Detecting the Untranslatable Colloquial Expressions of Japanese Verbs in Cross-Language Instant Messaging

Yuchang Cheng[1], Masaru Fuji[1], Tomoki Nagase[1], Minoru Uegaki[2], Isaac Okada[2]
[1]Speech & Language Technologies lab.,
Media Processing Systems Laboratories, Fujitsu Laboratories Limited., Kawasaki, Japan
[2]System Engineering Knowledge Improvement div.,
System Engineering Technology Unit, Fujitsu Limited., Tokyo, Japan
[1]{cheng.yuchang, fuji.masaru, nagase.tomoki}@jp.fujitsu.com,
[2]{uegaki.minoru, isaac-okada}@jp.fujitsu.com

Abstract

Using instant messenger in real time communication is widespread worldwide. However, in the communication of using cross-language (Japanese - other languages) instant messaging, the use of colloquial expressions usually degrades the efficiency of communication. The contributions of our research can be split into two parts: (1) we analyzed the in-house conversation logs of business correspondence to obtain the cause of the failure of translation in cross-language instant messaging conversations; (2) we proposed an automatic system to detect the untranslatable colloquial expressions of Japanese verbs that are the most significant cause of the failure of Japanese-Chinese (and Japanese-English) machine translation in instant message conversation.

1 Introduction

In recent years, using instant messenger in real-time communication has become common worldwide, and there have been advances in the use of machine translation for efficient communications in multi-lingual conversation. There are an increasing number of global enterprises that use instant messenger for real-time business correspondence. The users of instant messenger can use machine translation service to overcome the language barrier with foreign language speakers in real-time business correspondence. However, the unnecessary conversation arising from the discrepancy of intention increases when the quality of the machine translation is insufficient.

Recently, it has been concluded that the current technology of natural language processing can achieve a high level of accuracy only in processing standard linguistic expressions such as newspaper articles. However, non-canonical linguistic expressions are frequently used in instant message conversation. We believe that the translation quality of instant message conversation decreases due to the use of non-canonical language expressions.

In our research, we first analyzed a log of instant message conversation. The conversation log is in Japanese and it is collected from our in-house business correspondence. We observed the attributes in the conversation and calculated the appearance probability of the attributes. In order to observe the effect that the attributes have on the quality of machine translation, we translated the Japanese utterance[1] in the conversation log into English using Japanese-English translation software.

We found that the attribute "the colloquial expressions of Japanese verbs" is an important factor that causes machine translation quality to decrease. In Japanese, the string expression of a verb composes of two parts. One is the stem, which explains the core concept of the verb, and the other is a terminal expression that explains other information of the verb (such as voice type and tense/aspect). For example, the verb "買った(bought)" consists of the stem "買(buy)" and the termination "った(past tense)". The colloquial expressions of a verb usually occur as a non-canonical termination. Therefore the verb cannot be processed correctly because the termination is untranslatable. If verbs with the colloquial expressions cannot be processed correctly, the machine translation result will become impossible to understand. Therefore, to deal with collo-

[1] In this research, the "utterance" indicates the text that instant messaging users input into the massager.

28th Pacific Asia Conference on Language, Information and Computation pages 532–541

quial expressions, the verb is an important task for improving the quality of the translation in instant message conversation.

Automatic correction of colloquial verbs involves two steps: (1) detecting the colloquial expressions that cause machine translation failure; (2) replacing the colloquial expressions with corresponding formal expressions. Step (2) is difficult because the proofreading system should recognize the user's intention. Step (1) is a relatively easy task because the system only needs to detect the expressions that cause the degrading of the translation quality. In this paper, we proposed a method of detecting Japanese colloquial verbs that will cause the degrading of translation quality.

Detection of Japanese colloquial verbs involves two steps, namely detecting and inspecting the verb, which includes a stem and a terminal expression. First, the system will detect the range of the verb. Because the word order in Japanese is SOV and the length of the instant message is short, the verb is usually placed at the end of the sentence. Second, the system confirms whether the terminal expression of the verb will degrade the translation quality.

In Section 2, we describe related works that deal with the machine translation used in the instant message conversation. In Section 3, we describe an analysis of a log of instant message conversation that involves business correspondence. Section 4 describes our proposed method for detecting the colloquial expressions of Japanese verbs. Section 5 describes the experiment of our proposed method.

2 Related Works

Along with the rapid increase in the use of instant messaging, there are many pieces of research that deal with this topic from different angles. In (Yang, 2011), they introduced the design of the cross-language instant messaging with existing machine translation services.

Komine, Kinukawa, and Nakagawa (2002) discussed the feature that brought the influence to the accuracy of a Japanese-English translation in the Japanese chat conversation of the instant message. They explained that the colloquial expressions usually occur in Japanese chat conversation and the colloquial expressions are difficult to translate.

	Japanese conversations
The number of users	587
The number of conversations	2000
The total number of utterances	22715
The average number of the utterances in one conversation	11.3
The average length of the utterances	19.5 (characters)
The average response time between utterances	40.7 (sec.)

Table 1: Outline of our in-house conversation log

These observations are similar to ours, as we discuss in Section 3.

Saito, Sadamitsu, Asano, and Matsuo (2013) explained that twitter and other micro-blogging data are written in an informal style, so there are many types of non-standard tokens such as abbreviations and phonetic substitutions. They proposed a method for simultaneous morphological analysis and normalization using derivational patterns. Their method used a surface collection, which is expensive to collect. Also, their method was unable to deal with new non-standard tokens that are not included in the collection.

The research that is most related to our mention is (Miyabe & Yoshino, 2010; Miyabe, Yoshino, & Shigenobu, 2009). In (Miyabe et al., 2009), Translation correction plays an important role in multilingual communication using machine translation; it can be used to create messages that include very few translation mistakes. Miyabe and Yoshino (2010) explained the use of back translation for cross-language instant messaging. In their observation, using back translation for detecting the untranslatable text can improve the efficiency of cross-language communication. However, the observations are based on manual simulation. In contrast, our research proposed a procedure for detecting the untranslatable colloquial expressions automatically.

3 Analyzing the Log of Instant Message Conversation in Business Correspondence

It is necessary to analyze a monolingual instant message conversation that does not use machine translation in order to verify the influence of machine translation on the instant message conversation. In this section, we describe the log of the in-

Meanings of attributes	Appearance probability of the attributes	Acceptability in J-C translation	Acceptability in J-E translation
Using face marks	7%	4.29	4.33
Colloquial expression of Japanese verbs	8%	2.65	2.41
Other colloquial expressions	23%	3.58	3.26
Using named entities	16%	3.50	3.62
Incomplete utterances	12%	3.36	3.10
Omitted subject	3%	3.30	3.0
URL, source code, file path	2%	4.16	3.81
Using background knowledge	16%	3.52	3.62
Average (all utterances)		3.62	3.43

Table 2: Distribution and acceptability of the attributes in Japanese instant message conversation

stant message conversations of in-house business correspondence.

In this research, we adopted the communication application "Lync[2]" as the instant messenger. Table 1 shows the outline of our conversation log. The log of the instant message conversation acquired in this paper includes the content of the utterances and the transmission time, where 587 users used Lync in Japanese communications. The record period is three months and it includes 22,715 utterances.

Because the conversations are collected from in-house business correspondence, the definition of the boundary of a "conversation" is unclear. The users do not intentionally specify the beginning and the end of the conversations. We cannot divide the chunk of utterances as independent conversations. However, it is necessary to define the boundary of a "conversation" clearly in order to analyze the intention behind the conversation. We decided to define the "conversation" based on the time period between two utterances. We defined the verge of the conversation as the time of the utterance's transmission progressed ten minutes from the last utterance, and the log of conversations was divided to 2000 conversations.

In Table 1, the average number of the utterances in one conversation is 11.3. This means that when using instant messenger for the business correspondence, the users finished the business by the utterances of about five round trips. Moreover, the average length of the utterances is 19.5 (Japanese) characters. This result shows that the instant message sentences are shorter than other kinds of texts, such as news articles and the papers (Komine et al., 2002). We can also consider that the short sentences (instant message utterances) include simple structures. This means that we can analyze the influence of the factors that affect the machine translation quality easily and clearly.

Next, we analyzed the attributes of the instant message utterances in the conversation log, and we observed the effect that the attributes had on the quality of machine translation. In our research, we adopted the "acceptability" criterion for judging the quality of the machine translation (Goto, Chow, Lu, Sumita, & Tsou, 2013). We defined the quality of machine translation ranging 1 to 5, which corresponds to the acceptability from "F" to "AA".

We executed the machine translation in Japanese-Chinese and Japanese-English. In this experiment, we adopted the commercial translation software "J-Beijing 7[3]" and "ATLAS V14[4]" for the J-C and J-E translation. We can obtain the attributes that affect the quality of the machine translation in different language pairs. Table 2 shows the analysis results. Because there are many attributes that we observed in the analysis, Table 2 lists the major attributes that affect the machine translation quality. The column "Meanings of attributes" shows the observed attributes and descriptions thereof, and the column "Appearance probability of the attributes" shows the appearance probability of the corresponding attribute. The columns "Acceptability in J-C translation" and "Acceptability in J-E translation" show the acceptability in Japanese-

[2] Microsoft™ http://products.office.com/en/lync/

[3] KODENSHA™ http://www.kodensha.jp/

[4] FUJITSU ™ http://www.fujitsu.com/global/products/software/packaged-software/translation/atlas/

Chinese/English translation. The last row in Table 2 shows the average acceptability of all utterances in J-C and J-E translation.

It should be noted that the probability of occurrence of the attributes is independent in different attributes. It is possible that an utterance includes several attributes simultaneously. For example, the utterance "あのう、食べたらいかんぜよ (Uh…you should not eat that)" includes the interjection "あのう(Uh….)" and the colloquial[5] expression "いかんぜよ (you should not do…)" and we defined these expressions as the attributes "Other colloquial expressions" and "Colloquial expressions of Japanese verbs". The length of the string of utterances is not long, so most utterances include fewer than two different attributes.

The attribute with the maximum appearance probability is "Other colloquial expressions". Because the utterances of the attribute "Other colloquial expressions" includes various types of colloquial expressions (excepting the colloquial expressions of verbs), it is difficult to divide this attribute and to consider the measures for deal with various types of colloquial expressions. According to the results of translation acceptability, the translation quality of the utterances that has the attribute "Other colloquial expressions" is not significantly worse than the average acceptability of all utterances. Similarly, the acceptability of the attribute "Using named entities" and the attribute "Using background knowledge" are also not significantly worse than the average acceptability of all utterances. This means that these attributes (with high appearance probability) do not affect the translation quality in a significant way.

However, although the appearance probability of the attribute "Colloquial expressions of Japanese verbs" is 8%, which is less than the attribute "Other colloquial expressions" and the attribute "Using named entities", the acceptability of the utterances with this attribute (J-C:2.65, J-E:2.41) is significantly worse than the average acceptability (J-C:3.62, J-E:3.43). Similarly, compared to other attributes, the acceptability of the attribute "Colloquial expressions of Japanese verbs" is significantly worse than others.

[5] In our research, we didn't subdivide the types of colloquial expressions. The dialectal expressions (such as "いかんぜよ") are regarded to a type of colloquial expressions.

This observation means that if the utterance with the attribute "Colloquial expressions of Japanese verbs" appears, the translation quality will decrease. Moreover, because the average number of the utterances in one conversation is 11.3 (see Table 1), almost all conversations include at least one utterance with the attribute "Colloquial expressions of Japanese verbs". If an utterance with worse translation quality appears, the users will perhaps misunderstand each other. The effort to iron out misunderstandings is necessary. Therefore, the following conversation would be used for ironing out misunderstanding and the business correspondence would be suspended temporarily.

The suspending of the business correspondence using cross-language instant messaging is undesirable. In order to use cross-language instant messaging efficiently, we think that the important task is to deal with the colloquial expressions of Japanese verbs that degrade the translation quality. In the next section, we describe our proposed method for detecting the colloquial expressions of Japanese verbs.

4 Proposed Method - Using Back Translation to Detect the Colloquial Expressions of Japanese Verbs

4.1 The Proposed Method for Detecting the Colloquial Expressions of Japanese Verbs

In explaining our proposal, we used a metaphor to explain the concept of our idea: A personal computer cannot work well and we do not know which part of the PC is broken. To detect the broken part in the PC, we can replace a part of the PC randomly, such as the power supply, and then observe the operation of the PC. If the PC works better than it did when using the former power supply, we can declare that the former power supply is broken. If the PC works similarly to when using the former power supply, the former power supply is not broken.

The motivation of our proposed method is that using formal termination to replace the colloquial expressions of Japanese verbs can obtain a better translation result. According to the discussion in Section 3, the important feature that causes translation quality to decrease is the utilization of the colloquial expressions of Japanese verbs. The colloquial expressions of Japanese verbs usually occur in the termination of the verbs. If the expressions

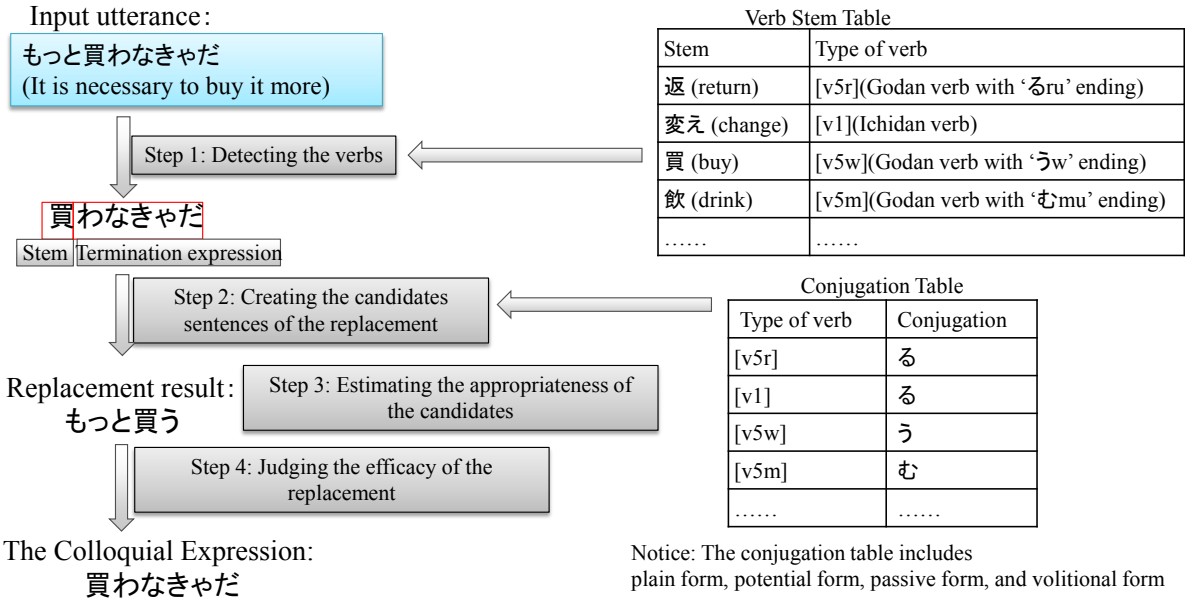

Figure 1: The process of the proposed method
The input utterance "もっと買わなきゃだ (It is necessary to buy it more)" is processed by the four steps and then the untranslatable colloquial expression "買わなきゃだ" is detected.

are untranslatable, the translation process cannot obtain an understandable result. However, using a formal termination of the verb to replace the colloquial expressions can obtain an understandable translation.

For example, the colloquial expression "買わねぇよ" (I don't buy it) cannot be translated into English by using a machine translation application because the termination "わねぇよ (do not)" is an untranslatable expression. If we replace the termination "わねぇよ (do not)" with "ない (do not)", the replaced sentence "買わない (I don't buy it)" can be translated.

Therefore, if a Japanese sentence that has worse translation quality obtains better translation quality remarkably after the replacement, we conclude that the replacement has fixed the untranslatable expressions. Figure 1 shows an example of detecting the untranslatable expressions by means of our idea. The input utterance "もっと買わなきゃだ (It is necessary to buy it more)" is processed by the four steps and then the untranslatable colloquial expression "買わなきゃだ" is detected.

Referring to Figure 1, our proposed method includes the following four steps:

Step 1: Detecting the verbs (comprising a stem and a termination expression)

First, the system detects untranslatable expressions that are extracted by referring to the verb stem table. The input utterance is analyzed into a sequence of morphemes. The verb stem table includes the stem of verbs and the type of verbs. The system detects the stem of verbs from the morpheme sequence and extracts the terminal expression that follows the stem. Because the terminal expression of a verb usually consists of the Japanese "Hiragana" characters, the system extracts the "Hiragana" characters that follow the stem as the terminal expressions.

In Figure 1, the system extracted the stem "買 (buy)" and the terminal expression (Hiragana characters) "わなきゃだ", which is an untranslatable colloquial expression.

Step 2: Creating the candidate sentences for replacement

In step 2, the system creates new utterances by replacing the suspect part of the original utterance. Refer to our idea; the system is not sure that the suspect part is untranslatable. Therefore, the system replaces the suspect expressions with the "correct" expressions. In Figure 1, the system refers to the conjugation table to replace the suspect expressions. The conjugation table is created from the Japanese textbook and the table describes the simple expressions of the verbs. The system used the conjugation table to replace the suspect expres-

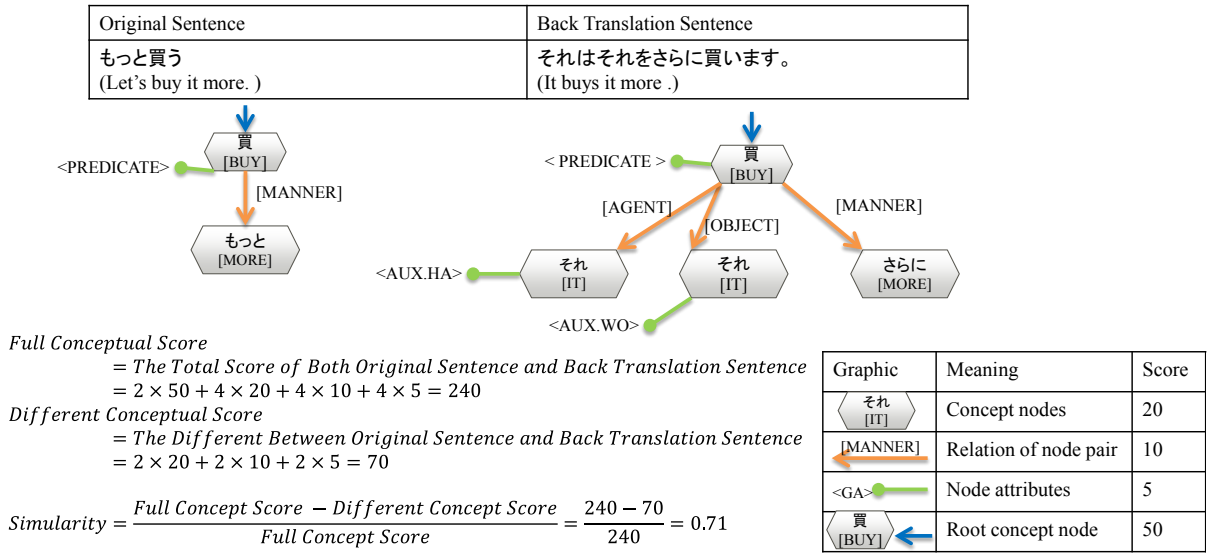

Figure 2: An example of calculating the similarity between the original sentence and the back translation sentence

sions "わなきゃだ" with the simple expression "う" according to the type of the verb, which is "v5w".

Step 3: Estimating the appropriateness of the candidates

After the process in step 2, the system had two candidates - the suspect verb expression "買わなきゃだ" and the simple verb expression "買う". Next, the system estimates the appropriateness of the candidates. The system calculates the similarity between the original sentence and the back translation sentence in each candidate. The system analyzed the concept structures of the original sentence and the back translation sentence of each candidate. Then the system compared the concept structure of the original sentence and the back translation sentence and calculated the similarity as the appropriateness of the candidates. Further detail will be described in Section 4.2. The result of the estimation in this step is that the appropriateness of the candidate "もっと買わなきゃだ" is 0.1 and the appropriateness of the candidate "もっと買う" is 0.71.

Step 4: Judging the efficacy of the replacement

In step 4, the system compared the appropriateness of the candidates and judged the efficacy of the replacement. If the appropriateness has increased tangibly after replacement, the system will judge that the replacement is efficient. The system will output the result of the judgment to indicate the untranslatable colloquial expressions. In Figure 1, the system judged that the replacement of the suspect expression "わなきゃだ" is efficient, and therefore the suspect expression "わなきゃだ" is untranslatable. More details are described in Section 4.3.

4.2 The Similarity between the Original Sentence and the Back Translation Sentence

In step 3 of the proposed method, we adopted the similarity between the original sentence and the back translation sentence for judging the efficacy of the replacement.

"Back translation" is the process of translating a sentence that has already been translated into a foreign language back to the original language (Miyabe et al., 2009) . Using back translation to check the quality of the machine translation is generally used by the users that do not understand the target language. Because the high-quality translation is correct semantically and grammatically, the translation result can be translated to the original language while maintaining a high level of quality. We considered that the system could compare the original sentence with the back translation sentence to judge the quality of translation.

The system uses conceptual structures to compare the original sentence and the back translation sentence. The conceptual structures can explain the semantics of the sentence while avoiding the effect

Case 1 (The untranslatable utterance): the verb in the original sentence includes the stem "買" (buy) and the termination "**わなきゃだ**" (should to do something)

Original (a)	もっと買**わなきゃだ** (It is necessary to buy it more.)
Translation (b)	..purchase.. [wanakyada] .. More
Back Translation (c)	購買wanakyada

The similarity of (a) and (c) : 0.1

Replacement:
[わなきゃだ] -> [う]

Original (d)	もっと買**う**
Translation (e)	It does buy it more.
Back Translation (f)	それはそれをさらに買います

The similarity of (d) and (f) : 0.71

Case 2 (The translatable utterance): the verb in the original sentence includes the stem "買" (buy) and the termination "おう" (want to do something)

Original (g)	もっと買**おう** (Let's buy it more)
Translation (h)	It does buy it more.
Back Translation (i)	それはそれをさらに買います

The similarity of (g) and (i) : 0.71

Replacement:
[おう] -> [う]

Original (j)	もっと買**う**
Translation (k)	It does buy it more.
Back Translation (l)	それはそれをさらに買います

The similarity of (j) and (l) : 0.71

Figure 3: The effect of replacing the verb termination
The difference of the similarity between the left part (before the replacement) and the right part (after the replacement) is 0.61 in Case 1 and 0 in Case 2. The replacement of Case 1 has greater impact than the replacement of Case 2.

of the variant expressions and the effect of the word order. For example, although the sentence "昨日(yesterday)私(I)はリンゴ(apple)を食べた(ate) (I ate the apple yesterday)" and the sentence "僕(I)は昨日(yesterday)林檎(apple)を食った(ate) (I ate the apple yesterday)" have different word order and use variant expressions, these sentences have similar conceptual structures and are translated into the similar English sentence "I ate the apple yesterday".

Figure 2 describes how to calculate the similarity between the original sentence and the back translation sentence. The system first obtains the back translation "それはそれをさらに買います (It buys it more)" of the original sentence "もっと買う (Let's buy it more)" by using a machine translation system. Then the system analyzes the conceptual structure of these sentences. The table on the lower-right side of Figure 2 explains the graphics that are used in this example. A hexagon is the concept node and it includes the surface string and the concept (shown in the square bracket). A hexagon with a blue arrow shows the root concept of the sentence. The arrows show the relations between nodes. The parenthesis shows the grammatical features of the node.

The right column of the table in Figure 2 shows the score of a node, node relation, root node, and features. We defined the scores heuristically. We assigned a high score to the root node because the root node is the core concept of the sentence. Also, the concept node has a higher score than the node relation because the concept node represents the meaning of the sentence.

The expressions for calculating the similarity are shown at the bottom of Figure 2. We calculated the full conceptual score of the two sentences and the different conceptual score and then calculated the similarity. The full conceptual score is the sum of the score of all root nodes, concept nodes, node relations and grammatical features. In this example, the full conceptual score is 240. The different score is the score sum of the different part of the sentences. As the node "それ (it)" in the back translation sentence did not occur in the original sentence, it is regarded as a different part. In this example, the different conceptual score is 70. Finally, we calculated the similarity using the expressions and the similarity of this example is 0.7.

4.3 Judging the Untranslatable Colloquial Expressions of Japanese Verbs

In step 3, the system estimates the appropriateness of the candidates by calculating the similarity of the original sentence and the back translation sentence. In step 4, the system judges the untranslatable colloquial expressions by comparing the appropriateness of the candidates.

Figure 3 shows an example that describes the judgment mechanism. The example includes two

cases: the untranslatable utterance "もっと買わなきゃだ (It is necessary to buy it more)" and the translatable utterance "もっと買おう (Let's buy it more)". Using the process in step 3, the system obtained the difference of the similarity in Case 1 (0.71 – 0.1 = 0.61) and the difference of the similarity in (0.71 – 0.71). In Case 1, the difference of similarity (0.61) can be seen as having tangibly increased the translation quality after replacing the suspect terminal expressions. Therefore, we concluded that the original colloquial expressions "わなきゃだ" is untranslatable. Alternatively, the difference of similarity in Case 2 (0.0) can be seen as having not changed the translation quality after replacing the suspect terminal expressions. Therefore, the original expression is translatable.

There are several calculation expressions and methods for the judgment. In this example, we used the simplest calculation to judge the untranslatable colloquial expressions. However, we can also adopt other factors for judging it accurately, such as comparing the conceptual structures. To minimize computational complexity, we used a heretical threshold to judge the untranslatable colloquial expressions of Japanese verbs.

	Utterances
Automatically detected untranslatable colloquial expression	2229
All utterances in our conversation log	14394

Table 3: The number of total utterances and the number of automatically detected untranslatable colloquial expression

Explain	The average acceptability
The detected untranslatable expression (2229 utterances) BEFORE manual correct	3.03
The detected untranslatable expression (2229 utterances) AFTER manual correct	3.46
The total conversation (14394 utterances) BEFORE manual correct	3.41
The total conversation (14394 utterances) AFTER manual correct	3.48

Table 4: The average acceptability in J-C translation before / after correcting the detected colloquial expression

Acceptability	Utterances	%
Increased	752	33.7%
Similar	1163	52.2%
Decreased	60	2.7%
Manually uncorrected	254	11.4%
Total	2229	

Table 5: The transition of the acceptability in J-C translation by correcting the utterances according to the judgment of the proposed system
The row "Total" means the number of automatically detected untranslatable colloquial expression.

5 Experiments

The automatic correction of colloquial verbs involves two steps: (1) detecting the colloquial expressions that cause machine translation failure; and (2) replacing the colloquial expressions with corresponding formal expressions. Out of the two steps, we have implemented and have given full discussion on detection step (1), while correction step (2) is yet to be implemented in our future works.

In order to estimate the effectiveness of detection module (1) used in the entire automatic correction flow, we firstly applied detection module (1) to Japanese-Chinese translation, and then we manually corrected the detected colloquial expressions. Since the user of translation cannot differentiate between translatable and untranslatable expressions, we have decided to manually correct all the machine detected expressions.

Table 3 shows the experiment data and the result of the automatic detection. Our system obtained the J-C back translation and the detected results are the untranslatable expressions in Japanese-Chinese translation. We used the proposed system to process the conversation log of the in-house business correspondence that we described in Section 3. The conversation log includes 22,715 Japanese utterances but there are a lot of duplicates. We deleted these duplicate utterances, and the remaining conversation logs include 14,394 utterances. The proposed system detected 2,229 untranslatable utterances (see the row "Automatically detected untranslatable colloquial expressions"). We believe that the users of instant messaging do not like the excess detection. Therefore we tune the system to reduce the excess detection and to require higher precession more than the recall.

Table 4 shows the average acceptability before / after the manual correction in J-C translation. We

considered the average acceptability of the automatic detected colloquial expressions (2,229) and all utterances in our conversation log (14,394). The average acceptability of automatically detected colloquial expressions increased from 3.03 to 3.46. The average acceptability of the total utterances increased from 3.41 to 3.48. Although the quality didn't increase dramatically, these results show that our system detected the untranslatable utterances effectively and helped users to increase the translation quality.

It should be noted that the system detected the untranslatable utterances. Therefore the translation quality of the detected utterances (3.03) is worse than the quality of all utterances (3.41). We can also claim that our system could be used for automatically evaluating the translation quality of the conversation.

Table 5 shows the number of the utterances that the acceptability increases / decreases / is similar. In this research, we manually corrected the untranslatable colloquial expressions that the system detected. In the manual correction, our operators try to "TRUST" the automatic detection and they tried to rewrite the untranslatable colloquial expressions with simple and synonymous expressions. If it is impossible to rewrite the expressions, the operators may decide not to correct the expressions. There are 254 (11.4%) utterances that the operators cannot correct any more. This result shows that the operators recognized the automatic detection.

After the manual correction process, we used the J-C translation application to translate the original utterances and the manual correction results. Then we evaluated the acceptability of these translation results. In our experiment results (see Table 5), 752 (33.7%) utterances are corrected and have their acceptability increased (see the row "Increased"). The acceptability did not change in 1,163 (52.2%) utterances, and the acceptability decreased in 60 (2.7%) utterances. This results show that the automatic detection can help the users to deduce the untranslatable utterances and our system achieved our expectation – detecting the untranslatable colloquial expressions in high precision and low degradation. We think that our system can be put into use.

6 The efficiency of communication

For the practical application of our system, the system is not only evaluated in the quality of translation, but also to be evaluated in the efficiency of communication. However, we didn't establish the method for evaluating the efficiency of communication.

The difficulty in evaluating the effect of our proposed method is that we cannot define the criteria of the success and the effectiveness of a cross-language instant messaging conversation. It can be considered that the number of the utterances in a conversation is useful information. If the conversation includes few utterances, it could be said that the instant massager users enjoy efficient communication. If the translation results cannot be understood, the users are forced to suspend the conveying of their intention and try to explain their intention to understand the unclear utterances (untranslatable utterances). This sort of communication will increase the number of utterances.

Table 6 and Table 7 show examples of cross-language instant messaging conversations. Table 6 has the correct translation results; therefore the users (user A and B) finished their intention transmission. In Table 7, the Japanese user (A) used the untranslatable colloquial expression "もっと買わなきゃだ" and the user (B) could not understand the intention of user A. Users A and B discussed the unclear translation (see the shaded row of Table 7). After the discussion, users A and B cleared up the misunderstanding. However, the number of the utterances increased in Table 7. Although there is only one untranslatable utterance "もっと買わなきゃだ" in Table 7[6], we thought that the effectiveness of the conversion of Table 7 is worse than Table 6.

Our system can detect the untranslatable colloquial expressions in Table 7 but does not detect the other utterances. Referring to our experiment, only one sentence will be counted as successful detection and the variation of the average acceptability is small. Therefore, the real effectiveness of our system is to reduce the non-effective utterance (see the shaded row in Table 7) but it cannot be evaluated fairly in this experiment.

7 Conclusion and future direction

In this research, we analyzed the in-house conversation logs of business correspondence to obtain the cause of the failure of translation in cross-language instant messaging conversations. Then

[6] In Table 6, the other utterances are translated correctly.

we proposed a method for automatically detecting the untranslatable colloquial expressions of Japanese verbs that are the most significant cause of the failure of Japanese-Chinese (and Japanese-English) machine translation in instant message conversation. The experiments result shows that our system can improve the average acceptability of all utterances from 3.41 to 3.48. The results of automatic detection can help users to reduce the untranslatable utterances with high precision. We also explained that the utterance unit criteria could not evaluate the effectiveness of our system. Therefore, one future direction is to create a criterion for evaluating the effectiveness of the conversation.

Another direction is to improve the calculation of the similarity of the original sentence and the back translation sentence. We used a simple method to evaluate the similarity but there are several related works that deal with the similarity of two graphs or tree structures. These methods can provide more credibility to the similarity of the original sentence and the back translation sentence.

Speaker (translation direction)	User (A)'s view (Japanese)	User (B)'s view (English)
A (J->E)	冷蔵庫にコーラがない。もっと買わないといけない	There is no cola in the refrigerator. It is necessary to buy it more.
B (E->A)	より多くのコーラも買いたいと思います。	I want to buy more cola, too.
A (J->E)	一緒に買いに行こう	Let's go for purchase together.

Table 6: Example of conversation without non-effective utterance

Speaker (translation direction)	User (A)'s view (Japanese)	User (B)'s view (English)
A (J->E)	冷蔵庫にコーラがない。***もっと買わなきゃだ***	There is no cola in the refrigerator. *More... purchase... [wanakyada]*
B (E->J)	「さらに。」「購入します」。? あなたは、より多くのコーラを買いたいと言っていましたか?	"More… purchase..."? Did you mean that you want to buy more cola?
A (J->E)	そうです、私はコーラをもっと買いたい	It is so, and I want to buy cola more.
B (E->J)	より多くのコーラも買いたいと思います。	I want to buy more cola, too.
A (J->E)	一緒に買いに行こう	Let's go for purchase together.

Table 7: Example of conversation with non-effective utterance
The translation direction "A (J->E)" means that the user (A)'s utterance is in Japanese and is translated to English

References

Goto, Isao, Chow, Ka Po, Lu, Bin, Sumita, Eiichiro, & Tsou, Benjamin K. (2013). *Overview of the patent machine translation task at the NTCIR-10 workshop.* Paper presented at the Proceedings of the 10th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access, NTCIR-10.

Komine, Hisashi, Kinukawa, Hiroshi, & Nakagawa, Hiroshi. (2002). *Sentence Extraction based on Document Frequency and Text Length.* Paper presented at the IPSJ SIG Notes.

Miyabe, Mai, & Yoshino, Takashi. (2010). *Influence of detecting inaccurate messages in real-time remote text-based communication via machine translation.* Paper presented at the Proceedings of the 3rd international conference on Intercultural collaboration, Copenhagen, Denmark.

Miyabe, Mai, Yoshino, Takashi, & Shigenobu, Tomohiro. (2009). *Effects of undertaking translation repair using back translation.* Paper presented at the Proceedings of the 2009 international workshop on Intercultural collaboration, Palo Alto, California, USA.

Saito, Itsumi, Sadamitsu, Kugatsu, Asano, Hisako, & Matsuo, Yoshihiro. (2013). *Extracting Derivational Patterns based on the Alignment of a Standard Form and its Variant towards the Japanese Morphological Analysis for Noisy Text.* Paper presented at the IPSJ SIG Notes.

Yang, Che-Yu. (2011). *Cross-Language Instant Messaging with Automatic Translation.* Paper presented at the Ubi-Media Computing (U-Media), 2011 4th International Conference on.