

国立国語研究所学術情報リポジトリ

An application system for Kanji lineprinter and computer 2

メタデータ	言語: jpn 出版者: 公開日: 2017-03-31 キーワード (Ja): キーワード (En): 作成者: 斎藤, 秀紀, SAITO, Hidenori メールアドレス: 所属:
URL	https://doi.org/10.15084/00001037

漢字プリンターを使用した

ターンアラウンドシステム II

斎 藤 秀 紀

0. 序

本稿は「電子計算機による国語研究 VI」で述べた漢字プリンタを使用したターンアラウンドシステム試案に関するプログラム実験の報告である。プログラムシステムはまだ確定的なものではなく、以後さらに改良の必要がある。しかし、システム設計に対する基本的態度は以後開発される一連の漢字プリンタの応用プログラムの共通思想ともなり、語彙調査等の単位切り処理についての従来の方法とは異なる処理の提案ともなるものである。

プログラムおよびシステム開発に関しては、使用機種の特性とそれらの開発指向に影響されるが、特に漢字プリンタについては、使用例、普及度も低く、出力装置としての目的以外に適用業務の中での位置も明確でないことが多い。これはハード、ソフト共に未解決の部分が多く、システム進行の障害となったことは否めない。また、これらは従来の入出力装置の開発が個々に開いた方向で進められ、入出力の統一的な扱いがまだ十分でないことを意味する。

しかし、このプログラムの開発によって、入力のための省力化と共に漢字処理をより人間作業形態に近づけることが可能となり、同時にターンアラウンド処理に対する問題点が明確になる。

1. 入出力問題

従来入出力に伴うターンアラウンド機能の開発はその必要性を認められつつも、使用者の目的にそった装置がメーカー側から常に提供されて来たとは言いがたかった。一つには使用者側のニーズに合った装置の応用とそれらについ

ての実際的な技術的要求が少なかったこともその要因の一つとなっていた様に思われる。しかし、最近の漢字プリンタのあいつぐ発表はコンピュータの入出力機器が欧文を基盤とした処理形態から漢字処理を基盤とした形態に移行する可能性が大きくなってきたと言ってもよい。

漢字処理のコンピュータ化は新聞社、通信社、また最近では電算機写植用に適用されてきたが、装置の大部分は入出力装置共に大型であり、非常に低速であった。これは、装置の一般化以外に比較的制限された人々によって使用されていたことも無関係ではないが、漢字固有の性質が最大の原因となっている。その一つは、字種が欧米のアルファベットに対し非常に多いこと、漢字の表意性からくる同音異義が多いこと、また一つには漢字の字体等にある種の美的観念が社会通念上暗目に要求されていることである。これらが、漢字表現上の簡略化に歯止めをかける要因ともなっていることは否めない。しかし、いずれにしても、出力装置に関しては一応の技術的な面での解決が見られたことは、今後の日本におけるコンピュータ処理に多くの可能性を期待できることになる。

漢字の入力問題としては、出力装置の場合と類似点が多いが、使用漢字の範囲からくる漢字盤内収容字数の制限等、通常のコンピュータの入力装置に比較し大型で低速であるという物理的制約の問題がある。その他、実社会における機器の普及度も同様である。また入力操作そのものは人手による方法以外に自動化の困難な点が多いため極力入力量を減少させる方法が必要となる。これらの問題のうち出力については従来とられてきた方法の一つに漢字テレタイプ付属の印字モニターの低速を補う手段として、高速のプリンターによるカナ印字の使用があった。今まで作成されてきた種々のカナ KWIC、ローマ字 KWIC 等もそのためのものである。通常、調査等の進行過程で増加するデータ量をこれらの方法でカバーするためには、英数字、カナ文字で表現可能な部分はプリンタ側にその処理を分担させることによって、漢字テレタイプへの直接の負担を少なくする方法が多くとられてきた。これは調査で使用される付加情報を記号化すること、また作業過程を段階的に進める様システム設計することが要求さ

れるが次のような問題が生じることになる。

- 1) 作業台帳作成の際、情報の転記ミスが多くなる。
- 2) 記録媒体の変更が繁雑になる。
- 3) 作業過程で中間データの再入力が必要となる。

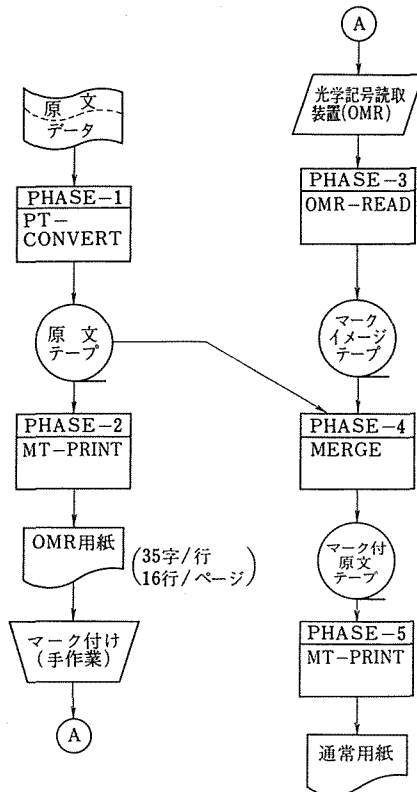
しかし、これらを解決するためには、ターンアラウンド機能を持たせることによって各媒体の共通化を図る。そのためには入出力媒体と装置の規格化が必要となるが、作業帳表、入出力媒体等を統一的に扱うことが可能となり、前述の問題点は一応解決できることになる。現状では、このターンアラウンド機能を持つものとしてはOCR、OMR、バーコード等があるが、バーコード以外には直接漢字処理を指向したものは開発されていない。また漢字プリンタについては出力装置としての単一機能を主体としてきたため印字文字の字間、行間共に微細な調整ができるものは少なく利用できる紙質、紙厚等も使用範囲の制限が多い。これは入力装置側の媒体に関する制限事項がそのままプリンタ側に要求される性能となる。

紙質に関するものとしては、通常のOMR用紙が90 kg～110 kg程度であることから、プリンタ側での任意の紙厚を使用できること、印字の定着による加熱、または定着液による紙の変形が少ないこと、その他、プリンタ側の紙送り機構の精度がOMRシート上の文字の位置ぎめ許容範囲内であること、紙の供給方式はロール状のものよりシート形態が望ましいこと。またこれは、紙の送り機構についてはスプロケット送りの方が精確なポジショニングが可能となる。その他、紙の反り、曲がりのないこと、OMRのマーク位置と一行当りの印字可能文字数が異なるため字間、行間を任意に調整できること。また、印字用紙は帳表設計のためプレプリントが可能であること、PCS(OCR規格で文字と紙の表面とのコントラスト化)が0.5以上であること。プリンター印字方式による紙のよごれが極力少ないこと、OMRマークサイズに相当する文字のポイント

(12~14ポイント程度)が指定でき、文字の大きさに大きなズレがないこと等がプリンタ側に要求される項目である。

その他現状のプリンタでの印字方式ではPPC (Plain Paper Copier の略) 方式の場合、普通紙が使用できることから、紙の費用を比較的安くおさえることができる。しかし、潜像保持媒体の耐久度、その他印字用紙全体に一律のトナー塗布によれ残存トナーの汚れが問題となる。また静電記録紙使用の場合にはPPCに比ベトナーの汚れは比較的少ないが、印字機構を簡単化できるため、逆に文字指定の微細な調整ができにくいものが多い。これらは OMR シートの

図1. テストフロー



マークポジションを出力側規格にあったものに設計するか、または直接出力側を入力規格の使用に固定するか、いずれかの処置が必要となる。

2. プログラムの概略

単位切り作業のためにさん孔された紙テープデータ(テストフロー図1参照)は PHASE-1 でイメージどおり磁気テープに情報が記録され原文資料マスターファイルが作成される。通常この部分では入力データに対し機械チェックを行なうが、処理対象としない場合、用意されているユーティリティルーチンの使用が可能である。ここで作成されたテープは PEASE-2 での単位切り作業台帳作成, OMR シートに印字するための入力テープ, そしてターンアラウンドされて再入力される PHASE-4 のための基本マスターテープともなる。

図2. 入力 (OMR) フォーマット

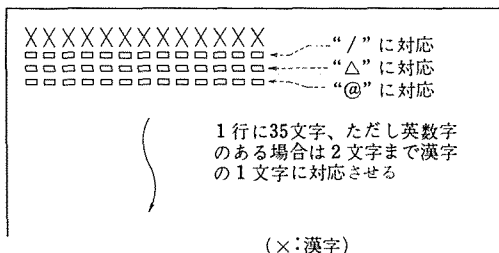
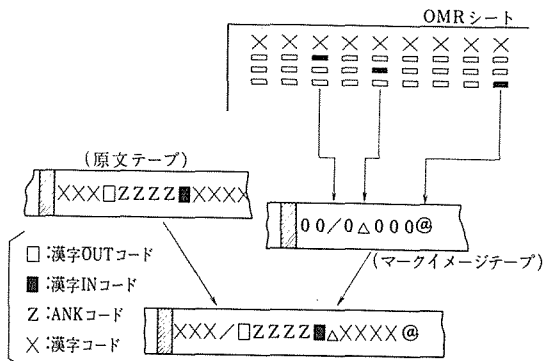


図3. 処理内容



(注)上図の1文字は磁気テープ上の2フレーム
(キャラクター)に対応

PHASE-2 の出力印字形態は図 2, 図 4 に示す形式であるが、ここで単位切りされた付加情報は PHASE-3 において OMR 装置によって再入力される。しかし PHASE-3 で入力される情報はマーキングポジションのみであり、漢字情報は入力されない。漢字情報については人間の作業に参考として使用されるのみであり、コンピュータ上の処理では PHASE-1 の出力マスターテープが使用される。図 3 のマークイメージファイル形式はこのマーキング情報のみのファイル内容を示したものである。ここで作成された磁気テープは PHASE-4 において PHASE-1 の原文イメージの単位切りされていないデータと単位の位置のみ記録されている PHASE-3 の出力との対応付けが行なわれる。この PHASE-4 の出力ファイルが一応単位情報が挿入された最終ファイルとなり、以後印字のための PHASE-5 の入力データとなる。ここでは通常シートに印字され、校正処理等に使用されるシートの作成が行なわれる。(図 5)

以上がプログラムの概略である。OMR シートの項目設計は便宜的に三項目としたが、これらは使用目的によって任意に変更が可能であり、出力側の制約は受けない。通常同時に処理される単位の種類によって決定され、本例では三種の単位切り処理の一例を示したものである。しかし、三項目は校正処理用のデータの挿入、削除等のチェックポイントとしても使用可能となる様配慮した。なお、個々の位置と単位の異なりを示す記号は(図 2, 図 5 参照)マーク位置により「/」、「△」、「◎」を使用した。その他 OMR 入力中に 1 漢字当り 2 マーク以上の指定が同時に記入されている場合「?」記号を出力し、アラームとしての役割を持たせている。これは、エラーとして解釈してもよいが、単位の異なった単語の切れ目が同一部分で切れて重なる場合もあるため、通常記号として扱ってもよい。

PHASE-4 については、照合処理は図 3 に示した形態で行なわれるが、漢字 IN コード、OUT コードは ANK モード(英数字等、半角扱い字種)へのソフトイン、アウトのためのものである。また PHASE-2, PHASE-5 における印字出力例を図 4, 図 5 に示したが、個々に入力コードが異なる原文の入力において

図 4. PHASE-2 OUTPUT

キタ・セクスアリス

金井潤君は哲学が職業である。

哲学者という概念には、なにが書物を書いているということがともなう。

金井君は哲学が職業であるくせに、なんにも書物を書いていない。文科大学を卒業するときには、外道哲学とSOKRATES前のギリシヤ哲学との比較的研究とかいう題で、よほどへんなものを書いたそうだと。それからというもの、なんにも書かない。

しかし職業であるから講義はする。講座は哲学史を受け持っていて、近世哲学史の講義をしている。学生の評判では、本をたくさん書いている先生があの講義よりは、金井先生の講義の方がおもしろいということである。講義は直観的で、ある物の上に強い光線を投じることがある。そういうときに、学生はい

図 5. PHASE-5 OUTPUT

キタ・セクスアリス／

金井潤君△は／哲学△が／職業△である／。／
 哲学△者△と／いう／概念△に△は／、／なに△が／書物△を／書いて△いる△と／いう／こと△が／ともなう／。／
 金井潤君△は／哲学△が／職業△である△くせに／、／なんにも△も／書物△を／書いて△い△ない／。／文科△大学△を／
 卒業△のする／とき△に△は／、／外道△哲学△と／SOKRATES△前△の／ギリシヤ△哲学△と△の／比較△的△研究△と△か／
 いう／題△で△、／よほど／へん△な／もの△を／書いて△た△そうだと／。／それ△から△と／いう／もの△は／、／なんにも△も／書か△ない／。／
 しかし／職業△である△から／講義△は／する／。／講座△は／哲学△史△を／受け△持△つ△て△い△て△、／近世△哲学△史△の／講義△を／し△て△い△る／。／学生△の／評判△で△は／、／本△を／たくさん／書いて△い△て△い△る／先生△が△あの△／
 講義△より△は／、／金井△先生△の／講義△の／方△が／おもしろい△と／いう／こと△で△ある／。／講義△は／直観△的△で△、／ある／物△の／上△に△強い／光線△を／投△げ△る／こと△が△ある／。／そう／い△う／とき△に△、／学生△は△い

は PHASE-2 において漢字プリンタ内の文字発生装置のコード対応表を書き換えることによって任意のコード体系を処理することが可能となる。この場合コード変換処理は、ほぼ半ハード的に行なわれる。これによって、かならずしも規定の漢字テレタイプを使用する必要はなく、基本的には任意のメーカーのものが使用可能となる。プログラム処理としては、さん孔された紙テープまたは磁気テープを直接 PHASE-2 において実行し OMR シート上に印字する。

その他、図 3 の対応はマーキングされたシートと磁気テープ上での記録内容である。4 桁のオクタル表示は個々にセパレート用記号に対応する。オクタルの "0000" は単語内の漢字の位置を示し、出力時に漢字コードと置き換えられる。

なお、このプログラムの中核である PHASE-1, PHASE-4 に関するプログラム条件書をこの稿の終りに示しておいた。また、この実験に使用したコンピュータの基本構成は次の通りである。ただし、プログラム作成の際に使用した通常の高速ラインプリンタ、カードリーダ等は省いてある。

漢字プリンタ装置	1 台
編集用計算機	1 台
磁気テープ装置	3 台
OMR 装置	1 台
紙テープ読み取り装置	1 台

3. 結 び

本稿におけるプログラムについては、今後さらに実験を行なう必要があるが、OMR, OCR 等の利用度の普及から見て、シミ、汚れ等の問題もさらに良くなるものと思われる。同様に漢字プリンターについてもターンアラウンド可能な機能が必要条件の一つとなることも考えられよう。また記録媒体としての紙の問題があるが静電記録紙を OMR 用紙として使用するテストケースとなったが、これもさらに安定した物、例えばマイクロフィルム等からの直接データ入力ができ得れば、データ蓄積、スペース等について現状の磁気テープよりはよくな

るものと思われる。また、このプログラムによって従来とられてきたパンチャーのための清書作業を省くことができ、入力パンチ量を $\frac{1}{3}$ ～ $\frac{1}{6}$ まで減らすことが可能となる。また80欄カード状の連続用紙を使用することもでき、調査帳の作成と結果の直接入力も同様の方法で処理可能である。データの位置ぎめの問題はこの場合は比較的自由に指定でき、特に80欄カードのOMR技能は従来のカード読み取り装置にもオプションとしているものが多いことから実用化には問題は少ないと思われる。しかし、いずれにしても実用化の場合、用紙に関しては相当の神経を使うことが要求され、マーキング等も赤色以外に目を疲れさせないものを使用する様、配慮しなければならないであろう。また用紙の大きさもあまり大きいものは折れ曲がり等の併害が出て来よう。

最後に、このプログラム作成に関しては NEC 官庁システム部の方々に大変お世話になりました深く感謝いたします。

参 考 文 献

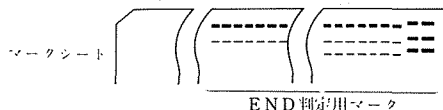
- 1) 斎藤秀紀 電子計算機による国語研究 VI (国語研究報告 51)
- 2) NEAC-シリーズ 2200 周辺装置取扱い説明書

光学マーク読取り装置 N240P-1A

プログラム条件書

1. OMR READ PROGRAM PHASE1

- 1) プログラム名 OMRSM
- 2) 入力条件 プログラム仕様書と同じ
- 3) END判定 マークシートの上段の後から18コマーク



- 4) 出力条件 プログラム仕様書と同じ
- 5) リスタート機能 このプログラムにはリスタート機能があり、その機能は $\bar{O}MR$ が $READ\ ERR\bar{O}R$ をおこしたときに動作を中止し、すべてのユニットを初期の状態にもどす。
- 6) MESSAGE
-
- ***** PARAMETER SW 2MD, BMD, AMD, $\bar{O}N$ ***
- $\bar{O}MR$ のパラメータスイッチ 2MD, BMD, AMD を $\bar{O}N$ にする。
-
- *** $\bar{O}UTPUT\ TAPE\ L\bar{O}GICAL\ ADDRESS\ N\bar{O}\ 1$ **
- マーク・イメージ・テープをロジカル・アドレス 1 にセットする。
-
- ** $\bar{O}MR, TAPE\ SET\ \bar{O}K\ \bar{O}K=TYPEIN\ G$ **
- $\bar{O}MR$ と TAPE の状態をチェックして、 $\bar{O}K$ ならコンソールタイプライターの G をおす。
-
- ***** $\bar{O}MR\ T\bar{O}\ TAPE\ PR\bar{O}GRAM\ START$ *****
- 処理の開始 ($\bar{O}MR\ READ\ START$)
-
- ** $\bar{O}MRREAD\ END\ TAPE\ \bar{O}UTPUT\ START$ *
- $\bar{O}MRREAD$ が終り TAPE にマークイメージを出力開始。
-
- ***** $\bar{O}MR\ T\bar{O}\ MT\ END\ SM$ ****
- 実行の終了
- 7) $ERR\bar{O}R$ $\bar{O}MR$ に $ERR\bar{O}R$ があった場合には次のメッセージが出力される。
-
- *** $\bar{O}MR\ ERR$ *** ☒ *** $RESTART\ TAPEIN\ G$ ***
- このメッセージは機械的エラーの時に出力される。
 - ☒ が 1 のとき ダブルフィード
 - ☒ が 2 のとき ジャムまたはセレクトエラー
 - リスタートするときには機械の状態をしらべて、コンソールタイプライターから G を入力する。
-
- *** $DATA\ ERR$ *** *** $RESTART\ TAPEIN\ G$ ***
- このメッセージはデータエラーの時に出力される。
 - ☒ が 5 のとき マーク記入ミス
 - ☒ が 6 のとき 未記入ドキュメント
 - ☒ が 7 のとき 不正長データ
 - リスタートするときにはドキュメントの状態をしらべてコンソールタイプライターから G を入力する。
- 8) 割込 END $\bar{O}MR$ のインタラプトスイッチ 1, 2 を $\bar{O}N$ にすると実行を中止することができる。このときのメッセージは
-
- *** $WARIK\bar{O}MI\ END$ ***
-

- 9) カウンタの印字 $\bar{O}MR$ が $\bar{C}LOSE$ されるとブロックカウントとファイルカウンタがラインプリンタに出力される。

カウンター内容

TOTAL	読み取った枚数
ACCEPT	アクセプトした枚数
REJECT	リジェクトした枚数
DFD-E	ダブルフィードを起した回数
DEV-E	デバイスエラーを起した回数
WNG-L	不正長の枚数
RED-E	読み取り不能のあった枚数
BLK-M	未記入の枚数
OTHER	DFD-E ~ BLK-Mに含まれないドキュメントの枚数(TOTAL, REJECT, カウンタには加算される)

(百分率は小数点第3位四捨五入)

2. MAS-MIT MERGE PROGRAM

- 1) プログラム名 TTTSM PHASE4
- 2) 入力条件 プログラム仕様書と同じ
- 3) END 判定 マスターテープおよびマークイメージテープの $\bar{E}OF$ 。
- 4) 出力条件
 - ANK 文字中に2つ以上のマークがあった場合には、最後のマークが有効となる。
 - 1ページには45文字、行で30行印字。

- 5) MESSAGE *** MAS TAPE LOGICAL ADDRESS N0 2 ***
 *** MIT TAPE LOGICAL ADDRESS N0 1 ***
 *** PIT TAPE LOGICAL ADDRESS N0 3 ***

• マスターテープのアドレス "2"
 マスターイメージテープのアドレス "1"
 プリントイメージテープのアドレス "3" にセットする。

*** TAPE ADDRESS SET OK? OK=TYPIN G **

• テープのセットが $\bar{O}K$ ならコンソールタイプライターの G をおす

*** MAS-MIT TAPE MERGE START ****

• マージ開始

**** TTTSM END *****

• 実行の終了

**** MIT TAPE READ ERR *****

• マークイメージテープのリードエラー

**** MAS TAPE READ ERR ****

• マスターテープのリードエラー

**** MAS TAPE E \bar{O} F ****

• マスターテープの E \bar{O} F が発見された。

**** MIT TAPE E \bar{O} F ****

• マークイメージテープの E \bar{O} F が発見された。

***** MAS-MIT TAPE MERGE END ****

• マージ終了

***** PIT TAPE \bar{O} UTPUT START ****

• プリントイメージテープに出力開始。