

国立国語研究所学術情報リポジトリ

国研用漢字テレタイプと同機利用の言語情報処理

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: 松本, 昭, MATUMOTO, Akira メールアドレス: 所属:
URL	https://doi.org/10.15084/00000985

国研用漢字テレタイプと同機利用 の言語情報処理

松 本 昭

- § 1 言語情報処理における入出力の方式をめぐって
 - § 2 国研用漢テレ設計上の基本的要求と、同機のコード構成及び盤面排列の概要
 - § 3 鍵盤上の文字排列と文字・記号群のコード的序列の概要
 - § 4 盤外字の入力方法と漢字の部首順排列のための理論コードについて
 - § 5 E. D. P. S. の入出力機器として見た、漢テレの問題点
- 附 国研用漢字鍵盤穿孔印刷機(漢テレ)仕様書 抜粋

§ 1 言語情報処理における入出力の方式をめぐって

Computer が記憶し、処理しうるのは、bit^{註1)}の形に転換された情報だけである。

そこで、E. D. P. S (電子情報処理組織)中の入出力装置とよばれる機器の中には、人間の使用する視覚的な文字・数字・記号などを、bit 情報の形に変換し(入力)、また逆に、本体で処理した情報を、視覚的な文字・数字・記号の形になおして印字する(出力)ための端末機器が必須とされる。(注2参照)

そもそも、Computer によって情報処理をおこなう場合には、

- 1) 処理される原始データが、どのような形で発生し、あるいは存在しているか。
- 2) そのデータに、どのような処理を加え、どのような結果を、どのような形で得る必要があるのか。

ということによって、E. D. P. S 導入の以前であれば、入出力機器の種類を選択の問題をふくめ、全体としての機器構成の組み方に工夫をこらす必要もおこるし、導入以後であれば、どのように、現存する入出力機器を利用するかという、入出力の方式についての検討、研究を十分おこない、総合的な見地に立って、最も効率の高い方式を発見し、その上で、全体としての処理過程の設計をしなければならない。

これは、まことに当然すぎる問題であり、また、あまり重要性を認められない事柄のようであるが、実用として役立つ入出力機器の種類にかぎりがあ
り、その性能にも、さまざまな限界がある現況においては、絶対にゆるがせ
にできない問題であり、また、必ずしも単純、容易な問題ではない。仮りに、
作業過程の機械化という観点だけから見れば、プログラム技術や、新しい機
器を最大限に活用し、人手をほとんど使わない非常にスマートなやり方であ
っても、結局、より多くの経費、時間をついやし、必ずしも満足できる結
果を得られない場合もあれば、逆に、甚だ「見てくれ」の悪い、いわゆる
「やりくり」に似た工夫による、機械力と人手とのくみ合せ利用といった方
式であっても、現段階においては、考えられる最高の効率を生むこともあり
るのである。

現在、普通の汎用 Computer は、検査用のパリティ bit を除いて、1 桁 6 bit
から、8 bit というのが大部分で、1 桁で区別できる文字・記号などの種類
は、6 bit なら 64、7 bit なら 128、8 bit なら 256 が最大限となる。ところ
が、日本語の書きことばは、大部分が、分ち書きなしの、漢字仮名まじり文
であるから、普通の E. D. P. S に備えられている入出力機器を用いたのでは、
原始データの形のままで、日本語の情報を入力することはできない。そこ
で、一般の適用例では、すべて、仮名またはローマ字の形になおして入力し、
処理することになっている。従って、少なくとも 1 度は、入力以前の段階で、
漢字仮名まじり文から、仮名、ローマ字だけの形に翻字することがおこな
われ、更に要すれば、出力後に、再び漢字仮名まじり文に復原することがおこ
なわれる。もちろん、一般の事務計算などの業務においては、若干の文字・
記号によるコードと数字が、情報の主体となるので、人名などの固有名詞を、
仮名またはローマ字になおすだけで、この方式で処理を進めても、あまり問題
はないようである。ところが、直接にことばを処理する、たとえば、自動抄
録、情報検索、索引づくり、語彙調査などの場合には、この翻字の段階に、
非常に多くの手間がかかり、全体としての処理能率のネックとなる可能性が
ある。また、日本語を仮名またはローマ字で表記した場合に、同表記となる
語の弁別も、人間であれば Context によって簡単にできるものでも、大量
のデータを迅速に処理する仕事において、Computer にそれをさせるの

は、——全く不可能というわけではないが——実用的に困難な面が多すぎるので、結局、翻字の段階で、人間が何らかの弁別情報を加えてやるか、あるいは、すべての語を用例つきで処理して、あとから、人間が弁別し、再整理するなどの手間をかけなければならない。また、翻字の作業そのものにおいてしばしば煩らわしい、かつ実用的には無駄の多い議論が必要とされ、時には、誤まった翻字をした結果、原表記のままであったら、起りえないような、あやまちを、最終の処理結果に与えることもおこりうる。

もちろん、このように、仮名・ローマ字などの表音文字に翻字して入力し、処理した場合には、各処理事項を発音順に分類排列する際、その作業を完全に自動化しうるといふ大きな利点が生ずる。

しかし、もし、表記とか用字とかについての情報もあわせて入力し、それに関する処理結果を得たい場合には、漢字仮名まじり文を、仮名・ローマ字に翻訳するだけでなく、漢字1字1字に相当する情報を、何らかの形で与えなければならない。それも、たとえば、中国で従来電報をうつ場合に用いていた、漢字1字を4桁の数字で表現する方法(電碼)のように、同一の文字に対しては、必ず同一の示し方をすることが可能なように、「対応表」を準備して、コードづけをしなければ、満足な処理結果はえられない。その点、四角号碼の方式を、そのまま利用することには問題がある。また、この場合、漢字を単にコードとおきかえるだけのやり方では、出力後、コードから漢字への再翻訳が終らないかぎり、一般には読むことができないことになるので、その漢字の読みもあわせて入力しておくことが望まれる。そうすると、「通う」「勉強する」などのデータは、「カヨ〇〇〇〇ウ」、「ベン〇〇〇〇キョウ〇〇〇〇スル」(〇〇〇〇は、かりに4桁の数字コードを示すもの)というような、かなり間びした形に翻字されることになるだろう。また、「反共」を「ハン(ソリ)キョウ(トモ)」、「反響」を「ハン(ソリ)キョウ(ヒビキ)」というぐあいに、弁別のために字訓をそえるやり方も、同一文字に対して同一の示し方をするためには、数字コードを与える場合と同様に、翻字作業者のための「対応表」を用意しなければならない。ところで、同音の漢字の極めて多い日本漢字音の実情を考えると、この種の「対応表」を作ること自体、決して容易な仕事ではない。さらに、今の例とは逆に、同訓異字のある漢字が、字訓に

よって使われている場合の処理法も、併せて考えておかなければならないのだから、「対応表」の内容及び機械処理の過程は、もう一段と複雑化するはずである。

国研(国立国語研究所)においては、このような困難を経ることなく、日本語の表記に関する情報を処理しうるシステムを構成することが望ましいとの判断に立って、漢字テレタイプ(以下漢テレと略称する)を、入出力の端末機器として備えることとした。^{注2)}

漢テレは、従来から、通信社・新聞社などで、通信や活字の自動鑄造に使われていたもので、鍵盤穿孔機と印刷機の両部分からなり、鍵盤上の文字を視覚的に確認して、直接それを打鍵することによって、その字に相当する1字分2桁、2列のコードが紙テープに穿孔され、同時に、その字を印字することもできるという点で、比類のない利点をもつものである。

この打鍵穿孔された紙テープを仲介とすれば、日本語の情報は、ほぼ原表記の形のままに、Computerに入力でき、処理した結果は、紙テープにそのまま穿孔して出力し、その紙テープを漢テレの印刷機にかければ、漢字仮名まじり文の形で印字され、直ちに読むことができるわけである。ただ、この場合には、何らかの形で、読みの情報も入力し、それによって処理を進めなにかぎり、異表記同語の類も、完全に別項目として扱われてしまうのであるから、語形あるいは、発音による分類をしたい場合には、そのための作業または、処理を予定しておかなければならない。

注 1) bit(binary digit の略)とは、有・無など2個の状態の1組によって保持される最大の情報量のことで、2進数の1桁に相当し、紙テープ上の一定位置に、1個の孔が有るか無いかといった形で表現され、Computerの内部では、電流のパルスや磁気などの有・無の2状態におきかえて、記憶し、処理して行くものである。

このような、情報の示し方は、最も単純素朴なもので、中国古代の易の卦にも全く同様のものが見られる。易の爻^まは、真中のときれの有(--), 無(-)によって、陰陽という2種類の情報を示すのだから、1bitに相当し、爻を3個重ねて8種類のちがった情報、(乾, 兌, 離, 震, 巽, 坎, 艮, 坤)を示す八卦(☰, ☷, ☱, ☲, ☴, ☵, ☶, ☸)は、3bit; 八卦を上下に重ねあわせて、64種類の卦を作るものは、6bitに相当するといえる。

注 2) 現在、国研に設置されているE. D. P. Sは、HITAC-3010で、その

機器構成は次のようになっている。

1. H-304 中央演算処理装置：1桁7 bit (1パリティ bit, 6情報 bit), 2万桁の高速(主)記憶部—High Speed Memory, 略して H. S. M. —と制御部とをそなえる。(この HSM は, 2桁1単位で7 μ s—秒間約28万桁—の速さで, 記憶している内容を制御部にとり出し, 再び同じ内容を HSM に書きこむ。制御部は, HSM に内蔵されているプログラムの解説, および処理の実行をする)
 2. H-382 磁気テープ補助記憶装置—Magnetic Tape, 略して MT—6デッキ。(秒間3.3万桁の速さで, HSMの内容をテープにかきこみ, またテープ上の内容を HSM によみとることができる)
 3. H-321 紙テープ読取り穿孔機(秒間100桁ずつの速さで, 紙テープに穿孔されている情報を HSM によみとり—入力—, また HSM 内の情報を紙テープに穿孔—出力—する)
 4. H-333 C 高速印字機—Line Printer—1台(HSMの内容を, 1行120字まで, 毎分600行—仮名をふくまぬ64種類の文字・記号を使う場合—または, 500行—仮名をふくむ96種類の文字使用の場合—の速さで直接印字する)
 5. H-177 万能入出力装置—フレクソ—3台(英字, 数字, 記号及び仮名をふくむ, 3段シフトの電動タイプライターで, 印字と同時に, 紙テープにコードが穿孔され, 逆に, 紙テープを読み取って, 印字もする。最大速度, 分間450字。)
 6. 漢字テレタイプ 3台 (この機能については, 後述)
- * 以上のうち, 2以下はすべて入力, 出力に関係する装置であるが, とくに4は On Line (本体と電氣的に接続している)の出力機器, 5, 6は Off Line (本体と電氣的に接続していない)の入出力, 端末機器となっている。
- * カード関係の入出力装置は, 全くふくんでいない。これは, 紙テープ関係のものと同く比べて, 価格が高い, 言語情報のように, 長さが不定のものをとり扱う場合, カードより, 紙テープの方が適応性が高い, 漢テレでは, 紙テープしか使えない, などの理由による。

§ 2 国研用漢テレ設計上の基本的要求と, 同機のコード構成及び盤面排列の概要

Computer への入出力機器としての漢テレであれば, その本体である Computer の機種, 性格のちがいに応じ, また今後それによって行うであろう処理内容や処理方式のちがいに応じて, それらに便利なように仕様を作りなおし, 標準品とは別に設計し, 製作する必要がある。

すでに述べたように、漢テレと Computer との間は、Off Line であるから、ただ紙テープを介してだけ情報の交換がおこなわれる。従って、漢テレに対する入力、鍵盤打鍵か、紙テープ読み取りかのどちらか一方であり、その出力は紙テープ穿孔、あるいは、印字のどちらか一方、または両方という形でおこなわれる。(この点は、Off Line の万能入出力装置も全く同様) この機能は、標準品の漢テレにそなわっているのであるから、仕様を改める主要点は、漢テレの出力の一つである紙テープが、Computer への入力でありえ、Computer からの出力である紙テープと同様の規格と精密さとをそなえること。コード上、盤面排列上の諸要求を十分満足させることという2点となる。

現在のところ Computer と情報の交換ができる程度の漢テレの製作に実績を持っているのは、沖電気と新興製作所の二つである。この両社の漢テレは、機能においては、ほとんど差がないようであるが、構造上にかなりのちがいがあつた。両社が国内向けに作っている標準品の漢テレは、ともに、6 bit・2列の12 bit で1文字分のコードとしている。従って、原理的には、64×64で4096通りのコードを作りうるのであるが、漢テレ自体の機構上の理由もあり、また国内需要において必要とされる文字数と、その使用頻度との関係から割り出される効率^(註1)、及び、機械の印刷部分の安定性と処理速度の確保のために、大体、沖で2400、新興で2500前後の文字数しか収めていない。沖の場合は、鍵盤上の key を一つ打つことによって、5 bit ずつ2列の10 bit が決定され、左右2個のシフトペダルによって、残りの2 bit を決定する。これに対して、新興の場合は、右手でうつつ文字 key 一つで8 bit、左手でおさえる14乃至16個の識別 key の中の1個で4 bit を決定する形である。従って沖の漢テレの文字 key 1個には、2 bit で識別される4文字までが、新興の場合は、4 bit で識別される16文字以内がおさめられるのである。(実際には、個々の user の用途に応じて、識別 key の数や文字 key 1個におさめられる文字数に変動がある。)

このような構造上のちがいは、各文字・記号、あるいは文字・記号群に対するコード配当や、鍵盤上の文字・記号の排列の仕方に、全く別の可能性と制約とを与えるものであるが、従来からの実績で、国研に導入された Com-

puter, H-3010 のメーカーである日立が、沖電気と密接な関係にあり、同機種による漢テレ・Computer の連動の経験もあるために、国研用の漢テレは日立を通じて沖電気に製作してもらうこととなった。従って、上記の点に関し、我々が工夫しうる範囲は、始めから、沖の型の中に限定されることとなった。

以下は、我々が、国研用漢テレの仕様を決定する段階において、想定した処理業務及び処理方式にもとづいて、提出した基本的な要求事項である。

- (1) 用字調査をふくむ、語彙調査の入力データーを作るため、調査対象の表記を、そのまま打鍵印字する場合を考え、文字数は、許容される最大限をおさめる。字種の選定は、打鍵の能率を考えて、使用頻度の高いものからという原則に従う。鍵盤上にない文字・記号の類は、仮名文字や別の漢字によって代用するのではなく、盤内に1字分の盤外字マークを用意し、それを利用して、表現する。
- (2) 盤外字の表現方法は、漢字を Computer において、分類排列する場合の方法とともに、別に考える。(後述)
- (3) 語彙調査の結果を発音順に分類排列する場合の便宜のため、漢テレ内の仮名文字に対するコードは、容易に漢字のコードと区別され、H-3010 によって分類した場合に、そのまま五十音順に排列されるようにし、同じ平仮名と片仮名のコード間には、bit 構成上、最大限の共通点をもたせ、容易に互換しうるようにする。
- (4) 打鍵の能率をあげるため、鍵盤上では、平仮名、片仮名の全体が可能なかぎり狭い範囲内にまとまり、同じ発音の平仮名、片仮名、及び清、濁、半濁の文字は互に連続してならび、かつ仮名の類全体としての排列も、五十音の順になっているようにする。
- (5) 6 bit 1 列の 3010 コードで表現され、ラインプリンター、フレクソによって印字可能な、英字、数字、特殊記号の類 48 種には、共通の弁別コードを第 1 列にあたえ、第 2 列のコードは、3010 コードと完全に一致させる。こうすることにより、3010 コードから漢テレコードへ、逆に漢テレコードから 3010 コードへの転換が、機械的に可能となるばかりでな

く、プログラムのデバッグ(検査、修正など)や処理の中間過程におけるデータの状態を、いちいち漢テレで印字しなくても、ラインプリンターなどに打ち出して、その大体を察することが容易になる。また、これらの文字・記号の鍵盤内の排列も、最下段に0から9、AからZという順序をくずさずに、左から1行にならべることとする。

(6) H-3010において、not used とされている(17)_s、(57)_s注2)のコード、及び、+0、EI、ED、EF、EB、Issなどに配当されている(37)_s、(52)_s、(55)_s、(56)_s、(72)_s、(75)_sなど、機械そのものの制御や、アセンブラーその他のシステムプログラムの制御用に使われているコードは、盤内の文字を表現するコードとしては使用しない。また、H-3010のもっている、言語情報処理の上で、極めて有利であると考えられる、DSL、DSR、LSL、LSR、「3010 命令語の説明」参照)などの命令を、有効に用いるために、可能なかぎり盤内文字用に使用するコードの種類を少くし、プログラム用に使えるコードの種類を多くする。特に、3010でスペースを表現する(12)_sなどは、言語単位の分割に利用するため、盤内文字用には使わない。

(7) 平仮名をふくめ、使用頻度の高い字には、ノンシフト(シフトペダルにタッチしない)で穿孔・印字できるコードを配当し、全体として、鍵盤上の文字・記号類の排列は、以上の諸要求に背反しない形で、未経験の操作者にも、探しやすいことを主眼として工夫する。(詳しくは後述する)

以上、とくに(3)以下の諸要求は、まず、漢テレの出力としての、従ってH-3010への入力としての、穿孔テープのコードの種類を制約し、同時に、漢テレに、その出力としての印字機能を果させるための入力としての、穿孔テープ(これはH-3010からの出力として穿孔されるもの)のコードの種類を限定するものである。

ところで、漢テレの印字機構にとって必要な入力情報は、印字機の活字ドラム中の、どの活字を選択印字するのかを決定するための情報としてのコードである。この活字ドラムは、縦に4字ずつならんだ活字の列が、24個横にならんだ形の矩形のパレットが、26枚集って、図1のようになったもの

である。印字すべき文字を決定するためには、**Ⓐ** 第何番目のパレットの、**Ⓑ** 第何列目の、**Ⓒ** 上から何字目に位する活字であるかという**Ⓐ****Ⓑ****Ⓒ** 3個の情報が与えられる必要がある。すでにのべた、6 bit 2列のコードというのは、印字機の側から見れば、**Ⓐ**を示す5 bit、**Ⓑ**を示す5 bit、**Ⓒ**を示す2 bit、の集まりでなければならないことになる。

国研用の漢テレの場合は、前記(3)、(4)、(5)の要求を実現する必要から、第1列の第1から第5までの5 bitが**Ⓐ**を、第2列の第2から第6までの5 bitが**Ⓑ**を示すことになっている。H-3010のコードの場合は、紙テープ上の穿孔のない部分が1を、穿孔のある部分が0を意味しているのに対し、この場合は逆で、穿孔のある部分が1、ない部分が0を意味し、それぞれ1~26、1~24までの順位と区別を示すことが必要なので、各単位のウェイトづけも、1, 2, 4, 8, 16, という純2進型である必要はなく、パレット番号を示す第1列は、1, 2, 8, 12, パレット内部の列番を示す第2列は、1, 2, 4, 8, 9, という形になっている。そこで、印字機のためには、表1のa, b, の印字機構の欄に見るようなコードを備えることが必要となる。ところが、これでは、H-3010への入力としての穿孔コードとしては、注の欄、*印の各個所において、前述の諸要求事項を実現する上で不都合を生じる。そこで、印字機内部に、表1の矢印の右側、穿孔テープコード欄にあるようなコードが、入力された場合でも、矢印左側のコードが入力された場合と同様の活字選択がおこなえるように、リレー回路を設けておき、紙テープのコードと、印字機の必要とする情報としてのコードとの間の矛盾を解決することにしてある。

一方、鍵盤穿孔機の方も、コード穿孔用の配線を単純化するためには、各行各列内の文字に共通コードを配当することが必要となる。その結果、盤面内の排列も、印字機内の活字ドラムの内容を大体において反映する形となる。

盤面は大きく、盤内keyと盤外keyとに二分される。盤内は、1 keyに4文字づつが標示され、横に25行、縦に24列の600 key 盤外は1 keyに1記号のわりで標示されて、盤内keyの右側に縦に排列され、漢テレ制御用の補助key 5, プログラム用のkey 19, スペースkey 1, からなっている。このうちの、印字動作をしない盤外keyのいくつかを除き、鍵盤key

topに標示された、すべての文字・記号の活字は、印字機内の26枚のパレットにおさめられており、そのうち、盤内最下段の行と盤外を除くと、あとの各行24個ずつのkey topに標示されている文字・記号の活字は、それぞれ1枚ずつのパレットに、そっくりおさめられている関係になっている。詳しくは、表1, aの鍵盤穿孔機の行番号と印字機のパレット番号とを対照されたい。

なお、盤内の各key topに標示されている4個の文字は、シフトペダルの操作との直観的な一致をはかるために、ヒシ形に配置されて、真下の字がノンシフト、左側の字が左シフト、右側の字が右シフト、真上の字が両シフトという形になっているが、コード的には、全体として、やや左側に傾け、左側の字を左下、真上の字を左上、右側の字を右上、真下の字を右下にもってきて、4個の字が正方形の各頂点に配置されているように考えると、より理解しやすくなる。(図2参照)。すなわち、第1列のコードを共通にする文字が、横に48つつならんで50行あり、第2列のコードを共通にする文字が、縦に50つつならんで48列あるわけで、横の1行、48文字間のコード的序列は、表1, bの穿孔テープコードのオクタル表示に見られるとおり、盤面の右端からかぞえて第1字目と第4字目とをのぞくと、左側から順に、上昇順にならんでいる。これに対して、縦、つまり、50の各行相互間のコード的序列は、表1, aのオクタル表示に見られるように、リレー回路によってコード変換をおこなっている4個の行をのぞき、1行おきに、まず偶数番目の行が上から下に上昇順にならび、ついで、奇数番目の行が上から下に、上昇順にならんでいる形である。このように、コード的序列のあり方が、縦と横とで著しく異っているのは、横の序列を決定する第2列目のコードについては、そのシフトbitを第1単位においたのに対して、縦の序列を決定する第1列目のコードについては、そのシフトbitを第6単位において、両者を一致させていないからである。(一般の標準品では、シフトbitの位置は、1, 2列とも第6単位にしているのが普通のようなものである。)このようにしたのは、前記(3)(4)の要求事項をほぼ満足させ、しかも、使用頻度の極めて高い、平仮名のすべて(全84字)を、ノンシフトのコードに配当するという(7)の要求を実現するためである。

なお、表1を見れば明らかなように、(6)の後半の要求は、決して十分満足にかなえられてはいない。すなわち、盤内字を表現するコードとして、第1列としても、第2列としても穿孔されないでいるのは、(17)₈。以下の特別のものを除外すると、3010コードでスペース、@、*、に配当されている(12)₈、(14)₈、(54)₈の、たった3種類で、第1列としては穿孔されていないのが、Jに配当されている(41)₈^{註3)}、第2列としては穿孔されていないのが、#、, (アポストロフ)、\$, =、に配当されている(13)₈、(36)₈、(53)₈、(76)₈、の4種類という状態である。

しかし、この点は、6 bit という条件であるかぎり、やむをえない結果であったと考えるわけにはいかない。従ってもしも、将来ともに、漢テレのように、文字数の多い入出力機器を利用する必要があるならば、Computer本体は、8情報 bit 1桁のものを採用し、漢テレの入出力となる紙テープも9単位のものを使うように設計した方が有利ではないかと考えている。

注1) 国研報告22号『雑誌九十種の用語用字、第2分冊漢字表』の第7頁、所載の使用度数分布表によれば、使用頻度の高い順にえらばれた1995字の漢字で、全標本中の98.6%の表記が可能であり、2,137字では、99%が可能である。しかし、これを100%にまで高めるためには、更に1,199字も字種を増やさなければならない。国研が、それ以前に実施した、婦人雑誌、総合雑誌の調査結果も、大体同様で、上位2,000字までで、前者の場合は98.6%、後者の場合は99%までの表記が可能であるという結果が出ている。従って、上手に選定された、約2,000字の漢字があれば、大体100のうち99まではまかなえると考えてよいことになる。

注2) (17)₈(51)₈などの表示は、octal(8進数)を示すもので、3 bit ずつを1桁(従って6 bit は2桁)として表示しうるので、各コード1 bit 構成を直観的に把握するのに便利である。下に、2進、8進、10進の対応を示しておく。なお2進の0は紙テープ上の穿孔部分に、1は穿孔のない部分に相当する。

2進	000	001	010	011	100	101	110	111	001	000	001	001	001	111	010	000	101	110	111	111
8進	0	1	2	3	4	5	6	7	1	0	1	1	1	7	2	0	5	6	7	7
10進	0	1	2	3	4	5	6	7	8		9			15	16		46			63

注3) 第1列のコードとして、(41)₈が使われないというのは、盤外の場合にも一貫しており、漢テレの印字機が、紙テープの読み取り印字をする際、常に、正しく第1列、第2列をくみあわせて1文字分のコードとして読み取るためのメカニズムに、この事実が利用されている。(附録、仕様書抜粋6-3参照)

§ 3 鍵盤上の文字排列と文字・記号群のコード的序列の概要

漢字の部は、上から第 20 行目までの全 480 key のすべて(4 × 480 = 1920 字)と、第 21, 22, 23, 24 行の全 96 key の上半の 2 字(2 × 96 = 192 字)の、総計 2112 字で、そのうちわけは、当用漢字 1845 (1850 字から、使用頻度の低い、朕、璽、脹、耆、弑の 5 字を除いたもの)、表外字 264 (雑誌 90 種の用字調査で、度数 11 以上であったもの)、及び○、々、盤外字マーク、をふくむ。

漢字の排列に際しては、コード上の序列もふくめて考慮すると、盤面上の排列に秩序をあたえることが困難となり、結局、文字を探しづらくなる恐れがあるので、コードの序列は全く無視し、次のように排列した。

(イ) 上から 20 行までは、書物の頁を開いたように、左右 12 列ずつに分けて、左上から右下に向って、漢字音による五十音順にならべる。その場合、盤外字マーク、○、漢数字をふくめて、使用頻度の高い漢字 480 を、まずノンシフトの位置に、五十音順に排列し、残りを、また同じ発音ごとにまとめて、左、右、両シフトの位置に、五十音順に配当して行き、全体として、表音部分の字形の似たものは、なるべく接近しておくように心がけたが、結局、2 系列の五十音順排列の形となり、同音の文字がすべて完全にまとめられているわけではない。

(ロ) 第 21, 22, 23, 24 行の上 2 字の部分には、訓だけでしか用いられない字、助字、動植物・度量衡単位など特殊な意味グループの語表記に用いられる漢字などを収めた。

仮名の部は、§ 2 の要求(3)(4)に従い、上から第 21, 22, 23, 24 行目の中で、盤面排列の序列と、コード的序列との一致を得られる範囲、つまり、左から第 22 列目まで(表 1 b の穿孔テープコード欄及び左からの列番欄参照)の 88 key の下の 2 字(2 × 88 = 176)の部分に収めることとした。実際には、普通大の片仮名、平仮名(清、濁、半濁)全 148 字と、小さい片仮名、平仮名の全 20 字及び長音符号、中黒、各 2 種で 4 個、総計 172 字でそれを、同種、同音の片仮名、平仮名を完全に 1 組として、前者を左シフト、後者をノンシフトの位置に排列していった。従って、同種同音の片仮名、平仮名のコードは、第 1 列は共通、第 2 列は前者が偶数、後者は、前者より 1.

だけ多い奇数という関係になっている。(表2参照)

その他の部は、盤面最下段つまり第25行の下の2字には、§2の要求(5)により、0~9, A~Z, &+・; : - / () , % など48個の文字・記号、また同じ行の上の2字には、a~z(英小文字) ゃゝ! ? 「 」 …… ← → < > など、48個の文字・記号がおさめられ、同種の英小文字と大文字とは、同一のkeyの上と下とに並び、コードとしては、第2列が完全に一致する形となっている。

第21, 22, 23, 24行の下半分で、片仮名、平仮名を排列した残りの部分には、§2の要求(3)に従うため、一般の漢字仮名まじり文において混用されることと考えられない種類の文字・記号類、αβγηε など20個をおさめた。(表2参照)

以上の文字・記号を、それらに配当されたコードの序列において排列すると、概略、次のようになる。以下12bitを4行の8進数で表示する。(8は省略)

(0000)~(0074) : 曲計に始まり等重に終る ; 盤面第9行目下半の漢字48字、

(0100)~(0174) : ゃゝに始まり}に終る ; 盤面第25行目上半の英小文字と記号など48種。(第2表参照)

(0200)~(3174) : 哀愛に始まり十惑に終る ; 盤面第1行から8行までと10行の下半分、21行24行の上半分、11行から20行までの下半分にふくまれる漢字1007字と盤外字マーク。

(3200)~(3574) : 21, 22, 23, 24行の下半分にある、片仮名、平仮名、172字に、発音符号の類20。(第2表参照)

(3600)~(3674) : 25行の下半分にある、3010コードと一致する文字・記号類、48種。(第2表参照)

(4000)~(7474) : 暁極に始まり呂巳に終る ; 第9行、つづいて第1行から8行、更に10行から23行までの上半分にふくまれる漢字1056字。

この序列が、所謂漢テレコードによるソート(分類排列)結果の序列でもあるわけだから、漢字をふくむデータを、漢テレで入力し、その結果をソートして出力する必要がある場合には、適当な方法、手順を加えて、別の序列にならべかえることをしなければ、一般的な意味での実用性のあるものを作

ることにはならない。

表2は、片仮名、平仮名の部を中心に、漢字以外のものについて、そのコード的序列を詳しく示したものである。なお盤面排列との対応を見るために、行番、列番を示す数字を左側と最上部とに記入してある。

これで明らかなように、仮名内部におけるコード上の序列が五十音順になっているのであるから、漢字で表記されたデータに読みの情報を仮名でつけて入力し、その読みを key として分類排列するならば、その処理項目を、ほぼ五十音順に排列して取り出すことが可能である。

ただ、この五十音順は、たとえば「か」の部が全部すんでから「が」の部に入るというような問題点があるために、特別のプログラムの手法を用いまいが、一般の国語辞典の排列と完全に同じ排列にすることはできない。また片仮名、平仮名を一応一貫してとり扱えるが、「いしゃ」が「いしや」に優先して排列されるので、「いしゃ」「いしゃがよい」「いしゃぎらい」……のあとに「いしや」「いしやきいも」が排列されるという形になる。(田中論文参照)

§ 4 盤外字の入力方法と、漢字の部首順排列のための理論コードについて

盤内におさめられた漢字は、2,110字(その他に、々、盤外字マーク)であるから、普通の文献を漢テレで入力しようとするとき、少なくとも100の漢字を打鍵する間に1回ぐらいの割合で、盤内にはない漢字が出てくるものと予想される(§2注1参照)

この盤外漢字を処理する方法として、

- 1) すべての盤外漢字に対して、単に盤外字マークだけをうって入力し、原稿中の該当文字に印をつけておき、機械の処理がすんでから、その原稿につけられた印を手掛りに、人間の手で処理をする。
- 2) 盤外字マークと、更に、その盤外漢字の種類に応じた情報・コードをあわせて入力し、それらの処理も機械にやらせる。

という二つが考えられる。処理の内容が、単に印字するだけであるとか、新聞社の場合のように、一連の記事を電送したり、自動活字鑄造機にかけて版をくむような場合には、1)の方法をとるのが最も現実的である。

しかし、大量のデータにもとづく、用字調査のような場合には、単に盤

内には漢字が全体でどれだけあったかという形でしか、自動的に結果が出てこない、1)のやり方をとることは好ましくあるまい。もちろん、データの量が少い場合には、入力前に要する手間が少いという利点があるのだから、1)の方法で処理した方がよいこともある。

2)の方法をとる場合も、§1でのべた、同一の文字に対しては常に同一のコードを与えるという原則が適用されなければならないので、存在しうるすべての文字に対するコードが準備されなければならないことになる。

現在の日本語の表記に用いられる文字の異り数は、大体3,000から4,000程度であると考えてよいと思うが、(雑誌九十種の用語用字の第二分冊参照)その内容となる字種を完全に固定してきめることは、必ずしも容易ではない。(たとえば、日本の少し大きな印刷業者は、6~7,000字種の活字をそなえている。)そこで、盤外漢字のコードとして、盤外字マークの次にうつ文字数も1(漢テレ)字におさえることはむずかしくなる。(もし1字とすると、盤内字以外に2,400字だけが入力可能ということになる。)しかし、2(漢テレ)字をもって、盤外漢字のコードとするならば、最大限、576万ものコードができることになるが、存在しうる漢字の数は、異体、略体、その他一切のものを別字として数えても、10万のオーダーを越えることは考えられない。そこで、盤外漢字のコード用の文字は、打鍵しやすいノンシフトの漢字(盤面第1行から20行までの各keyに1個ずつある全479字中の)450字にかぎることとした。表3は、そのコード用の漢字450字を、それぞれの漢テレコードによるコード的序列に従い、1番から450番までの番号をつけて並べたもので1計(0001)₈、2形(0003)₈、3型(0005)₈にはじまり、449料(3047)₈、450力(3051)₈におわっている。これら2字ずつの組合せで構成できるコードは最大202,500種類である。

次に、1)個々の盤外漢字に、どのようなコードづけをするかという問題と、

2) 漢テレ盤内字を3010で分類排列すれば、漢字については、曲、計、巾、形に始り、乃、那、呂、巳に終る、全く何の原理も法則も見られない序列で並んでしまうことは、前節までにのべたとおりであるが、これら盤内漢字と、盤外漢字との、コード的序列の関係をどのように調整するかという問

題 とが出てくる。

そこで、われわれは、1)、2)の問題を同時に解決する方法を考えた。つまり、盤内、盤外を問わず、計算機によって処理される漢字を、すべて、まず完全に部首画数順に配列し、それら1字1字に、一定のコード的序列をもつ理論コードを順次配当して行くのである。この理論コードは、現在存在している漢字だけでなく、将来発見されたり、作られたりする種類の漢字に対しても配当することができ、しかも、それによって、部首順にならぶという、序列に狂いを生ぜず、かつ重複したコードづけを絶体におこさないものとして工夫しなければならない。

そこで、

イ) 部首・画数順に排列された漢字リストの第一次的原簿として、現在の辞典中で最大の字数をおさめている、諸橋の漢和大辞典をそのまま利用する。

ロ) コードづけにあたっては、原簿である漢和大辞典の各見出し字に与えられている通し番号を、そのまま利用する。

ハ) 各部の先頭字に与えられた通し番号の下2桁を切りすてたもの、通し番号の下2桁が00の場合には、その番号から100を引いたものを、その部の基準番号とする。(以下、表4参照)

ニ) 各部の最後の字に与えられた通し番号から、その部の基準番号を引いた答が、450をこえる場合には、基準番号からかぞえて、450番までをその部の第1グループとし、以下同様に、450字ごとに第2、第3グループを構成して行く。こうすると、たとえば人の部は、通し番号344の「人」にはじまるグループと、751の「倆」にはじまるグループ、1201の「俊」にはじまるグループの3グループにわかれる。このようにすると、正式の部首の数は214であるが、全体で271の「部グループ」ができることになる。なお、正式の部の先頭でなく、第2、第3グループの先頭の場合は、その先頭字の通し番号から1ひいたものが基準番号となる。

ホ) 理論コード第1字の漢字は、271の「部グループ」を弁別し、序列づけるものとして、理論コード漢字表の1番の計から、271番の合までを、順次配当する。

へ) 理論コード第2字の漢字は、問題の字に、漢和大辞典が与えている通し番号から、その文字の属する部グループの基準番号を引いた答えの数値に相当する番号に位置づけられている理論コード用の漢字である。(表3,4参照)

以上のような手続きで作られる、漢字用の理論コードは、2(漢テレ)字からなるので、6 bit 4桁(オクタル表示で8桁)で構成され、一に対する計・計(00 01 00 01)。を優先順位の最も高いものとし、合・力(21 21 30 51)。を理論的に最も優先順位の低いものとしているから、コード数121,950(271×450)を最大限とする、すきまだらけのコード体系を構成していることになる。

しかし、とにかく、表3のような「理論コード用漢字表」と、表4のような「漢字コード基準表」とをそなえておけば、漢和大辞典の通し番号さえわかれば、すべての漢字に対して、一貫した理論コードを即座に与えることができるわけで、あらかじめ多数の漢字について、完全なコード表を作っておく必要はなく、比較的使用頻度の高い盤外漢字と盤内字についてだけ、漢字と理論コードとの対照表を作っておけばよい。

また、諸橋漢和大辞典にない(つまり番号の与えられていない)漢字や、同じ通し番号が、別の字に与えられている'(ダッシュ)つきの番号の漢字については、その部全体の最後につけ加えて、仮りの番号を与えて、コードづけを行い登録しておく。たとえば、リの部の剂には、2076'の通し番号がつけられているので、刀部の最後の2287番に1を加えて、2288番として扱い、理論コード第1字は、その部グループに共通の21番「東」をあたえ、第2字には、2288からそのグループの基準番号2250を引いた答えの38番に相当する「身」(表3参照)をあたえ、(0065 0241)。となるコードをつけるわけである。このようにして行けば、ほとんど考えられない事態であるが、仮りに、ある部に属する新しい漢字が大量に出て来たとしても、コード第1字が、次の部のコード第1字に優先しているのであるから、絶対に、全体としての部首の面数順排列の序列をみだすおそれはない。万一、コード第2字として、理論コード450番目の「力」まで使い果してしまった場合でも、漢テレコードで「力」のそれ(3051)。よりも大きいコードをもつ漢字を、臨時に、順次補充して使って行けばよいのである。補充として使用することのできる漢字は1,300

字以上ある。

かくして、一般の入力データーを作製する際には、盤面にある文字は、その文字をそのまま打鍵し(1字分、穿孔コードは2桁)、盤面でない文字については、まず、盤外字マークを打鍵し、ついでその文字に対して与えられている理論コードに相当する漢字2字を打鍵する(都合3字分、穿孔コードは6桁)わけである。機械内部で処理する際は、常に2桁を一まとまりに処理して行くのであるが、盤外字マークに相当する2桁のコード(3115)を発見した場合だけ、それに後続する4桁をあわせて、6桁一まとまりとして処理するように、プログラムを組んで行くことになる。

必要な処理がすんだのち、最終の結果を、部首順に分類排列する必要があるば、すべての盤内字コード2桁のあとに、その文字に与えられている理論コード4桁分を挿入し(これは、盤内字コードに対する理論コードのテーブルを作っておきさえすれば、極めて簡単なプログラム手法で行うことができる)、盤内字、盤外字をとわず、後の4桁を手がかりに、分類排列すれば、常に一貫した部首順のソート結果が得らる。

§ 5 E. D. P. S. の入出力機器として見た、漢テレの問題点

入出力機器としての漢テレのもつ問題点は、何といっても、その処理速度のおそいことにある。§1においても、のべたように、表記関係の情報をふくむ入力データーを作る場合などに、前編集の手間が著しく節約できる利点はあるが、前編集のすんだ原稿を打鍵するにあたり、普通のフレキソのように、原稿から目を離すことなく、完全にタッチ方式で連続的に打鍵することはできない。また、仮りに、連続的に打鍵可能な場合でも、印字をともなう場合は、秒間2字以上、穿孔のみの場合でも、秒間5.8字以上の速度で打鍵すると、後から打鍵された方の字については無効となり、穿孔も印字もされないことになる。(仕様 4—2, 6, 参照)実際の作業としては、平均、50~60字/分の速さでしか打鍵できないのが普通である。

なお、念のためにのべておくが、この速度は、全く機械的に打鍵することについてだけの計算で出したもので、漢テレで、完全な入力データーを作るために要する時間を計算するためには、漢テレ操作者のパンチミスを発見し

訂正する時間、また、漢テレの故障によっておこりうるミスパンチを発見し訂正するための時間なども考慮に入れなければならない。

しかし、とにかく、この入力データ作製の段階で要する時間は、現在のよりに人間の手で打鍵する方式であるかぎり、程度の差こそあるが、ある限度以上に縮めることは不可能であるし、漢テレの50字は、フレクソの100～150字には十分匹敵しうるのであるから、我慢しうるとしても、最大の問題は、処理結果の印字に要する時間である。

これは全く机上の計算にもとづくものであるが、今、最終の処理結果が磁気テープの中に、100万桁(50万漢テレ字分)の情報として入っているととして、それを人間の読める文字情報の形にする(印字する)のに要する時間を考えてみると、国研に入っているラインプリンター(333C型)のスピードは、1行120字で、500～600行/分、つまり、最大限、60000～72000字/分であるから、ラインプリンターで出力すれば、約17分乃至14分間で出力が完了し、カナまたは、漢字、数字、その他の記号の形で見ることができる。ただ、それが漢テレで入力された内容である場合には、英字、数字、など48文字の部分を除いて、2字ずつ、1組の暗号のような形でしか印字されない。(英、数字など48文字の漢テレコード第1列の(36)。は、333C型ラインプリンターにおいて、仮名から英、数字へのクリアコードとなっているため、漢テレで入力した英、数字の類は、ラインプリンターで印字しても1字分ずつの間隔をおいてではあるが、正確にもとの英、数字の形で印字される)

これに対して、同じものを漢テレで印字するためには、まずH-321の穿孔機(100桁/秒)によって、磁気テープの内容を紙テープの形にして、出力しなければならないが、それに要する時間は約2時間46分以上であり、この紙テープを漢テレの印刷装置によって、印字するには、更に約70時間を見なければならない。この時間には、紙テープのハンドリングに要する時間が全く無視されているので、実際には、もっと多くの時間がかかることになる。

もちろん、ラインプリンターによる印字と、漢テレによる印字とでは、その読みやすさの点で格段のひらきがあり、前者の場合は、いろいろと、後編

集をしないと解読できないのであるが、直接コンピューターを占有する出力時間が2時間半以上も短縮されることと、処理結果の大意が即座に知りうるという大きな利点がある。そこで、入力データを漢テレで作った場合でも、その処理結果の出力はできるかぎり、ラインプリンターを活用し、後編集を助けるために、その見出し部分や、key word 的な部分だけを紙テープに出力して、漢テレで印字し、ラインプリンターの印刷結果とあわせて利用するという出力方式を、今後、大いに研究、開発して行く必要があるだろう。

現に、筆者の開発した「表記(用字)統計プログラム」は、完全にこの方式によっており(第一資料研究室月報, '66 APR. 参照)、研究室全体で進めている、新聞の用語用字調査のプログラムにおいても、このような方式を、かなりとり入れて行くことになっている。

このほか、機械工学的な理由によって生ずる、誤印字、誤穿孔の問題や、穿孔テープの Computer による検査法などの問題があるが、言語そのものの問題から、はなはだしく離れてしまうので、ここではふれないこととする。

附 国研用漢字鍵盤穿孔印刷機(漢テレ)仕様書、抜粋

1 概 説

本装置は国語研究所に設置される HITAC-3010 のオフライン付属機器として使用されるもので、計算機との連絡はすべて8単位紙テープを介して行われるものとし、この8単位紙テープは HITAC-3010 にて読取り可能とする。入出力情報は1文字(ファンクション1動作とも)7単位2列の14単位を使用する。本装置は鍵盤部の文字キー選択打鍵によりその情報を紙テープに穿孔すると共に印字紙に印字することができる。又紙テープに穿孔された情報を読取り機にかけて上記のテープ穿孔及び印字を行わせることができる。なお本装置ではテープ複製のさい編集を行うことができる。

2 構 成

本装置の構成機器は下記の通りであり、各機器間は、接続コードにより連絡して使用する。

=, ', -, @, *, の19キーで、それぞれ専用コードを穿孔する。但し、※4, ※3, ※2, ※1は印刷機は不動作、他は各々の記号ないしは文字を印字する。

(4) 補助キー(ファンクションキー)

- a) 復改キー：打鍵印字の時はプラテンの行を改め、プラテンを第1字目印字位置にもどす、打鍵穿孔の時は漢テレ復改コードを穿孔する。
- b) スペースパー：打鍵印字の時は印字せずプラテンを1字分進める。(スペース)穿孔の時は間隔コードを穿孔する。
- c) 後退キー：プラテン及びテープを1字分後退させる——印字穿孔なし——
- d) 抹消キー：打鍵印字の時は抹消記号を印字する。穿孔の時はオールマーク(7つ孔2列)を穿孔する。
- e) 空白キー：このキーを1回押せばテープにオールスペース(スプロケットの孔だけあく)を2列分穿孔する。
- f) 連続キー：文字キー及びファンクションキーと同時に押すと連続動作を行う

——中略——

(6) 鍵盤アラームランプ及びアラーム断押ボタン

鍵盤タッチアラームすなわち約40ミリ秒以内には2ケのキーを押した場合は同時に押したものとみなして、穿孔(印刷部接続の場合は印字)がカットされてアラームを出す。即ち鍵盤右上方にあるアラームランプが点灯し右前方の放音窓内部のブザーが鳴動する。それ以上の間隔でタッチされた時は、先に押されたキーに対応する穿孔及び印刷が行われるが、穿孔速度(350字/分)(印刷機接続の場合は120字/分)を越えた場合のみアラームが出る。このアラームを解除するため「アラーム断」押ボタンが鍵盤部右下に設けられている。

——後略——

6 印 字 部

——前略——

6-3 読 取 部

8単位紙テープに穿孔された7単位符号を読取り、その信号を印刷部又は穿孔部に送信し、同テープの複製又はモニター印字を行わせるものである。テープは進行方向に対して横より挿入し、読取った符号は直ちに偶数パリティチェックを行い、もし読取った符号が奇数（オールスペース、オールマークは除く）であれば、読取動作を停止して制御部にあるパリティチェックランプを点灯してブザーを鳴動させる。XYチェックはX側読取りに於てJ符号(2, 3, 4, 5単位穿孔, (41)に相当する)があればこれであらうアラームを出す。(X側符号にJ符号はない。)^{注)1}

リセットはテープを訂正するか、正しい順序に掛けなおしてリセットボタンにて再スタートする。テープが無くなった時には、テープアウトコンタクトにより自動的に動作を停止する。オールマークを読取った場合は、印字も穿孔もしない。

—後略—

注) 1 漢テレコード7単位2列のうち、奇数番側(第1列)をX側、偶数番側(第2列)をY側として、常に奇数番、偶数番の順に、2列をくみあわせて、1漢テレ字のコードとしないと、正しい印字ができないので、紙テープのかけまちがいや、パンチミスをチェックするために、このXYチェック機能がある。

国研用漢字テレタイプと同機利用 の言語情報処理(附表図)

(表 1 a)

印 字 機 構 側				⑧	鍵 盤 穿 孔 機 側		
バレット 番号	バレット 指定 コード 1 2 3 4 5	シフト 6	3010 (オクタル)コード		リレー 転換	穿孔テープ 3010 (オクタル)	行番号 (上から)
25	1 0 1 1 1	0	42	K			1
		1	02	2			
24	0 0 1 1 1	0	43	L			2
		1	03	3			
23	1 1 0 1 1	0	44	M			3
		1	04	4			
22	0 1 0 1 1	0	45	N			4
		1	05	5			
21	1 0 0 1 1	0	46	O			5
		1	06	6			
20	0 0 0 1 1	0	47	P			6
		1	07	7			
19	1 1 1 0 1	0	50	Q			7
		1	10	8			
18	0 1 1 0 1	0	51	R			8
		1	11	9			
17	1 0 1 0 1	0	52	EI	**	← (マイナス)40	9
		1	12	スペース		← 0 (ゼロ) 00	
16	0 0 1 0 1	0	53	\$			10
		1	13	#			
15	1 1 1 1 0	0	60	∩			11
		1	20	&			
14	0 1 1 1 0	0	61	/			12
		1	21	A			
13	1 0 1 1 0	0	62	S			13
		1	22	B			
12	0 0 1 1 0	0	63	T			14
		1	23	C			
11	1 1 0 1 0	0	64	U			15
		1	24	D			
10	0 1 0 1 0	0	65	V			16
		1	25	E			
9	1 0 0 1 0	0	66	W			17
		1	26	F			

8	0 0 0 1 0	0 1	67 27	X G				18
7	1 1 1 0 0	0 1	70 30	Y H				19
6	0 1 1 0 0	0 1	71 31	Z I				20
5	1 0 1 0 0	0 1	72 32	EB +	*	←→	(15	21
4	0 0 1 0 0	0 1	73 33	,				22
3	1 1 0 0 0	0 1	74 34	% ;				23
2	0 1 0 0 0	0 1	75 35	Iss :	*	←→) 16	24
1	1 0 0 0 0	0 1	76 36(アポストロフ)	=				盤外(下) 1 (下)
26	0 1 1 1 1	0 1	41 01	J I	(*)			盤外(上) 1 (上)
(1 2 4 8 12)							空欄は印字側 と同じ内容	

(表 1 b)

印字機パレット内列指定				Ⓢ	鍵盤穿孔機盤面		
右から 列番	列番指定 コード 6 5 4 3 2	シフト 1	3010 (オクタル)コード		リレー 転換	穿孔テープ 3010 (オクタル)	左から 列番
24	1 1 1 1 1	1 0	00 01	0 1			1
23	1 1 1 1 0	1 0	02 03	2 3			2
22	1 1 1 0 1	1 0	04 05	4 5			3
21	1 1 1 0 0	1 0	06 07	6 7			4
20	1 1 0 1 1	1 0	10 11	8 9			5
19	1 1 0 1 0	1 0	12 13	スペース #	(*) (*)	←→ B ←→ C	22 23
18	1 1 0 0 1	1 0	14 15	(a) ((*) (*)	←→ D ←→ E	24 25
17	1 1 0 0 0	1 0	16 17) not used	(*) (*)	←→ F ←→ G	26 27
16	1 0 1 1 1	1 0	20 21	& A			6
15	0 1 1 1 1	1 0	40 41	(マイナス) -			13
14	0 1 1 1 0	1 0	42 43	K L			14

13	0 1 1 0 1	1 0	44 45	M N					15
12	0 1 1 0 0	1 0	46 47	O P					16
11	0 1 0 1 1	1 0	50 51	Q R					17
10	1 0 0 0 1	1 0	34 35	∴					12
9	0 1 0 0 1	1 0	54 55	* ED	(*)	↔	H	30	10
8	0 1 0 0 0	1 0	56 57	EF not used	*	↔	I	31	
7	0 0 1 1 1	1 0	60 61	∕	*	↔	+	32	11
6	0 0 1 1 0	1 0	62 63	S T	*	↔	.	33	
5	0 0 1 0 1	1 0	64 65	U V					18
4	0 0 1 0 0	1 0	66 67	W X					19
3	0 0 0 1 1	1 0	70 71	Y Z					20
2	0 0 0 1 0	1 0	72 73	EB ,	*	↔	(15	21
1	0 0 0 0 1	1 0	74 75	% Iss	*	↔)	16	22
	(9 8 4 2 1) ウエイト						空欄は印字 機側と同じ		23
									24

図 1

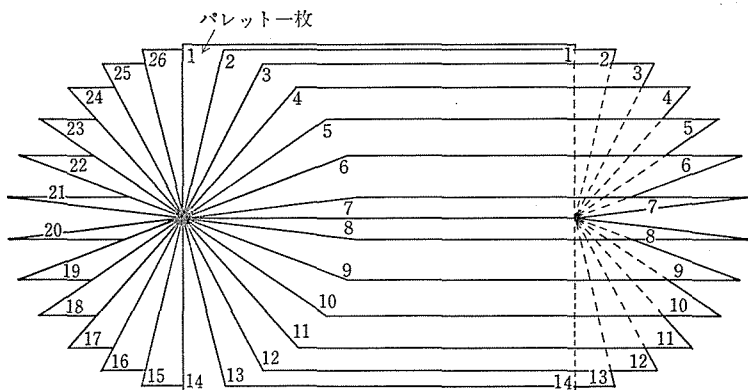


図 2

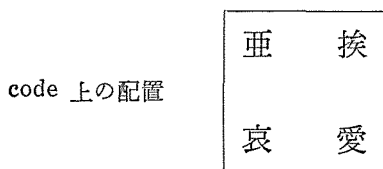
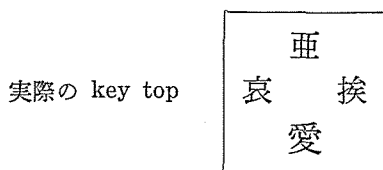


表 2

盤 面	列 番	1		2		3		4		5		6		7		8		9		10		11			
		左	右	左	右	左	右	左	右	左	右	左	右	左	右	左	右	左	右	左	右	左	右		
行	番号	00	01	02	03	04	05	06	07	10	11	15	16	20	21	22	23	24	25	26	27	30	31	32	33
25	01	ド	平仮名 長音	！	？	「	」	『	』	=	-	{	¶	…	a	b	c	d	e	f	g	h	i	→	
21	32	片仮名 長音	サ	ア	ザ	シ	ネ	イ	ジ	イ	ス	ハ	ウ	ズ	ラ	ウ	セ	ゼ	ガ	ソ	エ	エ	ゾ	おたふる	
22	33		ニ	ア	ザ	シ	ネ	イ	ジ	イ	ス	ハ	ウ	ズ	ラ	ウ	セ	ゼ	ガ	ソ	エ	エ	ゾ	おたふる	
23	34		に	ア	ザ	シ	ネ	イ	ジ	イ	ス	ハ	ウ	ズ	ラ	ウ	セ	ゼ	ガ	ソ	エ	エ	ゾ	おたふる	
24	35		め	ア	ザ	シ	ネ	イ	ジ	イ	ス	ハ	ウ	ズ	ラ	ウ	セ	ゼ	ガ	ソ	エ	エ	ゾ	おたふる	
25	36		1	ア	ザ	シ	ネ	イ	ジ	イ	ス	ハ	ウ	ズ	ラ	ウ	セ	ゼ	ガ	ソ	エ	エ	ゾ	おたふる	

盤 面	列 番	12		13		14		15		16		17		18		19		20		21		22		23		24	
		左	右	左	右	左	右	左	右	左	右	左	右	左	右	左	右	左	右	左	右	左	右	左	右	左	右
行	番号	34	35	40	41	42	43	44	45	46	47	50	51	60	61	62	63	64	65	66	67	70	71	73	74		
25	01	<	>	～	j	k	l	m	n	o	p	q	r	[s	t	u	v	w	x	y	z	(¶	↓	↑	↓
21	32	オ	タ	カ	チ	ガ	チ	キ	ツ	ギ	ツ	ク	ク	グ	コ	ケ	ゲ	ゲ	コ	ゴ	ゴ	ナ	ミ	□	↓	↑	↓
22	33	オ	タ	カ	チ	ガ	チ	キ	ツ	ギ	ツ	ク	ク	グ	コ	ケ	ゲ	ゲ	コ	ゴ	ゴ	ナ	ミ	□	↓	↑	↓
23	34	オ	タ	カ	チ	ガ	チ	キ	ツ	ギ	ツ	ク	ク	グ	コ	ケ	ゲ	ゲ	コ	ゴ	ゴ	ナ	ミ	□	↓	↑	↓
24	35	オ	タ	カ	チ	ガ	チ	キ	ツ	ギ	ツ	ク	ク	グ	コ	ケ	ゲ	ゲ	コ	ゴ	ゴ	ナ	ミ	□	↓	↑	↓
25	36	オ	タ	カ	チ	ガ	チ	キ	ツ	ギ	ツ	ク	ク	グ	コ	ケ	ゲ	ゲ	コ	ゴ	ゴ	ナ	ミ	□	↓	↑	↓

表 3 理論コード用漢字表

()内は4桁の8進数表示のコードで、以下次に4桁表示があるまで下の2桁は、01, 03, 05, のように累加され、73でおわり、上の2桁が変わって、同様にくりかえす。

10位 \ 1位	1	2	3	4	5	6	7	8	9	10
0	計 (0001)	形 (-03)	型 (-05)	経 (-07)	芸 (-11)	頭 (-16)	決 (-21)	結 (-23)	月 (-25)	建 (-27)
1	見 (-31)	件 (-33)	県 (-35)	伝 (-41)	電 (-43)	都 (-45)	土 (-47)	度 (-51)	当 (-61)	投 (-63)
2	東 (-65)	党 (-67)	島 (-71)	等 (-73)	愛 (0201)	悪	安	案	以	世
3	位	意	違	育	引	員	右	身	信	新
4	親	深	進	人	図	水	救	井	云 (0301)	運
5	英	映	影	石	衛	益	円	遠	演	応
6	横	生	性	正	成	青	声	制	西	政
7	勢	製	奥 (0401)	屋	音	下	化	選	果	加
8	何	花	家	火	夏	赤	切	折	接	設
9	説	千	川	先	船	戦	価 (0501)	過	歌	画
10	回	想	会	改	海	界	開	解	外	線
11	全	前	然	素	組	早	争	草	送	相
12	各 (0501)	客	格	学	楽	対	活	割	感	監
13	間	関	元	総	増	足	側	統	村	他
14	多	打	太	体	眼 (0701)	顔	気	記	起	男
15	帰	期	機	義	議	吉	久	待	隊	態
16	大	台	代	第	題	沢	達	団	球 (1001)	究
17	急	宮	居	調	共	供	協	強	教	橋
18	京	段	知	地	置	着	中	注	町	長
19	張	朝	興 (1101)	業	局	玉	近	田	金	銀
20	苦	空	君	軍	係	直	追	通	定	程
21	的	鉄	天	店	点	転	肩 (1301)	研	権	言
22	限	入	原	現	滅	古	個	呼	戸	同
23	動	道	働	特	得	読	内	南	難	日

24	後 (2001)	語	御	口	工	阪	公	広	行	交	
25		光	向	好	熱	年	能	農	派	配	売
26		買	白	発	反	校 (2101)	考	江	高	号	表
27		合	谷	告	国	黒	今	根	半	判	番
28		彼	非	飛	費	美	備	必	百	婚 (2201)	左
29		佐	座	才	福	妻	濟	裁	最	細	際
30		在	評	病	品	不	付	布	夫	父	婦
31		部	風	作 (2301)	昨	殺	山	参	歩	産	残
32		士	子	氏	止	支	物	分	文	聞	平
33		兵	米	別	編	変	返	仕 (2401)	市	死	糸
34		私	枚	使	始	指	思	紙	資	字	保
35		母	法	放	方	縫	望	北	僕	本	毎
36	次 (2501)	自	事	治	持	目	時	式	失	室	
37		質	実	写	万	満	味	民	務	無	名
38		命	明	面	木	社 (2601)	車	者	若	手	余
39		主	取	株	種	受	収	秋	門	間	夜
40		野	役	約	由	友	有	優	予	終 (2701)	集
41		住	重	出	立	術	春	初	処	所	書
42		女	用	洋	要	容	葉	様	来	落	利
43		理	喪	助 (3001)	小	少	昭	将	話	消	笑
44		商	勝	上	丈	条	流	両	良	料	力

表 4 漢字コード基準表

部首グループ (先頭字)	コード第1字 (グループ内に 共通)	グループ先頭 字の通し番号	グループの 基準番号	コード第2字 (先頭字に与え られるもの)
一	1 計	1		1 計
丨	2 形	67		67 声
丶	3 型	91		91 説
ノ	4 経	106	100	06 頭
乙	5 芸	161	100	61 横

ジ	6 頭	224	200	24 等
二	7 決	247	200	47 激
ㄥ	8 結	286	200	86 赤
人 1	9 月	344	300	44 人
(備) 2	10 建	751	750	1 計
(機) 3	11 見	1201	1200	1 計
儿	12 件	1336	1300	36 員
入	13 県	1415	1400	15 電
八	14 伝	1450	1400	50 運
口	15 電	1506	1500	06 頭
ㄱ	16 都	1565	1500	65 成
ㄴ	17 土	1607	1600	07 決
儿	18 度	1737	1700	37 右
口	19 当	1800	1700	100 画
刀 1	20 投	1845	1800	45 図
(翻) 2	21 東	2251	2250	1 計
カ	22 党	2288	2200	88 折
勺	23 島	2493	2400	93 川
匕	24 等	2570	2500	70 政
匚	25 愛	2595	2500	95 船
匚	26 悪	2672	2600	72 製
十	27 安	2695	2600	95 船
卜	28 案	2774	2700	74 屋
卍	29 以	2836	2800	36 員
厂	30 世	2890	2800	90 設
ム	31 位	3057	3000	57 円
又	32 意	3115	3100	15 電
口 1	33 違	3227	3200	27 安
(哪) 2	34 育	3651	3650	1 計
(曝) 3	35 引	4101	4100	1 計
(尊) 4	36 員	4551	4550	1 計
口	37 右	4675	4600	75 音
土 1	38 身	4867	4800	67 声
(堀) 2	39 信	5251	5250	1 計
士	40 新	5638	5600	38 身
女	41 親	5691	5600	91 説
(中 略)				
俞	271 合	48882	48800	82 花

漢テレ仕様付記

昭和42年度に、国立国語研究所で新しく発注した漢テレは、その仕様において従前のものと相違している点がある。その主な点について付記しておく。

1. 印字ずみ用紙まきとり部の付加。
2. リード・アフタ・パンチによるせん孔チェック機構の付加

上記のうち、2は重要な変更である。すなわち40年度に設置した漢テレにおいてはせん孔の誤り(機械自体の)が多く出たので、プログラム開発の上でもこの点に関する特別なチェック・プログラムを開発する必要があった。(本報告書所収 木村繁「漢テレ入力データのチェック」参照) このようなトラブルを減少せしめることを目的として、新しい漢テレについてはリード・アフタ・パンチによるせん孔チェック機構を付加することにした。これは次のような機能をもつ。すなわち、各6ビットパンチ終了後直ちに機械自身が検孔し機械自身による誤せん孔をチェックする。もし誤まりがなければ次に移る。誤りがあれば漢字1字分せん孔後全キーがロックしアラーム表示ランプがつく。このばあいせん孔者はアラームを解除し、後退キー、抹消キーを押すことによって誤せん孔を紙テープの上から消すことができる。

この機構付加の結果処理速度は次のように変わった。

- (1) けん盤部打けん せん孔 500列/分 (250字/分) $\pm 10\%$
 - (3) けん盤部機械読取りせん孔 500列/分 (250字/分) $\pm 10\%$
- 他の(2)(4)(5)は変わらない。(77頁4—2参照)