DECOMPOSITION OF JAPANESE SENTENCES INTO NORMAL FORMS BASED ON HUMAN LINGUISTIC PROCESS

Tsutomu Endo

Department of Information Science
and Systems Engineering
Oita University
Oita 870-11, Japan

Tuneo Tamati

Department of Information Systems
Graduate School of Engineering Sciences
Kyushu University
Fukuoka 812, Japan

A diversity and a flexibility of language expression forms are awkward problems for the machine processing of language, such as translation, indexing and question-answering. This paper presents a method of decomposing Japanese sentences appearing in the Patent Documents on "Pulse network", into normal forms. First, the linguistic information is analysed and classified based on the human linguistic process. Then, predicate functions, phrase functions and operators are introduced as the normal forms. Finally, the decomposing procedure and some experimental results are shown.

Introduction

One of the most remarkable features of natural language is a diversity and a flexibility of its expression form. Especially, Japanese appears to have a peculiar syntactic structure because it is an agglutinative language. This is an awkward problem for the machine processing of language, such as translation, subject indexing and question-answering. An approach to dealing with this problem is to transform the sentences into some normal forms if any. Proposals for such normalization have been made for some time, but there have been few attempts.[1,2]

The normal form needs to have every information which is contained in original sentences. Let us now consider what information the sentences contain. In human linguistic process, the objects to be exprssed are provided first, then the cognitive structure corresponding to them is formed, and lastly the language expression based on the cognitive structure is produced. In other words, the immediate basis of language expression is considered to be human cognitive structure. Therefore, the arrangement of words in sentences represents not only the relation among objects in the external world, but also the cognitions and the relations among them, which are relatively independent of the present objects.

This paper presents a method of decomposing Japanese sentences into normal forms based on such human linguistic process. First of all, the linguistic information necessary for decomposing process is analysed and classified from the above mentioned point of view. Then, predicate functions, phrase functions and operators are introduced as the normal forms. Two kinds of function describe the syntactic structure of the sentences and phrases. The operator describes the relationship among functions. Finally, the decomposing procedure and some experimental results are shown. Sample sentences are selected from the claim points of the Japanese Patent Documents on "pulse network".

Analysis of linguistic information

In this section, we analyse and classify the linguistic information necessary for decomposing Japanese sentences into their normal forms.

Classification of words

From the standpoint of the linguistic process, that is, objects, cognitions and expressions, all words are divided into objective expressions W_1 and subjective expressions W_2. W_1 is the set of expressions which reflect external objects, namely, conceptual expressions. On the other hand, W_2 is the set of cognitive expressions without conceptual process, and immediately represents the affection, judgement, desire, will and so on. The detail of the classification of words is summarized in Table 1. We give supplementary explanations about Table 1. Adjective II is the words which are called stem of adjectival verb in the traditional Japanese grammar. For inflectional words such as AA_n, V_n, TB_n and JJ_n, we specify n as 1, 2, 3, 4, and 5(6) according to inflectional forms, that is, negative, declinable word modifying, final, noun modifying, and conditional(imperative) form respectively.

Analysis of cognitive structure

In order to describe the content of words and the relation among words, we introduce the descriptive scheme M which consists of such five descriptors as follows;

$$M = < O, \Sigma, U, \Psi, \Lambda >.$$

(1) $O = \{s_{\nu 1}$(substance), $a_{\nu 2}$(attribute), $r_{\nu 3}$(relation)$\}$. O is the cognitive unit formed by separating and abstracting the external objects ideally, and is classified into three large categories, namely, substances, attributes and relations. The symbol νi specifies the variety and the abstracting level of each unit. Thus, O is regarded as the classification of concepts in the objective world (e.g., pulse network).

(2) $\Sigma = \{\sigma_1, \sigma_2, \sigma_3\}$. Σ describes the relationship between objects from the various view points. σ_1 is the relationship between substance and attribute, σ_2 is the relationship between substance and relation, and σ_3 is the

various connection of the same kind of objects.
(3) U represents the active cognitions which are relatively independent of concepts.
(4) Ψ specifies the cognitive behaviors how the speaker cognize the objects.
(5) $\Lambda = \{\lambda_1(\text{tense}), \lambda_2(\text{anaphora})\}$. Λ represents the relation between a speaker and objects.

A part of O, Σ, U and Ψ is tabulated in Table 2-5 respectively.

Definition of predicate function

In this section and following two sections, we define the normal forms of Japanese sentences.[3] Generally, a sentence expresses the property of an object, or the relationship among objects. The component which indicates such property or relationship, is the predicate of a sentence. So we introduce the function, the constants of which are the predicate and the case postpositions, and the variables of which are noun phrases just in front of case postpositions. This function is called predicate function and is expressed by

$$X_1 a_1 X_2 a_2 \cdots X_i a_i \cdots X_n a_n P$$

where X_i, a_i and P indicate the noun phrase, the case postposition and the predicate respectively.

[Example]

1. (SOOTI) GA (ZIZOKUHA) WO (PULSE) NI KAERU.
   X_1 a_1 X_2 a_2 X_3 a_3 P
   (A device converts continuous wave into pulse train.)
2. (DENATU) WO (TEIKOOKI) NI KUWAETA.
   (Someone applied voltage across a resistor.)
3. (DENRYOKU HENKA) GA TIISAI.
   (The variation in power is small.)

Table 1 Classification of words

Category				Symbol	Example
Objective expression (W_1)	Noun (N)	Common noun		NA	transistor
		Attribute noun	Dynamic	NBA	HASSIN(oscillation)
			Static	NBB	SEI(plus)
		Abstract noun		NC	MONO(thing)
		Pronoun		ND	KORE(this)
		Numeral		NE	ICHI(one)
	Adjective (A)	Adjective I		AA_n	OOKII(large)
		Adjective II		AB	KYUUGEKI(rapid)
	Verb (Vn)	Common verb		VA_n	KUWAERU(add)
		Special verb		VB_n	HASSIN-SURU(oscillate)
		Abstract verb		VCn	SURU(do)
	Uninflected noun modifier			RR	ARU(certain)
	Atribute adverb	Attribute adverb I		DA	SIDAINI(gradually)
		Attribute adverb II		DB	KIWAMETE(very)
	Prefix			HH	HU(non-)
	Suffix	Nominal suffix		TA	KA(-ize)
		Verbal suffix		TB_n	SASERU(make)
	Special symbols			EX	∿, (,)
	Compound word			RE	KAKAWARAZU (in spite of)
Subjective expression (W_2)	Auxiliary verb			JJ_n	YOO(will)
	Post-position (X)	Case postposition		XA	GA, NO, NI
		Dependent postposition		XB	WA, MO
		Adverbial postposition		XC	DAKE(only), NADO
		Conjunctive postposition		XD	TO, TEMO
	Conjunction			CC	OYOBI(and)
	Assertive adverb			YY	MOSI(if)
	Punctuation points	Comma		ZA	,
		Period		ZB	.

Table 2 List of descriptor O

Symbol	Descripter	Example
s	Substance	——
s_1	Material substance	MONO(thing)
s_{11}	Body(or Object)	TAI(body)
s_{111}	Functional body	HUKA(load)
s_{1111}	Circuit element	trannsistor, diode, condenser
s_{1112}	Circuit device	HASSINKI(oscillator)
s_{1113}	Device or System	KEISANKI(computer)
s_{112}	Material	HANDOOTAI(semiconductor0
s_{113}	Particle	DENSI(electron)
s_{114}	Abstract group	TUI(pair), GUN(group)
s_{12}	Phenomenon	——
s_{121}	Wavy phenomenon	pulse, SEIGENHA(sine wave)
s_{122}	Electric phenomenon	DENATU(voltage), DENRYOKU(power)
s_{123}	Functional phenomenon	SINGOO(signal)
s_{124}	Other phenomenon	ZATUON(noise)
s_{125}	Attribute phenomenon	drift
s_2	Ideal substance	KOTO(event)
s_{21}	View points	——
s_{211}	Characteristic	Zener TOKUSEI(characteristic)
s_{212}	Type	KATA(type), pattern
s_{213}	Quantity	RYOO(quantity)
s_{2131}	Impedance	TEIKOO(resistance)
s_{2132}	Wavy quantity	SYUUHASUU(frequency)
s_{214}	State	ZYOOTAI(state)
s_{215}	Property	SINRAISEI(reliability)
s_{216}	Shape	KUKEI(rectangle)
s_{217}	Degree	HODO(degree), level
s_{218}	Value	ATAI, TI(value)
s_{219}	Content	NAIYOO(content)
s_{22}	Number	0, 1, 2, 3, 4
s_{23}	Method	HOOHOO(way), SYUDAN(method)
s_3	Space	——
s_{31}	Direction	ZYUN-HOOKOO(foward direction)
s_{32}	Place	TEN(point)
s_{321}	Terminal	collector, emitter
s_{33}	Pass	pass, loop
s_{34}	Boundary	limit
s_{35}	Part of substance	HEN(side), peak
s_{36}	Scope	HANI(scope)
s_4	Time	——
s_{41}	Temporal point	TOKI(time)
s_{42}	Time interval	KIKAN(time period)
a	Attribute	——
a_1	Dynamic attribute	——
a_{11}	Operation	SURU(do), SASERU(make)
a_{111}	Concrete operator	ASSYUKU(compress),control
a_{12}	Event	——
a_{121}	Change	KAWARU(change)
a_{1211}	Process of change	ZOOKA(increase), KOOKA(drop)
a_{1212}	Effect of change	HOOWA(saturate), KOSYOO(trouble)
a_{122}	Remove	NAGARERU(float)
a_{1221}	Place of removal	TOORU9pass)
a_{123}	Input/Output	——
a_{1231}	Input	KUWAERU(add), ATAERU(give)
a_{1232}	Output	HASSEI(generate)
a_{124}	Continuation	ZIZOKU(continue)
a_{125}	Movement	DOOSA(work)
a_{1251}	Concrete movement	KYOOSIN(resonance)
a_{126}	Display	HYOOZI(display, show)
a_{127}	Start and stop	KAISI(start)
a_{13}	Generation of relation	——
a_{131}	Connection	SETUZOKU(connect)
a_{132}	Switching	HIRAKU(open), switch
a_{133}	Composition	KUMIAWASERU(combine)
a_{134}	Addition	TUKERU(attach)
a_{135}	Separation	BUNRI(separate)
a_{14}	Action	——
a_{141}	Usage	SIYOO, TUKAU(use)
a_{142}	Judgement	HANBETU(discriminant)
a_{143}	Determination	SADAMERU(settle), KIMERU(decide)
a_{15}	Passive	RARERU, HI
a_2	Static attribute	——
a_{21}	Possibility	DEKIRU(be able to)
a_{22}	Difference of quality	OOKII(large), YOI(good)
a_{23}	Property	SENKEI(linear), digital
a_{24}	Relational attribute	——
a_{241}	Positional relation	TYOKURETU(series)
a_{242}	Difference	HITOSII(equal), TIGAU(differ)
a_{243}	Conformity	MUKU(fit)
a_{244}	Dependency	DOKURITU(independent)
a_{245}	Possession	MOTU(have, own)
a_{246}	Opposition	HANTAI(opposite)
a_{247}	Necessity	HITUYOO(necessary)
a_{248}	Temporal relation	TUNE(always, usually)
a_{249}	Other relation	SOOHOTEKI(complementary)
a_{25}	Existence	ARU(exist), NAI(empty)
a_{26}	Comparison of degree	KIWAMETE(very, much)
a_{27}	Circumstances	ANZEN(safety), HEIKOO(balance)
a_{28}	Rank	SYU(main), KIZYUN(standard)
a_{29}	Abstract situation	SIDAINI(gradually)
r	Relation	——
r_1	Position	——
r_{11}	Front and rear	MAE(in front of, before)
r_{12}	Midway	AIDA, KAN(between)
r_2	Reference	——
r_{21}	Complement	HOKA(other)
r_3	Causal relation	——
r_4	Correspondence	——
r_5	Whole and part	——
r_{51}	Part	BUBUN(part)
r_6	Arithmetic relation	HI(ratio), HANBUN(half)

However, a predicate P has a variety of expression form in Japanese. For example, a verb is frequently connected with some auxiliary verbs(e.g., NAI(negative), TA(past)) or verbal suffixes(e.g., RARERU(passive), SASERU(causative)). Therefore, we decompose the predicate P into objective expression P_O and subjective expression P_S. Then, we define the basic predicate function as the function which consists of the following four kinds of predicate P_OP_S.

(1) P_O(Final form of verb) P_S(Zero element of speaker's judgement),
(2) P_O(Final form of adjective I) P_S(Zero element of speaker's judgement),
(3) P_O(Adjective II) P_S(Judgement expression "DA(be)"),
(4) P_O(Noun) P_S(Judgement expression "DA(be)").

The application of operators presented in next section, inflects the form of P_O or P_S. Other predicate functions are defined by the application of operators to basic predicate functions. Thus, the predicate functions are classified as follows.

Predicate function { Constant function (ideomatic expression) / Basic predicate function / Derivative function }

The predicate generally represents some attribute concept. Unlike substances an attribute does not occur alone. It arises accompanying substances. When we cognize an attribute as the concept, there exist some substances which accompany this attribute. The variables corresponding to these substances are called obligatory variables of the predicate, and the case postpositions, obligatory ones a_{ei}. On the other hand, one substance usually accompanies various kinds of attribute, and is related to other substances as a mediation of this attribute. In the predicate function, the variables corresponding to such attributes and substances are called facultative variables, and the case postpositions, facultative ones a_{oi}.

The variables of a predicate function have some domains of their own, that is to say, substitutable word classes. So we specify the domain of variables in terms of the descriptor O. Also, the relationship between the predicate and each variable is given by the descriptor Σ. These are summarized in Table 6.

Definition of operators

The operator produces a new function from one or two functions. They are classified into six groups, that is, modal(F_I), nominalization (F_{II}), embedding(f_{III}), connecting(F_{IV}), elliptical(F_V) and anaphoric operator(F_{VI}).

Modal operator

The modal operators consist of the objective expressions F_{I1}(e.g., abstract verb, verbal suffix, a part of prefix) and the subjective expressions F_{I2}(e.g., auxiliary verb, adverbial postposition). F_{I1} applies to P_O of the predi-

Table 3 List of descriptor Σ

Symbol		Descripter	Example
σ_1	$\sigma_{1,1}$	Simple connection between substance and attribute	(IMPEDANCE) GA TAKAI (The impedance is high.)
	$\sigma_{1,2}$	Object of action or operation	(SINGOO) WO ZOOHUKU-SURU (A thing amplifies signal.)
	$\sigma_{1,3}$	Starting point in action	(TRANSISTOR) KARA NARU (A thing consists of transistor.)
	$\sigma_{1,4}$	Finishing point in action	(SAIDAITI) NI TASSURU (A thing amounts to maximum value.)
	$\sigma_{1,5}$	Opponent in mutual action	(DIODE) TO KETUGOO-SURU (A thing connects with diode.)
	$\sigma_{1,6}$	Standard or reference	(ZATUON) NI YORU (A thing depends on noise.)
	$\sigma_{1,7}$	Way or means	(PULSE) DE HANTEN-SURU (A thing turns with pulse.)
	$\sigma_{1,8}$	Spatial positioning	(COLLECTOR KAN) NI ARU (A thing is between collectors.)
	$\sigma_{1,9}$	Temporal positioning	(TOKI) NI NAGARERU (A thing floats when·····.)
	$\sigma_{1,10}$	Others	(YOO) NI KETUGOO-SURU (A person connects a thing so as to····.)
σ_2		Substance with the relation	AIDA[r_{12}*σ_2*s_{32}, r_{12}*σ_2*s_{42}] (space or time between ···)
σ_3	$\sigma_{3,1}$	Order of substance	DAI (2) (TRANSISTOR) (second transistor)
	$\sigma_{3,2}$	Number of substance	(2) KO NO (TRANSISTOR) (two transistors)
	$\sigma_{3,3}$	Property of substance	(PULSE) NO SYUUHASUU (frequency of pulse train)
	$\sigma_{3,4}$	Material of substance	(HANDOOTAI) (DIODE) (semiconductor diode)
	$\sigma_{3,5}$	Unit of substance	(400 NM) NO (HATYOO) (wavelength of 400 nm.)
	$\sigma_{3,6}$	Various connection among attributes	(HEIRETU) (SETUZOKU) (parallel connection)

Table 4 List of descriptor U

Symbol	Descripter	Example of subjective expressions
u_1	Affirmative judgement	DA, ARU (be, do)
u_2	Negative judgement	NAI, NU (not)
u_3	Special judgement	WA, MO
u_4	Universal judgement	WA
u_5	Purpose or aim	——
u_6	Will	U, YOO (will, shall)
u_7	Assumption	BA, MOSI (if)
u_8	Certification	TA
u_9	Inference	U, TABUN (probably)
u_{10}	Desire	——
u_{11}	Natural judgement	BEKI (should, have to)
u_{12}	Instance (toward ideal premises)	NADO (···so on)
u_{13}	Limitation (toward ideal premises)	DAKE (only)
u_{14}	Exess (toward ideal premises)	MADE (even)

Table 5 List of descriptor Ψ

Symbol	Cognitive behavior
ψ_0	Cognizing object O faithfully
ψ_1	Cognizing attribute a as substance
ψ_2	Cognizing static attribute a dynamically
ψ_3	Cognizing dynamic attribute a statically
ψ_4	Cognizing causal relation of events backward
ψ_5	Cognizing U as objective substance ideally
ψ_6	Cognizing U as objective attribute ideally
ψ_7	Cognizing one object concretely and abstractly
ψ_8	Cognizing a degree of attribute as quantity
ψ_9	Cognizing one object from the various view points
ψ_{10}	Conjunctive enumeration
ψ_{11}	Disjunctive enumeration

cate, and varies the mode of the attribute which is expressed by the function. On the other hand, F_{I2} applies to P_s, and varies the mode of the judgement. An example of F_{I1} and F_{I2} are shown in Table 7-8 respectively.

Nominalization operator

The nominalization operators apply to one predicate function and nominalize it in the following way.

(1) f_{II1} : Cognizing one of the objects expressed by the predicate function, as the substance with attribute.

(DIODE) GA (HOODEN ZIKAN) WO HAYAMERU.
(A diode advances the time of discharge.)
→ (HOODEN ZIKAN) WO HAYAMERU DIODE
(A diode which advances the time of discharge.)

(2) f_{II2} : Recognizing the concrete event expressed by the predicate function, as substance ideally.

(HAKEI) GA NAMARU
(The wave form is blunted.)
→ (HAKEI) GA NAMARU KOTO (or NO)
(that the wave form is blunted.)

(3) f_{II3} : Transforming the predicate function into clauses which express the time, reason, state, effect and so on.

(DENKA) WO KENSYUTU-SURU
(A thing detects the electric charge.)
→ (DENKA) WO KENSYUTU-SURU TOKI
(when a thing detects the electric charge.)
→ (DENKA) WO KENSYUTU-SURU TAME
(in order to detect the electric charge.)
→ (DENKA) WO KENSYUTU-SURU YOO
(so as to detect the electric charge.)

Table 6 Example of basic predicate function

Predicate P	Variable	Case postposition		Σ	Domain
ATAERU (give)	X_1	a_{e1}	WO	$\sigma_{1,2}$	s_{12}
	X_2	a_{o1}	NI	$\sigma_{1,4}$	s_{32}
KENSYUTU-SURU (detect)	X_1	a_{o1}	GA	$\sigma_{1,1}$	s_{1112}
	X_2	a_{e1}	WO	$\sigma_{1,2}$	s_{12}, s_{213}
KOSU (exceed)	X_1	a_{e1}	GA	$\sigma_{1,1}$	s_{122}
	X_2	a_{e2}	WO	$\sigma_{1,8}$	s_{218}
SETUZOKU-SURU (connect)	X_1	a_{e1}	WO	$\sigma_{1,2}$	s_{111}, s_{32}
	X_2	a_{e2}	NI	$\sigma_{1,4}$	s_{111}, s_{32}
	X_3	a_{o1}	NI	$\sigma_{3,6}$	a_{241}, a_{27}
	X_4	a_{o2}	DE	$\sigma_{1,7}$	s_{111}
	X_5	a_{o3}	DE	$\sigma_{1,10}$	s_{31}
TAMOTU (keep)	X_1	a_{o1}	GA	$\sigma_{1,1}$	s_{1112}
	X_2	a_{e1}	WO	$\sigma_{1,2}$	s_{12}
	X_3	a_{o2}	NI	$\sigma_{1,6}$	s_3
DOOTUU-SURU (conduct)	X_1	a_{e1}	GA	$\sigma_{1,1}$	s_{1111}
	X_2	a_{o1}	DE	$\sigma_{1,7}$	s_{12}
ITTEI (constant)	X_1	a_{e1}	GA	$\sigma_{1,1}$	s_{213}, s_{216}
OOKII (large)	X_1	a_{e1}	GA	$\sigma_{1,1}$	s_{213}
TUYOI (strong)	X_1	a_{e1}	GA	$\sigma_{1,1}$	s_{1112}
	X_2	a_{o1}	NI	$\sigma_{1,6}$	s_{12}

N.B. In sample sentences, a substance "human being" is not considered explicitly, so the variable corresponding to it is omitted in this Table.

Table 7 Example of modal operators F_{I1}

Symbol		Operator	Content	Usage
F_{I11}	f^6_{I11}	SURU (make)	a_{11}	(IMPEDANCE) GA TAKAI (The impedance is high.) (MONO) GA (IMPEDANCE) WO TAKAKU SURU (A thing makes the impedance high.)
	f^7_{I11}	NARU (become)	a_{121}	(IMPEDANCE) GA TAKAI (IMPEDANCE) GA TAKAKU NARU (The impedance becomes high.)
F_{I12}	f^2_{I12}	RERU RARERU (be able to)	a_{21}	(IMPEDANCE) WO TAKAMERU (A thing increases the impedance (IMPEDANCE) WO TAKAME RARERU (A thing is able to increase the impedance.)
	f^3_{I12}	RERU RARERU (passive)	a_{15}	(DENATU) WO (TEIKOOKI) NI KUWAERU (A thing applies voltage across the resistor.) (DENATU) GA (TEIKOOKI) NI KUWAE RARERU (Voltage is applied across the resistor.)
	f^4_{I12}	SERU SASERU SIMERU (make)	a_{11}	(DENATU) GA HENKA-SURU (The voltage varies.) (SOOTI) GA (DENATU) WO HENKA SASERU (A device makes the voltage vary.) (SOOTI) GA (PULSE) WO HASSEI-SURU (A device generates the pulse train.) (SOOTI) NI (PULSE) WO HASSEI SASERU (A thing makes a device generate the pulse train.)
F_{I14}	f^1_{I14}	KA (-able)	a_{21}	KA SEIGYO (controllable) KA HOOWA (saturatable)
	f^2_{I14}	HI (-ed)	a_{15}	HI SEIGYO (controlled) HI SOKUTEI (measured)
F_{I15}	f^1_{I15}	KA (-ize)	a_{121}	KOGATA KA (making small) IC KA (integration)
	f^2_{I15}	TEKI	a_{23}	ZIKAN TEKI (temporal) DENKI TEKI (electrical)

Table 8 Example of modal operators F_{I2}

Symbol		Operator	content	Usage
F_{I21}	f^1_{I21}	DA, ARU (be)	u_1	(SWITCHING DOOSA) GA SEIKAKU DE ARU (The switching operation is correct.)
	f^2_{I21}	NAI, NU (not)	u_2	(KAIRO) GA (COIL) WO HUKUMA ϕ NAI ϕ. (A network does not contain a coil.)
	f^3_{I21}	U, YOO (will)	u_6	(OOKISA) WO HANTEI-SI ϕ YOO ϕ. (We will decide the size.)
	f^4_{I21}	BESI	u_{11}	HANTEN-SURU ϕ BEKI TRANSISTOR (a transistor to turn)
	f^5_{I21}	TA	u_8	(SOOTI) GA DOOSA-SI TE IRU (A device is working.)
	f^6_{I21}	TA (past)	λ_1	(TRANSISTOR) GA HANTEN-SI ϕ TA ϕ. (A transistor turned.)
F_{I22}	f^1_{I22}	HU	u_2	HU ITTI (disagreement) HU KANZEN (imperfect)
F_{I23}	f^1_{I23}	WA	u_3	(TRANSISTOR) GA (SINGOO) WO ZOOHUKU-SURU (The transistor amplifies the signal.) (TRANSISTOR) WA (SINGOO) WO ZOOHUKU-SURU (SINGOO) WA (TRANSISTOR) GA ZOOHUKU-SURU
			u_4	(TRANSISTOR) WA NOODOO SOSI DA (A transistor is an active element.)
	f^2_{I23}	MO	u_3	(PULSE) WO HASSEI-SURU (A thing generates the pulse train.) (PULSE) MO HASSEI-SURU (A thing generates the pulse train too.) (PULSE NO HASSEI) MO SURU (PULSE) WO HASSEI-SI MO SURU
F_{I24}	f^1_{I24}	NADO	u_{12}	(SOSI) GA (HASSINKI) NADO NI TEKISURU (The element is suitable for the oscillator and other things.)
	f^2_{I24}	DAKE	u_{13}	(PULSE) DAKE WO HASSEI-SURU (A thing generates the pulse train only.) (PULSE) WO HASSEI-SURU DAKE DA (A thing generates only the pulse train.)
F_{I25}	f^1_{I25}	MOSI (if)	u_7	MOSI (ZATUON) GA HASSEI-SURE BA, (If the noise generates,)
	f^2_{I25}	TATOE	ψ_4	TATOE (SYUUKI) WO KAE TEMO, (SINPUKU) WA ITTEI DA (Even if we vary the period, the amplitude is constant.)

(NYUURYOKU SINGOO) WO HENTYOO-SURU
(A thing modulates the input signal.)
→ (NYUURYOKU SINGOO) WO HENTYOO-SI TA SINGOO
(the signal which is modulated by the input signal.)

(4) f_{II4} : Cognizing the only attribute as substance.
(PULSE) WO HASSIN-SI WA SURU
(A thing generates the pulse train.)

(5) f_{II5} : Cognizing the event expressed by the predicate function, as substance immediately.
(ONDO) GA HENKA-SURU
(A temperature changes.)
→ ONDO NO HENKA, or ONDO HENKA
(A change in temperature.)

The clause or noun phrase which is produced by the application of the nominalization operator, is substituted in the variable of other predicate function by embedding operator f_{III}.

Connecting operator

A connecting operator joins one predicate function to another coordinately or subordinately. Generally, it corresponds to conjunctions and conjunctive postpositions. Some operators are related to modal operators, attribute adverbs, or variety of predicate. It is classified into following six groups.

(1) Conjunctive connecting operator(f_{IV1})
S_1 : (SYOOHI DENRYOKU) GA TIISAI
(The consumption power is small.)
S_2 : (SWITCHING ZIKAN) GA MIZIKAI
(The switching time is short.)
↓ $S_1 * f_{IV1} * S_2$
(SYOOHI DENRYOKU) GA TIISAKU, (SWITCHING ZIKAN) GA MIZIKAI (The consumption power is small, and the switching time is short.)

(2) Simultaneous conjunctive connecting operator(f_{IV2})
S_1 : (TRANSISTOR) WO KUDOO-SURU
(A thing drives the transistor.)
S_2 : (HOOWADO) WO SEIGYO-SURU
(A thing controls the saturation rate.)
↓ $S_1 * f_{IV2} * s_2$
(TRANSISTOR) WO KUDOO-SURU TO DOOZI NI (HOOWADO) WO SEIGYO-SURU (The moment a thing drives the transistor, it controls the saturation rate.)

(3) Disjunctive connecting operator(f_{IV3})
S_1 : (CONDENSER) WO SETUZOKU-SURU
(A person connects a capacitor.)
S_2 : (COIL) WO IRERU
(A person inserts a coil.)
↓ $S_1 * f_{IV3} * s_2$
(CONDENSER) WO SETUZOKU-SURU KA (COIL) WO IRERU (A person connects a capacitor, or inserts a coil.)

(4) Causal connecting operator(f_{IV4})
S_1 : (DENRYUU) GA (SYOTEITI) WO KOSU
(The current exceeds the fixed value.)
S_2 : (DENATU HENKA) GA SYOOZIRU
(The voltage changes.)
↓ $S_1 * f_{IV4} * S_2$
(DENRYUU) GA (SYOTEITI) WO KOSU TO (DENATU HENKA) GA SYOOZIRU (The voltage changes when the current exceeds the fixed value.)

(5) Concessive connecting operator(f_{IV5})
S_1 : (SYUUKI) WO KAERU
(A person changes the period.)
S_2 : (SINPUKU) GA ITTEI-DA
(The amplitude is constant.)
↓ $S_1 * f_{IV5} * S_2$
(SYUUKI) WO KAE TEMO (SINPUKU) GA ITTEI-DA
(Even if a person changes the periode, the amplitude is constant.)

(6) Modificatory operator(f_{IV6})
S_1 : (TEIKOO) WO KAISURU
(Through the resistor)
S_2 : (BASE) WO (DENGEN) NI SETUZOKU-SURU
(A person connects the base to the power source.)
↓ $S_1 * f_{IV6} * S_2$
(TEIKOO) WO KAISI TE (BASE) WO (DENGEN) NI SETUZOKU-SURU (A person connects the base to the power source through the resistor.)

Generally, more than one connecting operator is applied in the actual sentences. So we define the universal connecting formula as follows. Let f_{II} and f_{III} be the nominalization and the embedding operator respectively. An arbitrary predicate function A_i is expressed by

$$A_i = A_{i_1} * f_{IV1} * A_{i_2} * f_{IV1} * \cdots * f_{IV1} * A_{i_k} * f_{IV1} * \cdots * f_{IV1} A_{i_m}$$

where A_{i_k} is
(i) S_u,
(ii) $[A_i * f_{IVd} * A_j]$ (d = 2,3,4,5,6).

S_u is the basic predicate function, or the derivative function which is produced by the application of more than one modal operator, and is called unit predicate function. Moreover, the embedding operator is sometimes applied to S_u in the following way.

$S_u(f_{III}\text{-}A_1', A_2', \ldots, A_i', \ldots, A_n')$

where $A_i' = f_{II} A_i$.

Other operators

When one predicate function is produced by the application of the connecting operator to two functions, the elliptical operator omits the one of the same expression forms in the two functions and anaphoric operator replaces the one of the same expression forms with the pronoun.

Definition of phrase function

We introduce the phrase function in order to describe the structure of noun phrases or compound words. However, it is not easy to define the phrase function based on the word class, unlike the predicate function. So we classify the phrases according to their content, and define the phrase function based on this classification. An example of phrase function is listed in Table 9.

G_1 is the phrase connected in terms of such relational concepts as position(r_1), reference (r_2), and part(r_5). G_2 is the phrase formed by cognitive behaviors(Ψ), such as enumeration(ψ_{10}, ψ_{11}), cognition of one object from the various view point(ψ_9), concrete and abstract cognition of one object(ψ_7), and so on. G_3 is the phrase constructed in terms of the relationship(σ_1) between substance and attribute, and the various

connection(σ_3) of the same kind of objects. G_4 is other phrases.

Decomposition process

The new derivative functions can be produced by the application of the various operators to the basic predicate functions. This means that the sentences with complex syntactic structure correspond to one predicate function. Therefore, the normalization of sentences is the decomposition of the predicate function corresponding to these sentences, into a set of basic predicate functions, phrase functions and operators. In this section, we describe the decomposing procedure[4].

Machine dictionary

A machine dictionary consists of three elementary dictionaries, that is, word dictionary(WD), predicate function dictionary(PFD) and related concept dictionary(RCD). WD is utilized to acquire the basic linguistic information of each words in input sentences. PFD is given to the candidate word for predicate, such as verb, adjective, and so on, and is used to extract the predicate function from sentences and phrases. RCD is stored with the relation between concepts, and is used for not only the decision of embedded phrase but also the analysis of phrases. Table 10 shows an example of each dictionary.

Procedural description

General flow of decomposition process.

The general procedural flow and the data flow of decomposition process are shown in Fig.1 and Fig.2 respectively. Input Japanese sentences spelled in Roman letters are segmented word by word with spaces.

Each word is matched with entry words of WD. The word list(WLIST) is constructed based on the information from WD. The candidate for predicate (e.g., verb, adjective) is found by searching WLIST from the head of the list. Then, the modal operator (F_{I11}, F_{I12} and F_{I21}), embedding operator f_{III} and connecting operator F_{IV} are extracted by investigating the variety and the inflectional form of the predicate or the words which follow the predicate. The extracting method of these operators is shown in Fig.3. The extracted information is stored in FLIST 1 and CLIST. The variables of the predicate function are extracted by reference to PFD. At the same time, the modal operators F_{I23} and F_{I24} are extracted, if any. If the obligatory variable of the function is omitted, the word whose concept is coincident with the domain of the variable, is found from the extracted word string in WLIST. This is regarded as the application of the elliptical operator. When the embedding operator applies to the predicate, the variety of the nominalization operator and the embedded phrase are

Table 9 Example of phrase functions

	Symbol	Phrase function	Structure	Example
G_1	g_{101}	{RYOO|2|ε}w KAN	$w/w*r_{12}*P$	RYOO BASE KA (between bases)
		w_1 w_2 KAN	$w_1/w_2*r_{12}*P$	BASE COLLECTOR KAN (between base and collector)
	g_{103}	(TAHOO NO|TA)w_m	TAHOO$*r_{22}*w_m$	TAHOO NO KAIRO (another circuit)
	g_{105}	w{NO|ε}w_m	$w*r_{51}*w_m$	THYRISTOR NO GATE (gate of thyristor)
		w{NO|ε}w_m	$w_m*r_{51}*w$	TRANSISTOR KAIRO (transistor circuit)
	g_{110}	KAKU w_m	KAKU$*r_{24}*w_m$	KAKU DIODE (each diode)
G_2	g_{201}	w_{m_1}(TO|OYOBI|,|ε)w_{m_2}(TO|ε}	ψ_{10}-w_{m_1}/w_{m_2}	TEIKOO TO DIODE (resistor and diode)
	g_{202}	w_{m_1} MATAWA w_{m_2}	ψ_{11}-w_{m_1}/w_{m_2}	TEIKOO MATAWA CONDENSER (resistor or capacitor)
	g_{203}	w w_m	ψ_9-w/w_m	PULSE DENATU (pulse voltage)
	g_{204}	w w_m	ψ_7-w/w_m	ZOOHUKU SAYOO (amplifying operation)
G_3	g_{301}	DAI w w_m	w-$\sigma_{3,1}$-w_m	DAI 2 TRANSISTOR (second transistor)
	g_{302}	w{KO NO|TU NO}w_m	w-$\sigma_{3,2}$-w_m	2 KO NO TRANSISTOR (two transistors)
	g_{306}	KIZYUN w_m	KIZYUN-$\sigma_{1,1}$-w_m	KIZYUN DENGEN (standard power source)
	g_{308}	w{NO|ε}w_m	w-$\sigma_{3,3}$-w_m	PULSE NO SINPUKU (amplitude of pulse)
		w{NO|ε}w_m	w_m-$\sigma_{3,3}$-w	KOO IMPEDANCE SOSI (high impedance element)
G_4	g_{401}	{KONO|SONO}w_m	D-w_m	KONO TRANSISTOR (this transistor)

N.B. w_m indicates the main component of the phrase.

Table 10 Structure of machine dictionary

(a) Word dictionary (WD)

Entry word	Category	Code	Concept	Pointer
TRANSISTOR(transistor)	NA	300	$s_{1111}(\psi_0)$	- 1
GA(~)	XA	1	—	- -
SETUZOKU(connect)	VB	1010	$a_{131}(\psi_0)$	1 -
COLLECTOR(collector)	NA	410	$s_{321}(\psi_0)$	- 2
DENRYUU(current)	NA	376	$s_{122}(\psi_0)$	- n
SASERU(make)	TB_{34}	24	$a_{11}(\psi_0)$	- -
HENKA(change)	VB	1025	$a_{121}(\psi_0)$	2 -
HANDOOTAI(semiconductor)	NA	343	$s_{112}(\psi_0)$	- 3
HENTYOO(modulate)	VB	1018	$a_{111}(\psi_0)$	3 -
OOKIKU(large)	AA_{12}	1206	$a_{22}(\psi_0)$	4 -
BEKI(should)	JJ_4	32	u_{11}	- -
KONO(this)	RR	112	—	- -
DAI(large)	HH	1206	$a_{22}(\psi_0)$	4 -

(b) Predicate function dictionary (PFD)

NO.	Number of variable	Designator*	Case postposition	Number of domain	Domain	Character string of predicate
1	5	0	WO	2	s_{111}, s_{32}	SETUZOKU-SURU
		0	NI	2	s_{111}, s_{32}	
		1	NI	2	a_{241}, a_{27}	
		1	DE	1	s_{111}	
		1	DE	1	s_{31}	
2	2	0	GA	2	s_{12}, s_{21}	HENKA-SURU
		2	GA	1	s_{1112}	
3	2	0	WO	1	s_{12}	HENTYOO-SURU
		3	—	1	s_{12}	
4	1	0	GA	1	s_{213}	OOKII

* 0(obligatory variable), 1(facultative variable), 2(special variable due to f^6_{I11} and f^4_{I12}), 3(special variable due to f_{II3})

(c) Related concept dictionary (RCD)

NO.	Number	Variety	Direction	Level*	Related concept
1	3	r_{51}	→	0	410, EMITTER**, BASE**
	1	r_{51}	←	1	s_{1112}
2	1	r_{51}	←	0	300
	2	$\sigma_{3,3}$	→	0	376, DENATU(voltage)**
3	2	$\sigma_{3,4}$	←	1	s_{1111}, s_{1112}

* 0(code), 1(concept)
** The code is stored in actual dictionary.

decided. The extracted information is stored FLIST 1, and the word strings of the variables are stored in VLIST. These word strings are decomposed into basic predicate functions, nominalization operators and phrase functions, and then stored in FLIST 2 and GLIST. The above procedure are repeated for other predicate candidates. Finally, the connecting formula which indicates the relation among predicate functions are formed by reference to CLIST.

Processing of phrases. At first, the procedure finds the candidate for predicate, such as dynamic attribute noun, declinable word modifying form of common verb, prefix (e.g., "KOO(high)", "TEI(low)", "DAI(large)", etc.) and adjective II, from the word strings stored in VLIST. If the candidate is found, the basic predicate function, nominalization operator and embedded word are extracted. If not, the phrase function are extracted. They are classified into three types according to decision method.

[Type I] Phrase functions extracted by the features of their constant. The example are g_{101}, g_{201}, g_{301}, and so on, in Table 9. Their constants, such as "RYOO(both)", "KAN (between)", "TAHOO(another)", "DAI", "KO", etc., are given the priority based on the strength of the connectability to variable, and are stored in constant list. The phrase function of this type is extracted according to priority.

[Type II] Phrase functions extracted by using RCD. The examples are g_{105}, g_{308}, and so on.

[Type III] Phrase functions extracted by using the variety or level of word concept. For example, g_{203} is extracted by investigating whether the upper concepts of both words agree with each other or not, and g_{204} is done by investigating whether the concept of second word

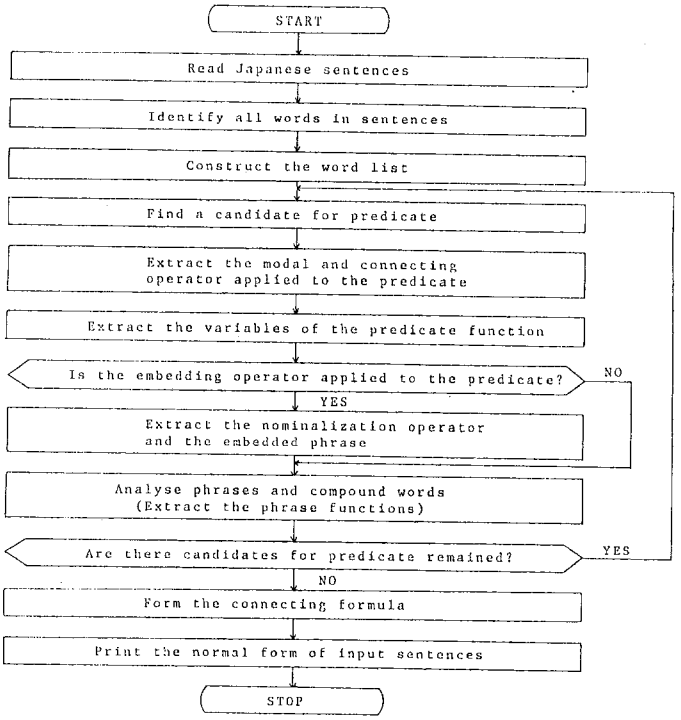

Fig.1 Decomposing procedure of Japanese sentences

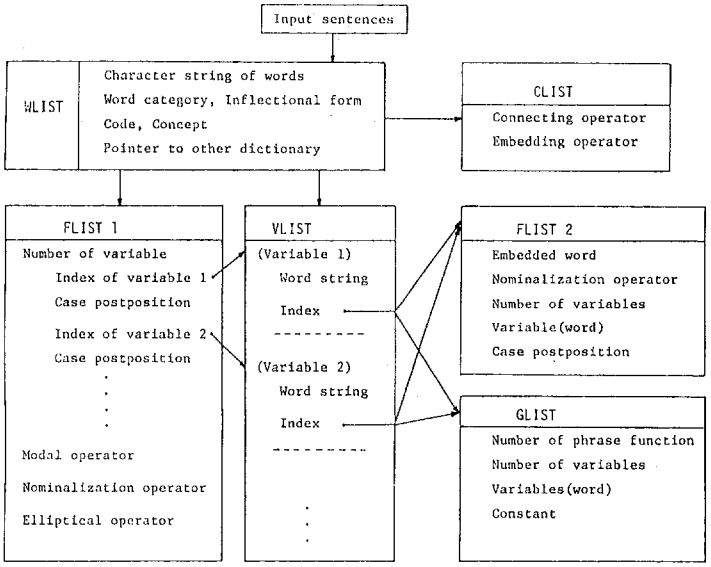

Fig.2 Data flow of decomposition process

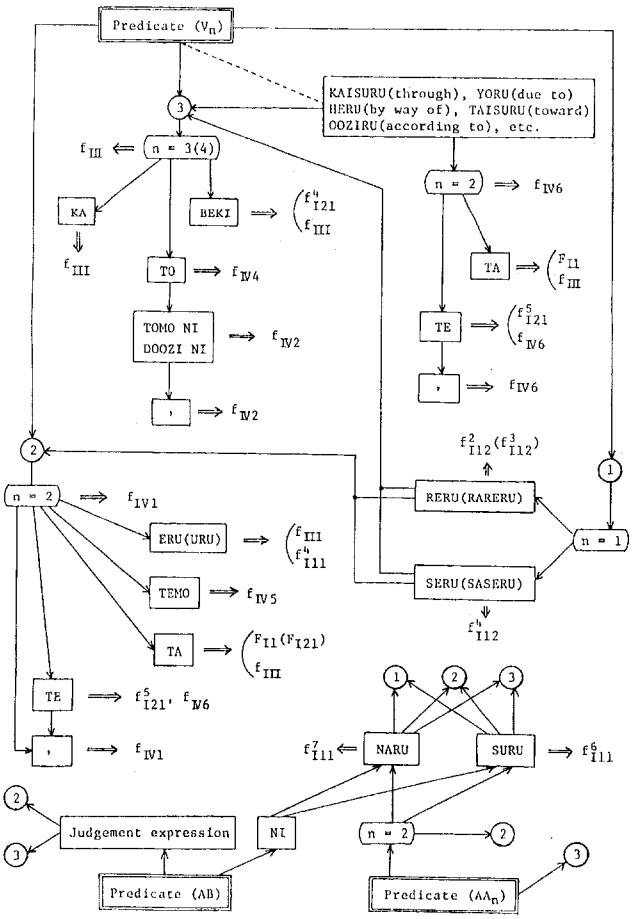

Fig.3 Extraction of modal, connecting and embedding operators

```
(1) Connecting point of resistor and inductance coil
    TEIKOO TO INDUCTANCE NO SETUZOKU TEN
  MAIN ELEMENT = TEN
  PREDICATE FUNCTION
    1 (INDUCTANCE) WO (TEIKOO) TO SETUZOKU-SURU
      N.OP. = F2.30      NOUN = TEN

(2) Output pulse with constant amplitude
    ITTEI SINPUKU SYUTURYOKU PULSE
  MAIN ELEMENT = PULSE
  PHRASE FUNCTION
    1 G3.08 : PULSE--CC3--SINPUKU
  PREDICATE FUNCTION
    1 (SINPUKU) GA ITTEI-DA
      N.OP. = F2.11      NOUN = SINPUKU
    2 (PULSE) GA SYUTURYOKU-SURU
      N.OP. = F2.12      NOUN = PULSE

(3) Voltage detecting device with high input impedance
    KOO NYUURYOKU-IMPEDANCE DENATU KENSYUTU KAIRO
  MAIN ELEMENT = KAIRO
  PHRASE FUNCTION
    1 G3.08 : KAIRO--CC3--NYUURYOKU-IMPEDANCE
  PREDICATE FUNCTION
    1 (NYUURYOKU-IMPEDANCE) GA TAKAI
      N.OP. = F2.11      NOUN = NYUURYOKU-IMPEDANCE
    2 (KAIRO) GA (DENATU) WO KENSYUTU-SURU
      N.OP. = F2.12      NOUN = KAIRO
```

Fig.4 Examples of phrase processing

is the upper concept of first word or not.

Experiments

The merit of above procedure is the combination of top-down processing and bottom-up processing. The formar finds a key word in sentences without reference to the word order. The latter analyses word string based on the key word. This is advantageous for the processing of Japanese sentences in which the word order variation and the embedding appear frequently.
The procedure was programmed by the assembly language of TOSBAC-40C mini computer. The experimental results for sentences in 30 documents confirmed the adequacy of our procedure. The examples of phrases and sentences processing are shown in Fig.4-5.

Conclusion

This paper have presented the method of decomposing Japanese sentences into normal forms. This method has following desirable advantages:
(1) The descriptive scheme M which describes the word content and the relation among words, is introduced based on the human linguistic process. This will be useful for language processing including the pragmatics in the future.
(2) The normal forms which consist of the basic predicate function, phrase function and operator, are interpreted according to the descriptive scheme M. This is useful for the semantic processing of input sentences.
(3) The structure of considerably long sentences can be described by the embedding and connecting operators.
(4) The structural description of phrases or compound words is useful to reduce the amount of storage for word dictionary.
(5) The normal forms of sentences can serve as input data for an automatic subject indexing or abstracting of documents in the information retrieval system[5,6].

The problems left unsolved are word segmentation of input Japanese sentences, detection of syntactic and semantic ambiguity, and semantic

```
***INPUT SENTENCE***
1 TUI NO PHOTO-TRANSISTOR NO COLLECTOR KAN NI TAGAI NI GYAKU-HEIRETU NO HAKKOO-D
IODE WO SETUZOKU SI TE , DOOTUU SURU BEKI PHOTO-TRANSISTOR WO HOZI SURU YOO NI H
IKARI-KETUGOO SASE , PHOTO-TRANSISTOR NO BASE NI KOOGO NI KUWAE RARERU HIKARI SI
NGOO NI YORI HANTEN SURU FF .  (A flip-flop in which light emitting diodes
connected in antiparallel are tied across collectors of a pair of photo
transistors; and in which they are photo coupled so as to keep the state of the
photo transistor to conduct; and which turns by alternately applying photo
signal to the base of photo transistor.)

***STRUCTURAL DESCRIPTION***
S1 : (X1.1) WO (X1.2) NI SETUZOKU-SURU
        OP. = F1.215(TA)
      X1.1 = TAGAI NI GYAKU-HEIRETU NO HAKKOO-DIODE
       MAIN ELEMENT = HAKKOO-DIODE
       PREDICATE FUNCTION
         1 (HAKKOO-DIODE) GA (TAGAI) NI GYAKU-HEIRETU-DA
           N.OP. = F2.110(NOUN = HAKKOO-DIODE)
      X1.2 = 1 TUI NO PHOTO-TRANSISTOR NO COLLECTOR KAN
       MAIN ELEMENT = KAN
       PHRASE FUNCTION
         1 G3.03 : 2--CC2--PHOTO-TRANSISTOR
         2 G1.01 : COLLECTOR/COLLECTOR*R12*POSITION
         3 G1.05 : PHOTO-TRANSISTOR*R51*COLLECTOR
S2 : (X2.1) GA DOOTUU-SURU
        OP. = F1.214(BEKI), F2.110(NOUN = X2.1)
      X2.1 = PHOTO-TRANSISTOR
S3 : (X3.1) WO HOZI-SURU
        OP. = F2.300(NOUN = YOO)
      X3.1 = X2.1
S4 : (X4.1) GA (X4.2) TO (X4.3) NI HIKARI-KETUGOO-SURU
        OP. = F1.124(SASERU), F5.100(PHOTO-TRANSISTOR), F5.100(HAKKOO-DIODE)
      X4.1 = PHOTO-TRANSISTOR
      X4.2 = HAKKOO-DIODE
      X4.3 = YOO
S5 : (X5.1) WO (X5.2) NI (X5.3) NI KUWAERU
        OP. =F1.123(RARERU), F2.110(NOUN = X5.1)
      X5.1 = HIKARI SINGOO
       MAIN ELEMENT = SINGOO
       PHRASE FUNCTION
         1 G2.03 : A09--HIKARI/SINGOO
      X5.2 = PHOTO-TRANSISTOR NO BASE
       MAIN ELEMENT = BASE
       PHRASE FUNCTION
         1 G1.05 : PHOTO-TRANSISTOR*R51*BASE
      X5.3 = KOOGO
S6 : (X6.1) NI YORU
      X6.1 = X5.1
S7 : (X7.1) GA HANTEN-SURU
        OP. = F2.110(NOUN = X7.1)
      X7.1 = FF

***CONNECTING FORMULA***
S1*F4.1*S4(F3-S3(F3-S2))*F4.1*[S6(F3-S5)*F4.6*S7]

N.B. 1 S1, S2, ... are basic predicate functions.
     2 X1.1, X1.2, X2.1, ... are variables of each functions.
     3 The symbol "OP." indicates the operator applied to the predicate.
     4 The symbol "NOUN" indicates the embedded phrase or word.
     5 The predicate function "S7" is as embedded one, but it is considered to
       be the independent function in connecting formula.
```

Fig.5 Example of sentences processing

description of sentences.

References

1. T.Fujita, H.Tsurumaru and S.Yoshida,"Machine Processing of Japanese—Decomposition of Japanese Sentences into Their Normal Forms—," Trans. IECE Japan, Vol.58-D, No.7, pp.405-412, July 1975.
2. F.Nishida and S.Takamatsu, "A Reduction of Restricted Japanese Sentences to Predicate Formulas and the Information-Extraction", Trans. IECE Japan, Vol.J59-D, No.8, pp.515-522, Aug. 1976.
3. T.Endo and T.Tamati, "Syntax Analysis of Japanese Text for Subject Indexing", Tech. Report of IECE Japan, AL77-46, Oct. 1977.
4. T.Endo and T,Tamati, "On a Structural Description of Japanese Text", Tech. Report of IECE Japan, AL79-37, July 1979.
5. G.Salton, "The SMART RETRIEVAL SYSTEM—Experiments in Automatic Document Processing—," Prentice-Hall Inc. 1971.
6. F.W.Lancaster, "Vocabulary Control for Information Retrieval", Information Resource Press, 1972.