Japanese View of the Future of Machine Translation

Makato NAGAO
Kyoto University
Kyoto, Japan

1. About the ALPAC Report

1.1 Summary of ideas and proposals in the ALPAC Report

The ALPAC Report is a report written by the Automatic Language Processing Advisory Committee established within the Division of Behavioral Sciences, National Academy of Sciences, U.S. National Research Council. The report consists of the committee's recommendations based on the results of an examination regarding the continuance of support funds for machine translation research, which has been promoted by the U.S. government to mechanize translation.

The ALPAC report is based on a broad-based survey and examination of translation and machine translation activities. The results are summarized in fifteen items. The last item comprises recommendations for tasks requiring concentrated efforts. The report also contains detailed considerations and data in nineteen categories as appendixes.

The ALPAC report points out the following as the condition of translation in U.S. government organizations in 1965.

(1) There is a sufficient translation supply.
(2) The translation market is small.

The report then points out the following as the condition of machine translation.
(3) The translation quality is poor.
(4) Machine translation does not necessarily lower the cost.

From these points, the report concludes the following.
(5) Machine translation is unnecessary.
(6) Machine translation will not justify the cost.
(7) The basic technology is not developed for machine translation.

The document then recommends that the following points should be emphatically promoted.
(8) Research in the area of computational linguistics
(9) Improvement in translation work

An outline of each item of the ALPAC Report is contained in Appendix 1.

1.2 Views on the judgment by the ALPAC Report

When the ALPAC Report came out, development work for machine translation had been going on for ten years in the United States. With this background, the ALPAC Report surveyed and examined the actual conditions of translation and machine translation-related work at U.S. government organizations.

The ALPAC report does not dwell directly upon the issue of financial support for machine translation. But by recommending changes for the items emphatically to be promoted, current financial support has been terminated for machine translation.

The ALPAC Report is based on the following insights. (1) There is no danger of a shortage of translations as a result of a scarcity of U.S. government translators, and (2) machine translation cannot be expected to be put into practical use in the near future.

The situation is radically different today. We must think about translation and machine translation from a broad, long-term point of view in terms of the large-scale, in-depth communication needed in international society.

Regarding machine translation technology, steady progress has been made in the following areas: proposals for a new computational linguistic theory, progress in natural-language processing technology, improvement and development of language data including dictionaries and grammar, the support system for development and maintenance, etc. Although it is true that the language phenomena are limited that the current machine translation technology can handle, it is clear that machine translation is steadily making progress. This can be seen from the fact that practical machine translation systems are appearing and being utilized in the United States, Europe, and Japan. We should also be looking at machine translation technology as the major technology supporting the highly advanced information society of the future.

Therefore, if we apply the ALPAC Report's concept and recommendations to our current situation, and if we assume that the report was a judgment for future prospects, we must conclude that the report was mistaken both in the prospects for an increase in the demand for translation, and in the progress of translation technology. Many of the concepts contained in the ALPAC Report, however, pointed out many problem areas for not only machine translation but also for technological development in natural language processing thus providing a guideline for research and development in this field.

The crux of the report is that the development of a machine translation system from a simple, materialistic point of view amounts to nothing without clarifying the technological phases and evaluation standard. In other words, the report emphasizes that we should not neglect activities for establishing basic theory and technology.

On the other hand, the ALPAC Report almost totally neglects discussions on how to utilize machine translation. With the progress of technology and internationalization, we cannot avoid resorting to machine translation. Therefore, we need to probe for the most efficient and effective method of utilization of machine translation that matches both the technological level and social conditions.

In sum, we should not be held captive by the conclusions and recommendations of the ALPAC Report. From now on, we must recognize the increase in the amount and variety of demand for translation, as well as the progress in the technology for machine translation. While paying attention to the true meaning, wisdom, and insight of the ALPAC Report, we must enlarge upon the capabilities of the practical system and improve upon the basic technology. This must be done with the cooperation of the countries of the world in reflection of societal demands and progress in related technologies.

1.3 The ALPAC Report viewed from the current status of technology

In the ALPAC Report, the results of a survey on translation activities in U.S. government organizations were used for making judgments on the proper research support of machine translation. During, however, the twenty years since the report was written, there have been dramatic changes in the demand for translation, the actual business situations dealing with translation, and the situation concerning machine translation. For these reasons, we the Machine Translation Investigation Committee of Japan Electronic Industry Development Association, decided to examine item by item the matters taken up by the ALPAC Report from todays technological point of view, and to comment upon each point. Detailed examination results on some of the items can be found in the appendixes: in Appendix 2, the changes in the general situation; in Appendix 3, the problem areas; and in Appendix 4, the comments. These are taken up for each topic included in the ALPAC Report.

a) Manual translation

The ALPAC Report points out that the efficiency of translation work can be improved by careful selection of the translators, improvement in the translation environment, use of better translation methods, etc. These recommendations are still valid. Especially, the use of a translators' work bench that has such functions as an electronic dictionary and machine translation functions, will considerably improve the efficiency of translation.

b) Types of translator employment

By the use of a translator work bench, a situation is emerging where the document to be translated can be received by the translator on-line and subsequently returned in the same fashion. This may herald drastic changes in the manner of translator employment. At least, whether machine translation will or will not be used widely will exert a large influence on the types of translators in the future.

c) English as the language of science

It is apparent that English, when compared to other languages, is the tongue most often used for writing scientific and technological documents. But the views appearing in the ALPAC Report mainly reflect American opinions. From the points of view of Japan and other countries—and also from the point of view of the technological expansion in most of the countries and the technology's precipitation into the daily lives of the people—we could hardly say that it is sufficient to have English as the sole scientific language. Contrariwise, present-day technological information is gleaned from documents written in the mother tongue of those who read the document.

d) Time required for scientists to learn Russian

The ALPAC Report states that, if one wants to read a document written in a foreign language, one should learn that language. That is one solution, but there are other solutions, such as translation, for overcoming the language barrier. Looking toward the twenty-first century, it can be argued that every language has its own value. This is a natural condition, and in this sense, the assistance from machine translation will be indispensable. Therefore, the expectation is great for a machine translation system capable of handling multiple languages.

e) Translations for the U.S. government

The ALPAC Report lists the translation costs in the U.S. government at the time the report was written. The costs, however, have greatly changed since then. For this reason it is necessary to re-evaluate the translation costs, including the latent demands that will materialize once machine translation becomes practical. Today, the demand is for translation of Japanese and European information, rather than Russian data as was the case twenty years ago. In this sense, Japan will need to expend more efforts in this area.

f) The number of government translators

The ALPAC Report lists the number of translators working in U.S. government organizations at the time the report was written. Both at the present time and in the future, an increased need is expected for English-Japanese translators in the U.S. This is especially true for experienced translators, and it is desirable for them to have skills in editing machine translation results.

g) Expenditures for translation

At the time the ALPAC Report was written, the U.S. government spent $2.2 million on translation—a very small figure in proportion to other funds disbursed by the government. Recently Washington established the Japanese Document Act, but the amount budgeted is very modest. Some Japanese corporations, on the other hand, are spending huge amounts on machine translation. In this sense, the United States should be spending considerably more for machine translation than the current amount.

h) Is there a shortage of translators or translations?

The ALPAC Report shows that at the time it was written there was no shortage of translators—either inside or outside the U.S. government. In this situation, times have certainly changed. The United States nowadays must pay more attention to foreign-language information. Therefore, it seems that the view toward translation has changed in the United States.

i) Access to translated documents

The ALPAC Report points out that many of the translated documents are not widely read. This is what happened when the quantity of translated documents was small. It is reasonable to assume that when the number of translated documents exceed a certain amount, the frequency of documentation utilization will increase at a rapid pace.

j) The crucial problems of translation

The ALPAC Report lists three crucial problems of translation: (1) translation quality, (2) translation speed, and (3) translation cost. The report compares manual translation and machine translation in these categories and concludes that manual translation is superior to machine translation. These three crucial problems still apply to the translation work done today. It has been proved, however, that machine translation is far superior to manual translation in speed. Also, it is expected that machine translation will be superior economically to manual translation.

k) The present state of machine translation

The ALPAC Report points out the following concerning machine translation. (1) There are limitations placed on the materials to be translated. (2) The translation quality is poor and this makes it necessary to post-edit the product. (3) Work economy does not yet exist. (4) There is no prospect for the technology ultimately to be usable. (5) There is no need for machine translation.

Presently, there has been great progress in (1) and (2), although problems still exist. There are very promising prospects for (3) and (4). Regarding (5), the situation has changed completely and the present needs for machine translation are very great.

l) Machine-aided translation at Mannheim and Luxembourg

The ALPAC Report shows an example of machine-aided translation at the time the report was written. Currently, because of progress in computers and computer-utilization technology, translation-supporting equipment such as word processors and electronic dictionaries have seen considerable progress. In the future, however, when translators' work benches are completed with a built-in machine translation system as a machine-aided translation process, we will have an improved system that will markedly improve both speed and efficiency. Such a system will be widely used by professional translators. Especially, it will become possible for us to turn out the final product or document quickly and at low cost by combining the machine-aided translation system with electronic publishing systems.

m) Automatic language-processing and computational linguistics

The ALPAC Report points out the following regarding machine translation systems at the time the report was written. (1) The machine-translated results do not justify the developmental investment. (2) The computational linguistics, that forms the basis for machine translation, are insufficiently developed. The report concludes that research on such computational linguistics must be promoted.

Clarification of the language mechanism is very difficult. It is a huge task that requires a research outlay equivalent to that for cerebral and space research. On the other hand, it would be implausible to conclude that the short-term results do not justify the research and development investment. In view of the fact that we are making steady technological progress—even if slowly—we only come to appreciate the enormity of the task. If one adopts such a point of view, then the results we are seeing today are indeed splendid.

n) Avenues to translation improvement

The ALPAC Report points out the following concerning machine translation. (1) We need research and development with a long-range orientation. (2) There is room for improvement in the translation-related work. (3) No established method exists for translation evaluation. These points evolve along with technological developments and social changes. The ALPAC Report's three points, however, still remain cogent today. Regarding (1), the long-term point of view for machine translation research, some people are calling for a machine translation system that is complete and practical by itself. But we need to stress the long-range view of machine translation system research and development, recognizing that we are achieving technological development in diverse stages. Regarding (3) evaluation, as yet we have not obtained a very universal evaluation method. But since many machine translation systems are coming into use, better evaluation criteria most certainly will emerge.

o) Recommendations

Based on the above considerations, the ALPAC Report makes the following recommendations. (1) Research in the field of computational linguistics should be promoted. (2) Research should also be encouraged for improvement of translation-related work.

When we consider the current situation, we can see that we must promote research and development in a multiplicity of areas. Especially, because machine translation has been put to practical use despite deficiencies in some of its functions, we must elucidate the technological development phases. At the same time, we should emphasize the need for promoting basic research to push the current phases on to higher levels. Regarding the improvements of translation-related work, we must completely change the work processes with the prerequisite of active utilization of machine translation systems.

2. **Current Status of Translation in Japan**

2.1 **Types of translation**

Based on our survey results, we will show examples of types of translation done in Japan. The details can be found in Appendix 6.

a) Translation systems

The translations are done internally within a company or they can be sent out. Of the translations done internally, some are done in the translation department or section, and some are translated individually. Those subcontracted are processed at translation companies.

Many companies having their own in-house translation sections also subcontract translation work. Some orders are placed overseas. Other companies merely order editing from foreign translation companies.

According to data of the Japan Translator Federation, there are about 700 translation companies within Japan. The exact figure is not known but it quite possible that there are at least twice that estimate. The sizes of translation companies vary considerably. The number of major translating firms, however, are relatively few and perhaps some twenty exist. In other words, there are a multiplicity of small companies.

The total sales of the largest translation companies range between ¥1 billion and ¥2 billion. Translation companies of this size number ten or twenty. The sales totals for small to medium-size companies are insubstantial.

The quantity of in-house translation done in a firm's translation department or section should be made clear. Concrete figures, however, are not obtainable. Neither can totals be obtained for the total amount of in-house translation work. When a large company has a division system, the total amount of translation done at the company as a whole is not usually known.

As we have seen, the commercial volume for the translation business as a whole is not accurately tabulated. Also, there seems to be a large amount of latent demand for translation. Of this, there is a great deal of information available in the form of databases. The latent translation demand in this form is incomparably large compared with the quantity translation currently being done.

b) Translators

Of the translators in Japan, 90 percent are Japanese nationals. The translators working at the translation companies were mostly free-lancers. In other words, they are subcontractors for translation companies.

Of the free-lancers, some have individual contracts with the translation companies and others have group contracts. Some of the groups that have contracts with translation companies are much larger than some of the smaller translation companies. Many translators are part-timers, of whom many are housewives.

There are schools that train translators. These institutions provide training in language used for translation as well as translation techniques. Many translators, however, had no such training.

The knowledge required of translators includes a knowledge of the language of the translation, knowledge in certain specialized areas, and skills for technical writing. Of these skills, many translators acquire their translation linguistic knowledge during the course of their schooling and occupational life. Others acquire language skills while sojourning overseas. In other words, few translators acquire their language skills solely from schools for translators.

As for the specialized knowledge needed for translating, the translators often acquire such knowledge while working in a specialized field or through on-the-job training as a translator. Many translators acquire their skills in technical writing while working as translators.

Translating skills, therefore, are mostly acquired through working as a translator, although some learn the art while being trained as a translator's assistant.

There is no national qualification for a translator. But private certification does exist. At some translation companies, the translators are rated into different classes: for example, A, B, and C, according to their capabilities. Their wages accord with these classifications, and these levels are often reflected in their annual employment income.

Translators are often registered according to their specialty. When subcontracting, the translation companies dispatch work to the translators according to their specialties.

As far as the quality of translation goes, the translators' technical writing abilities play a crucial role. Many of the documents to be translated require technical writing skills because they are written from a technical point of view and are not particularly user-oriented. Consequently, translators are often required to be proficient in areas in which they do not possess specialized knowledge.

As far as the technical terminology is concerned, a list of technical words is sometimes obtained from the client. This is of great assistance when specialized technical knowledge is involved.

There is no way but to depend upon the translator's skill regarding the expressions and style of the translated text. This is another area in which technical-writing expertise comes in handy.

When translating manuals for industrial products, specialized knowledge is required. Such specialized knowledge, however, often takes the form of a list of technical terms and this is often the extent of the specialized expertise required for translation. In other words, to translate a newspaper article or an academic paper about a specialized field the translator needs an ability to use a list of technical terms, but the knowledge level required for translation need not cover the whole area of a professional field or in-depth knowledge. Therefore, for translation of a highly technical document, it is necessary to employ a very specialized translation company.

It is often said that as far as a translator is concerned, the best result is obtained when the translator is translating into his/her mother tongue. But in Japan this presents certain problems. Because Japan is a country having a single language, foreign languages

are not used in daily life. This means that when translating into a foreign language, it is often difficult to find a translator if we adhere to the mother-tongue principle. Therefore, a Japanese translator proficient in the target language usually translates the document.

c) Translation work flow

Most of the original text to be translated is received by the translators in the form of a hard copy. A few of these take the form of an electronic file in a floppy disk. Certainly, we are seeing more and more electronic documents but we are still in a transitory period. When having the document in electronic form, certain problems arise. One of them is because various file formats are used by different word processors and computers, and the files often lack interchangeability when different types of machines are used.

Translation is often done by typing a rough translation into a word processor. In editing, corrections are sometimes done directly on the word processor. Sometimes the corrections are first done on paper and then typed on the word processor. The decision as to which form is used depends on the quantity of corrections, etc.

Most of the translated documents are delivered in the form of a hard copy, printed out from the word processor. Others are accompanied by an electronic file in a floppy disk.

Questions arising during the translation process are directed to the client. But there is no established pattern in this process.

d) Translating procedure

When beginning to translate, the translator considers the field and object of the document to be translated. Then a rough translation is made. When the rough translation is completed, the document is checked for missing portions and errors.

The translation is done sentence by sentence. No grammatical analysis of the original is made. The meaning is taken with a sentence as a unit and the sentence is translated. When doing this, a translation of the sentence can be made following the grammatical structure of the sentence. This will be more or less a direct translation. Translation by paragraph, i.e., by taking the meaning of the paragraph, is seldom done. Part of the reason for this is that when translating a long document with a team of translators, sentence-by-sentence translation better maintains the translation quality.

While translating, an experienced translator seldom uses a dictionary. In other words, it is difficult to stay in the translation business if one has to consult the dictionary all the time.

When the original document is well written, sentence-by-sentence translation is possible and plausible. When, however, such is not the case, a sentence may have to be divided into two or three parts.

At the final stage, a translation check is performed. A translation check contains the following processes: a check for missing portions, a check for special expressions and terminology, and a comparison check from the original to the translated document. Because some aspects of translation work are mechanical, sometimes a whole sentence is found to be missing.

e) Translation speed

An experienced translator can type into a word processor a rough translation almost as easily if he were copying the original text.

When making a rough translation from Japanese to English, a page of translation can be done in about twenty minutes, a page being approximately 150 words or double spaced 22 to 23 lines on A4-size paper. In a rough translation from English to Japanese, a sheet or 400 characters of Japanese can be made in about twenty minutes. This, however, represents a translation of manuals and the like, and the speed will be less than half when translating a serious literary text.

f) Quality control

Translation done in the translation department of a company is managed by specialized knowledge, technology, and experience. This is to say that the supervision of translation quality is done by an organization.   On the other hand, translation done by individuals

depends on their specialized knowledge, technology, and experience. In other words, translation quality management is not carried out in an organized manner.

A translation company tends to have considerable specialized translation knowledge, technology, and experience. Most translation companies have quality standards management and perform quality checks in the translation process. But this does not mean that translation companies have specialized knowledge of the contents of the translation document.

When a translation company receives a job, a coordinator glances at the manuscript to be translated to grasp its specialized field. Considering other elements such as the level of translation ease, the coordinator selects a translator.

As for the conditions placed on translations, there are the following types: translations with pre-editing and corrections, rough translations, translations with post-editing. etc. In pre-editing, Figures and Tables are separated from the text. In the pre-correction process, the original text is corrected from the technical-writing point of view. In post-correction, an experienced translator edits the rough translation. In post-editing, the Figures and Tables are reinserted and the format of the document is restored. Also, in some cases, the translated document is corrected for technical-writing quality.

g) Electronic equipment used in translation

Many free-lance translators use facsimile machines for receiving documents for translation. For actual translation work, word processors are usually employed. Faxes and word processors are now becoming almost necessities.

In addition, some translators are beginning to use electronic dictionaries for special terminology. A rare few are beginning to use machine translation systems. When machine translation is used, the translation is shown on the screen and post-editing is performed looking at the original document. That is, machine translation has not attained the translation quality in which post-editing is performed on the technical-writing level.

2.2 Problem areas in translation
a) The quality of the original document is poor.
- Poor logical construction in the contents of the document.
- Structural and contextual ambiguity
- Long sentences
- The original document requires technical writing.

b) Problems with the translator
- Ineptitude
- Lack of specialized knowledge

c) Problems pointed out by translators
- Poor quality in the original document
- Even when the quality of the original document is poor, high-quality translation is required.
- The technical terms and user-specific expressions are not well prepared.
- There are often particular requests for the translation styles.
- The delivery period is short.

d) Market competition
- Because of price competition, many poor translations are performed.

e) Conditions required of machine translation

The purpose of machine translation is to reduce translation costs. Therefore, machine translation should replace either the translation company or the translator in all or part of the translation work, and at a lower cost. By taking advantage of the fact that a computer is used for machine translation, it should be considered what humans cannot do during the translation, such as collection of new terms.

In the translation of manuals for machinery, etc., translation is done with the sentence as a unit. This level is satisfactory for the user, unless the job is a special translation requiring high readability levels. At the same time, the quantity of translation for manuals is considerable and straight translation is usually accepted. Therefore, by improving the current machine translation system, there might be many occasions where machine translation can be utilized effectively.

The biggest advantage of a machine translation system is the consistency in the use of technical terms. To utilize fully this advantage, it is necessary to improve and develop technical dictionaries. This should be done through machine-translation experience.

There is considerable demand for making the translated documents the same style as the original document. This includes that the Figures and Tables be placed in the same position as in the original document, and in the same format, or to have the same page layout. To achieve this, the machine translation system will use electronic files to the fullest extent.

By using communication lines to eliminate delays in document transportation, it is possible to shorten delivery times. In machine translation, the largest time period is spent on post-editing. To shorten the post-editing process, it is necessary to develop a translation support system. Currently, translation support system prototypes are beginning to appear.

3. Current Status of Machine Translation from the Japanese Point of View
3. 1 Trends in machine translation

Details are shown in Appendix 7. Examples of Japanese machine translation systems are given in Appendix 8 and those in foreign countries in Appendix 12.

a) Trends in Japan

Efforts to produce a practical machine translation system are being carried out by mainframers, electric appliance manufacturers, translation companies, research organizations, the national government and its branches, etc. Considerable high-quality work is being done towards the practical use of machine translation systems. This is a very unique situation internationally. Expectations are very high for establishing viable machine translation technology and reducing translation costs.

From the point of view that the technological impact of the core technology of machine translation will be dramatic—not to mention its side effects—work is proceeding to establish machine translation technology as the basic technology for future high-level information-processing. This is visualized in addition to the desire to develop a practical machine translation system.

The Mu project was executed as a four-year national project beginning in 1982, and is presently being continued as the Mu2 project for practical application. Other work is being done on the following topics: machine translation systems for the languages of nearby countries, an automatic translation telephone, and electronic dictionaries.

In addition, basic examination is being done for the possibilities of machine translation of patent summaries and Japanese Industrial Standards (JIS) documents.

The following Japanese computer companies and other corporations are involved in developing machine translation systems: Oki Electric, Canon, Sanyo Electric, CSK, Sharp, Toshiba, Japan IBM, Systran, NEC, Hitachi, Fujitsu, Bravice International, Matsushita Electric, Mitsubishi Electric, and Ricoh.

Also, KDD and NTT are engaged in machine translation system development, although these are not manufacturers.

Only a few of these companies have been making sales of systems for the mass market, but many have been experimenting with users as field tests.

If we include the portable translation machines of venture corporations, there are several more companies that can be added to the above list.

At the same time, as part of machine translation system research, universities and public research organizations are making experimental systems as part of their research activities.

Denjirin is an electronic dictionary system. Also, several compact-disk electronic dictionary files are now on the market.

In addition, there are pocket translation machines for overseas travel. Some of these are attempting to combine speech recognition and synthesis for input and output.

b) Trends at the CEC

The Commission of the European Communities has developed machine translation systems by applying the SYSTRAN system as a kernel part and has been providing a machine translation service. These systems are also actually being used in such countries as France, Germany, and Belgium.

The Commission of the European Communities is promoting a EUROTRA project for developing a machine translation system for interchangeably translating the nine official languages of the European Communities. A prototype system will be developed in 1990.

To perform Japanese to English machine translation, the Commission of the European Communities is carrying out a Japanese project using Japanese-English machine translation systems developed in Japan.

EURODICAUTOM is an electronic dictionary for technical terms.

c) Trends in the United Kingdom

In Great Britain, a project was completed in 1987 to develop an English-Japanese machine translation system as part of the ALVEY Project.

d) Trends in France

A national project was completed in 1987 with the goal of transplanting the technology for a machine translation system developed by the Grenoble University.

e) Trends in West Germany

The Translation Bureau of the West German government is developing LEXIS as a large-scale electronic dictionary for technical terms.

In conjunction with the University of Texas in the United States, Siemens has developed a machine translation system called METAL. The system is now being marketed The company has also developed TEAM as a large-scale electronic dictionary of technical terms.

Among universities, Saarland University has developed a machine translation system in a project called IAI. At Stuttgart University, a system has been developed to generate German.

f) Trends in the Netherlands

A machine translation system called ROSETTA is in the R&D stage at Philips. The BSO Corporation is developing a machine translation system known as DLT.

g) Trends in Canada

A complete automatic machine translation system called TAUM-METEO was created to translate meteorological information from English to French. The system has been in service 24 hours a day since the middle of the 1970s.

Also, at a research institute of the Telecommunications Agency, a machine translation system called CRITTER is being developed with support from the Translation Bureau.

TERMIUM and BTQ are terminology data banks.

h) Trends in the United States

The following commercial machine translation systems exist: SYSTRAN, WEIDNER, ALPS, LOGOS, SMART, etc. Also, as a practical system or as a system aiming at practicality but not commercialization, there are the PAHO system.

Machine translation systems are also being developed at the following universities: New Mexico State University, CMU, and BYU. The BYU system is a translator's workbench. Experimental systems are being developed at MIT, UCB, and IBM for research purposes.

i) Trends in Asia other than Japan

In Korea, machine translation systems are being developed in cooperation with American, French, and Japanese organizations.

In China, it is said that much experimental research is being done.

In Hong Kong, a translation support system called CULT has been developed and used.

Some of the Southeast Asian countries are engaged in machine translation system development in cooperation with Japan and other countries.

3.2 Utilization of machine translation in Japan

Some examples of machine translation utilization are included below. Details are found in Appendix 9.

a) Machine translation services include the following three types.
 (1) Rough translation using machine translation
 (2) Translation using machine translation
 (3) Translation VAN service utilizing machine translation

The first is a rough translation service, using machine translation. This has come to be used only recently and charges are made according to the machine translation processing quantity. This resembles a machine translation system rental service. The second and the third categories are translation services that use the machine translation system in the translation process. The difference between the two lies in the fact that in the second the translator makes post-corrections on the results of the machine translation, whereas the third is a translation service using machine translation via communication lines such as VAN. Furthermore, within the second classification, the following exist: a type that is a part of the translation service process; a type that is experimental and includes machine translation field testing; and a type that comprises an individual free-lance translator using the machine translation system. Also, there is individual use of machine translation for personal tasks.

b) The charges for machine translation services

There are only a few cases where the general public is considered as the clients for machine translation services. The charges, however, for machine translation are as shown below. But because machine translation is presently used only experimentally, the charges listed may not be the standard for machine translation.

As an example of rough machine translation service, English-Japanese translation is ¥180 per page and Japanese-English translation is ¥360 per page. These charges do not include pre-editing and post-editing. When we compare these charges with the ¥2,550 for manual translation, in English-Japanese work the charge for machine translation is one-fourteenth or 7% of manual translation, and for Japanese-English, one-tenth or 10%.

For machine translation — because pre-editing and post-editing are done manually, the proportion of the machine translation charges is not clear.

In VAN machine translation service, the tuning for technical terms is implemented just as in manual translation, and post-editing is added. For English-Japanese translation the charge is ¥2,100, and for Japanese-English ¥2,700.

Compared with the charges for manual translation, manual translation is more expensive: for English-Japanese it is ¥2,550 per page and for Japanese-English ¥3,600. For English-Japanese translation, the machine translation charge is 82 percent that of manual translation and for Japanese-English translation, 75 percent.

c) Translation process that includes machine translation

The translation process that includes machine translation proceeds as follows: (1) receipt of the original manuscript, (2) pre-processing, (3) pre-editing, (4) machine translation, (5) post-editing, (6) post-processing, (7) rewriting, and (8) delivery.

In translation service where rough machine translation is used, the client perceives the quality of machine translation. Therefore, confidence in machine translation is crucial.

In translation service where machine translation is utilized as part of the process, the machine translation results are not visible to the client. This means that translation with

corrections made by the translators are delivered to the client. Therefore, machine tansla-(ion only matters to the company using the translation service and the client has no qualms about the use of machine translation.

The pre-editing for machine translation is for dividing up sentences and changing for better punctuation. Sometimes parentheses are used to make the meaning clear. When, however, the quality of the original manuscript is such that it is very hard to understand, it is not sufficient to perform pre-editing only on part of the text. When using pre-editing, it is better to limit the pre-editing so that only such changes will be used that alleviate the load on the computer.

d) Use in the expansion of machine translation

Up to the present, more than 4,000 sets of machine translation systems have been sold. Of these, many are said to have been returned to the seller and some are not used and arc idle. The exact figures on these are not known.

In addition to the use of machine translation for the translation business, the machine translation system is often used for personal tasks. But the actual state of personal machine translation use has not been tabulated. Sometimes the machine translation system is used just for curiosity. At other times, the system is used for outline information at an early stage where fewer problems arise regarding the translation results. In these cases, rules are made that when it is necessary to ascertain the accuracy of information, one must inquire at the source of the information. The potential demand for this type of machine translation use seems to be great.

e) Effect of machine translation

Almost all the users point out as an advantage of machine translation that it has been made easy to unify technical terms and special expressions.

Because the client asks for the services of a translation company—knowing that machine translation will be part of the process—the client is prepared to make the original manuscript into an electronic file and to compile a user's dictionary for technical terms and special expressions. Because of this, a favorable condition for future improvement of translation efficiency is beginning to emerge as a side-effect of the use of machine translation.

By the use of communication lines, a more efficient use of time is often made possible. This leads to shortening the delivery time.

When using machine translation for a rough translation, the service provided is to obtain approximate information and for post-editing by the clients themselves, utilizing the machine translation's technical terms. By building upon this base, it might be possible automatically to translate the information sent from overseas countries via communication lines and compile the information into a data base.

The use of machine translation has just begun and we are still searching for better applications and cost accounting. Expectations, however, are great for side-effects in terms of efficiency and new economic activities—by making use of electronic files and using communication lines.

f) It is still too early to evaluate the effects of machine translation. But there are many possibilities for expanding its use. The reasons for this are as follows.

- There are many kinds of documents that are useful, and are composed of simple or complex sentences of exceedingly simple structure.
- By the better use of technical writing, there are many kinds of sentences suitable for machine translation.
- Rough translation is useful for making better translations.
- Machine translation results can be used to sort and edit technical terms.
- Page format, Figures, and Tables are arranged in a more efficient manner.
- An increased use of electronic files and document transmission is promoted.

The above items are necessary to improve translation efficiency and to lower translation costs, even when machine translation is not used. Thus, efforts to lower the cost of translation are in accord with the enlargement of the possibilities for machine translation. Consequently, the possibilities are substantial for expanding the use of machine translation, and the guidelines are clear in this direction.

3.3 Utilization of machine translation overseas

Below are examples of overseas utilization of machine translation. Details can be found in Appendix 13.

a) Utilization in Canada
Canadian government	TAUM-METEO
Canada GM	SYSTRAN

b) Utilization in the United States
U.S. government	SYSTRAN
NASA	SYSTRAN
U.S. Air Force	SYSTRAN
XEROX	SYSTRAN
Caterpillar	SMART
PAHO	

c) Utilization in Europe
CEC	SYSTRAN
	ATLAS
NATO	SYSTRAN
KFKS	SYSTRAN
Minitel	SYSTRAN

3.4 Evaluation of machine translation

a) Machine translation experiments

We have performed machine translation experiments using a practical Japanese machine translation system for both Japanese-English and English-Japanese. The results are shown in Appendices 10 and 11.

Some translations were done well. Others, however, were not translated or were translated incorrectly. In some cases, only fragments of sentences were translated and they are directly put into a sequence disregarding linguistic relationships among them. Even when this kind of work is output, the outline of the content can be guessed. In such cases, specialists in certain areas might be able to use this type of translation.

When translation is not possible, the causes for the most part lie in the difficulty of analysis. The failure to analyze is a result of an inadequate dictionary, multiple word meanings, complex syntax, overly complex sentences, ellipses, etc.

b) Basic theories of machine translation

Among the basic problems of machine translation, there are the following.
Computational linguistics
Dictionaries
Terminology
Software

c) Elemental technology of machine translation

In the core technology of machine translation, the following are included.
• Technology concerning text data
Corpus
Text data in multiple languages (in contrasting layout)
• Technology concerning word data
General word dictionary
Technical dictionary
Dictionary of multiple languages
Classification of meanings

. Technology concerning grammatical data
>   Sentence structure
>   Meaning
>   Anaphora
>   Ellipsis
>   Dialogue
>   Sentence structure contrastive data for multiple languages

• Technology concerning language processing
>   Analysis
>   Transfer
>   Synthesis

• Technology concerning systematization
>   Programming language
>   Communication function

• Technology concerning translation support
>   Electronic dictionary
>   Conversational translation system

• Technology concerning development environment

Depending on what is demanded as the capability of a machine translation system the method of compiling the core technology and contents of the list for detailed technologies is different. The core technology for machine translation includes artificial intelligence technology and ordinary information processing technology, and is not limited to language processing technology.

The evaluation of the core machine translation technology is shown in Appendix 14.

d) Level of achievements of core technology

Conversion of basic dictionary information into electric files has been completed. There are a few electronic dictionaries having some 300,000 words as head words which are actually used. Regarding the lexical item, however, both the content and quantity are inadequate. Only trivial surface-level information is included.

Regarding the syntactic component, basic grammatical rules are established. These are sufficient for translating a simple sentence with a embedded subclause and a complex sentence with easy-to-understand syntax. Therefore, when the clauses and phrases are understood correctly, translation is possible.

The language processing methods are sufficient for processing language information expressed in sentences with relatively simple structure. To gain processing speed, programming languages such as C are often used for implementing systems. The functions and systems for networking employ whatever computer networks are available. An electronic dictionary is used for translation support.

The results from varieties of software developments are used as facilities for the software environment of machine translation. But a development environment that includes special conditions in translation processing is yet to be developed.

The core technology achieved in the area of machine translation includes the following.

• Translation of simple sentences
• Relatively simple embedded sentences can be translated.
• Translation of simple, complex, and compound sentences
• Translation where unique translation is possible.
• Translation where even if multiple meanings are involved, the multiplicity can be resolved with a minimal addition of clarifying information.
• Translation where the syntax is clear. Syntactic dependency is assumed to the nearest phrase.
• Partial translation is done when translation of the entire sentence is impossible.

There are many manuals to be translated that can be satisfied with a translation capability of this level. In other words, the quantity of documents to be translated on this level is very large.

Accordingly, the possibility is quite good for lowering the cost of translation by utilizing the presently available machine translation technology. The use of machine translation has just started and we do not have very much experience. Therefore, effective uses have yet to be determined. Many people, however, intuitively understand through experience that there are effective applications for machine translation.

As for the practical use of machine translation, the crucial point is making use of the system as it exists. Even if an entire sentence cannot be translated, the clauses and phrases that can be translated should be output. There are many levels where clause by clause and phrase by phrase translation can be useful, and many machine translation systems work on this basis, even though they are far from complete.

The most salient effect of machine translation is the unification of technical terms in the translation. There are sufficient examples as to the usefulness of machine translation in this sense. Development, however, of technical terminology dictionaries in sufficient quantity has not been achieved. Through experimental use of machine translation, technical terminology dictionaries in different fields and for different users should be promoted.

e) Technological evaluation standards

The following evaluation criteria exist for translation quality:
• Evaluation standard of translation quality in the ALPAC Report
• Evaluation method for translation quality in the Mu Project

But the evaluation method for other core technology has yet to be established. It is necessary to work out an evaluation method corresponding to the technological development phases.

f) Users' evaluation

The largest response from users regards the implementation of technical terms and special expressions.

Another reaction from users is that there are large discrepancies between the sales talk on the seller's side and what actually can be achieved by the machine translation system. This basically represents an overevaluation on the buyer's side and insufficient knowledge of the current status of technology. Problems also exist on the seller's side of machine translation systems. In a way, it is inevitable that excessive expectations arise for the translation capability of machine translation systems if such names as machine translation system and translation support system are used.

The expectations expressed in the survey regarding the machine translation system are shown in Appendix 15.

4. Tasks of Machine Translation
The tasks of machine translation include both present and long-term considerations.

a) Present tasks of machine translation

Machine translation is beginning to be put to practical use in the translation process. Because of such practical use, some of the data insufficiencies are becoming manifest. Consequently, it is necessary for us to think about how to use the feedback from these manifestations to improve machine translation systems. By implementing these improvements over the long run, the systems will become markedly improved. For some users interested only in buying the system and using it, without thinking about improvements, their dissatisfaction with the presently incomplete system is great. In the end, these users may discontinue using the system. But with continued effort for improvements, the systems will continue to be upgraded.

It is beginning to be seen that there are many ways to utilize machine translation. By developing such additional functions as the use of communication lines and documentation processing by electronic files, we can improve the overall efficiency of the translation process.

b) Long-terms tasks of machine translation

The theory and technical bases supporting machine translation are computational linguistics theories and natural language-processing technology. These are further subdivided into the following: dictionaries, terminology data bank, grammar, sentence analysis, deep structure expressions, transfer of expressions of deep structure, sentence generation, etc.

Also included are these: the computer environment for development and maintenance, the operational environment for translation processing, and the supporting environment for pre-editing and post-editing.

Of these, basic technology has been developed to some extent. By improving upon the language data and improving upon the technology for further refined processing, it will be possible to enhance the ability of machine translation quite a bit.

The nature of technical language relates to those on controlled language. We must. not only reconsider such controlled language for supplementing the insufficiencies of machine translation technology, but such controlled language in terms of efficiently and unambiguously communicating information.

We also need to examine in detail the semantics information. If consistency is lacking throughout a dictionary the semantic information will do more harm than good.

The improvement and development of technical-term dictionaries should be executed over the long term through the experience gained in machine translation. We need to make progress in cooperative development with foreign organizations. We should seek to utilize the technical dictionaries developed by the Canadian government, the Commission of the European Communities, Siemens, and others. We should add Japanese to these dictionaries and make them available as an common property of mankind. The above-mentioned dictionaries have been developed over an extended period. Each would make a significant contribution, and from this point of view, we must cooperate and engage in developing terminology data bank. It is regrettable that such an effort has not been made in the past. An organization for carrying out the development should be established in Japan by the initialive of government organizations.

To constitute the theoretical background of machine translation technology, we should build upon the basic research in computational linguistics and the application of computational linguistics to machine translation. The theories governing machine translalion have yet to be clarified. Up to the present, we could say that development has been forcefully carried out without any serious theoretical consideration. But to secure efficient translation processing, we need a theoretical background. For this, many computational linguistics courses have been established at West German and British universities. In Japan, too, we should have university courses in computational linguistics and machine translation courses for research and researcher-training.

Recently, several new computational linguistics theories have been proposed: FUG LFG, DCG, GPSG, HPSG, JPSG, CUG, and PATR-II. Whether or not these theories in computational linguistics would be able to contribute to machine translation has not been examined.

In recent years, there have been more studies about semantics and dialog comprehension. These will be necessary for future advanced machine translation, and we must study possible contributions to machine translation core technology in more detail.

Further, when we translate more diverse material than manuals, it is necessary to process psychological information. Accordingly, we should study pragmatics and the language used when speaking.

The study of computational linguistics must be considered in a broader field and not only for machine translation. As a foothold for future machine translation development, we require such basic research.

To produce translation of satisfactory quality, we must use technical terms in each field correctly. Without a general knowledge in the field—even when having a list of technical terms—the translator cannot produce a smooth translation. This illustrates the need for knowledge in specific fields.

At the same time, we require a capability to make knowledge-based inferences. This is AI technology and we must examine machine translation as an AI system.

As the quantity of inferences increases, the quantity of computation also drastically rises. Therefore, we need to have a more efficient inference algorithm, as well as parallel processing capability.

It is also important to realize an interactive translation-support system utilizing the highest technology presently available. When an individual uses a machine translation system for translating a technical paper, or when we are writing a paper and translating it at the same time, such a support system would play a very important role.

5. Recommendations

a) Technical-writing education

To make better use of current machine translation technology, it is most important to have better original texts for translation. For this purpose we need technical-writing education, and we should go about this in an organized, systematic manner. In the United States, for example, technical-writing education has been growing in popularity. But because little technical-writing education exists in Japan, we need to establish the subject as soon as possible.

Technical writing is required not only for most effectively utilizing the present machine translation system, but also for linguistic-expression efficiency and cost effectiveness of communication. In fact, technical writing is essential for us, even without its application for machine translation. We must reflect and consider why technical writing had been neglected in the past, and do so as soon as possible. In the end this will promote cost effectiveness by mechanization, and minimize misunderstanding in communication.

b) Improvement and development of translation know-how

Fairly satisfactory level has been achieved in dictionary, grammar and analysis method. But no progress has been seen in transfer and generation methods. Especially lacking are language-contrasting data and sentence-generation technology. We must improve and develop data and technology by having knowledgeable translators whose mother tongue is the object language.

c) Improvement and development of a technical dictionary

A technical dictionary and an expression dictionary for different fields and users must be developed. This must be done in cooperation with machine-translation users. From this point of view, we should not pursue profits in the immediate future, but be content with making investments for the future. It is also necessary to develop technical dictionaries in such specialized fields as patents. When developing technical dictionaries, international cooperation is imperative.

d) Realization of a multilanguage machine translation system

We must create multilingual translation as an expansion of our current technology. The demand for multilingual translation in Japan is presently smaller when compared with Japanese and English translation. By developing multilingual translation, however, we might be able to elucidate the problems in the language field that were rather unclear between Japanese and English.

e) Establishing an evaluation standard

When a translator performs quality control, the principal task is comparing the original and the translation, with the sentences as the unit. This type of comparison, however, is insufficient as the standard for evaluating machine translation. We require an evaluation standard for a text as a whole, including the technical writing point of view.

f) Basic research for computational linguistics

To achieve dramatic advances in machine translation technology, we must develop technology with a backup from computational linguistics theories. Especially, to develop the technologies of handling anaphora, ellipses, and dialogues with very complex contextual relationships, we need theoretical background. Because we do not have the paths marked out for achieving these goals, we must perform basic research in parallel with expansion of machine translation technology.

g) When we acquire more information for translation processing, we will need more time for processing. To achieve in-depth translation, we require efficient processing algorithms. For this, we need theories and technologies for parallel computation.