

# 漢字情報処理システムの課題

——漢字セットの設計と漢字辞書運用システムについて——

田 嶋 一 夫

## 目 次

1. 序にかえて
  - 1.1 漢字情報処理システム発展の経過と課題
  - 1.2 漢字システムと漢字情報処理システム
2. 漢字セット設計上の問題点
  - 2.1 漢字の数
  - 2.2 分野間の漢字の層
  - 2.3 漢字における字体
  - 2.4 標準文字セットの基本
3. 漢字セット運用上の問題点
  - 3.1 漢字情報処理システムにおける内字と外字
    - 3.1.1 内字と外字の概念
    - 3.1.2 外字処理の機能
  - 3.2 漢字情報処理システムの為の漢字辞書
    - 3.2.1 漢字属性情報
    - 3.2.2 漢字シソーラス
4. むすびにかえて
  1. 序にかえて
    - 1.1. 漢字情報処理システム発展の経過と課題

漢字情報処理システムの発展の経過をたどってみると、1950年代の後半に漢字電信コード（1959、C O—59符号系）が制定され、共同通信社による漢字テレタイプ全国通信網が稼動しはじめたことが、実用システムとしての嚆矢とみ

られるであろう。一方、高品質文字の出力機構としての漢字プリンタの実用は、1960年代の中頃に入ってからのこととみられる。日本電子産業株式会社によるJEMシリーズの発表は、漢字テレタイプによる入力と、漢字プリンタからの高品質文字の出力を可能とし、出版物の編集、印刷業務への利用の道を大きく拓いたと言える。日本科学技術情報センターにおける「科学技術文献速報」の編集・印刷版下フィルム<sup>1)</sup>の作成が、はじめたのは、1968年のことであり、学習研究社における「グランド百科辞典」の編集・版下の作成がはじめたのは、1969年であった<sup>1)</sup>。いわば、1960年代の後半は、一部の特殊なユーザーのもとで、日本語の文章情報処理が行われ出すとともに、印刷業の一部において電算写植システムの利用がはじまった時と言えよう。

一方、1970年代の後半にはいると、それまで利用してきたカタカナのシステムを、漢字のシステムに変えようとする動きが顕著になってきた。一般の計算センターの中にも、漢字処理システムを導入し、漢字処理の受託を受けるようになるとともに、自治体等における窓口業務の漢字処理システム化に、大きな関心がよせられてきている。また一般のデータ処理分野の中にも、見やすさを重視してこれまでのカナ表記を漢字表記に変えたいとする要求が顕著になってきた。いわば特殊なユーザーにおいて特殊な使い方をされてきた漢字処理システムが、汎用化の方向に確実に歩み出していると言えるであろう。

しかしこの歩みは一般のコンピュータの発展からすれば、はるかに遅いものである。これはコンピュータが漢字を意識せずに、技術的に発展してきたことと、漢字の文字種が、アルファベット圏の言語に比べれば、極めて多いと言う漢字のもつ文字システムとしての特殊性から、その装置が特殊なものとして位置づけられ、他の周辺装置のような汎用性を持たなかったこととの二つが、発展を阻害していた主因であったろうと思われる。

私は漢字情報処理システムのめざすものは日本語文の自在な情報処理、もっと俗な表現をとれば、アルファベット言語圏におけるそれとの格差の解消にあると考える。またこのためには、アルファベット言語を中心としたコンピュータシステムの模倣ではなく、漢字システムを根源的に分析し、そこから体系的で、かつ現実に即応したシステムを構想し、その上で一つ一つ着実に実現していくべきであろう。その上で現実的にハードウェアの低価格化と、利用システムの標準化、汎用化をめざすべきであると思う。

本稿では主として、漢字システムの現状を分析し、漢字情報処理システムにおける文字セットのあり方、文字セットを設計する上での基本的問題、いわばハードウェア、ソフトウェア以前の基本的問題について考察する。

## 1.2. 漢字システムと漢字情報処理システム

まず基本的な問題から整理しておきたいと思う。今日、我々は文字として主に漢字やかなを用いている。そしてこの文字を利用したコミュニケーションの手段としては、手書の書類からはじまって、騰写印刷、和文タイプによる印刷、活版による印刷、写植システムによる印刷、或はコピー等の手段を用いている。この段階では、データの保存は、ノート或は印刷物或はそれらの二次的加加工物等によって行われる。この管理は基本的には手動方式である。仮にこうしたシステムを漢字システムと名づけておく。

一方、近年急激に増加してきたものは、漢字やかなを主要な表現媒体として情報の伝達を行うが、この情報を機械可読の媒体（磁気テープや磁気ディスク等コンピュータの周辺機器）に保存し、管理し、必要に応じ機械的に加工し、印刷する方法である。前者と異なる点は、機械的なデータ管理を伴うことであり、データの保存、伝達手段としての出力等が、一環したシステムとなっていることである。これを前者と識別する意味で、仮に漢字情報処理システムと名づける。言うまでもなくここではコンピュータが中心的役割を果たす。また両者を分かつ根本は、データが機械可読の形で管理され、出力されるかの一点にある。印字される字母の用意という点で、この二つのシステムを比較してみれば、前者にあっては、管理する字母の数はそう大量のものを要しない。なぜなら一般的に使用される漢字の数は、2000字程度で99%を超えるのである<sup>2)</sup>。ここには漢字（外字）は、必要に応じ用意し使用する方が効率的と言えるであろう。殊に近代活字以前の近世版本の印刷形態を考えれば明白である。ある版下にあられる文字はすべて外字処理のようなものであり、そのつど字母を作成していると見ることが出来る。版下の彫師にとっては、著者Aの使用する文字種と著者Bの使用する文字種とが、たとえ大きく異っていたとしても、ほとんど問題にならなかったと見做すことが出来るであろう。しかし、漢字情報処理システムにあっては、原則的にみれば、あるデータに出現する漢字は常に用意されていなければならない。いかに稀にしかあらわれない漢字でも、出現頻度の高い漢字と同様に、同じだけのメモリーを使って同じコード体系の中におさめておかなければならない。

このように漢字情報処理システムにおける漢字の問題をつきつめて考えてみれば、効率的な漢字セットの設計が重要な意味をもつことが明らかであると言えよう。

私が本稿において、文字セットのあり方、設計上の問題について考察しよう

とする根本は、これまで多くの漢字情報処理システムがそうであったように、制限され固定化された漢字セットでしか使えないのでは、真の意味での日本人のシステムにはなりえない。必要な漢字が必要な時に使えるシステムになっていなければならない。しかし漢字システムの実態を見きわめることなしに、漢字システムをより大きな混乱におとし入れるような方向で、漢字情報処理システムを設計し、運用すべきではない。この意味で漢字情報処理システムの設計にあたっては、漢字システムの実態をふまえ、その上で文字セットの問題をもっと考えてほしいと思うからである。

## 2. 文字セット設計上の問題点

漢字情報処理システムについて、そのシステム構成を考え、具体的に導入し実用化していくためには、文字セットをどのようなものにするかを決めなければならない。文字セットのうち、平かな、片かな、数字、ローマ文字、ギリシャ文字、ロシア文字等の有限個数のものについては、そのすべてを用意しておいたとしても、さしたる量にはならない。しかし漢字を用意するとなると簡単ではない。後にも述べるように漢字がいくつ存在するかも定かでないことからしても、そのすべてを用意することは論理的に不可能である。また当用漢字を用意ただけで十分であるか、どうかについては、誰もが無意識のうちに、当用漢字外の漢字も使用しているのが、実情であるから、ごく特殊な場合を除けば、不十分である。したがって漢字セットを作成する何らかの方法が必要となるのである。このために考えるべき点は、次の3点にある。その第1は、存在する漢字の数の問題であり、次で分野間の漢字の層の問題であり、第3には字体の問題である。

### 2-1 漢字の数

漢字が総数いくつあるかを数えることはほとんど不可能であろう。しかしおおよそどの位の数が存在するかを把握しておくことは、必要なことである。

世界で最大の漢字収録数をほこる『大漢和辞典』（諸橋徹次編、昭和30年～35年、大修館書店刊）には、個々の漢字に番号がふってある。最終番号は48,902の番号となっている。しばしばこの番号が『大漢和辞典』の収録漢字数と錯覚される場合が多いのであるが、同辞典の索引の部を見ると、「補遺」があり、ここに1,062字が収録されている。つまり、49,964字である。さらにこの他に同番で(ダッシュ)を付して区別したものがある。ダッシュがいかなるものか説明はない。おそらく編集上の不手際を補うためのものと、編集途上に新らし

い字体が認知された場合のものとの二種類と思われる。総画索引にあげられているものを数えると、50,292字であると言う<sup>3)</sup>。通し番号との差328は'（ダッシュ）の数かと思われるが、ダッシュの数は、当館漢字情報処理システムで利用している文字セット（これはJ I S<sup>4)</sup>の第1及び第2水準の文字に、J I S外で独自に使用する漢字を追加したもので、現在約900字にコードを与えている）を対象とした漢字辞書の中からもだけでも、322字を数える。おそらくはこれだけで数百はあるものと思われる。従って『大漢和』で収録漢字数が5万を超えることは確実なのであるが、正確に何字収録されているかは明確ではない。またこの5万余の漢字の中には、後に述べるような異体字がきわめて多く含まれているのである。

さらに今日日常的に使用している漢字の中で、『大漢和』に収録されていない漢字も多い。J I S漢字表（J I S C-6226<sup>-1978</sup>）の中には、私の調査で113にのぼる漢字が『大漢和』及び『新字源』に見い出せない<sup>5)</sup>。このうち151字は、『大漢和』にない。また古典の写本、版本などを見ていると、辞書に見い出せない漢字が多いことを実感させられる。また江戸時代には、『異字篇』『同文通考』『異体字辯』等異体字を整理、集成した辞書があらわれているが、その中の『異体字辯』（江戸中期の数学者中野元圭の編著）では、日用の字を数多く収集し分類しているが、その数は5,002に達している<sup>6)</sup>。この中には『大漢和』に収録されていないものも多く含まれている。このような事例をあわせ考えてみると、現存する（歴史的文献の存在も含めて）漢字の数は膨大な数にのぼるのであって、整理し集計することはほぼ不可能と考えなければならないであろう。

ところでこのような漢字の数が、歴史的にどう増加してきているかを、主要な辞典の収録文字数で見ると表-1のようになる。

表-1、辞典に於ける漢字収録字数の変化<sup>7)</sup>

説文解字	9,353
玉篇（六朝時代）	16,917
広韻（宋）	26,194
字彙（明）	33,179
康熙字典（清）	42,174

この表を見れば漢字の数が時代とともに増加しているさまが如実にあらわれているのである。漢字字種の増加の因は、新しい概念を表現するための漢字があらたに作られるという本来の漢字の文字機能に基づくものと、俗字や別体字が作られたことによる増加の2つであろう。これらの漢字辞書が日本でどう利用

されていたかについては、『字彙』が、江戸時代の漢字の辞書と言えば、典型的なものとして使われてい<sup>8)</sup> たという事実からしても、相当利用されていたことはまちがいない。

## 2-2. 分野間の漢字の層

次にこれらの漢字が実際にどの位使われているのかを考えてみよう。漢字の字種については、これまでさまざまな目的で、いくつかの調査が行われている。また私共のところでも、字種選定の目的や、文字セット使用効率調査等の目的のもとに調査を行っている。これらを整理してみると表-2のようになる。この調査の中では、データの扱い方、異体字の取り扱い方針・総文字数の提示等、統一はとれていない。しかしできるだけさまざまな分野にわたるデータを集めるよう心がけた。この表の中でAの日本基本漢字は、調査結果そのままのデータではないが、すでに基本的データとなっているので参考までにかかげたものである。Aを除いて20の調査結果がでているわけであるが、異り漢字数順に分布を見ると

4,500以上	2
4,000~3,501	5
3,500~3,001	4
3,000~2,501	6
2,500~2,001	1
2,000以下	2

となり、2,500から4,000までの間に大半が入っている。3,500を超す漢字のあらわれたものは、D<sub>1</sub>の郵便報知新聞の場合を除いて、母集団が大きいものが多い。E<sub>2</sub>、E<sub>3</sub>は総漢字数が記されていないが、いずれも膨大なページにわたって、「夫々の頁から漢字を全て抽出し、……計算機に入力して頻度の調査をした」と説明されているから、膨大な母集団であったと思われる。殊に上位3つの、B<sub>1</sub>、B<sub>2</sub>、B<sub>3</sub>が母集団の上でも上位を占めている。またF<sub>3</sub>のデータは、具体的には、昭和47年、48年、49年の国文学の論文タイトルを入力したものであるが、これを各年ごとに集計すると、2,271、2,355、2,338となっている。母集団の増加により異り漢字数も確実に増加しているのである。

また異り漢字数の多いもののデータの中味を見てみると、B<sub>3</sub>が、単行本、雑誌、辞書、百科辞書等とあるように、各種の分野にわたるデータである。B<sub>2</sub>も印刷に日常使われている活字の調査と思われるので、広い分野のデータであると思われる。

表-2 漢字の使用字数調査

	資 料 名	総漢字 数 (千字)	異り漢 字数	調査 種別	対 象	発表文 は調査 時
A	日本基本漢字(大西)	804	※3,000	△	国語読本, 文学書, 新聞等 分野	1941
B <sub>1</sub>	印 刷 局	3,280	3,948	△	第55,65,67帝國議会両院本 会議速記録	1941 1962
B <sub>2</sub>	印 刷 局	45,910	5,122	△	常用文字調査	~66
B <sub>3</sub>	凸 版 印 刷	5,300	4,520	△	単行本・雑誌・辞書・百科 辞書等	1976
C <sub>1</sub>	婦 人 雑 誌	170	3,048	㊦	主婦の友, 1950年1~12月	1953
C <sub>2</sub>	総 合 雑 誌	120	2,781	㊦	13誌, 1953.7~1954.6	1960
C <sub>3</sub>	雑 誌 九 十 種	280	3,328	㊦	5 部門90種, 1956.1~12	1963
D <sub>1</sub>	郵 便 報 知 新 聞	166	3,680	㊦	1877.11~1878.10	1967
D <sub>2</sub>	現 代 新 聞	990	3,213	㊦	3 紙朝夕刊, 1966.1~12	1976
D <sub>3</sub>	カナモジカイ(岡崎)	450	3,542	㊦	大阪毎日, 東京朝日等5 紙 60日分, 1935.1~12	1938
E <sub>1</sub>	姓名(日本ユニバ ック)	2,721	2,637	全	調査人数663,823人	1974
E <sub>2</sub>	官 報		3,616	全	官報1972, 計12,092頁	1974
E <sub>3</sub>	国土行政区画総覧		3,782	全	総覧4,006頁	1974
F <sub>1</sub>	古 典 書 誌 (1)	243	2,666	全	古典書誌目録約9,000件	1976
F <sub>2</sub>	" (2)	212	2,791	全	" 8,500件	1978
F <sub>3</sub>	文 献 目 録	340	3,234	全	国文学の論文約14,000タイ トル	1978
F <sub>4</sub>	論 文 抄 録	196	2,680	全	国文学論文の抄録830件	1976
G <sub>1</sub>	古典テキスト (1)	44	2,453	全	日本靈異記全	1975
G <sub>2</sub>	" (2)	23	1,584	全	保元物語全	1978
G <sub>3</sub>	" (3)	22	1,510	全	平治物語全	1978
G <sub>4</sub>	" (4)		2,702	全	万葉集全	1979

(調査種別は, △は不明, ㊦はサンプル調査, 全は全件の調査を示す)

## 資料注

- A 大西雅雄『日本基本漢字』1941三省堂  
 B<sub>1</sub>~B<sub>3</sub>, C<sub>1</sub>~C<sub>3</sub>, D<sub>2</sub> 林大「各種資料における漢字使用率の比較」(文献2)より部分転載  
 D<sub>1</sub> 進藤咲子「明治初期の新聞の用字」(『国立国語研究所論集』3, 1966)による。  
 D<sub>3</sub> 岡崎常太郎『漢字制限の基本的研究』1938, 松島三松堂  
 E<sub>1</sub> 日本ユニバック『漢字システム・デザイン資料(4)』  
 E<sub>2</sub>, E<sub>3</sub> 行政管理庁行政管理局・谷村株式会社新興製作所『行政情報処理用標準漢字の  
 選定に関する調査研究・報告書』1974  
 F<sub>1</sub>~G<sub>4</sub> 国文学研究資料館情報処理室及び筆者の調査による。F<sub>1</sub>, F<sub>4</sub>については、  
 国文学研究資料館報告1号『国文学研究資料館におけるコンピュータ及び漢字システム  
 に報告

また C <sub>3</sub> の雑誌九十種と C <sub>2</sub> の現代新聞をあわせると、	
新聞にのみ出現した漢字	382字
共通して出現した漢字	2,831字
雑誌にのみ出現した漢字	497字
計	3,710字

である<sup>9)</sup>という。これも分野が異れば、字種が確実にふえることを示している。

これとは逆に少いデータの方の G<sub>1</sub>, G<sub>2</sub>, G<sub>3</sub>は単独の古典テキストという限られた分野である。さらに『保元物語』と『平治物語』の場合は、両者を合わせると、1,919となり、増加ぶりが著じるしく、漢字の字種がデータ量と分野によって大きく影響されるものであることも如実に示している。

次に D<sub>1</sub>, D<sub>2</sub>, D<sub>3</sub>の新聞の漢字調査を検討してみる。D<sub>1</sub>の場合は、1877年11月1日から1878年10月31日までの1年分の「郵便報知新聞」を資料としたサンプル調査である。異体字と本字との間に字形や文字使用の上であまり差異の見られないものは、本字に合併するという明確な方針をとっている。異体字一覧には180あまりの漢字が挙げられている。D<sub>2</sub>は朝日、毎日、読売の3紙の1966年の1年分に含まれる語彙を母集団として60分の1の面積費でランダムサンプリングを行ったという。記事面の性格を考慮してか、政治、社会、経済、文化・家庭、運動・芸能、広告の六層に分けている。D<sub>3</sub>は大阪毎日、東京朝日、読売、報知、時事新報の5紙を対象とし、1938年の1年間内の60日分について政治面と社会面に限って調査したという。異体字の取り扱いについては収と収、萬と万、體と体等は1つに数えているが、箇と個、灯と燈、埼と崎、著と着、附と付等の11例については、別個のものとして数えている。3調査とも若干の違いはあるがおおよそ比較にたえるであろう。

D<sub>1</sub>は母集団が最も小さいにもかかわらず、また単独紙の調査であるにもかかわらず、3,680字であり、さらに異体字も加えれば、3,800余数数えることになる。D<sub>2</sub>, D<sub>3</sub>と比較したとき時代の反映と見られようか。D<sub>2</sub>はD<sub>3</sub>に比べ各分野の記事を収集しているが、それでも300あまり減少している。これは漢字を少くするという点で言えば当用漢字制定の効果と見ることができよう。しかし当用漢字1,850字の1.74倍の漢字が使われていることや、使用頻度上位、2,000字の中に表外の漢字250字が含まれている<sup>10)</sup>という事実からしても、日本人の漢字に対するしたたかな執着と見ることができよう。

以上の考察により漢字の字種が使用分野によって大きく異ること及び母集団の増加によって字種も漸次増加するであろうことの2点が確認できる。

このことは漢字セットのきめ方がいかにむつかしいかを何よりも雄弁に語っ



ているであろう。また標準的な漢字セットを決めようとする場合には、このことを十分に考慮する必要がある。当然のことながら全ユーザーを満足させる漢字セットなどあり得ない。いくつ標準するかという数の問題ではないのである。

## 2-3. 漢字における字体

第3の問題は漢字の字体の問題である。もっと明確に言えば、漢字システムにおける異体字の多さであり、また正字のゆれの問題である。（ここで筆者の使用する字体と言うことばは、古体、別体、俗体等のことばで言われるような文字の骨格表現のことである。明朝体、宋朝体、ゴシック体等のことばで表現される書形、書体の問題は別である。）

この字体の問題を明確に説明した資料はとぼしい。比較的整理された説明をし、かつ字体を明確に識別して編集した辞書は『新字源』であろう。本辞典には、親字という概念があり、これに旧字体及び異体字がある。異体字の中は、異体字とその親字との関係を、本字、古字、別体字（或体）、俗字、誤字（譌字）の五種に区別している。旧字（体）は、「当用漢字字体表」などによって改められる以前の漢字を言っている。また「異体字」については、「おのおの親字に通用している異形の同音同義の文字」と定義している。さらに「本字」については、「従来正字形として承認されているもの、またはそのなりたちから考えて正字形とすべきもの」と説明している<sup>11)</sup>。

このことは正字が絶対不変のものではないこと、正字が常に親字となるとは限らないことを示している。つまり本辞典では、収録されている親字の字体のうち、当用漢字については、「当用漢字字体表」（昭和24年内閣告示第1号）人名用漢字については、「人名用漢字別表」、補正漢字については「当用漢字補正案」によったことが明記されている<sup>12)</sup>。

林大は、上古以来の字体の変遷について略述し、いづれを正とし、いづれを俗とするかきめられたものではないと把握した後、正体と別体を区別して取捨の標準を示した中で、「正体とは、説文、干禄字書、康熙字典等で普通に正字としたもの、別体とは同じく古文、本字、省字、通用字、今字、俗字、訛字等すべて普通に正字としていないものを言う」と定義して<sup>13)</sup>いる。ここでも正字を本源的なところにさかのぼって、字源主義的に定めていないのである。

一方、中国古代の甲骨文や青銅器にきざまれた金文を手がかりに漢字の始源にまで遡り、個々の漢字についてその正しい形を説き示そうとする白川静の立場もある。この立場は、漢字の始源を明かにし、その上で正字（本来の字形）

を明らかにすることが可能であるとし、かつ自ら精力的にその正字を説き示している<sup>14)</sup>と言えるであろう。

私は、このような正字を求める立場が尊重されるとともに、正しい字形が一つでも多く明らかにされることを望む。しかしこれら正字体が明らかにされたとしても、現在の字形をすべて正字形に改めていくことは、もはや不可能な部分が多く、現実的に処していかなければならない。そして現在可能なことは、何らかの基準に基づいて、正体とすべきものと、それ以外の異体字とを整理し、階層づけておくことである。

この考え方を実行するために、私は漢字ソーラスを作成した<sup>15)</sup>。その詳細については、文献[15]にゆずるが、その概略を記せば

#### 1) 漢字ソーラスの定義

個々の漢字を対象として、同義又はそれに近い関連を有する漢字をグループリングし、漢字相互の関連を階層的に表現した漢字辞書である。

#### 2) 漢字ソーラスの対象範囲

ディスクリプタの範囲は、当館コンピュータシステムの文字セットであること。非ディスクリプタは『新字源』『大漢和』の収録漢字と、J I S漢字表(J I S C6226<sup>-1978</sup>)を対象としていること。

#### 3) 非ディスクリプタの種類

ディスクリプタ(親字)に対する非ディスクリプタ(異体字)の関係を、本字、古字、別体字、俗字、譌字、同字、旧字、簡略字の種類に整理したこと。

#### 4) ディスクリプタの基準

ディスクリプタになりうる語の基準は、新字源の親字の立項に準拠させ必ずしも正字をとっていないこと。

の4点にある。

この漢字ソーラスについて分析のため分類してみれば、表-3のようになる。ここでは非ディスクリプタ(異体字)の多さが問題となろう。ディスクリプタ5,699字に対し、J I S内で650の非ディスクリプタが存在し、さらにJ I S外の漢字も対称にすれば、異体字として5,689字の漢字が存在する。このことは、5,699で表現しうる概念に対し、さらに5,689のバリエーションが存在するのであるから、約二倍の冗長性をもっていることになる。

この事実の妥当性を考えるためには、より作業をすすめなければならないが、私の調べたところによれば9921字を収録している『新字源』において、異体字の数は、2,526字である(表4参照)。実に25.5%を占めている。収録字数

表-3 漢字ソースラスの分析

表3-1 J I S 漢字におけるディスクリプタと非ディスクリプタ

	数
ディスクリプタ	5,699
非ディスクリプタ	650
計	6,349

表4. 『新字源』における異体字の数

親	字	7,395	
異	体	字	2,526
	計		9,921

表3-2, ディスクリプタの分類

		数		
非ディスクリプタを持たないもの（一字種一字体のもの）		722		
非ディスクリプタを持つもの		4,917		
非ディスクリプタの内訳				
	本	字	719	
	古	字	1,405	
	別	字	504	
	俗	字	654	
	譌	字	51	
	同	字	1,531	
	旧	字	819	
	簡	略	字	6
	計		5,689	

が増えれば、異体字の占める割合も高まっていくことが考えられる。先の約二倍という数値の妥当性はともかくとして、漢字が冗長性にとんだ文字体系であることは確認できるであろう。（尚このソースラスの分析データは1979年10月15日現在のものである。その後データの整備を行っており、数字の変化はある）

この事実は、漢字情報処理システムに対し、次の二つの点を示唆する。一つは字体、字形の問題を無視して漢字セットを考えるならば、漢字セットの巨大化をおしすすめることになるであろうこと。二つには、漢字による表記を、漢字表記できることに重点をおくなら、つまり一字種一字体として設計するならば、現在存在すると思われる漢字の約半分のもので、間にあう可能性があると思われることである。

## 2-4 標準文字セットの基本

以上考察してきたように、

- 1) 把握しきれないほどの漢字の数
- 2) 分野間における使用字種の異り
- 3) 大量な異体字の存在

の3つの問題をあきらかにしてみると、漢字セットを設定するにあたっての間

題点もあきらかであろう。つまり

- 1) あらかじめ漢字セットのトータルがいくつになるかは、きわめて困難な推測をしなければならないこと。
- 2) 分野間によって使用字種が異なるのであるから、標準セットの作成はむづかしい。ユーザーごとに別の漢字セットを必要とする可能性が高い。
- 3) 大量の異体字が存在するのであるから、少々の漢字について、複数体セット化したとしても多くの効果はあらわれない。

の3点があきらかにされた。これらの点から考えて、標準的な文字セットを作成する場合には、第1にその文字セットは、全分野で共通に考えうるような、逆にどの分野でも冗長性を持たない適切な文字セットとなりうるものであること、第2に、異体字は完全に無視するか、或は大量に存在することを前提としてコード体系を作ることにあると考えられる。

### 3. 漢字セット運用上の問題点

漢字システムを、前章において分析したように把握するならば、漢字セットにフレキシブルなものが要求されることは自明であろう。完全に固定にしておくことが不可能であるとなれば、漢字情報処理システムに要求される機能として、外字が発生したときに、システムの外的に外字が処理できること、的確に外字であるか、内字であるかの判断のできるツールが用意されていること、そしてあらたに発生した漢字を辞書に登録し利用できるようにすること、つまり

- 1) 外字処理
- 2) 漢字セットのインデックス
- 3) 文字登録の機能

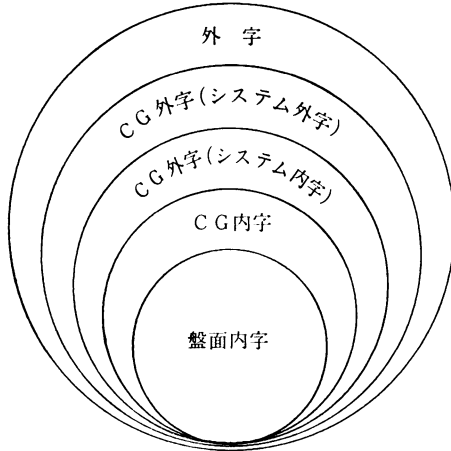
の3点が、漢字情報処理システムの運用上重要な問題であろう。ここでは1)及び2)の問題にふれる。3)については、登録の機能の他に文字フォントのデザインの問題、明朝体活字の歴史の問題など考えなければならないことが多いので、別稿で論ずることとする。

#### 3.1. 漢字情報処理システムにおける内字と外字

##### 3.1.1 内字と外字の概念

漢字情報処理システムにおいて、文字を出力（印字）する為には、何らかの形でその文字が CG (Character Generator, 文字発生装置) の中に収容されていなければならない。また収容されていない文字を出力するためには、一時的にもせよ CG を経過しなければならない。この原則に基づいて内字、外字等

図1 内字と外字の概念



の概念を図示すれば、図-1の様になる。またここで使用している用語は、以下の様に定義して使用する。

盤面内字とは入力機の盤面上にあり、通常の入力法で、盤面を指示することによって入力しうる文字である。CG内字は、原則としてCGの中に常駐している文字である。この文字セットはダイナミックにいれかえるとき以外は固定のものであり、使用頻度の高いものである。CG外字とは、システム内字とシステム外字に二分される。システム内字は通常存在するときは、中央側ディスクの中とか、CG以外のフロッピーディスクの中等、CGの外にあり、外字と見るべきである。しかし一つのシステムの中では必要に応じ、CG内字化して利用される。この内字化の方法は、一時的にCG内字の部分に外字エリアを設けて内字化する方法と、あらかじめそのエリアが設けられている場合とがありうる。システム外字とは、そのシステムにおいて文字の利用基準、追加基準に合致し、必要が生じればシステム内字化して使いうる文字である。最後の外字とは、ある一つのシステムの中では原則として使わない文字である。大きく内字と外字に分割する場合には、盤面内字とCG内字をあわせて単に内字と言い、CG外字を単に外字ということにする。

そして外字処理とは、このCG外字を内字化して使用する場合の処理方法のことである。

このように文字セットにおける内字と外字の概念を整理した上で以下の論をすすめる。

まずCGの中にはいる文字種が固定されていると、前章で考察したように、使用文字種は分野によって、また資料によって著じるしいばらつきがあるものであるから、利用効率に大きな影響を与える。ことにCGが大きければ大きいほど使用されない文字の数も増すことになる。

また標準文字セット（J I S）の問題がある。J I S C-6226<sup>-1978</sup>において標準化された文字セットは、第1水準の非漢字453種、漢字2,965字、第2水準の漢字3,384字、計6,802種である。現状のハードウェアは、CG内字として7,000字～8,000字位を収容するものが多いようである。仮に8,000字ということと数えれば、現在の標準化された漢字のすべてとユーザー固有の文字をいれろとすれば、約1,000位いまでしか余裕がない。

この文字セットが現状のシステムでどのように利用されているかを、国文学研究資料館のシステムに例をとって調査してみれば、表-5のようになる。

表-5 J I S漢字の使用率

字 種	異なり字数	使用率(A)	総 字 数	使用率(B)
漢 字	3,825	60.2	597,286	25.3
J I S 第1	2,617	88.3	583,978	
J I S 第2	1,052	31.1	12,955	
J I S 外	156	※ 4.08	353	※※ 0.06
漢 字 外	293		1,759,193	
合 計	4,118		2,356,479	

使用率A：各漢字セットに対する使用率

〃 B：総字数に対する漢字の使用率

※, ※※：出現した漢字に対するJ I S外漢字の占める割合

詳細は文献〔5〕で解説しているのて、そちらに譲るが、この対象となったデータは、国文学研究資料館で抜かうデータの中の、原本書誌データ、論文リストのデータ、保元物語及び平治物語の本文データの三種類を総合したものである。これによれば、第1水準の漢字2,965に対し、このデータで2,617字が出現したのであるから、88.3%の使用率を示している。第2水準の漢字は3,384に対して、1,052であるから、31.1%の使用率である。第1、第2をあわせたJ I Sの全体に対しては、60.2%である。J I Sの漢字156字は出現した漢字

3,825に対しては、4.1%を占める。

J I S外の漢字は、当館のシステムでは、これまではすべて内字化されているが、今後はC G外字（システム内字）のエリアに収めていかなければならない。ここで仮にJ I S内の文字をC G内字とし、J I S外をC G外字としたとすれば、前者の使用率60.2%に対し、後者は100%の利用率である。この仮定の段階で156の字は、J I Sの6,349に対し、わずかに2.5%にしかならないから、実際上は外字処理としてもそう大きな影響は与えないかも知れない。しかしこれが500字、600字という段階になってきて10%近くにもなってくると、システムの稼動効率に与える影響はきわめて高くなるものと思われる。つまり漢字情報処理システムにとって、外字の処理機能又は文字セットの設計運用方法が重要な役割を果たすことを意味している。また文字セットの管理も、システムの効率をあげる意味で大きな役割を荷っているのである。

### 3.1.1 外字処理の機能

漢字情報処理システムの重要なファクターとして、外字処理の問題がある。その内容は1)外字処理が発生した場合の処理速度（印字速度）の問題であり、2)は操作性、つまりいかようにすれば外字処理が可能となるかであり、3)は、外字登録、つまりシステム外字をシステム内字化する方法の問題である。

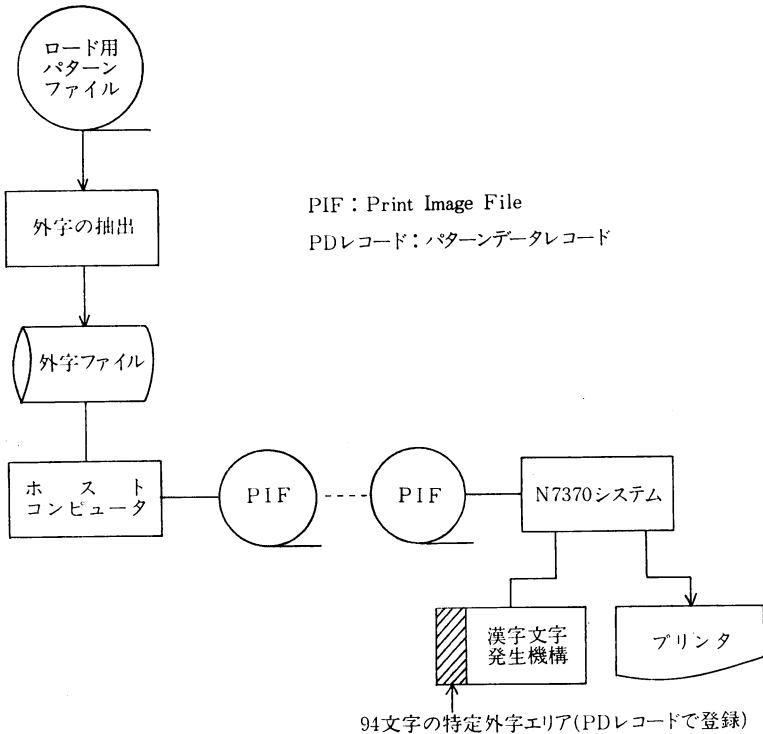
1)の問題は現在C Gの中に、あらかじめある単位の外字エリアを設けておき、そこに処理に先だってロードしておくか、外字の発生時点でロードして処理するかとの二つの方法が考えられよう。そしてこの外字エリアをどの位の大きさにしておくことが効率的であるかであり、それが操作にどのような影響を及ぼすかである。現在すでに稼動又は提唱されているシステムのうち、N7370システム（NEC）、J E Fシステム（ファコム）、H-8195及びH-8196（ハイタック）の3例について整理比較してみると、図-2に示したようになる。

N-7370システムの場合は、①ロード用パターンファイル（当該ユーザーの使用する全文字を収容）から外字だけを抽出し外字ファイルを作成する。②PIF（Print Image File）を編集・作成する段階でC Dレコード（文字データレコード）の中から外字の有無をチェックし、特定外字コードに割り当て、P Dレコード（パターンデータレコード）を作成する。特定外字エリアは94文字である<sup>16)</sup>。外字処理を実施する場合には、アプリケーションプログラム内で外字処理ルーチンを作成することになっている。

この方式の特色は、外字ファイルとC G内の特定外字エリアの2つで外字エリアを形成していること。特定外字エリアは、1ページの最大文字数が94文字

図2 外字処理方式の比較

図2-1 N7370システムの場合

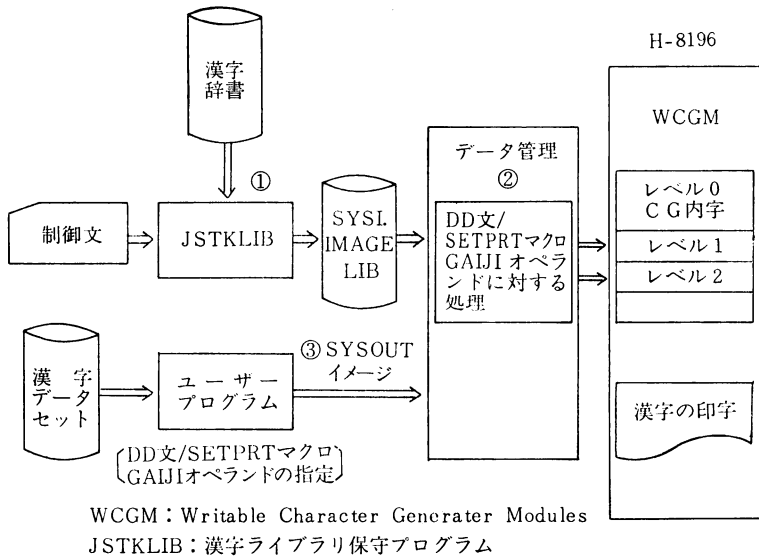


で、ページ単位で外字処理が行なわれること。実施はアプリケーションプログラムに委かされていること等にある。

HITAC の8195及び8196 (日立製作所) の場合は、外字処理にはプレロード処理とオンディマンド処理の2通りがある。その外字処理の流れを図示すれば、図2-2, 3, 4 になる。漢字辞書に対する外字の登録はすでに完了しているものとして図示している。H-8196 プレロードの場合は、①プレロード外字用の外字イメージを JSTKLIB により、SYSI. IMAGELIB データセット中に作成する。②データセットオープン時に、データ管理によりDD文の GAIJI オペランドで指定された外字イメージが漢字プリンタにロードされる。③外字を含むデータが漢字プリンタにより WCGM が参照され、文字が出力される。という流れになる。またオンディマンドの場合は、①外字を含むデータを漢字プ



図 2-2 H-8196におけるプレロード外字処理



リントに転送，②漢字プリンタより外字の報告，③③' 漢字辞書中の文字パターンをデータ管理が読み取り，漢プリに転送する場合と，外字出ルーチンに制御が渡り，このルーチンで漢字辞書から文字パターンを読み取り，漢プリに転送する。これにより漢プリから文字が印刷される。H-8195のオンディマンド処理の場合は，①外字を含むデータをSYSOUTに出力，②出力ライターが漢プリに転送 ③漢プリは外字の発生を出力ライターに知らせる ④出力ライターは文字パターンを読み取る ⑤出力ライターは読んだ文字パターンを漢プリに転送，漢プリは文字を印刷する。という処理になっている<sup>17)</sup>。(H-8195に於けるプレロード処理は省略)

このシステムでは，一度プレロード処理をしてWCGMに登録された文字は，次のプレロード処理まで一時的にCG内字化することになる。またオンディマンド処理は，1ページ内で一時的にCG内字化することになる。内字化された後は，外字処理と言えども，一般の処理に比し処理速度への影響はないで

図 2-3 H-8196におけるオンディマンド外字処理

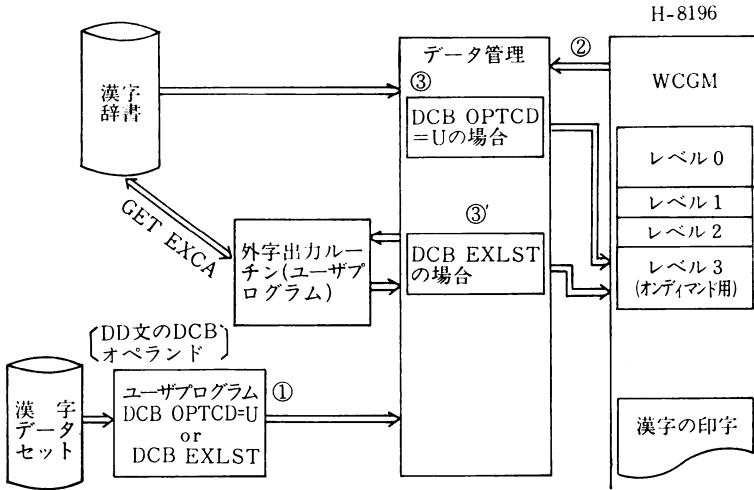
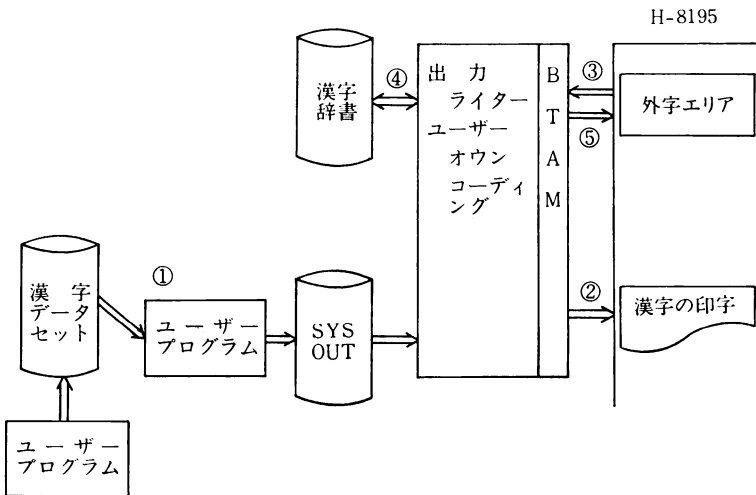


図 2-4 H-8195におけるオンディマンド処理



あろう。プレロードの処理に一つの処理が加わるだけということになる。オンディマンドの場合は、この点の処理は不要であるが、ページ内の内字化であ

るから、その発生頻度は高くなると思われるので、処理速度への影響は避け難いであろう。ユーザにとって方法の選択が可能であることは、操作性、使用効率を高める。また外字領域の構成も重要な側面であるが、H-8196のWCGMの構成は、レベル0から3までに区分されていて、CG内字領域、プレロード外字領域1,同2,オンディマンド外字領域となっている。最大登録CG外字の数は1,299字である。またオンディマンド用は256/頁である。H-8195の場合は、オンディマンド、プレロード共に領域を共有し、最大480/頁である<sup>18)</sup>。

次に『日本語情報システム概説書——JEF——』（富士通）によれば、このJEF（Japanese Processing Extended Feature）とよばれるシステムの中にはいわゆる外字処理と言う言葉はない。逆に文字セット管理の概念なるものが導入されている。この骨子はCGに収容する文字パターンを、利用者が扱うデータに使用される文字種に基づいての選択、使用文字種がCGに収容されているかどうかのチェック、未収容文字パターンのCGへの追加処理であるという。またこの文字パターンのローディング方法として、1)イニシャルローディング、2)プレローディング、3)ダイナミックローディングの3種類がある。1)は装置への電源投入時に行なわれるものである。2)は文字どうりプレローディングである。3)は先のオンディマンド処理と同種のものである。また、CGの領域は基本領域（イニシャルローディングされたCGの領域）と追加領域があり後者が追加文字パターンの領域であると説明されている。これはこれまで述べてきたところの外字処理である。そして文字セットの管理を行うソフトウェアはADJUSTと呼ばれるものであるという。この中には処理装置から出力可能性のある文字種を洗い出す機能もあるという<sup>19)</sup>。概説書を資料としているだけであり詳しいことはわからない。また追加領域の大きさもわからないので、実際の運用形態を予測することがむづかしい。しかし概説書から判断すれば、操作性が著るしく繁雑となるように思われる。

以上、3社で発表しているシステムについて紹介したが、システムにとって必要な機能は、内字であるか、外字であるかは、意識せずに使えることである。具体的に言えば、処理速度に大きな影響を与えないことであり、コード体系も一つのものであることである。このためにはハードウェア及びソフトウェアの設計上の問題に加えて、当初の文字セットをどう作成しておくか、また作成した文字セットをどう運用していくかといった総合的視点から見てゆかねばならない。

また、文字登録の機能の問題、つまり外字をシステム内字化する問題については、運用上から考えれば、その必要性が確認された時点（これは主に初期デ

ータの入力時である)で、即座にしかも簡便に登録できることである。この確認のために漢字辞書の活用がクローズアップされてくるのである。この登録の機能として必要なものは、既存の文字パターンを参照し、それに近いデザインが可能になることである。これは文字列として美しく見られ、かつ読み易くするためには、デザインの統一が不可欠だからである。このために、既存の文字をドットパターンで参照できる機能が必要である。これについては前記の文字登録の機能の問題であるので、ここではこれ以上ふれない。

### 3-2 漢字情報処理システムの為の漢字辞書

漢字システムの性格上、あらかじめ固定的な文字セットを定めることは不可能なものであることや、文字セットのすべての文字のコードを覚えておくことは不可能であることから、漢字情報処理システムでは、何らかの形で使用文字種のコードブックを持つことが不可欠である。しかも追加がある以上、メンテナンスが必要であり、しかも比較的頻繁に行なわれるので、何らかの形で機械可読のファイルとして保有することが必要である。また先ほどドットパターンの参照について指摘したが、このドットパターンも辞書の一部にすべきであろう。

つまりここでいう漢字情報処理システムの為の漢字辞書とは、ドットパターン情報(フォント情報)と漢字属性情報の二つを必要とする。前者は、漢字プリンタ及び漢字ビデオデータターミナル等の漢字出力機の外字用のフォントファイルとして使用し、かつ外字の新規作成に際してのフォント情報の参照ファイルとして使用するものである。後者はデータ入力に際して、盤面外字をコード入力する際のコードブックとして、また不明漢字の確認用として、さらに新規に文字を登録する際の確認用として使用する。

#### 3.2.1 漢字属性情報

今日、漢和辞典で漢字を検索しようとする場合、部首、総画、音訓の3つが主に使われているであろう。この三つの方法はそれぞれ長所と短所もっている。つまり部首索引はきわめて単純化された部首と部首内画数で検索するが、この部首は康熙字典での部首分類を行ってより、ほぼこれが踏襲されており、『大漢和辞典』では216種に分類し、最近の『新字源』、学研漢和大事典(藤堂明保編、学習研究社、1978)等では、245種に分類している。漢字における字体、字形の変化は、部首の認定をむつかしくしている場合があり、また部首内画数も変化がある。例えば“当”の字は“田”の部に収録されている場合が多い。しか

しここからは現行の“当”の字はひけない。このような例は、盡（皿部）→尽、豫（豕部）→予等がある。

総画索引の場合には、字体、字形が固定していれば、最も確実に検索する方法であるが、スピードの面で大きな障害がある。同時に省画による字形変化している漢字も多く、その対応がむづかしい。

音・訓の索引も読めることが前提であるから、一般性にとぼしい。

このように現在一般に使われている漢字の検索方法には、完璧なものがない。この点を考えると複数の検索方法を用意し、自由な選択ができる必要がある。つまり表-5に示すものが考えられる。使用漢字コードは、JISコードと体系の異なる漢字コードを使用している場合には、JISコードも必要となる。項番2～7まではすべて検索のためのものであり、8～9が管理上必要とするもので、8の字種は当用漢字、人名用漢字等のいわゆる政令漢字の情報が必要となる。これにより字体の変化が起こっているからである。

表-6 漢字属性情報

項番	収 容 情 報
1	使用漢字コード
2	辞典の検字番号(1)
3	” (2)
4	部 首
5	総 画 数
6	四 角 号 碼
7	読 み (音・訓)
8	字 種
9	作成・保守上の管理情報

辞典の検字番号は、漢字の戸籍簿である。

したがって漢字すべてを網羅した辞典の番号がほしい。それがあれば漢字システムを使用しているユーザー間で漢字の識別が確実にになり、不注意で漢字を作成するということがなくなるであろう。しかしこのすべての漢字と言う概念は、2-1で考察したように、あいまいであって捉えきれない。現状で最大の収録文字数があり、かつ漢字に番号がついているのは、『大漢和辞典』であるから、この番号となる。またハンディな利用も考えなければ

ならない。ここでは当用漢字の制定によって新しい字体が認定されたことを考慮にいれれば、当用漢字制定後に編集された辞典であることが必要条件となる。この意味で『新字源』（文献11）が適当と思われる。要するにここで指摘しておきたいことは、ハンディな辞書と収録字数のできるだけ多いものとの二つが必要であるということである。

部首、総画数については、学問的正しさのみならず、考えられる部首や総画数は、複数個つけておくべきであろう。

四角号碼索引とは、漢字の四すみの筆形、筆画をみて、それぞれあらかじめ約束されている0～9までの番号におきかえ、さらに付角をつけて漢字一字を5桁の数におきかえて排列していく方法である。これは四角号碼に限るもので

はないが、音・訓や部首等を無視しても、字形から漢字を検索できる手段がほしいという意味であげたものである。四すみの筆形・筆画が索引のポイントとなるため、これが普及すれば、漢字の字形への関心を高めることになるであろう。

このことの方が、漢字システムを考える時には、はるかに重要な意味をもつであろう。漢字情報処理システムの為の漢字辞書、そのための索引とは、言うまでもなく確実に対象（漢字）に達しうるものでなければならない。これが不完全であれば、入力ミスや新規に漢字を追加する際のコードづけにミスを犯すことになる。このことを重視して実用的なものを作成することが何よりも重要なことになるであろう。

尚、日本電子工業振興協会では、日本語処理に最も基本的なデータベースの一つという位置づけで、漢字辞書を実際に作成している<sup>20)</sup>。その仕様は網羅的なものである。筆者案はユーザーを対象としたものである。筆者案に基づく漢字辞書もすでに作成し、活用している。これは別図報告の予定がある（国文学研究資料館報告、第6号、1980年3月に刊行予定）のでご参照願いたい。また日立製作所でも作成<sup>21)</sup>しすでにユーザーに提供されている。比較参照願えれば幸いである。

### 3.2.2 漢字シソーラス

ところで、新規に文字を登録する際に漢字辞書を使用すると言ったが、この目的のために、漢字シソーラスが活用できればより良いものになる。異体字が存在する以上、そのすべてをコード化していくことは合理的ではない。漢字の新規登録の際には、その漢字をどの字体で登録すべきか、或は別の字体で登録されているかどうか等、十分に検討すべきであろう。そのツールとして私は、2.3で簡単に紹介したような漢字シソーラスを研究開発した。これを活用すれば、ある漢字に対する同義の漢字の一覧、すでにある漢字のコードとその字体、その他の関連字の情報等が一覧できるのである。すでに異体字が辞書の中に複数体存在する場合には、新たに登録しようとする場合に、より慎重な検討を要求するであろう。文字セット拡大への抑止効果はきわめて重要なものであろう。

以上のように豊富なツールを完備し、迅速な処理が行なえる環境の整備と、文字管理への真剣なとりくみが、漢字情報処理システム運用上の責任であろう。

#### 4. むすびにかえて

考えてみれば、従来の漢字システムはすべて外字処理で行ってきたようなものであると思う。これまで行なわれている一般の鉛活字による印刷や、写植システムの中では、新しい漢字（字母として用意されていない漢字）があらわれた時、その印刷の為の手書原稿に合わせて字母を作ることによって処理してきた。この点で漢字は個人のレベルでのきわめて弱い管理はあったとしても、体系的な管理はなされていなかったと言えるのである。近世における版本という印刷形態はその典型を示している。版木に書かれた原稿をそのまま、そのつど彫る方法は、いついかなる文字があらわれても何ら問題が生じなかったのである。ところが漢字情報処理システムとなって、使用する漢字のすべてにコードが与えられ、使用する漢字が漢字辞書で管理され、管理された文字を使用する状況になれば、無制限に新しい漢字を作成し、バリエーションを発生させることはなくなるであろう。これは漢字使用にとって革命に近い問題であろう。

漢字情報処理システムとは、このような崇高な課題を背負っているのである。漢字情報処理システムの課題としては、入出力システム、ソフトウェア体系等、本稿では論じ得なかった多くの問題がある。また本稿の展開として論ずべきものに、漢字のデザインの問題や、漢字コード体系の問題などがある。しかしこれらについては論じられなかった。いわばハードウェア、ソフトウェア以前の漢字システムの内部の問題の一部を論じたにすぎない。私があえてこれらの問題について考察したのは、従来漢字システムそのものを十分に分析、把握した上で、漢字情報処理システムを設計するということがなかったように思えるからである。またこうすることなしにシステムが存在しえたのは、特殊な機器として位置づけられ、きわめて部分的にしか使われていなかったことと、どうしてもシステム化しなければならない分野の、特殊な篤志ユーザーによって支えられていたことによるものであろう。

しかし現実には、このような時代を超えて、数年前の感覚からすれば、驚異的とも思えるほどの勢いで漢字情報処理システムの普及がはじまろうとしているのである。このような時に、現状の漢字システムを十分に分析し、把握し、その上でメーカーもユーザーもシステムを設計し実施に移していくことでなければ、数年後に高価な代償を払うことになるかも知れないのである。

(1979. 10. 19)

## 抄 録

漢字システムの現状を分析し、漢字情報処理システムの設計にあたって配慮しておかなければならない基本的な問題、それもハードウェア、ソフトウェア以前の問題を分析したもの。まず漢字の数として5万余存在するところまではあきらかであるが、その総数がかみ難いこと。各種の漢字調査を分析すると、分野により字種が異なり、母集団の増加は字種の増加につながるものがあきらかであること。漢字シソーラスの作成及びその分析から、字体の多様性が大きな問題であること。の3点、つまりオープンシステム化しすぎていることに問題があることを論証した。同時に効率良い文字セットを設計することの重要性を論証した。その上で漢字情報処理システムとして、外字処理の機能の重要性と、ユーザーとして望ましい方式の概要を述べた。またシステムの運用にあたっては、漢字辞書の活用を位置づけることと必要とする辞書検索の機能を考察した。また漢字情報処理システムが漢字を管理しなければならないという側面（漢字システムのクローズ化）で重要な意味をもつことを指摘した。

### 〔参考文献及び補〕

- 〔1〕 日本工業技術センター編：漢字情報処理システム資料集，（1979），及び関係者との談話による。
- 〔2〕 林大：漢字の問題，岩波講座日本語3，（1978）
- 〔3〕 〔2〕に同じ
- 〔4〕 J I S C-6226<sup>-1978</sup> 情報交換用漢字符号系，1978年1月制定
- 〔5〕 田嶋一夫：J I S漢字表の利用上の問題——漢字処理システムにおける漢字のデザインと管理——，情報管理21—10，pp753～761
- 〔6〕 杉本つとむ編：『異体字弁』の研究並びに索引，文化書房博文社，（1972）
- 〔7〕 白川静著：漢字百話，中央公論社（1978）
- 〔8〕 杉本つとむ編：異体字研究資料集成，第十巻解説，（1973～1975）
- 〔9〕 野村雅昭：新聞の文章に使われた漢字，言語生活，No.285，pp.27～36
- 〔10〕 〔9〕に同じ，
- 〔11〕 小川，西田，赤塚共編：『新字源』の凡例，角川書店，（1963）
- 〔12〕 〔11〕に同じ，
- 〔13〕 林大：当用漢字字体表の問題点，覆刻文化庁国語シリーズ漢字，教育出版社（1974）
- 〔14〕 〔7〕に同じ，
- 〔15〕 田嶋一夫：漢字シソーラスの作成—漢字情報処理システムの問題点とその対策，第16回科学技術研究集会発表論文集，日本科学技術情報センター，1980.3（予）
- 〔16〕 日本電気株式会社：シリーズ77，N7370高速漢字プリンタシステムシステム概説書（NDA01-1），1978初版，



- [17] 株式会社日立製作所ソフトウェア工場技術部，HITAC vos2/vos3 漢字プリンタ使用の手引，1979，第3版，
- [18] [17] に同じ，
- [19] 富士通，日本語情報システム概説書——J E F——，1979
- [20] 日本電子工業振興協会編：日本語情報処理の研究調査，日本電子工業振興協会，1979
- [21] 日立製作所：HITAC 標準提供漢字辞書，1978第1版