THE PRESENT SITUATION AND PROSPECTS FOR MT-VAN SERVICE

Tsutomu Tanaka
Fujitsu Facom Information
Processing Corporation
6-1-11, Shinbashi, Minato-ku, Tokyo, Japan
Zip Code 105

Introduction

The International Forum on Translation Technology will be held on the theme of cooperative work between humans and the computer. As an example of the cooperative work, the present situation and prospects for the machine translation value-added network (MT-VAN) service will be described below. In the MT-VAN, users, the machine translation systems and the translators are satisfactory linked by the computer network.

I. Outline

Four years have been past since the first machine translation system appeared in the Japanese market. Since then, various trials had been made within many companies. And, nowaday, a common usage of the machine translations are bound to be established, such as "tools for promoting efficiencies and qualities of human transla- tions works". The machine translation systems, at the same time, are used on trial basis at several translation companies and practi- cal knowledges for machine translation technique has been syst- ematically accumulated. On the other hand, while internationaliza- tion of societies are promoted, the translation market is now growing to be on the scale of 700 billion yen to 1 trillion yen annual. Problems of shortage of translators become more and more serious. At the same time, the personal computers and word processors become popular (sold estimated several millions), correspondingly, the usage of the electronic mail system is increased. As the result, the electronization of documents are rap- idly increased.

The MT-VAN service allows the interconnections between the machine translation system, electronic document, and translators. An attempt has been made in the translation business by several companies. That is, the translation works is divided between the human translator and computers to enhance the efficiency and good quality of translations. The machine translation systems do have the following advantages: (1) unification of terminology and style and (2) reduction of translation costs by mass processing and increased efficiency. However, in the current situation, a machine translation system required human help. In other words, by combina- tion with human skills, it will prow to be useful.

The outline of the MT-VAN systems is the following: 1) the user sends an electronic document to a mailbox in the electronic mail system. 2) the pre-editor modifies the document. 3) put the docu-

ment into the machine translation system. 4) the post-editor edits
the machine translation result. 5) the translation document is put
into the mailbox. 6) the user then takes the translation out of the
mailbox. There are variations in the processing in terms of the
pre- or post-editing between the several MT-VAN system.

For those which are not suitable for the electronic mail
system, that is, translating large amount of documents or such docu-
ments to be printed, etc. There are also services for making user's
private dictionary. Though, at present, the most of the transla-
tion systems are dealing only with English - Japanese or vice versa,
there are future plans to provide services for other languages.
Several computer manufacturers producing the machine translation
systems and computers which the machine translation systems are used
varies from personal computers to a large general purpose computers.

Currently, there are several companies that are providing
MT-VAN translation services: the Fujitsu FIP (an information proc-
essing company), Japan Science Foundation, IBS (a translation
company) and Oaks (a information processing company), etc.

One of the objects of the MT-VAN services is opening up a new
translation market for large and relatively recent technical
abstract and text data bases. Further, by expanding the continuous
practical use of the machine translation system, it is possible to
give feedbacks to the systems and to accumulate technical terms in
the technical term dictionaries. Also, by promoting division of
works and opening up new business opportunities in the field of
translation, it may be possible to promote the translation business
as a streamline and established business field. This is to say
that, currently, the translation business tends to be an attachment
to certain industries and many translation companies are
subsidiaries to other corporations.

In section II, MT-VAN services will be described in more con-
crete terms based on specific examples.

II. Examples of MT-VAN Services

The ATLAS-MAIL is a MT-VAN service which is operated by Fujitsu
FIP in corporation with Inter Group, a translation company.

1. The outline of the service

The translation services are provided involving Japanese to
English and English to Japanese. The translations are applicable in
the technical or industrial fields. In the near future, financial
and economic fields will also be considered. Currently, the types
of documents suited for the machine translation system are: manuals,
specification, technical report, technical papers and abstracts.
The machine translation system used under this service is called
ATLAS. And the computer, under the operations is Fujitsu's large
general purpose computer. The dictionaries provided are the basic
dictionary with 50,000 words and the technical term dictionaries

with 300,000 words which are divided into 13 categories. Also,
there are dictionaries developed for specific users. (user's pri-
vate dictionaries).

2. Process of machine translation

     The document to be translated go through the following
processes; the receipt of the manuscript, inputting, pre-processing,
translating in the machine translation system, post-editing and
retyping the post edited then, outputting of the completed
translation. As additional work, there are registration to and
modification of the dictionary.

     When the manuscripts are received, only 15 percent of these
documents will go through the electronic mail system. The reasons
why the electronic mail system is not widely use is that despite of
popularizations of electronic document processing media, still large
number of translation manuscripts sent in the forms of printed or
hand-written. The manuscripts stored in floppy disks and magnetic
tapes are only 20 percent but these are gradually increasing.

     The inputting of the manuscripts in the magnetic media will be
a large problem in terms of efficiency and cost. Currently, optical
character readers are used in some cases but most of the inputs are
done manually. For those reasons, the total cost of machine trans-
lation increases and delivery prolongs.

     In some cases, pre-processing have to be done after the docu-
ment is stored in electronic media, which include changing the docu-
ment format, extracting and/or converting of figures and tables.
However, the pre-editing the original document for the better fit
into the machine translation system is not carried out.

     The machine translation processes is done by a batch
processing. When the translated document can be categorized into a
certain technical field, a categorized technical term dictionary is
allocated. If the category cannot be specified, a general
dictionary with technical terms is used. When user's private
dictionaries are used, the priority for the use of these
dictionaries is designated as the highest.

     The machine translation results are output in the form of
corresponding lists. Then, the post-editor goes through the list
and corrects sentences and words. The post-editor also uses the
output words for making sentences in order to complete the
translation.

     The post edited version of the translation is finally input and
stored in the mailbox. The user can then take the translation out
of the mail box. When a large amount of the manuscripts and
translated documents need to be transported, often delivery service
is used.

     The users, during the translation process, can check the pro-
gress from to time, using the electronic mail system.

In the post-editing process, the technical terms are accumulated during the correction process. With necessary attribute information, the technical terms are added or altered to the technical term dictionary. For documents with indexes or with certain vocabulary lists, registered these terms in dictionary prior to processing in the machine translation system.

3. Processing efficiency

The processing efficiency is the most important factor which is reflected into the translation cost in the commercial translation service. However, the term "processing efficiency" used in the commercial service has different view points from those involving in the research and supplying the machine translation system. The researchers emphasizes "processing efficiency" on the quality and quantity of text which were translated. On the other hand, the system manufacturers emphasizes on the processes time necessary taken in view of the end users. However, in the commercial translation service, the whole process from the receipt of the translation manuscript to the deliver of translations are taken accounts. In this sense, the efficiency in long-term and overall process should be pursued in terms of facility and personnel assignment with the prerequisite of processing large amount of documents.

Distinctive features of this system is that there is no pre-editing. Of course, the quality of the translation can be improved in great extent if it has done. However, when processing a large amount and a variety of documents, it is difficult to establish an environment for pre-editing, including training pre-editors. Also, when experienced editors are assigned, the post-editing is far more effective in terms of the cost performance than the pre-editing. Basically, what is need to be done is to write clear and simple documents. A good technological documents are considered to be suited for machine translation. However, in a commercial service, it is based on the user written documents and it is difficult to regulate how the document is written. What is considered then is differentiating the prices depending on how well a document is fit for machine translation. This will promote making of simple and clear technical documents.

In general, the machine translation system is designed for those end users used it within an enterprise. However, in this system, as there is an prerequisite that a post-editor uses the system without any pre-editing, the alterations to the machine translation system was made. The major alterations are: for a long sentence, the corresponding word lists of the target language are output instead of the sentence that is translated, the dictionary identification number is displayed for the technical terms in order to confirm on the reliability; and corresponding list format is altered for editing the output list. By these, efficiency of the post-editing is improved and reduction of machine cost is achieved. Presently, automatic divisions of long sentences and improvement in efficiency by adding reliability scales on the output sentence are planned.

The maintenance of the technical term dictionary is very impor-
tant factor in order to improve the quality of machine translation
and the unification of terminology. However, the number of words
stored in the dictionary is 300,000 words, which is still
insufficient. However, to supplement this insufficiency in a short
period of time is impossible in terms of expenses and the volume of
works that are involved. This system improves upon accumulating the
technical terms over a certain period of time and registers them to
the dictionary, while carrying out the service. However, as ex-
plained in the processing section, extraction of the unregistered
words from the manuscript and registering them prior to putting the
manuscript through machine translation-has not been done due to cost
and operational considerations. On the other hand, providing user's
private dictionary in which user peculiar terminology and special
terminology which has a high probability for usage can be registered
help dedicating the improvement of the quality of translation.
Furthermore, in order to reduce the cost of registration to the
dictionary, a dictionary registration utility is provided in which
registration to dictionary is possible by making use of terminology
information in the existing dictionaries.

The increased efficiency in the post-editing processes is a key
factor of this system as well as the capability for translation.
Initially, it was difficult to reduce the number of processes in the
post-editing. The cause of this was the lack of know-hows and post-
editing experiences. Due to pre-set attitudes of the post-editors
toward the machine translation, it was difficult to change the
translation results. Currently, even those translated sentences
other than which are usable with little post-editing, results of the
dictionary referencing and partial translation will be used for
post-editing processes and it recognized as effective. For the
post-editors, those with aptitude for post-editing are selected
among the translators. However, those who go through post-editor
training courses at translation companies who acquire translation
capability are increasing.

The manuscripts suitable to be translated in the system sup-
plied currently are manuals and specifications. These manuscripts
include short sentences with clear meanings which are suitable for
the machine translation systems. By improving the efficiency of
post-editing through selection of technical terms and for the use of
translated documents for speedy outlining of a large quantity of
documents, magazines and abstracts in specialized fields can be put
through machine translation.

In order to improve the efficiency, those works such as receiv-
ing the manuscripts and delivery of translated manuscripts, which
are not directly related to the machine translation systems must
also be considered. The use of electronic mail system will be use-
ful for shortening the overall machine translation processes but the
service is not used so much currently. However, there are plans to
make more use of electronic mail system including businesses with
overseas.

As mentioned above, there are various measures for improving the efficiency. By MT-VAN service, total work processes in the translation which requires standard quality is reduced about 40 percent less than those by human translation. In about three to four years later, an estimated reduction of about 65 percent of the current work processes will be expected.

III. Further Tasks of MT-VAN Service

As the future tasks of MT-VAN service, standardization of documents, improving the level of machine translation systems, improving and developing dictionaries, and training of post-editors will be covered.

As for the standardization of documents, it is important that guidelines for writing technical documents, such as manuals, papers and abstracts should be widely established, and skills of technical writing should be introduced in the production of documents at corporations and public organizations. Over the long run, it is important that writing skills for simple and clear technical documents are to be taught in schools. Also, for making more electronic documents, the use of wordprocessor should be promoted. At the same time, the standardization of electronic media in terms of formats should be promoted.

In the machine translation system, the accuracy must be improved at a more basic level. The contest should be introduced and meaning analysis at a more higher level should be included. Through these, more complex and ambiguous sentences should be challenged. The translation among multiplicity of languages should be realized. The domain translation system in which sentences with different grammars are translated will be examined in practical terms. At the same time, although not related with translation capabilities in the direct manner, efficient ways of translation for tables and charts and partial translation in relation to other systems should be established in the system as a whole.

In terms of development of dictionaries, the number of technical terms sufficient for translation of scientific and technical documents are said to be five to six million words. Building and management of such a dictionary by a one corporation would be difficult especially in terms of costs involved. Open coordination system with both public and private sector organizations involved should be established for building, management, and provision of dictionaries and such policies should be promoted strongly.

The roles played by the post-editing in the machine translation system will become more and more important in the future. Along with the increase in the amount of translated materials and improvement in the quality of machine translation, training of capable post-editors with translation capability would be very important for the MT-VAN service in the future. It would also be necessary to establish the post-editing as a socially recognized occupation.

Conclusion

    The MT-VAN service as a new business tries to combine the machine translation with human translation, based on the possibility in the future and limitation at present. The service aims for combined effect of supplementing and coordinating capabilities of the machine translation system and human translation. The promotion of use of machine translation system in the actual business will contribute towards the improvement of the system. We hope this will also be a support for the researchers in this field. Also, we hope that those who are engaged in the translation business will be interested in machine translation. We would like to provide suggestion for better and more efficient use of machine translation. As the internationalization of our societies continue, we hope the machine translation will contribute to expansion of the translation market, which is the basis for the internationalization. At the same time, we hope the machine translation will contribute positively for the translation business.