

## Experiment Report of a Commercial Machine Translation in a Manufacturing Industry Domain

**Orie Fukutomi**

School of Languages and European Studies  
University of Wolverhampton  
Stafford St.  
Wolverhampton  
WV1 1SB  
UK  
O.Fukutomi@wlv.ac.uk

Keywords: user study, commercial machine translation, industry domain

### Abstract

The aim of this paper is to provide a report of an experiment using a Commercial Machine Translation (CMT) software in a manufacturing company in the UK, with particular reference to Japanese / English Machine translation. It presents the main difficulties involved in the translation of industrial documents from Japanese to English, and discusses how the productivity and quality of translation can be improved through the use of commercial Machine Translation (MT) software. This is an empirical data and is proposed from translators' point of view in a manufacturing factory. The survey focuses on a manufacturing organisation which does not have the resources needed to develop their own MT system. The globalization of the Japanese manufacturing industry makes it necessary for the translation of manuals and other documents to be as rapid as possible. In this paper, linguistic features of both English and Japanese are discussed from the evaluation experiment in order to make up writing rules for members of staff at Makita Manufacturing Europe. It also discusses viewpoint of British engineers when translated manuals are read.

### 0. Outline Structure

The quality of translations generated by MT often disappoints users and it even makes them stop using it for serious documentation. Although it is still impossible for MT to process natural language completely, the level achieved is significant. Yet, not many users are informed about the properties or strengths of MT. In this paper, one Commercial Machine Translation software package, Honyaku Adapter II (HAI) from NEC, has been chosen in order to demonstrate how it can be utilised in a medium-sized company, Makita Manufacturing Europe Ltd.(MME), where there is no translation department.

The author chose test sets from a Japanese organisation, Japan Electronic Industry Development Association (JEIDA). [Isahara, 95] However, this particular experiment only focuses on the manufacturing domain, and its language can be a sublanguage. As can be seen from the example below, test sets are not particularly relevant to the evaluation of sentences from the manufacturing domain.

Thus, instead of using actual sentences from JEIDA's test sets for the evaluation, the author used the test sets' categories for a CMT to translate actual texts from MME. In this way, the result of the evaluation offers immediate feedback to manufacturing companies that require translation services. The experiment procedure is as follows; Firstly five hundred twenty eight Japanese sentences are chosen from the typical types of document held by Makita Manufacturing Europe Ltd. (MME), and then these were translated

using the Commercial Machine Translation software, HAIL. Secondly, an evaluation was carried out using criteria from JEIDA's test sets. In this paper, linguistic features of both English and Japanese are discussed from the evaluation experiment in order to make up writing rules for members of staff at Makita Manufacturing Europe. Accordingly, recommendations are made concerning how a company seeking globalization could resolve communication problems by the employment of Commercial Machine Translation.

## 1 General Background of Commercial Machine Translation Products (Japanese/English) in the Market

The availability of CMT for non-professional translators increased tremendously over last five years. Ten years before, MT systems in Japan, like MT systems in other countries, were only run on large scale main frame computers or on UNIX workstation. The cost of the system was in the region of £20k and the market being targeted consisted of professional translators and translation agencies. [Johnson 97] Hence target customers were organisations where the volume of translation required was high, so the MT system is paid off.

Now, high quality MT packages are often sold under 9800 yen (approximately 100 US dollars) in the market. This is an incredibly low price compared to the price of equivalent MT software packages being sold in other countries. For example, SYSTRAN sells English to Japanese and Japanese to English MT software for 128,000 yen (approximately 1000 US dollars) on the Japanese market. According to Mr. Hutchins, a co-author of *Introduction to Machine Translation*, the quality of affordable MT in the European market is not as particularly high as Japanese counterparts. [Hutchins, 99]

Owing to the fact that any customer can pick up a CMT at any software shop in Japan, a user's guide is the most valuable information source for users, partly because it is also the only source of information. However, very few user's guides give a thorough explanation about the weaknesses of CMT and information about how to make the most of it. For instance, it asks users to cut long sentences into shorter sentences and to add subjects and objects if they are omitted. This may sound helpful, but unless an explanation is provided as to what the subject and object are in a sentence, this type of guidance is not practical at all.

If a user is an MT researcher, like the author who understands the weaknesses and strengths of the MT, this single guidance page may suffice. Yet, most users are not MT specialists. They require more information and an explanation in order to comprehend MT capabilities. The developer should remember that users of MT are not necessarily well educated or well informed in the field of linguistics. The lack of quality in user's manuals create disappointment among CMT users who had high expectations the product when purchasing. This report will enhance the use of MT in the manufacturing industry.

## 2 HAIL dictionary and Creating a User's Dictionary

Creating a user's dictionary is a vital part of customisation of MT. The dictionary of HAIL stores 100,000 words. A user's dictionary can be created in order to improve the quality of translation. NEC, a developer of HAIL, regularly updates the dictionary for newly coined words and offers it to customers over the Internet. Also, user's dictionaries are shared and the company makes them available to other users on the Internet. The company is hoping to offer a facility which enables their users to share the user's dictionary from the MT products of their competitors. [Kamei et al 97] Yet, it is not sufficient and the current situation require the building of user's dictionary.

It is easy to say that you can create your own dictionary and then you will receive a high quality translation, however creating a user's dictionary is not as simple as it may look. First of all, knowledge must be possessed about basic Japanese and English grammar in order to build a useful dictionary.

For example, if the user wants to enter a Japanese verb, he/she ought to know the type of verb: There are eight types of verbs in Japanese depending on how they conjugate. Therefore, it is essential for the user to acquire basic Japanese grammar as well as English grammar. It is vital for him/her to know a part of speech in both Japanese and English, to produce a user's dictionary. So, how can this be expected of a user who does not know much about English grammar? Probably he/she would claim that this is the reason why he/she bought this MT package. If they do, they would not need to use MT. The guidance is needed for a user to create a user's dictionary. Also, it has to be pointed out that native Japanese engineers do not necessarily have the sufficient knowledge of Japanese grammar.



**youni ki wo tsukeru.**  
 p. p. n. p. v  
 to care take

(Translation by a human translator:  
 e.g.1 Care should be taken for not letting  
 operators wait for the material.)

The expression in the second example is always  
 used in an operation manual. It is grammatically  
 wrong in Japanese because it uses the wrong  
 particle 'ni' (meaning 'at'), but it should use 'wo'  
 (object marker) instead.

e.g.2 **Hako ni narabe.**  
 n p v  
 box in lineup

(Place the carton on the line.)

The last word 'narabe' also has a missing  
 element It misses out the ending 'ru'. 'Narabe'  
 is the stem of the verb 'naraberu'. If the sentence  
 has to be rewritten in formal Japanese, it will be  
 'Hako wo naraberu'.

We can easily change the definition of a word in a  
 dictionary of HAI. Therefore, the author  
 changed the definition of a word when necessary.  
 For example, HAI translated the word 'okyaku  
 san' into 'a guest'. This word can be translated  
 into a few Japanese words, but 'a customer' is  
 more appropriate for MME document.

When HAI comes across with a new word, it  
 leaves it out. Consequently, the translated  
 sentence includes odd Japanese word in Japanese  
 characters. Since Makita uses special  
 terminology, whenever this case arises, it has to  
 be post-edited and register the word in the user's  
 dictionary. Here are a couple examples;

Table 2:

Japanese word (MJ document)	Translated word by HAI	MME term (English)
Shijisho	Indication book	Instruction manual
Shimetsuke toroku	Fastening-up torque	tightening torque

### 4.1.2 (1) Long Sentences

Inputting short sentences for MT is a typical  
 instruction written in the user's guide when we  
 buy an MT system. However, HAI could  
 manage to translate rather long sentences without  
 any pre-editing as in the following example.

[n: noun p: particle v: verb  
 adj. adjective adv. adverb]  
 e.g.3

**Akafuda-sakusen to wa,**  
 n pp  
 red card strategy as for

**akai fuda wo tsukatte**  
 adj. n p v  
 red tag attach

**kojo ni habikotteiru**  
 n p v  
 factory in be rampant

**aka wo**  
 n p  
 dirt

**dare ni demo**  
 n p p  
 everyone to even

**wakaru yo ni suru**  
 v p p v  
 understand in order to do

**seiri no yarikata desu.**  
 n p n p  
 sorting out of method it is

HAI produced the following translation for the  
 above Japanese sentence in a document. 'The red  
 card strategy is a way of the arrangement to make  
 anyone understand the dirt which is being  
 rampant at the factory using the red card.'

Although the translation was not very natural, it is understandable by British engineers and team leaders in a factory. It has to be noted that there are always about ten Japanese members of staff at MME, and they can explain if British members are not sure the meaning of the translation. However, the experiment shows that a sentence which does not have the subject and object with too many embedding phrases is not to be translated satisfactorily.

#### 4.1.2 (2) Changing Original Sentences

It is always the case that Japanese engineers do not receive the formal technical writing course, so they often write ambiguous sentences in. Also, the fact that a subject and object can be omitted in a Japanese sentence allows them to write more ambiguous sentences.

For example, there are many sentences without subjects and objects in Makita document, which need pre-edition in order to be processed by MT, as in sentence 87:

e.g.4

**Kojo dewa okyakusan ni**  
 n pp n p  
 factory in customers by

**yorokonde itadakeru shinamono**  
 v p n  
 feel happy (polite ending) products

**wo tukuru tameni,**  
 p v p p  
 make in order to

**mainichi isshoukenmei desu.**  
 n adv p  
 everyday hard

The above sentence misses the subject of the sentence, *watashi tachi*. (we). Therefore, the translated sentence generated by HAI did not convey the correct meaning. After adding the subject of the sentence, the MT produced the acceptable translation as follows:

‘At the factory, to make the product which it is possible for the customer to be glad about, we are strenuous every day.’

#### 4.1.2.(3) Imperative Sentences

Instruction manuals are full of imperative sentences, ‘-suru koto’ is the phrase used as a command form. ‘Koto’ is a nominalizer attached after a declarative statement. However, ‘koto’ also has a different meaning; ‘thing’ in Japanese. HAI translated ‘koto’ into the latter meaning, therefore the pre-editor ought to rewrite the whole imperative sentences where the sentence is rewritten using imperative form. (Te-form of verb + *kudasai*)

#### 4.1.2 (4) Phrasal Verbs and Idiomatic Expression

In case of translating idiomatic expressions, HAI has disadvantage. For example, in sentence 11, HAI broke up the idiomatic expression into individual words and translated each word separately. Therefore, the sentence is not meaningful and needs to be post-edited.

e.g 5 .....**ato wo tatanai...**

HAI translation is ‘*doesn’t cut off the back*’.

HAI could not recognise the phrasal verb ‘*ato wo tatanai*’ (never stop), so it translated word by word.

#### 4.1.2(5) Problems of Double Subjects

In most languages, a simple sentence only has one subjective case, whereas in Japanese, many adjectives and some verbs can dominate two surface subjective cases within a simple sentence. This is called double-subject construction which confuses MT when analysing sentences. [Oku 96]

## 5 Suggestions in Document Preparation (Writing Rules)

The quality of input sentences affect the quality of translation, so staff at MME should be provided with adequate information on this subject. It is well known that Natural Language is dynamic and flexible, and ambiguity is inevitable. We must alert CMT users to this point strongly, and educate them how to utilise CMT in their

organisations. CMT is too good to be ignored or abandoned, though it does not have human intelligence. Customers will be disappointed when a product fails to meet expectation. Therefore, adequate PR is needed for penetrating the market in order to enhance MT user awareness. The following are basic writing rules for MME.

Here are basic rules for members of staff at MME to make most of the commercial machine translation system, HAI.

1. *Input a grammatical sentence.*

(Do not omit the subject or object of a sentence, or do not input a fragment of a sentence.)

2. *Input the correct kanji.*

(a person who uses HAI in the company should check whether he/she is inputting the correct kanji or not.)

3. *Avoid using a phrasal verb.*

4. *Use a hiragana or kanji character, and avoid using katakana for nouns which are not borrowed words from foreign languages.*

(HAI sometimes does not recognise a word written in katakana character.)

5. *Avoid using a comma, but use a word 'to' (and).*

6. *Avoid using kanji for the verb 'okonau' because it tends to mistranslate.*

7. *Avoid a sentence with embedded clauses. (Make it simple and split up into more than one sentence.)*

8. *Specify singular or plural if necessary. (HAI treats a noun as a singular form.)*

9. *Avoid using 'koto' or dictionary form when making an imperative sentence. (Although it is a very typical way of making up an imperative form in Japanese, HAI has difficulty in translating these types of imperative sentences.)*

10. *Use a sentence from example translation database of HAI when translating the opening and closing part of a business letter.*

(Set phrases from typical Japanese business letters should be translated as a whole sentence, but not word by word.)

11. *Avoid using a word which has more than one meaning.*

(It is better to use the consistent terminology, so review the definition of well-used words within the company.)

12. *Input Makita terminology.*

(Most of terminology are not in the HAI dictionary, therefore it is necessary to input them in the dictionary.)

13. *Remember that HAI dictionary only carries one definition for one word.*

(It takes a while until the system is tuned. Since each entry of a lexicon from HAI dictionary cannot store more than one meaning, it is more advisable to use only in one domain such as translating operation manual.)

## 6 Summary of the Report

The result of the study can be utilised for enhancing the productivity and quality of translation within a small to medium sized organisations such as MME. As globalization of Japanese manufacturers becomes common, problems caused by lack of communication can seriously affect management. MT will help decrease time and cost for rough translation especially for those companies in the competitive manufacturing industry. British employees working at Japanese factories are used to listening to English spoken by Japanese staff. Translated documents can be understood even it is not flawless sentences. The introduction of the new system may look exciting and promising in a company, but we must make sure that we use it as a part of our work. The author recommends that MME should appoint one person who is responsible for MT and updates the system constantly. He/she must report regularly how they are utilised within the company. Also, he/she should give an induction for employees who wish to use the MT system. The induction includes training them to master the writing rules. In this way, MT systems can find their place in a company. Otherwise, the novelty will soon wear off and only few people can remember that MT is in the organisation.

## REFERENCE

- [Hutchins & Somers 92] Hutchins, W. J., Somers, H. L. An Introduction to Machine Translation. Academic Press, UK, 1992.
- [Hutchins 99] Hutchins, W.J., Interview with Mr. Hutchins 10 April 99 by e-mail]
- [Ikehara et al 94] Ikehara, S., Shirai, S., and Ogura, K. Criteria for evaluating the linguistic quality of Japanese to English machine translations. Journal for Japanese Society for Artificial Intelligence. Vol.9 1994
- [Isahara, 95] Isahara, H. JEIDA's Test-Sets for Quality Evaluation of MT Systems - Technical Evaluation from the Developer's Point of View - , Proc. Of MT Summit V, Luxembourg, July 19 - 22, 1995
- [Johnson 97] Johnson, I., Personal Translation Applications Proc. Of ASLIB 97 Translating and the Computer 19, November 13 & 14, 1997, London
- [Jones 98] Jones, I., A translation service to maximize Quality and efficiency. Proc. Of ASLIB 98 Translating and the Computer 20, November 12 & 13, 1997, London
- [Kamei et al 97] Kamei, S., Itoh, E., Fujii, M., Hirai, T., Saitoh, Y., Takahashi, M., Hiyama, T., Muraki, K., Sharable Formats and Their Supporting Environments for Exchanging User Dictionaries Among Different MT Systems as a Part of AAMT Activities. Proc. Machine Translation Summit VI, pp132-141, 1997, San Diego, USA
- [Nikkei 28 May 99] Tatakau Nihongo, (Fighting Japanese), p.1, 28 May 1999, Nikkei Shinbun Newspaper, Tokyo, Japan
- [Nagao 96] Nagao, M. (Editor and author), Sato, S., Kurohashi, S., Tsunoda, T., Shizen Gengo Shori (Natural Language Processing) in Japanese 1996 Iwanami Shoten Publishing Company, Tokyo, Japan
- [Oku 96] Oku, M, Analysing Japanese Double-Subject Construction having an Adjective Predicate, COLING 96 Proc. Pp. 865-870 1996]
- [Povlsen et al, 98 ] Povlsen, C., Underwood, N., Music, B., Neville, A. Evaluating Text-type Suitability for Machine Translation: a Case Study on an English-Danish MT System Proc. Of the First International Conference on Language Resources and Evaluation, Granada pp.27-31, 28-30 May, 1998, Spain,
- [Shirai et al 98] Shirai, S., Yokoi, A., Okuyama, N, Kawamura, M, Ikehara, S., The Support System for end users to create dictionary of English-Japanese Combining Pattern. Proc. Of the Fourth Annual Meeting of the Association for Natural Language Processing in Japan pp.568-571, Kyushu, Japan, March 23-26, 1998
- [White 97] White, John, MT Evaluation, Tutorial given at MT Summit VI, San Diego: Association for Machine Translation in the Americas, 29 Oct.- 1 Nov. 1997, San Diego, USA,