解　説

パターン合成による漢字入出力処理*

長谷川　実郎**

1. はじめに

現在，世界における文字文明としては，ローマ字文化の傘の下の識字人口が約 10 億人と最も多く，漢字文化の傘の下の識字人口約8億人がこれについでいるといわれており，漢字が用いられている国の数は少ないが，その識字人口は世界の人口の約 1/3 を占めている．このローマ字情報と漢字情報との情報化社会に対する適応性を比べてみると，26 種の基本文字からなるローマ字情報は，コンピュータシステムにとって，ハードウェア，および，ソフトウェアの面でも非常に扱い易い言語であり，数十のキイからなる小形で安価なタイプライタというユニークな入出力装置が用いられているのに対し，漢字情報の場合は，文字種が約 2,000～10,000 字種と極めて多く，その入出力処理はローマ字と比べ，経済的，技術的に大きなハンディをもたらされており，タイプライタのような簡潔な入出力装置の利用は期待できそうにない．

漢字情報処理システムにとって，言語上のいくつかの大きな問題点はあるが，基本的には漢字の文字種が極めて多く，かつ，その集合の大きさが不定であるという漢字そのもののもつ特質に問題は起因している．

文字種が多いということにより，入力システムでは文字キイ数が多くなり，訓練に時間を要すると共に入力速度はローマ字に比べ約数分の一になる．また，出力システムでは一文字の絵素数もローマ字に比べ大きく，文字数の増加につれて文字の記憶容量が増大し，装置が高額になる．漢字情報とローマ字情報の入出力特性の比較を**表-1** に，主な各種漢字表と収録文字数を**表-2**（次頁参照）に参考として示す．

漢字情報処理に与えられている最大の課題は，入出力処理の効率の向上と，装置の低価格化にある．このためには，多数の漢字を，すべて同じ条件の一文字として取扱う方法では，文字種の増加につれてそのまま入出力処理効率が落ち，装置が高価格となるので，漢字のもつ特性を利用し，その特性に応じて条件を変えて取扱かい，文字種の増加に対処する方法が考えられる．これらに有効な漢字のもつ特性としては，漢字の使用率と，漢字のパターンがいくつかの共通した素子から組合せて作字されるという漢字構造表現法の二点をあげることができる．

漢字は表-2 に示されるように，一般に約 2,000～10,000 字種が用いられているが，各分野に共通して用いられる機能度の高い漢字は約 2,000～3,000 字種で，国立国語研究所の新聞・雑誌に用いられる漢字調査による文字種と使用率の関係は，**表-3**（次頁参照）に示されるように，2,500～3,000 字種で 99.9% の使用率を示している[1]．したがって，使用率の高い文字と，低い文字との取扱いを変えて，全体の漢字を処理する方式は，多種文字の問題を解決する有効な手段で

表-1　ローマ字・漢字システムの比較

	ソフトウェア					オペレーション	ハードウェア		
	A文字種	B絵素数	A×B メモリ量	語表現	コード表現	入力速度	手動タイプ価格	ラインプリン価格	端末システム価格
ローマ字	26	5×7	910	2.8 字	1 バイト	英文タイプA級 250字/分	30,000円	レンタル 250,000円	2,000,000円
漢　字	2,600	18×18	842,400	1 字	2 バイト	和文タイプ1級 50字/分	150,000	レンタル 1,000,000	20,000,000
漢字 / ローマ字	100	10	約 1,000	$\frac{1}{2.8}$	2	$\frac{5}{1}$	5	4	10

* KANJI Input and Output System by Method of Unit Pattern Construction by Jitsuro HASEGAWA (Japan Electronic Mfg. Co., Ltd.)

** 日本電気漢字システム(株)システム開発部

表-2 各種漢字表と収録文字数

漢字表	発行年	編集	文字数
1. 辞(字)典等			
1.1 大漢和辞典	1960	大修館	49,964
1.2 大字典	1917	啓成社	14,942
1.3 明解漢和辞典	1927	三省堂	6,488
1.4 新漢和辞典	1963	大修館	8,028
1.5 国語辞典音訓総覧	1966	講談社	5,805
1.6 新字源	1968	角川書店	9,921
1.7 現代漢字辞典	1968	小学館	3,885
2. 業界標準漢字表等			
2.1 実用漢字等級表	1932	日下部	6,478
2.2 活字帳	1958	毎日新聞	5,991
2.3 標準活字目録		日本活字鋳造(株)	8,400
2.4 常用漢字目録	1968	全日本活字配列協議会	4,000
2.5 標準コード用漢字表(試案)	1971	情報処理学会	6,100
2.6 行政情報処理用基本漢字(案)	1974	行政管理庁	約3,000
3. 統計的調査報告			
3.1 日本基本漢字	1941	大西三省堂	3,000
3.2 本邦常用漢字の研究	1941	印刷局	3,950
3.3 新聞の漢字	1941	カナモジカイ	3,542
3.4 雑誌九十種の用語用字	1963	国語研究所	3,505
3.5 現代新聞の漢字調査	1971	国語研究所	2,879
3.6 姓の漢字	1969	野村	2,382
3.7 地名の漢字	1968	林	2,433
4. 主要実用システム漢字表			
4.1 日本科学技術情報センタ	1968		3,071
4.2 内閣官房内閣調査室	1973		3,960
4.3 国立国会図書館	1970		5,028
4.4 学習研究社	1970		4,992
4.5 明治生命保険	1971		7,847
4.6 国立国語研究所	1966		7,474
4.7 和文タイプライタ		日本タイプライタ	3,059

表-3 新聞と雑誌の使用率分布の比較

	新聞	雑誌		新聞	雑誌
上位の 10字	10.0%	8.8%	全体の 80%	499字	638字
50	27.5	25.5	85	615	777
100	39.9	37.1	90	781	992
200	56.4	52.0	95	1,068	1,358
500	80.0	74.5	96	1,156	1,479
1,000	94.1	90.0	97	1,269	1,617
1,500	98.4	96.0	98	1,421	1,832
2,000	99.7	98.6	99	1,661	2,157
2,500	99.9	99.5	100	2,879	3,328
3,000		99.9			

ある.

また，その場合の漢字の取扱いの表現法としては，従来からもいくつかの研究報告[2),3),4),5)]がおこなわれているが，漢字パターンがいくつかの素子（偏，旁など）から構成されているという性質を利用した，パターン合成による方式が有効である.

この解説は，以上の考え方を基本として，多種文字の漢字入出力処理の問題である処理効率の向上と，装置の低価格化が期待できるより現実的なパターン合成方式について述べている．この方式による入力処理では，使用率の高い基本の2,000字以外の外字入力に対しては，3〜4ストローク入力によるパターン合成入力方式を適用し，出力処理では，約770種の基本パターンから合成作字をする方式により，約1/8に漢字パターンメモリを圧縮し，3,000字/秒の高速出力の性能を有している[16)].

2. 入力処理

2.1 入力方式の現状と問題点

漢字情報処理システムのプロセスの中では，入力処理が最大のネックになっている．それは，漢字情報の内部処理や出力処理がハードウェアおよび，ソフトウェアの技術によって，多種文字の問題をカバーしているのに対し，入力処理ではオペレーション上の問題として解決しなければならないからである.

現在，発表され，実用化されている日本語情報の入力方式には，ローマ字のタイプライタ鍵盤のような絶対的な方式はないが，各種の特長をもった方式のものがあり[6)]，漢字を打鍵により入力するオペレーション上からみた方式分類を表-4に示す.

表-4に示されているように，入力速度と，入力の容易さとは背反関係にあり，アプリケーションのシステムに応じて入力方式が選択採用されている.

また，収容文字種は各方式共，標準は2,000〜3,000字種であり，何れも，収容文字種外の外字処理[7)]が必要となっている．このように入力方式の問題点としては，

（1） 入力操作性（訓練度と入力速度）

（2） 外字処理

表-4 漢字入力方式（打鍵方式）

種類		方式	訓練度	入力速度	備考
フルキイ方式	①多数シフト式(漢テレ)	4〜15 シフト	中	40〜80 字/分	漢字入力装置として実績が古く最も多い
	②和文タイプ式	QMR 方式 活字バー方式 ホロタプレット方式	中	30〜50	和文タイプのオペレータがそのまま使用できる
	③ペンタッチ式	磁界方式 容量方式 光走査方式	中	30〜70	方式としては最も新らしい
カナ・タイプ式	④2ストローク式	配列方式 ラインプット方式	大	60〜100	記憶訓練がむずかしい収容字数は少ない
	⑤対話式	音訓入力 部首入力	小	20〜30	一般素人向き

をあげることができる.

入力操作性は入力文字の位置を探す目視操作の訓練度の容易さによって決まり，探す操作を一部入力装置側が負担しているものは（表-4 の⑤）訓練の必要性は少ないが入力速度は遅い．高速入力・熟練者用であり探す操作が記憶による方式（表-4 の④）を除いて，一般に最も使用されているのはフルキイ方式であるが，これらの収容文字種，及び，配列が標準化されていないので，オペレータの共通の教育の点で問題が生じている.

漢字の標準化に関しては，現在 JIS の標準漢字符号化が進行中であるので，その成案ができれば，それを基本にした入力装置の文字盤の標準化も期待できる.

第2の問題である外字処理は，漢字を扱う場合にはさけられない問題点であり，一般に入力方式の如何に拘らず外字入力に対しては，前もって準備された漢字のコード表（多くの場合，10 進4桁のコード）を索引し，そのコードナンバーを4桁入力し，入力処理のソフトエェアによってその漢字コードに変換する方式が採られているが，そのコード表を索引するのに時間を要し問題となる．例えば，漢字のみ（かな，英数字等を除いて）の入力の場合，漢字 2,500 字収容の入力装置を用い，漢字の使用率を表-3 の値であると仮定し（2,500 字で平均 99.7% とする）平均 60 字/分の入力速度とすると，約5分に1回の割合で外字処理が必要となり，外字コード索引と入力操作に平均30秒を要すると，約10% の入力効率の低下となる.

漢字の索引法としては，主として

（1） 音訓順，画数順

（2） 部首順，画数順

の2つの方法が用いられているが[8]，入力外字処理の場合は，収容文字（2,000～3,000 字種）以外の使用率の低い文字が対象であり，オペレータが容易にその音訓を記憶していることを期待できない場合が多いので，部首順が一般に用いられると考えられる．しかるに，部首順の分類にも多くの問題を含んでいる．部首順分類の基本となっている 214 部首の康熙式では，和（口部），相（目部）当（田部）のように，字源的な分類のため，字形から判断できない場合がある．このため，最近，康熙式に修正を加え，より視覚的な判断がつき易い新部首（140 部）等が実用されているが，標準化，統一化がおこなわれていない．このように，漢字の索引には検索しやすい絶対的なものがないので，外字索引コード表には音訓順，部首順とも夫々重複をいとわず編集しておく必要がある.

2.2 パターン合成入力方式

ここに解説する新しいパターン合成入力方式の特長としては

（1） 使用率の高い基本漢字は入力効率の高いフルキイ方式とする.

（2） 外字に対しては直視的に判断できるパターン合成方式とする.

という点にある.

漢字を構造面からみて，いくつかの構成素子に分解し，これらの組合せ方を記述して表現するというパターン合成の考え方は従来から研究されており，主として当用漢字を含む 2,000 字程度の漢字に対し，素子の組合せ位置関係を示すオペレータを 10 種，素子としての部分パターンに 250 種類をとり記述する方式や[2),3)]，同じように，組合せ関係を示すフレームを 18 種，素子としての部分パターンを 590 種で記述する方式[4)]や構成素子をさらにストロークに分解して，ストロークの組合せにより記述する方式[5)]等があるが，何れも漢字記述の表現が長く，主として漢字を出力することを考慮した方式である.

漢字の特性として，その構成素子を少なくして表現する場合には，その組合せの記述が複雑となり，入力キイの数は少なくてすむが，入力タッチ数が増加し，入力の判断が困難となる．一方．その構成素子を多くして，表現する場合には，その組合せの記述は簡単になり，入力キイの数は多くなるが，入力タッチ数はへり，入力の判断は容易となる．一般に，2,000 字種程度までのフルキイ入力方式は入力効率もよいので，パターン合成入力の方式としては，素子数を多くして，入力タッチ数が少ない方が有効である．この考え方に基づき，本方式では次のようなパターン合成入力方式をとった.

（1） 入力法としては，**図-1** に示すように，合成パターンを第1に入力し，それに引続き，その構成素子を順に入力する.

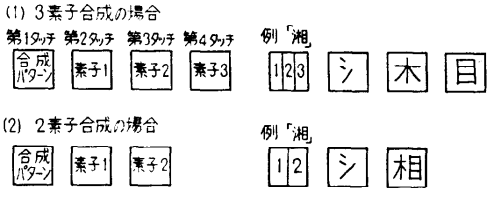

図-1 パターン合成入力方式

図-2 合成パターン（8種）

（2） 合成パターンは漢字の構成素子の組合せ方法を規定するもので，従来の構造表意的な部首という考え方でなく，より直視的な構造パターンとして，筆順にあうように，左上からの構造順序とし，図-2 に示すように，2．或いは3合成の8種の基本合成パターンとした．

（3） 構成素子の決定に対しては，10,000 字種を対象にして調査し，各字に対し，考えられるいくつかの合成法を考慮した．例えば，図-1 に示すように，「湘」の入力に対し，2合成，3合成の何れでも入力できるようにした．

（4） 入力文字盤はフルキイ方式の何れの方式でもよく，使用率の高い基本漢字 1,712 種，合成パターン8種，合成素子 336 種を配列した．

（5） 基本漢字の入力は直接その文字のキイ入力により，鍵盤外の外字入力は鍵盤上の基本漢字と合成素子をその構成素子として利用したパターン合成入力を標準とするが，基本漢字に対してもパターン合成入力が可能とした．例えば「松」は文字盤に配列されているが，⿰，木，公の入力でもよい．

（6） 合成素子 336 種は，部首（艹，辶，忄など），素子（夫，夬，昜など）及び，使用率の低い基本漢字以外の文字（而，升，蜀など）からなっている．

（7） 約 8,000 字種の漢字をパターン合成入力する場合に用いられる素子は，合成素子 336 種の他，基本漢字約 650 種で，合計で約 1,000 種が用いられている．

（8） 合成パターンは，上下，左右の素子組合せ位置関係の形式になっているが，実際には，あし，かまえ，たれ，にょう等の位置関係の素子があり，これを合成パターンとして設けるのも一つの方法であるが，本方式ではそれぞれを上下，左右の何れかの位置のものと約束した．

例えば「广」は左の位置とし，庄は⿰，广，土の入力法とした．

これらのパターン合成入力法によって入力されたデータは，電算機システムの入力処理の漢字辞書索引プログラムによって，その漢字コードに変換される．

ミニコンピュータを用いた場合，その最小構成は図-3 のように 8 kW の中央処理装置で，漢字辞書は磁気ドラムに収容されており，8,000 字の漢字に対し

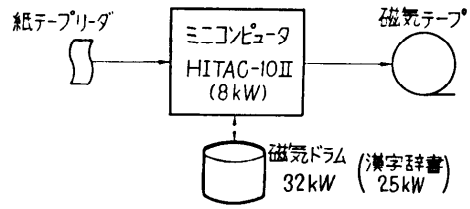

図-3 漢字辞書索引システムの構成

約 25 kW の記憶容量である．漢字辞書索引プログラムは約 1 kW のパッケージで，1パターン合成の索引は平均して，1.5回のドラムアクセスで約 15 ms を要するが，パターン合成入力の文字の頻度は平均して1％以下であり，紙テープリーダの入力速度は数 ms/字程度なので，入力と索引の並行処理により，入力速度は殆んど影響されずに処理されている．

2.3 考 察

本方式による入力は数か所での実用化の実績を有しており，次のような効果が得られている．

（1） 統計的な比較データは得られていないが，パターン合成入力はコード表索引というキイタッチ・オペレーションの中断はなく，直視的判断による連続したキイタッチ・オペレーションがおこなわれ，入力速度は平均して1件5秒程度で外字コード表索引の方式に比べ，外字の入力に対しては数倍以上の入力速度が期待できる．

（2） 漢字情報処理の応用分野ごとに異なる文字種の相異に対し，入力鍵盤の標準化がおこなわれ，オペレータの教育が共通しておこなわれる．

本方式では前述したように，使用率の高い漢字は直接その文字のキイ入力により，使用率の低い外字はパターン合成の入力によっておこなわれるが，外字の中でパターン合成入力方式の判断のつきにくい文字もありうるので，外字コード表も併用している．

パターン合成方式の問題点としては，漢字の字体の問題[9]が最も大きい．合成用の素子の種類を少なくするために，

（1） 「木→木」，「⻊→足」，「金→金」のように一部のみ異なり判断しやすい部首等は元の漢字を用いる．

（2） 当用漢字の字体表で略体化されたパターンと元のパターンは同一のものとした．例えば「艹→艹」，「曾→曽」，「靑→青」，「僉→僉」のように略体化の判断のつき易いもので，「會→会」，「廣→広」のような大きな略体化のものは別とした．

以上のような例外の約束事があり，必ずしも全ての外字が直視的なパターン合成でおこなえない場合もありうる．

3. 出力処理

3.1 出力方式の現状と問題点

入力処理の項でも述べたように，日本語を電子計算機で扱う場合の最大の問題点は，日本語を構成する多種文字からなる漢字の特質そのものであるが，その漢字処理システムのプロセスの中では出力処理の技術レベルは，各種記憶素子等のハードウェアの技術革新にともなって，これらの問題点を克服し，実用期の段階に到達しているということができる．

文字を出力するシステムの基本となる技術は，文字パターン発生であるが，これをとりまく文字出力システムのプロセスを図-4 に示す．

図-4 に示されるように，ハードウェア的には文字パターン発生部は出力印字部の方式と密接した関係があると共に，出力システムとして出力文字情報が用いられるアプリケーションから文字発生部に必要な性能が決定されてくる．そのアプリケーションの大きな分類としては，印刷植字に用いられる高品質の文字情報と，一般情報処理に用いられるプリント文字情報とにわけられ，漢字パターン発生システムとしては表-5 に示されるような各種の特長ある方式が用いられている[10]．漢字パターン発生方式は単に文字を発生する技術の評価だけでなく，漢字の文字種が極めて多く，外字処理がさけられないため，文字パターンメモリの製作は複雑で，相当な時間と経費が必要である点から文字パターン製作の性能を含めて評価されなければならない．

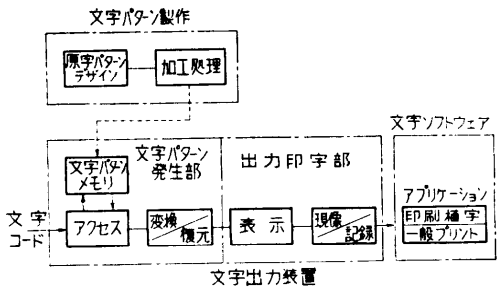

図-4　文字出力システムダイアグラム

表-5　漢字パターン発生方式 (A 写植用 / B 一般情報用)

- 漢字パターン発生方式
  - 字母式
    - 活字ドラム式 …… B
    - 移動文字板式
      - 字母露光式 …… A
      - 字母撮像式 …… B
    - 固定文字板式
      - フライング・スポット式 … A, B
      - ホログラム式 …… B
  - デイジタル式
    - ドット式 …… A, B
    - ストローク式 …… B

この漢字パターン発生システムに関連ある出力印字に現在実用されている方式を表-6 に示す[6]．

漢字パターン発生部の性能を決定する印字文字ソフトウェア[10]としては，文字の種類，文字の品質，文字の書体，及び，文字の大きさがあり，それぞれ漢字パターン発生部に関連するデータとして，表-7 に分野別平均使用文字種，表-8 に文字品質の要因，表-9(次頁参照) に印字品質に対する平均要望値，表-10(次頁参照) に分野別使用文字大きさを示す．

漢字パターン発生方式はこれらの点を考慮して設計

表-6　漢字出力方式

方　式		記録方式及媒体	関連文字発生方式
インクトンパ式	活字打鍵	インクリボン打鍵，普通紙	活字
	ニードル	インクリボン打鍵，普通紙	ドット
ノン・インパクト式	サーマル	感熱，感熱紙	ドット
	静　電	多針電極，静電記録	ドット
	CRT	乾式電子写真，普通紙	字母撮像
		銀塩安定化，安定化紙	フライングスポット
		湿式電子写真，電子写真紙	ホログラム
			ドット
	CRT	銀塩，フィルム印画紙	写植用
	字　母	銀塩，フィルム印画紙	写植用

表-7　分野別平均使用文字種

分　野	平均使用文字種
人名・地名を含んだ事務処理関係	1書体　6,000～10,000
印刷出版関係	1書体　3,000～ 5,000 総合3書体 5,000～16,000
新　聞　社	1書体　2,500～ 5,000 総合2書体 4,000～ 8,000
一般情報処理	1書体　2,000～ 5,000
放　送	1書体　2,500～ 3,000

表-8　出力文字品質の要因

解像度	文字パターン発生部の絵素数 出力印字部の解像力（文字のきれ）
濃　度	均一性 絶対値 コントラスト
幾何学的歪	字並び 直線性 大きさ歪
品　位	文字のデザインの美しさ デザインのフォント内のバランス 線画の太さの均一性

表-9 印字品質に対する要望値

分類	解像力	文字寄り引き	濃度
一般用 モニタ	4本/mm 以上	1.0 mm 以下	2.0～2.5
〃 軽印刷	10 〃	0.2 〃	2.2～2.5
植字用 商用印刷	20 〃	0.1 〃	2.5±0.1

表-10 分野別使用文字大きさ

使用分野	本文文字の大きさ	見出し文字の大きさ
新聞	約7ポイント	10～20～48ポイント
雑誌	8～9 ポイント	10～20ポイント
タイプ印刷	10.5～12ポイント	12～14ポイント

されなければならないが，ローマ字に比べ印字文字に要求される文字ソフトウェアの性能ははるかに複雑であると共に，表-1 に示されたように，パターン発生に必要なメモリ容量は約1,000倍であり，これらが漢字出力装置を高額にし，漢字処理システムの普及の大きな壁となっている．このことからも，漢字出力システムに与えられている最大の課題はシステムの低価格化という点にあり，次いで，印字出力の高速化と高品質化にある．この対策として，入力方式の項でものべたように，漢字の使用率の特性を利用して，漢字パターン発生のアクセス方式を2つに分け，高使用率の文字を高速アクセスのメモリに記憶し，低使用率の多種文字を低価格の低速アクセスのメモリに記憶して発生させる方式が実用されているが，基本的には，大容量，高速，低価格のメモリ素子の開発，及び，漢字パターンメモリの圧縮技術の開発による低価格化が期待されている．

漢字出力システムのもう一つの問題点は，入力システムと同じように外字処理があげられ，この出力システムに標準として収容されていない外字を印字する際に，外字パターンを文字発生部に収容する外字処理の容易性が大きな課題となっている．

これらの課題に答える将来性のある漢字パターン発生方式としては，現状ではホログラム方式と，ドット方式ということができる．ホログラム方式はディジタルな磁気記録よりもメモリの高密度化が期待できる点で，ドット方式はメモリとして，電子計算機メモリと同じものが使用でき，将来の開発進歩の点と，次節以下でのべるパターンメモリ圧縮技術の利用の点が期待できるからである．

3.2 パターン圧縮

漢字パターン発生システムに与えられている最大の課題である低価格化の対策としては，パターンメモリに使用するメモリ素子の技術開発，及び，漢字パターンメモリ圧縮技術の開発があげられており，前者では，高速不揮発性半導体メモリの他，CCD（電荷結合素子）や磁気バブルメモリ等が期待されている．

低価格への積極的な対策としての漢字パターンメモリ圧縮に関しても，従来から，いくつかの研究が報告されているが，効果のある実用的なものは少なかった．漢字パターンのメモリ圧縮技術には次の2つの方向がある．

（1） 漢字を文字としてではなく，一般のパターンとしてとらえて圧縮する方法，

（2） 漢字を漢字としてとらえ，漢字パターンのもつ特性を利用してメモリ圧縮する方法．

パターンメモリ圧縮の方式評価は単に圧縮率だけではなく，その復元，変換方法の簡易さと，出力文字に要求される文字ソフトウェアの機能をいかに備えているかという総合的な評価が必要である．

まず，（1）の場合の，漢字を一般のパターンとしてみたメモリ圧縮の技術に関しては，いくつかの調査，検討が報告されている[11),12),13)]．これらによると，18×22 のマトリックスでは Run-length code による方法では効果はなく，漢字イメージのサブパターンを利用した方法では 70% 程度で効果は少ないといわれており，図-5 に示すように，漢字イメージを 4 ドット

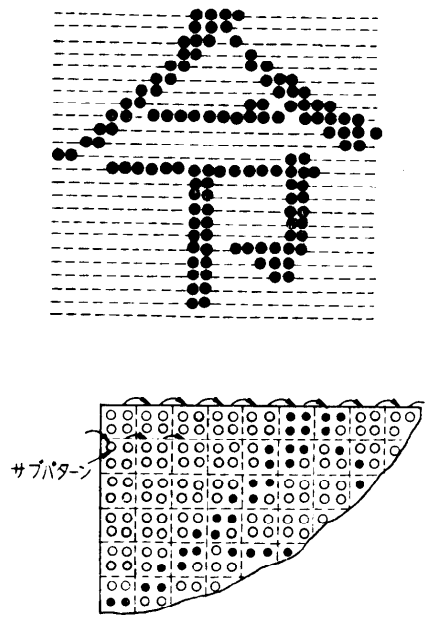

図-5 漢字パターンの圧縮符号化

程度の区割で区切り，そのパターンのかたまりと，前後の連続性を16個のサブパターンを用意して，Huffman法によるコード化をおこなった結果，約50%の圧縮率がえられたと報告されている．

なお，植字用の文字パターンの場合には，Run-length code方式によって90×90程度のマトリックスで，33～50%程度の圧縮率がえられているが[14]，漢字パターンの場合は，1字内のマトリックスが小さく，特に一般処理の場合には，1字のマトリックスが18×18，24×24，32×32ドット程度で，冗長性も小さく，ファクシミリにおけるパターン圧縮と同じような方式では，同じような大きな効果は期待できない．

次に，(2)の場合の，漢字を一般のパターンとしてではなく，漢字としてとらえ，漢字パターンのもつ特性を利用したメモリ圧縮方式には興味あるいくつかの方式が発表されている．

その1つは，内容が詳しく発表されていないので推定になるが[15]，漢字パターンの横線が縦線よりも多く，独立した1ドットの点はありえない等の論理によって，文字パターンメモリとしては16×32ドットで記憶し，出力する場合に，補間法による論理で32×32ドットに復元し，50%の圧縮率をえている．

他の方法は，**2.2** パターン合成入力方式の項で紹介したように，漢字を構造面からみて，いくつかの構成素子に分解し，これらの組合せによって表現するというパターン合成の考え方で，メモリには構成素子のパターンと，各文字の各素子の組合せ方式の記述等を記憶させておけばよいので，パターンメモリ圧縮には最も効果のある方式であると期待できる．

前述のように，従来発表されているものは，主として当用漢字を含む2,000字程度の漢字に対し，素子の組合せの位置関係のパターンを示すオペレータ[2]，あるいはフレーム[4]と部分パターンとから記述する方式でプログラムにより合成するが，パターンメモリ圧縮の効果は25%[2]，10%[4]，と大きいが，出力印字文字の品質が自然性をかき，かつ，出力速度が遅く，実用化はされていない．

3.3 パターン合成出力方式

漢字をいくつかの基本となる構成素子パターン（以下基本パターンという）を組み合せて合成作字する場合の実用化の問題点としては，前述したように，

（1） 自然性をもち品位のある合成文字であること．

（2） 出力速度が速いこと．

である．

品質のよい文字を合成作字するためには，一定の大きさの基本パターンを組み合せるだけでは不充分である．例えば，同じ「木」でも，つくりの簡単な「村」の場合と，複雑な「欄」の場合とでは「木」の横幅を変える必要がある．また，基本パターンをストローク情報で表現すると，それを出力する場合，安価のものではXYレコーダがあるが，その出力速度は遅い．また，高速化のためCRTを用いた場合，XY軸の高速偏向が必要となり，安定性，及び価格の点で実用性に問題を生ずる．

ここに解説するパターン合成方式は，これらの点を改良し，先ず第1に，約10,000字の文字を対象に実用性のある品位の高い文字を，高速に合成作字することを目標とした．

このため，本方式では，従来の合成方式で用いられていた基本パターンの定形的な組み合せ（若干，組み合せのウエイトを制御しているが）を規定するオペレータ[2]やフレーム[4]のようなものを用いず，実用されている標準的な文字の形にあうように，各基本パターンの位置，大きさを制御して組み合せるような方式をとった．各基本パターンの表現には，位置，大きさを容易に制御できるストローク情報を用い，出力の高速化，簡易化のために，ストローク情報によって合成作字する場合に，ファームウェア化された論理回路によって，ドットパターンに変換し，普遍性のある安価なドット式の出力印字方式を用いるようにした．

文字の絵素表現は解像度及び表現性を高めるために**図-6**に示すように，標準となる記憶パターンの座標の大きさを32×32ドットとし，基本パターンは原字のストローク近似によりなるべく細かく忠実に表現するようにした．

基本パターン以外の文字は，これらの基本パターンからの合成により**図-7**（次頁参照）に示すように，合成作字データは各基本パターンの原点位置と，X, Y方向それぞれの縮小率とで表現することにし，基本パターンの選定に当っては次の点に留意しておこなった．

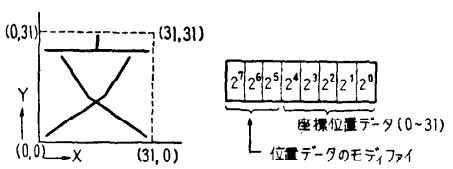

図-6 ストローク文字デザイン例とデータ表現

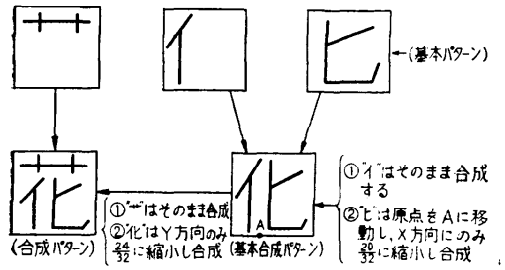

図-7　基本パターンによる合成作字

（1） 基本パターンの選定は約 10,000 字種を対象に調査して求めた．

（2） 合成作字に当っては，あまり合成数を多くすると，作字した場合の不自然さが増えると共に，作字の高速化が妨げられるので，1文字の合成は基本パターンの3合成以内になるように基本パターンを設けた．

（3） 使用頻度の高い合成パターン例えば者（耂＋日）や化（亻＋匕）等は図-7 に示すように，基本合成パターンとし1文字の合成に1回だけ基本パターンの代りに用いられるようにした．

（4） 前述のパターン合成入力方式では「村」，「床」の「木」と「木」は同一素子としたが，出力の場合は，形が異なるものは印字品質の向上のため，別パターンとした．同じように，「勝」と「前」の月等は「はね」が異なるので別パターンとした．

（5） 文字の表現上，同形の基本パターンではあるがいくつかの文字を合成して表現する場合に，字形と不自然さが生ずる場合は別に基本パターンを設けた．

例えば，「問」と「闇」に使われる「門」や，「則」と「資」に使われる「貝」の場合で，前者は合成する他のパターンの大きさが異なり，一つの基本パターンを変形して用いられない場合であり，後者は用いられる位置が「へん」や「かまえ」等で一つの基本パターンからの変形方向が異なり合成した場合不自然さが増す場合である．

基本パターンのストロークデータと，合成文字の作字データとから文字を合成するハードウェアのブロックダイヤグラムを図-8 に示す．これらの合成作字，ストローク/ドット変換はすべてファームウェア化された高速の論理回路で構成されており，約3,000 字/秒の出力速度である．

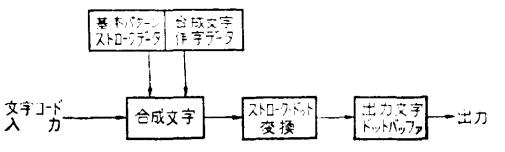

図-8　方式ブロックダイアグラム

本方式に用いられている基本パターンの数は表-11 に示すように，約 7,000 字の漢字を対象にした場合約 770 種で，英数字，ひらがな，かたかな，記号・約物を対象に含めるとすると，基本パターンの数は約 1,000 種程度になる．

基本パターン及び，合成基本パターンの中，使用頻度の高いパターンをそれぞれ表-12，表-13 に示す．

参考までに約 10,000 字種を収容している市販の漢字辞書で，使用頻度の高い部首を表-14（次頁参照）に示す．

本方式を採用した漢字ラインプリンタ[16]の場合，英数字，平がな，片かな，記号・約物，及び，漢字を含めた標準 5,000 文字種に対し，基本パターンは，前述のように約 1,000 種で，パターンメモリ容量は約 23k

表-11　基本パターンの種類

	基本パターン	合成基本パターン	合　計
文　字	280	700	980
非文字	490	190	680
合　計	770	890	1,660

表-12　主要基本パターン

順位	パターン	使用数	順位	パターン	使用数	順位	パターン	使用数
1	氵(汁)	354	11	竹(笑)	121	21	土(地)	82
2	⺾(花)	309	12	辶(込)	115	22	疒(病)	81
3	木(札)	265	13	月(明)※	110	23	土(去)※	80
4	扌(打)	222	14	心(志)	104	24	王(珍)	75
5	亻(化)	215	15	日(早)	103	25	宀(字)	74
6	言(信)※	182	16	口(占)	101	26	木(休)※	73
7	口(叶)	172	17	虫(蚊)	90	27	田(男)※	73
8	金(針)	164	18	一(丙)※	90	28	攵(攻)	72
9	糸(紀)	160	19	女(如)	86	29	鳥(鳴)※	71
10	忄(忙)	138	20	石(砂)※	86	30	魚(鮭)	67

（※印は文字で無印は非文字）

表-13　主要合成基本パターン

順位	パターン	使用数	順位	パターン	使用数	順位	パターン	使用数
1	者※	26	11	令※	17	21	軍※	14
2	合※	24	12	台※	17	22	昷	14
3	易	20	13	奇※	17	23	交※	13
4	分※	19	14	辟※	17	24	周※	13
5	肖※	19	15	公※	16	25	禹※	13
6	兪※	18	16	共※	16	26	番※	13
7	丰※	18	17	占※	15	27	慮※	13
8	各※	18	18	京※	14	28	章※	13
9	疒	18	19	加※	14	29	龍※	13
10	昜	18	20	区※	14	30	参※	13

（※印は文字で無印は非文字）

表-14 辞書に用いられている主要部首パターン

順位	パターン	使用数	順位	パターン	使用数	順位	パターン	使用数
1	氵	479	11	虫	206	21	目	119
2	木	461	12	土	193	22	鳥	118
3	++	397	13	火	173	23	魚	117
4	忄, 忄	353	14	女	167	24	犭	111
5	口	346	15	月	162	25	王	105
6	亻	345	16	辶	159	26	刀, 刂	104
7	扌	333	17	竹	157	27	馬	104
8	糸	250	18	日	145	28	山	102
9	言	245	19	礻	137	29	石	102
10	金	215	20	足	123	30	疒	102

(新字源より)

バイトを要し，5,000 文字種を合成作字する場合の基本パターン及び，合成基本パターンの合成数は一字当り平均して 2.5 回で，その作字データに要するメモリ容量は約 57 k バイトであった．

これにより，本方式では 5,000 文字種に対し，約 80 k バイトのメモリ容量でよく，平均して約 16 バイトで表現され，32×32 ドットの単純なドット方式に比べ，約 1/8 にメモリ圧縮することができた．

3.4 漢字デザイン法

本方式による漢字のデザインに関しては**図-9** に示すように，8 kW のミニコンピュータを用い，ストレージ方式の簡易グラフィックディスプレイを数台並列に接続し，会話形式で CRT 上に合成パターンの原点位置と，変形大きさを任意に指示して作字させ，適正なデザインが得られるようにした．

3.5 考　察

本方式と他のドット方式とのパターンメモリ容量の比較を**表-15** に示す．

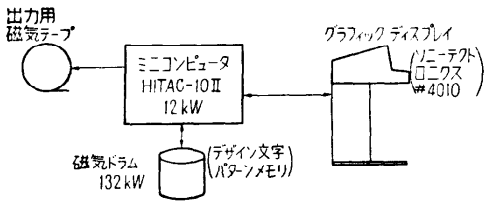

図-9 漢字デザインシステム構成

表-15 本方式とドット方式とのパターンメモリ容量比較

ドット構成	1字に必要なバイト数	5,000 字の場合の記憶容量	本方式を1とした場合の比較
32×32	128 バイト	640 k バイト	8
24×24	72	360	4.5
22×20	55	275	3.4
18×16	36	180	2.2
32×32(本方式)	16	80	1

本方式の主な特長を下記に示す．

(1) 表-15 に示すように，32×32 ドットパターンで比較してみると，約 1/8 にパターンメモリ圧縮ができ，5,000 字種で約 80 k バイトのメモリ容量ですみ，高速のメモリ素子を用いることができ，かつ，低価格が十分期待できる．

(2) メモリ容量が 5,000 字種で 80 k バイトと小容量ですみ，物理的に約 300×450 mm のメモリ基板 1 枚の大きさなので，装置全体が小型化できる．

(3) 文字の記憶情報がストローク情報なので，出力文字の大小制御及び長体，平体が容易におこなえる．本方式を採用した漢字ラインプリンタでは 1 フォントの漢字パターンメモリから 8 ポイントより 16 ポイントまでの文字出力が可能である（普通のドット方式では不可能である）．16 ポイントの場合，文字のドット構成は 48×48 ドットで，単純なドット方式に比べ約 1/18 にメモリ圧縮したことになる．

(4) 文字種を 5,000 字から 2 倍の 10,000 字に増加した場合，メモリ容量は 80 k バイトから約 140 k バイトに約 1.75 倍の増加ですみ，これら低使用率の文字の増加に対し，全体の出力装置価格に比べ，価格増が少なくてすむ．

(5) 合成作字方式なので，いわゆる外字に対しても，標準的に収容している基本パターンを用いて，合成作字することが可能である場合が多く，簡易な外字処理方式が期待できる．

次に，本方式には大きな問題点はないが，

(1) 合成作字とストローク/ドット変換のハードウェア論理に 24×24 ドットの場合，平均して約 150 μs/字の復元時間を要する．

(2) パターンメモリの節約のために，基本パターンを少なくしすぎると，文字のバランスや品位が悪くなるので注意を要する．

(3) 文字の大小出力をする場合，線間のバランスがくずれる場合がある．

などに留意する必要があるが，基本パターンの増加は全体のメモリ増加に対する影響はそれ程大きくないので，基本パターンを増加し，出力文字の品位の向上に処した方がよい．

4. ま と め

漢字を処理する場合，漢字を一つのパターンと解釈せず，漢字のもっている特性を利用して，構造的にパターン合成によって表現する方式は，以上で述べたよ

うに漢字処理の問題点を解決する有効な手段の一つであると考えられる．但し，漢字を入出力する場合，入力処理ではオペレータがその合成法を判断し，出力処理ではハードウェアがその合成を実行するので，本解説の方式のように，パターン合成の方法は，それぞれ，入力・出力に適した別の形にした方がよいといえる．

漢字をパターン合成する場合の基本的な問題点としては，漢字の字体が統一されていない点である．当用漢字に対しては略体化がおこなわれ，政令によって当用漢字字体表が制定されているが，当用漢字外の字体に関しては，当用漢字でとられた略体化の方式をとる場合と，とらない場合とがある他，新字，旧字，俗字別体字等，字体の使用の標準化がないので，それぞれの基本構成パターンを用意する必要が生じ，パターン合成を複雑にしている．

なお，本方式は我が国の漢字のみならず，中国で用いられている簡体字をもとにした漢字や，韓国のハングル文字等にも有効な方式として応用することが期待できる．

参考文献

1) 野村雅昭：現代の漢字，国立国語研究所の歩み・8，pp. 81～96，国立国語研究所 (1974)
2) 坂井利之，長尾真，寺井秀一：部分パターンによる漢字の合成，情報処理，Vol. 10, No. 5, pp. 285～293 (1969)
3) T. Sakai, S. Sugita, H. Fujita: Some experiments on KANJI I/O system, Preprint Seminer I/O System Japan Chinese Characters, Tokyo, pp. 232～252 (1971)
4) H. Ishida, S. Furukawa: Synthesis and Display of Kanji by unit Construction, Preprint Seminer I/O System Japan Chinese Characters, Tokyo, pp. 276～282 (1971)
5) O. Fujimura, R. Kagaya: Structural patterns of characters, Preprint Seminer I/O System Japan Chinese Characters, Tokyo, pp. 172～189 (1971)
6) 日本語情報処理技術動向調査委員会：日本語情報処理の技術動向調査報告書，日本情報処理開発センター (1973)
7) 大倉信治：漢字入出力装置による標準外字の取扱い，情報処理学会マン・マシン・システム研資料 74-15, pp. 1～7 (1974)
8) 林 大：漢字コード，電子計算機の国際標準化 ISO の動きとわが国の歩み，第4章，pp. 83～97，情報処理学会 (1971)
9) 林 大：当用漢字字体表の問題点，文化庁国語シリーズ漢字，VI, pp. 253～364，教育出版 (1973)
10) 長谷川実郎：漢字パターン発生システム，画像電子学会誌，Vol. 4, No. 1, pp. 31～43 (1975)
11) 禰津孔二：漢字表示とディスプレイ，テレビジョン，Vol. 27, No. 5, pp. 401～404 (1973)
12) K. Ohmori, K. Nezu, S. Naito, T. Nanya: An application of Cellular logic for high-speed decoding of minimum redundancy codes, AFIPS Conf. Proc. Vol. 41, ptl, pp. 345～355 (1972)
13) 小田，能勢：漢字イメージ・データの圧縮，情報処理学会第14回大会予稿，No. 158 (1973)
14) 日本新聞協会工務委員会：電算植字部門，新聞印刷 CTS 編，日本新聞協会，pp. 61～232 (1974)
15) 金田，伊藤：T 4100 漢字情報処理システム，エレクトロニクス，Vol. 17, No. 1, pp. 79～85 (1972)
16) 長谷川実郎：新しい漢字出力処理方式，画像電子学会第2回全国大会予稿 No. 18 (1974)

(昭和50年5月29日受付)