effort in collaboration by offering more and more information until their partner confirms understanding.

4.3 Efficiency of Mutual Acceptance Process

To see whether efficiency differs between conversations in English and those in machine translation-mediated communication, we performed a repeated measures analysis of variance (ANOVA) on the proportion of basic exchanges, using condition order and language conditions as repeated factors As discussed in section 2.5, basic exchanges, in which a speaker presents a referring expression and the addressee accepts it, are the most efficient way to identify a referent.

Table 5 Average Proportion of Basic Exchanges in the First and Second Trials of Each Condition.

	First trial	*Second trial*
English (Common Language)	0.24	0.70
Machine Translation (Native Language)	0.23	0.46

The proportion of basic exchanges increased significantly in second trials ($F[1,7]=68.60$, $p<.001$). There was no main effect of language condition. We found a slight language by trial interaction ($F[1,7]=4.47$, $p=.07$). Consistent with *H3*, participants had trouble identifying the referents in basic exchanges with machine translations even in their second trials.

5. DISCUSSION

The results provide insight into the effects of machine translation on communication in referential communication. From our experiment, we found that (1) echoing, an important tool for ratification process in lexical entrainment [23], is disrupted by asymmetries in machine translations; (2) the process of shortening referring expressions is also disrupted because the translations do not translate the same terms consistently throughout the conversation.

In the following, we first discuss on how the results might vary with different experimental settings. Next, we discuss on how the results might contribute in building better machine translation tools for communication use.

5.1 Impacts of Experimental Settings

As for asymmetries in machine translation, we found that the addressee's echoing is disrupted. Since the participants could not efficiently indicate that they understood each other's messages through echoing in machine translation-mediated communication, addressees tended to wait for speakers to provide sufficient information before confirming the referent. Although such a communication style (one person offering more and more information until his/her partner understands) may be useful when the information to be transmitted is already determined, it may be less suitable when the communicational exchange is extremely important, such as in negotiation, or when there is time pressure.

As for inconsistency in machine translation, we found that shortening of referring expressions is disrupted. The phenomenon reminds us of Krauss's study [18] where participants also failed to abbreviate referring expressions under audio delay. Participants in our experiment often used exactly the same referring expressions throughout the process to ensure consistent translation. Identifying a figure using the same expression might be useful in cases where the participants continuously work together and frequently refer to the same referents, as in our experiment. Although people have remarkably good memories for the expressions they entrain on [12], we speculate that it would be difficult for participants to remember lengthy expressions exactly when working intermittently.

Also, identifying a figure using the distinctive part of expressions in the prior trial ("narrowing") would be useful for groups where discussion members are fixed. However, it would be difficult for a participant to guess what the reference indicates without sharing the process. Thus, "narrowing" is probably unfit for collaborative work where members change frequently and work intermittently.

5.2 Machine Translation for Communication Use

Earlier studies of machine translation have focused almost exclusively on translating written (unidirectional) documents. Many natural language processing researchers have become experts on developing high quality translation algorithms of certain language pairs in one direction. Thus, most research in machine translation has not taken into account interaction (dual-directional) factors. Also, machine translations have commonly been evaluated by the adequacy and fluency of translated single sentences.

Asymmetries in machine translation result from the constitution of machine translation; machine translation systems consist of an aggregation of unidirectional translation systems. Asymmetries cannot be resolved by improving the translation quality of single sentences. As was shown in Figure 5, communication breaks down even when the translation quality of each individual sentence is high. In order to support natural referring behavior, it is important that machine translations of each pair of languages coordinate and resolve asymmetries.

Similarly, inconsistencies cannot be resolved through improvement of translation quality on single sentences, since inconsistencies are a matter of context.

We believe that there is a need to consider a new definition of translation quality that improves machine translation-mediated communication.

6. CONCLUSION

Even though multilingual communities using machine translation to overcome language barriers are increasing, we still do not have a complete understanding of how machine translation affects communication. Research literature in English on CSCW has focused almost exclusively on computer-mediated collaboration in English [16, 22, 29]. It would be instructive to examine how people establish common ground using machine translation tools.

In this paper, we considered the effects of machine translation on referential communication among pairs using different languages. In particular, we compared referential communication between speakers of different native languages using machine translation and their shared second languages (without machine translation).

The results show that in machine translation-mediated communication, participants have trouble efficiently identifying a referent through basic exchanges; lexical entrainment is disrupted in machine translation-mediated communication because echoing is disrupted by asymmetries in machine translations. The process of shortening referring expressions is also disrupted because the translations do not translate the same terms consistently throughout the conversation.

To overcome asymmetries and inconsistencies in machine translation-mediated communication, participants tried to minimize exchanges (addresses waited for the speaker to provide further information until they can identify the figure) and used exactly the same referring expressions throughout the experiment.

Since such an unwieldy conversational style would not be useful in general conversation, there is a need to support natural referential behavior in machine translation-mediated communication. For example, support that creates correspondences among references (or keywords) between the two languages may help. Also, support that creates correspondences among referring expressions before and after shortening may help.

Our next step is to investigate more carefully what happens when there are more than two trials. Are participants able to overcome the disruption in machine translation and manage somehow to share the same perspective later? We are also interested in investigating communication patterns of machine translation-mediated communication with other types than information transmission. Also, we will build a system to support symmetry and consistency into machine translation and examine its effects.

7. ACKNOWLEDGMENTS

This research is partially supported by International Communications Foundation.

We would like to express our sincere gratitude to Susan Fussell for providing significant help in improving this paper. We want to thank the anonymous reviewers who provided constructive comments and suggestions. We thank Keiji Hirata for his careful comments on an earlier draft of this paper, and Takashi Yoshino for letting us use the multilingual chat system ("AnnoChat") in our experiments.

8. 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] Aiken, M. Multilingual Communication in Electronic Meetings. *ACM SIGGROUP,* Bulletin, 23, 1, 2002.

[3] Brennan, S. E. Lexical Entrainment in Spontaneous Dialogue. *Proceedings of International Symposium on Spoken Dialogue*, 1996, 41-44.

[4] 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.

[5] Clark, H. H. and Marshall, C. E. Definite reference and mutual knowledge. *Elements of discourse understanding*, 1981, 10-63.

[6] Clark, H. H. and Wilkes-Gibbs, D. Referring as a collaborative process. *Cognition*, 22, 1986, 1-39.

[7] 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.

[8] Danet, B. and Herring, S.C. Introduction: The Multilingual Internet. *Journal of Computer Mediated Communication*, 9, 1, 2003.

[9] Fujii, K., Yoshino, T., Shigenobu, T. and Munemori, J. Development of an Intercultural Collaboration System with Semantic Information Share Function. *Proceedings of Ninth International Conference on Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies*, 2005, 425-430.

[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] Garrod, S. and Doherty, G. Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition*, 53, 1994, 181-215.

[13] Higgins, E. T. The 'communication game': Implication for social cognition and communication. E. T. Higgins, C. P. Herman & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium*, 1, 1981, 343-392.

[14] Isaacs, E. A. and Clark, H. H.: References in Conversation Between Experts and Novices. *Journal of Experimental Psychology*, 16, 1, 1987, 26-37.

[15] Ishida,T. Language Grid: An Infrastructure for Intercultural Collaboration. *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, keynote address, 2006, 96-100.

[16] Jackson, M., Anderson, A., McEwan, R., and Mullin, J. Impact of Video Frame Rate on Communicative Bahavior in Two and Four Party Groups. *Proceedings of CSCW*, 2000, 11-20.

[17] Kim, K.J. and Bonk, C.J. Cross-Cultural Comparisons of Online Collaboration. *Journal of Computer Mediated Communication*, 8, 1, 2002.

[18] Krauss, R. M. and Bricker P. D. Effects of transmission delay and access delay on the efficiency of verbal communication. *Journal of the Acoustical Society of America*, 41, 1967, 286-292.

[19] Krauss, R.M. and Glucksberg, S. The development of communication: Competence as a function of age. *Child Development*, 40, 1969, 255-256.

[20] Krauss, R. M. and Weinheimer, S. Changes in reference phases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, 1, 1964, 113-114.

[21] Krauss, R. P. and Fussell, S. Mutual knowledge and communicative effectiveness. Intellectual Teamwork: Social and Technological Foundations of Cooperative Work, 1990, 111-146.

[22] McCarthy, J., Miles, V., and Monk, A.: An Experimental Study of Common Ground in Text-Based Communication. *Proceedings of CHI*, 1991, 209-215.

[23] Metzing, C. and Brennan, S. E. When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49, 2003, 201-213.

[24] Nomura, S., Ishida, T., Yamashita, N., Yasuoka. M. and Funakoshi. K. Open Source Software Development with Your Mother Language: Intercultural Collaboration Experiment 2002. *International Conference on Human-Computer Interaction (HCI-03)*, 4, 2003, 1163-1167.

[25] Ogden, B., Warner, J., Jin, W. and Sorge, J. Information Sharing Across Languages Using MITRE's TRiM Instant Messaging. 2003.

[26] Ogura, K., Hayashi, Y., Nomura, S. and Ishida, T. User Adaptation in MT-mediated Communication. *The First International Joint Conference on Natural Language Processing (IJCNLP-04)*, 2004, 596-601.

[27] Takano, Y. and Noda, A. A temporary decline of thinking ability during foreign language processing. *Journal of Cross-Cultural Psychology*, 24, 1993, 445-462.

[28] Tung, L. L. and Quaddus, M. A. Cultural differences explaining the differences in results in GSS: implications for the next decade. *Decision Support Systems*, 33, 2, 2002, 177-199.

[29] Veinott, E., 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, 302-309.

[30] Yamashita, N. and Ishida, T. Automatic Prediction of Misconceptions in Multilingual Computer-Mediated Communication. *Proceedings of IUI*, 2006.