ＮＥＡＣ100による和文英訳

谷口道興

最近オフィス・オートメーション（ＯＡ）がブームになっている。このＯＡでは事務処理上のあらゆる作業をその対象としている。販売・仕入・経理，在庫管理，顧客管理，給与計算等。また外国文献の翻訳，外国向けビジネスレターの作成等も含まれる。前者についてはコンピュータ化が著しく進んでいるが，後者のコンピュータを使用しての機械翻訳（二言語間あるいは多言語間）については，まだ個人の手作業に負うという感じがある。本稿では機械翻訳のための大型専用コンピュータではなく，小型の汎用コンピュータである長野大学 NEAC システム100M80II を使用して，日本語を英語に翻訳する方法論を述べる。使用言語は COBOL である。

１．和文英訳，機械翻訳のアルゴリズム

計算機（コンピュータ）は人間のように文章を読んで，その意味を理解するという水準まで達していない。計算機にそれを実行させるためには文章を符号化，数値化しなければならない。文章を符号化，数値化するために，文章を構成する各品詞をコード（code）によって表現することにする。

英語の品詞は名詞，代名詞，形容詞，副詞，動詞，前置詞，接続詞，間投詞の８品詞に分けられる。このうち動詞は動詞と助動詞の２つに分類する。これらの品詞に対して，名詞は N，代名詞は PRON，形容詞は A，副詞は ADV，動詞は V，助動詞は AUX，前置詞は PREP とコード化しておく。さて，英文を構成する主構成要素は主語，目的語，補語である。この４つを主構成要素と呼ぶことにする。この主構成要素のうち，目的語は直接目的語と間接目的語の２つに分類する。この５つの構成要素がどのように配列されて英文が構成されるかを次の５つの文型で表わす。

文型１：　主語＋動詞
文型２：　主語＋動詞＋補語
文型３：　主語＋動詞＋目的語
文型４：　主語＋動詞＋間接目的語＋直接目的語
文型５：　主語＋動詞＋目的語＋補語

この５つの文型の中で，すべての主構成要素４つを含んでいるものは文型５である。文型５の目的語を間接目的語と直接目的語の２つに分類したものを

文型６：　主語＋動詞＋間接目的語＋直接目的語＋補語　とする。

この文型はすべての構成要素を含むもので，その構成要素がどのような順序で配列されるかを示している。そのため実際の文では，この中のいくつかが抜けることもある。

文型６について，それぞれの構成要素になり得る品詞を当てはめてみる。主語に対して名詞または代名詞が，動詞に対しては動詞が，間接目的語に対しては代名詞が，直接目的語に対しては名詞または代名詞が，補語に対しては名詞または代名詞または形容詞が当てはまる。すなわち名詞または代名詞＋動詞＋代名詞＋名詞または代名詞＋名詞または代名詞または形容詞となる。これが文型６を品詞で表わしたものである。これをコードで表わすと，N または PRON＋V＋PRON＋N または PRON＋N または PRON または A となる。

さらに文の各構成要素には，それぞれ修飾語がつけられる。修飾語としては副詞，形容詞，前置詞，助動詞がある。これらの修飾語の機能を定義しておくと，副詞は動詞や形容詞または他の副詞を修飾する語である。しかし，本稿の段階では動詞または形容詞を修飾する語と限定している。形容詞を修飾する場合は，形容詞の前に置かれるものであり，動詞を修飾する副詞は英文の並び方の

関係から文の最後にくる場合がある。形容詞は，名詞または代名詞の前に置かれてそれらを修飾する語。前置詞は，名詞または代名詞の前に，また名詞または代名詞が形容詞や副詞で修飾されている場合には，さらにその前に置かれる語。助動詞は，動詞の前に置かれてそれを修飾する語。

これらの修飾語を，定義に従って品詞で表わした文型6に当てはめると，

ADV＋A＋N または PRON＋AUX＋V＋PREP＋ADV＋A＋PRON＋PREP＋ADV＋A＋N または PRON＋N または PRON または A＋ADV となる。これが品詞で表わした文型6の応用である。最後の N または PRON または A に対して修飾語が付いていないのは，補語には修飾語がほとんどないと思われるからである。

品詞で表わした文型6の応用をコードで表わすと図1となる。

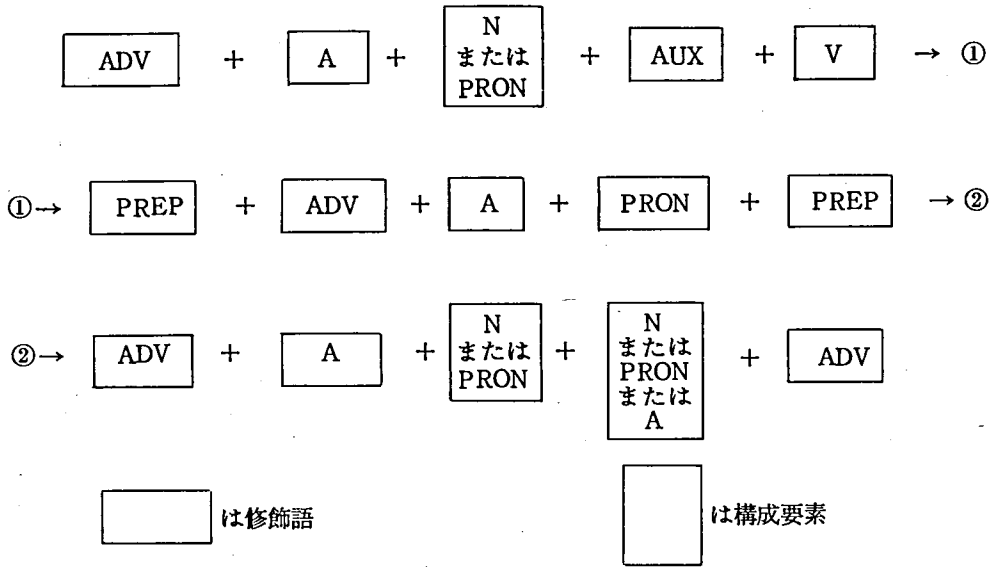


図1．　英語の文型

これに対して，日本語の品詞の分類は名詞，代名詞，形容詞，形容動詞，副詞，連体詞，動詞，接続詞，感動詞，助動詞，助詞となっている。また日本(語)文を構成する主な構成要素は，主語＋述語＋動詞である。このように日本語の品詞と構成要素の分類は，英語のそれとはかなり違っている。しかし機械翻訳を行うには，英語の場合と同じ品詞と構成要素を使用して，英語の場合と同じ方法で日本語の文型を作る。そのため形容動詞は形容詞に含め，助詞は英語の前置詞に相当するものとして考える。

英文の場合，4つの主構成要素をすべて含むものは文型5であった。日本語になった場合の主構成要素の並び方は，主語＋目的語＋補語＋動詞である。目的語を間接・直接目的語に分けて，主語＋間接目的語＋直接目的語＋補語＋動詞とする。これらの構成要素になり得る品詞は，英語の場合と同じで主語に対して名詞または代名詞，間接目的語に対して代名詞，直接目的語に対して名詞または代名詞，補語に対して名詞または代名詞または形容詞，動詞に対して動詞である。つまり名詞または代名詞・代名詞・名詞または代名詞・名詞または代名詞または形容詞・動詞となる。これらの品詞に修飾語が付くのであるが，日本語の場合には英語の場合と違う点がいくつかある。(例えば前置詞の位置が「学校へ」は“TO SCHOOL”と前置詞の位置が前後逆になる。また助動詞の位置も「行くだろう」が“WILL GO”というように前後逆である。さらに動詞を修飾する副詞は文の最後ではなく文中に入る。その他の修飾語については，英

語の場合とほゞ同じである。)

これらのことからそれぞれの品詞に修飾語をつけてコード化したものが図2である。日本語を英(語)文に翻訳するためには，日本語の文型の配列を文型6の応用の配列に並びかえればよい。しかし品詞で表わした文型を使用したのでは，計算機で実行はできない。計算機で実行させるためにはコード化した文型を使用することになる。

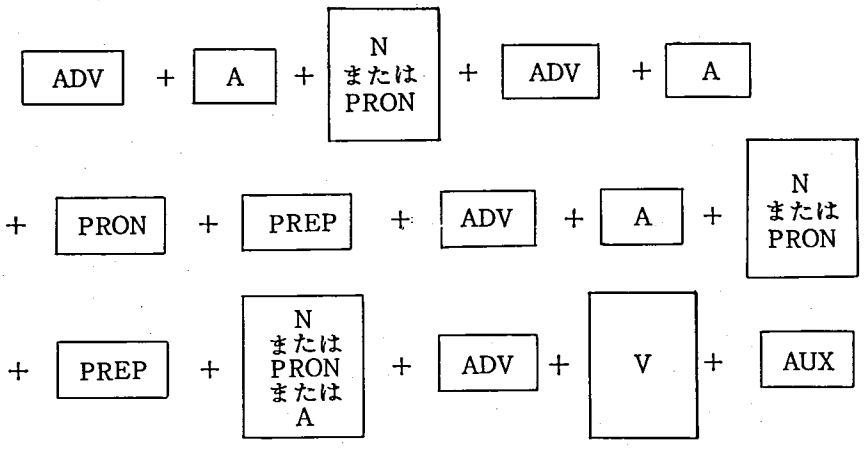


図2．　日本語の文型

2．ファイルについて

計算機に翻訳をやらせるための必要な要素の一つとして辞書がある。この辞書に相当するものがファイルである。ファイルはレコードの集合によって構成される。さらにレコードは項目の集合によって構成される。本稿のファイルは一つのレコードを3つの項目によって構成している。(例)「ガッコウ SCHOOL N」これは片仮名で表わした日本語の部分『ガッコウ』と，それに対応する英語の部分『SCHOOL』と『SCHOOL』という名詞をコード化した部分『N』の3つによって構成されていることになる。

上例において，3つの項目をそれぞれJAPANESE, ENGLISH, M-CODEと呼ぶことにする。翻訳の不完全さを補うためにファイル内のJAPANESEを詳しく記述することにする。

ある一つの英単語についても，日本語の言い回しで違った言い方をすることがある。例えばHAPPENという単語は，オコル，オキル，オコリマス，オキマスというように幾通りもの表現がある。したがってHAPPENという一単語に対してレコードが4件分ほど必要ということになる。逆にそれが日本語についてもいえる。また「エキデ」とか「クルマデ」とかのように場所を示すか方法を示すかの違いで，日本語の「で」という言葉が「AT」または「BY」と変わってくる。次に同音異義語について。日本語の「オキル」という語に対しては，WAKE UP, HAPPEN, RISE等の英語が当てはまる。また，アクセントの判別ができないので「カエル」という日本語に対しては，FLOG, CHANGE, COMEBACK等の英語が当てはまる。

このように，日本語の表現をできるだけ詳しくファイルに記述しておくことが必要で，言いかえればファイルを整備しておくことが重要となる。これらのファイルは必ずしも五十音順，またはアルファベット順に並んでいる必要はない。しかしプログラムの都合上，同音異義語だけは連続して並んでいなければならない。また，M-CODEの部分で名詞(N)は単数形(N−S)と複数形に，動詞も現在形(V)と過去形(V−P)に分けてある。これは翻訳のためのプログラムが高度になれば当然これらの識別が必要となるからである。しかしこの実験的なプログラムではその識別は除外し，次稿で検討することにしている。

3．入力部について

プログラムで使用する変数について。TABLE 1：ファイルから読み出された英単語の記憶場所。 TABLE 2：ファイルから読み出されたコードの記憶場所。 TABLE 3：コードの識別のために付けられたウェイトの記憶場所。TABLE 4：ファイルから読み出された日本語の記憶場所。 TABLE 5：ファイルから読み出された英語を正しく並びかえたものの記憶場所。BAN：TABLE 1，2，3の位置を示す。CNT：TABLE 4，5の位置を示す。 ACP：キーボードから入力された仮名文字の記憶場所。WEIGHT：ウェイトの計算場所。 ENGLISH 1とCODE 1：同意異義語とそのコードの記憶場所。 ENGLISH 2とCODE 2：同音異義語とそのコードの記憶場所。 W 1，W 2，W 3：単語の順序を入れかえるための記憶場所。

プログラムを実行させると，これらの変数がゼロまたは空白でクリアされる。次にBANに1が加算される。ついでカナ文字の入力が可能となる。このときの入力方法は一単語づつ区切って行なう。入力された単語はACPに格納される。次にファイルが1件づつ読まれてACPに格納されているカナ文字と一致するレコードを捜す。その際一致するレコードがファイルに無い場合は，画面上に“UNDISCOVERD IN MASTER-F”の文字が表示されるので，キーボードから必要な英単語と，その品詞を示すコードを入力することになる。これらはそれぞれTABLE 1とTABLE 2のBAN番目に格納される。またACPに格納されているカナ文字と一致するレコードがファイル内にあれば，一致したレコード内のENGLISHとM-CODEを，それぞれENGLISH 1，CODE 1に格納する。つぎに同音異義語の有無を調べるために，ファイルをもう一件読む。そのレコード内のJAPANESEとACPが一致しなければ，ENGLISH 1とCODE 1に格納されている英語とコードが，TABLE 1とTABLE 2のBAN番目に格納される。またこのときACPに格納されていたカナ文字もTABLE 4のBAN番目に格納される。（以下，流れ図で説明）

4．処理部について

このプログラムでは品詞の判別はコードによって行なっている。しかし一つの文中に名詞，形容詞等が複数個含まれているとNとN，AとAという具合に判別ができない。このような場合はN 1，N 2，A 1とA 2とすれば判別が可能である。コードの後につけるこの数字をウェイトと呼ぶ。本節は，すべてのコードにウェイトをつけることから始まる。まずBANをゼロクリアする。BANに1を加算して，BANが1になる。コードが格納されているTABLE 2のBAN番目のコードは何であるかを調べる。もし主語となりうるNまたはPRONであったなら，WEIGHTに1を加算する。WEIGHTはゼロクリアされているので，0＋1＝1となる。NまたはPRONでなければWEIGHTへの加算は行なわない。再びBANに1が加算されBANが2となる。TABLE 2のBAN番目のコードは何であるかを調べる。もしNまたはPRONであれば，WEIGHTに1を加算する。WEIGHTをTABLE 3のBAN番目に格納する（以下，流れ図で説明）。各コードにウェイトを付け終ったならば，正しい英文になるように単語を並びかえる。まず主語が最初に置かれなければならない。主語となり得るものはNまたはPRONである。まずBANをゼロクリアする。BANに1を加算する。TABLE 2のBAN番目のコードが何であるかを調べる。NまたはPRONでなければBANに1を加算して，TABLE 2のBAN番目について調べる。もしNまたはPRONであれば今度はTABLE 3の，BAN番目のウェイトが何であるかを調べる。ウェイトが1であれば，そのNまたはPRONが主語であると判断する。そしてその主語はAまたはADVで修飾されているかを調べる。AはNまたはPRONの前に置かれてそれを修飾するのであるから，BANから1を減算してTABLE 2の（BAN－1）番目のコードを調べる。もしAでなければ，主語は修飾されていないものと考えて，主語となるNまたはPRONをTABLE 5の1番目に格納する。もしAであれば，そのAはさらにADVによって修飾されているかを調べる。ADVはAの前に置かれてそれを修飾するのであるから，BANから1を

減算してTABLE 2の(BAN－1)番目のコードを調べる。もしADVでなければ，Aは修飾されていないものと判断する。そして主語を修飾するAに対応するTABLE 1の英語をTABLE 5の1番目に，主語となるNまたはPRONに対応するTABLE 1の英語をTABLE 5の2番目に格納する。もしADVであれば，そのコードに対応するTABLE 1の英語をTABLE 5の1番目に，Aに対応するTABLE 1の英語をTABLE 5の2番目に，主語となるNまたはPRONに対応するTABLE 1の英語をTABLE 5の3番目に格納する。この1，2，3……番目を示す変数がCNTである。CNTは1から始まり，TABLE 5への格納があるごとに1づつ加算される。

動詞の格納について。　BANをゼロクリアしてから(BAN+1)をくり返しながら，TABLE 2のBAN番目のコードを調べる。VはAUXを伴っているかを調べる。日本語の文型では，AUXはVの後に置かれている。そのためBANに1を加算してTABLE 2の(BAN＋1)番目のコードを調べる。AUXでなければ，TABLE 1のBAN番目の英語を，TABLE 5のCNT番目へ格納する。もしAUXであれば，VとAUXの位置を入れかえてから，TABLE 5に格納する。その方法は，TABLE 1，2，3のBAN番目に格納されている英語，コード，ウェイトを，いったんW 1，W 2，W 3に移し，TABLE 1，2，3の(BAN＋1)番目に格納されている英語，コード，ウェイトをTABLE 1，2，3のBAN番目に格納する。そしてW 1，W 2，W 3に格納されていた英語，コード，ウェイトをTABLE 1，2，3の(BAN＋1)番目に格納する。これでAUXがVの前に入れかわる。その後，TABLE 1のBAN番目の英語をTABLE 5のCNT番目に，TABLE 1の(BAN＋1)番目の英語をTABLE 5の(CNT＋1)番目に格納する。

目的語について。　目的語になり得る品詞は，NまたはPRONである。(BAN+1)をくり返しながら，TABLE 2のBAN番目のコードを調べてゆく。もしNまたはPRONがあれば，TABLE 3のBAN番目のウェイトを調べてみる。ウェイトが1のNまたはPRONは，主語として使用されているので，この場合はウェイトが2のものを捜す。ウェイトが2のNまたはPRONが見つかれば，それを目的語と判断する。この目的語は，Aまたは ADVによって修飾されているかを調べる。Aは目的語の前に置かれて，それを修飾するのであるから，BANから1を減算して(BAN－1)番目のコードを調べる。もしAでないならば，目的語は修飾されていないものと判断して，今度はPREPを伴っているかを調べる。日本語の文型では，前置詞はNまたはPRONの後に置かれている。そのため，さきほど1が減算されたBANに1を加算して元に戻し，さらに1を加算して(BAN+1)とする。TABLE 2の(BAN＋1)番目のコードがPREPであるなら，PREPとNまたはPRONの位置を入れかえる。入れかえの方法は前述した方法による。そしてTABLE 1のBAN番目の英語を，TABLE 5のCNT番目に，TABLE 1の(BAN＋1)番目の英語を，TABLE 5の(CNT＋1)番目に格納する。もしPREPを伴っていない場合は，目的語だけが格納される。もしこの目的語がAによって修飾されている場合は，さらにAがADVによって修飾されているかを調べる。もしADVによって修飾されていなければ，Aに対応するTABLE 1の英語を，TABLE 5のCNT番目に，目的語をTABLE 5の(CNT＋1)番目に格納する。もしADVによって修飾されている場合は，そのコードに対応するTABLE 1の英語を，TABLE 5のCNT番目に，Aに対応するTABLE 1の英語をTABLE 5の(CNT＋1)番目に，目的語をTABLE 5の(CNT＋2)番目に格納する。目的語は，間接目的語と直接目的語として，1つの文中に2回含まれる場合がある。そのため，ウェイトが3のNまたはPRONについて，再度この処理をくり返す。さらに以上の処理の中に含まれなかったADVがあるなら，それはVを修飾するADVと判断して，そのコードに対応するTABLE 1の英語を，TABLE 5の文末に格納する。以上で英単語の並びかえは終了する。処理結果はプリンタで入力データ（日本語）と共に印字する。

5．結論

このプログラムを実行させた結果，文型1から

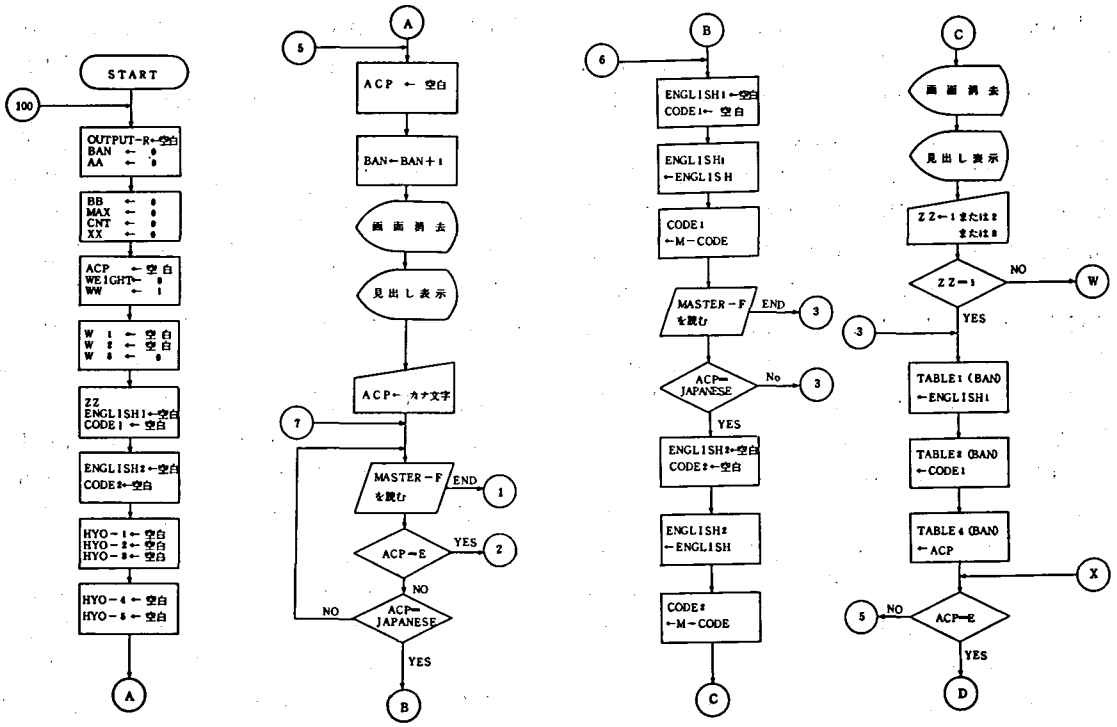
文型5の文例の翻訳は成功した。従って本論で述べた方法論の範囲内の文であれば正しく翻訳ができる。逆に，予め用意されたファイルと構文に合致する文でなければ翻訳はできない。即ち，大きなファイルと複雑な文法情報が用意されなければこれらの問題は解決できない。

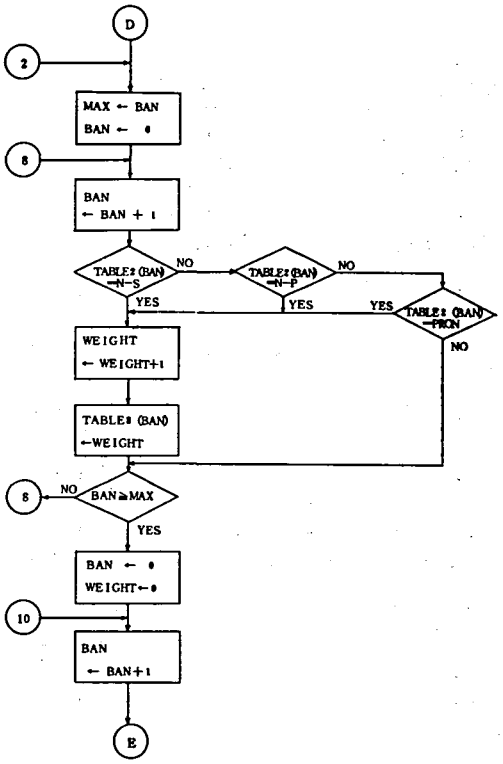
（謝辞）　本稿の作成にあたり，小林誠氏に多大なる御協力をいただきましたことを感謝いたします。

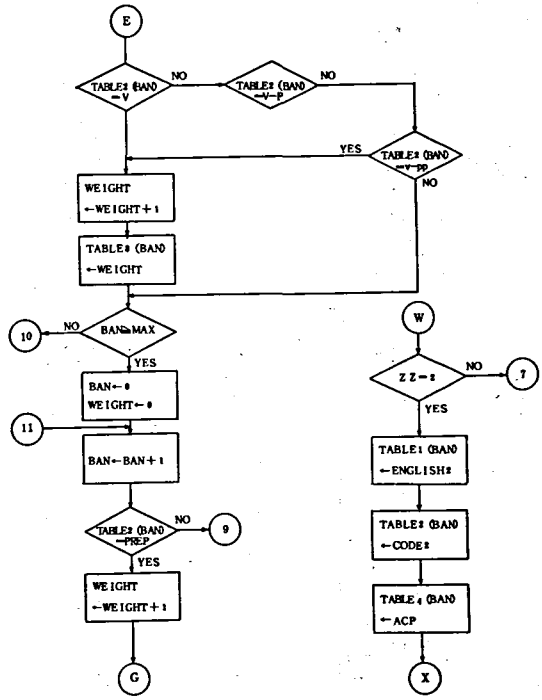
（文献）

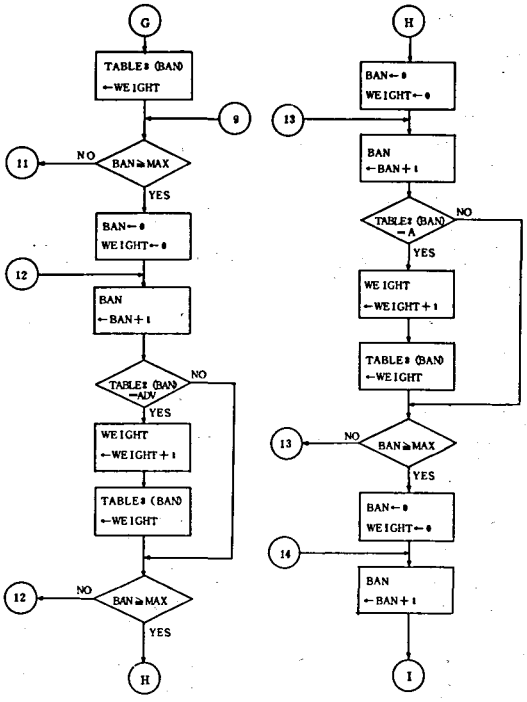
1）山崎貞：新自修英文典，研究社（1972）
2）長尾真：着実に進展している機械翻訳，科学朝日（Oct. 1980）

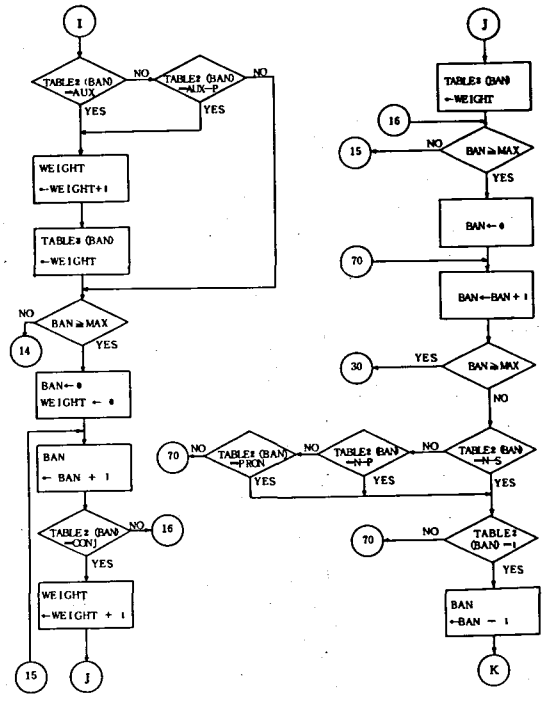
〔処理部の流れ図〕

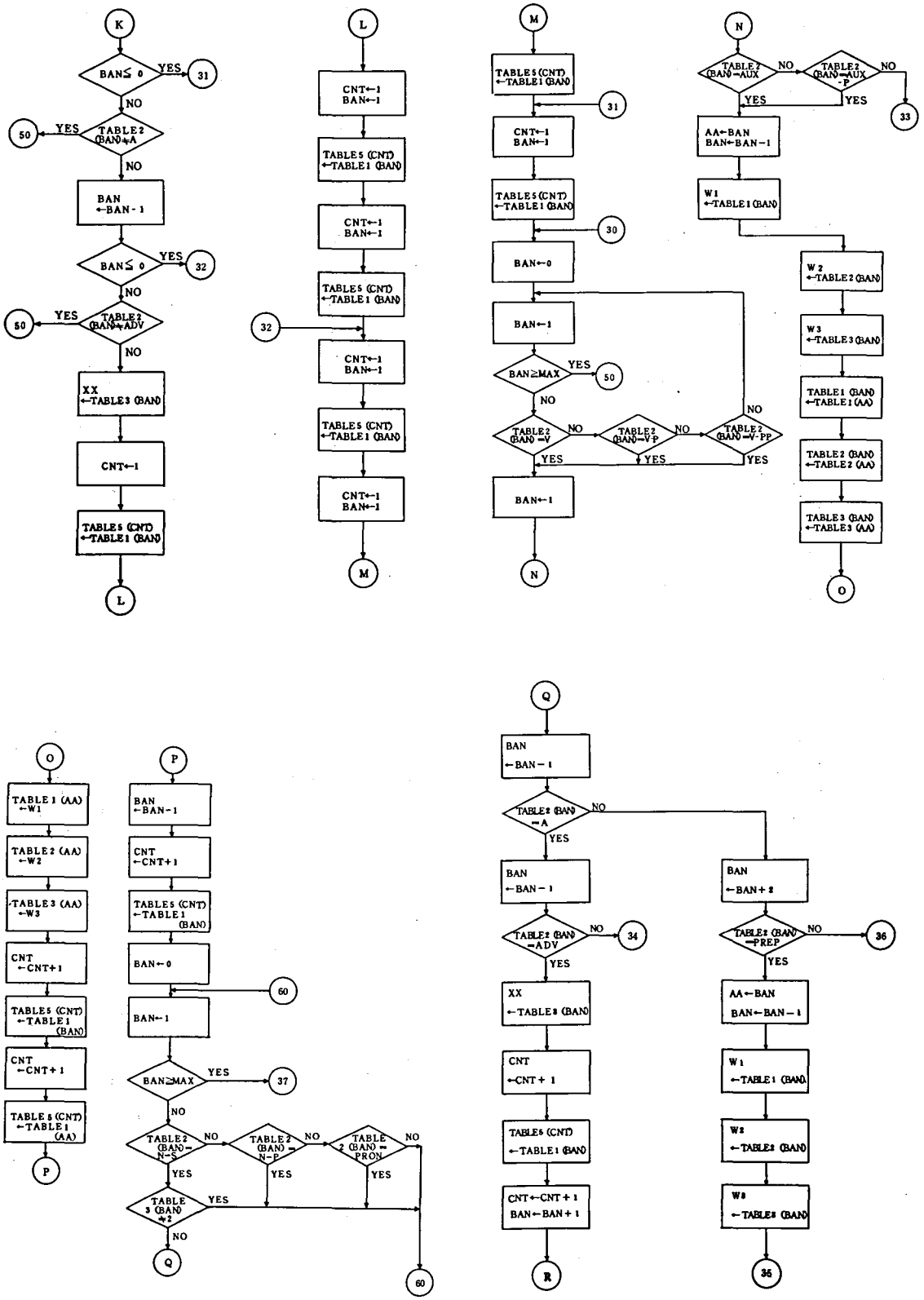


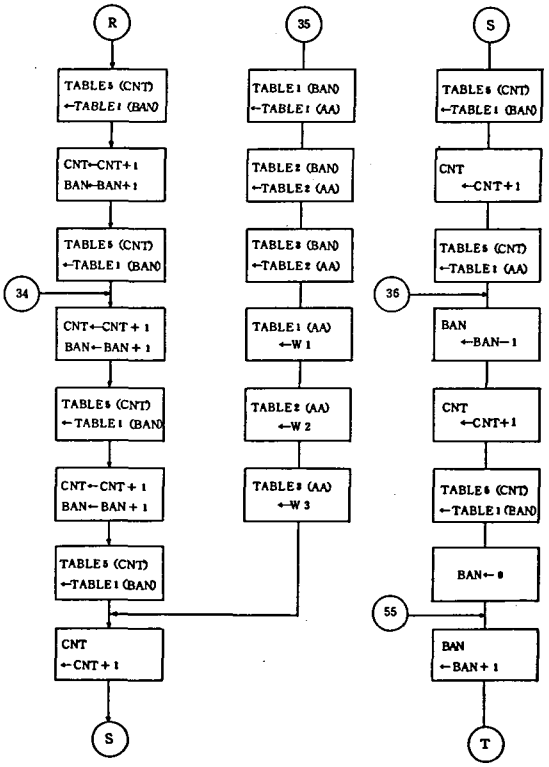

D
2
MAX ← BAN
BAN ← 0
8
BAN
← BAN + 1
TABLE# (BAN)
=N-S
NO
TABLE# (BAN)
=N-P
NO
YES
YES
YES
TABLE# (BAN)
=PRON
NO
WEIGHT
← WEIGHT+1
TABLE# (BAN)
←WEIGHT
8
NO
BAN≧MAX
YES
BAN ← 0
WEIGHT←0
10
BAN
← BAN+1
E

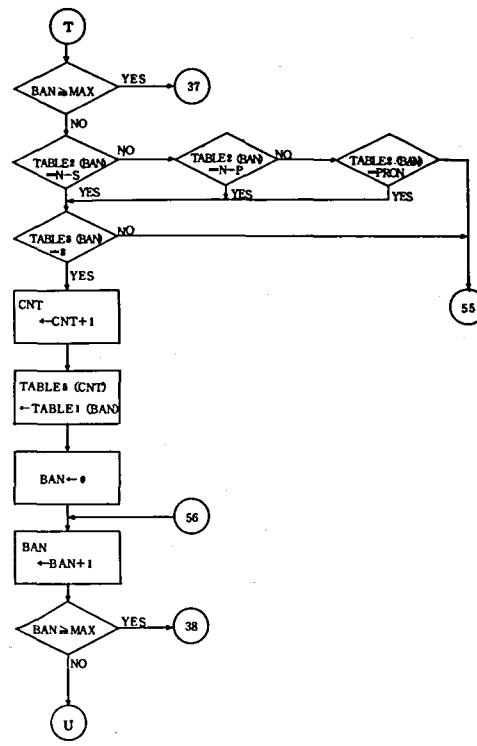

E
TABLE# (BAN)
=V
NO
TABLE# (BAN)
=V-P
NO
YES
TABLE# (BAN)
=V-PP
NO
WEIGHT
←WEIGHT+1
TABLE# (BAN)
←WEIGHT
10
NO
BAN≧MAX
YES
BAN←0
WEIGHT←0
11
BAN←BAN+1
TABLE# (BAN)
=PREP
NO
9
YES
WEIGHT
←WEIGHT+1
G
W
ZZ=#
NO
7
YES
TABLE1 (BAN)
←ENGLISH#
TABLE# (BAN)
←CODE#
TABLE4 (BAN)
←ACP
X

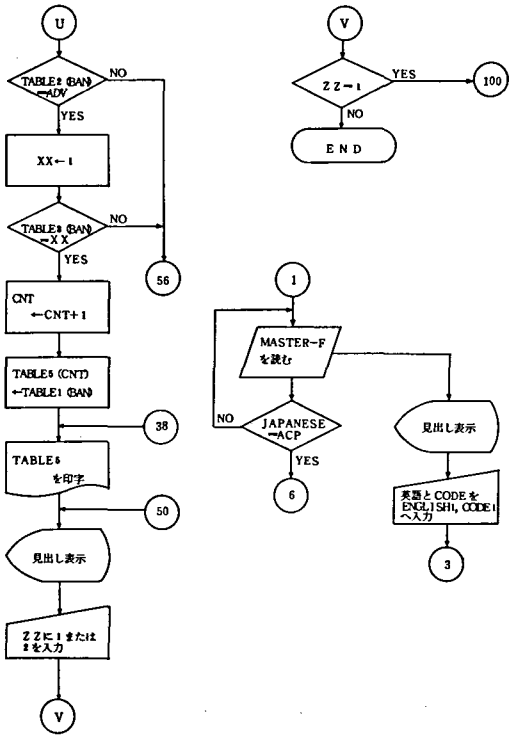

G
TABLE# (BAN)
←WEIGHT
9
11
NO
BAN≧MAX
YES
BAN←0
WEIGHT←0
12
BAN
←BAN+1
TABLE# (BAN)
=ADV
NO
YES
WEIGHT
←WEIGHT+1
TABLE# (BAN)
←WEIGHT
12
NO
BAN≧MAX
YES
H
H
BAN←0
WEIGHT←0
13
BAN
←BAN+1
TABLE# (BAN)
=A
NO
YES
WEIGHT
←WEIGHT+1
TABLE# (BAN)
←WEIGHT
13
NO
BAN≧MAX
YES
BAN←0
WEIGHT←0
14
BAN
←BAN+1
I


I
TABLE# (BAN)
=AUX
NO
TABLE# (BAN)
=AUX-P
NO
YES
YES
WEIGHT
←WEIGHT+1
TABLE# (BAN)
←WEIGHT
NO
14
BAN≧MAX
YES
BAN←0
WEIGHT ← 0
BAN
← BAN + 1
TABLE# (BAN)
=CONJ
NO
16
YES
WEIGHT
←WEIGHT + 1
15
J
J
TABLE# (BAN)
←WEIGHT
16
15
NO
BAN≧MAX
YES
BAN←0
70
BAN←BAN+1
30
YES
BAN≧MAX
NO
70
NO
TABLE# (BAN)
=PRON
NO
TABLE# (BAN)
=N-P
NO
TABLE# (BAN)
=N-S
YES
YES
YES
70
NO
TABLE# (BAN)
=I
YES
BAN
←BAN − 1
K


K
BAN≦0
YES
31
NO
50
YES
TABLE2 (BAN)=A
NO
BAN ←BAN-1
BAN≦0
YES
32
NO
50
YES
TABLE2 (BAN)=ADV
NO
XX ←TABLE3 (BAN)
CNT←1
TABLE5 (CNT) ←TABLE1 (BAN)
L
L
CNT←1 BAN←1
TABLE5 (CNT) ←TABLE1 (BAN)
CNT←1 BAN←1
TABLE5 (CNT) ←TABLE1 (BAN)
32
CNT←1 BAN←1
TABLE5 (CNT) ←TABLE1 (BAN)
CNT←1 BAN←1
M
M
TABLE5 (CNT) ←TABLE1 (BAN)
31
CNT←1 BAN←1
TABLE5 (CNT) ←TABLE1 (BAN)
30
BAN←0
BAN←1
BAN≧MAX
YES
50
NO
TABLE2 (BAN)=V
NO
TABLE2 (BAN)=V-P
NO
TABLE2 (BAN)=V-PP
NO
YES
YES
YES
BAN←1
N
N
TABLE2 (BAN)=AUX
NO
TABLE2 (BAN)=AUX -P
NO
YES
YES
33
AA←BAN BAN←BAN-1
W1 ←TABLE1 (BAN)
W2 ←TABLE2 (BAN)
W3 ←TABLE3 (BAN)
TABLE1 (BAN) ←TABLE1 (AA)
TABLE2 (BAN) ←TABLE2 (AA)
TABLE3 (BAN) ←TABLE3 (AA)
O
O
TABLE1 (AA) ←W1
TABLE2 (AA) ←W2
TABLE3 (AA) ←W3
CNT ←CNT+1
TABLE5 (CNT) ←TABLE1 (BAN)
CNT ←CNT+1
TABLE5 (CNT) ←TABLE1 (AA)
P
P
BAN ←BAN-1
CNT ←CNT+1
TABLE5 (CNT) ←TABLE1 (BAN)
BAN←0
60
BAN←1
BAN≧MAX
YES
37
NO
TABLE2 (BAN)=N-S
NO
TABLE2 (BAN)=N-P
NO
TABLE2 (BAN)=PRON
NO
YES
YES
YES
TABLE3 (BAN)=Z
YES
NO
Q
60
Q
BAN ←BAN-1
TABLE2 (BAN)=A
NO
YES
BAN ←BAN-1
TABLE2 (BAN)=ADV
NO
34
YES
XX ←TABLE3 (BAN)
CNT ←CNT+1
TABLE5 (CNT) ←TABLE1 (BAN)
CNT←CNT+1 BAN←BAN+1
R
BAN ←BAN+2
TABLE2 (BAN)=PREP
NO
36
YES
AA←BAN BAN←BAN-1
W1 ←TABLE1 (BAN)
W2 ←TABLE2 (BAN)
W3 ←TABLE3 (BAN)
35


R
TABLE5 (CNT)
←TABLE1 (BAN)
CNT←CNT+1
BAN←BAN+1
TABLE5 (CNT)
←TABLE1 (BAN)
34
CNT←CNT+1
BAN←BAN+1
TABLE5 (CNT)
←TABLE1 (BAN)
CNT←CNT+1
BAN←BAN+1
TABLE5 (CNT)
←TABLE1 (BAN)
CNT
←CNT+1
S
35
TABLE1 (BAN)
←TABLE1 (AA)
TABLE2 (BAN)
←TABLE2 (AA)
TABLE3 (BAN)
←TABLE3 (AA)
TABLE1 (AA)
←W1
TABLE2 (AA)
←W2
TABLE3 (AA)
←W3
S
36
55
T


T
BAN≧MAX
YES
37
NO
TABLE2 (BAN)
=N-S
NO
TABLE2 (BAN)
=N-P
NO
YES
YES
YES
TABLE3 (BAN)
=0
NO
YES
55
CNT
←CNT+1
TABLE5 (CNT)
←TABLE1 (BAN)
BAN←0
56
BAN
←BAN+1
BAN≧MAX
YES
38
NO
U


U
TABLE2 (BAN)
=ADV
NO
YES
XX←1
TABLE3 (BAN)
=XX
NO
YES
56
CNT
←CNT+1
TABLE5 (CNT)
←TABLE1 (BAN)
38
TABLE5
を印字
50
見出し表示
ZZに1または
2を入力
V
V
ZZ=1
YES
100
NO
END
1
MASTER-F
を読む
NO
JAPANESE
=ACP
YES
6
見出し表示
英語とCODEを
ENGLISH, CODE1
へ入力
3