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.

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 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 $\sim$ is the words which are called stem of adjectival verb in the traditional Japanese grammar. For inflectional words such as $A\alpha_n$, $V_n$, $T\beta_n$ and $J\gamma_n$, 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;

$$M = < O, E, U, \sim, \Lambda >$$

(1) $O = \{ s_u, a_u, r_u \}$, where $s_u$ is the substance, $a_u$ is the attribute, and $r_u$ is the relation. The symbol $\sim_i$ specifies the variety and the abstracting level of each unit. Thus, $O$ 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.

(2) $E = \{ c_1, c_2, c_3 \}$. $E$ describes the relationship between objects from the various viewpoint. $c_1$ is the relationship between substance and attribute, $c_2$ is the relationship between substance and relation, and $c_3$ is the
Definition of predicate function

We define the normal forms of Japanese sentences, so 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 constans 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

$$X_1{a_1}X_2{a_2}...X_n{a_n}P$$

where $X_1$, $a_1$ and $P$ indicate the noun phrase, the case postposition and the predicate respectively.

[Example]
1. (SOOTI) GA (ZIZUKUSA) WO (PULSE) NI KAERU.
   $X_1{a_1} = \text{device}$
   $X_2{a_2} = \text{converts continuous wave into pulse train.}$

2. (DENATU) WO (TEIKOOKI) NI KUWAETA.
   $X_1{a_1} = \text{someone}$
   $X_2{a_2} = \text{applied voltage across a resistor.}$

3. (DENRYOKU HENKA) GA TISSAI.
   (The variation in power is small.)

Table 1 Classification of words

| Category | Symbol | Example |
|----------|--------|---------|
| Objective expression ($X_1$) | | |
| Noun | Common noun | NA transitor |
| | Attribute noun | Dynamic, Nominal |
| | Dynamic | HABA(SIN) (oscillation) |
| | Static | NOD(SIN) (sinusoidal) |
| | Nominal | HABANANA(SIN) (oscillation) |
| Pronoun | HABANANA(SIN) (oscillation) |
| Adjective | Adj ective | AD (large) |
| | Adjective II | AB (tidy) |
| | Abstract | VDN (voltage) |
| | Special | VDN (voltage) |
| Uninflected noun modifier | RK | ARHI (current) |
| Prefix | NO | HIN (noon) |
| Suffix | Nominal suffix | TA | KAI (the) |
| | Verbal suffix | TAN | SAFER (make) |
| Special symbols | | |
| Compound word | RE | KURURUKO |
| Auxiliary verb | TT | YO(1) |
| Post-position | Case postposition | XA | GA, MD, N1 |
| | Dependent postposition | XD | DA(N1), NAD |
| | Adverbial postposition | XD | TD(TO), DON |
| | Conjunctive postposition | XD | DON |
| Assertive adverb | CC | OPORI (and) |
| Punctuation marks | | |

Table 2 List of descriptor O

| Symbol | Descriptor | Example |
|--------|------------|---------|
| A | Substance | NITTO 
| B | Potentiul substance | SMART 
| C | Substance object | DAIU 
| D | Functional body | AHAI, DONAP 
| E | Circuit element | HANAMAI, DIAD, INO 
| F | Circuit device | HANASEM(ionizer), KIBANAI, KOMASU |
| G | Device of system | KISANAI, KOMADONAI |
| H | Material | KOMANAI, KOMADONAI |
| I | Particle | DONAI, DONAI, DONAI |
| J | Abstract group | NITEKI |
| K | Phononym | XUAI |
| L | Paradigm | XUAI |
| M | Word | XUAI |
| N | Phoneme | XUAI |
| O | Abstract phoneme | XUAI |
| P | Functional phoneme | XUAI |
| Q | Other phonemes | XUAI |
| R | Abstract function | XUAI |
| S | Ideal substance | XUAI |
| T | View position | XUAI |
| U | Characteristic | XUAI |
| V | Component | XUAI |
| W | Quantity | XUAI |
| X | Arguement | XUAI |
| Y | Argument | XUAI |
| Z | State | XUAI |
| AA | Degree | XUAI |
| AB | Value | XUAI |
| AC | Constant | XUAI |
| AD | Method | XUAI |
| AE | Process | XUAI |
| AF | Function | XUAI |
| AG | Place | XUAI |
| AH | Process | XUAI |
| AI | Format | XUAI |
| AJ | Event | XUAI |
| AK | Attribute | XUAI |
| AL | Possibility | XUAI |
| AM | Possibility | XUAI |
| AN | Substance | XUAI |
| AO | Object | XUAI |
| AP | Operation | XUAI |
| AQ | Concrete operator | XUAI |
| AR | Event | XUAI |
| AS | Change | XUAI |
| AT | Change | XUAI |
| AU | Condition | XUAI |
| AV | Effect of change | XUAI |
| AW | Source | XUAI |
| AX | Place of removal | XUAI |
| AY | Input/output | XUAI |
| AZ | Input/output | XUAI |
| BA | Output | XUAI |
| BB | Continuation | XUAI |
| BC | Movement | XUAI |
| BD | Movement | XUAI |
| BE | Movement | XUAI |
| BF | Movement | XUAI |
| BG | Movement | XUAI |
| BH | Movement | XUAI |
| BI | Movement | XUAI |
| BJ | Movement | XUAI |
| BK | Movement | XUAI |
| BL | Movement | XUAI |
| BM | Movement | XUAI |
| BN | Movement | XUAI |
| BO | Movement | XUAI |
| BP | Movement | XUAI |
| BQ | Movement | XUAI |
| BR | Movement | XUAI |
| BS | Movement | XUAI |
| BT | Movement | XUAI |
| BU | Movement | XUAI |
| BV | Movement | XUAI |
| BW | Movement | XUAI |
| BX | Movement | XUAI |
| BY | Movement | XUAI |
| BZ | Movement | XUAI |
| CA | Attribute | XUAI |
| CB | Dynamic attribute | XUAI |
| CC | Dynamic attribute | XUAI |
| CD | Dynamic attribute | XUAI |
| CE | Dynamic attribute | XUAI |
| CF | Dynamic attribute | XUAI |
| CG | Dynamic attribute | XUAI |
| CH | Dynamic attribute | XUAI |
| CI | Dynamic attribute | XUAI |
| CJ | Dynamic attribute | XUAI |
| CK | Dynamic attribute | XUAI |
| CL | Dynamic attribute | XUAI |
| CM | Dynamic attribute | XUAI |
| CN | Dynamic attribute | XUAI |
| CO | Dynamic attribute | XUAI |
| CP | Dynamic attribute | XUAI |
| CQ | Dynamic attribute | XUAI |
| CR | Dynamic attribute | XUAI |
| CS | Dynamic attribute | XUAI |
| CT | Dynamic attribute | XUAI |
| CU | Dynamic attribute | XUAI |
| CV | Dynamic attribute | XUAI |
| CW | Dynamic attribute | XUAI |
| CX | Dynamic attribute | XUAI |
| CY | Dynamic attribute | XUAI |
| CZ | Dynamic attribute | XUAI |
| DA | Relational attribute | XUAI |
| DB | Relational attribute | XUAI |
| DC | Relational attribute | XUAI |
| DD | Relational attribute | XUAI |
| DE | Relational attribute | XUAI |
| DF | Relational attribute | XUAI |
| DG | Relational attribute | XUAI |
| DH | Relational attribute | XUAI |
| DI | Relational attribute | XUAI |
| DJ | Relational attribute | XUAI |
| DK | Relational attribute | XUAI |
| DL | Relational attribute | XUAI |
| DM | Relational attribute | XUAI |
| DN | Relational attribute | XUAI |
| DO | Relational attribute | XUAI |
| DP | Relational attribute | XUAI |
| DQ | Relational attribute | XUAI |
| DR | Relational attribute | XUAI |
| DS | Relational attribute | XUAI |
| DT | Relational attribute | XUAI |
| DU | Relational attribute | XUAI |
| DV | Relational attribute | XUAI |
| DW | Relational attribute | XUAI |
| DX | Relational attribute | XUAI |
| DY | Relational attribute | XUAI |
| DZ | Relational attribute | XUAI |

Table 3 List of descriptor O

| Symbol | Descriptor | Example |
|--------|------------|---------|
| A | Substance | NITTO |
| B | Substance | NITTO |
| C | Substance | NITTO |
| D | Substance | NITTO |
| E | Substance | NITTO |
| F | Substance | NITTO |
| G | Substance | NITTO |
| H | Substance | NITTO |
| I | Substance | NITTO |
| J | Substance | NITTO |
| K | Substance | NITTO |
| L | Substance | NITTO |
| M | Substance | NITTO |
| N | Substance | NITTO |
| O | Substance | NITTO |
| P | Substance | NITTO |
| Q | Substance | NITTO |
| R | Substance | NITTO |
| S | Substance | NITTO |
| T | Substance | NITTO |
| U | Substance | NITTO |
| V | Substance | NITTO |
| W | Substance | NITTO |
| X | Substance | NITTO |
| Y | Substance | NITTO |
| Z | Substance | NITTO |

---

Various connection of the same kind of objects.

3. $U$ represents the active cognitions which are relatively independent of concepts.

4. $V$ specifies the cognitive behaviors how the speaker cognize the objects.

5. $\Lambda = (\iota_3(\text{tense}), \nu_3(\text{anaphora}))$. $A$ represents the relation between a speaker and objects.

A part of $O$, $Z$, $U$ and $V$ is tabulated in Table 2-5 respectively.
However, a predicate P has a variety of expressions in Japanese. For example, a verb is frequently connected with some auxiliary verbs (e.g., NA (negative), TA (past)) or verbal suffixes (e.g., RAREHU (passive), SASERU (causative)). Therefore, we decompose the predicate P into objective expression P_0 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_0P_s:

1. P_0 (Final form of verb) P_s (Zero element of speaker's judgement)
2. P_0 (Final form of adjective I) P_s (Zero element of speaker's judgement)
3. P_0 (Adjective II) P_s (Judgement expression "DA(be)"
4. P_0 (Noun) P_s (Judgement expression "DA(be)"

The application of operators presented in next section, inflects the form of P_0 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 (ideographic expression)
- Basic predicate function (simple connection between substance and attribute)
- Derivative function (expand the form of basic predicate)

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 the attribute. The variables corresponding to these substances are called obligatory variables of the predicate, and the case postpositions, obligatory ones aei. On the other hand, one substance usually accompanies various kinds of attribute, and is related to other substances as a mediation of attributes. In the predicate function, the variables corresponding to such attributes and substances are called facultative variables, and the case postpositions, facultative ones agi.

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 Σ. Also, the relationship between the predicate and each variable is given by the descriptor Σ. These are summarized in Table 6.

**Table 3 List of descriptor Σ**

| Symbol | Descriptor | Example |
|--------|------------|---------|
| 01.1   | Simple connection | (IMPULSE) CA TAKAI (The impulse is MAX.) |
| 01.2   | Object of action or operation | (SINDO) KO RISSHIDORU-SU (A thing amplified signal.) |
| 01.3   | Starting point in action | (TRANSITOR) KARA MAN (A thing consists of a transistor.) |
| 01.4   | Finishing point in action | (SAYUTTARI) NI TASSERU (A thing reaches minimum value.) |
| 01.5   | Opponent in mutual action | (NO) JE KITUROI-SU (A thing depends on noise.) |
| 01.6   | Standard or reference | (POLUB) DE MAster-SU (A thing depends on pulse.) |
| 01.7   | Way or means | (KASEKU) IN KU (A thing in between collectors.) |
| 01.8   | Spatial positioning | (OTEKI) DI MAGAI (A thing floats whatever...) |
| 01.9   | Temporal positioning | (YO(U)) NI KITUROI-SU (A person connects a thing as it is...) |
| 01.10  | Others | |

| Symbol | Substance with the relation | Example |
|--------|---------------------------|---------|
| 02.1   | Order of substance | DA (1) (TRANSISTOR) (second transistor) |
| 02.2   | Number of substance | (2) KO NO (TRANSISTOR) (two transistors) |
| 02.3   | Property of substance | (MUKI) SI NAI-SU (frequency of pulse train) |
| 02.4   | Material of substance | (HEIRETU) (SETUZOKU) (wavelength of 400 nm.) |
| 02.5   | Unit of substance | (YO0) NI KETUGOO-SU (A thing depends on noise.) |
| 02.6   | Various connection among attributes | |

**Table 4 List of descriptor U**

| Symbol | Describer | Example of subjective expression |
|--------|-----------|---------------------------------|
| 01     | Affirmative judgment | DA, ARU (be, do) |
| 02     | Negative judgment | NA, NI (not) |
| 03     | Universal judgment | VA, ND |
| 04     | Purpose or aim | |
| 05     | Will | EI, YOO (will, shall) |
| 06     | Assumption | DA, SOSI (if) |
| 07     | Certification | VA |
| 08     | Inference | U, TANIN (probably) |
| 09     | Desire | |
| 10     | Natural judgment | HUREI (should, have to) |
| 11     | Instance | HAN (should, have to) |
| 12     | Limitation | (toward ideal) |

**Table 5 List of descriptor V**

| Symbol | Describer | Cognitive Behavior |
|--------|-----------|--------------------|
| 48     | Cognizing object | O faithfully |
| 49     | Cognizing attribute | A in substance |
| 50     | Cognizing static attribute | Dynamically |
| 51     | Cognizing dynamic attribute | Statically |
| 52     | Cognizing causal relation | Events backward |
| 53     | Cognizing an objective object | Ideal |
| 54     | Cognizing an object | Ideal |
| 55     | Cognizing one object | Ideationally |
| 56     | Cognizing the degree of | Quantity |
| 57     | Cognizing one object | From the various view points |
| 58     | Conjunctive enumeration | |
| 59     | Disjunctive enumeration | |

-494-
cate, and varies the mode of the attribute which is expressed by the function. On the other hand, $F_{12}$ applies to $P_3$, and varies the mode of the judgement. An example of $F_{11}$ and $F_{12}$ 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) $F_{11}$: Cognizing one of the objects expressed by the predicate function, as the substance with attribute.

$$F_{11} \text{ (object) } \text{ predicate function}$$

(2) $F_{12}$: Recognizing the concrete event expressed by the predicate function, as substance ideally.

$$F_{12} \text{ (event) } \text{ predicate function}$$

(3) $F_{13}$: Transforming the predicate function into clauses which express the time, reason, state, effect and so on.

$$F_{13} \text{ (clause) } \text{ predicate function}$$

Table 6 Example of basic predicate function

| Predicate P | Variable | Case position | $\lambda$ | Domain |
|-------------|----------|---------------|-----------|---------|
| KUSEI (give) | $x_1$ | $y_1$ | $w_1$ | $x_2$ |
| KUSEI-SURI (detect) | $x_3$ | $y_1$ | $w_1$ | $x_4$ |
| KUSEI (received) | $x_5$ | $y_1$ | $w_1$ | $x_6$ |
| KUSEI-SURI (connect) | $x_7$ | $y_1$ | $w_1$ | $x_8$ |
| TANGU (keep) | $x_9$ | $y_1$ | $w_1$ | $x_{10}$ |
| DOMASHI-SURU (conduct) | $x_{11}$ | $y_1$ | $w_1$ | $x_{12}$ |
| ITTEI (constant) | $x_{13}$ | $y_1$ | $w_1$ | $x_{14}$ |
| OKU (large) | $x_{15}$ | $y_1$ | $w_1$ | $x_{16}$ |
| TATE (strong) | $x_{17}$ | $y_1$ | $w_1$ | $x_{18}$ |

Table 7 Example of modal operators $F_{11}$

| Symbol | Operator | Content | Usage |
|--------|----------|---------|-------|
| $F_{11}$ | SURI (make) | $a_1$ | IMPROVING THE TIME  (The time is high.) |
| $F_{11}$ | NARI (become) | $a_2$ | IMPROVING THE TIME  (The time becomes high.) |
| $F_{11}$ | RAHEI (be able to) | $a_3$ | IMPROVING THE TIME  (A thing is able to increase the time.) |
| $F_{11}$ | SHIHI (passive) | $a_4$ | IMPROVING THE TIME  (A thing is able to increase the time.) |

Table 8 Example of modal operators $F_{12}$

| Symbol | Operator | Content | Usage |
|--------|----------|---------|-------|
| $F_{12}$ | DA, AB (be) | $u_1$ | SWITCHING (ON/OFF)  (The switching operation is correct.) |
| $F_{12}$ | MATU (be) | $u_2$ | IMPROVING THE TIME  (The switching operation is correct.) |
| $F_{12}$ | U, YOSHI (do) | $u_3$ | IMPROVING THE TIME  (The switching operation is correct.) |
| $F_{12}$ | RENJU (be) | $u_4$ | IMPROVING THE TIME  (The switching operation is correct.) |
| $F_{12}$ | TA (be) | $u_5$ | IMPROVING THE TIME  (The switching operation is correct.) |
| $F_{12}$ | MADO (be) | $u_6$ | IMPROVING THE TIME  (The switching operation is correct.) |
| $F_{12}$ | DAKE (be) | $u_7$ | IMPROVING THE TIME  (The switching operation is correct.) |
| $F_{12}$ | TATE (be) | $u_8$ | IMPROVING THE TIME  (The switching operation is correct.) |
(NYUURYOKU SINGO) WO HENTYO-SURU
(A thing modulates the input signal.)
→ (NYUURYOKU SINGO) WO HENTYO-SI TA SINGO
(the signal which is modulated by the input signal.)

(4) fII4 : Cognizing the only attribute as substance.
(PULSE) WO HASSTIN-SI WA SURU
(A thing generates the pulse train.)
÷ (NYUURYOKU SINGO) WO HENTYO-SI TA SINGO
(the signal which is modulated by the input signal.)
(ONDOKA) 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 fIII.

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(fIV1)
S1 : (SYOOGI DENRYOKU) GA TISAI
(The consumption power is small.)
S2 : (SWITCHING ZIKAN) GA MIZIKAI
(The switching time is short.)
÷ S1*fIV1*S2
(SYOOGI DENRYOKU) GA TISAI, (SWITCHING ZIKAN) GA MIZIKAI
(The consumption power is small, and the switching time is short.)

(2) Simultaneous conjunctive connecting operator(fIV2)
S1 : (TRANSISTOR) WO KUDOO-SURU
(A thing drives the transistor.)
S2 : (HOOWADO) WO SEIGYO-SURU
(A thing controls the saturation rate.)
÷ S1*fIV2*S2
(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(fIV3)
S1 : (CONDENSER) WO SETUZOKU-SURU
(A person connects a capacitor.)
S2 : (COIL) WO IKERU
(A person inserts a coil.)
÷ S1*fIV3*S2
(CONDENSER) WO SETUZOKU-SURU KA (COIL) WO IKERU
(A person connects a capacitor, or inserts a coil.)

(4) Causal connecting operator(fIV4)
S1 : (DENRYU) GA (SYOTEITI) WO KOSU
(The current exceeds the fixed value.)
S2 : (DENATU HENKA) GA SYOOZOKU
(The voltage changes.)
÷ S1*fIV4*S2
(DENRYU) GA (SYOTEITI) WO KOSU TO (DENATU HENKA) GA SYOOZOKU
(The voltage changes when the current exceeds the fixed value.)

(5) Concessive connecting operator(fIV5)
S1 : (SYUUKI) WO KAIERU
(A person changes the period.)
S2 : (SINPUKU) GA ITTEI-DA
(The amplitude is constant.)
÷ S1*fIV5*S2
(SYUUKI) WO KAIERU TEMO (SINPUKU) GA ITTEI-DA
(When a person changes the period, the amplitude is constant.)

(6) Modificatory operator(fIV6)
S1 : (TEIKOO) WO KAIERU
(Through the resistor)
S2 : (BASE) WO (DENGEN) NI SETUZOKU-SURU
(A person connects the base to the power source.)
÷ S1*fIV6*S2
(TEIKOO) WO KAIERU 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 fII and fIII be the nominalization and the embedding operator respectively. An arbitrary predicate function Ai is expressed by
Ai = Ai1*fIV1*Ai2*flV1*...*fIV1*Ai m
where Ai k is
(i) Su,
(~) [Ai*fIVd*Aj] (d = 2,3,4,5,6).
Su 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 Su in the following way.
Su(fIII-A1, A2, ..., Ai, ..., An)
where Ai = fIIAi.

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.

G1 is the phrase connected in terms of such relational concepts as position(r1), reference (r2), and part(r3). G2 is the phrase formed by cognitive behaviors(f), such as enumeration(φ10, φ11), cognition of one object from the various viewpoint(φ9), concrete and abstract cognition of one object(φ9), and so on. G3 is the phrase constructed in terms of the relationship(φ) between substance and attribute, and the various
connection($G_o$) of the same kind of objects. $G_o$ 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.

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 word 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 9 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 ($P_{II}$, $P_{I2}$ and $P_{I2}$), embedding operator $P_{II}$ and connecting operator $P_{I}$ 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 WLIST1 and CLIST1. The variables of the predicate function are extracted by reference to PFD. At the same time, the modal operators $P_{I2}$ and $P_{I2}$ 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 |
|--------|----------------|----------|---------|
| $P_{I1}$ | $V_{I1}$ | $V_{I1}$ | $V_{I1}$ |
| $P_{I2}$ | $V_{I2}$ | $V_{I2}$ | $V_{I2}$ |
| $P_{I3}$ | $V_{I3}$ | $V_{I3}$ | $V_{I3}$ |

### Table 10 Structure of machine dictionary

#### (a) Word dictionary (WD)

| Entry word | Category | Code | Concept | Pointer |
|------------|----------|------|---------|---------|
| TRANSISTOR | NA | 300 | $V_{I1}$ | -1 |
| GE(-) | NA | 1 | - | - |
| SETSUKI | VB | 1010 | $V_{I1}$ | -1 |
| COLLECTOR | NA | 410 | $V_{I2}$ | -2 |
| DENTYOO | NA | 376 | $V_{I2}$ | -2 |
| SIDEN | VB | 36 | $V_{I3}$ | -3 |
| HEN~% | VB | 1025 | $V_{I3}$ | -3 |
| NOITAI | NA | 343 | $V_{I3}$ | -3 |
| DOU | VB | 1018 | $V_{I3}$ | -3 |
| BAI | NA | 410 | $V_{I3}$ | -3 |
| KONI | VB | 1018 | $V_{I3}$ | -3 |
| NOITAI | NA | 343 | $V_{I3}$ | -3 |
| KONI | VB | 1018 | $V_{I3}$ | -3 |

#### (b) Predicate function dictionary (PFD)

| No. | Number of case | Case postposition | Number of domain | Domain | Character setting of predicate |
|-----|----------------|------------------|-----------------|--------|-------------------------------|
| 1   | 1              | 0                | 0               | 0      | $V_{I1}$, $V_{I2}$, $V_{I3}$ |
| 2   | 2              | 0                | 2               | 2      | $V_{I1}$, $V_{I2}$, $V_{I3}$ |
| 3   | 3              | 0                | 3               | 3      | $V_{I1}$, $V_{I2}$, $V_{I3}$ |

#### (c) Related concept dictionary (RCD)

| No. | Number | Variety | Direction | Level | Related concept |
|-----|--------|---------|-----------|-------|-----------------|
| 1   | 1      | 0       | +         | 0     | $V_{I1}$, $V_{I2}$, $V_{I3}$ |
| 2   | 2      | 0       | +         | 0     | $V_{I1}$, $V_{I2}$, $V_{I3}$ |
| 3   | 3      | 0       | +         | 0     | $V_{I1}$, $V_{I2}$, $V_{I3}$ |

* $O$ (obligatory variable), $I$ (optional variable), $I$ (special variable due to $P_{I1}$ and $P_{I2}$). $I$ (special variable due to $P_{I2}$).
decided. The extracted information is stored FLIST I, 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)", "TSI(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 8101, 8201, 8301, and so on, in Table 9. Their constants, such as "NYOOG(both)", "KAN (between)", "TAHOO(another)", "DAI", "GO", 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 8105, 8308, and so on.

[Type III] Phrase functions extracted by using the variety or level of word concept. For example, 8203 is extracted by investigating whether the upper concepts of both words agree with each other or not, and 8204 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

---
**Experiments**

The merit of above procedure is the combination of top-down processing and bottom-up processing. The former finds a key word in sentences without reference to the word order. The latter analyses word string based on the word order. This is advantageous for the SPC processing of Japanese sentences which embody the order variation and the embedding appear frequently.

The procedure was programmed by the assembly language of TOSBAC-60C 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.

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. 

The problems left unsolved are word segmentation of input Japanese sentences, detection of syntactic and semantic ambiguity, and semantic reference to the word order. The latter analyses word string based on the word order. This is advantageous for the SPC processing of Japanese sentences which embody the order variation and the embedding appear frequently.

The procedure was programmed by the assembly language of TOSBAC-60C 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.

**Reference**

1. T.Pujita, H.Tsurumaru and S.Yoshida,"Machine Processing of Japanese—Decomposition of Japanese Sentences into Their Normal Forms—", Trans. IECE Japan, Vol.59-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.59-D, No.8, pp.515-522, Aug. 1975.

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. P.W.Lancaster, "Vocabulary Control for Information Retrieval", Information Resource Press, 1972.