[*Translating and the Computer 7*. Proceedings of a conference… 14-15 November, ed. Catriona Picken (London: Aslib, 1986)]

Machine translation: Japanese perspectives

Toyoaki Nishida and Shuji Doshita

Department of Information Science, Kyoto University, Kyoto, Japan

This paper overviews Japan's approach to machine translation (MT). It gives a survey of technical developments in machine translation, as well as a few case studies of experience with early products. It especially points out that most of the current MT systems are rather general purpose. This partly reflects the fact that Japan has few existing mass-production lines for technical translation into which computers may be incorporated. Instead, our approach is to design the entire environment in which humans and computers can co-operate using the state-of-the-art technologies.

(1) **INTRODUCTION**

It has been a common recognition among MT researchers in Japan that deep analysis is needed to achieve a realistic machine translation system between the Japanese and European languages, since these languages differ from each other both in linguistic and conceptual structures. As a result, Japanese researchers have long been working on basic problems.

Recently, however, as social demand for machine translation has increased and technology for natural language processing in general has made some progress, intensive efforts have been made to produce commercially available products for machine translation. Since 1984, major computer industries started to announce the release of their first commercial products.

This paper is intended to overview machine translation activities in Japan, both from the technical and practical viewpoints. It is organised as follows: Section 2 gives a brief introduction to the Japanese language. Those issues which are thought as relevant to machine translation are listed and difficult issues are distinguished. Section 3 attempts to present some common features of activities in research and development in Japan. Section 4 overviews the technical aspects behind state-of-the-art products. Various techniques for overcoming difficulties in language translation between completely different language families are summarised. Section 5 provides some case studies on the experiences with early products and Section 6 presents a number of new research programmes directed towards more advanced machine translation.

(2) BRIEF INTRODUCTION TO THE JAPANESE LANGUAGE

The Japanese language differs from Western languages both in linguistic and conceptual structure. Some prominent aspects of the Japanese language are listed below, from the easiest to the most difficult.

1. Large character set. The character set for Japanese can be divided into Kana, Kata-Kana and Kanji characters. JIS (Japanese Industrial Standards) specifies standard 16-bit code for these characters. The first level contains about 3,000 characters and the second level 3,400 more. This aspect is not a problem at all, as most Japanese computer products involving microcomputers support Japanese I/O (input/output) as well as fancy graphics. See [ICTP 83] for technical aspects.

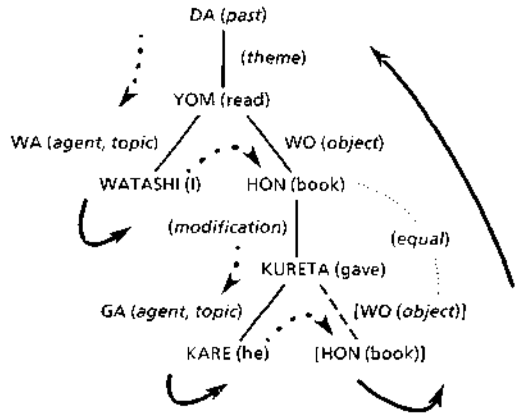
2. No delimiters between adjacent words (but punctuation does occur between sentences and optionally between long phrases). Though we have breath-group notation, we are not generally familiar with it. Thus, word boundary is not a stable notion among ordinary Japanese people.

   This aspect is problematic for machine translation as more than one segmentation is usually possible for a given sentence and this squares the complexity of analysis. However, Japanese computer industries succeeded in handling this problem by taking an engineering approach. Therefore, this delimiter-free problem is thought of as a relatively minor one to the current machine translation researchers.

3. Word order roughly corresponds to postfix notation or 'reverse Polish notation'. (Prefix notation is sometimes called 'Polish notation' because it was introduced by a Polish logician. Postfix notation is called 'reverse Polish notation' to emphasise the fact that it came from the same principle as prefix notation.) Thus, if the meaning of a sentence is represented as a case structure tree as in Figure 1(a), the word order in Japanese may be obtained by traversing it in postorder. Mood, tense, and aspects are indicated by left-embedding.

   This is not a serious problem for the current technology, because we can use efficient algorithms for parsing sentences. It raises a problem only

(a) Case structure tree which means "I read the book which he gave (me)."



(b) Rules for traversing a case structure tree

(a) start from the top node.
(b) go along the tree until any terminal node is reached.
(c) during going downward, neglect words.
(d) when you arrive at a terminal node, read it, and go upward.
(e) during going upward, read words.
(f) when you come to any node with any branch you have not yet traversed, traverse them in turn, and then go back to the upper node.
(g) finish traversing when you return to the top.

(c) Result: Japanese sentence for (a)

WATASHI WA KARE GA KURETA HON WO YOM DA.

Figure 1. Case Structure Tree and Word Order in Japanese: Example.

when the concrete syntactic analysis fails for the input sentence, for we cannot use the word order of the input sentence as a default.

4. Surface case labels are explicitly attached to case elements as post positions (called *joshi's)* to make word order flexible.

Freedom in word order makes typical phrase structure analysis awkward, but this specific aspect can be handled by semantic processing. The real difficulty arises when we try to obtain a deep case relation from a surface case structure. This is common to Western languages in that one surface label usually represents more than one deep case label.

5. Word ending inflects. This brings about no essential problem except that it slightly complicates the analysis and synthesis of Japanese sentences, for inflections are completely rule-based. All that is needed is inflection tables and dictionary entries which indicate inflection type.

6. Ellipsis is allowed as far as information is recoverable from the context. In general, short, concise, and insightful sentences are preferred in Japan

to rather long, detailed, logical sentences. This allows writers to eliminate as many words/phrases as possible.

This phenomenon is a real problem to computers. In order to recover eliminated words and phrases, a large chunk of common-sense knowledge is needed for the subject field. If this is not available, the machine translation program must be able to use a number of heuristic paraphrasing techniques to generate target language expressions without referring to unrecoverable information.

7. Conceptual differences. It is often impossible to find a corresponding word or phrase when we translate sentences between Japanese and Western languages. Some features are given below with examples.

(a) Word level. It is sometimes hard to find a single word in the target language which can cover the entire meaning of a given source language word. For example, there seems no single adequate Japanese word for ‘take’, ‘integrate’ nor any two words distinguishing ‘ambiguity’ and ‘vagueness’.

(b) Phrase level. For example, the Japanese language does not allow prenominal negation. Therefore a phrase like ‘few researchers are working on this issue’ should be paraphrased as: ‘the number of researchers working on this issue is very small’.

There are a number of ways to translate an indefinite determiner in English into Japanese. However, the corresponding adjectives are not neutral; they may represent additional incorrect information:

" KARE	WA	INU	TO	SAMPO	SHITEITA"	
he	<subj>	(a) dog	with	walking	was	; he was walking with a dog

=> indefinite determiner for "INU" is not indicated; neutral.

"...	IPPIKI-NO	INU	TO..."
	one	dog	with

=> the speaker seems to stress the number of dog is one; not two or more.

"...	ARU	INU	TO..."
	certain	dog	with

=> the dog seems to be extraordinary; the speaker may start a story about the dog.

(c) Sentence level. Japanese does not allow a non-animate subject. For example, ‘the wind opened the door’ should be paraphrased as, ‘the door opened due to the wind’.

The way of capturing events also differs between Japanese and European languages. The meaning conveyed by the English sentence, ‘he touched me on the shoulder’ must be expressed in Japanese something like, ‘he’, ‘touched’, ‘my’, ‘shoulder’ (direct, phrase-by-phrase translation from Japanese).

(d) Discourse level. Japanese generally like stories to proceed from the concrete to the abstract. Hence, strictly speaking, sentence-to-sentence translation is not a good translation.

(e) Social level. The Japanese language has honoratory expressions. To generate them, social relationships between the speaker, the hearer, or the referent should be taken into account. This issue, though hard, is important in the business field; if the honoratory expression is not generated adequately, the result would be completely unacceptable.

Nitta [85] discusses these issues from somewhat different viewpoints.

(3) **CHARACTERISTICS OF JAPANESE RESEARCH AND DEVELOPMENT IN MACHINE TRANSLATION**

It is now evident that the translation between Japanese and European languages cannot be achieved by any simple method. Before going into some details, we will attempt to extract some common features in research and development in Japan.

Historical notes [Nagao 82a,85a] {Tamati 85}

From the early days, the so-called first generation method [Hutchins 78], by direct, word-by-word replacement, was thought inadequate for translation between Japanese and European languages. From the beginning, researchers in Japan aimed at building systems by second or third generation methods [Tamati 63], {Yamada 64}, {Wada 65}, [Sakai 66,69], [Sugita 68], {Kurihara 73}, {Shudo 73,77}, [Uchida 80], [Nagao 80b], {Miyazaki 83}, {Ikeda 84}, [Yoshida 84], [Muraki 85].

The 1960s and 1970s were mostly spent on basic research. Hardware and software for Japanese input and output were investigated on the practical side. At the same time, adequate computational theory for MT was sought by building and scrapping small prototypes on the basic research side.

The years around 1980 were a turning point for the following reasons:

- the progress in basic research in natural language processing
- significant improvement of cost performance in computer facilities
- the success of Japanese language word processors.

From the practical point of view, the last issue is significant, because the success of the Japanese language word processors implies success in Kana-Kanji conversion, which indeed is an artificial intelligence (AI) problem in the sense that no complete solution is expected unless the machine can understand the input completely. (See the word processor section of [ICTP 83] for technical details.) The developers of the Japanese language word processors solved this problem by an engineering method.

In 1982, Japan's Agency of Science and Technology, a minor organisation of Japan's Ministry of International Trade and Industry (MITI), launched a three-year project (later extended to a four-year project) {Nagao 83d, 85c}[Tsujii 85], called the Mu project, to develop a practical system by integrating the state-of-the-art technologies. Around this time, some of the major computer industries had finished the initial developmental stage and started evaluation through internal use. Central Research Institute for Electrical Power Industries (CRIEPI) started to evaluate ATLAS-I, Fujitsu's early product for English-Japanese translation {Terano 85}. In the Fifth Generation project, machine translation was one of the early subjects and basic research focused on dictionaries and conceptual analysis of translators' knowledge and skills {Tanaka 84}. Major research projects and investigators are listed in the appendix.

Japanese perspectives

As a whole, the Japanese approach differs from that of Western countries in several respects:

1. The language pair demanded for translation is starwise: one to many. The investigation by the Japanese Electronic Industrial Development Association (JEIDA) revealed that Japanese language is almost always included at one end of desired language pairs. As to its partner, demand for English is most significant, followed by French, Chinese, and Spanish {JEIDA 82,83,84}. This is contrasted with the European Community, (EC) where all pairs between seven and nine languages are required.

2. Unlike the EC, we have had few industry-like mass-production lines for translation. This means that we do not have a mother system into which computers may be incorporated. This is the major reason why Japanese industries have not started building machine aids for human translators.

3. Therefore, the situation is quite product-driven. Potential requirements have not been explicit until the users have seen actual products. Since the demands for machine translation were rather widespread and vague, the computer industries had to develop a rather general-purpose, powerful translation engine for it to be usable in unexpected environments.

Thus, it is evident that we have to wait a few more years to adapt these systems to individual environments. But it is anticipated that a number of new environments which are suitable for state-of-the-art machine translation systems will be invented, quite like Japanese word processors, in the near future.

(4) TECHNOLOGIES FOR MACHINE TRANSLATION – SURVEY

The target of the current research and development in Japan can be characterised as a full implementation of so-called second generation machine translation systems featuring:

- the separation of linguistic data from the program
- the use of intermediate representation
- the use of semantic information.

No current machine translation product can be categorised as an interlingua system; no interlingua is defined which is completely independent of any specific language in its vocabulary, syntax, semantics and concepts. Therefore, all of these systems have somehow to transform a source-language-oriented representation into a target-language-oriented one. Thus, they are transfer systems. They differ in the level at which the transfer is performed. In the remainder of this section, we will review technical aspects from this viewpoint.

Morphological analysis

The morphological analysis of English is mainly devoted to the detection of word stems and the recognition of compound words consisting of relatively few words. It is fairly safe to separate the morphological analysis phase from the syntactic analysis phase. However, this is not so for the Japanese language, as the delimiter-free problem cannot be handled independently from syntactic and semantic analysis. No method developed so far can perfectly separate the unsegmented Japanese input into a word sequence. Hence, if the Japanese analyser is organised in tandem, from morphological analysis to syntactic analysis, then the accuracy of the whole system is bound by that of the morphological analyser. On the other hand, if the morphological analysis and syntactic analysis are done simultaneously in a uniform manner, then the complexity of the analysis will increase tremendously beyond manageable limits.

The approach taken by the Mu project was, like other methods, to incorporate local syntactic analysis into the morphological analyser to increase accuracy. They used a table-driven method to store word-continuation information. They also used some heuristics to cope with failures that may occur in the 'left-to-right dictionary-look-up and cut-a-

Word' method. If there still is ambiguity, parallel results are passed to the subsequent phase {Sakamoto 83}.

Syntactic and semantic analysis

The general problem here is how to resolve ambiguities that arise from a purely syntactic analysis. The usual method is to introduce selective restriction or preference. The most common approach to this problem is to use syntactic and semantic markers (or equivalently, features). Another approach is taken by the KDD group, who defined the notions of forbidden trees, recommended trees and exclusive trees to capture the semantic constraints rather as exceptional cases to pure syntactic analysis. They proposed a method of managing the (probably very large) set of these trees so that those trees are stored and retrieved efficiently {Sakaki 82,83,84}.

In ATLAS-II, a Fujitsu product, the notion of a world model is used to handle structural transformation at the conceptual level [Uchida 85]. Logically speaking, this approach seems to be essentially the same as the semantic marker method. However, from the engineering point of view, it is good for avoiding complexity, because the knowledge for structural transformation is separated from the main processing flow as an independent module and is managed by an ordinary database management system.

What should be realised in building a practical system, however, is that ambiguities may still remain unresolved or equally preferred, as it is almost impossible to incorporate into computers all semantic and common-sense knowledge for a full account of ambiguities. This aspect should be handled either by interactive or heuristic methods.

The Hitachi group uses shallow trees to allow certain ambiguities to remain unresolved [Nitta 82,84]. The shallow tree will be modified when further information is available from the semantic analysis. The English-Japanese version of their system, called ATHENE/E, does not use a chronological backtracking regime to recover from misdirected analysis; they do this instead by rewriting the structure when failure occurs {Okajima 85}. The Toshiba group similarly incorporates syntactic preference to cope with such cases [Amano 85].

The Mu project accumulates cues from multiple levels into a single structure called an Annotated Tree Structure (ATS) [Tsujii 84]. This makes it easy to give preference by considering the combination of cues at various levels. Nagao {83a} reported in detail their attempts at analysing conjunctive noun phrases.

Yiming Yang, a graduate student at the authors' group, takes a similar but more specific approach to give preferential interpretations to Chinese sentences [Yang 84,85a,b].

Transfer phase

It is generally believed that there are basically two types of approach to the transfer phase. One type is to use rather shallow structures, such as phrase structures augmented by features. The advantage of this approach is that many cues, such as information about surface subject or word order, are preserved which might not be able to be conveyed adequately by deeper representation. It is obvious, though, this approach suffers from a complication of the transfer phase. Accordingly, the other type is to use deeper, conceptual representation as an intermediate representation.

However, it seems that there is no substantial difference between the two. It is somewhat a matter of preference and trade-off. Almost any transformation that can be done by the deeper method can be simulated by the shallower method. This observation has led to a third approach in which physical structure does not play any essential role; a rather neutral structure is used as a physical structure and the structural information is virtually conveyed by annotations.

1. Using shallow structure as an intermediate representation The IBM Japan group takes this approach to translate computer manuals {Tsutsumi 82}. The background of this approach is their observation that the use of syntactic structures does not seem to complicate the structural transformations required in this field as much as had been thought, because the sentence types in computer manuals are, and should be, rather restricted.

2. Using deep structure as an intermediate representation The basic method is to determine the target language expression by examining the case structure and semantic markers [Ishihara 74a,b] [Nishida(F) 80,82]. In spite of the use of deep, conceptual structures, this approach still requires an enormous amount of structural transformation due to conceptual differences between languages. Since translation at the conceptual level does not seem to be governed by a small number of universal rules, lots of word-specific rules should be prepared as lexical knowledge for conceptual transformation. Tanaka's active dictionary {Tanaka 85} and the NTT group’s frame system {Nomura 82}[Iida 84]{Iida 85}{Shimazu 85} are attempts to allow sophisticated word-specific rules in each lexical entry that requires complex conceptual transformation. This approach needs to be supported by user-friendly, highly sophisticated software to develop and maintain dictionary entries. However, this has hardly been addressed.

3. Using physically neutral structure The Mu project uses physically rather neutral structures in which syntactic, semantic and contextual information are virtually embedded by means of annotation. This allows both surface and deep information to be accessed rather uniformly. Accordingly, it has advantages of both deep structure approach and shallow structure approach. Additionally, this ensures that translation at a deep level gracefully degrades to that at syntactic or default level, when analysis at a deep level fails.

To decrease the complexity in the structural transformation, the Mu project's system has two more subphases for the transfer phase: the pre- and post-transfer loops [Nagao 84a]. At the pre-transfer loop, source-language-oriented conceptual structures are paraphrased into more neutral ones. At the post-transfer loop, an attempt is made to paraphrase target-language theoretically independent structures into target-language-oriented expressions. These substages are expected to be useful when the system is extended to multilingual translation.

Generation phase

On the one hand, if shallow information of the source language is preserved in the transfer stage, then it is somewhat easy to generate target language expressions, in the sense that the discourse structure of the source language is preserved in the transfer phase and serves as a default discourse structure which may not completely fit in the target language, but which might be much better than nothing. On the other hand, if only a conceptual structure is passed to the target language, the discourse structure should be generated to make the translation understandable. This, however, has not been fully investigated yet and is left for the future. Therefore, current machine translation systems somehow preserve surface information and transfer it to the target language.

The Mu project divided the transfer phase into two: core syntax generation and style generation [Nagao 84a]. The core syntax generation subphase generates phrase structure from annotated tree structure. It in particular generates lexical or morphological entries for tense, mood or aspect from the underlying information in annotations. The style generation subphase adjusts styles. A long infinitive subject, for instance, will be replaced by an 'it-that' construct.

Morphological synthesis

There is no additional complexity in generating Japanese sentences for current computers.

Dictionary problems

The quality of a machine translation system depends heavily on the quality of dictionaries. The dictionary-related problems are divided into two issues: lexical acquisition and dictionary maintenance.

1. Lexical acquisition Work has been done on the construction of dictionary databases. Early work by Professor Nagao was related to the automatic construction of dictionary databases from ordinary dictionaries, Longman's for instance [Nagao 80a]. Now some terminology databases easily interfaceable to microcomputers have become available. Other researchers have been working on automatic collection of technical terms from existing document files. But there still remains much to be done before these can be used for machine translation.

Therefore, dictionaries have to be constructed by hand coding. Problems arise as to how to manage quality and coherence. The Mu project designed special forms for each part of speech and detailed manuals to fill them out [Sakamoto 84]. However, they were still faced with errors and incoherencies.

2. Dictionary maintenance It is a hard problem to maintain a large number of dictionary entries and to tune them up for a given situation. This issue has hardly been worked out.

Human—machine environment

Sophisticated interactive tools are needed to make the machine translation system effective in an integrated human-machine environment. There have been relatively few studies in this area, too. The Mitsubishi group has done some preliminary work on an interactive machine translation system {Fukushima 84, Arita 83}. Masaru Tomita at Carnegie-Mellon University developed an interactive front end for disambiguation which only requires the user to have knowledge of the source language [Tomita 85]. Tomita implemented a number of versions, including the one for the authors' system {Tomita 84}. Okajima *et al.* at Hitachi discussed a linguistically motivated editor for post-editing {Okajima 83}. The editor for pre-editing is discussed rather in the context of controlled languages than in a general context.

Software environment

Most computer companies in Japan preferred efficient implementation, sometimes using very low level programming languages such as C or a macro assembler. This would work when the goal is rather restricted and high performance is desired. However, it is hard to extend such systems. The NTT group, the Mu project, and the Hitachi group each take an approach using a high level grammar description language to achieve high quality translation. The NTT group uses a frame language to write semantically-oriented sophisticated rules for translation {Kogure 84}. The Mu project developed GRADE, a high level tree manipulation language [Nakamura 84]. Grammatical rules as well as lexical rules for analysis, transfer and generation have been coded in GRADE. The Hitachi group has started to use a similar language {Kaji 85}.

Evaluation method
The Mu project designed a qualitative evaluation method to measure accuracy and understandability {Nagao 85a}. This method was also applied to evaluate ATLAS-1 [Sawai 82] at CRIEPI. However, no quantitative method for evaluation has yet been proposed in Japan.

(5) MACHINE TRANSLATION SYSTEMS AT A PRACTICAL LEVEL

Titran
This system can be thought of as the first system that reached the practical level. This system, developed by Prof. Nagao's group, was designed to translate titles of scientific and technical literature. There were three versions. The earliest system was for English-Japanese [Nagao 82b]. Two other versions were developed for Japanese-English and Japanese-French. The English-Japanese version {Nagao 83c} was tested against some 3,000 titles and the accuracy rate was 80 per cent on average. This version was originally implemented at Kyoto University and later it was transported to RIPS (Tsukuba Research Information Processing System) of Japan's Agency of Industrial Science and Technology, where extensive evaluation was carried out. The Japanese-English version was tested against more than 6,000 titles and a 90 per cent accuracy rate was achieved. For the Japanese-French version, no extensive evaluation was carried out [Hubert 83]. Through these evaluations, it was experimentally proved that a relatively small set of patterns is sufficient for translating titles.

Commercial systems
Current systems, available from major computer companies in Japan, are not restricted to simple phrases like titles. They are summarised in Table 1.

Published experiences with early commercial products
Some published experiences with commercial systems are summarised below.

1. **Fujitsu-CRIEPI** CRIEPI (Central Research Institute for Electrical Power Industries) introduced ATLAS-I in 1982 {Terano 85}. They wanted to use the system to provide a database service in which abstracts

Table 1. Commercial Machine Translation Systems in Japan

Industry	Name of the Product	Source Language	Target Language	Time Available
FUJITSU	ATLAS I	English	Japanese	9-1984
	ATLAS II	Japanese	English	7-1985
NEC	PIVOT	Japanese	English	(12-1985)
		English	Japanese	(9-1986)
TOSHIBA	TAURUS	English	Japanese	(12-1985)
SHARP	?	English	Japanese	(?-1985)
BRAVIS	MediumPack	Japanese	English	6-1984
	MicroPack	Japanese	English	8-1985

of English literature are made available in Japanese. Terano used the evaluation method of the Mu project to evaluate the results {Nagao 85b}. Though the results were only partially published, it seems that they think ATLAS-I usable for their service. Terano's plan for the future is to build an environment for a full, automatic translation facility for abstracts. The human beings are supposed to be involved only in offline dictionary updates.

2. Fujitsu-Fellow Academy Fellow Academy is a school which teaches students translation skills. They introduced ATLAS-I in 1984 and started an evaluation. Seminars for advanced students are held to let them have practical experience with state-of-the-art technology. Their short-term goal is to evaluate the limitation of current machine translation systems from the translator's point of view {Oka, personal communication} .

3. Systran-Technical Service Technical Service introduced Systran and formulated a computer network for translation services {Yamaguchi 85}. This was intended to cope with the overwhelming amount of raw translation coming out from a machine translation system. Through the computer network, drafts (raw outputs from the computer) are sent immediately to translators at workstations to be returned to the chief editor after post-editing. Some early results, if urgently required, will also be sent to customers through the computer network.

Many other joint projects between computer companies and translating companies have started recently. Some interesting results are anticipated as they can design the entire environment by taking advantage of the state-of-the-art technology.

(6) RESEARCH PROGRAMMES TOWARDS NEW GENERATION MACHINE TRANSLATION

At the research level, different groups are taking different approaches.

Machine translation by formal semantics

Viewed as a software system, a machine translation system should be among the most complex. The difficulty with designing a machine translation system is to achieve the following two rather contradicting demands at the same time:

- the system should be powerful enough to cope with various linguistic phenomena
- the system should be understandable enough to make it easy to maintain and improve.

The authors' group proposed to solve this by combining functional programming and object-oriented programming methodology [Nishida(T) 82, 83a, 83b, 84]. Each lexical entry is implemented as an object in an object-oriented language. Each object can interact with others by sending and receiving messages. Like Tanaka's active dictionary {Tanaka 85}, each lexical entry is active, in the sense that it is not simply a linguistic datum to interpret but it can embody procedures to manipulate linguistic structures with which it has semantic relations. This makes it possible to write arbitrarily detailed, word-specific rules for exceptional words or for those which have wide usage and which require complex judgement to assign adequate expression in the target language. Functional programming principles are used to prevent grammar writers from writing *ad hoc* rules. Formal semantics proposed by Montague serve as a theoretical basis of this research. The notion of semantic type introduced by Montague semantics gives a rigid protocol for interactions among word-specific rules. A prototype was implemented and was tested against scientific and technical literature.

Translation by analogy

This approach is motivated by the observation that (a) it is a painstaking task to construct a precise and detailed dictionary for machine translation, (b) success in detailed syntactic analysis does not ensure the quality of the translation, and (c) human beings are guided by examples. A method proposed by Prof. Nagao [Nagao 81] is based on breath group analysis, use of ordinary dictionaries, examples involved in them, and thesaurus look-up.

Use of a controlled language

Full automatic, high-quality translation – a dream – is strongly desired in some fields where rapid translation is desired in spite of an overwhelming amount of information flow. Use of controlled language seems to be a practical approach for the current development of technology. As it is hard for untrained humans to write a text in any controlled language, the development of dedicated word processors is necessary. Approach in this direction was taken by Prof. Yoshida at Kyushu University {Yoshida 83}, Prof. Nagao at Kyoto University, and Fujitsu. Prof. Nagao proposed MAL (Machine Acceptable Language) and special purpose word processor for MAL {Nagao 83b, 84b}. The design principles behind MAL involve the domain independence and readability for humans. What MAL controls is the use of punctuation. When the word processor detects syntactic ambiguity, it will ask the user which modification is intended. A number of Nagao's papers are actually written in MAL. However, it has not yet been proven how MAL is effective in machine translation. Fujitsu is attempting to design another controlled language for Japanese to achieve industrialised translation {Hayashi, paper presented at the International Symposium on Machine Translation, 14 October 1985, Tokyo, Japan}.

More basic issues related to machine translation are investigated in the natural language processing field by many groups. This field, natural language understanding in particular, has become very active in Japan. Most reports from these groups, though they are in Japanese, can be found in Transactions of the Information Processing Society Japan (IPSJ), WG Preprints from the Working Group on Natural Language Processing of IPSJ, and convention records of the IPSJ. Some can be found from proceedings of IJCAI or COLING (International Conference on Computational Linguistics). It is strongly anticipated that basic results in natural language understanding (story understanding and generation techniques; see for example Schank [77]) will contribute to third generation machine translation.

SUMMARY

Japan has recently begun to see the first products in machine translation. More products will appear soon. A number of contrasts can be found in these products. Some were developed by domestic companies and others are modified versions of commercial products which were already available for languages other than Japanese. Some use large time-sharing computers and others networked personal computers. Some take a general approach and others a rather domain-specific approach. Some aim at HAMT (human-assisted machine translation), in which computers take initiative, asking human beings for help when they come into difficulty; others aim at MAHT (machine-assisted human translation) which is a human-centred organisation. Thus, there are many perspectives currently competing with each other in Japan.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Jun'ichi Tsujii, Prof. Jun'ichi Nakamura, Dr John Bateman, Mr Oka, Mr Hiroshi Kushima, and Mr Takao Terano for useful and informative discussions.

REFERENCES

Note: [ ] refers to those written in English; { } to those in Japanese.

[Amano 85] Amano, S., Toshiba Machine Translation System, *Proc. International Symposium on Machine Translation,* 45-48, 1985.

{Arita 83} Arita, H. and Fukushima, M., Another Approach to a Machine Translation System from Japanese to English, *WG Preprints,* WGNL 48-7, IPSJ, 1983, (in Japanese).

{Fukushima 84} Fukushima, M. and Arita, E., WGNL 41-5, 1984, (in Japanese).

[Hubert 83] Hubert, J., Ibuki, J., Kume, M. and Nagao, M., A System of Japanese-French Automatic Translation for Paper Titles, Records *of Convention of IPSJ,* 1983.

[Hutchins 78] Hutchins, W. J., Progress in Documentation, *J. Documentation,* Vol. 34, No. 2, 1978.

[ICTP 83] *Proceedings of lCTP'83,* CLCS-IPSJ, Tokyo, Japan, 1983.

[Iida 84] Iida *et* al., A Case Analysis Method Cooperating with ATNG and its Application to Machine Translation, Proc. *COLING-84,* 1984.

{Iida 85} Iida, H., Ogura, K. and Nomura, H., On Semantic Relations in English Complex Nominals and Handling These Relations by Computer, *WG Preprint,* IPSJ, 1985, (in Japanese).

{Ikeda 84} Ikeda, N., Machine Translation by Phrasal Rule Convertion Method, Proc. *Symposium on Natural Language Processing Technology,* IPSJ, 107-114, 1984, (in Japanese).

{Ishihara 74a} Ishihara, Y. and Tamati, T., On the D-Tree Model and Language Analysis for English-Japanese Machine Translation Based on the Model, *Trans. IECE Japan,* Vol. 57-D, No. 7, 435-442, 1974, (in Japanese).

{Ishihara 74b} Ishihara, Y. and Tamati, T., On an English-Japanese Machine Translation System Based on the D-Tree Model and its Experiment, Trans. IECE *Japan,* Vol. 57-D, No. 7, 443-450, 1974, (in Japanese).

{JEIDA 82, 83, 84} JEIDA, *Survey on Machine Translation Activities,* JEIDA (Japan Electronic Industry Development Association), 1982, 1983, 1984, (in Japanese).

{Kaji 85} Kaji, H., Grammar Design Language for Japanese-English Machine Translation System ATHENE/N, *Records of the 31th Convention of IPSJ,* 2H4, 1985, (in Japanese).

{Kogure 84} Kogure, K., Yokoo, A. *et al*., The Frame Editor for Dictionary Editing, *WG Preprints,* WGNL 45-1, IPSJ, 1984, (in Japanese).

{Kurihara *73}* Kurihara, T., Machine Processing of Natural Language, *Information Processing,* Vol. 14, No. 4, 267-281, IPSJ, 1973, (in Japanese).

{Miyazaki 83} Miyazaki, T. and Yasuhara, H., Generation of English Sentences from Semantic Representation Based on the Case Grammar for Japanese, Proc. *Symposium on Natural Language Processing Technology,* 67-71, 1983, (in Japanese).

[Muraki 85] Muraki, K., NEC Machine Translation System VENUS: Two-phase Machine Translation System, *Proc. International Symposium on Machine Translation,* 39-44, 1985.

[Nagao 80a] Nagao, M., Tsujii, J., Ueda, Y. and Takiyama, M., An Attempt to Computerized Data Bases, *Proc. COLING-80,* 1980.

[Nagao 80b] Nagao M. et al., A Machine Translation System from Japanese into English, *Proc. COLING-80,* 414-423, 1980.

[Nagao 81] Nagao, M., The Framework of a System of Mechanical Translation between Japanese and English by Analogy Principle, presented at NATO Symposium: Artificial and Human Intelligence, Chateau de Chapeau Cornu, Oct. 26-30,1981.

[Nagao 82a] Nagao, M., A Survey of Natural Language Processing and Machine Translation in Japan, 1980-1982, in Kitagawa (ed.), *Computer Science and Technologies,* North Holland, 64-70, 1982.

[Nagao 82b] Nagao *et al.,* An English-Japanese Machine Translation System of the Titles of Scientific and Engineering Papers, *Proc. COLING-82,* 245-252, 1982.

{Nagao 83a} Nagao, M. *et al*., Conjunctive Constructions in Japanese Abstracts and Their Analysis Procedures, *WG Preprints,* WGNL 34-4, IPSJ, 1983, (in Japanese).

{Nagao 83b} Nagao, M., An Approach towards a Controlled Language, *Proc. Symposium on Natural Language Processing Technology,* IPSJ, 91-99, 1983, (in Japanese).

{Nagao 83c} Nagao, M., Tsujii, J., Ibuki, J. and Ueda, Y., A Japanese-English Translation System of Titles of Scientific Papers, *WG Preprint,* WGNL 36-4, IPSJ, 1983, (in Japanese).

{Nagao 83d} Nagao, M., Outline of a Machine Translation Project of the Japanese Government, WG Preprints WGNL 38-2, 1983, (in Japanese).

[Nagao 84a] Nagao, M., Nishida, T. and Jun'ichi Tsujii, Dealing with Incompleteness of Linguistic Knowledge on Language Translation, *Proc*. *COLING-84,* 1984.

{Nagao 84b} Nagao, M., *et al.,* Document Production Assistant System Based on Controlled Language, *WG Preprint,* WGNL 44-5, IPSJ, 1984, (in Japanese).

[Nagao 85a] Nagao, M. and Tsujii, J., Machine Translation, in H. Buhlered., *Proceedings of Xth World Congress of FIT,* Wilhelm Braumuller, Wien, 1985.

{Nagao 85b} Nagao, M., Evaluation of the Quality of Machine-Translated Sentences and the Control of Language, *J. Information Processing Japan,* IPSJ, Vol. 26, No. 10, 1197-1202, 1985, (in Japanese).

{Nagao 85c} Nagao, M., *et al*., Outline of Machine Translation Project of the Science and Technology Agency, *J. Information Processing Japan,* IPSJ, Vol. 26, No. 10, 1203-1213, 1985, (in Japanese).

[Nakamura 84] Nakamura, J. *et al*., Grammar Writing System (GRADE) of Mu-Machine Translation Project and its Characteristics, *Proc. COLING-84,* 1984.

[Nishida(F) 80] Nishida, F. *et al.,* English-Japanese Translation through Case-Structure Conversion, *Proc*. *COLING-80,* 447-454, 1980.

[Nishida(F) 82] Nishida, F. *et al.,* Japanese-English Translation Through Internal Expressions, *Proc*. *COLING-82,* 271-276, 1982.

[Nishida(T) 82] Nishida, T. and Doshita, S., An English-Japanese Machine Translation System Based on Formal Semantics of Natural Language, *Proc*. *COLING-82,* 277-282, 1982.

[Nishida(T) 83a] Nishida, T. and Doshita, S., An Application of Montague Grammar to Machine Translation, *Proc*. *ACL Conference on Applied Natural Language Processing,* Santa Monica, 156-165,1983.

[Nishida(T) 83b} Nishida, T., *Studies on English-Japanese Machine Translation System based on Formal Semantics,* Doctoral Dissertation, Kyoto University, 1983.

[Nishida 84] Nishida, T. and Doshita, S., Combining Functionality and Object-orientedness for Natural Language Processing, *Proc COLING-84,* 1984.

[Nitta 82] Nitta, Y., *et al*., A Heuristic Approach to English-Japanese Machine Translation, *Proc*. *COLING-82,* 283-288, 1982.

[Nitta 84] Nitta, Y., *et al*., A Proper Treatment of Syntax and Semantics in Machine Translation, *Proc*. *COLING-84,* 1984.

[Nitta 85] Nitta, Y., Problems of Machine Translation Systems -- Effect of Cultural Differences on Sentence Structure, *Proc. International Symposium on Machine Translation,* separate issue, 1985.

{Nomura 82} Nomura, H., Shimazu, A. and Iida, H., Artificial Intelligence Approach to Machine Intelligence, *WG Preprint,* WGNL 30-5, IPSJ, 1982, (in Japanese).

{Okajima 83} Okajima, A., Nitta, Y. and Yamana F., Editing Functions in Machine Translation, *WG Preprints,* WGNL 40-5, IPSJ, 1983, (in Japanese).

{Okajima 85} Okajima, A., Yamano, F. and Okamoto, E., On Recovery Processing in an English-Japanese Machine Translation System, *Records of the 31th Convention of IPSJ,* 1H2, 1985, (in Japanese).

{Sakai 66} Sakai, T., and Sugita, S., Mechanical Translation of English into Japanese, *J. of IECE Japan,* Vol. 49, No. 2, 46-53, 1966, (in Japanese).

{Sakai 69} Sakai, T., Sugita, S. and Watanabe, A., Mechanical Translation from Japanese into English, *J. of IPS Japan,* Vol. 10, No. 6, 418-427, 1969, (in Japanese).

{Sakaki 82} Sakaki, H. and Hashimoto, K., A Method for English-Japanese Machine Translation Using Pattern Matching, *WG Preprints,* WGNL, IPSJ, 1982, (in Japanese).

{Sakaki 83} Sakaki, H., Hashimoto, K. and Nogaito, I., A Parsing Method Utilizing Forbidden Tree Patterns, *WG Preprints,* WGNL 37-1, IPSJ, 1983, (in Japanese).

{Sakaki 84} Sakaki, H., Hashimoto, K., Suzuki, M. and Nogaito, I., A Parsing Method Having Filtering Procedure, *WG Preprint,* WGNL43-1, IPSJ, 1984, (in Japanese).

{Sakamoto 83} Sakamoto, Y., Design for Japanese Morphological Analysis in Machine Translation, *WG Preprint,* WGNL 38-3, 1983, (in Japanese).

[Sakamoto 84] Sakamoto, Y., *et al.,* Lexicon Features for Japanese Syntactic Analysis in Mu-Project-JE, *Proc. COLING-84,* 1984.

[Sawai 82] Sawai, S., Fukushima, M., Sugimoto, M., and Ukai, N., Knowledge Representation and Machine Translation, *Proc. COLING-82,* 351-356, 1982.

[Schank 77] Schank, R. C. and Abelson, R. P., *Scripts, Plans, Goals, and Understanding,* Lawrence Erlbaum, Hillsdale, New Jersey, 1977.

{Shimazu 85} Shimazu, A., Naito, S. and Nomura, H., Semantic Relations in Japanese Noun Phrases and Their Analysis, *WG Preprint,* 1985, (in Japanese).

{Shudo 73} Shudo, K., On Machine Translation from Japanese into English for a Technical Field, *J of IPS Japan.* Vol. 14, No. 9, 661-668, 1973.

{Shudo 77} Shudo, K., Tsurumaru, H. and Yoshida, S., A Predicative Part Processing System for Japanese-English Machine Translation, *Trans. IECE Japan,* Vol. J60D, No. 10, 830-837, 1977, (in Japanese).

[Sugita 68] Sugita, S., *A Study on Mechanical Translation from English into Japanese,* Doctoral Thesis, Kyoto University, 1968.

[Tamati 63] Tamati, T., Kurihara, T. and Yoshimura, A., The Translation Process and the Design of an Automatic Translator as an Information Processing Machine, *Information Storage and Retrieval,* Vol. 1, No. 1, 1963.

{Tamati 85} Tamati, T., A Historical Overview of Machine Translation, *J. Information Processing Japan,* Vol. 26, No. 10, 1140-1147, 1985, (in Japanese).

{Tanaka 84} Tanaka, H., *et al.,* Approach to More Natural Translation I , *Proc. Symposium on Natural Language Processing Technology,* IPSJ, 115-121, 1984, (in Japanese).

{Tanaka 85} Tanaka, H. and Mizoguchi, F., Semantic Based Machine Translation System, *J. Information Processing,* Vol. 26, No. 10, 1191-1196, 1985, (in Japanese).

{Terano 85} Terano, T., *An Experiment of a Machine Translation System and Some Remarks on its Practical Use,* Economic Research Center Report No. 584008, CRIEPI, 1985, (in Japanese).

{Tomita 84} Tomita, M., Nishida, T. and Doshita, S., A Front-End for Disambiguation in an Interactive English-Japanese Machine Translation System, *Proc*. *Symposium on Natural Language Processing,* IPSJ, 1984, (in Japanese).

[Tomita 85] Tomita, M., *An Efficient Context-free Parsing Algorithm for Natural Languages and Its Applications,* CMU-CS-85-134, Department of Computer Science, Carnegie-Mellon University, 1985.

[Tsujii 84] Tsujii, J., *et al.,* Analysis Grammar of Japanese in the Mu-Project, *Proc*. *COLING-84,* 1984.

[Tsujii 85] Tsujii, J., Science and Technology Agency's Machine Translation Project: Mu Machine Translation Project, *Proc*. *International Symposium on Machine Translation,* 49-68, 1985.

{Tsutsumi 82} Tsutsumi, T., On the English-Japanese Machine Translation of Computer Manuals, *WG Preprints,* WGNL 32-3, IPSJ, 1982, (in Japanese).

[Uchida 80] Uchida, H. and Sugiyama, A Machine Translation System from Japanese into English, Based on Conceptual Structure, *Proc*. *COLING-80,* 455-462, 1980.

[Uchida 85] Uchida, H., Fujitsu Machine Translation System: ATLAS, *Proc*. *International Symposium on Machine Translation,* 29-38, 1985.

{Wada 65} Wada, H., Mechanical Translation, Present Status and its Prospect, *Information Processing,* Vol. 6, No. 3, 1965, (in Japanese).

{Yamada 64} Yamada, S., Analysis of the Input Japanese Text in Mechanical Translation, *Information Processing,* IPSJ, Vol. 5, No. 3, 134-141, 1964, (in Japanese).

{Yamaguchi 85} Yamaguchi, S., An Example of Automatic Translation of Practical Level and Translation Distribution Service, May Issue of *Office and Management,* 14-17, 1985, (in Japanese).

[Yang 84] Yang, Y., Nishida, T. and Doshita, S., Use of Heuristic Knowledge in Chinese Language Analysis, *Proc. COLING-84,* 222-225, 1984.

[Yang 85a] Yang, Y., Doshita, S. and Nishida, T., Partial Constraints in Chinese Analysis, *Proc. IJCAI-85,* 1985.

[Yang 85b] Yang, Y., Studies *on Constructing Chinese Analysis System,* Doctoral Thesis, Kyoto University, Japan, (forthcoming).

{Yoshida 83} Yoshida, S., Tanaka, T., et al., Controlled Japanese for Technical Writing, *WG Preprints,* WGNL 37-4, IPSJ, 1983, (in Japanese).

[Yoshida 84] Yoshida, S., A Consideration on the Concepts Structure and Language in Relation to Selections of Translation Equivalents of Verbs in Machine Translation Systems, *Proc. COLING-84,* 1984.

Appendix: Major Machine Translation Projects and Investigators in Japan

1. The Government
    - The Mu project
        - Kyoto University:
            - Prof. Makoto Nagao, Prof. Jun'ichi Tsujii, Prof. Jun'ichi Nakamura,
        - Electro-Technical Laboratory (ETL):
            - Dr. Yoshiyuki Sakamoto,
        - Japan Information Center of Science and Technology (JICST):
            - Mr. Tsuyoshi Toriumi, Mr. Masayuki Sato
        - Tsukuba Research Information Processing Center (RIPS)
            - {Nagao 83d, 85c}, [Nagao 84a], [Nakamura 84], [Tsujii 84, 85], {Sakamoto 83}, [Sakamoto 84]
2. Universities and Research Institutes
    - Kyoto University
        - Prof. Makoto Nagao, Prof. Jun'ichi Tsujii, Prof. Jun'ichi Nakamura
            - [Nagao 80a,b, 81, 82b, 85a]
        - Prof. Shuji Doshita, Prof. Toyoaki Nishida, Ms. Yiming Yang
            - [Nishida(T) 82, 83a,b, 84],[Yang 84, 85a,b]
    - Kyushu University
        - Prof. Sho Yoshida, Prof. Toru Hitaka [Yoshida 84]
        - Prof. Tuneo Tamati {Ishihara 74a,b}
    - Fukuoka University
        - Prof. Kimiaki Shudo {Shudo 77}
    - Tokyo Institute of Technology
        - Prof. Hozumi Tanaka {Tanaka85}
    - University of Osaka Prefecture
        - Prof. Fujio Nishida, Prof. Shinobu Takamatsu [Nishida(F) 80, 82]
    - Toyohashi University of Technology
        - Prof. Tadahiro Kitahashi, Prof. Makoto Hirai
    - Electro-Technical Laboratory
        - Dr. Shun Ishizaki, Dr. Hitoshi Isahara
        - Dr. Naoshi Ikeda {Ikeda 84}
    - ICOT (the Fifth Generation Project of Japan)
        - focus is attentioned on basic issues; see {Tanaka 84}
3. Computer Industries
    - NTT
        - Dr. Hirosato Nomura, Dr. Akira Shimazu, Dr. Hitoshi Iida, Dr. Katagiri, Dr. Ogura, Dr. Kogure, Dr. Naito
            - {Nomura 82},{Shimazu 85},{Iida 85}
    - FUJITSU
        - Dr. Hiroshi Uchida, Dr. S. Sawai [Uchida 80, 85], [Sawai 82]
    - TOSHIBA
        - Dr. Shin'ya Amano [Amano85]
    - NEC
        - Dr. Kazunori Muraki, Dr. Ichiyama [Muraki 85]
    - HITACHI
        - Dr. Yoshiyuki Nitta, Okajima, K. Kaji [Nitta 82, 84, 85]
    - KDD
        - Dr. Hiroshi Sakaki, Dr. Kazuo Hashimoto {Sakaki 82, 83, 84}

MITSUBISHI
Dr. Eiichi Arita, Dr. Masatoshi Fukushima {Fukushima 83}, {Arita 84}
OKI
Dr. Hiroshi Yasuhara, Dr. Toshihiko Miyazaki {Yasuhara 83}
SHARP
Dr. Suzuki
IBM Japan
Dr. Taijiro Tsutsumi {Tsutsumi 82}

AUTHORS

Toyoaki Nishida, Department of Information Science, Kyoto University, Sakyo-Ku, Kyoto 606, Japan.
Shuji Doshita, Department of Information Science, Kyoto University, Sakyo-Ku, Kyoto 606, Japan.