

## Six Pre-edit Techniques for Enhancing Japanese to English Machine Translations (日英自動翻訳改善のための6ステップ)

June-ko Matsui and David Magnusson  
(明海大学 Meikai University)

*On-line machine translations are one of the most powerful tools available at the disposal of non-fluent language users. Machine translations provide good general descriptions, but are not necessarily accurate at detailed translations, especially for languages like English and Japanese, which have different grammatical structures. This pedagogical report introduces a six-step process where users revise their Japanese input, creating more accurate and understandable machine translations to English. Paying attention to the grammar and lexicon can significantly enhance the output, maximizing communication capabilities over a wider range of topics.*

### 1. Introduction

Numerous on-line machine translation systems provide Japanese users the means to write more readily in English. Although significant technological progress has resulted in vast improvements, differences in grammar etc. still result in unnatural or incomprehensible translations. Martin Kay (n.d.) states “Many factors contribute to the difficulty of machine translation, including words with multiple meanings, sentences with multiple grammatical structures, uncertainty about what a pronoun refers to, and other problems of grammar.”

Methods to overcome such problems are not widely available to English learners. For instance, Japanese frequently does not require subjects, numbers, or determiners, whereas English usually does. This results in inaccurate and misleading passages for Japanese to English machine translations. For example a Japanese sentence which means “I was very tired.” is translated “It became tired very much.” by the free on-line Google machine translator.

“totemo (=very) tsukaremashita (=was tired)”

---

MATSUI Junko & David Magnusson, “Six Pre-edit Techniques for Enhancing Japanese to English Machine Translations,” *Interpreting and Translation Studies*, No.11, 2011. pages 173-184. © by the Japan Association for Interpreting and Translation Studies

「とても疲れました。」

“It became tired very much.”

Users in the study were told to insert sentential subjects which are missing in the Japanese passages. After inserting the subject, *watashi* (=I) (わたし), something closer to a correct translation “I’m very tired.” is obtained from the Google free on-line translator.

*watashi* (=I) *wa* (=particle) *totemo* (=very) *tsukaremashita* (=was tired)

「わたしはとても疲れました。」

“I’m very tired.”

Each language poses its respective difficulties in the translation process. As Kay (n.d.) points out, “Consider the following question, stated in French: *Ou voulez-vous que je me mette?* It means literally, ‘Where do you want me to put myself?’ but it is a very natural translation for a whole family of English questions of the form ‘Where do you want me to sit/stand/sign my name/park/tie up my boat?’ In most situations, the English ‘Where do you want me?’ would be acceptable, but it is natural and routine to add or delete information in order to produce a fluent translation. Sometimes it cannot be avoided because there are languages like French in which pronouns must show number and gender, Japanese where pronouns are often omitted altogether, Russian where there are no articles, Chinese where nouns do not differentiate singular and plural nor verbs present and past, and German where flexibility of the word order can leave uncertainties about what is the subject and what is the object.”

Various researchers have compared vying on-line machine translators (Syahrina, 2011). This study focuses on the structural and lexical differences between two languages - English and Japanese, and attempts to set out a six-step pre-edit method especially geared especially for English language learners who want to translate in the relatively difficult direction – from Japanese to English translation. The procedure trains English learners to add or edit the missing information etc. in Japanese – information that is easy for a human to understand, but still eludes the current free on-line machine translation systems. This study deals with general topics, rather than specific technical areas at which the machine translators excel. The process empowers language learners, providing them a tool which can expand the scope and accuracy of their communication.

## 2. Method

Eighty-eight English learners were asked to write a Japanese passage on a memorable event. They were next instructed to translate it into English on their own, then translate it using the three most popular on-line machine translation systems on the Japanese Google search engine: Excite, Google, and Yahoo. They were then given the following six-step pre-edit instructions to make their machine translations more natural and accurate:

1. Make longer sentences shorter.
2. Insert subjects such as “I, you, he, she, and they”.
3. Replace words that have multiple meanings. (ex. 「通っていた」(went to) → 「習っていた」(learned))
4. Clarify whether nouns are singular or plural.  
(ex. 「男」(man/men) → 「男たち」(men))
5. Add in determiners for nouns. (ex. 「先生に」(teacher) → 「その先生に」(the teacher))
6. Check the spelling of proper nouns.

The same users were then requested to fill out a survey on the machine translation and their English studies in the two-part process (instruction/survey) which took place in July, 2011. Two raters (the authors) judged the translations based on the six steps.

### 3. Results

#### 3.1 Six Step Pre-edit Machine Translation Manual

Results show improvements before and after the six-step pre-edit instruction, indicating even one session of training provides English learners a powerful tool to boost their communication skills. Fuller training and practice could potentially expand the scope of users' communication skills far beyond their inherent capabilities.

The following are excerpts from the results which demonstrate improvements for each of the six pre-edit steps:

##### 3.1.1 Making Longer Sentences Shorter

The first technique calls for shorter sentences. It is evident that shortening the original Japanese sentences resulted in greater translation accuracy. For example, one user wrote:

Haruyasumi ni guamu ni itte, totemo tanoshii omoide ga dekita.

The machine translations produced the following:

- (1) A very happy for went to Guam at the spring vacation memories were able to be done.  
(Excite)
- (2) Go to Guam at the spring vacation break, it was very pleasant memories.(Google)
- (3) I performed it in Guam in spring vacation, and it was possible for a memory fun at all.  
(Yahoo).

Upon suggestion to reduce sentence length, the user produced two sentences out of the original one:

Watashi wa haruyasumi ni guamu ni itta.

Sokode tanoshii omidega dekita.

The machine translations produced the following:

- (1) I went to Guam at the spring vacation.  
Then very happy memories were able to be done. (Excite)
- (2) Spring break I went to Guam.  
So we have very happy memories. (Google)
- (3) I went to Guam in the spring vacation.  
Therefore a very pleasant memory was possible. (Yahoo)

### 3.1.2 Insert Subjects Such as “I, You, He, She, and They”

The second technique of inserting the subject results in perhaps the greatest improvement to the machine translations. Since Japanese sentence sentences often omit the subject, the machine translations often make up for this by inserting the subject as “it,” which often results in odd translations. However, with a specified subject, the machine translations translated the original text with much greater accuracy. For example, another user wrote:

Kotoshino haruyasumi ni gogaku kenshuu toshite furorida ni ikimashita.

The machine translations translated this original Japanese text into English as follows:

- (1) It went to Florida as language study training at the spring vacation of this year.  
(Excite)
- (2) Florida went on spring vacation this year as language training. (Google)
- (3) I went to Florida to take a rest in this spring as the language study training. (Yahoo)

Upon the suggestion of the second technique to insert the subject, the user produced:

Watashi wa kotoshi no haruyasumi ni eigo wo benkyou surtameni furorida ni ikimashita.

The machine translations translated this original Japanese text into English as follows:

- (1) I went to Florida to study English at the spring vacation of this year. (Excite)
- (2) I went to Florida for spring break this year to study English. (Google)
- (3) I went to Florida to take a rest in this spring to study English. (Yahoo)

As one can see, the revision produced a much more accurate translation.

### 3.1.3 Replace Words that have Multiple Meanings

The third technique to replace a word with numerous meanings with a synonym with fewer meanings was difficult for the users to apply, but a few language learners attempted to use this technique.

For example, one user produced the following Japanese sentence:

Soshite futsuka-me niwa, sendai no yuumena kankouchi wo mawatta.

The machine translations produced the following:

- (1) And it is surroundings [ ] on the second day as for a famous sightseeing spot of Sendai. (Excite)
- (2) And on the second day, the anniversary was a famous sight in Sendai. (Google)
- (3) And on the second day, it is lap (tsuta) at a famous sightseeing spot of Sendai. (Yahoo)

The meaning of these sentences is not clear. However, by changing the word, “mawatta,” (to go around or to go in circles, or surroundings) which has numerous meanings with “itta” (to go), the machine translations produced the following:

- (1) And it went to a famous spot of Sendai for the second day. (Excite)
- (2) On the second day and went to the famous sights in Sendai. (Google)
- (3) And on the second day, I went to the famous sightseeing spot of Sendai. (Yahoo)

If the user had also entered the subject, “I,” Excite and Google would have produced a more accurate translation. Nevertheless, the change from “mawatta” to “itta” produced a much more precise translation.

### 3.1.4 Clarify whether Nouns are Singular or Plural.

The fourth technique, which is to clarify the number of the nouns as singular or plural, also helps produce more accurate translations. For example, one user produced the following sentence:

Kono haru yasumi ni tomodachi to dhizunii ni tomarimashita.

The machine translations translated this as follows:

- (1) I went to the friend and Disney by staying at the spring vacation of this year. (Excite)
- (2) This year’s spring break, we went to Disney with friends at night. (Google)
- (3) It reached friend and Disney by night watch on a holiday this spring. (Yahoo)

Using the technique of inserting a subject, shortening the sentence, and inserting the number of the noun, the user produced:

Watashi wa haruyasumi ni dhizuni ni ryokou ni itta.  
Watashi wa tomodachi gonin to tommatta.

The machine translations then produced:

- (1) I went on a trip to Disney at the spring vacation. I stayed with five friends. (Excite)
- (2) I went to spring break to Disney. Five friends and I stayed. (Google)
- (3) I went to Disney for a trip in spring vacation. I stayed with five friends. (Yahoo)

It is clear that as a result of concurrently applying three techniques, the translations have vastly improved. The number of friends was specified in the revision. In the original version, the word *tomodachi* could refer to one friend or numerous friends. There is no way for the machine translation to know which it was. By specifying the exact number, the meaning became clearer.

#### 3.1.5 Add in Determiners for Nouns

The fifth technique of inserting determiners and pronouns also helps the machine translation process. For example, one user wrote:

Tanjoubi no atta shuu ni tomodachi ga sapuraizu de purezento to gamen kehki wo shite kuremashita.

This resulted in the following machine translations:

- (1) The friend did the present and the face cake by the surprise on the week with the birthday. (Excite)
- (2) The week was a birthday present me with a cake and surprise you friends face. (Google)
- (3) In the week that there was of the birthday, a friend did a present and a face cake by a surprise. (Yahoo)

Upon applying techniques of inserting the determiners/possessive pronouns and of inserting the subject, the user produced the following sentence:

Watahino tanjoubino atta sono shuu ni watahino tomodachini, sapuraizude takusanno purezentoto gamen kehki wo shite kuremashita.

- (1) My friend did me a lot of presents and the face cakes by the surprise the week with my birthday. (Excite)
- (2) The week was my birthday, I have my friends, me and a lot of surprise gifts and cake face. (Google)
- (3) In the week there was of my birthday, I did a present and the face cake which my friend did not entrust me with by a surprise. (Yahoo)

The revision resulted in greater accuracy, but the user would have succeeded even more by breaking up the long sentence into two shorter sentences.

### 3.1.6 Check the Spelling of Proper Nouns

The sixth technique of improving machine translations, which is to check the spelling of proper nouns results in better translations. For example, one user produced:

Ichinichi-me wa shii de asobi, hoteru mirakosuto ni ippakushite tsuginohi wa rando de tomari-mashita.

The machine translation by Excite produced the name of Hotel MiraCosta as “hoterumirakosuta” and could not recognize it as a proper noun. The user in her self-translation, accurately referred to it as “Hotel MiraCosta.”

### 3.1.7 Different Translation Systems

The approaches used by Excite, Google, and Yahoo are different. Google is a statistical machine translator, while Excite and Yahoo are rule-based. This makes it more likely that Google will add in the first person pronoun as a subject. Meanwhile, the rule-based programs also have subject-inserting features. However, these systems are not sophisticated enough yet to follow a thread of thought, and accurately change the topic, and the accompanying subject, for instance switching from the first person (I) to the second and/or third person (you/he/she/it/they), then back to the first person again in quick succession. Inserting the subject of a sentence for Japanese to English translations yields the most accurate result, using the existing free on-line translators.

However, improvements in the systems take place constantly, and in the future, free on-line machine translators may be able to accurately pick up the actor of each sentence.

## 3.2 Survey

In the latter half of the survey, language learners responded to a survey on the machine translation systems they had just used.

Most language learners feel the machine translation systems are useful, and they would like to use them in the future. However, a large number of language learners feel their own translations are better than the machine translation. Results suggest machine

translations have a limited but potentially important role to play in language communication for users at various levels of language proficiency.

### 3.2.1 Usefulness

A large majority (88%) of language learners feel the machine translations are useful.

#### 1. Were the machine translation systems useful?

1. Very useful
2. Somewhat useful
3. Not possible to say either way
4. Not very useful
5. Not useful at all

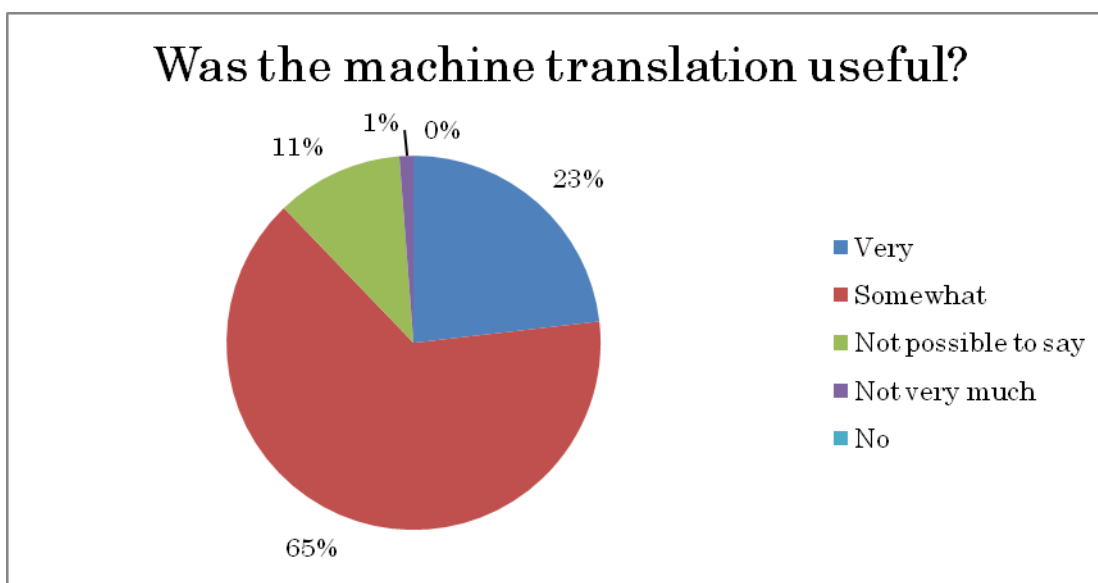


Figure 1: Percentage of Respondents who feel the Machine Translation was Useful

Results indicate users feel the machine translations systems could enhance their communication capacity.

### 3.2.2. Future Use

Almost all of the respondents (94%) say they want to use machine translation in the future.

#### 2. Would you like to use machine translation systems in the future?

1. Yes, definitely
2. Sometimes
3. Not possible to say either way
4. Not really
5. Not at all



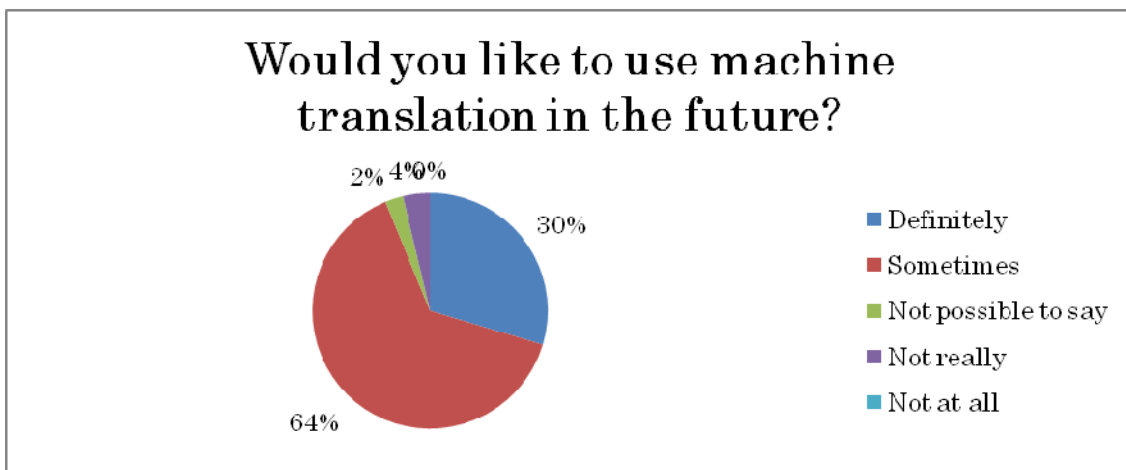


Figure 2: Percentage of Respondents who would like to Use Machine Translation in the Future

Results indicate machine translation systems will play a larger role in the future, even for intermediate-level English learners.

### 3.2.3 The Best Machine Translation System

The largest percentage of respondents feel the Excite translator was best, followed respectively by Google and Yahoo, The appraisal of the translating systems is, interestingly, the same as the ranking in the Google search engine.

#### 3. Which machine translation system was best?

1. Excite
2. Google
3. Yahoo

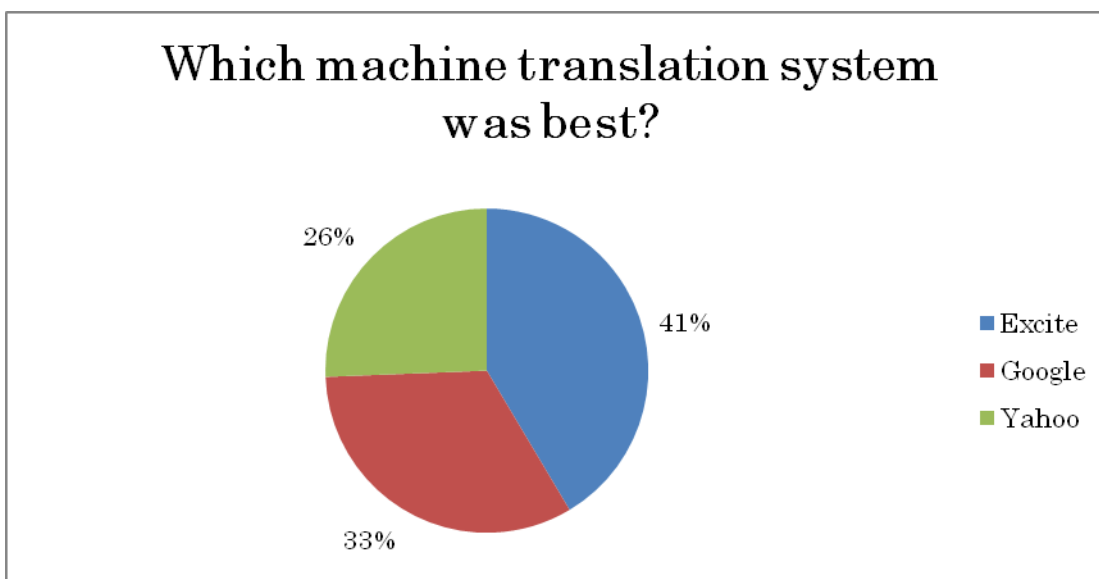


Figure 3: The Machine Translation Respondents Feel is best

There was a marginal correlation ( $r=0.26$ ) between the evaluation of the best machine

translation system, and whether respondents felt their own translation is better than all the other machine translation systems combined. Results show people who feel the Excite system is best are also likely to feel their own translations are better than the machine translation systems. On the other hand, respondents who feel the Yahoo system is best are more likely to feel that the machine translations are better than they are. Results indicate that respondents who are sensitive to the most popular machine translation systems on the Google search engine are more confident about their English skills.

### 3.2.4 Time Spent on English Studies per Day

The amount of time spent on studying English every day peaked at 30-60 minutes and 2-3 hours.

4. Including your studies at the university, how much time do you spend per day studying English?

- 1. 0 minutes    2. 1-15 minutes    3. 15-30 minutes    4. 30-60 minutes    5. 60-90 minutes
- 6. 90-120 minutes    7. 2-3 hours    8. 4-5 hours    9. 6-7 hours    10. 8-9 hours
- 11. 10 hours or more

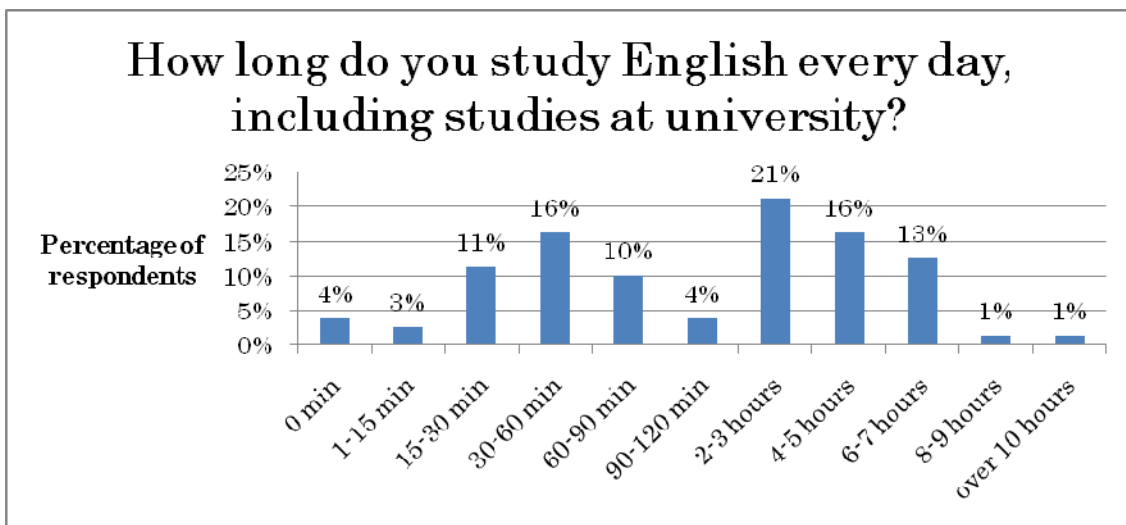


Figure 4: How Long Respondents Study English Every Day

The time spent on studying English every day was not correlated with any other item. This indicates that both enthusiastic English learners as well as moderately interested learners feel that machine translations are useful, and that they will use machine translations in the future. Both types of English learners also tend to think that Excite is the best system, followed by Google and Yahoo. Results appear to show that machine translations are an attractive communication tool for lower and intermediate level English learners.

### 3.2.5 Own Translation vs. Machine Translation

Over a third of the respondents (35%) feel their own translations are more understandable than the machine translation. The remaining two-thirds chose one of the machine translations as the best translation. The three translation systems (Excite, Google, and Yahoo) all have about the same percentage (21-23%) of first-choice appraisals.

5. Which was easier to understand, your own translation, or the machine translations?

1. Your own translation
2. Excite machine translation
3. Google machine translation
4. Yahoo machine translation

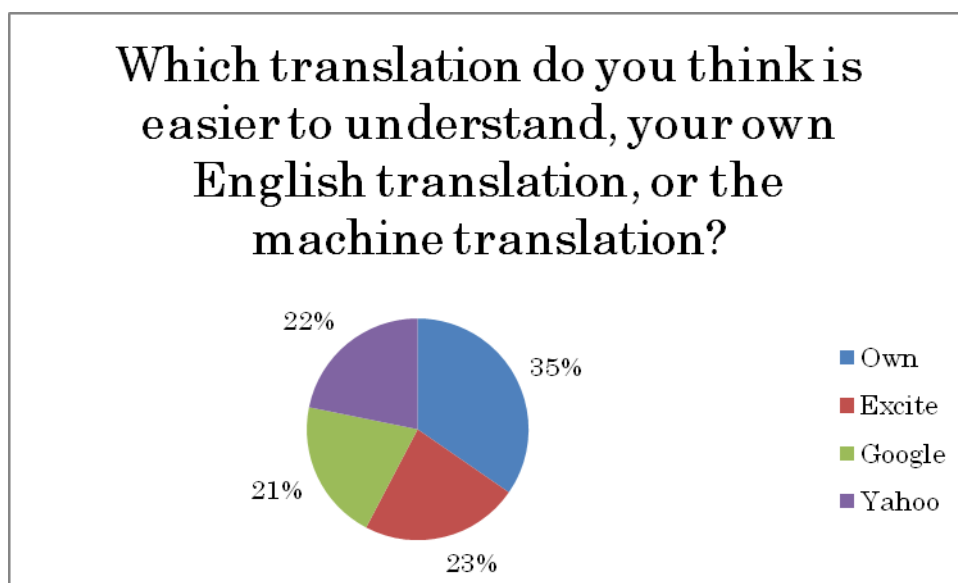


Figure 5: Which Translation is Easiest to Understand

A significant number of respondents feel their own translation is easier to understand than the machine translation. The large percentage of high self-evaluations is believed to stem in part from the soft expressions used in the original Japanese which are especially hard for machine translation systems to process. The machine translations could be more useful for elementary English users who have trouble forming basic sentences, in this category. However, two thirds of the respondents recognize the machine translations as the superior translation, indicating machine translation systems have an important role to play, even for intermediate level English learners.

## 4. Conclusion

Machine translations provide an essential service in a global society. However online translations are far from perfect, and often misleading. Following the six pre-edit steps outlined in this study leads to a considerably more understandable and accurate machine translation, and appears to be a promising tool for lower and intermediate level language learners.

Many users feel that their own translations are better than the machine translations. Results indicate language learners who have had a few years of exposure to a foreign language sometimes beat out the machine translations when it comes to softer, non-formulaic expressions. However, machine translations could be more helpful for users who are not very familiar with English, or users who wish to translate more technical or mechanical contents which are far beyond their reach.

Machine translations are popular with the majority of respondents in the survey. About 90% of the respondents say the machine translation systems are useful, and they would like to use them in the future. Further advances in the software will lead to a greater role for machine translations in the future as the accuracy of the systems increase. In the meantime, machine translations are not infallible. Learning the strengths and the weaknesses of the systems, and applying the six-step pre-editing approach outlined in this paper can minimize the faults and maximize the benefits of machine translations, and provide a wide level of language learners a tool which allow them to communicate in domains they previously never felt was possible.

.....

#### **Acknowledgements**

This research was made possible, in part by a fund from the Ministry of Education, Science, Sports and Culture Grant-in-Aid for Scientific Research (A) – the Tono Kaken, and a grant from the Meikai University Miyata Fund. The authors would like to thank everyone who provided valuable comments on the paper.

#### **About the Authors**

June-ko Matsui (松井順子): Meikai University Associate Professor (明海大学准教授)。

David Magnusson: Meikai University Integrated Education Center Urayasu Campus  
Communication Skills Development Program Lecturer

.....

#### **References**

Kay, M. (n.d.). “Machine Translation”. *Linguistic Society of America*.

<http://www.lsadc.org/info/ling-fields-machine.cfm>

Syahrina, Alvi (2011). “Online Machine Translator System and Result Comparison – Statistical Machine Translation vs. Hybrid Machine Translation”. University of Borås.