

## 専門分野を対象とした日英機械翻訳について†

首 藤 公 昭†

### Abstract

Difficulties in investigating a Machine Translation (MT) system or a more general machine processing system of natural languages are caused by various ambiguities contained in natural languages.

We are developing a MT system from Japanese into English in which ambiguities based on manifold meanings of words can be reduced to some extent, restricting the field that the system deals with to a technical one (especially to the field of transistor circuits).

The main feature of our MT process is to grasp relations between words through the medium of **case** and some **cause-and-effect** relations between predicates in the input sentence and to construct the output sentence on the bases of them. In this paper, the basic strategy of the semantic processing and informations required for it are shown along with examples.

### 1. まえがき

電子計算機の誕生いらい機械翻訳 (Machine Translation; MT) の研究が各国で行なわれてきた。しかしこれらの試みには構文論的あるいは形態論的な処理を中心とするものが多く、実用上満足な結果を得るには至らなかった。自然言語は豊かな表現力を持っているがあまりに複雑多岐であり、種々のあいまいさのため、機械処理の対象とするには困難な点が多い。しかしながら近年、格 (case) を媒介とした意味の取扱い<sup>3)</sup> や述語論理の真偽決定問題と関連した人工知能からの接近<sup>4)</sup> 等が論じられるに至り、各国において自然言語処理系が考察の対象となりつつある。われわれは人間の学習過程と関連して、思考、言語の領域を一般分野と専門分野とに分け、分野間のリンクを考慮する方法で自然言語処理系および人工知能の研究を行なっている<sup>5), 6), 10)</sup>。また、これらの研究における実験の第一段階としてトランジスタ回路に関する分野における書物、文献を対象とした日英 MT システムの開発を進めている。本システムは体言、用言の意味を大まかに定義して設定された概念系の近似モデルを利用した意味処理を中心とするものであり、現在基本方針を決定し実験の段階に入りつつある。一般に語の意味は見方

により多くの側面を持つが、専門分野では語の意味の側面は分野固有の見方により限定され、意味の取扱いが比較的容易となる。文の深層構造と表層構造とを明確に分離するのは困難であるが、大まかに言えば本システムの特徴は入力文より深層構造を抽出し、これに基づいて出力文を合成する過程にある(図1)。これは入力文の表層構造を出力文の表層構造に直接変換する従来の方法とは本質的に異なった翻訳過程である。本論文では本システムでの意味処理のための各種の情報と意味処理を中心とした翻訳過程の概略を述べる。

本文で述べない日本語に関する基礎的事項、構文分析については文献 7), 10) 等を、文法処理と意味処理とのつなぎの処理については文献 8) 等を参照されたい。ただしこれらの処理は専門分野を対象とする本システムの場合には簡単化することができるることを付記しておく。本文で用いる記号の意味は次の通りである。

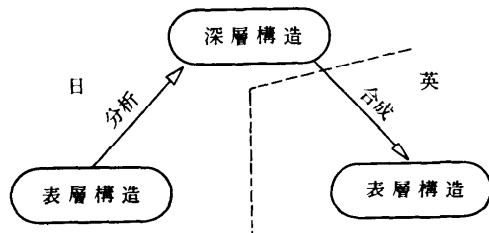


Fig 1 Fundamental notion of processing

† On Machine Translation from Japanese into English for a technical field, by Kimiaki SHUDO (Fukuoka University, Faculty of Engineering)

†† 福岡大学工学部電子工学科

- N: 体言を自立部とする日本語文での文節。  
 P: 用言または体言に‘である’の接続したものを自立部とする日本語文での文節。  
 E(x): 日本語文での表現 x に対応する英文での表現。

## 2. 語類の設定

日本語文における文節間の関係<sup>†</sup>を意味を考慮して決定するのがわれわれの MT での分析の中心過程である。そこで見方で規定される語の意味上側面(以後意味の側面と呼ぶ)によって語を分類し、語のクラスを用いて付属語を介した語の間の関係を規定しておくことが必要である<sup>††</sup>。本システムでは格関係の決定を分析の中心と考えるので主として格を通じて他の語との関係に基づいて約 1000 語を対象に意味分類を行ない、図 2 に矩形で示すような語類<sup>†††</sup>を約 70 個設定している。分類作業で留意した点を列挙すれば次の通りである。1) 格関係があいまいさなく決定されること。2) 訳語、英語の基本構文および前置詞が意味的に決定されること。3) 連語の意味の側面がある程度把握されること。4) 格以外の依存、並列関係もある

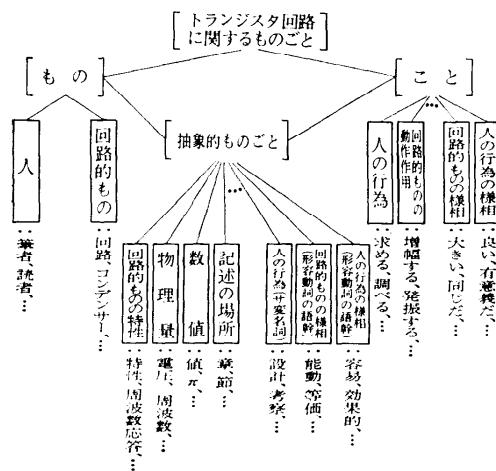


Fig. 2 Semantic word classes

<sup>†</sup> 文節間の関係は、その意味に基づいて次のように分類しているが本稿では特に依存関係の処理について述べる。

格関係  
文節間の関係  
依存関係 修飾関係  
並列関係 広義の因果関係

<sup>††</sup> すべての文節間の関係を 2 項関係として規定するのでは分析でのあいまいさを減らす事は難かしく、合成過程も繁雑となるので 3 で述べるように、或る種の格関係については 2 項 NP から 5 項 N<sub>1</sub>N<sub>2</sub>N<sub>3</sub>N<sub>4</sub>P の範囲で P の取る格のパターンとして規定している。

<sup>†††</sup> 1 つの意味の側面をもつ語の集合。

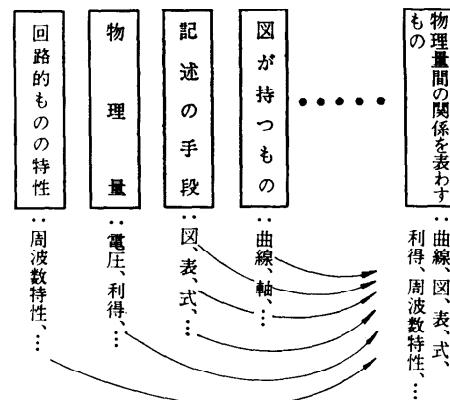


Fig. 3 Generation of new word classes

る程度決定されること。5) 品詞分類を包含すること。分類作業のある段階以降では図 3 に例示するような既設の語類から語を抽出して語類を定める過程が必要となるため、語類は互いに素ではない。また、語の追加等で意味の側面を新たに考慮する必要が生じた場合も上記の過程で語類を新設することができる。この意味でわれわれの分類は開いているが、分類の拡張と辞書、テーブル類の更新等については学習過程として考察中である。語が属する各語類を単語辞書に与えてそおくことで語の意味の各側面が定義される。訳語句はそれぞれの側面に対して与えられ、文中の他の文節との関係から語の意味の側面が決まると適切な訳語句が求まることになる。

連語を単語として登録しておくのは語数の点で困難であるため連語の意味の側面を規定する手段が必要となるが、本システムでは語類変換の規則を定めておく。これは次のような一般形の文脈自由形の 2 項書きかえ規則とする。

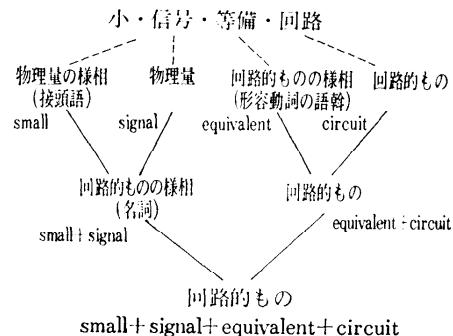


Fig. 4 An example of processing Rengo

Table 1 Word classes specified for each word

語	語類
小 信 等 回	物理量の様相(接頭語), … 物理量, … 回路的ものの様相(形容動詞語幹), … 回路的もの, …

Table 2 Rules for processing Rengo

$\alpha$	$\beta$	$r$	$t$
物理量の様相 (接頭語)	物理量	物理量, 回路的もの の様相 (名詞)	$E(\alpha') + E(\beta')$
回路的ものの様相 (形容動詞語幹)	回路的もの	回路的もの	$E(\alpha') + E(\beta')$ ただし, $E(\alpha')$ は形容詞形
回路的ものの様相 (名詞)	回路的もの	回路的もの	$E(\alpha') + E(\beta')$

$\alpha', \beta'$  は  $\alpha, \beta$  に対応した表現。

$$\alpha\beta \longrightarrow r(t).$$

ここで  $\alpha, \beta$  および  $r$  は語類,  $t$  は語順, 挿入語などの訳語句合成のための情報である。連語処理の例を図 4 に示す。ただし用いられる語類情報、変換規則および訳語句情報は表 1, 2 の通りである。

接頭語、接尾語および名詞からなる連語の処理を目的として約 50 個の変換規則と訳語句情報を定めている。

### 3. 意味的基本構文<sup>8), 10)</sup>

われわれは文を支配(govern)する P とその格をなす N とが作る意味的構造(格構造と呼ぶ)が文の深層構造の中核をなすと考える。そこで格構造の処理を主な目的として第 1, 第 2 の意味的基本構文(semantic basic structure of sentence; sbss)を設定している。これらは文献 10)に述べられているものと基本的には同一であるが、第 2 の sbss(以後 sbss-2 と書く)で代入可能性が規定される格は本システムでは次の規準(1)あるいは(2)で選ばれたものとする。(この種の格を 1 次的な格と呼ぶ。)

- (1) 英語の基本構文†を決定するために必要なもの。
- (2) 訳文で例外的な前置詞、または前置詞相当語句を要求するもの。

一般に 1 個の P に対して複数個の sbss-2 を与え、それぞれの sbss-2 に適切な訳語動詞と訳の基本構文を対応づけておけば、P の意味を規定することができ、意味を考慮した構文対構文の翻訳が可能となる。動詞

† 文献 11)に基づき約 20 種を定めている。

‘求める’に対する sbss-2 および訳の基本構文の 2 例を示す。

主体〔人〕が対象〔回路的もの、…、方法〕を…訳の基本構文: (主体) + determine + (対象)。主体〔人〕が対象〔回路的ものの特性、…、数値〕を…訳の基本構文: (主体) + calculate + (対象)。

ここで(主体)、(対象)等は日本語文での主体、対象等の格をなす語句に対応する訳語句を表わす。

P が形容詞、形容動詞または名詞に‘である’が接続したものの場合は、同一の sbss を取るものが多いので、訳の基本構文および文のタイプ(後述)が適切に指定できる範囲で、P の語類に対して sbss を与えておくのが実際的である。次に形容詞に対する sbss-2 および訳文情報の例を示す。

主体 1 [回路的もの]は主体 2 [回路的ものの特性]が P 形容詞 [回路的ものの様相]

…(主体 1) + have + (P) + (主体 2)。

sbss-2 は本システムにおける意味処理での最も重要な役割を持つものであり、現在約 400 種の P に対して約 2,000 個の sbss-2 および訳文情報を定めている。

### 4. 2 次的な格関係の処理

P が取る格†のうち sbss-2 で語の代入可能性が規定されない格には訳文では前置詞句を対応づけることができる。(この種の格を 2 次的な格と呼ぶ。) 2 次的な格については代入可能な語として同じ語類を要求する P が多いため、前置詞が適切に指定できる範囲で P をまとめ、P の語類に対して規定しておくのが実際的である。また 2 次的な格関係は 2 項関係として規定†して処理を簡単化している。表 3 に 2 次的な格関係と前置詞の規定例を示す。ここでの規定は約 50 個である。

Table 3 Examples of secondary case relations expressed by ‘で’ specified with word classes

N の語類	P の語類	格の種類	前置詞
記述の場所	人の行為	場所	in
物理量、数値、…	回路的ものの動作、様相、…	抽象的方法	at

### 5. PP 関係の処理

P と P との依存関係のうち修飾関係と見なせないものをここでは PP 関係と呼ぶ。これらの関係は広い意味での因果関係を考えることができる。P は文を支配

† 計 15 種に分類している。

†† 規定する辞書を NP 辞書と呼ぶ。

するものであるから PP 関係は文対文の関係と捉えることが必要である。本システムは文をこの分野での基本的な因果関係に基づいて分類し、34種の文のタイプを定め、これを用いて PP 関係を定義しておくことにしている†。現段階では文のタイプは P とその主体、対象である N の大まかな意味に基づいて定めているが、処理をより精密にするため、文のタイプをさらに細かく定めることや個々の PP 関係に固有の分類を設けることなどを次の段階の処理として検討中である。文のタイプと PP 関係の規定例を図 5、表 4 に示す。たとえば表 4 の◎印から、タイプ《人がものごとを求める》の文とタイプ《人がものごとを知る》の文とは助詞‘と’を介して因果関係をなし得ることがわかる。PP

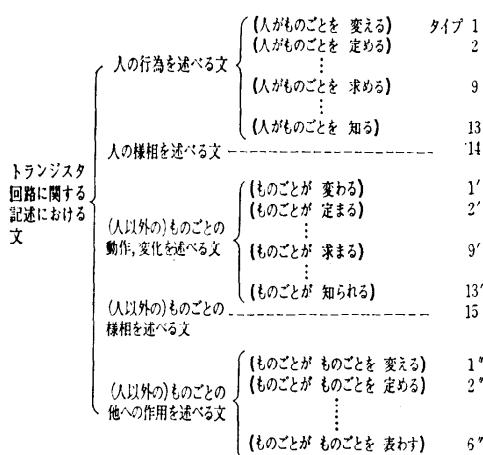


Fig. 5 Semantic sentence types

Table 4 Examples of PP relations expressed by 'と' specified with sentence type

<sup>†</sup> 規定する辞書を PP 辞書と呼ぶ。

处理

関係の規定は約 20 種の標準的な付属語に対して行なうので付属語の標準化が必要となる場合がある。たとえば、文‘周波数を 変えても 利得は 変わらない。’は‘周波数を変えると 利得が 変わる。’と変換して分析する。

文のタイプはそれぞれの sbss-2 に対して与えておくので原則として 1 次的格構造が求まると同時に文のタイプが決まるが、ある種の P の付属語<sup>†</sup>に対してはタイプの変換が必要となる。たとえば、文「電圧を上昇させる。」のタイプは標準形文「電圧が 上昇する。」のタイプ 1 を、付属語「せる」に基づいて変換して 1' または 1'' と決められる。この種の付属語には変換規則を与えている。

## 6. 修飾関係の処理

文が修飾関係を含む場合は、文を格構造のみを持つ  $N_1N_2\cdots N_nP$  形の標準形文の論理結合に分解し、格関係の処理に帰着させるのが修飾構造処理の基本方針である。NN 関係の処理方法は次の通りである。ただし文頭側の N を  $N_1$ 、文末側の N を  $N_2$  と書く。

a.  $N_2$  がサ変名詞、または用言の名詞化されたものの場合は  $N_2$  を P に変換し、 $N_1$  の付属語を手がかりに  $N_1$  が P の格となるかどうかを調べ、格となる時 NN 関係と前置詞が決まる。たとえば表現 ‘回路での減衰’ 対しては文 ‘回路で 減衰する.’ を構成して格関係を調べることで修飾関係が決まり、名詞句 ‘attenuation+in+circuit’ が求まる。 $N_1$  の約 60 個の付属語に対して格を推定するための規則を設けていく。

b. a以外の場合は次の処理を行なう。

b-1) 訳文の構成のための規則の例を表5に示す。

b-2) 語類間の2項（あるいは3項）関係として規定しておく

NN 関係処理の基本的な方法は上記の通りであるが N<sub>2</sub> がある種の複合名詞の場合など、例外的な処理が

大約 20 種不同

† 指定する辞書を NN 辞書と呼ぶ。

Table 5 Examples of rules for processing  $N_1N_2$  relations expressed by 'の', where  $N_1$  and  $N_2$  make cases of some P latently.

構成する文			訳語句	例
主体(が)	対象(を)	P		
$N_1$	$N_2$	持つ	$E(N_2) + of + E(N_1)$	回路の特性
$N_1$	$N_2$	表わす	$E(N_2) + from + E(N_1)$	図の利得

必要となる場合がある。

つぎに PN 関係では P を言い切りの形に、 N を適当な付属語の接続した形に変換して NP の格関係を調べ、 N が P の格となるとき PN 関係を決定する。訳文では主として関係詞節による修飾が対応する。

連用修飾関係には形容詞、形容動詞または副詞による P の修飾があるが、これらの関係は簡単のため語類間の 2 項関係として規定している†。ただし P の付属語と呼応関係を持つ陳述の副詞については別種の処理を考慮している。

## 7. 翻訳過程

現在定めている翻訳過程の概略と各段階で用いる辞書および規則を図 6 に示す。ただし入力日本語文は意味的に可能な文節ごとに分かち書きされた文とする。意味分析のある段階で結果が一意的でない場合はすべ

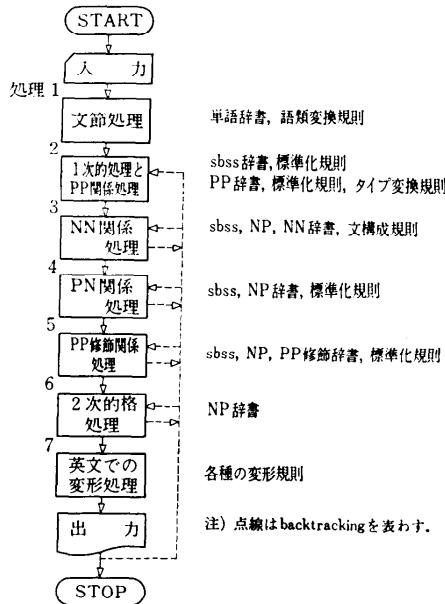


Fig. 6 General flow of our MT experiment

† 規定する辞書を PP 修飾辞書と呼ぶ。

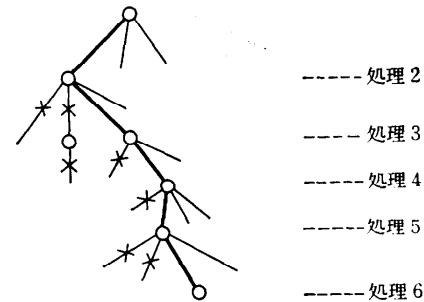


Fig. 7 A part of the "tree" produced by the analythis algorithm

ての結果を列挙し、そのうちの 1 個を選択して次の処理へ進む。分析結果が得られない場合、または出力文が完成した場合はそれまでの処理で分析結果にあいまいさの残っている最後の部分へ backtracking を行ない、選択をやり直して処理をくり返す。このように分析過程では木の探索アルゴリズム<sup>4)</sup>を用いて意味的な文節間の関係の可能なパターンをすべて求め、それぞれに対して訳文を出力する(図 7)。また分析過程では、文節間の関係が交差しないこと、PP 関係は文を分離すること等の構文上の特徴を利用して処理を簡単化している。処理 1, 2, 7 の流れをそれぞれ図 8, 9, 11 に、処理 3~6 については流れを一般化して図 10 に示す。

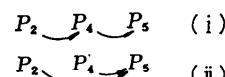
図中、 \*印は文法情報に基づく処理を表わす。実際にはこれらの処理のほかに接続語句(接続詞等)の処理、並列関係の処理、熟語の処理、種々の例外的処理等を行なうが本稿では省略する。また日本語文での言い換えの問題についても検討中である。ここで簡単な例について翻訳過程の大じじを示す。入力文 S は次の通りとする。

'增幅器は 信号と 雑音とを 識別する 能力を  
持たないから 増幅器で 生じる 雑音も 増幅  
されて 出力端子に 到達する'

処理 1-1, 1-4) 結果を表 6 に示す。

処理 1-3) 結果を表 7 に示す。

処理 2-2) S は表現  $S_2, S_4$  および  $S_5$  に分離され  
次のような PP 関係のパターンが得られる。



$S_2 : N_1 \ N_2 \ N_3 \ P_1 \ N_4 \ P_2$

$S_4 : N_5 \ P_3 \ N_6 \ P_4$

$S_5 : N_7 \ P_5$

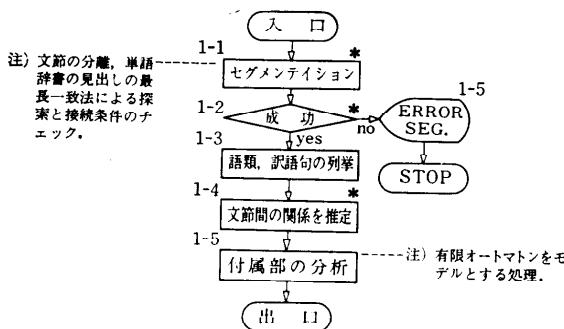


Fig. 8 Process 1

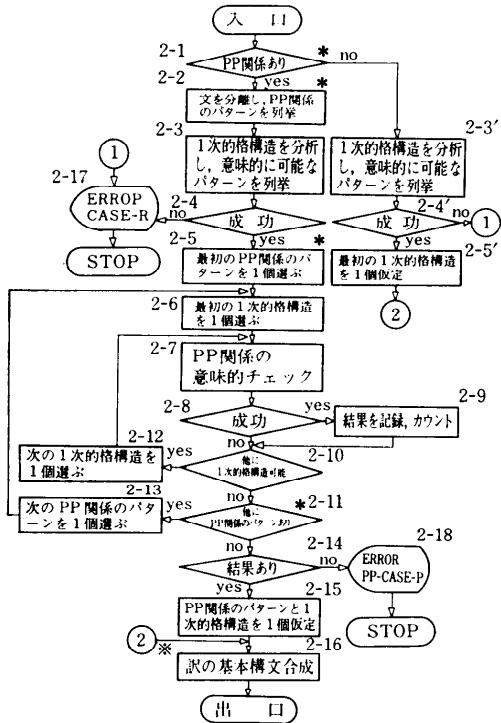
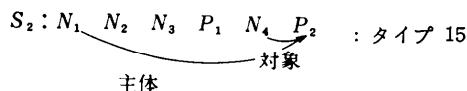


Fig. 9 Process 2

処理 2-3)  $S_2, S_4, S_5$  の内部で 1 次的格構造と文のタイプを求めるところのようになる。



ただし  $S_4$  のタイプは付属語 ‘れる’ の変換規則を適用して得られたものである。 $N_6$  の付属語は ‘が’ と標準化されて  $P_4$  の主体となる。また  $S_5$  の主体は  $S_4$  の中から選ばれる。入力文での各  $P$  に対する  $sbss$  辞書の内容を表 8 に示す。

処理 2-5~2-14) 表 9,10 に示す PP 辞書の内容からパターン (i), (ii) はどちらも可能であることがわかる。

以後パターン(ii)に対する処理について述べる。

処理 2-16) 結果は次の通りである。

$E(S) : as + E(S_2) +, + E(S_4) + and + E(S_5) +.$

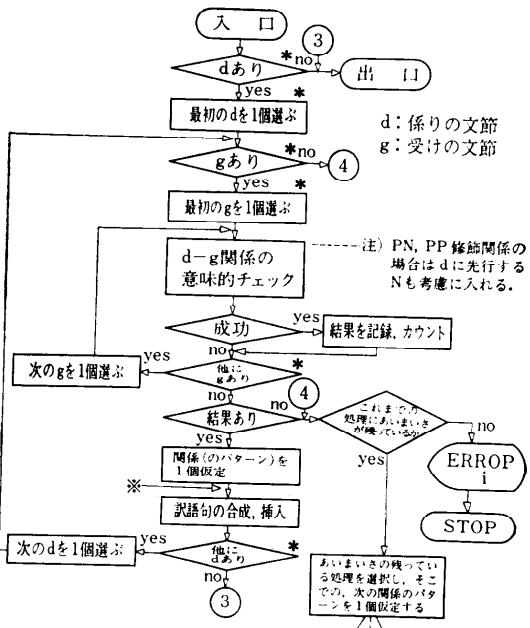


Fig. 10 Process i ( $i=3, 4, 5, 6$ )

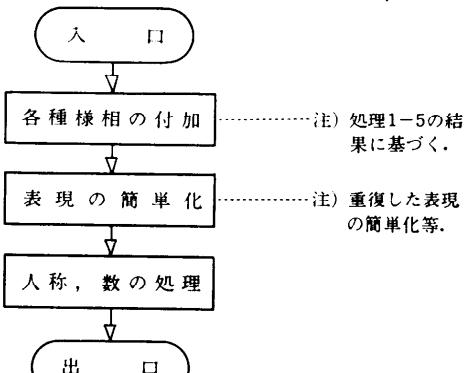


Fig. 11 Process 7

Table 6 Segmentation of each Bunsetsu and possible relations between Bunsetsu's

分離された文節	N, P の別	dependent としての機能
増幅器・は	N <sub>1</sub>	格
信号・と	N <sub>2</sub>	格, 並列
雑音・と・を	N <sub>3</sub>	格
識別・する	P <sub>1</sub>	終止, 遷体修飾
能力・を	N <sub>4</sub>	格
持た・ない・から	P <sub>2</sub>	PP
増幅器・で	N <sub>5</sub>	格
生じる	P <sub>3</sub>	終止, 遷体修飾
雑音・も	N <sub>6</sub>	格
増幅・さ・れ・て	P <sub>4</sub>	PP
出力端子・に	N <sub>7</sub>	格
到達する	P <sub>5</sub>	終止, 遷体修飾

Table 7 Word classes and corresponding English equivalents

N	語類	回路的もの	物理量	回路的ものが持つもの	回路的ものの部分	人が持つもの
増幅器	○ amplifier			○ amplifier		○ amplifier
信号			○ signal	○ signal		
雑音			○ noise	○ noise		
能力				○ capability		○ capability
出力端子					○ output + terminal	

N	語類	人の精神行為	回路的ものと様相を表わす動詞	現象	回路的ものの動作
識別する	○				○
持つ			○		
生じる				○	
増幅する					○
到達する					○

E(S<sub>2</sub>) の基本構文: amplifier + have + capabilityE(S<sub>4</sub>) の基本構文: noise + be + amplifiedE(S<sub>5</sub>) の基本構文: noise + reach + to + output

Table 9 PP relations and conjunctions specified for 'から'

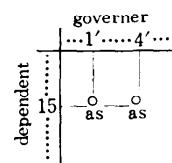
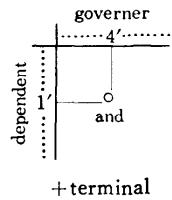


Table 10 A PP relation and a conjunction specified for 'て'



+ terminal

処理 4) 結果は次の通りである。ここで用いるNP辞書の内容を表 11 に示す。

$N_2 \ N_3 \ P_1 \ N_4$  ..... capability + to  
 対象 1 対象 2 抽象的方法 + discriminate  
 + between + signal + and + noise +, + noise + that + originate  
 + and + noise

$N_5 \ P_3 \ N_6$  ..... noise + that + originate  
 場所 主体 + in + amplifier

ここで訳文の概形が次のように求まる。  
 as + amplifier + have + capability + to + discriminate  
 + between + signal + and + noise +, + noise + that + originate + in + amplifier + be + amplified + and  
 + noise + reach + to + output + terminal +.

処理 7) S<sub>2</sub>, S<sub>4</sub> での様相の付加, S<sub>5</sub> での重複した主語の削除, 動詞の語尾処理等を施せば次の出力文が得られる。

$As + amplifier + has + not + capability + to$   
 + discriminate + between + signal + and + noise +,

Table 8 sbss-2's

P	主 体		対 象		対象 1		対象 2		場 所		文のタイプ	訳 文 情 報
	語 類	付	語 類	付	語 類	付	語 類	付	語 類	付		
識別する	回路的もの	が			物理量	と	物理量	を			3'	(主体) + discriminate + between + (対象 1) + and + (対象 2)
持つ	回路的もの	が	回路的ものが持つもの	を							15	(主体) + have + (対象)
生じる	物理量	が							回路的もの	で	4'	(主体) + originate + in + (場所)
増幅する	回路的もの	が	物理量	を							1''	(主体) + amplify + (対象)
到達する	物理量	が							回路的もの	に	4'	(主体) + reach + to + (場所)

+ noise + that + originates + in + amplifier + is  
+ amplified + and + reaches + also + to + output  
+ terminal +.

Table 11 A NP relation and a preposition

Nの語類	Pの語類	付属語	格	前置詞
人が持つもの、回路的ものが持つもの	回路的ものの動作		で	抽象的方法

## 8. むすび

自然言語の機械処理システムを開発するにあたり、言語を一専門分野に限定することの主な利点は、語の意味が分野固有の見方による側面に限定され述語の定義域が制限されること、したがって概念系のモデルの設定が比較的容易となることにある。本稿で述べた格および因果関係の処理を中心とした分析過程により、語の多義性に基づくあいまいさをある程度減らすとともに意味を考慮した構文から構文への翻訳が可能である。しかしながら单一の文内で分析を行なう現段階の処理過程では論理的な矛盾を把握することは困難である。代名詞の意味の詳細な分析や冠詞の意味的な決定の問題と同じく、論理的矛盾を把握するためには文連続分析 (discourse analysis) に基づく詳細な推論過程や環境による制御の過程が必要となり、今後の課題である。本論文で述べた専門分野を対象とした日英MTシステムは将来開発される各種の自然言語処理システムに対する1つの目やすを与えるものと考えられる。当面の実験では語数約1,000、意味処理のための規則は計4,000程度とし、7で述べた処理過程を採用するが実験結果を検討しつつ処理過程の修正、意味情報の充実をはかり、語数、分野を拡大してゆく予定である。

また、処理の複雑化に対処するため、連想記憶の利用についても検討中である。本稿で述べなかった辞書、テーブル類、合成過程等の詳細は実験結果とともに稿を改めて報告したい。なお本稿で用いた図表は物理的なデータ構造を表わすものではない。

本研究は文部省科学研究費、特定研究(1)“学術情報の言語構造論的研究”的一部である。

おわりに、御指導をいただいた故九州大学工学部栗原俊彦教授に深甚の謝意と哀悼の意を表すとともに日頃助言を得ている九州工業大学吉田将助教授、九大栗原研の諸兄、福岡大学工学部石津湛助教授に感謝いたします。

## 参 考 文 献

- 1) N. Chomsky: *Syntactic Structures*, Mouton & Co., Publishers, the Hague, (1957).
- 2) R. B. Banerji: *Theory of Problem Solving, An Approach to Artificial Intelligence*, Am. Elsevier, (1969).
- 3) D. T. Langendoen: *Essentials of English Grammar*, Holt, Rinehart & Winston, (1970).
- 4) N. J. Nilsson: *Problem-Solving Methods in Artificial Intelligence*, McGraw-Hill, (1971).
- 5) 栗原、吉田、鶴丸、藤田: 概念構成と思考行動過程について、信学会オートマトン研資、A-70-63, (1970).
- 6) 栗原、長沢、田中: 自然言語を利用した専門分野における情報処理について、信学会オートマトン研資、A-70-63, (1970).
- 7) 吉田: 二文節間の係り受けを基礎とした日本語文の構文分析、信学会論文誌、Vol. 55-D, No. 4, (1972).
- 8) 吉田、栗原、松尾: 日本語の機械処理-構文分析から意味分析へ-, 信学会論文誌、Vol. 55-D, No. 6, (1972).
- 9) 首藤、栗原: トランジスタ回路に関する分野における日英機械翻訳(第1報), 九州大学工学集報、Vol. 45, No. 5, (1972).
- 10) 栗原: 自然言語の機械処理(解説), 情報処理誌、Vol. 14, No. 4, (1973).
- 11) 岩崎、小稻: 新英和中辞典、三訂版、研究社、(1972).
- 12) 池上、萩原、星野: 電子回路I、オーム社、(1965).

(昭和48年4月12日受付)

(昭和48年6月25日再受付)