

## 電子計算機による和文英訳\*

坂井利之\*\* 杉田繁治\*\* 渡辺明正\*\*

## Abstract

This report describes a method and experimental results of mechanical translation from Japanese into English by a medium size digital computer.

Input sentences, which are fed into the computer in separate phrase form in Japanese letters, are analyzed with the "Table Look-up" method. It generates English sentences which include operators such as ED- or DO-. By using transformation rules on this sentences, normal English sentences are obtained.

For experiment, we used 200 Japanese idioms, 8,000 words, and 600 grammatical patterns. Input Japanese sentences may have any embedded clauses, and any form of sentences such as interrogative, negative, or imperative cases are accepted.

For the improvement, we are investigating as a next step to deduce grammatical relations by introducing semantical information, and to subclassify the parts of speech more in detail.

## 1. まえがき

計算機による自然言語の自動処理を完全に行なうには、本質的な難点が明らかに存在するが、人間にとって事後処理をする箇所が少なく、常識のある人に受け入れられる程度の質のものでもよいから欲しいという要求は、ますます増加している状況である。

計算機による言語翻訳は 1946 年ごろ A. D. Booth や W. Weaver が論じて以来、数々の試みがなされ、日本においても、1958 年ごろから電気試験所、九州大学、京都大学で、主として英文和訳を対象として実験が行なわれ、特に京都大学のもは、何重かの複文にまで適用可能であり、構文的には相当複雑な文も翻訳可能である。この論文は、その英文和訳に用いた基本思想のもとに、和文英訳をかなりの規模で行なったものの報告である。日本語は印欧語と語族が大きく異なっているので、いわゆる Chomsky の syntax-theory がどのように適用されるかは興味ある問題である。

言語のモデルとしては、直接構成要素モデル (Immediate constituent model) を用い、品詞記号を 4 つ、3 つ、2 つと連ねた 4 項則、3 項則、2 項則のパ

ターンに優先順位を設けている。われわれの方法の特徴を挙げると、(1) 直接構成要素モデルの使用、(2) Table Look-up 方式 (パターン方式)、(3) 左→右スキャン方式、(4) 分析・合成同時方式、(5) 変形文法の導入、(6) 入力と和文に構造的制約がない。(7) 翻訳時間が短い、などである。また、プログラムは辞書類と分離されているので、プログラムはパターンの影響を受けず、パターンの追加・修正によって、複雑な文の翻訳がプログラムの変更なしに行なえる特徴を持っている。

## 2. 日本語の言語理論

## 2.1 文の構成

文は文節を要素として成り立ち少なくとも述語文節が必要である。要素の定義は、次のように書くことができる。

〈自立語〉: =〈体言〉|〈用言〉|〈副詞〉|〈連体詞〉|〈接続詞〉

〈付属語〉: =〈助詞〉|〈助動詞〉|〈付属語〉+〈付属語〉

〈文節〉: =〈自立語〉|〈自立語〉+〈付属語〉

〈文〉: =〈文節〉|〈文〉+〈文節〉

ここに用いた記号は ALGOL の場合と同じく、: = は右辺の如きものを左辺のように名づけることを意味し、右辺の縦棒 (|) は OR の意味である。

日本語には文型といわれるような固定したものはほとんどなく、以下のような規則で配列される。

(1) 文は文節を基本構成要素としている。

\* Mechanical Translation from Japanese into English, by Toshiyuki Sakai, Shigeharu Sugita, Akimasa Watanabe (Faculty of Engineering, Kyoto University)

\*\* 京都大学・工学部

- (2) 述語文節は文の最後に現われる。
- (3) 主語は一種の連用修飾語であり、文の成立の必須条件ではない。
- (4) 複文は連体形の述語文節により自由に構成できる。
- (5) 修飾語は常に被修飾語の前にくる。

以上の原則にあてはまる限り、語順はかなり自由であるが、一応標準的な語順は、主語+目的語+述語文節、の形であり、副詞語句は位置の変動が大きく許される。また、文節の結合は単純に合わされるのではなく、時枝文法でいう“入子型構造”という名がよくその結合法を説明しているように、典型的には次々と前のものが後のものを修飾し、明らかに Phrase-structure 的取扱いの正しさを示している。なお、文節を単位と考えると橋本文法的図を得るが、両者とも句構造であることには変わらない。

2.2 品詞分類

構文分析は日本語の各単語自体を見て行なうのではなく、その品詞記号に基づいて行なうのであるから、品詞分類はきわめて重要である。さらに、英訳を目的としているため、その品詞分類は、単に日本語における働きのみでなく、英語になった場合の働きをも考察しなければならない。

2.2.1 自立語の品詞分類

Table 1 に自立語の分類を示すが、それはきわめてあらいものである。細分化によって微妙な翻訳が期待できるのは当然であるが、そのために文法規則が膨大なものになり翻訳時間が長くなる。また、これ以上の細分化には多かれ少なかれ意味の問題がはいってくるため、構文に基づく方式にとっては、労多くして功少なしということになる。

用言としては、形容詞・自動詞・他動詞を採用し、

その細分類は活用形によった。未然形・連体形・命令形は、他の活用形や語幹などに一致するので、独立した品詞とはしていないが、文脈によりその働きがはっきりする場合には記号の書換えを行なう。この活用形による分類は、翻訳対象となる文章を文節に分割して人力する方式をとったためであり、もし、入力単語単位ならば語幹の分類のみでよい。形状名詞 (N2) は英語でもおもに形容詞となるもので、「～の」という形や、いわゆる形容動詞も含む。動的名詞 (N5) と呼んでいるものは、英語になったときに動詞で表現される可能性のある語である。

2.2.2 付属語の品詞分類

付属語は日本語において、語の相互の関係、時制や態などを決定する重要な機能を持ち、英文中にはそれに対応する単語は現われないこともあり、語順の転倒をひき起こしたり、補助語の要求をすることがある。この付属語を正確に把握することが構文分析のポイントとなる。したがって、付属語はかなり細かい分類が必要である。Table 2 に付属語分類を示す。Xは助詞、Zは助動詞の大分類記号である。助詞はそれ自体の訳語を持たないものが多い。しかし、それが他の自

Table 2 Examples of dependent words

記号	名称(代表機能)	例	記号	名称(代表機能)	例
X1	一般助詞	から、まで	XF	假定助詞	ば
X2	所有格	の	Z0	補助用言	ている
X3	主格	が	Z1	一般助動詞	たい、ます
X4	主格	は	Z4	述語化	だ、です
X5	目的格	に	Z7	命令化	なさい
X6	副詞化	へ	Z8	否定化	な
X7	副詞化	で	ZA	比較	より
X8	目的格	を	ZV	助動詞を持つ	のようだ
X9	接続	と			
XA	”	ので	..	ピリオド	
XB	”	のに	⌘	文頭、文末	
XC	”	ても	,,	コンマ	

Table 1 Examples of independent words

記号	名称	例	記号	名称	例
AA	形容詞	語幹	N5	動的名詞	すき、きらい
AB	”	連用形	N6	副詞的名詞	今日、来年
AC	”	終止形	N7	同格名詞	みな、みんな
B1	副詞	たいへん、かなり	R1	連体詞	この、わが
B2	”	(名詞にもつく)	IA	自動詞	歩か (ARUK)
B3	”	(文全体を修飾)	IB	”	連用形
CN	接続詞	そして、また	IC	”	終止形
CT	”	とき	ID	”	音便形
N1	名詞	山、川	TA	他動詞	語幹
N2	形状名詞	あいまい、静か	TB	”	連用形
N3	代名詞	私、彼	TC	”	終止形
N4	形式名詞	こと、ところ	TD	”	音便形
					読み
					読む
					読ん

立語と結合する場合、その文節がどのような構成要素になるかを明確に示す。「太郎は」は訳語としては「太郎」と同じであるが、その文節が主語となりうることを示す。一方、助動詞は明確な訳語を持つものが多い。「～したい」は「want to」となる。ZO は補助用言であり、高々時制の情報を持つにすぎない。Z4 で示す述語化助動詞は断定の「だ」「です」などであり、名詞と結合して述語文節を形成する。ZV はZ4 に近いが動詞補充の必要のないものである。また、複合付属語にもこれらのいずれかの品詞が与えられる。

### 2.2.3 句記号 (Cluster symbol)

**Table 3** に示す記号は句、節、文、または特別の状態を示すもので Cluster symbol と呼んでいい。いわゆる non-terminal symbol であり、樹枝状構造での枝分れ点の記号にあたる。これを単語に対する品詞記号と同列に扱うことにより、単文・複文の区別なく構文分析が可能となる。また、構文分析を行なっていく場合、句記号を1つ見るだけで、その記号が代表している構成要素を大体知ることができる。つまり、句記号によって少数の記号列を見るだけでも、長い品詞列を一度に見たのと同程度の効果が期待できる。Table 3 に示してあるものは代表機能であり、翻訳実験を行ないつつ適当な句の抽出を行なう。以下の議論では、句記号のことも品詞記号と呼んでいる。

Table 3 Examples of cluster symbols

記号	機能	例	記号	機能	例
BP	副詞句	N1+X1	SY	節	MW+VY
ME	"	N1+N6	SZ	"	N1+X2+VY
MG	主語	N1+X3	UY	動詞句	ID+ZY
MI	目的語	N3+X5	VA	"	TB+Z1
MO	"	N3+X8	VP	"	MO+VA
MW	主語	N1+X4	VX	命令動詞句	TA+Z7
RY	節	MW+UY	VY	連体動詞句	TB+ZY
SG	主節	SY+X3	VZ	連用動詞句	TD+X7
SO	目的節	SY+X8	VF	仮定動詞句	TB+XF
SS	文	MW+VP	KB	文末の助詞"か"	KA+結
SX	命令文	結+VX			

## 3. 機械翻訳のための文法

### 3.1 書換え法則 (Rewriting rule)

自然言語はもともと文脈に関係する context-sensitive なものであるが、機械で取り扱う場合複雑な手順を必要とする。一方、context-free と考えると、標準パターンとの一致を調べるだけで簡単に構文分析が可能となる。ここでは、パターンを構成する品詞数を英和訳の際より増加させることにより、context-free

の形で context-sensitive な働きを行なわせている。すなわち

$$\begin{aligned} \alpha \cdot \beta \cdot \gamma \cdot \delta &\rightarrow \varepsilon (\xi, \sigma) && [4 \text{ 項則}] \\ \alpha \cdot \beta \cdot \gamma &\rightarrow \varepsilon (\xi, \sigma) && [3 \text{ 項則}] \\ \alpha \cdot \beta &\rightarrow \varepsilon (\xi, \sigma) && [2 \text{ 項則}] \end{aligned}$$

上のような一般形で示される規則を用いる。ここで  $\alpha, \beta, \gamma, \delta, \varepsilon$  は品詞記号、 $\xi$  は訳語の語順、 $\sigma$  は挿入補助語を示す。上式の意味は左辺の品詞列を右辺の  $\varepsilon$  記号で置換することを意味し、右辺のカッコ内は翻訳文生成のための情報である。

構文分析において parsing を文頭から始めるか、文末から始めるかに関しては、日本語では修飾する場合、常に修飾語句が前に立つ関係なので、文頭から始める左→右 parsing の方法をとった。

### 3.2 文節形成の文法

文節を形成する品詞列は、その結合の度合いが強いものでは一義的に定まる。つまり、特定の品詞の組合せが存在すれば、その結合に対して例外は少なく、ほとんどが2項則で処理可能である。

- (1) 名詞+助動詞 名詞に連体修飾語がある場合は、左→右 parsing によって、すでに結合されているので、名詞+助動詞は無条件に結合できる(ただし、助動詞がX2の場合は例外)。結合の結果生成された文節が、どのような要素となるかは、全く助動詞の種類による。
- (2) 名詞+助動詞 「あれは学校だ」、「彼は学生でした」のように、動詞を含まず述語文節を形成する場合であり、助動詞としてはZ4, ZV などである。
- (3) 動詞+助動詞 動詞句を形成するが、訳語としては英語での助動詞の他に複雑な組合せのものが与えられていることがある。時制・態などの情報を含む。

以上が代表的なものであるが一語で一文節を形成する語もあり、「の」(X2)のように文節を形成できる場合でもこの段階では結合を行なわないものもある。

### 3.3 単文構成の文法

- (1) 連体修飾 名詞+名詞、形状名詞+名詞、連体詞+名詞、特定の副詞(B2)+名詞、などであり、名詞+X2+名詞、代名詞+X2+名詞、などX2を含む場合もこれである。名詞+X2+名詞、に対しては「of」を使っているが、品詞の細分化を行なえば、それ以外の前置詞を入れることもできよう。

(2) **連用修飾** 副詞および副詞が動詞句を修飾する場合、 $B1+VA \rightarrow VA$ ,  $BP+VP \rightarrow VP$ , などの構造を持ち、訳語において語順の転倒をひき起こす。B1, BP の位置に形容詞の連用形やN2がはいることもある。ただ副詞句は他の要素と比較して、位置の自由度が大きいため、動詞句のまじかにあることが少なく、他の要素にさえぎられることがある。この場合でも何段階かに分けて規則を適用すれば、結果的に英語での正しい位置を占めるように規則が構成されている。

(3) **主構成要素の結合** ここで主構成要素というのは英語で、主語・述語・目的語・補語などに相当するものである。二重目的語の場合、 $MI+MO$  (～に～を) は  $MO+MI$  となったり、両者の間に副詞がはいったりするので、 $MI+MO \rightarrow MO$ ,  $MO+MI \rightarrow MO$ ,  $MO+BP+MI \rightarrow (MO+BP)+MI \rightarrow MO$  のパターンによって  $MO$  にしておく。そして  $MO+VA \rightarrow VP$  のパターンによって動詞句を形成する。

$MO$ ,  $BP$  の処理が終わると、残る記号は  $MW$ ,  $MG$ ,  $VA$ ,  $VP$  などであり、文はこれらの組合せとなる。たとえば、 $MW+VA \rightarrow SS$ ,  $MG+VP \rightarrow SS$  を使用する。 $VA$  が文頭記号(##)の直後にきた場合は ##+VA  $\rightarrow SS$  とし、英語においては何か主語を補充する。命令文の場合は述語部が  $VX$  で表現されているから、##+VX  $\rightarrow SX$  として主語を補充する必要はない。疑問文は  $SS+KB \rightarrow SS$  のパターンで処理し、operatorを付加し、後に変形文法によって正しい文とする。受身・否定・推量などは、すべて  $VA$ ,  $VP$  の中で処理できるが、仮定の場合は  $VF$  としておき、節を形成するときに接続詞を補充する。

3.4 複文構成の文法

文章を単文と限定すれば、パターンの数はかなり少なくてよいし、その構成法も簡単である。しかし、言語翻訳を考えるときは、どうしても述語部を複数個含む文を考察しなければならない。重文の場合は単文を構成した後に単純に結合すればよいので問題はないが、句や節を連体形で結合する場合は重要で、英語では関係詞を要求することが多い。

(1) **主格の関係代名詞** 述語部が連体形のもの+名詞、という組合せであるが、述語部が動詞のみで構成される  $IY+N1$ ,  $TY+N1$  は不定詞の形容詞的用法を使えばよい ( $IY$ ,  $TY$  は  $IC$ ,  $TC$  が文脈により書き換えられた記号)。たとえば、「教会へ行く道」は「way to go to church」となる。しかし、 $UY+N1$ ,  $VY+N1$  の場合は、述語に付属語がついているので関係代名詞

を使う。先行詞となる  $N1$  について他の情報が無いので、関係代名詞として  $that$  を使う。たとえば、「昨日死んだ男」は「man that died yesterday」となる。

(2) **所有格の関係代名詞** これは日本語ではあまり出現しないが、述語として形容詞の用いられる場合に時々現われる。「屋根が赤い家」は「house whose roof is red」, 「鼻が短い象」は「elephant whose nose is short」となる。しかし、動詞がくると目的格の場合と区別し難い。

(3) **目的格の関係代名詞** 名詞にかかる語句が節であり、その連体形が他動詞である場合で、 $SY+N1$ ,  $SY+N4$  が対応する。たとえば、「私が昨日食べた果物」は「fruit that I ate yesterday」となる。

(4) **関係副詞** (3)に対して動詞が自動詞の場合には関係副詞と考える。 $RY+N1$  の形となるが、 $N1$  の意味分類が無いため、関係詞としては「that」で代用する。「水が流れる川」は「river that water flow」となる。また、この形で  $RY+N4$  のときは名詞節を形成する。「彼が出席することを知っている」は「know that he attend」となる。

以上の例中にも出てくるように、節中の主語は「が」を伴った  $MG$  の形をしているので、 $MW$  (～は) の場合と異なり混乱は少ない。ただし、 $SY$ ,  $RY$  は  $MG$  が出現した時点においてのみ形成されるものである。

3.5 オペレータを用いる変形文法

変形文法 (Transformation grammar) は Chomsky によれば、核文 (Kernel sentence) に対して変形をほどこして、われわれが通常使用する様々な文型を生成するものである。この考えを利用して翻訳作業の単純化をはかる。すなわち、構文分析時に直ちに通常の訳文を得ることは困難な場合もあり、また煩雑である

Table 4 Operators

記号	機能	例
'S	所有格に変形	I+'S $\rightarrow$ MY
3-	目的格	3-+HE $\rightarrow$ HIM
4-	独立所有格	4-+I $\rightarrow$ MINE
ED-	過去形	ED-+SAY $\rightarrow$ SAID
EN-	過去分詞	EN-+WRITE $\rightarrow$ WRITTEN
ING-	現在分詞	ING-+TAKE $\rightarrow$ TAKING
-ER	比較級	SMALL+-ER $\rightarrow$ SMALLER
-EST	最上級	BIG+-EST $\rightarrow$ THE BIGGEST
-LY	副詞	STRICT+-LY $\rightarrow$ STRICTLY
BE	人称による変形	YOU+BE $\rightarrow$ YOU ARE
BEED	人称による	HE+BEED $\rightarrow$ HE WAS
DO-	疑問文に変形	DO-+I+BE $\rightarrow$ AM I
NOT-	否定文	NOT-+GO $\rightarrow$ DO NOT GO
NOTED-	否定文	NOTED-+HAVE $\rightarrow$ HAD NOT

Table 5 Examples of operator application

疑似英文	変換された英文
(1) I ED- LOVE 3- SHE ING- TURN TO RIGHT	I LOVED HER TURNING TO RIGHT
(2) HE BE BOY I BE EN- LOVE BY 3-SHE YOU BEED SCHOOLGIRL	HE IS BOY I AM LOVED BY HER YOU WERE SCHOOLGIRL
(3) I NOT- BE GIRL. SHE NOTED-READ BOOK	I AM NOT GIRL SHE DID NOT READ BOOK
(4) DO- YOU BE DOCTOR? DO- SHE WILL COME SOON? DO- MARY BE WHERE? DO- YOU ED-SEE WHAT?	ARE YOU DOCTOR? WILL SHE COME SOON? WHERE IS MARY? WHAT DID YOU SEE?

こともあるため、一旦オペレータを含む疑似英文を生成し、後にそれに変形規則を適用して通常の英文にする。オペレータはこの変形の指示のための目印である。オペレータとその適用例を **Table 4, Table 5** に示す。

#### 4. 機械翻訳のアルゴリズム

計算機によって打ち出された翻訳過程図を **Fig. 1** に示した。この図の左端に示した過程を径て翻訳される。

#### 4.1 和文読込

原文入力的手段として、現在はまだ漢字が容易に使用できないので、カナ文字を用い、文節分ち書きで行なう。文節を使用する理由は、文法が文節を基本単位としていること、単語に分割するより文節分割の方が、われわれにとって容易であると考えられることによる。また、タイプライタで上下段シフト記号を利用することにより、ローマ字表記、もしくは英語名詞を入れること、並びに促音・撥音(Q)の表現も可能である。入力は内部ではすべてローマ字表記に変換される。これは活用形の処理が容易にできるためである。

#### 4.2 分割処理

文節ではいつてきたものを、単語のレベルにまで分割する必要がある。これは、いわゆる分ち書き問題である。通常は単語辞書をひきながら、最長一致法で分割するが、ここでは、付属語をあたかも語尾のように考え、後方より複合付属語を含めて、一度で分割する方法をとる。これは通常の方法と比較して、付属語表を見るだけでよいため、時間が早く、また、複合付属語がしばしばまとまって1つの文法単位となるなど有

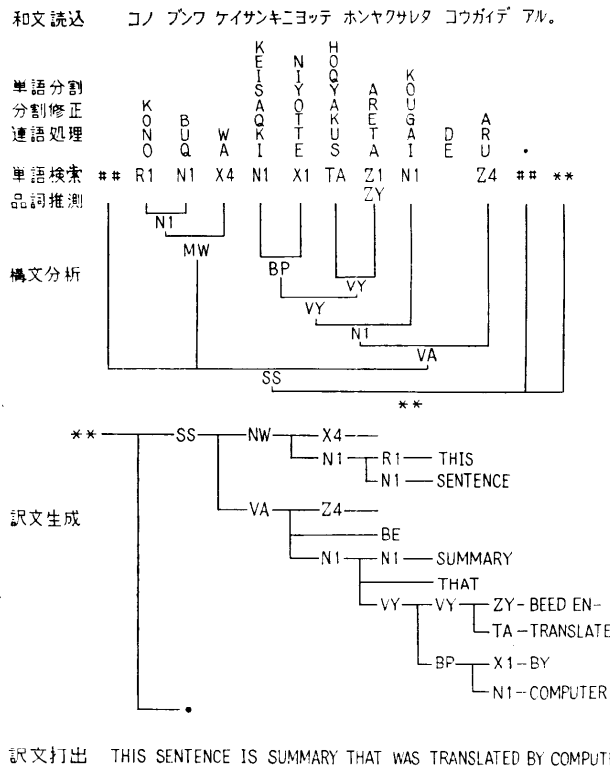


Fig. 1 Process of mechanical translation

Table 6 Correction of word division

誤 分 割	修 正 分 割
KO-NO	KONO
WA-KATTA	WAKAT-TA
ZUT-TO	ZUTTO
MA-TA	MATA
MOHA-YA	MOHAYA
ZYUU-YOUNA	ZYUUYOU-NA
SOSI-TE	SOSITE
-NAITO	-NAI-TO
-ANAIGA	-ANAI-GA
-DATO	-DA-TO

利な点が多い。付属語としては200語ばかりであり、そのうち助詞・助動詞の数が、活用形も含めて100語近いことを考えると、数が少ないにもかかわらず、かなりよく分割できる(付属語の例: NO, WA, NODE, YOUNA, ANAKATTA,...)。

4.3 分割の修正処理

前記の分割は辞書を見ないので、分割誤りがあり、また、複合付属語をさらに分割する必要のあるときもあるので、これを修正する。これは分割された特定の組合せに対して、再分割・分割修正・再結合を行なうもので、例を Table 6 に示す。現在150組くらいである。

こうして得られた各単語は折重ね法により5文字に圧縮し、固定長にしておく。

4.4 連語処理

これは英語の idiom に相当するようなものである。熟語または慣用句は、その言語にとって特別な意義があるため、個々の単語から正しく翻訳できない場合が多い。さらに、日本語には補助用言があり、これを助動詞的に扱うことが必要となる。これらのために連語テーブルを用いて、いくつかの単語のかたまりを、あたかも一語であるようにみなして処理する。現在約200組の連語があるが、その1/3は補助用言に関するものである(連語の例: -NISIKA+SUGINAI, YAKUNI+TATU, DOU+KAQGAE+-TEMO, -NO+NAKA+-DE,...)。

以上の各処理の様子を具体列で示す(ローマ字表記

をし、圧縮しない形で示す。Qは撥音を示す)。

- (1) 入力文: KONO HOQWA ZYUUYOUNA BUQKENDE  
ARU
- (2) 分割処理: KO-NO HOQ-WA ZYUU-YOUNA BUQKEN-DE  
ARU
- (3) 修正分割: KONO HOQ-WA ZYUUYOU-NA BUQKEN-DE  
ARU
- (4) 連語処理: KONO HOQ-WA ZYUUYOU-NA BUQKEN  
DEARU

4.5 単語検索

単語辞書に登録されている見出し語は、すべて折重ね法によって5文字に圧縮されている。これは原単語を5文字ごとに分割してそれらを単純に数として加え合わせて得られるもので、単語数8,000語に対して、同一見出しとなる単語はわずかに3語であった。辞書に登録される情報は、固定長化した見出しとその品詞、および訳語である。1つの単語が持つ品詞は最大2つであり、訳語は一品詞につき1つ与えられる。未登録単語の品詞は一応名詞(N1)とされ、訳語は入力単語そのままが用いられる。辞書の例を Table 7 に示す。

4.6 品詞推測

単語に対して2つの品詞が与えられている場合、どちらかに決定しなければならない。付属語はほぼ一義的に決定できるので、問題となるのは自立語である。しかし、自立語と付属語の結合は必ずしも自由ではなく、ある程度の拘束がある。したがって、二品詞語が現われた場合、次の品詞との結合をパターン表で調べ可能なものを採用する。ただし、どの組合せも可能、もしくは不可能の場合は、第1品詞を優先させる。たとえば、<WATASI>という単語はTBとN3の品詞を持つが、

$$(TB \text{ or } N3)+X4 \Rightarrow N3+X4$$

$$(TB \text{ or } N3)+Z1 \Rightarrow TB+Z1$$

私 は  
渡した

その後にX4がきた場合には、名詞N3(私)とし、Z1がきた場合は動詞TB(渡し)とする。

また、用言では連体形と終止形が同形であるが、次

Table 7 Examples of word dictionary

日 本 語	見 出 し	品 詞	英 語
GASUSUTOBU	051R4	N1	GAS-STOVE
HAI TATUSURU	06: P@	TC	DELIVER
HAISEQ	0AISE	N1	DEFEAT IN WAR
HAQBUQ	0AQB	N2	HALF
TAMERAI	4#MER	TB N1	HESITATE+HESITATION
KYUUSOKU	9/QUR	N2 N1	RAPID+REST
WATASI	&ATAS	TB N3	PASS+I

Table 8 Examples of patterns

品詞列		置換	語順	補助語	品詞列		置換	語順	補助語
A <sub>4</sub>	## MO TB XF	MW	1 3 2 0	3 0 0 0	B <sub>4</sub>	## MG VP KB	SS	4 2 3 0	> 0 A 0
	N1 ZA AA Z8	VY	4 3 2 1	1 0 0 0		## MO TB ZS	SS	1 4 3 2	I 0 0 A
	BP CN N1 X5	BP	1 2 3 0	0 0 0 0		MW B1 MG VA	SS	1 4 3 2	0 0 0 0
	BP VP ., VP	VP	2 1 4 0	0 G 0 0		VP MW VA ##	SS	4 1 2 3	I 0 0 0
A <sub>3</sub>	B1 B1 X9	RP	1 2 0	0 0 0	B <sub>3</sub>	MI MG VA	SS	2 3 1	0 0 0
	MG TY N1	N1	3 1 2	2 0 0		MG VP KB	SS	3 1 2	> 0 A
	MG AY N4	N1	3 1 2	2 1 0		MW CN VP	SS	2 1 3	0 0 0
	MG TY TY	SY	1 2 0	0 0 0		MW N3 Z1	SS	1 3 2	1 0 0
A <sub>2</sub>	AB N2	N2	1 2	G 0	B <sub>2</sub>	CN VA	SS	1 2	I 0
	AB VF	VF	2 1	0 0		MW VA	SS	1 2	0 0
	KA ##	KB	0 0	0 0		SS KB	SS	2 1	> A
	B1 AY	AY	1 2	0 0		SS SS	SS	1 2	0 0

補助語の例: 1: BE, 2: THAT, 3: TO, A: ?, G: AND, I: I(WE, IT), >: DO-

にくる品詞によって、両者の区別ができるときがある。

- (1) TC + .. → TC + .. (終止形)
- TC + N1 → TY + N1 (連体形)
- (2) Z1 + CN → Z1 + CN (終止形)
- Z1 + N4 → ZY + N4 (連体形)

上例で TC, Z1 は次にくるものによって終止形であったり、連体形であったりする。

4.7 構文分析

前段までに用意された品詞列に対して、文法パターンを準備し、Table Look-up 法により英文和訳の場合と同様、構文分析と部分訳文合成を同時に行なう。パターンの構造は Fig. 2 に示すとおりである。第1の部分は比較されるべき品詞列、第2の部分はそれと置換される品詞、第3の部分は訳文合成時の語順、第4の部分は補助語の種類・順位を示す。Table 8 にパターンの例を示す。

パターンは大きく A, B の2種類に分類し、各々に4項則、3項則、2項則がある。AはBに優先し、4項、3項、2項の順に優先順位が下がる。結局、順序

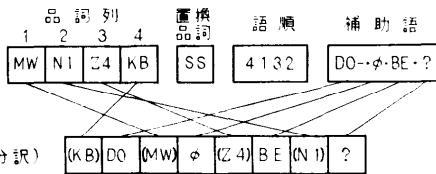


Fig. 2 Structure of pattern

づけられた6個の群になり、各群内においては優先順位はない。A群は句、節を形成するものが主体であり、どちらかといえば結びつきが強く、context-freeであり、B群には文を形成するものがはっている。パターンの適用方法を Fig. 3 に従って説明する。

まず、品詞列の先頭より連続する品詞 P<sub>1</sub>P<sub>2</sub>P<sub>3</sub>P<sub>4</sub> を取り [A-4] 群を調べる。一致するものがなければ

(品詞列)

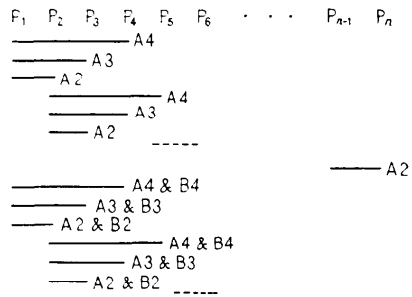


Fig. 3 Algorithm of parsing

P<sub>1</sub>P<sub>2</sub>P<sub>3</sub> を [A-3] 群で調べる。なければ次に P<sub>1</sub>P<sub>2</sub> を [A-2] 群で調べる。これも不一致ならば P<sub>2</sub>P<sub>3</sub>P<sub>4</sub> を取り、同様の順で各パターン群を調べる。いま、仮りに p<sub>i-1</sub>p<sub>i</sub>p<sub>i+1</sub> が [A-3] 群にあれば、そこに示された置換品詞 p<sub>j</sub> で p<sub>i-1</sub>p<sub>i</sub>p<sub>i+1</sub> を置換し、p<sub>i-1</sub>, p<sub>i</sub>, p<sub>i+1</sub> が持つ訳語 Q<sub>i-1</sub>, Q<sub>i</sub>, Q<sub>i+1</sub> を指示された語順に並べ、それらの訳語間に指示された補助語を挿入する。こうしてできた訳語群を p<sub>j</sub> の訳語と考える。この操作の後にもまた文の最初から同様の手順を繰り返す。A群に一致するパターンがなくなれば、A群とB群の両方を用い、[A-4], [B-4], [A-3], [B-3], [A-2], [B-2] の順で同様の手順を繰り返す。構造分析の過程は必要とあれば Fig. 1 に示すように、樹枝状構造として打ち出すことができる。

4.8 訳文生成

訳文は分析に使ったリスト構造を逆にたどって、push-down-store を使って順次生成されるが、この出力英文はオペレータを含む疑似英文であるため、オペレータ処理による語形変化、および語順の変更が必要である。これは transformation rule の一部と考えてよい。Fig. 1 には疑似英文(樹枝状表現の右端)

Table 9 Examples of MT results

<p>(1) 私は走った。 I RAN. それは我々の学校です。 IT IS OUR SCHOOL. あなたはボールを持っていた。 YOU HAD BALL. 彼は私に本を与えた。 HE GAVE ME THE BOOK. 我々は彼を JOHN とよびます。 WE CALL HIM JOHN.</p> <p>(2) 彼は正直でなかった。 HE WAS NOT HONEST. 我々は米を食べなかった。 WE DID NOT EAT RICE.</p> <p>(3) あなたは医者ですか。 ARE YOU DOCTOR? 彼女はまもなくくるでしょうか。 WILL SHE COME SOON? あなたは何を見ていたのか。 WHAT DID YOU SEE?</p> <p>(4) 食堂へ行きましょう。 LET US GO TO DININGROOM あなたの手紙を書きなさい。 WRITE YOUR LETTER.</p> <p>(5) 私の母はこぼしてばかりいる。 MY MOTHER IS COMPLAINING. 我々は英語を研究しつづめる。 WE ARE INVESTIGATING ENGLISH.</p> <p>(6) 彼女は私より美しい。 SHE IS MORE BEAUTIFUL THAN I. 彼は組で一番背が高い少年です。 HE IS THE TALLEST BOY IN CLASS.</p> <p>(7) 我々は皆に愛されている。 WE ARE LOVED IN (AT) EVERYBODY.</p> <p>(8) 私は仕事を終えたところで。 I HAVE FINISHED THE WORK. あなたはフランスへ行ったことがありますか。 DO YOU HAVE GONE TO FRANCE?</p> <p>(9) 質問していいですか。 MAY I(WE, IT) ASK? 彼女は金持ちだったに違いない。 SHE MUST HAVE BEEN RICH.</p> <p>(10) 我々はお目にかかれて大変うれしい。 WE ARE VERY GLAD TO SEE YOU 我々は生きるために食べる。 WE EAT IN ORDER TO LIVE.</p>	<p>(11) 右へまがると、郵便局が見つかるでしょう。 TURNING TO RIGHT, POST OFFICE WILL BE FOUND. 国境の長いトンネルを抜けると雪国であった。 PASSING THROUGH LONG TUNNEL OF BORDER, IT WAS SNOW COUNTRY.</p> <p>(12) あなたはすぐ駅へ行ったほうがよい。 YOU HAD BETTER GO TO STATION AT ONCE. 私があなただったら私はそれをしただろう。 I WAS YOU I WOULD DO IT.</p> <p>(13) 私は馬鹿だ、しかし、私は愛することを知っている。 I AM FOOL BUT I KNOW TO LOVE. 私は少年だ。そしてあなたは少女です。 I AM BOY AND YOU ARE GIRL.</p> <p>(14) 彼女は父によって与えられたリンゴを食べている。 SHE EAT APPLE THAT WAS GIVEN BY FATHER 我々は彼が昨日会った男を知っている。 WE KNOW MAN THAT HE SAW YESTERDAY 屋根が赤い家は学校です。 HOUSE WHOSE ROOF IS RED IS SCHOOL. 我々は水が流れている川を渡る。 WE CROSS RIVER THAT WATER FLOW. 私は彼が正直であることを知っている。 I KNOW THAT HE IS HONEST. 音が空気の振動であることは読者諸兄のよく知るところと思うが、しかし、我々が話をしたり、音楽を楽しんだりする音は、我々が使用する空気振動のごく一部分にしかすぎない。 I(WE, IT) THINK THAT READER KNOW WELL THAT SOUND IS VIBRATION OF AIR, BUT SOUND THAT WE DO STORY AND ENJOY MUSIC IS ONLY ONE PART OF AIR VIBRATION THAT WE USE. ここで、これからお話しする超音波は、もはや、我々の耳には聞えない周波数の高い振動である。 SUPER SONIC TO TELL HERE FROM THIS IS HIGH VIBRATION OF FREQUENCY THAT IS NOT HEARD, NOW, IN(AT) OUR EAR. ON-LINE TIME SHARING SYSTEM では人間と機械が直接会話することができるが、その文法は実際にはまだ不完全である。 HUMAN BEING AND COMPUTER CAN CONVERSE WITH DIRECTRY IN ON-LINE TIME SHARING SYSTEM, BUT THE GRAMMAR IS IMPERFECT INDEED YET. 自然言語と計算機言語との間のギャップをうめるためには人工の人間言語を創造することが望ましい。 TO CREATE ARTIFICIAL HUMAN BEING LANGUAGE IN ORDER TO BURY GAP OF THIS NATURAL LANGUAGE AND COMPUTER LANGUAGE IS DESIRABLE.</p>
--	---

と、最終出力英文が下半分に示されている。

## 5. 実験の結果と検討

### 5.1 実験システム

使用した電子計算機は NEAC 2200/200 で、コア容量は 16 Kch (6 bit/ch)、サイクルタイム 2 $\mu$ s の可

変語長汎用電子計算機である。周辺装置として、プログラム用磁気テープ、辞書類磁気テープ各1本を使い、入出力に紙テープ読取り装置、ラインプリンタを使用した。プログラムはアッセンブリー言語である EASYCODER で書かれており、実行文としては約 1,800 ステートメントである。



翻訳時間は 30 文節くらいの文で、40 秒足らずである。なお、この時間のうちで約 1/3 が磁気テープからのデータ移送の時間である。

## 5.2 翻訳例

翻訳例を Table 9 に示す。名詞は単数・複数の区別がなく、動詞も 3 人称単数変化をしない。冠詞は特別の場合を除いては考慮していない。また、I (WE, IT) は日本語に主語がなかった場合に仮りに挿入するもので、適当に選択すればよい。前置詞に関しても意味的に正しいものを決定することは困難であるため、仮りに IN(AT) などとなっている場合があるが、それ以外の場合もありうる。以下、各例題番号に従って説明する。

- (1) 基本文型 英語の 5 文型に対応する。日本語の動詞は 2 分類にすぎないが、いろいろな場合に対応して訳出できる。
- (2) 否定 2 種類の否定形がある。NOT オペレータは助動詞「ない」に対応するが、形容詞「ない」については、多少異なる考慮が必要である。
- (3) 疑問 DO オペレータの働いた結果であり、各種の変形能力があることを示す。
- (4) 命令 動詞句が文頭に来た場合で、一般の文と処理方法は変わらない。
- (5) 進行形 直訳体の日本語を進行形にすることはやさしいが、進行形の意識は日本語では薄い。しいていえば、「～している」などの補助用言がその働きをするが、常にそうではなく、接続する動詞により変わるので、動作動詞と状態動詞の区別が必要となる。
- (6) 比較 比較級は助詞「より」、最上級は「いちばん」などを目印としているが、「より」が動詞と関連する場合は処理を誤る。
- (7) 受身 「～によって」と明確に表現されていない場合には、前置詞が正しくはいらぬ場合がある。
- (8) 完了 日本語では明確に完了形を示さないので過去とあまり区別はない。「～したところだ」などの表現があったときには完了形となる。
- (9) 助動詞 日本語と英語の助動詞が一致している場合のほかは、「なければならぬ」、「かも知れぬ」、「ことができる」などの連語が対応する場合が多い。
- (10) 不定詞 名詞的用法は「すること」の形で

あるので明確であるが、他の用法は文脈から判断する必要がある。

- (11) 分詞構文 日本文で主語がなく、「と」で区切られている文に限定して使っているが、曖昧なことが多い。独立分詞構文は連語処理を使用するときのみ有効な翻訳が期待できる。
- (12) 假定 仮定の接続詞は必ずしも存在するとは限らないので、仮定形の助動詞、助詞を中心として決定する。BE 動詞を WERE とするには、別の考えが必要である。
- (13) 重文 単文の分析が終了した後に結合すればよいが、等位接続詞は、文+文以外に、語+語、句+句のレベルでも使用されるので、その接続詞の及ぶ範囲を明確にしないと誤る。
- (14) 複文 複文構成は節の結合が要点であり、節の結合が成功すれば、直ちに単文のレベルになる。ただ、日本語での連体型結合はほとんどの場合、関係詞を使って翻訳されるので、たどたどしい文体になる。

## 5.3 翻訳結果の検討

いろいろなタイプの文を数多く翻訳してみると、パターン不足のために、翻訳できない場合もあるが、このときは、他のパターンと矛盾しないように注意して、新しいパターンを追加すれば、アルゴリズムそのものは、何ら変更することなく、翻訳ができるようになる。

訳語が不適當である場合の中には、同音異義語と多義語の問題があるが、前者に関しては、漢字を導入すれば、解決できる部分がかかなりある。多義語の場合は、前置詞の用法も含めて大量の用例集を用いる必要がある。「電報を打った」は「HIT」ではなく「SENT」であり、「皆に愛される」は「IN(AT)」ではなく「BY」にするとかは意味の細かい分類が必要になる。同じようなことが、MW・MW・VA という形の文についてもいえる。「私は仏語は話せる」において、どちらが主語なのかははっきりしない。日本語では「仏語は私は話せる」ともいえるからである。さらに、「私は彼は好きだ」、「彼は私は好きだ」となれば、意味分類をいくらしても、この文だけからは確定した翻訳はできない。これらの意味に関係したことで、何らかの細分類ができてしまえば、この論文で述べた句構造文法的取扱いによって、構造的には翻訳が可能となる。

## 6. あとがき

この論文では、日本語から英語への機械翻訳の研究について述べた。この研究を通して、句構造の概念が印欧語と全く異なる族の日本語に対しても、多少の例外を除いて、充分あてはまることがわかった。このことは従来、非論理的言語と評された日本語が、句構造というレベルで考えるならば、機械的に取り扱える程度に、構造的論理性を持つことを示している。また context-free grammar は自然言語の近似でありながら、取扱いの工夫によっては、context-sensitive grammar に近い能力を持ち、機械翻訳システムには、きわめて有効であるといえる。

和文英訳の問題点としては、意味情報を含めた品詞の分類とその方法が、今後の重要な課題となる。しかし、意味といっても、まずは構文分析を主体とした場合のことから手をつけていけばやりやすい。さらに、高度の意味の問題については、翻訳過程のどこかの点において、人間が介入することを必要とし、また、そうしないでは能率のよい、良質の機械翻訳システム

は構成できないように思われる。

#### 参 考 文 献

- 1) N. Chomsky: "Syntactic Structure" Mouton & Co., The Hague (1962).
- 2) 橋本進吉: "国語法研究", 岩波書店 (昭 23).
- 3) 時枝誠記: "日本文法国語篇", 岩波書店 (昭 25).
- 4) 森岡健二, 他編集: "国語文法講座 2", 明治書院 (昭 40).
- 5) 坂井, 杉田: "機械による英文和訳", 電気通信学会誌, 第 49 卷, 2 号 (昭 41).
- 6) 植村俊亮: "機械翻訳 (和文英訳) の一方法について", 京都大学修士論文 (昭 41).
- 7) 杉田繁治: "A Study on Mechanical Translation from English into Japanese", 京都大学博士論文 (昭 43).
- 8) 坂井, 他: "計算機による和文英訳", 電子通信学会専門委員会資料 (昭 44-01).

(昭和 44 年 5 月 17 日 受付)