# Automatic Prediction of Misconceptions in Multilingual Computer-Mediated Communication

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
Multilingual communities using machine translation to overcome language barriers are showing up with increasing frequency. However, when a large number of translation errors get mixed into conversations, users have difficulty completely understanding each other. In this paper, we focus on misconceptions found in high volume in actual online conversations using machine translation. We first examine the response patterns in machine translation-mediated communication and associate them with misconceptions. Analysis results indicate that response messages to include misconceptions posted via machine translation tend to be incoherent, often focusing on short phrases of the original message. Next, based on the analysis results, we propose a method that automatically predicts the occurrence of misconceptions in each dialogue. The proposed method assesses the tendency of each dialogue including misconceptions by calculating the gaps between the regular discussion thread (*syntactic thread*) and the discussion thread based on lexical cohesion (*semantic thread*). Verification results show significant positive correlation between actual misconception frequency and gaps between syntactic and semantic threads, which indicate the validity of the method.

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

## General Terms
Human Factors, Experimentation, Measurement.

## Keywords
Multilingual Groups, Machine Translation, Computer-Mediated Communication, Misconception.

## 1. INTRODUCTION
As computer-mediated communication allows more collaboration across broad distances, collaboration involving people speaking different languages and across cultures is starting to play a significant part in our lives. To date, however, the research literature in English on Intelligent User Interfaces (IUI) and Computer Supported Cooperative Work (CSCW) have focused almost exclusively on computer-mediated collaboration in English, neglecting populations communicating in other languages [9].

In multilingual groups where the native languages of members differ, communication typically takes place in one language, thereby requiring that some members communicate in a non-native language. However, members required to communicate in their non-native language frequently find communication difficult [23, 1, 16], and so such collaboration tends to be ineffective [2, 24].

One solution to such problems is machine translation. Although machine translation lacks accuracy, participants may not require perfect translation as long as comments can be understood [2]. In fact, we have seen several actual Internet communities communicating via machine translations[1]. Such communities are centered in Eastern Asia, where their lingua franca is English, but few are actually proficient in the language. The number of such communities is expected to grow in the future [8].

However, pitfalls exist in communication relying on machine translations, especially when the translation quality is low. One pitfall is the occurrence of misconceptions among participants. When translation quality is low (as in Eastern Asia due to grammatical construction dissimilarity), the burden on participants to "guess" the mistranslated part of comments increases. Under such circumstances, participants may falsely assume that others are speaking and understanding on the basis of the same information and interpretations. In this paper, we refer to such a phenomenon as "misconception."

As described in Section 2, we found a considerable amount of misconceptions in our study, where Chinese and Japanese members reached a consensus through discussions via machine translation. Even though Chinese and Japanese members were speaking interchangeably and answering as if understanding each other, we found many instances where they actually did not. The problem appears to lie in the nature of machine translation-mediated communication, where discussion members have


*IUI'06*, January 29-February 1, 2006, Sydney, Australia.


[1] http://bbs.enjoykorea.naver.co.jp/jaction/list.php?id=enjoyjapan_8

difficulty confirming what has been transmitted or whether comments were correctly transmitted to other members.

Such misconceptions can be very problematic, especially in group work. For example, in collaborative situations where discussion members are unaware of the existence of misconceptions, it is possible for expectations to be fractured, sending repercussions into subsequent conversations.

Thus, it is important to consider ways to prevent such misconceptions. As Fussell and Krauss point out, we must provide mechanisms that detect and correct communication errors (including misconceptions) [12].

However, research on misconceptions or communication breakdowns [21, 4] has largely drawn upon hand-built analyses, such as ethnographic or sociolinguistic studies. A method that automatically detects misconceptions or communication breakdown has yet to be proposed.

As a first step to solve the problem, we propose a method that quantifies misconceptions caused by translation errors. The method automatically predicts the occurrence of misconceptions in each dialogue. By extending the method into a system, we may alert discussion members to the existence of misconceptions.

Below, we first explain how actual misconceptions occur from translation errors by drawing an example of misconception from our experiment. Then, we discuss the differences between misconceptions caused by translation errors and misconceptions mainly treated in previous research. In Section 3, we analyze machine translation-mediated communication at the micro level by comparing direct responses with and without machine translation. We then associate the relationships between these characteristics with misconceptions. In Section 4, we expand the unit of analysis from the micro level (direct responses) to the macro level (entire dialogue). Based on previous results, we focus on the gaps between discussion threads based on header information and lexical cohesion and consider how the gaps relate to misconceptions. We then propose a method that quantifies misconceptions caused by translation errors by combining the micro- and macro-level analysis results. Finally, we verify the validity of our method. We conclude the paper in Section 5 followed by a brief comment on future work.

# 2. MISCONCEPTIONS IN ACTUAL COMMUNICATION

## 2.1 Case: Asia Broadband Project

We chose the Asia Broadband Project as a case study. The Asia Broadband Project was conducted in 2003 by Chinese and Japanese universities and research institutes with the support of the Japanese government.

Eighteen Japanese members from three universities and two research institutes, and sixteen Chinese members from two universities joined the project. Most members in dispersed locations (see Table 1 for details) had never met before the project. All members had specialized skills in and knowledge of computer systems.

Their mission was to discuss and settle on an implementable tool to encourage intercultural collaboration within one month. During the project, all discussions took place on a machine translation-embedded BBS, which automatically translates Chinese and Japanese messages and displays both original and translated messages. Since none of the members understood both Chinese and Japanese, all members posted and read the messages in their native languages. During one-month discussions, 1,106 messages were posted on BBS: 649 from the Japanese members and 457 from the Chinese members.

Machine translation software embedded in BBS is a commercially available product. From evaluation results of Web pages, the machine translation's quality from Japanese to Chinese was evaluated as "Good" (within four scales of "Very Good," "Good," "Not Bad," and "Bad") and Chinese to Japanese as "Not Bad"[2].

**Table 1. Number of members at each location.**

| **Japan** | **L1** | **L2** | **L3** | **L4** | **L5** |
|---|---|---|---|---|---|
| **# of members** | 10 | 3 | 1 | 3 | 1 |
| **China** | **L6** | **L7** | | | |
| **# of members** | 6 | 10 | | | |

Discussion logs in the Asia Broadband Project are well suited for the analysis of misconceptions caused by translation errors: (1) Since the discussion topic was software development and all members were specialized in the area, misconceptions were unlikely to reflect cultural differences or backgrounds among members; most of the misconceptions were caused by translation errors. (2) Discussion types in the project can be categorized into "consensus building" or/and "decision making," which typically requires mutual understanding without misconceptions. (3) Since both messages before and after translation were available from the BBS, we could analyze how translation errors caused misconceptions.

## 2.2 An Example of Misconception

In this section, we report on misconceptions found in the Asia Broadband Project. We examined all 1,106 messages before translation (original messages in their native languages) and after translation (messages translated by machine translation). A human Chinese-Japanese translator translated the original messages, and thus we were able to identify where mistranslations (by machine translation) and misconceptions took place.

In discussions, when members presented an objection, they often showed polite respect for the opinions of others and/or agreed with some parts of such opinions before stating their own thoughts. However, since translation errors got mixed into their conversations, distinguishing the preface from the main part of the message was difficult. As a result, it seemed that members were focusing on affirmative words, and so they were mislead into believing that others agreed with their opinion without realizing that others were actually opposed to their opinion. Indeed, we found many instances where users were mutually agreeing on different topics. The example below captures, to a small extent, this tendency:

[Message written by a Chinese member - message translated by a human translator]: *I think your idea is very good and important.*

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

*(However, Omni-directional camera is not enough to actualize your idea. For example, ) We cannot collect the images of new products each time they arrive. The key technology to actualize your idea is to use the image segmentation method. Omni-directional camera must be compatible with the technique.*

[The same message as the above – message translated by machine translation]: *I think that it is very good by all means in and am very important. Because we add the product that the shop is new each time is impossible, it collects an image again. Such a way is to have supported Omni-directional Camera because the essential point which realize such an idea uses the partition method of the good image comparative-ly or it creates it.*

[Response message written by a Japanese member - message translated by a human translator]: *I can't understand what you are saying. Do you want to know what products are displayed in stores? For such purposes, ordinary web pages are well-suited. What I want to know is the stores located in shopping areas. Thus, I think Town Digitizing is well-suited.*

In the example above, a Japanese member is responding to the first message posted by a Chinese member. To understand what he really wanted to say, we translated the Chinese message into English. The second message is an automatically translated version of the Chinese message. In the experiment, the message was automatically translated into Japanese. However, to share an image of the automatically translated message in this paper, we further translated the Japanese message into English, referring to the automatically translated results of the Chinese message into English (Here, the machine translation quality from Chinese to English used was "Not Bad"). All Japanese members read the translated Japanese message. The third message is a response from a Japanese member. As for the first message, we translated the Japanese message into English.

From the first message, we see that the key interest of the Chinese member is using an image segmentation method. Investigating new products is one application where the method has been found useful. However, machine translation mistranslates the message. Consequently, the content is changed, suggesting that the Chinese member's main interest is to check out new commercial products and that image processing is one solution to perform that task. From the third message, we see that the Japanese member mistakenly thought that the Chinese member's main interest was to check out new commercial products.

## 2.3 Accumulation of Misconceptions

We found a large number of misconceptions between Chinese and Japanese discussion members like those introduced in the previous section. Counting the number of such misconceptions by hand (see section 4.4 for details), we found that approximately one out of twenty messages included misconceptions.

However, more than half of the participants (12 Japanese and eight Chinese) answered in English interviews that they could "often" understand the general outline of the translated messages (Table 2).

The accumulation of such misconceptions caused serious communication breakdowns. For example, the perceptions of the Chinese and Japanese members differed concerning what they thought they had accepted as conclusions of the one-month discussion. In our interviews, we asked both Chinese and Japanese members about such inconsistent conclusions, and they claimed that they did not agree with the conclusions of their counterparts. Interestingly, they did not even know that there was an "understanding gap" between the conclusions reached between Chinese and Japanese members.

**Table 2. Member evaluations of understandability of translated messages: how frequently they understood general outline**

| Understandability | Always | Often | Some-times | Rarely | Never |
|---|---|---|---|---|---|
| **# of Japanese members** | 0 | 12 | 4 | 2 | 0 |
| **# of Chinese members** | 0 | 8 | 4 | 4 | 0 |

## 2.4 Misconception Types

Based on our observation, we found that most misconceptions in Asia Broadband Project had different characteristics than those mainly treated in previous research.

Previous research has mainly focused on misconceptions occurring when a speaker falsely assumes that discussion members share mutual knowledge and proceed to leave much unstated. Misconceptions of this type are usually found in elliptic discourse, where people "believe" that they share a lot of knowledge. Most of the misconceptions in our daily lives occur in such a context. For example, while a speaker assumes that listeners know what is meant by a particular speech act, listeners may interpret the situation quite differently and carry out a different set of actions than what the speaker assumed [23]. In this paper, we refer to such misconceptions as "*unstated misconceptions.*"

Although most misconceptions in usual conversations occur from the unstated part, most misconceptions found in the Asia Broadband Project occurred from the *stated* part. Such misconceptions are a sort of mis-meaning between discussion members, typically occurring from clearly stated comments. For example, while a speaker is talking about T, listeners may think that the speaker is talking about something else. This type of misconception is rarely found in common conversation; even if it did occur, discussion members can detect it in subsequent conversations and correct it themselves. However, discussion members in the Asia Broadband project were unaware of the misconception, leading to serious communication breakdowns. Misconceptions of this type are usually caused by translation errors. In the following, we focus on misconceptions occurring from the stated part and refer to them as "*stated misconceptions.*"

# 3. MICRO-LEVEL ANALYSIS

In the Asia Broadband Project, we found massive amounts of stated misconceptions. They frequently occurred among Chinese and Japanese members. Since machine translation-mediated communication is one of the biggest discourse features of the Asia Broadband Project, insights about stated misconceptions may be gleaned by analyzing the characteristics of machine translation-mediated communication.

For this investigation, we first analyze conversations from the Asia Broadband Project at the micro level where the basic unit of analysis is response pairs. We analyze the response pattern of

machine translation-mediated communication and investigate how such patterns generated confusion among members. In the following, we call response pairs *parent-child pairs*, where child indicates a response to the parent.

## 3.1 Method

A basic assumption underlying our investigation is that people guess the meaning of others' messages based on words and guess how messages are related based on lexical cohesion between messages, especially when discussing via low-quality machine translation. Stated misconceptions occur when such guesses are wrong. Thus, insights about how machine translation generated stated misconceptions may be gleaned using lexical cohesion analysis [14].

Lexical cohesion analysis is also suitable for our research because the analytic process can be automated, since our goal is to develop an automatic technique that shows the tendency that creates stated misconceptions.

In this paper, we measured lexical cohesion between messages based on shared content words and synonyms using a Japanese thesaurus [27]; we focused on the Japanese version of each message. We refer to the content words and synonyms shared between messages as cohesive content words.

We first gathered message pairs from direct responses. Next, we divided the pairs into two groups based on whether parent and child are both posted from the same country (that is, Chinese-Chinese or Japanese-Japanese) or from different countries (that is, Chinese-Japanese or Japanese-Chinese). We compared the two groups' response patterns using lexical cohesion.

## 3.2 Responses of Low Relevance

We compared parent-child pairs posted from the same country and from different countries by the number of cohesive content words shared between parent-child pairs (Figure 1).

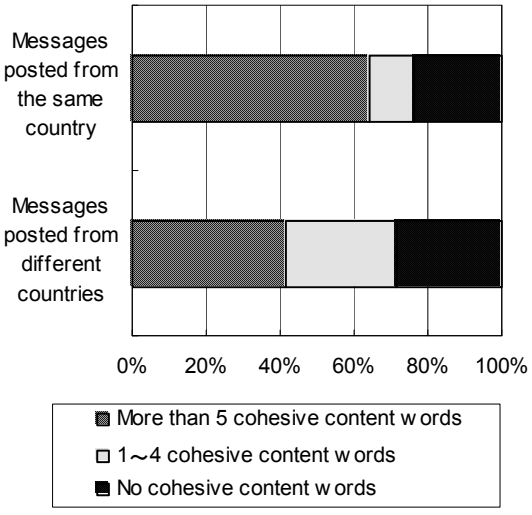


**Figure 1. Number of cohesive content words shared between parent-child pairs.**

Figure 1 indicates that the ratio of parent-child pairs where cohesive content words are not shared at all is similar (10 to 15%) to those posted by members from the same country and those posted by members from different countries. From further detailed analysis on such parent-child pairs, we found that the content of the child message tended to be simple, such as greetings, agreement, short comments, etc.

Meanwhile, the ratio of parent-child pairs including more than five cohesive content words differed significantly between those posted by members from the same country (around 60%) and those posted by members from different countries (around 40%). Further t tests proved that parent-child pairs posted from the same country share significantly more cohesive content words than parent-child pairs posted from different countries ($F=16.078$, $p=0.000$). Low cohesiveness between parent-child pairs posted from different countries indicates that the messages have low relevance [26].

From further analysis of such low cohesive parent-child pairs, we found that many child messages "trip" on the wording of the parent messages. Such responses tended to be incoherent and inadequate as a response, even if the parent-child messages shared some cohesive content words.

## 3.3 Responses Focusing on Short Phrases

In general, the translation quality of a sentence decreases as its length increases. Thus, it is difficult to fully understand translated messages when they contain long sentences. We expect that members especially respond to short phrases, since the translation quality of short phrases is relatively high and thus understandable.

To investigate our hypothesis, we compared parent-child pairs posted from the same country and from different countries by the distribution of the parent message's length of sentences; the sentence and the child message share cohesive content words.

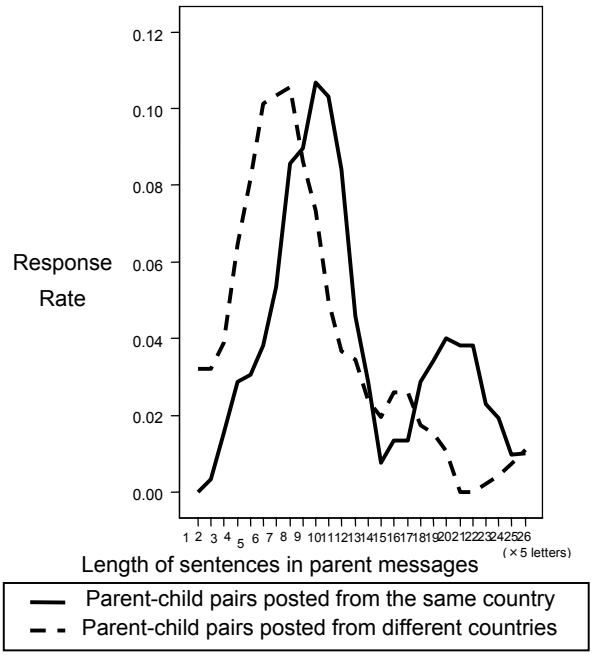


**Figure 2. Distribution of responses to parent messages**

Figure 2 shows that when a sentence is too long in a message, responses to the message from other countries tend "not" to mention the long sentence. Further t tests proved that the average

length of sentences where the sentence and the child message share cohesive content words is significantly different between the two groups; response messages from members in other countries are significantly more focused on short phrases of its parent message than the response messages of members from the same country (F=4.816, p=0.029).

The result supports our hypothesis. Members tended to respond only to short phrases that are easy to understand.

# 4. MACRO-LEVEL ANALYSIS

From the above analysis, we showed that parent-child pairs posted from different countries via machine translation have low cohesion, and such child messages tend to be incoherent and only partially connected to the parent messages. Since such inadequate responses reflect the danger of misconception, we may be able to assess the tendency of stated misconception using the strength of coherency between parent-child messages; when coherency between parent-child messages is weak, stated misconceptions tend to occur.

However, merely measuring coherency between parent-child messages is insufficient to assess the tendency of stated misconceptions. For example, as we have seen in Section 3.2, lexical cohesion between parent-child messages is also weak when the content of a response message is just a greeting or a few simple comments. Such responses are coherent and appropriate as a response and thus not a source of misconception.

To overcome this weakness, we expand the unit of analysis and consider ways to distinguish coherent and incoherent responses when lexical cohesion between parent-child messages is weak. The basic unit of analysis in this section is an entire dialogue.

## 4.1 Syntactic and Semantic Threads

We focus on discussion threads in each dialogue. A discussion thread indicates a series of postings and responses (and responses to responses) on a common subject [19]. Discussion threads have been widely used for expressing discussion flows, and many ways to construct discussion threads have been previously proposed. In this paper, we consider two types of previously proposed discussion threads: *syntactic* and *semantic*.

A syntactic thread is a discussion thread constructed on references in "References" and/or "In-Reply-To" fields of email standards. Discussion flows are usually constructed according to syntactic threads.

A semantic thread is a discussion thread that typically uses lexical cohesion to reconstruct actual discussion flow. Many ways to construct discussion flows using lexical cohesion have been proposed [18, 20, 11]. Semantic threads reflect actual discussion flows better than syntactic threads [18].

## 4.2 Gaps between Syntactic and Semantic Threads

To distinguish between coherent and incoherent responses when lexical cohesion between parent-child messages is weak, we focus on the gaps between syntactic and semantic threads. Since response pairs in semantic threads reflect actual responses, we consider a response pair incoherent when its lexical cohesion in a syntactic thread is low and also when it is not a response pair in a semantic thread.

Gaps between syntactic and semantic threads may be interpreted as follows. Response links in syntactic threads represent the intention of those who posted the message concerning that to which they intend to respond[3]. Meanwhile, a semantic thread is automatically generated from the messages, and thus response links in semantic threads may represent readers' (or other discussion members') impressions about which pairs are response pairs. Thus, a gap between syntactic and semantic threads represents a mismatch between intention and understanding.

Since stated misconceptions occur when members misunderstand the intention of the message poster, misconceptions occur when there is a gap between syntactic and semantic threads.

A large gap between syntactic and semantic threads indicates many mismatches between members of which they are not aware. Thus, conversations with large gaps between syntactic and semantic threads tend to include many stated misconceptions, resulting in serious communication breakdowns.

## 4.3 Measuring the Gaps

Based on the previous discussion, we propose a method that automatically predicts the occurrence of stated misconceptions in dialogues.

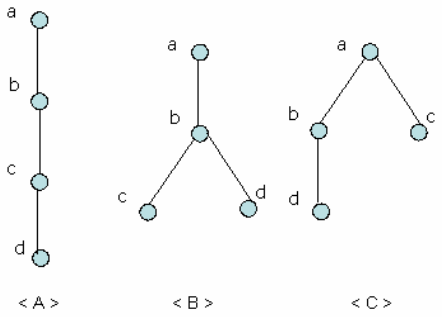


**Figure 3. Examples of syntactic and semantic threads**

One way to measure gaps between threads is to calculate their edit distance. For example, consider a case where <A> is a syntactic thread and <B> and <C> are semantic threads, as in Figure 3. Parent-child pairs in <A> are *ab*, *bc*, and *cd*, whereas parent-child pairs in <B> are *ab*, *bc*, and *bd* and parent-child pairs in <C> are *ab*, *ac*, and *bd*. Common parent-child pairs <A> and <B> share *ab* and *bc*, while parent-child pairs that belong to either <A> or <B> are *cd* and *bd*. Similarly, common parent-child pairs <A> and <C> share *ab* while parent-child pairs that belong to either <A> or <C> are *bc*, *cd*, *ac*, and *bd*. Since <A> and <B> share two parent-child pairs while <A> and <C> have one parent-child pairs in common, the gap between <A> and <C> is larger than <A> and <B>.

In the following, we present an algorithm to calculate the gap between syntactic and semantic threads of dialogue with $n$ messages, followed with a detailed explanation of the algorithm.

[3] When discussion members are pressing toward a unique goal on BBS or mailing lists (as in the Linux mailing lists), they pay attention to the indication of which message they respond to, so that others can easily track the discussion (Yamauchi 2000).

Step 0 [Preparation]: Label each message with an ID so that "message $i$" is the $i$ th message posted in time sequence.

Step 1 [Construct a Syntactic Thread]: Define the relation between messages $i$ and $j$ in a syntactic thread as

$$x_{ij} = \begin{cases} 1 : \textit{when messages } i \textit{ and } j \textit{ are response pairs in syntactic thread}, \\ 0 : \textit{otherwise}. \end{cases}$$

Step 2 [Construct a Semantic Thread]: Define the relations between messages $i$ and $j$ in a semantic thread, $y_{ij}$, from the following procedure.

Step 2-0. Initialization: Set a threshold $\alpha$. For all $(i, j)$, set $y_{ij} = 0$.

Step 2-1. For all $(i, j)$ that satisfies $x_{ij} = 1$, execute the following:

Step 2-1-0. If lexical cohesion between $(i, j)$ is stronger or equal to $\alpha$, then $y_{ij} = 1$.

Step 2-1-1. If lexical cohesion between $(i, j)$ is weaker than $\alpha$, then execute the following:

For message $k$ ($1 \le k < j$), examine lexical cohesion between $(k, j)$.

Step 2-1-1-0. If there are multiple messages $k$ where lexical cohesion between $(k, j)$ is stronger or equal to $\alpha$, then choose the most recent message $k$, which has been posted before $j$. Set $y_{kj} = 1$.

Step 2-1-1-1. If such message $k$ was not found, then $y_{ij} = 1$.

Step 3 [Calculate the Gap between Syntactic and Semantic Threads]: Calculate the gaps between a syntactic thread and a semantic thread as

$$G = \frac{\sum |x_{ij} - y_{ij}|}{2(n-1)}.$$

Step 2 is the procedure to construct a semantic thread. Parent-child pairs in semantic threads must be coherent and strongly related by content. Usually, messages are strongly related when their lexical cohesion is high [26]. However, as previously explained, coherent responses exist even when lexical cohesions are weak. Thus, we decided to combine evidence from syntactic threads and lexical cohesion to construct a semantic thread. Concretely, we infer that messages $i$ and $j$ are response pairs if they are response pairs in a syntactic thread and their lexical cohesion is strong. However, if lexical cohesion between the messages ($i$ and $j$) is weak, we search for a more appropriate message $k$ for message $j$'s parent. If we find a message $k$ whose lexical cohesion with message $j$ is strong, then we consider that message $k$ is a more appropriate parent than message $i$. However, if we cannot find such a $k$, then we infer that messages $i$ and $j$ are response pairs even if their lexical cohesion is weak; when message $j$ is a simple comment or agreement on message $i$, then not only the lexical cohesion between messages $i$ and $j$ but also message $j$ and any other messages posted before message $j$ are weak.

For threshold $\alpha$, we need to select a value at which it is appropriate to infer that message pairs are strongly related when their lexical cohesion is stronger than $\alpha$. To select such a value, it might be helpful to use examples from lexical cohesion between response pairs posted from the same country, since such response pairs posted from the same country and strong cohesion are always strongly related. For example, in the Asia Broadband Project, more than five cohesive content words existed between response pairs posted from the same country and with strong cohesion. Thus, in this case, we set $\alpha = 5$.

In Step 3, we calculate the gap between syntactic and semantic threads. $G$ takes a value from 0 to 1; syntactic and semantic threads are identical when $G = 0$; all response pairs differ between syntactic and semantic threads when $G = 1$. For example, in Figure 3, $G = 2/(2*3) = 1/3$ between <A> and <B>, and $G = 4/(2*3) = 2/3$ between <A> and <C>.

## 4.4 Verification

We expect $G$ to take a large value when a conversation is riddled with stated misconceptions.

- Verification Data: Messages from the Asia Broadband Project were used for verification.
- Verification Procedure:
  (1) We first collected dialogues that simultaneously satisfy the following two conditions: a) its syntactic thread includes more than fifteen messages; b) its syntactic thread includes more than five messages posted from Chinese members and more than five messages posted from Japanese members.
  (2) Next, for each dialogue collected from Procedure (1), we calculated the *misconception frequency*, which in a dialogue is calculated by counting the number of stated misconceptions in the dialogue and dividing by the total number of parent-child pairs of its syntactic thread.
  (3) Third, for each dialogue collected from Procedure (1), we calculated $G$.
  (4) Last, we examined the correlation between misconception frequency and $G$.

For Procedure (1), we considered that conditions a) and b) are rough indications of dense discussions between Chinese and Japanese members. Twelve dialogues were extracted that satisfied these conditions. The minimum number of messages in the dialogues was 15, and the maximum number was 52.

In Procedure (2), one author counted the number of misconceptions in each dialog by hand, using the criteria given in the literature [29]. For example, when a child message does not include "acceptance/refusal" content when the parent message is

about "request," then there is a misconception between the parent-child pairs. From this procedure, 7 misconceptions were found between parent-child pairs posted from the same country, and 32 misconceptions were found between messages posted from different countries.

- Verification Results:

Figure 4 indicates the correlation between *G* and misconception frequency. Each dot in the figure indicates a dialogue.

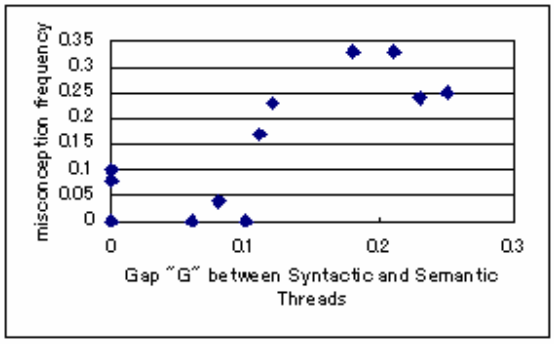


**Figure 4. Correlation between "G" and misconception frequency**

From Figure 4, we see that misconception frequency rises as *G* takes a larger value. Spearman's correlation between the two values appeared significantly positive (r=0.785, p=0.003).

## 5. CONCLUSIONS AND FUTURE WORK

Mutual understanding is largely influenced by communication media [5, 6, 7], and a number of studies have shown that reaching mutual understanding is more difficult via computer-mediated communications systems than in face-to-face settings [10, 13]. Reaching mutual understanding over lean media using low-quality machine translation is a difficult task; in particular, machine translation poses confirmation hurdles for discussion members receiving information and understanding it correctly. When members do not fully understand others' comments, they tend to speak and understand on the basis of their own information and interpretation of the situation, falsely assuming that the other speaks and understands on the basis of that same information and interpretation [3].

Based on our experimental research, we offer two key practical lessons on misconceptions occurring over machine translation-mediated communication.

- Response messages via low-quality machine translations tend to be incoherent. Such incoherent response messages tend to be only partly related to the original message, and they focus on the short phrases of the parent message.
- The gaps between discussion threads based on header information and lexical cohesion represent a mismatch between discussion members.

By combining these findings, we proposed a method that automatically predicts the occurrence of misconceptions in each dialogue. We conducted verification experiments using actual conversations; verification results show that the correlation between actual misconception frequency and syntax-semantics gaps are significantly and positively correlated, indicating that the method is valid.

Since we only analyzed the Japanese version of each message, we need to conduct the same analysis for the Chinese versions as well. We predict that we will attain similar results, since most commercially available machine translation systems do not automatically omit or complement content words; every content word tends to appear in the translated sentence because most machine translation systems translate each sentence separately without considering context. As long as machine translations do not automatically omit or complement content words, our method should be valid in machine translation-mediated communication.

The validity of our method also lies in the phenomenon that users iterate the same wordings over and over when a large amount of noise gets mixed into conversations. The same phenomenon (people repeating the same words over and over) is also found in cell phone conversations when loud noise gets mixed in conversations [29]. Thus, the method may be also valid in predicting misconceptions in such hostile environments. Conversely, in chat-like communications where people tend to avoid overlaps and omit words [15], our method of using lexical cohesion may not help infer relations between messages.

This study's main limitation is the lack of verification: it was only conducted on a small data set. The proposed method needs further verification on a larger and different set of data. We also need to investigate how much information the method can actually contribute to the participants compared to what they can deduce themselves during attempts to comprehend their discussion partners' utterances.

Our next step is to overcome limitations by expanding the proposed method into a system that can alert discussion members to the existence of misconceptions. Offering a reliable predictor of misconception to all participants may help make coherent conversation easier to navigate. By alerting members about the occurrence of misconceptions, they may work toward fixing them. We will continue to examine the process of how discussion members actually fix misconceptions.

## 6. ACKNOWLEDGMENTS

We would like to thank Keiji Hirata for his contributions. The form of the paper itself greatly benefited from discussions with him. We also acknowledge the careful comments of Kumiyo Nakakoji. Most importantly, we thank Tomoko Koda for organizing the Asia Broadband Project and providing us with log data.

## 7. 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] Blakar, R. Toward a theory of communication in terms of preconditions: A conceptual framework and some empirical explorations. *Recent Advances in Language, Communication, and Social Psychology*, 1985, 10-40.

[4] Churchill, E. F. and Bly, S. Cultural Vultures: Considering Culture and Communication in Virtual Environments. *Proceedings of ACM SIGGROUP*, 21, 1, 2000.

[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] Clark, H. H. and Brennan, S. E., Grounding in communication. *Perspectives on socially shared cognition,* 1991, 127-149.

[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] Danet, B. and Herring, S.C. Introduction: The Multilingual Internet, *Journal of Computer Mediated Communication*, vol. 9, no. 1, 2003.

[10] Doherty-Sneddon, G., Anderson, A. H., O'Malley, C., Langton, S., Garrod, S. and Bruce, V. Face-to-face and video mediated communication: a comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied*, 3, 1997, 105-125.

[11] Donath, J. A Semantic Approach to Visualizing Online Conversations. *Communications of the ACM*, 45, 4, 2002.

[12] Fussell, S. R. and Krauss, R. M. Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, 62, 1992, 378-391.

[13] Fussell, S. R., Kraut, R. E. and Siegel, J. Coordination of communication: Effects of shared visual context on collaborative work, *Proceedings of CSCW*, ACM Press, 2000, 21-30.

[14] Halliday, M.A.K. and Hasan, R. Cohesion in English, Longman, London, 1976.

[15] Herring, S. C. Interactional Coherence in CMC, *Proceedings of the Hawaii International Conference on Systems Sciences*, 1999.

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

[17] Kraut, R. E., S. Lewis, S., and Swezey, L. Listener responsiveness and the coordination of conversation. *Journal of Personality and Sociological Psychology*, 43, 1982, 718-731.

[18] Lewis, D. D. and Knowles, K. A. Threading Electronic Mail: A Preliminary Study. *Info. Proceeding and Management*, 33, 1, Elsevier, 1997, 209-217.

[19] McDaniel, S., Olson, G., and Magree, J. Identifying and analyzing multiple threads in computer-mediated and face-to-face conversations. *Proceedings of CSCW*, ACM Press, 1996, 39-47.

[20] Newman, P. Exploring discussion lists: Steps and directions. *Proceedings of JCDL 2002*, ACM Press, 2002, 126-134.

[21] Rogers, Y. Ghosts in the Network: Distributed Troubleshooting in a Shared Working Environment. *Proceedings of CSCW*, ACM Press, 1992.

[22] Rogers, Y. Coordinating computer-mediated work, *Computer Supported Cooperative Work*, vol. 1, pp. 295-315, 1993.

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

[24] 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.

[25] Yamauchi, Y., Yokozawa, M., Shinohara, T. and Ishida, T. Collaboration with Lean Media: How Open-Source Software Succeeds. *Proceedings of CSCW*, ACM Press, 2000, 329-338.

[26] Grice, H. P. Logic and conversation. *In Cole, P. and Morgan*, 1975, 41-58.

[27] Ohno, F, Hamanishi, T, Kadokawa Thesaurus, Kadokawa Shoten, 1981.

[28] Schegloff, E. A. and Sacks, H. Opening up closings. *Semotica, 8*, 1973, 289-327.

[29] Yoshida, Y. and Ohga, J. Effect of S/N ratio on comfortableness of telephone dialog. *Acoustic Conference,* 2004.