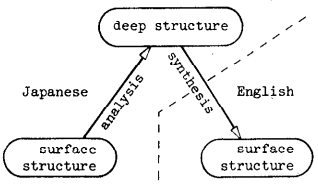
On Machine Translation from Japanese into English for a Technical Field

Kimiaki Shudo*

1. INTRODUCTION

The difficulties in development of Machine Translation (MT) or machine processing of natural language are caused by various ambiguities in natural languages. Without efficient semantic processing, one can not overcome them. One of the applications of our research on the natural language processing system is the development of a MT system from Japanese into English, in which the manifoldness of the word-meaning is reduced to some extent. Our MT is designed for the field of transistor circuits for the reason that the word-meaning can be narrowed by the proper viewpoints and the semantic process is simplified in the specialized field.

fig. 1 The notion of processing



The system is characterized by processing through the deep structure shown in fig. 1. In this paper, we give a brief description of our semantic approach and informations prescribed for it. For the syntax analysis of Japanese sentences and for the process linking the syntactic process to the semantic process, one should refer to [1,4] and [5], respectively. The notations used in this paper are as follows. N: a bunsetsu (i.e. sentence element composed of an independent part and an annex part) whose independent part is a noun; P: a bunsetsu whose independent part is a yougen (verb, adjective or adjective verb) or a noun annexed by "*dearu*"; E(x): an expression in English which is equivalent to the expression x in Japanese.

2. SEMANTIC WORD CLASSES

The main process in the analysis is to grasp the semantic relations between bunsetsu's on the basis of cases. To prescribe these semantic relations in machine dictionaries, we have set up approximately 70 semantic word classes on approximate 1 000 words. In the course of classification, great cares were taken so that decisions on

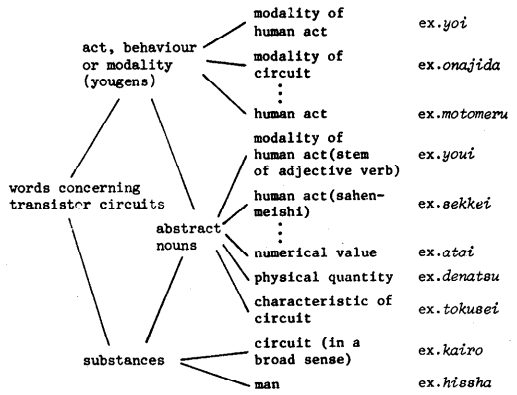
This paper first appeared in Japanese in Joho－Shori (Journal of the Information Processing Society of Japan), Vol. 14, No. 9 (1973), pp. 661～668.

* Faculty of Engineering, Fukuoka University, Fukuoka, Japan

the following matters can be made in the analysis without ambiguity: 1) case-relations; 2) equivalents, verb patterns and prepositions in English; 3) meanings of rengo's (i.e. compounds). (See fig.2) In general, the classes are not disjoint because the word-meaning has several sides.

We prescribe the word-meaning in our Word-dictionary by specifying all classes containing the word, each of which is associated with a proper equivalent. In order to decide the meaning and equivalents of rengo's composed of prefixes, suffixes and nouns, we use approximately 50 rewriting rules of the form $\alpha \cdot \beta \rightarrow \gamma$ (t), where α, β or γ is a word class and t represents the attached informations for the synthesis. The rule means that a word in α can be annexed by a word in β and the resulting compound should be in γ.

fig. 2 The outline of the word classification



3. SEMANTIC BASE STRUCTURE OF SENTENCE

In our system, the main part of the case-structure is prescribed as Semantic Base Structure of Sentence (SBSS). In SBSS-dictionary, we describe the cases that can be taken simultaneously by each yougen in a sentential expression, and annex words implying the cases and also lists of word classes which can play the roles in the cases. The notion of SBSS is identical to that which is described in [1,5] with the exception that we only specify the 'primary' cases which are essential for deciding the verb patterns or unusual prepositions in the synthesis. One verb pattern required by the equivalent verb is attached to each SBSS. Two examples are shown in table 1. In our MT, we use approximately 2 000 SBSS's for 400 yougen's.

4. SECONDARY CASE RELATIONS

The cases we call 'secondary' which are not entered in SBSS correspond to the prepositional phrases. Such cases and their word substitution capabilities are specified in NP-dictionary not for a single P but for a class of P's for the reason that

table 1 Two examples of SBSS's for "*motomeru*", where (SUB) ((OBJ)) is an equivalent that plays as the subject(object) case

SUB (subject)		OBJ (object)		informations for synthesis
word class	annex word	word class	annex word	
man	*ga*	circuit,•••,method	*o*	(SUB)+*DETERMINE*+(OBJ)
man	*ga*	characteristic of circuit, •••,numerical value	*o*	(SUB)+*CALCULATE*+(OBJ)

many P's require the same word classes for a case. Table 2 shows three examples of the specification.

table 2 Examples of secondary case relations expressed by "*de*" specified with word classes

word class of N	word class of P	case	preposition
place of description	human act	PLACE	*IN*
physical quantity, numerical value	movement of circuit, modality of circuit	ABSTRACT METHOD	*AT*
possessions of man, possessions of circuit	movement of circuit	ABSTRACT METHOD	*WITH*

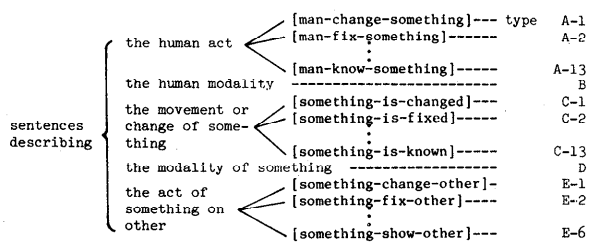
5. PP-RELATIONS

The dependency relations between two P's are called 'PP-relations'. These relations are regarded as the cause-effect relations in a broad sense, which are specified as binary relations between two sentences in PP-dictionary. We have classified sentences to make the above specifications according to their rough meanings and set up approximately 40 sentence types. (See fig.3) Some annex words are transformed into one of the 20 standard ones for which the relations are specified. For example, the sentence "*shuuhasuuo kaetemo ritokuwa kawaranai.*" is transformed into "*shuuhasuuo kaeruto ritokuga kawaru.*" and then analyzed. Though each sentence type is designated for each SBSS, some annex words require the transformation of types. For example, the type of the sentence "*denatsuo joushousaseru.*" is decided from the type of "*denatsuga joushousuru.*".

6. MODIFICATION RELATIONS

The processings of modification relations are mostly reduced to those of the case relations. NN-relations are processed as follows, where we assume that N_1 precedes N_2 in the input sentence: A) When N_2 is a bunsetsu changed from a yougen (e.g. sahen-meishi), the N_1N_2-relation and the corresponding preposition can be decided if N_1 can be some case of the P which is the original of N_2. For example, the relation in the expression "*kairodeno gensui*" is recognized by examining the case relation in "*kairode gensuisuru*" and the noun phrase "*ATTENUATION+IN+CIRCUIT*" is obtained. We have set up the rules to presume the cases from approximate 60 annex words; B) In cases other than A), the following two kinds of process are required: (i) After filling up the unexpressed P, we examine whether N_1 and N_2 can be respective cases of it. For example, the relation in "*kairono tokusei*" is recognized by examining the case relations in "*kairoga tokuseio motsu.*" and the noun phrase "*CHARACTERISTIC+OF+CIRCUIT*" is obtained. Several P's (i.e. "*motsu*","*arawasu*") and two cases (

fig. 3 The outline of the sentence classification



subject and object) suffice this process and the preposition "*OF*" is used in the output; (ii) We consult NN-dictionary in which the relations are prescribed as binary or ternary relations between word classes.

The PN-relation is decided when N can be a case of P and corresponds to a relative clause. Other modification relations (i.e. adjective → P, adjective verb → P and adverb → P) are prescribed in respective dictionaries as binary relations between word classes. For some specific adverbs which are usually uttered in concert with some annex words of P, exceptional processes are considered but they are not mentioned here.

7. OUTLINE OF MT PROCESS

The general flow of our MT process and the rules or dictionaries used on each step are shown in fig.4. The input sentence written in wakachigaki way (i.e. a way of writing sentence by leaving one space between two bunsetsu's) is assumed to be semantically possible. By using the tree search method, we can find all possible patterns of semantic relations between bunsetsu's and synthesize the output sentences respectively. In the analysis we simplify the semantic process by making use of the syntactic features of Japanese sentence (e.g. that the arcs expressing the dependency relations in the sentence do not cross). Connectives, idioms and parallel relations are also taken into considerations in our processing but they are omitted here. In the remainder of this chapter, we outline

fig. 4 The general flow of MT experiment

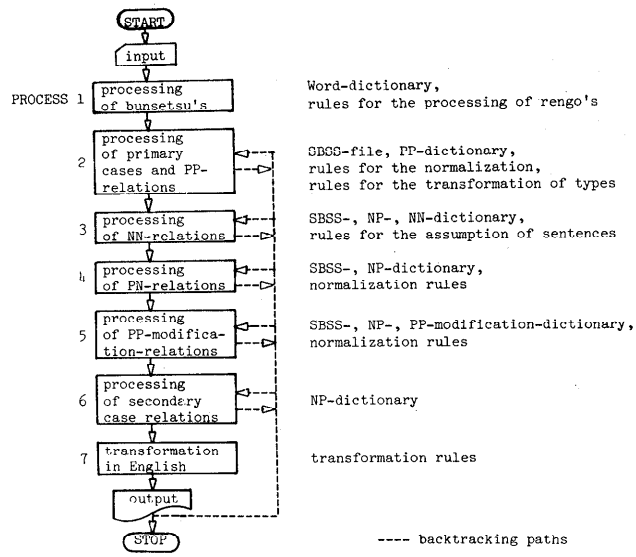


table 3 Segmented bunsetsu's and their estimated functions

bunsetsu	name	functions as a dependent
zoufukuki·wa	N_1	case
singou·to	N_2	case, parallel
zatsuon·to·o	N_3	case
sikibetsu·suru	P_1	terminate, rentai
nouryoku·o	N_4	case
mot·a·nai·kara	P_2	PP
zoufukuki·de	N_5	case
shouj·iru	P_3	terminate, rentai
zatsuon·mo	N_6	case
zoufuku·s·a·re·te	P_4	PP
shutsuryokutansi·ni	N_7	case
toutatsu·s·uru	P_5	terminate, rentai

table 4 Word classes and equivalents

word class / N	circuit	physical quantity	possessions of circuit	part of circuit	possessions of man
zoufukuki	*AMPLIFIER*	----	*AMPLIFIER*	----	----
singou	----	*SIGNAL*	*SIGNAL*	----	----
zatsuon	----	*NOISE*	*NOISE*	----	----
nouryoku	----	----	*CAPABILITY*	----	*CAPABILITY*
shutsuryokutansi	----	----	----	*OUTPUT TERMINAL*	----

word class / P	mental act	modality of circuit	phenomenon	movement of circuit
sikibetsusuru	*(DISCRIMINATE)*	----	----	*(DISCRIMINATE)*
motsu	----	*(HAVE)*	----	----
soujiru	----	----	*(ORIGINATE)*	----
zoufukusuru	----	----	----	*(AMPLIFY)*
toutatsusuru	----	----	*(REACH)*	----

table 5 SBSS's

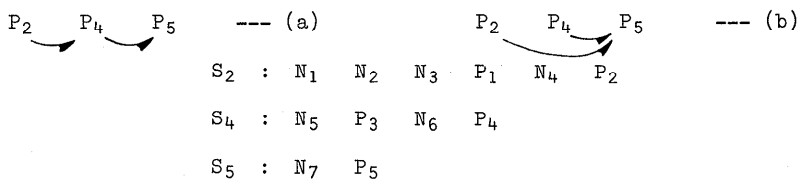
	SUB		OBJ		OBJ-1		OBJ-2		PLACE		type	informations for synthesis
	word class	annex word	word class	annex word	word class	annex word	word class	annex word	word class	annex word		
sikibetsusuru	circuit	*ga*	----	--	physical quantity	*to*	physical quantity	*o, too*	----	--	C-3	(SUB)+*DISCRIMINATE+BETWEEN* +(OBJ-1)+*AND*+(OBJ-2)
motsu	circuit	*ga*	possessions of circuit	*o*	----	--	----	--	----	--	D	(SUB)+*HAVE*+(OBJ)
shoujiru	physical quantity	*ga*	----	--	----	--	----	--	circuit	*de*	C-4	(SUB)+*ORIGINATE+IN*+(PLACE)
zoufukusuru	circuit	*ga*	physical quantity	*o*	----	--	----	--	----	--	E-1	(SUB)+*AMPLIFY*+(OBJ)
toutatsusuru	physical quantity	*ga*	----	--	----	--	----	--	circuit	*ni, e*	C-4	(SUB)+*REACH+TO*+(PLACE)

the semantic process by means of an example.

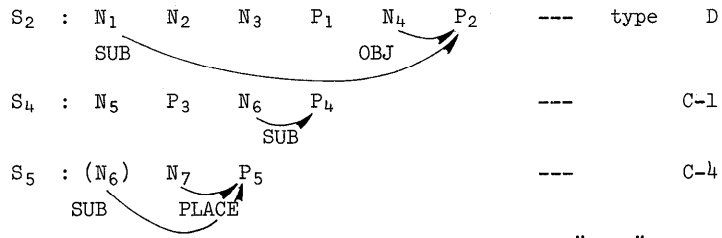
INPUT SENTENCE S: *zoufukukiwa singouto zatsuontoo sikibetsusuru nouryokuo motanaikara zatsuonmo zoufukusarete shutsuryokutansini toutatsusuru.*

PROCESS 1) Table 3 and 4 show the results.

PROCESS 2) S is separated into three subsentences S_2, S_4 and S_5, and the following patterns of PP-relations are obtained.



The primary case-structures extracted from S_2, S_4 and S_5 and their sentence types are as follows. Table 5 shows the SBSS's applied here.



The type of S_4 is decided on the basis of the annex word "*reru*" and the annex word "*mo*" in N_6 is changed to "*ga*". And the subject lacking in S_5 is substituted by N_6.

Next, PP-dictionary is looked up and both of the patterns (a) and (b) are shown to be reasonable. (See table 6) We only describe the processing of pattern (b) from now on.

The framework of the output sentence is obtained as follows.

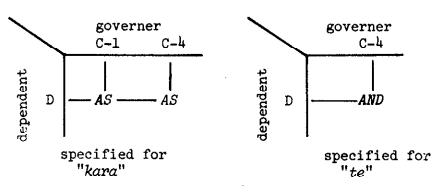
E(S) : *AS* + E(S_2) + , + E(S_4) + *AND* + E(S_5) + .

VERB PATTERN OF E(S_2): *AMPLIFIER+HAVE+CAPABILITY*

VERB PATTERN OF E(S_4): *NOISE+BE+AMPLIFIED*

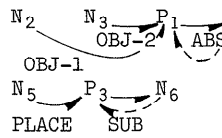
VERB PATTERN OF E(S_5): *NOISE+REACH+TO+OUTPUT +TERMINAL*

table 6 PP-relations and conjunctions



PROCESS 3) Skipped.

PROCESS 4) The modification structures are decided as follows, where the secondary case relation obtained here is shown in table 2.



--- OUTPUT EXPRESSION: *CAPABILITY+TO+DISCRIMINATE+BETWEEN+SIGNAL+AND+NOISE*

--- OUTPUT EXPRESSION: *NOISE+THAT+ORIGINATE+IN+AMPLIFIER*

PROCESS 5,6) Skipped.

PROCESS 7) After adding the modalities in S_2 and S_4 and deleting the redundant subject in S_5 and also inflecting the verbs, we obtain the following output sentence.

OUTPUT SENTENCE E(S): *AS + AMPLIFIER + HAS + NOT + (CAPABILITY + TO + DISCRIMINATE + BETWEEN + SIGNAL + AND + NOISE) + , + (NOISE + THAT + ORIGINATES + IN + AMPLIFIER) + IS + AMPLIFIED + AND + REACHES + ALSO + TO + OUTPUT + TERMINAL + .*

8. CONCLUDING REMARKS

Our MT outlined in this paper is designed to cope with the ambiguities based on the manifoldness of the word-meaning and to accomplish the syntax-for-syntax translation through the deep structure. By our analysis procedure, however, we can not take care of the logical inconsistency. In order to analyze the meaning of pronoun in detail or to decide the adequate article in the synthesis in addition to overcome the above difficulty, we have to establish the processes of detailed inferance and environmental control based on the discourse analysis, which are left for a future study. In our initial experiment, approximately 1 000 words and 4 000 rules for semantic processing are considered. We hope that the semantic process and informations will be improved and enriched in the course of experiments.

ACKNOWLEDGMENT

I would like to acknowledge the continuing guidance and suggestions of Prof. T. Kurihara blessed memory of Kyushu University and Prof. S. Yoshida of Kyushu Institute of Technology.

REFERENCES

1. Kurihara, T., "Machine Processing of Natural Language," (Commentary), Journal of Information Processing Society of Japan, Vol. 14, No. 4, 1973.
2. Langendoen, D. T., "Essentials of English Grammer," Holt, Rinehart & Winston, 1970.
3. Rustin, R.,(ed.), "Natural Language Processing," Algolithmic Press, 1973.
4. Yoshida, S., "Syntax Analysis of Japanese Sentence Based on Kakariuke Relation

between Two Bunsetsu," Trans. of the Institute of Electronics and Communication Engineers of Japan, Vol. 55-D, No. 4, 1972.

5. Yoshida, S.,Kurihara, T. and Matsuo, F., "Machine Processing of Japanese - From Syntax Analysis to Semantic Analysis," Trans. of the Institute of Electronics and Communication Engineers of Japan, Vol. 55-D, No. 6, 1972.