Japanese-to-English Title Translation System, TITRAN —Its Outline and the Handling of Special Expressions in Titles—

JUN IBUKI*,**, MASAKO KUME*, MAKOTO NAGAO* and JUN-ICHI TSUJII*

A machine translation system, TITRAN, is constructed, which translates the titles of Japanese scientific papers into English. The paper explains the system outline, and the specific parts of the system which handle many special expressions and notations used in title sentences. These are the treatment of unknown words, negation expression, special characters and some other linguistic expressions. The system translated about 86% of the titles correctly in a collection of about 3,300 sample titles by neglecting the handling of articles.

1. Introduction

Machine translation has obtained great interests among computational linguists and in the society in general these days. Its substantial demand has been increasing due to the enormous amount of business and technical documents published every day. There are a few experimental and half-practical machine translation systems, which however have lots of problems to be solved for the practical usage. These systems are still all imperfect ones.

The system we describe in this paper tries to be practical by restricting the range of source texts to a very narrow field of the titles of scientific papers. Titles of scientific papers are much easier, in various points, to translate than ordinary text. Our system, named TITRAN, translates Japanese titles of scientific papers in the field of information science into English. The system was written by PL/1 on FACOM M200 at the computer center of Kyoto University. It was tested by 3379 titles from the Journal of Information Processing Society of Japan. The basic structures of the machine translation system and the handling of many specific expressions which typically occur in the title will be described in detail. Other ordinary components are not described because they have no new points worth describing.

2. Outline of TITRAN

Machine translation of titles has several advantages in the simplicity of the algorithm compared to machine translation of ordinary texts. The reasons are the following:

(1) Polysemous words are rarely found in titles. One of the main difficulties in translation is the problem of translating polysemous words, which have more than one translation equivalents in the target language. Generally speaking, ordinary words are more polysemous than technical words. It is obvious that scientific and technical documents, especially their titles, contain less ordinary words than everyday life texts such as newspapers, novels, etc. We can determine almost uniquely the translation equivalents of technical words in titles, though certain words are still polysemous.

(2) Most titles are noun phrases. We can avoid the burden of syntactic analysis of complicated sentential structures. Verbs still appear in titles, but almost all of them appear in embedded sentences. Furthermore we found that there is at most one verb in a title. We can safely omit the analysis of conjunctive constructions of sentences, deeply embedded sentences, and so on, which are the major troubles in the analysis procedures of ordinary texts.

(3) The constructions of noun phrases are simpler than those in ordinary texts. In ordinary tests we often encounter long conjunctive noun phrases which consist of several of these. Such constructions never occur in titles.

(4) As the consequence of the above properties of (2) and (3), the modifier-modified relationships in titles are expected to be much simpler than ordinary texts. Adverbial phrases, for example, usually modify the nearest predicative words (verbs, adjectives, etc.), and adjective phrases modify the nearest nouns.

These characteristics of titles imply that the analysis procedures can be much simpler and efficient. For example, the titles can be analyzed almost deterministically.

The basic configuration of TITRAN is shown in Fig. 2.1. It consists of two major components, i.e. morphological analysis component and analysis-translation component. Morphological analysis component analyzes a sequence of Japanese characters and converts them into the sequence of words, each with its translation equivalent of English and its syntactic features such as the part of speeches. The analysis-translation component analyzes the word sequence syntactically and transforms the analysis results into English titles.

These two components carry out their tasks by using

*Dept. of Electrical Engineering, Kyoto University.
**At present, Fujitsu Laboratory Ltd.

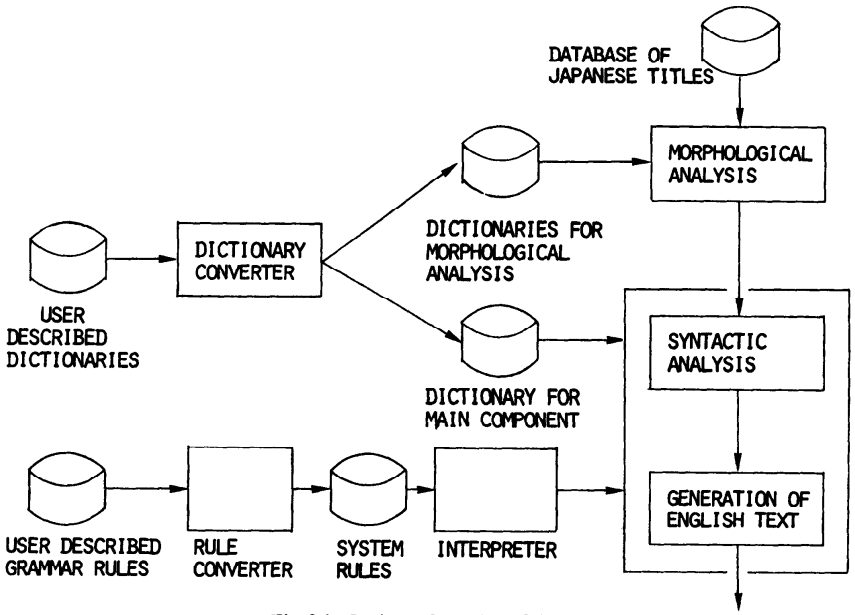

Fig. 2.1 Basic configuration of the system.

their own dictionaries, one for the morphological component, and the other for the analysis-translation component. While the morphological analysis component is directly coded by PL/1 program, the analysis-translation process is driven by the grammatical rules written by a meta-language. A PL/1 program converts the grammatical rules into internal forms. The analysis-translation component interprets these internal forms and applies them to input titles. Another meta-language is also provided for dictionary descriptions. These meta-languages were very useful and effective in the sense that the grammar writers can modify the grammatical rules and dictionary descriptions very easily.

3. Main Flow of Translation

3.1 Morphological Analysis

Since Japanese sentences have no definite separators like spaces in English which mark word boundaries, the first step of the morphological analysis is to segment the source text into individual words. Most of the stems of content words (noun, verb, adjective, etc.) are written in Chinese characters, while the function words such as case particles and inflexional suffixes of verbs and adjectives are written in KANA characters. Therefore these classes of characters give us important clues about how to segment texts into words. We can expect that a content word begins at the position where the kind of character changes from KANA to Chinese character. Certain exceptions exist naturally, almost all of which are content words written in mixed forms of KANA and Chinese characters. We prepared a special dictionary which contained the content words which are often written in mixed forms. By using the information of these and some others a title sentence is first segmented into a sequence of BUNSETSU, which consists of a stem of a content word followed by several inflexional suffixes and function words. BUNSETSU is again segmented into the component words, where the grammatical connectivity test is performed between the adjacent two words.

3.2 Process of Analysis and Translation

The analysis-translation component analyzes a sequence of words syntactically and transforms them into the target language (English). The whole process is

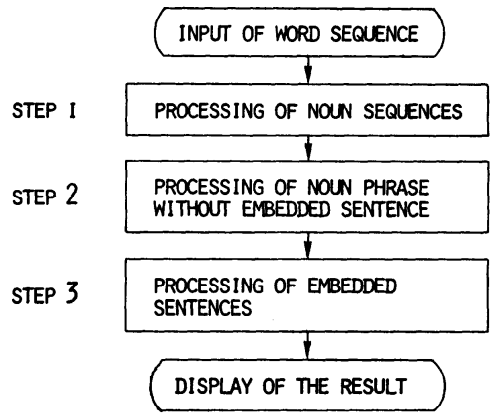

Fig. 3.1 Algorithm for the processing of analysis/translation component.

carried out in the same stage without any intermediate expressions, which means that the translation is given by replacing partial structures of the source text to those of the target. The outline of the process is shown in Fig. 3.1. Step 1 and 2 are the steps for processing noun phrases without embedded sentences. Step 3 is the final phase, where the structures of embedded sentences are analyzed and the whole title is transformed into an English title. Some details are given in the following.

(1) A noun sequence can be translated in most cases by just putting English equivalent of each noun in the same order as Japanese. There are, however, some exceptions, one of which is shown in Ex. 3.1. In this case it will be better to place the last word in Japanese in the first position with a preposition in English. For the processing of such exceptional noun sequences we provide a special description in the dictionary as shown in Ex. 3.2.

Ex. 3.1
TITLE NUMBER=#001092
OBJECT TITLE IS.....
米国における自動翻訳概観
ITS TRANSLATION IS.....
A GENERAL SURVEY ON AUTOMATIC TRANSLATION IN THE UNITED STATES OF AMERICA

Ex. 3.2
```
12900    概観, がいかん
12910    GAIKAN
12920              N    = A GENERAL SURVEY
12921              O-KAKU=ON
12930
```

Table 3.1 Classification of Japanese Parts of Speech.

Classification				examples
verb	SAHEN-verb			処理する 動作する
	other verb			はしる 読む
Auxiliary verb				れる た ている
modifier	adjective		RENTAI-SHI	諸
			KEIYO-SHI (in RENTAI-KEI form)	高価な
	adverb		FUKUSHI	特に
			KEIYO-SHI (in RENYO-KEI form)	適切に
CASE PARTICLE			ADVERBIAL PARTICLE	を で
			ADJECTIVE PARTICLE	に対する の
			CONJUNCTIVE PARTICLE	と 及び
Noun	Chinese word		complex word	複合名詞
			others	動作
	others			コンピュータ

(2) Noun phrases without embedded sentences are processed next. This step determines the interrelationships between nouns followed by an ADJECTIVE-PARTICLES (see Table 3.1) and nouns modified by them. The whole phrase is transformed into a noun phrase with appropriate prepositions inserted between nouns. An example of the conversion of a structure is shown in Fig. 3.2. The prepositions are selected in the following way.

Table 3.2 Classification of generation pattern.

```
1:  N to V    [···]
2:  N of Ving [···]
3:  N for Ving [···]
4:  N in Ving  [···]
5:  N by Ving [···]
6:  N Ving···, (or Ving N)
7:  N Ven···  (or Ven N)
        where N: noun
              V: verb, infinitive
              Ving: verb, ing-form
              Ven: verb, past participle
```

If a preposition is specified in the dictionary entry of a noun, it is selected as the translation. In the case of Ex. 3.2 the preposition 'on' is specified in the noun entry of GAIKAN. Otherwise a preposition associated with each ADJECTIVE-PARTICLES is selected.

(3) Processing of embedded sentences is done as follows. When a keyword like TOKINO, TAMENO, TA follows a verb, an English generation pattern is selected from Table 3.2, which is specified for each keyword. An example is the following.

···+VERB+TAMENO+N→N+FOR+VERB-ING+···

If there is no such keyword the noun modified by the embedded sentence is examined to see if it is a noun of a specific kind. If so, an appropriate generation pattern is selected based on the noun. When these conditions are not satisfied, and when the verb is accompanied by a passive auxiliary verb, their generation patterns are determined by considering their surface cases of the subject(s) as follows.

S-GA+VERB-SARERU+N→N+WITH+S+V-ED
VERB-SARERU+N→N+V-ED

An example is given in Ex. 3.3.

Ex. 3.3
OBJECT TITLE IS.....
ソート・アルゴリズムにより生成されたバイナリーツリーの最適化
ITS TRANSLATION IS.....
OPTIMIZATION OF BINARY TREE GENERATED BY SORT ALGORITHM

When the text is not passive, the generation pattern is determined by considering its surface cases of the subject(s) and the object(o). Table 3.3 shows the patterns.

Table 3.3 Generation pattern in case of active voice Japanese text.

Cases		Verb-type vt	vi
S	O	N with S Ving O	—
S	—	N Ven by S	N with S Ving
—	O	N Ving O	—
—	—	N Ven	N Ving

Where S: SUBJECT, O: OBJECT.

4. Handling of Some Special Linguistic Phenomena

4.1 Unknown Words

Handling of unknown words is inevitable in a practical

machine translation. When an unknown word is detected in the morphological analysis, a tentative translation equivalent and some syntactic features are given by the following algorithm.

(1) Consecutive characters of the same type are regarded as one word or a stem of a word.

(2) If the consecutive characters of (1) are followed by an inflected form of verb "SURU", the whole sequence is regarded as a verb, and the unknown part is regarded as the stem of a SAHEN verb.

(3) If the consecutive characters of (1) are followed by an inflected form of verb "NARU", the whole sequence is regarded as an adjective. If not, the whole segment is regarded as a noun.

(4) When there is a sequence of alphabetic letters or Arabic numerals separated by special characters, they are connected together and are regarded as a noun. It is transferred to English as it is without any transliteration. Examples are given in Ex. 4.1.

Ex. 4.1
KDC—I のプログラミングについて→
ON PROGRAMMING OF KDC—1
National-Elliott 802 における時分割方式→
TIME SHARING SYSTEM IN NATIONAL-ELLIOTT 802

4.2 Negation Expressions

There are several kinds of negative expressions in titles. Because the translation of negative expressions depends on surface environments of them, we will describe each case in the following.

(1) Negative expressions in predicate phrases.

We will restrict our consideration here to those represented by auxiliary verb "NAI". The system translates Japanese sentential expressions either in the form of sentences or in the form of particle clauses (to-infinitive, ing or en-forms) in English. In the former case, auxiliary verb "do" and adverb "not" are inserted before the verb, and appropriate inflexion form is selected. In the latter case, a negative adverb "not" is inserted before the particle or to-infinitive. Examples are given in Ex. 4.2.

Ex. 4.2
FORTRAN を使用しない場合→
CASE NOT TO USE FORTRAN
計算をしない計算機→
COMPUTER THAT DOES NOT DO COMPUTATION

(2) Negation in adjective phrases.

There are many negative expressions by prefixes. They are treated in the same way as those by auxiliary verbs. When an attributive adjective is negated, and if there is a description of its antonym in the dictionary of the adjective, the antonym is used instead of explicit negation. If no antonym is given in the dictionary, the prefix "non-" is attached to the adjective. Examples are given in Ex. 4.3.

Ex. 4.3
正常でない終了→ABNORMAL END
数学的でないデータ
非数学的なデータ NON-MATHEMATICAL DATA

When the specific expressions such as "≥ NAI-BAAI-NO" follow the adjective, "not" is placed before the adjective like "in case···is not (adjective)". The negation of adverbs are generally expressed in Japanese by prefixes such as "HI-" or "FU-". In this case the negation is realized by putting another adverb "not" before the adverb. Examples are given in Ex. 4.4.

Ex. 4.4
誤差が小さくない場合の計算→
COMPUTATION IN CASE ERROR IS NOT SMALL
不正確に計算する計算機→
COMPUTER THAT COMPUTES NOT PLECISELY

4.3 Processing of Special Characters

Many special characters are used in titles to represent special relations. While they provide a strong power of expression in a relatively simple form, they give complications in the computer interpretation of their roles. Some of them are used not as a syntactic component but as a splitter of a title, and as an indicator of volumes, editions, etc. Therefore it is necessary to process the special characters before the main phase of syntactic analysis.

(1) Parentheses

There are several special symbols which are used as parentheses. They are usually used in pairs to express various sorts of concepts such as stress, apposition, insertion, and indicator of volumes, editions, etc. The parentheses for volumes, numbers etc. are usually placed at the end of a title or before a big break. The expressions enclosed by parentheses in this kind of usage are rather simple. They are usually numbers either in Arabic, Latin, or Chinese. They can be repeated, sometimes separated by commas, and NAKAGURO's (middle point between nouns). Examples are given in Ex. 4.5.

Ex. 4.5
OBJECT TITLE IS.....
情報処理システムにおける待ち行列理論の応用(1)
ITS TRANSLATION IS.....
APPLICATION OF QUEUING THEORY IN INFORMATION PROCESSING SYSTEM (1)
OBJECT TITLE IS.....
日本経済シミュレーション実験(1)・(2)
ITS TRANSLATION IS.....
SIMULATION EXPERIMENT OF JAPANESE ECONOMICS (1)・(2)

Parentheses are used to express stress of a word or a phrase, or to show that they are used in restricted or altered meaning. The special symbols "and" alone are used for this purpose in actual titles. There seem no obvious restrictions about where these parentheses are inserted in titles. Almost all words and phrases can be parenthesized. Examples are given in Ex. 4.6.

Ex. 4.6
"Unate" な論理関数→
"UNATE" LOGICAL FUNCTION
OBJECT TITLE IS.....
"密集した" 誤りを訂正する符号
ITS TRANSLATION IS.....
CODE THAT CORRECTS "CROWDED" ERROR

Apposition and insertion are also parenthesized. They are usually noun phrases in titles. Round brackets are used for this type. An example is given in Ex. 4.7.

Ex. 4.7
A—D (アナログーディジタル)変換器の動特性→
DYNAMIC CHARCTERISTICS OF A—D (ANALOG TO DIGITAL) CONVERTER

Mathematical functions sometimes appear in titles, in which case the function arguments are bracketed. This usage is generally distinguished by the property of the arguments in the brackets. They are usually single alphabetic letters separated by commas. An example is given in Ex. 4.8.

Ex. 4.8
I(x, a, b) 計算とその応用例→
I(x, a, b) COMPUTATION AND ITS APPLICATION EXAMPLE

Sometimes a long sentential expression is quoted such as shown in Ex. 4.9. In this case quotation marks in Japanese are converted to double quotation marks in English.

Ex. 4.9
「算術式の直接計算について」に対する回答→
ANSWER TO "ON DIRECT COMPUTATION OF EXPRESSION"
OBJECT TITLE IS.....
論文「スパース行列の確率的考察」について
ITS TRANSLATION IS.....
ON PAPER "PROBABILISTIC CONSIDERATION OF SPARSE MATRIX"

(2) Other special characters

Hyphen, space, comma, period and NAKAGURO's are used in the writing of single words, conjunctive phrases, and to indicate big breaks in titles. Hyphen is often used as a part of a word. They are categorized into two groups according to their surface environments. One is the usage in the sequence of alphabetic characters, and Arabic or Roman numerals. Another is the usage in KATAKANA characters. The former construction is usually used to form a single noun. It introduces an edition or version number of systems such as ALGOL 60 and PL/1, combines strings into a disjunctive (conjunctive) phrase such as CAD/CAM, or shows an abbreviation such as I.F.I.P.

NAKAGURO is often used in KATAKANA strings to represent the spaces in English phrase. They are regarded as a transliteration of the spaces. NAKAGURO is also used to represent the conjunction of nouns, especially nouns written in Chinese characters. Comma is also used for the same purpose.

Colons and hyphens are used to indicate big breaks in titles such as supplementary explanation and subtitle. Because of their roles, they do not appear at the head of the titles. They are often placed at the end when they are used in pairs.

The recognition of these different cases was achieved with reasonable accuracy. Examples are shown in Ex. 4.10.

Ex. 4.10
OBJECT TITLE IS.....
YOHPAC—2100A におけるマイクロプログラミング技術の実際
ITS TRANSLATION IS.....
ACTUAL CONDITION OF MICROPROGRAMMING TECHNOLOGY IN YOHPAC-2100 A
OBJECT TITLE IS.....
超小形プリント・システム
ITS TRANSLATION IS.....
MICROMINIATURE PRINTED SYSTEM
OBJECT TITLE IS.....
ALGOL—部分集合について
ITS TRANSLATION IS.....
ALGOL—ON SUBSET
OBJECT TITLE IS.....
文献情報の蓄積・検索を行なうプログラム
ITS TRANSLATION IS.....
PROGRAM FOR DOING ACCUMULATION, RETRIEVAL OF BIBLIOGRAPHIC INFORMATION

4.4 Modifiers

Problems in the processing of modiffers can be summarized as follows. English modifiers can be categorized into a hierarchy of exclusive classes, and modifiers of each class should be located at their specific positions in the whole sequence of modifiers, if several coexist. Since Japanese language does not have such a strict ordering of phrases, we must rearrange them in the process of translation from Japanese to English.

Some English modifiers have effects upon the inflexion of the succeeding noun. For example, plural form is not explicitly expressed in Japanese, and so their inflexion in English must be decided in the process of translation from Japanese to English. English adjectives and adverbs have the forms of superlative and comparative, which are expressed by certain adverbial phrases in Japanese. We should carefully treat such adverbial phrases to reflect them in the form of English adjectives.

To solve these problems we have classified nominal modifiers as shown in Fig. 4.1. Japanese nominal modifiers are classified into two categories, KEIYO-SHI (adjective) for modifiers having their own case structures and inflected forms, and RENTAI-SHI (noun modifier) for other modifiers.

The relationship between the modifiers and nouns is determined after the noun sequences in texts are

Japanese parts of speech	SYMBOL	examples
RENTAI-SHI	$ RENTAL	ある，諸
KEIYO-SHI	$ K	美しい，確かな

English parts of speech	SYMBOL	examples
pre-determiner	pre	all, both, half, etc.
main-determiner	are	the, a(n), some, etc.
post-determiner	post	few, last, etc.
other adjective	adv	

Fig. 4.1 Classification of modifiers.

processed. When a certain expression like "TOKI-NO", "BAAI-NO", follows the modifiers, they are translated in the forms of embedded sentences. Otherwise, they are translated as the modifiers preceding the noun.

When the adjectives are used as modifiers of noun phrases, the recognition of the case structure of an adjective like subjective and comparative cases (those marked by CASE PARTICLE "YORI") are done, and the proper translation is given. When the adjectives have certain attributes which decide the number property of the modified noun phrase, it is inflected to have the same number property. Ex. 4.11 is a typical example. "SAMA-ZAMA-DA" and '*M*' require the modified noun to be plural and singular respectively. When there is a comparative case or an adverb like "MOTTOMO", which indicates superlative, the inflexion of the adjective is determined appropriately. Then finally the reordering of modifying words is done according to their parts of speeches, in the sequence of pre-determiners, main-determiners, post-determiners, and other adjectives. Examples are given in Ex. 4.12. When adjectives are used as complement (see Ex. 4.13), they need not be reordered.

Ex. 4.11
さまざまな方法→VARIOUS METHODS
一方法→ONE METHOD

Ex. 4.12
最も大きなX→THE LARGEST X
1より小さいX→X SMALLER THAN 1
新しい一方法→ONE NEW METHOD

Ex. 4.13
xが小さい場合のf(x)→F(X) IN CASE X IS SMALL

4.5 Conjunctive Phrases

Although the structures of titles are simpler than those of full sentences, conjunctions do cause a serious problem of many ambiguous interpretations. Particular difficulty exists when the stems of SAHEN-VERB (verb stem + SURU) and KEIYO-SHI are conjuncted together such as shown in Ex. 4.14.

Ex. 4.14
算術演算における誤りを検出または訂正する2進符号
マトリクス形メモリにおけるアクセスおよびサイクル時間

For the recognition of these conjunctions we have classified nouns (especially those written in Chinese character or KATAKANA and the stems of SAHEN-VERB) into the categories shown in Fig. 4.2.

class	example
stem of SAHEN-verb	処理, 統合
stem of KEIYO-SHI	最少
others	

Fig. 4.2 Classification of Japanese nouns.

The processing of conjunctive phrases is performed between the special character handling and the main analysis/translation phases. The process goes as follows.

(1) Among others, SAHEN verb and KEIYO-SHI are processed first, for their conjunctive forms are rather simple compared to those of nouns. This type usually conjuncts two words just before and after the conjunctive particle. When both words connected by a conjunctive particle are SAHEN verb or KEIYO-SHI, the sequence of these words is replaced by a new phrase whose translation is the concatenation of their translation equivalents.

(2) We will process the conjunctive phrases whose scope can be easily recognized by their surface forms. When two CONJUNCTIVE PARTICLES "TO" are used in a pair, the right conjuncted element is determined as the phrase between the two PARTICLES. Ex. 4.15 is an example where the scope of conjunction can be determined by this criteria.

Ex. 4.15
人間と機械との関係→RELATION BETWEEN MAN AND MACHINE

When a keyword specified as an inducer of conjunctive phrases exists, such as "NO", the right element of the conjunction is determined to be the part between the CONJUNCTIVE PARTICLE and the keyword. Examples are shown in Ex. 4.16.

Ex. 4.16
人間と機械の結合→UNIFICATION OF MAN AND MACHINE
人間と計算機の関係→RELATION BETWEEN MAN AND COMPUTER

When a CONJUNCTIVE PARTICLE is followed by a RENTAI-SHI "SONO", the whole left part of the conjunctive particle is supposed to be in the scope of the conjunction. An example is given in Ex. 4.17.

Ex. 4.17
OBJECT TITLE IS.....
語の中の bit を扱う Operation とその回路
ITS TRANSLATION IS.....
OPERATION HANDLING BIT IN WORD AND ITS CIRCUIT

(3) When the scope can not be decided by these steps, it is determined so that the lengths of the conjuncted phrases make balance. We mean here the length of a noun phrase as the number of ADJECTIVE PARTICLES in it. When one element of a conjunction has been determined, the other element is determined to have the equivalent (or ± 1) length in this sense.

The determination of conjunctive phrases is quite difficult even with these heuristic processes. We still have more than thirty percent mis-interpretations. Conjunction of adjectives and verbs seldom appear in titles, and even if they appear, they have rather simple structures compared to those of nouns. Unlike those of noun phrases there are no CONJUNCTIVE PARTICLES between conjuncted parts. Instead, they are marked by the inflexion of the first verb (adjective), or by a comma which follows the first verb (adjective). So when there is a pair of verbs (adjective) separated by commas, they are replaced by a new verb (adjective) whose translation is the concatenation of the translation equivalents of the two verbs (adjectives). An example is given in Ex. 4.18.

Ex. 4.18
応用に富み，安価な方法→FLEXIBLE LOW COST METHOD

4.6 Handling of Compound Stems of SAHEN Verb

Compound stems of SAHEN verbs are to be treated carefully, because there is no simple one-to-one correspondence for them between Japanese and English. English modifiers which correspond to the left component of a compound stem in Japanese must be generated in appropriate forms. Most of the left component can be regarded as case elements, which have lost surface case markers. Some examples are given in Ex. 4.19.

Ex. 4.19
情報処理をする→PROCESS INFORMATION (OBJECTIVE CASE)
計算機処理をする→PROCESS BY COMPUTER (INSTRUMENTAL CASE)

There are some other cases where the modifiers should be translated as adverbs. An example is shown in Ex. 4.20.

Ex. 4.20
情報の自動処理をする→PROCESS INFORMATION AUTOMATICALLY

SAHEN verbs has three types of expressions as shown in Fig. 4.3. One is the case where the stem of a SAHEN

Pattern	Example
N-O N_V-SURU	情報を処理する
N-NO N_V-O SURU	情報の処理をする
NN_V SURU	情報処理する

Fig. 4.3

verb is directly followed by an inflexional form of verb "SURU". An example is given in Ex. 4.21. Another is the case where the stem of a SAHEN verb is followed by the sequence of CASE PARTICLE "WO", and an inflexional form of a verb like "SURU", "OKONAU", etc. In this case a SAHEN stem becomes a noun. An example is given in Ex. 4.22, which has the same meaning as the example of Ex. 4.21.

Ex. 4. 21
計算機において情報を自動的に処理するためのアルゴリズム
Ex. 4.22
計算機における自動的な情報の処理を行なうためのアルゴリズム

The third one is that the stem is a compound noun of two components. It is noticeable from these examples that a phrase expressed by an adverbial phrase can be expressed by an adjective. That means that ADVERBIAL PARTICLES are changed to the corresponding ADJECTIVE PARTICLES, and adverbs are also transformed into adjectives. So it will be one of our concerns to treat those two forms equivalently.

In principle the second type is converted to the first type. For this purpose we have developed the following algorithm. The algorithm is executed before the main analysis/translation phase.

We see first which form of the above two a SAHEN verb takes. When it takes the latter form (i.e. a noun form allowed by CASE PARTICLE "WO"), an ADJECTIVE PARTICLE "NO" is transformed to ADVERBIAL PARTICLE "WO", since it is most likely to fit an objective case. An example is shown in Ex. 4.23.

For the third type the following judgement is made. When the word N has an adverb as one of its translation

Ex. 4.23
文献情報の検索をするプログラム→
PROGRAM FOR RETRIEVING BIBLIOGRAPHIC INFORMATION

equivalents, then it is taken as an adverb with the specified translation. An example is in Ex. 4.24. Other-

Ex. 4.24
情報の自動処理をするプログラム→
PROGRAM FOR PROCESSING INFORMATION AUTOMATICALLY

wise the word N is regarded as a case element without case markers. The omitted case marker is assumed as "WO" (objective case). When the objective case explicitly exists in the title, the whole sequence of SAHEN verb is replaced by a new verb with the concatenation of their translation as a new translation. Examples are shown in Ex. 4.25.

Ex. 4.25
情報処理をする→PROCESS INFORMATION
情報の割り込み処理をする→INTERRUPT-PROCESS INFORMATION

5. Evaluation of the System

The results of our latest experiment over 3379 titles from the Journal of Information Processing Society of Japan are summarized in Table 5.1. It obtained the correct interpretation of about 86% for these titles by the condition that such features as articles are not considered in this system and ignored in this statistics. Typical failures are as follows.

Table 5.1 Result of Title Translation.

Number of Titles Tested	3,379	100%		
Success	2,911	86%	3,044	90%
Failure	468	24%	335	10%
Segmentation errors	31	6.6		
Program errors	64	13.7	≠	
Rules not prepared	83	17.7		
Syntactic errors in conjunctive phrases	106	22.6		
Syntactic errors in other phrasal relations	51	10.9		
Misinterpretation of special symbols	26	5.6		
Bad translations	38	8.1		
Errors in original text data	49	10.5	≠	
Words not found in the dictionary	20	4.3	≠	

*) Errors marked ≠ were easily recoverable. When these are added to the success, the success rate comes to 90%.

There are no conjunctives of clauses in Japanese RENYO-KEI, and an inflexional form of verbs or adjectives is used to connect various sorts of clauses including both co-ordinate and sub-ordinate clauses. Problems are caused by the fact that the same form, RENYO-

CHUSHI, expresses various relationships among clauses. We must recognize relationships of the deep meanings among clauses to decide whether they are to be translated in English as co-ordinate or sub-ordinate clauses, and in the case of sub-ordinate clauses, which conjunctives or constructions like participial clauses by -ing should be used to express them.

Our system recognizes some sub-ordinate constructions by regarding RENYO-KEI of certain verbs as ADVERBIAL PARTICLES. An example is given in Ex. 5.1. In this example "WO-MOCHIITE" is regarded as an adverbial particle as a whole, though it is taken in traditional grammars as an independent phrase with RENYO-KEI of the verb "MOCHIIRU". Other cases can not be handled in the system.

Ex. 5.1
計算機を用いてダイヤグラムを作る Algorithm
→ALGORITHM FOR MAKING DIAGRAM USING COMPUTER

Errors in conjunctive noun phrases are found in about 20% of the failures in our system. Since there are few surface evidences for deciding their scopes, certain semantic considerations must be introduced to improve the results. It is, however, rather difficult to decide what kind of semantic features is useful for this purpose.

In spite of our assumption of noun phrases as titles, few sentential expressions appeared in titles. Therefore we modified the system to deal with simple declarative sentences. Interrogative and imperative sentences are, however, not processed by our system.

Selection of an appropriate article is completely neglected in our system. In general it requires a complicated process to decide whether noun phrases are definite or indefinite from the surface evidences in Japanese. Certain heuristics such as follows will be useful to decide them.

Certain sequences of keywords and prepositions seem to entail the attribute (definite, indefinite) for the succeeding noun phrases, such as "for the purpose of (definite)···". Some nouns are used with only definite (or indefinite) articles in most cases, such as "future (definite)". Some adjectives seem to entail definite (or indefinite) articles, such as "main (definite)". A noun with a long list of modifiers seems to have definite articles. These heuristics should be added to the next version.

6. Conclusion

Through the development and experiment of this system, we recognized that the restrictions on the source text of titles are very effective for reducing the burden on the grammatical rules and the English dictionaries.

When our system is extended to other subject areas, it is expected to be done without any major modifications in grammars. The only thing required might be to supplement the dictionary of technical terms. It may, however, cause another problems of choice from many translation equivalents depending on subject fields. We cannot estimate at present how ambiguous our technical terms might become, and how we could manage such ambiguities by using subject area codes.

References

1. NAGAO, M., TSUJII, J., YADA, K. and KAKIMOTO, T. An English Japanese Machine Translation System of the Titles of Scientific and Engineering Papers, *Proc. COLING* **82**, (July 1982).
2. UEDA, Y. Japanese-English Machine Transalation of Titles of Scientific Papers (in Japanese), MS Thesis, Kyoto University, Dept. of Electrical Engineering, (Mar. 1982).
3. NAKAMURA, T. Morphological Analyzer for Japanese Text (inJapanese), BS Thesis, Kyoto University, Dept. of Electrical Engineering, (Mar 1982).
4. HUBERT, J., IBUKI, J., KUME, M. and NAGAO, M. A System of Japanese-French Automatic Translation for Paper Titles, *26th Annual Convention of the Information Processing Society of Japan*, (Mar. 1983).

(Received July 26, 1983; revised October 24, 1983)