

国立国語研究所学術情報リポジトリ

漢字かなまじり文を全文カナ書き・ローマ字書きに 変換するシステムについて

メタデータ	言語: Japanese 出版者: 公開日: 2017-03-31 キーワード (Ja): キーワード (En): 作成者: 田中, 章夫, TANAKA, Akio メールアドレス: 所属:
URL	https://doi.org/10.15084/00001000

漢字かなまじり文を全文カナ書き・ローマ字書きに変換するシステムについて

田 中 章 夫

0. はじめに

0. 1. 日本語の言語データを電子計算機で処理する場合、最大の障害は、やはり漢字である。

これを、漢字OCR（光学的文書読み取り装置）や、漢字ラインプリンターの開発など、入出力機器の面から機械的に解決していくのも一つの道には違いないが、情報当りのコストや処理スピードからみて、まだまだ手軽に使える段階ではない。

そのため一般には、オンラインのレベルでの処理は、あきらめてしまって、漢字テレタイプや漢字テレファックスにたよらざるをえない現状である。

そこで、考えられるのは、この能率の悪い入出力機器をなるべく使わないですませ、そのかわりに、情報当りのコストが安く、しかも処理スピードの最も早い電子計算機を用いて、入出力を肩がわりさせていく方式である。さきに述べた機械的な解決に対して、プログラム的な解決とでも言うべきものである。

日本語の言語データを、電子計算機で処理してみると、必ずしも、そのすべてに漢字出力が必要なわけではない。特に、処理の中間段階においては、漢字出力でなくてもよいから、急いでアウト・プットしたいという場合が、きわめて多い。ところが、こうした場合も、電子計算機が、一旦、漢字かなまじりの形式でデータを記憶している以上、やはり印字スピードのおそい漢字テレタイプなどの世話をしなくてはならない。こういう場合、計算機が自分で、漢字まじりデータを、カナやローマ字の形になおして、ラインプリンターに打ち出してくれたら、たいへん便利である。

また、その逆に、入力は、すべて、カナやローマ字で、すませておいて、出力に漢字まじり文を必要とする場合には、電子計算機自身が、漢字まじりデータを作ることができれば、入力面で漢字テレタイプのキイをたたく必要がなくなってしまう。

すなわち、前者においては、電子計算機が自分で漢字を解読して、データの中の漢字の音や訓を決定していくわけであり、後者においては、データの中のカナやローマ字を、つぎつぎに適切な漢字におきかえていくわけである。前者は、いわば、漢字のヨミガナつけであり、後者は書き取りの答案をつくっていくことにはかならない。これを「漢字＝カナ」あるいは「漢字＝ローマ字」の相互変換方式のプログラム・システムという。

このプログラム・システムは、なにも、漢字の入出力の面からの必要性にこたえるばかりではない。電子計算機で漢字まじりのデータを扱う場合、そこに、一つの大きな問題として、同表記別語・異表記同語の問題がある。電子計算機は、データの表記が違えば、別語とし、表記が同じならば同一語として処理してしまう。たとえば、「工夫」という表記は、それが、「こうふ」という語であっても、「くふう」という語であっても、全く区別せずに同じ語として処理してしまう。もし、これを区別する必要があるなら、電子計算機に入る前に「工夫（クフウ）」「工夫（コウフ）」というように、情報を与えておかなくては、このような同表記別語の処理はしてくれない。また、この逆に、同じ語であっても「桜／サクラ／さくら／SAKURA」は全く別の語として処理されてしまう。このような異表記同語の場合は、原データに「桜（サクラ）／サクラ（サクラ）／さくら（サクラ）／SAKURA（サクラ）」のように、一定の情報をつけなくては同一語として処理してもらえない。

こうした情報つけも、現在のところでは、すべて人間が行なわなくてはならない。そのため、用語索引の作成や辞書の見出しの整理あるいは語彙調査などをはじめ、データの照合・検索をともなう場合には、この入力前の手作業に、多くの時間と労力が費されてしまう。

漢字＝カナ（漢字＝ローマ字）相互変換システムの目的の一つは、この前処理（プレ・エディッティング）の自動化である。これは、入出力機械が、いかに

改善されても、漢字かなまじりの日本語データを電子計算機で処理する場合には、やはり無視することが出来ない。

0. 2. 漢字＝カナ（漢字＝ローマ字）相互変換システムとは言っても、その基本は、各々の漢字について、漢字とヨミ（音訓）との間の関係法則を確定していくことがある。それには、個々の漢字が持っているいくつかのヨミの中から、ある特定な一つのヨミが、文脈にしたがって選定されていくプロセスを明らかにしておかなくてはならない。これは、各漢字について、その選定過程のアルゴリズム（演算規則）を確定しておくことにはかならない。したがって、漢字＝かな（漢字＝ローマ字）相互変換システムの出発点は、漢字の解読すなわち、ヨミガナつけのシステムの開発にある。この観点にたって、「漢字解読システム」を作成した。そして、現在は、そのつぎの段階として、以下に述べるような「漢字のかな・ローマ字への変換システム」すなわち漢字かなまじりのデータを、全文かな書き、あるいは、全文ローマ字書きとして、ラインプリンターに出力するシステムの作成にとりかかっている。

こうした漢字処理のシステムを作成するということは、前述のように、各漢字のヨミの決定過程を追究するものであるから、単に、電子計算機のプログラム上の問題にとどまらず、漢字をめぐるさまざまな問題について研究の糸口を与えると同時に、従来の漢字についての研究や資料に大きく依存するものもある。特に、漢字とヨミガナの間にある一定の関係法則、各漢字の個々のヨミガナの使用頻度、各漢字の現われる文中の場面的な問題、（たとえば、送りがな）あるいは、個々の漢字の個々のヨミガナについて、その選定過程の難易（言いかえれば漢字の読み方のむずかしさ）の測定などは、すぐに浮かび上ってくる問題である。

1. システムのあらまし

1. 0. 基本的な考え方

このシステムの基本的な考え方は、漢字かなまじり文を、全文カナ書きあるいは全文ローマ字書きに変換していく場合、その変換のしかたが、原文に漢字

が連続して現われる部分と、漢字が単独で孤立して現われる部分とでは、はっきり異なってくるというところから出発している。言いかえれば、連続して使われている漢字と、単独で孤立して使われている漢字とでは、そのヨミ（音訓）の選ばれ方に、はっきり異なった傾向がみられるという点に着目したシステムである。

連続して現われる漢字は、どちらかというと、音（オン）読みされる確率が高い。それに対して、単独で孤立して現われる漢字は、訓で読まれる傾向が強いという前提にたって、漢字の解読すなわち、カナ・ローマ字への変換を進めていくものである。

1. 1. 処理の目標

このシステムの、さしあたっての目標は、国立国語研究所の漢字テレタイプに収容してある2110字の漢字（盤内漢字）が、データの中に現された場合、それを適切なヨミ（音訓）に変えていくことである。したがって漢字テレタイプに収容しない漢字（盤外漢字）にヨミを与えることは、考えていない。

また、盤内漢字ではあっても、人名・地名など、固有名詞として使われた場合、これを適切に変換することは、目ざしていない、「大阪（おおさか）」「奥羽（おおう）」など、試験的に試みた特例をのぞいては、ヨミの選定方法について、特別な処置をしていない。したがって、固有名詞についても、一般的のデータの場合と同じ扱いで、何らかのヨミガナに変換はするが、正解は期待しえない。

連濁・連声・音の交替・あて字・熟字訓など、特殊なヨミは、漢字テーブル（辞書）に余裕がある場合に、いくつかの漢字について試験的に試みた程度で、本格的な処理は、目ざしていない。したがって、一般には、連濁が清音のままのヨミですまされたり、「煙草（えんそう）」というような結果になることが多い。

今回、試験的に認めた特殊なヨミの例をあげると、つぎのようなものである。

端（ばた）・帆（ぼ）・舶（ぱく）・般（ぱん）・応（のう）・縁（ねん）

・教（おそ）・酒（さか）・稻（いな）・上（うわ）・迷（まい）・勝（かっ）・卒（そつ）・弟（で）・九（ここの）・伯（お）・期（ご）・除（じ）
・星（じょう）・主（ず）・載（だい）・納（なっ・なん）……

連濁・連声などの処理としては、たとえば「舶（はく）」や「応（おう）」を前の音の関係から、その都度「舶（ぱく）」「応（のう）」に変化させる方法も、ないではない。しかし、この方法は、きわめて能率が悪いので、今回のシステムでは、「舶」については、「はく」とは別に「ぱく」を独立の読みとして扱ってしまっている。全く同様に、「教わる」「卒直」などの「教」「卒」を処理するために、さきにあげたように「教（おそ）」「卒（そつ）」などのヨミガナを用意した。また、先の例に「期（ご）」「除（じ）」「納（なっ・なん）」など、風変わりなヨミガナがあるのは、「最期」「掃除」「納豆」「納戸」などの処理を試みるためである。

つぎに、ヨミガナと送りがなとの接続面についてであるが、これも、まだ本格的な処理を目指してはいない。漢字テーブル（辞書）にヨミガナを収容していくさいには、頻度の高い、もっとも一般的な送りがなにもとづいて、ヨミガナの形を決めていった。しかし今回のシステムの中には、データの送りがなにしたがって、ヨミガナの形を調整する機能は、試験的にやってみたいいくつかの漢字の場合をのぞいては、そなわっていない。したがって、実際のデータの処理においては、送りがなと一部分がダブったヨミガナ、その逆に、送りがなにつながらない「不足分のあるヨミガナ」などが現われることは避けられない。送りがなとヨミガナとの接続面の処理としては、たとえば「届」のヨミガナが、「欠席届」のときは「とどけ」、「届け出」は「とど」となるようにした例や、「語」のヨミガナが、「語る」「語り合う」などのときは「かた」、「物語」などのときは「かたり」となるようにした例など、いくつかの漢字について試みたにすぎない。

1. 2. システムの内容

このシステムの一つの特徴は、データの漢字かなまじり文の中に現われる漢字について、その漢字と前後の1字、すなわち、3字の範囲内から得られる情

報だけを手がかりとして、漢字のヨミを決定していくところにある。プログラミ的に表現すれば、データの中の、ある漢字をPとすると、その前の文字は $(P - 1)$ 、そのつぎの文字は $(P + 1)$ だから、 $(P - 1) \leqq (P + 1)$ の範囲ということになる。この $(P - 1) \leqq (P + 1)$ の範囲すなわち、漢字Pとその前後の1字からなる3文字の文字列のことを、以下「漢字Pの環境」と呼ぶことにする。

では、この「漢字Pの環境」から、どんな情報を得ていくかというと、それは、つぎの2種類である。

その一つは、字種についての情報である。漢字Pの前あるいは後、すなわち $(P - 1)$ あるいは $(P + 1)$ が、漢字であるか、漢字でないかという情報である。これによって、1.0.にのべたように、漢字Pが、単独で現われているか、他の漢字とつながって現われているかがわかるわけである。

もう一つは、漢字Pの前後に、どんな文字が現われているかという情報である。「漢字Pの環境」から得られる上記二つの情報、すなわち、字種についての情報と、文字についての情報にもとづいて、漢字Pのヨミがなを決めていくのである。

つぎに、漢字のヨミガナ変換は、盤内漢字2110字の各漢字について、1字当たり最大6個のヨミガナを収容してある漢字テーブル（磁気テープ）とのコレートによって行なう。漢字テーブルは、いわば一種の漢字辞典であり、電子計算機に、この辞書すなわち漢字テーブルをひかせていく。

漢字テーブルは、辞書とはいっても、単に漢字と、そのヨミガナだけを收めているわけではない。漢字テーブルには、一つ一つの漢字について、ヨミガナとともに、つぎの3種の情報が入れてある。

第1は、さきに述べた「漢字Pの環境」から得られる情報（字種情報・文字情報）を使用してヨミガナを決定していくか、使わずに決定していくかを示す情報である。字種情報や文字情報を、全く使わずにヨミガナが決められるということは、辞書をひいただけでヨミガナが決められる場合であり、これは、ヨミが、一種類しか存在しない漢字の場合である。

第2は、漢字Pの環境から得られる字種情報のみにもとづいて、いくつかあ

るヨミガナのうちから一つを選択させるための情報である。

これは、言いかえれば、漢字Pの前後に、漢字が、どのように現われているかを、電子計算機に調べるように指示するための情報であり、その結果にもとづいて、ヨミガナを選ばせるわけである。

第3は、漢字Pの前または後に、特定な文字が出てくるかどうかを、電子計算機に調べさせるための情報である。その結果、指定した文字が存在していれば、ただちに、ヨミガナを決定するが、それが存在していない場合には、上に述べた第2の「字種情報」にもとづいてヨミガナを決定することになる。

したがって、漢字の変換は、すべて、この漢字テーブルに入れてある上記三種類の情報によって進んでいくわけである。第1の情報は、いま解読しようとしている漢字の処理のしかたを指示する情報（処理表示）であり、第2の情報は、その漢字の環境についての、演算のやり方すなわち性格の決め方を指示するための情報（演算表示）である。そして、第3の情報は、その漢字が、特定な環境、すなわち特定の文字の前や後に現われている場合の、いわば特例的な処理を指示する情報（環境表示）である。

以上述べたように、データの漢字について、漢字テーブル（辞書）をひきさえすれば、あとは、漢字テーブルが、解読のプロセスを完全にコントロールしてしまう。この点も、このシステムの一つの特徴といえよう。

2. システムの構成

2. 0. 全体の流れ

「漢字→かなローマ字変換システム」は、つぎにあげる7つのプロセスに、大別することができる。

- A) 入力選択過程
- B) 0 1 変換過程
- C) 字種ふり分けルーチン
- D) 盤内漢字処理過程

- E) 盤内非漢字コード処理過程
- F) 盤外コード処理過程
- G) 出力選択過程

上の(D) 盤内漢字処理過程は、さらに

- D₁漢字テーブル・コレート
- D₂転写回路
- D₃環境演算回路
- D₄指定環境処理回路

の4つに分れる。

以上について、ゼネラル・フロー・チャートを示すと、表1の通りである。

2. 1. 入力選択過程（表1のA）

このシステムの、入力関係については、漢テレ・コードによるデータならどんな形のものでも、なるべく巾広く扱えるように、種々の制限は、できるかぎり、はずしてある。

まず、入力方式としては、紙テープ、磁気テープのどちらでも、入力が可能なようにしてある。そして、その選択は、電子計算機のインターラプト・ボタンによって行なう。紙テープで入力する場合は、インターラプト・ボタンをOFFとし、磁気テープの場合はONとすることにより、入力方式は、どちらにも定めることができる。

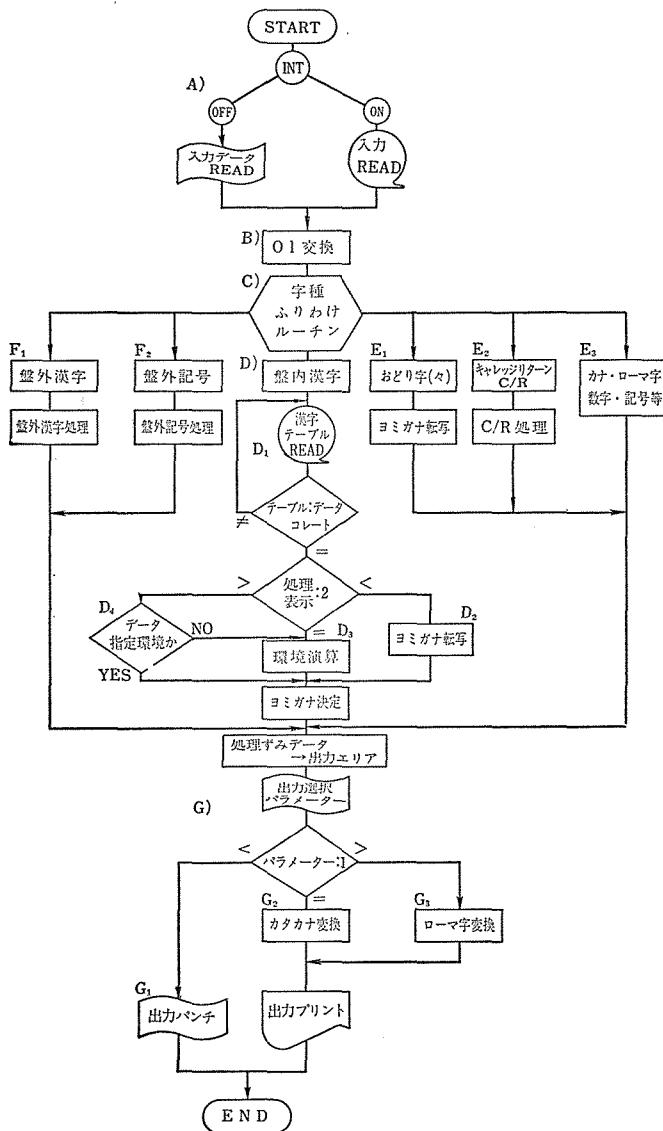
入力データの内容については、つぎの2条件さえ、そなえていれば、固定長のデータでも、可変長のデータでもよい。

○データの長さが、1ブロック（1バッチ）当り、漢テレ2000字（4000ch）以下であること。

○データ末に、紙テープ入力の場合は「=E/F」が1ブロック、磁気テープ入力の場合は、「E/F」「E/D」が各1バッチ存在すること。

したがって、レコード末またはブロック末（バッチ末）についてのフォーマット上の制限は、まったくない。しかし、これらの個所にエンド・マーク（=E/I, =E/F, E/I, E/Fなど）が存在しているときは、そこまでを一連

表1 全体の流れ図



のものとして扱かう。エンド・マークが存在していない場合は、後続のレコードあるいはブロック（バッチ）と連続するものとして扱かっていく。

2. 2. 「01変換」過程(B)

漢字かなまじり文のデータが、読みこまれると、データの中の漢字を1とし、それ以外を0におきかえた変換表が、中央記憶装置（メモリー）の中に、作成される。つぎの例文のようなデータが読みこまれると、その下に示すような形の変換表が作られる。

データ 変換	⑤我々の大学は、この町から約20km西の山のふもとにある。 0 1 1 0 1 1 0 0 0 0 1 0 0 1 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0
-----------	--

データ 変換	⑥(¶MOポン発UP I=共同)西ドイツの財界人と接触しつつ、来C/R 0 0 0 0 0 0 0 1 0 0 0 0 1 1 0 1 0 0 0 0 1 1 1 0 1 1 0 0 0 0 1 0
-----------	--

月末◆定学滞在する予定の経済使節団は、……… 1 1 1 1 1 1 0 0 1 1 0 1 1 1 1 1 0 0 ………

例文の中の「¶MO」は、漢テレ盤内にない記号（盤外記号）を表わすものであり、「◆定学」は、盤外漢字の「迄」を表わすコードである。また、「我々」の個所でわかるように、このプロセスにおいては、「おどり字（々）」は漢字扱いとなり、「1」に変換される。

この変換表は、さきに、1. 0.において述べた「漢字が連続して現われているか、単独で孤立して現われているか」を示すものに、ほかならない。すなわち、「漢字Pの環境」において、(P-1) P (P+1) の3文字列の変換値が、111や110または011ならば、漢字がつながっているわけであり、変換値が010ならば「漢字P」は単独で現われているということになる。

2. 3. 字種ふり分けルーチン(C)

このルーチンは、データの中の文字を、漢テレ・コードにしたがって、つぎの6種類に類別し、それぞれの処理過程に送り出すルーチンである。

盤内漢字／盤外漢字／盤外記号／おどり字(々)／キャレッジ・リターン
(改行復改コード)／その他

最後の「その他」は、出力のプロセスまでは、まったく処理をする必要のないコードで、カタカナ・ひらがな・ローマ字・算用数字・盤内記号・スペースなどである。これらは、このルーチンから、そのまま「出力エリア」に送り出される。

2. 4. 盤内漢字処理過程(D)

2. 4. 1. 漢字テーブル・コレート (D₁)

データの中に盤内漢字が現われると、ただちに「漢字テーブル・テープ(辞書)」を読みこみ、テーブルの中の漢字と、データに現われた漢字とのつき合わせ、すなわち「辞書引き」を行なう。

盤内漢字2110字は、すべて、表2に示す3種のフォーマットの中の、いずれかの形で、漢字テーブル・テープ(磁気テープ)に収容してある。辞書引き(コレート)の結果、データの中に現われた漢字が、処理表示1(表2のア)のものであれば、その漢字の処理は、「転写回路(D₂)」にまわされる。また、処理表示2(表2のイ)の場合は「環境演算回路(D₃)」に、処理表示3(表2のウ)の場合は「指定環境処理回路(D₄)」に、それぞれ、処理がうけつがれていく。

2. 4. 2. 転写回路 (D₂)

表2の(ア)のフォーマットになっている漢字は、ヨミガナが、1種類しか存在しない漢字であり、この場合の処理表示は1になっている。したがって、データに現われた漢字と、漢字テーブルとのコレートすなわち辞書引きの結果、その漢字が、処理表示1のものであれば、ヨミガナは決定したことになる。この場合には、漢字テーブルに収めてあるヨミガナが出力エリアに転写され、処理

表2 漢字テーブルテープのフォーマット

(ア)	
	例:
(イ)	
	例:
(ウ)	
	例:

(Q: 環境演算表示記号 / R: ヨミガナ番号 / S: 指定環境表示記号)

は完了する。表2のアに示した「院」の例で言えば、データの中に、この漢字が出てきたときは、ヨミガナ「いん」が、そのまま出力エリアに移されるわけである。

この回路で処理する漢字は「院・党・族・宇・批……」など、約700字である。

2. 4. 3. 環境演算回路 (D₈)

さきに、2. 2において述べたように、「0 1 変換過程」によって、データの中の漢字は1、漢字以外は0とした変換表が作られている。この変換の結果、データの中に出でた漢字Pの環境は、表3の4通りの中のいずれかになっている。

表3

P - 1	P + 1	
0	0	前後とも非漢字
0	1	後の字のみ漢字
1	0	前の字のみ漢字
1	1	前後とも漢字

この $(P-1)$ と $(P+1)$ の0や1を、プール代数によって演算し、その演算結果にしたがって、ヨミガナを選定するのが、「環境演算回路」である。

表2の(イ)に示すフォーマットで、漢字テーブルに収容してある漢字は、その処理表示が、2となっている。そして、各ヨミガナのあたまに、演算表示記号(R)が、1ケタ(1ch)についている。

辞書引きによって、データに現われた漢字Pが、処理表示2である場合には、演算表示記号にしたがって、漢字Pの環境すなわち $(P-1)$ と $(P+1)$ について、論理演算を開始する。そして、その演算結果が0のときに、そのヨミガナが採用され、ヨミガナは出力エリアに写される。演算結果が1の場合には、そのヨミガナは不採用となり、つぎのヨミガナについての演算に移る。

$(P-1)$ と $(P+1)$ についてのプール代数による演算結果は、表4に示す16通りがあり得る。この表の上欄「a A・b B……h H」は、それぞれの場合の演算表示記号を示している。

表4 環境演算の結果

漢字の現われ方	漢字Pの環境		環境 演 算 の 結 果							
	P-1	P+1	A a	B b	C c	D d	E e	F f	G g	H h
前後トモナシ	0	0	0 1	0 1	0 1	0 1	0 1	0 1	0 1	0 1
後ダケアリ	0	1	1 0	0 1	1 0	0 1	1 0	0 1	1 0	0 1
前ダケアリ	1	0	1 0	1 0	0 1	0 1	1 0	1 0	0 1	0 1
前後トモアリ	1	1	1 0	1 0	1 0	1 0	0 1	0 1	0 1	0 1
漢字 1 :	0 : 非漢字		0 : ヨミガナ記入 / 1 : ヨミガナ無記入							

たとえば、漢字「互」について言えば、この漢字のテーブルは、表2(イ)に示したように、

互(aご)(Aたがい)

となっている。表4でわかるように、「演算表示記号A」は、“(P-1)と(P+1)とが、ともに0の場合すなわち「互」の前と後とがともに漢字でない場合にのみ、演算結果が0となり、ヨミガナ「たがい」が採用される”という演算を指示する。「演算表示記号a」は、その逆に、(P-1)と(P+1)、

のうちの、すくなくとも一方が1すなわち漢字の場合は、すべて、ヨミガナ「ご」が採用されることを示す表示である。したがって、漢字「互」は、単独で孤立して現われた場合のみ「たがい」のヨミガナが選ばれ、他の漢字と並んで出て来たときは、常に「ご」のヨミガナが選ばれるわけである。

演算表示Aが指示する演算は、言うまでもなく「論理和(ロジカル・オア)」であり、aは、その否定演算である。

テーブルが

崩 (c ほう) (C くず)

となっている漢字「崩」は、表4に示すように、(P-1)と(P+1)とがが、ともに漢字の場合、および(P+1)のみが漢字の場合にかぎって「ほう」のヨミガナが採用される。それ以外の場合のヨミガナは、すべて「くず」となる。したがって「明治天皇崩御」「土砂の崩壊」のような形で「崩」が使われているときは「ほう」のヨミガナが採用され、「山崩れ」「がけが崩れる」のような形で使われていれば、これらのヨミガナは、すべて「くず」となる。

漢字「模」のテーブルは、つぎのようになっている。

模 (F も) (f ぼ)

したがって、漢字「模」の前の文字だけが漢字の場合にのみ、「ぼ」のヨミガナが選ばれ、他の場合は、すべて「も」になる。その結果「大規模な」「模」は「ぼ」と処理され、「声帯模写」「飛行機の模型」「花を模して……」などの形で出て来た「模」は、すべて「も」として処理される。

また、漢字「板」を例にとると、これは、

板 (b ばん) (B いた)

の形で、テーブルに収容されている。

この字の場合は、「板」の前後がともに漢字のときと、(P-1)すなわちこの字の前だけが漢字のときとに限って、ヨミガナ「ばん」が採用される。その他の現われ方をした場合は、すべて「いた」のヨミガナになる。したがって「看板娘」「黒板ふき」のような形で「板」が出てくれば、「ばん」のヨミガナが出力エリアに送られ、「~~板~~板前さん」「杉の板を切る」などの形で出てきたときは、つねに「いた」のヨミガナが送られることになる。それゆえ、「杉

板を切る」「床板が抜ける」のように、前が漢字の場合は、「ばん」として処理されてしまう。また「求ム板金工」は、「いた」として処理してしまうこととなる。

テーブルが、つぎのようになっている「軽」の場合

軽 (c けい) (Cかる)

「けがは軽い」のように前後とも非漢字のとき、「身軽な人」のように前の字だけが漢字のときは、いずれも「かる」のヨミガナが採られる。それに対して「日本製軽自動車」のように前後とも漢字のとき、および「(軽金属) 関係」のように後だけが漢字のとき（前は記号「ペーレン」）は、すべて「けい」と処理される。そのため、「昔の軽業師」の「軽」は「けい」として、「※軽トラック」の「軽」は「かる」として扱われる結果になる。

つぎに、ヨミガナが、3個以上のものについて例をあげてみよう。

たとえば、漢字「尋」のテーブルは、

尋 (c じん) (Aたず) (Fひろ)

となっている。したがって、(P-1) と (P+1) が、ともに漢字でないときは、「じん」のヨミガナが採用され、(P-1) が漢字のときに限って「ひろ」と処理される。それ以外の場合は、すべて「じん」となる。

また、漢字「御」については、テーブルが

御 (b ぎょ) (g ご) (Aお) (hみ)

となっている。表4からわかるように、環境演算表示記号「h」は、すべての環境について、そのヨミガナは不採用となることを示す。したがってヨミガナ「み」は、テーブルには収容してあるが、現在のところ使用しないということになる。その他のヨミガナについては、「ぎょ」は、「制御する」「崩御。」のように、前の文字のみが漢字のときに採用され、「ご」は、「朝御飯」のように前後とも漢字のときと、「遠い御先祖様」のように後の文字のみが漢字のときの二つの場合に採用される。「お」は、前後が、ともに漢字でない場合にのみ採用される。そのため、「御手紙」などは「ご」となってしまう。

以上述べたように、環境演算回路は、テーブルの中の漢字がもっている一つ

一つのヨミガナについて、その環境演算表示にもとづいて演算を進め、データの中の漢字のヨミガナを決定していく回路である。したがって、環境演算表示は、表4にあげた16通りの演算方式（厳密に言えば、A～Hの8通りと、その否定とを合わせた9通り）の中から、そのヨミガナについての演算方式を指定するものであり、演算の結果、採用となったヨミガナは、そのまま出力エリアに移されていく。この回路のみで、ヨミガナが決定できる漢字は、盤内漢字2110字のうち、約500字である。

2. 4. 4. 指定環境処理回路 (D₄)

これは、データの中の漢字について、辞書引きを行なった結果、漢字テーブルの処理表示が、3となっている場合の処理をする回路である。処理表示3の漢字の場合、漢字テーブルは、表2の(4)のフォーマットになっている。このフォーマットの特徴は、漢字とヨミガナのあとに、最大4個の範囲内で、環境指定表示が、ついている点である。「環境指定表示」は、処理しようとする漢字の前後、すなわち(P-1)と(P+1)に、特定な文字が現われた場合の処理のしかたを示すものである。

たとえば、漢字「荷」のテーブルは、つぎのようになっている。

荷 (1 b か) (2 B に) * M重2 / M初2 / N担1 / N重1

上の*以下すなわち「M重2」「M初2」「N担1」「N重1」が、指定環境表示である。頭の「M」「N」は、それぞれ(P-1), (P+1)を示す。したがって「M」は、漢字「荷」の前を、「N」は後を指定するものである。二番目の「重」「初」「担」「重」は、その位置すなわちM・Nに現われる文字の指定である。言いかえれば、(P-1), (P+1)の内容である。最後の「1」「2」は、ヨミガナを指定するもので、この場合ならば、「1」は「か」、「2」は「に」である。

結局、「M重2」は、「荷」の前が、「重」ならば、ヨミガナ2すなわち「に」を選べという指示を与えるものであり、「N担1」は、後の文字が「担」ならばヨミガナ1すなわち「か」を選べという表示である。また「M初2」「N重2」は、それぞれ「初荷」「重荷」の形で、漢字「荷」が出現したとき

の処理を示すものである。それ以外の場合は、「 $b - B$ 」の環境演算で処理される。

つぎに、表2にあげた漢字「騒」のテーブルは、

騒 (1 c そう) (2 C さわ) * M物1 / Nぎ2

である。この例の指定環境表示「M物1」は、「騒」前の文字が「物」となっている場合の特別な処理を指定したものであり、「Nぎ2」は、後の文字が「ぎ」となっているときの処理を指示するものである。すなわち、「M物1」は、「物騒」の場合には、ヨミガナ2「そう」を採用せよという指示を、また、「Nぎ2」は、「大騒ぎ」「無駄騒ぎ」などの場合には、ヨミガナ2「さわ」を選べという指示を与えるものである。「騒」の前後の文字が、これに該当する場合には、それぞれ指定されたヨミガナを出力エリアに送り出す。これに該当しない場合には、ヨミガナの選定は、「c そう / C さわ」の環境演算にゆだねられる。

テーブルが

構 (1 a こう) (2 A かま) * Nえ2

となっている漢字「構」の場合、もし、この漢字を「 $a - A$ 」の環境演算だけで処理してしまうと、「構」が、他の漢字と並んで現われたときは、「気構え」も「心構え」も「門構え」も、すべて、「こう」のヨミガナで処理してしまう。これを救うのが、指定環境表示「Nえ2」である。この表示により、「構」の後の文字が、「え」である場合には、無条件にヨミガナ2すなわち「かま」を採用することになる。これ以外の場合に限って、「 $a - A$ 」の環境演算に移っていくわけである。

同様なものに、

国 (1 a こく) (2 A くに) * M御2 / N々2

位 (1 a い) (2 A くらい) * MS1 / M氣2

などがある。上の「N々2」は、後の文字が、オドリ字(々)である場合の処理を示すものであり、「MS1」の「S」は「0~9」の算用数字を一括し示す表示記号である。したがって「MS1」は、「位」の前の文字が、「算用数字である場合には、ヨミガナ1すなわち「い」を採用せよという指示である。

漢字「渋」の場合、テーブルは、

渋（1 h じゅう）（2 H しぶ）* M難1／N滞1

の通りである。表4に示したように「H-h」の環境演算では、Hは、すべての環境について、常にそのヨミガナが採用され、hは、そのヨミガナは環境のいかんにかかわらず絶対に採用されない。しかし、「渋」の場合は、「M難1」「N滞1」の指定環境表示があるために、「渋」の前が「難」のときと、後が「滯」のときとに限って、ヨミガナ1すなわち「じゅう」が採用される。それ以外の場合は、環境演算にもちこまれ、その結果は常に「H」のヨミガナすなわち「しぶ」が採用されることとなる。

ヨミガナが三つ以上のものとしては、たとえばつぎのようなものがある。

逃（1 c とう）（2 Aに）（3 Fのが）* Nげ2／Nれ3

由（1 h ゆ）（2 b ゆう）（3 A よし）（4 g ゆい）* M経1／N来1

色（1 d ショク）（2 g しき）（3 C いろ）* M異1／M特1／M景

2／N々3

率（1 b りつ）（2 g そつ）（3 A ひきい）（4 h そつ）* M引4／
M統4／M軽4

以上述べたように、指定環境処理回路は、指定環境指示表の指示にしたがって特定の環境におけるヨミガナの決定を、優先的に扱う回路である。この回路で扱われる漢字、すなわち処理表示3の漢字の数は、約900字である。

2. 4. 5. 以上、述べたように、このシステムでは、ヨミガナが一種類の盤内漢字については、漢字テーブルからのヨミガナの転写によって、また、2種以上のものについては、環境演算あるいは「指定環境処理と環境演算の併用」によって、データの中の盤内漢字には、すべて、なんらかのヨミガナが、選定される。こうして選定されたヨミガナは、「出力選択（G）」が、紙テープの場合には、つぎのような形で出力エリアに入れられる。

国 [こく④④] 立 [りつ④④] 国 [こく④④] 語 [ご④④④] 研 [けん④
④] 究 [きゅう④] 所 [しょ④④]

一方、出力選択が、ライン・プリンター出力の場合には、ヨミガナだけが、

こくりつこくごけんきゅうしょ
の形で、出力エリアに記入される。
いずれの場合も、出力エリアに記入されるのは、漢テレ・コードのままである。

2. 5. 盤内非漢字コード処理過程(E)

「字種ふり分けルーチン(C)」で、ふり分けられた「漢字以外の盤内コード」のうち、カタカナ・ひらがな・ローマ字・算用数字・盤内記号については、さきに2・3において述べたように、まったく処理の必要がないために、そのまま出力エリアに送り出される。したがって、処理を必要とする「漢字以外の盤内コード」は、漢字の「オドリ字(々)」と「キャレッジ・リターン(改行復改コード)」の二つだけであるが、いずれも「出力選択(G)」が、紙テープ出力(漢テレ印字)の場合とライン・プリンター出力の場合とによって処理のしかたが、やや異なる。

2. 5. 1. オドリ字(々)処理(E₁)

オドリ字(々)については、その前の漢字の処理で採用されているヨミガナを、出力エリアに転写することになる。しかし、前の漢字が、盤外漢字である場合には、紙テープ出力においては、「々〔\$P\$P\$P\$P〕」の形で、出力エリアに入れる。ライン・プリンター出力においては、盤外漢字コード(♦)のみを、出力エリアに入れておく。

2. 5. 2. キャレッジ・リターン(C/R)処理(E₂)

入力データのキャレッジ・リターンは、ヨミガナ処理の結果、出力のさいには、位置がずれてしまう。そのため、入力データのC/Rは、つぎの四つの場合をのぞいては、出力側にもってくる必要はない。

- ①キャレッジ・リターンの連続(C/RC/R……)
- ②ピリオドとキャレッジ・リターンの連続(. C/R)
- ③スペースとキャレッジ・リターンの連続(\$C/R) (C/R\$)

①は、行送りだけの場合であり、②は文章段落における改行である。③は、

行の長さについてのコントロールをしているC/Rである。

これらのC/Rは、その位置も入力データとずれないし、出力データにおいても入れておく必要がある。そのため、上の四種のC/Rだけは、例外的に、出力エリアに入れることにしてある。

したがって、紙テープ出力（漢テレ印字）においては、上記のキャレッジ・リターン以外の復改は、すべて漢字テレ・タイプの自動復改機能に依存することになる。

2. 6. 盤外コード処理過程(F)

盤外コードすなわち、盤外漢字と盤外記号は、紙テープ出力（漢テレ出力）においては、コードの後に〔\$P\$P\$P\$P〕をつけた形で出力エリアに送りこむ。

したがって、漢テレ印字においては、つぎのような印字となる。

◆定学〔スペース〕

¶MO〔スペース〕

このスペースは、印字結果について、もとの文字・記号を記入するための空白である。

一方、ラインプリンター出力においては、コードのあたま1字（2ch）のみを出力エリアに送りこむ。したがって、この場合は、盤外漢字については、「◆（盤外漢字マーク）」のみが、盤外記号については「¶（盤外記号マーク）」のみが、出力エリアに送りこまれる。

2. 7. 出力選択過程(G)

このシステムの出力は、パラメーターにより、つぎの三種類のいずれでも可能なように予定している。

G₁ 紙テープ出力（パラメーター=0）

G₂ カタカナ書きライン・プリンター出力（パラメーター=1）

G₃ ローマ字書きライン・プリンター出力（パラメーター=2）

上記の出力方式は、データの読みこみに先だって、パラメーターとして「0」「1」「2」のいずれかを1桁（1ch）読みこませることによって選択され

る。ただし、ライン・プリンター出力については、現在、システムを設計中である。

2. 7. 1紙テープ出力 (G₁)

これは、出力エリアに、はいっている処理済みデータを、そのまま紙テープにパンチ・アウトするものである。出力エリアには漢テレコードで、処理結果が記憶されているので、この紙テープを漢字テレタイプで印字すれば、全文、漢字かなまじり文で、文中の盤内漢字には、ヨミガナがついている形で、結果が得られる。テスト・ランの処理結果の一部を示すと、つぎのとおりである。

<昭和41年1月3日・朝日朝刊・第1面>から

ベトナム問〔もん〕題〔だい〕をめぐる米〔べい〕国〔こく〕の“平〔へい〕和〔わ〕外〔がい〕交〔こう〕攻〔こう〕勢〔せい〕”は、年〔とし〕が明〔あ〕けても引〔いん〕続〔ぞく〕き行〔ゆ〕われ、北〔ほく〕爆〔ばく〕停〔てい〕止〔し〕も続〔ぞく〕行〔こう〕している。しかし問〔もん〕題〔だい〕のカギを握〔にぎ〕る北〔きた〕ベトナムは、これまでのところ従〔じゅう〕来〔らい〕通〔とお〕りのきびしい反〔はん〕応〔のう〕しか示〔しめ〕しておらず、ワシントンでは悲〔ひ〕観〔かん〕的〔てき〕な見〔けん〕通〔つう〕しも現〔あらわ〕れ始〔はじ〕めた。ドゴール仏〔ぶつ〕大〔だい〕統〔とう〕領〔りょう〕も一〔いち〕日〔にち〕、「関〔かん〕係〔けい〕諸〔しょ〕国〔こく〕の組〔そ〕織〔しき〕的〔てき〕な接〔せつ〕触〔しょく〕、外〔がい〕部〔ぶ〕からの武〔ぶ〕力〔りょく〕干〔かん〕渉〔しょう〕排〔はい〕除〔じよ〕」を呼〔よ〕びかけ、米〔べい〕中〔ちゅう〕接〔せつ〕触〔しょく〕、米〔べい〕軍〔ぐん〕撤〔てつ〕退〔たい〕の必〔ひつ〕要〔よう〕を示〔じ〕唆〔さ〕した。一〔いち〕方〔ほう〕、南〔みなみ〕ベトナムでは新〔しん〕年〔ねん〕早〔そう〕々〔そう〕、激〔げき〕戦〔せん〕が行〔ゆ〕われ、サイゴンでも交〔こう〕渉〔しょう〕へのきっかけがつかめるとの見〔けん〕方〔ほう〕は少〔すくな〕ないようだ。=関〔かん〕係〔けい〕記〔き〕事〔じ〕、3面〔めん〕に
來〔き〕ない”よい返〔へん〕事〔じ〕”米〔こめ〕なお望〔のぞ〕み捨〔す〕てず
〔ワシントン=松〔まつ〕山〔さん〕特〔とく〕派〔は〕員〔いん〕ニ〔に〕日〔に〕ち〕發〔はつ〕】

ベトナム戦〔せん〕争〔そう〕解〔かい〕決〔けつ〕を求〔もと〕める米〔べい〕国〔こく〕の平〔へい〕和〔わ〕攻〔こう〕勢〔せい〕は、その後〔うしろ〕も拡〔かく〕大〔だい〕しているが、北〔ほく〕爆〔ばく〕停〔てい〕止〔し〕後〔ご〕九〔きゅう〕日〔にち〕たっても共〔きょう〕産〔さん〕側〔がわ〕からまだ“よい返〔へん〕事〔じ〕”がないため、ワシントンにはそろそろ悲〔ひ〕観〔かん〕論〔ろん〕が出〔しゅつ〕始〔はじめ〕ている。米〔べい〕政〔せい〕府〔ふ〕筋〔すじ〕は「北〔ほく〕爆〔ばく〕停〔てい〕止〔し〕」は別〔べつ〕に時〔じ〕間〔かん〕制〔せい〕限〔げん〕があるわけではない」と忍〔にん〕耐〔たい〕強〔ごう〕くハノイの反〔はん〕応〔のう〕を待〔ま〕つ姿〔し〕勢〔せい〕を示〔しめ〕しながらも、同〔どう〕時〔じ〕に「共〔きょう〕産〔さん〕側〔がわ〕の態〔たい〕度〔ど〕に変〔へん〕化〔か〕のない場〔ば〕合〔あ〕にも“無〔む〕期〔き〕限〔げん〕”に停〔てい〕止〔し〕しておくという意〔い〕味〔み〕ではない」ことを明〔あき〕らかにしており議〔ぎ〕会〔かい〕開〔かい〕会〔かい〕までに手〔て〕がかりが得〔え〕られなければ、軍〔ぐん〕事〔じ〕面〔めん〕での大〔おお〕きな発〔はつ〕展〔てん〕が起〔お〕る可〔か〕能〔のう〕性〔せい〕も十〔じゅう〕分〔ぶん〕にあるようだ。

しかし、ホワイトハウスは目〔もく〕下〔か〕のところ、「こんどの働〔はたら〕きかけが失〔しつ〕敗〔はい〕した後〔うしろ〕、エスカレーション〔段〔だん〕階〔かい〕的〔てき〕〕拡〔かく〕大〔だい〕」にもどるというような決〔けつ〕定〔てい〕はまだ何〔なに〕もしていない」と非〔ひ〕公〔こう〕式〔しき〕に述〔の〕べ、あくまで平〔へい〕和〔わ〕外〔がい〕交〔こう〕攻〔こう〕勢〔せい〕の成〔せい〕功〔こう〕に期〔き〕待〔たい〕をつなぐ構〔かま〕えを見〔み〕せている。

<昭和41年1月3日・朝日朝刊・第4面>から

東〔とう〕京〔きょう〕都〔と〕新〔しん〕宿〔しゅく〕矢〔や〕来〔らい〕町〔ちゅう〕／

振〔ぶり〕替〔たい〕東〔とう〕京〔きょう〕八〔はち〕〇〔れい〕八〔はち〕／
新〔しん〕潮〔ちょう〕社〔しゃ〕

<昭和41年1月3日・朝日朝刊第6面>から

新〔しん〕風〔ふう〕をはらむコンサイス・シンボル

発〔はつ〕刊〔かん〕以〔い〕來〔らい〕44年〔ねん〕半〔はん〕世〔せい〕紀〔き〕の
伝〔でん〕統〔とう〕につちかわれた<10訂〔てい〕版〔はん〕>コンサイスは常〔つ
ね〕に新〔あたら〕しく現〔げん〕代〔だい〕に息〔いき〕吹〔ふ〕く

<昭和41年1月3日・朝日朝刊・第7面>から

ことしの景〔けい〕気〔き〕

経〔けい〕営〔えい〕者〔しゃ〕アンケート

不〔ふ〕況〔きょう〕さらに深〔ふか〕まる

¶1〇〔 〕倒〔とう〕産〔さん〕がいぜんふえ続〔つづ〕けていることなどから
みて、不〔ふ〕況〔きょう〕はまだ底〔てい〕入〔い〕れしていないと思〔おも〕う。
これからまだまだひどくなろう。一〔いち〕般〔ばん〕に夏〔なつ〕すぎから上〔じよ
う〕向〔こう〕くという観〔かん〕測〔そく〕が有〔ゆう〕力〔りょく〕なようだが、
私〔わたくし〕はそうは思〔おも〕わない。不〔ふ〕況〔きょう〕は予〔よ〕想〔そう〕
以〔い〕上〔じょう〕に深〔しん〕刻〔こく〕で、今〔こん〕年〔ねん〕いぱついは
本〔ほん〕格〔かく〕的〔てき〕に上〔じょう〕向〔こう〕く見〔けん〕込〔こ〕みは
立〔た〕たない。

¶2〇〔 〕これといった指〔し〕標〔ひょう〕はない。中〔ちゅう〕小〔しょう〕
企〔き〕業〔ぎょう〕などの経〔けい〕営〔えい〕状〔じょう〕態〔たい〕、輸〔ゆ〕
出〔しゅつ〕の見〔けん〕通〔つう〕し、政〔せい〕府〔ふ〕の景〔けい〕気〔き〕対
〔たい〕策〔さく〕など自〔じ〕分〔ぶん〕がジカに感〔かん〕じたことから總〔そ
う〕合〔ごう〕的〔てき〕に判〔はん〕断〔だん〕した。

¶3〇〔 〕政〔せい〕府〔ふ〕に頼〔たよ〕るというより、経〔けい〕営〔えい〕
者〔しゃ〕の自〔じ〕省〔しょう〕が必〔ひつ〕要〔よう〕だ。他〔た〕力〔りき〕本
〔ほん〕願〔がん〕ではこんなきびしい不〔ふ〕況〔きょう〕は乗〔じょう〕切〔せつ〕
れない。結〔けつ〕局〔きょく〕、自〔じ〕分〔ぶん〕でやらなければだめだ。政〔せい〕
府〔ふ〕も、われわれも自〔じ〕覚〔かく〕を新〔あらた〕たにし、分〔わ〕に応
〔おう〕じ、反〔はん〕省〔せい〕しながら慎〔しん〕重〔ちょう〕に対〔たい〕処
〔しょ〕していくべきだ。

自〔じ〕動〔どう〕車〔しゃ〕についても横〔よこ〕ばいで行〔ゆ〕ければ成〔せ
い〕功〔こう〕と考〔かんが〕えている。野〔の〕放〔ほう〕図〔ず〕な計〔けい〕画
〔かく〕をたててはいかん。

2. 7. 2. ライン・プリンター出力 (G₂・G₃)

この部分については、現在、システムを作成中であるが、ほぼ、つぎのような方針で設計している。

このプロセスは、出力エリアに入っている「処理済みデータ（漢テレ・コード）」を、片端から、ライン・プリンターのコードに変換してプリント・エリアに送りこみ、ライン・プリンターに印字させるプロセスである。

ライン・プリンター出力の場合すなわち「出力選択パラメーターが1または2の場合」には、出力エリアには漢字コードが一切はいってこないようにする。したがって、この場合には、出力エリアには、入力データの中の盤内漢字は出力エリアに送りこまれず、そのヨミガナのみが、出力エリアに送りこまれる。

また、盤外コードは、2. 6で述べたように、◆（盤外漢字マーク）と引（盤外記号マーク）のみが、出力のエリアに送りこまれる。

漢字の「オドリ字（々）」については、すでに2. 5に述べたように、オドリ字の前が盤内漢字のときは、そのヨミガナのみを、また盤外漢字のときは◆マークのみを出力エリアに入れる。

以上の結果、ライン・プリンター出力の場合、出力エリアには、カタカナ・ひらがな・ローマ字・算用数字・盤内記号・キャレッジ=リターン・スペース・◆マーク・引マークに対応する漢テレ・コードだけがはいっている。このうち、スペースだけはライン・プリンターのコードと一致しているので特別な変換を行なう必要はない。また、ローマ字・算用数字と大部分の盤内記号も、漢テレ・コードのY軸コードが、ライン・プリンターのコードと一致しているので、プログラム的に簡単にコード変換ができる。盤内記号の一部分、例をあげると{}「」『』→←などは、対応するライン・プリンター・コードがないため、一括して一定のコードに変換する。いまのところライン・プリンターのspに変換する予定である。

漢テレ盤外コード、すなわち◆と引は、すべてライン・プリンターの#に変換する予定である。

出力エリアにあるカタカナ・ひらがなについては、HSMの中に2種類のテ

一ブルを用意しておいて、ライン・プリンターのコードに変換していく。2種類のテーブルのうち、どちらのテーブルを選ぶかは、2. 7に述べた出力選択パラメーターの情報による。すなわち、このパラメーターが1の場合には、カタカナ変換用テーブルが選ばれ、出力エリアのカタカナとひらがなは、すべてライン・プリンターのカタカナのコードに変換され、全文カタカナ書きとなる。また、出力選択パラメーターが、2の場合には、ローマ字変換用テーブルが選ばれ、出力エリアのカタカナとひらがなの漢テレコードは、すべて、ライン・プリンターのローマ字コードに変換される。そのさい、カナのおどり字（ゝ、ゞ）、長音符号（ー）および促音符号（っ、ッ）については、特別な処理をしなくてはならない。

まず、カナのおどり字（ゝ、ゞ）は、前の文字についての変換結果を、そのまま転写する。長音符号は、ローマ字変換においては、前の文字の変換結果の末尾の一宇（すなわち母音）を転写するだけですむが、カタカナ変換においては、母音変換のテーブルを用意しなくてはならない。促音符は、カタカナ変換では、ライン・プリンター・コードの「ッ」に変換すればよいが、ローマ字変換では、つぎの文字についての変換結果のあたま一字（すなわち子音）を当てることとなる。

出力エリアの処理ずみデータに、以上述べたような変換処理を施しつつ、データの末尾あるいはC/Rコードまでをプリント・エリアに送りこみ、印字していく。もちろん、C/Rコードそのものは、プリント・エリアには送らず、その個所で、プリントの、行を送ることにする。したがって、印字面は、改行となる。

3. テスト・ランの結果

すでに述べたように、現在の段階では、「ライン・プリンター出力(G)」の回路は、まだ完成していない。この段階での最初のテスト・ランの結果は2. 7に示した通りである。ここに示したデータについての処理時間は、約7分間であった。このシステムのデータ処理時間は、言うまでもなく、漢字含有率の高いデータほど、長くかかる。それは、データの中の漢字と、漢字テーブ

ルとのコレートすなわち辞書引きに最も処理時間を食うからである。

しかし、漢字含有率の高いデータ言いかえれば、漢字を使い得る個所には、必ず漢字が使用されている文章ほど、解読の正解率は高くなる。これは、漢字の現われ方を目やすにして、処理を進めているので、当然のことである。たとえば、「頭蓋骨」と漢字で出してくれば、「ズガイコツ」と変換されるが、「頭がい骨」と出でくると「アタマがいホネ」となってしまう。したがって、小学生向きの童話などよりも、漢字がたくさん使つてある学術論文などの方が、処理結果は良好である。

漢字の多いデータとして、尾高朝雄氏の論文「多数決の原理」を用いたテスト・ランでは、総文字数6200字で、処理時間は、約18分であった。正解率というよりは、1. 1に述べた目標への到達率は、データに現われた漢字（延べ字数）の86%程度であった。このテスト・ランの結果の一部をつぎにあげておく。さきの新聞記事の場合の処理結果にくらべると、かなりよくなっているようである。

多 [た] 数 [すう] 決 [けつ] の原 [げん] 理 [り]

多 [た] 数 [すう] 決 [けつ] の原 [げん] 理 [り] には、確 [たしか] かに相 [そ] う] 対 [たい] 主 [しゅ] 義 [ぎ] 的 [てき] な意 [い] 味 [み] がある。甲 [こう] 論 [ろん] 乙 [おつ] ◆日送 [] の意 [い] 見 [けん] の対 [たい] 立 [りつ] がある場 [ば] 合 [あ] に、神 [かみ] ならぬ人 [にん] 間 [けん] の知 [ち] 性 [せい] をもつてしては、その中 [なか] のどれを選 [えら] ぶべきかを絶 [ぜつ] 対 [たい] の確 [かく] 信 [しん] をもって断 [だん] 定 [てい] しうる者 [もの] はない。それを、なおかつひとりの絶 [ぜつ] 対 [たい] の権 [けん] 威 [い] をもって断 [だん] 定 [てい] するという制 [せい] 度 [ど] は、独 [どく] 裁 [さい] 主 [しゅ] 義 [ぎ] である。独 [どく] 裁 [さい] 主 [しゅ] 義 [ぎ] に走 [はし] って、人 [にん] 間 [けん] の合 [ごう] 理 [り] 性 [せい] を◆空処 [] ◆君儻 [] し國 [こく] 民 [みん] の運 [うん] 命 [めい] を奈 [な] 落 [らく] の底 [そこ] に転 [てん] 落 [らく] せしめる◆係要 [] を踏 [ふ] むまいとする以 [い] 上 [じょう]、多 [た] 数 [すう] の意 [い] 見 [けん] を採 [さい] 用 [よう] するという方 [ほう] 法 [ほう] を採 [と] らざるを得 [え] ない。それは、相 [そう] 対 [たい] 主 [しゅ]

義〔ぎ〕であると同〔どう〕時〔じ〕に便〔べん〕宜〔ぎ〕主〔しゅ〕義〔ぎ〕である。少〔しょう〕數〔すう〕の意〔い〕見〔けん〕でも、ひとりの意〔い〕見〔けん〕でも、正〔ただ〕しいものは正〔ただ〕しいに相〔そう〕違〔い〕ないのであるから、いやしくもそれが正〔ただ〕しければ、少〔しょう〕數〔すう〕の意〔い〕見〔けん〕、ひとりの意〔い〕見〔けん〕であっても、それを採〔さい〕用〔よう〕し、それを断〔だん〕行〔こう〕するのが、政〔せい〕治〔じ〕の理〔り〕想〔そう〕にかなった最〔さい〕善〔ぜん〕の方〔ほう〕法〔ほう〕であるに相〔そう〕違〔い〕ない。しかし、その最〔さい〕善〔ぜん〕の方〔ほう〕法〔ほう〕によろうとすれば、寡〔か〕頭〔とう〕政〔せい〕治〔じ〕となり独〔どく〕裁〔さい〕政〔せい〕治〔じ〕となる以〔い〕上〔じょう〕、次〔じ〕善〔ぜん〕の策〔さく〕、三〔さん〕善〔ぜん〕の策〔さく〕ではあっても、多〔た〕數〔すう〕の支〔し〕持〔じ〕するところによって行〔こう〕動〔どう〕するほかはないという便〔べん〕宜〔ぎ〕主〔しゅ〕義〔ぎ〕が、多〔た〕數〔すう〕決〔けつ〕原〔げん〕理〔り〕の中〔なか〕に内〔ない〕在〔ざい〕していることは、否〔ひ〕定〔てい〕し得〔え〕ないところであろう。

しかしながら、多〔た〕數〔すう〕決〔けつ〕原〔げん〕理〔り〕が全〔ぜん〕然〔ぜん〕底〔てい〕のない相〔そう〕対〔たい〕主〔しゅ〕義〔ぎ〕であり、決〔けつ〕定〔てい〕のためのみに決〔けつ〕定〔てい〕を求〔もと〕める便〔べん〕宜〔ぎ〕主〔しゅ〕義〔ぎ〕にすぎない、と言〔い〕うのも、また極〔きょく〕端〔たん〕な一〔いち〕面〔めん〕観〔かん〕であると言〔い〕わなければならない。多〔た〕數〔すう〕決〔けつ〕原〔げん〕理〔り〕が言〔げん〕論〔ろん〕の自〔じ〕由〔ゆう〕の精〔せい〕神〔しん〕と深〔ふか〕く結〔むす〕びついており、したがって少〔しょう〕數〔すう〕意〔い〕見〔けん〕にも十〔じゅう〕分〔ぶん〕にその主〔しゅ〕張〔ちょう〕を吐〔と〕露〔ろ〕する機〔き〕会〔かい〕を与〔あた〕え、多〔た〕數〔すう〕意〔い〕見〔けん〕にも影〔えい〕響〔きょう〕を及〔およ〕ぼす道〔みち〕を開〔ひらく〕ものであるということは、これまた否〔ひ〕定〔てい〕すべからざる真〔しん〕実〔じつ〕である。それによって、多〔た〕數〔すう〕意〔い〕見〔けん〕もみずから反〔はん〕省〔せい〕を重〔かさ〕ねることがあるか、否〔いな〕、さらに進〔すす〕んで、正〔ただ〕しい少〔しょう〕數〔すう〕意〔い〕見〔けん〕がしだいに大〔おお〕せいを動〔うご〕かして、ついに逆〔ぎゃく〕に多〔た〕數〔すう〕の支〔し〕持〔じ〕を得〔え〕るにいたるか、あるいは、多〔た〕數〔すう〕が數〔かず〕の暴〔ぼう〕威〔い〕をふるって少〔しょう〕數〔すう〕意〔い〕見〔けん〕を圧〔あつ〕倒〔とう〕し、一〔いち〕党〔とう〕一〔いち〕派〔は〕の利〔り〕益〔えき〕をば理〔り〕非〔ひ〕

[ひ] を問〔と〕わざに貫〔つらぬ〕こうとするかは、むしろ民〔みん〕主〔しゅ〕政〔せい〕治〔じ〕の運〔うん〕用〔よう〕の問〔もん〕題〔だい〕である。

人〔にん〕間〔けん〕の歴〔れき〕史〔し〕を顧〔かえり〕みると、正〔せい〕義〔ぎ〕が実〔じつ〕力〔りょく〕によって◆空処〔 〕◆君働〔 〕された事〔じ〕実〔じつ〕は、数〔すう〕限〔かぎ〕りないほどに多〔おお〕い。過〔か〕去〔こ〕においてそうであったばかりでなく、現〔げん〕在〔ざい〕においてもそうであり、將〔しょう〕來〔らい〕もおそらくそうであろう。けれども、それと同〔どう〕時〔じ〕に正〔せい〕義〔ぎ〕が邪〔じや〕惡〔あく〕を克〔こく〕服〔ふく〕し、理〔り〕想〔そう〕が現〔げん〕実〔じつ〕を動〔うご〕かし、少〔しょう〕数〔すう〕の正〔ただ〕しい意〔い〕見〔けん〕が最〔さい〕後〔ご〕の勝〔しょう〕利〔り〕を獲〔かく〕得〔とく〕した場〔ば〕合〔あ〕も、またけっして少〔すくな〕なくない。現〔げん〕在〔ざい〕もそうであるし、將〔しょう〕來〔らい〕はますますそうあらしめなければならない。歴〔れき〕史〔し〕の進〔しん〕歩〔ほ〕への道〔みち〕は、進〔しん〕歩〔ほ〕の可〔か〕能〔のう〕性〔せい〕を信〔しん〕じ、そのために必〔ひつ〕死〔し〕の努〔ど〕力〔りょく〕を傾〔けい〕注〔ちゅう〕する者〔もの〕!前〔まえ〕に開〔ひら〕かれる。

ただ、具〔ぐ〕体〔たい〕的〔てき〕な場〔ば〕合〔あ〕にあたって、どこに正〔ただ〕しい進〔しん〕歩〔ほ〕への道〔みち〕があるかは、もとより容〔よう〕易〔い〕には決〔けつ〕定〔てい〕しがたい。これについて議〔ぎ〕論〔ろん〕を重〔かさ〕ねるのは、共〔きょう〕同〔どう〕生〔せい〕活〔かつ〕に関〔かん〕与〔よ〕するすべての人〔にん〕々〔にん〕の義〔ぎ〕務〔む〕であると同〔どう〕時〔じ〕に、権〔けん〕利〔り〕である。言〔げん〕論〔ろん〕の自〔じ〕由〔ゆう〕が重〔おも〕んぜられなければならないゆえんは、そこにある。しかも議〔ぎ〕論〔ろん〕を重〔かさ〕ねて、なおかつ帰〔き〕一〔いち〕点〔てん〕を見〔み〕いだし得〔え〕ないとき、いたずらに移〔うつ〕り行〔ゆ〕く事〔じ〕態〔たい〕の前〔まえ〕に◆月玉〔 〕立〔りつ〕することが許〔ゆる〕されない以〔もつ〕上〔うえ〕、なるべく多〔おお〕くの人〔にん〕々〔にん〕の納〔なつ〕得〔とく〕する方〔ほう〕針〔しん〕を探〔と〕るよりほかはない。そこに、多〔た〕数〔すう〕意〔い〕見〔けん〕必〔ひつ〕ずしも正〔ただ〕しからず、少〔しょう〕数〔すう〕意〔い〕見〔けん〕がかえって道〔どう〕理〔り〕にかなっている場〔ば〕合〔あ〕も、まれでない。その結〔けつ〕果〔か〕は、やがて現〔げん〕実〔じつ〕の事〔こと〕のなりゆきのうちに現〔あらわ〕われてくる。その

結〔けつ〕果〔か〕を見〔み〕て、かつての決〔けつ〕定〔てい〕に同〔どう〕意〔い〕した多〔た〕数〔すう〕の人〔にん〕々〔にん〕も、虚〔きょ〕心〔しん〕◆身急〔じんきゆう〕懐〔かい〕、前〔まえ〕には多〔た〕数〔すう〕決〔けつ〕によって否〔ひ〕定〔てい〕された第〔だい〕二〔に〕案〔あん〕を支〔し〕持〔じ〕し、これをこころみて失〔しつ〕敗〔はい〕の救〔きゅう〕済〔さい〕を図〔はか〕るというは、多〔た〕数〔すう〕決〔けつ〕の倫〔りん〕理〔り〕である。さればこそ、少〔しょう〕数〔すう〕意〔い〕見〔けん〕の者〔もの〕も、「みずから省〔はぶ〕みて直〔なお〕くんば、千〔せん〕万〔まん〕人〔じん〕といえどもわれ往〔ゆ〕かん」の概〔おおむ〕をもってその見〔けん〕解〔かい〕を主〔しゅ〕張〔ちょう〕し、多〔た〕数〔すう〕に圧〔あつ〕倒〔とう〕されて敗〔やぶ〕れても、これを將〔しょう〕來〔らい〕の再〔さい〕檢〔けん〕討〔とう〕のための伏〔ふく〕線〔せん〕とし、共〔きょう〕同〔どう〕財〔ざい〕として残〔のこ〕すというは多〔た〕数〔すう〕決〔けつ〕の道〔どう〕義〔ぎ〕である。

ジョン＝スチュアートミルは、真〔しん〕理〔り〕は共〔きょう〕同〔どう〕の論〔ろん〕義〔ぎ〕・檢〔けん〕討〔とう〕によってのみ發〔はつ〕見〔けん〕せられ、誤〔ご〕◆興入〔こうにゅう〕は真〔ま〕◆果違〔かくち〕な批〔ひ〕判〔はん〕・◆客派〔きゃくはい〕磨〔ま〕を通〔とお〕じて初〔はじ〕めて是〔ぜ〕正〔せい〕せられるという確〔かく〕信〔しん〕のもとに、高〔こう〕貴〔き〕なる言〔げん〕論〔ろん〕の自〔じ〕由〔ゆう〕を説〔と〕いた。ミルを生〔い〕んだイギリス国〔こく〕民〔みん〕の政〔せい〕治〔じ〕道〔どう〕徳〔とく〕の高〔たか〕さが、その國〔くに〕の民〔みん〕主〔しゅ〕政〔せい〕治〔じ〕を不〔ふ〕断〔だん〕に向〔こう〕上〔じょう〕せしめ、多〔た〕数〔すう〕決〔けつ〕による世〔せ〕論〔ろん〕立〔りつ〕法〔ほう〕の範〔はん〕を世〔せ〕界〔かい〕に示〔しめ〕したのは、けだしゆえあることと言〔い〕わなければならぬ。

ミルもまた、人〔にん〕間〔けん〕の判〔はん〕断〔だん〕がすこぶる誤〔あやま〕りやすいものであることを認〔みと〕める。これこそまちがいなく真〔しん〕実〔じつ〕だと思〔おも〕うことが、実〔じつ〕はやはり誤〔ご〕◆興入〔こうにゅう〕であることは、きわめて多〔おお〕い。だからこそ、一〔ひと〕つの判〔はん〕断〔だん〕だけを絶〔ぜつ〕対〔たい〕の真〔しん〕理〔り〕として人〔ひと〕に強〔きょう〕要〔よう〕し、これに反〔はん〕対〔たい〕する意〔い〕見〔けん〕の發〔はつ〕表〔ひょう〕を許〔ゆる〕さない態〔たい〕度〔ど〕は、けっしてとてはならない。(ミル「自〔じ〕由〔ゆう〕論〔ろん〕」)思〔し〕想〔そう〕および言〔げん〕論〔ろん〕の自〔じ〕由〔ゆう〕論〔ろん〕)

[ゆう] の主 [しゅ] 張 [ちょう] は、ここから出 [しゅつ] 発 [はつ] する。だから、その出 [しゅつ] 発 [はつ] 点 [てん] を見 [み] れば、ミルの見 [けん] 解 [かい] もまた一 [いち] 種 [しゅ] の相 [そう] 対 [たい] 主 [しゅ] 義 [ぎ] である。自 [じ] 己 [こ] の信 [しん] 念 [ねん] にも誤 [ご] ◆興入 [] がありうるし、理 [り] 解 [かい] しがたい他 [た] 人 [じん] の主 [しゅ] 張 [ちょう] が正 [ただ] しいこともありうる。対 [たい] 立 [りつ] するいくつかの意 [い] 見 [けん] のうちのどれか一 [ひと] つが常 [つね] に絶 [ぜつ] 対 [たい] に正 [ただ] しいといふことは、いかなる偉 [い] 人 [じん] も賢 [けん] 者 [しゃ] も断 [だん] 言 [げん] することを得 [え] ない。そう考 [かんが] えるのは、一 [ひと] つの相 [そう] 対 [たい] 主 [しゅ] 義 [ぎ] である。けれども、ミルによれば、その中 [なか] のどれが正 [ただ] しいかは永 [えい] 遠 [えん] に知 [し] られ得 [え] ないといふのである。その中 [なか] には、必 [かなら] ず正 [ただ] しい意 [い] 見 [けん] があるのであり、その意 [い] 見 [けん] の正 [ただ] しさは、やがて事 [じ] 実 [じつ] によって立 [りつ] 証 [しょう] される。まず論 [ろん] 議 [ぎ] が尽 [つ] くされ、それに經 [けい] 驗 [けい] が加 [くわ] わることによって、やがて真 [しん] 理 [り] と誤 [ご] ◆興入 [] との別 [べつ] が判 [はん] 然 [ぜん] と示 [しめ] される時 [とき] が来 [く] る。かように、論 [ろん] 議 [ぎ] と經 [けい] 驗 [けい] とによって真 [しん] 理 [り]、すなわち、人 [にん] 間 [けん] の進 [すす] むべき正 [ただ] しい道 [みち] が◆化調 [] 握 [あく] されうると信 [しん] ずる点 [てん] では、ミルはけっして単 [たん] なる相 [そう] 対 [たい] 主 [しゅ] 義 [ぎ] 者 [しゃ] ではない。

(尾 [び] 高 [こう] 朝 [ちょう] 雄 [ゆう])

4. おわりに

われわれが、漢字を解読していく場合には、文章全体から、さまざまな手がかりを得ながら読んでいく。ある場合には、文章の内容から、ある場合には読み手の知識・経験から、またある場合には前後の単語・文字・送りがなからといったように、解読の要因は、多種多様であり、これを現在の段階で、電子計算機にわかる形に整理していくことは、きわめて困難である。

そこで、今回のシステムにおいては、漢字の解読すなわちヨミガナ決定のための、手がかりを得る範囲としては、おそらく最小の範囲である「漢字の前後

1字」に限定し、そこから得られる手がかりだけを用いて解読を試みたわけである。それでも、前節に述べたように、一応の成果をあげることができた。

今回のシステムにおいても、漢字テーブルのヨミガナの数（最大6個）をふやしたり、指定環境表示の制限（最大4個）をとりのぞいたりすれば、処理スピードはともかくとして、結果は、なお向上する余地がある。また、漢字かなまじり文のエントロピーの計算結果などを用いて、それぞれの漢字の前後に、どんな文字が現われやすいかを、適確に把握できれば、処理能率も、精度も、さらに改善しうる。

しかし、ヨミガナの選択・決定のための情報を、このようなせまい範囲だから得ていたのでは、やはり限度がある。たとえば「天下／天下り」の「天」あるいは、「喫茶店／茶店」の「店」のヨミガナを決めるためには、「天下り」の「り」あるいは「喫茶店」の「喫」を手がかりにしなくてはならない。しかし、このように前後2字目に現われる解読の手がかりは、今回のシステムではもはや生かすことができない。「盛り場／盛り上り」の「盛」や「滑らか／滑らない」の「滑」あるいは「魚市場／青果市場」の「場」なども同様である。このような場合を処理できるように、手がかりを得る範囲を、その漢字の前後2字、3字というように広げていくのも、計算機的には、一つの方向には違いないが、基本的な考え方は、今回のシステムと同一のものである。このやり方でも、すべての場合について手がかりを得る範囲を広げず、ある漢字は前後2字、あるヨミガナは前後4字というようにやっていけば、処理能率を、それほど落さずに、精度をよくしていくこともできよう。しかし、この種の考え方では、手がかりとなる文字を決めにくいくらい多くの同表記別語「仮名（かな／かめい）」「工夫（こうふ／くふう）」「通（とお／かよ）って」「逃（に／の）がす」などの処理には、なかなか近づけないのでなかろうか。

そこで、考え方を一転して、解読の手がかりを、単語単位あるいは文節単位に求めていくというやり方も、当然出てくる。前後の一定のセンテンスを使うということを考えられないではない。どうせ範囲を広げるのなら、こうした方向に広げていった方が、すくなくとも同表記別語や熟字訓の処理には有効なはずである。

ヨミガナの選択・決定を、機械で処理する場合の情報源として、どんな言語単位が適切か、そして、そこからどんな情報を、いかにして得ていくか、これを見つけるのが、今後の大きな課題である。それは、また、機械処理ばかりでなく、漢字の解読過程の研究に、示唆を与えるものもある。