An application system for Kanji lineprinter and computer

漢字プリンターを使用した ターンアラウンドシステム

斎　藤　秀　紀

§0　序

現在国立国語研究所（以下国研）では現代語の語彙の実体を知る目的を持って計算機を使用した大量の新聞の語彙調査を行ってきた。これらは昭和41年に起案され，昭和48年度をもって終了する。計算機を使用した文字処理は当初種々の問題があったが，その一番の難問は入出力装置に関するものであった。また我々が使用した計算機は，現在では旧型機となっているが，ＨＩＴＡＣ-3010計算機であり，入出力装置として漢字テレタイプライターと付属のモニター装置のみであった。これは従来の計算機用に開発された専用出力装置と比較した場合，非常に低速であり，計算機とのインターフェースは全て紙テープによるオフラインで行なうものである。これは今までの計算機処理の中核が数値計算に力点がおかれていたため，入力出力装置についても一応この方面の需要を満たすことに重点がおかれていたとしてもやむをえないことであろう。これらは我々にとっては計算機処理のメリットを減ずる結果となったことは否めない。しかし，この2～3年漢字処理に対する需要が増し，特にダイレクトメール等の宛名印刷，計算機と写真技術を使った自動組版処理，また文献編集等，直接漢字情報を扱う処理が増加し，それと共に入出力装置の開発が種々のメーカによってなされる様になってきたのも事実である。本稿では，これらの点を考慮し，漢字の出力装置の一つである漢字プリンタ装置の国研作業に対する応用の試案と問題点について述べるものとする。なおＨＩＴＡＣ-3010は48年度中に新型の計算機ＨＩＴＡＣ-8250（98ＫＢ）にリプレースされる。また新機種への移同と共に49年度から新プロジェクトが発足する。本稿での試案は，このプロジェクトに対しほぼ入力部分をおおうことになると思われるが，その他間

接的にシステム設計に対する方向付けを示すことに繋がる。

§ 1　語彙調査システムの概略

図1は計算機を使用した語彙調査システムの作業行程の概略である。作業の中で人手にたよる部分は前処理と入力作業，中間処理と再入力処理の二部分四ヵ所である。ここに，付加情報，単位切り等調査の対象となる全ての情報付けとそれにともなうパンチ作業が集中して行なわれる。さらに中間処理による再パンチ作業がデータの扱いを複雑にしている。問題点はこの部分に生じやすく，全ての付加作業を人の判断で処理することは微細な所まで目がとどく反面，対象が多量である場合，作業方法そのもののあり方まで考慮しなければ，発生する誤りに対する制御がむずかしくなる。また，再入力のさいに出力された情報について清書，パンチ作業と前処理と同様の処理を行なわれなければならないことは，中間で発生するデータの誤りの問題から出力されたデータ数と再入力されるデータのマッチングをとる操作がむずかしくなる点をさけることができない。本稿で述べるシステムについては，この部分の省力化とエラーデータを極力おさえるための方法について述べるものである。

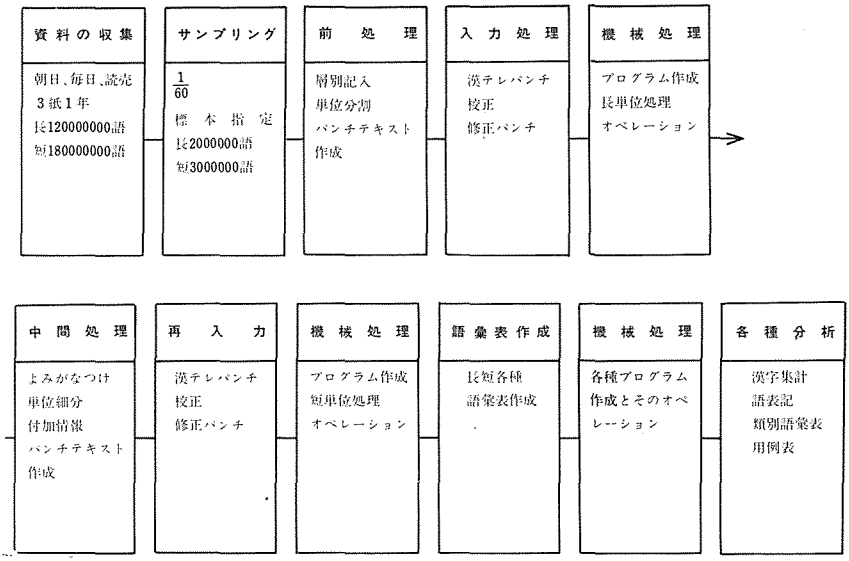


図1　機械処理による用語調査の過程

まず調査は，新聞のサンプリングを計算機で処理し，標本指定された部分を原稿用紙に清書する。そのさい，新聞名，発行日，朝夕刊の判定記号，区画ブロック，層別として何面の記事か，等が単位切りの分離記号と共に付加情報の清書作業として行なわれる。以後は数回の校正作業の後，漢字テレタイプライターを使用し原稿通りのパンチが行なわれる。パンチ済みデータはモニターにかけられ校正漏れのチェック，打鍵のさいのパンチミスの有無がさらに調べられ，この後計算機処理に入る。以上が語彙調査システムの前段階作業の概略である。以後，後処理として中間処理（これを我々は短単位処理と呼び前述の処理を長単位処理と言う）のためのMT等に記録されている情報を編集し漢字テレタイプを使って作業台帳を作成する。中間処理からの作業は短単位処理の対象となる品詞，語種，読みがな，語構成情報，活用形の有無等が手作業として作業台帳に記入される。次に前処理同様，清書に廻される。

以上が従来の作業の概要である。これらはシステムの長所ともなっている部分であるが，原文以外の他の全ての調査対象となる情報を計算機入力以前に人手の作業の形でプレエディットを行ない，清書の後にこれをパンチするという二重の操作を必要とした。当然の結果であるがこの段階で発生する誤りを減らすためには大きな労力を必要とする。本システムにおける目的の一つはこの部分の省力化を目ざすものであることは前にも述べたが，これは入力時の原文のみのパンチを行なうことが可能になることによって従来のプレエディットで必要であった付加情報作業を入力時に処理対象からはずすことが可能となり，次の章で述べる利点が得られる。

§ 2　OMRシートの特徴と利点

OMR (Optical Mark Reader) はある特定の位置にあるマークポジション上につけられたマークの有無を光学的に読み取る装置で従来の入力方式がカードまたは紙テープに情報をさん孔する操作が必要であったのに対し，ペンまたは鉛筆等で必要な部分にマークをつけるのみで良い等，パンチ装置の必要がなくなる利点がある。また特別のパンチ装置が不必要であることは，入力データ作

成のさい，シートを直接付加情報付け作業の帳表として使用でき，異った場所において，例えば調査者個々人に対し独立に情報付加ができることになる。本稿では，この入力媒体であるOMRシートを逆に計算機の出力媒体に使用すると同時に一種の記録媒体としても共用することによって，必要情報のある部分は直接計算機処理の対象とし，入力作業の縮少とエラー修正，自動化，プレエディットの消力化その他計算機のリプレースに伴なうデータコンバートの仕事量を減らすことを目的とする。OMRと類似したものにOCR (Optical character Reader) があるが，これは直接印字されている英数字を光学的に読み取ることが可能である点を除き，パンチ作業の省力化を図る点ではOMRに類似している。本システムにおいてはOCR，OMR共に本質的な違いはない。

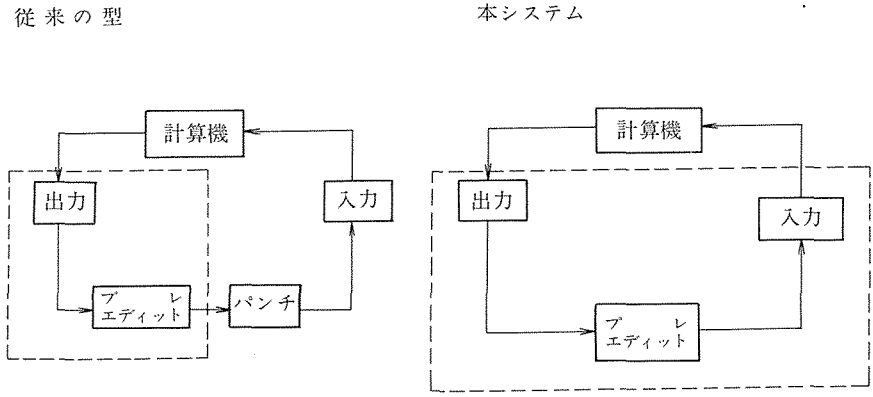


図2

従来の型は新聞の語彙調査に用いられた方法に近い。国研の場合，出力は通常のLPまたは漢テレに出力されたものを使用し，入力は全て紙テープにパンチされた後，計算機処理に廻される。出力媒体の一部，漢テレで出力されたものが作業台帳として使用されるのみである。本システムにおいては，前述の通り出力媒体，入力媒体エディットのための作業台帳が同一媒体で共通化される。

§3　新システムの概要

図3に示したシステムフローの概略を説明する。システムは三つのルートに

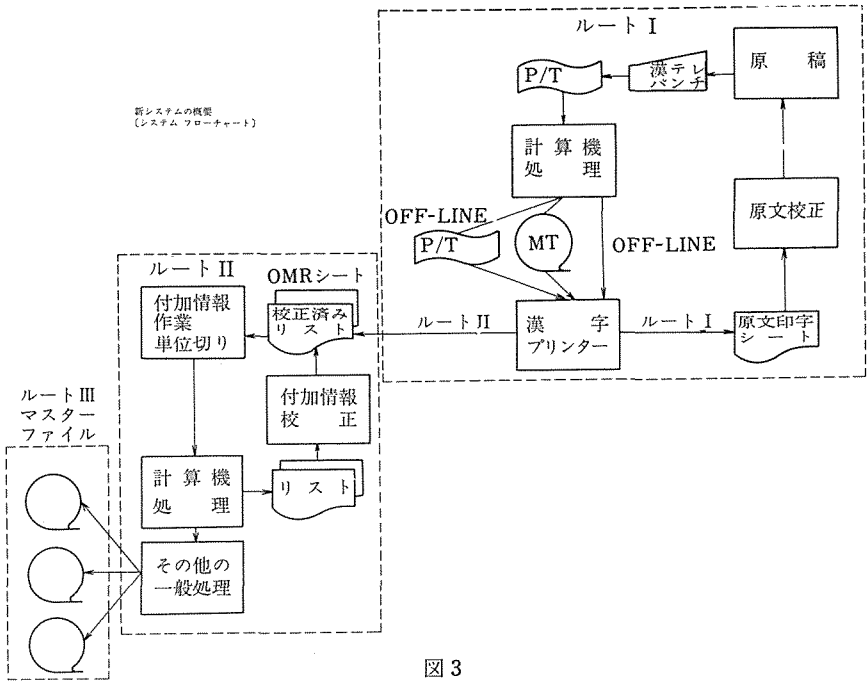


図3

分けられているが，調査システムを対象とした場合，データ入力部と付加情報作業部分をカバーする。ルートIはこの入力部分のうちデータ入力と漢テレさん孔による打鍵ミスの修正作業を主として行なう。データ入力の方法は前章に概略を示したが，調査資料をそのまま手を加えずパンチ業務に廻す。従来の処理は付加情報の挿入作業もプレエディットとしてパンチ以前にまとめて処理しようとしたため清書作業を省くことができなかったが，原文のみの入力ではこれらの清書作業を必要としない。実験では，単行本程度のものであってもほとんどパンチャーの負担とはなってはいない。むしろ，原文入力の方法により，清書作業が省略となるため，付加情報の挿入を必要とせず漢テレで入力されるデータを大巾に減少させることが可能であること。また，それと共に清書のさいの転記ミスとパンチミスを防ぐことが可能となる。

次にこれらの入力データのエラーチェックの問題を述べる。漢字プリンタで出力された校正用データはOMRシートまたは通常の出力用紙に印字する。校正作業はこれらのシートを作業台帳に処理されるが，ここで修正方法は次の場合が考えられよう。

1) シート上の校正結果を参照し紙テープの修正をオフラインで行なう。

2)出力のさい校正処理をあらかじめ想定し,オンラインで修正処理を行なう。

1) の場合は従来の調査で用いられてきた手段であり,比較的人間よりの作業となる。しかし校正処理の速度は作業者の人数とそのための装置の数に制限される。

2) の場合は,修正箇所を文字,語,文等,修正単位を比較的自由度のある指定ができ,データの削除,挿入が前者に比べ処理しやすい利点があり,より強力であると言える。しかし,そのために使用される計算機の使用時間は大巾に増加する欠点を持つ。これは後述するが,漢字プリンター内のミニコンを利用し,メインコンピュタから独立したシステム構成内で修正機能を持たせることが可能となればこれは一応解決する。またディスプレイ等を接続することによってman-machineの相互制御機能を付加し,これを修正機能と結びつけることが将来可能となれば漢字プリンターとしての機能から,より拡大された処理能力を持つことが期待できよう。また2) を満足するためには,プリンターの出力形式を修正しやすい様,編集と帳表形式を考慮すべきであろう。

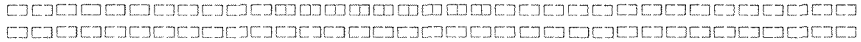
次にルートIIではルートIで校正されたデータをOMRシート上に出力し,付加情報としての単位切り作業処理のため図4の2の形式で出力する。単位切り処理は4の方式で行ない,計算機に再入力する。ここでルートIIで対象となる作業は二種ある。その一つは,単位切り作業であり,他の一つは,付加情報の処理である。これらの処理に対処するためにKWICまたはKWOCの特徴を述べる。まずKWICの場合,一センテンス内での用例の中に見出し語をうめこむ形で印字される。この形式では印字用紙の一行当りの文字数で固定されることから一文の用例の収録範囲が限定され,この巾を越える用例は切り捨てとなる。二行以上にわたって印字する場合もあろうが,KWICの特徴である見出し語の見やすさという利点は失なわれる。また調査分析のための付加情報はソートキーの対象となるにしても,見出し語に対する論理コードの一種として表面には出ない。印字の必要性がある場合は次の頁に見出し語と対応可能な方法を考慮しなければならない。KWOCの場合,見出し語は用例文の先頭に

（1）原文（漢テレで入力される形）

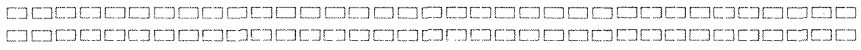
東京で開いている関税貿易一般協定（ガット）閣僚会議の日本政府代表団筋が十三日朝明らかにしたところによると，同会議の集会である開発途上国問題について……（以下略）……　　（昭和48年9月13日朝日夕刊より）

（2）OMR シートに出力された形

東京で開いている関税貿易一般協定（ガット）閣僚会議の日本政府代表団筋が



十三日朝明らかにしたところによると，同会議の集会である開発途上国問題に



ついて～（以下略）9月13日朝日夕刊

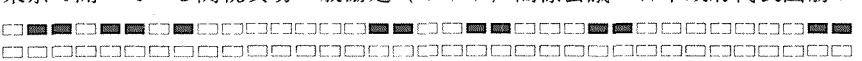
（3）長単位による単位切り処理（｜は分離記号）

東京｜で｜開い｜て｜いる｜関税貿易一般協定｜（｜ガット｜）｜閣僚会議｜の｜日本政府代表団筋｜が｜十三日｜朝｜明らかに｜し｜た｜ところ｜に｜よる｜と｜，｜同会議｜の｜集会｜で｜ある｜開発途上国問題｜に｜つい｜て｜……（以下略）……

（昭和48年9月13日朝日夕刊より）

（4）OMR シート上での同一単位切り処理

東京で開いている関税貿易一般協定（ガット）閣僚会議の日本政府代表団筋が



十三日朝明らかにしたところ

図　4

出典情報，付加情報と共に印字される形式であるため他の情報との対応付けはKWICよりも有利であろう。ただし分類処理の第一キーによって配列される論理的同一集団の面的パターン構造からの比較概念の優利性は失なわれる。以上の点から，KWOCの場合はあくまで情報記録の作業台帳としての性格が強く，調査そのものの作業に対し情報の付加等の記録または整理台帳としての機能を主としている様に思われる。反面KWICは単なる分類処理または分類概念から見出し語のある規則のもとに分類された集団の中に個々の他のセンテンス内で使用されている語の用法と広がりを見ることが可能であり，分類概念から計量化の過程である比較概念のレベルで処理することが可能となる。もちろん記録その他の作業を主体とした場合分析作業の道具としてKWICの必要性は変わらない。以上がルートIIでの概略と印字形式の長短の問題点である。

次にこれらの単位切り処理の修正方法の概略を述べる。単位切りの誤りを修正する場合，切るべき所を切らなかった場合と逆に切りすぎた場合が考えられる。前者に対する処理は改めてその位置にマークを挿入することで解決する。後者の場合，基本的には，前に処理されたマークを消すことによって修正できるが，この場合再度シート単位の単位切り処理をしなおさなければならない。他の方法としては，修正処理を次の様に考えることによってシート上の修正ポジションを付加する方法がとれる。まず修正は単位を切ること，切りすぎた単位の分離記号を消去することの作業であるとし，シート設計のさい印字されるデータの個々に別の行を与えておく。

1）東京|で|開い|て|いる|関税貿易一般協定|～

切らなかった場合の例

2）東京で|開いて|いる|関税貿易一般協定|～

この場合必要な個所に分離記号を挿入することで解決する。実際は図4で示された単位切り情報付加の欄にマークを印すのみでよい。

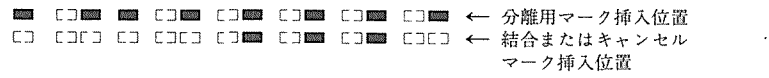
切りすぎた場合の例

3）東京|で|開い|て|いる|関税|貿易|一般|協定|～

この場合「関税貿易一般協定～」までを一単位と認め中間の分離記号を削除す

る。これを次の様に処理する。

4)|開い|て|いる|関税|貿易|一般|協定|～



これで「税」「易」「般」の位置の分離マークは下部のキャンセルマークによって削除され「関税貿易一般協定」は一単位となる。

次に付加情報に関する修正であるが，形式的な組合せまたは対応記号の範囲等から判定可能なものは計算機によるチェックを行なうが，内容面に関するものには，自動的に処理する方法はまだない。自動構文解析処理が可能となれば，この方面の自動チェックへの可能性もあろうが，その間は人手処理による確認にならざるを得ない。むしろ当初は，情報付加に対する作業規則の方法を，対象にし，極力形式的に決定できる様アルゴリズムの明確な記述法とそれらの構造を考慮すべきであろう。これは誤りを減少させるための手段として有力な方法の一つになりうる。またOMRシート上に設計される調査項目の設定は，明らかに調査そのものの作業過程と項目間の論理的結合関係を暗目に含んでいるためであり，その点では調査そのもののアルゴリズムを反映しているためである。また他の特徴は，類似した調査全体にも言えることであるが，調査と計算機の関係，特に計算機の問題をまったく考える必要のない点である。調査の論理性は全てシート上に反映されると見るならば，調査精度のポイントは全てシート上の項目設計作業と作業規則書の作成に重点が置かれ，全ての情報は処理結果としてシート上にダンプされると同様だからである。この場合計算機は単なる情報編集と印字装置として見ることが可能となる。これは計算機の持つハード，ソフトの問題と類似したレベルでとらえることができよう。計算機の場合，問題が与えられ，それらをプログラム可能な手順に分解し，機械的な処理として実行させるという操作をとる。ここに「問題」をプログラム可能な手順に分解する操作を，調査に対する作業手順または規則，また計算機の特性をシートの持つ二次元的特質「面」と対比させて見ることが可能であろう。これは，シートの情報記述方法と階層構造の情報表現がシート設計方法と作業進行

の順序による面が多いためである。当然，調査効率はシート上の設計と作業規則のたて方に影響する。

次にルートIIIにおける概略を述べる。

ルートIIIは分析または出版を目的とした編集作業を主とする。我々の場合，語彙表の作成に主力が置かれるが，必要に応じマイクロフイルム等従来と異った記録媒体への記録を行ない，ハードコピーされたものに対しては，将来機械検索と任意の情報の複写処理の可能性を残す様配慮しておいた方がよい。現在ではMT上の不安定な状態で記録の保存を行なうよりは少なくともビジブルでありかつ安定度は非常に高い。資料に対する基本的な態度としてはこれらの点で将来のことも合わせて考えておく必要があろう。これらを一般的に見るならば，ルートIIIはデータの保存と自動化処理また分析の方向を目ざし，ルートI，IIでは自動化のための分析と調査そのものに力点がおかれる。資料の保存はシートIの記録によってもなされているが，これは調査そのものの記録であって，ルートIIIの開いた状態に対し，閉じたものであると言うことができよう。

次に漢字のよみがな付けの問題であるが，文字処理の場合，文字配列，意味，単位分割等，直接漢字のよみと関係することが多い。同語異語の判別また索引等の語配列の順序等も見やすさ，引きやすさの面から無視できない問題であり，これらの大部分は漢字のよみの決定に関するものと言ってもよい。対象によっては比較的機械的に決定できる場合もあるが，例えば漢字の代表音を事前に定め一対一の対応のみで処理できる場合等は，あくまで配列順序程度の場合に許されるにしても読みやすいものではない。以下，本システムにおけるよみの処理方法を示す。基本的には前述の単位分割，情報付加処理と変わらない。漢字の複数個のよみの選択を，シート上の記号のマークの選択と一致させることにより，よみの選択を行なうものである。なおシート上のよみは事前に辞書等によって，原文の漢字によみを付けシート上にダンプする。辞書の作成は，漢テレ盤内字を基本とする。表外字については，漢字プリンターの外字処理サイクルにあわせ，辞書のメンテナンスを行なうならば未収容字種の問題はほぼ解決する。辞書引きの処理は簡単なマッチングですみ技術的にはあまり問題とはな

らないむしろ，よみの選択はそれらを処理する作業者の能力と判断資料の問題が大きい。次にその一例を示す。これはデータ「東京」に対する処理である。

辞書引きの後

東〔トウ，ヒガシ，アズマ〕

京〔ケイ，キョウ，ミヤコ〕

シート上の印字と選択

東〔トウ，ヒガシ，アズマ，〕

□□□□□■□□□□□□□□□□□□□□□

京〔ケイ，キョウ，ミヤコ，〕

□□□□□□□□□□■□□□□□□□□□□

上記の例では，よみの選択はマークされたシンボルのカンマ以前のよみが選択されているが，本質的には他の処理によってもよい。これは対象となる漢字のよみが多い場合，若干選択子が長くなる欠点がある。しかし従来このよみの問題を前処理からはずす場合，シートまたは語番号によってよみを挿入する方法が多かったが，再度の機械処理と再入力による校正の問題はさけられない。

以上のシステムの構成をさらに要約する。このシステムの機能は次の五種にまとめることができる。

1）単位切り作業

2）付加情報作業

3）エラー修正作業（原文データ，付加情報，単位切り　～ｅｔｃ）

4）パンチ作業

5）分析作業

システムの特徴の概略は前述したが，システム構成のモジュール化は，同一モジュールに（例えばルートⅠ，Ⅱ）異なった処理機能を重複してもたせることが可能であり，システム及びプログラム全体の単純化を進めることができる。また作業面からはエラー修正であっても，対象となるエラーの種類が異なれば他のモジュール上での処理対象となり，段階的に，仕事を小さなまとまりとし

て分散させることが可能となる。同時に作業スケジュールも個々の作業行程は比較的独立したものとして見ることができる。これは従来の作業形態が縦の流れの中で逆ピラミット型であったのに比べ，最終作業まで平均的な量となることを意味する。実際処理での特徴は前述の点と重複するものもあるが，次の四点にあると思われる。以下，それらの周囲の問題にふれてみる。

1) データ入力は付加情報なしのまま原文のみの入力でよい
2) 単位切り作業，その他の付加情報作業は全てOMRシート上のマークによって処理されるため入力のための清書の必要性，付加情報のパンチの必要がない。
3) 校正処理が簡単となり，原文以外の付加情報の校正は計算機処理からはずすことが可能となる。
4) 情報をシート上に印字またはダンプすることによって，計算機のリプレースにともなうデータコンバートの省力化が可能。

1) ～3) までの問題は，システムの概要で説明してきたが，4) については，現状の計算機の発展とそれらの周辺装置に対する開発の状態の方向を見る必要がある。通常の業務または一調査のサイクル（語彙調査の場合，新聞で約9年）に合わせて，計算機または周辺装置の変更は作業の進行面では有利となろうが，調査そのもの，特に現状の調査の問題点のかなりの部分が機械によって制約されている場合は，ある程度中間レベルでのリプレースを考慮せざるをえない場合が多い。

これらのリプレースは当然計算機以外の入力装置の場合も同様であり，コード上の問題をさけることができない。方向としては内部，外部コード共に標準化にあるとしても，漢字の場合，調査対象となる資料によっては漢テレ内の基本収容字種の問題が常に生じ，後述の漢字出力装置の収容字種の問題と共に重要である。

以上の点からコード・コンバート問題は非常にむだな労力を必要とするが，本システムではこれらを原文または単位切り済みファイルにとどめることができ，従来の様な大量のMTレベルでの交換の必要はまったくなくなる。(国研の

場合，ＨｉＴＡＣ-3010にともなうＭＴコード変換は現用機が非常に古いため，約100本程度のコード変換に６ヶ月以上かかっている。）

これは処理の基本が，入力された原文に対し一度計算機処理を行ない，ＯＭＲシート上に出力した後，付加情報作業シート上に行なう。これにより，原文に付加される全ての情報はシート上の収容情報の許す限りシート上に記録され，従来のマスターテープが作業の進行と共に内容が変化し種類が多くなるのに対し，シート上にダンプされた情報は直接には，マスターファイル上に関係しないためである。

シート上に印字された漢字情報は，再入力の際，直接読み込み不可能であってもシート上の文番号，または単語番号等によって，原文イメージとマッチング可能であり，この点から漢字の直接情報読み取りの必要性はない。

将来，データがＭＴの形で手に入りやすくなるか，漢字ＯＣＲの開発が進み実用可能となるならば，これはさらに強力な方法となろう。また，プログラムに関しては，語彙調査の様な特定の業務を対象とした場合，若干の編集機能を持つ，印字プログラムのみでルートＩ，IIに関する処理は全て満足できる利点がある。

§ 4　漢字プリンター使用上の注意点

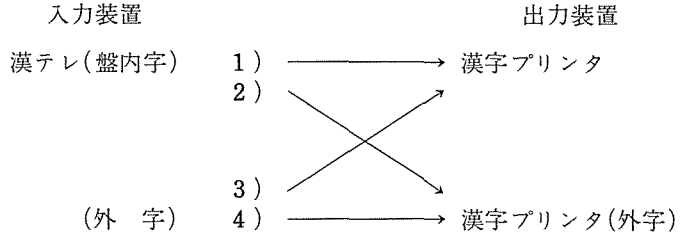
漢字プリンター利用の際，初歩的な問題として，入出力装置個々の収容字種の違いによって生じる外字（サパートされていない漢字記号類，通常は別処理で文字パターン発生のためのソフトが用意されているが，文字デザインの問題がある。）処理がある。漢字プリンターの使用対象によっては，自動組版等で最も印字品質の良いもの，（通常，媒体はフイルムで出力される）また，軽印刷用の版下程度で良いもの（文字品質は比較的良いが媒体は紙の場合が多い）その他，入力データ等のモニター用，宛名印刷のレベルで使用され，一応読める程度で良いものの三種に分けられる。次に印刷時に外字処理の扱いが版下等直接印刷を目的とする場合問題となるが次の方法でさけることが可能であろう。

1）プリンター側の収容字種としてないものは，そのまま従来の文選処理と同

様ゲタをはかせる。また特別の記号類で置き換える。

2)事前にデータ内容の出現字種を調べ外字扱いとなる字種のチェックを行ない出力時に外字の登録を行なう。

これら二種の方法で，大部分はさけることができるが，この問題は単に出力上のものだけではなく入力処理そのものとの関係を無視できない。これは次の様に対応ずけられよう。



メーカー二社の収容字種の調査では 1) 2) については問題とはならない。3)については，出力側は一万字程度の保有があるためこれもほぼ解決できよう。しかし 4)については，入力と出力共に調査対象となる資料に非常に影響をうける。これは調査対象が現代語からはずれてゆくに従い，外字処理の扱いも増加する。特種な用途を対象とする場合は，収容字種の根本的なレベルまで考慮し，資料の基本字種の概略を常に把握しておく必要がある。これは本システムについてもシート上に印字する際に起りうる現象である。

以上で，新システムの概略と外字処理の問題点の説明を終える。プリンター一般の問題であるが，開発中のものを省いて現有の機種は，印字出力を主体に開発されてきたために，情報のターンアラウンド方式を満足できるものはほとんどない。一部ＯＣＲ文字を印字できるものがあるにすぎない。これは今後の問題として解決されなければならないであろうが，国研におけるアプリケーションも，このターンアラウンドの方向で進めることが必要となろう。以下直接問題となる点を列挙しそれらの関係について述べる。

1)　紙質に関するもの。

2)　印字用紙の形式に関するもの。

3) 印字用紙のプレプリントの可否。

4) 漢字プリンターの印字方式に関するもの。

1) に関するものとして，通常のＬＰ用紙として使用される 70 kg～130 kg シートが使用可能であれば，入力側としては問題とならない。2) については，ほとんどの漢字プリンターがロール紙を使用するため，用紙のあまり厚いものが利用できない欠点がある。また入力側に用紙の湾曲による誤動作も問題となろう。これは通常のシート状の用紙，またはミシン目形式のものを利用できれば解決する。しかし，ミシン目のケバをとるためのカッターを使用する必要が生じよう。これはロール紙の場合も同様である。次に 3) の場合，印字の方法がゼログラフィー等の乾式とその他の湿式の二種が代表的であろうが，湿式の場合，プレプリント不能となる。乾式の場合は，逆にコピー時のトナーの汚れが問題となろう。これは 4) の印字方式と直接関係がある。トナーの種類を赤，緑等，OMR装置でセンスされない色であればこれはさけられる。

4) の印字方式では，傾向としては，湿式の場合ブラウン管等の電子的に文字発生ができるものに多く，いわゆる写植機としては，第三世代のレベルのものが多い。これは一般に写植用に開発されたものであるため，文字品質，図形処理等の性能は非常によい。しかし 3) のプレプリント上の問題と重なる。乾式の場合，ピン電極を使用したドットによる文字イメージ発生方式が多いが，使用するドット数により表示される漢字の種類が制限される。また直接用紙に電界破壊方式で焼き付けを行なう方式がある。用紙の汚れの面では，この方式が最っとも良いが，あまり文字品質がよくないという欠点がある。ポイント指定が 9 ポイント程度の小さなものであれば文字品質はあまり目だたなくなる。

その他の入出力装置に関しても（通常のＬＰ，ＣＲ，ＣＰ，ＰＴ等）同様であるが，入出力装置の開発の方向が少なくとも，個々に開いた方向で進められ，記録媒体相互の共通利用可能なものは非常に少ない。これは今後入出力装置の多様化と共にこれらを効果的に利用するために，記録媒体形式の標準化とターンアラウンド可能な装置機能の開発要求が増大するであろう。またこれから，おのずと開発のポイントを変えざるを得ないであろう。最後に漢字の読みがな

付けの問題は，ＯＣＲ文字としてカナ文字処理可能な装置の開発が望まれるが，国産機ではまだ開発されていない。OMR，ＯＣＲ併用方式と共にこれらの処理が可能となれば，本システムもより簡単な方法をとることができる。

§ 6　結び

語彙調査を考える場合，データの蓄積の問題が重要である。データを確実にかつ大量に保存可能な形で考えるならば，現在使用可能な方法としては，磁気ディスクまたはマイクロフイルム等が考えられよう。他の記録媒体に比較してビット当りの価格の点ですぐれているのが磁気テープであるが，長期の安定したデータの保存の点では問題なしとは言えない。通常の保存の場合，一定の周期をおいて複写等の処置がとられるのが普通である。しかし，この複写によってデータの正確度を保っていく方法は，磁気テープの量によっておのずと限界が生じるはずである。我々が新聞の調査で使用した磁気テープの量は，総数2,000本（1,200フィート556ＢＰＩ）になる。通常の処理と平行してこの磁気テープのメンテナンスをかねていくためには非常に多くの計算機使用時間と労力を必要とする。これらはデータを効果的にかつ大量に蓄積するという点では磁気テープを基礎媒体とする資料センターはとうてい不可能に近いと言わねばならない。

また語彙調査およびそれに付随する語彙研究に関してこれらが本格的に研究される様になったのは1950年以後であるとされ，言語または国語の分野においても比較的新しい領域となっている。その意味において方法論，それらを裏ずけるための理論面の本格的研究が必要とされる。また同様にコンピュータによる調査を今後続けて行くならば装置上の開発もそれらをささえるものとしては無視できないであろう。我々には，ある道具によって何ができるかというアプローチの方法も実際処理上では必要となるからである。本稿での内容は今後実験を重ね実用化のための問題もさらに探ることが必要である。またこの試案の「単位切り処理」に関する部分は自動単位切り問題が解決できるまでの繋ぎとして臨時的なものである。最後に，シート上の記録は広義には，読み取り可能

な永久記録媒体であると考えることができ，前述の記録媒体よりは，より安定度は高くなる。しかし紙特有の短所はさけられない点が問題となろう。またこの報告のシステムは人間—機械系のうち，人間系を中心とした処理体系に偏したきらいがあるがこれらは，当然次の段階として自動処理の問題にポイントが移向するであろう。これはそのための前処理用道具として使用できるであろう。

参考文献

1) 斎藤秀紀　電子計算機による語彙調査（国立国語研究所報告 34）

2) 同上　電子計算による語彙調査II（報告 39）

3) 同上　電子計算機による語彙調査III（報告 49）

4) H-8258　マークシート読取装置ＨｉＴＡＣハードウエアマニアル

5) H-8252-2-4形　光学文字読取装置ＨｉＴＡＣハードウエアマニアル