

# The Role of Annotation in Intercultural Communication

Tomohiro Shigenobu<sup>1</sup>, Kunikazu Fujii<sup>2</sup>, and Takashi Yoshino<sup>3</sup>

<sup>1</sup> Language Grid Project, National Institute of Information and Communications Technology, 3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan  
shigenobu@nict.go.jp

<sup>2</sup> Graduate School of Systems Engineering, Wakayama University,  
930 Sakaedani, Wakayama, Japan  
s075065@sys.wakayama-u.ac.jp

<sup>3</sup> Faculty of Systems Engineering, Wakayama University,  
930 Sakaedani, Wakayama, Japan  
yoshino@sys.wakayama-u.ac.jp

**Abstract.** In intercultural communication, there are large barriers when the languages and the cultures are different. It is undoubtedly preferable for people to have smooth communications using their mother language. Therefore, we have developed a chat system called AnnoChat. AnnoChat has an annotation function for smooth intercultural communications. We applied AnnoChat in experiments with Japanese, Chinese, and Korean speakers. The results of the experiments showed that about 70% of the added annotations were reusable as intercultural knowledge information. About 20% of the added annotations were used to supplement information that could not be described while chatting. It is thought to be an effective example of applying annotation in intercultural communications.

**Keywords:** Intercultural Communication, Machine Translation, Annotation, Computer-Mediated Communication.

## 1 Introduction

Opportunities for intercultural communication are increasing due to the spread of the Internet. The number of Internet users in Southeast Asia keeps on increasing, and about 65% of the users are non-English speakers<sup>1</sup>. Mutual understanding of the partner's language is the largest barrier to intercultural communication, since the users have a variety of mother tongues.

If a pair of users speaks different mother languages, they try to communicate using a mutually understandable language such as English, but this is seldom satisfactory. Therefore, collaborative work under tends to be ineffective [1,2]. Machine translation systems are an effective solution. However, the accuracy of most machine translation systems is not sufficient, and misunderstanding due to mistranslation is common [3].

---

<sup>1</sup> Global Reach: <http://global-reach.biz/globstats/>

Although machine translation lacks complete accuracy, imperfect translation may be acceptable if the messages can be understood. Actually, several communities are now communicating via machine translation<sup>2</sup>, and such communities are expected to increase [4].

We think that besides communication using a common language, a method is necessary that will allow people from different countries to communicate easily. Machine translation technology has been developed as a way to overcome language differences. With the continued advances in machine translation technology, highly accurate translations have been achieved in specific fields. If translation accuracy were excellent, mutual communication through machine translation would be possible [5]. Mail, chat, and bulletin board systems using machine translation have been developed to support intercultural communication. AmiChat is a chat system that can translate chat messages into other languages through machine translation [6]. AmiChat has a machine translation engine that can translate two or more languages. This system can display the original message entered by the user as well as the message translated into two or more languages at the same time. TransBBS is a bulletin board system using machine translation and is utilized as a daily discussion space [7]. This tool provides translation services in Chinese, Japanese, Korean, Malay, and English. This system was used as a communication tool when researchers in Asia jointly developed software as one of the experiments at ICE2002 (Intercultural Collaboration Experiment 2002<sup>3</sup>).

An accurate translation result is generally not obtained in intercultural communications that handle a variety of messages. If people are not familiar with the other person's cultural background, there may be a lack of understanding about a message. Proper nouns, adjectives, etc., often contain culturally-specific meanings. To deal with such cultural problems, annotations added to culturally-specific words and phrases might help to improve mutual understanding. Various systems that aim to accumulate and share knowledge information have been developed [8,9,10]. Users create annotations that are shared as knowledge information in the systems. These annotations are often asynchronously applied to static documents. However, in multilingual communication, the effectiveness and availability of annotations have not been sufficiently examined.

We have developed a multilingual chat system with an annotation function and have applied it in intercultural communication situations. This system supports communication between people from different countries who each use only their native language. However, because machine translation cannot completely prevent mistranslations, this system has a back translation function to improve the machine translation output. In this paper, we describe the effectiveness and availability of annotations in intercultural communication.

## 2 Multilingual Chat Tool AnnoChat

### 2.1 Design Policy

To support intercultural communication, we have developed a chat system called AnnoChat that has a function to create annotation data. The design policy of this system is shown below.

---

<sup>2</sup> Enjoy Korea: <http://www.enjoykorea.jp/>

<sup>3</sup> Intercultural Collaboration Experiment: <http://www.ai.soc.i.kyoto-u.ac.jp/ice/>

1. Supporting function of annotating to words or phrases

In intercultural communication, people with different cultural backgrounds communicate with each other. Therefore, even if a message is translated accurately and formally, people may not have the same understanding of its meaning. Proper nouns, adjectives, etc., often contain culturally-specific meanings. If a user does not know the other person's cultural background, it may result in a lack of understanding about that person's message. We should consider the possibility that words and phrases in a given message can be understood differently. Therefore, a function to add an appropriate meaning as an annotation is necessary for intercultural communication. We think that a function to add annotations to arbitrary words and phrases of a message will increase the user's understanding of machine-translation-dependent communication.

2. Supporting function of inputting a machine-translatable message

Chat messages often contain typographical errors, omissions, or euphemistic expressions. Also, some spoken languages omit the subject of the sentence, and these languages are often used in chatting online. It is difficult to translate such messages accurately using existing machine translation technology. Users can confirm whether or not the translation has succeeded by showing the translated result in their mother language before sending the message. If the translated result is not good, the user can revise it to come up with a message that is suitable for machine translation. For example, if a Japanese person writes a message to a Korean person, this function first translates the Japanese message into Korean and then translates the result back into Japanese. This method enables the translation accuracy to be confirmed in the input language. Back translation allows a user to write a sentence that is machine-translatable using only the user's mother language [11,12].

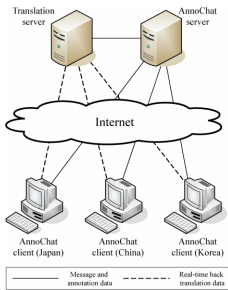


Fig. 1. System configuration of AnnoChat

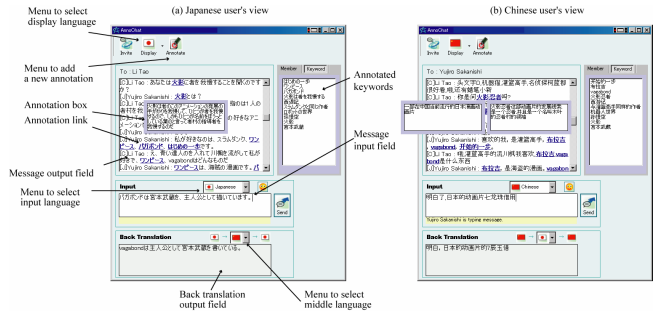


Fig. 2. An example screen of AnnoChat client

2.2 System Configuration

This system is a client-server model; the server consists of an AnnoChat server and a machine translation server (Fig. 1). The AnnoChat server receives data of messages and annotations from the AnnoChat client, and the data is translated into the other languages by machine translation. Then, the server sends the translated data to all participants of the same session. The data of the real-time back translation is communicated directly between the AnnoChat client and the translation server. The AnnoChat

server stores the message and the annotation as log data. We have used J-Server<sup>4</sup>, which was developed by Kodensha, as a machine translation server. J-Server is available to translate between Japanese and Chinese, Japanese and Korean, and Japanese and English. We executed multi-hop translation through Japanese for languages that could not be translated directly (e.g., between Chinese and Korean).

### 2.3 Functions of AnnoChat

AnnoChat has a multilingual input and display, a real-time back translation function, and an annotation function. The operation procedure and screen layout are like that of an instant messenger. A sample screen of an AnnoChat client is shown in Fig. 2. Details of the functions are below.

#### 1. Multilingual input and display

When the button to select the display language is pushed, the available language list is displayed as a menu. Displayed messages, the annotated keyword list, and the content of annotations are switched to the selected language when a user selects an arbitrary language. If a message input field is empty, selection of the input language is also switched to the selected language. Fig. 2 depicts a screenshot from a chat session between a Japanese user (a) and a Chinese user (b).

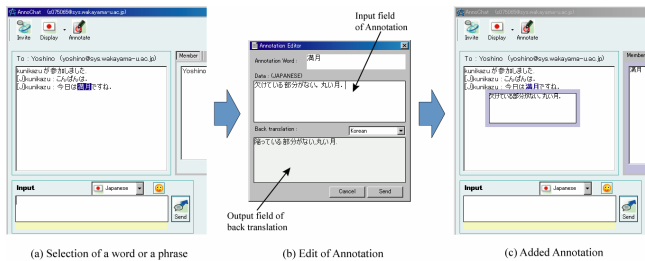


Fig. 3. Procedure for editing Annotation

#### 2. Real-time back translation function

The client executes the back translation by using the message entered by the user at intervals of a few seconds. The result is displayed in the back translation output field. A user corrects the original message while confirming the result of the back translation displayed in the user's mother language.

#### 3. Annotation function

The underlined bold text in the message output field shows the existence of an annotation, called an Annotation link. When the user scrolls over the Annotation link with the mouse cursor, the content of the annotation is shown in the current display language in the Annotation box. Additionally, the annotation edit window is displayed by clicking the Annotation box, and a user can edit the annotation content. The procedure of creating a new annotation is shown below (Fig. 3).

<sup>4</sup> KODENSHA <http://www.kodensha.jp/>

- (a) The user selects a word or phrase that he/she wants to create an annotation for and clicks the button to create a new annotation.
- (b) The user enters a detailed explanation of a word or a phrase in the annotation edit window. The user can repair a sentence by referring to the back translation result.
- (c) The user clicks the send button, and the annotation data is delivered to all users.

One word or phrase often has two or more meanings. Thus, we thought that it was better to be able to create two or more annotations if necessary. This system makes it possible to create some annotations for arbitrary single words or single phrases in a chat message. Additionally, the annotation function in this system displays the annotation links to all words and phrases that appear when chatting. Words, phrases, and the content of annotations created with the AnnoChat client are translated into each language and delivered to all participants through the server.

### 3 Experiments and Results

#### 3.1 Experimental Procedure

We applied the developed system in intercultural communication and examined the annotations created at the initiative of participants. The experiment was carried out as part of Intercultural Collaboration Experiment 2005 (ICE2005<sup>5</sup>) that research institutes from five Asian countries (China, Korea, Thailand, Malaysia, and Japan) jointly hosted in 2005. Participants in this experiment were Japanese, Korean, and Chinese. In each experiment, the participants chatted with a partner one-on-one. Chat communication experiments were carried out with 19 pairs: three pairs were Japanese and Korean; six pairs were Chinese and Japanese; and ten pairs were Korean and Chinese. The number of experiments was different with each combination because the number of participants at ICE2005 varied in each country. Participants were undergraduate and graduate students, and they had no personal acquaintance with each other. The experimental task was as follows.

1. Participants decided the chat theme mutually.
2. Participants mutually communicated a culture concerning the theme for 20 minutes using their native language.
3. After the chat task, participants created five annotations for their own messages.
4. After the experiments, participants answered a questionnaire.

In this experiment, participants created annotations after 20 minutes of chatting. Normally, it is more natural to create annotations in parallel with the communication. However, to create the annotations as a controlled experiment, we thought they should be done after some messages had been accumulated. We prepared simple themes for chatting that did not need any domain knowledge about a specific field. The participants chose one from among the six prepared themes (popular destinations for tourists, introduction of local specialty foods, etc.). We prepared the questionnaire form in English and Japanese. The participants wrote their answers in their mother tongue or in English in the description column.

---

<sup>5</sup> <http://ice.kuis.kyoto-u.ac.jp/ice/ice2005.htm>

In this experiment, we installed the translation server and the AnnoChat server at Wakayama University in Japan. Each client accessed both servers through the Internet.

**Table 1.** Classification of annotations

Classification type	The number per experiment			Ratio (%)
	Japanese/ Korean	Japanese/ Chinese	Korean/ Chinese	
Dictionary	8.3	8.2	6.5	73.5
Conversation supplementation	1.7	1.0	2.0	16.4
Translation confirmation	0.0	0.7	1.5	10.1

### 3.2 Results of Annotation

We examined the details of the annotations by using the annotation data created in the experiments, and the questionnaire results. Table 1 shows classified results. The investigated results are shown below.

#### (1) Classification of annotations

We classified the annotations created by the experiments into various usage types based on the meanings of words or phrases, the content of the annotations, and the questionnaire results. These different usage types can be used like a dictionary to explain aspects of intercultural communications. Table 2 shows examples of annotation data in experiments.

##### (a) Dictionary type

The participant can understand the meaning even when the annotation is added to the same words and phrases in other communications.

##### (b) Conversation supplementation type

The participant cannot understand the meaning even if the annotation is added to the same words and phrases in other communications. This annotation was created to supplement the content of the chat. In other words, this is dependent on the context.

##### (c) Translation confirmation type

This annotation does not actually explain the content of words and phrases. The annotation is created for the reason “I have not understood the meaning of words and phrases.” The participant was asking the other party the meaning of words and phrases because he/she was not able to understand the result of the translation.

#### (2) Ratio of concordance between created annotations and requested annotations

The participants created five annotations in each experiment. The criterion for selecting them was that the user felt they were words or phrases that the partner would not understand. We investigated the difference between the created annotations and the requested annotations. The method we used to evaluate the requested annotations was a questionnaire survey after the experiments. The participants selected five words and

phrases that needed annotations in order to understand their partner’s message accurately. The selected words and phrases were the requested annotations. Table 3 shows the relationship between the created annotations and the requested annotations.

**Table 2.** Classification example of annotations

Type	Word or phrase	Content
Dictionary	Nami Island	This site is place-name that had filmed representative drama “Winter Sonata” in Korea.
	One Piece	This is a comic and features pirates.
Conversation supplementation	I watch TV on a mobile phone.	A new service of mobile communication service carriers. Users can watch TV programs on mobile phone.
	Man district	Comic book. (* Participant corrected a typographical error.)
Translation confirmation	World edition	I can’t understand this meaning of a phrase.
	Listen	This meaning of word is “listening,” but this is not right contextually.

**Table 3.** Ratio of concordance between the created annotations and the requested annotations

	Japanese/ Korean	Japanese/ Chinese	Korean/ Chinese	Total
Number of created annotations	30	59	100	189
Number of requested annotations	30	60	100	190
Number of concordance	12	21	30	63
Ratio of concordance (%)	40.0	35.6	30.0	33.3

## 4 Discussion

### 4.1 Effectiveness of Annotations

In the questionnaire after the experiments, we questioned the participants on whether the annotations were useful for mutual understanding of the message. Table 4 shows the results of the questionnaire. Participants evaluated each item on the questionnaire on a 5-point scale (1: Strongly disagree, 2: Disagree, 3: Neutral, 4: Agree, 5: Strongly agree). Each value shows an average rating. The results are given for each pair combination (Japanese/Korean, Japanese/Chinese, and Korean/Chinese). Table 5 shows the participants’ comments about the annotation function.

The participants evaluated annotations highly that were effective in helping them understand the message (Japanese/Korean pair: 3.8; Japanese/Chinese pair: 4.5; Chinese/Korean pair: 3.7). In the participants’ impressions of the annotation function, a lot of affirmative comments such as, “The annotation was useful for understanding the message” were indicated. However, some participants expressed the opinion “Because the problem is solved by the conversation, the annotation is unnecessary.”

**Table 4.** Results of questionnaire survey

Questionnaire item	Japanese/ Korean	Japanese/ Chinese	Korean/ Chinese
(1) I could communicate smoothly with a partner.	4.0	3.3	2.5
(2) I could understand a partner's message through the machine translation like own mother language.	3.8	2.8	2.0
(3) I think that annotation is useful to understand a message mutually.	3.8	4.5	3.7
(4) I think that annotating task is difficult.	2.0	2.8	2.3

**Table 5.** Comments about annotation function

– I want to create annotation while communicating without creating annotation after chat. (Japanese)
– When the annotation was added to proper noun of partner's country, I can understand the unknown thing easily. (Japanese)
– The annotation is useful. It is necessary for various users' communications. (Chinese)
– I thought that I could use the system easily. I can understand the content well by using the annotation. (Chinese)
– I can explain in the conversation. Should I explain by using annotation? (Korean)
– Because the translation quality was bad, annotations were useless. (Korean)

We expect that many trifling questions are generated in intercultural communication. A lot of questions in a conversation hinder smooth communications.

The classification result of the annotations showed that 16% of the created annotations were the conversation supplementation type for supplementing the content of communication (Table 1). We think that the users use the annotation function when necessary to help make the communications smooth.

## 4.2 Details of Annotation

In this experiment, we entrusted the object of the annotations to the participants' judgment. The evaluation value of the difficulty of the annotation-making process was less than three points (Table 4). In other words, the participants did not have the impression of difficulty. In the annotation classification, about 70% of the given annotations were the "dictionary type" that explain certain words and phrases as in a dictionary. Annotations of this type can be shared by accumulating them in the server. For instance, when the same words and phrases are input in another chat, the system can automatically display the annotation. We believe that these are useful annotations that are reusable as knowledge information.

About 20% of the created annotations were the "conversation supplementation type." Because some individual participants had felt that explanations given in the conversation itself were insufficient, they had to supplement the explanation. In other words, the annotation was used as a "second chat channel." We think that it is natural to make such chat supplementations while chatting. This result shows that the annotation function is effective in multilingual communications.

In this experiment, we entrusted the object of the annotations to the participants' judgment. The ratio of which the created annotations and requested annotations correspond was about 30% (Table 3). In intercultural communication, we confirmed that there was a big difference in the annotations that are mutually demanded. Thus, a new function to demand the addition of the annotation is necessary.

### 4.3 Communication Through Translation

We conducted a questionnaire survey to examine whether or not participants were able to communicate smoothly. Message understanding was evaluated highly for machine-translated messages between the Japanese and Korean languages (Table 4-(1), (2)). In the evaluation of machine translation accuracy at ICE2002 [7], a Japanese and Korean pair received the best evaluation. The evaluation of messages translated to and from Japanese and Chinese was also good. The translation accuracy of the Chinese/Korean messages was the worst. This is because the translations between Chinese and Korean used multi-hop translation through Japanese, which reduced the translation accuracy. The machine translation engine used in this experiment was the same as that used at ICE2002. Essentially, smooth communication depends on the accuracy of the machine translation. Therefore, communication supported by annotations is important because annotations enhance the translation accuracy.

## 5 Conclusion

In this paper, we described the multilingual chat tool called AnnoChat, which was developed for intercultural communication using machine translation. We have experimented in adding annotations to words and phrases communicated in messages in Chinese, Korean, and Japanese. In the classification of created annotations, about 70% of all annotations were the "dictionary type" that explain certain words and phrases as in a dictionary. This type can be used like a dictionary to explain unfamiliar phrases in intercultural communication. About 20% of all annotations were used to supplement information that could not be described while chatting.

We have confirmed the possibility of which the annotation makes smooth communications from participants' impression. In the future, we will support multilingual communications by improving the usability of the annotation.

## References

1. Tung, L.L., Quaddus, M.A.: Cultural differences explaining the differences in results in GSS: implications for the next decade. *Decision Support Systems* 33(2), 177–199 (2002)
2. Takano, Y., Noda, A.: A temporary decline of thinking ability during foreign language processing. *Journal of Cross-Cultural Psychology* 24, 445–462 (1993)
3. Yamashita, N., Ishida, T.: Automatic Prediction of Misconceptions in Multilingual Computer-Mediated Communication. *International Conference on Intelligent User Interfaces (IUI-06)*, pp. 62–69 (2006)

4. Salvador, C., Joaquim, M., Antoni, O., Miriam, S., Imma, S., Mariona, T., Lluisa, V.: Bilingual Newsgroups in Catalonia: A Challenge for Machine Translation. *Journal of Computer Mediated Communication*, 9, 1 (2003)
5. Milam, A.: Multilingual Communication in Electronic Meetings. *ACM SIGGROUP, Bulletin*, 23, 1 (2002)
6. Flournoy, R.S., Callison-Burch, C.: Secondary Benefits of Feedback and User Interaction in Machine Translation Tools, Workshop paper for “MT2010: Towards a Roadmap for MT” of the MT Summit VIII (2001)
7. Nomura, S., Ishida, T., Yamashita, N., Yasuoka, M., Funakoshi, K.: Open Source Software Development with Your Mother Language: Intercultural Collaboration Experiment 2002. *International Conference on Human-Computer Interaction (HCI-03)* 4, 1163–1167 (2003)
8. Wojahn, P.G., Neuwirth, C.M., Bullock, B.: Effects of interfaces for annotation on communication in a collaborative task. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 456–463 (1998)
9. Weng, C., Gennari, J.H.: Asynchronous Collaborative Writing through Annotations. In: *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pp. 578–581 (2004)
10. Cadiz, J.J., Gupta, A., Grudin, J.: Using Web Annotations for Asynchronous Collaboration Around Documents. In: *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pp. 309–318 (2000)
11. Yokoyama, S., Kashioka, H., Kumano, A., Matsudaira, M., Shiokizawa, Y., Kodama, S., Ehara, T., Miyazawa, S., Murata, Y.: An Automatic Evaluation Method for Machine Translation using Two way MT. In: *Proceedings of the 8th MT Summit Conference* (2001)
12. Frederking, R.E., Black, A.W., Brown, John Moody, R.D., Steinbrecher, E.: Field Testing the Tongues Speech-to-Speech Machine Translation System. In: *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)* (2002)