技術部教育課

47-R 001

日本語情報処理の技術動向調査報告書

昭和48年3月

JIPDEC

財団法人 日本情報処理開発センター

この報告書は、日本自転車振興会から競輪収益の一部である機械工業振興資金の補助を受けて昭和47年度に実施した「日本語情報処理の技術動向調査」の成果をとりまとめたものであります。

JIPDEC

昭和 47 年度事業成果報告書要約

日本語情報処理の技術動向調査報告書　47－R－001

情報処理の分野では，これまでアルファベット，数字，記号などが入出力表示の中心となっている。しかしこのような表示方法は，日常日本語を使用しているわれわれにとって，あまり便利な表示方法とはいい難い。

このような背景から，最近に至って、古くから印刷部門の技術革新という形で始まった日本語文（漢字かな混り文）の機械処理技術を，電算機の処理能力に結びつけようとする研究，あるいは電算機システムで文字をジェネレートしようとする研究などが活発化し，日本人にとって，より便利で，より広範な利用分野を生み出そうとする努力が始まってきた。その具体的なあらわれは、たとえば新聞，出版，印刷の分野における文選，編集，組版作成のシステムや，官公庁の情報検索システムなどである。

このように，技術的に重要な黎明期にあたって実施した本調査は，日本語による情報処理が当面するさまざまな問題，あるいは具現化しつつある技術動向などを体系的にさぐりながら，この分野における将来の姿の展望を試みたものである。

とりまとめた調査報告書は，次の７章で構成してある。

1　日本語情報処理のニーズと社会的影響

前半では日本語情報処理の背景となる社会的動向と技術的動向を述べ，さらにカナ字文ローマ字文に始まるこの技術の発展過程を解説している。後半は日本語情報処理の適用分野についての現状と今後の動向を概観的にまとめ

財団法人　日本情報処理開発センター

東京都港区芝公園３－５－８（機械振興会館内）　電話 434－8211（大代表）

本報告書についてのご照会は，内容については技術課（村上，内 536 ）に，入手については，庶務課（内 470 ）へご連絡下さい。

てある。

2. 日本語の特質と言語情報処理

この分野に携わるためには，日本語そのものの特質を十分理解しておかねばならない。そのため本章では，言語情報処理の立場からみた日本語の特質をまとめている。

主な内容は，諸外国の言語との比較における日本語の特質，言語処理上の問題点に関する解説であるが，特にこの分野で最大のウエイトを占める漢字処理について，分野別の使用字種，書体，字形，配列，音訓など多角的に検討を行なった。また最後に，機械処理の立場からみて望ましいこれからの日本語に関する意見について付言している。

3. 日本語情報処理の入力システム

まず，システムのなかで入力機器をどのように使うのか，どのような位置づけにあるのかを，実例によって考察している。次に現在商用化されている各社の製品についての原理，構造，性能等の体系的な解説を行なうとともに，漢字を構成要素に分解する方式，手書き文字入力方式，パターン認識，音声入力など，内外各所さまざまな試みを紹介した。

4. 日本語情報処理の出力システム

出力システムの面からとらえた文字，文字品質，文字発生方式などの解説とハードウェアの説明を行なっている。ハードウェアに関しては，文字出力システムとしてのハードプリント，ソフトプリント，COMプリント，電算植字，あるいは音声出力としての録音編集方式，パラメータ合成方式について，各社の商用機種と研究結果について述べている。

5. 日本語情報処理のための利用技術と標準化

広範な利用を促進するためには標準化は欠かせない。ここではハード，ソフト上の標準化に伴なう問題点を指摘し，若干の分析を行なった。

6. 日本語情報処理の応用

現在稼動中の代表的なシステム例を，各利用分野別に細く紹介するとともに，ユーザー覧表を掲げてある。

7. 日本語情報処理の開発動向

来るべき情報化時代における日本語情報処理の技術動向と利用動向について，その見通しを体系的にまとめてある。

序

情報処理の分野では，これまでアルファベット文字を中心とする，データの入出力が一般に行なわれてまいりました。しかし，アルファベット以外の文字を国字とする国の人々にとっては，すべてアルファベットで記述しなければならぬという不便を感じさせてきたことは事実であります。

こうした背景から，最近わが国では日本語文（漢字かな混り文）のままで情報処理を可能ならしめようとする研究がすすめられております。

その具体的なあらわれとして，ここ数年来このための各種の機器あるいはシステムがあいついで発表され，将来の自然言語による情報処理への道の一つがひらけつつあることは，まことによろこばしいことであります。

この調査は日本語による情報処理が当面する，さまざまな問題あるいはその動向をさぐりながら，この分野における将来の姿を展望しようとするものであります。

本報告書が各方面で利用され，わが国の情報処理普及に寄与できますよう念願いたしております。

最後にご多忙のなかを，この調査に種々ご協力いただいた委員ならび関係者各位に厚くお礼申し上げる次第であります。

昭和48年３月

財団法人　日本情報処理開発センター

会長　難波捷吾

日本語情報処理技術動向調査委員会

（敬称略50音順　・印小委員会委員を兼ねる）

委員長	・高橋達郎	（特）日本科学技術情報センター
委　員	相澤輝昭	日本放送協会
	伊木俊昭	日本アイ・ビー・エム株式会社
	・石綿敏雄	国立国語研究所
	磯部光朗	日本電気株式会社
	稲場公誌	ペティ株式会社
	後藤榛男	株式会社学習研究社
	斎藤秀紀	国立国語研究所
	篠田達郎	富士通株式会社
	・長井　坦	株式会社日立製作所
	西村靖雄	内閣官房内閣調査室
	・長谷川実郎	日本電子産業株式会社
	船崎武男	行政管理庁
	宮地　仁	（社）共同通信社
	武笠二郎	工業技術院
	森　健一	東京芝浦電気株式会社
	・安田寿明	東京電機大学
	保原　信	電気通信大学
	矢田光治	電子技術総合研究所
	渡辺龍雄	通商産業省
	篠崎　敬	（財）日本情報処理開発センター
	中嶋栄之助	（財）日本情報処理開発センター
協力者	坂井利之	京都大学
	長尾　真	京都大学
	高木　晃	株式会社特許データ・センター

事務局　財団法人　日本情報処理開発センター技術部技術課

目　　次

1.　日本語情報処理のニーズと社会的影響 …… 9

1.1　日本語情報処理の背景 …… 9

1.1.1　日本語情報処理の概念（社会的動向） …… 9

1.1.2　日本語情報処理の機能と問題点（技術的動向） …… 11

1.2　日本語情報処理技術の発展過程 …… 14

1.2.1　ローマ字文 …… 14

1.2.2　カナ文字文 …… 15

1.2.3　和文タイプライタ …… 15

1.2.4　活版印刷 …… 16

1.2.5　写真植字機 …… 17

1.2.6　漢字印刷電信装置 …… 18

1.3　日本語情報処理の適用分野 …… 18

1.3.1　情報産業 …… 18

1.3.2　事務機械化 …… 22

1.3.3　言語処理 …… 22

1.4　まとめ …… 23

2.　日本語の特質と言語情報処理 …… 25

2.1　日本語の特質と言語情報処理 …… 25

2.1.1　日本語の特質 …… 25

2.1.2　日本語の特質と言語情報処理上の問題点 …… 31

2.2　漢字 …… 36

2.2.1　字種 …… 36

2.2.2　書体 …… 39

2.2.3　字形 …… 39

2.2.4　漢字の機能 …… 42

2.2.5　漢字の配列 …… 43

2.2.6　音訓 …… 44

2.3　これからの日本語 …… 45

3. 日本語情報処理の入力システム …… 49
3.1 日本語情報入力システムのソフトウェア …… 50
3.1.1 システムの流れ …… 50
3.1.2 入力手順 …… 50
3.1.3 校正サイクル …… 51
3.1.4 エラー …… 52
3.1.5 入力字種，コードおよび字体 …… 52
3.2 日本語情報入力システムのハードウェア …… 53
3.2.1 エントリー形式の分類 …… 53
3.2.2 人手を介するエントリー方式 …… 55
3.2.3 人手を介しないエントリー方式 …… 69

4. 日本語情報処理の出力システム …… 75
4.1 文字 …… 75
4.1.1 文字種および文字コード …… 76
4.1.2 字体 …… 78
4.1.3 字形 …… 79
4.1.4 文字品質 …… 80
4.2 文字発生方式 …… 81
4.2.1 活字打けん方式 …… 81
4.2.2 ドット方式 …… 82
4.2.3 ストローク方式 …… 83
4.2.4 字母アナログ方式 …… 83
4.2.5 字母光学方式 …… 85
4.2.6 その他の方式 …… 85
4.2.7 外字処理 …… 86
4.3 文字出力システム …… 87
4.3.1 ハードプリントシステム …… 87
4.3.2 ソフトプリントシステム …… 89
4.3.3 COMプリントシステム …… 90
4.3.4 電算植字システム …… 90

4.4　音声出力システム …… 92
4.4.1　録音編集方式 …… 92
4.4.2　パラメータ合成方式 …… 92

5.　**日本語情報処理のための利用技術と標準化** …… 101
5.1　鍵盤配列と符号化規則 …… 101
5.2　情報交換用漢字符号 …… 103
5.3　日本語情報処理のためのＣＰＵ …… 105
5.4　漢字情報ファイルの編集 …… 106
5.5　配列プログラムについて …… 107

6.　**日本語情報処理の応用** …… 109
6.1　JICSTにおける科学技術文献速報編集システム …… 109
6.1.1　導入の経緯 …… 109
6.1.2　字種の選定 …… 110
6.1.3　漢字入力装置 …… 111
6.1.4　漢字出力装置 …… 112
6.1.5　抄録誌編集システム …… 113
6.1.6　今後の問題点 …… 118

6.2　工業技術情報に関する日本語情報処理の一例 …… 120
6.2.1　日本語による技術情報処理の背景 …… 120
6.2.2　システムの設計 …… 122
6.2.3　GIIPSYシステム構成概要 …… 123
6.2.4　漢字入力データ・ファイル管理システム …… 126
6.2.5　漢字出力データ・ファイル管理システム …… 131
6.2.6　システムの運営 …… 134
6.2.7　システムの評価 …… 135
6.2.8　将来のシステム展開 …… 138

6.3　出版社における自動編集システム……139
6.3.1　沿革……139
6.3.2　電算植字システムの開発……140
6.3.3　学研電算植字システムの概略……141
6.3.4　漢字入力装置……144
6.3.5　漢字出力装置……145
6.3.6　電算植字システムの利用状況……147
6.3.7　システムの評価……151
6.3.8　今後の方向……151

6.4　日本語によるオンライン情報検索システム……152
6.4.1　沿革……152
6.4.2　システム設計……152
6.4.3　システム概要……156
6.4.4　漢字入出力機器の概要……160
6.4.5　オンラインシステムの評価と問題点……162

6.5　物品管理システムの漢字出力の試み……163
6.5.1　沿革……163
6.5.2　システムの設計……163
6.5.3　システムの概要……170
6.5.4　運営状況……177
6.5.5　システムの評価……180
6.5.6　将来の構想……180

6.6　生命保険会社における日本語情報処理……182
6.6.1　沿革……182
6.6.2　システム設計……184
6.6.3　システムの概要……187
6.6.4　漢字入力装置……188
6.6.5　漢字出力装置……189

6.6.6　運営状況およびシステムの評価……190

6.7　電算写植センターにおける文字組版の事例……191
6.7.1　漢字情報処理における文字組版の特徴……191
6.7.2　日本船舶明細書の組版事例について……192

6.8　新聞業界における日本語情報処理……200
6.8.1　沿革……200
6.8.2　現在の新聞製作工程……201
6.8.3　システムの特徴……202
6.8.4　システムの概要……203
6.8.5　ＪＰＳによる新聞製作工程……207

6.9　日本語記録システム……211
6.9.1　日本語記録システムの必要性……211
6.9.2　システムの設計……212
6.9.3　システムの構成……214
6.9.4　漢字入力装置……215
6.9.5　漢字出力装置……220
6.9.6　運営状況とシステム評価……221
6.9.7　将来の構想……221

6.10　電子計算機による国語研究……222
6.10.1　沿革と概要……222
6.10.2　機器構成……222
6.10.3　用語用字調査……223
6.10.4　各種言語処理の研究……228

6.11　計算機によるカナ漢字変換の一手法……232
6.11.1　まえがき……232
6.11.2　入力文の形式……232

6.11.3　同音異字語処理……233
6.11.4　試作システムとその能力……234
6.11.5　今後の課題……238

6.12　日本語情報処理ユーザー覧表……240

7.　日本語情報処理の開発動向……257
7.1　情報化時代と日本語情報処理……257
7.1.1　データ通信と計算機ネットワークの時代……257
7.1.2　情報の各種形態……257
7.2　入力システム……259
7.2.1　日本語文字に対する符号の統一……260
7.2.2　けん盤入力……260
7.2.3　印刷文字の自動読取り……261
7.2.4　手書き文字の読取り……261
7.2.5　音声認識……262
7.3　出力システム……262
7.3.1　漢字プリンタ……263
7.3.2　文字ディスプレイ……263
7.3.3　フィルム出力……263
7.3.4　音声出力……264
7.4　処理システム……264
7.4.1　ファイル……264
7.4.2　言語分析……265
7.4.3　言語処理ソフトウェア……268
7.5　応用の将来……268

参考文献……271

1．日本語情報処理のニーズと社会的影響

1. 日本語情報処理のニーズと社会的影響

1.1 日本語情報処理の背景

1.1.1 日本語情報処理の概念（社会的動向）

1970年代初頭における情報化社会論の熱狂的なブームも過ぎ去り，落ち着いた状況のもとに情報化社会への移行は着々とその歩みを進めている。情報化社会の特徴は物的生産を中心とする第二次産業から，知的生産を中心とする第三次産業への移行であり，また脱工業化社会ともいわれている。

第二次産業はその大量生産性によって特色づけられるが，これを可能としたものは，技術的には物質の究明と，大量生産を可能とする安定した制御技術の発達である。そして人間はエネルギーの捕捉によって重労働から開放され，制御技術の獲得によって，手先の技能とカンから開放されるに至ったのである。この制御技術の高度化が，工業化社会から情報化社会へのかけ橋と考えられよう。これは別のことばでいえば，情報処理技術の高度化である。

制御技術を支えるものは，情報処理技術と通信技術である。これらの発達によって情報のフィードバックが迅速となり，またこれに対する応答も正確かつ迅速となり，情報化社会を特色づける多種少量生産も可能となってきた。これは消費者側からみて，選択の多様化または消費者主導型といわれているものである。

一方，計算機と通信技術の発達は，人間の社会活動の面にも大きな影響を及ぼしつつある。情報の迅速な伝達という面のみでなく，従来マスコミという形で情報を一方的に受けるのみであった立場から，自己の意思によって，積極的に情報を獲得するということが可能となりつつある。これは情報ユーティリティという形で実現されるものと期待されているが，歴史的にみると特定少数の情報源との双方向通信の時代から，不特定多数の情報源からの一方向通信の時代を経て，不特定多数の情報源との双方向通信の時代へと入りつつあるといえよう。

以上のように情報化社会は未だその揺らん期にあり，今後の発展の予測をすることは困難であるが，その一端をかいま見ることはできよう。ただここではっきりしていることは，計算機と通信技術との結びつきがその主役を演ずるということである。このことは将来これらの技術に対して，過酷なまでの多くの要求が出されることを意味している。本報告書で取り扱う日本語情報処理の技術もその一つである。

計算機による情報処理は10進数のみによる数値計算から始まり，その後，英文字（主として大文字のみ）を組み入れ，現在キャラクタセットとして英数字が定着している。これはハードウェアの発達によって取り扱いが可能となった面も見逃がせないが，最大の原因は英文字は文字字種が少なく，技術開発に多くを要しなかったためである。すでにタイプライタおよび PCS として，ある程度の処理が可能となっていたこともまた一因である。

しかし英文字情報の真の意味における情報処理となると話は別で現在までのところほとんど見るべ

き成果はあがっていない。この問題はいわゆる言語処理の問題であって，用いる文字に関係なく世界共通の課題である。またこの問題はかつてエネルギー，物質，コントロールとつぎつぎに手に入れてきた人間が，情報化社会において次に挑戦すべきメインテーマの一つである。したがって，この点に関しては，細部の小異はあるにしても，日本語も英欧語も同一のスタートラインにあると考えてよかろう。

現在の情報処理における英文字の役割りは，もっぱら人間と機械との情報上での，インターフェースの改善に置かれている。その最も端的な現われはプログラム言語である。単なる記号としての使用を除いて，リテラルにしろマンマシン応答における双互のメッセージにしろ，インターフェースとしての役割りに止まっている。

上述のような状況から現時点における日本語情報処理の概念としては，言語処理の問題を除いた範ちゅうが当面の目標となる。日本語が英欧語と異なる最大の点は文字字種が多いことである。字種が多いことはこれらを互に識別するために，文字のパターンが複雑となり，二重の負担となってくる。これらの問題を解決するには，入出力とも英数字とは比較にならない高度な技術開発を必要とし，これが日本語情報処理の開発を妨げていたといえよう。

日本語情報処理の適用分野としては，現在日本語が日常使用されている分野がすべて対象となるものであるが，当面，最も関心の高まっているのは印刷，印書に関連した分野である。これらの分野はマスコミュニケーションの手段として，早くから工業化が進んでいたが，工業的には第二次産業的部分，つまり機械化によって解決し得る部分のみに止まっていた。必然的に情報処理的部分，つまり文選，組版，植字の工程も対象として考えられていたが，日本語情報処理システムの未開発のために，今日まで実現し得なかったのである。

印書として最大の分野は速記である。出版産業，国会，裁判所などで行なわれている速記の量は，日々膨大な量にのぼっている。欧米と異なり高速の印書機械を持たないわが国としては，この印書の労力を計算機システムによって解決しようという試みは，ほとんど願望に近いほどである。

また最近，日本語情報処理の導入の機運が高まってきた分野としては，あて名印刷がある。これは日本語による高速印字という面とともに，顧客の管理を日本文字によって行なうという，管理上の問題が大きなウェイトをしめている。つまり情報をありのままの形でとらえて，情報処理の精度を向上させようという試みである。

このように日本語情報処理の分野は着々として拡大しつつあり，マン—マシン間のコミュニケーションを日本語で行なおうという方向をめざして，開発が進められている。以上の適用分野については後に詳しく述べることとする。

冒頭にも述べたように，わが国における情報化社会論も，そのマスコミ的存在の時代を過ぎて，やっと足が地に着いてきた感があるが，これらの情報化社会論の中で，日本語の問題を掘り下げて論じたものが見当らないのは，不思議なことである。一般的傾向として情報化社会という場合は，その背景にコンピュータ・ユーティリティを暗黙のうちに想定しており，これに対して知識産業という場合には人間の知的労働ということを念頭において論じている場合が多い。

しかし日本語を取り扱わないコンピュータは，今後コーティリティとしてはまったくその存在さえ危うく，また知識産業も産業として存続するためには，少なくとも情報処理システムを利用しなければならないであろう。そのため今後，言語情報処理のウェイトが高まると考えられるが，日本語情報処理技術はその基礎的技術として，欠くことのできないものなのである。

このように見てくると，日本語情報処理技術が発展し，日常的な情報処理の中に深く根をおろした時点こそ，現在の米国人が感じ取っていると同じような意味において，われわれもコンピュータを身近なものとして，感じ取るのではなかろうか。

1.1.2　日本語情報処理の機能と問題点（技術的動向）

日本語情報処理を論ずるには，次の五つの項目に大別して考察を進める必要がある。これら個々の項目については，次章以下で詳しく論じられるが，ここに簡単にまとめておくことにする。

(1)日本語の特質

日本語の特質はその語義と構文を別とすれば，その文字字種が多いことにある。英数字，かな，カナ，記号については，すでに標準化が行なわれているので問題は少なく，漢字にあらゆる問題が集中している。まず文字字種が多く，しかもその数が不定である。一応，当用漢字としてのわくはあるが，これは固有名詞と過去の漢字に対してなんらの拘束力も持たない。これらを含めての漢字の字種の限定と，それらのコードの標準化について，情報処理学会を中心として討論が進められているが，未だ結論は出るに至っていない。

日本語情報処理技術を推進するためには，まずこの字種を限定し，外字に対する処置を明確にすることが，第一条件であるといえよう。次に問題となっているのは，漢字を含む文字の配列順である。これはコードを定める上でも不可欠の要件であり，また配列順を一義的に定めることによる計算機側のメリットは，はかりしれないものがある。

この二点が明確になれば，日本語処理の問題点の大部分が解決したことになるのであるが，これは国語自身の問題であり，最後まで持ち越されるものと思われる。しかし一方，情報処理側からこの問題に対して，はっきりした意見を提出することもまた必要なことである。

(2)入力システム

英数字の場合には計算機が開発される以前から，すでにタイプライタや PCS が普及していたため，入力システムとしては，それら既成のシステムを流用すればよいという，好条件に恵まれていた。これらの技術は，当時すでにほぼ完成に近い域に達していたため，その後大きな変化はなく今日に及んでいる。これは計算機にとって非常な幸運であったといえよう。

英数字の場合の入力システムは，すべてフルキー方式をとっている。フルキー方式とは，必要とするキャラクタがすべてキーとして用意されている方式である。これは英数字のキャラクタセットの字種が少なく，技術的およびオペレーション上で大きな問題がなかったためであろう。このフルキー方式をそのまま，日本語情報処理の入力機器として実現させたものが，漢字けん盤さん孔機といわれるものである。

この種の機器は計算機による日本語情報処理が開始される以前に，すでに日本語情報のデータ伝送

用として開発され，多くの実績を持っていた。したがって計算機による日本語情報処理が始まると，その入力用として漢字けん盤さん孔機がいち早く採用され，今日に及んでいる。現在ほとんどのシステムが，これと同等のフルキー式の入力機器を使用している。フルキー方式はそれ自身最も自然な姿であり当然のことであるが，計算機の入力システムとしてみると，多くの問題点をかかえている。

一般的にいえることは，入力システムの必要条件としては，入力速度の早いことおよび入力ミスの少ないこと（またはミスの修正の容易なこと）があげられる。英数字の場合においては，フルキー方式は明らかにこの条件を満しており，さらにカード入力の場合にこれが顕著である。しかるに日本語の場合には，字種が多いという根本的な原因によって，この条件が失なわれている。つまり，多くの問題点がここに起因しているのである。

一方，言語の面からみると，一般に表音文字は字種が少なく書きやすく，表意文字は字種が多くパターンに訴える力が大きいために，読みやすいといわれている。したがって日本語の場合，書きやすい表音文字（たとえば，カタカナまたはローマ字）で入力し，読みやすい漢字かなまじり文によって出力すれば，理想的な言語処理システムが構成できるであろうという考えも成立する。この方向に沿った開発も着々進められており，すでに二，三発表されているシステムもある。

以上のように日本語情報処理をきっかけとして，入力システムに対する基本的問題点が明るみに出てきた感がある。人間と機械とのインターフェースという面から，この際これらの問題点を原点に戻って，落ち着いて十分検討する必要があろう。

(3)処理システム

日本語処理における計算機内部での処理は，いわゆる数値的な取り扱いはまったくないといってよかろう。つまりその大部分は配列，照合，転送などの，字単位また語単位の処理である。現在のところソフト的にクローズアップされているのは，主として植字のためのエディティングであるが，近い将来の問題としては，カナ—漢字変換にみられるような，ある程度の構文をふまえた語いの処理があろう。

しかし，漢字コードの標準化も行なわれておらず，日本語の正書法も確立されておらず，あいまいな点の多い日本文の処理を，計算機のみにまかせることは，コストの面からもまた正確さの面からも問題が多い。また植字における編集処理のように，グラフィックアーツの面からの要求もあり，これらの面を解決する一方法として，最近マン—マシン応答システムを取り入れた，ソフトウェアの開発が盛んに行なわれている。

簡単な定型的な処理は別として，語義的または構文的判断を要する場合では当面，人間が介入するシステムが先行することとなろう。日本アイ・ビー・エム社が発表した JPS (Japanese Publishing System) は，その一つの現われである。これは主として新聞の編集処理を対象として開発されているため，ソフト的に大規模であるが，その他の出版物ではより小規模なシステムで十分であろう。かかるシステムは計算機側の処理能力のレベルが向上するにつれ，人間側の負担が小さくなり，それにともなって応答システムも小規模ですむという関係にあり，将来は非常に簡略化された編集チェックシステムとして，普及することとなろう。

主として入力用に用いられるカナ—漢字変換のソフトも，当面の大きな開発目標となっているが，これには計算機以前の言語学的ソフトの開発が必要である。漢字—カナ変換（音声出力用など）も含めて，これらのいわゆる翻字システムは，機械翻訳などと同じく，研究体制の問題として取り上げる必要がありそうである。

(4)出力システム

出力システムは入力と比してよりハード的である。漢字プリンタなりディスプレイ装置の開発によって，目的を達せられる場合が多い。現在の動向としては，漢字プリンタは植字機とラインプリンタに分化し，前者はさらに高速用と低速用とに分化している。ディスプレイ装置はドット式の専用機の開発が進んでいるが，一方グラフィック・ディスプレイ装置を，必要のつど流用する方式も十分可能性がある。

ここで注意すべきことは，英数字のみによる情報処理の場合，わが国では植字機はまったく計算機とは無縁な存在であったことである。しかるに日本語情報処理技術の発展にともない，ラインプリンタと並行して，植字機の開発も進められ，このことがわが国における言語処理の発展に大きく貢献している。その最大の恩恵をこうむっているのが出版，印刷の分野である。

高速植字機は電子的に作動するため，その植字速度がきわめて早く，一方，植字量の関係から光学式の低速植字機も普及しつつある。これらの植字機をフルに活用するには，その前段階である入力，編集システムなどの採算性が問題となるところである。

漢字ラインプリンタは，文字をドットの集まりで表現する方式にほぼ固まりつつある。ただし実際にこれを印字するメカニズムについては各社各様で，一定しておらず，新たな方式が開発される可能性もある。これらプリンタは日本語処理用として，漢字，カナ，ひらがな以外にも英数字，記号などを包含しており，字種の面からみると，従来のラインプリンタの機能を包含するものであり，別な面からみると，国際万能プリンタとでもいうべき性能を持っている。この意味から漢字プリンタの開発は，プリンタ技術の向上に大きく貢献したものといえよう。

漢字プリンタ開発にともなうもう一つの恩恵は，プリンタの無音化が実現したことである。多字種によるメカの複雑さがノンインパクト方式へ指向し，その結果として無音化が実現した。このように多くの特徴を持つ漢字プリンタが，将来その量産によってコストが下がり，万能プリンタとして現在のラインプリンタにかわる時代がくれば，少なくとも出力に関する限り，字種の制約を念頭に入れずに，情報処理システムを設計できるようになろう。

漢字ディスプレイ装置は，わが国における情報ユーティリティの本命機種と考えられているものである。しかし，現在英数字用にみられるような大衆化を実現するには，コスト面での制約があまりにも大きすぎる。これを技術的に解決するには，未だ多くの時間を必要とすることであろう。

(5)アプリケーション

日本語情報処理の適用分野については，本章の最後で詳述するので，ここでは省略する。

1.2　日本語情報処理技術の発展過程

日本語情報処理に対するニーズ，その適用分野，またこの技術が内包している可能性などについて考察し，また将来を推測するためには，これまでに先人がたどってきた足跡をふり返ってみることも，また意義のあることである。ここでは情報処理ということばを拡大解釈して，印刷，印書の技術にまで踏み込んで，明治以後の技術の発展過程をたどってみることにする。

1.2.1　ローマ字文

ローマ字は16世紀のなかごろ，キリスト教とともに日本に入ってきた。そのころ出版されたローマ字の日本語本は，ポルトガル語のつづり方であったが，その後幾多の変遷を経て，慶応3年ヘボンが，「和英語林集成」を出版するに至ってつづり方も確立し，ヘボン式ローマ字として現在に至っている。その後明治18年に田中館愛橘博士らが日本式を主張し，両案の論争が始まったが，昭和29年の内閣告示で最後の決定が下され，日本式を部分的に改訂した訓令式が確認され，ヘボン式も特殊な除外例として認められている。

わが国において国字をローマ字と定め，すべての日本語をローマ字によって表現しようという主張を，ローマ字論という。この運動は早くも明治維新とともに始まり，明治18年には「羅馬字会」が創立され，この運動の最盛期には，会員数も1万人を越したといわれている。

ローマ字運動の当初の目的は，早急に文盲をなくそうということであった。しかしこれは義務教育の普及によってその理由が失なわれ，後には英文タイプライタによる能率化が主張の中心となっている。今日では日本文をローマ字によってつづろうという考えは，ほとんど影をひそめているが，皮肉なことに日本文字（カナも含め）を扱えない計算機処理の中では，最も幅をきかせている。

このローマ字運動の盛衰は，他山の石として多くの示唆を与えている。これを要約すると言語における能率化とは，書きやすさと読みやすさとの両者を合わせた，トータルシステムとしての能率化が重要なことで，単に一方の能率化だけでは受け入れられないことを示している。ローマ字化は日本語における固有名詞を，外国人に伝えるためには必要不可欠なものであるから，それなりの存在意義はあるが，それを日本文に適用しようとしたところに，飛躍があったとみるべきであろう。

ここでローマ字文についての二，三のデータを示しておく。いま平均的な漢字かなまじり文をローマ字化すると，その字数は約3倍弱となる。したがって入力速度が約3倍ないと同等とならない。英文タイプの技能検定試験における基準速度は次のとおりである。

A級　250字／分（50語）

B級　200字／分（40語）

C級　150字／分（30語）

D級　125字／分（25語）

和文タイプの1級で50字／分であるから，B級以上で英文タイプが上回り，ローマ字文のほうが入力速度が早いことは確かである。このことから，前に述べたカナ—漢字変換のかわりに，ローマ字—漢字変換というシステムも考えられよう。

1.2.2　カナ文字文

カナ文字運動とは，日本語をすべてカタカナで表現しようという運動で，前項のローマ字運動とは

ぼ同じ経過をたどっている。まず慶応2年，前島密の「漢字御廃止之儀」によって口火が切られ，明治16年「かなのくわい」として大同団結し，その後幾多の変遷を経て今日の「カナモジカイ」に至っている。

その目的とするところおよび経過は，ローマ字運動とほぼ同じであるが，一つの大きな相異は，今日においてもある程度の支持を得ていることである。とくに事務能率の向上という面から，全社的にカナモジを採用している企業がある。このようにカナ文字がローマ字よりやや受け入れられているのは，より読みやすいという一語につきよう。しかし総体的にみればカナ文字文もまたローマ字文と同じく，わが国に定着するには至らなかった。その理由は表音文字のみの文が，漢字かなまじり文と比べて読みやすさの点で格段に劣っているためと考えてよいであろう。

これらの点から判断して，今後，日本語情報処理を進めるにあたっては，入力は別として出力の点では，漢字かなまじり文をつらぬくことを大前提とすべきであるといえよう。

今日においてはカナ文字という場合，カナタイプライタが前提となっている。カナタイプは英文タイプの英文字を，カナ文字に置き換えたもので，キーの配列の問題以外とくに技術的な問題はない。カナタイプの考案は，大正11年，当時住友総本社の専務理事であった山下芳太郎氏が，アンダーウッド社に依頼して製作し，発売したのが初めで，その後レミントン，ロイヤルなどが続いて発売を始めた。

一方，英文字とカナ文字のいわゆるコンビネーションタイプは，昭和4年に東大病理学教室の緒方知三郎博士が，丸善と相談してロイヤル社に試作させたのが初めである。その後，各社からいろいろなキー配列のタイプが発売され，その種類は10を越えるに至っている。

これら英文タイプからの系列とはまったく別に，わが国における電信装置の開拓者である黒沢貞次郎氏は，明治34年にカナタイプを製作し，大正5年には電信用カナタイプを完成した。

カナタイプのタイプ速度は，だいたい英文タイプと同等とみてよいが，一方，漢字かなまじり文をカナ文字化すると，その字数は約2倍となる。したがって日本文をタイプする場合，ローマ字と比してカナ文字のほうが，はるかに有利であることがわかる。非常にすぐれたタイピストであれば，講演の速記も可能であると報告されている。

1.2.3　和文タイプライタ

和文タイプライタは，わが国独特のタイプとして開発され，現在あらゆる分野において印書機械として活躍している。さらにオフセット印刷の普及にともなって，植字機（清打）としての役割も大きい。和文タイプの構造は英文タイプとはまったく異なり，文字盤に収容されている活字を選択して取り出し，印字する方式をとっている。これら文字盤および活字の位置はまったくフレキシブルで，これにより字種に制限されないタイプとしての機能を発揮している。したがってこの意味から，和文タイプというより世界共通に通用する，万能タイプというべきである。

和文タイプの歴史は大正4年，杉本京太氏による発明に始まるが，その基本的な機構はそのまま今日まで受け継がれている。すなわち可動文字盤の中から所要の活字を選び出し，円筒形プラテンに巻き付けた用紙に印字するという機構は変わっていない。なお和文タイプの発明は英文タイプに遅れる

こと50年である。

今日，企業において，和文タイプを利用していないところはないといってよいほど普及しているが，それに反して個人の利用は皆無といってよいほどである。これには機械の価格が高いとか，手で書くより遅いと思われているとか，いろいろの理由があろうが，その最大の理由は，メーカが事務機としての面にとらわれ過ぎているためであろう。個人用として字種をさらに制限した，小型低廉な機種の開発に努力をそそぐべきであろう。

この個人用としての普及が遅れていることが，日本語情報処理の開発を遅れさせたとも考えられる。つまり日本語の機械的な取り扱いは，非常に非能率でコストがかかるという，先入観を植え付けてしまったといえよう。しかし手書きがその後の工程である印刷に，いかに大きな負担をかけているかは，例をひくまでもなく明らかなことである。

和文タイプは英文タイプにおけるタッチメソッドと異なり，文字盤を見ながら文字を選択していくため，印字速度ははるかに遅い。技能検定試験における基準速度は次のとおりである。

1級　５００字／10分

2級　４００字／10分

3級　３５０字／10分

4級　２５０字／10分

これを前掲の英文タイプの速度と比較すると１／５である。なお競技会における最高記録は，大体８００字／10分である。

欧米の例をみると，既存の英文タイプにそのままさん孔機能を付加して，計算機の入力装置として活用している。これに従えばわが国においては，和文タイプにさん孔機能を付加すればよいことになる。日本電気のＣ－5110漢字入力装置がこれに相当する。しかし和文タイプ自身が電動化されていないため，能率上の問題があり，また字種が多いためコード発生機構が複雑となり，価格が高くなるなど多くの問題をかかえている。とくに電動化の問題は，和文タイプ自身今後の開発課題として取り組まねばならないことである。このほか盤面の文字配列が機種によって違っている点も問題となる。

1.2.4　活版印刷

わが国における近代的活版印刷は，明治２年に本木昌造氏が長崎に活版伝習所を設立し，日本文字の活字を製作したのに始まるといわれている。本木氏の考案によって，号数活字が生まれ長く使用されたが，現在では各国共通のポイント活字が用いられている。

活版印刷において日本語情報処理と関係が深いのは，モノタイプ（鋳植機）である。欧文用の鋳植機は，欧米では早くから発達したが，わが国においては字種が多いため開発が遅れ，大正９年に至って初めて杉本京太氏（和文タイプの発明者）によって製作された。しかしこの機械はあまり利用されず，モノタイプの使用が本格化したのは，第二次世界大戦以後である。

モノタイプの機構は文字を選択する文字盤（あるいはキーボード）と，活字を鋳造する鋳植部とからなる。機械の構成としてはこれらが一体となったものと，キーボードが独立し，紙テープにさん孔し，このテープによって鋳植部をコントロールするものとがある。とくに後者は各機能が独立し，情

報処理システムとしての形態をとっている。このキーボードは，後に開発されたデータ通信用の，漢字けん盤さん孔機と機能的にはまったく同じである。

現在，稼動している漢字情報処理システムの多くは，印刷を目的とした自動編集・植字システムである。したがってその最終製品である印刷物は，多かれ少なかれ既存の印刷物と比較されまた批判される。今後自動編集・植字システムの技術を一層向上させるため，明治以来１００年間にわたる先人の築きあげたノーハウを，積極的に吸収するとともに，また計算機処理による新たな印刷技術の開拓へと進まねばならない。

1.2.5　写真植字機

ネガの文字盤から文字を選択し，レンズ系によって感光材料上に植字していく装置で，欧米においては１９１０年ごろより多くの考案が試みられたが，実用には至らなかった。その後１９５０年ごろより，ライノフィルム，フォトセッタ，モノフォト，フォトン，フィルモタイプなどが現われ，製版に盛んに利用されるようになった。

わが国では大正１３年に石井茂吉，森沢信夫両氏により，和文用の写植機が開発され，その後多くの改良が加えられて今日に至っている。その機構は和文タイプとよく似ており，可動文字盤上に配列されたネガ母型から文字を選択し，レンズ系を通して印画紙に植字する。レンズは約２０本あり，その選択によって２０種の大きさの文字が得られ，また変形レンズを用いて，長体，平体，斜体などの変形文字も得られる。

写真植字機はもともと，植字のみを目的として開発されたものであるが，オフセット印刷の発達によって，和文タイプも植字機として利用可能となり，現在では両者が競合している形である。それぞれ一長一短があるが，写植機の長所は文字の書体，サイズ，形などの点で，フレキシビリティが大きいことである。短所としてはモニタ印字がとれないこと，およびミス印字の修正がめんどうなことである。

この短所の克服および写植機の機構上の利点が結びついて，ここに自動写植機が生まれることとなり，これが日本語情報処理実現への第一歩となったのである。わが国では昭和３７年にＳＡＰＴＯＮ－Ｆ（数字専用），同４０年にＮ（新聞用），同４３年にＰ（ページ物用），同４４年にＡ（万能型）と，写真植字機研究所から各種の自動写植機が発表になった。とくにＳＡＰＴＯＮ－Ａは編集用としてＰＤＰ－８ミニコンを用い，完全に日本語情報処理システムの形態をとっている。ただ編集用ソフトが限られ，汎用の処理システムとなっていないだけである。

前項の出力システムのところでも述べたように，自動写植機にはこのほかに高速の全電子式写植機も開発されており，今後ますます高性能化へ進むものと思われる。なお欧米における状況も，わが国と同じような経過をたどりつつある。

1.2.6　漢字印刷電信装置

電信回線によるデータ伝送は，わが国においても明治維新とともに開始されたが，これらはすべて英数字およびカナ文字によるもので，漢字によるデータ伝送が始まったのは，第二次世界大戦以後のことである。漢字によるデータ伝送を行なうためには，送信側の入力用キーボードつまり漢字けん盤

さん孔機（一般に漢字テレタイプライタ）と，受信側の出力用プリンタつまり漢字印刷機（一般に漢字テレプリンタ）とが必要である。

これらは新聞社，通信社などでの必要性が高かったために，まずそこで開発が進められた。たとえば朝日新聞社では昭和２２年に研究が開始され，昭和３３年に試作機が発表されている。また読売新聞社と防衛庁は，共同開発になる漢字テレプリンタを昭和２９年に発表している。なお前にも述べたように，漢字けん盤さん孔機は，モノタイプのさん孔キーボードと機能的には同じである。これらの成果を取り入れて，新興製作所および沖電気工業が漢字印刷電信装置として市販し，幾多の改良を経て今日に及んでいる。

漢字によるデータ伝送を行なうにあたってなによりも重要なことは，送信側と受信側とで，同一の文字コード体系を用いねばならないということである。またこのコード体系は，モノタイプにとってその機構上有利な構成であることが望ましい。このため共同通信社とその加盟新聞社の代表は，昭和３４年６月に「漢テレ・全自動モノタイプ符号統一会議」を発足させ，約２か月，約５０回に及ぶ討論を重ねて同年８月にＣＯ－５９統一符号を定めた。

以上のような経過をたどって，漢字によるデータ伝送および全自動モノタイプ方式が，新聞社を中心として普及し，これが漢字情報処理をスムースに発足させた原動力となったのである。とくに漢字情報処理システムの入力機器としては，この漢字けん盤さん孔機，またはこれにある程度の改良をほどこした機器が中心となっている。なお漢字印刷機は全機械式であるために印字速度が遅く，その他多くの問題点があるため，出力機器としてはほとんど利用されていない。また漢字印刷機自身も最近では，感熱紙を用いたサーマルプリンタなどが発表されている。

1.3　日本語情報処理の適用分野

1.3.1　情報産業

情報産業ということばは，今日，一種の流行語となって，多くの人々によって論じられているが，ここではマハルハープのいう知識産業と同義としてとらえてみると，今後，計算機システムがあらゆる面において，この知識産業をバックアップするために，活用されるであろうことは確かである。その中でここでは日本語の言語情報を主とする産業に焦点を当てて，その適用分野について考察することにする。

マハループが知識産業に含まれるものとしてあげている産業の中から，これに該当すると思われるものをあげると，次の二つのグループに大別できる。

放送，印刷，出版

情報提供サービス

もちろんこれ以外の産業も日本語とは多かれ少なかれ関係するが，あえて焦点をしぼれば上の二つのグループになる。また上にあげた産業でも，こと情報の生産（創造的生産）に関しては当面，人間の仕事である。

この二つのグループは，その機能においてきわ立った対照を示している。すなわち前者は情報を大

量に複製して，一方的に流すマスコミ型であり，後者は情報需要に応じて，必要とする情報のみを提供するカスタム型である。また日本語情報処理システムは，前者においては，大量の情報媒体を生産するプロダクション型であり，後者は情報をコントロールする制御型を示している。

（1）印刷，出版，放送

日本語情報処理システムといわれるべきものが，初めて実用化されたのは，すでに述べたとおり印刷，出版業である。ここでは従来の印刷技術の技術革新という形で誕生した。この間の事情を活版印刷を例にとって考察してみよう。活版印刷の作業は組版と印刷の二つに大別できる。組版はさらに文選と植字の２工程に区分される。文選とは原稿に従って必要な文字を集める作業で，植字とは原稿に従って，句とう点，込め物などを入れて１行ずつに組み，段落を整えて組版を作る作業である。

モノタイプを用いる場合には，キーボードで必要な文字を打けんすることにより，その文字が鋳造され，活字に手を触れないで文選が行なえる。以上の工程を情報処理の立場から分析してみよう。まず従来の活版印刷では，常に「文」という情報と活字という印刷に必要な物理的物体とが，一体となって行動しているのが特徴である。したがって文選とか植字とかの情報処理的作業が，組版を媒体として行なわれ，システム的に非常にハードな形となっている。

したがって，たとえば校正において大幅な修正ができないとか，組版を直す場合に非常な手間がかかるとかの形で現われている。印刷術は金属活字の発明により，従来と比して非常にフレキシビリティに富んだシステムとなったが，実際には未だ不十分な点が多いのである。これを解決するには情報処理的な部分は，はっきり区分して処理システムにまかせ，その最終段階において，印刷に必要な形で植字をすればよいことがわかる。つまり従来の植字工程を編集と植字とに分け，文選・編集と植字とし，前者を情報処理システムとして組立て，後者をそのアウトプットとして得られるように，構成すればよいことがわかる。この構成は計算機の導入によって可能となり，必然的に日本語情報処理システムとなったのである。かくて文選は打けんに，組版はソフトウェアによる編集に，植字は出力装置としての植字機により自動化されることとなった。

以上のシステムは出力装置としてモノタイプを用いれば，その後の工程は従来どおりとなるわけであるが，上に述べた以外の多くの理由によって，出力装置としては写真植字機が用いられている。前述のようにこの装置には光学式の低速のものと，全電子式の高速のものとがあり，用途によって使い分けられている。このシステムを最近では CTS(Cold Type System または Computer Typesetting System）と呼んでいる。

このような考え方は，米国のフォトン社などから明確に打出されたが，わが国においては，写真植字機研究所（現　写研）が先導の役割を果した。写研ではＳＡＰＴＯＮシリーズとして，多くの機種を開発しているが，その採用状況の一端を次に示そう。

昭和４１年　日本社会党（社会新報）

昭和４２年　朝日新聞社北海道支社（地方版）

　　佐賀新聞社

昭和４３年　毎日新聞社（案内広告）

朝日新聞社（見出し用）

昭和４４年　日本海新聞社

昭和４６年　神奈川新聞社，上毛新聞社

西日本新聞社，北日本新聞社，高知新聞社

山梨日日新聞社　など

一方，高速の全電子式写植機も国産化が軌道にのり，日刊スポーツ新聞社，学習研究社，日本科学技術情報センターなどで，印刷，出版を主とした日本語情報処理システムが実動しつつある。

これらのシステムのうち，地方紙ではほぼ全紙面のＣＴＳ化が行なわれているが，中央紙では案内広告欄のような編集の比較的単純な紙面に限られている。中央紙の大組のように編集が複雑で，しかも短時間のうちに紙面の変更を行なう必要があるようなものでは，バッチ処理では不十分で，マン－マシン応答の可能なオンラインシステムが必要となってくる。日本アイ・ビー・エム社が発表した自動編集組版システムＪＰＳは，ＩＢＭ社と朝日新聞社および日本経済新聞社が共同で開発したシステムでこのソフトウェアに，朝日，日経の両社はそれぞれユーザ作成のソフトを追加して，ＮＥＬＳＯＮおよびＡＮＮＥＣＳシステムとして発表している。かかるシステムが一般的に実動を始めた時点が，本格的なＣＴＳ時代のはじまりといえよう。

このようにＣＴＳ化を目標とした日本語情報処理システムが，着々として実動を開始しているが，そのメリットとしては次の点があげられよう。

１）省力化，作業場スペースの節約

２）熟練工の不要

３）高熱，騒音からの解放

４）鉛公害の防止，職場環境の改善

つまり編集，植字に関しては，情報処理としてオフィス作業となるということである。

また編集面からもいろいろなメリットが考えられる。

１）複雑な編集を迅速に処理できる。

２）組版のフレキシビリティが大きく，改版が容易である。（電話帳のような索引においてとくに顕著である。）

３）情報の配列のような処理を編集システムに組み入れられる。

４）情報内容を他のシステムで活用することが可能となる。（出版しつつデータベースを構成できる。）

５）情報の保存，再版が容易である。

一方，デメリットとしては，情報内容が可視化されていないため，人間との間に常にインターフェースが必要で，また情報へのアクセスが，常にリモートコントロールとなることである。これは計算機処理の宿命で，将来はディスプレイ装置などで改善されるであろうが，当面はこのための思わぬトラブルが発生することが多い。とくに入力情報の校正処理にこのデメリットが現われている。

最後に放送関係では日本語情報処理システムは，報道の迅速化に役立つものと考えられている。と

くにテレビにおいて，漢字ディスプレイ装置の画面を直接聴視者へ送る方法は有効である。これはすでに選挙の開票速報においては実行されているが，その他の応用面は今後に残されている。

本項においてはとくに印刷関係に焦点を当てて詳しく述べたが，もともと印刷産業は情報産業の一種として，だれも認めるところであるが，わが国においては日本語情報処理の開発が遅れたため，これまで取り残されていた感がある。しかし今後は急速にシステム化が進み，執筆者がタイプライタで原稿を打てば，直ちに編集，植字されて印刷物ができ上るということも，そう遠い将来のことではなさそうである。

（2）情報提供サービス

情報提供サービスは今後の情報化社会における，花形産業として喧伝されているものであるが，その市場性に未だ難点があり，現在検討段階のものが多い。わが国において情報提供サービスを行なう場合，その内容が日本語情報であれば，ありのままの形で提供しなければ，受入れられないという可能性が強い。これは明治初年におけるローマ字運動やカナ文字運動の例をみても，考えられるところである。計算機が英数字しか扱えないということで，不特定多数のユーザを納得させられるとも思えない。

情報提供サービスにおいても，情報の創造的部分は当分人間の分野であるとすれば，ここにおける日本語情報処理は，それほど高度なものとも考えられず，情報の出力側つまり情報の伝送，印字，ディスプレイなどの面で，ユーザの要求に合った形の情報を，いかに早く，安く，しかもスマートに提供するかにかかっているといえよう。

このための当面の課題は端末装置である。これまで開発されている装置は，主として大量のバッチ処理を目標にしたものが多いが，今後は個人のユーザが使用する，簡便な装置の開発が要求されよう。英数字の分野では現在がちょうどその時期にあたっているが，ユーザの要求のエスカレートによって，将来必ずその時期が来るものと思われる。

現在情報提供サービスとして，どのようなものが予測されているかをみるために，(財)日本情報処理開発センターのＮＩＳ小委員会がまとめた，「情報処理産業の発展予測」から，その業種を次に示そう。

時事情報　天気予報　株価情報

所在情報　予約　レジャー情報

マクロ経済情報　ミクロ経済情報

海外情報　企業信用情報　消費者信用情報

特許情報　判例情報　医学情報

リサーチ情報　科学技術情報　一般知識情報

これらの業種はその内容が千差万別で，一概に論ずることは困難であるが，日本語情報処理の面からみると，大量の言語情報を蓄積したデータベースをもととして，その中からユーザの要求する情報を検索して提供するタイプの，いわゆるデータバンク的提供サービス業が，当面現われて来よう。これらの代表的なものは時事情報，特許情報，判例情報，科学技術情報などであるが，これらが実用化

される時期は昭和５０年代になると予測されている。

1.3.2　事務機械化

わが国におけるコンピュータの普及はめざましいものがあるが，この処理の特色は数値計算が主体で，ほとんど言語処理をともなわないことである。また取扱っている言語も「文」のレベルではなく，「語」のレベルに止まっている。また計算機システムもキャラクタセットして，英数字およびカナのみであるから，「語」はすべてローマ字またはカナ書きで処理している。

このような状況から，事務処理の分野における日本語情報処理の当面の課題は，「語」を本来の姿である日本語に置換えることに集約されよう。このことは結局，入出力装置の問題となる。まず出力装置としては文字に制約されない万能型プリンタを用意すればよい。この種の万能型プリンタの技術開発は現在盛んに行なわれているから，近い将来現在のラインプリンタよりも高性能な，万能型プリンタが出現する可能性が強い。次に入力装置としては，入力情報が大量の場合には，情報産業が実用化している方式でよく，また入力のみを計算センターに委託すればよい。また少量の場合には適宜，便宜的方法をとればよい。

以上の他に，日本語情報処理が多くの分野に取入れられるためには，各種技術の標準化が大きな問題となる。とくに字種とそのコードの標準化は緊急の課題であろう。現在事務処理分野における日本語処理としては，あて名印書を目的とした住所，氏名の漢字化，また人事管理における情報の漢字化など二，三の実施例が散見されるのみであるが，近い将来この他の分野にも急速に普及するものと思われる。また日本語情報処理が可能になることによって，これまで以上に事務機械化の適用分野が拡大することも考えられる。すでに見積書，契約書の自動作成システムを設計している例もある。

1.3.3　言語処理

言語処理は日本語のみとは限らないが，計算機処理の一つの到達目標である。それぞれ各国の特殊事情があり別個に研究が進められている。とくに日本語においては表音文字と表意文字を共用していることもあって，多くの興味ある課題が山積している。計算機を用いて日本語の言語処理をするためには，その基礎に日本語の入出力ならびに処理などの技術が確立していることが必要である。この言語処理は人間の知的労働を大幅に軽減する要素を秘めているもので，たとえば機械翻訳を想定すればこのことは明らかであろう。

現在開発が急がれているものに，カナ—漢字変換，漢字—カナ変換がある。前にも述べたように，表音文字は書きやすく，表意文字は読みやすいという性質を持っている。また音声はすべて表音的であるから，日本語では文章情報と音声情報とは，表意—表音の関係になっている。上述の変換システムは，これら両者の間の変換を自動的に行なおうとする試みである。

まずカナ—漢字変換の応用としては次のようなものが考えられる。カナは書きやすいという性質を利用して，カナ文字文で入力を行ない，このシステムによって変換して漢字かなまじり文とするもので，これをカナインプット方式と呼んでいる。入力速度が速く，装置が安く，タイピストも得られやすいなどの，メリットを持っている。

速記の自動化もこのシステムの応用となる。もともと音声は表音的であるから，速記の記録を見な

がらカナタイプを打つとか，直接聞きながらタイプを打つとか，ソクタイプで直接入力するとかして，このシステムで変換すればよい。さらに将来音声を認識して，このシステムによって漢字かなまじり文に変換することも考えられる。

漢字—カナ変換は日本語にふりがなづけをするとか，日本語情報をカナプリンタで出力するとか，さらには日本語を発声させるなどの応用分野を持っている。これまでは以上のようなシステムを開発するのに，日本語情報処理システムが未熟であったため，その開発を遅らせていた面があるが，今後はその技術と研究体制とが車の両輪となって，開発のテンポを早めることであろう。これらシステムの最終目標としては，機械翻訳と一体となって同時通訳をめざしている。

言語処理としては「文」の創造が目標であろうが，その中間段階としても，自動索引，自動抄録，自動編集など多くの実用性の高い研究テーマがあり，またこれらの基礎データを得るために，現存する日本語情報を分析，評価することもまた重要である。これらの研究に日本語情報処理システムは，今後大きな力を発揮することであろう。

1.4　まとめ

日本語情報処理はわが国の情報化社会の進展にともなって，必然的に生まれたシステムである。かつて電話がわが国へ上陸したときは，幸いなことに日本語の問題は生じなかったが，電信の場合には簡単ではなく，長い間人々はいわゆる「電報文」に悩まされたものである。戦後漢字印刷電信装置が実用化されるまでには，半世紀以上もの長い年月を必要とした。

言語の改革には独裁者と文盲が必要条件であるといわれているが，幸か不幸かわが国はこの条件に合わなかったとみえて，本来の日本語の地位はますます確固たるものがある。戦後米国を中心として計算機時代が現出し，わが国でもだいたい１９６１年ごろより計算機が社会の中に普及し始め，すでに１０年以上を経過した。わが国の計算機設置台数はその間に米国についで世界第二位となり，その応用分野も多岐にわたっている。

しかも約１０年にして，計算機と日本語とのドッキングが完成したわけであるが，これは技術的成果はさることながら，その社会的影響は測りしれないものがある。これまで計算機というと，ごく限られた人々にしか理解しえないものとして，一種の拒絶反応すらあったものである。情報化社会，ＭＩＳ，コンピュータユーティリティ，データバンクなどといっても，なにか遠い国の物語りという感じがぬぐえなかった。

しかし今日ではすでにそれらがすべて，日本語によって処理されるということで，ぐっと身近な感じを持つに至った人々も多いことであろう。一方計算機側からみてもこれまで概念的には把握されながら，日本語が扱えないために足ぶみしていたシステム，たとえばＣＡＩなども今後急速に実用化へ向うと思われる。これまで日本語が扱えないために生じていた一種の計算機公害からも解放され，人間本位のシステムがぞくぞく開発されるであろう。じゃ口をひねれば出てくる水道の水のように，情報が流れ出すシステムには，各国の民族に固有の不可欠な要素の解決が要求されるのである。

最後に日本語情報処理システムを広い立場からながめてみると，これは多字種処理システムという

こともでき，また日本語が処理できるということは，その字種数の多いことから世界中のどの国へ持っていっても通用することになり国際情報処理システムといい換えてもよかろう。この技術においてわが国が先端を歩んでいることは，将来必ず国際的に注目される時代がくることであろう。

2. 日本語の特質と言語情報処理

2.　日本語の特質と言語情報処理

言語情報処理の観点から現代日本語について考えてみようとするとき，まず問題にすべきは，その焦点となる時点をどこにおくべきかということである。というのは，言語自身きわめて複雑な構造をもち，言語処理の方法は長期間にわたる研究開発を必要とし，したがって今後時日の経過と共に次第にレベル・アップしていき，問題の焦点がうつってゆくと考えられるからである。言語の構造と機械処理技術からみて，はじめは文字の入出力の処理が中心的な話題になり，次第に文法や意味を問題にせざるを得なくなるのであろう。将来は文章や意味が中心課題になると考えられる。そこで，時点をまずだいたいきめてかかる必要がある。

この報告書ではだいたい今後５年間くらいのことを話題の中心とすることに定めている。そうすると，取り扱うべき問題は主として文字の入出力の部分に限られてくる。そこでこの「日本語の特質と言語情報処理」でも，文字のこと，とくに漢字のことを重点的に取り扱っていきたい。言語処理全体からすると，ほんらいシンタクスやセマンティクスこそかなりの紙数を投ずべき部分であるが，ここではそれについて多くふれないことにする。ただ，最近カナ入力，機械による漢字変換ということが広く考えられてきている（たとえば本報告書所収のＮＨＫ技研の例）ので，それに関係したごく狭い範囲でのシンタクスなどの事項について，ふれておこうと思う。

このような事情で，ここでは漢字にかかわることを中心に，広く日本語の特質についてまずとりあげ，次に漢字の問題を中心としてとりあげ，最後に，それに関連してこれから日本語の問題についての意見を紹介しておく。

2.1　日本語の特質と言語情報処理

2.1.1　日本語の特質

言語の特質について問題にするばあい，その要素である，音声，文字，語い，文法，表現などについて取り扱うのが便利であろう。そのあとで，これらの特質と日本語の情報処理との関係をまとめておくことにする。

(1)音声

音素の数え方にもよるが，ある見方によれば，日本語（現代東京語）の音素の数は子音 14，半母音２，母音５，拍的音素２のように数えられる。[注1)] すなわち，

子音(C)／ｈ，ˈ，ｋ，ｇ，ｐ，ｂ，ｔ，ｄ，ｃ，ｚ，ｓ，ｒ，ŋ，ｍ，ｎ／

半母音(S)／ｊ，ｗ／

母音(V)／ｉ，ｅ，ａ，ｏ，ｕ／

拍的音素／Ｎ，Ｑ／

拍（モーラ）の構造は，　CV, CSV, V, N, Q

の五種のタイプしかないのである。この点からみると日本語は，音素の数も少なく，拍の構造がきわ

めて簡単で，母音を多く含む言語であることがわかる。ちなみに，英語では，母音は8，子音は24ほどはあり，モーラの構造は CVC があることからきわめて種類が多くなる。日本語には英語の next／nekst／（cvccc），ロシヤ語の ВЗГЛЯД ／v z g l a d／（c c c c s v c）のような音節は存在しない。

日本語のアクセントは高低アクセントであって，強弱アクセントではない。このためにアクセントによって発音語形がゆがめられることがない。英語ではこの影響が非常に強く，アクセントのない音はほとんど発音されないことがあるほどである。同じ語で，強形，弱形をもつこともある。たとえば Have という動詞は，文中での重要度に応じて，〔 hæv,hov,əv,v 〕のようにいろいろに発音される。

このように音の数，その組みあわせ方，アクセントとの関係などからみると，日本語の発音は簡潔で明確な構造をもっているということができる。

(2)文字

これに対して，文字の点では，大きなハンディキャップをもっている。すなわち日本語を表記する文字はいくつかの種類があり，そのなかのあるもの（漢字）は多くの字種をもっている。そしてこれらを漢字かなまじり文という形で使用するのがふつうであり，漢字は使用法がきわめて複雑である。

まず文字の種類としては，漢字，ひらがな，カタカナを通常とし，ローマ字も用いる。数字はアラビア数字，漢数字，ローマ数字のいずれも用いる。通常の日本語の書き方はいわゆる漢字かなまじり文であって漢字とひらがなを用いるが，外来語についてはカタカナを用いるのがふつうであり，また最近では PCB, PPM, STOL などアルファベット書きの表記の固定した用語も著しい増加を見せているので，上述四種の字はすべて通常の表記文字として必要である。

漢字を中心としてもう一つの表記の問題点は，日本語の正書法が確定していないということである。正書法が確立しているヨーロッパの諸言語にあっては，文字の連続はほとんどのばあい，語に対応する。日本語では一つの語を漢字で書くこともあり，かなで書くこともあり，ふだん漢字やひらがなで書く語をカタカナで書いて目立たせるようなこともあり，その書き方は全体としては一定していない。何人かの人に共通な書き方がないばかりでなく，同一人でも必ずしも一定しない。これは文字でなくことばを処理しようとする機械内部の処理を複雑にする。

さらに，これらのいくつかの文字（漢字，かななど）を同時に用いることによって，用字法全体を複雑にしている。たとえば活用語を書き表わすのに，漢字とかなを用いると，どこまでを漢字で書き，どこからかなで書くかというような問題が生ずる。これが送りがなの問題で，たとえば「表わす」か「表す」かというような問題である。

漢字には単に字種が多いというだけでなく，その他にたとえばその用法が複雑であるという性質もある。このことは，かなの用法とも合わせて考えなければならないし，あとで述べるように，文法や意味とも関連がある。かなを漢字に直すばあいの困難点も，ここに原因があるといえよう。

かなづかいとしては現在は歴史的かなづかいが少なくなっているが，少なくとも一部の人はこれを使用している。現代かなづかいとあわせて大別して二様のかなづかいがあるわけである。ローマ字も

つづり方に三方式があって，いずれもよく用いられている。

また漢字かなまじり文では分かち書きの習慣がない。

かな専門書きやローマ字書きのばあいも，分かち書きにいくつかの方式があって必ずしも一定していない。分かち書きの点でも，日本語には問題が多い。[注2)]

漢字は俗に表意文字といわれるように，英字などに比して情報の量の多い文字であり，日本語と英語を比べて調査した結果でも，日本語の方が同一内容を表現するのに少ない字数ですんでいる，ということがわかっている。[注3)]

(3)文法

日本語は助詞助動詞のような接尾語的な要素が名詞動詞などの自立語のうしろについて文法的な諸機能を果たすという，いわゆる膠着語的な特質を有する言語であって，現代のインド，ヨーロッパ系の諸言語などとはかなり異なった構造をもっている（文法的には，ハンガリー語，トルコ語，モンゴル語，朝鮮語などと比較的よく似た構造をもつ)。すなわち名詞自身には格変化などがなく，形態的には比較的単純な構造をもつ。その文法関係は主として助詞の助けを借りて表わす。(たとえば主語になるときは助詞「が」を伴うのがふつうである)。動詞自身には活用と呼ばれる語形変化があり，(多くの）助動詞を伴うことがある。用言の「活用」は次にどのような語（自立語，付属語）が続くかによってきまる，といういみでヨーロッパ語とかなり違っている。文の構造では述語の前に主語があり，被修飾語の前に修飾語がくるという語順はかなり厳格に守られる。英語やフランス語のように文の主語が常に明示されず，文脈に依存して，示さなくてもわかるばあいは示さないのがふつうである。

(4)語い

日本語の語いも多くの特徴をもっているが，これらが現段階の情報処理と直接関係する点はかなり広範にわたり，文字と並んで重要な分野である。

	英　語	フランス語	スペイン語
最初の　1,000語で	80.5％	83.5％	81.0％
〃　5,000語で	93.5％	96.0％	92.5％

	日本語
最初の　1,200語で	62.0％
〃　7,200語で	85.0％
〃　15,400語で	93.0％
〃　22,000語で	96.0％

表2.1　上位何語でテキストの何パーセントをカバーするか

まず単語の種類がかなり多いということがよくいわれる。表2.1は英語，フランス語，スペイン語，日本語について，単語を使用率順に並べたばあい，上位何語でテキストの何パーセントをカバーするかを示したものである。これによれば，フランス語は少数の語でテキストの用語をカバーする率

が最も高く，日本語は他の三言語に比してきわめて低くなっている。このばあい，日本語の例としては国語研で行なった〈雑誌九十種の用語用字〉の数字をあげたのであるが，この調査では助詞助・動詞をカウントの対象にしていない。そういうわけで異なる言語の間で数字をあげての比較は困難な点があるが，もし助詞・助動詞を入れてもおそらく日本語の単語の種類の多さを立証することができると考えられる。このことにはいろいろな原因が考えられる。たとえば〔手〕を意味することばとして，フランス語では main あるいは結合語基として manu- （この二つはもちろん関連した語であり，語形にも類似性がある）があるだけであるが，日本語では「テ」（「手にもつ」「手作業」「手さげかばん」「手引き」），「シュ」（手術」「手芸」「運転手」），「ハンド」（「ハンドバッグ」「マジックハンド」），「マニュ」（マニキュア」「マニュアル」）のように類義語が多く存在するということも一つの原因であろう。

そして，そのことから，類義語の多いことが予想される。単に数が多いだけでなく，その使い分けもむずかしい。たとえば同じことを示すのに「火事」「火災」という二つのことばがある。「火事」は一般の人が会話のときに用いる，ふつうのことばである。近所に火災が起きたばあい，まず「火事」というのがふつうである。「火災発生」というのは消防署か警察であり，「火災」の方はややおもおもしいひびきをもち，書きことば的である。だから「火災保険」というのがふつうで，「火事保険」とはいわない。「小使いさん」と「用務員」なども同様である。「いいなずけ」と「フィアン

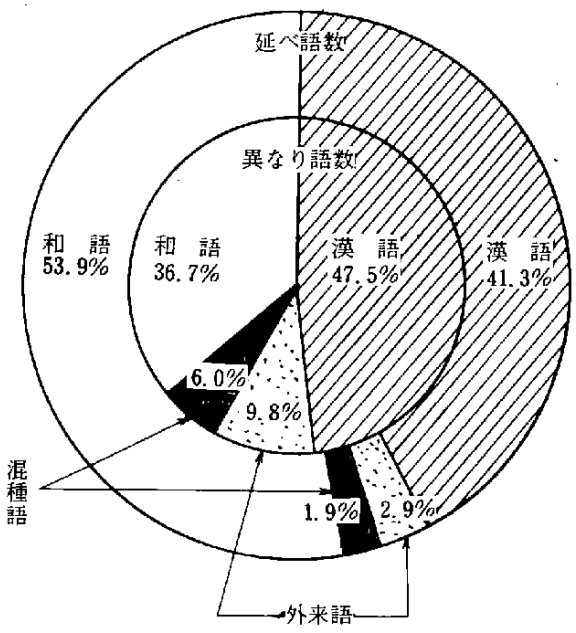


図2．1　語種・品詞別の異なり語数と延べ語数(雑誌90種)

セ」では後者の方が新しい感じがするといえよう。[注4)]

日本語の単語を和語，漢語，外来語と分けてみると，最も多いのは漢語である（図 2.1 参照）。

これは辞書的にみたばあい（異なり語数）と使用の実態（延べ語数）でみたばあいとで異なっているが，漢語が最も多くなるのは辞書的にみたばあいである。雑誌のなかで雑誌の種類別にみたばあい，婦人雑誌には漢語が最も少なく，実用，通俗科学に最も多くなっている（表 2.2 参照）。[注5)]

これは政治，経済，法律，自然科学，医学などの啓蒙誌であるが，これに漢語が多いのはそれらの分野のテクニカル・タームが多いからであろう。そしてそれは明治以降の日本が西欧文化を採り入れる際，漢語の訳語を作ったことに原因があろう。[注6)] 漢語も本来の和語に比べて，結局外来語的要素を多くもつものである。

一方，現代の日本語は多くの西欧語からの外来語を受け入れている。モードやスポーツの用語にも多いが，専門の分野にも多く，外来語までにいたらない，ほとんどなまのままの外来語も多く用いられている。現代の外来語を急速に受け入れているという点は，現代ではいわば世界的な傾向ということができるが，日本語はそのなかでも最も著しい傾向をもつ言語の一つであろう。

異なり語数

	語種				品詞			
	和語	漢語	外来語	混種語	名詞	動詞	形容詞	感動詞
評論・芸文	39.9	51.8	5.0	3.3	71.6	15.7	11.8	1.0
庶民	35.9	54.3	5.7	4.0	75.3	13.4	10.5	0.9
実用・通俗科学	28.8	60.3	7.0	3.9	79.0	10.9	9.3	0.7
生活・婦人	44.7	39.1	9.9	6.2	73.8	14.4	10.8	0.9
娯楽・趣味	41.3	45.7	8.3	4.7	73.1	14.1	11.7	1.1

（数字は横に 100 %）

延べ語数

	語種				品詞			
	和語	漢語	外来語	混種語	名詞	動詞	形容詞	感動詞
評論・芸文	58.9	40.0	1.5	1.6	56.0	27.1	15.1	1.8
庶民	55.1	41.2	1.9	1.8	59.8	25.3	13.1	1.8
実用・通俗科学	36.7	59.3	2.1	1.8	69.9	18.2	10.0	1.3
生活・婦人	56.3	35.3	5.7	2.5	65.0	22.3	11.1	1.6
娯楽・趣味	60.7	34.7	2.7	1.9	57.4	30.0	14.0	2.6

（数字は横に 100 %）

表 2.2　雑誌の種類と語種・品詞別異なり語数および延べ語数

漢語，外来語をあわせると，日本語のなかの外来的要素は相当なものになる。フランス語や中国語など外来語を受け入れることが少ない言語に比べて，日本語のなかで外来的要素の多いことは，日本語の一つの特質ということができる。外来的要素の多い言語としては英語，インドネシア語などがある。

外来語が現代で多く使われており，新しい外来語が多い，ということは，ことばの変化が早い，ということを意味する。たとえば「あいびき」ということばは明治時代には使われたが，その後すた

れ，大正昭和の時代になると，「ランデブー」ということばが使われた。しかしこのことばも生命が短かく戦後の若い人々は「デート」ということばをしきりに用いる。戦前は「オールド・ミス」といったのに戦後は「三十娘」と和語的表現に改まり，それもやがてすたれていまでは「ハイ・ミス」といっている。このようほ，現代ではことばのはやりすたり，入れ替えがまことにはなはだしく，かつ急テンポで行なわれるようになってきている。[注7] フランス語などに比べると，日本語の方が，使い捨ての用語が多いといえるだろう（前述の用語の多さとも関連）。

外来的要素は，和語などと異なった表記形式をもっている。外来語は特にかたかなを用いるのが普通であるという点で，和語漢語と異なっている。もちろん「ピカピカ」とか「ボク」のような表記はあるが例外的だろう。漢語は漢字で書かれることがほとんどである（「もちろん」「あいさつ」など当用漢字外の字があるばあい，最近ではかながきの語が出てきてはいる）。もっとも和語も，いわば訓読み式の漢字表記がかなりあるから，この方は外来語に比べて，区別が十分ではない。

漢語が多いということの影響の一つに，同音語が多いということがある。同音ではあるが，漢字で書くと異なり，もちろん意味用法も異なる，というばあいが多い。たとえばコウコウという音に対して，小さな辞書でも11の異なる語を載せている。「異なる」あるいは「別の」語であるかどうか問題なものもある（たとえば「基準」と「規準」，「追求」と「追及」と「追究」）。もっとも和語でも，異なるかどうか問題なものはある（「上げる」と「揚げる」と（「挙げる」）。

文字の上で分かち書きをしない，ということから，単語の認定が困難だという事情が生まれてくる。単語について分かち書きをしている現代ヨーロッパの諸言語に比べて，これは著しい特質である。「板の間」「茶の間」「くすの木」「はんの木」「あくる日」「仲よく」などは，どちらかといえば一語的であろう。さらに漢語などを長くつないで，「第一部業種別株価指数」などといったり「東証全銘柄最終値利回り表」「アポロ月着陸計画」などといったりする。

このような，単語の境界がはっきりせず，単語の結合がかなり自由である，ということは，あとで述べるようほ，ＩＲ（情報検索）におけるキーワードとして問題が生ずることになる。

(5)表現

文法よりもやや大きな見方で，日本語独特の表現あるいは発想法というような点から取り扱ってみる。

「演芸会は比較的寒い時に開かれた。年はようやく押し詰まってくる。人は二十日足らずの眼の先に春を控えた。市に生きるものは忙しからんとしている。越年の計〈はかりごと〉は貧者の頭〈こうべ〉に落ちた。演芸会はこのあいだにあって，すべてのどかなるものと，余裕あるものと，春と暮の差別を知らぬものとを迎えた。」（夏目漱石「三四郎」十二）

この文全体が日本語のふつうの表現からはずれているが，たとえば「越年の計は貧者の頭に落ちた」とか「演芸会は……ものとを迎えた」などにいちじるしい。「計」のような抽象名詞が主語になることは日本語ではほとんどない（「勤勉さが彼を成功させた」のような表現は日本語らしくない）。また「演芸会が人を迎える」という言い方よりも，人間を主体にした表現の方がふつうである（「多勢の人が演芸会に集まった」など）。

天候，自然現象を表現するとき，日本語で自然現象そのものを示す語が主語となり，それに対応する自動詞が述語となる。「雨が降る」「雪が降る」「雷が鳴る」「氷がはる」など。これに対して英語では it rains などという。ドイツ語では es を用いて， es regnet, es schneit, es donnert, es friert などというのがふつうである。たとえばこのような点も言語の間の著しい発想の相違がみられ，具体的で人間を中心とした表現を好む日本語の特性がみられよう。

2.1.2 日本語の特質と言語情報処理上の問題点

以上，音声，文字，文法，語い，表現の各分野についてひとわたり日本語の特質についてふれてみた。述べるべき点は非常に多いが簡略に言及することを目的として，いくつかの例を示したにすぎないところが多い。

次にこれらの特質のうち，言語情報処理と特に関係の深い分野について取りあげ，その問題点をあげ，解決の方向を考えてみることにする。

その前に一つだけ問題にしておかなければならないことがある。日本語を扱うということは言語を扱うということの一部であるけれども，日本語に限定せず，およそ自然言語処理であるかぎり，他のコンピュータ処理（たとえば事務計算，科学計算）と基本的に異なっているところがあるはずである。この点は日本語の特質という以前に，言語処理の段階で問題にすべきものである。

そのことでもいくつかの問題点があるが，そのなかでどうしても取りあげなければならないものとして意味の問題がある。言語は意味をもつものであって，形式だけの処理では，ほんらい言語の取り扱いとはいいがたい，という点である。科学計算でも事務計算でもデータの数値そのものはそのひとつひとつが何かを指しているということはないが，言語データとしては，語形だけよりも，その語形によって指されるなにものか，が重要である。英語でＤＯＧというと，ＤとＯとＧが単にその順に並んでいる，というだけでなく，それがある動物を指し，しかもこの種の単語の一性質としてｓがついていないということが，「言語」としては最も重要な点である。このように語形（文字に表記された形，ないしはビット構成）によって指し示されるなにものか，が重要であり（ソシュールの言語学でいう signifié), いわば言語学の基本原理である。[注8)]

ところが，このような部分を，計算可能な形で具体化，あるいは形式化することは，ひじょうに困難である。（言語学自身も，意味については，どちらかといえばいままであまりふれてこなかった。）そこでいままでのコンピュータによる言語処理はこの点については主として別な角度から，いわば近似的に取り扱ってきた。また実用上の見地から，必ずしも意味をもったものとしての取り扱いを必要としないような分野もあった。

文章を，意味をもつものとしてでなく，単に文字の連続として処理するというような処理にあっては，このような配慮は必要ではない。例をあげると電算植字のようなばあいがそれであって，このばあいには，コンピュータによる処理の本質は数字の連続を印字，印刷するのと本質的には相違はない。

ただ文字の数が漢字を含むのでかなり多いという点，あるいは文字の位置に若干の制限があるなどの点において異なるのみである。

次に文字の連続を単語の段階でとらえるという応用分野がある。例をあげればいわゆるキーワード

による情報検索システムである。この分野で意味をとり入れているのは，いわゆるシソーラスであって，キーワードシステムによる文献検索はまさに典型的に，意味を単語の段階で処理するものである。そうしてこの段階までの言語処理は実用化がすでに行なわれている。

単語以上の段階での意味を取り扱うものとしては，機械翻訳，やや進んだ段階における事実検索，やや進んだ段階における文字変換（たとえばかな漢字変換など）などがある。言語がコミュニケーションを行なう姿として最もふつうな存在は，文章であるが，このような，文あるいはその順序づけられた集合としての文章の意味を扱うのはむずかしく，それ故にまだ言語のコンピュータ処理として実用化されていないわけである。いわば研究中のものでありいわゆるコンピュータ言語学でも，この意味の点，あるいは意味と文法の交差する点が，目下の最も中心的な研究課題として研究されている。[注9]

さて，以上のことを前提として，日本語の諸特質と言語処理上の問題点について考えてみたい。

音声の特徴についての問題で取りあげるべきは音声の入力であろう。現状では音声認識が，アルゴリズム的にも問題が多いようであるから，近い将来の実用ということは考えられないだろう。しかしそのような条件がととのったばあい，日本語は実用化が容易で問題点が少ないということができよう。[注10]

文字の点についていえば，文字の種類が多く，そのなかに漢字を含み，その漢字の字種が豊富であることがとくに入出力機器の構造に大きな負担をかけることになる。この点については「2.2 漢字」でくわしく述べる。

日本語の正書法が確立していず，語の表記法がまちまちである，という点は，言語情報処理の各分野に大きな影響を与えるのであるが，たとえば漢字処理をふくむばあいシソーラスの構成なども，このことを考慮に入れなければならないだろう。もし正書法が確立していれば，この点での負担がかなり軽減されることは事実である。漢字を使うか使わないか（「モト」と「元」など），送りがな（「起こり」と「起り」），外来語（「インディケータ」と「インジケーター」），人名地名（「ハマーショルド」と「ハムルスキョール」）など，機械のなかで完全に別な表現になってしまうので，効率を著しく悪くするものであろう。（もっとも，このなかには，正書法とひとくちにいえないものもまじっている。）

文字の部分で漢字かな混用についてのべ，語いの部分で，漢語が主として漢字で書き，和語は漢字でも書くが，どちらかといえばかなで書くことについて述べた。これにつき漢字とかなの分布あるいは連続確率を利用して，分かち書きされていない日本語文の分かち書きをする（単語を recognize する）こと，あるいは文字変換（漢字をかなに改めるなど）をするというアルゴリズムを考えたプログラムがあり，すでにいくつかできている。このようなプログラムも，もし正書法が確定していれば効率のよいものになるであろう。表記がまちまちであることが，このプログラムの処理結果をその分だけ不安定にしてしまう。漢字の用法が複雑であることも，この種のプログラムの障害になっている。（たとえば「方」をホウと読むか，カタと読むかなど）。

ただ，日本語の表記では，字種が多い，ということは必ずしもすべて欠点の方にばかり作用するとは限らない。語いのところで述べたように，漢語は漢字で書くことが多く，和語でも自立語は漢字で

書くことが，どちらかといえば多い。外来語はカタカナで書き，なまの外国語は英字で書くことが多い，そしてこれらの用語は，分野によってはテクニカル・タームであることが少なくない。ということは文字の字種（カナ，漢字）を検査することによって，単語の性質の大まかなふり分けが，ある程度可能であるということを意味する。上に述べたような，単語の recognition に文字の連続を利用するプログラムの成立する根拠は，この点にあるといえよう。したがってこのような事実をうまく使えば，ある程度の近似的な処理が可能になってくる。

文法の問題に移って，上述のように，現段階の実用面で，文法が問題になっていることは比較的少ないが，たとえば主語を表現しないという日本語的な表現法は，言語処理を著しく困難にすることがありうる。主語補充の自動化を考えなければならないだろう。[注11] 日本語の用言の活用の性質を利用し，自立語＋付属語などの性質を利用して同形語の判別をある程度することもできよう（ＮＨＫ技研の例）。

語いについては文字の段階で関連していくつかふれたが，単語の多いこと，類義語，関連語が多いことは，言語処理の上でひとつの問題をひきおこすことであろう。英語の文献でなら information，ひとつですむところを，日本語の文献中では「情報」「インフォーメーション」「information」の三つの形で現われる可能性がある。シソーラスの内容が増加してしまう。

このことに関連して，単語の流行のはげしいこと，単語がはやく興隆，交替，衰亡すること，がやはり問題になるであろう。たとえば，ひところの「電子計算機」に比べて現在では「コンピュータ」が著しく多くなってきている。

また，科学技術の研究開発が速い速度で行なわれる結果，知識やものそれじたいの変化がはげしい。それに対してつぎつぎと新しいたくさんのことばを用意していくという傾向が日本語にはある。

このようなことはたとえばシソーラスのメンテスンスなどにも影響を及ぼすことになるのではないかと思う。

日本語の文が分かち書きされていないことは，結果として日本語の単語の長さの実際的な取り扱いを不安定にしてしまう。このことは情報検索で分かち書きの問題としてしばしば問題にされてきている。分かち書きの長さが，そろわなければ，キーワードのマッチングができないからである。さしずめ，キーワード・マッチングによるＩＲを設計するとすれば，シソーラスをふやすか，分かち書きの原則を定めるかしなければならない。いずれの操作も，必ずしも容易でないだろう。

語いの分野で，同音語が多いことについて述べた。これは現段階の言語処理として，きわめて重要である。それは一つには，カナやローマ字では処理しきれないことがあること（たとえば会計学で，「減価」と「原価」が同音になるが，これはいずれも意味の違う重要な用語である），その結果結局漢字を使用することになる（生命保険などでの漢字処理にはこのことも原因の一つになっていよう）ということがあり，もう一つには，漢字入力が未解決の問題として存するところから，カナ入力，漢字変換が考えられているが，[注12] その中心となる問題点が，同音語の処理ということである。これは結局，漢字の問題に結びついているので，ここでややくわしく取り扱ってみたい。

この問題を解決するためには，まず同音語の実態を調べなければならない。同音をもつ造語基とし

ての漢字はたとえばコウとかショウとかが考えられるが，これらをもった語形を単に辞書のなかで求めるだけでなく実際に使われている，それぞれの分野での実態を調べてみることが大切である。

ここではその例の一端を示すことにする。国立国語研究所で行なった新聞用語調査の，カナモジによるラインプリンタ出力のＫＷＩＣはこの目的にこたえてくれるので，この全体量300万のうち$\frac{1}{20}$の15万語分について調べてみた結果を利用した。まずショウコウという例をみると全体で5例あるが，「将校」(3)，「商工」(1)，「小康」(1)という分布で，「昇降」「商港」はなかった。全体で5例というのは少ないので，このうちわけをそのまま拡大して全体を推定するわけにはいかないが，その用例をみると，「将校」はすべて「青年将校」で3例，同一のサンプリングブロックのものであり，「商工」は「名古屋商工会議所」，「小康」は「一応の小康状態をたもち」のような使用例であった。青年将校は小説の例で特殊であるが，それを入れても，その使用用例まで考えに入れると，すべてきわめてありうべき用法である。この判別のためには使用法を検討することがまず第一ではないかと考えられる。

次に「ジンケン」の例を調べてみると全26例のうち「人件」(2)，「人絹」(24)で，後者のうち20例は「人絹糸」であった。新聞における経済欄の用例が多い。このことは，この数字に新聞というマスコミのメディアがもつ特殊性が現われたものといってよいだろう。分野によって同音語の分布が変わるものと思われる。ちなみに「人件」(4)はすべて「人件費」という形で使用されていた。新聞が一つの特殊な分野であることは，「ジュウコウ」3例のうち，「重厚」「銃口」は0ですべて「重工」(3)であったことでも判明する。

「ジカン」68例では「時間」(64)，「次官」(4)で，後者の用例は「外務次官クラス」「外務次官」「大蔵省の次官折衝」「竹内次官」であった。「時間」の方は「数字＋時間」が16例，あとは「時間がかかる」「時間を短縮」「時間を消費」あるいは「教師の時間」あるいは「時間短縮」などの形，すなわち「Ｎ（名詞）＋ヲ＋Ｖ（動詞）」「Ｎ＋ニ＋Ｖ」「Ｎ＋ノ＋Ｎ」「Ｎ＋Ｎ」のような形で用いられていることに注目すべきだろう。

「コウコウ」の例をあげると，前述のように小型辞書でも11語あり，「口腔」「硬膏」「膏肓」「抗口」「鉱坑」「香香」「高工」「高校」「航行」「港口」「煌煌」と並んでいるが新聞の例では81あり「港口」(1)，「高校」実に80であった。この結果は「望星高校の時間」などラジオ番組欄などにまとまって出てくる点，やはり新聞の特色が出ているものであろう。「港口」の例は「これをむかえるため横浜港ではコウコウ付近の航路を深くしようと浚渫船がけんめいに……」のようである。「高校」は「高校講座」「高校教育」「高校卒業」「高校三年」「高校生」あるいは「高校へすすむ」「高校にいけない人」「高校に赴任」などの文脈でよく使われている。ことばの単位の長さがちがうと別な同形語が生まれる。「ミタ」に「三田」と「見た」がありうるし，「キタ」に「北」と「来た」がありうる。

以上のことから同音語判別のストラテジーを考えてみると，プログラムのシステム，アナリシスのばあい人間の判断にまかせる（ディスプレイなどを利用した人間との交信による）ことがまず考えられるが，もし計算機による処理をある程度考えてゆくとするならば，アプリケーションの行なわれる

分野での同音語の実態調査をまず行ない，その結果プログラム化の容易なもの（その実態はまだあまりよくわかっていない）から計算機処理を行なえばよい。もしかなり全面的に計算機処理を考えるとすれば，たとえば「する」がつくかどうかとか，接頭接尾語がつくかどうかというだけでなく（いくつかのプログラムで使用されてはいるがそれらがそれほど多くの範囲をカバーするものでないことは上に示した例でもわかる），文脈それ自体を全面的に調べる，という方向にむかうべきである。これらはたとえば「N＋ガ＋V」「N＋ヲ＋V」「N＋ニ＋V」「N＋ノ＋N」「N＋N」のような基本的なパターンとそれに属する名詞や動詞の分類という方向に向かうことになるであろう。[注13)] そのことは任意に選んだ上の同音語の実態から考えられることである。

ただ問題によっては文の総合的な解析を必要とするものもでてこようし，さらに一つの文にとどまらず文章（discours）にまで情報を求めることもでてこよう。実用的なあるいは教育的見地からの機械と人のやりとりを，どこの段階でも考えることが必要で，いたずらに100％の解決をのみ目指すことは有効でなかろう。

ヨーロッパ語でも計算機で言語を処理するばあいこの同形語の判別が問題になる。アメリカのある用語調査（WORD FREQUENCY の調査）の例では，CAN（できる）と CAN（缶），LEFT（左）と LEFT(LEAVE の過去）などを区別しないで処理してしまった。そのような処理は "beyond the reach of computer technology" と考えたのである。[注14)] ドイツ語でも SEIN（彼の）と SEIN（ある），LIEBE FRAU と ICH LIEBE と DIE LIEBE のように同形語，多品詞語が多数存在する。ドイツのザールブリュッケンの研究者たちはドイツ文解析自動処理を考え，7人のアナリストと十数年の時間を費して "die Erkennung mehrteiliger Formen und die Auflösung syntaktischer Homographer, sind praktisch bereits gelöst" というまでに到達した。[注15)]

この結果が現在の機械システムですぐ具体化できるかどうかはともかくとして（ドイツの例はともかく動いている），この研究を複雑で時間がかかるという理由で放置せず，実際に進めてゆく必要性を感じさせるものである。

最後に，表現についてふれるが，これがとくに問題になるのは機械翻訳のばあいである。しかし紙数の関係で，ここではふれない。

以上，日本語の特質と言語処理の関係を考え，いろいろの問題点を指摘してきた。ここでいっておきたいことは，日本語の特質を考えるときは音声，文字，語い，文法，表現などと関連はありながらも別々にみることが可能であったが，言語処理との関連とその問題点およびその解決を考えるにあたって，これらをばらばらのものとみることが有効でないことが考えられたことである。文字を扱っても，単に文字として扱えるのはごく一部の問題あるいは分野であって，多くのアプリケーションにおいて，それは用語の問題と関連してくる。そしてたとえばカナ入力漢字変換を考えるばあいには，文法の問題を見すごすわけにはいかなくなってくることである。言語はその要素がばらばらに無関係に存在しているのでなく，互いに緊密な関係をもって存在するものであるから，当然のことである。

2.2　漢字

以上，日本語の特質と言語処理の問題点とについて略叙したが，そのなかで，現時点として最も多くの問題を含むものは漢字である，としてこれをまとめて述べることにしてあった。ここではそれを取り扱いたい。取り扱う内容は「2.2.1 字種」，「2.2.2 書体」，「2.2.3 字形」，「2.2.4 機能」，「2.2.5 配列」，「2.2.6 音訓」である。

2.2.1　字種

康熙字典は約5万字を収めているといわれるが，その大部分は中国の昔の文献に現われる特殊な字であって，現代の日本語で実際に使用している文字は大体4けたの数字におさまるであろう。現在までに行なわれた，漢字の使用実態調査としては，次のようなものがある。

大西雅雄「日本基本漢字」（昭和16）

朝日新聞調査研究室「新聞文章実態調査」（昭和24）

毎日新聞大阪本社印刷局「本社使用活字使用数調査」（昭和28）

国立国語研究所「婦人雑誌の用語」（昭和28）

同「総合雑誌の用字」（昭和35）

同「現代雑誌九十種の用語用字」(2)（昭和38）

同「現代新聞の漢字調査」（中間報告）（昭和46）

このほか結果が単行書の形で公刊されていないが，科学技術情報センターの科学技術文献で速報作成自動化のための調査[注16]，沖電気・毎日新聞調査，その他の文献がある。

各調査で得られた延べ字数と異なり字数を表の形にまとめると表2.3のようになる。さらにほかのデータについて言及すると，地名に関しては約4500の文字があれば足りるという。

これらの調査結果で，全体の字数も問題の一つであり，情報処理学会の漢字コード委員会では約6000の字種を試案として選定されたそうであるが，どの文字がそこに選ばれるか，がこの問題になろう。それと関連して，各調査の結果の字数がふぞろいなのも一つの問題をふくむと考えられる。

	延べ字数	異なり字数
婦人雑誌	17 万	3048
総合雑誌	11 〃	2781
現代雑誌90種	28 〃	3328
毎日新聞調査	32.2 〃	2613
朝日新聞調査	1600 〃	2308
科学技術文献 (JICST)	99,771	1378

表2.3　各調査の延べ字数と異なり字数

その説明のために，国語研究所で行なった雑誌と新聞の例を比較してみると，新聞の方が少なくなっている。これは一つには，調査対象の時点の相違も考えられるが，もう一つには，新聞社じたいで新聞記事の用語用字に制限を加えている事実が見のがせないだろう。

しかしもう一つの問題点も含まれている。それは同じ「現代雑誌九十種の用語用字」の調査のなかでも，分野によって使用漢字数が異なる点である。

「現代雑誌九十種の用語用字の調査」では，昭和31年の雑誌を5種に分けて調べているが，その種類によって漢字の使用状況が異なるのである。五種類というのは①「世界」「中央公論」「新潮」など，総合芸文誌グループ，②「週刊朝日」「サンデー毎日」「文芸春秋」など一般家庭雑誌のグループ，③「科学読売」「実業の日本」「時の法令」など実用通俗科学といったような半専門雑誌グループ，ただし完全な学会誌は不採，④「主婦の友」「婦人公論」などの婦人雑誌，⑤「映画の友」「オール読物」「旅」「囲碁」などの趣味，娯楽の雑誌グループである。その漢字使用の結果は表2.4のとおりであるが，ここでまず注目されるのは③の実用通俗科学のグループである。この類の雑誌は一般によく漢字を使用していて，漢字含有率からいうと実はこれが最も多い。しかしそこに使用されている漢字はほとんどが当用漢字であって，当用漢字以外の漢字を用いる率は五種類中最も少ない。婦人雑誌は漢字の延べ数はそれほど多くないが，表外字は三層より多い。一層と五層はとくに表外字が多いが，これは小説などに出てくる文字が多いことに起因しているようである。これは分野によって漢字使用の実態が必ずしも同一でないことを示すものであり，漢字だけでなく，どんな用語が存在して，それが漢字使用を要請しているかについても考えてみる必要がある。

	異なり字数			延べ字数		
	全体	当用漢字	表外字	全体	当用漢字	表外字
評論・芸文	2,286 (100%)	1,624 (70.9%)	663 (29.0%)	29,715 (100%)	28,148 (94.7%)	1,567 (5.3%)
庶民	2,421 (100%)	1,739 (72.2%)	673 (27.9%)	50,910 (100%)	49,049 (96.3%)	1,861 (3.7%)
実用・通俗科学	1,980 (100%)	1,629 (82.3%)	351 (17.7%)	57,911 (100%)	57,081 (98.6%)	830 (1.4%)
生活・婦人	2,199 (100%)	1,626 (74.0%)	673 (26.0%)	47,813 (100%)	44,767 (93.6%)	3,046 (6.4%)
娯楽・趣味	2,761 (100%)	1,755 (63.6%)	1,006 (36.4%)	93,745 (100%)	88,525 (94.4%)	5,220 (5.6%)
全体	3,328 (100%)	1,835 (55.1%)	1,493 (44.8%)	280,094 (100%)	267,570 (95.5%)	12,524 (4.5%)

注）斉賀秀夫「現代における漢字・漢語の分析」文学語学41号による

表2.4　雑誌90種の部門別に見た当用漢字・表外字の使用状況

国語研究所で行なっている新聞用語調査でも，分野による相違が見られる。たとえば新聞にのせられる広告を「案内広告」と商業広告に分けてみると，

	異なり字数	延べ字数
商業広告	1779	26768
案内広告	1554	48353

のようになり，案内広告は漢字の延べ字数は多いが字種の異なりでは商業広告の方が多いという結果になっている。[注17] これは案内広告がいわゆる求人や土地家屋などの，かなりきまりきった型や表現をもつのに対して，商業広告ではそのようなきまった型が少なく，バラエティのある表現や用語が存在していることによるのであろう　表2.5は新聞各欄のなかでの上位九字までの使用度数の累計をパーセントで示したものである。[注18] これでみると次のようなことがわかる。

1）スポーツ欄，経済欄，案内広告などは100字で50～60％に達するが，商業広告や文化欄は，40％前後である。

2）案内広告と婦人欄を除いた他の区分では，使用度数の多い漢字の上位から40～50パーセントで，90％の累計に達している。

3）案内広告では496字（異なり字の32％）で全体の90％に達するが，100％に達するのは1554字である。すなわち比較的少ない字数で全体をおおうことができるが，字種とするとかなり少なく，度数の少ない字が多い。

4）婦人欄では，全体の90％に達するのに568字（5割強）を必要とするが，100％になるには1007字である。このことは，婦人欄は，ある程度の使用率に達するためには比較的多くの字数

	政治・労働	外交	経済	社会・地方	国際	文化	スポーツ	婦人・家庭	芸能・娯楽	小説・漫画	商業広告	案内広告
最初の 10字	11.9％	14.9％	15.1％	11.1％	13.2％	10.1％	16.8％	11.0％	10.8％	20.6％	9.0％	11.3％
20字	18.7	22.9	22.6	17.4	20.2	16.5	25.6	17.6	17.6	29.9	15.0	18.8
30字	24.3	29.0	28.5	22.4	25.7	21.3	31.9	23.2	23.2	37.0	19.7	25.1
40字	28.9	33.9	33.3	26.6	30.7	25.1	36.9	28.1	27.7	43.3	23.5	30.6
50字	32.9	38.2	37.5	30.4	35.0	28.5	41.2	32.0	31.4	48.5	26.7	35.4
60字	36.5	42.2	41.1	33.7	38.6	31.5	44.9	35.0	34.7	53.1	29.6	39.6
70字	39.6	45.7	44.3	36.6	41.9	34.3	48.2	38.7	37.6	57.0	32.4	43.3
80字	42.7	48.9	47.2	39.3	45.0	36.8	51.0	41.5	40.3	60.9	34.9	46.6
90字	45.4	51.8	49.9	41.7	47.7	39.2	53.5	44.0	42.8	64.0	37.3	49.6
100字	48.0	54.4	52.4	43.9	50.3	41.4	55.8	46.4	45.2	66.6	39.5	52.3
200字	66.1	73.2	70.0	59.8	68.5	58.6	71.0	63.4	62.0	85.3	55.6	69.8
500字	88.9	95.6	90.6	82.8	89.5	83.0	90.1	87.2	84.2		80.3	90.1
1000字	93.6		99.1	96.3	98.8	96.9	99.9	99.9	96.7		94.5	98.2
1500字				99.7							99.0	99.9
全体の 80％	335字	256字	302字	445字	320字	445字	300字	374字	416字	159字	495字	302字
85％	414	307	376	551	400	542	381	458	519	197	607	380
90％	529	377	480	693	514	674	498	568	661	236	766	496
95％	711	482	660	910	696	890	691	729	885	274	1,034	695
96％	768	511	709	978	746	947	743	784	952	282	1,138	761
97％	836	538	777	1,064	817	1,011	822	840	1,020	289	1,207	849
98％	920	565	864	1,170	890	1,126	901	895	1,114	297	1,337	,966
99％	1,071	594	973	1,312	1,030	1,242	1,055	951	1,279	305	1,511	1,155
100％	1,240	622	1,172	1,582	1,171	1,358	1,214	1,007	1,414	313	1,779	1,554

表2.5　新聞層別区分による使用度数分布

を必要となるが，全体としてはそれほど多くを必要としないことを意味する。

以上のようなことから，漢字の使用はその分野分野でかなり相違のあることが考えられる。漢字使用は結局どんなことばが用いられているかにも関連し，どんなことばがどのように用いられるかということは，その分野，そのシチュエーションでどんな種類とコミュニケーションが行なわれるか（案内広告と商業広告の相違など）にかかわるところが大である。ある信託銀行では8,900の字種を使用し（記号類を除くと8,000），明治生命では6,500字種でやっているが不足があるとのことで人名を含めて8,000種必要であろう。漢字使用の実態はこのようなコミュニケーションの性質と関連していることは，機械処理システムの設計の上で，重要な考慮事項であるといわなければならない。

2.2.2　書体

漢字を中心として，明朝，清朝，宋朝，教科書体，ゴシックなどいくつかの字体がある。[注19] 清朝は名刺など，教科書体は教科書に，ゴシックは文中の特定の語句などに用いて，それぞれ使用範囲が限定されている。それ以外の，一般の使用には明朝が用いられる。それぞれ独特の様式をもったデザインである。字のくずし方として行書草書，楷書などあるが，行書，草書などにはふれなくてよかろう。

2.2.3　字形

これにはいくつかの問題がある。まず機械認識という観点からみた，漢字の字形。これはもっぱら図形としてみた漢字の形の問題である。次に字形についてのゆれの問題。最後に字形とミスパンチ（人間による認識行動の問題点）あるいは字形と情報量の問題（これは間接的に機械認識の問題に結びつく）

まず図形としての漢字の問題。漢字には部首があり，漢字の構成原理の一つに形声というのがあって，これに属する漢字の数がきわめて多いことから，当然それを部分に分解できる。たとえば「語」という字は「ゴ」という音（中国語でもともっていた語形—— significant）を表わす「吾」と，意味を示す「言」との二つの合成である。このようにまず分離して存在するいくつかの要素に分解する。この部分の数を限定し，その組み合わせ（オペレータを用いる）で，文字の形を記述しようとする方法（京都大学坂井研究室）（図2.2参照）と，その要素をさらにストロークに分解し，ストロークの組み合わせ方に規定を定めて字形を記述しようとする方法（東大藤村研究室）との二つの大きな考え方がある（図2.3参照）。このような記述法は漢字コードにもかかわりを持ちうるが，これからの漢字のインプット，アウトプットの機械の設計にも関係があると考えられる。

次に字形のゆれについて。漢字は手書きのばあいと活字のばあいとで字体に少しではあるが相違のあることがある（たとえばしんにゅう）。また当用漢字の新字体表にもとづいても，デザインの上で少しの違いを生じうることがある（せりふなど）。当用漢字の外では，字形の部分として当用漢字内のものと，活字上の形のちがうことがある。たとえば「艱難」とか「四辻近く」など。

このようなこまかな相違も，機械による認識，あるいは文字発生のときあるいは漢字の現段階での通常の入力のとき，一つの問題点となることがある。システム設計の側では，統一，標準化を望む声がある。

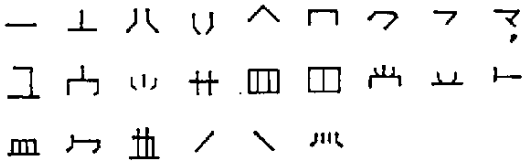
印刷などのばあい，ものによっては当用漢字の新字体でなく，旧字体を用いることを要請されるこ

(1)

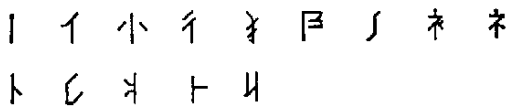
部分パターン		
ノ	又	ヒ
㇂	女	氏
㇆	子	礻
一	厂	ネ
丨	土	示
ノ	忄	疒
丶	寸	癶
㇇	宀	祭
乚	小	士
亠	屮	矢
八	尸	木

(2) (クラス A)

(3) (クラス B)



(4) (クラス C)



図A　漢字のデータ(部分パターン)の例

operator	意　味	例
H	左右(離)	休
V	上下(離)	電
P	左右(接)	非
Q	上下(接)	否
I	貫　通	中,丹,孝
S	包　囲	困
K	冠	字
L	ニョウ	進
J	タスキ	司
G	タ　レ	歴

表A　漢字の記述オペレータ

基本方式
図Aに示す漢字の部分パターンを表Aで示す漢字の記述オペレータを用いて結合する。
たとえば「林」はH(木, 木)
「高」V(亠, 口, S(冂, 口)

注)　坂井利之「情報工学」(筑摩書房1970, P.69)より
なお，このほか，次の文献を参照
Preprints for seminar on input / output systems for Japanese and chinese characters, 1971, US, Japan Committee on Scientific Corporation.

図2．2　京都大学坂井研究室における漢字字形の記述法

Table 1 Strokes

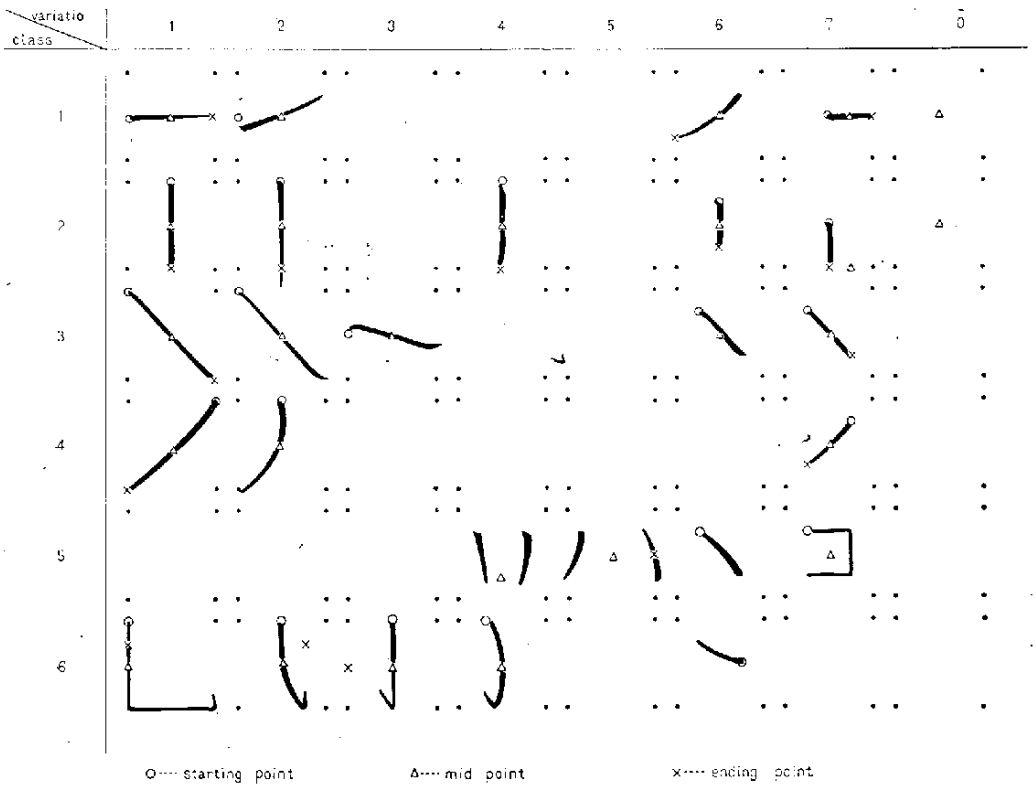


Table 2 Concatenators and Pseudo-Concatenators

	name	stroke 1	stroke 2	example
A	aim at center	μ	ω	/11, 17, 11A21/: 王
C	center start	μ	α	/11X42C36/: 大→大
E	end cross	ω	μ	/11C26E11/: 工
P	pass to start	ω	α	/11P41X32/: 又
S	start at start	α	α	/22S17P63P17-/: 月→月
T	terminate	ω	ω	/21S17P21A17T17/: 日
X	cross	μ	μ	/11X22/: †
—	repeat	μ	μ	(=X) The second stoke when identical to the first, is omitted in the i-rep. /21S17P21A17-T17/: 目→目
,	comma	μ	μ-2	Place the second stroke parrallel to the first and separate the μ's by two units in the direction perpendicular to the stroke lines. The two strokes must be of the same class. A comma must be followed by C or /

基本方式

Table 1のStrokeをTable 2のConcatenatorで結合させる。たとえば「王」は「一」(11), 「一」(17), 「一」(11)の三つの横線と「丨」(21)が中央で接する(A), のようになっている。

(注) 藤村靖「文学認識に関する2, 3の考察」(「情報と制御の研究」1963年4号)より

図2．3 東京大学藤村研究室における文字の記述の研究

とがある。(古典の資料の複刻など)。

最後に字形と情報量について。漢字はその構成上から字形が全体としてかなり複雑であり多くのストロークからなるので，そのことから種々の問題が生ずる。漢字入出力のばあい，このことがいくつかの障害を引き起こすのである。まず漢字を認識することについては，アルゴリズムがむずかしいかあるいはコストが高くなることである。出力についても，もしドット方式あるいはドットに分解する方式であれば字画が多いとメッシュの数がふえてしまう。現段階のもう一つの問題点としては，ミスパンチが多いということもある。入力に漢テレを用いるばあいでも例に示すように，誤りが多いが，漢テレを用いないばあいには，さらに誤りがふえていく可能性がある。漢字入力に誤りが多いことは，漢字処理システム全体に大きな影響を与え，データの保守についての，いろいろな段階でのソフトウエアを要求することになる。これは印刷の場合の校正とかなり似た面がある。例として国語研究所で入力した用語用字についてのミスパンチの例を示すが（表 2.6 参照），その最大の問題点はストロークが多いにもかかわらず字形の類似が多いことである。「宮」と「官」，「大」と「太」などで，これはすべてがキーパンチャの罪に帰せられるわけのものでなく，まぎらわしい字でパンチ前の原稿を書いた人の責任に帰するものが多い。漢字だけでなく，漢字とかな，あるいはかなどうしなどの間の相似ももちろんある（漢字の「力」「又」とかたかなの「カ」「ヌ」，かたかなとひらがなの「へ」「へ」，「リ」「り」など。

(誤)	(正)	(誤)	(正)
文	文	急及落	急反落
且つ	且つ	執行使	執行吏
子婦	主婦	報造書	報告書
北東	北京	作春	昨春
入気	人気	衆議院議院	衆議院議員
肉休	肉体	間合せ	問合せ
配遇者	配偶者	訪間	訪問
失取	失敗	集国防衛	集団防衛
印札	切札	強庄	強圧
厚案	原案	他合い	地合い

表 2.6 漢字インプットにおけるエラー・データの例

2.2.4 漢字の機能

漢字は日本語の表記のなかでは主として自立語を表わすのに用い，付属語を表わすために用いられることはない（以前は「迄」「也」のような表記が用いられたが，現代は特殊なばあいを除いてほとんどないといってよい)。

現代の表記法は戦後当用漢字を用いたことで変革を経てきたものが，二十数年を経て，かなり安定してきた部分があるといえる（「居る」などはたいてい「いる」と書いている)。動詞，形容詞などよく用いられる語が，かながきとなり，全体として戦前に比べてかながきの傾向がふえてきた。という書き方のなかで，一応の安定を示したものである。

そのなかで，漢字はいぜんとして全体としては語を書き表わすために用いられているといってよい

状態にあり，いわば分かち書きの役目を果している。しかし正書法が確立しているわけでなく，かな書き漢字書きが不安定な語もあり，人による相違も少なくない。そこで漢字がどの程度 word recognition の役を果たしているかを追求してみることも必要である。そのことが， word recognition のプログラムの作成にどのように役立てられるか，がプログラムのシステム・アナリシスにとって重要であるからである。[注20]

そこで，表記の上でハンドブックなどをつくりある程度人為的操作を加えようとしている新聞についてではあるが，その実態について調べたものがあるので，ここでその結論だけを紹介しておく。[注21]

1）文節は，漢字で始まり，ひらがなで終わる傾向が強く，とくに，漢字の連続のあとにかなの連続が続くという類型をもつものは，漢字とひらがなからなる文節の約60％をしめる。

2）ひらがなで始まる文節のうち，接続詞・連体詞・副詞などは，大部分がかなのみからなる類型に属し，文のはじめや，読点のあとに用いられることが多い。

3）ひらがなではじまる文節のうち，名詞・動詞・形容詞は，直前に，ひらがなのくる確率が高い。

4）ひらがなで表記される，名詞・動詞の大部分は，出現頻度の高い語によって占められ，その語を含む出現確率の高い文字列を作りやすい。

5）ひらがな表記の動詞・形容詞の直前には，漢字やかたかなを含む文節が書きやすい。

6）ひらがなの中には，文節の終わりに用いられることの多いものがあり（「を」）ひらがなが連続する場合に，意味の切れ目を示す機能を持つことが推測される。

このなかには文節にまたがって漢字が連続するばあいのことなど，システムアナリシスのためデータとしては必ずしも十分でない点もあるが（というのはもともと国語問題のための分析なので）。以上のことから，このような文字の連続のしかたに注目して単語の認識をすることは，かなり有望であると考えられる。もちろん，これだけにたよるならば，非常に高い近似度をねらうことには，問題があるかもしれないが。

2.2.5　漢字の配列

一般に文字には，なんらかの配列がある。かなは50音順，英字はＡＢＣ順にならべられる。漢字にはこのような，だれにでも手軽にならべられるユニークな並べ方がない（いままでなかった）。このことは漢字の，辞書その他での検索をめんどうにする。

一般には漢字字典の部首順，漢字の代表順などが用いられている。漢和辞典では漢字の訓による索引，総画数による索引なども用いられる。電話帳では独特のならべ方をしている。中国では四角号碼も用いられる。その他なおいくつかあるが，一般に普及していないものは，ここでは省く。そしてそれぞれに問題がある。漢字の配列は，キーボード上の配列とも関係があり，機械の設計，コードブックの作成など幅の広い関連問題を蔵している。

部首順は一般に康熙字典の方式を準用している。しかしこれには必ずしも検索に便利とはいえないところがある。「相」「和」など，つくりで引いたりする意外性がある。[注22] この点をわかりやすくした，康熙字典式の修正版もすでに出ており，かなり普及してきている。しかしこのような修正が行なわれ

てもなお，部首順による配列は必ずしも引きやすいという感じは与えていない。

部首順のよい点は，たとえば康熙字典などによることにすれば，漢字の配列がユニークにきまることである。

代表音による50音順という配列もしばしば用いられる。このばあい，音のないもの「樫」「搖」，音を使わないもの「扱」「垣」などをどうするか，また二つ以上音のあるもの，「生」のセイ，ショウ，「石」のセキ，シャク，コクなどをどうするか，などに問題がある。ことに後者の二つ以上音のある文字はその数が必ずしも少なくない（いわゆる漢音，呉音などの相違がある）。音による索引のもつもう一つの欠点は同音の漢字が数多く並ぶことである（カン，ショウなど）。

これに対して訓による索引は，漢和字典などでは有効に使用される。同訓になるものの数が少ないからである。ただ訓は音よりさらに多く，一字について存在するから，漢和字典の字の検索のばあいは有効であっても，逆に重複出現をこのまないばあいにはぐあいのわるい配列になってしまう。

電話帳の配列は一種独特のものであるが，全体の大きな配列は文字の読み方を重視しつつも，一字一字の読みが同じものについては，その字のレベルで同じ漢字を集めてしまうというような原則である。名簿などでは，これにならうものが少なくない。

このように配列にいくつかの方式があることは，機械化に際して不便を生ずることがある。まずコードとの関係があり，途中でコード変換が可能とはいうものの，一応配列と関係があり，検索可能であることも必要であろう。この点から，標準化を望む声が強い。

2.2.6　音訓

音訓についてはある程度配列のところで述べたが，別の面から補足しておく。漢字音は中国語の音をそのままの形で日本語のなかに移したものであるが，借用の際日本語的な受容のしかたをし，単純化したために，かつ日本語のなかでの音韻変化があったために（「甲」カフ→カウ→コウ，「広」クワウ→コウ，「工」コウ，「交」カウ→コウその他），同音の漢字が増加し，第二音節のバラエティもきわめて少なくなっている（ンイウオキクチツの8種）。

訓は中国語の意味の Japanese equivalent であるため，種類が多くなり，かつ必ずしも十分安定していない面がある。また，同訓異字の書き分け（たとえば「あげる」に対して「上」，「揚」，「挙」など）も必ずしも安定していない。

漢字の読み方としては一字についていくつかあるだけでなく（「間」あいだ・ま・ケン・カン，「方」ホウ・かた），文字の連続についてもおこりうる。「時雨」しぐれ，「五月雨」さみだれ，「五月蠅い」うるさい，など。つまりこの段階ではいわゆる語に対応して存在するただし，一般のばあいにはいわゆる語よりも小さいもの（logöid,あるいはmorpheme）に対応している（「必」ひつ・「要」よう・「性」せいなど。このように，漢字とその読みの関係は必ずしも単純でなく，いくつかの層（stratum）に重複して存在することがありそこに特質があるとみるべきである。

また漢字一字のばあいだけでなく，二字になっても読み方に ambiguity のあるばあいがある（「一行」いっこう・いちぎょう，「上手」かみて・うわて・じょうず，など）。

漢字の読み方，いわば日本語のなかでの使用法は，このようにきわめて複雑な面がある。漢字か

な，かな漢字のコンピュータによる変換を行なうばあいには，この点についての見通しを立てなくてはならない。[注23]

2.3　これからの日本語

以上，漢字を含めての日本語処理と日本語の特質についての問題について述べたが，次に，このような機械処理の対象となる日本語の将来についての意見を，簡単に紹介しておきたい。

まず，カナモジと漢字についてである。最近では各種の請求書など，電算機処理による帳票数が増加してきており，それが必ずしも多くの障害を与えてはいないようである。今後もこの傾向はふえていくことであろう。一方，漢字処理も次第に本格化してきており，普及を始めている。おそらく一方が他を制するということは急速にはおこらず，両者ともよく発展し普及していくことであろう。カナモジによる印字には手軽さと速さと安さという便利さがあり，これを生かしたコミュニケーションをすることも考えられてよいと思う。漢字を利用することが有効な言語伝達の分野では，漢字処理の本格化による利益が大きいことであろう。

次に，そのことと関連して，文字を中心とした日本語の将来を，機械処理と関係づけて，どのように考えていくかという点は大きな問題であり，これについてはさまざまな意見と動きがある。まず意見であるが，互いに反する二つのものがある。一つは機械に合わせて人間の生活を規制することに対する反論であって，このばあい，人間を機械に合わせるのは考え方が逆であって，機械を人間の要求にあうまで高めようというものである。故時枝誠記氏はこのようにいわれていた。もう一つの意見は，この際漢字をなくして，かなにしようというものである。[注24] 漢字を扱う機械を作るのはやはり考え方として時代逆行的であって，この機会を利用して国字を合理化しようというのである。

この二つの中間にいろいろの意見がある。たとえば漢字を処理する方向でおしすすめながら，しかしあまり使用しないような漢字はこの際淘汰してもよいのではないか，というのであり，あるいは機械あるいはコスト的な理由から，ある特定のせまい事務処理分野などでの条件を考えて，漢字の数を制限したシステムを作り，これを動かしているところも，いくつか出てきている（ＪＩＣＳＴなど）。漢字処理は始まったばかりなので，今後どのような方向に動くか，十分予測はしかねるが，上述のような行き方も一つの行き方として，今後の勢力の一つになっていくことが考えられ，文字の数をいくつかにおさえるか，ということは，その分野分野で事情が異なるから，かえってきめがたいこともあろう。むしろ全国民的には，全体の動きをどの方向に向けるべきかについて考えるのがようのではなかろうか。

最後に，漢字使用の言語として中国語があり，今後は国交回復による交流が行なわれるようになろうということから中国語との関係も考えるべきである。中国では漢字の字形を大幅に簡略化して，日本のそれと異なるものがかなり存在する結果となった。日本語では日本のなかでの諸事情があり，中国語では中国のなかでの諸事情が存在する。中国のなかでの言語事情は，最近のそれが必ずしもよく分かっていない面もある。日中の早急な文字の調整については悲観的な見方も行なわれているようである。しかし言語の機械による処理が大幅に進歩してきた今日，やはり，この問題を放置することが

できない事情が今後生じてくることも考えられる。

引用文献

注1．服部四郎「言語学の方法」岩波書店 1960 cf. p. 360

注2．高橋達郎「日本語の機械処理」東洋経済新報社 1970 cf. p. 100

注3．注2に同じ p. 89

注4．松尾拾，西尾寅弥，田中章夫「類義語の研究」秀英出版 1965

注5．石綿敏雄「現代語の語い」(『講座国語史，語い史』大修館 1971 所収)

注6．注5に同じ p. 368

注7．注5に同じ p. 387

注8．ソシュール著，小林英夫訳「一般言語学講義」岩波書店 1972

注9．たとえば T. Winograd : Understanding Natural Language, Edingburgh University Press 1972 p. 195

注10．坂井利之「電子計算機」(岩波新書) 岩波書店 1968 p. 225

注11．中井浩・高野守正 論理と情報の世界」ダイヤモンド社 1970 cf. p. 256

注12．松下温「かなやローマ字の入力をいかに漢字かなまじり文に変換するか」『ソフトウエア科学』1972-3

注13．石綿敏雄「人間の行動を意味する動詞の用法」(「電子計算機による国語研究Ⅴ」国語研報告 秀英出版 1973 所収)

注14．Kucera, Henry : Computational Analysis of Present Day American English. Brown Univercity 1967.

注15．H. Eggers u.a. : Elktronische Syntaxanalyse der deutschen Gegenwarts sprache Max Niemeyer 1969

石綿敏雄「ドイツのコンピュータ言語学」『計量国語学』1972-62

注16．注2に同じ cf. p. 136

注17．野村雅昭「新聞使用漢字の試行的分析」(『電子計算機による国語研究Ⅱ』国語研究報告 34 秀英出版 1969 所収)

注18．注17に同じ

注19．林大「当用漢字字体表の問題」(国語シリーズ 53) 光風出版 1963

注20．江川清「漢字かな混り文の『自動単位分割』に関する一研究」『計量国語学』1967-43144

注21．野村雅昭「漢字かなまじり文の文字連続」(『電子計算機による国語研究Ⅳ』国語研究報告 46 秀英出版 1972 所収)

注22．長沢規矩也「三省堂新漢和辞典」三省堂 1967

注 23．田中章夫「漢字かなまじり文を全文カナ書き・ローマ字書きに変換するシステム」（『電子計算機による国語研究Ⅱ』国語研報告 34　秀英出版 1969 所収）

石綿敏雄「新聞用語調査の用例印字プログラム COBOL-KWIC″」（のなかの，4．漢字解読ルーチン）（『電子計算機による国語研究Ⅲ』国語研報告 39　秀英出版 1971 所収）cf. p. 54

注 24．たとえば，杉本つとむ「歴史から学ぶべきもの――日本語への提言――」『言語生活』1972．11 月）

3．日本語情報処理の入力システム

3．　日本語情報処理の入力システム

電子計算機は，飛躍的な進歩を遂げたが，データエントリー (data entry) のメディア (media) の多様性はあるにしても，依然として人手を必要とするし，計算機言語が多種考案されてもソフトウェアの作成はいまだに大量の人，時間，費用を要しているのが実情である。英数字を基とするこのような計算機システムによって，日本語の情報処理を行なおうとすると，前章までに述べられたように，日本語の特殊性からくるさまざまな障害があることは明らかである。これは，ほとんどが入出力に集中していて，従来の機器と本質的に異なる点は，入出力装置である。[注1] 日本語の特殊性による装置の問題点は，[注2]

1）字種の多さにともなう装置の複雑化

2）装置の操作性の問題点

3）印刷様式（たて書き，よこ書き）

4）印刷または表示上の品質

5）字体の装置への適合性

6）文字コードの割当て

などである。入出力装置のうちとくに入力装置は，つぎの特殊性があるため考慮すべき点が多い。

1）漢字は字種が多い上に，一字を語 (word) と見た場合，英字では語がアルファベットで構成されるのに対して，確かに“へん，つくり”などの構成要素に分解はできるが，要素の変形で漢字一文字が形成されるので，欧文タイプライタと異なり印字が機械にのりがたい。

2）上記の理由で，いわゆる“探す”目視操作が必要なため，原稿目視だけで打鍵ができない。

3）熟練にかなりの時間を要する。

4）誤入力となる要素が多い。

5）入力装置は，字種が多く，英文タイプライタのように普遍性がないことなどから高価となる。キーボード，紙テープ機器のみの機成でも高価である。

装置の価格とどんな入力システムを採るかが，システムを考える場合に，コストパフォーマンス (cost performance) の観点から大きく全体を左右する要素となる。たとえば，校正システムの入力の場合，モニタプリントまたはディスプレイを併用して目視チェックを行ないながら入力するシステムと，目視チェックをまったくなしにいわゆる“めくら入力”する場合とは，装置の種類，台数，オペレータ数，オペレータの質などまったく入力システム構成，費用を異にする。日本語情報処理の入力システムは，従来の入力システムにおいて考慮すべき点に，日本語の特殊性を加えて，本質的にどうあるべきかを検討する必要に迫られている。参考のために従来の入力システムでの要点[注3] を述べると次のとおりである。

1）正しいデータの入力は全システムの大前提である。

2）実際の入力システムの内容は，誤り検出と訂正の作業である。

3）ヒューマンエラーの大部分は，データの記録とコーディングにある。

4）エラー検出の具体的方法は，目検法，重複法，バッチトータル法，コントロールトータル法である。

入力システムのハードウエアについては，数字コードシステム，フルキーボード方式から始まって多くの考案が行なわれてきたが　いまだに決定版が出現するに至っていないのが現状である。人手を介しない漢字認識，音声入力なども，まったくその研究が緒についたばかりである。

3．1　日本語情報入力システムのソフトウエア

ここでいうソフトウエアとは入力システムにおいて，システム中で入力機器をどのように使うかを指し，全システムの中でどのような位置づけにあるかを実例によって考察しようとするものである。

3．1．1　システムの流れ

入力システムは，大きく分けると入力作業そのものと，訂正および訂正入力の二つになる。また方式的に見ると，訂正をターンアラウンド（turn around）で行なうか，または対話形式で行なうかの二つの形態となる。とくに日本語入力となると，われわれにとって，“理解できる”内容であり，機械的なベリファイ法は非能率で行なわれていない。ターンアラウンドで行なうか，対話形式で行なうかは，

1）要求されるデータ校正サイクル時間と全処理システム時間との関係。

2）入力システム装置の投資費用の経済性

などに左右されるが，後者は日本語入力を考える場合，大きな要素である。

(1)　ターンアラウンド方式の例

現在，最も大量に日本語入力を行ない，かつターンアラウンド方式を採用している代表的例に，日本科学技術情報センターの文献速報自動作成システム[注4]がある。詳細に関しては第6章に述べられている。ターンアラウンドに要する時間は校正サイクルで，1サイクル1週間とされている。

(2)　対話形式の例

これは，1971年11月に日本アイ・ビー・エム社で完成された自動編集組版システム（Japanese Publishing System;JPS）[注5]の入力システムである。これも前例と同様に第6章に述べられている。新聞などの即時性を要求されるトータルシステムには不可欠である。とくにこの例では，校閲のみでなく，組版をも対話形式で入力している。

この二つの入力方式の比較は前記条件によって決定されるが，校正，修正処理および処理時間，費用などに，入力方式における問題点がクローズアップされ，[注6]対話形式または，目視チェックを強化する目的で漢字ディスプレイ装置による校正実験が行なわれている。[注7]対話形式のさらに拡張された例としては，日立製作所がその研究の一端を発表したカナけん盤漢字入力装置[注8]がある。この詳細は次の項で述べる。

3．1．2　入力手順

入力手順は次のような手順で行なわれる。

1）原稿の用意（原稿フォーマット）

2）入力装置による入力（一般に紙テープに出力）

3）校正マスターテープ（校正未完のテープ），校正トランザクションテープ（校正情報入力のテープ）への紙テープから磁気テープへの変換。

4）校正と訂正。次項に述べる校正サイクルにしたがって，校正と訂正を行ない，訂正部分を所定の手順によって指定して，訂正データの紙テープを作る。

5）電子計算機によるマスターテープとトランザクションテープのマージ。校正サイクル中で行なわれるが，最終的には完成したテープとなる。

以上の手順の中で，最も重要なのは校正，訂正の具体的手法と，そのサイクルである。

3．1．3　校正サイクル

図3.1にＪＩＣＳＴで行なわれている校正サイクルの概念図を示し，図3.2（＝図6.8.3），に日本アイ・ビー・エム社で行なわれているＪＰＳのシステム図を示す。

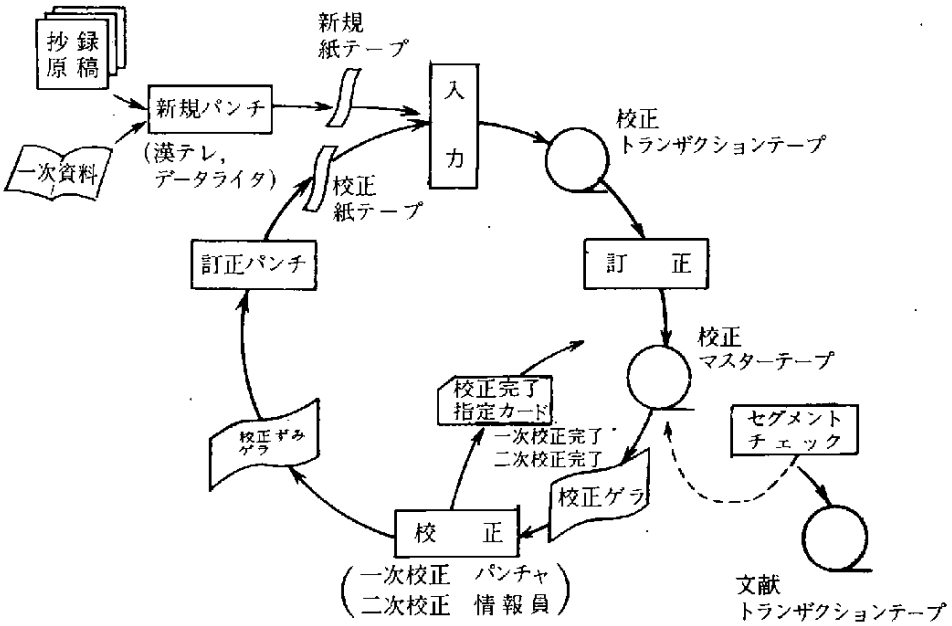


図3．1　校正サイクル概念図（JICST）

校正サイクルは，ＪＩＣＳＴの例では原稿と校正ゲラとは内容と順序を同じで見やすくしてあり，訂正をする場合には，原稿と見比べられてのち朱筆がいれられた原稿により，図3.1のごとく訂正パンチ作業が行なわれる。訂正レコードを指示するためには，

1）原稿番号

2）セグメントの種類

3）セグメント中のレコード番号

を必要とし，2字訂正するのに26字パンチするとされている。訂正するのに要する情報量および，訂正の訂正誤りを避けられるような運用が望ましい。訂正に要する校正サイクルの収れん速度がきわ

めて重要で，これをとくに考慮に入れた入力システムもある。[注9)]

3．1．4　エラー

日本語情報処理システムで，エラーの発生する個所はおもに次のように考えられる。

1）原稿

2）オペレータ

3）入力機器

4）校正者

5）ソフトウエア

6）電子計算機

7）出力機器

このうちで，エラーの出やすいのはいずれも人間に介在するところで，1)，2)，3)，4)，5）である。入力エラーは，入力条件で大巾に異なるが，表3.1に各種データの比較を示す。

エラーに対する対策は，入力システムの良否，関係する要員のワークマンシップ（workmanship）にまでさかのぼる必要がある。エラーに関して日本語入力は表3.1に示すように多かれ少なかれむずかしさが存在する。

入力手段	エラーの状況	入力条件	備考
電話ダイアル	8／100件誤ダイアル	7桁ダイヤル	参考文献 3）P. 469 ～P. 470
キーカードせん孔機	1／1,500ストローク	経験半年	
カナタイプ検定	1級1／250，2級1／200，3級1／150字		
タイプ後の目視チェック	80％発見可能		
電信オペレータ	1／3,200字	昭和40年	
電信時チェック	9.8／10,000→2.8／10,000字	〃 38年	
漢字ミスパンチ	1／100字		4）P. 12
実験校正	発生率（件）64％→3％		7）P. 22

表3.1　各種入力エラーの比較

3．1．5　入力字種，コードおよび字体

一般的に日本語情報処理システムに使われる漢字の字種については次の条件を考慮しなければならない。

1）当用漢字　1,850字

2）新聞　2,600〃

3）雑誌　3,400〃

4）大漢和字典（服部）　8,626〃

5）康熙字典　49,030〃

使用字種の選定については，入力，出力機器とも装置ハード・ウエア仕様を決定する上で，細密な解析調査が必要である。ＪＩＣＳＴにおいては，このような解析が行なわれ，結果として表3.2，表

3.3（＝表6.1.1）のようになった。

この例は科学情報と範囲が限られている場合で，さらに行政体組織，民生業務などに範囲を拡げると，最低3000字種は必要とされている[注10]

コードは，字種数の使用分野における違いなどから，新聞社5社の協定によるＣＯ―59があるだけで，漢字コード体系はばらばらで，けん盤配列と同様に，標準化がきわめて困難である。ただし今後，日本語情報処理システムが普及するにしたがって，標準化，規格化が強く望まれるところである。入出力機器において，日常，よくわれわれの目にふれる字体についてもバラエティに富んでおり，下記のような系列に分かれている。

1）新聞，雑誌，印刷所固有の字体

2）印刷装置メーカによる字体

3）字体家（シタリング・デザイナー）による字体

4）電子計算機に適した字体（たとえばドット，折線字体など）

入力装置に，人間工学的に見てどの字体がよいかなどの定説はない。

編	抽出標本記事数	抽出延べ字数	異なり字数	異なり字数／延べ字数
物理	136	8,625 ⑥	687 ⑨	0.08
機械	208	16,760 ②	954 ②	0.06
化学	370	19,227 ①	1,008 ①	0.05
電気	120	10,580 ⑤	788 ⑤	0.07
金鉱地	122	10,837 ④	783 ⑦	0.07
土建	83	6,705 ⑦	824 ④	0.12
経営	75	5,934 ⑨	784 ⑥	0.13
原子力	53	6,484 ⑧	701 ⑧	0.11
国内化	130	14,324 ③	883 ③	0.06
小計（当用）	1,297	99,476	全編 1,378	
（表外）		295	150	
合計	1,297	99,771	1,528	

注）○付き数字は各欄数値の編別順位を示す

表3.2　JICSTにおける編成異なり漢字表

3．2　日本語情報入力システムのハードウエア

この項では，これまでに開発，製品化または発表された入力システムの装置のまず方式的な面から見た分類を述べ，つぎにハードウエアを詳細に，原理，特長などについて説明する。

3.2.1　エントリー形式の分類

表3.4に日本語情報入力の方式分類を示す。

日本語情報入力は二つに分かれて，人手を介する方式と人手を介しない方式に分かれる。現在，市場にある製品または具体的に発表されたものは前者が大部分で，後者は主として研究的段階にある。

(1) 数字コード方式

これは，前もって漢字を数字によって，コード化してインプットするもので，コードへの変換はテーブルなどを用いて人間が行なうもので，テンキーボードなどを入力機器として使用しうる。

(2) フルキーボード方式

最もありふれた入力方式で，けん盤のキイトップに文字を配置し，シフトキーにより一つのキートップのうちの一つの文字を指定する。4シフト位のシフト数から，10シフトまである。

(3) テーブル・ルックアップ方式

文字を表に配置する点では，外観的にはフルキーボードに類似するが，全文字をA3版大にまとめ，表におけるその文字をペンまたはスタイラスで指示することにより入力を行なうもので，平面的に表を設けるものと，円筒状のものがある。目視すべき面積が少なくなる。

(4) マルチストローク方式

フルキーボードやテーブル・ルックアップのように，全文字種から一字を選ぶ困難さを多くの指のストローク・タッチで行なうものである。カナ文字入力によって，電子計算機で所定の漢字に翻訳する。直接全部の音，訓読みを入力するものと，電子計算機と対話を行なって，なるべく少ないキータッチで入れようとする対話型がある。文字を“へん”“つくり”などの構成要素に分解して入力するもの，音によりそれの属するファミリー全部を表示し，その中から所定の一個を指定するものもある。表3.4に字形式とあるのは，文字の特長によって，ファミリーにまとめた写真植字に採用されているような方式をいう。

- 日本語情報入力方式
  - 人手を介するエントリー方式
    - 数字コード方式
    - フルキーボード方式
      - 少数シフト
      - 多数シフト
    - テーブルルックアップ方式
      - 平面テーブル
      - 円筒面テーブル
    - マルチストローク方式
      - カナ文字入力
        - 直接型
        - 対話型
      - 文字分解ストローク型
      - ファミリー分類型
        - 音訓式
        - 字形式
    - 人手介在音声入力(速記入力式)
    - オンライン手書き漢字入力方式
  - 人手を介しないエントリー方式
    - 印刷漢字認識入力
    - 音声入力

表3．4　日本語情報入力方式の分類

(5) 人手介在音声入力（速記入力方式）

会話や議事を聞いて速記するかわりに，特殊キーボードで入力するもので，きわめて長い訓練を要するが入力速度は早い。

(6) オンライン手書き漢字入力方式

書かれる文字のストローク，書き順などを用紙の下に置かれたタブレット装置で座標変化して送り，文字認識を行なうものである。

(7) 印刷漢字認識入力

前項は書かれつつある文字を自動認識するのに対して，この入力は印刷された漢字を認識し入力するものである。

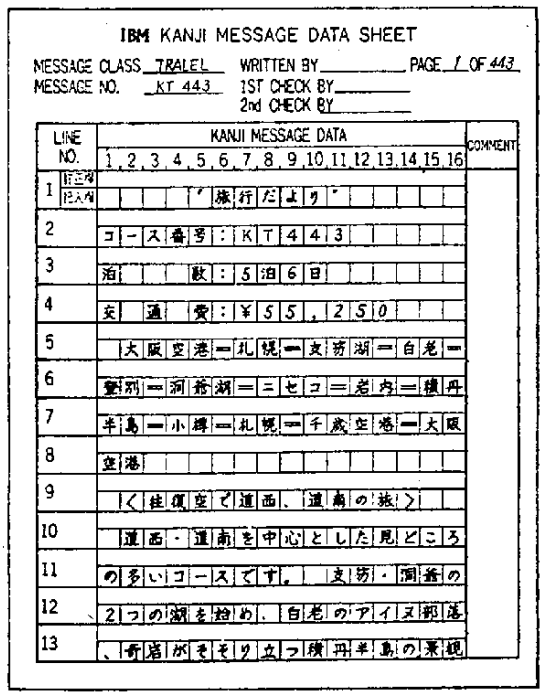
(8) 音声入力

文字どおり人間の発する音声を認識し日本語情報をインプットしようとするもので，まだ研究の段階であるが，その動向を述べる。

3.2.2 人手を介するエントリー方式

(1) 数字コード方式

この方式は，前もって準備された漢字と数字コードの対応表を使用して，人手により漢字を10進コードで変換する。これをさらにテンキーボード，キーカードさん孔機を使って，紙，磁気テープまたはカードに記録する。この例を図3.3，図3.4に示す。[注11] 10進数4桁を使えば最高1万字種を取り扱いうる。コードを人間が記憶するのに便利なようにしておけば，翻訳作業の限界はある程度軽減しうる。コードの決め方は次のふたとおりの考え方がある。



IBM KANJI MESSAGE DATA SHEET

MESSAGE CLASS 7RALEL WRITTEN BY ＿＿＿ PAGE 1 OF 443
MESSAGE NO. KT 443 1ST CHECK BY ＿＿＿
2nd CHECK BY ＿＿＿

LINE NO.	KANJI MESSAGE DATA 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16	COMMENT
1	行先 住人名 ゛旅行だより゛	
2	コース番号：KT443	
3	泊数：5泊6日	
4	交通費：¥55,250	
5	大阪空港ー札幌ー支笏湖ー白老ー	
6	登別ー洞爺湖ー三セコー岩内ー積丹	
7	半島ー小樽ー札幌ー千歳空港ー大阪	
8	空港	
9	〈往復空で道西、道南の旅〉	
10	道西・道南を中心とした見どころ	
11	の多いコースです。支笏・洞爺の	
12	2つの湖を始め、白老のアイヌ部落	
13	、所沢がそそり立つ積丹半島の景観	

図3．3 漢字文字印刷レイアウトの例

IBM KJ IMAGE CODE MESSAGE SHEET

MESSAGE CLASS: TRAVEL	WRITTEN BY:	PAGE 1 OF 443
MESSAGE NO.: KT 443	KEY NO.:	REL. BLOCK ADDR.

MESSAGE DATA

NO.	DATA FIELD 1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61	CTL 65	RESERVED 66 72	LENGTH 73 74	PCB 75	REY NO. 76 78	SEQ NO. 79 80
1	9679067101219857983898409679	1		44	T	443	01
2	9710966797130423047396749611962099039903990 2			44	T	443	02
3	1304 0271967499041304990501 05			40	T	443	03
4	0410 0176 0523967496619661990496779901990499 09			52	T	443	04
5	0104057803841186967812173730967805212669168196780183061896 78			64	T	443	05
6	0772028496781712174416819678972297749710967807230186967808051395			64	T	443	06
7	0397025896780152204596781217373096780189083203841186967801040578			64	T	443	07
8	0384 1186			08	T	443	08
9	9655127407740384986002320382968402320421982506719672			56	T	443	09
10	0232038296540232042198450110016598209812981601159861981098 43			64	T	443	10
11	9825021698029710966797139860981396 83 0521266996541712174498 25			64	T	443	11
12	9901981898251681984504539834968401830618982597019702972301510351			64	T	443	12
13	9684108807239847981598159840014998180805139503970258982507520654			64	T	443	13
14	9684344012462105081898080772028407171056982501531426051398219861			64	T	443	14
15	0597023098471612980798319814984696 83	1		36	T	443	15
16	0181 1118 0105011397019701965497639667965497049733			56	T	443	16
17	0595 1652 0105011304100176041101 72			40	T	443	17
18	0105011302260136103617560133 9904966899099904966896779906990 9			64	T	443	18
COMMENT							

図3．4　漢字コードに変換した文章例(図3．3のレイアウトを変換したもの)

メーカー名	富士通株式会社	富士通株式会社	谷村(株)新興製作所	谷村(株)新興製作所	日本電子産業㈱,谷村㈱新興	日本アイビーエム株式会社	高千穂交易株式会社	沖電気工業株式会社
機種名	F6801A	F6802B	S－4000	SCK－200	KB202A	IBM漢字鍵盤機構付29ぜん孔機	T-4112, T-4011, T-4012	漢字鍵盤さん孔機
文字キーの数	480＋192(オプション)	240	48	192	192	240 (space bar含まず)	256	650
シフトキーの数	8	12	1～3	MAX 15	14	15	15	4
文字キー当りの文字数	8 文字/キー	12 文字/キー	文字/キー	MAX 15 文字/キー	14 文字/キー	15 文字/キー	12 文字/キー	4
文字種	3,840種	2,880(細目はユーザーによる)	細目はユーザー指定	細目はユーザー指定	2,688種	3,500種	3,072種	2,600種
漢字	2,808				1,912	2,950	2,808	
ひらがな	86				85	87	83	
カタカナ	103				87	86	83	
英数字	62				62	62	62	
記号	232				150	170	42	
特殊文字	549				392		10	
その他						ユーザーの選択した145字		
外字	1,536				4,500	10,000(MAX)	5,120	
入力可能文字種	5,376		4,096種	2,880種	7,188		8,192	5,200
記録媒体	紙テープ	紙テープ	紙テープ	紙テープ	紙テープ	IBMカード	カセット型磁気テープ	紙テープ,ディスプレイ
コード パリティの有無	8 bit 無	6 bit 無	6 bit , 8 bit パリティ可	6 bit , 8 bit 無	8 bit 無	EBCDICカードコード (2 Col /字)	8 bit 有	6 bit , 8 bit 有
モニター	無	無	可 (SC－4000)	プリント可(SCP－200)	無	無	ディスプレイ	ディスプレイ
打鍵速度(MAX)	80 文字/分	120 文字/分	400 文字/分	400 文字/分	100 文字/分	通常60～80文字/分	120 文字/分	500 字/分
外字インプット方法	システムによる 一例 番号入力	テンキーによる番号入力			パターン入力	外字用連数字キー (100種) を2度Key inする事によって入力	コード入力方式	
使用環境			常温常湿	常温・常湿		通常のオフィス環境	T-4011:紙テープせん孔 T-4012:ディスプレイモニターなし	
温度	5 °C～35 °C	5 °C～35 °C			10 ℃～30 ℃		15 ℃～35 ℃	5 ℃～35 ℃
湿度	40%～80%	40%～80%			35 %～85 %		20 %～80 %	30 %～70 %
備考	コード 富士通漢字コード 書体 明朝/ゴシック切換あり	コードCO－59 制御方式 IC						
価格 買取 万円	140 万円 160 万円(オプション付)	未定		110万円	140万円	約612万円	488万円	160～180万円
レンタル 万円/月						12万円/月		
発売年月	昭和45年3月出荷	昭和47年12月出荷	昭和47年8月	昭和39年	昭和42年		昭和46年4月	昭和43年12月
販売	富士通株式会社	富士通株式会社	谷村(株)新興製作所	谷村(株)新興製作所	日本電子産業(株)	日本アイビーエム株式会社	高千穂交易株式会社	沖電気工業株式会社

表3．5 日本語情報入力装置（その1 フルキーボード方式の代表的な例）

1）漢字の形をもととして決める。

2）漢字の音訓をもととして決める。

記憶のしやすさの点からは後者が有利とされている。数字コード方式はどれだけの漢字のコードを記憶できるかにかかっている。この方式を使用する場合，処理システム内ではメモリ必要量を節約するため，10進4桁よりたとえば12ビットに変換する必要がある。

(2) フルキーボード方式

日本語の入力機器としては最も歴史も古く，さまざまなものが発表されている。

表3.5は現在市場にある製品の代表的なものを示している。由来から見ると次の三つがある。

1）漢字テレタイプ，いわゆる“漢テレ”をそのまま入力として使う。

2）自動モノタイプ用のけん盤を流用する。

3）漢字ディスプレイとか，漢字システム専用に設計され開発されたもの。

自動モノタイプ，漢テレの開発は1920年に始まり，[注12] 長い経路をもつため採用しやすい。表3.4に示すように，少数および多数シフトがあるが，操作上，価格上の優劣はつけがたい。キーボードの大きさに関しては，明らかに多数シフトが小型になる。図3.5に新興製作所製漢字けん盤さん孔機の文字配列を示す。さらに大規模な面積，機能を持った例としては FACOM 6801 A 漢字入力キーボードがある。その入力可能文字数は3,840種であるがオプションで5,376種まで拡張できる。またT 4112漢字鍵盤ディスプレイ・レコーダーではその名のとおり，入力した漢字64文字をブラウン管に表示し，目視チェックができる。出力は従来一般的に行なわれて来た紙テープ出力ではなくて，最近英数字の入出力媒体として流行しているカセットテープに出す。つぎに漢字けん盤をキーボードさん孔機を取り付けて，漢字入力装置にした例を写真3.1に示す。これはＩＢＭより発表されたもので，ＩＢＭ29穿孔機（カタカナ付）を母体として漢字データ，英数字データ入力ができる。主要性能については表3.5のとおりである。漢字けん盤上にない文字を入力するには，外字処理用連数字キーが用意され，2けた連数字を2回キーインすることにより，4けたの文字番号を入力する。

(3) テーブル・ルックアップ方式

表3.4に示すように，平面および円筒面テーブル型がある。従来の和文タイプライタの簡単な改造によって，この二つの型を実現できることが提案されたが，[注2] その変形は最近日本電気から発表されたC 5110漢字入力装置である。写真3.2に外観，図3.6に文字盤を示す。この入力装置の特長は，素人でも直ちに扱いうることと，日本全国に10万人に近い習得者がいることであるとされている。原理的には片手でＸＹの座標に配置された文字盤の入力希望の文字を探して文字指示具をもってきて，左手で印字レバーを押すとトリガが働くと同時に活字庫から指定した活字が拾われて印字が行なわれるようになっている。文字のＸＹ座標は，レーザ光源を使うホロタブレットや高周波による電子タブレットにより文字コードとして送られる。[注13], [注14]

次にＩＢＭが，EXPO'70 で発表したスタイラスによる漢字入力装置を述べる。これは次の点を考慮して設計された。[注11] すなわち従来のものは，

1）3段より最高12段にわたるシフトキーと文字キー両手操作は相当の熟練および専門化を必

図3．5　漢字テレタイプの文字盤配列(新興製作所)

要とする。

2）コード体系が確立していない。

3）不特定多数の人が使用できるオフィスマシンではない。

この装置のキーボードの表面形状は，約84×33cmサイズにまとめられ，使用頻度にしたがって320，1,600，1,600文字の3区画に分けられている。漢字コードは，スタイラスといわれる鉄筆のような形をしたペンを使って，所定の文字位置を軽くおさえることによって，2桁のカードコードまたは

写真3．1　IBM漢字けん盤機構付29穿孔機

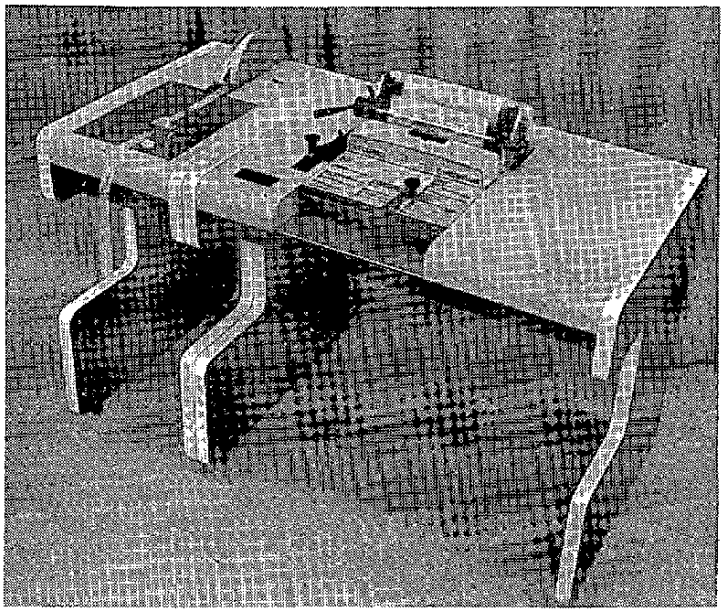


写真3．2　NEC C—5110　漢字入力装置

内　容　ひらがな　75字、カタカナ　75字、数字　30字
英字　52字、記号・符号　35字、予備欄　60字
当用漢字　1850字、補正漢字　28字

●標準盤面活字表（2205字）

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
	ア	亜	愛	哀	悪	握	圧	扱	安	案	暗	引	印	因	員	院	.	‥	…	:	・
	イ	以	衣	囲	位	医	委	胃	移	異	意	飲	姻	陰	隠	韻	&	!	?	／	½
	遺	依	威	尉	偉	為	違	維	慰	緯	芋	域	育	逸	沿	鉛	$	#	No.	%	@
	ウ	羽	雨	宇	雲	運	衛	詠	鋭	易	益	液	駅	疫	炎	延	＋	－	×	÷	＝
30	エ	永	泳	英	栄	営	映	影	閲	悦	越	謁	縁	塩	園		○	◎	△	※	～
29	オ	ヲ	王	往	皇	応	黄	横	沖	屋	憶	音	恩	温	遠	演	宴	援	煙	々	
28	汚	央	押	欧	殴	翁	奥	桜	虞	乙	卸	穏	価	佳	活	括	割	渇	轄	穀	
27	カ	火	化	加	可	何	河	仮	花	荷	果	菓	華	架	掛	滑	且	株	刈	渦	
26	貨	夏	科	課	家	嫁	禍	箇	寡	蚊	歌	暇	過	干	刊	汗	肝	甘	巻	看	
25	回	快	改	界	海	介	灰	各	角	画	革	完	官	管	寒	冠	寛	貫	慣	敢	
24	戒	械	開	会	悔	怪	絵	格	拡	覚	核	間	閑	関	簡	感	憾	歓	観	勧	
23	皆	階	解	塊	貝	懐	壊	郭	隔	較	確	幹	乾	勘	堪	陥	漢	館	換	喚	
22	カ	我	賀	外	害	劾	該	閣	獲	穫	嚇	愚	棺	款	監	緩	環	還	艦	鑑	
21	芽	雅	餓	街	慨	概	学	楽	額	岳	岩	岸	眼	願	顔	丸	含	渥	気	汽	
20	キ	己	危	忌	鬼	規	既	飢	企	季	器	岐	祈	喜	貴	帰	希	奇	寄	騎	
19	紀	記	起	基	期	棋	旗	揮	輝	軌	棄	机	幾	機	共	供	京	享	協	脅	
18	菊	吉	詰	喫	客	却	脚	級	給	糾	去	居	挙	許	凶	叫	恐	教	強	橋	
17	弓	久	及	吸	旧	丘	求	救	球	巨	拒	距	虚	拠	況	恭	胸	競	境	鏡	
16	休	朽	泣	急	宮	究	窮	曲	局	極	均	斤	近	禁	狂	狭	峡	郷	驚	響	
15	キ	技	宜	欺	偽	戯	疑	擬	義	儀	犠	虐	逆	琴	緊	菌	筋	勤	謹	矯	
14	ク	句	苦	駆	空	屈	掘	繰	君	議	牛	魚	行	業	仰	凝	暁	玉	銀	吟	
13	グ	具	愚	偶	遇	軍	群	訓	薫	勲	欠	穴	血	大	件	見	券	研	肩	圏	
12	ケ	兄	系	係	敬	景	茎	径	経	軽	慶	決	傑	剣	倹	検	険	験	建	健	
11	刑	型	形	契	啓	恵	傾	継	掲	携	警	憩	潔	軒	堅	献	権	憲	兼	謙	
10	ゲ	下	芸	迎	鯨	劇	激	撃	元	言	原	鶏	結	遣	顕	賢	絹	繭	懸	渓	
9	コ	古	故	枯	工	功	巧	攻	江	貢	源	減	玄	幻	弦	限	現	厳	告	谷	
8	固	個	湖	戸	口	向	后	厚	項	孔	好	考	孝	酵	康	侯	候	国	黒	克	
7	呼	庫	孤	弧	交	効	郊	香	肯	光	坑	抗	航	控	拘	校	耕	刻	穀	酷	
6	雇	顧	誇	鼓	甲	幸	公	更	硬	荒	慌	恒	港	降	高	稿	興	骨	込	洪	
5	ゴ	午	後	語	誤	互	広	鉱	鋼	綱	紅	絞	構	講	購	今	混	困	紺	婚	
4	サ	左	佐	差	査	呉	悟	娯	碁	護	拷	剛	豪	獄	衡	根	恨	魂	墾	懇	
3	砂	詐	唆	鎖	才	災	再	妻	細	債	載	作	昨	酢	咲	削	刷	山	産	算	
2	菜	採	彩	祭	際	最	済	砕	宰	催	栽	裁	策	索	錯	搾	殺	蚕	散	賛	
1	ザ	座	在	材	財	罪	剤	雑	残	暫	斎	歳	冊	札	察	擦	撮	酸	惨	桟	

	V	W	X	Y	Z	1	2	3	4	5	6	7	8	9	0	,	a	b	c	d	e
	昭	和	年	月	日	/	2	3	4	5	6	7	8	9	0	¥	あ	い	う	え	お
	都	道	府	県	市	1	2	3	4	5	6	7	8	9	0	合	か	き	く	け	こ
	区	郡	町	村	字	一	二	三	四	五	六	七	八	九	十	計	さ	し	す	せ	そ
	丁	目	番	地	金	壱	弐	参	拾	百	千	万	億	兆	円	銭	た	ち	つ	て	と
	号	殿	様	御	中	士	仕	至	止	祉	歯	司	伺	詞	嗣	飼	な	に	ぬ	ね	の
	シ	旨	指	脂	死	示	視	資	賜	氏	紙	支	枝	糸	紫	雌	は	ひ	ふ	へ	ほ
	史	使	私	始	姉	姿	子	志	誌	思	師	刺	施	詩	試	諮	ま	み	む	め	も
	失	室	質	写	車	社	勺	尺	手	主	首	酒	収	舟	秀	式	や	ゆ	よ	わ	を
	疾	執	赦	斜	者	煮	借	釈	種	朱	珠	殊	衆	集	臭	識	ら	り	る	れ	ろ
	漆	湿	舎	捨	射	謝	爵	守	狩	取	趣	修	習	終	就	芝	出	春	俊	瞬	ん
	処	書	暑	署	諸	緒	囚	周	週	秋	愁	州	酬	醜	襲	祝	叔	淑	宿	縮	粛
	初	所	庶	小	少	抄	称	渉	升	昇	匠	床	色	植	殖	心	臣	身	辛	針	進
	召	沼	招	詔	紹	照	肖	消	硝	章	障	彰	食	飾	触	審	深	森	新	薪	親
	唱	晶	松	訟	証	症	祥	詳	賞	償	焦	礁	嘱	織	職	信	診	申	伸	神	紳
	承	象	商	勝	将	奨	笑	掌	焼	粧	傷	衝	鐘	真	慎	振	娠	震	侵	浸	寝
	ジ	チ	次	寺	治	弱	受	授	旬	准	盾	順	純	潤	女	上	冗	丈	冑	尚	酌
	耳	自	似	事	辞	若	需	儒	殉	準	循	巡	遵	徐	如	状	条	常	辱	仁	刃
	侍	時	持	児	軸	寂	寿	充	銃	柔	従	縦	熟	除	序	情	場	城	浄	迅	陣
	慈	滋	磁	璽	実	邪	樹	住	重	獣	渋	述	術	叙	助	乗	剰	蒸	錠	尋	尽
	ス	水	吹	炊	推	粋	帥	衰	穂	遂	垂	睡	錘	据	杉	嬢	譲	醸	畳	汁	壌
	ズ	ソ	図	随	髄	酔	数	枢	崇	寸	夕	石	昔	惜	籍	川	泉	線	先	洗	銑
	セ	生	性	姓	牲	星	井	西	世	勢	斥	析	赤	席	跡	占	専	染	船	扇	旋
	正	政	征	整	声	制	製	省	畝	婿	瀬	隻	責	積	績	宣	戦	浅	践	遷	選
	青	静	清	晴	精	請	成	盛	誠	聖	誓	切	折	接	拙	摂	潜	鮮	薦	織	斉
	ゼ	是	税	舌	絶	全	前	善	然	漸	禅	繕	雪	節	窃	設	説	奏	則	側	測
	ソ	素	阻	粗	疎	塑	早	争	倉	創	壮	荘	装	相	想	霜	総	喪	足	束	息
	祖	組	租	措	訴	礎	草	走	送	宗	僧	層	遭	双	巣	操	燥	葬	促	速	即
	ゾ	造	像	増	憎	贈	蔵	臓	俗	族	属	続	賊	桑	窓	捜	掃	騒	卒	存	尊
	タ	太	体	待	対	耐	怠	態	替	胎	滝	沢	拓	託	達	丹	胆	担	探	孫	損
	他	多	貸	袋	帯	滞	泰	逮	隊	退	宅	択	卓	但	単	炭	短	淡	端	嘆	誕
	ダ	打	妥	惰	堕	大	台	代	第	題	諾	濁	脱	奪	男	断	団	談	段	暖	鍛
	チ	池	知	痴	逐	秩	窒	茶	虫	忠	仲	衷	長	帳	張	脹	腸	庁	朕	弾	壇
	置	値	致	竹	畜	着	嫡	直	抽	注	柱	駐	鳥	弔	頂	彫	調	聴	陳	賃	珍
	恥	遅	稚	築	蓄	貯	著	勅	宙	昼	鋳	超	跳	朝	潮	澄	徴	懲	沈	鎮	挑

	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
	や	ゆ	よ	っ	ー	ッ	ッ	津	追	墜	通	痛	坪	釣	的	笛	敵	滴	摘	適	天
	が	ぎ	ぐ	げ	ご	テ	定	弟	底	低	抵	呈	程	帝	締	迭	鉄	撤	典	店	点
	ざ	じ	ず	ぜ	ぞ	停	訂	貞	廷	庭	艇	堤	提	邸	逓	哲	徹	展	転	添	亭
	だ	ぢ	づ	で	ど	デ	田	伝	電	泥	東	凍	豆	登	頭	燈	痘	闘	峠	特	偵
	ば	び	ぶ	べ	ぼ	ト	斗	吐	徒	当	党	等	筒	答	塔	投	到	倒	得	徳	
	ぱ	ぴ	ぷ	ぺ	ぽ	途	塗	渡	刀	冬	湯	悼	透	陶	稲	踏	統	匿	督	篤	
	、	。	（	）	「	」	ド	土	島	討	盗	桃	逃	唐	糖	謄	騰	突	届	豚	
	ナ	ニ	ヌ	内	南	軟	難	努	度	奴	怒	同	動	働	堂	童	導	銅	胴	毒	
	ネ	寧	熱	念	燃	粘	尼	肉	入	乳	尿	人	任	妊	忍	認	独	読	鈍	曇	
	ノ	納	能	農	悩	脳	濃	敗	俳	廃	博	薄	箱	畑	反	坂	板	版	販	飯	
	ハ	波	破	派	拝	配	背	肺	排	白	伯	拍	発	髪	半	判	伴	畔	煩	頒	
	バ	パ	馬	婆	倍	培	陪	杯	輩	迫	泊	舶	帆	犯	班	般	搬	範	繁	藩	
	ヒ	皮	彼	被	疲	賠	買	売	梅	媒	麦	爆	縛	伐	閥	抜	罰	晩	蛮	盤	
	比	批	非	悲	秘	否	妃	肥	費	卑	碑	飛	罷	氷	評	表	俵	票	漂	標	
	ビ	ピ	美	備	鼻	尾	微	秒	病	苗	猫	敏	避	匹	必	泌	筆	姫	品	貧	
	フ	不	父	夫	扶	婦	負	富	赴	浮	布	怖	付	附	符	腐	膚	浜	賓	披	
	ブ	プ	武	部	侮	舞	仏	物	文	聞	敷	賦	普	譜	復	複	腹	覆	払	沸	
	ヘ	平	丙	兵	並	併	柄	閉	陛	癖	風	封	伏	服	副	幅	福	分	粉	紛	
	ベ	ペ	米	別	弁	便	勉	幣	弊	壁	片	辺	返	変	偏	遍	編	慎	墳	噴	
	ホ	歩	保	浦	捕	補	舗	方	放	倣	芳	訪	包	抱	砲	胞	飽	崩	縫	奮	
	ボ	ポ	墓	募	亡	忙	乏	宝	法	邦	豊	奉	報	峰	北	本	奔	翻	倹	堀	
	母	簿	慕	暮	忘	望	剖	防	坊	妨	肪	傍	紡	房	冒	帽	某	謀	棒	貿	
	マ	麻	摩	魔	毎	妹	枚	埋	幕	膜	又	末	満	漫	慢	暴	膨	牧	撲	墨	
	ミ	ム	未	味	魅	密	脈	妙	民	眠	無	夢	矛	務	霧	娘	没	凡	盆	僕	
	メ	モ	名	銘	命	明	盟	迷	鳴	滅	面	免	茂	毛	耗	網	木	黙	匁	朴	
	ヤ	ャ	夜	野	矢	役	約	訳	薬	躍	厄	綿	模	盲	猛	門	問	紋	戻	猶	
	ユ	ュ	油	愉	諭	輸	唯	友	右	由	有	勇	遊	郵	幽	雄	裕	憂	優	誘	
	ヨ	ョ	余	予	預	与	誉	用	幼	葉	養	庸	曜	羊	容	陽	要	窯	揺	融	
	ラ	裸	来	雷	頼	落	絡	酪	卵	乱	欄	濫	覧	洋	溶	揚	腰	踊	謡	擁	
	リ	里	理	裏	立	律	率	粒	隆	柳	了	良	両	料	涼	浴	欲	抑	翌	翼	
	吏	利	痢	履	離	陸	略	留	流	硫	量	糧	寮	僚	療	力	緑	林	臨	隣	
	ル	類	涙	累	塁	礼	例	励	麗	旅	虜	慮	領	猟	陵	漁	厘	倫	輪	竜	
	レ	令	冷	零	鈴	齢	霊	隷	歴	暦	列	烈	裂	劣	連	練	錬	廉	恋	湾	
	ロ	ワ	路	露	炉	老	労	浪	朗	郎	廊	楼	漏	録	論	話	賄	惑	腕		

図3．6　NEC C—5110　漢字入力装置の文字盤

12ビットの計算機コードを出す。文字スイッチには，EDST(elastic diaphram switch technology)という信頼性の高い長期間の使用に耐える新技術を採用している。[注15)，注16)] つぎに日立製作所が発表した磁気ペン方式漢字入力装置について述べる。[注9)，注17)] 図3.7にその原理図を示す。この装置は，multi-scan F.S.S. による漢字ディスプレイの入力部として開発されたもので，2,400種の文字の入力ができるにもかかわらず大きさは52 W × 41 D × 14 Hcmと小さく，文字表はA 3版程度で小型である。原理は，ペンの先につけられた小型永久磁石が，入力しようとする文字表中の字に近づけると，図のように配置されたリードスイッチのONにより，トランス型ROMよりコードを発生するものである。磁気ペンが文字表をなでただけでは入力しないようになっている。表3.6にこれらの各種の入力装置の代表例を示す。

これまで述べたテーブル・ルックアップ方式の入力装置はすべて平面的な文字盤または文字表によ

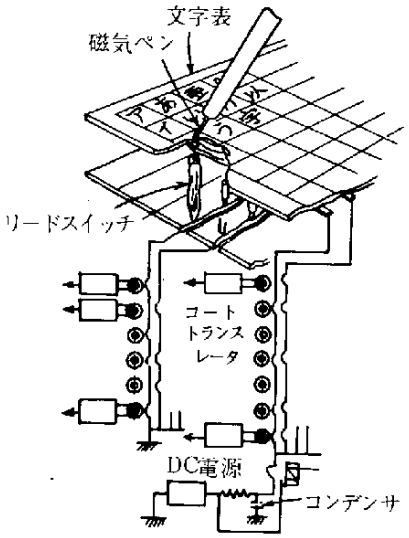


図3．7　磁気ペン入力装置の原理図

っているが，前述のドラム式タイプライタの概念によるものもある。これは沖電気から発表されたもので，和文タイプライタの変形ではないため，モニタ印字は行なわない。操作は漢字の表をドラムに巻きつけて，左手でハンドルホイルを回転しながら左右に移動し，所定のところへ指針を持ってきて，右手でユニバーサルキーを押すとコードが発生する。テーブル・ルックアップ方式とフルキーボードとの比較は困難ではあるが概念的には次のことが考えられる。

1）入力装置をオペレータ専用にするか，素人にも向くようにするかによって装置の設計思想が変わるが，日本の社会事情から見て，オペレータ専用とならざるを得ない。

2）10数段シフトにもなるとキーボードの大きさは，文字表の大きさに近づき記憶にも便利となるが，シフトのため次の項で述べるマルチストローク方式に近くなる。

3）原稿を見ないで打鍵することを考えると，テーブル・ルックアップ方式はまったく困難である。目が原稿と文字表の間を往復する。

4）フルキーボード方式の場合は両手が使える。

メーカ名	日本電気(株)	日本アイ・ビー・エム(株)	(株)日立製作所
機種名	C-5110漢字入力装置	(試作)	(試作)
文字キーの数	2205	3520	600
シフトキーの数	0	0	4
文字キー当りの文字数	1文字/キー	1	4
文字種	2,145	>3520	2,400
漢字	1,878	320種頻度大なるもの 1,600 2区画	1,900
ひらがな	75		81
カタカナ	75		81
英数字	82(数字 10種×3組)		62
記号	35		115
特殊文字	0		161
その他			
外字	60(文字盤面に固定的にセット)		
入力可能文字種	技術的にはほぼ無制限		
記録媒体	紙テープ	カード	紙テープ
コード パリティの有無	8bit 有		12bit
モニター	タイプ印書(標準に装備)		
打鍵速度(MAX)	75文字/分		
外字インプット方法 外字インプット方法	上記外字は打鍵入力, それ以外はテンキーにより コード番号を数字で入力		
使用環境			
温度	5°C～35°C		
湿度	45%～85%		
備考	電子タブレットによる 文字表座標入力	elastic diaphram switch による座標入力	リードスイッチによる 座標およびシフトによる入力
価格 買取 万円	160		
レンタル 万円/月			
発売年月	昭和47年3月		
販売	日本電気(株)	日本アイ・ビー・エム(株)	(株)日立製作所

表3．6 日本語情報入力装置(その2，テーブル・ルック・アップ方式)

(4) マルチストローク方式

ア．カナ文字入力

この方式は，漢字かなまじり文を見ながらカナタイプでカナ文字としてインプットし，計算機の中で漢字かなまじり文に変換しようという方式である。この方式はいかに簡単な装置で容易にインプットするか，および計算機の中でいかに正しく，効率よく，変換するかである。一般的には次の方法が考えられる。

1）漢字かなまじり文をそのままタイプする。

2）一般の分かち書き程度の手を加えて入力する。

3）漢字で書かれている部分に特定の記号をつけるとか，音訓で注釈を付けるとかして，漢字への変換を容易にしてやる。

4）計算機と対話を行ないながら，カナ入力のあいまいさを避ける。

5）漢字読み音どおりに入力し，あいまいさを避ける追加の情報を入力する。

カナ文字入力の主体となるものは前述のとおり、カナ文字の入力列と文脈により，なるべく正しい日

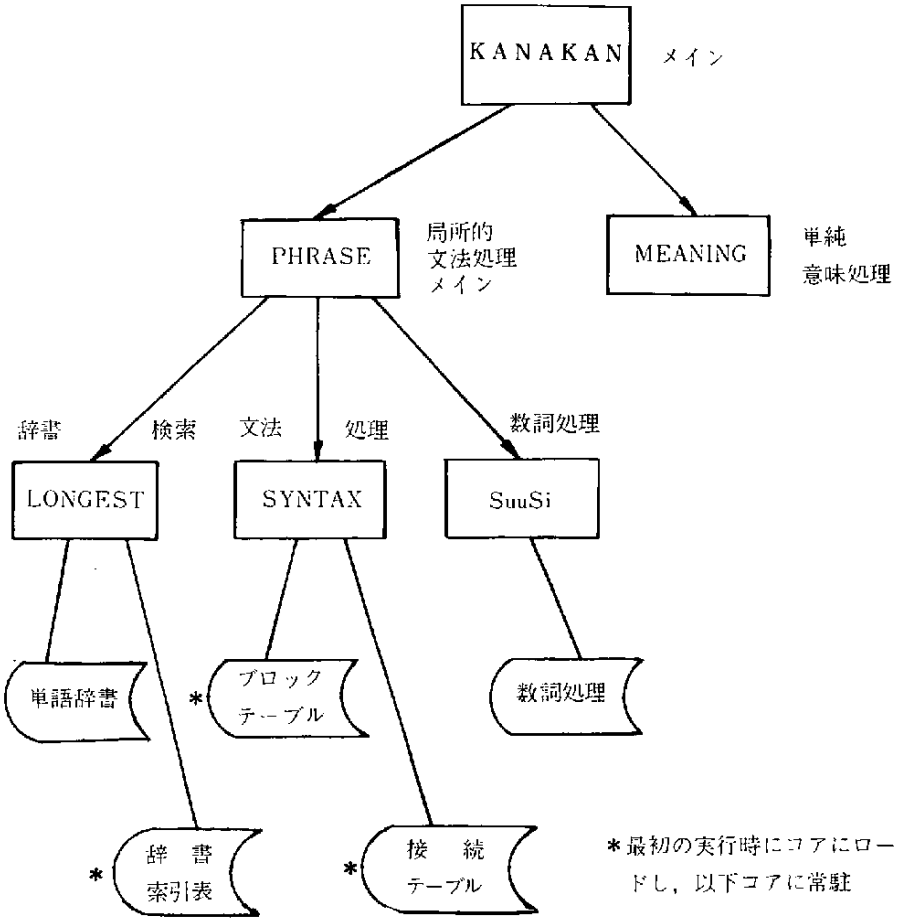


図3．8　カナ漢字変換部のプログラム構成 (NHK)

本文に直すことであって，古くから多くの研究が続けられ，[注18)，注19)，注20)，注21)，注22)，注23)] 成果が上っている。図3.8に代表的なプログラム構成を示す。前述5）項のうち，1），2），3）はひと口にいうとカナ文字列の日本文としての自動認識で理想とされているところであるが，漢字への変換率，入力に要する速度，計算機費用などの問題が考えられる。その一つの解決法として考えられるのが4）項である。日立製作所が発表[注24)] したものは，対語において，フィード・バックとして次の情報を使用している。

1）3字目要求

2）音要求

3）部首要求

4）特殊要求（つくり，濁点など）

図3.9にけん盤の配列を示す。対語に要するメモリ総量は約12万ビットで，半導体メモリの使用により価格を下げうるとされている。

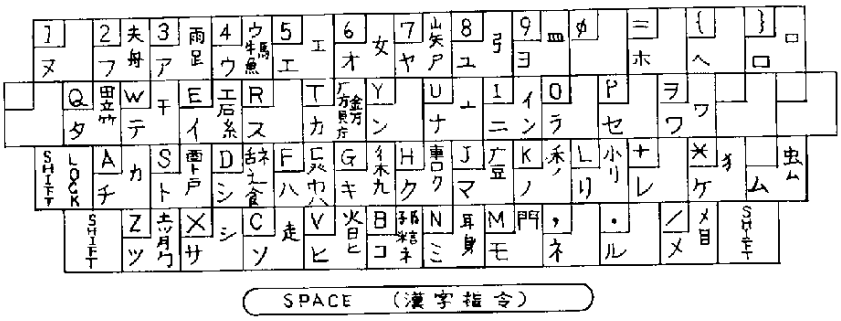


図3．9　対話型カナ文字けん盤の配列（日立）

次に東芝の発表したもので，日本語1字を入力するのに必ず2打けんするという原理の下に，ラインプット入力装置なる従来とまったく異なる鍵盤配列を持つ入力機器と，2週間から3カ月間の訓練で記憶される入力の約束の組合わせで行なわれる方式がある。この方式は6章に詳述されている。図3.10（＝図6.9.2）にラインプット入力装置のけん盤配列を示す。

イ．文字分解ストローク型

漢字を分解し構成要素のストロークに分けて入力する方法であるが，漢字の読み書きになじみの少ない米国で主として研究が行なわれている　図3.11にＲＣＡで開発された Ideographic Composing Machine のブロックダイヤグラムを示す。[注25)] この装置はほんらいは Composing のために開発されたもので，選択された21個のストロークと，20個の完全なシンボルによってフレキソライタから入力する。その後さらにＲＣＡから円板の円周上に配置された多数のホログラムによる Composing machine が発表され，[注26)] 同じ入力法を使っている　その他，米国ＩＴＥＫ社で開発された Chicoder，[注27] ＩＢＭと Mergentheler 社が開発した Sinowriter[注28)] がある。これは漢字を“へん”“つくり”などに分解し，図3.12のようなキートップを持つＦＬＥＸＯＷＲＩＴＥＲを使用し，3ストロークで入力が行なわれ，オペレータの前面にあるスクリーンに，同じ字画を持つ漢字のファミリーの所望の一つを2ストロークの打けんで指示する。この方式は，漢字のなじみの深いわれわれにとっては回り

くどい入力の方法といわざるを得ない。国内ではこの種の研究，試みは見られない。

ウ．ファミリー分類

ここでいうファミリーとは，音訓によるものと，写真植字のように字画にしたがったグループをい

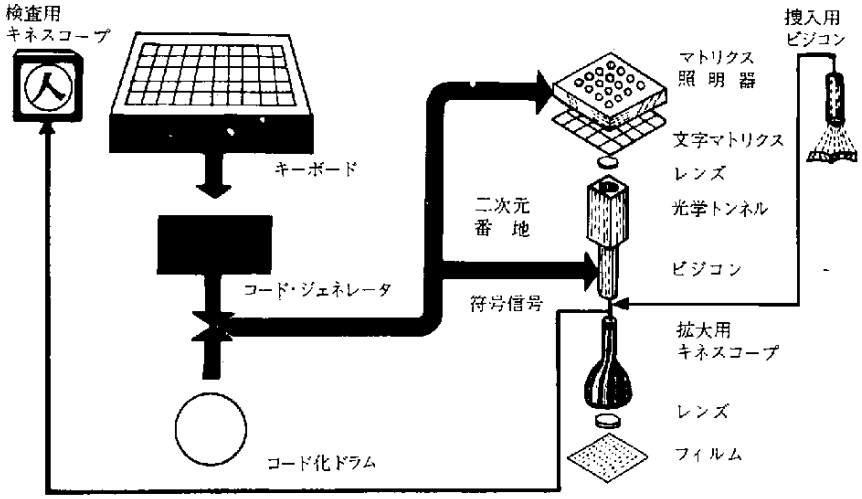


図3．11　Ideographic Composing Machine

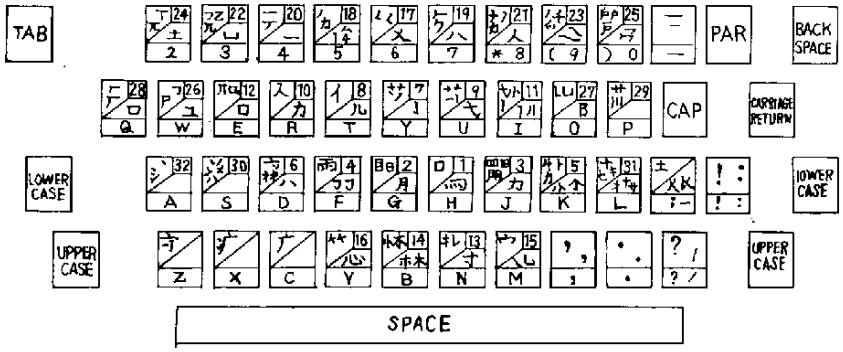


図3．12　Chicoder Sinowriter のけん盤配列

う。後者の方は分解して考えずにその字画をもつ一字として考える。入力機器としては，ほとんどがカナ文字タイプライタを使用するが，前途のカナ文字入力と異なるところは下記の入力サイクルを行なう点である。

1）漢字の読みにしたがったカナ入力を行なう。

2）その入力に対応するファミリーがオペレータに表示される。

3）その中から入力希望の文字をなんらかの手段で指定する。

4）指定された特定文字のコードを出力する。

次に工業技術院電子技術総合研究所で開発した漢字入力装置について述べる。図3.13にその原理を示す。[注29] 直径14cmの円板のまわりに180枚のマイクロフィルムが配置され，打けんされた漢字のファミリーが光源の前に止まり50倍に拡大されて漢字表示盤に投影される。表示盤上で所望の文字

が投影されたキーを押すとコードが発生する。表示数は33文字である。同じような原理のものが，電電公社武蔵野電気通信研究所からも発表されている。[注30] この他に，ＩＢＭが発表した2760光学映像装置も漢字入力装置の機能をそなえている。この装置は電算機の知識のない人も，指示ペンを使って会話モードスクリーンに映る映像の助けにより会話モードで使用できる。映像はフィルムカートリッジで発生されるが，ファミリー分類型として使える。（写真3.3）[注31]

(5) 人手介在音声入力（速記入力方式）速記用のけん盤装置を漢字入力装置として使用する試みは内外で行なわれているが，独特のけん盤を持ち，音声を聞いて両手で必要なだけ同時に打けんしうる。

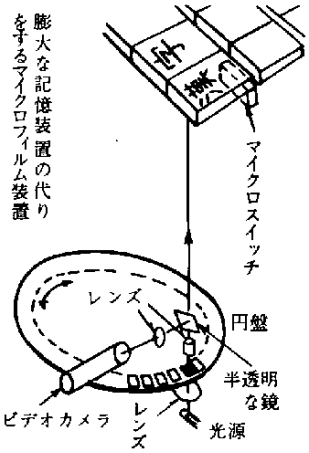


図3.13 電総研漢字入力装置の原理図

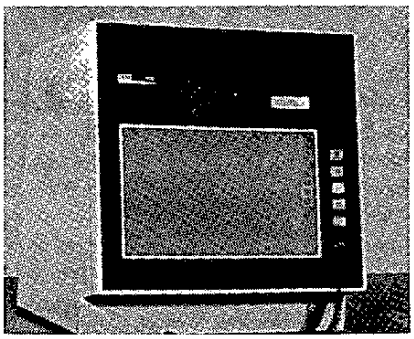


写真3.3 IBM2760 光学映像装置

装置は音がないのが特徴で，日本語の速記の場合はローマ字表に基づいて，母音と子音の略号を決め速記記号を構成している。オペレータは中枢神経，運動神経を最大限に利用し，音声をリアルタイムで図3.14のような出力印字テープに記録する。従来はこのテープを記号を解読しうる速記者が平常文に直し（反訳）ていたが，電算機で自動反訳が試みられた。これは最高裁判所で行なった例であるが，反訳に要する速記時の6～10倍の所要時間の問題が解決された。図3.15にシステム概要を示す。反訳のアルゴリズムは次の機能からなり立っている。[注9]

1）速記符号を対応する訳に反訳し，分かち書きされてない仮名文字列に変換する。

2）連糸辞書を探索し，単位語認定処理を行なって，漢字化，カタカナ化，数詞化をへて漢字仮

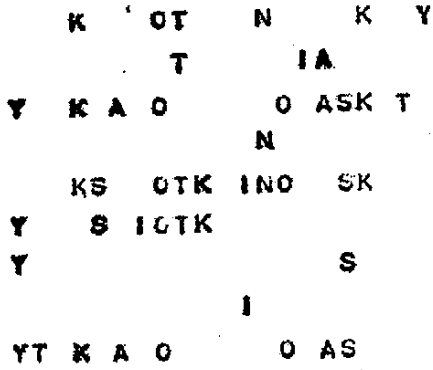


図3.14 ソクタイプ出力印字テープ

名混じり文に変換する。最長の一致法を用いて誤りを少なくしている。

3）同音異義語のうち，前後の環境で一意に決定できる語は漢字を選択している。

4）日本語文章を速記録の様式に編集して漢字プリンタに出力する

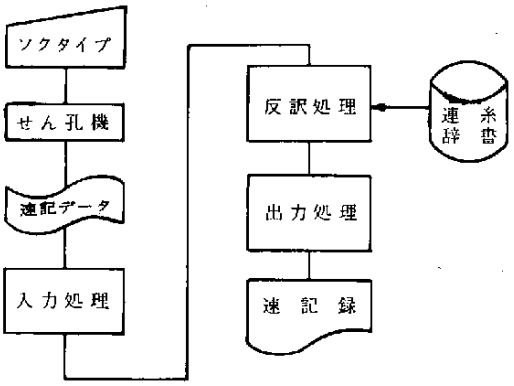


図3.15　ソクタイプ速記自動反訳システム処理概要（日立）

以上，実時間の音声を取扱うので，入手介入音声入力として分類した。

(6)　オンライン手書き漢字入力方式

1966 ～ 1967 年頃から研究が発表された。[注32] これは米国の The RAND Corporation の G.F. Groner などによるもので，ＣＲＴディスプレイ，ランドタブレットシステムを使って，リアルタイムで手書き漢字を認識することを試みた。この方法では，タブレット上に漢字を書くと，それがＣＲＴ上に最大 21 文字類似文字が表示され，そのうちから正しい字を選択する方法をとっている。日立製作所が発表したものでは，認識できる文字は，楷書体の教育漢字 881 字と濁点を含むカタカナ文字の計 952 字である。入力装置で楷書体を書くと次のステップで認識が行なわれる。[注33]

1）入力装置は筆点座標入力により筆点の座標が刻々計算機に送られる。

2）送り込まれたデータからプログラムによりまずストロークを分離抽出して，表 3.7 の 11 種のどの基本ストロークに相当するか決定する。

3）画数は入力装置のペンの圧着回数を数え，基本ストロークの内の特定のものの存在と，その個数の組み合わせによって表 3.7（b）のように大分類する。一義的に決らないものは文字の特徴をさらに詳細にたどって，最終決定を行なう。

4）カタカナと漢字とは，非常に形が似ているので，図 3.16 のごとく付加情報を特定の領域に書き込む。

5）全文字の書き終りを合図すると認識が開始され，平均 250 m s で終了し結果を表示する。

3.2.3　入手を介しないエントリー方式

(1)印刷漢字認識入力

日立製作所の発表した研究結果について述べる。漢字認識として次の二つの方式が考えられ実験が

図1　文字記入枠（Fig.1）

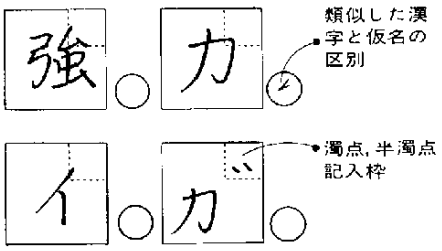


図3．16　文字記入枠と特定領域の記入

(a)　基本ストローク(Table 1)

コード		形　状
単純ストローク	A	一 ノ
	B	丨 亅
	C	丿 ㇀
	D	㇏ ㇔
複合ストローク	P	㇇ ㇇ ㇇ ㇇
	Q	乙 乙 乙
	R	了 了
	S	㇜ ㇟ ㇟ ㇟
	T	㇉
	U	㇋
	V	㇋

(b)　分類例(Table 2)

3	none	工土士才上下大川千三 エオキケサシチツテヲ
4	P	支日月中内方欠友反戸 今分水ネ
6	PS	死池地両安西
9	QO	飛
12	SSP	結 給
15	PP	輪 敵 課 論 諸 確

(c)　類似文字(Table 3)

仮名	漢字	仮名	漢字	仮名	漢字
ア	刀	タ	夕	ミ	三
エ	工	チ	千	ロ	口
オ	才	ナ	十		
力	力	ニ	二		
シ	三	ハ	八		

表3．7　基本ストロークと文字分類

行なわれた。[注34]

1）　Pattern Information Reduction

図3.17のごとく漢字をX，Y軸の方向に投影して，濃度，漢字パターンの空間周波数を分析して認識する。

2）　Macro to Micro Multistage Processing

漢字を二次元空間フィルタロジックを通して認識する。すなわち，図3.18のごとく4×4，16×16，32×32，と Viewing mesh で階層的に行ない，各階層で標準パターンとマッチングテストをやって認識する。

実験結果は881字の認識が可能であった。

(2)　音声入力

一般的な音声入力または認識は，学問的な興味から数多く研究されているけれども，実用を目的として研究されるには至ってないのが現状である。音声で入力できるとなれば，システム的にまったく異なる発想ができる。しかし実用的には現在の技術からはほとんど不可能のように見える。その困難さの原因は次のようなものであるといわれている。[注35]

1）話者の変化による音声の特長の広範な変化

2）音声の内容（音素環境）の変化による音声の特長の広範な変化

本項では，音声認識の内外の現状を表3.8に示すにとどめる。

以上日本語情報処理の入力システムについてソフト，ハードウエアの面からシステムについて述べてきたが，これらをまとめるとつぎのようなことがいえよう。

1）操作性，システムおよび機器の価格についていまだに英数字システムにおよばない。

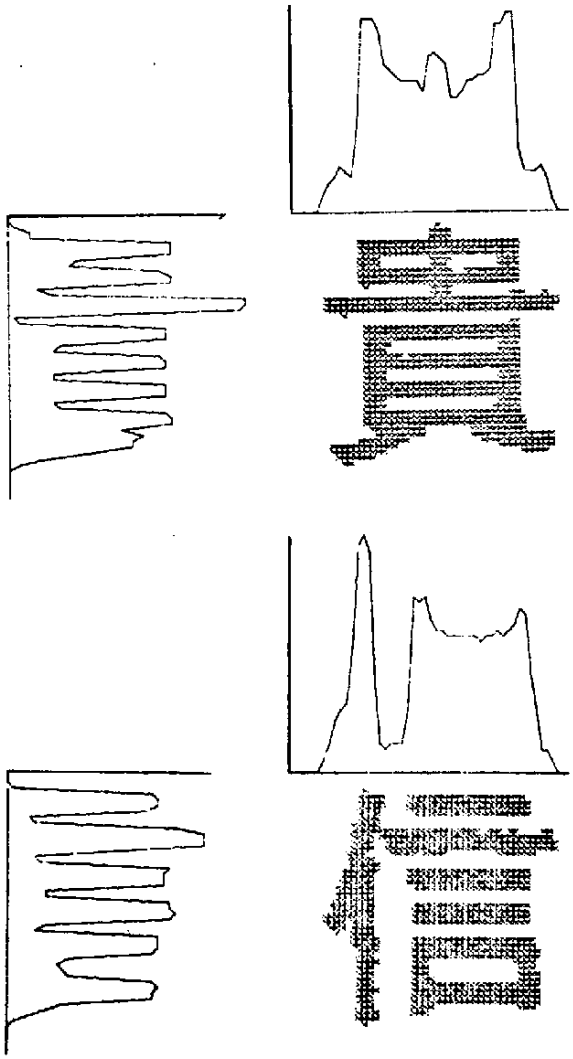


図3.17 Example of Projected Patterns on X and Y axes.

2）日本語入力機器として新規のものがまだ考えられる段階で，いわゆる"枯れた"機器が出てこない。

いずれも，つきつめて考えて見ると表意文字の問題であることがわかる。さいわいわれわれは表意文字使用民族のうちでも，中国と異なり，カナ文字がある程度問題を緩衝してくれている。　カナ文字を逐行しようとして，国際性を破ったけん盤配列まで出現した。カナ文字文から，漢字まじり文へ

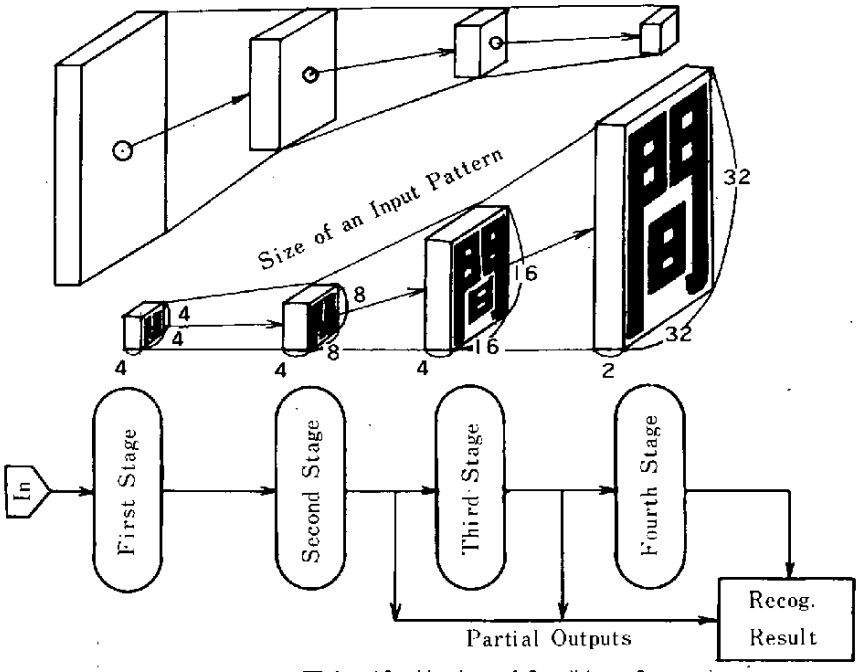


図 3．18　Number of Candidate Categories

項目 研究機関	語　　彙	話　者	パラメータ	認識率	目　　的	備　　考
電電公社通研	1桁数字音	男　24名	kパラメータ	99.8％		最尤推定，学習
KDD研究所	1桁数字音	男　10名	18チャネルスペクトル	誤：なし Reject：3％		書き替え規則の応用
NHK総合技研	5　母　音	不特定多数	10チャネルスペクトル	96％以上		主要因分析
日本電気中研	限定FORTRAN (約50語)	男　2名	8チャネルスペクトル	99.6％	音声による計算機入力	DPの利用＊ ほぼ実時間
富士通	制御語（約10語）	不特定多数	10チャネルスペクトル		機械の制御	万　博　用
日立中研	1桁数字音	男　1名	ケプストラム kパラメータ	99.8％		
東北大通研	都市名（13個）	男　1名	4チャネルスペクトル	96％	純　研　究	単語辞書の利用．
京大工学部	1桁数字音	男　16名	18チャネルスペクトル	91～94％	純　研　究	シンボル系列の処理
ベル電話研	1桁数字音	不特定多数	音声エネルギー	100％	自動ダイヤル	特殊な方法＊＊
IBM研究所	1桁数字音	男　11名 女　2名	8チャネルスペクトル	初回：87.9％ 訓練：96.4％		パターンマッチ
RCA	数　字　音				小包区分機の制御	現在中止
ソ連科学アカデミー	ALGOL60用 （203語）	男　2名	4チャネルスペクトル	95.2％	計算機入力	ノボシビリスク， DP使用
スタンフォード大学	192の限定文章	男　2名	3チャネルスペクトル と零交差回数	～85％	ロボットの制御	

＊個人別に認識

＊＊厳密には音声認識とはいえない

表 3.8　音声認識研究の内外の現状

の変換も集積回路と，メモリの進歩によって異なる局面にきているように思われる。このような時代にあって，カナ文字と，世界の標準語である英語との組合わせで，合理的な入力システムの出現が期待できないだろうか。

注 1．高橋達郎：日本語の機械処理，東洋経済新報社 p.173～188.

注 2．長井　坦：漢字入出力装置の動向と技術的問題点，情報処理， SEPT. 1969 p.320～332.

注 3．小沢暢夫：インプット研究，日本経営出版会 p.467～496.

注4．不破，西尾：文献速報自動作成システムについて，情報管理， SEPT. 1969 p.1～27.

注5．高橋義毅：ＩＢＭ漢字情報処理システム，ビジネス・コミニケーション '72 VOL.9 NO.10 p.58～62.

注6．高橋達郎：日本科学技術情報センター，5）と同誌 p.52

注7． JICST 電子計算機室：漢字インプット・データのデイスプレイ装置を用いた校正実験総合報告書

注8．吹抜他：あいまいさ種別のフィードバックを有するカナ鍵盤漢字入力装置，電子通信学会，電子計算機研究会資料 NO.EC 72－5(1972－05)

注9．鎌田素明：日立の日本語情報処理，5）と同誌 p.72～75.

注10．清正　清：行政システムにおける漢字処理システム，5）と同誌 p.45.

注11．石橋正治：言語からみたコンピュータと人間のインタフエース，オートメーション，第15巻，第7号， p.45～52.21

注12．安田寿明：わが国の新聞社における漢字情報処理，情報処理， SEPT. 1969 p.340～347

注13．大倉信治：日本電気の漢字情報処理システム，5）と同誌 p.65～68.

注14．同　　上：日本電気に於ける漢字情報処理システム，第2回I/O研究会資料，Ⅱ－3漢字情報処理システムのハードウエア， p.70～78.

注15． M. Krakinowski 他： Depressible Diaphram Overlay Switch for Display. U.S.PAT. 3,560,675, patented Feb. 2 1971.

注16．松本文雄：ダイヤフラムキーボードスイッチ，電子材料1972年8月 p.101～105.

注17．長井，村井：漢字ディスプレイについて，昭和46年度電子通信学会全国大会予稿， S2－6 p.115～116.

注18．黒崎悦明：カナ鍵盤漢字表示方式，昭和43年度，情報処理学会大会予稿， p.117.

注19．黒崎他：カナ文の漢字まじり文への変換，43年度連大予稿

注20． H.Clnanaga：Transformation of phonetic sentences into ideographic sentence in Japanese, 日米科学ゼミナー 1971年1月予稿 p.264～275.

注21．相沢輝昭：計算機によるカナ・漢字変換，情報処理における日本語（漢字）の取扱い，p.1～11. 日本経営科学協会ゼミナ

注22．斉藤他：漢字の機械への入力，情報処理， SEPT 1969 p.294～298.

注23．保原他：和文入力に関する一考案，情報処理， FEB 1970 p.108～111.

注24．吹抜他：あいまいさ種別のフィードバックを有するカナ鍵盤漢字入力装置，電子計算機研究会資料C 72－5

注25． Pamphlet：Chinese Ideographic Photo-composer RCA

注26． W.F.Heagerty：Ideographic composing machine 20) と同稿 p.73～93.

注27． ITEK：Annual Report 1964.

注28．桜井宣隆：情報検索用機械，情報処理， Nov. 1966, P.346,

注 29．朝日新聞：かなのキーで漢字もＯＫ，電算機に新しい入力装置，46 年 7 月 8 日 10 版

注 30．日刊工業新聞：3355 字を収容，電電武蔵野通研が漢字入力装置試作 47．12．4．

注 31．IBM：System Reference Library, IBM 2760, Optical Image Unit Component Description.

注 32．G.F.Groner他：On line computer classification of handwritter chinese character as a translation aid[注20] と同文献 p.53～72.

注 33．日立パンフレット：漢字・カタカナ，オンライン手書き文字認識装置，p.4～5. 日米コンピュータ会議展示会，

注 34．K.Nakata others：Problems in Chinese Character Recognition, First USA-JAPAN Computer Conference 1972 p.296～301

注 35．中田和男：音声の認識と発生，情報処理，APRIL 1972 p.247～255.

注 36．倉石武四郎：漢字の運命，岩波新書

4．日本語情報処理の出力システム

4. 日本語情報処理の出力システム

電子計算機を主役とした技術革新により今日社会全般が情報化社会へと飛躍的に発展しつつあるが，さらにデータ通信網の利用により広く情報が容易にえられるようになると，日本語による情報の処理および入出力に対する必要性はますます高まってきた。電子計算機システムそのものが米国において開発され，英数字情報の処理に適したソフトウエアおよび入出力装置のハードウェアが設計され，そのままの形でわが国において広く利用されてきたことと，英数字情報に比べ日本語を構成する文字が多種であるとともに，印刷編集が複雑であるなどの日本語情報の特殊性により，その出力システムのソフトウェア，ハードウェアの開発および利用が技術的，経済的な面で遅れをもたらしたためである。日本語情報処理の出力システムの処理内容については次の三つのレベルにわけて考えることができる。

(1)データ処理のレベル

(2)情報編集処理のレベル

(3)言語処理のレベル

(1)のデータ処理のレベルはラインプリンタやタイプライタによって英数字がデータとして出力処理されると同じように，データを漢字でプリントするレベルをいい，この場合漢字は文字としてまた文章言葉としての処理はされていない。このレベルでの出力システムは漢字処理として最もふるく，昭和30年頃から機械印字方式の漢字テレタイプライタが受信端末機として新聞社を中心に現在に至るまで長く使われてきた。昭和42～43年にかけて日本電子産業・日立製作所の両社から電子式の高速漢字プリンタが発表され，昭和45～47年にかけて富士通・高千穂交易・日本電気・沖電気などの大手各社からも漢字ラインプリンタが発表され，このレベルでの日本語情報処理の出力システムは文字種，コード，字形の標準化などの問題を除いて実用期に入ってきた。

(2)の情報編集処理のレベルは漢字をデータとしてではなく意味のある文字として処理するレベルをいい，文字を集合体としての文章，組版，画像などとしてとらえ，文字の大きさ，組み方向，書体，字詰め，禁則などの行編集処理，ページ割り付け処理などの日本語情報としての編集処理ソフトウェアおよびハードウェアとして電算植字装置などを必要とする。これらの電算植字装置は機械式のものとして昭和44～45年に富士通，写研から，高速電子式のものは昭和42～43年頃に日本電子産業，RCA，IBM から発表され実用化されてきた。

(3)の言語処理のレベルは漢字の文字としての問題の他に日本語としての表記法の複雑さ，および音声処理の技術的な困難さなどから研究段階のものが多く，これからのレベルのものであり，カナ漢字変換，音声入出力，和欧文翻訳，情報検索などがこのレベルの処理システムに含まれる。

4.1 文字

日本語情報の出力システムの大きな問題点は日本語情報を構成する文字の複雑さにある。

まず日本語の情報は多種文字からなる漢字とカナなどの和文情報と英数字からなる欧文情報が混在して用いられており，日本語の情報処理は世界の文字情報処理について考えなければならないという点にある。また，その漢字のもつ文字種の多さ，複雑さなどの特殊性から文字種，字体，コード体系，伝送処理などの標準化がおくれており，発展する情報処理システムの必要性に対応できなくなるおそれもある。したがって従来文字が文学的，教育的，社会的な面から論じられてきたが，これらの問題点について広く情報工学的な面からも考慮しなければならない時期にきていると考えられる。

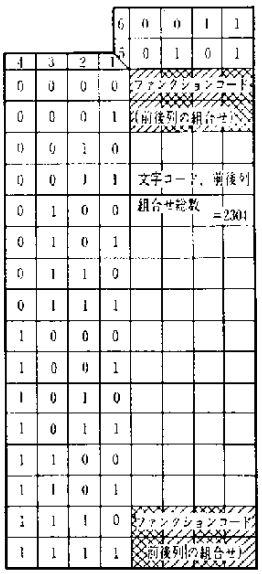
4.1.1　文字種および文字コード

日本語情報で一般に用いられている文字種の数は応用分野によってまちまちであるが，その概略の値を表4.1に示す。それぞれの分野において文字の出現頻度もまたまちまちで，平均した値として示すことはむずかしいようであるが，およそ2,000字種で99％近くの値を示すといわれている。

これらの多種文字からなる日本語情報の文字システム内に用いられている文字や記号を含めた文字

分　　野	使用文字数（字）
新聞報道記事等	約2,500
一般の書籍雑誌等	約4,500
日本の地域名（都道府県市町村大字）	4,300
氏名等を含む情報および辞書百科事典など	8,000 ～ 10,000

表4.1 分野別使用文字数



▩ のファンクションを前列にして後列文字コードの組合せで文字コードにし得る。

図4．1　CO—59コード表

群もまた多種多様であり，外国文字がかなり頻度の高い割合で用いられており，全体の文字種を5,000字とした印刷分野における文字群の一例を表4.2に示す。

日本語情報の文字の中心となる漢字に対しての文字種の標準化の例は少なく，したがって漢字文字コードの標準化が遅れており，日本語情報処理の発展の大きなネックとなっているが，国および団体としてのいくつかの標準について表4.3に示す。

この文字種の標準化とあわせて漢字文字コードの標準化が必要となってきており，JIS化の努力が情報処理学会を中心としておこなわれているがいまだ成案をみるに至っていない。

漢字文字コードの標準として有名なものは昭和34年に共同通信社を中心として各新聞社との間で作られたCO-59コードのみであり，これは図4.1に示されるように6ビット2列の計12ビットで1字を表わすコード体系をとっている。

現在，各種の漢字プリンタが発表されているが，これらはほとんど8ビット2列の計16ビットで

文字群	文字数（字）
漢字	4,200
ひらがな	99
カタカナ	97
数字	30
英字	52
ギリシヤ	49
約物記号	199
予備	274

表4.2　文字種5,000字のハードプリンタの文字構成の一例

種類	文字数	審議機関	制定時期
当用漢字	1,850	国語審議会	昭和21年 注1)
人名用漢字	92	〃	〃 26年 注1)
当用漢字補正案	29	〃	〃 29年　注2)
常用漢字	4,000	全日本漢字配列協議会	〃 43年
標準コード用漢字（試案）	6,100	情報処理学会	〃 46年

注1.）　いずれも内閣告示による制定。

2.）　当用漢字から28字を削り，新しく28字を加える。別に音訓を加える字1，字体を改め音訓を加える字1字がある。なお，これは国語審議会の第20回総会の審議報告であり，これによって当用漢字そのものが変更されたわけではない。

表4.3　おもな漢字標準案

1字を表わしているが，そのコード体系の内容は統一がとれていない。

漢字文字コード化上のシステマチックな問題点としては，文字種を限定することにより将来の文字種の改廃に対する冗長性の問題と，文字の分類上のウエイトづけに部首順（康熙字典式，新部首式）音訓順，画数順などが応用分野によっていろいろ用いられ，一定の基準がないことなどがあげられる。

4.1.2　字体

漢字の特殊性は文字種の多さとともにその字体の複雑さにある。昭和21年に国語審議会から1,850字の当用漢字が発表され，ひきつづき昭和24年に当用漢字字体表が発表され，字体が確実に決められ統一がとられてきたが，それ以外の字体については規準がなく必要に応じていくつかの異体字が用いられている。表4.4に字体の分類を示す。

当用漢字字体表の発表により旧字体から新字体へと多くの簡略化がなされ，一般に画数が減少し，とくに横線の多かった字も是正された。その簡略化の種類を表4.5に示す。

分　類	内　容	例
新　字	当用漢字制定時に字体が新しくなったもの	剣　挙（恪）（恭）
旧　字	当用漢字制定時に新字ができた旧字体	劍　舉
本　字	従来正字形として認められているもの	
古　字	説文解字などに所載の古文を楷書形にしたもの	劒
別体字	ある体のもの，または異なったしくみをもつが同音同義の字とすべきもの	
俗　字	本字がくずれた形で通用しているもの	劔　舉
誤　字	部分的に通用しているが誤まった字形で使用してはいけないもの	舉

表4.4　字体の分類の一例

当用漢字以外の文字の字体については規準がなく，字体の選定も各種各様におこなわれ混乱をおこしている。オーソドックスな考え方としては当用漢字は新字体を用い，当用漢字以外の文字については新字体でとられた簡略化をしない字体を用いる。この場合，当用漢字と当用漢字外との間でたとえば，「花」と「芯」，「辺」と「辻」のように同じ部首のパターンが異なってくる。このようなオーソドックスの考え方に対し，最近は当用漢字に採用された簡略化を当用漢字外の文字の字体に対しても採用する動きが多くなってきた。しかしこの場合でもその簡略化の範囲は表4.5の簡略化の種類のうち，1と2に属する簡略化は一般に受入れられても，表に示すようにその他の簡略化に対しては使用者の使用に対する抵抗が大きい。

したがって，一般に新字と旧字，あるいは正字と異体字が同一の文字について要求される場合があるので，文字種が字体の点からも増加する傾向となるばかりでなく，文字を出力する場合の選定に規準がなく混乱が増加されている。

また，字体とはいえないが，日本語情報文の特殊性として，たて書き，横書きがあり，カナのよう音ときつ音および句読点などの約物の字体がそれぞれ異なり，これも文字種の増加の一因となっている。普通，たて書きと横書きで用意しなければならない約物の種類は 70 種程度である。

	分　　類	例
1	デザイン上のアクセントを改良したもの	者 ← 者 巻 ← 卷
2	部首の形を簡略化したもの	花 ← 花 辺 ← 邊
3	素子の形を簡略化したもの	仏 ← 佛 独 ← 獨
4	略字体を採用したもの	体 ← 體 会 ← 會

表 4.5　字体簡略化の分類

4.1.3　字形

日本語情報処理の出力システムとして用いられている漢字プリンタの出力文字に対するデザイン上の字形には，モニタプリント用としての簡易型と，一般印刷に用いられている電算植字用の字形とがある。

簡易型はもっぱらドットの形で文字が作られ，その素子数は 15 × 18 ～ 32 × 32 といろいろのものがある。ドット表現の簡易型では細ゴシック調であるが，24 × 24 以上の素子数になると明朝体調のドット文字をデザインすることができる。

電算植字用の文字の書体としては明朝体，ゴシック体，教科書体，新聞書体が一般印刷用とおなじく用いられている。一般に文字種が約 10,000 字種の場合，

明朝体	5,000 字種
ゴシック体	3,000 字種
教科書体	900 字種
その他	1,100 字種

のような文字書体のセットとして用いられる例が多い。またこれらの文字書体は文字の大きさによって明朝体の場合は細線と太線の割合ゴシック体の場合は線の太さをデザイン上変えて用いている場合が多く，したがって電算植字用では使用する文字の大きさの範囲が広くなると，それだけ必要とする文字種が多くなる。

新聞書体の場合は特殊で，本文用の 1 倍文字は 80 %の偏平文字を用い，見出し用の 1.5 倍以上の文字は別にデザインされた普通の正方文字を用いている。電算植字による新聞を発行している日刊ス

ポーツ社の場合は電算植字装置に収容する文字種の増大を防ぐために，正方の1書体のみを収容し，本文用の1倍文字は文字発生装置で偏平にして植字をおこなっている。

なお，わが国におけるプリント文字の大きさには用途によって表4.6に示すような3種類の単位が用いられているが，これらも標準化の必要な対象となっている。

文字のデザインは文字の印字品質を決める大きな要素の一つで，日本の文字は手書き書体から活字書体にかけていくつかのグループがあったが，現在残っている代表的なものにモトヤ書体，岩田書体，日本活字書体がある他，各新聞社専用の書体がある。以上のホットタイプの書体に対し，写真植字用の書体として，石井書体，モリサワ書体がある。

これらの書体は長い間用いられてきたが，この数年来，活字，写真植字共，新しいデザインの書体作りの動きがあり，モトヤ書体を改良したアポロ書体，写真植字用としてまったく新しい明朝体とゴシック書体をあわせたようなタイポス書体などが開発されている。

いずれにしても，書体は印字機構によってその品質が影響されるので，ハードプリンタ，電算植字装置の各機構にあわせて，書体の設計がおこなわれている。

単　位	使用分野	mmとの換算
ポイント	一般印刷活字	0.3518
級	写真植字機	0.25
倍	新　聞	2.235

表4.6　活字の単位

4.1.4　文字品質

英数字にくらべ日本語情報に用いられる漢字の素子数は8～30倍もあり，それだけ出力文字に対する解像力の文字品質が要求される。

一般に出力プリントに要求される文字品質は用途によって表4.7に示される値が必要であるといわれている。

一例として商用印刷用の12ポイントの漢字1字に対しては10,000ビットを必要とし全体で5,000

分　類	解像力	文字寄り引き	濃　度
モニタ	4本／mm以上	1.0mm以下	2.0 ～ 2.5
軽印刷	10　〃	0.2　〃	2.2 ～ 2.5
商用印刷	20　〃	0.1　〃	2.5 ± 0.1

表4.7　文字品質の要求度合

字の場合には50,000,000ビットの大量の文字記憶が必要となる。

文字の印字品質は文字の発生方式，印字方式，記録現像方式，記録媒体などの総合的な性能で決定されるが，可読性などの要素として文字のデザイン設計上の問題は大きく，ハードプリント用，電算

植字用の文字はそれぞれのハードウェアの特性に合うように設計されている。

4.2　文字発生方式

出力システムの代表的なものとしての漢字プリンタの一般的な構成を図4.2に示す。このなかでは文字記憶部と文字アクセス部をあわせた文字発生装置が多種文字からなる複雑な漢字を，いかに速く，安定して，経済的に安価でプリントできるかを決める重要なものである。

その文字発生方式には各種のものが考案され実用化されており，その分類を図4.3に示す。

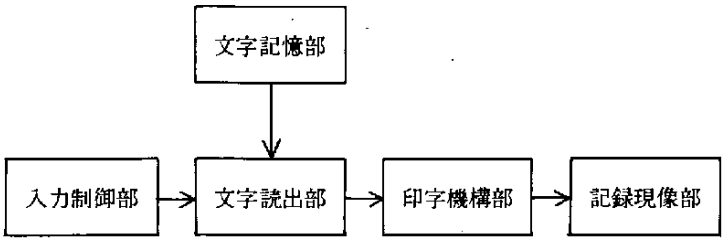


図4．2　プリンタの一般構成のブロック図

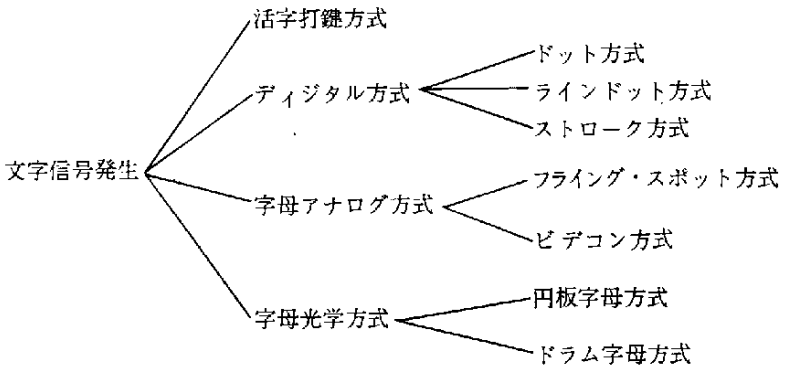


図4．3　文字信号発生方式の分類

4.2.1　活字打けん方式

この方式は漢字プリンタとして最初に用いられ現在に至るまで広く使用されているもので，活字をハンマで打けんする純機械式のものである。

この方式のものとしては沖電気と新興製作所の2機種が有名である。

文字の収容方式に二つあり，沖電気製の古い型のものはタイプパレットに活字を2～4段で24～48列の配列で収容し，活字を差動歯車による加算機構により所定位置に移動させ打けんする。もう一つの方式は活字ホイールにタイプ活字を収容し，常時一定速度で回転させ，ホイール位置を選択し，打けん印字するものである。

この方式のものは収用文字数は2,500字前後で，プリント速度が120字／分程度である。

4.2.2　ドット方式

ドット方式は漢字をドットに分解して，各ドットをビットに対応させてディジタル信号として記憶させ，出力時に記憶部から読み出して出力する方式で，漢字プリンタとしては最も多く用いられている方式である。

漢字1文字を表現するのに必要なドットの数は出力文字品質に要求される解像度によって決まり，表4.8に示すような値が用いられており，ドット文字の一例を図4.4に示す。

一例としてハードプリント用のドット文字の場合，18×18のドット構成では当用漢字では「曇」「警」「響」「襲」の4字は表現できず略体となる他，約4,000字の漢字に対しては約30字の漢字が略体となる。文字のドット構成が24×24以上になるとほとんどの文字に対して略体を使うことなく

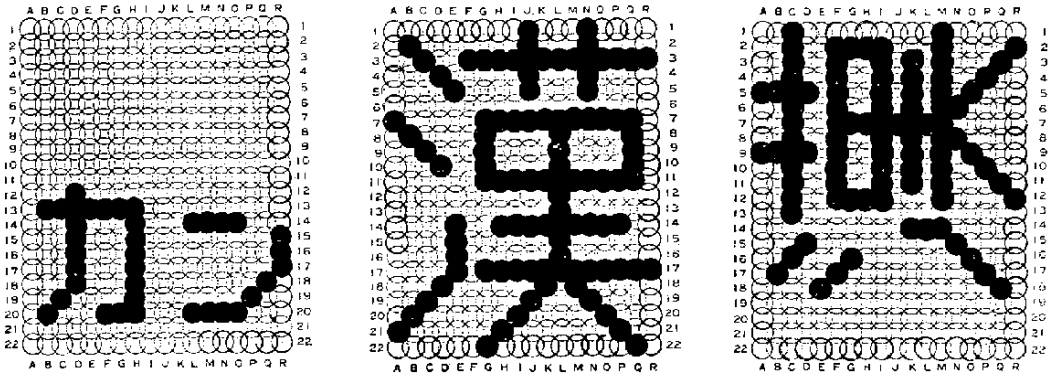


図4．4　英数字(7×9)，たて書き活字(18×22)，横書き活字(18×18)ドット文字の例

分　　野	ド　ッ　ト　構　成	
ハード プリンタ 英　字		5×7 7×9
ハード プリンタ 漢　字	沖	18×16
	日　　電	24×24
	高　千　穂	32×32
	富　士　通	18×15
	I　B　M	18×18(22)
電算植字 漢　字	8ポイント	60×60
	12ポイント	105×105
	(ただし25本/mmのとき)	

表4．8　ドット文字の構成例

表現できるようになる。

ドット方式の場合の文字のドット信号の記憶素子としては磁気コア，ドラム，ディスク， MOS IC メモリなどが用いられている。磁気ドラム，デイスクは大容量で多種文字の記憶素子として適当で1ドット当りの価格も安価であるが，アクセスが遅いので出力速度が100字／秒以下になる。これよりも現在は1ドット当りの価格が高価であるが，高速アクセスが可能で，近い将来の価格が安価になる期待のもてる半導体メモリの MOS IC シフトレジスタや RAM が有望視されている。

ドット方式の文字記憶の方法で特徴のあるものに高千穂交易の T4100 漢字プリンタでとられている方式がある。 T4100 ではプリント品質として1文字当り 32 × 32 ドットであるが，文字の記憶は 16 × 32 ドットの半分の記憶容量で，記憶部から読み出す際に漢字のパターンのもっている特性を利用し，特殊な論理回路によって補正し， 32 × 32 ドットに変換している。

電算植字用のドット文字の場合は表4.8に示すように，1文字で5,000 ～ 10,000 ドットを必要とし，多種文字の場合には記憶部は大容量となるので，ファクシミリの帯域圧縮の方式と同じような文字の線素のドット位置を記憶する方式をとっている。

この方式はラインドット方式ともよばれ RCA 社のビデオコンプでとられた方式である。

一般にドット方式は後述の字母方式などのアナログ処理のものに比べ，ディジタル処理されるので，出力位置（字並び）や濃度の一様性などの特性の安定性にすぐれている。しかし，文字品位を上げるためには，記憶容量が高価になることや，文字の大きさの可変に対してはそれぞれの大きさの文字の記憶を必要とすることなどの欠点があるが，将来性や機器の設計のしやすさから最も有望視されている。

4.2.3　ストローク方式

ストローク方式もドット方式と同じくディジタル方式で，文字の線素の情報のみをディジタル量として記憶している。このストロークの記憶方法の主なるものとして二つの方法がある。第一の方法は文字の各線素に対して，始めの位置と，方向と，長さを記憶するもので，第二の方法は各線素に対して，始めの位置と終りの位置を記憶するものである。

ストローク方式で文字の分解能をドット方式の 32 × 32 ドットに対応させて設計すると，1ストロークに約 16 ビットの記憶量となる。漢字の平均的な字画数は約 12 画で，これをストロークに換算すると約¼の記憶容量ですむことになる。

しかしストローク方式の場合は各文字の記憶容量がそのストローク数によってそれぞれ可変となる他，出力装置に対しストローク信号で動作することが要求され，装置が複雑で高価になる欠点があるので，現在はグラフィック表示機能をもった CRT ディスプレイや， COM 装置などに用いられている。

4.2.4　字母アナログ方式

字母アナログ方式には図4.3に示すように，フライングスポット方式とビデコン方式の二つがある。

フライングスポット方式による文字発生の原理を図4.5に示す。図にみられるように，平板上の写

真乾板に文字を多数配列して撮影した文字マトリックス板を設け，これを高解像度フライングスポット CRT のビーム偏向走査により，必要な文字を選択走査し，文字画像の走査ビデオ信号をうる。この出力信号をフライングスポット CRT の走査信号と同期した CRT 上に表示し，プリントやディスプレイに用いられている。

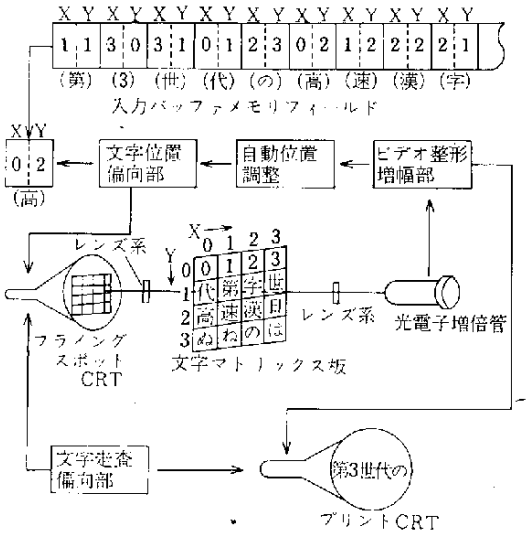


図4.5　フライングスポット方式原理図

フライングスポット方式を用いた実用機としては，日本電子産業のハードプリンタや電算写植装置があり，この方式による文字の解像力は高い。同システムの場合，ひと組のフライングスポット系で，ハードプリント用として，50 × 50 ドット相当の品質で約 10,000 字を，電算植字用として 100 × 100 ドット相当の品質で約 5,000 字を発生させている。

フライングスポット方式の場合，出力文字品質が高いほか，文字のアクセスが速く，したがって出力速度も速い。また，文字発生がアナログ式のため，文字の変形出力（長体，偏平）や，文字の大小出力制御が簡単であり，高速に良質で多種文字の出力をうるのに適している。しかし他の方式に比べ，装置が大きくなり，アナログ式のため，ディジタル方式に比べ，出力文字の字並びや濃度の安定性が劣る欠点を有している。

字母アナログ方式のもう一つのビデコン方式は富士通と電電公社武蔵野通信研究所によって共同開発されたもので，図4.6にそのブロック図を示す。

ビデコン方式では直径 160 mmのプラスチック円板に同心円状に約 5,760 字の文字パターンが配列されている。この円板は約 130 回転／秒で回転しており，フラッシュランプのせん光によりビデコン上に 1 列 15 字の文字パターンを蓄積し，この中から必要な 1 字を選択走査して，文字の走査ビデオ信号をえている。

このビデコン方式はフライングスポット方式に比べ速度は 80 字／秒と遅く，解像度も悪くなるが，偏向系が簡単で，現状では最も価格的に安い方式である。フライングスポット方式に比べ解像度は劣

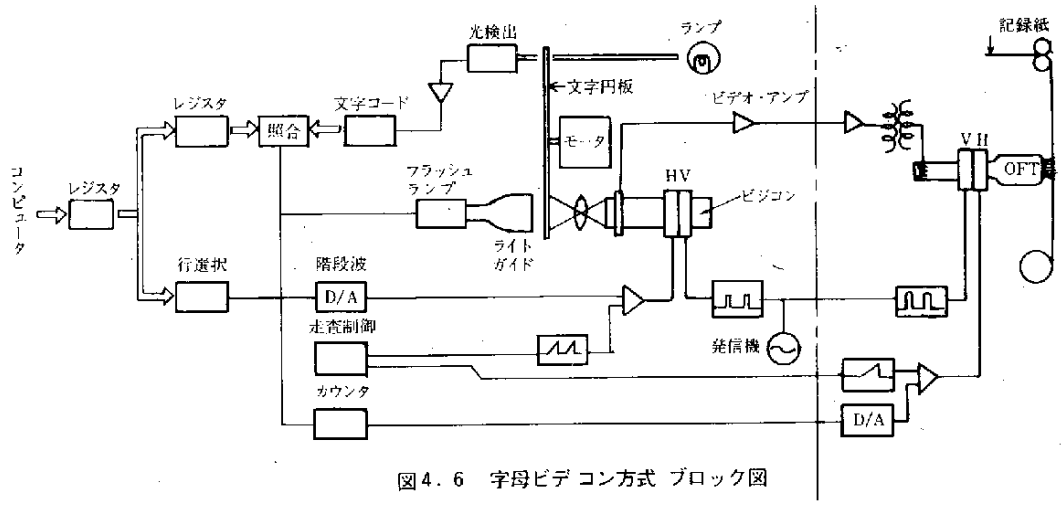


図4．6　字母ビデコン方式 ブロック図

るが，ドット方式の文字に比べ，アナログ式のため，自然な文字が出力され可読性は良い。

4.2.5　字母光学方式

字母光学方式は機械式の写植装置に多く用いられている方式で，富士通，写研のそれぞれのシステムで採用されてをり，その原理図を図4.7，図4.8に示す。

両社の方式とも，回転している文字盤をフラッシュランプで照射し，必要とする文字をレンズ系によりプリント面に撮影する方式であるが，図にみられるように，構造が異なっている。

富士通の方式は回転文字盤にドラムを使用し，1系列に約4,000字を収容し，最大3系列まで組み込めるようになっている。文字はフラッシュランプにより，ロータリミラで選択されて照射され，コリメーションレンズ系を通して感光面に投射される。文字の出力速度は約20字／秒である。

写研の方式は回転文字盤に円板を使用し，1枚に約3,200字を収容し，最大4枚まで組み込めるようになっている。文字はフラッシュランプにより，プレートセレクター，オプチカルトンネルなどの光学系によって選択されて感光面に投射される。文字の出力速度は約8字／秒である。

これらの字母光学式は電子式の写植方式に比べ，出力速度は遅いが，安価で，良好な印字品質をうることができる。

4.2.6　その他の方式

以上の方式は漢字情報の汎用性のある文字発生方式であるが，特定の限られた漢字情報の文字発生方式としては各種のものが実用化されている。たとえば電光ニュース，交通案内表示板，駅の発車行先案内，みどりの窓口における切符情報の発生，定期券等の切符情報の発生等がある。

なお，日本語の文字発生に対しては外国でも多くの開発研究がおこなわれ，実用機としては前述のドット方式の RCA，IBM の電算写植システム，あるいは富士通の字母光学式に似たフォトン社の機械式写植装置がある。その他 Linotron（Mergenthaler Linotype社），Stromberg-Carlson 社，ベル電話研究所などの研究および文献がある。

4.2.7　外字処理

日本語情報処理の出力システムで文字に関連して生ずる大きな問題に外字処理がある。これは収容

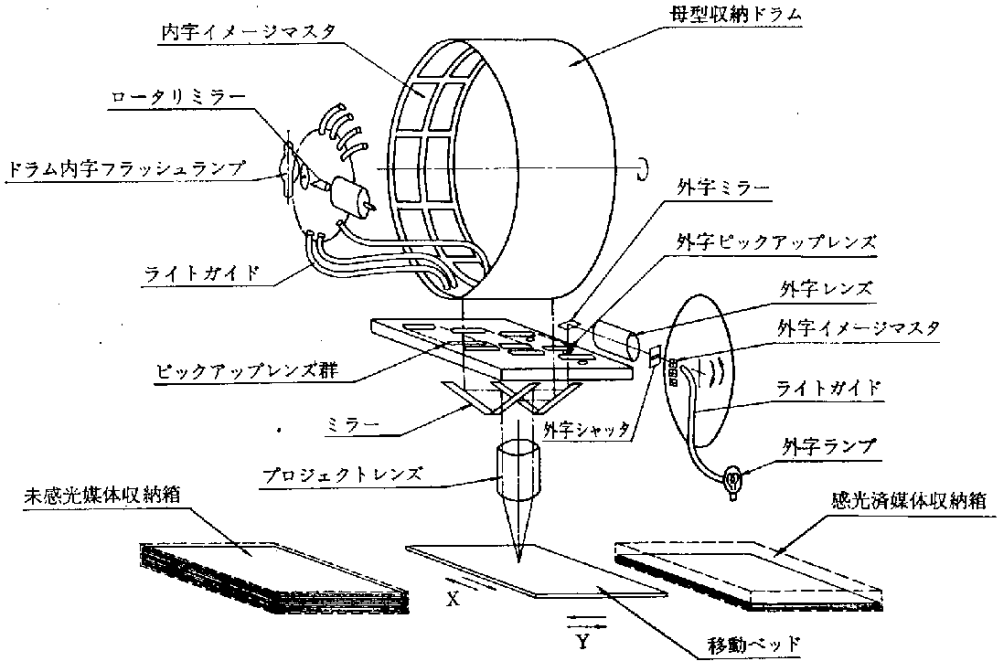


図4．7　富士通F6531C　写植装置原理図

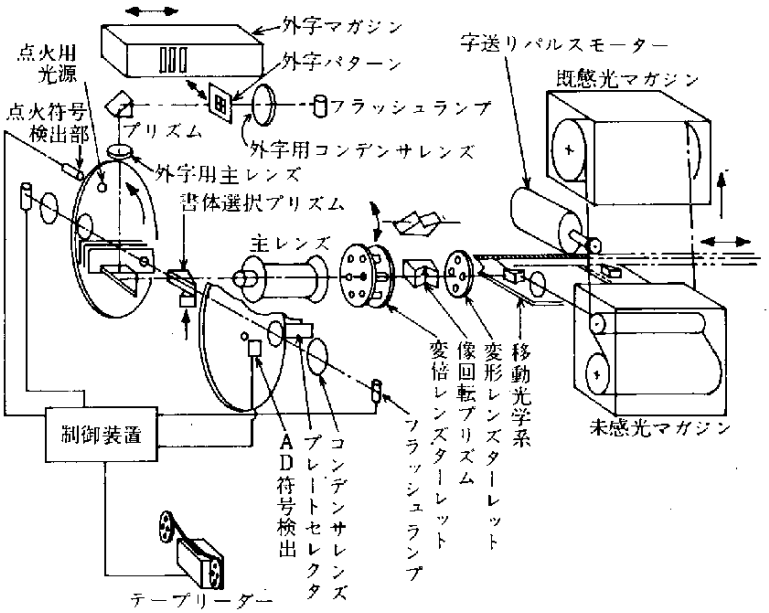


図4．8　写研SAPTON　写植装置原理図

文字種を限定して設計されなければならない出力装置にとって，現状ではさけることのできない問題点である。

それぞれの文字発生方式でも，標準の文字種に収容されてない外字パターンを用意して，必要に応じて外字文字発生部から発生させる方式をとっているものもあるが，その場合でも，用意してある外字にない文字が要求された場合は処理不能となる。

この外字処理はシステム的に不可欠の問題であるので，システム設計上，経済性を考慮した対策をほどこしておく必要がある。

4.3 文字出力システム

前述のように，日本語情報は漢字のもつ特殊性により，英数字情報の出力装置に比べ，高額で開発も遅れていたが，この1～2年来，需要と相まって急激に各社から漢字プリンタなどが発表され，実用期に入ってきた。

文字出力システムは応用の分野から，ハードプリント，ソフトプリント，COM プリント，電算植字の四つに分けることができるが，このなかでは，ハードプリントと電算植字の分野の応用が最も進んでいる。これらは前述のシステムとしての処理内容のレベル1とレベル2のデータ処理と情報編集処理のレベルのものであるが，システムの利用機能面からみると，通信回線の開放を迎えて，日本語情報処理用の使いやすいオンライン端末機の出現が望まれている。

4.3.1 ハードプリントシステム

従来の機械式カナ英数字ラインプリンタにかわる電子式漢字ラインプリンタも各種発表され，速度も早く，価格も低額のものが現われ，使いやすくなってきた。

ハードプリンタの性能上のキーポイントは図4.2の構成図に示される文字信号発生部と印字部とにあり，高速電子プリンタとしては，それに使用される用紙の種類とコピーの可能性が問題となっている。現在発表されている代表的なハードプリンタの性能一覧表を表4.10に示すが，プリンタ制御装置として，ミニコンピュータを用い漢字のもつ特殊性をソフトウェアで処理するケースが多くなってきた。ミニコンピュータはほとんどが16ビット1ワードの構成で漢字コード処理に適しているほか，ビット処理の必要な漢字情報処理の面からもプリント前処理用として使いやすい。

表4.9に各種のハードプリンタの分類を，表4.10にその性能を示す。

(1)活字打鍵式

前述の文字信号発生の項で説明したように，代表的なものは沖，新興製のもので，50 BPS 通信回線の受信端末機として，古くから新聞社等で広く用いられている。

(2)ドット機械式

ドット機械式のものとしては，IBM と日本電気製のものがあり，いずれも印字方式にインパクトなワイヤプリント方式をとっており，いずれもコピーがとれることが特長である。

IBM 製のものは計算機の周辺装置の漢字ラインプリンタとして設計され，1行の印字数は16字と少ないが，プリント速度は300～333行／分と高速である。

	方式名	印字方式	記録方式	代表機種名
インパクト方式	活字打鍵式	活字	インク	沖漢テレプリンタ，新興漢テレプリンタ
	ドット機械式	ワイヤピン	インク	日電(C5220)，IBM(2245)
ノンインパクト方式	ドット電子式	多針電極	静電記録	富士通(6501B，6502A)，日電(C5210)
	〃	FO CRT	湿式電子写真	高千穂（T4100）
	〃	イオン流電界制御	インク	沖エレクトロプリンタ
	字母ビデコン式	FO CRT	銀塩安定化	富士通（6521A）
	フライング・スポット式	FO CRT	銀塩安定化	日本電子産業（3100／750）
	〃	CRT	乾式電子写真	日本電子産業（3100／800）

表4．9　ハードプリンタの方式の分類

日本電気製のものはヘリカルドラムを用いた特長のあるワイヤプリント方式で，プリンタも小型で高速通信回線受信端末用として設計されている。

以上(1)，(2)の機械式のプリンタはプリント速度は以下の電子式に比べおそいがプリント用紙に普通の紙が使え，コピーがとれるところに特長がある。

(3)ドット電子式

200 BPS 通信回線受信端末用として沖，東芝製のものがあり，沖製は記録方式に感熱式を用い，東芝製は昭和44年に共同通信社との共同で開発されたもので，静電記録方式を用いている。

富士通の新聞用漢字プリンタは新聞社の集信プリントシステムとして設計されたもので，50～200 BPS の8～32回線をＦＡＣＯＭ２７０－２０計算機で集約制御し，静電記録方式でモニタプリントを出力する。このプリンタの特長として，並列に3枚のコピーがとれ，重ね綴じの処理がおこなはれて出力される。

計算機周辺装置の漢字ラインプリンタとしては富士通，日電，沖製のものがある。富士通，日電製のものはいずれも多針電極による静電記録方式であるが，富士通製は電子式ノンパクトプリンタとして唯一の2枚同時コピーがとれる特長をもっており，日電製のものは1,600行／分と高速である。沖製のものは最近発表されたもので，特殊なインクミスト印字方式により，普通紙に4,000行／分の超高速でプリント出力する。

高千穂交易製の漢字プリンタはドット文字としては32×32ドットと高品質で，明朝体調の文字を高速に FOCRT を用いて，乾式電子写真方式により，プリント出力する特長を有するほか，磁気テープ入力によるオフライン方式で，各種のアプリケーションソフトウェアをそろえている。

(4)字母ビデコン式

富士通の字母ビデコン式は文字発生に簡潔な同方式を用い， FOCRT により銀塩安定化方式で印字記録している。文字発生はアナログ式なので印字品質は自然で，5,000字を収容し80字／秒と高

速ながら価格が安い。

(5)フライングスポット式

日本電子産業製の漢字システムはフライングスポット方式を用い，20,000字種の文字を高品質で高速にプリントできる特長を有する他に，文字の大小，変形，可変文字ピッチ，可変行送り等，高度な機能をもっている。この漢字システムはミニコンを制御装置内にもち，CRT印字方式を用いたプリンタの他，次節以下でのべるディスプレイ，COM装置も接続できる漢字情報処理用のターミナルコンピュータ的な設計がなされている。

漢字プリンタとしてFOCRTによって銀塩安定化紙にプリントする機種と，CRTから乾式電子写真方式によって一般普通紙に高速でプリント出力できるプリンタを有しており，電子式のプリンタが一般にプリント用紙が高価で，消耗品経費がかさむ問題点を解決している。

(6)応用システム

日本語情報処理のデータ処理レベルのシステムとしては，ハードウェアの価格面でまだ問題点はあるが，性能面では実用的にほとんど問題のないところまでにきており，最大の問題は出力システムのそれよりも，入力とソフトウェアにある。

応用システムとしては，まず従来カナあるいはコードで出力されていた氏名，住所，品名などを，漢字で出力するケースと，宛名印刷に用いられる応用が多い。いずれにしても氏名の印字には多文字種を必要とし，外字処理がさけられないという問題はあるが，現在広くおこなわれているカナ書きの宛名印刷が漢字にかわることは，配達の能率化の他，情報の確実さおよびユーザへの大きなサービス向上となる。

新聞社における漢字処理は最も古くからおこなわれ，もっぱら受信端末のモニタプリントに用いられているが，ハードプリンタの文章出力を軽印刷用にする応用も多い。

いずれにしろ，漢字情報の入力は高価であるので，重複利用されるもの，改訂の多い情報処理などに適しており，各種の応用に対するソフトウェアの開発が必要となってくる。

4.3.2 ソフトプリントシステム

ソフトプリントの代表的なハードウェアとしてはCRTディスプレイ装置であるが，英数字情報用のディスプレイシステムも，応用面ではこれからのものである。したがって漢字情報用としても一般的なシステマティックな製品がなく，沖電気，東洋通信機，日立製作所，東芝，日本電気などの各社から発表はされているが，実用機としては日本電子産業製のものが数例あるにすぎない。表4.11に沖電気と日本電子産業製などの性能表を示す。

ディスプレイ文字の表示方式としてはドット，ストローク，字母走査の方式があるが，今後はドット，ストロークの方式が有望である。表示CRTとしてリフラッシュタイプとストレージタイプのものがある。日本電子産業製は字母走査方式（フライングスポット方式）により，放送用はリフラッシュタイプ，情報処理用にはストレージタイプを使用している。

ディスプレイの応用分野として，

(1) 情報照会，検索用などのオンライン端末利用

(2) 情報編集校正用

(3) テレビ放送用

などがあげられる。

(1)のオンライン端末の例としては，漢字ディスプレイが16台用いられている明治生命の例のみで，これからの分野である。この場合には漢字入力指示の方法も含めて，各応用システムごとにマンマシンインタフェース機能の設計は慎重に取り組む必要がある。

(2)の情報編集校正用は大量情報，とくに文章情報の入力校正に訂正場所，訂正法，訂正内容を指示してプログラムでおこなうことが，一般に採用されているが，訂正入力ミスなどのため，実際的にその正確さに問題があり，ディスプレイによる校正システムの検討が有望視されている。

(3)のテレビ放送用としては選挙速報などの電算機による直接オンライン放送が有名である。

この他，漢字を画像として，グラフィックディスプレイで表示処理する応用や，漢字入力装置のスライド投影など各種のものがある。

4.3.3 COM プリントシステム

COM 装置の利用も，まだマイクロフィルムシステムそのものがシステマティックなレベルにきていないので，これからの分野であり，英数字用の COM も国産機にはなく，漢字 COM としては日本電子産業製のもののみで，その性能を表4.12に示す。

外国製のもので，グラフィック機能をもつ COM 装置で漢字を画像として処理することができるものがある。

日本電子産業製のものはプリンタ制御部は同社のハードプリンタとおなじもので，フライングスポット方式の文字発生により，高品質の文字を英数字の場合4,000行／分，漢字の場合，2,000行／分の高速で，16㎜あるいはマイクロフィッシュにプリントアウトする。

同社製のものの応用例としては，キヤノン，総理府統計局があるが，本格的なＣＯＭとしては住友信託銀行の証券代行業務にコピー・フローと組み合わせて使用されている。

4.3.4 電算植字システム

日本語情報処理として，電子計算機によって本格的に処理された最初の例は，昭和43年に実用化された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システムである。印刷分野においては，人手による活字拾い，版組みなど，経験的な技能を必要とし，これらの労働力の不足，および生産の合理化に対応したシステム化の要望が進み，電算植字システム導入の例が多い。

電算植字システムのハードウェアの方式分類を図4.9に，その代表的なものの性能を表4.13に示す。第2世代機としては，機械式の字母光学式文字発生による写研，富士通のシステムがあり，植字速度は10～20字／秒である。第3世代機としては RCA, IBM, 日本電子産業のシステムがあり，RCA, IBM はドット方式，日本電子産業はフライングスポット方式の文字発生で，印字記録はいずれも CRT 方式により70～150字／秒の高速植字である。

写研のものは SAPTON システムと呼ばれ， HITAC－10 ミニコンピュータによる編集処理で，1段組みの棒組みフィルム出力をおこない，小型システムとして，おもに新聞社および印刷会社に数

多く採用され，最も実績のあるシステムである。

富士通の PETY システムは FACOM270－20 あるいは FACOM－R の電算機によってページレイアウト編集され，最大A４版のシートフィルムに出力する本格的なシステムで，大手の凸版印刷をはじめ，協業化あるいは共同利用の電算植字センタに多く利用されている。

日本電子産業の JEM システムは HITAC－10 ミニコンピュータによってページレイアウト編集され，最大 250 mm幅のロールフィルムに高速で出力する全電子光学式システムで，出版社，新聞社などで利用されている中型システムである。

IBM のシステムはグラフィックディスプレイによるページ割付け，組版制御などを含んだ大型の電算編集植字システムで，朝日新聞，日本経済新聞などに納入され，システム化が進んでいる。

RCA のビデオコンプは欧米諸国で実績のあるシステムで，印字部の CRT は前２社のものが行ごとに植字するのに対し，ページ植字をおこなう特長をもっており，ある大新聞社に納入される予定といわれている。

電算植字システムはまだハードウェアにも若干の問題はあるが，システム全体としてみると，編集処理ソフトウェアの充実，インプットの簡易化，有効な校正の方法の確立，ページレイアウト処理の確立など，ハードウェアの周囲の問題が大きく残っている。このなかで当面の問題となるものは有効な校正法である。ファンクションの校正を含め，校正入力ミスによるエラーの増加など有効な対策が望まれているが，原始的ではあるが，正確な入力法の確立が第一の対応策であるとされている。いずれにしろ，これまでシステマティックでなかった印刷出版分野における電算植字システムの成果は，これらの問題点のシステム化にかかっているということができる。

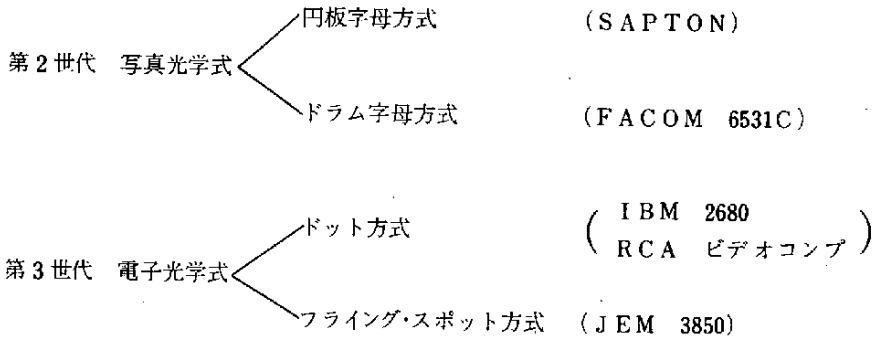


図４．９　電算植字システムの方式分類

なお電算植字システムの性能として最も重要視される文字品質は文字のデザインを含めたハードウェアの性能の他，フィルムの性能，現像処理，版材の製作，印刷工程などの総合システムの性能で決まってくるものである。

4.4　音声出力システム

音声出力システムも文字出力システムとおなじように音声テープに記録するハード出力と、音声として出力するソフト出力の二つの形式があるが，実用的な用途としてはソフト出力が一般的である。音声出力は文字出力のシステムに比べ，空間的に広い場所で沢山の人に一度に情報を伝達する場合や，少量のデータでない応答の情報出力などに適している。

音声出力システムとしては長い研究期間を過ぎて，電電公社の時報サービス，電話計算サービスの応答に用いられている初歩的なものから，成田新国際空港における案内放送システムなどと実用的なシステムがいくつか利用されるようになってきた。

音声出力システムの出力方式としては

1）録音編集方式

2）パラメータ合成方式

3）法則合成方式

がある。前二者は実用化の段階にきている。が，法則合成方式は研究の段階である。しかしながらいずれも実用はこれからの技術で，

1）出力語い数の増加

2）出力品質の自然性，明瞭性の向上

3）安価な方式

などが一般的な条件として要望されている。

4.4.1　録音編集方式

録音編集方式は前述の実用例に用いられている方式で，磁気ドラム上にあらかじめ標準の発声者によって，そのシステムで必要とする音声の単語および文節を録音しておき，処理装置からの制御信号によって切換えて合成編集し，文章を組立てる方式である。

録音言語が自然語なので，自然性は他の方式に比べもっとも高いが，各組合わせ単語間の自然性や時間的な間隔などの技術的な問題がある他記憶情報量が多く，語い数に制限があり，数百語以下が実用的である。

この方式の実用例としては前述の電電公社の応答サービスの他，富士通の成田新国際空港の案内放送，松下通信の札幌市地下鉄の案内放送などがある。

4.4.2　パラメータ合成方式

録音編集方式における語い数を増加するために，記憶する情報を自然の音声ではなく，音声の素片を分析し，音声の特徴要素（パラメータ）を材料としてドラムに記録し，制御信号により必要なパラメータを演算で合成し，音声を作り出す方式がパラメータ合成方式である。

この方式の応用例としては日立製作所が開発している国鉄の座席電話予約システムなどで，この場合，語い数は1,500語をこえるという。

パラメータ合成方式の中には音声素片のパラメータの種類によって，IBM などのチャネルボコ・

ダ方式や電電公社武蔵野電気通信研究所の PARCOR 方式などが他にある。

最後に，日本語情報処理システムは日本語のもつ漢字の文字種の多さ，複雑さ，文章としての編集の複雑さなどの特殊性により，欧文情報処理システムに比べ大きなハンディをもって，その研究開発がスタートした。しかしながら，現状の技術的進歩のレベルは，一般の技術レベルの向上と相まって数年前に考えられたより以上のレベルに，これらの問題点を克服して到達していると考えられる。

欧文の情報処理装置の代表的なタイプライタは入力と，出力用のモニタが同時にとれる，コンパクトな装置で，これに匹敵する漢字処理装置の出現は当分考えられないが，文字の絵素で約 10 倍，種類で約 100 倍以上の漢字出力装置が，価格的には欧文用の装置の約数倍以下のレベルにまできている。

日本語情報処理の出力システムのみの問題ではないが今後，文字種，コード，字体，伝送制御手順などの標準化が望まれる。

メーカー名		沖	沖	沖	日本電気
型式名		漢字印刷受信装置	T 300	エレクトロプリンタ	C 5210
収容文字	最小単位		1,850	3,000	2,209
	拡張単位			1,000	2,209
	最大収容字数	2,600	10,000	8,000	8,836
文字発生方式		活字打鍵	ドット	ドット	ドット
ドット文字	ドット数		18×16	18×16	24×24/15×10
	大きさ(mm)		4.0×3.6	4.7×4.0	4.23×4.22/2.7×1.8
	記憶素子		ワイヤ or コア	ワイヤ or コア	
字母方式	形状				
	1枚収容字数				
	1字大きさ				
印字方式		活字ドラム	多針電極	イオン流電界制御	多針電極
記録方式		インキ	感熱	インキ	静電記録
プリントフォーマット	行大きさ(字)	10.5P×19字	10P×66字	12P×66字	12P×80字
	字間隔 mm	1/5″	1/5″	1/5″	1/5″
	行間数 mm	1/3″ or 1/6″	1/3″	1/3″	1/3″ or 1/4″
プリント文字大きさ		3.66×3.66mm	4.0×3.6mm	4.7×4.0mm	4.23×4.23/2.7×1.8mm
プリント速度	行/分		6	4,000	1,600
	字/分	120	400	264,000	
用紙	大きさ	8インチ	8～15インチ	16⅞インチ	Y18-T16/Y16-300
	スプロケット	無	有	有	有
	給紙	ロール	シート	シート	ロール/折畳
	種類	普通紙	感熱紙	普通紙	静電記録紙
	コピー枚数	可			

表 4.10 ハードプリンタ性能表 (1)

日本電気	日本電子産業	日本電子産業	谷村新興	谷村新興
C 5220	3100 / 750H	3100 / 800	SCP−200	SC 4000
2,960	5,000	5,000	2,544	4,096
1,000	2,500	2,500		
3,960	20,000	20,000		
ドット	フライング・スポット	フライング・スポット	活字打鍵	活字打鍵
18×18				
4.23×4.23				
	平角板	平角板		
	100字	100字		
	1×1mm	1×1mm		
ワイヤ	FOCRT	CRT	活字ドラム	活字ドラム
インキ	銀塩安定化	乾式電子写真	インキ	インキ
26字	12P×48字	12P×56字	12P	12P
1/5″	1P毎可変	1P毎可変	5.08mm	5.08mm
1/3″	1P毎可変	最小 11P 1P毎可変	8.5 mm	8.47mm
4.23×4.23	4〜24P	10〜24P	12P	12P
100	平均 400	平均 750		
			180	180
214 mm巾	A4, B5巾×180mm〜500	B5, A4, B4	A4巾	A4巾
無	無	無		
ロール	ロール	シート	ロール	ロール
普通紙	安定化紙	普通紙	普通紙	普通紙
			3枚	5枚

メーカー名		東芝	東芝	富士通	富士通
型式名		KF100/KF200	KF400/KF3000	F6501B	F6502A
収容文字	最小単位	2,362	2,362	2,688	4,800
	拡張単位	480	1,734	384	
	最大収容字数	2,842	4,096		7,200
文字発生方式		ドット	ドット	ドット	ドット
ドット文字	ドット数	24×22	24×22	16×15	18×15
	大きさ(mm)	4.2×3.84mm	4.2×3.84mm	3.8×3.8	4.0×3.8
	記憶素子	変成器形リードオンリメモリ	磁性線ソレノイドメモリ	磁気ドラム	ROM
字母方式	形状				
	1枚収容字数				
	1字大きさ				
印字方式		多針電極	多針電極	多針電極	多針電極
記録方式		静電記録	静電記録	静電記録	静電記録
プリントフォーマット	大きさ(字)			15字	68字
	字間隔mm	1.5mm	5.08/4.65mm	5.5mm	5.08mm
	行間隔mm	12mm	12/8mm	12.5mm	8.47/6.35mm
プリント文字大きさ		4.2×3.84mm	4.2×3.84mm	3.8×3.8mm	4.0×3.8mm
プリント速度	行/分	26/53	107/600	121.5	600
	字/分	400/800	1,600/12,000		
用紙	大きさ	130mm×100m	仕様による	182mm×250m	7～17インチ
	スプロケット	無	〃	無	有
	給紙	ロール	ロール	ロール	ロール
	種類	静電記録紙	静電記録紙	静電記録紙	静電記録紙
	コピー枚数			並列3部	2部

表4．10　ハードプリンタ性能表（2）

富士通	高千穂	IBM
F6521A	T4100	2245
	2,048	7,525
	512	
5,376	8,192	10,000
字母ビディコン	ドット	ドット
	32×32	18×22 / 18×18
	8～12P	4.6×5.8 / 4.6×4.6
	LSI	磁気ディスク/ドラム
円板		
5,376字		
0.8×0.8mm		
OF CRT	OF CRT	ワイヤ
銀塩安定化	湿式電子写真	インキ
9P×55字	64字	16字
3.2mm	3mm	5.08mm(1/5″)
8.47 / 4.23mm	6mm	8.47mm(1/3″), 6.35mm(1/4″)
3×3mm	3～12P	4.6×5.8/4.6×4.6
		300, 333
3,900	72, 115	
210mm	B5 or A4	90～480mm巾
無	無	有
ロール	ロール	シート
安定化紙	電子記録紙	普通紙
		可

メーカー名		沖	東洋通信機	日本電子産業	日本電子産業
型式名			D1501	2420	3100
収容文字	最小単位	2,578	550		2,500
	拡張単位				2,500
	最大収容字数		4,000	3,000	10,000
文字発生方式		ドット	ドット	フライング・スポット	フライング・スポット
ドット文字	ドット数	18×18	18×18		
	大きさ(mm)	6×6 mm	7×7 mm		
	記憶素子	コア			
字母方式	形状			平角板	平角板
	1枚収容字数			100字	100字
	1字大きさ			1×1 mm	1×1 mm
表示性能	素子	CRT	CRT	CRT	ストレージCRT
	画面大きさ	160×230 mm	150×210 mm	160×220 mm	160×240 mm
	フォーマット	32字×16行	32字×8行	18字×8行	32字×16行
	文字大きさ			10×8 mm	8×6 mm
	フレームレイト			40	—
	色		P 39	P 24	P 1
指示	種類			カーソル	カーソル
	コントロール		ライトペン	キー	キー
入出力制御	オンライン入出力		1,200BPS	2,400BPS	2,400BPS
	入力キーボード		漢字キー	2,688字漢字キー	2,688字漢字キー
	入力機器			紙テープリーダ	紙テープリーダ
	出力機器			紙テープパンチ	紙テープパンチ
	ハードコピ				有
編集機能			字挿入，削除 行挿入，削除 全面消去等	字挿入，削除 全面消去等	字挿入，削除 入替・等
接続台数				3台	32台

表4.11 ディスプレイ性能表

メーカー名		日本電子産業
型式名		3100/150
収容文字	最小単位	5,000
	拡張単位	2,500
	最大収容字数	20,000
文字発生方式		フライング・スポット/ドット
ドット文字	ドット数	20×16
	大きさ(mm)	OCRB フォント
	記憶素子	MOS IC
字母方式	形状	平角板
	1枚収容字数	100字
	1字大きさ	1×1mm
印字方式		CRT
記録方式		銀塩写真
プリントフォーマット	ページ大きさ	AN 132字×64行 漢字 66字×40行
	字間隔(mm)	1CP 可変
	行間隔(mm)	1CP 可変
	位置指定	
	グラフィック機能	線引き可能
プリント文字大きさ		8～12CP
プリント速度	行/分	AN 5,000 漢字 2,500
	字/分	AN 660,000 漢字 150,000
使用フィルム	ロール	16～105mm
	フィッシュ	105mm巾

表4.12 COMプリンタ性能表

メーカー名		写研	富士通
型式名		SAPTON－A	6535
収容文字	最小単位		
	拡張単位		
	最大収容字数	6,400字	7,680
文字発生方式		字母光学式	字母光学式
ドット文字	ドット数		
	大きさ(mm)		
	記憶素子		
字母方式	形状	回転円盤	回転円筒
	1枚収容字数	3,200字	
	1字大きさ		2.8×2.8mm
印字方式		フラッシュランプ露光	フラッシュランプ露光
記録方式		銀塩写真	銀塩写真
プリントフォーマット	行大きさ(字)	240mm	210mm
	字間隔(mm)	1/4級	0.25P
	行間隔(mm)	0.5級	0.25P
	文字大きさ	9～18級 6種	6～14P 8種
	文字変形	可	可
	たて/よこ	可	可
プリント速度	行/分		
	字/分	900	1,000
版下フィルム出力	モニタプリント		
	大きさ	10インチ×100フィート	8インチ×100フィート
	フィルム送り	ロール・片パーフォ	ロール・アンパーフォ
	供給形状	ロール 100フィート	ロール 100フィート
	種類	フィルム	安定化処理紙, 印画紙フィルム
入力装置		紙テープ	紙テープ
編集システム	コンピュータ	HITAC-10 8KW	FACOM U-200 12KW
	プログラム	SAPCOL-H	FCL-3
	入力装置	紙テープ	紙テープ
	出力装置	紙テープ	紙テープ
	補助記憶	磁気ドラム 32KW	磁気ドラム 65KW

表4.13 電算植字プリンタ性能表(1)

メーカー名		富士通	日本電子産業	RCA	IBM
型式名		6531	3850	ビデオコンプ	2680
収容文字	最小単位		5,000	10,000	自由
	拡張単位		5,000	10,000	自由
	最大収容字数	12,762	20,000	40,000	外部記憶装置の容量による
文字発生方式		字母光学式	フライング・スポット	ラインドット	ラインドット
ドット文字	ドット数				
	大きさ(mm)				
	記憶素子			磁気ディスク	
字母方式	形状	回転円筒	平角板		
	1枚収容字数		50字		
	1字大きさ	2.8×2.8mm	1.2×1.2mm		
印字方式		フラッシュランプ露光	CRT	CRT	CRT
記録方式		銀塩写真	銀塩写真	銀塩写真	銀塩写真
プリントフォーマット	行大きさ(字)	210mm	240mm	296mm	210mm
	字間隔(mm)	0.25P	0.5P	$\frac{1}{200}$ P	
	行間隔(mm)	0.25P	0.5P	$\frac{1}{32}$ P	2回以上の分割出力で1インチ以上も可
	文字大きさ	6～12P 6種	4～24P	4～96P	4～18P
	文字変形		可	可	可
	たて/よこ	可	可	可	可
プリント速度	行/分				
	字/分	1,200	10,000	4,200	135,000(英文8ポ)
版下フィルム出力	モニタプリント		有 安定化紙		
	大きさ	A4, B5, A5, B6	250mm巾		9.44インチ巾
	フィルム送り	XY送り	ロール・パーフォ無		
	供給形状	シート 300枚	ロール 100m	ロール	ロール
	種類	フィルム	フィルム	フィルム	フィルム/印画紙
入力装置		紙テープ	磁気/紙テープ	磁気/紙テープ	
編集システム	コンピュータ	FACOM270-20 (32K)	HITAC-10 8KW	HITAC-8400 (131K)	IBM 360-40, 370-155など
	プログラム	FCL-7	EDIC-1	OCL PAGE 1-J	JPS/OLS
	入力装置	紙テープ	紙テープ	紙テープ	紙テープ
	出力装置	紙テープ	磁気テープ	磁気テープ	磁気テープ
	補助記憶	磁気ドラム131KW	磁気ドラム131KW	磁気ディスク	磁気ディスク

表4.13 電算植字プリンタ性能表(2)

5．日本語情報処理のための利用技術と標準化

5　日本語情報処理のための利用技術と標準化

日本語情報処理そのものの歴史の浅さもあってか，そのためのソフトウェアは，まったく体系化されておらず，基本的な考え方すらも固まっていない。また，ソフトウェア・パワーとしての蓄積量も，現在のところ，まだまだ見るべき充実ぶりを示すまでには至っていない。

日本語情報処理への関心だけについてみてみるならば，わが国における電子計算機普及の初期段階から，一部には，かなり先進的な挑戦もおこなわれており，その点では，日本語情報処理の歴史も，浅からぬものがあるといえよう。しかしながら，世間一般の関心を集めるようになったのは，1970年代にはいってからであり，それまでは，ごく少数の研究者，実務技術者によって，ほそぼそと実験的開発が進められていたにすぎない。

従って，そのソフトウェアの内実は，質，量ともに蓄積がないのは当然のことであって，むしろ，入出力機器や処理応用システムが，ある程度の普及をみせているが，これから以後に大きな発展が予想される分野である。

しかしながら，その体系化にあたっては，十分，かつ慎重に考慮を払わねばならない問題点が，すでにいくつか存在していることを，これまでの試験的開発成果が教えてくれている。その根源的な問題点は，国字（漢字を含む）という多数のキャラクター・セットを処理対象にしなければならないという宿命的なものであり，そこに既存のハードウェア・ソフトウェア両面の技術との整合をはかるのに特有の困難さがある。以下，そのいくつかの点を指摘し，若干の分析を試みることによって，今後の体系化のための参考に供したい。

5.1　鍵盤配列と符号化規則

漢字鍵盤入力方式としては，現在，かずかずの方式が提案されており，第6章での報告にも見られるとおり，カナ漢字変換方式，ラインプット方式などのユニークな試みもある。しかしながら現在のところ，キートップと漢字とを1対1で対応させるか，あるいは，シフト・キー操作によって対応させていく，いわゆる「漢字テレタイプ鍵盤方式」が主流を占めている。

漢字テレタイプ鍵盤方式の場合，鍵盤上の文字配列と符号構成とは，相互に密接な関連があり，それぞれ，そのどちらをとるかによって，入力以後のソフトウェア処理への負担の度合いが大幅に異なってくる。すなわち，鍵盤上のキートップ配列を音訓順にするか，あるいは，出現ひん度を考慮した配列にするか，によって，文字符号の構成規則は，決定的な影響を受けることになる。

ふつう，入力段階でのキーパンチャの操作性を重視し，スピード・アップが最重要の目標となる組版処理などの入力鍵盤では，出現ひん度を重視した配列となる。すなわち，漢字テレタイプ鍵盤では，盤内収容文字種が一般に2,000字前後であるため，盤面は大型化し，両手操作に頼らざるを得なくなる。

打鍵時の手の動作範囲を最小にするためには，日常の日本語文で，最もよく使用される文字群を中

央部に配置した方が望ましく，CO　59符号をはじめとする新聞社，印刷会社で用いられる漢字鍵盤は，この方式を採用している。ふつう350～600字種の文字群を高ひん度出現字種とし，盤面中央部から手前に配置し，操作性の向上をはかるとともに，打鍵速度のスピード・アップを期しているわけである。

しかしながら，この方式では，後述のような理由によって，文字符号の音訓順による構成は，ほとんど不可能となる。印刷処理のように，入力された一連の文字ストリングの配置レイアウトのみが対象となるソフトウェア処理では，これでもほとんど問題は生じない。従って印刷処理用鍵盤が，操作性を最重要視して，出現ひん度別配列方式をとるのは，そこに理由がある。一般に，350字種を打鍵者の手元近くに配列した場合，通常の日本文の75％は，それによってカバーすることができ，配列位置の習熟記憶も可能である。残り25％が，盤面上に手を大きく移動させ，場合によっては，目視による捜索が必要とされるわけである。このような配列方式の模式図を描けば，図5.1のようになる。

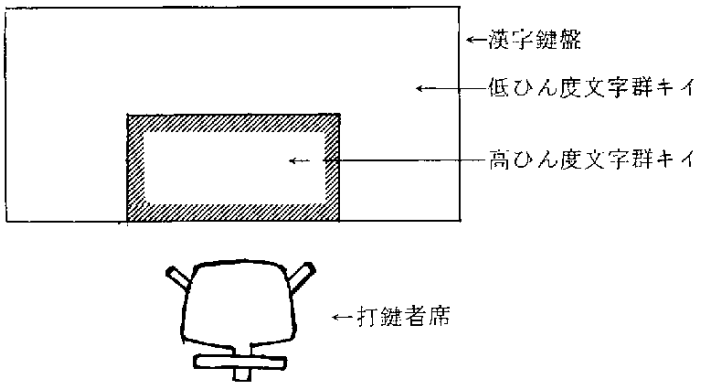


図 5·1　ひん度順配列方式

この場合，当然のことながら，符号構成と音訓順との対応は，ほとんどとれなくなってしまう。たとえば，英文テレタイプの鍵盤の場合，鍵盤文字配列と符号構成が最も独立的に定めやすいのはコードバー符号化方式である。同種のコードバー型漢字テレタイプ鍵盤では，8列のコードバーに対し，192個のキートップを配し，1キートップに12～15文字を収容する方式を採っている。

12～15字種という1キートップ上の文字集合ごとに，たとえば音訓順ビット重みづけをとるなどという論理関係の対応をとることは不可能である。また，マトリクス・スイッチ方式でも，文字配列に無関係に符号構成をとろうとした場合，その必要とされる論理素子は，かりに包括される文字集合の大きさを2^nとすると$\sum_{i=1}^{n} i$が最低限の数となってくる。これは，現状のハードウェア生産技術では，機械装置としての入力機器の価格を非常に高価なものにしてしまうおそれがあり，まったく採用されていない。

以上のことから，単なる文字ストリング単位の処理ではなく，その内容にまで立ち入りたい場合，たとえば，文字単位，語単位の五十音順分類などという，ごくごく簡単な処理例を考えてみても，ひん度順別配列での入力符号では，そのままでは内部処理にかかれないことになる。

従って，単なる文字ストリング処理ではなく，ある程度の内容処理，たとえば五十音順分類などの

程度の処理を対象とする場合でも，鍵盤配列には，音訓順配列キートップを採用している。たとえば，国立国語研究所，電子技術総合研究所などで採用されている漢字テレタイプ鍵盤がそれである。この場合，キートップの位置は完全な音訓順であるため，いきおい，盤面上での必要文字キーの捜索範囲は大となり，打鍵速度は著しく低下する。

通常，内部処理を考慮しない印刷関係では高ひん度別配列が用いられ，打鍵速度をそれほど問題としない研究機関では音訓順配列が用いられてきたわけだが，日本語情報処理の普及に伴い，この面についての基本的な整合を考慮すべき時点にさしかかっているということができよう。

すなわち，印刷関係以外でも，こと実務分野に関する限り，打鍵速度が重視されるとともに，音訓順符号構成もまたなおざりにできない問題である。たとえば，五十音順の住所氏名録作成など，日常的な応用業務となる可能性が強い。

こうした場合，高ひん度別配列をとるか，音訓順をとるかは大問題である。作業性を考慮すれば，高ひん度別配列方式がベターであるが，後処理のためには音訓順符号変換が必要とされる。最近，提案されているカナ漢字変換方式，ラインプット，ソクタイプ漢字変換方式なども，広義の意味での符号変換を必要とするので，作業性重視の高ひん度別配列方式の流れをくむものとみてよかろう。

いずれにせよ，こうした符号変換は，目下のところ，全面的にソフトウェアに頼っている。しかも，その変換のアルゴリズムには，対象キャラクター・セットが2,000～6,000字種，あるいはそれ以上に及ぶため，それなりのくふうを要する。たとえば，通常の表引きルックアップ手法では，処理速度の低下のおそれが強く，変換対象符号の収容されているアドレスをいかにして早く見つけ出すかが，ソフトウェア作成上のテクニックとなってくる。一般に，高ひん度別配列方式では，ビット重みの符号構成の規則性は，まったく乱れているので，いかにしてアドレス算出のためのアルゴリズムを見つけ出すかが，プログラム作成でのコツになる。

なお，ISOならびにJIS情報交換用符号体系では，図形文字（英数字・記号）の符号構成のビット重みはABC順に準拠し，推奨鍵盤配列では，ECMA規格事務用タイプライタの操作性を考慮した配列を採り入れている。その間には，まったく独立の関係となっているわけだが，コードバー方式の鍵盤では両者の整合は容易に実現することができ，またマトリクス・スイッチ方式では，LSIエンコーダに頼っている。わが国にあっても，現状はソフトウェア依存の符号変換方式も，いずれは規格化され，LSIによるハードウェア変換を真剣に検討しなければならない時期が到来するものと思われる。

5.2 情報交換用漢字符号

前節では，配列方式からみた符号構成についての考察を試みてみたが，情報交換のための符号構成についてもゆるがせにできない問題である。現在，わが国では十数種の漢字符号体系が存在しているが，そのうち，異なる事業主体間での共通性を，いくぶんでも考慮し，かつ交換実績を持っているのは，1959年に共同通信社を主務機関として新聞社間で協定されたCO－59符号だけである。

それ以外の符号体系は，すべて実施当事者にとってのみ，最も便利なよう配列方式と符号構成を制定しているため，符号体系間の互換性はまったくない。また，これによって，共通ソフトウェアの開

発促進を著しく阻害しているということもできよう。

さらに，通信線による情報交換を考慮しているのも，いまのところCO－59符号のみである。CO－59符号は，専用電信線網による新聞記事の送配信を前提として制定されたものであるため，直接の伝送制御には関与しないものの，Stop, Bell, Cariage Return & Line Feedなどの機能符号は，国内6単位加入電信符号に準拠している。他の符号体系で，あらかじめ電子計算機への入力を前提として制定されたものには，数種の機能符号を備えているものもあるが，JIS情報交換用符号の伝送制御符号とは，まったく無関係である。

こうしたことから，漢字符号は，将来，公衆通信網によってデータ伝送を実施することが不可能となる事態も予想されるわけで，ソフトウェア開発面での共通基盤の確立とともに，伝送基準面でも，情報交換用漢字符号体系の制定が急務とされるわけである。JIS情報交換用符号の母胎となったISO/R646では，6単位および7単位の文字符号が制定され，これに準拠してJIS/C6220－69の7単位，8単位符号が制定されていることは周知のとおりである。これには32種の制御符号と94種の図形文字が含まれ，JISの8単位符号の場合，94種を1群としてG0，G1のふたつのキャラクターセットを制定，SI（シフトイン），SO（シフトアウト）の選択によって，英数字，記号，カナ文字のキャラクターセット選択をおこなっているわけである。

ISO，およびJISの規格では，このSO，SIのキャラクターセットを基本として，NUL，SO，SI，CAN，SUB，SP，DEL，ESCの制御符号を除いて，図形文字セットの拡張ができることになっている。この場合，ESC（エスケープ；拡張）制御符号によって，図形文字（この場合は国字）セットのシーケンスを呼び出していくわけであるが，その拡張方法の標準化については，なおISO技術委員会で審議中である

しかし，いずれ近い将来，その標準方式も制定され，それとともに，国字の標準符号化が急がれることはまちがいない。ひとつの試案として注目されるのは，71年10月に情報処理学会漢字コード委員会が発表した「標準コード用漢字表」である。これには，一応6,100字種の漢字を収容しているが，まったく問題がないわけではない。

その第一に，はたして6,100種もの国字の標準化が必要であるかどうかという点。現在のCO－59符号体系で表現できない文字の出現字種は，新聞記事の場合，100万字について380字種程度であり，その出現確率は，いずれも,10万分の1を越えることはない。6100番目の出現確率はきわめて低いものであり，そのオーダーは10^{-7}程度であると思われるが，少なくとも，そこまでカバーする必要があるかどうかということ。

第二に，その選定の基準であり，辞典，各種文字統計調査を参考に選定されているが，当用漢字1850字種の選定についても，国語審議会への批判が存在する現在，字種の増大は批判の自由度を増大さすことになる。

これらを総合すると，将来に想定される情報交換用標準符号は，音訓順，あるいはそれに準じるなんらかの規則性をもたさねばならないが，その際，標準字体表との整合性を確立するのに，著しく困難が伴う。また配列との関連を考えると，その整合性の確立は絶望的といってもよかろう。

したがって，情報交換用漢字符号としては，日常的に必要欠くべからざる字種，たとえば教育用漢字表程度の字種を最大限度として制定し，それ以上の拡張キャラクターセットは，標準拡張にのっとって，ユーザ自体に選択をまかせた方が無難であると思われる。

むしろ，そうした方が，ハードウェア的には，入出力機器，あるいは，符号変換用論理素子の量産体制も整いやすく，かつ，ソフトウェアの標準化も推進しやすい。量産化・標準化によるコスト・ダウンを目標とするなら，情報交換用漢字符号体系の制定には，十分に慎重な態度であるべきであるといえよう。

なお，現状で最も普及している漢字符号体系は，6単位2列さん孔による計12ビット／字がふつうであるが，上述のような情勢から8単位2列16ビット／字コードも普及をはじめており，さらには多重バイト・キャラクターセットの方向も十分，考えられることである。

5.3 日本語情報処理のためのCPU

現在，実用化されている日本語情報処理システムに用いられている電子計算機中央処理装置には，中・小型機では，一般にバイト・マシン，キャラクター・マシンの汎用事務用計算機が用いられている。また大型機でのアプリケーションもいくつかあるが，これにはワード・マシンやバイト・マシンが使用されている。

しかし，台数から見た場合，最も数が多いのは，PDP－8，HITAC－10，FACOM－Rなど，いわゆるミニコンピュータを利用したシステムである。しかも，システムの処理内容としては，常識的には大型機の分野での適用業務とされているところにまで，大いに活用されているのが特徴である。たとえば，人名・住所録などの五十音順分類編集や，辞典の項目分類や整理などに，ミニコンピュータが進出しているのが，その典型例であるということができよう。

その理由は，おおよそ，次のように考えることができる。つまり，漢字テレタイプ符号は，6単位制の場合，12ビット／字，8単位制であるなら16ビット／字であるが，これに対し，ミニコンピュータは，一般に1語12ビットか16ビットである。つまり，1語と1字が対応し，きわめて簡素化されたプログラムを作成しやすくなる。

内部処理は，一般にミニコンピュータの場合，語単位で行なわれるので，処理速度も大型機に対し，遜色ない。そこで，たとえば組版処理などというジョブの性格がはっきり定まり，システムの目的自体が，わりあいに単能的である場合，ミニコンピュータの方が，総合性能としては偉力を発揮することが多い。

現在，ミニコンピュータ・システムが，総数的には，最も普及している理由も，このあたりにあり，主として組版処理に用いられているのがめだっている。もちろん，磁気テープなどの外部補助記憶装置を付加すれば，前述の通り，項目別分類などの高級処理も十分に可能である。しかし，その際に考慮せねばならないのは，一般に，ミニコンピュータ関連の周辺機器は，低速・低容量がふつうであるので，システムとしての総合性能は，周辺機器の処理能力によっておさえられる傾向がある。比較的にジョブの対象がはっきり定まっており，かつ，単能的・小規模である場合，ミニコンピュータ

・システムが，圧倒的に有利となる。

バイト・マシンの場合，2バイトがちょうど漢字1字分に相当し，半語長（2バイト語）単位のアドレス機構を持っていると，内部処理では，ミニコンピュータに匹敵する能力が期待できる。しかし，汎用事務目的の小型バイト・マシンでは，むしろミニコンピュータと同列に考え，単能的に利用し，しかもプログラムは，機械語かアセンブラ・ベースで考えた方がベターである。

中型機以上のバイト・マシンは，汎用目的のための漢字情報処理システムに適合するが，その際注意せねばならないのは，コンパイラの構造によっては，ある特定の処理ステップに対し，処理時間がきわめて大きくなるという傾向がある。漢字情報処理では，1回の処理作業量が200万字を突破することも，往々にして生じるので，こうした大量データ処理の場合，思わぬ時間を必要とすることがある。そうしたことから，中型以上のバイト・マシンにあっても，プログラム言語は，アセンブラ・レベルに依存した方がベターといえよう。

内部処理速度だけに着眼するならば，大型ワード・マシンが最優位となる。ところが，その語長単位は，必ずしも12ビット，あるいは16ビットの整数倍とはならない。こうした場合，n文字を1語にパックし，残るビットにダミー・データをおくという手間を必要とする。処理段階の進行に応じて，プログラム的には，パック・アンパックをひんぱんに繰り返さなければならないことが多いが，その際，ダミー・データの介在によって，プログラム手順は，かなり複雑になるおそれがある。

こうした状況から考えて，現状では，漢字情報処理に最適のCPUシステムは存在しないと断言してよかろう。小規模・単能的システムに対しては，ミニコンピュータの適用が最適であり，大規模・汎用目的に対しては大型バイト・マシンが適しているというぐらいのところであろうか。実用システムにおいても，期せずして，このような傾向をたどっているようである。

5.4 漢字情報ファイルの編集

漢字キャラクタ・セットは，たとえばパンチ・カードのHセットなどにくらべ，圧倒的に字数が大きいのが特徴である。したがって，その入力情報ファイルには，ミス・パンチや，その他の原因による誤字の存在する確率も，圧倒的に大である。また，入力データは，一般に文章情報であるので，文字単位，パラグラフ単位，の追加・訂正・削除要求も，ひんぱんに発生する。

当然，ファイル中のある任意のストリング単位で迅速な更新を実施するプログラムの存在が必要となってくる。すなわち，一般に「エディター（編集プログラム）」と呼ばれているものを漢字向けに機能強化したものである。

現在，日本語情報処理システムで，数多く存在する実用システムのうち，最も弱体なのは，このエディター・プログラムである。これは，主として，プログラムの能力にあるのではなく，周辺装置，さらにシステムの運用能力を加えた，いわゆるファームウェアに，まだまだ大きな問題点が存在しているためである。

エディターには，オンライン処理とオフライン処理の二種類がある。このうち現段階では，最も進んでいるといわれているオンライン処理についてみてみると，この場合，現状では非常に高価な周辺

機器である漢字ディスプレイ装置の存在が絶対条件となる。

すなわち，コアメモリ，またはランダム・アクセス・ファイル中の任意のストリングを漢字ディスプレイ画面上に表示し，付属した漢字入力装置を対話モードで操作しつつ，追加・訂正・削除ルーチンとともに対話モード管理のためのソフトウェアを付加する必要がある。また，処理はほぼ完璧に行なわれるが，大量データの連続処理には不向きである。さらに，ディスプレイ操作に，特別な訓練を終えた専門オペレータを必要とするなどの欠点がある。少量のデータを対象として、絶対にまちがっては困るという作業対象に適したシステムであるといえよう。それにしては，周辺装置の価格，関連ソフトウェアの開発負担があまりにも大きいのが現状である。

オフライン処理の場合では，次の処理が必要となる。つまり対象入力ファイルのプルーフ・リストの高速漢字プリンタによる打ち出し，人手による校正・訂正情報の入力，プルーフ・リスト再プリントによる確認の手順である。このうち，問題があるのは訂正情報の入力である。

一般に，訂正情報は，次のようにして作成される。すなわち，該当するファイル名，ストリング所在位置の所在情報，訂正情報の三つによって，1トランザクションを形成する。ストリング所在位置は，行番号，文字番号をアイデンティファイヤとして用いる。その際，追加される訂正情報の性質から，当然，漢字入力装置が使用されるわけだが，たとえば，1字単位の追加・訂正・削除がひん発した場合，そのさん孔作業量はぼう大なものとなる。

このような点から，エディターのソフトウェア的機能は，ソフトウェア能力よりも，むしろファームウェアへの依存度が高く，周辺装置の改良・低価格化をも含めて，処理システム，とりわけ人員システムをも含む，総合的なシステムの見地から，解決をはかっていくべきであろう。

5.5 配列プログラムについて

一般に日本語情報処理の対象となるデータは，文章情報が主体であるといってよい。従って，その出力のプリント様式は，フォーマッティングに，かなりのウェイトがかかってくるのは当然である。そのフォーマッティングについても，アルファ・ニューメリックのラインプリンタ・アウトプット様式を漢字・かなによって打ち出した程度の簡単な書式から，高級商業印刷の版下となる複雑なものまで，その形態はさまざまである。

いずれにせよ，アウトプット用ソフトウェアとしての必須条件のひとつに禁則処理ルーチンがある。国字のキャラクタ・セットのなかには，文字ストリング中で，行頭，行末には存在してはならない文字群がある。句読点やカッコ類がそれであり，これらのサブセットを「約物」と通称している。行頭行末に約物が存在する場合，ストリング中の位置関係を1字分シフトし，不足字詰め分を文字間空白によって調整しなければならない。この調整マージンは，周辺ハードウェア能力に大きく左右されるが，その能力限界まで活用するようストリングのフォーマッティング処理をするのが禁則である。

また，印刷用版下作成のための処理ルーチンは，一般に植字プログラムと呼ばれている。これは，1ページという限られた領域内にストリングを十分に考慮されたフォーマッティングによって配列処

理するもので，その使用目的に応じて種々の形態がある。一般に，いくつかの副集合でページ単位にひとつの集合形態をとる新聞と，場合によってはページ単位での完結した集合形態をとらない書籍形態とに大別することができる。

この場合，便宜的にページ単位の処理形態を考えてみると，文字によって構成されるストリング・セットだけではなく，ケイ線，写真，カットなどのサブ・セットも処理対象として存在することになり，また余白も印刷デザイン上では無視できないサブセットである。

これらのサブセットが，それぞれ閉じた領域（例・コラム文，写真，カットなど）と開いた領域（例・本文，余白など）を形成し，相互の位置関係によって，デザイン的感覚を表示しているのがつねである。従って植字プログラムは，ある意味ではCADの変形であると見なすことも可能である。

従って，その処理ソフトウェアにもCADと同じくグラフィック・ディスプレイ表示による対話モード・ソフトウェアと，図形アウトプット・テクニック併用によるオフライン処理とがある。いずれも，各サブセットに共通なエレメント領域（これの定義基準は，ソフトウェア開発者によって異なる）を設定し，そのエレメントをサブセットごとに集積配置していくことが考えられる。その際，ページ・セットの大きさ，各サブセットの大きさ，種別などのいずれかを目的関数とし，残りを制約条件とみなすと，最適化アルゴリズムが存在するはずである。

しかし，この点については，いまだ完全に解明されておらず，完璧な意味での自動化配列ソフトウェアは出現していない。オンラインの場合，グラフィック・ディスプレイ・ターミナルによって，オフラインの場合，座標化パラメータによって，それぞれ，人間が各サブセットの位置関係を入力指示し，処理しているのが現状で，芸術的創造性は低いにもかかわらず，人間の介在に大きく依存している分野である。

6．日本語情報処理の応用

6.1　JICSTにおける科学技術文献速報編集システム
6.2　工業技術情報に関する日本語情報処理の一例
6.3　出版社における自動編集システム
6.4　日本語によるオンライン情報検索システム
6.5　物品管理システムの漢字出力の試み
6.6　生命保険会社における日本語情報処理
6.7　電算写植センターにおける文字組版の事例
6.8　新聞業界における日本語情報処理
6.9　日本語記録システム
6.10 電子計算機による国語研究
6.11 計算機によるカナ漢字変換の一手法
6.12 日本語情報処理ユーザー一覧表

6　日本語情報処理の応用事例

6.1　JICSTにおける科学技術文献速報編集システム

6.1.1　導入の経緯

日本科学技術情報センター（以下JICSTという。）がコンピュータを導入して、文献情報の処理を全面的に機械化する計画を開始したのは，昭和41年のことである。JICSTにおけるコンピュータ化の最終目標は，科学技術情報のいわゆる総合情報提供サービスシステムの確立であることには異論がなかったが，実行面からみると多くの問題が山積していた。

情報提供サービスには，大別して二つの形態がある。一つは時々刻々発生する情報を整理して，できるかぎりすみやかに提供するカレント・アウェアネス・サービスで，もう一つは要求により過去の情報を探索して提供する遡及検索サービスである。前者はマスコミ形，後者はカスコミ形で，ともに情報検索の一環をなしている。

情報サービス機関にはこの両者が要求され，JICSTも昭和33年以来，前者の形として抄録誌（科学技術文献速報）およびその索引を刊行し，後者の形として依頼による調査サービスを続けてきた。したがってコンピュータシステムも，この両者を包含したシステムとして形成され，そのメリットをあげねばならないという前提があった。

もっとも将来はファクシミリも含めて，オンラインで，この両者が総合されたシステムとなるであろうが，当面はコスト面から困難である。

さて上述の考え方にたってコンピュータ化を進めていた情報機関の代表は，米国国立医学図書館（NLM）であった。索引誌としてIndex Medicusを刊行しつつ情報を蓄積し，MEDLARSという医学情報検索システムを着々と整備しつつあった。このシステムは現在すでに実用に供されている。

NLMの一つの特色は，索引誌の発行のためにコンピュータの出力装置として，Photon ZIP 900という高速電算植字機を導入し，コンピュータで直接印刷用の版下を作成していたことである。この方式の成功がJICSTのシステム設計に，もっとも大きな影響を与えたのである。

JICSTでは，すでに世界各国から収集した科学技術情報の，日本語の抄録を作成し，抄録誌を刊行していたが，日本において不特定多数に情報提供サービスを行なうには，日本語であることが絶対に必要な条件であった。

かくしてカレント・アウェアネス・サービスを含む機械システムを計画すれば，必然的に日本語処理に立ち向かわねばならない状況となった。日本語のインプット機器については，当時すでに既成のものがあったので，焦点は高速の出力装置，しかも植字機にしぼられることとなった。

欧米においては，当時すでに前出のPhotonをはじめ，Digiset, Videocomp, Linotronなどが開発途上にあり，これらの日本語への適用可能性について調査中のところ，日本電子産業K.Kの全面的な協力が得られ，後述するJEM3800を導入することに決定し，JICSTのシステムが完成したのであ

る。

現在すでに抄録誌編集システムは軌道にのり，日々大量の版下がアウトプットされており，次の段階の遡及検索システムの実用化に努力をそそいでいる。しかし日本語処理についてはなお今後改善を要する問題点が多く，本システムの評価は後日をまたなければならないであろう。

6.1.2 字種の選定

本システムはインプット装置として，新興製作所の漢字鍵盤せん孔機を用いることにしたので，ハードウェアの面から使用できる字種は2,496字に限定された。この制限内で最も効率のよい字種を選定するために各種の調査が行なわれた。

まずすでに発行された抄録から実際に使われた漢字字種の調査を行なった。調査対象は昭和40年の抄録誌各編から層別抽出で1,279件を抽出し，そこで使用された漢字約10万字を調査した。その結果1,528字種が求められ，そのうち150字が表外漢字（当用漢字外）であった。

抄録の原稿はもともとJICSTで作成したものであり，表外漢字は極力制限していたので当然の結果であるが，この結果から本システムの前途は明るいものとなった。標本数が少ないことを考慮に入れ，種々検討の結果，最終的に当用漢字から1,781字種を選定し97字を削除した。

一方，科学技術文献を植字するためには英欧文字や記号類は欠かすことはできない。これらは別途調査し，必要と思われるものは全部収容することにした。このほかカタカナ，ひらがな，数字は全部収容し，なお余裕があるので表外漢字をとりあえず80字種収容した。

表外漢字で最も必要なものは，文献の著者名のための人名漢字である。このため日本人の著者名索引からの字種調査を行ない，その他地名および雑誌名の調査と合わせ，また唯一の例外として，「凸，凹」の2字を加えて80字種を選定した。

以上の結果をまとめて字種一覧表を表6.1.1に示す。

字種	インプット	アウトプット	
漢字	1,861	1,861	（明朝）
カタカナ	81	162	（明朝，ゴシック）
ひらがな	77	154	（〃）
英欧文字	65	195	（ローマン，イタリック，ボールド）
ロシア文字	66	198	（〃）
ギリシャ文字	33	99	（〃）
アラビア数字	10	30	（〃）
ローマ数字	20	40	（ローマン，ゴシック）
記号	199	248	（ローマン，一部イタリック，ボールド）
スペース	6	6	
予備	78	78	
合計	2,496	3,071	

表6・1・1 使用字種一覧表

表中予備欄は将来の拡張，たとえば医学文献の処理などが予想されるので，空白にしてあるものである。実際の使用経験からみると，著者名に少しの欠字（かなでインプット）があるほかは何らの障害も生じていない。

以上のほかに植字においてはさらに書体（ゴシック，イタリックなど），上つき，下つき文字(x^2，x_2など），合成（$\sqrt{x}$など）が必要であるが，これらを始めから鍵盤に収容しておくことは不可能であったので，漢字プリンタの機能を拡張し，インプットに際しては，ファンクションコードで処理する方法を採用した。この機能については後に詳しく説明する。したがって表6.1.1でアウトプット側は異書体を含めて延べ3,071字となっている。

6.1.3 漢字入力装置

コンピュータのハードウェアシステムはMT8台，ディスク2台の，FACOM230－50を中心とする普通の構成であり，とくに説明することはないであろう。このほかに日本語処理用としてインプット用に漢字鍵盤せん孔機，アウトプット用にオフラインの漢字プリンタが用意されている。

JICSTで使用しているインプット装置は，新興製作所製の，「SCK－201形漢字鍵盤せん孔機」である。本機は鍵盤部，パンチ部，リーダ部，インジケータ，カウンタ，見台，機械台からなり，8単位紙テープ（幅25.4mm）に，文字一字につき8単位2列のコードをせん孔する。

鍵盤は図6.1.1のように3種のグループからなる。

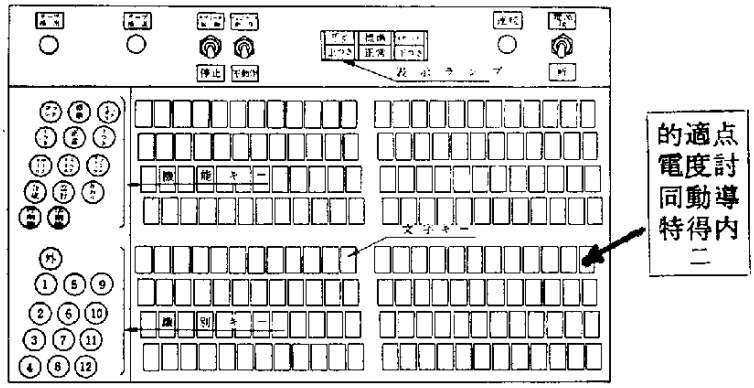


図6・1・1　SCK—201の鍵盤部

文字キー	24列×8行	192個
識別キー		13個
機能キー		14個

文字キーにはそれぞれ13個の文字が収容され，全部で13個×192：2,496字が収容されている。文字の選択は，その文字が収容されている文字キーとその中から必要な文字を選択する識別キーを，両手で同時に打鍵してせん孔を行なう。

紙テープへのせん孔は図6.1.2のとおりで，これは文字「点」のせん孔である。

2,496文字を表わすには12ビットを要するが，図のX1～X6，Y1～Y6の12ビットが文字を表わし，X8とY7ビットはそれぞれX列，Y列であることのコードであり，したがってパリティビットは設けて

ない。

このようにX，Y列の判定ビットを設けた理由は，システム導入の初期においてハードエラーによるXまたはY列の脱落がひん発したためで，これを目検によっても発見できるように工夫したもので，また紙テープリーダでの読み取りにあたってもこのビットを利用してチェックを行ない，インプットの万全を期している。しかし，さらに万全を期してY7をパリティビットとする方法もある。

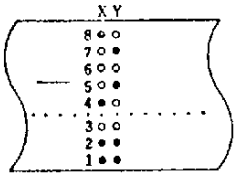


図6-1-2 紙テープのせん孔

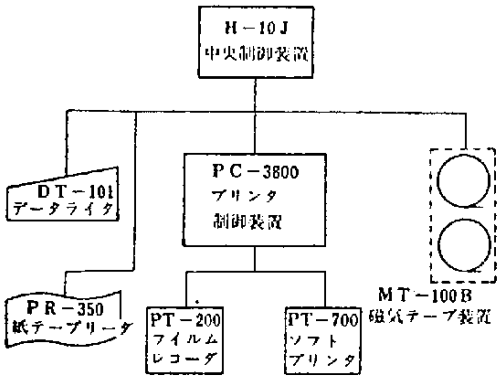


図6-1-3 ＪＥＭ3800の構成図

機能キーとしては，とくに設計して次の3種の機能を設けてある。

① 書体シフト（ゴシック，標準，イタリック）

② 上下シフト（上つき，正常，下つき）

これらの機能キーを用いると，それに続く文字は次のシフトキーを押すまでは，指定された形でプリントアウトされる。

③ 合成シフト

この機能キーを用いると，それに続く文字は直前の文字に重ねてプリントされる。以上の3種の機能は重複使用が可能である。たとえば

イX日2合下*i*

とせん孔すると「X」とプリントされる。

機能キーはそれのみで8単位2列のせん孔が行なわれる。なお機能キーはそのコードをせん孔するのみで，それらの機能に従ってプリンタを作動させるのは，コンピュータ側のソフトと漢字プリンタのハードウェアとである。

6.1.4 漢字出力装置

日本語用のアウトプット装置としては，日本電子産業K.K.の高速漢字プリンタ・システムJEM3800を用いている。このシステムの構成を図6.1.3に示す。FACOM230－50ですべて編集が終ったMTをかけて，プリント・アウトのみを行なう。

JEM3800は文字発生部にフライングスポット方式を用いた全電子式の高速植字機である。文字発生部は，PC－3800プリンタ制御装置に内蔵され，文字コードによって偏向されたフライングスポットにより文字板上の所定の文字を走査し，光電子増倍管により映像信号に変換し，プリンタCRT（ブ

ラウン管）上に文字を画かせ，これを撮影してプリントするものである。

文字板を持ち写真方式によってプリントを行なう点では，機械式の写植機とまったく同じ考え方であるが，文字の選択を電子式に行なう点が大きな相違で，これにより200字／秒以上の高速なプリントを可能としている。

プリンタにはPT－200フィルムレコーダと，PT－700ソフトプリンタの2種があり，それぞれ切り換えて使用できる。前者は感材としてフィルムを用い印刷版下の植字用，後者は銀塩安定化紙を用いて即時に現像するもので，校正ゲラ用として用いている。両者の性能の概要を表6.1.2に示す。

	フイルム・レコーダ PT－200	ソフト・プリンタ PT－700
記録媒体	フイルム	銀塩安定化紙
媒体幅	92㎜	282～257㎜
媒体長さ	100 m	同　左
文字サイズ	6，7.5，10.5ポ	10.5ポ
1行の長さ	max 30パイカ (126㎜) 6ポ：60 em 10.5ポ：34 em	max 30パイカ (126㎜) 10.5ポ：34 em
行間隔	0.5㎜単位可変	5.0㎜固定
印字速度	6～12行／秒 (200～600字／秒)	6～12行／秒 (200～350字／秒)
現像	外部	湿式内蔵
解像力	20本／㎜以上	8本／㎜以上

表6-1-2　漢字プリンタの性能

印字の1行の長さは30パイカ（12ポの文字30字分）で，1行分の印字時間は一定であるので，文字が小さくなると印字速度は早くなり，6ポの場合は10.5ポに比して約2倍となる。また印字の状況は，プリンタ制御装置に設けたモニタ用CRTによって，目検が可能である。

6.1.5　抄録誌編集システム

次にJICSTが編集発行している抄録誌「科学技術文献速報」の，コンピュータによる編集システムの概要を紹介する。

JICSTでは，国内および国外で発行された科学技術文献を網羅的に収集し，これをもとに日本語で抄録を作成しているが，この抄録原稿がそのままインプット原票となる。原稿には抄録のほかに分類コード，キーワード，雑誌番号，巻，号，ページ，発行年，使用言語，文献に掲載されている図，表，写真，参考文献の数などが記入されている。

この原稿をもとに漢字鍵盤せん孔機でインプット用の紙テープを作成するが，一方，国外誌の原文標題，著者名には日本語が使用されていないから，これらは別途原文献から英字せん孔機で紙テープを作成している。図6.1.4に編集システムのフローチャートを示すがこれは大別して，漢字コード変換，校正処理，編集処理の3つの工程に分かれる。

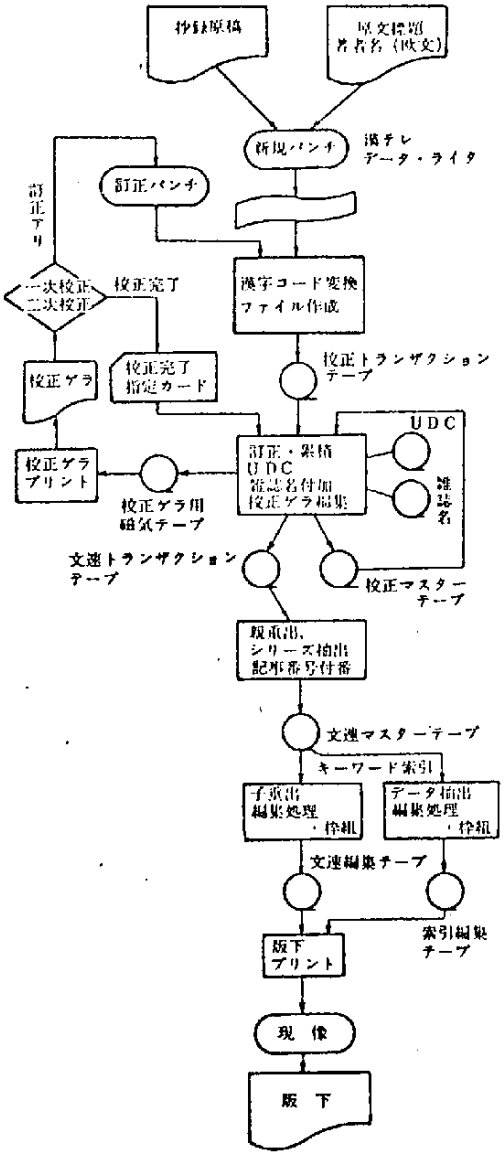


図 6-1-4　編集システムフローチャート

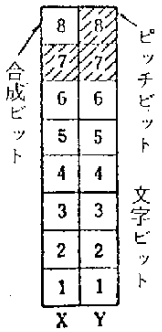
(1)　漢字コード変換とファイル構成

インプットには2種類の紙テープがあり，また前述のように機能コードを使用しているので，これらをプリンタ側で要求されている漢字コードに変換する。漢字コードは図6.1.5に示すように1字16ビットで構成されている。X1～6，Y1～6の12ビットは文字コードで，これは書体シフト「標準」の場合のせん孔機のコードと同じである。

書体シフトがイタリックまたはゴシックの場合は，書体が異なることはプリンタ側では別文字となるので，それぞれ新たな文字コードに変換される。上下シフトと合成はX7,8,Y7,8のファンクション

ビットを図6.1.5に従ってオンにする。

また本システムは最終的に版下を植字するため各文字に文字幅（ピッチ）が与えられているが，これをファンクションビットによって指定する。ピッチ6／6は全角（縦と横の長さが同じ）で日本文字はすべて6／6である。



X7	Y7	Y8	
0	0	0	ピッチ6／6
0	0	1	〃 5／6
0	1	0	〃 4／6
0	1	1	〃 3／6
1	0	0	〃 2／6
1	0	1	〃 1／6
1	1	0	上ツキ
1	1	1	下ツキ

図6・1・5 漢字コード

英字せん孔機でインプットされた文字のコードは，同じ文字の漢字コードに変換される。以上の変換によりすべての文字は内部漢字コードとなりこのコードによって最終的にプリンタへ送られる。

ファイル構成は文献情報には可変長のデータが多いため，セグメント方式をとっている。かかる方式をとったのは，FACOM230－50がワードマシンであるためである。文字からファイルまでの構成を図6.1.6に示す。

セグメントは文献の内容項目により定められており，インプット段階から指定されている。

A　原稿番号など

B　雑誌番号，巻，号など

C　分類コード

D　ページ

E　著者名

F　原文標題

G　和文標題

H　キーワード

I　抄録

(2)　校正処理

ファイル化の終ったデータは，UDC（分類コード）と雑誌略名を付加して校正処理に入る。本システムでは，せん孔にあたって検孔を行なっていないので，データのミスの訂正はすべて校正処理によって行なう。

まずゲラのフォーマット編集を行ない，PT－700で図6.1.7に示すゲラをプリントアウトする。これはセグメント，レコード別に表示され，さらに各レコードの中を10文字ずつのアイテムに区切ってある。図6.1.7に示したように，1レコードの情報領域は20Wからなり40字分であるから，各レコード

は4つのアイテムからなっている。

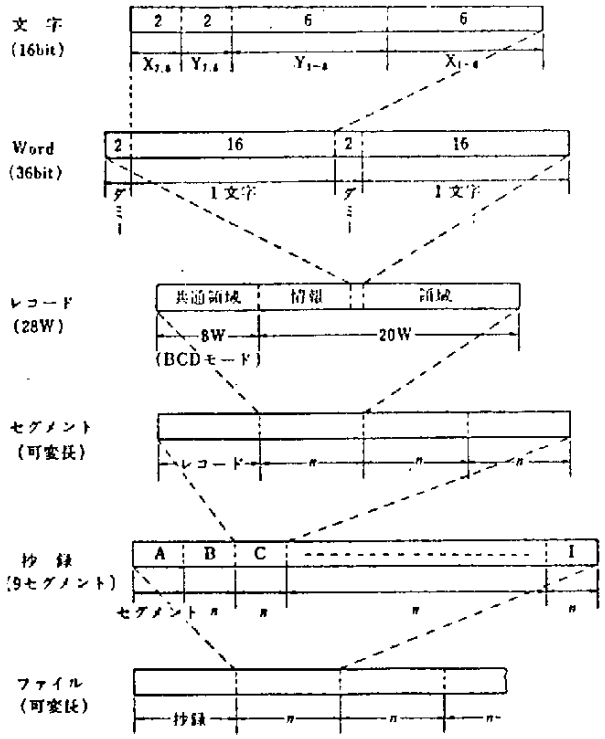


図6-1-6 ファイル構成

校正はゲラの目検によって行なう。図ではI 010レコードのⓀアイテムに誤りが発見されている。これを訂正するには誤りの個所（アドレス）と正しいデータを，紙テープにせん孔してインプットする。本システムではアイテムを単位とし，アイテムの内容全部を正しいデータと入れ換える方式をとっている。

したがって訂正のせん孔は次のようになる。

SM	AⒶ03769002Ⓑ538201Ⓒ5
SM	I 010Ⓚレス鋼線の一般的

1行目はこの原稿のアドレスで，1字の訂正に36ストロークのせん孔が必要となっている。

訂正紙テープをインプットしてファイルを訂正した場合には，その部分のみのゲラ（訂正確認ゲラ）を出力して確認を行なう。校正は2校まで行ない，また校正処理は2日分をまとめたバッチ処理で行なっている。

校正が完了すると，1件ごとに校正完了指定カードによって編集工程に入る。

(3) 編集処理

親重出：文献速報は化学，機械などの8シリーズの分冊で発行されるので，2つ以上のシリーズに掲載される記事は複製される。

シリーズ抽出：版下作成日がくると文献速報1冊分の記事が抽出される。

記事番号付番：1冊分の記事を分類コードによって掲載順に配列し，一連の記事番号を付ける。

子重出：1つのシリーズの中で，記事の内容から2つ以上の分野に関係ある記事を複製する。

編集と枠組：印刷形式に整える段階で，前もって定められたフォーマットに従って，必要なデータを各セグメントからもってきて，1行単位の印刷イメージ形式のレコードを作る。処理としては文字ピッチの計算と行末そろえ（1行は全角31文字分），図形スペースの確保，欧文のハイフン付け，禁則処理，化学記号などの分離禁止処理，見出しの漢字コードの挿入，文字の大きさの指定などが含まれる。

```
SM A Ⓐ 0 3 7 6 9 0 0 2 Ⓑ 5 3 8 2 0 1 Ⓒ 5 |  Ⓓ 0 5 Ⓔ 3 6 9
SM B 0 1 0 |  |  |  Ⓕ 7 1 0 2 1 0 Ⓖ |  |  |  Ⓗ c 1 Ⓚ |  Ⓛ G 8 6 6  Ⓜ J |  |
|  |  |  |  |  |  |  |  Ⓝ |  |  Ⓟ |  |  Ⓡ 7 0
SM B 0 1 1 |  |  |  J |  ▲ねじの世界RM
SM B 0 2 0 |  |  |  Ⓕ 0 0 0 0 Ⓖ 0 0 4 Ⓗ 0 0 6 Ⓚ 0 0 0 Ⓛ 1 4 . 下期別冊RM
SM D 0 1 0 |  |  |  Ⓕ 2 6 - 3 3 RM
SM E 0 1 0 |  |  |  Ⓕ せきね |  |  |  |  |  |  Ⓖ 関根義行RM
SM G 0 1 0 |  |  |  Ⓕ 特集：ステンレス鋼ねじの生産状況 |  ねじ Ⓗ 用ス
|  |  |  |  |  |  |  |  テンレス鋼線の選Ⓚ択RM
SM I 0 1 0 |  |  |  Ⓕ ステンレスねじの成形Ⓖ加工で最も問題の多い Ⓗ 冷
|  |  |  |  |  |  |  |  間圧造ねじ用ステンⓀレス鋼線の一般的な様▲
SM I 0 2 0 |  |  |  Ⓕ 質と，ステンレスねじⓖの内包する問題につい Ⓗ て触
|  |  |  |  |  |  |  |  れる。鋼線の種類Ⓚとその特性（SUS5▲
SM I 0 3 0 |  |  |  Ⓕ 1．SUS24．SUⓖS27・62・63．Ⓗ 18Cr9NiCu）Ⓚ，
|  |  |  |  |  |  |  |  鋼線の仕上げ処理．▲
SM I 0 4 0 |  |  |  Ⓕ 冷間圧造技術の開発とⓖ鋼線の課題．の各項目 Ⓗ 別に
|  |  |  |  |  |  |  |  解説RM
SM H 0 1 0 |  |  Ⓕ |  Ⓖ ねじ製造RM
SM H 0 1 1 |  |  |  Ⓕ ネジセイゾウRM
SM H 0 2 0 |  |  Ⓕ |  Ⓖ ステンレス鋼（線材）RM
SM H 0 2 1 |  |  |  Ⓕ ステンレスコウRM
SM C 0 1 0 |  |  Ⓕ |  Ⓖ M K 0 8 0 0 1 1 0 ▲機▲ 6 2 1 . 9 9 |  RM
SM C 0 2 0 |  |  Ⓕ |  Ⓖ G A 1 0 0 0 1 1 0 ▲金▲ 6 6 9 . 1 4 . 0 1 8 . 8 |  RM
SM C 0 3 0 |  |  Ⓕ * Ⓖ G A 1 0 0 0 0 5 8 ▲金▲ 6 6 9 . 1 4 - 4 2 6 / - 4 2 7 |  RM
```

図6-1-7　校正ゲラ

禁則処理は主として行頭および行末の体裁を整えるための，禁止事項の処理である。以上の処理で1行ごとにレコード化され，JEM3800側ではこのレコード単位にプリントするのみでよい。行の編集が終ると文献速報1ページ分ごとの，ページ枠組の処理を行なって完了する。

以上が抄録誌としての編集処理であるが，この他に記事番号付番に続いて，キーワード索引の編集も行ない文献速報の末尾に掲載される。また1年分の記事をまとめた年間索引の編集も行なっている。

35 図6.1.8にJEM3800で植字した記事のサンプルを示す。年間の処理件数は抄録約40万件である。

621.311.2/.4.004　621.311.16.004　**42026388**
電力系における水・火力発電所間の負荷配分決定の2方法〔a①〕
〔Int〕 F
Deux methodes pour determiner la repartition de la charge entre des centrales hydrauliques et des centrales thermiques dans un systeme electro-energetique.　POŽAR H. KEGLEVIĆ J : H323 Trans World Energy Conf 7th〔15〕197-225('69)
累積貯水池を備えた水力発電所間の負荷配分を決める方法を示した。両方とも,同一条件から始めて,最小経費を得るような方法で負荷配分を決めるという同一目的をもっている。それには漸近的にできる限り最適な解答を得るように進めなければならない。最適の解答への接近のみならず,所望の繰返し数に関しても数値的実例を示して両者を比較した；表11参4

図6-1-8 記事の写植サンプル

6.1.6 今後の問題点

冒頭にも述べたように，文献速報の自動編集システムは，JICSTの総合情報提供サービス・システムの一環であり，情報検索という面からみるとその蓄積工程にあたっている。したがってこの蓄積テープを用いて，分類またはキーワードによる情報検索が可能である。

しかし現状では，一般に漢字モードによる処理は困難であるので，情報検索用としては上記蓄積の中から，抄録以外のデータを，BCDコードまたはEBCDICコードに変換したファイルを作成し，実験的提供を開始している。

わが国においてはコンピュータによる情報検索は，やっと実用化の時代を迎えたばかりで，その評価もまちまちであるが，実用化には2つの問題点がある。1つは磁気テープに収録されている情報の内容が，従来印刷された形態で存在していた内容よりも希薄であること，1つはわが国の特殊事情として日本語の取り扱いが困難であることである。

この2つはともにコンピュータ側の制約に由来するものであるが，前者は今後ハードウェア上の大きな発展があって，初めて解決される性質のものである。後者については，コンピュータの制約を打破するか，または日本語を欧米なみの表音文字による表記に変革するかで，議論の分かれるところである。

しかしこの問題については，表音文字つまりカタカナまたはローマ字でインプットして，必要な場合には漢字かなまじりの日本語でアウトプットするという，巧妙な方法の目鼻がついているので将来としては大きな問題とならないと思われる。しかしいずれにしろ，日本語という面からみて，現在の情報検索には問題点が多い。

次に本論で述べた文献速報の編集システムにおける問題点は，校正システムである。ゲラ出し→校正→訂正パンチ→ゲラ出し…と繰り返される工程のため，校正完了となるまでのデータのコンピュータ滞留時間が非常に長く，1カ月以上かかるものもある。

このネックを解消するため，漢字ディスプレイ装置を用いた校正実験なども行なわれているが，コ

スト面から実用化は当面困難である。理想的にはインプット以前にミスを取り除けばよいわけで，その方面からのアプローチもいろいろと考えられているが，植字システムという面からは1回の校正は必要であろう。したがってそのときまでに，機械的なミスが絶無となるような方式が目標であろう。

将来の人件費の上昇，マシンコストなどから考えて，日本語処理を確立するためにも，最も努力を傾けねばならない課題である。

最後に本システムも本格稼働から約3年を経過し，文献速報の編集植字も軌道に乗ってきており，各号のキーワード索引，年間索引の発行期間の短縮など，ユーザ・サービスに大いに役立っていると確信している。

またその間，環境公害文献集の発行などにも，機械化システムの威力を発揮し，現在日本語シソーラスの編集にも活躍中である。

6.2 工業技術情報に関する日本語情報処理の一例

6.2.1 日本語による技術情報処理の背景

昨今，社会情勢の変革する加速度は大きくなりつつあり，1970年代に至って激動の時代とまでいわれている。技術の進歩は単に物の盛衰のみにとどまらず社会情勢をも左右し，さらにそれがフィードバックされて再び科学技術の進展に関係するという複雑な輪廻の様相を呈することになる。製品とは直接関係のない製造途上の処理工程で生ずる外部環境との接点での問題も従来無関心であったものが，昨今公害問題として提起される次第となっている。

種別 / 年別 件数	特許 件数	特許 指数	実用新案 件数	実用新案 指数
昭和7－11年平均	15,532 (6,817)	100.0	37,476 (1,120)	100.0
32	33,188 (7,896)	213.7	58,759 (1,145)	156.8
33	38,518 (9,613)	284.0	70,719 (997)	188.7
34	41,537 (11,587)	267.4	68,102 (1,069)	171.7
35	43,484 (13,659)	280.0	63,077 (1,383)	168.3
36	48,417 (16,101)	311.7	65,685 (1,264)	175.3
37	60,127 (17,914)	387.6	81,858 (1,424)	218.4
38	71,790 (19,424)	462.2	99,627 (1,712)	265.8
39	74,980 (21,127)	482.7	102,995 (2,118)	274.8
40	81,923 (23,084)	527.4	108,553 (2,071)	289.7
41	86,046 (23,643)	554.0	119,061 (1,909)	317.7
42	85,634 (25,596)	549.6	111,060 (2,033)	296.3
43	96,710 (28,454)	622.7	114,785 (2,463)	306.3
44	105,586 (30,309)	679.8	124,170 (2,369)	331.3
45	130,831	842.3	142,036	379.1

注：表中の（　）内の数字は外国人の出願件数

表 6・2・1　昭和7～11年平均に対する出願件数指数表

種別 / 年別 / 月別	特許 44	特許 45	実用新案 44	実用新案 45
1	7,032	**8,415**	8,318	**9,710**
2	7,990	**8,588**	9,369	**9,826**
3	9,315	**9,943**	10,827	**10,942**
4	8,897	**9,839**	10,889	**11,623**
5	9,194	**9,578**	11,036	**10,926**
6	9,034	**10,500**	11,260	**11,872**
7	8,905	**10,140**	10,882	**11,409**
8	8,246	**8,946**	9,647	**10,142**
9	9,133	**9,574**	10,854	**10,409**
10	9,508	**10,314**	10,288	**11,247**
11	8,380	**9,577**	9,890	**10,688**
12	9,952	**25,417**	10,910	**23,272**
総計	105,586	**130,831**	124,170	**142,066**
対前年比	109.2	**123.9**	108.2	**114.4**
年間平均	8,799	**10,903**	10,348	**11,839**

表 6・2・2　月別出願件数表

また，一方では企業の国際化，労働力不足，通貨関税問題，さらには資源枯渇による原料コスト高あるいは入手困難という複雑な社会環境の中で，企業が発展していく所産として生ずる膨大な量の情報をいかに処理していくかが企業の盛衰のひとつの大きな鍵として問題となっている。現に最近の技術の特許出願を見ても増加の一途をたどり，テンポの早い社会情勢変化に鑑み日本でも1971年を境に公開公報と公告公報の2本立てとなり，出願および公告の件数の年度別推移ならびに最近2カ年間の月別の推移は表6.2.1および表6.2.2のようになっている。この様子をグラフで示すと図6.2.1のとおりである。年度によって増加率は負になっていることもあるが，全般的な傾向として変化率は正勾配となっているのが明瞭である。

このような大量な情報に対する索引作成等情報サービス業務が，従来と同様人力に頼っていたのでは，仕事量の増大にしたがってデータの信頼性が失われる危惧が生じてくる。またその処理を機械化していっても，これまでのように常用言語表現をわざわざコーディングし直してカナ化していたので

は，なおさらエラーが増大し，かつ同音異語，同音異人，同一語の読み違いによる同一語異語発生，読み難さ，音標文字による直感力欠除等，カナ文字処理の欠陥がかなり目立っている。さらにユーザ・サイドの情報処理機械，機材，形態が多様化している。

そこでユーザ・サイド，ドナー・サイド双方によって，一貫した情報処理，サービス管理の行なえる新しいシステム・インターメディアを用いたシステムが要求されるようになった。

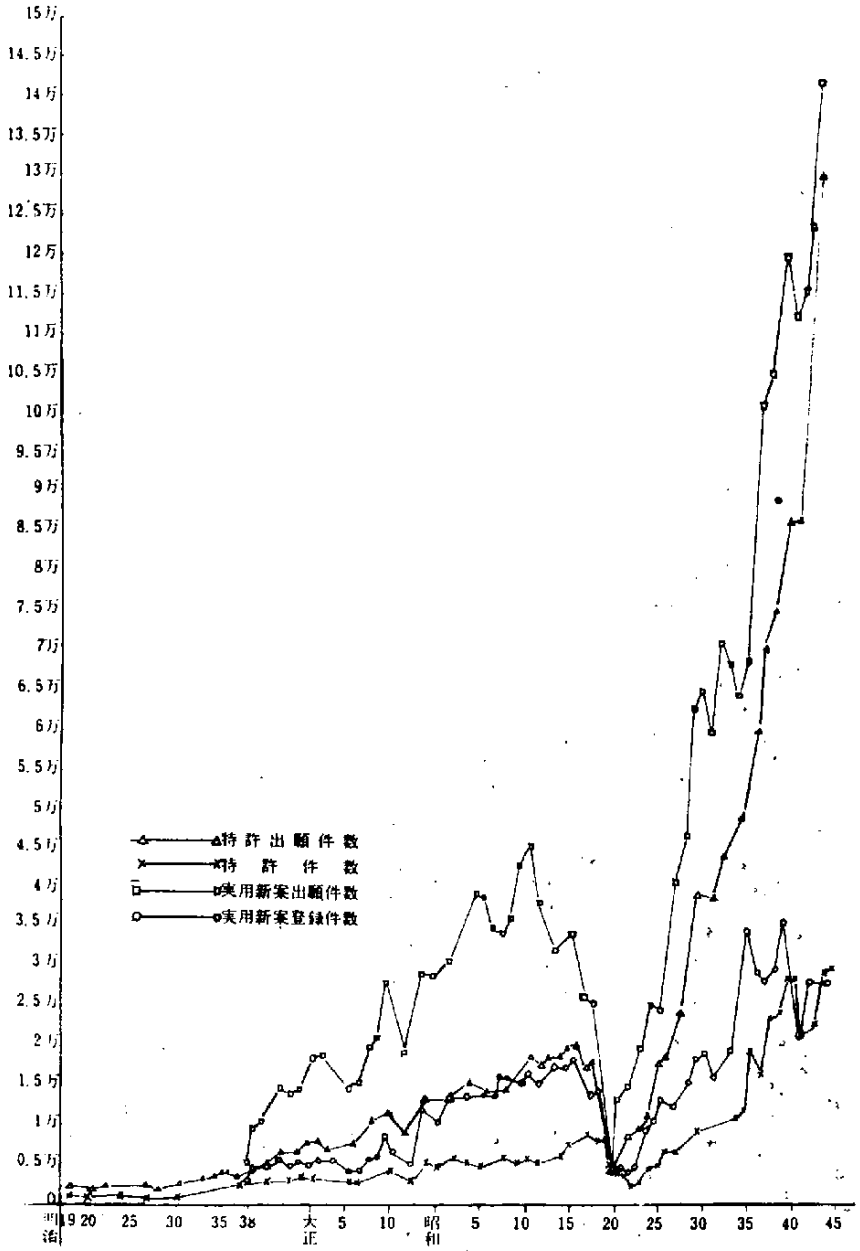


図 6・2・1　特許、実用新案の出願および登録の推移

しかしながら，ある程度まで自動化された漢字文字処理機材は，本来情報処理機材としてよりもむしろ図6.2.2に示すような経路をたどって組版印刷の自動化が行なわれてきた所産として生れてきたものだけに，それを情報処理インターメディアとして汎用するためには種々の工夫を必要とし，問題点も多くあった。

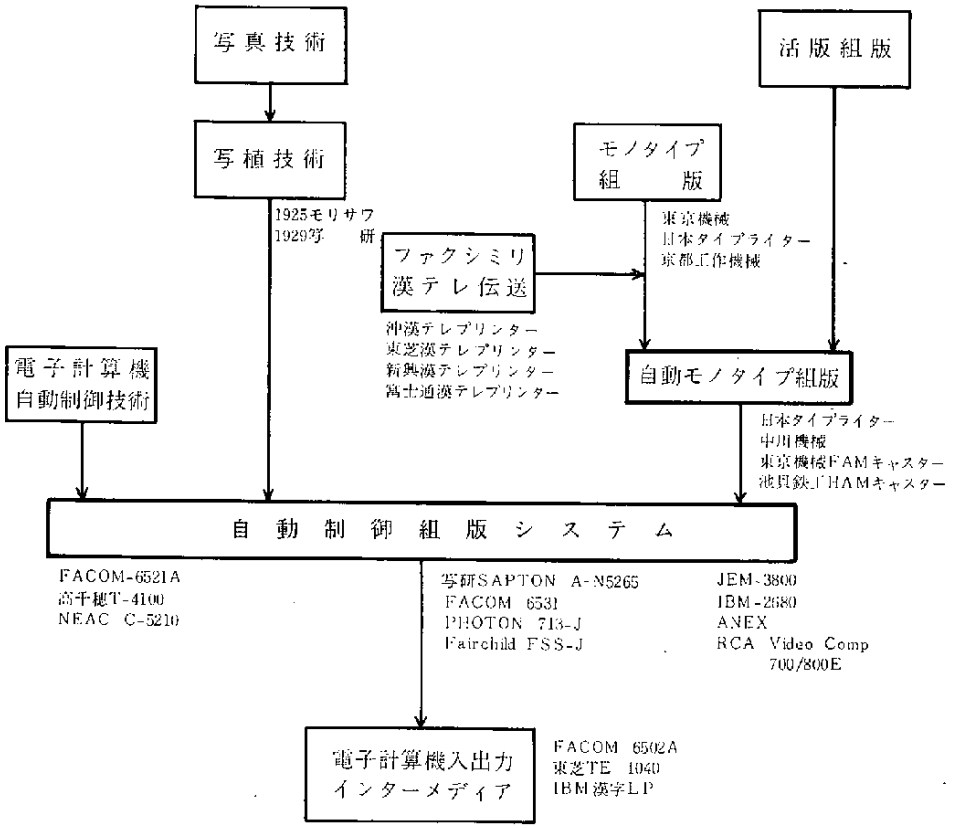


図6・2・2　自動組版システムと情報処理インターメディア

6.2.2　システム設計

(1)　システムの特徴

㈱特許データセンターでは，組版印刷の自動化のために発展してきた漢字文字処理機材を，新しい情報処理のインターメディアと見なし，技術情報サービス処理および管理システムに組み入れ，大量な特許，実用新案技術情報の索引作成，情報サービス，販売管理等のデータ汎用システムを，1970年夏に企画し，1971年1月のデータからこのシステムに切り換えた。

この工業技術情報汎用情報処理システム（GIIPSY：Generalized Industrial Information Processing System）のシステム設計上の重要なポイント，は図6.2.3に示すような3項目の最適化条件に基づいて計画された。

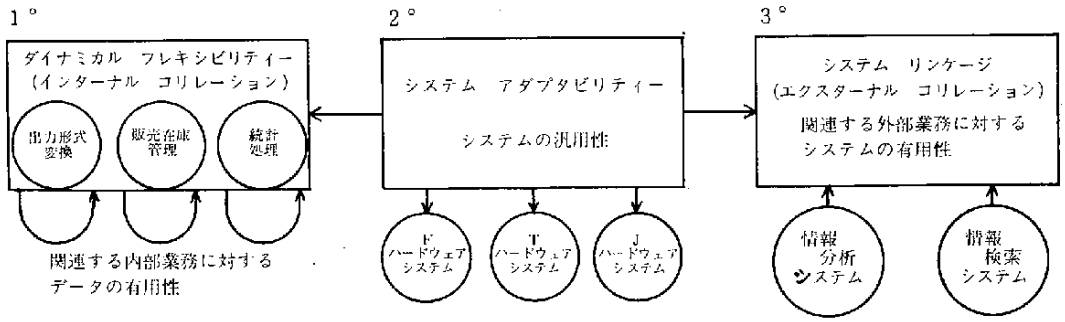


図6-2-3 システム設計における最適化条件

その才1は，システムで扱うデータを単に情報サービスのためのインターメディアとして用いるだけでなく，その情報サービスに関連する業務の管理にも役立て，システムのダイナミカル・フレキシビリティを高めること。

才2にユーザの情報処理システムの多様性に対してもデータが十分役立つような出力形式をとることができ，処理ハードウエアのシステムの限定を極力避けること。

才3にこのGIIPSYで作成されたシステムが関連するデータも処理でき，書誌的事項に内容データとしての単語ないし要約が付加され得る汎用データ処理システムとして，見出し語レベルを越えた常用言語（含漢字）による技術情報分析・検索システムとシステム・リンクされ得ること等である。

システムを企画した当初は，漢字文字処理の行なえる商用機として可能性のあったものはF社のシステムのみであった。それゆえシステム構成の焦点を，組版印刷用に稼動されているシステムを，いかに汎用の情報処理システムの中で有利に用いるかということに向けながら検討を加えた。その主な点は，作業能率，コスト・パフォーマンス，信頼性，他社ないしは他の形式の漢字文字処理システム出現時の流動性であった。その結果，組版用に設計されたFCLシステムの持つ頁処理用中型計算機F－270－20および出力装置制御用ミニコンF－R等の電子計算機は，ユーザに公開されないシステムインターフェイスを合わせるためのみに用い，編集機能等は汎用の電子計算機で，データ処理として扱った。したがって入力からすべての処理を汎用機を用いて実行するようにし，MTベースでシステム間インターフェイスを合わせた。

(2) 文字種の選定

このGIIPSYで必要とする文字種は特許という特殊な情報の性格から，かなり種類を多く要求され，出力時には科学記号の表現まで可能でなければならなかったが，その要求が自動組版の機能性とマッチしてかなりの必要度がリカバーされていたので，他の入力機器の選択性を許されない環境下で必然的に決まってしまった格好であった。それでも後続の他社のシステムよりも好都合であった。字種の問題よりもむしろ，ベリファイ機能のない入力システムで入力したデータをどの様にメンテし，データの信頼度を高めるかという，ソフトウェア上にもたせる機能の設計に苦慮した。

6.2.3 GIIPSYシステム構成概要

GIIPSYシステムの使用するハードウェアおよびそのソフトウェア構成はそれぞれ次のようになっ

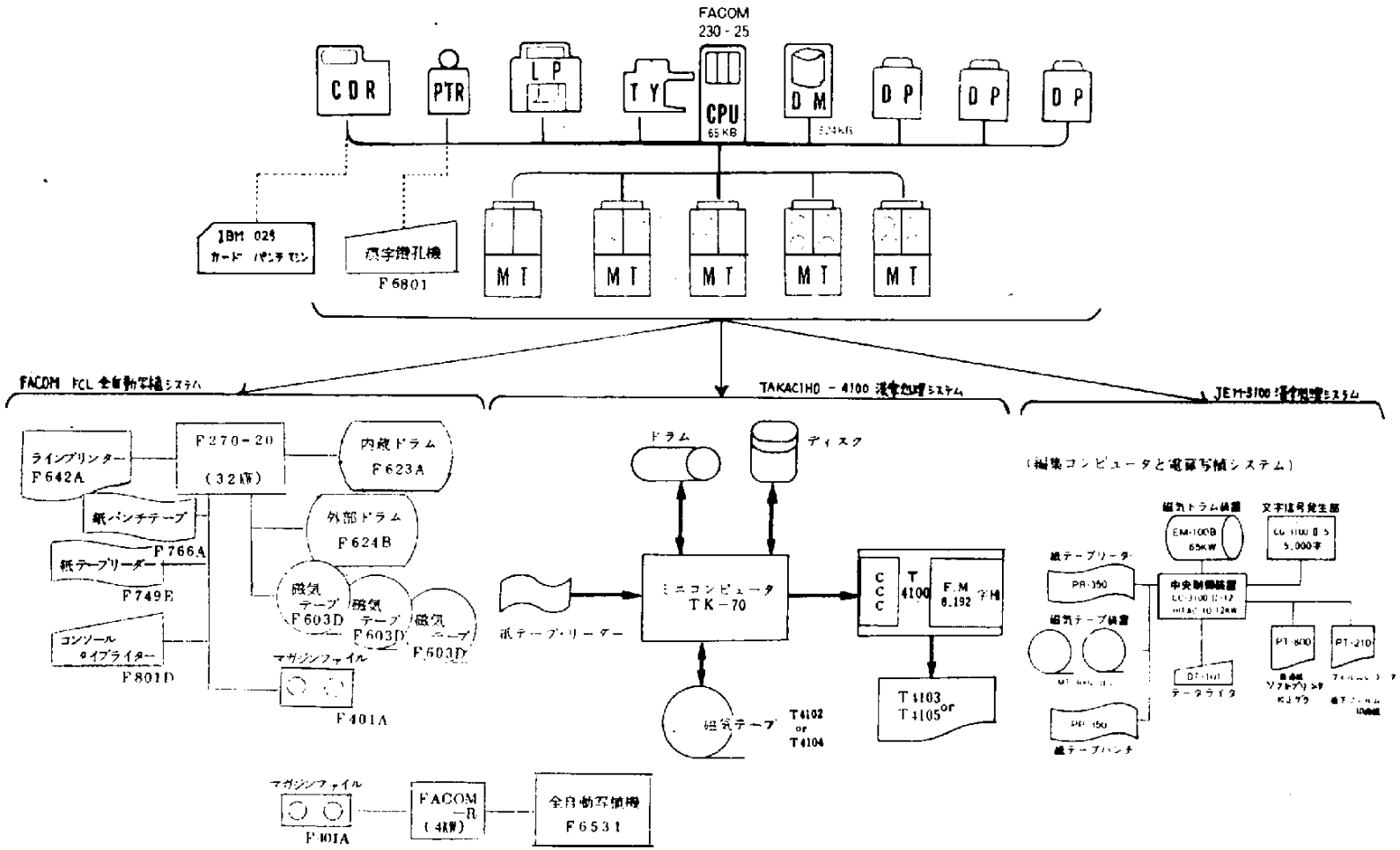


図6・2・4 P.D.C.-GIPSYシステムで用いるハードウェア構成

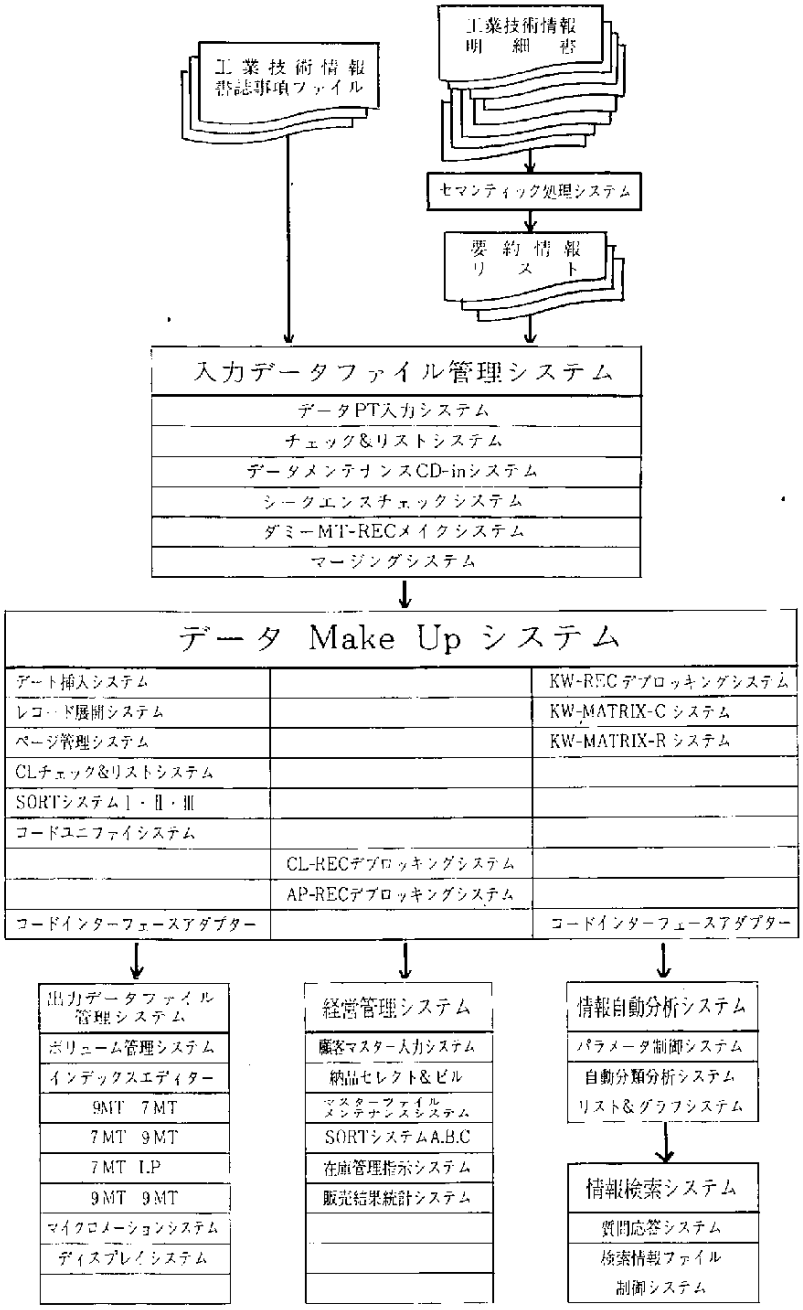


図 6・2・5 工業技術情報の漢字文字処理による汎用情報処理システムのソフトウエア構成

ている。

(1) ハードウェア構成

GHPSYシステムでは，データおよびシステムの汎用性を持たせるために，プログラム・ソフトウェアおよびプロセスを大別して〝データ入力〟，〝処理〟，〝出力〟の3ステップに分類し，現在商用ベースの漢字文字処理システムが，出力，印刷を主体としたオフライン機構で汎用データ処理機能を持たないことから，すべてのデータ処理を行なう部分と出力部分とをとくに分離した。

入力処理，データ処理，出力処理に用いるハードウェア構成は，図6.2.4に示すA.B.C.D.4つのブロックからなっている。PT／CD漢字データ入力，データ管理，出力管理等データ処理はすべてFACOM230－25をベースにしており，出力は適宜用途やデータの必要となる時間的タイミングに合わせて，FACOM－FCL－AFTSシステム，T－4100システム，JEM－3100システムの各システムへデータを落すようになっている。その概要は次のようである。F－FCL－AFTSは印刷用下版フィルム，T－4100はゲラないしはペーパー・コピー出力，JEMはMICRO－FILM出力用に用いている。

(2) ソフトウェア構成

このシステムは図6.2.5に示すように，データおよびシステムを種々の角度から利用できるように，全体で5ブロックのプロセスからなっており，各プロセスは処理を円滑に進めるために，図6.2.5の各プロセスのテーブルに示すようなサブシステムからなっている。5ブロックのシステムは

① 入力データ・ファイル管理システム

② データMake Upシステム

③ 出力データ・ファイル管理システム

④ 経営管理システム

⑤ 情報分析・検索システム

である。これらのシステムはすべて汎用機FACOM230－25を主体として作られている。使用言語はアセンブラおよびコボルを主としているが，情報分析システムはフォートランを基本としている。

システム中では処理を効率的に行なうため，およびコンパイラでは稼動させ得ない部分に対しては，適宜アセンブラとのリンクを行ない，システム作成の合理化，速度向上等に効果を発揮している。漢字文字処理の結果の出力の主流を占めるのは，インデックス作成に関する出力データ・ファイル管理システムである。図6.2.6のフローチャートは，〝入力〟，〝Make Up〟，〝出力〟の処理ルートを通って索引データ処理を行ない，索引を編集作成するプロセスである。

6.2.4 漢字入力データ・ファイル管理システム

(1) 漢字入力装置

入力装置の選定は，企画当初意図的に選択できる商用ベースの漢字処理システムとして，一貫した機能を有するものがとくになかったが，字種，使いやすさの点からFACOM－6801Bを選定した。始めはこの紙テープさん孔装置のみによってすべてのデータを扱うつもりであったが，穿孔ミス等によるデータ・エラーの修正等のことを考慮して，少量のデータを扱う場合は，IBM－029パンチ・マシンをも使用し，カードによる入力も可能にした。したがって入力装置としては，F－6801B漢字せん

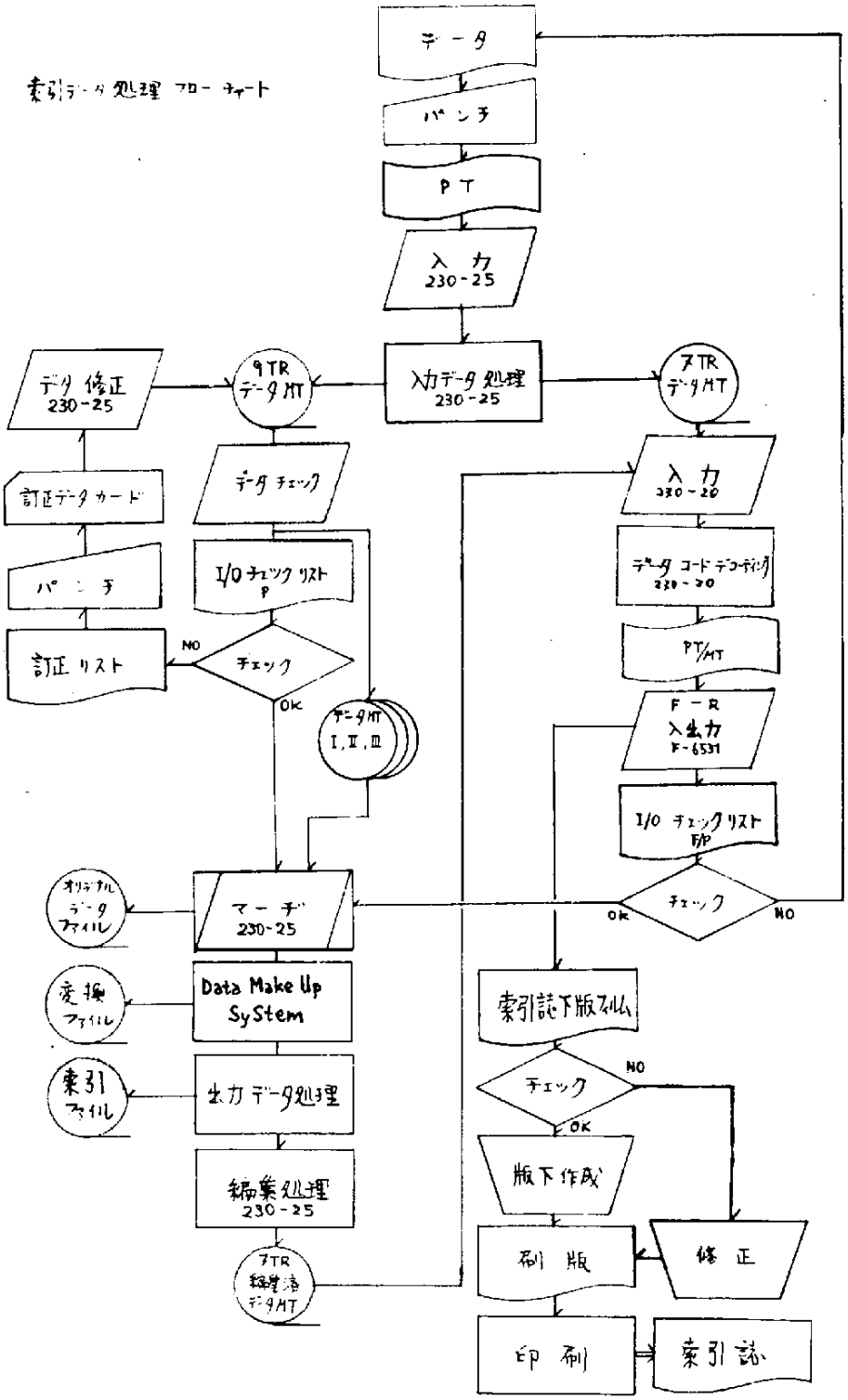


図 6・2・6　索引データ処理フローチャート

孔機およびIBM－029カードパンチ装置をシステムに組み入れた。とくにALPHA－かな－NUMERIC表現のデータ・アイテムの場合には，カードで入れた方が能率的でベリファイ機能もあるので，エラーをあらかじめ防ぐことができ，信頼度が高く，速度も速い。漢字せん孔機の中でベリファイ機能のあるものがないので，入力データのチェックに負担がかかる。パンチ能力は3000～3500字／時間（6カ月以上の経験），2500～2800字／時間（3カ月程度の経験）である。入力データレコード数に対するエラー・レコード数は，漢字：数字／30：30ほどの文字比で漢字のエラーの方が率が高い。ちなみに図6.2.7に最近1年間のデータ件数に対するエラー・データを含むレコード数の比を示した。

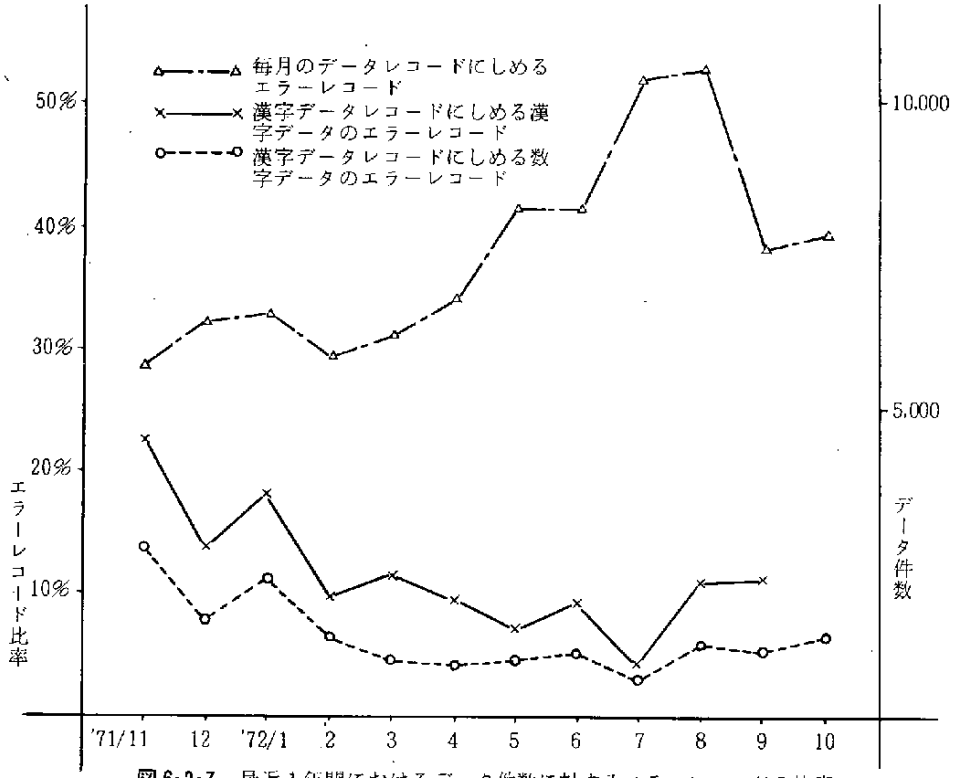


図6・2・7　最近1年間におけるデータ件数に対するエラーレコードの比率

(2)　入力データ・ファイル管理システム

種々のデータ・ソースに対し，多角的汎用化を可能にし，関係のある一連の業務の機械化処理のソース・データに役立てることができるように，入力システムに機能的な柔軟性を持たせてある。したがってこの入力システムでは，書誌的事項以外の文章内容を有する漢字データも入力でき，システム相互の関連をはかった。

図6.2.8は書誌的事項入力リストの一例であり，要約情報入力リストの一例はつぎのようなものである。

1001.　銑鉄，噴霧，ガス流，衝突，帯域，炉床，排出孔，スラグ，向流，鉄粒子，酸化鉄。

1002.　噴霧，精製，耐火，ライニング，損耗，溶銑，ガス流，炉壁，排出，冷却，霧化。

この入力データ・ファイル管理システムは次の6セクションのサブシステムで構成されており，

データがただちに業務に利用できるようデータメンテナンスまでこのシステムで行なう。

4-1727　昭和47年1月11日発行の実用新案出願公告目次　47-8

公告番号	分類	願書番号	実用新案の名称	出願人 都府県国籍	出願人 氏名(名称)
580	〃	41 — 88889	電動機鉄心の盗難用保持具	大阪	三洋電機株式会社 外1名
581	55 A 05	40 — 20305	回転電機	東京	三菱電機株式会社
582	55 A 051	19860	大容量直流機の整流子カバー装置	同	株式会社日立製作所
583	〃	23281	回転電機のカバー装置	同	同
584	55 A 054	59492	抜け止め装置を備えた機	同	同
585	55 A 311	38 — 99025	単相同期電動機	大阪	松下電器産業株式会社
586	55 A 44	37 — 74180	インダクタ型多極同期電動機	同	同
587	55 B 1	39 — 52854	自励複巻発電機励磁調整回路	神奈川	富士電機製造株式会社
588	55 C 2 108 D 11	41 — 89234	電動機を信号毎に定量回転せしめる装置	愛知	愛知時計電機株式会社
589	60 B 011	38 — 26624	コードリール	大阪	松下電器産業株式会社

図6・2・8　書誌的事項入力リスト

① データ紙テープ入力システム

② チェック&リストシステム

③ データメンテナンスカード入力システム

④ データ・シークエンス・チェックシステム

⑤ ダミーMT－REC作成システム

⑥ マージング・システム

入力システムの中でとくに綿密な機能を考慮したのは，入力システムとデータ・チェックシステムである。2 Byte 1文字という漢字コードの特殊性から，パンチ・マシンのスルー・エラーや入力エラーによる後続文字データへのデータ・キルを避けるため，バイトずれ検出・修復機能にとくに注意してあり，図6.2.9に示すように，数文字のキルで後続データのリカバーが可能である。

チェックシステムでは，人間の視覚を補助するための種々のデータ・チェックを自動的に行なって，リスト・アウトするようになっている。資料の図6.2.10，図6.2.12にその一例を示した。図6.2.9は入力ソース・データ・チェック・リストで，レコード番号582の右の〝N〟は出願人アイテムにエラーがあること，〝?〟はこのレコードに外字があることを示し，583の〝#〟は582と583の間にレコードがひとつ，ないしそれ以上抜けていることを示している。ちなみに図6.2.11の漢字チェック・リストを見ると，レコード番号582の583データの出願人アイテムにエラーがあることがわかる。図6.2.13はメンテ・データのチェック・リストで，図6.2.12と合わせてみて，修正した項目のフラグがリスト上になければ，正しく修正されていることを示す。マージ後のリスト図6.2.14を見ると正しくメンテされている。

▶3824▶126F3▶134B261▶460209▶40-97004▶衣服、反物文庫▶東京▶東レ株式会社
▶3825▶126G01▶460209▶41-74653▶雑品引掛具▶大阪▶堀井和夫
▶3826▶126J7▶460209▶79993▶梯子又は類似物▶端内▶ジョーン・ルイス・ラヴアラード
▶3827▶126K220.6▶67D31▶128F23▶90C92▶67M09▶460209▶43-97812▶湯沸器の空焚防止装置▶静岡▶土屋勝次
▶3828▶126K220.6▶67E9▶67M09▶460209▶100189▶通隔操作装置付バルブを有するガス風呂用安全器▶東京▶[illegible]
[illegible]

16 6 25 14 25 0 SAMPLE 15 PAT000 JOB1 2

DFF2

DF2F/3F/6 9 5F/78

78F6

▶3829▶126K512▶132C51▶460209▶38 51946▶家庭紙の容器▶大分▶大分製紙株式会社
▶3830▶126L61▶126E203.4▶460209▶41-35330▶米びつ▶東京▶丸山工業株式会社
▶3831▶126M203▶460209▶40 49928▶扇風機首振制御装置▶大阪▶シャープ株式会社
▶3832▶126M204▶460209▶15645▶扇風機の支柱取り付け装置▶東京▶株式会社日立製作所
▶3833▶・▶460209▶52267▶扇風機▶同▶同

図6-2-9 バイトずれ検出のためのリスト

579		580	*				41-88889
580		581	55	A05			40-20305
581		582	55	A051			19860
582	N ?	583	*				23281
		584>	55	A054			59492
583	#	585	55	A311			38-99025
584		586	55	A44			37-74180
585		587	55	B1			39-52854
586		588	55	C2	108	D11	41-89234
587		589	60	B011			38-26624
588		590	60	B4			14119
589		591 <<<	60	C02			43-106344

図6-2-10 入力ソース・データチェックリスト

▶580▶電動機鉄心の磁装用保持具▶大阪▶三洋電機㈱外1名
▶581▶回転電機▶東京▶三菱電機㈱
▶582▶大容量直流機の整流子カバー装置▶〃▶㈱日立製作所
▶583▶回転電機のカバー装置▶〃▶〃584
▶585▶単相同期電動機▶大阪▶松下電器産業㈱
▶586▶インダクタ型多極同期電動機▶〃▶〃
▶587▶自動複巻発電機励磁調整回路▶神奈川▶富士電機製造㈱
▶588▶電動機を信号毎に定量回転せしめる装置▶愛知▶愛知時計電機㈱
▶589▶コードリール▶大阪▶松下電器産業㈱

図6-2-11 漢字チェックリスト

40	BG N	#	578	#				##
41	BGM	#	583	#				##
42		?	584	55	A054			59492
43	BG N	#	598	#				##
44	BGM	#	601	#				##

図6・2・12 データチェックリスト

▶578▶空気圧式圧力計の手動一自動切換装置▶ ▶
▶583▶ ▶ ▶同
▶584▶抜け止め装置を備えた[illegible]▶同▶同
▶598▶ケース軸受用圧入洗い・流し弁を具備した強軸タービンポンプ▶ ▶
▶601▶ ▶ ▶大阪ゴム工業株

図6・2・13 メンテ・データチェックリスト

580			580	*				41-88889
581			581	55	A05			40-20305
582			582	55	A051			19860
583			583	*				23281
584		?	584	55	A054			59492
585			585	55	A311			38-99025
586			586	55	A44			37-74180
587			587	55	B1			39-52854
588			588	55	C2	108	D11	41-89234
589			589	<<< 60	B011			38-26624

図6・2・14 マージ後のリスト

6.2.5 漢字データ・ファイル管理システム

(1) 漢字出力装置

GIIPSYの出力装置は入力装置をFACOMに選んだことと，出力リストが単物と複物とがあって，複物は印刷を必要とすることから，印刷にもむいているシステムということで，F－6531自動写植装置を主体としたものとなっている。しかしながらF－6531は少量で簡易印刷下版作成を行なうには非常に精度の高いもので，コスト・パフォーマンスが良くないということもあって，他社の漢字文字出力装置を用いる試みの必要性が生じた。そこで出力システムは，図6.2.4のB.C.D.のブロックにある次の3社のハードウェアを共用することができるようにした。

ブロックB. FACOM全自動写植システム

F－270－20，F－R－6531，F－KLP，F－KPR

C. 高千穂漢字処理システム

T－4100

D．日本電子産業漢字処理システム

JEM－3100

図6.2.15はFACOMシステムで出力したもの。図6.2.16は同じデータを高千穂システムで出力した索引誌の例である。

第 55 類

実用新案　　昭和47年出願公告細分類別総目録

分　類	公告番号	公告日	願書番号	実用新案の名称	出願人
55 A 0	1567	1.20	43 61403	小型電動式加圧ロール装置 ☆103 C 238	インターナショナル・ポラロイド・CORP
☆ 55 A 02	579	1.11	39 58834	鉄心溝を備えた電気機器	三菱電機㈱
☆ 55 A 02	580	1.11	41 88889	電動機鉄心の挿装用保持具	三洋電機㈱外1名
☆ 55 A 034	1149	1.17	43 7928	回転電機の電気接続端子	日本電装㈱
☆ 55 A 034	1150	1.17	43 19767	モーターのブラシ保持装置	三洋電機㈱
☆ 55 A 034	1779	1.21	42 10099	ワイパーモータのブラシ装置 55 A 42	㈱三ツ葉電機製作所
☆ 55 A 05	581	1.11	40 20305	回転電機	三菱電機㈱
☆ 55 A 051	582	1.11	40 19860	大容量直流機の整流子カバー装置	㈱日立製作所
☆ 55 A 051	583	1.11	40 23281	回転電機のカバー装置	㈱日立製作所
☆ 55 A 051	1151	1.17	43 17554	電動機点検カバー 63(5)H 303　80 C 0	㈱日立製作所
☆ 55 A 051	2093	1.25	42 55686	電動機 55 A 122	㈱安川電機製作所
☆ 55 A 051	2094	1.25	43 18480	カバー類装置 53 E 3　53 E 4	㈱日立製作所
55 A 052	1633	1.20	42 72955	電動機 ☆ 54 B 12　53 A 21	三洋電機㈱外1名
55 A 053	599	1.11	44 16167	電気掃除機 ☆ 63(3)F 11　53 E 3	松下電器産業㈱

図6・2・15 FACOMシステムによる出力

第 55 類

実用新案　　昭和47年出願公告細分類別総目録

分　類	公告番号	公告日	願書番号	実用新案の名称	出願人
55 A0	1567	1.20	43－ 61403	小型電動式加圧ロール装置 *103 C238	インターナショナル．ポラロイド．CORP
* 55 A02	579	1.11	39－ 58834	鉄心溝を備えた電気機器	三菱電機（株）
* 55 A02	580	1.11	41－ 88889	電動機鉄心の挿装用保持具	三洋電機（株）外1名
* 55 A034	1149	1.17	43－ 7928	回転電機の電気接続端子	日本電装（株）
* 55 A034	1150	1.17	43－ 19767	モーターのブラシ保持装置	三洋電機（株）
* 55 A034	1779	1.21	42－ 10099	ワイパーモータのブラシ装置 55 A42	（株）三ツ葉電機製作所
* 55 A05	581	1.11	40－ 20305	回転電機	三菱電機（株）
* 55 A051	582	1.11	40－ 19860	大容量直流機の整流子カバー装置	（株）日立製作所
* 55 A051	583	1.11	40－ 23281	回転電機のカバー装置	（株）日立製作所
* 55 A051	1151	1.17	43－ 17554	電動機点検カバー	（株）日立製作所

図6・2・16 T4100システムによる出力

ハードウェア上の問題点としては，出力装置そのものが完全なデータ処理機能を持たず，中央演算装置と直結されているケースが少ないので，システムの体系にかなりの工夫を要求されることであろ

う。

(2) 出力データ・ファイル管理システム

このブロックの機能は次の6つのサブシステムから構成されている。

① 索引編集システム

② ボリューム管理システム

③ 7トラックMT／9トラックMT相互変換システム

④ ファイル変換システム

⑤ マイクロメーション・システム

⑥ ビデオ・システム

このブロックのシステムは，データMake Upシステムによって作成されたファイルを，種々の用途にマッチした形式の出力を得るための機能を果す。

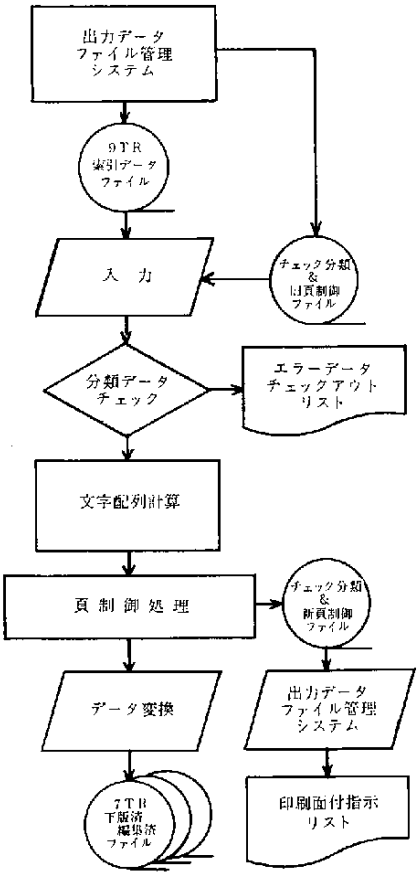


図6-2-17 編集システムフローチャート

索引を編集するインデックス・エディタでは，図6.2.17のフローチャートに示す手順でバイトマシンによって索引誌編集を行ない，キャラクター・マシンにデータを渡す処理を行なう。本システムで

使用する3種の出力システム・ハードウェアの中で，FACOM-FCLは7TR-MTを必要とするので，バイト／キャラクターのデータ変換を行なうことになる。データ量が増加する場合のためのマイクロメーション出力システムに，JEM-3100を稼動させる。VTR，EVRが普及すると，MICRO-EVR-VTR相互利用することが必要になると考えられる。

6.2.6 システムの運営

GIIP・SYSTEMは1971年1月のデータから入力，出力を実施した。データ入力は月単位で区切り，マンスリー・インデックスを発行している。処理は汎用電子計算機で出力できる部分は極力それを用いて行なうため，データの性格上漢字部分と数字部分のデータを，それぞれ別途に扱うことができる。運営状況を項目別に説明する。

(1) 入力データ量

毎月，特許・実用新案公告の平均8,000件から10,000件の書誌的事項を，目次リストにより入力する。1件当りの平均文字数は50～60文字程度で漢字：数字／3：2くらいである。

(2) 出力データ量

索引作成に用いるデータは，分類項目にしたがってデータ展開を行ない，15,000～23,000レコードのデータ量に転写し，図6.2.18のごときフォーマットのB5版サイズで編集・頁管理されて出力が行なわれる。

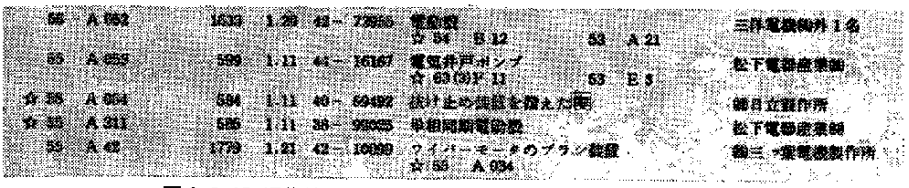


図6・2・18 編集済フォーマット

毎月の頁数は平均700～1,000頁となる。印刷は限定150部でK－オフ方式で刷られる。印刷の段階で出力下版図6.2.18に罫線をマスキングして，図6.2.15のような索引誌として出版する。毎月，当月のデータが翌月15日までに出版される。

(3)データMake Upシステムの機能

このブロックのシステムは，入力データ・ファイル管理システムで入力メンテされたデータを，このブロックのシステムに連携する次のような

① 出力データ・ファイル管理システム

② 経営管理システム

③ 情報分析・検索システム

の3つのシステムでそれぞれデータが利用できるように，オリジナルデータに加工をほどこし，ファイルを編成し直す機能を有する。出力機器に対応するコードに変換するのもこのシステムの中のサブシステムである。

(4) 経営管理システムの機能

索引の編集印刷の目的で入力されたデータを，公報の販売管理・在庫管理に用い，販売情報の検索選出;伝票作成;販売指示書，在庫管理指示書等の作成をする。このシステムを付加し，索引作成マンスリー・データを転用することによって，月間処理を行なっていたものを，バイマンスリー15日区切りで入力管理をする必要にせまられたが，逆に索引作成上のデータ管理がスピード・アップされ，スムーズに実行されることになった。また従来3万枚ほどの伝票を人手で作成していたものを，索引データを転用して自動作成することができる。したがってこれまでこの3万枚の伝票を数日ごとに集計手計算し，さらにCDパンチしていたものがほとんど割愛されるに至った。

(5) 情報分析・検索システム

書誌的事項のデータ処理を行なっている入力システム，データMake Upシステムをそのまま応用し，技術情報の内容を入力し，ファイルを作成する

▶1001▶銑鉄。噴霧。精製。ガス流。衝突。帯域。炉床。排出孔。スラグ。向流。鉄粒子。酸化鉄。▶ ▶
▶1002▶噴霧。精製。耐火。ライニング。損耗。溶銑。ガス流。炉壁。排出。冷却。霧化。▶ ▶
▶1003▶溶鉄。連続。精製。噴霧。貯蔵。銑鉄。スラグ。電磁。運搬。ガス流。添加剤。▶ ▶
▶1004▶精製。ガス流。噴霧。液状。銑鉄。衝突。帯域。石灰。自由。落下。吹付け。ランス。霧化。▶ ▶
▶1005▶機械。加工性。溶接性。時効。硬化性。ニッケル。クロム。鉄。合金。原子量。リング。圧力。容器。高温。応力。▶ ▶

図6・2・19 入力システムを用いて入力した意味データ

0012	000002	001002	0001	000012	000003	001003	0004	000012	000004	001004	0003
0013	000099	001410	0005	000014	000040	001080	0011	000015	000041	001081	0011
0016	000020	001060	0008	000016	000044	001084	0010	000016	000060	001100	0010

図6・2・20 分析用データファイル

図6.2.19は入力システムを用いて入力した意味データ，図6.2.20はMake Upシステムで処理された分析用データ・ファイル，図6.2.21は分析処理前，図6.2.22は分析後の関連データ・ファイル例である。相互に類縁関係の強い情報から順序立てて出力される。

従来のカナ－ALPHA－NUMERIC表現のデータ・ファイルでは，データとシステムを相互に連携させた有機的なシステム利用が困難であるが，漢字文字処理による場合には，システム・パフォーマンスを体系的にまとめることが可能である。

6.2.7 システムの評価

(1) GIIPシステムのメリット

漢字を含む技術情報を，電子計算機で処理することが従来の方法と比較して，どの程度のシステム・パフォーマンスを持っているかを検討した結果，次のような点にメリットがあるといえよう。

① 従来手作業で行なっていた索引作成に要する時間の1／4～1／6で発行でき，処理速度が速い。

② 機械処理を行なえば，一度の入力でデータの分解，結合，転写，展開等データの変形加工が容易で，手動タイプ打ちより入力量も少なくてすむ。

③ データの有機的な転用が可能。

④ データの信頼性が高い。

⑤ 保管蓄積データのスペースが小さい。

DATA#	M.CL	S.CL			TITLE
1	102 E 21	114 A 512			磁気テープ保持装置
2	102 E 21	55 C 23 102 E 0	54 B 451	97(5) B 3	回転ヘッド制御方式
3	99 F 71				蓄積装置
4	102 E 21	97(5) B 3			ら旋案内テープさんえ装置
5	102 E 4	114 A 512	102 E 5	105 B 0	データ貯蔵装置
6	102 E 31	102 E 21			テープレコーダーのテープリフター装置
7	102 E 0	102 E 5	102 E 9	54 B 11	磁気録音装置
8	99 F 123	99 F 120.1			キネスコープおよびその制作方法
9	102 E 5				磁気ヘッド
10	100 D 0	104 G 0			レーザー装置
11	102 E 8	102 E 0	55 C 23	97(5) B 0	磁気媒体転写方式
12	12 A 41	12 A 48	12 A 5	62 E 51	高耐熱性を有する電気絶縁性被膜をけい素鋼板表面に生成しむる方法
13	102 E 21				回転ヘッド制御方式
14	97(5) L 01	99 F 123			カラーテレビジョン受信機
15	114 A 516	104 G 0			電気光学的情報変換装置
16	102 E 32	97(5) B 3	97(5) D 24		記録再生方式
17	102 E 51	62 B 7			立方結晶構造を有する焼結酸化物磁性材料からトランジューサを製造する方法
18	94(5) B 6	102 E 21			電話自動応対録音装置
19	100 D 0				高周波による多電極放電モードを利用したレーザ装置
20	102 E 21	55 C 23 102 E 0	97(5) B 3	97(5) B 3	回転ヘッド制御方式
21	102 E 21	55 C 2	97(5) B 3	102 E 0	磁気記録再生方式
22	99 F 71				直視型蓄積管
23	100 D 0	104 G 0			ジャイアンツパルスレーザ装置
24	102 E 5				磁気ヘッドの製造方法
25	98(5) E 2	98(5) B 1			自動周波数誘引発振器
26	102 E 8				磁気媒体転写方式
27	100 D 0	104 G 0	99 G 0		ガスレーザー管
28	97(5) A 22	114 A 511 97(5) B 1	97(5) A 0	104 G 0	光学的伝送装置
29	102 E 5				空隙高さを精密に規定した複合磁気ヘッドの製造方法
30	102 E 32	94(8) A 0 97(5) H 6	102 E 21 97(5) H 4	97(5) B 3	複数信号のジッター補正及時間差補正方式
31	102 E 5				電磁ヘッドの製造法
32	102 E 21	55 C 23 102 E 0	97(5) B 3	54 B 45	回転ヘッド制御方式
33	102 E 21	55 C 2	97(5) B 3	102 E 9	磁気記録再生装置
34	102 E 32	102 E 23 54 B 8	102 E 5	114 A 512	情報記憶装置
35	102 E 21				テープレコーダの反転機構
36	102 E 91	102 E 21			磁気テープカートリッジ
37	102 E 21	102 E 0	97(5) B 3		多ヘッドビデオテープレコーダのテープガイドエラー検出方式
38	97(5) D 24	97(5) B 3	102 E 32		磁気記録再生方式
39	102 E 21	114 A 512			磁気記憶装置
40	102 E 0	102 E 21	102 E 5		磁気記録再生装置
41	102 E 21	55 C 23 102 E 0	54 B 45	97(5) B 3	回転ヘッド制御方式
42	102 E 24				磁気シート録音再生装置
43	114 A 511	101 E 9			表示装置
44	99(5) K 2	99(5) K 10	99(5) F 32		固体映像表示装置

図 6・2・21 情報分析前順不同情報

DATA	M.CL	S.CL			TITLE
27	100 D 0	104 G 0	99 G 0		ガスレーザー管
14	97(5) L 01	99 F 123			カラーテレビジョン受信機
8	99 F 123	99 F 120.1			キネスコープおよびその制作方法
19	100 D 0				高周波による多電極放電モードを利用したレーザ装置
3	99 F 71				蓄積装置
12	12 A 41	12 A 48	12 A 5	62 B 51	高耐熱性を有する電気絶縁性被膜をけい素鋼板表面に生成しむる方法
22	99 F 71				直視型蓄積管
10	100 D 0	104 G 0			レーザー装置
15	114 A 516	104 G 0			電気光学的情報変換装置
25	98(5) E 2	98(5) B 1			自動周波数掃引発振器
11	102 E 8	102 E 0	55 C 23	94(5) B 0	磁気媒体転写方式
26	102 E 8				磁気媒体転写方式
44	99(5) K 2	99(5) K 10	99(5) F 32		固体映像表示装置
23	100 D 0	104 G 0			ジャイアンツパルスレーザ装置
39	102 E 21	114 A 512			磁気記憶装置
1	102 E 21	114 A 512			磁気テープ保持装置
5	102 E 4	114 A 512	102 E 5	105 B 0	データ貯蔵装置
28	97(3) A 22	114 A 511 97(5) B 1	97(3) A 0	104 G 0	光学的伝達装置
43	114 A 511	101 E 9			表示装置
9	102 E 5				磁気ヘッド
24	102 E 5				磁気ヘッドの製造方法
29	102 E 5				空隙高さを精密に規定した複合磁気ヘッドの製造方法
17	102 E 51	62 B 7			立方結晶構造を有する鉄磁酸化物強性材料からトランジューサを製造する方法
31	102 E 5				電磁ヘッドの製造法
34	102 E 32	102 E 23 54 B 8	102 E 5	114 A 512	情報記憶装置
4	102 E 21	97(5) B 3			ら旋案内テープさュえ装置
36	102 E 91	102 E 21			磁気テープカートリッジ
42	102 E 24				磁気シート録音再生装置
35	102 E 21				テープレコーダの反転機構
7	102 E 0	102 E 5	102 E 9	54 B 11	磁気録音装置
6	102 E 31	102 E 21			テープレコーダーのテープリフター装置
16	102 E 32	97(5) B 3	97(5) D 24		記録再生方式
40	102 E 0	102 E 21	102 E 5		磁気記録再生装置
38	97(5) D 24	97(5) B 3	102 E 32		磁気記録再生方式
21	102 E 21	55 C 2	97(5) B 3	102 E 0	磁気記録再生方式
37	102 E 21	102 E 0	97(5) B 3		多ヘッドビデオテープレコーダのテープガイドエラー検出方式
33	102 E 21	55 C 2	97(5) B 3	102 E 9	磁気記録再生装置
18	96(3) B 6	102 E 21			電話自動応対録音装置
30	102 E 32	96(8) A 0 97(5) H 6	102 E 21 97(5) H 4	97(5) [illegible] 3	複数信号のジッター[illegible]去及時間差補正方式
13	102 E 21				回転ヘッド制御方式
32	102 E 21	55 C 23 102 E 0	97(5) B 3	54 B 45	回転ヘッド制御方式
2	102 E 21	55 C 23 102 E 0	54 B 451	97(5) B 3	回転ヘッド制御方式
41	102 E 21	55 C 23 102 E 0	54 B 45	97(5) B 3	回転ヘッド制御方式
20	102 E 21	55 C 23 102 E 0	97(5) B 3	97(5) B 3	回転ヘッド制御方式

図 6-2-22 情報分析後の情報相関線型距離関係
表中にある情報に注目した場合、上下に離れるほど相関は低くなる。したがって最上端と最下端の情報がいちばん相関が小さい。

(2) システムの問題点

メリットと同時にデメリットとなる点もいくつか挙げることが出てきて，その対策を考慮する。

① 手作業に比較して，データ修正に時間，経費，手間がかかる。（時間の流れの中で小まわりがきかない。）

② データ・チェックを厳重にする必要がある。（小まわりのきかない点のフォローのため）

③ 経費が割高となる。

（データおよびシステムのアベイラビリティを高めることにより帰雪する。）

④ 出力装置のオフライン稼動によるタイム・ロスがあること。

6.2.8 将来のシステム展開

これまで述べたGIIPシステムは，既存のハードウェアの機能のアベイラビィリティの低さを，ソフトウェア的な面から工夫を加えて汎用化を試みてきたが，データ構造が新しいため，従来蓄積されているデータ，ないしは出力装置の使用ルートを持たないユーザに対するデータ利用の無用な点等を解決するために，従来のカナ表現データと漢字文字表現データとの相互利用のはかれるカナ／漢字翻訳システムを開発し，この一連のシステムの中に組み入れて，さらに利用度を高める方向に進めたい。

現在，カナから漢字，漢字からカナへの翻訳システムについて，種々研究されているが，まだ実用に耐えるシステムは開発されていないのが現状のようである。もしカナ・データが漢字に直せれば，漢字文字処理システムは一段とその普及度を早めることと考えられる。

6.3 出版社における自動編集システム

6.3.1 沿　　革

数万枚のカードに，1枚.1枚，項目を記入する，50音順にカードを配列する，昨日も，今日も，明日も，重苦しいほど単調な日々……。

しかし，百科事典や辞典の編集にとっては，さけることのできない大切な仕事だ。当時（昭和39年）この単調な手作業から編集者を解放し，内容の充実により力を向けることができるかというのが編集部門の大きな課題であった。この課題の解決方法として，我々はコンピュータを利用できないかと考えた。

一般に編集業務におけるコンピュータの利用は，以下のようなものが考えられる。

ア．編集製作における製作日程管理，統計管理への利用

イ．編集製作における情報の活用

（情報検索，データバンク）

ウ．編集製作技術への利用（自動編集）

当社では，コンピュータによる編集業務の合理化として上述したうちのイ．とウ．の業務を手はじめに手がけることにした。まず始めは，昭和40年に開発した「学研古語辞典」作成に関するパンチカード利用がその第一歩である。

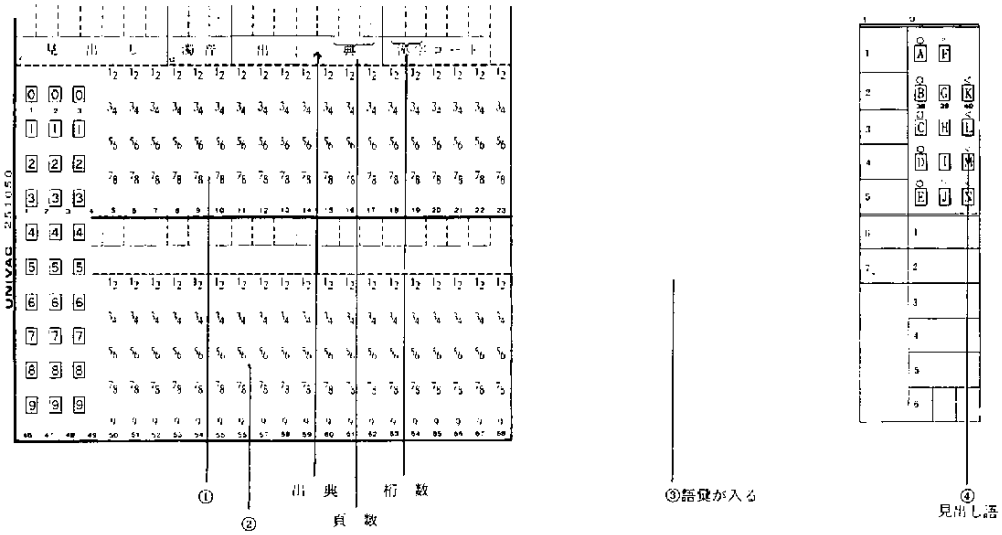


図6・3・1　古語辞典編集用パンチカード

これは，古語辞典の用例の収集，検索を目的としたもので，図6.3.1のように④の部分に書き込まれた見出し（カナ文字）を①の部分にパンチする。①には，その他，出典，頁数，桁数などがパンチされている。この①のパンチされた部分をOUK－1004で数字および英字に変換し，分類機にかけると50音順に配列される。③には用例を複写機でプリントしておく。このカードは，現在，45万枚に達

している。

その利点は，

① 形式が定まり，規則的かつ体系的に保管できる。

② コンピュータおよびPCM（パンチカードマシン）が使用できるので，分類，追加，削除，検索が，また，50音配列も容易である。

③ 保管，管理が比較的容易である。

などというメリットを生んだ。

しかしながら上記の方法は，間接法といわれ，文字情報を直接コンピュータのメモリーに記憶するのではなく，パンチカードに印字されているのにすぎない。

一方，この方法とは別に直接コンピュータに文字情報を記憶させる直接法という方式を考えた場合，漢字プリンタが必要となる。この直接法は，間接法に比べて

① I／Oのスピード

② データの保管

③ 検索スピード

などいずれも優れているし，またデータを磁気テープ，磁気ディスクに記憶させることによって将来の情報産業に直結することができる。

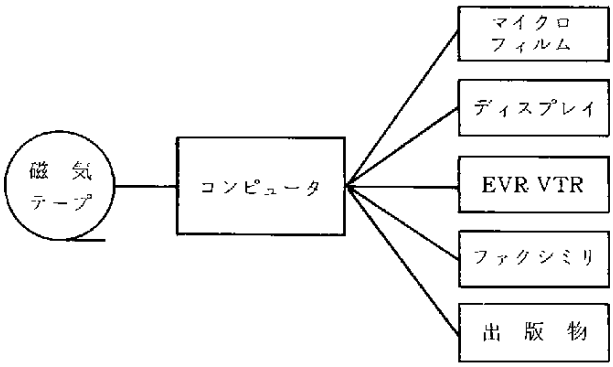


図6・3・2 出版事業におけるコンピュータ処理の概念

当社もカード利用の次の開発目標として，漢字プリンタ（電算植字システム）の利用にふみきった。

6.3.2 電算植字システムの開発

前述したようにパンチカードシステムの開発を一応終了し，次に漢字プリンタを中心とした電算植字システムの開発にふみ切った。

我々としては，最終の目的として，編集の合理化と編集データのデータバンク利用という面を考えていた。その場合，必要条件として漢字プリンタを開発する必要があった。

しかしながら，昭和41年当時は，我々の要求にこたえる漢字プリンタはなかった。

当時，我々の望む漢字プリンターの条件は，

ⓐ コンピュータのI／Oとして十分なプリント速度をもつこと。

ⓑ モニタ（校正用ゲラ）がとれること。

ⓒ 文字の縦横，大小，ピッチ付，行間の自動処理など文字を自由に扱えること。

ⓓ 印刷用のシャープな版下がとれること。

などである。

一方，発表されていた漢字プリンタの方式には

- 印字を目的にしたもの（単なるモニタプリント用）
  - ① 純機械方式
  - ② 光学方式
  - ③ ドット方式
- 印刷用の植字を目的にしたもの
  - ④ 光学－機械方式
  - ⑤ 全電子方式

などがある。

漢字プリンタの方式＼当社の条件	ⓐ	ⓑ	ⓒ	ⓓ
㋐	×	○	×	○
㋑	△	△	×	○
㋒	○	○	×	×
㋓	×	×	○	○
㋔	○	△	○	○

○ 条件をほぼ満たすもの
△ 条件をやや満たすもの
× 条件を満たさないもの

図6-3-3 漢字プリンタの評価

当社の要求する漢字プリンタの条件を，漢字プリンタの方式にあてはめた。その結果，図6.3.3に示されるように，全電子方式のものがもっとも条件に満足することがわかった。したがって，その焦点を電子方式にしぼることを決定した。

その結果，当時，電子方式として日本科学技術情報センタで稼動していたJEM3800に注目し，メーカと4年間の共同研究開発の結果，昭和45年8月ここに商業印刷用漢字システム，学研電算植字システムの導入にふみ切った。

6.3.3 学研電算植字システムの概略

(1) ハードウェア

当社における電算植字システムは，JEM3850漢字プリンタを中心にして，つぎのようなハードウェア群によって構成されている。

- 漢字鍵盤さん孔機　　15台
  （SCK－200D－4型）

- 手動写真植字機　　1台

  (MD－C型)

- HITAC－10汎用システム

  16KW，MT (3台)，DR (131KW)，PTR，CR

- IBM370－145汎用システム

  256KB，DS (9台)，MT (7台)，PR (2台)，CR (2台)，PTR，PTP

- JEM－3850電算植字システム

  8KW，MT，PTRモニタ・プリンタ，ハードプリンタ

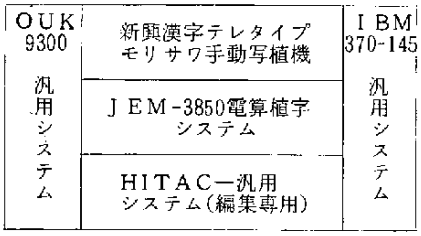


図 6・3・4　電算植字システムの構成

(2)　ソフトウェア

ハードウェアの開発とともに我々が苦心したのは，ソフトウェアの開発である。それは，編集という頭脳作業のパターン化から始まった。

以下，代表的なものを述べてみたい。

ア．　GEMS

GEMSとは，ゼネラル・エディトリアル・マネジメント・システムの略で，コンピュータを介した漢字情報処理システムを，効率よく運用するためのソフトウェア・システムである。

現在，実用化されているGEMSは，HITAC－10またはOUK－9300の磁気テープファイルシステムによって運用している。GEMSは，漢字情報をコンピュータをもって正確かつ迅速に処理し，漢字による情報のデータバンク化をはかり，その情報の提供サービスを行なうことを目的とする。

とくにGEMSとは，当社の編集部門と共同研究で開発したもので，あらゆる出版物，名簿などの作成・検索にすぐれた効果を発揮するように，総合的な角度から検討された汎用ソフトウェアである。

さらにJEM3850高速漢字処理システムとの結合によって，漢字を含む印刷物の版下や名簿・顧客台帳などの汎用プリントも可能である。GEMSのおもな機能は，次のとおりである。

①　入力処理（GEMS－Ⅰ）

漢字テレタイプからインプットされた紙テープのデータに対して，文字種変換，文字の大小，並びのたて，よこなどのファンクション処理および英数字・特殊文字などのピッチ付加処理などを

行なう。

② 棒組調整処理（GEMS－Ⅱ）

入力処理を得たデータは，ここにおいて1行単位をもって字詰，禁則，ルビ，ジャスティフィケーション，行頭，行末，中央揃え，分離禁止，数式合成，などの処理がなされる。

③ 組版処理（GEMS－Ⅱ）

ここでの処理は，文章全体の編集処理であり，図版，写真，見出し，ページ割，ノンブル，柱，処理などが行なわれる。

④ 校正ゲラ・版下作成処理（GEMS－Ⅲ）

校正用，版下プリントを中心にした，各種モニタおよび台帳などのプリント処理を行なう。

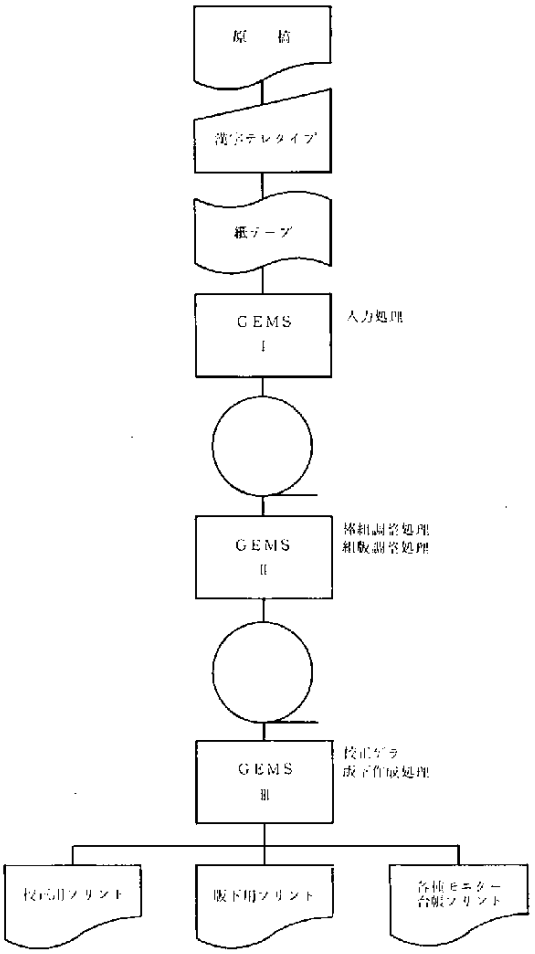


図6-3-5 ソフトウエアGEMSの機能

イ． 項目検索「50音項目整理」システム

事典・辞典を中心にした漢字カナ混り文項目を，50音シーケンス，分野，図版情報，グレード情報，人・地名情報などの条件により検索するシステムである。

コンピュータにより，カナ文字（20桁），漢字（5桁）の漢字カナ混り文を濁音，半濁音，促音，拗音，清音，などを加味し，7桁の完全な50音変換数としてアウトプットするものである。システムの概要は

① 訂正処理：編集部門からの原稿を漢字テレタイプでパンチし，漢字プリンタによってモニタを取り，校正を行なって完全な形で磁気テープに保管する。

② 外字処理：原稿にある文字で漢字テレタイプ上にない文字種の処理。

③ 50音配列処理：項目（カナ・漢字・キーコード）を数字の形に変換し，コンピュータ・サイドで50音分類できる形にする。

④ 項目・分野処理：総項目の50音分類，分野，担当者ごとの50音分類をアウトプットする。

このシステムによって整理されている項目は，現在約50万語ある。

ウ．そ の 他

漢字かな混り文を中心にしたSORT，LISTER，各種CODE，CONVERSION，IR汎用パッケージ，カナ文字→漢字かな混り文変換システムなどがある。

6.3.4 漢字入力装置

当社が電算植字システムの研究を始めたのは41年であるが，この当時の代表的な漢字入力装置は，両手でインプット操作する谷村新興方式と両手，両足でインプット操作する方式がよく使用されていた。この両機を検討した結果，パンチスピード，性能，どれをとっても前者の方が勝っているという結論に達した。

インプット文字種に関しては，社内外の文献を調査し，出版向きに検討した結果，合計4,992字の文字を選定した。

その結果，漢字テレタイプ上のけん盤配列も，当社独自のものになっている。ところで日本語の文字種は，4,992字（自由に交換できる文字192を含む）ではとうていたりない。

字　種	インプット	アウトプット
ひらがな	164	254(明朝、アンチ、細ゴジ)
カタカナ	84	188(明朝、アンチ)
欧文文字	50	180(ローマン、イタリック、ボールド)
アラビア数字	40	60(　〃　)
記号類	101	229
スペース	4	4
漢字	3885	3885
外字	192	192(自由に定められる)
計	4520	4992

表 6・3・1　入出力文字種の内訳

しかし，一般の出版物に対しては，4,992字の文字種をもっていれば，約99％以上カバーできる。そのため，ない文字に対しては，手動の写真植字機で切りぬけることにした。SCK漢字テレタイプは，すでに約5年の使用経験がある。一方，MD－C型手動写真植字機は2年目である。最近，漢字イ

ンプット機器も

- コンピュータと直結してディスプレイインプットする方式
- 手動写植機を基本とするものでモニタもとれる方式
- けん盤キーをペン方式のもので指示する方式
- システムタイプを変形した方式

など，各種のものが開発されている。しかし

- パンチスピード
- インプット文字種の大小
- コストの問題
- マシンの機構上の問題
- パンチャー育成の問題
- 漢テレコード不統一の問題

などを100％満足する機種は，まだまだであると思われる。いずれにしても，上述の問題点を満足させるようなインプットマシンが早く開発されることを望みたい。

6.3.5 漢字出力装置

当社導入の漢字出力装置・JEM3850導入の経過については6.3.3で述べたとおりである。このJEM3850の使用経験も3年目を迎えている。ここでJEM3850の機能について述べておきたい。

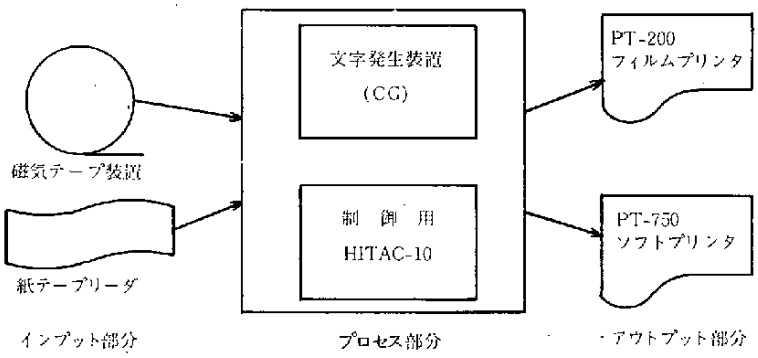


図6・3・6 漢字出力のプロセス

〔インプット〕

- 磁気テープ装置
- 紙テープリーダ

〔プロセス〕

- 文字種 4,992字 （交換可能文字種 192字）
- 文字サイズ 6級～32級
- 文字の幅 8／8～1／8cm
- 送り量 空間・行間とも0.5mm

2002-1-620　あまり深くない砂泥底にすむ海水魚。夜行
2002-1-630　性であるらしく、昼間は海底に穴をほって
2002-1-640　その中にとぐろをまいていることが多い。
2002-2-010　底引網にしばしばはいるが、食用価値はほ
9002-2-020　とんどない。全長：40cm内外。分布：南日
2002-2-030　本沿岸から中国の広東付近。（硬骨魚綱　ス
2002-2-040　ズキ目　アカタチ科　*Acanthocepolu krusen*
2002-2-050　*sterni*）
2002-2-051390025000※　①　②
2002-2-060　**アカデメキン**……22
2002-2-070　デメキンの3品種の1つ。ふつうのうろこを
2002-2-080　もち、体色の赤い品種である。→デメキン
2002-2-081390027000※　①　②
2002-2-090　**アカニシ**……112、145
9002-2-100　内湾の深さ約20mの砂地や小石のある海底
2002-2-110　にすむ大型のごつい巻貝。大きくなると殻
2002-2-120　の口の中がまっ赤になる。食肉性で、ほか
2002-2-130　の貝をおそうので、養殖上の害貝のひとつ
2002-2-140　にされている。卵のうはその形からナギナ
2002-2-150　タホオズキとよばれ、小石の上や貝殻の上
2002-2-160　などにうみつけられる。肉は食用となり、
2002-2-170　やわらかくておいしい。殻高：15cm。分布
9002-2-180　：日本各地。（軟体動物　腹足綱　新腹足目

（初　校）

3A246-1-410　全変態の昆虫では成虫化後の短時間内に
3A246-1-420　崩壊してしまう。
3A246-2-010■
3A246-2-011390018445230※理化金井一裏遠隆一　（花岡謹一郎）　※00Z0
3A246-2-020　α　**ぜんさようホルモン**【前胸腺ホ
3A246-2-030ルモン】昆虫の前胸腺から分泌されるホ
3A246-2-040ルモン。脱皮ホルモンともいい、幼虫お
3A246-2-050　よびさなぎの変態を誘導する。略称
3A246-2-060　ＰＧＨ(prothoracic gland hormone)。▽遠
3A246-2-070　では前胸腺は頭部第一環節の左右両側の
3A246-2-080　気門の直下にある帯状の器官である。熱
3A246-2-090　源を結紮して二分すると、前胸腺を含ま
3A246-2-100　ない後半は正常に発育し、さなぎを用いた
3A246-2-110　場合も同様にして後半は成虫化を起こさ
3A246-2-120　ない。幼虫ではまた、ＰＧＨがアラタ体
3A246-2-130ホルモンと共存して脱皮を誘導する。後
3A246-2-140　者は幼虫形態の存続を決定し、脱皮白体
3A246-2-150　はＰＧＨによって誘導される。ＰＧＨの
3A246-2-160　分泌は、脳の神経細胞から分泌される脳
3A246-2-170ホルモン(ＢＨ)によって支配されてい
3A246-2-180　る。◇植物　からもこれと同様の働きを
3A246-2-190　もつ類似化合物が発見された。
3A246-2-200■
3A246-2-201390018445700※有馬遠隆一　（有馬久雄・花岡謹一郎）　※00Z0

（再　校）

じんめい【人名】⇨姓名

じんめいかんじょう【人名勘定】企業会計における人的勘定の一種。商品の売買にともなう取引先ごとの債権・債務の記録は、その取引先の名称を付した勘定をもって行なう。そこでこのような勘定を人名勘定と呼んでいる。▽取引先が多数になると、人名勘定を総勘定元帳に設けることは煩雑になる。そこで人名勘定は仕入先元帳や得意先元帳に設け、総勘定元帳にはこれらの統制勘定として買掛金勘定や売掛金勘定を設けるようになっている。→人的勘定　（小林健吾）

しんめいき【申命記】『旧約聖書』の「五書（トーラー）」中第五番目の書。▽モーセがカナンを眼前にしたヨルダン川の東で行なった訣別説教の形をとり、三部三四章からなっているが、前七世紀の地方の憂国の士がモーセの名において建国の理念を強調した説教である。最終的に編集されたのは前五五〇年ごろとされている。▽一一～一八章の律法の部分は本書の中心をなし、ことに一二章の礼拝集中と偶像禁止に関する部分は、前六二一年のユダ王ヨシヤの宗教改革運動の支柱となった。また、神への愛を強調した六章四～九節、一一章一三～一一節は新約時代のユダヤ人の日常生活に決定的な影響を与えた。◇『新約聖書』には「創世記」「イザヤ書」「詩篇」とともに最も多く引用されている。　（高橋正男）

しんめいしゃ【神明社】東京芝の神明社のように、天照大神の神霊を祭った神

＋　＋
〈381―3〉　①書驛②聾韓③楚榎④歳膿

（校　了）

図 6-3-7　グランド百科辞典用アウトプットの例

- 組体裁　縦・横混植可
- 変形プリント　長平体および上つき下つき
- プリント速度　150字／秒
- 文字書体　モリサワ

〔アウトプット〕

- PT200フィルムプリンタ

  ラインプリント方式，プリントサイズ（30パイカ），100mm×30m印画紙
- PT750ソフトプリンタ

  ラインプリント方式，B4×100m印画紙，200～400mmカット可

である。

漢字入力装置の項で前述したように，アウトプット文字は4,992字あり，うち192字が自由に文字を交換できる。漢字プリンタにない文字は，最終アウトプットの段階でアウトプット用紙にゲタで表示し，手動写真植字機でタイプした文字を貼り込み完成させる。そのため手動写真植字機は，漢字プリンタにない文字を約10,000字もっている。つまり，何百字に1回しか出てこない文字をコンピュータに所有させたり，インプットで苦労するよりも，手動写真植字機で貼り込みしたほうが，コスト的にも安いためである。

現在我々がかかえているアウトプット問題点は，

- 文字品質の向上
- プリントスピードのアップ
- 文字変換手段の簡便化

などである。

6.3.6　電算植字システムの利用状況

当社において，編集部門へのコンピュータ利用は

① 編集の合理化

② 編集情報の収集・検索

の二大目的であることは前に述べた。前者においては，コンピュータシステムを利用することによって編集作業のスピード化がはたされ，可能なかぎりの業務をコンピュータに代行させることによって，編集者に時間の余裕を与える。そこで得られた時間の余裕は，編集者の本来業務である創造的な企画業務などに還元して，製品の内容の充実をはかることができるなどが最大のメリットであろう。

一方，後者では，データをコンピュータの媒体である磁気テープに収集することにより，常に最新のデータを蓄えることが可能になり，日々に蓄積される情報に対する検索，分類，追加，削除なども自由自在に実施できる。さらに大量の情報の蓄積が果されれば，将来は教育データバンクとしていろいろな面での応用のみちが開かれることになる。

当社の出版物には，雑誌と書籍の二本柱がある。これまでのような諸条件でのメリットを得るために，現在このシステムで処理する業務としては，基本的な情報の蓄積が可能であり，将来のデータバ

ンクとして活用することが最適である百科事典，辞典，などを中心としている。学研電算植字システム利用による最初のアプリケーションは，昭和45年当時，当社，創立25周年記念出版として出版される百科事典「グランド現代」をのせることであった。

この「グランド現代」百科事典は，全21巻，そこに収録される総項目数は，実に25万項目にも達するぼう大な情報量をもるものである。その刊行に要する期間は45年10月に第一巻を発行してから，全巻完成までには2年半かかる。また，最終25万項目を選定するためには，実に40～50万項目の基礎情報を必要とする。これを従来の手作業で分野別処理，専門家および編集者による検討，50音配列，相関処理，といった一連の編集作業を行なうとなると，それは気が遠くなるような作業量である。

そこでまず，このぼう大な情報量をもつ「グランド現代」百科事典の作成目的をしぼった。その第一歩として，昭和43年5月から50音項目整理がコンピュータ処理によって開始された。現在，電算植字編集システムで処理されている業務は，「グランド現代」百科事典，「学習ワイド図鑑」，「ホーム百科事典」，「平家物語索引集」などを中心に，進行中の辞典，名簿，顧客台帳，社内各種情報など約25点近くの業務を処理している。これに必要なインプットは，1カ月約400万～450万であり，JEM3850，HITAC－10，他のコンピュータの利用状況は，おのおの約300～350時間におよんでいる。

次にJEM3850を利用した「グランド現代」百科事典の業務概略を述べる。

(1) データ作成処理

1編集部門よりインプットされた原稿は，漢字テレタイプによって紙テープにせん孔される。せん孔された紙テープは，校正しやすいようにHITAC－10を使用して，シーケンス付，ピッチ付，ルビ処理，行末，行頭，などの処理を行なう。紙テープより磁気テープ（初稿用）に変換されたデータは，JEM3850により初稿モニタとしてアウトプットされる。この校正用モニタは，各項目ごとに50音順キー，分野，担当者，図版，グレードなどのキー（編集部門からインプットされる原稿に記入されてくる）および各行にコンピュータによって付加された番号がついている。

校正用モニタは，編集部門に返され，赤字校正が行なわれる。訂正は，校正用モニタについている行の頭の数字（キー番号）ごとに行なわれる。

キー番号によって起こされた訂正原稿は，紙テープにさん孔され，完全なデータになるまでチェックされる。完全なテープとして完成された訂正テープは，磁気テープ（訂正テープ）に書き込まれ，初稿テープと組み合わされコンピュータにかけられる。コンピュータはキー番号で文中1行分の区分を行なう。つまり，初稿，訂正，両磁気テープのキー番号が一致したとき初稿テープの内容を訂正テープの内容と入れ替える。これが訂正処理である。以上の作業が従来の活版での赤字にしたがっての文選と直しにあたる作業である。

(2) 自動編集処理

訂正処理の終了したデータは，校了済磁気テープとして，全データに対する自動編集処理が行なわれる。つまり組版，棒組，図版処理などである。（前述した「GEMS」を参照）

(3) 版下作成処理

自動編集処理の終ったデータは，各ページ，段ごとに版下としてアウトプットされる。JEM3850

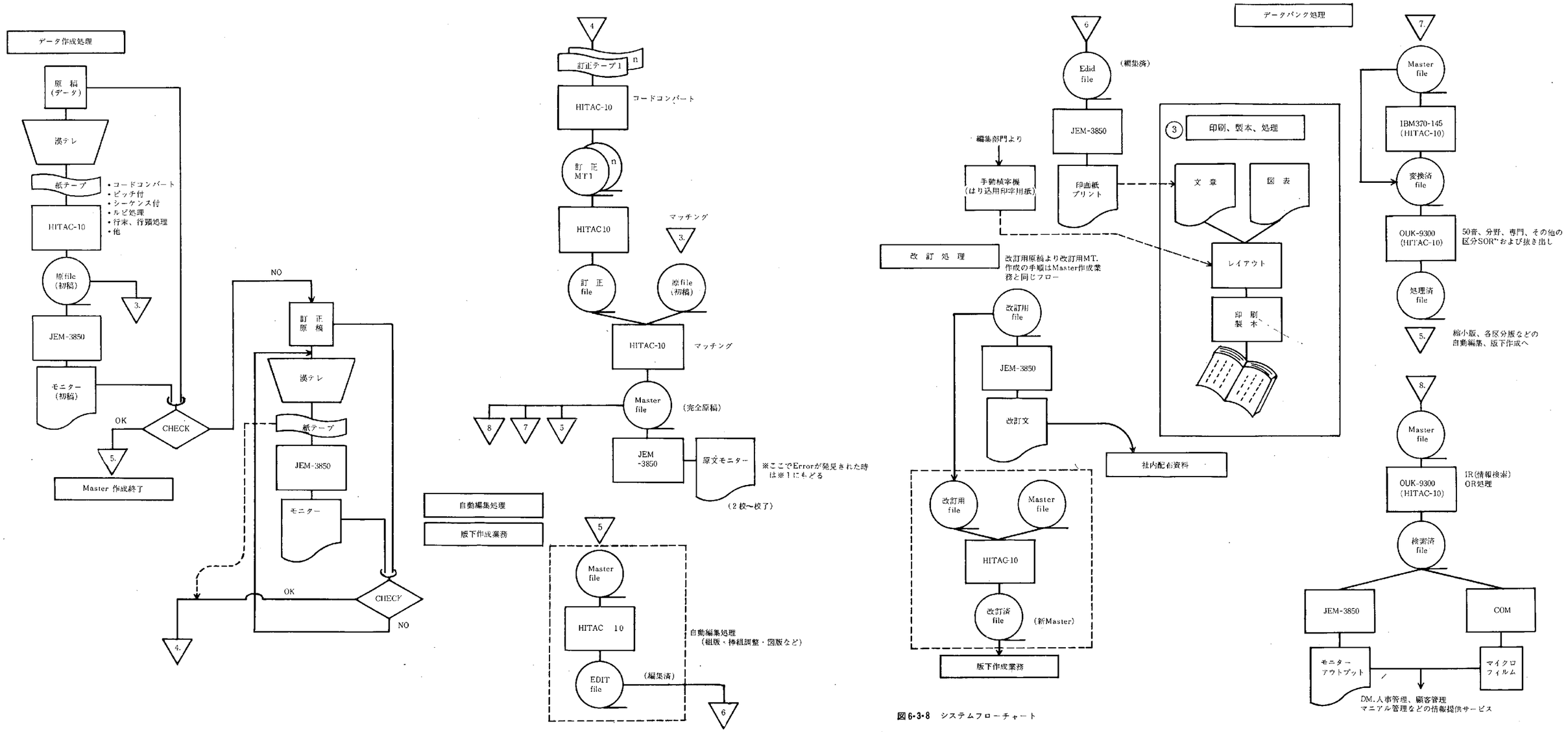


図6-3-8　システムフローチャート

にない文字は，ゲタで処理をしておき，その部分の文字は手動写真植字機で別途作成する。

(4) データバンク処理

編集出版の終了した磁気テープは必要に応じて，分類，検索される。たとえば，百科事典であれば，50音，分野，専門に分類され，縮小版，各区分版などの版下に利用される。

6.3.7 システムの評価

学研電算植字システムによって得られるメリットを，「グランド現代」百科事典について述べてみると以下のようになろう。

① 編集データを豊富に蓄積し，必要に応じて迅速に検索できるシステムが完成する。

② 入稿してから校正し，レイアウトし，製版するまでの工程が一貫作業となるため，製作日程がたやすくたてられるようになり，作業上のムリ，ムダ，ムラを大幅に短縮することが可能になる。

③ 校正に際して，文字の加筆，修正，削除，処理が自由，迅速，正確である。したがつて，印刷用版下の作成が容易である。

④ 辞典，事典の改定版，年度版の作成がきわめて容易で低コストにできる。また，対象別，分野別，地域別などあらゆる専門事典の作成も容易である。

ただ一方において，訂正処理に従来よりも人手がかかっている。この解決が今後の課題であろうと思われる。

6.3.8 今後の方向

当社の電算植字システムの現状をいろいろ述べてきたが，我々にはさらに研究開発を続けなければならない問題がいくつかある。

① 訂正処理の改良

② 文字品質の向上

③ 各種アプリケーションソフトの開発

などである。

マシン導入3年弱であるが，改良に改良をかさね，一応の成果はあがっている。現段階では十分にシステムを使いこなしている自信はある。しかし，ハード，ソフトの技術はそれが終りになることはない。コンピュータの欠点，全電子方式漢字プリンタの欠点，ソフトの欠点をいかにカバーし，コスト的にも，システム的にも最上の成果を得るかは，今後さらに研究を続ける必要があろう。

そのためには，デイスプレイや他の周辺機器の活用が今後の大きな課題になろう。

最後に将来の方向として，このシステムの最終目的は，印刷版下の作成ではなく，あくまでも編集の合理化であり，脱活字のための人手不足の対処という副産物であって，データバンクの完成こそ，我々のめざす大きな目的である。将来は，デイスプレイおよびファクシミリ，マイクロフイルムによって，辞典又事典のあり方も大きく変わってくるであろう。

日本に漢字のあるかぎり，漢字による情報伝達を必要とする分野は，無限に存在すると思われる。今後ともコンピュータによる漢字処理の有効な利用のために，電算植字を含めた副広い漢字入出力装置をおおいに利用したい。

6.4 日本語によるオンライン情報検索システム

ここに紹介する内閣調査室の計算機システムは，昭和48年1月より稼動のもので，十分運用の経験を経ているものではない。このシステムは，基本的には，日本語によるバッチ処理の情報検索システムとして設置することを承認されているものである。

しかし日本語処理の問題を効果的に解決するため，関係メーカとの間にオンライン情報検索の共同研究開発を行ない，その第一次成果として，CRO－V0を作成した。このシステムによって，実務の運用を行ないつつ，さらにCRO－V1の研究開発を行なっている。

6.4.1 沿　　革

(1) 資料処理の問題点

資料処理の最も単純なサイクルを考えると，初段階は，あらかじめ分類要求者によって定められた基準に従って分類整理することであり，次には，逆にこの分類を組み合わせて，指定した分類項目に属する資料を再収集し，関係資料群を得ることであり，最終的には，この群から目的に応じて，所要の形態に資料を再構成し情報として提供することである。このサイクルについて，第2段階の再収集は，機械処理に適しているが，なお初段階は，事前処理すなわち，機械入力のための記入作業として残り，全作業力の軽減にならないのである。また最終段階は，資料内容の主観的編集処理の機械化が困難であり，事後処理として手作業に依存することになる。

(2) 機械適用処理の拡張

この様な事前，事後の手作業を軽減する方法として，資料処理者が，常時機械を利用できる状態にあることが必要である。すなわち手作業時間帯に合致した機械処理が可能で，その作業を軽減した形で，容易に機械操作に転換できる機能が必要である。さらにこの機械処理は，「考えながら操作」するために機械占有が長時間にわたり，また，逐次条件の変更を指示し，機械と応答できるように機能的に配慮されていることが望ましい。

オンライン会話処理のシステムは，以上の諸点の設計に適合するものとして，採用されたものである。なお蓄積データの検索効率を高めるためには，データに付与された索引語の修正が必要である。蓄積データの情報価値は，索引語によって抽出されるかどうかにかかっている。情勢の変化に応じ適切な索引語の追加修正を行なうことは，関係ファイルのメンテナンスを意味している。しかも資料処理者の立場から実施できるためには，オンライン・ファイルメンテナンスの機能を必要とする。このような問題は，CRO－V1の課題として残されている。

6.4.2 システム設計

(1) 業 務 の 流 れ

処理業務を大別すると，入力フロー（資料収集－機械入力データ整備－機械入力）と，検索フロー（質問要求－機械入力質問語作成－機械処理－回答確認と試行），および統計フロー（統計作業要求－データ整備－処理）に分類される。

入力フローは，機械室内の共用入出力端末機で，せん孔した紙テープデータを，センターバッチで

仮編集ファイルに編集するか，または資料処理者の専用端末機を使用して，直接データを伝送し，仮伝送ファイルに入力するかを，選択できる。またデータ伝送の段階では，指定した入出力端末に，単純な転送も考慮している。なおマイクロフィルムへの撮影も並行して行なう。

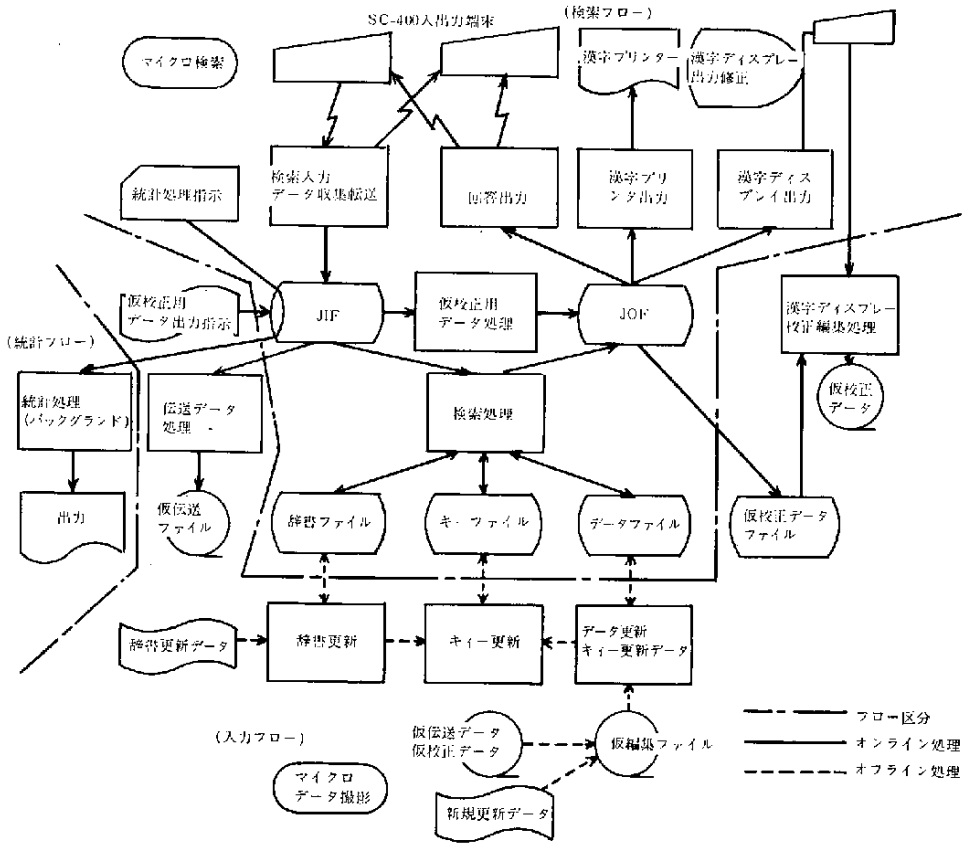


図6・4・1　処理業務のブロックチャート

仮伝送データを含む，各種の更新データを総合し，仮編集データを作成，これより，データ・ファイル，キー・ファイル（索引語とデータ番号の対応），辞書ファイル（索引語と分類上の代表語の対応）に対して，新規入力に伴うメンテナンス，あるいは，更新入力，辞書分類の更新に伴うメンテナンスを実施する。

検索フローは，原則的に入出力端末機から随時実施できるが，質問者コードによって，回答を制限している。また回答出力は，質問開始時に出力機器を指定し，任意の入出力端末機，センターの高速漢字プリンタ，あるいは，漢字ディスプレイ等により出力する。この段階で出力データに記入されているマイクロフイルムロケーション番号によって，マイクロリーダから映像サービスおよびハードコピーの提供を受けることができる。

回答の編集は，質問の段階で指定し得る4種の形式に限定されているが，最終要求者の需要に合致

するよう，いわゆるQA処理に結合して，編集処理を開発し，選択指定できるよう研究している。この課題は，ファイルメンテナンスと同様に，データ形式，データ・ベースの問題に還元される。

統計フローは，バックグランド・ジョブとして，行なっているが，この限りでは，漢字情報処理と数値情報処理の問題は発生しない。これを，CRO－V0のモニタ・システムであるCRJE（Conversational Remote Job Entoryの略）の本来の機能の下に，端末よりの伝送，処理，回答を期待すると，基本的には，単純なこのフローが，漢字モードと英数モードの変換，データ・ベースの設計等の複雑な問題となり，CRO－V1に残された研究となっている。

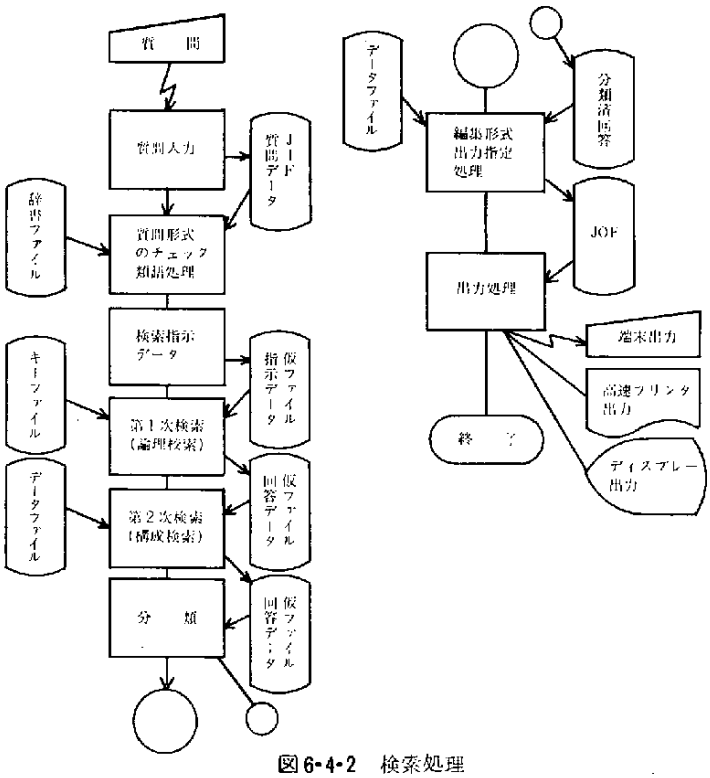


図6-4-2　検索処理

(2)　ファイルの形式

ア．情報検索用データ

情報検索用のデータの形式は，標準形式として，標題部，識別部，構成部，索引部の各セグメントから構成される。標題部は，資料提供者に関する事項とデータの概括的な性格を，28個の漢字（または数字）で表現した資料番号よりなる。識別部は，文章形式でのデータ内容であるが，これを20～79までの約50個の識別セグメント符号で分類し，内容の基本的な区分を行なったもので出力の単位となるものである。構成部は，データ主題の附属事項で，出典名称，出典時点，マイクロ番号等である。本来この部分は，識別部の各識別セグメントごとの内容に対応するものであるが，現状ではデータの全内容に対応している。検索部は，情報項目，索引語等の名称で分類される検索用キー・ワード群で

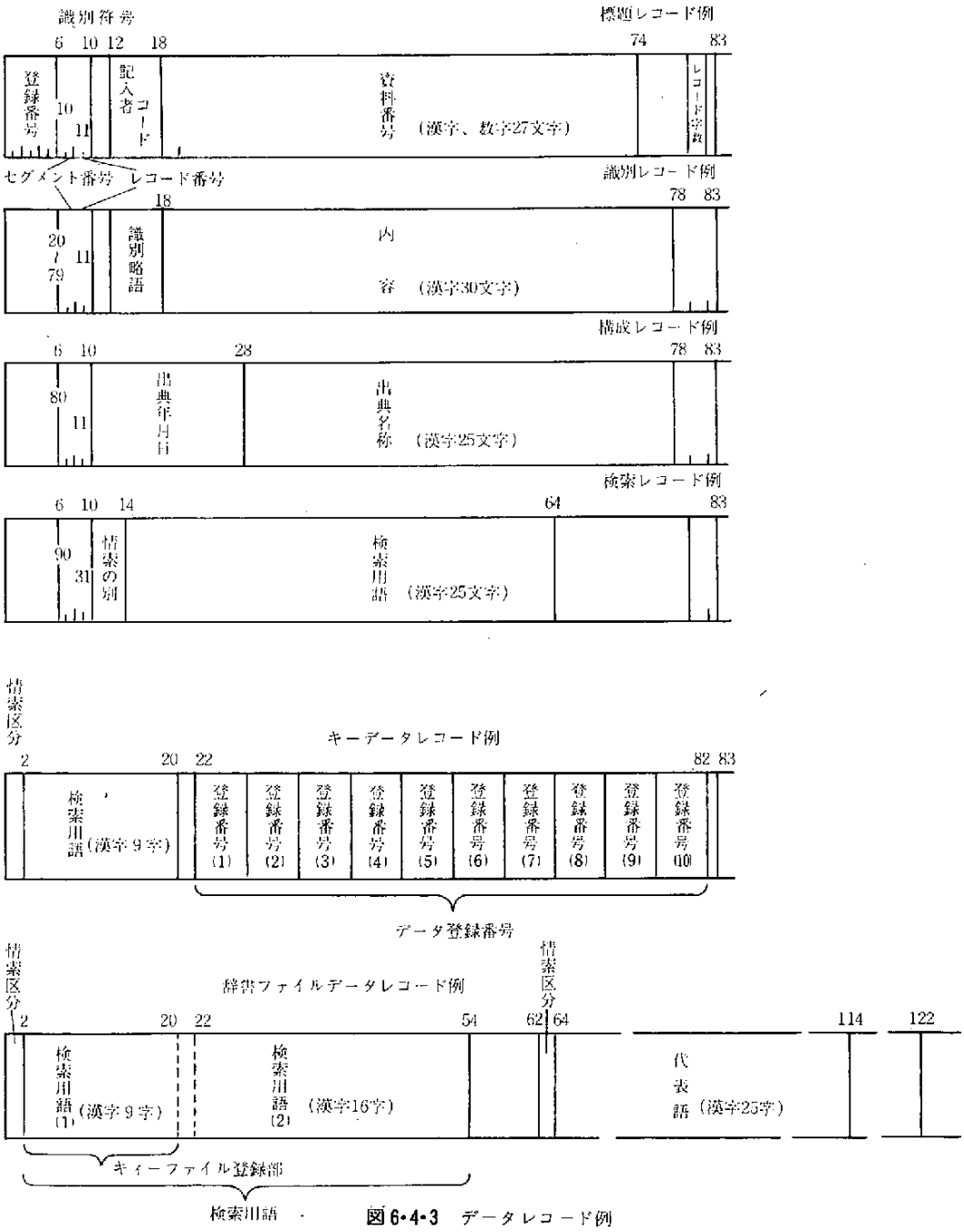


図6・4・3 データレコード例

ある。

識別部の1セグメントに許される最大レコード数は35であり，同一セグメント符号は，1データ内で35回まで重複記入が可能である，1レコードはコード部分を除き30漢字より構成される。検索部の検索コードは，少なくも1個の選択記入を必要とし，最大25個を限度とする。標題部，構成部の各セグメントの構成は単一である。またデータ最大文字は4,000漢字まで可能である。

イ．検索コード

検索コードは，情報項目と，索引語よりなる。情報項目は，国際十進分類表（UDC）等を参考として作成されたもので，主分類表と補助分類表（国籍，地域コード）があり，資料の客観的な分類基準を示すものである。しかしながら，資料処理者は，なお主観的な基準で，動態的な資料の細分類を必要としており，この為，細分類を随時索引語として使用し，段階的に情報項目体系との体系的関連を辞書管理に定義して組み入れていく。これらの階層分類の体系に従って，質問時の情報項目とデータに付与された情報項目との関連を随時拡張指定することを考慮している。また資料の記述表現を再現する形で選定するのが，索引語であって，当初自然語処理の必要性から，類語の処理を行ない，代表語を指定している。これら代表語の体系的な分類のために一部は情報項目との関連，一部は新たな体系基準いわゆるシソーラスが必要となる。

辞書ファイル，データ・レコードは，25字の検索用語の前半9字を第1次，後半を第2次の索引対応とし，これに代表語を付属する。当初代表語には階層的機能は与えられていない。

キー・データ・レコードは，検索用語の前半9字のコードに対応して，データ登録番号が10個まで与えられる。同一コードのレコードは，約1,000レコードまで許容されている。

(3) 設　計　目　標

以上のような設計思想から，CRO－V0の設計目標として，まず，オンライン・ジョブとして実施したものは，

①端末より検索の処理要求伝送

②QA処理（検索処理）回答の一部編集

③各種漢字出力機器への出力

④QA処理回答のディスプレイ映像を漢字鍵盤より修正編集改行等を初期に完了し，続いて，

⑤端末機器より直接データ伝送，

⑥受信データの転送，

⑦各種データのディスプレイによる修正編集と再入力

等を行なう。なおオフライン・ジョブとして，各種ファイル・メンテナンスの処理，統計処理を実施し，これらのオンライン化は，CRO－V1の課題としている。

6.4.3　システム概要

(1) 機器の構成と機能

機器の運用は，大別して，4群で行なっている。A群は，センター内の処理であり，NEAC－2200－500の基本構成からなり，中央処理装置，付加記憶装置を合わせ，CRJEシステム　約40万ステッ

プ，ジョブプログラム　約15万ステップの処理を可能とする容量を持つ。N3200は，高速漢字プリンタC5210および漢字ディスプレイの制御を行なっている。集団ディスク装置N274Bは，8スピンドルで，約2.8億キャラクタの容量を持つが，このうちシステム制御ファイル　2，用語管理ファイル　1，障害予備　1を充当し，データ・ファイルは4，すなわち，7千万漢字の容量となる。磁気テープN204D－2は，各種ファイル・メンテナンス伝送用仮編集に7台を設備する。コンソール・タイプライタおよびラインプリンタは主としてジョブ・システム運用管理用に用いる。

B群は，谷村新興SC－4000入出力端末装置を設置し，共用端末として，データ作成，質問検索処理を行なう。この時マイクロ撮影処理も行なう。C群は，緊急検索，各種データの編集のために漢字ディスプレイを中心とした処理を行ない，またD群は，SC4000を専用端末として設置するグループである。（6.4.4図参照）

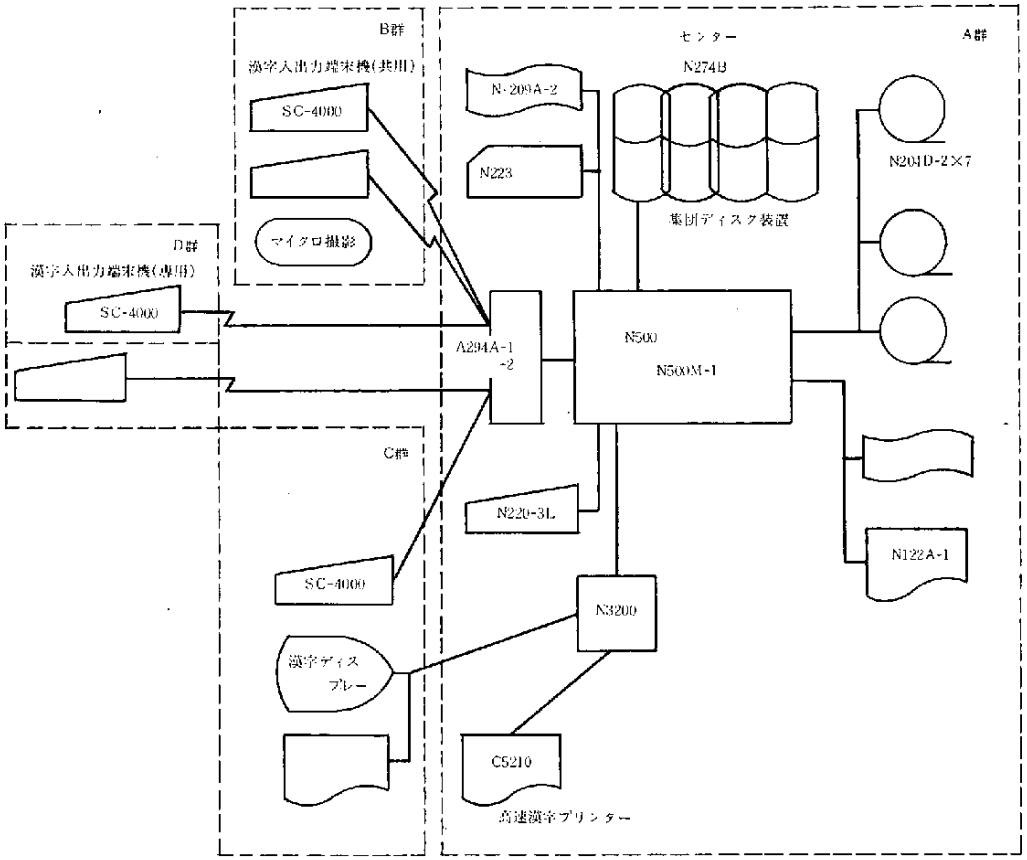


図 6-4-4　機器構成

(2)　ソフトウエアの機能

CRO－V0システムは，CRJEシステムを適用している。CRJEシステムは，RJE機能，狭義のCRJE機能，実時間処理機能，利用者管理機能等に大別されている。このうち，狭義のCRJE機能は，CRJE　データ・ベース管理機能を基本として作成されたファイル処理機能をオンラインで操作し，各種

のファイル管理，登録，修正，削除等が行なわれる。しかしながら，CRO－V0では，検索処理の設計にBIRD[注1]システムの思想を基本として改造したものであるため，システムの対象となるファイル構成は，CRJE　データ・ベースを適用せず，BIRD方式でのファイル設計が行なわれている。

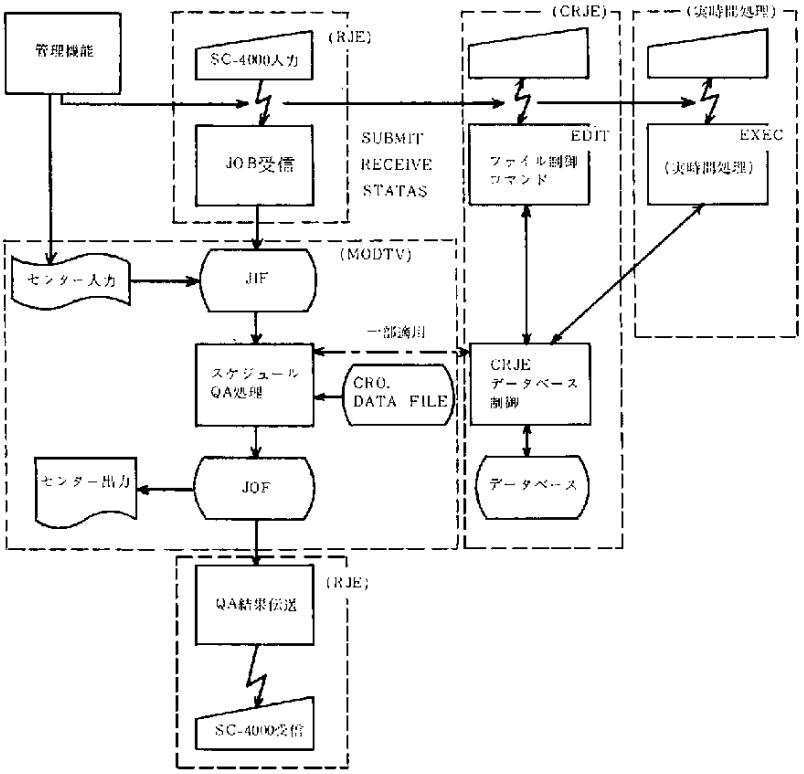


図6-4-5　CRO－VOのCRJE適用機能

CRO－V0システムに適用されるのは，基本的には，RJE機能と，MODⅣベース機能が中心となっている。今後の開発の目標として，出力編集処理，ファイルメンテナンス処理，さらに，数値情報の実時間処理等のオンライン処理のために，狭義のCRJE機能の適用がシステムの発展を確保するために必要であり，標準形データ形式の再検討とともに，CRJE　データ・ベース　への総合処理を考慮している。（図6.4.5参照）

（注1）BIRDはBibliographic Information System Using Random Access Deviceの略である。

入出力端末装置からの伝送メッセージは，まず通信制御機能の制御を受ける。DSR（端末，回線制御）の段階で，12ビット漢字外部コード系より6ビット内部コード系に変換される。ラインMCPによって，COMMIO FILE上のI／O処理待ちキューRTPQ（Real Time Processing Queueの略）に登録させる。漢字コマンド解析プログラムの処理（コマンド　アナライザ）により，コマンド，パラメータの解析を行ない，結果は，MCPに再度与えられる。ロールI／Oコントローラは，会話待時

間，伝送データ量処理等の管理のため，ロールI／O処理を通して各種コマンドの並行処理を制御する。かくて，検索要求コマンド・プロセッサ等に制御が移ったとき，各種コマンドの処理が実施される。

検索要求コマンドプロセッサはCOMMIO FILE上のメッセージからJIFを作成する。スケジューラによって処理可能なジョブをJIFから選択する。トランジショナルモニタはジョブ・リソースを収集して検索ジョブ・プログラムのロード要求をする。そして検索ジョブ・プログラムがロードされ，検索実行に移る。

検索実行プログラムは結果をJOFに出力する。回答要求コマンドを受けて，回答出力がJOFから回答要求コマンドプロセッサによってCOMMIO FILEのCOQに一時登録される。このあとDSRを経て検索出力要求端末装置に出力される。

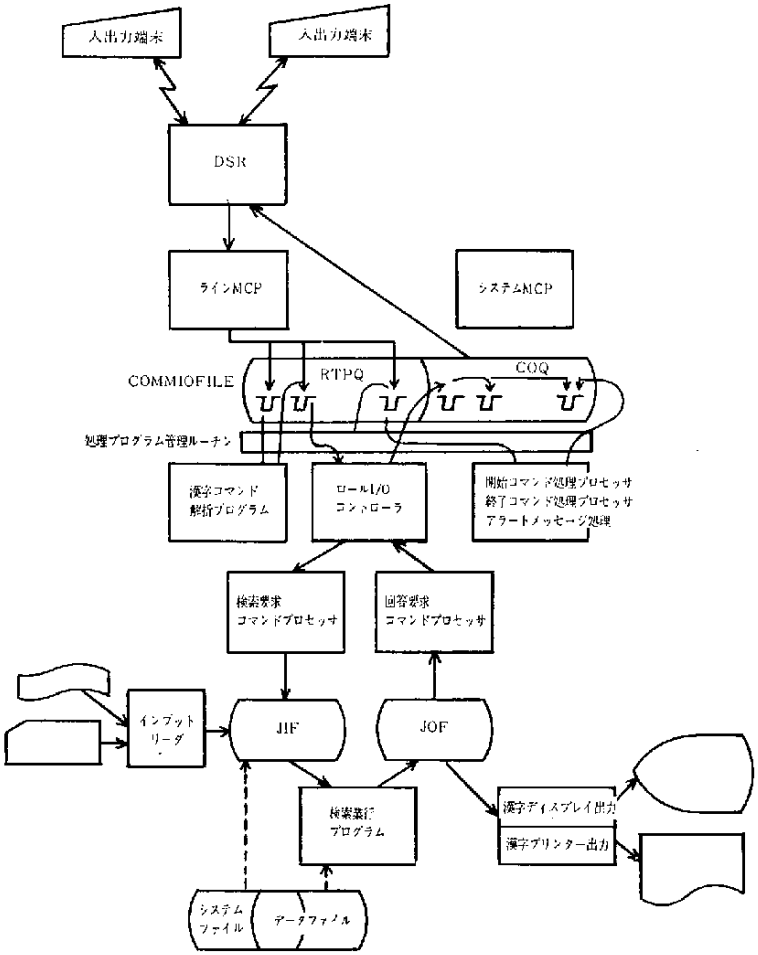


図6・4・6　メッセージの制御

なお，漢字ディスプレイあるいは漢字プリンタ出力は，それぞれのアウトプット・ライタ出力を経て出力される。また，伝送データのメッセージ交換は入力データとしてRTPQ登録後，転送コマンドの処理を受け，COQの伝送端末指定を変更する措置を行なう形で実施する。（図6.4.6参照）

(3) 検　索　手　順

標準的な検索処理手順を示すと以下のようになる。

◀開始

▶どうぞ

▶指示は　検索要求△T，L

▶準備完了

◀01　質問

02　氏名　山田一郎

03　標題　国際通商会議の開催状況

04　情報　01＝国際，02＝政治，03＝通商

05　索引　04＝通商協議

06　論理　（01＋02＋03）＊04

07　出力　会議

08　時点

▶QA0726*001

▶指示は　◀回答要求　A

▶回答件数　QA001＝01

＊回答出力＊

▶指示は　◀終了

▶了解

（注）◀端末からの入力　▶端末への出力

以上のようにコマンド処理を簡易にするために，できるだけ単純な会話処理としている。またコマンドおよびメッセージを一部漢字で表現している。

6.4.4　漢字入出力機器の概要

(1)漢字入出力端末装置

ア．通信伝送方式

漢字入出力端末装置はCO－59（全国新聞社メッセージ送信符号）を基準とし，これに拡張文字を選択し，合計3,800の文字符号を用意している。なおこの拡張文字の選択にあたって，頻度分析のほかに，処理担当者の主観によって選定した文字を加えている。

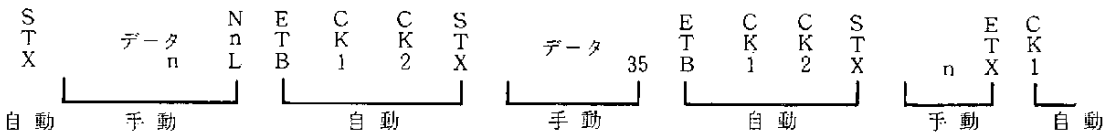
さらに伝送制御用として，SOH（宛先開始），STX（テキスト始，ブロック始符号），ETB（データブロック終符号），ETX（テキスト，終符号，ブロック終符号），EOT（送信終），ENQ（応答要求），ACK（受信側確認応答），NAK（不正受信応答），INT（割込み）等を使用し，このほか文字フ

ァンクションとして，ESC（コード変換），操作用としてCR（復改），BC（後退），CAN（削除）等を装備する。

通信開始時，端末装置の開始キーの押下げによりENQシーケンスを送信し，計算機よりのACK，またはNAKを受信して，送信または再度開始手順を行なう。端末装置よりの割込みは，INTキー押下げにより割込み信号に対するEOTの受信で送信権を得る。またCPUは，EOTを送信して，端末側に送信権を与え受信可能状態にする。通信の終結は，CPU側より，EOTを送信し，端末側からEOT受信後，再度CPUがEOTを送信して完了する。以上の処理は，伝送制御上6種のパッケージに分割し，それを随時組合わせることによってSC－4000漢字端末を制御している。

なお，アラーム処理は，端末受信中の場合には媒体アラーム・コードをCPUに伝送する。以後端末側は，アラーム原因を復旧させ，復旧キーを押し，CPU側よりアラーム・データを再受信する。端末送信中アラームは，CPU伝送を停止し，原因復旧後，開始を押し，現在入力中の1ブロックデータを再入力する。

電文伝送形式には，標準形データ入出力ブロックサイズが，35漢字を最大とする方法をとっているため，35漢字をもって自動復改し，自動的にETBが付加されて伝送単位としての1データブロックを構成する。なお，第1文字を読み取りSTXを自動発生し，そのデータを送信する。手動，自動の別は下記のようになる。



イ．機　器　性　能

漢字入出力端末装置は，谷村新興製，SC－4000型漢字印刷電信装置である。鍵盤部，印字部，せん孔部，読取部，制御部を有し，TC－50伝送制御装置を付属する。通信速度50BPS，調歩同期方式であり，半二重専用回線に接続されている。出力印字幅35字で，これによってA4版出力帳票を作るが，識別コードを除き内容印字幅は30字である。

鍵盤部は，48個の文字符号キーと，17個のファンクションキーを有する。文字符号は，2操作により，2,304種の文字を作成し，さらにESCキーとの組み合わせにより1,500文字を拡張可能である。その機能は，ESC 1とESC 2の組み合わせでは次の文字のみ拡張，変換してESC 3は次にESC 0がくるまでこの状態を保持する。

標準的な経常打鍵速度は，1分間平均100字程度である。

(2)　漢字プリンタ

NEAC－C5210S高速漢字プリンタの名称である。高速漢字プリンタ制御部・文字発生部増設付加機構，外字記憶部，印字文字拡張部等より構成される。字種は4,096を可能とする。印字速度は，1,600行／分であり，印字方式は，多針電極による静電記録方式をとる。なお連続折りたたみ式の静電記録

紙を採用，文字の構成は，24×24ドット，文字の大きさ 12ポイントである。

(3) 漢字ディスプレイ

NEAC－C5310S漢字ディスプレイの名称である。制御部，表示部，キーボード部，ハードコピー部から構成される。画像表示部は11インチ，画面表示字数21×21＝441文字である。表示部は，N3200－10制御部のディスク上のドットパターンを移送して表示する。

漢字入力装置は，文字盤拡張付加機構と，テンキー盤よりなる。文字盤は，2,205文字のNEACコードを操作し，そのほかは，テンキーによる外字処理を行なう。ハードコピーは，A4版でドライシルバ・ペーパーである。

ファンクションキーは16種あって，受信は，表示の指示，送信は画像完了後ジョブに転送，復元は，元の画像の復元，削除は，行面共にカーソルの処理で指定して実施する。挿入はキーボード入力，置換は，入れ替えを示す。障害発生時には，JOFから再出力して復旧する。

6.4.5 オンライン システムの評価と問題点

CROシステムの基本目的は，データ収集，管理作業をデータ提供者の機械操作に還元して，デスクから，機械に対応できるシステムを設計することであった。しかしこのような運用方式に対するデータ提供者の利用意志と，検索要求者の評価が，システムの目的と調和できるまで，かなりの時間を要すると思われる。なおCRO－V0の設計上に関連した問題点を列挙する。

文字符号，伝送符号の標準化は，システムのコスト，端末機器のコスト・ダウンに貢献するものである。

CROシステムでは，アラーム処理と，割込機能が問題となった。これらは，ユーザ側の運用に関することであるが，電電公社の標準仕様の中である程度ユーザ間の標準基準を設定することが必要である。

通信制御機能の標準化は可能ではないかと思われる。他の端末装置間のメッセージ交換は，コード系，伝送手順が標準化されれば問題はないが，重要なことは，周辺機器を介しての制御の機能と範囲が明確にされることである。

オンラインデータ・ベースの問題は，日本語特有のファイル形式と開発されるモニタ・システムの管理方式下にあって，多様性をおびてきている。このような点から，データ・ネットワークの設定のためには，ネットワーク集団とその内でのデータ・ベースの互換性が必要となる。

システムの開放に伴いデータ保護の問題が重要となり，これらは，運用上の取り決めや，これより発生する諸問題の法制上の措置が考慮されねばならない。

6.5 物品管理システムの漢字出力の試み

6.5.1 沿　　革

いまのところ電子計算機で処理できる文字は，英字，数字，カナ文字などが一般的である。しかし事務処理などの場合，漢字が扱えることができれば，非常に有効である。

現在電子計算機利用に関する技術研究会の物品システム開発研究班が(注1)，物品管理資料作成サブシステムを作成していて，その際多品種の種々雑多な物品の管理を計算機で行なう作業が試みられている。この際，物品名，業者名，内容などを漢字で表わすことができれば，物品管理者にとって，物品の保管維持が容易になることをねらって，漢字プリンタを使用してみた。この物品管理システムは，全官庁を対象にして共通に使用できることを目的にしている。しかし各官庁で使用している計算機にはメーカも大きさもいろいろあり，また同一であってもシステムの構成は大分異なっているのが普通である。したがっていろいろな機種に共通に使用できるようにソフトウェアを考慮しなければならない。さらに各官庁がいままでに物品管理のやり方の違いなどをどのように処理するかなどの問題がある。

このような物品管理システムのもとで漢字プリンタ出力は，やはりいろいろな機種に共通に使用できるように考慮することが必要であり，そのソフトウェアを作成しテストして，経験を得たので報告する。

物品管理システムでは従来どおり，英文字，カナ文字による出力形式のまま磁気テープにファイル化されるが，これを漢字出力する場合，あらかじめ作成した漢字コード・テーブルを参照しながら漢字出力イメージ・テープに変換すると同時に，漢字プリンタにて出力結果が得られるように編集用ソフトにて編集処理して結果を得るようにしている。

(注1)政府省庁の電子計算機利用上の技術的諸問題を研究することを目的とした研究会で，現在27省庁が参加している（上田陸奥夫主査:社会保険庁)。

6.5.2 システム設計

(1) 物品管理システム

官庁の物品管理は，物品管理法等に基づく物品管理の適正かつ効率的な運用をはかるために，正確かつ迅速な管理が要求されている。このため電子計算機の利用により，事務の簡素化，合理化をはかり，また不用物品，遊休物品，代替物品の活用など物品の効率的使用という付随的な効果をねらって設計したもので，将来性に富んだ標準システムであり，汎用性，拡張性および有機性をもっている。

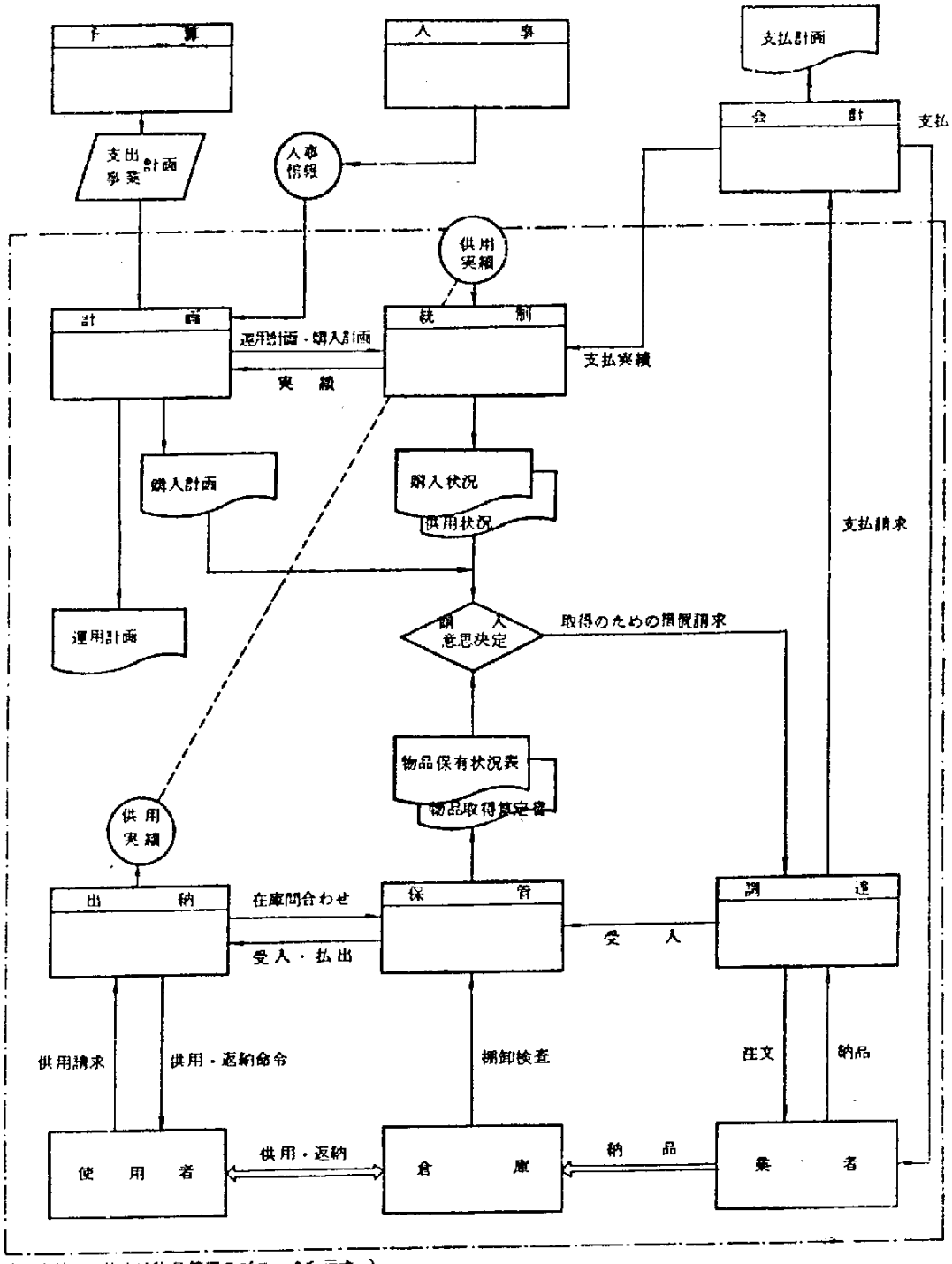
このシステムは5つのブロックからなり，それぞれの目的を分担し，相互の独立性を保ちながら有機的に結合している。各ブロックの目的，機能，内容を表6.5.1に，各ブロック関係を図6.5.1に示す。図6.5.2はシステムフローチャートである。

(2) 字　　種

電子技術総合研究所で使用している物品管理のための字種は表6.5.2～6.5.10のようなものである。

ブロック	保管	出納	調達	計画	統制
目的 機能	物品の保有状況の把握 照会システム 単一備品毎の管理 オン・ライン化	物品の供用および返納 法定帳簿、命令書、報告書の作成 帳票類の一元化	物品の調達と納期管理	物品の需給計画の立案 物品の運用計画の立案	計画と実績との比較
内容	◦保管マスターの初期ファイル作成 ◦物品の点検、棚卸 ◦物品の保有状況表の出力 ◦物品の保管現在高の照会 ◦物品に対する問合せ照会 ◦単一物品毎の管理に適した保管マスターファイル ◦物品管理官、物品出納官、物品供用官、部課、使用職員別に管理 ◦保管マスターはダイレクト・アクセス・ファイル	◦保管マスター二次引当ファイルの更新 ◦物品の受入、払出 ◦物品の異動状況の記録 ◦物品の売払・廃棄	◦受入マスターの初期ファイル作成 ◦物品の取得 ◦物品の修繕・改造 ◦物品の貸付・寄託 ◦物品の保管 ◦納期遅れに対する督促	◦需要計画の作成 ◦運用計画の作成 ◦購入計画の作成 ◦購入限度額の設定 ◦物品カタログの作成	◦購入計画と購入実績との比較 ◦運用計画と運用実績との比較
入力帳票	物品保有状況表（チェック済） 棚卸差異分析表（調整済） ※物品管理簿 ※物品供用簿 棚卸検査報告書	※供用請求書 取消請求書 ※返納報告書 供用換通知書 ※分類換通知書 ※管理換通知書 契約措置済通知書 ※不用決定承認書 物品取得査定書	※管理換通知書 ※取得通知書 契約措置済通知書 納品書		
ファイル	保管マスター 初期ファイル	二次引当ファイル 本日分異動状況ファイル 累積異動状況ファイル	受入マスター 本日分異動状況ファイル 累積発注受入ファイル	支出計画ファイル 事業計画ファイル 需給計画ファイル 運用計画ファイル 購入計画ファイル	
出力資料	物品保有状況表 棚卸差異分析表 物品取得算定書 物品払出予定通知書	※物品出納命令書 ※分類換通知書 ※物品管理簿 ※物品出納簿 ※物品供用簿 ※使用者別物品供用簿 ※物品増減及び現在額報告書 部門別備品請求額表	物品受入予定表 督促状 購入先一覧表	需給計画表 運用計画表 購入計画表 物品カタログ	運用状況 購入状況

（※印の入出力資料は法定帳簿及び命令書、通知書である）　**表 6・5・1**　物品管理のためのブロックと目的機能



（一点鎖線の枠内は物品管理のブロックを示す。）

図 6・5・1　物品管理システム関連図

電　総　研　　文字てーぶる

文字コード表　　プルーフ．リスト

キー	文字
4CI	アルファメーター
3EV	アーク安定度測定装置
3PB	暗視装置
5AD	安定度測定器

表 6・5・2　物品コードの例

単　位	コード	単　位	コード
び　ん	01	食	30
木　箱	02	ロ　ー　ル	31
センチメートル	03	セット(揃)	32
か　ん	04	枚	33
ボール箱	05	⋮	⋮
⋮	⋮		

表 6・5・3　単位一覧表

事　由　内　容	コード
事　務　用　品	01
運　　搬	02
不　　用	03
余　　り	04
使　用　不　能	05
⋮	⋮

表 6・5・4　事由コード

物　品　区　分	コード
重　備　品	1
普　通　備　品	2
消　耗　品	3

表 6・5・5　物品区分コード

場　所	コード
永　田　町	100
木　挽　町	200
田　　無	300
大　　阪	400

表 6・5・6　場所コード

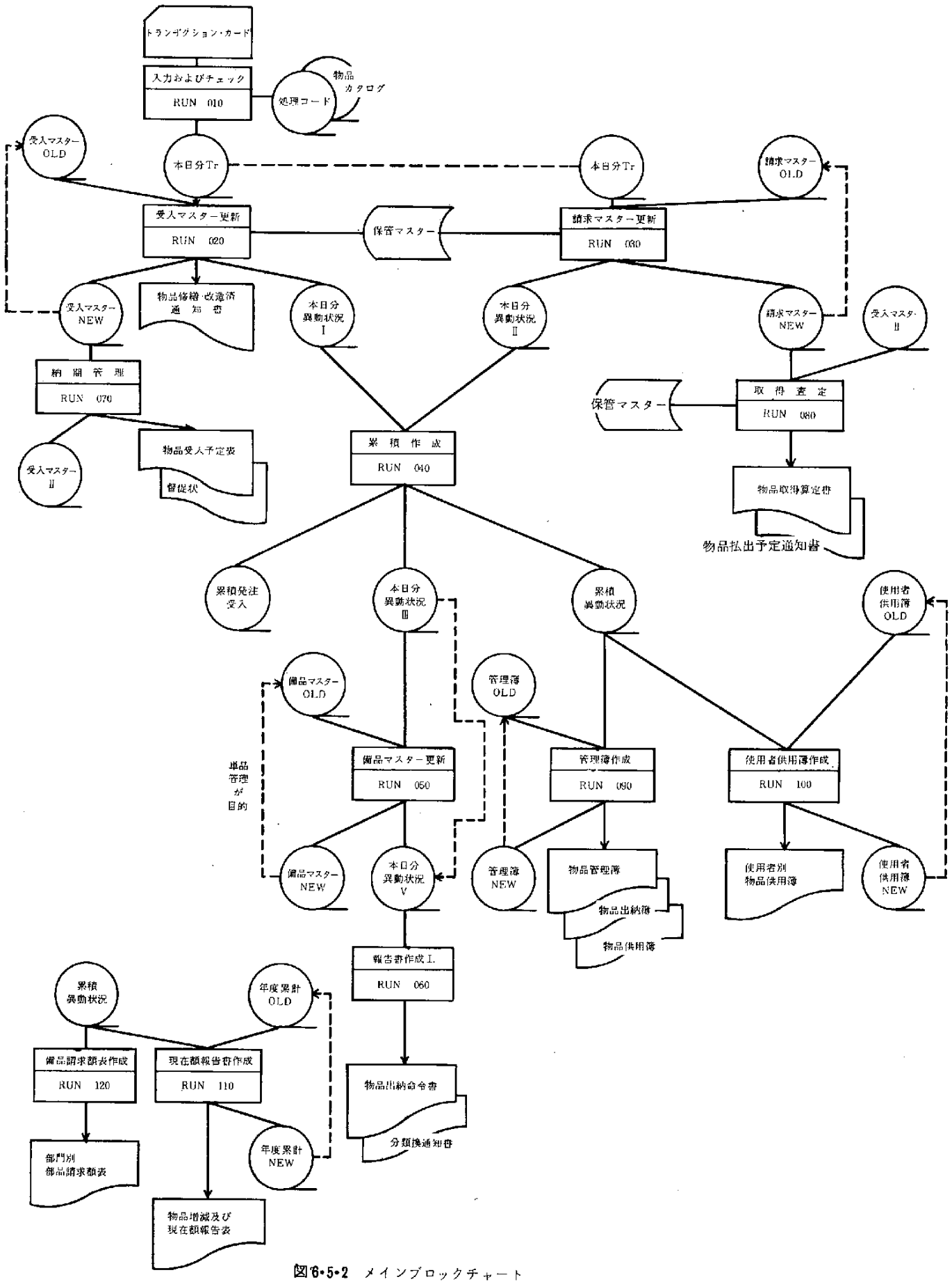


図6・5・2 メインブロックチャート

電　総　研　　文字てーぶる

文字コード表　　プルーフ．リスト

9999

キー　　文字

0001　　石川島　磨重工業（株）

田無市向台町3－5

0424－62－2111

巻田　実

0002　　ゼネラル（株）

中央区日本橋通り2－2東京支店

271－5392

塩田　史郎

表6･5･7　業者コードの例

整理区分	コード	整理区分	コード
購入	01	月計	30
寄付	02	累計	31
譲受	03	翌年度へ繰越	32
借受	04	前年度より繰越	33
生産	05	雑件	99
⋮			

表6･5･8　整理区分コード

会計分類	コード	会計分類	コード
一般会計	01	アルコール専売事業	32
造幣局	10	輸出保険	33
印刷局	11	機械類賦払信用保険	34
資金運用部	12	木船再保険	35
貴金属	13	⋮	
⋮			

表6･5･9　会計コード

省庁名	コード	省庁名	コード	省庁名	コード
会計検査院	01	大蔵省	40	中央労働委員会	76
内閣	02	国税庁	41	公共企業体等労働委員会	77
内閣法制局	03	印刷局	42	建設省	80
人事院	04	造幣局	43	自治省	85
国防会議	05	文部省	45	消防庁	86
総理府	06	文化庁	46		
⋮		⋮			

表 6・5・10 省庁コード

6.5.3 システムの概要

(1) ハードウェア

ここで使用する漢字プリンタはフォントメモリに文字パターンを記憶し，これをアドレス化しておいて，磁気テープから入力されるデータと一致する文字を，ファイバーチューブ・ディスプレイによって，電子記録紙に印刷する装置で，高千穂交易(株)のT4100漢字情報処理システムである。

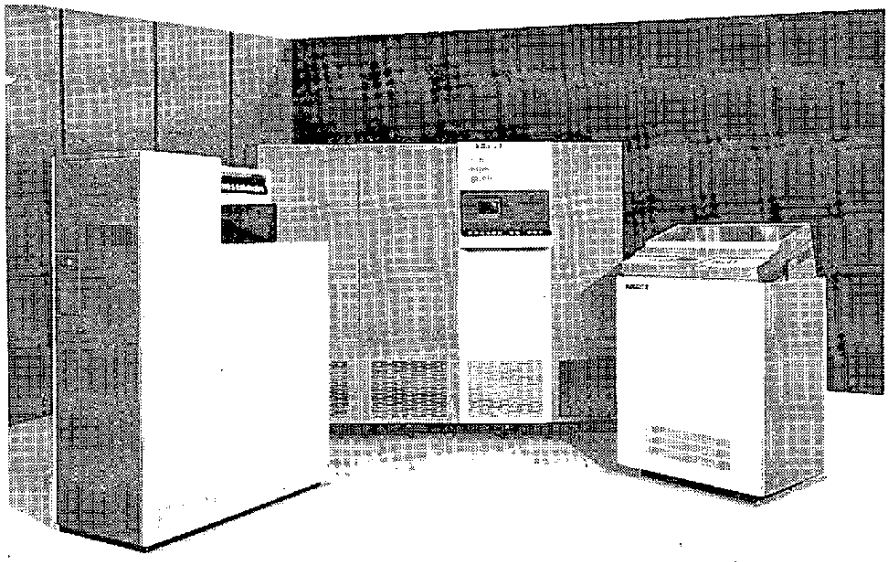


写真 6・5・1 T4100漢字プリンタシステム
(左・磁気テープ装置　中・フォントメモリ装置　右・プリンタ)

ア．フォント・メモリ装置

フォント・メモリ装置はT4100システムの中核をなすもので文字パターンを記憶する。このフォント・メモリにはLSIを用いて文字単位にランダムアクセスができる。文字パターンの高速読み出し，書き込みが可能になったため，小容量のフォント・メモリで多量の字種の取り扱いが可能である。これをフローティング・アドレス方式とよんでいる。

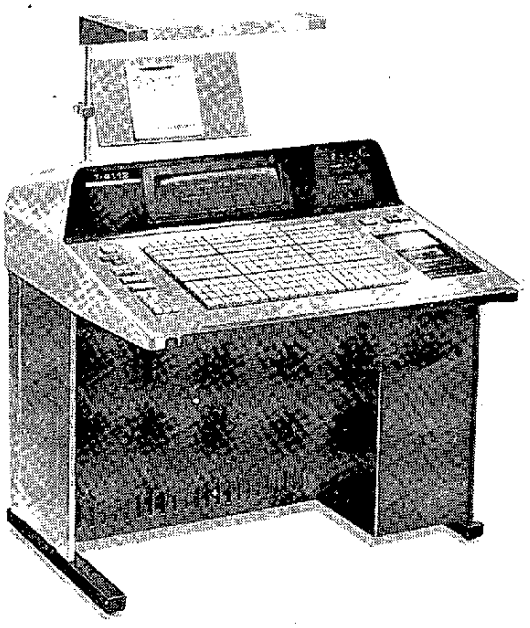


写真 6-5-2　T 4112漢字鍵盤ディスプレイ・レコーダ

フォント・メモリに記憶された文字を読み出すために，文字ごとにアドレスがついており，このアドレスは2バイト（＝16ビット，ヘキサデシマル）でコード化されている。

したがって，入力データ・テープにはこのアドレス（文字コード）が1文字当り2バイトで記録されているし，漢字鍵盤穿孔機のキーボードから打鍵することによっても，これらの文字コードが得られるわけである。

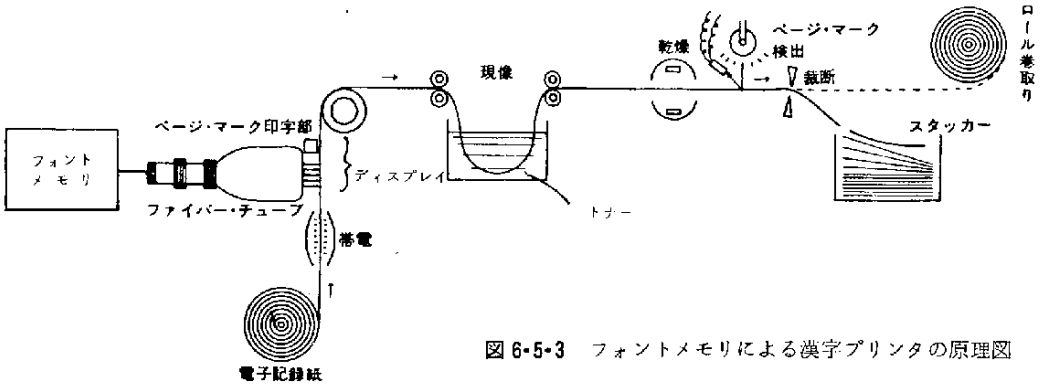


図 6-5-3　フォントメモリによる漢字プリンタの原理図

イ．磁気テープ装置

磁気テープ装置は，フォント・メモリ装置に，文字パターンと文字コードを入力する装置である。磁気テープは9トラック，800BPIを使用する。

ウ．プ　リ　ン　タ

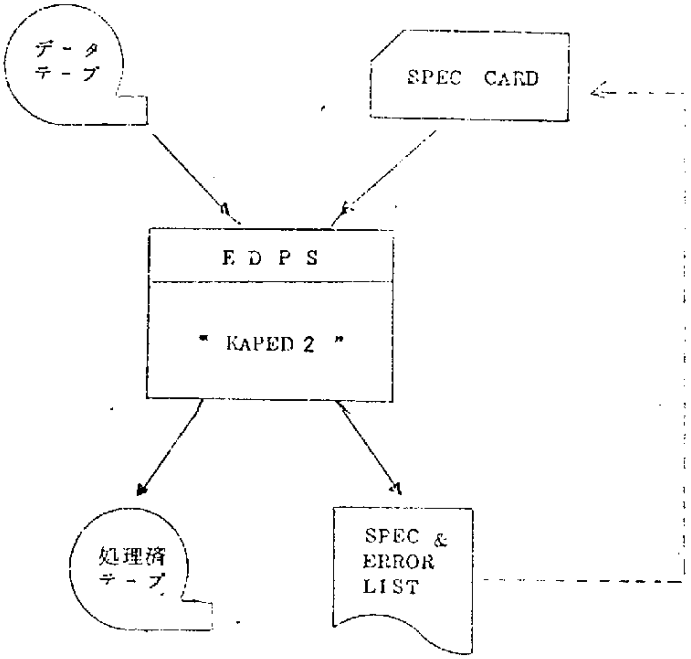


図6・5・4　KAPED-2

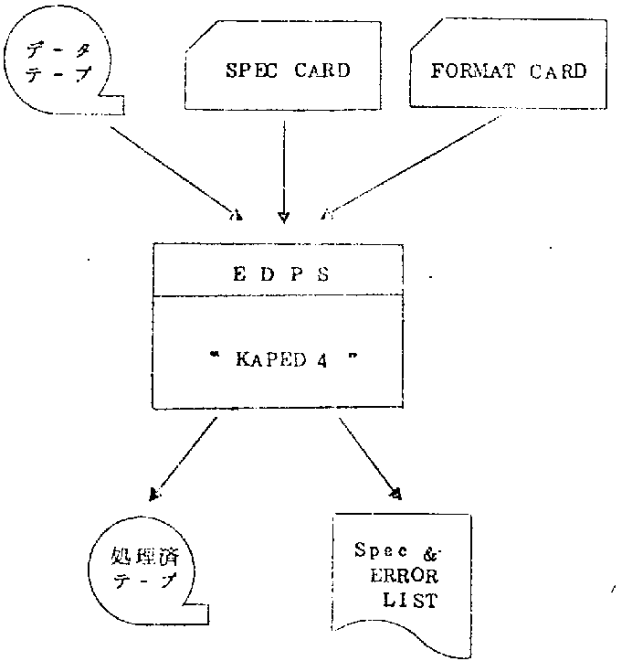


図6・5・5　KAPED-4

プリンタはオプティカル・ファイバーチューブを用いた電子記録方式による高速度プリンタで，印字結果をページ単位に裁断してアウトプットする。

(2) ソフトウェア

T4100システムのユーティリティ・プログラムとして用途別に次のようなものがある。

ア．FG (FONT GENERATOR)

文字パターンを磁気テープ（フォント・マスタ・テープ）に記録する。

イ．FMMP (FONT MASTER MAINTENANCE PROGRAM)

FONT MASTER TAPEの内容の追加，変更，削除を容易に行なうためのプログラム。

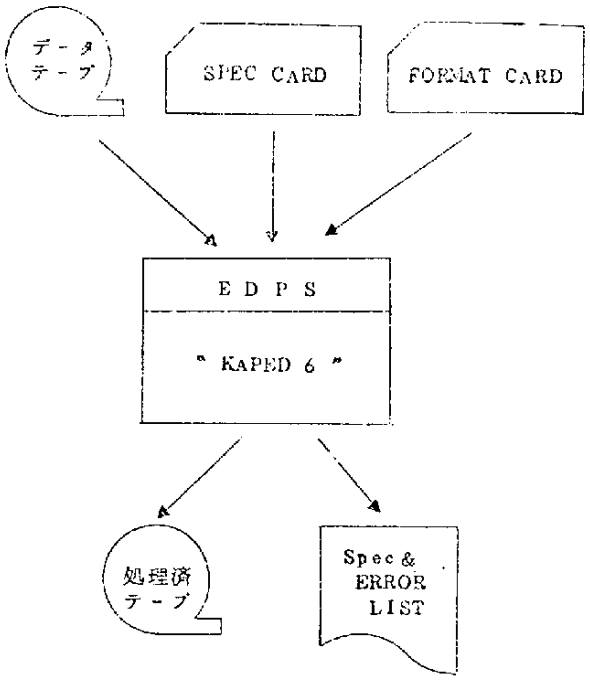


図6-5-6 KAPED-6

ウ．KAPED－2

レコード上の任意の部分を編集指示カード（SPEC CARD）でページ単位に編集する。宛名ラベル，電話帳，名簿等の作成に適する。

エ．KAPED－4

一定のフォーマット（固定項目，枠組，表題等，各頁に共通のもの）に，任意の場所にSPEC CARDの指示でデータを挿入しページ単位に編集する。伝票等の編集に適す。

オ．KAPED－6

KAPED－4の機能を拡張して汎用性を持たせたもので連続的にリストできる。一覧表，統計表等の編集に適する。

(3) ファイル

ア．文字コードテーブルファイル

各コード別にディスク上にファイルとして作成した。ファイルのレコード形式は次のとおりである。なお1文字は2バイトである。

i　業者コードだけは次のようにした。

コード	業者名	住　所	TEL	代表者名
4バイト	20文字 40バイト	25文字 50バイト	14文字78バイト	9文字18バイト

ii　その他のコードはすべて次のような共通フォーマットにした。

コード	文字部
4バイト	14文字78バイト

iii　ディスク内におけるファイル名とテーブルの対応は表6.5.11のようにする。

イ．漢字処理システムへのインプットファイル

漢字処理システムへのインプットファイルは物品管理システムのラインプリンタへの出力を磁気テープに変換したファイルで，ラインプリンタのコントロールバイトが先頭につく。（図6.5.7参照）

各テーブル名	ディスク中に登録されているファイル名
物品名 TABLE (9001)	▼DK−01▼
省庁名 TABLE (9002)	▼DK−02▼
会計 TABLE (9003)	▼DK−03▼
人名 TABLE (9004)	▼DK−04▼
分類、細分類 TABLE (9005)	▼DK−05▼
単位 TABLE (9006)	▼DK−06▼
物品区分 TABLE (9007)	▼DK−07▼
整理区分 TABLE (9008)	▼DK−08▼
事由 TABLE (9009)	▼DK−09▼
場所 TABLE (9010)	▼DK−10▼
所属 TABLE (9011)	▼DK−11▼
業者 TABLE (9999)	▼DK−12▼

(　　)内のコードはダイレクト・アクセス用キーである。

表 6・5・11 ディスク内におけるファイル名とテーブルの対応

(4) 作業手順

① まず出力する文字の種類，内容と各種のコードについて指示する。

② 文字イメージを作成し，コード化してフォントメモリにロードする。同時に各種コードを文字コードでテーブルを漢字テレタイプで作成する。

③ 提供されたテーブルを媒体変換する。

④ 出力ルーチンから得られる英文字，カナ文字の出力イメージを，テーブルを参照して文字出力，テープファイルを作成する。

なお出力ルーチンで漢字コードテーブルを参照して一気に漢字出力テープファイルを作成してもかまわない。

⑤ 提供されるユーティリティプログラムで，漢字処理装置で処理できるように編集する。

物品管理簿

コントロール	省庁コード	会計コード	管理官コード	分類コード	細分類コード	物品コード	品　目	単　位	空	物品区分	作成年月日
1	2	2	4	2	2	3	20	2	66	1	10

コントロール	異動年月日	整理区分	証書番号	異動高					現在高				
				増加数量	増加価格	減少数量	減少価格	その他	供　用	貸付寄託	保　管	合計数量	合計価格
1	10	2	7	6	17	6	17	6	6	6	6	7	18

物品出納簿

コントロール	省庁コード	会計コード	管理官コード	出納官コード	分類コード	細分類コード	物品コード	品　目	単　位	空	物品区分	作成年月日
1	2	2	4	4	2	2	3	20	2	18	1	10

コントロール	異動年月日	整理区分	証券番号	異動高		現在高			
				増加数量	減少数量	供　用	保　管	計	
1	10	2	7	6	6	6	6	7	20

図6・5・7　インプットファイルのフォーマット

⑥ 編集されたファイルを漢字処理装置にセットして，オフラインで結果を得る。

(5) 出 力 例

物品管理簿，物品出納簿の例を表6.5.12，表6.5.13に示す。

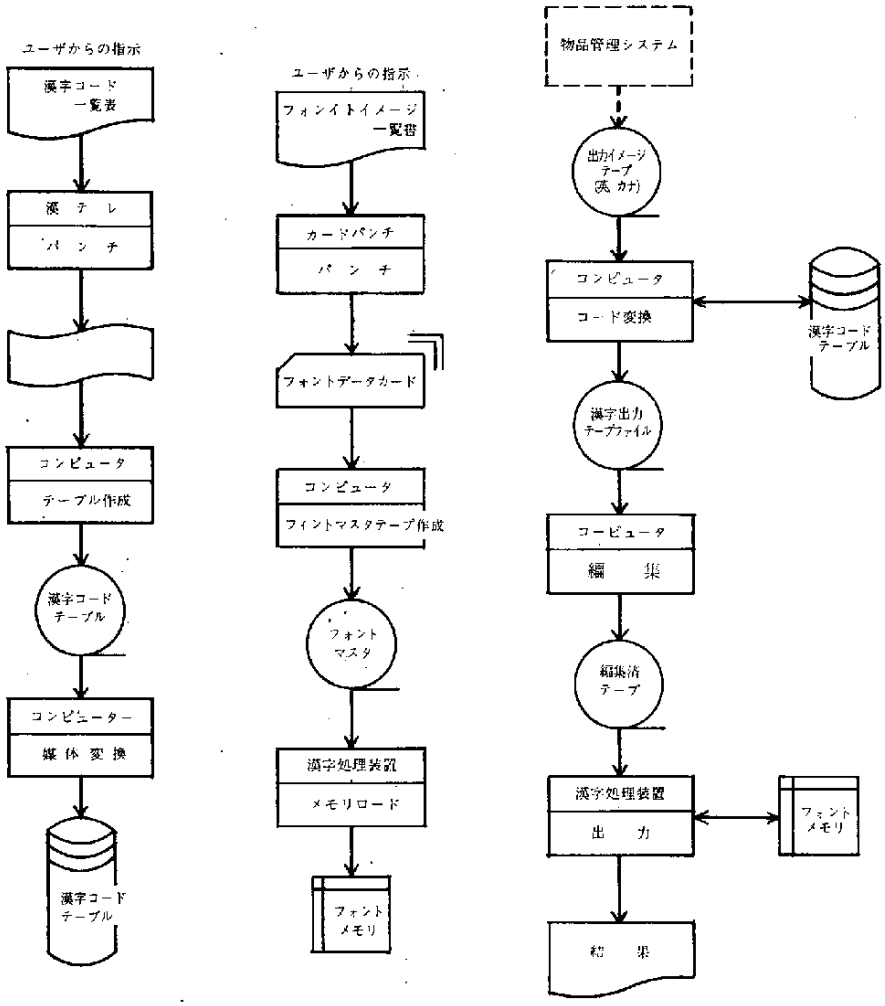


図 6-5-8 漢字処理システム

6.5.4 運 営 状 況

漢字プリンタを使用するため，一昨年7月から電総研2名，高千穂交易2名が参加して実質期間2カ月位で所期の目的の出力を得ることができた。その結果，漢字プリンタのユーティリティルーチンもいくつか整備され，今後漢字プリンタを使用するユーザは，コード変換プログラムが約300ステップ，媒体変換プログラムが約100ステップのコボルプログラムで作成できるようになった。

文字テーブルが作成されれば，漢字の入力はなくなり，文字テーブルは修正だけになる。これは年度初めに行なわれるが，量は100品目以下である。

様式番号 308

物　品　出　納　簿

省庁名	会計名	管理官	出納官
工業技術院	一般会計	吉江清	鈴木芳雄

物品区分	作成年月日
普通備品	S. 43 3 3

分類	細分類	物品名	単位
工業技術院用品	備品	プラズマ発生器	一式

異動 年月日	整理区分	証書番号	異動高 増加数量	異動高 減少数量	現在高 供用	現在高 保管	現在高 計
S. 43 3 30	譲受		1		1		1
S. 43 4 1	前年度より繰				1		1
S. 44 4 1	前年度より繰				1		1
S. 45 3 31	雑件				1		1
S. 45 3 31	雑件				1		1
S. 45 4 1	前年度より繰				1		1
S. 46 1 20	譲受		1		2		2
S. 46 1 20	雑件				2		2
S. 46 4 1	前年度より繰				2		2

表 6-5-13 物品出納簿出力例（原寸大）

様式番号 307

物品管理簿

省庁名	会計名	管理官	分類	細分類	物品名	単位
工業技術院	一般会計	吉江清	工業技術部組織	工業技	プラズマ発生器	一式

物品区分	作成年月日
組 1	S. 43 3 30

異動年月日	整理区分	訂番号	異動高 増加数量	増加価格	減少数量	減少価格	その他	現在高 供用	貸付等託	保管	合計数量	合計価格
S. 43 3 30	親受		1	1,050,000				1			1	1,050,000
S. 43 4 1	前年度より繰越							1			1	1,050,000
S. 44 4 1	前年度より繰越							1			1	1,050,000
S. 45 3 31	雑件			80,000				1			1	1,130,000
S. 45 3 31	雑件			602,600				1			1	1,732,600
S. 45 4 1	前年度より繰越							1			1	1,732,600
S. 46 1 20	親受		1	5,237,550				2			2	6,970,150
S. 46 1 20	雑件			600,000				2			2	7,570,150
S. 46 4 1	前年度より繰越							2			2	7,570,150

表 6・5・12 物品管理簿出力例（縮小≒½）

物品管理は月一度のバッチ処理であるが，備品数は約2,000，レコード数は40,000以上で，これらすべてが処理される。漢字システムへの入力は，システムからラインプリンタ・イメージで，磁気テープファイルより行なわれる。

6.5.5 システムの評価

(1) 出 力 装 置

1文字ずつアドレスがついていて，1文字32ビット×32ビットなので，文字の形を自由に変えたり，文字体系を自由に定めることができる。

また浮動アドレス方式を用いることができるので，特殊な文字も処理でき，字種が豊富に扱える。

(2) ソフトウェア

漢字編集用アプリケーションがそろっているので，出力編集が文字だけでなく，枠組み，絵などの用途が広く自由である。

また文字テーブルはキーと単語で作成でき，英数字コードでテーブル処理が行なえるので，従来のシステムを漢字処理に変更することなく適用でき，英数字コードから漢字コードへの変換も，どんなデータでもキーで処理できるので容易にすみやかに行なうことができる。

(3) コ ー ド 体 系

データのコード体系は単語を単位として定めることができるので，従来システムとは変わらない。

しかし，文字1字について考えるなら，多量で，種々の体系があるので容易に定めることができなく，文字テーブルを容易に作成できなくなる。

6.5.6 将 来 の 構 想

研究情報管理システムは図6.5.9のようなハイアラキシステムが考えられている。現在実務担当者レベルで，会計システム，人事システム，物品システム，技術情報システムの一部が稼動している。

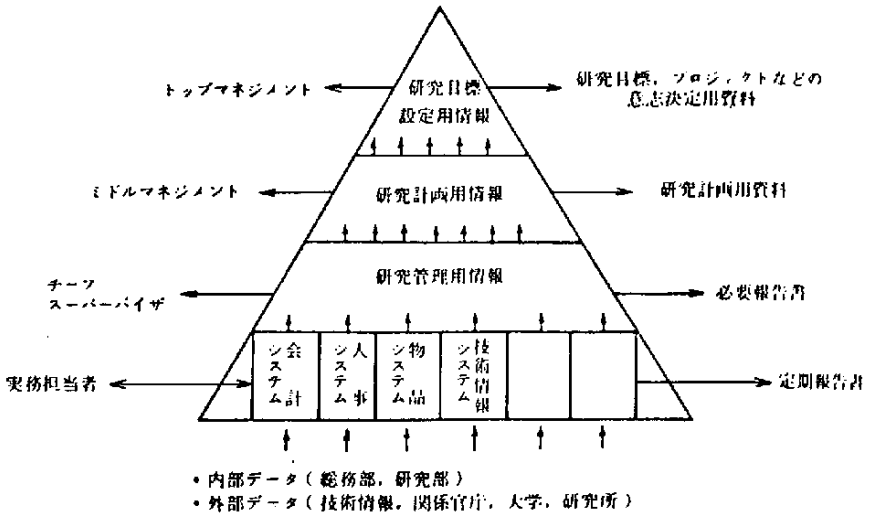


図6・5・9 研究情報管理システムのハイアラキ

しかし物品管理を除いた他のシステムはまだ漢字処理を行なっていない。

漢字処理システムは英数字で処理した後に漢字処理を行なうので，現在稼動しているシステムおよび上級レベルのシステムすべてに適用することは容易である。

したがって，このシステムの基本は，従来のプログラムで処理された結果を漢字にして出力するだけであるが，今後の問題として入力の段階から漢字処理することが検討され，研究所の科学用データバンクへの適用についても調査している。

6.6 生命保険会社における日本語情報処理

6.6.1 沿　　革

(1) 保険契約管理事務システム

生命保険会社における日本語情報処理の対象としては，保険契約に関する顧客の氏名・住所に関する漢字情報処理が主であり，各生命保険会社において顧客情報のカナ文字化あるいは漢字化が計画実施の段階にきているが，昭和46年に生命保険業界で最初に漢字情報処理システムの実用化をおこなった明治生命の事例について紹介する。[注1), 注2)]

生命保険契約管理システムの基本となる事務は，保険契約を20年，30年という長期にわたって維持管理し，保険金支払いを可能とすることにあり，そのため長期にわたる保険料の払込み記録の完備体制が必要となる。

これらの徴収保金事務は当社の場合，全国に散在する支社（約100店），営業所（約1,500店），代理店（約2,000店）などの集金機関の事務組織で遂行されており，図6.6.1に示すように新契約，継続分の契約内容の異動等の契約記録の更新は膨大な事務量となっている。

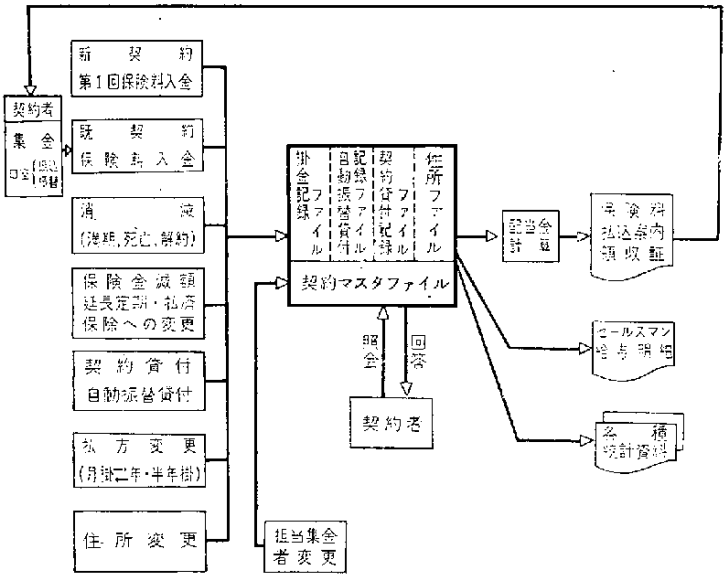


図6-6-1　生命保険契約管理事務の概要

当社の普通保険契約の件数は，普通の月掛契約が330万件，年掛・半年掛契約が120万件，集団月掛契約が110万件，合計560万件，このほかに延長・払済契約や団体保険，企業年金保険などを合計すると1,000万件にもおよぶ。

これらの膨大な件数の契約管理のために，当社では昭和37年にPCSからコンピュータ処理に切り換え，昭和42年にIBM 360／40を，昭和44年に同360／65を併設し，図6.6.2に示されるシステム構成によって処理されている。

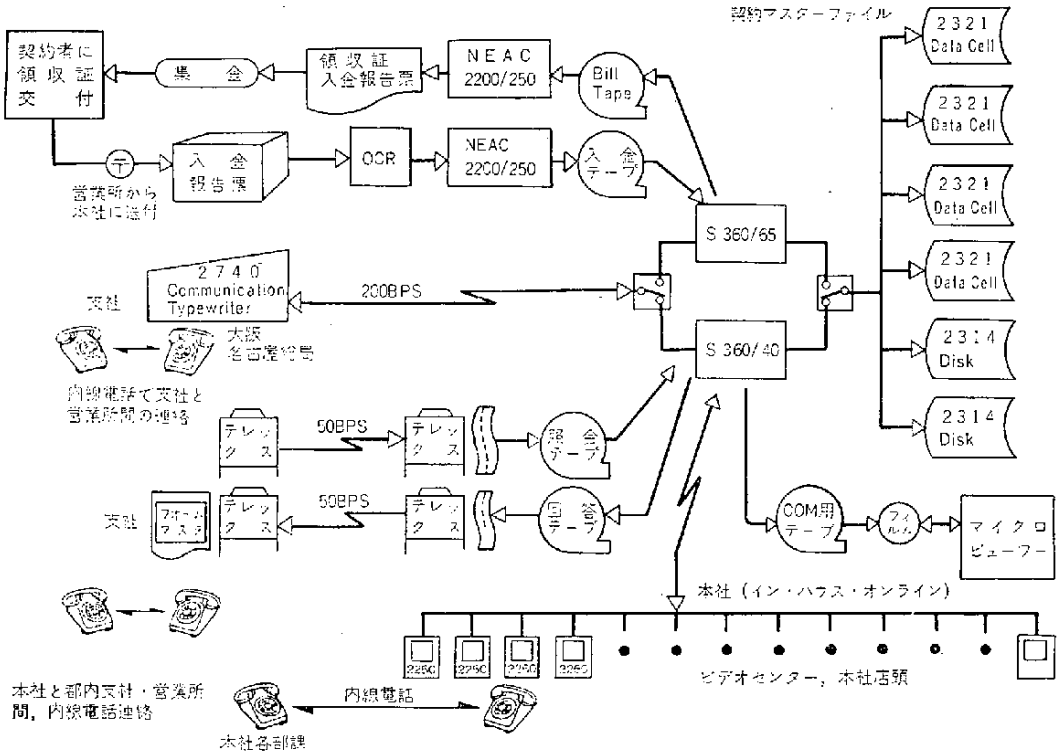


図6・6・2　契約管理システム

とくにこのうち，契約マスターファイルは本社店頭，大阪・名古屋両総局店頭で，ビデオ・ディスプレイ，タイプライタ通信装置にオンラインで接続され，契約内容その他の自動的な照会応答システムが確立されている。

(2)　漢字情報処理システムの必要性

一般に生命保険会社におけるデータ・ベースの一つである顧客の契約マスターファイルは証券番号（数字コード8桁）を基本としたインデクスド・シークェンシャル（IS）ファイルの構成で，証券番号がなくてはファイルの検索ができない。

店頭あるいは電話などによる顧客からの契約内容の照会は，ほとんどが氏名だけによって要求される場合が多く，現在では顧客氏名索引カード[脚注]を手作業で抽出し，証券番号を得てから契約内容の検索をおこなうという方法がとられている。

ところが顧客氏名そのものが漢字ファイルとして完備されると，次のようなことが可能となってくる。

①　顧客氏名を直接漢字テレタイプ端末機から漢字そのままのイメージでキー・インして照会でき，しかも漢字でビデオ画面に応答表示することにより，証券番号が不明でもただちに契約内容の照会回答が得られる。

②　一契約ごとの顧客氏名を名寄せすることにより，同一契約者の複数加入契約が一元的に把えら

れ，保険金の加入最高限度調査がコンピュータで求められる。

③　カナ文字による種々の払込案内，領収書，配当金支払明細が漢字化され，顧客サービスの向上ができる。

④　その他，現在数字項目を機械プリントした資料に顧客氏名を手書きしてきた種々の帳票は，すべてコンピュータでプリントされるようになる。

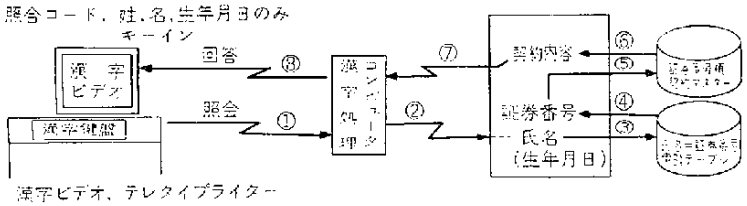


図6-6-3　漢字による契約照合応答の流れ

以上のような単なる事務的な見地からした漢字情報処理の必要性のほかに経営情報面での次のような活用が考えられる。

コンピュータによる事務処理の機械化が定着してくると，省力化後の機械化は徐々に経営情報面への活用に指向される。この場合，コンピュータのアウトプットの経営情報資料が英・数字・カナ文字だけでプリントされているとすれば，きわめて見難いものである。

これらの資料がコンピュータから直接親しみやすい漢字まじりの日本語でプリントされ，もしくはビデオ・ディスプレイ上に要求した情報が表示されることは利用者にとってきわめて理解しやすく，デシジョン・メーキングの助けとなることができる。

また，当社のもう一つの大きな必要性として，年・半年掛保険料の徴収は郵便（銀行）の窓口振込みや代理店，支社扱いであるため，昭和36年から契約者住所，氏名のカナ文字ファイルをもっているが，一般の月掛契約は担当セールスマンによる集金が主体のため，顧客氏名の本社ファイルがなかったので，カナ文字の氏名ファイルを作成する労力を，一足飛びに漢字で持つ結論に達したこともあげられる。

（注1）生田　勝彦："漢字のインプット・アウトプット方式の設計"
事務管理，Vol. 11，No. 8，PP65～73（1972）

（注2）生田　勝彦："明治生命における漢字情報システム"
ソフトウェア科学，Vol. 4，No. 3，PP37～45（1972）

（注3）被保険者名や契約者名から証券番号を索引できるカードで，同姓同名を区別するために，生年月日，その他契約日，住所などの記載をもつカード。

6.6.2　システム設計

当社では前述の要求に対処しうる汎用性のある漢字情報システムを調査検討の結果JEM－3100S[注3)]1セットおよびF6502A漢字ラインプリンタ付FACOM230／25[注3)]1セットを導入することを決定し，システム設計を行なった。

(1)　文字種の選定

使用文字の対象は主として氏名・住所で，氏名に関しては6,000～10,000字種が必要であるといわれており，規定をすることができない。

一方，住所に対しては，都道府県，市町村字までで約4,400字種の漢字が必要とされており氏名・住所をあわせて6,000字種以上が最低必要である。

JEMシステムの場合，文字の収容能力は20,000字種まで拡張できるので，初めは漢字に対しては，住所地域に使用される漢字を含むJEMの標準6,500字種を採用し，その後必要に応じて文字種を拡張することとした。

その結果，文字種としては下記の7,847字種を第1次として決定した。

明朝漢字	6,451	明朝両仮名	204
ゴシック漢字	520	ゴシック両仮名	204
英欧字	228	約物・記号	260
数字	20		

(2) 漢字コード

漢字コードはJEMシステム標準のEXEKコード[4]とし，IBMシステム内も漢字はEXEKコードを用い，FACOM漢字プリンタでプリントする場合はFACOMシステム内でEXEKコードとFACOM漢字プリンタコードとの変換を行なっている。

(3) FACOMプリンタシステムの用い方

JEMシステムのプリンタは，後述するように銀塩安定化紙を用いているので，コンピュータ出力のようにフォーム印刷をほどこすことができないので，FACOM漢字プリンタシステムの特徴を活かし，対外フォーム印刷用紙への漢字アウトプットの専用機として用いることとした。

(4) JEMシステムの用い方

JEMシステムは漢字処理用のターミナルコンピュータとして，次の三つの用い方をもつシステム設計とした。

ア．入力紙テープから磁気テープへの変換

後述するようにオフライン漢字鍵盤でさん孔された紙テープを，IBMコンピュータで編集処理するための磁気テープへの入力処理を，JEMシステムで行なう。入力処理の主な内容は，さん孔紙テープのEXOKコード[4]から内部コードのEXEKコードへの変換，および漢字辞書処理（6.6.4漢字入力装置，参照）である。

イ．ディスプレイ・システムとして

ディスプレイ・システムの用途としては，システム入力用のさん孔紙テープを，ディスプレイによる検定を行ないながらさん孔する入力用と，窓口ターミナルとして契約照会ディスプレイの出力用としての二つの用途がある。

前者はJEM－3100システム内でのオペレーションで，各ターミナルで鍵盤入力→ビデオ表示検定→紙テープさん孔の入力オペレーション，あるいは，紙テープ入力→ビデオ表示検定→紙テープさん孔の校正オペレーションが行なわれる。

後者ではJEMシステムはIBMシステムと2,400BPSの通信回線でオンラインで結合され，IBMシステム内の2260／2848ディスプレイ装置と同等の取扱いができIBMのオンライン用プログラムで2260と同じオペレーションがなされるようにJEMシステムが設計されている。

ウ．プリンタ・システムとして

JEMシステムのプリンタは特殊な銀塩安定化紙を用いているので，コンピュータのラインプリンタ出力のようにフォーム印刷をほどこすことはできないが，そのプリント品質は鮮明で，ゴシック体や文字の大小，行間隔の編集等，フォーマットが編集できるので，各種の経営情報資料のプリント出力や，原稿との照合校正プリントに用いることにした。

この場合，プリント入力磁気テープは，IBMシステムで編集した磁気テープを使用するようにした。

（注4）伊達　巌："JEM－3100／4100システム"
印刷雑誌Vol. 55, No. 3, PP57～62 (1972)

（注5）早瀬　敏夫："JEM－3100漢字情報システム"
情報化設計Vol. 2, No. 8, PP39～44 (1972)

（注6）海住喜彦他："FACOM6502A漢字ラインプリンタ"
FUJITSU, Vol. 21, No. 7, PP145～168 (1970)

（注7）JEMシステム標準の漢字コードシステムで，Expanded Edited Kanji Codeの略で，漢字の編集に適した内部コードで，2バイト（16ビット）で漢字1字を表わし，20,000字種まで表現できる。

（注8）JEMシステム標準の入力漢字コードシステムで，Expanded Original Kanji Code の略で，漢字鍵盤のハードウェアに適した入力コードで，2バイト（16ビット）で漢字1字を表わし，3,600種まで表現できる。

6.6.3 システムの概要

当社のシステムはJEMシステムとFACOM漢字プリンタシステムとからなり，それぞれの構成，主要性能は下記による。

(1) JEMシステム

ア．構　　成（図6.6.4参照）

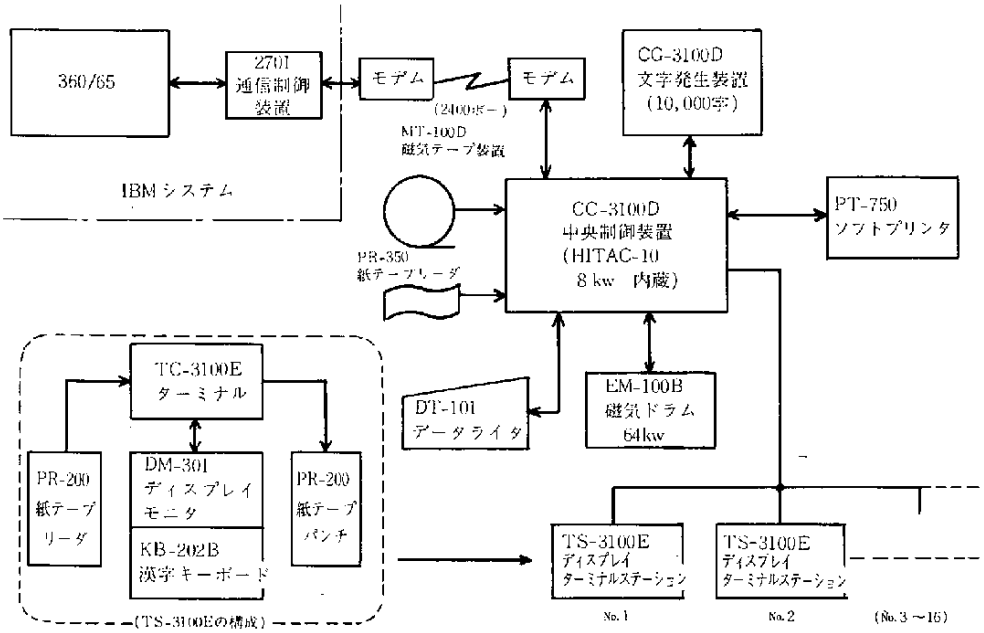


図 6•6•4 JEM-3100S システム構成

(2) FACOM漢字プリンタシステム

ア．構　　成（図6.6.5参照）

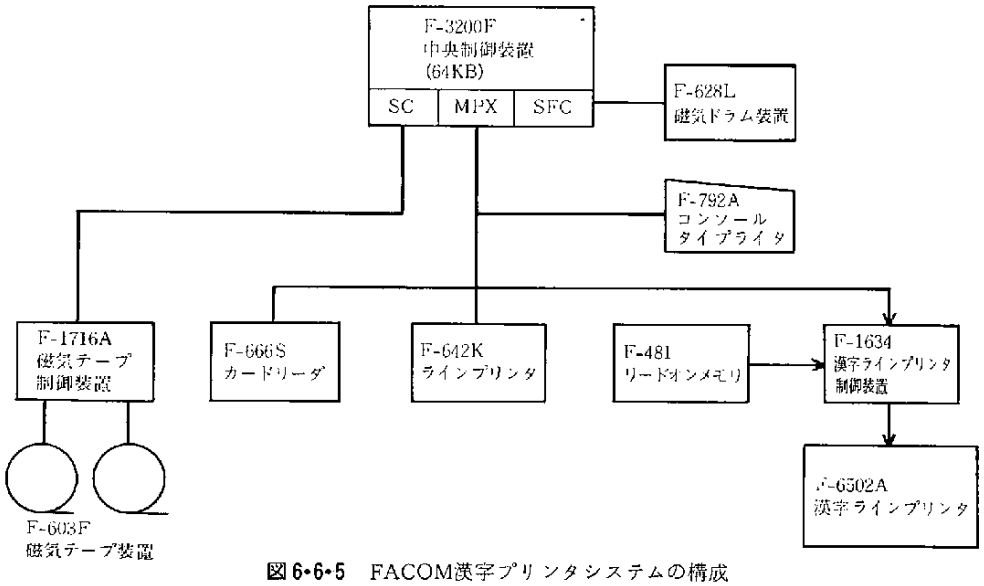


図 6•6•5 FACOM漢字プリンタシステムの構成

（注９）長谷川実郎："高速漢字プリンタ"情報処理，Vol. 10，No. 5，PP279～284（1969）

6.6.4 漢字入力装置

システムへの原始データを入力する装置としては，システムとはオフラインのKB－202漢字鍵盤，あるいは，JEM－3100システムのディスプレイ・ターミナル・ステーションの前記漢字鍵盤を組み込んだ漢字ビデオ表示装置付漢字鍵盤が用いられる。

写真6・6・1　漢字ビデオ照会装置付テレタイプライタ

この漢字鍵盤は谷村新興製作所製のものをJEMシステム用に改造したもので，前者はオフラインのさん孔入力用に，後者はビデオディスプレイによる表示検定を行ないながらのさん孔入力用に用いられる。

この漢字鍵盤は全体で192キーあり，１キーに14字が配列され，全体で2,688字を入力することができるが，その内訳は下記のようになっている。

漢字	1,963	両がな	96
部首	144	英・欧・ギリシャ	100
パターン	48	数字	10
ファンクション	56		
氏名	50		
約物	221		

この入力漢字鍵盤は，JEMシステム特有のパターン入力方式を採用しており，鍵盤上に直接ある漢字1,963字と合成パターン素子48字と部首144字とによって，盤外文字約7,000字を容易に入力することができるようになっている。

パターン入力方式の入力方法は，例を図6.6.6に示したが，「湘」の字を入力する場合，図のように合成パターンキーをまず打鍵し，続いて「氵」，「木」，「目」の順にキーインすることにより，その四つのコードが紙テープにさん孔され，そのさん孔紙テープを入力したときJEMシステム内の磁気

ドラム内の漢字辞書プログラムにより，「湘」のコードが索引されるようになっている。この方法によれば，読めない外字に対しても容易に入力オペレータがパターンを分解することによって，全体の速度を落とさず，さん孔入力することができる。

なお，当社の場合，氏名の入力が頻度的に多いので，前述のように頻度の高い姓50種をあげ，文字キーの1文字と置換し，これもJEMシステム内の磁気ドラムにテーブルを設け，入力は1タッチで行なえるように考慮した。

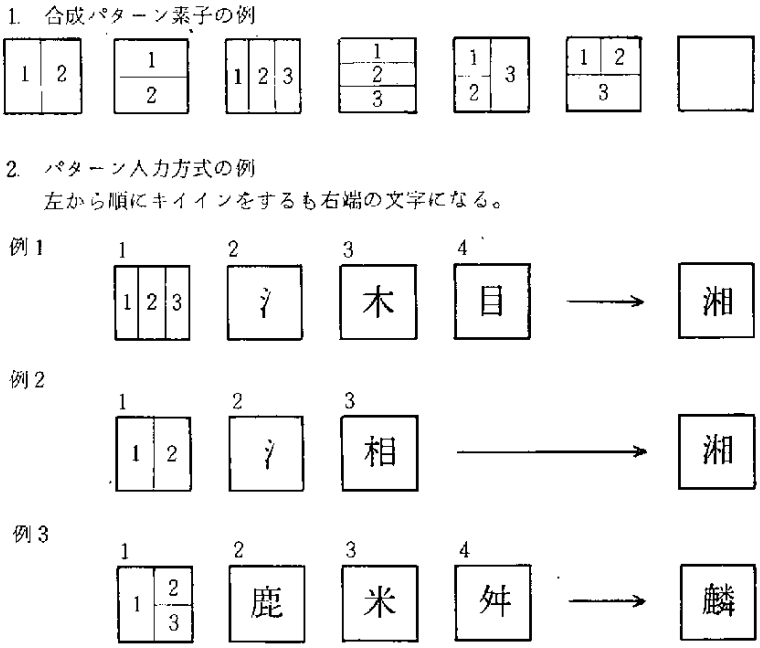


図6・6・6　パターン入力方式

6.6.5　漢字出力装置

(1)　FACOM漢字プリンタ・システム

前述のように本システムは対外用のプリント出力に用いられ，次のような特徴を有している。

- プリント文字は15×18のドット方式で，小型文字の英数字・カナに対しては7×9のドット文字が用いられるほか，OCR文字もプリントできる。
- 印刷速度が600行／分，（40,800字／分）と高速である。
- 用紙はJEMシステムと異なり，特殊な静電記録紙を用いているが，あらかじめフォーム印刷ができるほか，2枚までの複写がとれる。
- 用紙幅が17インチと広くとれる。
- システムの内部標準漢字コードEXEKとFACOMプリンタコードとの変換は，230／25の中央制御装置のコアメモリ内でテーブル変換される。

(2)　JEMプリンタ・システム

前述のように本システムは，社内の経営情報資料のプリントに用いられ，次のような特徴を有している。

- プリント品質は，解像度8本／mm程度で軽印刷用として用いられる鮮明さである。
- 給紙はスプロケットなしの100mロール紙で，1分間にA4版サイズで5枚の速度でカット・アウトされて出力される。
- 書体，文字の大きさ，行間隔，字間隔が可変で，適当なフォーマット編集ができる。

(3) JEMディスプレイ・システム

本システムはディスプレイ検定をしながらの入力用，および契約照会応答出力用に用いられ，次のような特徴を有している。

- IBMシステムの2260ディスプレイ・コントロール用のソフトウェアが，そのまま使用できるようにJEMシステムのソフトウェアが組みこまれている。
- JEMの中央制御装置（8KW）で，最高32台までのディスプレイ・ステーションを同時処理する能力をもっている。
- ディスプレイCRTがストレージ方式なので，ちらつきがまったくない。
- 表示文字の大きさを可変にすることができる。
- 表示情報の挿入，削除，訂正等の校正編集をすることができる。

6.6.6 運営状況およびシステムの評価

当社では，とりあえず普通保険金契約について，証券番号と顧客氏名を結びつけた漢字ファイルの完成が当面の課題であり，まず顧客台帳から証券番号と被保険者姓名（生年月日），契約者姓名（生年月日）を紙テープにせん孔する作業から始めている。せん孔された紙テープは，JEMのソフトプリンタで校正プリントし，修正ランの過程を経て完全なデータとなってから，IBMの契約管理マスターファイルにEXEKコードのまま収納し，徐々に漢字氏名ファイルの蓄積がなされている。

昭和47年5月には対外用として初の試みで，集団月掛契約の漢字氏名ファイルを使って，同契約の配当金支払明細書約60万件分を漢字プリントし好評を得ている。従来の方法であれば，普通のラインプリンタで数字部分を印刷しておき，そのうえで契約台帳から60万件について，1件ごとに顧客氏名を手書きで転記するという膨大な事務量があったわけで，これらが漢字プリンタにより省力化できたことは，大きなメリットとしてあげることができる。

システム全体としての評価は，システムが現在なおビルド・アップ中であるので，これからの問題であるといえる。

6.7　電算写植センターにおける文字組版の事例

6.7.1　漢字情報処理における文字組版の特徴

漢字情報処理として，一般印刷用文字組版のアプリケーションを考えたとき，そこには一般の情報処理とは異なるいくつかの特徴が見出される。

第一に，昭和21年より漢字制度が施行されはしたが，一般書籍関係では，相変らず一万数千字におよぶ文字種が使用されており，さらにシステム収容外の特殊な文字については，いわゆる外字処理が要求され，入力される文字種がきわめて多いということである。

第二に，一般の事務処理では，入力データは固定長で，出力形式も固定のものが多いが，文字組版では，入力データは可変長で，その出力形式は多彩をきわめている。

文字組版における出力制御としての字詰・割付などは，外部指示によることが多いが，文字の字体（明朝，ゴシック，アンチックなど）や種類（ルビ文字，上つき下つきなどの添字），文字サイズ，位置関係を制御するファンクションコードは，原稿入力時に，文字と関連させてパンチする内部指示による。

そのために，ファンクションコードは，一般事務処理に比べて，はるかに多く必要となり，文字種が多いことと相まって複雑な入力データの作成を要求されることである。

写真6.7.1に示す FACOM 6801A 漢字入力キーボードでは，総収容文字数は 10,752 文字で，ファンクションとして最大114の機能がもうけられている。

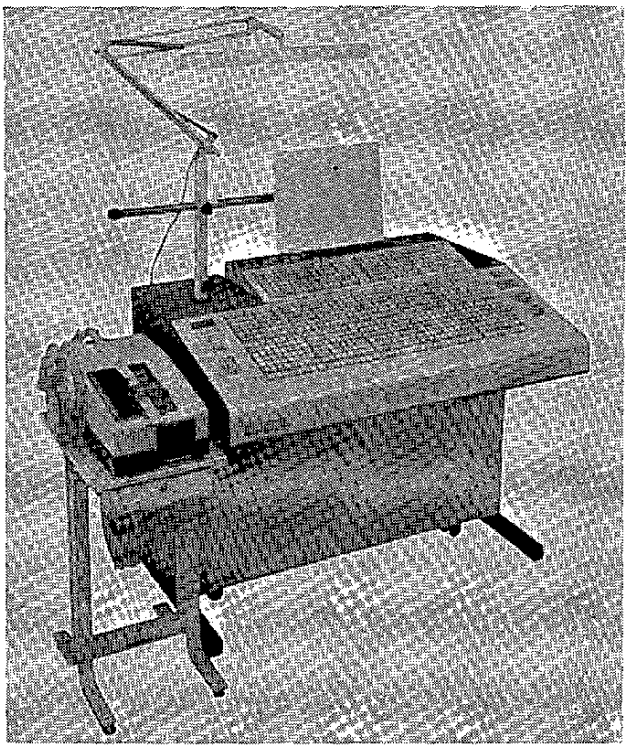


写真6．7．1　FACOM 6801A 漢字入力キーボード

第三に，以上の事柄から，入力に際しての配慮は，一般事務処理よりも多く要求される。なぜならば誤入力の訂正は単なる文字情報の訂正以外に，出力制御情報の訂正がからんでくるからである。したがって訂正処理は重要な処理工程となり，システムを効率よく運営するためのキーポイントとなっている。また従来の活字組版では，組版の仕上がり経過を確認しつつ作業が進行させられたが，電算写植では，この処理過程が機械内部のものとして取扱われることである。

第四に，単に組版ができればよいというわけではなく，それが製版され，印刷されてはじめて評価されるわけで，製版工程が容易になるとともに，印刷物として高い印字品質を有していなければならないことである。印字品質としては，漢字パターンの忠実な再現と共に，字並び精度，印字濃度の均一さなどの要素がからみあっている。

以上のように電算写植システムは，原稿の入力から始まり，印刷までの一連の工程を含む巾の広く，底の深いシステムの基礎をなすものであり，新しいハードウェアとソフトウェアがつぎつぎと開発され，漢字情報処理システムとして，新しい市場を創造しつつある。

印刷の需要は，年々増加しており，反面労働集約型の文字組版関係では，熟練を要する組版工は不足し，かつ鉛活字を使用することによる鉛害，熱気，騒音などをともなう作業環境の改善のため，今後，印刷業界は活字を用いないいわゆるコールドタイプ組版にすすむであろう。

この文字組版の自動化のため開発された FACOM 全自動写植システムは， FACOM 6801A 漢字入力キーボード， FACOM 270－20/30 編集システム， FACOM 6531 写植システムのハードウェアとソフトウェアとして FACOM Composition Language (FCL－7) よりなり， FCL－7 は，組版機能の拡張，システム運用効率の向上のために，数度にわたるレベルアップが行なわれており，ページアップされたフィルムまたは印画紙に出力され，その印字品質はきわめて高く評価されている。

6.7.2　日本船舶明細書の組版事例について

電算写植センターにおける文字組版の事例として，日本船舶明細書の組版処理経過について紹介する。

(1)　日本船舶明細書について

発行所　　社団法人　日本海運集会所

日本国籍を有する 100 総トン以上の鉄鋼船（ただし 500 総トン未満の第 1 種，第 2 種漁船を除く）の毎年 6 月末日における実態を表 6.7.1 に示す各項目について調査し，集録したもので，毎年 1 回発行される約 2,000 頁の出版物である。

わが国は，実質的には世界第一位の船腹を保有する海運国であり，この日本船舶明細書は，船舶の総合ガイドブックとして貴重な出版物であり， 1973 年度版では約 9,000 隻のデータが，総トン数順に配列されている。

組版形態は，1 隻分データが見開き頁にわたる複雑な表である。

(2)　電算写植化された理由，目的

この日本船舶明細書の出版が電算写植によって行なわれるようになった理由，目的の第一は，この本文が，総トン数順に配列されており，新造，売買，修繕などで，毎年配列が変わり，しかも 1 隻分

のデータが見開き頁にわたる表のため，置版（いったん組版した版を保管しておき，再版のとき使う。）されず，毎年新しく組版され，全データについて校正が行なわれていたこと。しかし実際の訂正は全データの10％程度であるので，組版を電算写植化することによって，部分訂正を行ない，訂正のあったところだけを出力して校正する方式をとり，校正作業の省力化を計ることである。

第二は，部分訂正することによって，全データを新組する場合より，よりデータの正確性を保てることである。

第三は，総トン数順の配列は，電算機にて処理することにより，従来行なわれていた原稿の並べかえ作業をはぶき，原稿をランダムに入力することで，組版作業を短縮し，発行期日を早め，新しいデータをより早くユーザに提供することにある。

第四は，組版処理の結果得られる船舶マスターファイルより，船腹統計処理を行なうことである。

第五は，さらに船舶マスターファイルの利用による各種の情報処理サービスが考えられたことである。

第六は，目次，索引（船主所有船表）を自動的に作成することである。

以上のいろいろの目的と理由のもとに，(社)日本海運集会所は1971年4月に，本出版物を電算写植化することに決定した。

(3) 電算写植処理の経過

発行所（社)日本海運集会所の電算写植化の決定により，入力および印刷を担当した㈱天理時報社組版処理を担当した㈱日本コンプ（日本活字㈱が協力），ソフトウェアの作成を担当したペティ㈱（当時は東洋インキ製造㈱ PETY 事業部）の各社の協力により，電算写植によって最初に組版されたのが1972年度版日本船舶明細書である。

翌，昭和47年に，2回目の組版処理をするにあたり，わが国保有船腹の詳細を，的確，迅速に把握するための統計用資料として，日本船舶明細書のデータを使用することが，運輸省ならびに（社）日本船主協会とで決定された。

そのためデータの正確性を期すべく，明細書記載の船舶について，運輸省の協力を得て，船舶原簿との照合が行なわれた。

かくてよりデータの正しくなった1973年度版では，本文と索引である船主所有船表（表6.7.2）が電算写植にて組版され，組版処理の結果得られた船舶マスターファイルより，わが国で初めての電算機による船腹統計処理が，(社)日本船主協会にて実施され，昨年11月に表6.7.3にその一部を示す日本商船船腹統計の第1号が刊行された。

1974年度版については，現在計画が進行中である。

(4) システム概要

システムは，4つのサブシステムより構成され，図6.7.1にそのフローを示す。

1. インプットデータ収集システム
2. マスターファイル作成システム
3. 写植システム

日本海運集会所編纂

『日本船舶明細書』用船舶明細記入用紙（新造船及び購入船）

No.

	1 英文船名（計画年次）（船種）（前名）船名 船舶番号 船主（英文）	2 甲板数 航行区域 信号符字 船籍港	3 起工 進水 竣工 船級	4 総トン数 純トン数 満載排水量（t）	5 載貨重量（夏期）英トン キロトン 載貨容積（m³）ベール グレーン	6 燃料消費量 航海 碇泊（t） 種類別 燃料吸容量（t）	7 （連力）（kg/m）満載 最高 航続距離（M）	8 寸法（m） 全長 登録長 垂線間長 型幅 型深	9 平均吃水（m） 満載 軽荷	10 タンク フレッシュウォーター（t） ウォーターバラスト（t） カーゴ・ディーゼル（m³） 船底構造	11 旅客 乗員	12 船倉口 ハッチカバーの種類 船倉数	13 （貨）船倉 寸法（m）及び数 （油）タンク別容積（m³）	14 荷役装置 デリック ウインチ 貨油ポンプ型式能力（m³/h）及び数 主貨油連絡管口径（吋）	15 船体製造者 船型 建造番号 次期検査年月 中間 1種 2種 定期	16 主機製造者 主機製造年月 型式種類 連続最大出力	17 無線装置 送信機（W） 受信機 船舶電話	18 1 汽罐型式・圧力（kg/cm²）・数 2 発電機の種類・容量 3 航海機器・特殊装置 4 運航者・運航形態
記載例	Tanaka ㉑ 鑑（吉田丸）田中丸 99423 山田海運 Yamada Kaiun	3 層 遠洋 JXYZ 東京	72.1 72.4 72.8 NK LR	9,203.45 5,274.36 18,340.00	11,495.31 11,676.25 17,604.60 18,975.47 ㉑206〒82 ⑤465	㊀ 28 ㊁ ㊂5.8 ㊃1.2 A 285.1 B C 1523.5 計 1808.6	15.7 18.2 12000	151.25 143.12 140.00 19.40 12.20	8.77 3.15	F.W. 434 W.B.6316 C.D. 1429 全部二重底	11 ㊀13 ㊁35	無 調 6	① 7.53×4.50 ② 12.68×7.00 ③ 10.40×7.00 ④ 8.80×7.00 ⑤ 12.80×7.00 ⑥ 8.80×7.00	電動 デ 6×10 20×2 ウ 3×10 5×2	山里造工 三島型 S No. 1234 中 ① 72年10月 ② 71年10月 定 74年10月	山里重工 1965年9月 S9ZW75/150 D 12000PS	短 1kW 中距 500W 短 16球 長中 8球 全 13球 31-1234	1 鑑機電用鑑 10kg×2 2 AC 450 V D 340 kW×2 3 ㊀回 ㊁回 ㊂回 4 山田汽船定
			年 月 年 月 年 月		L.T. K.T. B G	㊃ ㊄ ㊅ ㊆ A B C 計		全 登 垂 巾 深		F.W. W.B. C.D.	㊀ ㊁			主機駆動・電動・電動油圧・内燃機駆動・汽動	S No. 中 ① 年 月 ② 年 月 定 年 月	19 年 月		

船主名（英文名又はローマ字綴）	郵便番号 住所 電話	購入先	氏名	住所 電話	運航形態	自営
						定期傭船先
共有船の相手方	郵便番号 住所 電話	購入年月				裸傭船先
						運航委託先
外航スクラップ アンド ビルド船 有 無 年度		内航登録船 許可番号				

お願い：ご多用のところ恐縮ながらご記入の上は同封返信用封筒にて　迄に当所に必着するようご返送方お願いします。

社団法人 日本海運集会所 東京 神戸

表6．7．1　日本船舶明細書記載例

船名	総トン数	重量トン数	頁数
千代田丸	1,753.01	983.3	193
津軽丸	1,662.30	1,181.6	194
天草丸	359.20	193	459
瀬戸丸	213.70	133	542
計 4	3,988.21	2,490.9	
日本道路公団 ☎ 03-503-0111			N2750
〒105 東京都港区新橋1-1-13			
第一豊予丸	986.32	—	249
第二あさかぜ丸	617.08	300	302
第二あさぎり丸	607.44	306	303
第二あさしお丸	602.87	304	304
あさかぜ丸	294.97	135.85	498
若鳥丸	263.47	126.64	519
あさしお丸	215.61	121.34	541
厚岸丸	140.12	50	745
計 8	3,727.88	1,343.83	
日本液化ガス輸送(株) ☎ 03-212-8211			N2760
〒100 東京都千代田区丸の内3-12			
第一えるびい丸	1,077.59	653.7	220
日本沿海フェリー(株) ☎ 03-573-1911			N2765
〒105 東京都港区新橋1-10-7 大和銀行新橋ビル			
しれとこ丸	7,875.33	3,264.45	81
えりも丸	7,858.33	3,172	81
計 2	15,733.66	6,436.45	

船名	総トン数	重量トン数	頁数
栄興丸	198.99	—	596
瑞興丸	198.96	—	596
明興丸	192.88	—	657
計 8	1,900.66	—	
日本海重工業(株) ☎ 0764-37-9271			N2855
〒931 富山市西宮71			
日海丸	111.73		783
日本海汽船(株) ☎ 03-571-997			N2860
〒104 東京都中央区銀座5-9-19			
あどりあ丸	96,760.30	183,572	5
大海丸	31,950.52	28,848.00	25
栃木丸	17,430.78	27,156	39
神奈川丸	17,423.47	27,152	39
祥海丸	11,217.56	18,473	50
若葉山丸	10,013.88	15,858	58
春海丸	8,317.20	13,055.71	77
朝海丸	8,310.46	13,025	77
晴海丸	7,938.00	11,568.30	81
大鵬丸	6,356.78	9,384	92
大南鵬丸	6,300.75	9,384	93
計 11	222,019.70	357,476.01	
日本海商(株) ☎ 03-434-7561			N2870
〒105 東京都港区新橋5-14-6			
銀海丸	4,027.66	6,100.05	112

表6．7．2 船主所有船表(索引の一部)

1．船種別構成

船種別	総船腹量				うち内航登録船		
	隻数	総トン	重量トン	構成比	隻数	総トン	重量トン
油送船 ①	1,978	12,527,786	23,081,580	43.7	1,439	691,406	1,288,565
油/乾貨物兼用船 ②~⑤	34	1,881,518	3,246,844	6.1	1	497	1,083
オア・バルクキャリア ⑥~⑬	295	6,068,629	10,331,814	19.5	69	121,653	217,396
木材専用船 ⑭~⑯	211	1,743,156	2,520,283	4.8	16	16,945	32,260
カー・バルクキャリア ⑰~⑱	143	1,097,676	1,530,690	2.9	55	52,115	39,978
その他専用船 ⑲~㉔	640	789,290	1,171,304	2.2	299	200,473	339,259
化学薬品船 ㉕	313	147,137	222,901	0.4	246	100,049	150,571
液化ガス船 ㉖	112	478,081	541,255	1.0	79	55,889	46,647
フルコンテナ船 ㉗	28	642,003	572,837	1.1	2	27,009	26,759
一般貨物船 ㉘	3,396	5,849,682	9,297,793	17.6	2,011	872,741	1,669,443
旅客船 ㉙~㉚	583	570,803	201,580	0.4	17	10,263	3,816
その他特殊船 ㉛	457	95,361	14,440	0.0	72	13,828	586
不明	48	129,138	130,590	0.2	17	5,369	9,973
合計	8,238	32,020,260	52,863,911	100.0	4,323	2,168,237	3,826,336

(注) ①船種別欄の数字は下表基礎分類船種の集合を示すもので，2，3，4，5，6表のいずれも同じ。②構成比は重量トンによる。

表6．7．3 日本商船船腹統計例((社)日本船主協会発行1972年度版より)

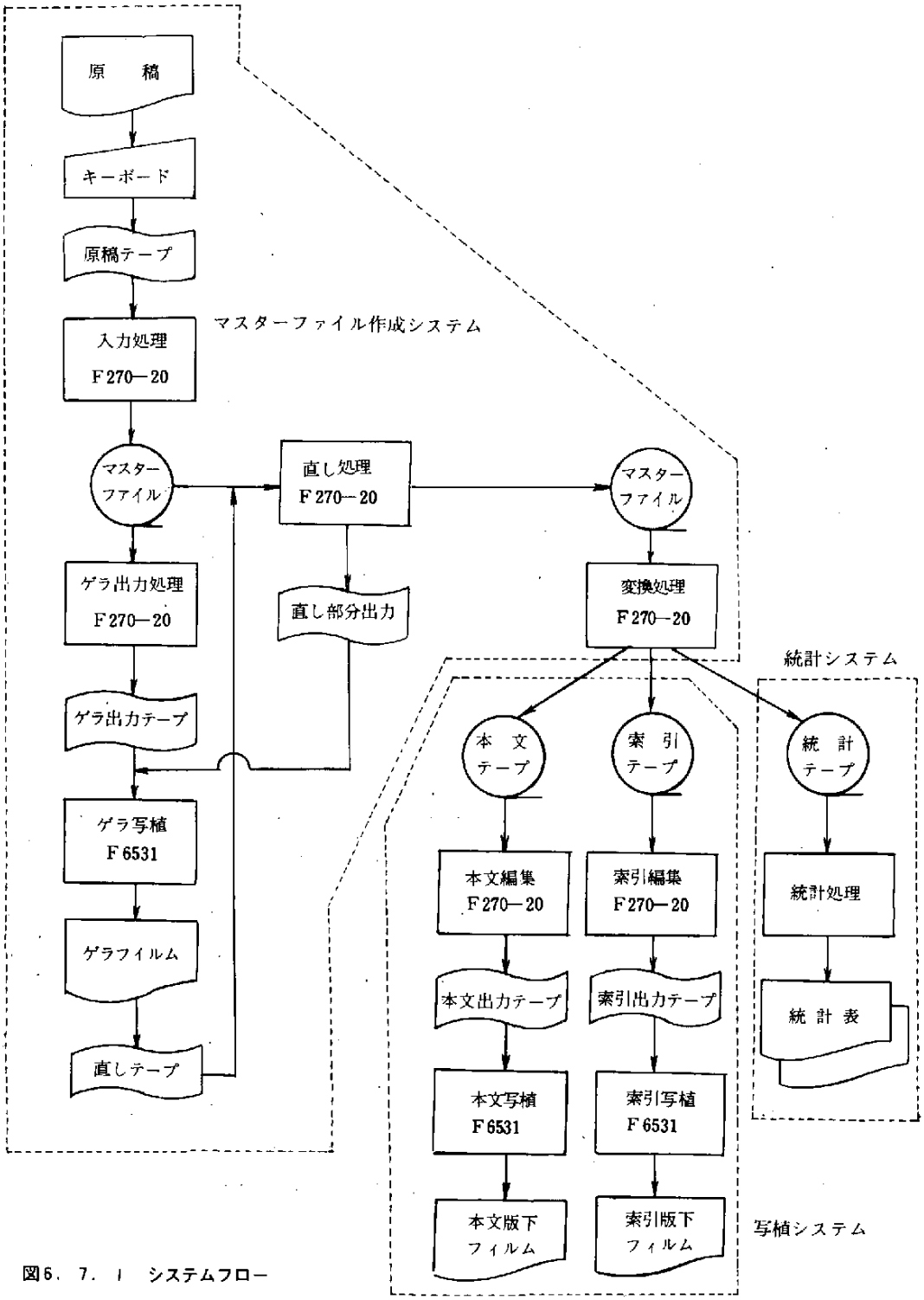


図6．7．1　システムフロー

4. 統計システム

ア. インプットデータ収集システム

システムにとって，良質のデータを短時日に収集することは，最も大切な要件である。本事例では，(社)日本海運集会所において，表6.7.1に示すように，既に集録された船舶については，前年度のデータを台紙に貼りつけ，約6,000社の船主に送付し，補正データの記入返送をもとめる。新造船，購入船については，同じ様式で，記載例をつけたアンケート用紙を送付し，船主よりのデータを入手するとともに，各造船所からもデータを入手して，チェックする方式をとられている。船主の中には一隻船主で，しかも家族ぐるみで船上生活をしている場合があり，データの収集は困難をきわめ，ある時点を限って，全国にちらばっている船主に対し直接電話をかけたり，訪問したりしてデータの収集にあたるという努力がはらわれた。

船舶統計処理にからんで，船舶原簿との照合が行なわれたことは，この点で大きな力となっている。

イ. マスターファイル作成システム

(社)日本海運集会所にて収集されたデータは，入力を担当した㈱天理時報社にて，出力制御用のファンクションコードのチェックが行なわれ，FACOM 6801A 漢字入力キーボードによってパンチされる。パンチされた原稿テープは，㈱日本コンプで入力処理され，船舶マスターファイルが作成される。

ついで校正用ゲラ作成のために，FACOM 270－20 にてゲラ用に出力処理がされ，FACOM 6531 写植システムに，その出力テープをかけることにより，棒組形式のゲラ用フィルムが得られる。

マスターファイルは，図6.7.2に示すように，1隻分のレコードは区切Aのファンクションコードで区別され，レコード識別用に入力順にレコード番号が割当てられる。レコードを構成する18コの大項目，52コの中項目，125コの小項目は，各々のレベルごとに区切B，区切C，区切Dのファンクションコードで区別されている。大項目は，1レコード中必ず18コあるが，中項目，小項目は大項目ごとに異なっている。

入力時には，明朝，ゴシックの書体の切り替え，ポイントサイズの指定，出力位置の指定などは行なわず，すべてプログラムで処理することとし，ファンクションコードとしては，組版制御用に最小限必要なものと，レコードおよび各項目を区切るファンクションコードのみを入力した。

直しは，小項目単位に処理できる。校正用ゲラは，本文の表組形式でなく，棒組出力とし，直し指示のために，レコード識別番号，大項目番号，および入力した各種のファンクションコードは，コード変換をしてすべてゲラに出力し，文字とファンクションを同時にチェックする方式をとった。

この方式では，文字校正は割合に容易であるが，ファンクションコードによる組版制御の規則を習得するのに若干の時間を要した。しかし，本事例のよう，量がまとまっている場合には，規則をよりかんたんにし，出力形式をよく検討すれば，複雑な表組の場合でも，運用可能と考えられる。

校正済ゲラより，一度直し指示のコーディングが，㈱天理時報社で行なわれた。それはやはりファンクションの入力エラー防止のためである。作成された直しテープにより，FACOM 270－20 にて

マスターファイルが修正されるが，直し処理の結果は，直しが行なわれた部分のみを出力して，校正作業の合理化を計った。この部分出力は，効果があり，電算写植のメリットとして，十分その目的を達している。

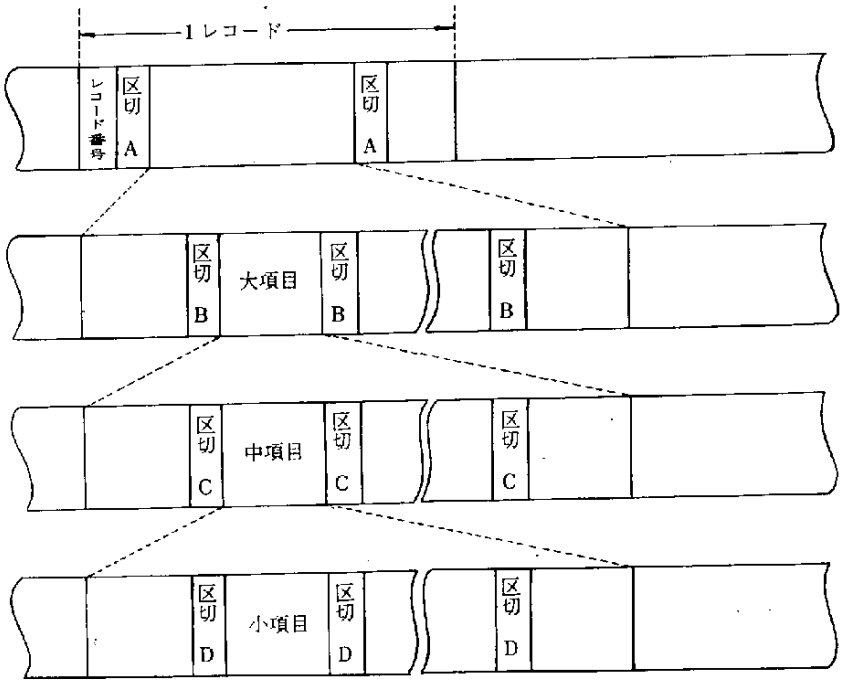


図6．7．2　マスターファイル　レイアウト

ウ．　写植システム

作成されたマスターファイルは，本文出力用には総トン数順に分類され，組版処理にまわされる。組版処理は，レコードが見開き頁にわたり，しかもかなり複雑ではあるが，同一形式であるので専用プログラムで処理された。

専用プログラム開発の段階で，表組としての標準化が（社）日本海運集会所の協力によってなされ，処理の統一化が行なわれた。標準化されなくては，専用とはいえプログラム化は困難であり，この標準化が棒組出力ゲラによる校正処理とともに，電算写植化を成功させた要因の一つとなっている。

索引出力用には，マスターファイルより船主所有船表作成に必要な項目を抽出し，分類し，これも専用プログラムにて処理され，表6.7.2に示す船主所有船表が作成された。

電算写植による索引の自動作成は，原稿の作成が不要となるばかりでなく，内容も正しく，電算写植の効果的利用方法である。

写植出力は，FACOM 6531 写植システムにより逆ポジの印字フィルムが出力され，罫線は全頁同一形式のためマスキングすることとし，オフセット製版時に処理された。

エ．　統計システム

統計用として必要な項目が，マスターファイルより抽出され，コード変換処理されたファイルより，それぞれ隻数，総トン数，重量トン数などの統計処理が（社）日本船主協会にて行なわれた。

総計処理を行なうにあたって，運輸省の船舶原簿との照合が行なわれたが，さらに統計用として必要なデータを付加する作業が，一般の直し処理と同時に行なわれた。そのためマスターファイルには，本文出力以外の統計用だけの項目が存在しており，ゲラ出力時にはこれらもすべて出力して直しが行なえるシステムとなっている。

船腹統計としては，11種類の統計表が作成され，今後も必要に応じて，新しい統計表が作成されることになろう。

(5) システムの評価

本事例を成功させるに最も大きな力となったのは，発行所の（社）日本海運集会所をはじめ，関係各社の絶大な協力が得られたことであり，電算写植システムの実用化に対し，一つの方向が示されたものと考えられる。

システムの開発は，専用組版プログラムの開発に加えて，システムの発展，運用性の改良などにより，当初の計画を上まわる開発工数を要したが，ソフトウェアの開発により，電算写植化の初期目的は，逐次はたされてきた。

組版制御用のファンクションコードの入力エラーの防止のため，直し指示では，コーディングする方式をとったが，直接入力する方法の検討を要する。

マスターファイルとしては，ほぼ完成されており，船腹統計以外の多角的利用の検討段階にあり，新しい展開が期待される。

(6) むすび

日本船舶明細書の処理は，電算写植の応用としては，多くの新しい試みがなされ，それが関係各社の一致協力にて処理され，成功に結びついた事例であり，関係各社に対し心から敬意を表する次第である。

しかし，システムはこれで完成したわけではなく，さらにより一層の検討を加え，よりすぐれた漢字情報処理として発展することが期待されている。

6.8 新聞業界における日本語情報処理

6.8.1 沿　　岸

15世紀の中頃ドイツ人グーテンベルグにより，鉛合金の活字を並べて版を作り，紙に印刷する活版印刷の方式が完成されたといわれる。この活版印刷システムは，それ以来約500年を経た今日でも，基本的にはほとんど変わらず，あらゆる文字印刷物の基礎となっている。

このように長い間，活版印刷システムが印刷方式として存続してきた大きな理由の一つとして，人手と活字による組版が，字の差し替え，挿入，版の組みあげ，組みかえなどに対して，非常に融通性に富んでいることがあげられる。この活版印刷のメリットは，時間の制約のなかで最新のニュースを挿入する機動性，多彩なレイアウトなど新聞業界における要求に対しても，大量の人手作業に支えられて，それにこたえてきた。

しかし，今日，社会環境の変化とともに，大量の人手作業と鉛を主体とした活版印刷システムは問題となりはじめてきた。すなわち，

① 若年労働力の不足により，このような経験とカンに頼る作業における熟練作業者の養成や確保が困難となってきたこと

② 高熱，騒音，重量作業，鉛毒で代表される作業環境の問題

③ 情報化社会に呼応して，年々増大し複雑多様化する情報の迅速な処理と多面利用の要請に対して，活版印刷システムはその限界にあること

など，一言にしていえば，経営近代化に対処できなくなってきていることである。

新聞製作工程の場合，そのある部分は戦後急速に機械化が行なわれてきた。1960年代の初めに確立された，漢字テレタイプとキャスター（自動活字鋳造機）による，本社と本社間，本社と地方支局間のニュースの送受信と自動活字鋳造の一貫したシステムなどはその代表例である。しかし新聞にとって重要な編集あるいは組版の部分は，活字による工芸的性格のため，ほとんど手つかずのままでいるのが現状である。

このような問題を抱えた新聞・印刷業界において，最近CTS（Cold Type System，あるいは，Computerized Typesetting System，以下CTSという）が脚光をあびている。その大きな理由としては，全自動写真植字装置の登場により，これまで大量の人手を要していた，漢字カナまじりの文章の編集・組版などの処理，および商品として使える品質の版の作成を，コンピュータで行なうことが可能となったこと。これに伴い，騒音，高熱，重量作業，汚れなど鉛につきものの欠点を除去し，快適な作業を行なえる職場環境へ整備できること。さらには，コンピュータの導入により，将来，複雑多様化する一方の情報の蓄積・加工・多面利用への可能性が開けてくることなどがあげられる。

このように，多くの問題を抱えて大きく変革しようとしている新聞印刷業界に対して，その解決の手段として開発されたのが，自動編集組版システム（Japanese Publishing System，以下JPSという）である。JPSは，CTSの実現に対話形式をとり入れたトータル・システムであり，編集と製作にコンピュータを本格的に利用し，原稿をコンピュータに入力して，校閲，レイアウトを行ない，写真

植字装置で印刷用フィルムを出力するまでの作業を，オンライン・システムで一貫，同時並行処理するシステムである。

6.8.2 現在の新聞製作工程

図6.8.1は，現在の新聞製作工程の例を示す仕事の流れ図である。以下にその概略を説明する。

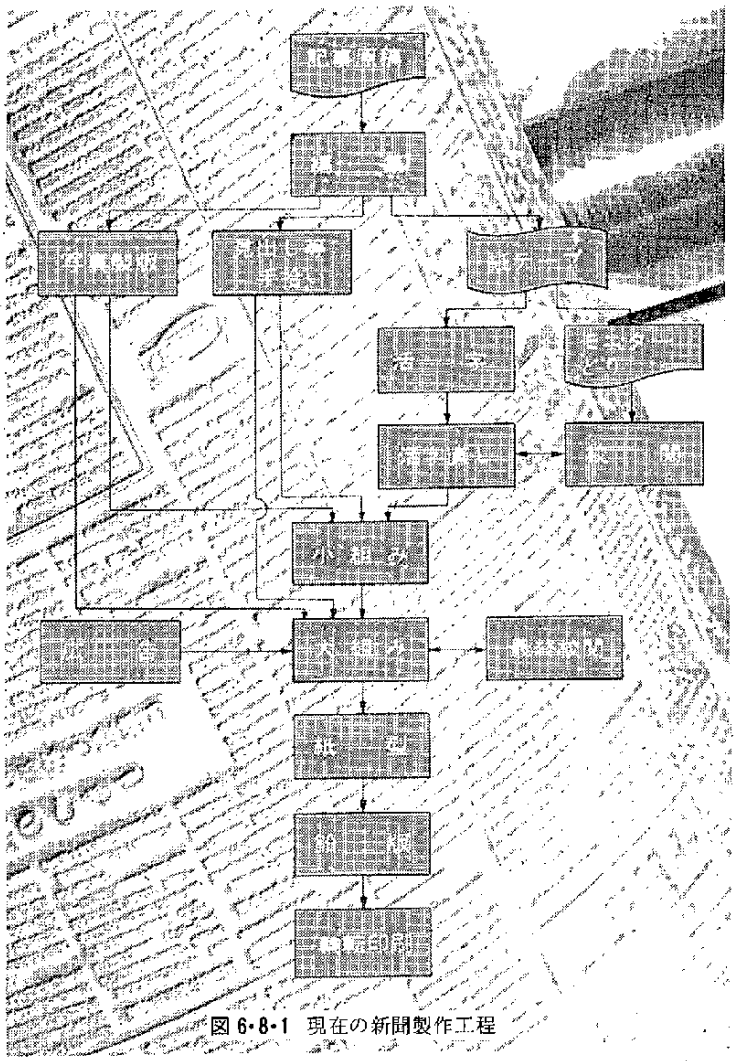


図 6・8・1 現在の新聞製作工程

新聞の紙面に載せられる記事は，そのほとんどが記者による取材か，通信社からの提供によって作られる。これらの記事は，一度出稿部門に集められ，そこで内容の重要性がチェックされ，ある程度取捨選択が行なわれる。

出稿された記事は，漢字テレタイプにより紙テープにせん孔され，必要に応じて他本社にも送られる。一方，支局，支社，他本社で取材された記事が，やはり紙テープに変換されて送られてくる場合もある。このように，記事を紙テープに変換することにより，漢字カナまじりの文章が電気信号の形で処理され，通信回線を介して遠隔地へ記事を送ることや，キャスター（自動活字鋳造機）による活

字の製作，そして記事のプリント（モニタ・プリント）など一連の作業が同時に進められる。モニタ・プリントは，すぐ原稿と読み合わされ（校閲），紙テープせん孔時の誤り，出稿上の誤りがモニタ・プリント上に記入されて，整理・編集部門にまわされる。

整理・編集部門では，ページ内の記事量を考えながら，記事の扱い方を検討し，適切な「見出し」をつけたり，一部を箱形などに小組みするように指示したり，締切時間のギリギリまでレイアウトの検討が行なわれる。

整理・編集は，新聞製作工程の中で，もっとも重要，かつ複雑で手間のかかる作業である。校閲により記事のモニタ・プリント上に訂正がなされる上，編集担当の指示により，すでに作られた活字を新たに組み直したり，またその結果の検証を行なったりする。活字の直しにしても，小組みにしても，鉛でできている活字を配列する作業には，禁則処理と呼ばれる作業上の定石など，多くの規則があり，かなりの熟練を要する。1ページ大の組み（大組）が行なわれる段階で，写真，広告が組み合わされ，さらに，その組み版上にインクがつけられ，数枚のコピー（ゲラ刷り）が最終校閲用として作成される。活字が人手によって動かされる都度，誤字，誤植を防ぐため校閲作業が並行実施される。

1ページに仕上げられた組み版から「紙型」をとり，その紙型に溶解した鉛が流しこまれて半円筒上の輪転印刷用鉛版が作られる。

このようにして刷られたいくつもの鉛版が，新聞社特有の高速輪転印刷機に装着され，新聞が作られていく。

6.8.3 システムの特徴

JPSのシステムのおもな特徴として，次のような点をあげることができる。

- コールド・タイプ・システム
- オンラインによる一貫・同時並行処理システム
- 対話形式のシステム
- 柔軟性のあるシステム

6.8.1沿革で述べたごとく，新聞印刷業界において，鉛活字を用いないですむ，コンピュータと写真植字装置によるCTSの実現は，求人難への対処，職場環境の整備，経営近代化などに対する解決を与えるものである。

オンラインによる一貫・同時並行処理とは，原稿をコンピュータへ入力してから以降，校閲，レイアウトを行ない，写真植字装置へ最終の印刷用フィルムとして出力するまでの一連の作業を，たとえば，工程の中間結果の出力，校閲，訂正データの紙テープへのせん孔，再入力などの手作業をなくして，コンピュータ内部ですべて処理することである。これは，人手の介在により生ずる誤り，作業の繁雑，非能率化をできるだけ少なくすることを可能にする。

この一貫・同時並行処理を実現するためには，コンピュータと人間が互いに対話しながら作業を進めることが必要である。すなわち，新聞を製作するには，大きく分けて，次のような作業工程が考えられる。

・記事自体の価値判断とその取り扱いに関する指示

・誤りのない文章の作成

・紙面の割付け，および定められた規則にしたがって文章を流していくコンポジション

・高品質な原版の作成

これらの作業工程の一つ一つは，本質的には，すべて人間が直接手をくだし，作業進行上の仲介者となっていくものであるが，次のような作業は，人間の指示さえあれば，コンピュータ・システムが処理できる作業である。

・文章というデータ・ストリームに対する更新／追加

・指定された区域に文字，カット等，まちまちな大きさを持った要素を，定められた規則にしたがって配列していくこと

・最終結果を印刷原版上に出力すること

JPSは，これらの作業を処理するソフトウェアと，人間からの指示入力，および各種の情報処理装置を統合して構成されたシステムであり，人間とコンピュータが，それぞれの役割を分担して，対話形式により作業を進めていくシステムを実現している。ことに，編集，組版は，新聞製作のうえで，最も重要かつ，複雑な作業であり，この部分にディスプレイ装置を用いた対話形式により，コンピュータを本格的に使っているのがJPSである。

最後に柔軟性あるシステムについては，新聞製作工程は，大まかには各社同じであるといえるが，実際には，細部において各社それぞれ異なっているのが実情である。そのため，システムとしては，基本的な部分に付加して，それぞれ各社の工程の特徴を盛りこむことのできる柔軟なシステムであることが望ましい。JPSのソフトウェアでは，100以上のパラメータにより，禁則処理の方法と該当文字，命令の形式，レイアウトにおける紙面構成要素の基本的な性質，あるいは，データ・ファイルの構成配置の方法などを，使用者側で指定できるので，非常に柔軟な機能を持っているといえる。さらに，新聞製作システムは当然のことながら，漢字を扱うことを前提としているので，漢字の多様性に起因する，文字の数，外字の処理，字体，文字サイズ，文字のデザイン，文字コードなどの問題に対処できることが必要である。JPSでは，このため，文字をディジタル・データとして保有し，文字の生成，追加，削除，文字コードと文字のディジタル・データとの関係などを，ソフトウェアで制御することにより，上記に対する自由度を提供している。

6.8.4 システムの概要

(1)JPSの主要端末装置

JPSを構成する端末装置のうちおもなものは次のとおりである。

ア.IBM4280映像発生装置／IBM4510映像表示装置

JPSで校閲用，レイアウト用に使用されるのがIBM4510映像表示装置（以下DTVという）で，IBM4280映像発生装置に接続されて働く。4280映像発生装置は，16×16のドット・マトリックス・パターンで表わされた4,096の文字イメージを収容し，また内蔵した2台の固定ヘッド・ディスクに画面を構成するデータを蓄える。なお4280映像発生装置には14台までの4510DTVが接続可能である。

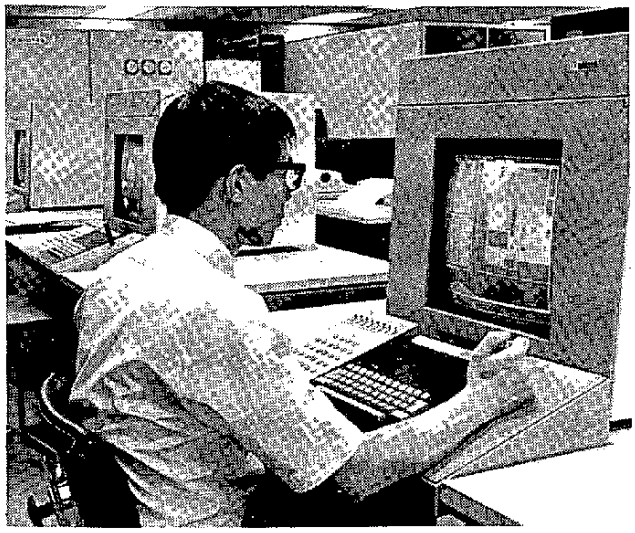


写真 6・8・1　ＩＢＭ―4510映像表示装置

4510DTVは，21インチのCRTを使用し，１画面に2,916文字の表示が可能である。また，英数字キーボード，32キーのプログラム・ファンクション・キーボード，８箇のトグル・スイッチを持ち，さらに漢字キーボードの接続も可能である。DTVには，ライト・ペンのかわりにジョイ・スティクという操作レバーがあり，この操作により，画面上のカーソルをあやつり，位置の指定を行なう。

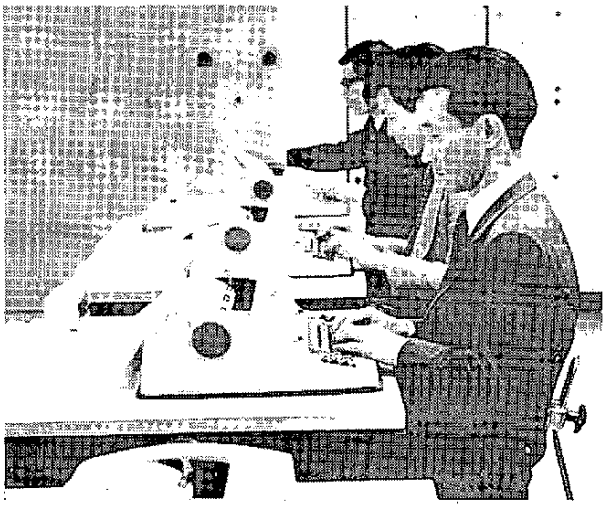


写真 6・8・2　ＩＢＭ―2740タイプライタ通信装置２型

イ．IBM2740タイプライター通信装置－２型

JPSで4510DTVと同様にレイアウト用として使用される。

ウ．IBM2680写真植字装置

最終出力をフィルムまたは感光紙上に印字するための装置である。2680で使用される漢字・かな文字は，ディジタル・データとして，フォント・ファイルあるいはコアに蓄えられている。これらの文

字イメージは，必要の都度CRT上に組みあげられ，レンズを通してフィルム上に印字される。

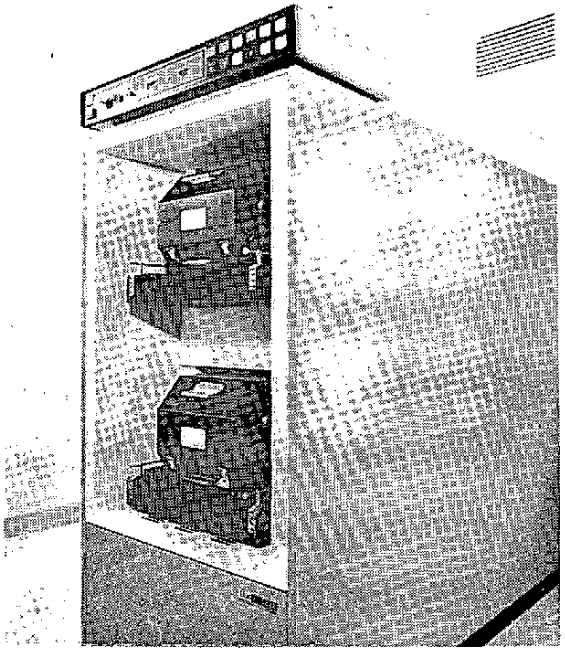


写真 6-8-3　IBM—2680写真植字装置

2680写真植字装置の機能は次のとおりである。

- 印字スピード：　英文8ポ　　2.250字／秒
- 印字文字の大きさ：　1回の出力で4～18ポ，2回以上の分割出力で1インチ以上も可能
- 印字文字の形態：　自由
- 収容文字数：　外部記憶装置の容量による。
- 印字方法：　CRT上に表示したものを行印字
- 使用感材：　フィルム・印画紙（ロール状）で9.44インチ幅
- 印字密度：　800スキャン・ライン／インチ

図6.8.2はJPSのハードウェア・システムの構成例を示すものである。

(2)　JPSのソフトウェア

JPSのソフトウェアは大きくわけて，オンラインで使用するプログラム（JPS／On-Line System，以下JPS／OLSという）とオフラインで使用するプログラムがある。JPS／OLSはDOS（Disk Operating System）のもとで働き，約750のプログラム・モジュール，約10万ステップからなっている。最も大きな特徴は，システムの特徴のところで述べたとおり，汎用性のあるシステムにしてあることであるが，このほかの特徴としてパターンとユーザ・プログラム・インターフェースが挙げられる。

パターンとは，ある程度決まった形を持つ箱物・見出しなどに対して，単独機能を持つレイアウト・コマンドを用いて，レイアウトを定義するコマンド群を作り，これをあらかじめカードにより入力

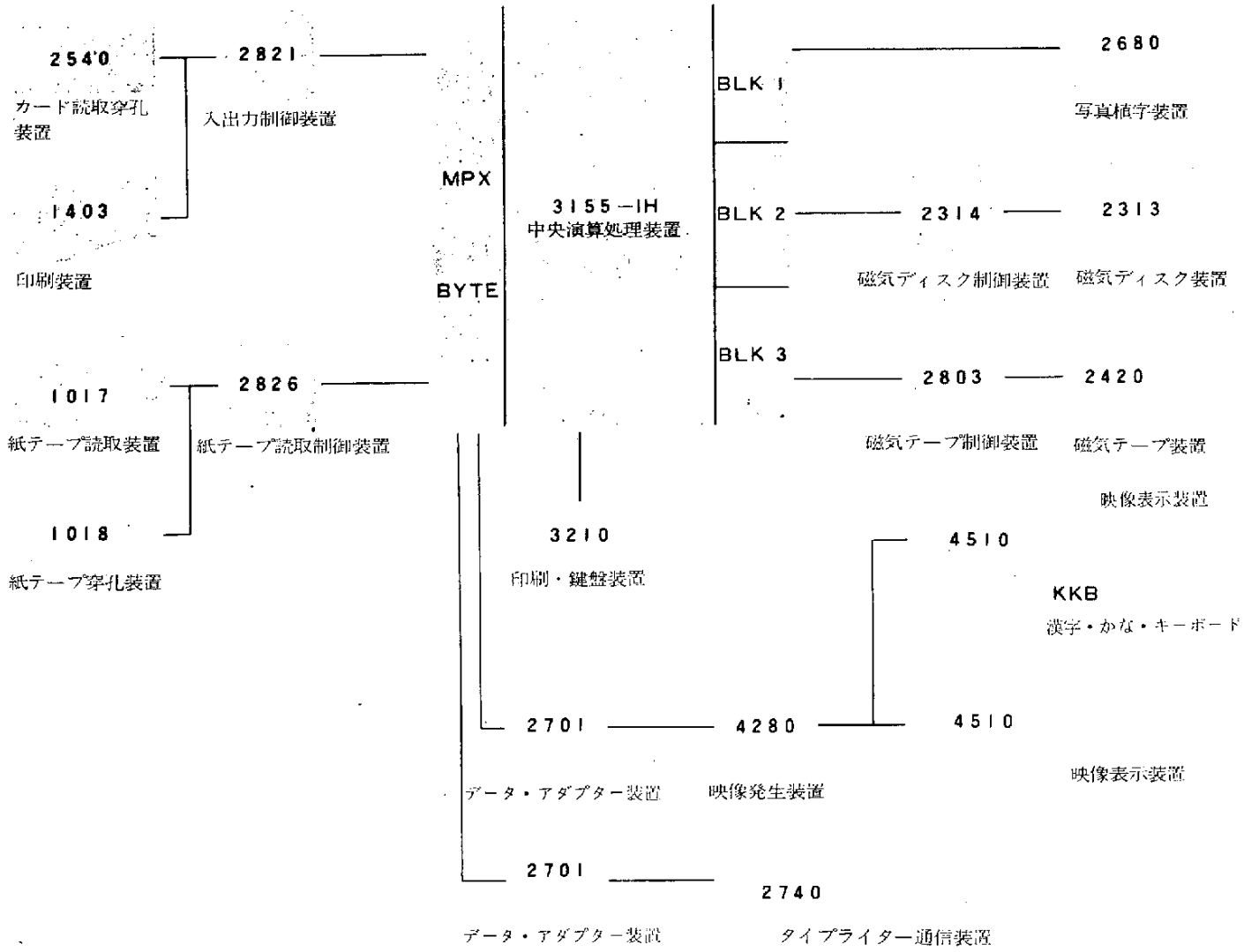


図 6-8-2 ＪＰＳのハードウェア・システム構成例

し，ファイル（システム・パターン・ファイル）に記憶したものをいう。このパターンは，レイアウト時に1箇のコマンドにより呼び出され，パラメータを与えることにより，箱物・見出しなどを一挙に作りあげる。このパターンの使用により，コマンド入力ミスを防ぎ，作業を簡略化することが可能である。

ユーザ・プログラム・インターフェースは，ユーザにより異なる記事の入出力などを自由に処理するため，ユーザ・プログラムをJPS／OLSの中に組み込み，JPSの他のプログラムとまったく同じように働かせることを可能にする。

JPS／OLSは次のようなサブシステムから構成される。

- ターミナル・サブシステム：4510DTV，2740ターミナルからのコマンドの処理，ディスプレイ画面の制御，校閲オペレーションの処理を行なう。
- インプット／アウトプット・サブシステム：記事テープの処理を行なう。
- コンポジション／レイアウト・サブシステム：紙面のレイアウトの処理を行なう。
- フォートプリンタ・サブシステム：2680写真植字装置による印字の処理を行なう。
- そのほかにキュー・サービス，データ・サービス・ユーティリティなどのプログラムがある。

これらのサブシステムは，JPS／OLSのSCP（システム・コントロール・プログラム）の制御のもとに働く。

JPSのオフライン・プログラムとしては，2680写真植字装置のフォント作成・維持のフォント・ジェネレーション・プログラム，JPS／OLSのシステム生成サポート・プログラム，その他がある。

6.8.5 JPSによる新聞製作工程

図6.8.3はJPSによる新聞製作工程を示す流れ図である。

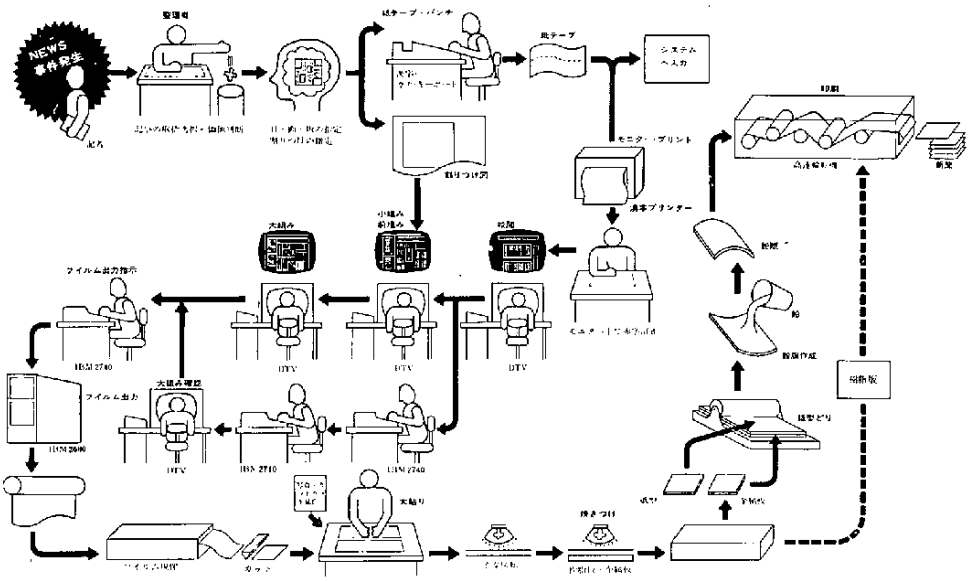


図6・8・3　ＪＰＳによる新聞製作工程

(1)データの入力

各取材部門から出された記事原稿は，取捨選択，日，面，版の指示が行なわれたあと，紙テープにせん孔される。紙テープにせん孔されるデータは，記事本文のみでなく，組み方の指示，文字の大きさ・形の指示等も含まれる。また組み方指示データのみを入力することもでき，これはとくにレイアウト・パターン（JPSのソフトウェアの項を参照）の入力として有効である。

せん孔された紙テープは，IBM1017紙テープ読取装置によりシステムに読みこまれ，外部記憶装置に記憶される。

(2)校　　閲

校閲オペレーションは，4510DTVを使って行なう。各DTVは校閲用とレイアウト用に使われ，どちらのモードで作業を始めるかは，オペレータが作業開始時にモードの指定をすることにより決められる。

DTVを校閲モードに指定し，入力時に記事本文に与えた識別番号をDTVから入力すると，漢字かなまじりの記事と必要情報が画面に表示される。オペレータは，校閲から出された訂正指示書にもとづき，画面上の記事を読んで，DTVの各種キーボードを利用し，文字の挿入，削除，入れ換え，移動などの訂正のためのデータ入力を行なう。訂正はまずジョイ・スティックで訂正箇所を指し，プログラム・ファンクション・キーにより命令を与える。この命令（コマンド）は，JPSのコマンド解析プログラムにより解析され，コマンド実行の該当プログラムにコントロールが渡され，そのプログラムにより，呼び出されている記事本文が訂正される。

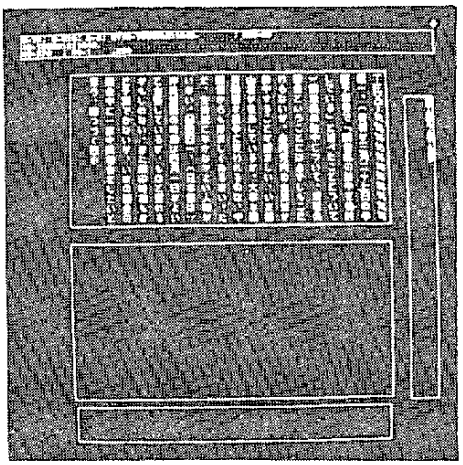


写真6・8・4　DTV画面（校閲）

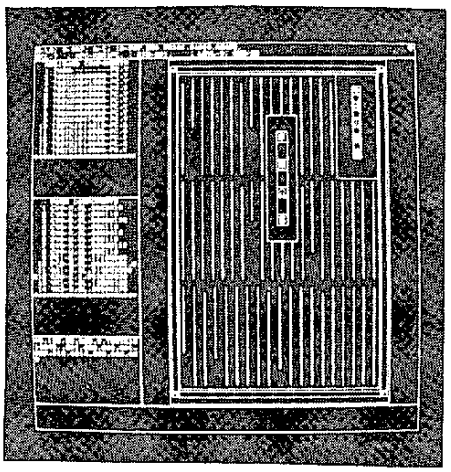


写真6・8・5　DTV画面（小組み）

(3)レ イ ア ウ ト

記事データが校閲（訂正）されたあと，1ページの紙面を構成するための，レイアウト・オペレーションを行なう。まずDTVまたは2740ターミナルをレイアウト・モードに指定して，あらかじめ整理の時点で作成された割り付け図に従ってレイアウトを行なう。レイアウトの方法には2種類あり，1つは単独機能を持つコマンドを入力し，1ステップごとにレイアウトしていく方法である。他の方

法は，あらかじめファイルに記憶されているレイアウト・パターンを呼び出して，必要なパラメータを与えて一挙にレイアウトする方法である。後者の方法は，形の決まっている見出しや，箱物に対しては有効で，オペレーションの簡略化がはかれる。

すでに校閲ずみの記事は，レイアウト時に作られる指定された区域に流される。このときに，禁則処理，ルビ処理，倍数変換処理等の操作が同時に行なわれる。オペレータは，その結果がうまくいったかどうかを確認し，もし記事が枠内で余ったとき，または，はみ出したときには，記事データを表示して記事の追加，削除をするか，レイアウトをやりなおして枠の拡大，縮小をはかる。小組みの結果は，いったん外部記憶装置（システム・コンポージング・ファイル）に記憶され，以後識別番号を使うことにより，必要なときに呼び出すことができる。新聞1ページ大のレイアウト（大組み）時には，小組みの結果が呼び出され，割り付け図に従って仕置づけが行なわれる。写真，カット，見出しなどは，レイアウト時にそのスペースの割り付けが行なわれる。

これらの作業は，人間が大組み台上で目で確認しながら並べていく作業と基本的には同じであるが，人間の場合に発生しやすいミス（活字の脱落やひっくり返りなど）が防げるという大きな利点がある。

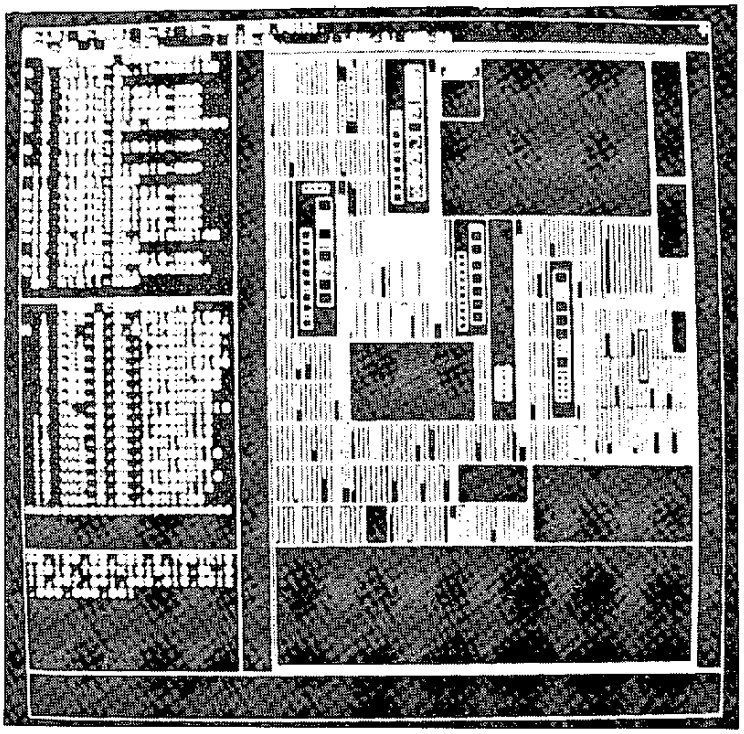


写真 6-8-6　ＤＴＶ画面（大組み）

(4)フィルム出力

大組が終わると，オペレータによる出力の指示により，該当レイアウト面（1ページ組み，小組み）が呼び出され，2680写真植字装置で出力する順序に並べかえられ，フォント・ファイルから文字イメージを一つ一つ取り出しながら，ロール状のフィルムに印字される。カットや写真の部分は余白のまま残される。印字は，フィルム上に新聞1ページの1/3，すなわち5段分ずつ行なわれるので，現像処理後，1ページ大に貼り合わせ，カット，写真などの貼り込みをし，ネガに反転したうえ原版を作る。それ以降印刷のための必要な処理が行なわれる。

ここに掲載された写真は，朝日新聞社ならびに日本経済新聞社の好意により提供されたものである。

6.9 日本語記録システム

6.9.1 日本語記録システムの必要性

現代は情報化時代であるといわれ，情報の生成，記録，伝達，変換，検索，複製などの技術がめざましい発達をとげつつある。産業に第一次産業，第二次産業，第三次産業の区別があるように，情報処理にも原材料となる情報の生成，記録，伝達を中心とする第一次情報と，電子計算機などを用いて加工した第二次情報およびサービスを主目的とする検索などの第三次情報が考えられる。現在は電子計算機などの強力な道具があるために第二次情報システムが多く開発されつつある。しかし豊富で正確な第一次情報があって，はじめて第二次情報システムが力を発揮することを理解する人は少ない。

日本語記録システムは第一次情報の処理を中心的な目的にしている。人が考えたり，発言したりした原始情報を，正確に速く記録，伝達し，第二次情報システムで加工が容易な形に整理することが必要である。欧米では第一次情報処理はタイプライタの普及によってほぼ解決され，作家は考えると同時にタイプで清書し，口述をタイプすることさえ行なわれている。テレタイプライタを用いれば打けんと同時に遠隔地で記録をとることも，また電信符号の紙テープを作成し，第二次情報処理の入力データにすることも可能である。ここで重要なことは，欧文タイプライタは正確に速く入力し，記録出力をとるすべての機能をそなえている点である。

一方，日本では第一次情報処理は欧米に比して非常に遅れている。これは主に漢字という字種の多い国字を用いていることに起因して，欧文タイプライタまたはテレタイプライタのように第一次情報の生成，記録，伝達に便利な装置が発達しなかったためである。

日本語記録システムは第一次情報処理の中で①議会，裁判所，その他の会議における発言の正確な記録を早く作成すること，②電子計算機による日本語情報の第二次的加工処理に適した正確なデータを大量に供給すること，③伝送回線網を介して日本語情報を安価に配布し，記録再生することなどを応用目的としている。

①の応用について従来行なわれている機械化としては，録音機械，あるいは速記機械の採用があり，最も古くは昭和25年頃から始まった。しかし，これらの機械は発言者のことばを正確にとらえてはいるが，日本語情報として完全に記録され，他の情報に容易に変換できる形にはなっていない。従来この部分は人手で多くの時間をかけて行なわれていた。この作業は「反訳」と呼ばれている。1時間の録音，または速記に対し反訳の作業は最低8時間を必要とし，会議録ができあがるまでに通常14日を要している。日本語記録システムの開発が考えられたのは，これらの問題点を解決し，完全な記録となった日本語情報を，大量に迅速に得るシステムを確立することが強く望まれたからであった。本システムには，

i　反訳の作業時間を従来の半分以下にすること。

ii　反訳と同時に形式のととのえられた記録が得られること。

iii　また同時に日本語情報の第二次的処理に適した符号化された情報も得られること。

iv　発言者の加筆，訂正などの赤字校正が容易に行なえ，完全な日本語情報としての記録が得られ

ること。

v　大きな計算機などは使わずに，コスト性能比のすぐれたシステムであること。

などの仕様を満たすことが望まれた。これらの仕様が満たされれば，②，③の応用事例にも同様に利用することができる。

6.9.2　システムの設計

(1)　字 種 の 選 定

基本文字として2,304文字が選ばれた。2,304の数は後に述べる入力装置が，1文字をあらわすのに必ず2打けんするという原理を用いていることと，48キーの欧文型タイプライタを用いているので，48×48＝2,304となることに基づいている。同時に現在の新聞社の6社協定コード（CO-59）が同数の文字種を内字として採用していることも考慮されている。

2,304文字の内訳は次のとおりである。

漢　　字	1,987字
平 仮 名	78字
片 仮 名	81字
英　　字	56字
数　　字	20字
符　　号	82字

漢字は当用漢字1,850字から補正案で削除が提案されている28字を除外し，県名に使われる文字，人名として使われるもので高頻度に出現する文字やその他の文字で，当用漢字に含まれていない165字を加えた。新たに，加えた文字を表6.9.1に示した。

崎	仙	伊	媛	潟	阜	縄	阿	阪	奈	梨	鎌	栃	杉	斉	巳	柴	鹿	塚	埼	岡	辻	藤	茨	幌
熊	磯	曽	幡	苫	槻	靖	挑	尻	棟	鞍	隅	俸	洞	嵐	朴	嶺	那	脇	碧	荻	猿	睦	洒	琉
梶	股	羅	旭	渦	秦	稼	笹	畿	橘	竜	洲	函	鯛	汁	敦	栗	篠	茅	禎	釣	戻	鶴	祐	機
溝	斐	昌	菅	堺	虎	磨	宏	弘	牟	柏	槍	渥	偵	掘	樋	棚	亀	鷹	桂	洪	錦	酌	之	釜
矯	榎	庵	晃	樽	庄	窪	桐	麿	苑	籠	駒	淀	芦	欣	呂	又	垣	云	蘇	葛	尚	吾	辰	綾
鷲	薩	笠	智	亭	檜	堀	釧	萩	湘	灘	湧	柏	殻	僕	泥	嘉	鯉	彦	臼	哉	筲	晋	也	渓
爺	磐	厄	弥	壌	鍋	銚	楠	須	湊	淵	蒲	披	岬	々										

表 6・9・1　基本文字として採用した当用漢字以外の漢字

平仮名，片仮名には小文字，濁音，半濁音文字も含み，英字は大小文字，数字はアラビア数字とカッコ付数字を含んでいる。

これ以外の文字として

組み合わせ漢字	約10,000字
符　　号	約　500字

機 能 符 号　約　200字

を採用している。組み合わせ漢字とは，基本文字に含まれる1,987字の漢字以外の漢字で，システムの構成をかえることなく，任意の数だけ入力文字として追加することができる。他の漢字入力装置では基本文字以外の漢字を入力する場合には，辞書を見て目的とする文字コードを1字ずつさがす方法が一般的である。この方法では外字をさがすのがめんどうであり，非能率的である。日本語記録システムではオペレータが基本文字さえ知っていれば，他の組み合わせ漢字はほとんど辞書を参考にせずとも迅速に入力できる手段が考案され，この問題は完全に解決されている。その方法については6.9.4で詳述する。

機能符号はイタリック体，ゴシック体などの字体を指定する符号，ギリシャ文字，ロシア文字などの字種を指定する符号，ルビ，クワタ，上ツキ，下ツキその他の編集用の符号などを含んでいる。最高240種まで現在のシステムでは指定できるが，実際には約200種が使用されている。

(2) システム設計のポイント

(1)に述べたようなシステム設計に必要な仕様を具体的に実現するために，次のような点に考慮が払われた。

ア．速　　度

従来のシステムでは，印刷まで入れて最も速くて2日，通常14日かかっていた作業（録音－再生－反訳－漢テレけん盤－印刷，の工程で，録音時間1時間分で約1万字が含まれている作業）を1日に短縮するには，最も時間を要している反訳および漢テレけん盤入力の部分を，まったく新しい方法で抜本的に解決する。

イ．システム・コスト

情報の第二次加工では電子計算機などの高価な装置が利用されているが，第一次情報処理においては，情報当たりの経費を最小限にすることが強く望まれる。そのためコスト性能比のすぐれた機器を選定し，バランスのよいシステムを構成する。

ウ．校正・ベリファイ

原始情報には冗長な情報や，表現が不足しているため，後から挿入する情報などが含まれ，発言者または原著者による赤字加筆訂正をすることが非常に多い。また第二次情報処理システムの原始データ作成に本システムを応用する場合には，データの正確さが強く望まれるので，入力オペレータのミスを最小にするベリファイ機能をシステムに含まねばならない。本システムではベリファイは入力機器で行ない，校正は日本語記録システムで得られた記録に赤字加筆訂正されたものを用い，漢字ディスプレイなどの高価な装置を用いずに校正する方式を開発する。

エ．コ　ー　ド

日本語情報は生成された場所で記録されるばかりでなく，伝送回線にのせて遠隔地で記録再生することも多いので，コードは新聞社で用いられている6ビット・コードを考慮した。しかし別に第2次情報処理では8ビット・コード（JISコード）体系が使用されることも多いので，その互換性があるように十分考慮する必要がある。

オ．編集その他の機能

本システムの当面の目的は日本語情報の第一次処理にあるが，当然，第二次さらに第三次処理も将来考えられるので，入力機器の段階で編集，検索，写真植字などのための種々の機能指定を入力できるように考慮されていることが望ましい。とくにシステムに余分なボタンや機能を追加せずに，このような将来のための拡張性が含まれるように考慮する。

6.9.3 システムの構成

日本語記録システムで，1日50万字を処理するケースについての機器構成では，20台のラインプット入力装置，3台の東芝漢字ファックス装置，3台の東芝電磁誘導型タブレット付入力装置，1台のTOSBAC－10ミニ計算機が必要である。

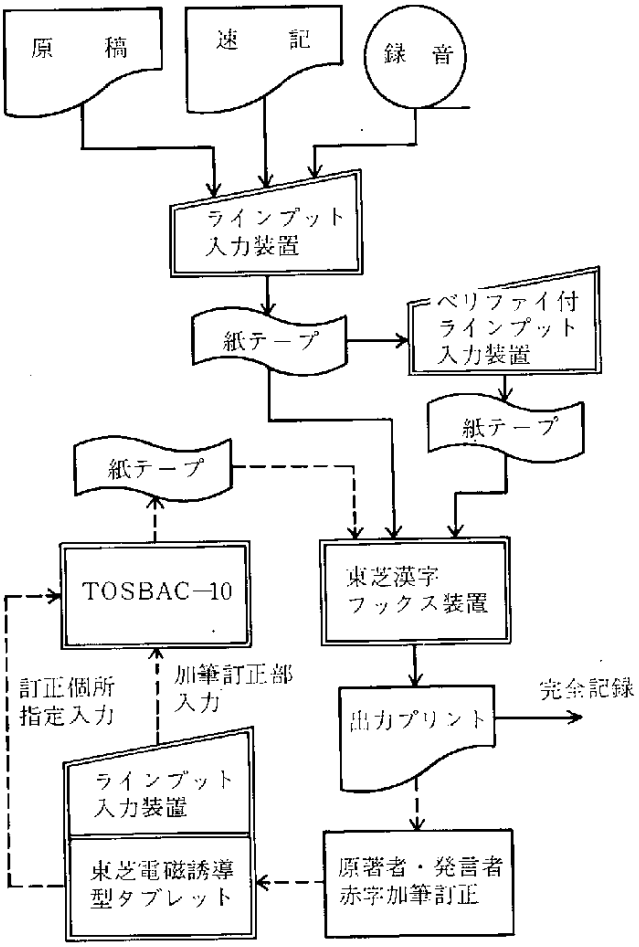


図 6・9・1　システム構成と流れ図

現在，実線の部分が稼動中であり，点線の部分が実験中である。ソフトウェアとしては，第二次情報処理のためのシステムで用いている漢字コードが種々あるので，この間の符号変換を行なうプログ

ラム，タブレット付入力装置を用いた校正を管理するプログラムが用意されている。

システムの作業の流れを図6.9.1に示した。

録音，速記，原稿，あるいは印刷物の形で与えられた日本語情報は，ラインプット入力装置で入力され，必要に応じて別のオペレータによりベリファイされる。入力の結果は紙テープ（将来はカセット磁気テープ）に一時記録され，東芝漢字ファックス装置で形式をととのえて印刷される。この印刷結果が情報源（発言者，著者，速記者あるいは校正者）に戻され，必要な赤字加算訂正が加えられる。訂正原稿と紙テープをタブレット付入力装置にかけ，TOSBAC－10の校正管理プログラムのもとで校正され，完全な日本語情報となって，再び紙テープおよび漢字ファックス装置に出力される。

6.9.4 漢字入力装置

(1) ラインプット入力装置の開発

さきにシステム設計のポイントで述べたように，現在の人手による反訳，漢テレけん盤入力の隘路を打破するには，まったく新しい入力装置の開発が必要であった。

入力速度を向上させ，誤入力率を小さくし，基本文字以外の約1万字の漢字を，オペレータが辞書をほとんど参照しなくても正しく入力できるようにするために多くのくふうがなされ，ラインプット漢字カナ入力装置が開発された。

入力けん盤としては図6.9.2のように配列されたカナ・ローマ字けん盤をもったタイプライタを用いている。このけん盤はタッチメソドで入力できるために，オペレータは原稿のみを見たまま，高速，正確に入力できる。すべての基本文字は48けん盤の2打けんによって入力される。タイプライタにおいては，左右の手を交互に使う場合が最も高速に入力できることが知られている。そこで日本語情報の50～70％を占める平仮名と出現頻度の高い漢字，数字は，交互打けんとなるようにけん盤配列が工夫されている。また5本指の中で人さし指は最も機敏に動き，子指やその他の指は動きが鈍いことから，指の負担率にも注意がはらわれた。

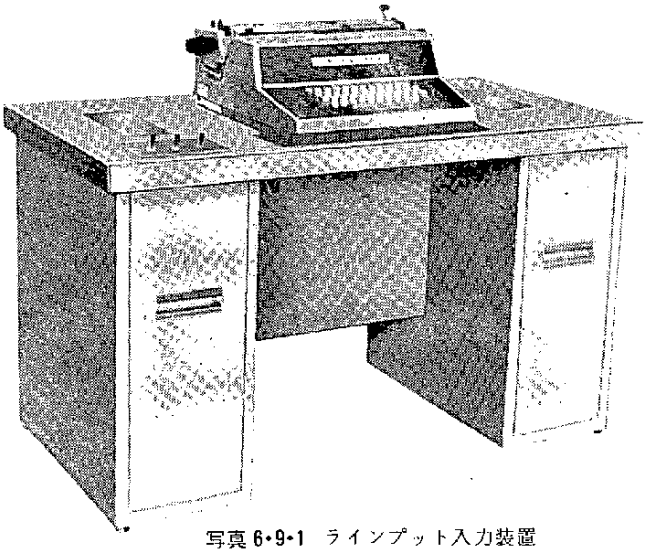


写真6・9・1 ラインプット入力装置

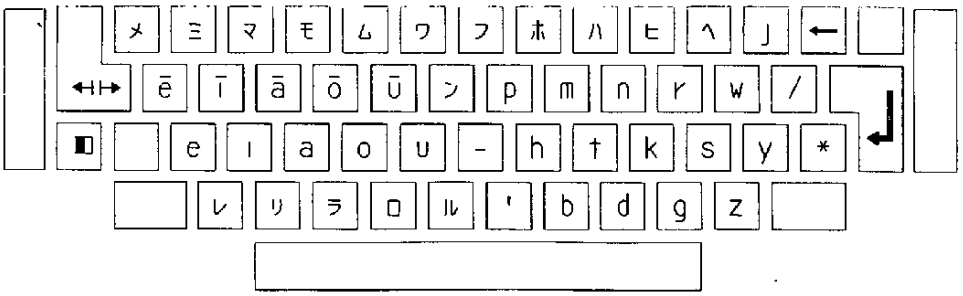


図 6・9・2　ラインプット入力装置の鍵盤配列

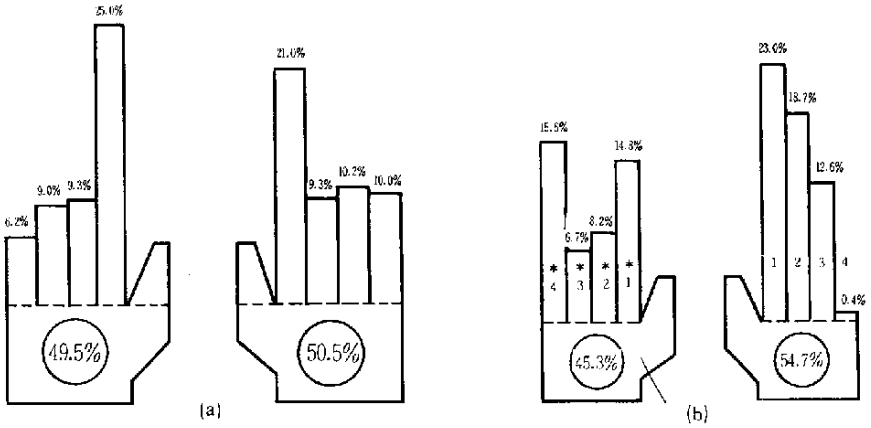


図 6・9・3　ラインプット入力装置(a)と欧文タイプライタ(b)の指の負担率の比較

図6.9.3には，ラインプット入力装置で漢字カナ混り文字を入力したときの平均的な指の負担率と，英文タイプの場合を対比して示したが，明らかにラインプット入力装置の方が優れた配列になっている。

基本文字は2回の打けんで1字があらわされる。ある文字に対し，どのけん盤の組み合わせで打鍵すればよいかをオペレータが容易に記憶でき，正確に高速に入力できることがキーポイントになる。平仮名，片仮名は訓令式のローマ字表現になっており，右手で子音を，左手で母音を打けんする。

図6.9.2のけん盤配列でもわかるように，右手の子音は濁音，半濁音を含めたすべての日本語の子音を含んでいる。下から2段目の母音と子音を組み合わせたときが，平仮名をあらわし，下から3段目の母音を組み合わせたときが片仮名をあらわす。ア行の子音としてはアポストロフィー（'）を用い，促音の「っ」は（WU）で，撥音の「ん」は（n～」で，拗音の「ゃ，ゅ，ょ」は（ja，ju，jo）であらわす。

このような配列のけん盤を用いると，JISけん盤のようにカナ清音の1字が1打けんであらわされる場合に比べ，はるかに高速に入力ができる。ラインプット配列でカナ文字を打けんすると，JIS配

列より指の動きが滑らかになり，交互打ちのために間断のない操作が可能となる。朝鮮の諺文(ハングル)はやはり母音と子音の組み合わせで表現されており，諺文タイプライタとJISカナタイプライタとの速度競技会で，圧倒的に諺文タイプの方が高速であった事実からも同じ結論がみちびかれる。またタイプを学習する場合，ラインプット配列の方が上達が早いことが立証されている

（注1）ウスイユキオ；「日韓親善盲人タイプコンテストとハングルタイプライタ」，カナノヒカリ，598巻，pp13～15，1972年6月

'a 'i 'u 'e 'o あ い う え お		eo え
ka ki ku ke ko か き く け こ	ga gi gu ge go が ぎ ぐ げ ご	
sa si su se so さ し す せ そ	za zi zu ze zo ざ じ ず ぜ ぞ	
ta ti tu te to た ち つ て と	da di du de do だ ぢ づ で ど	
na ni nu ne no な に ぬ ね の		
ha hi hu he ho は ひ ふ へ ほ	ba bi bu be bo ば び ぶ べ ぼ	
ma mi mu me mo ま み む め も	pa pi pu pe po ぱ ぴ ぷ ぺ ぽ	
ya yu yo や ゆ よ		ja ju jo や ゆ よ
ra ri ru re ro ら り る れ ろ		
wa wu wo わ う を		n- ん

図6・9・4 ラインプットにおける入力文字の組み合わせ

漢字を打けんするときは48のキーを図6.9.2に示すように，カナ清音に対応させる。平仮名，片仮名の表現に用いた2打けんの組み合わせを除外して，残りの組み合わせが漢字，英数字，および記号に対応づけられている。漢字のうち出現頻度が高いものの多くには，通常用いる読みの上の2音がそのまま用いられている。（例　山（ヤマ），海（Uミ）上（Ue），下（Zk…シタ）など），また扁や部首のあるものには常に同一の音が与えられており，系統的に記憶できるようになっている。（例　クニガマエには（ロ）が対応しているので，回（ロロ），国（ロタ…クニガマエの中に玉タマがある），

圏（ロケ）とあらわされる。）その他の漢字の2音にも漢字の意味や，形，熟語などの連想しやすい記憶の手かがりと関連が与えられているため，機械的に漢字の音を憶える場合に比べ，比較にならないくらい容易に憶えられ，また一度憶えると忘れないことになる。これはちょうど，英語の単語の意味を単語帳から全部記憶しようとしても容易なことではないが，単語とともにその単語を使う短文も憶えるようにすると記憶がらくになることが知られているのに似ている。

またラインプットの練習では，文字に対する2音を頭で憶えることが主眼ではないことに注意しなければならない。われわれが自分の考えたことを手で文章に書くとき，一つ一つの文字の筆順などはほとんど意識に上らずに，いわば手がひとりでに頭で考えた内容に即した文字をつぎつぎに書いていくように，ラインプットの打けんでは，原稿の文字をみると指がひとりでに対応する2音を打けんしていくような感じである。漢字混り文を紙の上に書くときには，一文字について平均4～5画程度書かねばならないのに対し，ラインプットでは2回指を動かせばすむので高速に入力できるのは当然のことである。それゆえ，漢字に対応した2音を頭で憶えるというより，筆指で憶えるといった方が適切である。

熟練した英文タイピストでも，まったく意味をもたない単語からできた文章をタイプさせると，素人と変わらない打けん速度となることはよく知られていることである。すなわち通常の文章であれば，タイピストは一文字ずつを意識して打けんするのではなく，一連の指の動きとして単語を打鍵していくのが普通である。従来の邦文タイプライタや漢テレけん盤は，いわば一字ずつ意識しながら打けんしているわけであるから，原理的にも英文タイピストの高速打けんとは異質の作業であり，神経も使うことになる。ラインプットでは上にも述べたように，オペレータはほとんど一字ずつの打けんを意識することなく，原稿を読みとばしながら打けんしていくので，英文タイピストにきわめて近い性質をもつことになる。ラインプットで入力した原文と入力文の例を図6.9.5に示した。

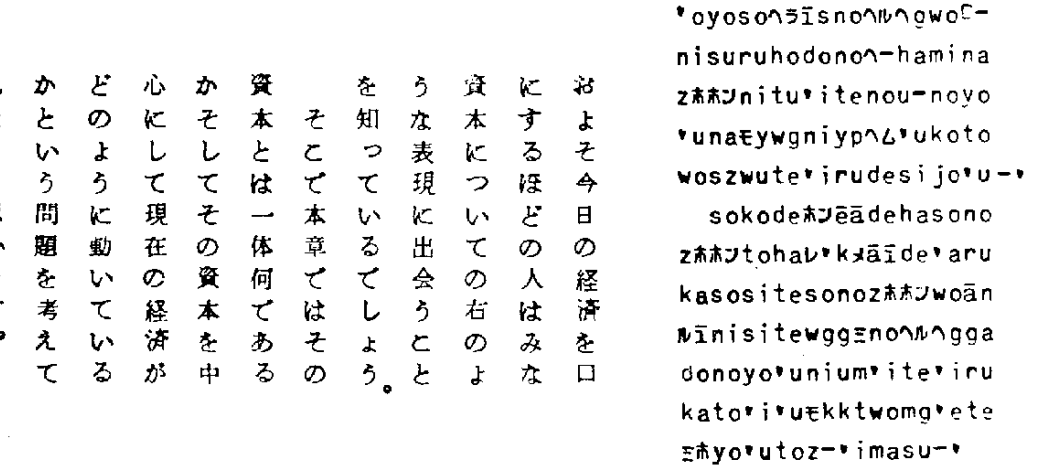


図6・9・5 ラインプットで入力した原文と入力文の例

基本文字以外の漢字約1万字に対しては，4打けんで1文字を入力するのを原則としている。その打けんの仕方はふたつの方法がある。

ア．部分文字の組み合わせ

たとえば晦，森などは基本文字には含まれていない。これらの文字に対してはその部分文字である基本文字の日，毎，雨，林の2打けんを続けて行ない，合計4打けんであらわす。この方法を用いると，基本文字の2音読みを知っていれば約5,000字の基本文字以外の漢字を，辞書などまったく見ることなく入力することができる。

イ．字画の組み合わせ

字画の種類に適当なカナ文字を対応させておき，文字を左上から書き順にしたがって分解して，4回打けんする方法で残りの約5,000字を入力する。たとえばクニガマエは「ロ」，サンズイは「ミ」などのカナ文字が対応づけられており，囮という文字は「ロイヒ・」と打けんする。最後の点「・」は4打けんに数を合わせるためのものである。

機能符号も2打けんであらわされている。機能符号の第一打けんは，シフトけん盤と同時に5種の母音のいずれかを打けんし，2打けん目との組み合わせで機能の種類が指定される。総計で5×48＝240種の機能符号がつくれるが，現在は約200種を用いている。

(2) ラインプット入力装置の使用経験

ラインプット入力装置のオペレータは，専門学校で3カ月の教育をうけるが，その間に約300時間の練習が行なわれる。26人の生徒についての学習曲線を図6.9.6に示した。

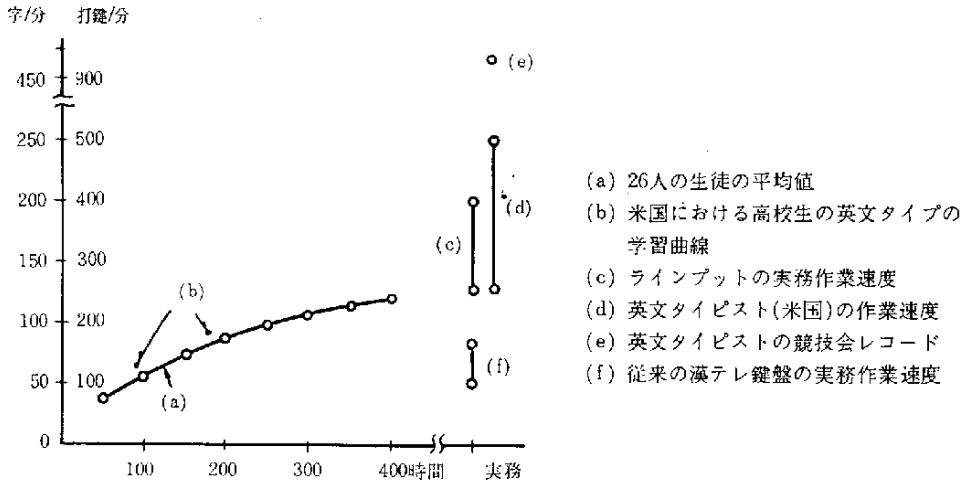


図 6・9・6 初心者に対するラインプット入力装置の学習曲線

これは新聞や小説などの試験文章を与えての成績で，誤打けんについては5倍して総打けん数から差し引かれている。図6.9.6には米国の高校生が英文タイプライタを習ったときの平均的な成績も同時に示されており，ほとんど大差のない曲線であることがわかる。現在，実務についている人が漢字混り文を平均170字／分で打けんしている。英文タイピタトの平均が50語～80語（1語平均5打けんであるから250～400ストローク／分）であるから，2打けん1字に換算すると漢字混り文の125～200字／分に相当し，この点でもよく一致する。英文タイピストの熟練者は短時間であれば120語／分が打てるので，ラインプット入力装置でも300字／分の熟練者があらわれても不思議ではない。

誤入力率についても英文タイピストとほぼ同じ程度である。ラインプット入力装置では普通のタイプライタと同じ機構を用いているので，打けんと同時にカナ・ローマ字混りのモニタ印刷が得られる。オペレータは自分で気がついた誤入力を，その場ですぐ訂正することができる。また計算機の入力データにするような場合には，ベリファイ機構のついた入力装置で紙テープ上の情報の照合訂正を行ない，より正確な原始データとすることができる。

(3) 校　　正

日本語の第一次情報は原著者みずからがタイプで清書する習慣がないので，必然的に加筆訂正する量が多くなる。従来の第二次情報処理では，漢字ディスプレイ装置を用いて校正を行なうものが多かったが，コスト的にあまりにも高くつきすぎ実用的でなかった。本システムでは東芝電磁誘導型タブレット装置を用いて，校正個所を指定する方式を採用した。この方法はコスト的に最も安く，校正個所が正確に指定できる点ですぐれている。校正の結果再び後述する漢字ファックス装置で印刷し，出力される。

(4) 録音テープからの入力

日本語情報の原始データは文字ばかりでなく，口述，会議録音などの音声である場合も考えられる。録音の品質，発言者の発音速度にもよるが，足スイッチを用いて録音テープの前進，後退再前進，停止などを制御しつつ，イヤホーンで聞いてラインプット入力装置で入力するのに，録音時間の4倍以内であれば余裕をもって行なうことができる。録音内容がよく整理されたものであれば2～3倍でも可能である。

(5) 速記原稿からの入力

会議の記録の方法として，録音以外に速記記録の技術が発達している。手書きの速記の方法にいくつかの方式があり，さらに速記機械としてソクタイプが開発され，裁判所その他で活用されている。これらの速記原稿を反訳する場合，反訳者は単に逐語的に記録文に変換するばかりでなく，発言者の区別や発言とともに示した図面や物の説明を適当に加えたり，発言者のいい直しなどを適当に修正しつつ句読点を加えて文章化し，さらに記録としての形式をととのえるという，種々の作業を並行して行なうのが普通である。また発言の中には，日本語の特徴として多くの同音異義語が含まれているが，これに適切な漢字をあてて記録化することは，反訳作業の重要な点である。これらの反訳作業は，現在の電子計算機による翻訳技術を用いても，まったく見通しが立たないほど人手によらねば解決できない問題である。ラインプット入力では，速記をとった時間の3～4倍をかければ，上に述べたような反訳の作業を完全に行なった記録文を得ることができ，手書き反訳原稿をさらにタイプする作業などはまったく不要となる。

6.9.5 漢字出力装置

漢字の出力装置として機械式のプリンタもあるが，コスト性能比ですぐれている東芝漢字ファックス装置を出力装置として採用した。標準的に内蔵している文字パターン数は基本文字＋αの2,600字であるが，必要に応じて増設することも可能である。現在のシステムでは，文字パターンに含まれていない文字では「ゲタ」を印刷している。この出力装置は縦書き，横書きいずれの書式でも印刷で

き，1行24文字で600行／分の印刷速度をもっている。1台で約6～8人分のオペレータの入力情報を出力印刷できる能力をもっている。日本語記録システムの出力装置としては手頃な速度と価格である。この出力装置で得た記録紙上に直接赤字加筆訂正したものをフィードバックして校正するので，システム構成に無駄がなく，バランスがとれたシステムであると考えている。

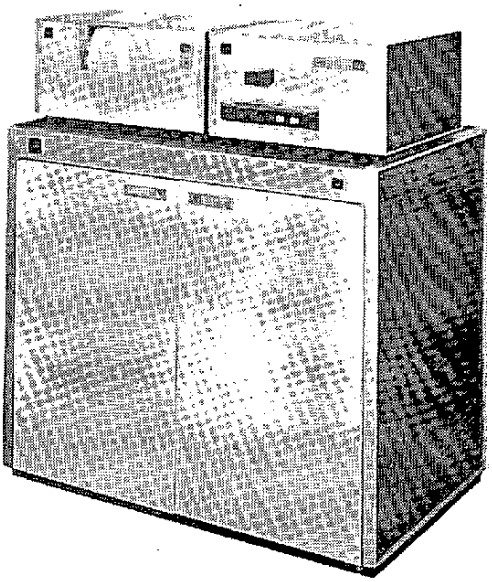


写真 6・9・2 東芝漢字ファックス装置

静電記録方式を用いた漢字出力装置であるため，コピーの枚数，用紙に特殊紙を使うことなどの問題点があるが，複製は別の方法でも容易に行なえる。出力装置はさらに小型化し，安価することが望まれている。

6.9.6 運営状況とシステム評価

この日本語記録システムの前半部を用いて，現在，1日20万字程度の処理を行なっている。オペレータの入力速度は125字～200字／分が実務平均速度である。従来の漢テレけん盤などでは想像できないような高速，正確な入力と，迅速な記録化が達成されている。オペレータは待ち時間，準備時間，休憩時間を考慮して，一日の適正作業量は実働4時間で2万～2万五千字程度が年間平均した場合の適正な量であると考えている。これでもシステム設計の目標であった従来の機械化システムの2倍以上の能率向上はらくに達成されている。さらに6.9.1で述べた本システムで目標とした五つの項目は，だいたい満足できる程度に達成されている。

6.9.7 将 来 の 構 想

今後システムの改良点として残されている点をいくつか列挙すると，

① さらにコスト性能比を向上させる。

② 各構成装置をデスク型から卓上型にまで小型化を進める。

③ 第二次，第三次の情報処理システムとの連結をはかり，機能の拡充を行なう。

などが考えられている。新しいくふうと使用経験の蓄積により，さらにすぐれたシステムに育てるべく努力をしている。

6.10 電子計算機による国語研究

6.10.1 沿革と概要

国立国語研究所は国語問題解決のための資料を作成するということを，その主要な任務としている。国語問題は広い範囲にわたる問題であるが，その最も重要な部分は，漢字，かなづかい，送りがななど，表記に関連した書きことばの問題であろう。そして，それら表記の問題は，必然的に用語の問題とも結びついているので，全体として用語用字の問題と呼ばれている。この用語用字問題解決のための基礎資料としては，現代における実態を調べたものが，まず，第一に必要であろう。そこで国立国語研究所においては，長期にわたって用語用字に関する実態調査をすすめてきた。現在までに「婦人雑誌」「総合雑誌」「現代雑誌90種の用語用字」のような調査が行なわれた。これらはいずれも20万から40万の規模のものであって，規模としてはそう小さいものではない。

しかし，用語の数はかなり多い。「広辞苑」でも約20万語を含んでいるという。のべ数10万の規模では，各単語の出現度数がきわめて低く，それをもとに審議の参考資料としようとしても，その絶対数が不足なのである。そこで，調査の規模をさらに多くして，数百万の用語用字の調査を行なおうとした。そのための人員の確保は不可能なので，調査のために機械を導入することをはかり，調査工程に電子計算機を使用することを企図した。同時に，漢字かなまじり文をそのままの形で処理することが，研究目的の上から要請されているので，漢字テレタイプを使用して，電子計算機との連結をはかることにした。使用機種は，計算機はHITAC－3010，漢字テレタイプは沖電気製打けんせん孔機，印字機であり，計算機と漢字テレタイプはオフラインで，紙テープを介して情報が授受される。

この機構により用語用字調査が進められた。現在，現代の新聞用語約300万語の用語の表が作成刊行されることになっている。

このような用語調査の進行と同時に，その用語調査で電子計算機をより有効に使用するために，したがって全体をより多く自動化するために，各種の研究を行なっている。これらは電子計算機で自然語を扱う言語情報処理の一部をなすと考えられるが，このような言語情報処理一般に関しても研究をすすめている。

以下，6.10.2機器構成，6.10.3用語用字調査の内容，6.10.4言語情報処理の研究，について説明したい。

6.10.2 機器構成

- 電子計算機：HITAC－3010(主記憶装置20,000キャラクター)，高速印字装置（LP－ALPHANUMERIC，カナ），磁気テープ読取装置，（1,200フィード・6デッキ），紙テープ読取装置，紙テープせん孔装置
- 漢字テレタイプ（沖電気製）

収容文字：当用漢字1845（朕，顰，脹，壱，弐の5文字を除く），表外字264字（雑誌90種の用字調査で度数11以上であったもの）ほかにかたかな，ひらがな英数字特殊記号など291字で合計2400字

せん孔機構：4段シフト600キー

印字機構：読取せん孔速度　120キャラクター／分

盤外字の処理方法――盤内3字で構成する。第一字目は盤外字マーク，◆。第二字目と第三字目は，ノンシフトの漢字450字で構成。この漢字自身を一つのコードのように使用し，その字のもつHITAC－3010コードによる表現により，大小順に並べる。一字目を部首，二字目をそのなかの漢和字典中の順序として表現した。このコードは盤内漢字をふくめて漢字全体をコード化部首順配列がそのまま可能であるようにし，台帳としては大修館の大漢和字典を使用する。たとえば「芥」→◆供組など。このエンコード・デコードはコードブックを作成して人間が行なう。

6.10.3　用語用字調査

現在行なっているのは新聞300万語の処理であり，朝日，毎日，読売の三紙について昭和41年の朝夕刊全紙面の1／60のサンプルについての調査である。調査は昭和41年から着手した。用語用字調査の現行の手順は図6.10.1に示したとおりである。

処理のあらすじをいえば，まず，かなり大がかりなプリエディトを行なう。すなわち文章の種類（新聞記事の多面的な分類）の情報を書き入れ，日本語は分かち書きしてないのでそれを行なうなどである。そして紙テープにパンチし，入力する。機械処理の後一度出力して人間が中間処理をする。漢字に読みがなをつけ，長い単位を短く切り，それに語種（漢語，和語，外来語など），品詞（名詞，動詞など）などの情報をつける。そして再入力し，機械処理をし，各種の語い表を作成する。たとえば語い表としては

五十音順長単位表

総度数順別度数つき長単位表

各層内長単位順位表

総度数順短単位表

総度数順類別短単位表（語種，品詞など）

用例表としては

KWIC（カナ[注1]，漢字）

KWOC

カード形式

漢字表としては

総度数順層別度数つき漢字表

部首順総度数つき漢字表

部首順長単位用例つき漢字表

その他である。

現在までに次のような報告書が刊行された。

電子計算機による新聞の語い調査[注2]

〃　　II

〃　　Ⅲ

現代新聞の漢字調査

語い調査の規模としては，長い単位でのべ200万，短い単位で300万語におよぶ。上記の調査（「電子計算機による新聞の語い調査」）はその１／３の大きさでの中間報告である。完全な語い表は昭和48年３月に出，長単位異なり18万語を収容している。漢字調査はやはり全体の１／３の大きさで異なり約2879字の表である。

この処理方法での最大の問題点は，全体を人間が行なった場合のレベルへ近づけないことである。たとえば文字連続で「かき」とあった場合，人間なら文句なく前後の文脈から判断してこれが「柿」なのか「牡蛎」なのか，「書き」という意味なのか，「掻き」なのか，「書き」という意味なのか，「掻」という意味なのかを判断する。しかし機械に，実験室的研究の規模でなく，業務のレベルでこれを行なわせるというのは，現状では不可能に近い。まず第一に自動処理を行なうとしても十分な用意がない。用意があっても現在の計算機の記憶容量の制限でどのくらいまで有効な処理ができるかは疑わしい。

この場合の第一の理由と第二の理由とは区別してかからなければならない。第二の理由は，ハードウェアの進歩によって，解決する時期がくるかもしれない。しかし第一の理由はハードウェアが進歩しても，全く先へは進めないのである。全くソフトウェア，あるいは基礎的な言語研究の問題である。

われわれの用語調査では前述の点が解決できなかった。機械処理は人間処理よりも低いレベルでしか大量処理ができない。そして後でもいうように，同形語判別など問題は用語調査のみならず，カナ漢字変換のときも重要な問題である。「カキ→柿，牡蛎，書き，掻き」など，端的な例である。

外国語にはカナ漢字変換ということはないが，音声タイプライタを考える際には同様な問題が出てこよう。アメリカで機械による用語統計が計算機を使用して行なわれた場合も，この問題が解決できなかった。たとえばcan（できる）とcan（缶）のごとき，bear（熊）とbear（生む）とbear（運ぶ）のごとき，重要な単語にかなりあるのである。

このような重要な未解決問題をのこしたままわれわれの語い表はできあがっている。「いっ」という文字連続は「言っ（た）」「行っ（て）」「いっ（本）」などの区別をつけることができず，全体をまとめた数字が示されているのみである。「あっ」も「あっ（た）」のか，感動詞なのか，そのなかにどれくらい，それぞれが含まれているかわからないまま，処理を終了した。これを区別するためのポストエディトはあと数年を要する。

このように，問題点はあるが，大量のデータを処理して，その結果を公開したので，利用価値はあると思う。ことに昭和48年３月刊の約18万語の語いを載せた語い表は，いろいろな意味で有効に使用されよう。

（注１）カタカナのKWIC用例表については表6.10.4にその一例を示した。

（注２）新聞の語い調査については表6.10.1にその一例を示した。

資料の収集 → サンプリング → 前処理 → 入力作成 → 機械処理Ⅰ → 中間処理 → 再入力 → 機械処理Ⅱ → 語彙表作成 → 機械処理Ⅲ → 各種分析

資料の収集
朝日、毎日、読売
3紙1年
長　1億2千万語
短　1億8千万語

サンプリング
標本指定
長　２００万語
短　３００万語

前処理
層別記入
単位分割
パンチテキスト作成

入力作成
漢テレパンチ
校正
修正パンチ

機械処理Ⅰ
プログラム作成
長単位処理
オペレーション
漢テレアウトプット

中間処理
よみがなつけ
単位細分
付加情報
パンチテキスト作成

再入力
漢テレパンチ
校正
修正パンチ

機械処理Ⅱ
プログラム作成
短単位処理
オペレーション

語彙表作成
長短各種
語彙表作成

機械処理Ⅲ
各種プログラム作成とそのオペレーション

各種分析
漢字集計
語表記
類別語彙表用例集

原資料についてサンプリング箇所の指定
ブロック番号層別の記入

七日ぶり開通

北海道向け輸送 七割確保

いた奥羽本線大鰐―石川間（青森県）は、十九日午後六時ごろ、一週間ぶりに開通することになった。

るが、奥羽本線の開通で北海道方面への輸送難はなかば解消することになる。

徐行区間が残っているため、十九日夜は全般的に変則ダイヤになるが、二十日かわは一部のローカル列車を除き、全列車を通常ダイヤの七割程度が確保できる見通しだ。

れで北海道向け物資の輸送は平常

と国鉄ではいっている。また貨物列車は二十日から一日片道貨車五百両を通すことになっており、こ

東北本線も23日までに開通

【青森】国鉄東北支社青森出張所は十九日、奥羽本線開通後の同本線短区間の貨客輸送と、東北本線の復旧の見通しなどについて次のように発表した。

奥羽本線復旧後五、六日間は、大館―陣馬間で一日上下各一本分、弘前―碇ケ関間は下り二本分、

上り三本分のバスを運転してお客をさばく。貨物車輌は、しばらくは四百五十両を用意、五十両分を自動車にかえて運送する。

また、東北本線浅虫―野内間の不通個所も遅くても二十三日までに復旧できる見通しがついた。貨物は東北本線開通後も、当分トラック輸送と室蘭―八戸間の船による代行輸送で補う。

＝関係記事、7面に

ブロック番号
Ｊ１２３５２３
Ｊ１：朝日新聞夕刊
　　６月～12月

層別
Ｐ１：見出し
Ｐ４：本文
Ｓ１：無署名記事
Ｔ５：社会
Ｇ１：ニュース

長単位　｜東北本線｜も｜23日｜まで　に｜開通
短単位　｜東北｜本線｜も｜2｜3｜日まで｜に｜開通

前処理ずみインプット見本

（Ｊ123523）（Ｐ1Ｓ1Ｔ5Ｇ1）＃□東北本線□も□23日□までに□開通□（Ｐ４）＃□【□青森□】国鉄東北支社□青森出張所□は□十九日□，□奥羽本線開通後□の……

青森出張所
青梅署
青柳美枝子
税制簡素化
税務当局
税務相談所

（機械処理Ⅰがすんで漢テレで語形を印字したもの）

中間処理

（単位切り、かなつけなどの人間作業）

Ｐ568
ｎ青〔あお〕森〔もり〕（Ｗ800）
ｐ出〔しゅっ〕張〔ちょう〕（Ｔ100）
ｓ所〔じょ〕（Ｔ600）
ｎ青〔おう〕梅〔め〕（Ｗ800）
ｓ署〔しょ〕（Ｔ100）
ｎ青〔あお〕柳〔やぎ〕（Ｗ800）
ｓ美〔み〕枝〔え〕子〔こ〕（Ｗ800）
ｎ税〔ぜい〕制〔せい〕（Ｔ100）
ｐ簡〔かん〕素〔そ〕（Ｔ100）
ｓ化〔か〕（Ｔ600）
ｎ税〔ぜい〕務〔む〕（Ｔ100）
ｓ当〔とう〕局〔きょく〕（Ｔ100）
ｎ税〔ぜい〕務〔む〕（Ｔ100）
ｐ相〔そう〕談〔だん〕（Ｔ100）
ｓ所〔じょ〕（Ｔ600）

（各長単位語について語構成、漢字の読み、付加情報、短単位分割をしたもの）

語構成　ｎ 前部分　ｐ 中部分　ｓ 後部分
漢字の読み〔　　〕のなか
付加情報　例　出張（Ｔ100）
　Ｔ漢語　１名詞　０活用がない　０活用の種類に関係がない
短単位　改行で示す

図 6·10·1 機械処理による用語調査の過程

全体 順位	全体 出現率	見出し	語種・品詞	度数	部分 順位	部分 出現率
		*集まる	和・動	10	5415	.023
		*集る	和・動	18	3431	.042
5713	.013	淵美	別・名	12		
1566	.068	集め	(和・動)	55		
		*集める	和・動	62	1167	.144
9648	.006	集める	(和・動)	6		
3140	.027	圧力	漢・名	25	2656	.058
6367	.014	当て	(和・動)	13		
4271	.018	宛	和・名	17	3591	.039
2223	.039	あて	～	37	1890	.086
6142	.012	あて先	和・名	11	5054	.026
		*あてる	和・動	6	7686	.014
9648	.006	あてる	(和・動)	6		
3140	.027	跡	和・名	25	2656	.058
332	.263	後	和・名	247	229	.578
412	.221	あと	～	206	294	.482
9648	.006	アドバイス	外・名	6	7686	.014
9648	.006	アナ	～	6	7686	.014
2699	.032	アナ	外・名	30	2288	.050
3353	.024	穴	和・名	23	2822	.063
5024	.015	アナウンサー	外・名	14	4197	.032
449	.198	あなた	和・代名	186	327	.431
5357	.014	貴方	和・代名	13	4436	.030
8664	.007	貴女	和・代名	7	6947	.016
9648	.006	阿國	固・名	6		
6142	.012	兄	和・名	11	5054	.026
6615	.011	姉	和・名	10	5415	.023
2631	.033	あの	和・連体	31	2231	.072
1721	.053	アパ	外・名	50	1444	.116
1641	.056	アパート	外・名	52	1374	.121
7196	.010	暴れ	和・名	9	5862	.021
8664	.007	あばれ	和・名	7	6947	.016
3234	.026	浴び	(和・動)	24		
7196	.010	アピール	外・名	9	5862	.021
		*浴びる	和・動	24	2303	.063
8664	.007	アフター	外・名	7	6947	.016
7876	.009	アフタヌーン	外・名	8	6347	.019
		*あぶない	和・形	10	5415	.023
7876	.009	あぶない	(和・形)	8		
9648	.006	脂	和・名	6	7686	.014
2862	.030	油	和・名	23	2417	.055
6142	.012	油谷	別・名	11		
1202	.077	アフリカ	固・名	72		

全体 順位	全体 出現率	見出し	語種・品詞	度数	部分 順位	部分 出現率
		あふれ	(和・動)	9		
		あふれる	(和・動)	11		
		*あふれる	和・動	20	3153	.046
7196	.010	安部	固・名	9		
3462	.023	阿部	固・名	22		
8664	.007	アベック	外・名	7	6947	.016
9648	.006	アポロ	固・名	6		
7196	.010	アマ	外・名	9	6862	.021
		甘い	(和・形)	19		
		*甘い	和・形	25	2656	.058
		*あまい	和・形	5	8567	.012
10844	.005	アマゾン	固・名	5		
6615	.011	アマグ	固・名	10		
9648	.006	アマチュア	外・名	6	7686	.014
7876	.009	天辺	固・名	8		
9648	.006	天満	固・名	6		
8664	.007	アマノ	固・名	7		
5713	.013	天野	固・名	12		
5713	.013	余	和・名	12	4723	.028
8664	.007	余り	～	7	6947	.016
1137	.081	あまり	～	76	938	.176
8664	.007	あまり	副・副	7	6947	.016
7196	.010	余	副・接尾	9	5862	.021
9648	.006	編み	和・名	6	7686	.014
		*編む	和・動	9	5862	.021
616	.146	雨	和・名	136	472	.316
368	.242	アメリカ	固・名	228		
7196	.010	綾子	固・名	9		
10844	.005	あやまっ	(和・動)	5		
5024	.015	誤っ	(和・動)	14		
5718	.013	誤り	～	12	4723	.028
		*あやまる	和・動	6	8567	.012
		*誤る	和・動	14	4197	.032
8664	.007	歩み	～	7	6947	.016
10844	.005	あゆみ	固・名	5		
		*歩む	和・動	6	7686	.014
4086	.019	新井	固・名	18		
9648	.006	洗い	(和・動)	6		
6615	.011	荒井	固・名	10		
6615	.011	洗場	和・名	10	5415	.023
7876	.009	洗う	(和・動)	8		
		*洗う	和・動	29	2339	.067
7876	.009	あらかじめ	和・副	8	6347	.019
1936	.047	荒川	固・名	44		
7196	.010	荒木	固・名	9		

全体 順位	全体 出現率	見出し	語種・品詞	度数	部分 順位	部分 出現率
2699	.032	嵐	和・名	30	2288	.070
7876	.009	アラスカ	固・名	8		
6615	.011	アラ石	混・名	10	5415	.023
9648	.006	争い	和・名	6	7686	.014
6615	.011	争い	～	10	6415	.023
		*争う	和・動	15	3973	.035
7196	.010	争っ	(和・動)	9		
2171	.045	新た	和・名	38	1839	.088
9648	.006	新球	固・名	6		
4271	.018	改め	(和・動)	17		
4271	.018	あらため	(和・動)	17		
7876	.009	改める	(和・動)	8		
		*改める	和・動	25	2656	.058
		*あらためる	和・動	17	3591	.039
5357	.014	読っ	(和・動)	13		
7196	.010	アラビア	固・名	9		
2760	.031	アラブ	固・名	23		
6615	.011	荒磯	固・名	10		
2952	.029	あらゆる	和・連体	27	2303	.063
9648	.006	現わし	(和・動)	6		
9648	.006	あらわし	(和・動)	6		
		*現わす	和・動	6	7686	.014
		*あらわす	和・動	9	5862	.021
3735	.021	現れ	(和・動)	20		
7875	.009	現われ	(和・動)	8		
3735	.021	あらわれ	(和・動)	20		
		*現れる	和・動	22	2914	.061
		*現われる	和・動	10	5415	.023
		*あらわれる	和・動	24	2727	.066
3284	.026	有り	～	24	2727	.066
3853	.024	あり	和・名	23	2822	.063
154	.803	あり	～	523	80	1.206
5713	.013	あり	(和・動)	12		
		*ありがたい	和・形	12	4723	.028
7196	.010	ありがたい	(和・形)	9		
9648	.006	ありがとう	～	6	7686	.014
7193	.010	有沢	固・名	9		
10811	.005	有高	固・名	5		
9648	.006	アリナミン	外・名	6	7686	.014
5713	.013	有馬	固・名	12		
146	.581	有	(和・動)	546		
5357	.014	有る	(和・動)	13		
		*有る	和・動	561	72	1.301
9648	.006	ある	和・連体	6	7686	.014
44	2.449	ある	～	2303	10	5.341

表 6・10・1 語彙表

6.10.4 各種言語処理の研究

このように，人間作業による作業結果と機械処理による作業結果との間にレベルの差ができてしまったので，これをどのように解決するかが，問題として残されるようになった。またプリエディト，中間エディトに非常に多くの時間を要するのでその合理化をどのようにするかも，関心がもたれることになった。

まず入力は漢字テレタイプだが，出力としては導入当時高速漢字プリンタがなかったため，ラインプリタを利用しなければならないことがあった。（漢字プリンタが入手できれば，漢字のままの大量出力が可能になる。）とくに用例表のごとき，ぼう大な量の出力の場合漢字プリンタがなければラインプリタを使わざるをえない。そこで漢字→かなの変換という，他のユーザにはあまりないプログラムが必要になった。このため，二種のプログラムがつくられた。

その一つは漢字の前後関係に目をつけ1字ずつ変換する方法である。

現代語では「研」は一つの読みしかない。「生」は多くの読みがある。このように一つしかないものはその読みを利用し，いくつかあるものはその環境を調べて読みを入れる方法である。（表6.10.2参照）もう一つは単語を読み方の基本とし，単語単位で漢字を読む読み方である。（表6.10.3参照）大量の辞書を利用するが，辞書引きをプログラム的に工夫して高速化する。後者は正解率が高く（漢字当たり99％以上）実用プログラムとして使用しすでに60万語以上を実際処理した。

この逆の，かな漢字変換については，漢字かな変換システムのと類似の方式によるよみがな方式がある。これは目下開発中である。

多〔た〕数〔すう〕決〔けっ〕の原〔げん〕理〔り〕

多〔た〕数〔すう〕決〔けっ〕の原〔げん〕理〔り〕には，確〔たしか〕かに相〔そう〕対〔たい〕主〔しゅ〕義〔ぎ〕的〔てき〕な意〔い〕味〔み〕がある．甲〔こう〕論〔ろん〕乙〔おつ〕◆日送〔　　〕の意〔い〕見〔けん〕の対〔たい〕立〔りつ〕がある場〔ば〕合〔あ〕に，神〔かみ〕ならぬ人〔にん〕間〔けん〕の知〔ち〕性〔せい〕

表6・10・2 計算機によって解読された漢字（環境方式）

日本語では分かち書きをしないのがふつうの文の書き方であるから，計算機に入力されたものは，あらかじめ分かち書きが人為的にほどこしてないかぎり，計算機自身が行なう必要がある。これが自動分かち書きである。

この問題についても二種類のものが作られた。その一つは，文字の連続環境に注目して大きな辞書にあわせて文を切り，その切り方をシンタクティクに検定する方法である。（表6.10.5参照）

後者のばあい，自動構文解析のプログラムを内蔵している。前者の場合単位切りにミスが後者に比して多く出，後者は構文解析に必要な文法規則がほとんどできていない。いずれもまだ実用段階にいたっていない。（後者は漢字→かな，かな→漢字，非当用漢字非現代かなづかい→当用漢字現代かな

①原文

（K001129）（G1P1S1T5） ＝ 大谷重工 深川工場 ボヤ （P4） ＝ 三日 午後 一時 十五分 ごろ，東京 墨田区 柳原町 一○ ， 大谷重工 深川工場 の 機械工場 南側 の カベ から 煙 が 出 て いる のを 従業員 が みつけ た 。 ＝ およそ 十平方メートル の カベ を こがし た だけで 消しとめ た 。 （P1） ＝ 埋立地 に 男 の 変死体 （P4） ＝ 〔 船橋 〕 三日 午前 三時 十分 ごろ ， 千葉県 船橋市 日の出町 一○ ， 第一コンクリート会社 正門前 埋立地 の 臨海道路ぎわ に 若い 男 が うつ伏せ に なっ て 死ん で いる のを 通行人 が みつけ 船橋署 に 届け出 た 。

②漢字解読

K0011290101015 01 ＝ 大〔おお〕谷〔たに〕 重〔じゅう〕工〔こう〕 深〔ふか〕川〔がわ〕 工〔こう〕場〔じょう〕 ボヤ ＝

K0011290104010502 ＝ 三〔みっ〕日〔か〕 午〔ご〕後〔ご〕 一〔いち〕 時〔じ〕 十〔じゅう〕 五〔ご〕 分〔ふん〕 ごろ ， 東〔とう〕京〔きょう〕 墨〔すみ〕田〔だ〕 区〔く〕 柳〔やな〕原〔はら〕 町〔ちょう〕 一〔いち〕 ○〔れい〕 ， 大〔おお〕谷〔たに〕 重〔じゅう〕工〔こう〕 深〔ふか〕川〔がわ〕 工〔こう〕場〔じょう〕 の 機〔き〕械〔かい〕 工〔こう〕場〔じょう〕 南〔みなみ〕側〔がわ〕 の カベ から 煙〔けむり〕 が 出〔で〕 て いる の を 従〔じゅう〕業〔ぎょう〕 員〔いん〕 が みつけ た 。 ＝

K0011290104010503 ＝ およそ 十〔じゅう〕 平〔へい〕方〔ほう〕 メートル の カベ を こがし た だけ で 消〔け〕しとめ た 。 ＝

K0011290101010504 ＝ 埋〔うめ〕立〔たて〕 地〔ち〕 に 男〔おとこ〕 の 変〔へん〕 死〔し〕体〔たい〕 ＝

K0011290104010505 ＝ 〔 船〔ふな〕橋〔ばし〕 〕 三〔みっ〕日〔か〕 午〔ご〕前〔ぜん〕 三〔さん〕 時〔じ〕 十〔じゅっ〕 分〔ふん〕 ごろ ， 千〔ち〕葉〔ば〕 県〔けん〕 船〔ふな〕橋〔ばし〕 市〔し〕 日〔ひ〕の出〔で〕 町〔ちょう〕 一〔いち〕 ○〔れい〕 ， 第〔だい〕 一〔いち〕 コンクリート 会〔がい〕社〔しゃ〕 正〔せい〕門〔もん〕 前〔まえ〕 埋〔うめ〕立〔たて〕 地〔ち〕 の 臨〔りん〕海〔かい〕 道〔どう〕路〔ろ〕 ぎわ に 若〔わか〕い 男〔おとこ〕 が うつ伏〔ふ〕せ に なっ て 死〔し〕ん で いる の を 通〔つう〕行〔こう〕 人〔にん〕 が みつけ 船〔ふな〕橋〔ばし〕 署〔しょ〕 に 届〔とど〕け出〔で〕 た 。 ＝

表 6-10-3 計算機による漢字解読（単語辞書方式）

づかいなどの汎用の処理も同時に可能)。

そのほか，やはり試作的であるが，単語の品詞を計算機で推測すること，同音語を環境的に見分ける（かな漢字変換と同じ問題）ことなどいくつかのプログラムも実験されたが，いずれも開発中途でとまっている。このような研究開発の内容については，国語研報告書「電子計算機による国語研究」「同Ⅱ」「同Ⅲ」「同Ⅳ」「同Ⅴ」などに発表されている。

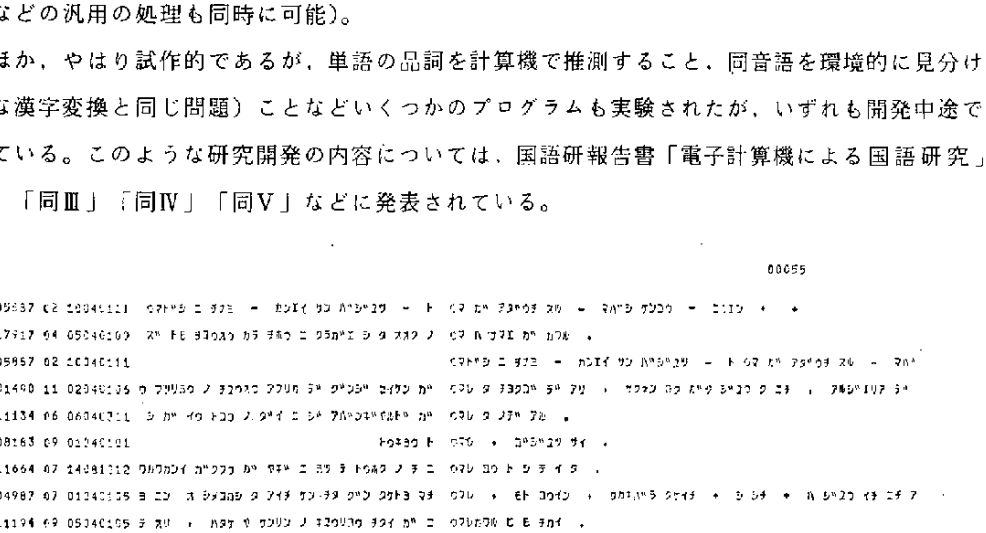


表 6・10・4 新聞用語KWIC（毎日新聞夕刊・単語辞書方式による）

入力原文

谷川の岸に小さな学校がありました。

さわやかな9月1日の朝でした。

青ぞらで風がどうと鳴り，日光は運動場いっぱいでした。おわり

```
**** DISPLAY PHRASE-STRUCTURE ****

SEG1-NO  0015    SEG3-OUT-NO   3--01


01  KOZORA          X N ------------------------------------ X V ---------- X KD --------- S KD
                                                                 I                I                I
02  LP              B PA ..................................                  I                I
                                                                 I                I                I
03  KAZE            X N ---- ------------------- X V --------                  I                I
                                                 I                                I                I
04  GA              B PA ....................                                    I                I
                                                 I                                I                I
05  DOOTO           X ADV -------- X V --------                                  I                I
                                  I                                             I                I
06  NARI            X V ---------                                               I                I
                                                                                I                I
07  .                                                                           I                I
                                                                                I                I
08  NIQKOC          X N ------------------------ X KD ---------------------                I
                                                 I                                                I
09  WA              B PD ....................                                                I
                                                 I                                                I
10  UNDPOZYOO-IGPAI X KD --------- X KD --------                                                I
                                  I                                                             I
11  DESI            X COP ------                                                                I
                                                                                                I
12  TA                                                                                          I
                                                                                                I
13                  = PERIOD ------------------------------------------------------------------
```

表 6-10-5 計算機による自動分かち書き

6.11 計算機によるカナ漢字変換の一手法

6.11.1 ま え が き

カナ文またはローマ字文を漢字かな混り文に自動変換するシステムは，海外ニュース電文の漢字かな混り文変換，計算機への漢字かな混り文の入力手段などに直接応用できるほか，日本語情報処理技術を確立する上での里程標としての意味も持つと考えられる。そのため1966年頃から，九州大学[注1]，[注2]沖電気[注3]，[注4]を中心に研究が始められ，最近では国立国語研究所[注5]日本ソフトウェア[注6]などの結果も一部公表されている。

NHK総合技術研究所は，ニュース文などの情報伝達を旨とする日本語文の機械処理技術を確立するための最初の課題として，1969年頃から計算機によるカナ漢字変換システムを取り上げてきたが，このほど，試作システムの評価実験などを終えたので，概要を紹介する。なお，このシステムは，NHKの海外支局から打電されてくるローマ字ニュース電報の漢字かな混り文変換への応用を念頭に開発したものであるが，まだ実務に使用するまでには至っていない。

カナ漢字変換における主要な問題は，同音異字語の処理である。以下では，そのための手法と，それによってシステムに課せられる使用上の制約を中心に述べる。

（注１）栗原俊彦，黒崎悦明：カナ文の漢字混り文への変換処理について，昭和43年度電気四学会連合大会予稿（昭43－３）。

（注２）栗原俊彦，稲永紘之：カナ漢字変換（1），九大工学集報，第42巻第６号，p.880～884（昭45）

（注３）黒崎悦明：カナけん盤漢字表示方式，昭和43年度情報処理学会大会予稿（昭43－12）。

（注４）沖電気：漢字カナ混り文変換プログラム，電子通信学会誌，Vol. 54，No. 11，ニュース欄（昭46）。

（注５）田中章夫：ヨミガナ方式によるカナ（ローマ字）の漢字変換，計量国語学，No. 55，p.29～37（昭45）。

（注６）藤井純，原哲夫，木村繁：カナから漢字への変換，昭和46年度電気四学会連合大会予稿（昭46－10）。

6.11.2 入力文の形式

入力のカナ文は次に示す二種の単位に分ち書きされていることを前提とする。それらを'分ち書き単位'と総称する。その特徴は，自立語が一つしか含まれないこと，自立語に続く付属語群は，すべてその自立語にくっつけて書くことの二点である。

① 通 常 単 位

接頭語－自立語－接尾語－付属語群

② 数 詞 単 位

数詞－助数詞－接尾語－付属語群

ここで，接頭語，接尾語，付属語群は実際にはなくてもよい。次に分ち書き単位の例を二，三掲げ

る。

ヘンカンースルーニハ（変換するには）

カクーグタイーサクーヲ（各具体策を）

19－ニチーニ（19日に）

9－カイーゴーカラ（9回後から）

ふつう自立語といえば片カナ表記の外来語や，地名・人名等の固有名詞も含まれるが，本システムではそれらはいずれも変換対象からはずした。一般語からなる文字通りの通常文を変換処理の当面の主眼としたからである（外来語や固有名詞の処理の見通しについては6.11.5今後の課題で述べる）。そのため，さらに次の2種の特殊単位を設けた。

③　強制片カナ変換単位

特殊記号*で囲まれた単位で，その中味を片カナのまま出力する。

④　強制平かな変換単位

特殊記号'で囲まれた単位で，その中味を平かなで出力する。ナンセンスな漢字に変換される恐れのある固有名詞の類を平かなで出力するのに用いる。

個々に分ち書き単位同士を分ける分離記号としては，ふつうは空白を当てるが，コンマやピリオドなどの句読点，(),「」などの引用符，また上記③④の単位構成に用いられる*や'もすべて分離記号と見なしておいた方が便利である。

実際の新聞記事を上述の入力形式に合うように分ち書きした例を次に示す。『ところが，ニクソン米大統領の訪中決定など，最近の国際情勢の大きな変化によって，佐藤内閣といえども，もはや中国代表団の入国を拒否するような態度はとれないだろうと判断し，招待状を出すことに踏切った。』（朝日新聞，昭和46年8月1日朝刊）トコロガ，*ニクソン*　ベイ　ダイトウリョウノ　ホウチュウ　ケッテイナド，サイキンノ　コクサイ　ジョウセイノ，オオキナ　ヘンカニヨッテ，'サトウ'　ナイカクト　イエドモ，　モハヤ　チュウゴク　ダイヒョウダンノ　ニュウコクヲ　キョヒスルヨウナ　タイドハ　トレナイダロウト　ハンダンシ，ショウタイジョウヲ　ダスコトニ　フミキッタ。（この入力文に対する変換結果は6.11.4「試作システムとその能力」に示す）

なお6.11.4で述べる試作システムでは，入力のカナコードとしてEBCDICコードを使用しているので，濁点・半濁点も独立した1字となるし，小さい文字（ッ，ャ，ュ，ョ）も通常の文字で代用される。とくに後者は同音異字語の原因になる（賞と使用，病院と美容院など）。

6.11.3　同音異字語処理

変換処理の主要部は同音異字語を除くための文法処理である。これは以下に述べる三つの部分に分かれる。

ア．基　本　分　析

ここでは分ち書き単位がひとまず次の形をしていると仮定して分析を始める（通常単位の場合を考える。数詞単位についてもほぼ同様である）。つまり分ち書き単位には接頭語も接尾語も含まれていないと仮定するのである。

自立語－付属語１－付属語２－…－付属語k　（k≧０）

このとき次の三つの操作を行なう。

①　与えられた分ち書き単位（カナ文字列）を上のような自立語と付属語列の形に区切る。

②　最後の単位（上の例でいえば付属語k）が分ち書き単位の末尾になり得るか否かを調べる。

このうち①は単語（自立語）辞書と付属語辞書を用いて調べる。②は単位間の接続をまとめた接続行列を用いて行なう。③に必要な情報は付属語辞書に収容しておく。

例として「アメラシイカラ」という分ち書き単位をとり，その分析過程を図6.11.1にしたがってたどってみよう。

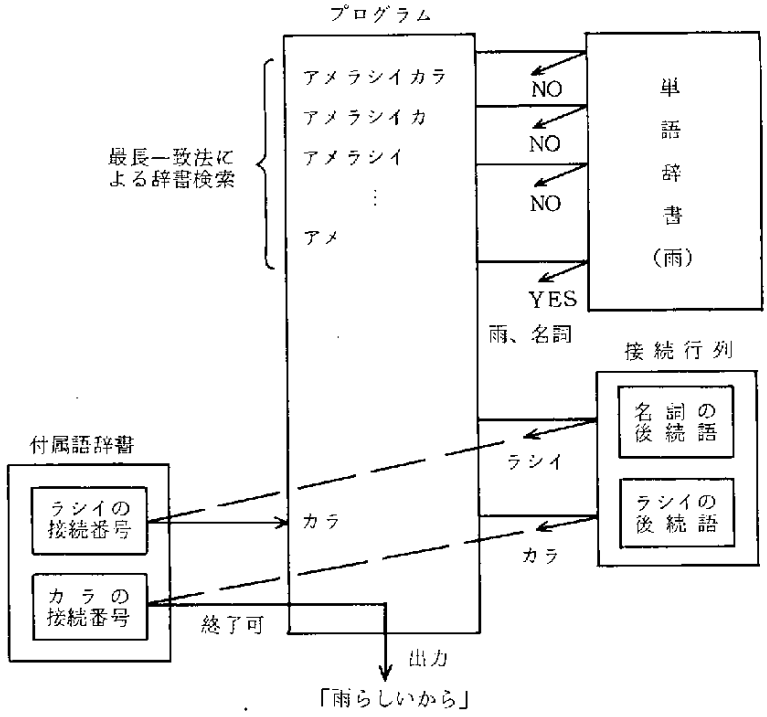


図6-11-1　「アメラシイカラ」の分析過程

まず「アメラシイカラ」という単位が，全体として一つの自立語であるか否かを単語辞書によって調べる。もちろんこれは自立語ではない（したがって単語辞書にはない）。そこで末尾の１字を削って次は「アメラシイカ」を調べる。これも自立語ではない。このように末尾から１字ずつ削っては，残りの部分が自立語か否かを調べていく（いわゆる最長一致法による辞書検索）。

この場合は「アメ」になって「雨」という自立語が得られる。このとき雨という漢字とともに，それが名詞であるという品詞情報も同時に得られる。そこで次に「雨」と残余部「ラシイカラ」の文法分析にうつる。そのためにまず接続行列の「名詞の後続語」欄にしたがって，名詞の後続単位を探していく。一般にはそのようなものは多数見つかるが，その中で残余部の頭の方と一致するものを選ぶ。この場合は「ラシイ」が見つかる。

そこでこんどはラシイの接続のパターン番号（接続行列の「ラシイの後続語」欄の番号）を付属語辞書から読み取り，「ラシイ」を削った新しい残余部「カラ」について，接続行列の「ラシイの後続語」欄を上と同様に調べていく。

そして「カラ」が得られる。「カラ」は「アメラシイカラ」という分ち書き単位の末尾だから，それが分ち書き単位の末尾になれるか否かを付属語辞書によって調べる。もしなければこの分ち書き単位の文法分析は首尾よく終ったことになり，「雨らしいから」という出力が得られることになる。

もし「アメ」に対応する自立語が「雨」以外にもあった場合には，それらについてもそれぞれ上のような分析を行ない，うまく分析できたものだけを出力するようにして，可能な限り一意的な漢字変換をはかるわけである。

次に，このような基本分析に用いる諸表について簡単にふれておこう。

単語辞書は[注9]，自立語に関する種々の情報をまとめたものである。自立語選定には国立国語研究所の新聞の語い調査の結果を基礎資料とし，その中の「五十音順短単位表」から，固有名詞の大部分・片カナ表記の外来語・助詞・助動詞・算用数字・記号類を除いた6,460語を抽出した。辞書中の各単語に収容した情報は，見出しのカナ表記，漢字表記，カナ表記と漢字表記の対応をつける情報，品詞情報，出現率（同音異字語が一つにしぼれなかった場合はこの順に出力する），その他である。

付属語辞書には，助詞，助動詞，形式名詞などのいわゆる付属語を，活用のある場合には活用形ごとに，必要な情報とともにまとめたものである。この辞書の現在の大きさは

9バイト×291語＝2,619バイトである。

接続行列は，自立語の各品詞別（活用する品詞はさらに活用形別），および各付属語別に，それに後続する付属語を記述したものである。まとめ方を工夫することによって現在の接続行列は約3,200バイトの大きさに収まっている。くわしいことは文献「分ち書き単位分析表」[注10]にゆずる。

イ．接尾語処理

基本分析の過程で，自立語は見つかるが，それと残余部との文法分析がうまくいかないという状況になったとき，残余部の頭に接尾語がついていると考えて接尾語処理に入る。その大要は，接尾語表と残余部をつき合せ，一致するものがあればそれ（接尾語）を残余部から削り，その接尾語と新しい残余部との文法分析を行なうことからなる。

この接尾語処理もうまくいかなかった場合には最長一致法による単語辞書の検索に戻り，新しい自立語を探し直して再び基本分析に入る。

現在の接尾語表には，資料「電子計算機による新聞の語彙調査[注7,注8]」から抽出した45語の接尾語が入れてある。

ウ．接頭語処理

最長一致法による辞書検索のすべての段階で，基本分析も接尾語処理もうまくいかなかった場合には，分ち書き単位の冒頭に接頭語がついていると考え，接頭語表によりそれを削ってから，残余部をあらためて分ち書き単位とみなして再び基本分析（および接尾語処理）を行なう。

現在の接頭語表には，上記資料から抽出した26語の接頭語表が入れてある。

（注７）国語研究所：電子計算機による新聞の語彙調査，秀英出版（昭45）。

（注８）国語研究所：電子計算機による新聞の語彙調査（Ⅱ），秀英出版（昭46）。

（注９）相沢輝昭，江原暉将：カナ漢字変換のための単語辞書，NHK技研研究報告（昭47）。

（注10）相沢輝昭，江原暉将：分ち書き単位分析表，NHK技研研究報告（昭47）。

6.11.4　試作システムとその能力

NHK総合技術研究所設置の計算機IBMシステム／360モデル40Gに対して作成した全システムは図6.11.2のとおりである。使用言語はすべて360アセンブラである。

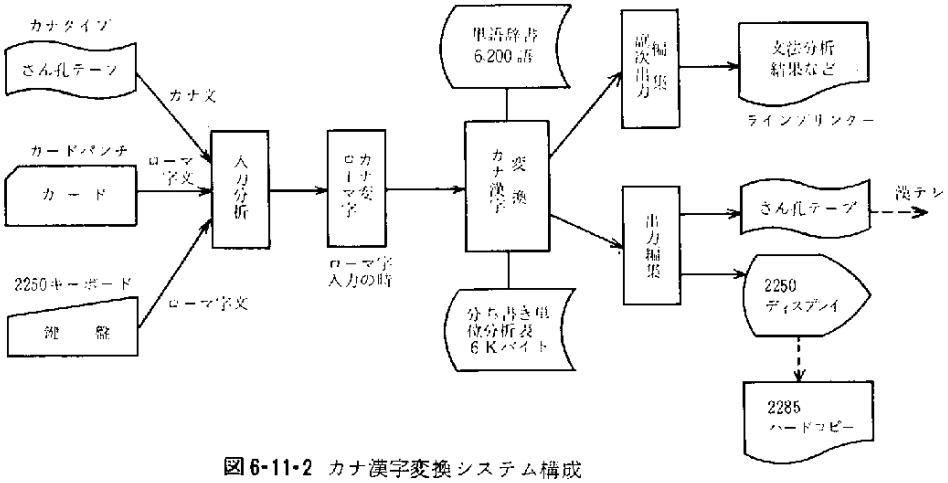


図6-11-2　カナ漢字変換システム構成

ステップ数は約4,000，コア容量は約60Kバイトであるが，紙テープ入出力の最小システムにすると約2,000ステップ，24Kバイト程度にまとめられる。

IBM2250キーボード入力，2250出力の変換例を次に示す。

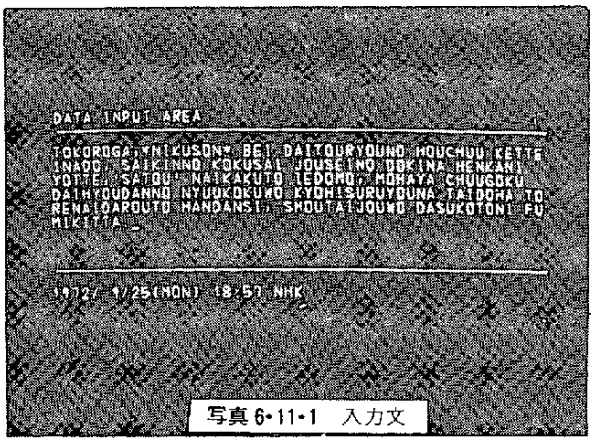


写真6-11-1　入力文

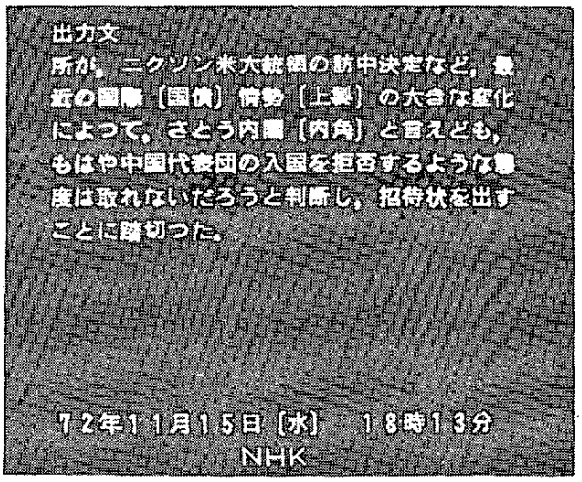


写真6・11・2　出力文

次に示すのは漢テレ出力の一例であるが，同音異字語がしぼれなかった場合にも第1順位のみ出し，他のものは抑制してみた。

社会党の中国側に対する訪日招待には，具体的に名前を上げることはしないで，人選は一切相手に任せているが，同協会の責任者である　おうこつけん　じよめい　おうぎようりん　各市等対日関係者数人の来日を期待している。社会党は去る5月，既に中日友好協会代表団の招待を決めたものの，さとう内閣が入国を拒否すれば中国側に失礼な事になるとして政府からはつきりした入国許可の保証をとりつけてから招待状を出そうと言う慎重な構えを取つてきた。
所が，ニクソン米大統領の訪中決定など，最近の国際情勢の大きな変化に依つて，さとう内閣と言えども，もはや中国代表団の入国を拒否するような態度は取れないだろうと判断し，招待状を出すことに踏切つた。

表6・11・1　変換結果の一例（第1順位のみ出力）

システムの漢字変換能力を見るために次のような測定を行なった。なお測定にあたっては2250を用いた会話型専用ルーチンを作成し，測定の能率と精度を高めるようにした。

サンプル文は朝日新聞昭和46年8月1日～6日の43記事。全分ち書き単位数は7,506で，そのうち変換対象単位は6,842であった。それらを外交から家庭までの7分野に分類統合した形で表6.11.2に示す。

処理速度は，以上43記事をローマ字文でカードから読み取り，漢テレ出力用の紙テープをさん孔し終えるまで約25分かかっている。したがって，分ち書き単位では約0.2秒かかる勘定になる。

表6.11.2の測定結果に関して次のことがいえる。

① N_3の原因の100%，N の原因の約82%が「単語辞書になかったため」となっている。したがっ

て単語辞書の改良という簡単な手段により，まだかなり変換率の改善が可能であるように思える。

② N_4の原因は約18%は最長一致法による辞書検索に伴う原理的な誤変換である。

なお，処理は読込み領域（1Kバイト）がいっぱいになるまで入力文を読み込んでは，それらの漢字変換を一括して行ない，その結果をまとめて出力している。読込み領域に入る分ち書き単位数は一定しないので純粋処理時間の算出は困難であるが，それを平均約60と見ると，読込み終了から出力開始までの平均時間が約5秒であるから，1分ち書き単位当りの純粋処理時間は約0.08秒となる。

対　象	N	$\frac{N_1}{N}$	$\frac{N_2}{N}$	$\frac{N_3}{N}$	$\frac{N_4}{N}$	$\frac{N_5}{N}$	記事数
外交・外政	1376	80.8%	90.6	3.0	3.7	3.0	9
内　政	1421	76.4	92.4	2.0	2.2	3.6	7
経　済	908	76.5	89.0	5.0	3.2	2.9	6
社　会	1116	78.3	91.1	3.4	3.8	1.9	7
スポーツ	876	73.3	86.8	5.3	5.1	3.0	6
科　学	448	78.7	92.2	3.2	3.2	1.6	3
家　庭	697	76.9	86.7	5.6	5.2	2.6	5
全　体	6842	77.5	90.1	3.7	3.6	2.8	43

N ： 変換対象分ち書き単位数
N_1： 一意的に変換された単位数
N_2： 正しいのが第1順位にきた単位数（N_1 を含む）
N_3： 無変換であった単位数
N_4： 誤変換された単位数
N_5： 正しいのが第2順位以上にきた単位数

表6・11・2 変換能力の測定結果

6.11.5 今後の課題

実用的で，しかもコンパクトなカナ漢字変換システムとなると，上述のように，対象とする分野をある程度しぼってそれによく適した単語辞書を用意し，それに分ち書き単位内のみの文法分析を行なう，というような形に落ち着くのではないかと思われる。将来の問題点として次のようなことが考えられる。

① 固有名詞処理は入力文に適当なマークをつけることで解決するのが一番手取り早い。欧米諸言語の正書法でも固有名詞は大文字で書き出すことが行なわれているから，それほど負担にならないであろう。さらに地名，人名等の区別をつけることで変換率がどのように変わるかは調査する価値がある。

② 複合語はどのように分ち書きしてもよいようにすることで，入力文作成時の分ち書きの仕方の制限をゆるめること。計算機への漢字かな混り文の入力手段としてこのシステムを見た場合，「読む

通りに入力すればよい」という利点がある。その利点をより一層生かすためにも，複合語処理などを行なって，分ち書きの仕方に残っている制限をゆるめる必要があろう。

6.12 日本語情報処理ユーザー一覧表

前節までは代表的と思われる応用事例についてのべてきたが，そのほかの例を以下の表で示す。この表であげている各事例は，文献調査，アンケート調査ならびに面接調査などをもとにとりまとめたものであるが，記載もれも多数あることをお断りしておく。

（一覧表をみるにあたっての注意）

(1)システム構成の欄で使用している記号は次の機器を示す。

①入力装置関係

F 6801 A	漢字入力キーボード（富士通）		KB 202 B	漢字けん盤さん孔機（谷村新興）	
F 603 F	磁気テープ装置	（ 〃 ）	SCK 200	〃	（ 〃 ）
F 801 D	コンソールタイプライタ	（ 〃 ）	SC 4000	漢字入出力装置	（ 〃 ）
F 804 D	データライタ	（ 〃 ）	C 5110	漢字入力装置（日電）	
F 749 A	紙テープ読取装置	（ 〃 ）	N 109 A	紙テープリーダ	（ 〃 ）
TS- 3100B	ディスプレイターミナルステーション	（日本電子産業）	OKIGT 3303	紙テープリーダ（沖）	
TS- 3100E	〃	（ 〃 ）	OKIGE 6148	コンソールタイプライタ	（〃）
TS- 103	〃	（ 〃 ）	OKIGL 5800	漢字けん盤さん孔機	（〃）
OSK	オプティカル・スキャンニング・リーダ	（ 〃 ）			

②出力装置関係

JEMシリーズ			IBM 2680	写真植字装置（IBM）	
PT- 101, 150	マイクロフイルムプリンター	（日本電子産業）	OKIGR 100A	高速テープさん孔機（沖）	
PT- 200, 210, 220,	フイルムプリンタ	（ 〃 ）	OKIGJ 7301	高速ラインプリンタ	（〃）
PT- 700, 750	ソフトプリンタ	（ 〃 ）	F 6802	漢字入出力キーボード（富士通）	
PT- 800	普通紙ソフトプリンタ	（ 〃 ）	F 6502A	漢字ラインプリンタ	（ 〃 ）
DM- 110	放送用ディスプレイモニター	（ 〃 ）	F 6521A	漢字プリンタ	（ 〃 ）
JEMH- 10T	編集用コンピュータ	（ 〃 ）	F 6531C	全自動写植装置	（ 〃 ）
SAPTONP-MA	写真植字装置（写研）		F 401A	マガジンファイル装置	（ 〃 ）
SAPTON-A	写真植字装置	（ 〃 ）	F 6501B	漢字プリンタ装置	（ 〃 ）
SAPTON-N	〃	（ 〃 ）	F 767A	紙テープパンチ	（ 〃 ）
C 5210	高速漢字プリンタ（日本電気）		T 4100	漢字プリンタ（高千穂）	
C 5220	漢字プリンタ	（ 〃 ）			
C 5310	漢字ディスプレイ	（ 〃 ）			
N 110A	紙テープパンチ	（ 〃 ）			

なお周辺機器は日本語情報処理のために使用するものについてのみあげている。

(2)文字種の数は英数字，記号等を含めたもの。また（　　）内は漢字のみ。

(3)書体の欄でドットマトリクスの（　　）内はドットの数を示す。

(4)ソフトウエアの欄は日本語情報処理システムのために特に開発したものについてあげてある。

(5)備考欄のうち(1)は稼動状況であり，(2)は将来の構想である。

(6)なお時間的な都合で一覧表に収録できなかったが，この他，日本語情報処理システムを導入しているユーザには次のようなものがある。

ユーザ名	適用分野	システム・機器
日本経済新聞社	新聞の編集，版下作成	IBM 2680
神奈川新聞社	〃	SAPTON-N
高知新聞社	〃	〃
日本海新聞社	〃	〃
北日本新聞社	〃	〃
サンケイ新聞社	箱組，ラテ番組作成	FACOM 230/10D
大日本印刷㈱	一般商業印刷物の組版，版下の作成	FACOM全自動写植システム
ダイヤモンド社	雑誌の編集，組版，版下の作成	SAPTON-P
黒船印刷	印刷物の組版，版下作成	SAPTON-A
†三菱商事	人事情報検索のプリント	T 4100
†丸紅	〃	〃
†三菱電機	〃	〃
†三井金属	〃	〃
†国税庁	名簿のプリント	〃
キャノン	在庫管理，個人記録	キャノンJCOM
高千穂計算センター	計算サービス	T 4100
†特許デイタセンター	工業技術情報の検索プリント	F-FCL, T 4100, JEM 3100

（注）†は外注による処理

（事務処理関係）

ユーザ名	住友信託銀行	総理府統計局	東洋信託銀行
システム名	漢字情報処理システム		証券代行業務システム
適用分野	証券代行業務の宛名印刷および各種帳票作成	国勢調査、各種統計資料の作成	宛名印刷、各種帳票作成
システムまたは業務の概要	株主管理として所有株式・株数・地域・税金の扱い、配当金支払方法等の各種データのほか住所・氏名などの漢字情報を電算機処理し、宛名印刷および管理用帳票を作成する。	統計表で使われている文字、表題、表側けい線などの情報を入れた磁気テープをあらかじめ作成する。このテープとメイン・コンピュータで処理された統計データ(数値)をマッチさせて、各種統計表をマイクロフイルム、印刷用フイルム、ゲラ刷用に出力する。	東京本店証券代行部および、大阪支店証券代行部において委託株式の代行業務として、 株式総会招集通知 配当金領収書 新株申し込み書 などの作成を漢字によって行っている。
システムの構成 入力機器	漢字テレタイプ		F 6801A×4
システムの構成 出力機器	JEM3100（PT-150×1、PT-800×1）	JEM 3820（PT200×1,PT700×1、PT100×1）	F 6502A×4
システムの構成 メイン・コンピュータ	IBM 360、370×1	NEAC 2200/400×1、IBM 360/40×1	FACOM 230/25×2
システムの構成 周辺機器	コーデル400、copyflol1		MT×12、ドラム×2、ディスク×2
文字種 入力用	約8,500字	2,688字(2,260字)	5,376字
文字種 出力用	約8,500字	2,688字(2,260字)	5,285字(5,174字)
書体	中ゴシック	明朝	ドットマトリクス(15×18)
ソフトウエア			
備考		(1)入力：約100万字/年、6時間/日 出力：730ページ/年、8時間/日（1ページ1万字） (2)データバンクの基礎資料作成に積極的に利用していきたい。	

ユーザ名		日立製作所	三菱信託銀行	明治生命保険
システム名		日立ドキュメンテーション・エディットシステム	宛名印刷	漢字処理システム（I）
適用分野		更新のあるドキュメントの組版／更新	証券代行業務	顧客氏名の印刷、統計資料の作成
システムまたは業務の概要		マニュアル等のドキュメントの組版、更新のためのシステムで組版機能、リスト出力機能、更新機能から構成されている。組版機能では漢字けん盤せん孔機の出力紙テープよりマスタ用磁気テープをつくる。リスト機能では、マスタテープから出力される初校リスト、版下リストの印字、トランザクションテープより、2校・3校リストの印字を行う。更新機能では、紙テープにパンチされた修正情報をいったんトランザクションテープに格納して、マスタテープの更新をおこなう。	名義書換により発生した株主住所氏名の変更訂正（含新規株主）IBM29型穿孔機（漢字用）を使ってパンチし、IBM 370/155に入力して期末株主の住所氏名を確定する。 確定後の住所ファイルをIBM 2245漢字プリンターの入力として株主宛発送物（委任状、配当金領収証等）の宛名印字を行なう。	数百万件にのぼる契約者について、 配当金支払明細書 課税所得控除 保険料証明書 各種管理資料 などの氏名記入を一括　字プリントによって行なっている。
システムの構成	入力機器	新興製漢テレ×10	IBM29型穿孔機（漢字用）×1	F 603 F×2
	出力機器	JEM3800（PT-700×1、PT-200×1）	IBM2245型漢字プリンター×1	F6502A×1
	メイン・コンピュータ	HITAC 8210×1	IBM 370/155×1	FACOM 230/25×1
	周辺機器	MT×6、ディクス×2	MT DISK	ドラム×1
文字種	入力用	2,400字(1,900字)	3,600字	7,200字(3,840字)
	出力用	2,938字(1,900字)	10,000字(Max)	7,200字(約7,000字)
書体		明朝、ボールド、ゴシック	ドットマトリクス（18×22）	専用書体
ソフトウエア		DOPS（管理プログラム）、HDES（ドキュメンテーション用プログラム）	ASSEMBLER, COBOL のUSER program とIBM MACRO INSTRUCTIONを使い通常のOS下で稼動	BOS 漢字ユーティリティ（磁気テープをプリンターに出力させるプログラム）
備考			(1)システム総合Test中 入力は大半外注パンチ委託	(1)50時間(2000件)/月 年2回300時間(120万件) 4半期10時間(2,000件) (2)契約物の印刷への応用 名寄せ等による顧客情報システムへと発展させたい。

（印刷・出版関係）

ユーザ名	明治生命保険	旭電算製版センター	オール出版印刷㈱
システム名	漢字情報処理システム(Ⅱ)		
適用分野	保険契約内容照会 各種経営情報のプリント	印刷の組版処理 版下作成	新聞および一般商業印刷の組版
システムまたは業務の概要	メイン・コンピュータで処理された各種経営情報、顧客管理情報を漢字プリンタで印字する。 また保険契約内容の照会応答はオンラインで接続された漢字ディスプレイ(TS-3100E、B)でサービスされている。	印刷会社4社協同で設立された写植センターである。原稿をセンターに持ち寄り漢字キーボードにより書体、文字の大きさ等指定して紙テープにさん孔し原テープを作成する。校正テープを入力して組版処理を行ない、完全テープを作成して写植機にかけ、印刷用の版下として出力する。 （なお計算機処理は外注である。）	業界紙などの新聞の印刷を主体としてきたが、これと並行してオフセット印刷による一般商業印刷の比率を高めようとしている。 新聞の組版作成にはサプトンN・システムを用いて出力する。この場合見出しなどは手動の写植機を用いている。大組みは貼込みを人手により行なっている。一般商業印刷ではちらしやカタログ、雑誌などがあり、サプトンP・システムで出力する。この場合もページアップは人手により行なう。
システムの構成 入力機器	TS-3100E×8 TS-3100B×8	SCK-200D-3A×8	漢字けん盤さん孔機×15
システムの構成 出力機器	JEM3100S(PT-750×1)	SAPTON-A-5260×1	サプトン-P、サプトン-N
システムの構成 メイン・コンピュータ	IBM360／40×1 IBM360／60×1		HITAC-10(内蔵)×1
システムの構成 周辺機器	MT×1、ドラム×1		コダック自動現像機×1
文字種 入力用	2,688字(2,000字)	2,880字	2,790字
文字種 出力用	7,847字(6,451字)	3,072×2字	3,344字
書体	明朝、ゴシック	明朝、ゴシック	明朝、ゴシック
ソフトウエア			サプコール（新聞組版、一般印刷組版用）
備考	(1) 250時間／月	(1) 写植機 約2時間／日 なお現在のところ試行段階である。	(2)将来はかなりの部分を自動写植に移行させたい。しかしシステムの柔軟性の問題から、それが全面的にホットタイプにとってかわることはないだろう。

ユーザ名		学習研究社	竹田印刷㈱	千代田コンピュータ印刷
システム名		学研植字編集システム		
適用分野			一般商業印刷の組版	印刷の組版処理 版下作成
システムまたは業務の概要		事典、一般書籍などの当社の出版物の組版、版下作成のほか、他の出版社等からの漢字情報処理の受託サービスを行なう。 また組版等のソフトウエアの開発サービスも行なっている。	ポスター、カタログ、カレンダーなどの一般商業印刷を行なっているが、サプトンシステムでは年鑑、名簿などが主体となっている。 端物でも急ぎの場合、自動写植を用いることもある。この場合の時間的なメリットは非常に大きい。	中堅印刷会社6社の協同組合でキーボードを各社が保有し各社で作成されたインプットテープの編集処理を行なう。
システムの構成	入力機器	KB202B×15、MD−C×1	漢字けん盤さん孔機×8	F6801A×25、F804 D×1
	出力機器	JEM3850(PT−210×1 PT−750×1)	サプトン−A 5260	F749A×1、F6531C×1 FACOM−R×1
	メイン・コンピュータ	IBM 370／145×1 HITAC10×1	HITAC−10(内蔵)×1	FACOM270／20×1
	周辺機器	MT×3、DR×1	自動現像機×1	ドラム×1
文字種	入力用	4,520字(3,885字)	3,000字＋外字	10,752字(7,944字)
	出力用	4,992字(3,885字)	3,000字＋外字	12,288字(9,480字)
書体		明朝、アンチック	明朝、ゴシック	明朝、アンチック、ゴシック
ソフトウエア		①GEMS（漢字処理関係の管理プログラム ②項目50音項目整理（漢字カナ混り文項目を50音シーケンス、分野、情報などの条件で検索する。	サプコール（組版専用プログラム）	PCL−7（印刷版下自動編集プログラム
備考		(1) コンピュータ 約500時間／月 JEM3850 約300時間／月 (2) 書籍等の記事情報をデータバンクとしてファイルし、必要の都度引出して処理することを考える。	(1) 月間2,500ページ 〜10,000ページ (2) 需要が高まっているので、近い将来システムを拡張する計画である。	

ユーザ名	東京ジェーピー	凸版印刷	カシヨ印刷株式会社
システム名			
適用分野	印刷の組版処理 版下作成	一般書籍用文字組版	印刷の組版処理 版下作成
システムまたは業務の概要	主として一般書籍を取扱う印刷企業19社が、協同して設立した電算写植センターである。会員各社は受注した印刷物原稿に組版指示等を行なった上センタに持込む。センタは漢字キーボードにより入力からハードコピーの出力までの処理を行なうほか一般の計算機で処理された磁気テープ等から漢字処理出力を行ない版下作成を行なう。	活字にかわる文字製版を電算機と全自動写植機とを結びつけて行なう。原稿編集作成に当っては ランダム入稿 ソート 版下作成 索引の自動作成 のほか統計処理用のMTより版下作成を行なう機能を有する。またDMのファイルメンテナンスおよび宛名印刷も行なっている。	原稿ならびに字体、文字の大きさ、行送り字詰等の指定をさん孔機でテープにさん孔する。組版用プログラム（SAPCOL－D5）を用いて、このテープをSAPTON－A5260にかけて版下作成をする。 おもに一般書籍、法規等の印刷に使用している。
システムの構成 入力機器	F6801A×40、F804D×3	F6801A×56、F804D×6	新興製漢字けん盤さん孔機（3A型）×8
システムの構成 出力機器	F749A×1、F401A×3	F6531C×6、F6521A×1	SAPTON－A×1
システムの構成 メイン・コンピュータ	FACOM－R×3 F6531×3 FACOM 270／20×1	F270／20×2、F230／15×1	PDP－8／I×1
システムの構成 周辺機器	MT×3、ドラム×2	MT×10、ドラム×3	
文字種 入力用	10,752字（7,944字）	10,752字（7,876字）	6,144字（2,340字）
文字種 出力用	12,288字（9,480字）	12,288字（9,412字）	6,144字（2,340字）
書体	明朝、アンチック、ゴシック	明朝、アンチック、ゴシック	明朝、ゴシック
ソフトウエア	FCL－7（自動組版システム）	FCL－7（印刷版下自動編集） ランダム入稿処理システム 索引自動作成システム	SAPCOL－D5 （容量8Kのミニコンのため組版用のみのプログラム）
備考	(1)FACOM 270／20 350時間／月 写植機 200～300時間／月 (2)ディスプレイ等を使用した校正システムの開発印刷に関連した出力処理サービス		(1)46年7月から稼動 従来通り活版印刷も行なっているが、7割はこのシステムを使用している。 初校まで⅓～¼の時間であがる。

（放　送　関　係）

ユーザ名	日本コンプ	ペティ	東京放送
システム名			
適用分野	印刷の組版処理 版下作成	印刷の組版処理 版下作成	テレビによる選挙速報
システムまたは業務の概要	印刷会社20社を会員とする電算写植センターである。漢字入力キーボードを会員各社が保有し、パンチした紙テープを持ち込み、編集処理を行なうと共にパンチより版下作成まで一括処理サービスも行なっている。また他システムの汎用コンピュータで処理されたMTなどからの出力サービスも行なっている。	一般書籍の文字組版のサービス及び他システムで処理された磁気テープ等の漢字を含む出力サービスを行なう。 また当社は文字組版用ソフトウエアの開発も行なう。	選挙速報等に用いられるもので、漢字放送用ディスプレイモニタ上に表示された漢字情報をカメラに撮りテレビ放送する。入力はターミナルステーションのキーボード、紙テープにより入力する。
システムの構成　入力機器	F6801A×7、F804D×1	F6801A×3、F804D×2	TS－103×1、PR－350×1
システムの構成　出力機器	F749A×2、FACOM－R×2 F6531C×2	F401A×1、F749A×1 FACOM－R、F6531C×1	JEM2410Ⅲ(DM－110×3)
システムの構成　メイン・コンピュータ	FACOM270／20×1	FACOM270／20×1	IBM360／40×1
システムの構成　周辺機器	MT×2、ドラム×2	MT×3、ドラム×2	
文字種　入力用	10,752字(7,944字)	10,752字(7,944字)	2,688字(2,200字)
文字種　出力用	12,288字(9,480字)	12,288字(9,480字)	2,688字(2,200字)
書体	明朝、アンチック、ゴシック	明朝、アンチック、ゴシック	専用書体
ソフトウエア	FCL－7（自動組版システム）	FCL－7（自動組版システム） 字詰編集機能 体裁処理機能 ファイル処理機能	
備考	(2)将来漢字総合処理センターへと発展して行きたい。		

ユーザ名	日本教育テレビ	日本テレビ放送網	日本放送協会
システム名			選挙システム
適用分野	テレビによる選挙速報	テレビによる選挙速報	テレビによる選挙の開票速報
システムまたは業務の概要	選挙速報放送等に使用されているもので、漢字放送用ディスプレイ上に表示された情報をカメラに撮り、テレビ放送する。	選挙速報等の漢字放送用ディスプレイ上に表示された漢字情報をカメラに撮り、テレビ放送を行なう。入力はシステム内のターミナルステーションの漢字キーボードあるいは紙テープから行なわれる。	選挙の際の開票速報（選挙区ごとの各候補者の票数）、党派別当選、当確者など直接漢字でテレビ放送する。コンピュータによって制御される漢字表示装置の画像を直接テレビカメラに撮る。
システムの構成 入力機器	TS－103×1、PR－350×1	TS－103×1、PR－350×1	TS－103×3
システムの構成 出力機器	JEM2410Ⅲ（DM－110×3）	JEM2410Ⅱ（DM－110×3）	JEM2410（DM－110×3）
システムの構成 メイン・コンピュータ	IBM360／40×1 IBM360／20×1	IBM360／40×1 HITAC8400	IBM360／50I×1
システムの構成 周辺機器			
文字種 入力用	2,688字（2,200字）	2,516字（2,200字）	2,496字
文字種 出力用	2,688字（2,200字）	2,516字（2,200字）	2,496字
書体	専用書体	専用書体	丸ゴシック
ソフトウエア			
備考			(1) 選挙の際、約2日間フル稼動

（新　聞　関　係）

ユーザ名	フジテレビ	朝日新聞社	朝日新聞社
システム名		漢字集信システム	NELSON
適用分野	テレビによる選挙速報 野球放送	新聞原稿の受信、モニタプリント、キャスター用紙テープの出力	新聞紙面の編集 組版・写植
システムまたは業務の概要	選挙速報等で漢字放送用ディスプレイモニタ上に表示された漢字情報をカメラに撮り、テレビ放送する。	50BPS、24回線から入力された新聞原稿をF 270 / 20で電文組立編集等を行ない、モニタープリント、キャスター用紙テープを出力する。	原稿を紙テープにパンチし、入力する。DTV画面上で文字の削除、挿入、さしかえなど校閲（校正）を実施する。挿入のため漢字キーボードを連結している。禁則処理は自動的に行なわれる。見出し、前文、箱組み、大組みなどの特殊組みはDTV（または、2740）を使用して実施し、DTV画面上で確認する。大組み終了後、写植機2680で打出し、あとは合せ写真やカット類の込みで版下を作りネガフィルムとする。
システムの構成 入力機器	TS－103×1	F 749 A×1 24通信回線／1システム	SCK 200
システムの構成 出力機器	JEM2410(DM－110×3)	F 6501 B×6、F767A×9	IBM2680電子写植機
システムの構成 メイン・コンピュータ	IBM 360／40×1	FACOM 270/20×6	IBM360／40H(2040H)
システムの構成 周辺機器			IBM4510(DTV)×5 IBM2740×2
文字種 入力用	2,496字(2,200字)	2,304字(2,304字)	約2,300字（キーボード）
文字種 出力用	2,496字(2,200字)	2,304字(2,304字)	約45,000字
書体	専用書体	ドットマトリクス（15×16）	明朝、ゴシック
ソフトウエア		モニタ（システム全体の制御） 入側処理（入力制御、禁則処理外字変換） 出側処理（紙テープパンチ、漢字プリンタ装置への出力）	ディスプレイ上での校閲、訂正プログラム ディスプレイ上での編集 レイアウトプログラム タイプライタ式入力機（2740）での上記プログラム
備考		(1)CPU　18時間／日 入出力　15,000行／日 （1行15字）	(1)1971年11月から稼動、開発作業を主として実施中、現在はパイロットプランである。

ユーザ名	上毛新聞	中日新聞社	日刊スポーツ印刷社
システム名		漢字集信システム	NEPS
適用分野	新聞の組版処理、版下作成	新聞原稿のオンラインによる受信、新聞原稿のモニタプリント	新聞案内広告の組版
システムまたは業務の概要	原稿および書体、文字の大きさなどの指定をさん孔機でテープにさん孔する。 モニタに打ち出して、校正を行ないながらSAPTON 5265でわく組みをつくる。 小さいわく組みはSAPTON 5265で行なうが大組のレイアウトは人間がはりつける。	50, 200 BPS 回線で送られて来た新聞原稿をF 270-20で電文組立、編集等を行ないモニタープリント、紙テープパンチを出力する。	各新聞社の案内広告欄の版下を作成し、凸版、紙型の形で新聞社へ渡す。小組・校正過程は8単位2フレーム一字とし、紙テープに入力、組版しゲラとしてクイックペーパーに出力する。またそのペーパー上で校正する。校正部分を紙テープで再入力し、マスタテープの更新を行なう。大組、版下過程は総段数をカードによってインプットし、小組・校正過程の出力データをマッチングして印刷用テープを作成する。このテープをJEM 3850にかけてポジフイルムを出力する。
システムの構成 入力機器	沖製漢テレ SCK 200 D×4、2340×4	F 749 A×1, 50 B, 200 B通信回線	新興製漢テレ×8
システムの構成 出力機器	SAPTON 5265 × 1	F 6501 B× 2	JEM 3850(PT 750×1,PT200×1, PT 220×1)
システムの構成 メイン・コンピュータ	HITAC- 10 × 1	F 270/20 × 1	OUK 9200 × 1 (32 KB)
システムの構成 周辺機器			MT× 4, ディスク× 3, OSR × 1
文字種 入力用	2304 × 2字	2, 304字 (2, 304字)	約5, 000字
文字種 出力用	2304 × 2字	2, 304字 (2, 304字)	約5, 000字
書体	明朝、ゴシック	ドットマトリクス (15×16)	明朝
ソフトウエア		モニタ(システム全体の制御) 入側処理 (入力制御) 編集処理 (禁則処理、外字変換)	小組・校正用プログラム 大組用プログラム ファイルメンテナンス用プログラム
備考	(1)48年8月から、SAPTON 5265稼動 人員確保、技術の短期習得、職場環境の改善等の理由により、従来の活版印刷から切りかえた。		(1)CPU 350時間/月 出力 50段/日 (2)案内広告のCTS化完了後、一般記事の処理へ拡張して行きたい。

ユーザ名	毎日新聞社	毎日新聞社	山梨日々新聞社
システム名	漢字コード変換システム	Mainichi Automated Composing system	
適用分野	新聞の原稿編集、組版処理	新聞の原稿編集、組版処理、版下作成	新聞原稿組版処理、版下作成
システムまたは業務の概要	通信社配信原稿の処理上、必要な漢字コードの変換（例えば共同通信社 CO-59 から毎日コードへ）をプログラム制御により自動的かつ、高速度で行なう。その他本社内ホットタイプコードからコールドタイプコードへの変換、海外用コードから国内用コードへの変換字詰変更および削除、加筆などの編集処理もあわせて行なう。	コンピュータと自動写植機を組み合わせて、組版処理の高速化と高能率化を目的としたシステムで、将来の大型システム導入のための研究システムの役割りを担うとともに、案内広告、特集面などの部分的実用を行なう。	原稿を紙テープにパンチし、テープリーダからシステムに入力してモニタをとり、校正箇所を再度入力して、校閲ずみテープを作成し、写植機にかけて印刷用の版下として出力する。 なおシステムの空き時間を利用して名簿、事典等の印刷も行なっている。
システムの構成 入力機器	ディジトロニクス社製 PTR 2540 × 1	OKIGT 3303 × 2, OKIGE 6148 × 2, OKIGL 5800×6	SCK-200D× 7
システムの構成 出力機器	ディジトロニクス社製 PTP 1560 × 2	SAPTON-PMA×2, OKIGR 1001 A×1, OKIGJ 7301 × 1	SAPTON-A-5265 × 2
システムの構成 メイン・コンピュータ	TOSBAC 40 × 1	OKITAC 7000 × 1	HITAC-10 × 1
システムの構成 周辺機器		MT× 3	ドラム× 1、DR× 1
文字種 入力用	2,314 字（2,036 字）	2,396 字	3,376 字（1,974 字）
文字種 出力用	2,400 字（2,020 字）	3,072 字	3,376 字（1,974 字）
書体		明朝、専用書体	明朝、ゴシック
ソフトウエア	MCCS-71-40（各コード間の変換、字詰変更プログラム） MCCS-72-40（ラジオ・テレビ番組の編集処理）（行単位の削除、挿入、1字単位の差替え）	AADP 01～91（案内広告、組版） APTP 01～99（ラテ番組組版） ASTP 01～91（本文組版）	SAPCOL－H3 （印刷版下自動編集プログラム）
備考			(1)A.M.10：00～PM.11：00 47年5月より全紙面写植化。 (2)選挙速報を完全に電算植字システムにのせたい。

（情 報 検 索 関 係）

ユーザ名	読売新聞社	外務省	国立国会図書館
システム名	高速漢字通信処理システム		国会会議録索引システム
適用分野	新聞原稿の集配信処理	国会議事録、調査等の情報検索	国会会議録の索引誌編纂および検索
システムまたは業務の概要	50および200ＢＰＳ通信回線で送られてきた新聞原稿の情報をＦ270/20で電文組立て、編集処理等を行ない紙テープパンチモニタ出力を行なう。なお大阪本社では自動変換システムを有し、ＦＡＣＯＭ－Ｒで電文の選択、集配信を行なっている。	外務省関係の国会議事録、調査等に関する情報をバッチ処理で検索サービスする。あらかじめ現物探索の手がかりとなる内容インデックス、書誌的事項を表わす額表インデックス、資料の内容を表わすサマリー、在り場所を示す資料アドレスをファイルしておく。検索はその内容ならびに額表インデックスをもとにして行ない、資料アドレスとサマリーを出力する。	(1)国会会議録索引編纂システム ・入力データは漢字テレタイプにより紙テープにせん孔 ・この紙テープを磁気テープに変換し文法チェックやリードを行なって基本ファイルおよび辞書ファイルを作成する。 ・この二つのファイルから編纂用データを作成し、漢字プリンタで出力する。 (2)国会会議録索引検索システム ・あらかじめ定めた12項目の質問事項を組合せまたは否定やマスク方式により検索データを入力する。 ・基本ファイルと質問ファイルを照合して所要の回答データをラインプリンタまたは漢字プリンタに出力する
システムの構成 入力機器	スキャナ装置×2、F 749×2	沖製漢テレ×4 新興製漢テレ×1	新興製漢テレ×2
システムの構成 出力機器	F 6501 B×2、F 767 A×12	JEM3100（PT－750H×1）	JEM3800B（PT－200×1 PT－700×1）
システムの構成 メイン・コンピュータ	FACOM 270／20×2	HITAC8400×1	HITAC8400／65×1
システムの構成 周辺機器		MT×10、ディスク×1、TRP×1	MT×6、ディスク×4、TRP×1
文字種 入力用	2,304字＋α	4,103字（3,611字）	2,496字（1,968字）
文字種 出力用	2,304字＋α	4,103字（3,611字）	5,028字（3,965字）
書体	ドットマトリクス(15×16)	明朝、ゴシック、センチュリーカタログ	明朝、ボールド、ゴシック
ソフトウエア	漢字プリンタ出力プログラム電文組立て、編集処理プログラム、自動変換プログラム	校正ストレージ、検索、インデックス修正、索引法関係、標準漢字プリントのプログラム	索引編集システム（テーブル作成、辞書作成等のプログラム46本から構成される。） 情報検索システム（試験的段階であるが、プログラム3本からなる。）
備考		(1)ＣＰＵ、平均2時間／日 利用状況、約10件／日 データ量、500件／日 (2)オンラインによるＩＲシステムの編成と適用業務を拡大していきたい。	(1)入出力、994,000字／年 ＣＰＵ、約455時間／年 (2)目録編纂の機械化、さらには文献情報サービスに発展させて行きたい。

ユーザ名	通産省	内閣官房内閣調査室	日本科学技術情報センター
システム名	MITI情報検索システム	内閣調査室総合情報管理システム	
適用分野	一般経済情報記事、情報処理ニュース、法律条文などの検索	文献検索、記事情報検索	文献速報の作成 漢字情報による情報検索
システムまたは業務の概要	漢字を含んだ対象情報を自然文のまま磁気テープに蓄積しておき、利用者の需要に合わせてこれらの中から必要なものだけを電算機を利用して検索提供する。この検索の際は漢字CRTディスプレイ装置を利用して電算機との会話を進めながら、質問の変更、検索結果のディスプレイ上への表示、プリンターへの印刷などの処理ができる。	情報システムで、質問データを漢字入出力端末装置よりオンライン入力する。検索結果は質問を投入した漢字入出力端末装置あるいは漢字ディスプレイのいずれかの装置に出力する。このシステムで使用するデータは、すべて日本語（漢字）の自然文の形で記述されている。	内外の科学技術に関する文献速報とこれらの索引等の発行を主な業務としている。 文献速報の発行に当っては、まず抄録原稿を漢テレ等で紙テープにした後、漢字プリンタで出力して、校正等のための修正作業を行なう。次に分類項目毎に配列、重出などの編集を行ない、MTに蓄積しまた漢字プリンタからフイルムにプリントアウトし、印刷原版とする。 索引の作成に当っては、MTに蓄積された抄録をもとにして、著者別、国別、雑誌別、キーワード別などの索引を作る。 これらの蓄積は情報検索用のファイルとなり、種々の検索サービスに利用されている。
システムの構成 入力機器	新興製印字さん孔タイプライタ×2	SC4000×6、C5110×1 C5210×1、C5220×1 C5310×1、(SC400×6)	新興製漢テレ×33 データライタ(英・露文)×10
システムの構成 出力機器	JEM3100(PT－750A)		JEM3100(PT－200×1 PT－700×1)
システムの構成 メイン・コンピュータ	NEAC－2200／400 HITAC－10	NEAC2200／500×1	FACOM230／50×1 JEMH－10J×1
システムの構成 周辺機器		MT×7、ディスク×8	MT×2
文字種 入力用	2,571字(2,169字)	3,960字(3,350字)	2,496字(1,861字)
文字種 出力用	2,571字(2,169字)	3,960字(3,350字)	3,071字(1,861字)
書体	明朝	ゴシック	明朝
ソフトウエア	簡単なキーワードの組み合わせで必要な情報が検索できる検索プログラムなど。	データ収集、管理プログラム 検索プログラム 回答の編集出力プログラム	文献速報作成、索引作成 情報検索用システムの各種プログラム
備考	(1)蓄積情報(7種類)の更新約20万字／日　必要に応じて検索サービスを行なっている。 (2)現在は検索業務が中心であるが、来年度より磁気テープ装置、フイルムプリンタの使用が可能になるので電子計算機による集計表を印刷するシステムなども開発する。	開発中	

（漢字処理総合センター）

ユーザ名		防衛庁	朝日ビジネスコンサルタント	デンショク
システム名				
適用分野		情報検索と印刷	漢字処理システム用入力サービス	漢字情報処理サービスと一般商用印刷版下作成
システムまたは業務の概要		庁内の各種資料、速報等の印刷、検索出力を目的とするもので、メイン・コンピュータで検索編集処理されたプリント情報を磁気テープに出力する。この磁気テープを入力として、PT－700 ソフトプリンタで印刷プリントし、庁内に配布している。	出版社等の漢字処理を行う顧客からの原稿をラインプット装置ならびに漢字けん盤入力装置によりパンチし、さらに顧客の要望するコードに変換しMTで渡すサービスである。なおエラーチェックのためあらかじめ外部センタの漢字出力システムによりモニタ出力を行ない原稿とのチェックを行う。	汎用コンピュータで打ち出されたデータをもとにして、各種統計表およびダイレクトメール、請求書送り状等の宛名印刷のサービスを行なう。また印刷会社の利用センターとして会員制によって頁物の印刷の版下を作成サービスを行なっている。
システムの構成	入力機器	神漢字キーボード	ラインプット×10、SC－59×4	
	出力機器	JEM3800（PT－700×1）		JEM3850（PT-220×1、PF-750×1）
	メイン・コンピュータ	HITAC－8400×1	FACOM230／25（コード変換用）	JEM H－10J×1
	周辺機器		MT	MT×3、ドラム×1
文字種	入力用	2,496字（1,900字）	12,000字	
	出力用	2,496字（1,900字）		16,446字（10,450字）
書体		明朝		明朝、ゴシック
ソフトウエア				
備考				

（そ　の　他）

ユーザ名	東洋漢字処理センター	ペ　テ　ィ	大倉商事
システム名			和文自動組版ディジセット ビデオコンプシステム
適用分野	FACOM漢字処理システム用入力サービス	FACOM漢字処理システムの出力	出版印刷、新聞検索、文献検索、記事情報等
システムまたは業務の概要	FACOM関係の各機種の入力用テープの受託作成のサービスを行なう。	FACOM漢字処理システムの総合センターであり証券代行業務、ダイレクトメールサービス、人事管理業務等の受託サービスを行なう。	社内の印刷物の版下作成を行なっている。また各部に対しオープン使用の形で新聞制作、出版印刷の版下作成文献検索、記事情報等のサービスを行なっていると同時にディジセットとビデオコンプの製品の販売を行なっている。
システムの構成　入力機器	F6801A×38		漢字さん孔機×1
システムの構成　出力機器		F 6502 A×1、F 6521 A×1	ディジセット×1、ビデオコンプ×1
システムの構成　メイン・コンピュータ		FACOM230／25×1	HITAC8400×1
システムの構成　周辺機器		MT×5	MT×6
文字種　入力用	10,752字(7,944字)		制限なし
文字種　出力用		4,800字(約4,200字)	7,000字(6,000字)
書体	明朝、アンチック、ゴシック	専用書体	明朝、ゴシック
ソフトウエア		各種専用システム	汎用組版プログラム 禁則処理、校正等プログラム
備考			

ユーザ名		共同通信社	国立国語研究所	ラインプット社
システム名		ローマ字、漢字かな変換処理システム		日本語記録システム
適用分野		翻訳	用語用字実態調査研究	会議録作成、原稿編集
システムまたは業務の概要		変換処理システムと辞書作成処理システムから構成されている。辞書作成処理では変換処理で使用する各辞書、テーブル類を作成、保守更新を行なう。変換処理は打電してくるローマ字電文を単語辞書、入力変換辞書、文法表などのファイルを参照して漢字かな混じり文に変換する。	用語用字実態調査等を行なうために、新聞、雑誌などに使用されている用語用字研究資料を蓄積し、そのデータを国語研究の分析に利用している。	原稿（録音、速記、手書、印刷物を含む）から高速入力用に工夫された紙テープパンチ（ラインプットシステム）により紙テープを作成（120字～200字／分）し、校正ベリファイをした後、印刷物にする作業を行なうシステムである。
システムの構成	入力機器	N 109 A－1×1	沖製漢テレ×8	ラインプット×10、東芝電磁誘導タブレット×1
	出力機器	N 110 A－3×1	〃	東芝漢テレファックス×1
	メイン・コンピュータ	NEAC2200／15013×1	HITAC 3010×1	TOSBAC40×1
	周辺機器	MT×4、ディスク×1	MT×6	
文字種	入力用	112字(0字)	2,400字(2,109字)	12,000字(11,000字)
	出力用	2,304字(1,976字)	2,400字(2,109字)	3,000字(2,750字)
書体			明朝	明朝
ソフトウエア		電文変換プログラム 辞書作成サブシステム(辞書、文法表の作成、データの入替え等)	用語用字の統計調査プログラム用語調査の自動化のプログラム 用例表（KWICなど)作成のプログラム	コード変換、入力管理プログラム、校正管理プログラム
備考			(2)日本語の広範囲な用語用字調査および言語情報処理の基礎的研究	(1)入力　20万字／日 (2)速記、出版界の需要に対応できるようにしたい。

7. 日本語情報処理の開発動向

7　日本語情報処理の開発動向

7.1　情報化時代と日本語情報処理

7.1.1　データ通信と計算機ネットワークの時代

電気通信網がデータ通信に使用され，だれでもが種々の端末機器を通信網に結合できるようになった。とくにこの電気通信網に電子計算機を結合することによって任意の場所から電子計算機を利用し，また電子計算機どうしがお互いに情報を交換しあうことができる時代になった。そして各種の情報の交換がますます盛んとなってきている。このようなシステムが構成されると各地に存在する計算機をあたかも自分のもののように使用することができ，それぞれのファイルも利用できるいわゆるResource Sharing System（情報資源共有システム）が実現するわけである。

現在のアメリカではこの種のシステムが各地に作られつつあり，米国全体にわたって最も大きな規模のものとしてARPA（Advanced Research Projects Agency）networkという計算機ネットワークが作られている。これは米国全土にちらばった30カ所あまりの計算機を相互に結合し，お互いに高速通信回線を介して使用しあえるようにしたシステムで，さらに同種のシステムがもう一つ作られようとしている。またイギリス，カナダ，フランス等でもこのような目的の計算機ネットワークが検討されている。さらに遠隔地から計算機にアクセスできるシステムとしては，航空機の座席予約システム，在庫品に関するIBMのAAS（Advanced Administration Systemの略）があり，これらは世界中に広がったシステムである。

日本ではこの種のものとしては，国鉄・日航の座席予約システムやホテル予約，バンキング，職業紹介，その他多数のオンラインシステムが存在するが，計算機ネットワークはまだ存在しない。しかし通信回線の計算機をつないだ計算サービスはすでに行なわれているので，遠からず計算機ネットワークが形成され，また在来の電話・電信とデータ通信を統合した総合通信網が整備されるものと期待される。

情報の蓄積を1カ所の大規模な計算機で集中し処理するやり方がよいか，あるいは分散して各所の端末にかなりの能力をもたせるのがよいかは議論のあるところである。しかし計算機ネットワークが発達し，そのためのソフトウェアもすぐれたものができてくれば，処理能力は分散されていくだろう。そして各所の情報処理装置はそれぞれ特色をもち，相互利用が発達するものと思われる。

7.1.2　情報の各種形態

このような総合通信網が整備されてくると種々の情報が伝送され，処理され，交換されるようになる。情報としては，音声，文字，図形などのほかにも，心電図波形のような医学データの送信などあらゆる情報が考えられる。しかしこれらのうちで最も広く用いられるのはいうまでもなく自然言語であり，我々の場合は日本語である。

言語を情報形態からみると次のようにわかれる。

(a)　音　声（アナログ波形）

(b) 文　字（ディジタル符号）

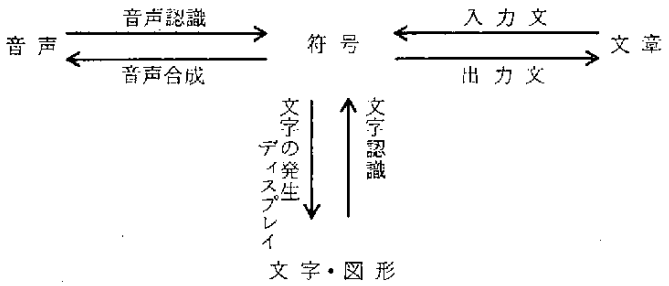
(c) 文　字（手書き文字，波形）

(d) 文　字（TV走査，波形）

そしてこれらの情報形態に対する入出力の機器としては次のようなものがある。

(a) 音　　声	電話，マイクロフォン，スピーカ，インターホン
(b) （符　　号）文字	タイプライタ
(c) （手 書 き）文字	座標入力装置，XYプロッター，CRT出力
(d) （書かれた）文字	図形走査装置，TV出力，CRT出力

しかしほんとうに重要でむずかしい問題は，これらの情報相互間の変換であって，これは次のような形に表現される。



従来の通信は符号の伝送，音声の伝送，図形の伝送といった同一の情報形態間での伝送であったのに対し，これからは文字で入力したものを音声に変換して聞かせたり，文字を自動的に読みとって符号にするなど，異種の情報媒体内での情報の変換がさかんになっていくことはまちがいない。とくに文字読取りはかなりの進歩をし，数字，アルファベット，カタカナなどの自動読取りは実用の域にはいりつつある。

言語が主役を占める使われ方としては主として次のような分野がある。

(a) 通　　信

(b) 情報検索，自動抄録

(c) 機　械　翻　訳

(d) 質問応答システム

(e) 言語による教育

これらが計算機のもつ強力な処理能力と巨大な情報記憶能力とによって今後どのように発達していくかは，これからの情報化社会において重要な意味をもつ。しかしこのようなシステムが十分に発達するためには，情報の機密保持が重要な問題となる。情報がすばやく取り出せること，情報の機密が十分に保持されること，情報内容が正確なこと，情報の更新が容易なことなどの問題がさらに研究されねばならない。しかし大量の情報を蓄積してもそれらがどれほど有効に利用されるかには疑問もあ

る。

このような各種の応用において，日本語がどのような形で使用されるかを予想することはむずかしい。どのような入出力機器が開発されるかによってかなりその形態がかわってくることは当然予想される。しかしながら漢字の入出力のむずかしさは将来とも依然としてのこるであろうから，漢字かなまじりの普通の日本語文がどこまでも用いられるということはありえない。現在，電話計算サービスで行なわれているように，プッシュホンから記号入力をし，答を合成音声で得るというような，いろいろな形態の組合わせが考えられる。これらの組合わせとしては次のようなものが考えられるであろう。

入　　力	出　　力
数字・記号	電話音声応答
	テレビ電話への文字ディスプレイ（カタカナ）
	カナ文字出力
	漢字カナまじり文出力
カナ文字入力	カナ文字出力
	漢字カナまじり文出力
	文字CRT出力
	音声出力
アルファベット入力	アルファベット出力
	カナ文字出力
	漢字カナまじり文出力
漢テレ入力	漢テレ出力
	写真植字出力

このように入出力関係は種々の組合わせが考えられ，従来のように入力と出力とは同一情報の機器であるといった概念はなりたたなくなっている。入力を行なう場所と出力をする場所とが異なっている場合は当然入出力機器は異なるし，たとえ同じ場所であっても性質の異なった入出力機器が一つにまとめられる場合は多い。漢字テレタイプライタが広く用いられるためには価格の低廉化が大切であるから相当な年月が必要であろうし，たとい使われたとしても，格段に安い数字，記号，アルファベット，カナ入力機器の方がはるかに多く用いられるであろう。すなわち，カナ文字による入出力，情報のやりとりは当分つづくものと思われる。

7.2　入力システム

人間が直接手でタイプライタを打つなどの手段で情報を入力することは，たとい文字の自動読取りが発達しても，つねに存在するものである。この場合に日本語の文字システムとアルファベット文字システムとの間の本質的な差があらわれることになる。すなわちキーの数が数十個ですみ，単語がわかち書きされるシステムと，キーの数が数百におよび単語がわかち書きされないシステムの情報処理

における根本的な差である。これにくらべると8単位符号と16単位符号との差はそれほどむずかしい問題ではない。αNシステム（アルファニューメリック・システム）と日本語文字システムの差は今後の情報処理システムの発展に根本的な影響を与えずにはおかないだろう。

この困難を克服するためにとりうる道はまだまったく未知であるが，一つは日本語をカナ文字，分かち書き表現すること，他は漢字入力けん盤を非常に使いやすく，しかも安価に作ることであろうが，どちらも非常にむずかしい問題である。その中間としてカナ入力・漢字出力のためのカナ漢字変換方式の開発であるが，これも簡単ではない。

7.2.1 日本語文字に対する符号の統一

日本語で使われる文字は片仮名，平仮名，漢字，英語，アルファベットに至るまで一万字以上となる。これらに対していまだに符号が統一的に与えられていないことは大きな問題である。その結果，ある漢字テレタイプライタによって入力されたデータは他の機器で出力ができず，データ，プログラム等の互換性に問題が生じている。現在この問題に対してはJISを作るべく努力がはらわれているが，理論的でシステマティックな方法はないようである。符号の統一における第一の考え方は仮名，漢字，アルファベット等に対して12ビット，あるいは16ビットのコードを対応させることである。したがって漢字テレタイプライタを作る際にはけん盤内の漢字としてなにをとるかは自由であるが，出力のテープ上のコードはつねにこの統一コードが出てくるようにコード変換装置をつけねばならないことになる。文字のストローク数による番号づけ，偏，傍などによる配列などの試みがあるが，既存の装置などのことも考えると問題はむずかしくなる。中国では漢字に4けたの数字を割当てている（四角號碼）ようであるが，それとの関係も考慮する必要があろう。

さらに注意をしなければならないことは将来の略字化，当用漢字の数と種類の増加の可能性，新しい漢字の導入の可能性に対しても漢字のコード化の問題が矛盾なく解決されねばならない。したがって，もし偏と傍で符号化するとすれば，たとえば「さんずい偏」のある文字には一連の番号をうち，いくつかの「さんずい偏」の新文字のためにつづき番号をあき番にしておくといった考慮が必要となる。

7.2.2 けん盤入力

カナタイプの入力は英語のタイプライタと同じ程度の努力で習得でき，熟達すれば紙に手書きをしているよりは速い。しかし広く普及するというところへはとうてい行きそうもない。それは日本語が文節単位に区切られないため「かな」では非常に読みづらいからであり，漢字のもつコンパクトさと意味表現の能力が要求されるのは当然である。出力もこれまではカナだけでがまんしてきたが，これからはますます漢字出力が要求され，これをみたす技術が開発されるであろう。

これに対して入力の方は依然として目立った進歩がない。カナタイプは習得が容易であるが，漢字がはいらないこと，漢字カナまじり文をカナタイプで打つ場合は漢字の正しい読み方がわからねばならないことに問題がある。カナタイプを用いる場合に漢字を入力する方法としては，カナで入力した文章が漢字カナまじり文になるよう適切なカナ漢字変換を行なうことであり，すでにいろいろな研究が行なわれてきている。その基本的手法は文章を正しく単語に分割しそれと同時に単語の仮名を漢字

になおすものであるが，同一発音の単語で異なった漢字に対応するものは非常に多く，一意に決定することは非常にむずかしい。これを一意に決定するために，漢字の訓読や音読をつけ加えて入力する方法もあるがめんどうである。文法的知識や話題の分野をかぎっても現在90％以上正しく漢字を復元することはむずかしい。将来文章の意味がうまく取りあつかえるようになったとしても95％以上を望むことはむずかしいと想像される。

漢字入力をけん盤で行なう方式も最近いろいろと改良が重ねられてきているが，将来とも抜本的な方式はむずかしいであろう。

7.2.3 印刷文字の自動読取り

最近になってようやく漢字読取装置の研究が盛んになってきた。教育漢字881字，当用漢字1,850字を対象として研究が進んでいる。1種類のフォントで印字の質が良好な場合には，おそらく95～98％程度までは比較的容易に達することができよう。問題は印字の質，フォントの種類の異なり，装置の値段である。仮名やアルファベットの場合にはある程度印字の質がわるくても，雑音をのぞいたり骨格線を出したりすることができるが，漢字のように単位面積あたりに文字ストロークが密にはいる場合には雑音の影響が非常に大きく，雑音をのぞくこともむずかしい。フォントとしては主としてゴシック体の読取りが研究されているが，明朝体の場合の細い横線の検出が問題となる。ゴシック体と明朝体の字体の相異は非常に大きいが，同じ明朝体でも文字の外郭の縦・横比や字体の細部がいろいろと異なる。一種類のフォントでも数千の文字があるのであるから，いくつもの異なったフォントに対処することは非常にむずかしい。

しかし漢字の読取りは現在の文字読取装置に関する技術を十分に利用し，判定に時間がかかってもよいとすれば実現可能であろう。問題は経済的に十分成り立ちうるような装置を，要求される判定精度，判定速度で作りうるかという問題であり，今後数年の努力に期待したい。またこれを実現するためには明確な漢字文字読取装置の導入分野，適用業務を設定することが最も大切である。たとえば郵便の宛先の県，市，郡名などは印刷漢字の場合読取りの対象となりうると考えられる。漢字の使用が必要な他の分野は多くのターン・アラウンド・ドキュメントである。住所，氏名など現在はすべてカタカナに頼っているが，これを漢字にすることが必要である。このような分野では漢字の読取りと漢字の計算機出力とが密接な関係をもつ。したがって出力漢字の品質を十分向上させ，また各漢字に補助情報をつけ加えることによって，読取りが確実になるようにすることが必要となろう。

漢字のように複雑な文字を多数読取るためには非常に分解能の高い入力装置が必要である。アメリカのI.I.I.(Information International Incorp.)のフライングスポットスキャナーは32,000×32,000という高分解能のものであり，漢字の読取りの研究もしている。

7.2.4 手書き文字の読取り

手書き文字には二つの種類があって，すでに手書きされてしまっているものを対象とする場合と，手書きしている運筆を実時間で入力し読取るという場合とがある。前者はかなりむずかしい問題を含むが，楷書で一定の枠内にていねいに書かれた文字の場合には十分に自動読取りの対象となりうるだろう。たとえば専門のガリ版をきる人の文字などについて積極的に研究を進めるべきであろう。後者

のオンライン手書き文字は，たとえば銀行の窓口とかIR端末で署名によって印鑑のかわりをするというシステムが考えられる。この場合には署名の認識の研究が必要であり，とくに他人が似せて書いた署名と本人のものとの区別を厳重に行なわなければならない。そのためのキーポイントとなるのは筆跡の速度であり，幸いオンライン手書き入力の装置には筆速が入力できるものがあるので，これを用いることが必要となろう。字体の特徴と共に筆速情報を用いることによって，署名の認識が確実となる。署名の場合は必ずしも文字そのものが認識されなくてもよいが，オンラインの漢字認識はすでに書かれているものの認識よりも，筆順，筆速，ならびに処理時間に余裕があり，これらを利用できるので十分に実用化される可能性をもつと考えられる。この場合，入力装置は形があまり大きくなくできるだけ安いものとし，数十のこれらの入力端末装置を1台のミニコンピュータで管理するようなシステムとすれば，1使用者当りのコストは非常にさがり，実用の可能性がふえよう。

7.2.5 音 声 認 識

音声認識研究も最近かなりの進歩をみせたが，個人の発声を訓練によって規格化するということがほとんど不可能であるので，文字の場合のように標準形を設定して制定を簡単化し正読率を向上させることができない。しかし個人の声をあらかじめ，または使用の直前，機械に記憶しておき（これは10～20秒でできる），これをもとにして音声認識をするというシステムは実用の可能性があると考えられる。この方向の研究としては電々公社武蔵野電気通信研究所でよい成果をあげている。

一方，自然な会話文を認識しようという研究がアメリカで1971年から数カ所の研究所で大規模に開始されている。これは約1,000単語を対象にし，数人から10人程度の人が話す文をあつかうことを当面の目的としているもので，音声学的な特徴抽出，判定とともに，言語学的知識を用いるところに特徴がある。すなわち会話文を構造分析，意味分析し，話題に関してもっている内部知識を用いることによって次に発話される単語や文を予想し，その文脈の中で分析をある限られた範囲において精密化することができるようにしているところに特徴がある。

7.3 出力システム

文字入力と文字出力はつねに相ともなって発達することが望ましい。入力をけん盤からする場合にはかならずモニタ出力が打けんと同時に得られることが必要である。この点アルファベット，カナの場合には入力と出力とが端末装置でバランスがとれているが，漢字の場合には問題となる。従来の漢字テレタイプではほとんどの場合モニタ出力を簡単に便利に得ることができなかった。最近のものはこの部分に多くの改良がみられる。とくに漢字入力をカナけん盤から入力する場合やSyno typeのように，ストローク情報をけん盤から入力する場合には，選ばれる漢字がただちに表示されることが必要である。多くの場合それらは複数であるから，それらから目的のものを選ぶ操作を次に行なわねばならない。いずれにしても入力と出力とはつねに組合わせられる必要があろう。文字出力には，いわゆるハードプリントとソフトプリントの区別がある。ソフトプリントは一時的で，ある時間表示された後は消去され，別の文字がつづいて表示されるようなもので，ブラウン管にディスプレイされる文字の大きさ，一度にディスプレイできる文字の数，文字の種類などが使われる場によって種々異な

る。ハードプリントでは主として文字の大きさと，きれいで読みやすいかどうかということが問題となろう。また同時に何枚コピーをとることができるかも大きな要素である。

7.3.1 漢字プリンタ

漢字プリンタとして最近いろいろな種類のものが市場に出されてきた。値段はまだかなり高いが，だんだんと安くなることが期待される。これらは大きくわけると次のようになる。

(a) 漢字の活字ケースから機械的に選択して打つもの

(b) ドット・プリンタによって活字に対応する信号によって打つもの

(c) 活字の図形を電子的に走査して出力するもの

(d) 写真植字の方法によるもの

などが主なものである。これらの方式のうち，(a)，および( )で機械的なインパクト方式で出力するものは，コピーが何枚もとれるという利点をもつ。電子計算機からの文字出力としては，速度が速く，値段が安くなる可能性が大きいものは機械的に細いピンをならべたドット方式によるものであろう。この場合の問題は，印字品質と紙として特殊なものを用いなければならないかどうかということである。これらも今後の改良によって十分実用になるものができると期待される。ドットのかわりにCRTに信号を出力しこれを焼きつける方式もあるが，当分は機械的なドット方式の方が広く用いられるものと考えられる。ゼロックスのように静電式のコピー技術の将来は注目する必要がある。

人間の手でキー入力するときのモニタ出力には，従来からある，活字をひろって打つ方式のほかに，CRT上に電子的に文字を出力する方式もありうる。現在の段階では人間のキー入力の速度に合わせて安いものとしては活字によるものであろうが，将来はCRT出力も十分考えられる。この場合はその後のハードコピーをとりたいという要求があるから，その装置としては，別のものを考えねばならないだろう。モニタ出力としては，打けんした文字すべてが表示されなくとも，直前に打った1文字だけが表示される。

国際電信電話会社研究所の開発したようなベクトルによる文字出力装置も将来性がある。

7.3.2 文字ディスプレイ

CRTへの文字出力としては，ストロークの構成によるもの，ドットによるもの，および漢字パターンを走査してそのパターンを出力する方式などがある。この場合もread only memoryによるドット方式が結局は安くでき品質もまずまずのものが得られると考えられる。CRTとしてはTVと同じ走査方式で家庭用TVのブラウン管も使えるので，その面からも安くなるであろう。液晶やプラズマ・ディスプレイの場合には分解能が将来どこまであげられるかが問題であろう。1画面内になるべく小さな文字でなるべく多数表示することが必要となるから，リフレッシュ・タイプでなく，記憶方式のディスプレイがどうしても必要となる。1文字か，せいぜい数文字表示すればよい場合には，光学的にマスクパターンを選んで表示するという方法も十分使える。ホログラフィ技術のこの分野への応用はこれからの研究にまたねばならない。

7.3.3 フィルム出力

フライング・スポット管によるものと光学的なものとがある。漢字のマスクパターンがきれいであ

れば，出力は十分に良質のものが得られる。文字の並び，文字の大きさの変化，その他フォーマットに関してはフライング・スポット方式が自由度があってよく，この方式は今後広く用いられるであろう。ただし値段がもっとさがることが望まれる。

写真植字は現在広く用いられているが，計算機と連動してフォーマット制御まで完全にされているものはそれほど多くない。この種の機械は今後もっとのびるだろう。しかしこの場合も入力データがモニタでき，修正が簡単に行なえることがキーポイントとなる。計算機のソフトウェアとしてテキスト・エディタの便利なものが作られる必要がある。

7.3.4 音 声 出 力

音声出力として，録音音声の編集方式はすでに実用されている。この方式の今後の問題は音声データのビット数を能率よく圧縮することができないかということであるが，Predictive Codingの手法を用いれば，なまの音声の録音情報の1／30程度には圧縮することが可能であろう。

音声出力のもう一つの方式は規則による音声合成方式であるが，これは現在まだよい音質（自然性）のものは得られているとはいえない。将来もっと研究されるべきテーマである。音声合成における問題は文全体としてのイントネーションであり，これは文の構造の認識を必要とするが，イントネーションをうまく与えるための構造分析は使用する分野をかぎれば比較的容易に行なうことができるだろう。音声応答は今後ますます広く用いられることになるだろう。一般に音質がわるい場合には聞き手は二度，三度と同じことをくりかえし聞くから音声回線の保留時間が長くなる。これは多数の端末装置に同時にサービスをしたいときには障害となる。したがってシステムは少し複雑でも十分によい品質の音声応答を行なう必要があり，今後もっと改良する必要があろう。

7.4 処理システム

日本語情報処理がほんとうに発達するためには，入出力機器の問題とともに言語の処理プログラムがいろいろと作られる必要がある。そのためにまず大容量のファイルが利用でき，各種の辞書とともに日本語情報が大量に蓄積されねばならない。そして日本語の分析処理のためのアルゴリズムの研究，それをささえるソフトウェア・システムなどの整備が必要である。これらはまだ研究段階を出ないが，将来のことを考え汎用性のあるシステム的考え方にたって研究を発展させねばならないだろう。

7.4.1 フ ァ イ ル

最近になって電子計算機の外部記憶装置の容量が急速に大きくなってきた。とくに集団磁気ディスクの使用によって，10^{10}ビット程度の容量は比較的簡単にもつことができる。ホログラフィによる記憶は密度を非常に高くとることができるが，書きこみが簡単でないという欠点がある。したがって固定された情報の読出しだけでよい分野にしかむかない。

日本語情報処理に必要な辞書的な情報としては次のようなものが考えられる。

(1) かな文字（ローマ字）単語辞書

(2) 活用語語尾変化辞書

(3) 付属語辞書

(4) 接頭語辞書

(5) 同音異字辞書

(6) 漢字単語辞書

(7) 漢字の特徴による辞書(画数，偏，傍などによるもの)

(8) 単語の音声情報辞書

辞書は目的によってそれぞれ固有のものが作られる必要があるから，見出し語の数，各見出し語につけるべき情報の積類などは場合によってすべて異なる。現在日本語に関しては計算機で使える辞書としてよいものはない。国立国語研究所では新聞文章を磁気テープに収録し，そこから異なり語を出し，辞書を作ったのが唯一であろう。辞書は作成するのに大きな費用がかかるから，一つの辞書がいろいろな観点から多数の人が使えるような形に作られる必要がある。

大量情報の記憶としては依然としてマイクロフィルム方式が用いられるものと思われる。図書館がますます大量の書籍，資料を収集するようになるにつれて，この問題は真剣に検討される必要がある。その試みとしてMITの図書館で行なっている研究にIntrex Projectがある。これは工学系の論文2万件あまりを1,500枚のマイクロフィルムにおさめ，キー・ワードによって検索し，目的のマイクロフィルムをとり出し，これをフライング・スポット・スキャナーで走査し，ブラウン管ディスプレイ装置に表示したり，コピーをとったりすることのできるものである。マイクロフィルムへの圧縮率はまだまだ高くなる可能性がある。

今後マイクロフィルム方式，電子的記憶方式，ホログラフィ方式のいずれが発達していくかはわからないが，少なくともマイクロフィルム方式と電子的方式とについていえば，それぞれ独自の分野を占めていくと考えてよい。電子的方式は情報のアクセスの時間が速く，情報の変更が容易であるが，まだまだ図書館の本全部が入れられるほど大容量のものにはなり得ない。そして既存の情報をすべて電子的情報に変換して入力することは不可能といってよい。また文の中に挿入される図面，写真などを電子的に記憶することはまずむりといわねばならない。これに対してマイクロフィルム方式はこれとはまったく逆の性質をもっているということができる。このように利用率が高く，変更が要求されるもので図面など複雑なものが含まれていない情報は電子式になるだろうが，百科辞典的情報はマイクロフィルムによらざるを得ないだろう。このような立場から情報ネットワークにおける通信形態を考える必要がある。すなわち文字のディジタル通信と共にファクシミリのような図形伝送の技術がもっと進歩するだろう。

7.4.2 言 語 分 析

すでに述べたように日本語は欧米諸国の言語とは異なった特色を数多くもっており，これが情報システムであつかう際の種々の困難さの原因となっている。この節では日本語の言語学的な面と情報処理との関係について論じよう。

(1) 辞　　書

言語情報処理を行なううえで基本となるものは辞書である。日本語の単語辞書は単語の簡単な意味

の説明はあっても，その文中での使い方についてはまったく説明がないということ，および適切な引用例がほとんどない（とくに現代語について）ということなど種々の不便があり，英語などの辞書とくらべて大きな欠点がある。しかしこれを改善し，よい辞書を作ることがこれからの課題であり，最近そういったものがあらわれてきた。英語においては辞書がそっくりそのまま磁気テープにおさめられ計算機が使用できるようになっているものがある。たとえばSystem Development Corporation（SDC）が作ったWebsters Seventh New Collegiate Dictionary, Merriam Webseter Pocket Dictionaryなどはその例である。これらは単語の意味の研究の基礎となっている。

日本語の辞書を計算機で利用できるように磁気テープにおさめることは近い将来ぜひとも必要なことである。その際は日本語の単語に関して，基本語を決定すること，同義語，反義語，派生語，関連語などを整理する必要がある。これはシソーラス作成に理論的基礎を与える。

(2) 言　語　統　計

従来行なわれてきた言語統計は単語を構成する素子としての文字の出現頻度や，文字のならびに関するマルコフモデルとしてのものであった。これからの言語統計はもっと大きな単位，たとえば単語，文体などに関するものである必要があろう。単語に関する頻度統計によって基本単語選択の基礎を与える。文節の付属語に関する統計，付属語の文中での並び方に関する統計などは日本語文の解析に大きな寄与をする。とくに動詞が要求する文節の型の統計（たとえば「行く」という動詞は「（時間）に」，「（場所），から」，「（場所）へ」という文節をとりやすい。）をとることは重要である。

現在国立国語研究所には新聞の文章が約300万語磁気テープに記録されているが，このデータから上記のような統計をうることは十分可能であろう。ただし重要なことは，アプリケーションの需要を考えるとまず科学技術の分野の文献の処理が考えられるから，統計をとるべきデータはそれぞれ分野別（たとえば物理学，化学，電気工学等）になっていることである。

かな一漢字変換はこの種の言語知識を十分とり入れて行なわれる必要がある。

(3) 文および文章の構造分析

従来から行なわれてきた文脈自由型あるいは文脈規定型句構造文法は，英語文章に対しても行詰まりにきたと考えることができよう。とくに日本語の構造は英語のようにはっきりした構造をもっていないため，このような句構造文法での方法では不十分であることは明らかである。英語においても句構造文法のように規則の厳密なものではうまく解析ができないという考えから，もっと柔軟性のある分析の方法が試みられてきている。これは意味を中心とした解析である。日本語に対してもこのような考え方からの解析が試みられるようになるだろう。文の解析の次に文の連続である文章（discours）の解析も必要である。文章を読んで内容をはあくすることは，自動抄録，内容分析などに連なっている。現在のところ単語の使用度数によるアプローチが多いが，構文分析を土台にした文章解析も，理論的には考えられるし，一部実験もされている。それは(4)の意味分析にもつながっている。

日本語の文章の解析において，第一に問題となるのは連続した語の列を文節単位に区切ることである。ローマ字書き，またはかな文字文のときに文節単位に切ることは非常にむずかしく，これを成功させるためにはどうしても漢字かなまじり文または漢字部分の指示をした文を入力とする必要がある

ものと考えられる。漢字かなまじり文の場合，漢字の直前で文節区切りを行なったとき，50%近くの誤りがある。

構文分析は依然として困難であるが，話題をある分野に限って，比較的素直な文体で書けばかなりの程度まで正しく分析することができ，ある程度使用にたえるものが作れる可能性がある。

(4) 意 味 分 析

言語の意味分析はいくつかにわけて考えることができる。

その第一は単語のもつ意味で，これは従来の辞書をある程度利用できる。

第二は単語がいくつか集まってひとつの句文を構成しているとき，その句文が表現している意味の問題がある。これは次に述べる質問応答システムにとくに関係をもつ。

第三はプラグマティックス，あるいは現実の社会の知識をどのように反映させるかの問題である。百科辞典的知識の記憶と利用である。これらのうち，第一のものはかなりの程度行なえる見通しがあるが，第二のものは当分は研究，実験の段階にとどまらざるをえないであろう。第三のものについても同様である。これらは記憶装置を非常に大きく占めるが，現在ではハードウェア的な問題でなく，原理的なところに問題があるということがはっきりしてきている。

(5)論理学との関係

自然言語で話される内容はしばしば論理的に筋が通っていて，論理の世界に表現しなおして処理するのがよいと考えられる場合がある。文章内容が論理表現になれば，そこから種々の演繹をすることができるため，有効な質問応答システムを作れる可能性がある。またこの論理表現じたいが一種の中間言語となり，言語の翻訳の可能性への橋渡しをすることができよう。

しかし問題は第一に自然言語の文章を論理式表現に変換することが非常にむずかしいということにある。とくに述語論理で用いる述語変数の概念を自動的に文章から決定し取り出すことはむずかしい。

第二に現在の論理学は第一階述語論理についてはよく研究されているが，自然言語の文章を表現するためにしばしば必要とされる二階以上の述語論理については，まだほとんどよくわかっていないことである。また文章が表現している時制，様相などについて論理学はまだ十分な発達をしていない。このようなことから，今後この方向での種々の試みは必要であり行なわれるであろうが，これだけでなく自然言語じたいをうまく記憶し，質問応答などができるシステムの開発が必要であろう。

(6) 文 章 の 発 生

文章を発生することは情報処理の種々の分野で必要とされることである。対話型のプログラミングシステムと自動翻訳，質疑応答システムなどはその代表である。自動翻訳の場合は文章を発生させる源となる文は，入力として与えられるが，その他の場合には内部で発生させる文章の源をもっていなければならない。従来は文章の発生は比較的簡単であると見なされているが，実際には分析と同様に非常にむずかしいことがわかる。対話型システムの場合などほとんどの場合は，出力すべき文の型，文の原形を内部に用意しておき，わずかにその中のいくつかの単語をさしかえるだけで出力するという内容である。ほんとうの文章の発生は表現すべき内容がいわゆる深層構造として概念構造が内部に

形成され，それがいくつかの段階をへて文法的に正しく詳しい正確な表現となって出力されるべきものである。この方向の研究は現在ほとんどない。質問応答システムの発展にテンポを合わせて将来発展する可能性は十分ある。

(7) 内 容 分 析

内容分析は，一団の文が全体としてどのようなことを，どのような態度で表現しているかを把握するものである。日本では現在までほとんど研究されてきていないが，情報検索の問題，その他と関連して将来は研究がもっと活発になる分野であると考えられる。アメリカでは数年前からいろいろと研究されてきた。その方法は，主として用いられる単語の範囲と頻度などによるものあるが，構造的要因，断定，推定，肯定，否定などの様相に関する要因を加える必要がある。

(8) 機械のための日本語

まったく自由な自然言語を対象としていては，いつまでたっても文の完全な分析，理解をすることは機械にはできないであろう。そこで実用上考えられるのは，機械が正しく解釈できるように文の構造を制限することである。たとえば副詞はそれがかかっていくものの直前におくとか，文の構造が複雑になるときには適切な句点を挿入するとか，ある種の文体を禁止するとかである。これは近い将来ぜひとも研究開発を行なうべきものである。

7.4.3 言語処理ソフトウェア

言語処理のソフトウェアとしてはCOBOLが広く用いられている。情報検索，KWICを作るようなプログラミングシステムも COBOL で書かれたものが多い。しかし現在の日本の計算機の多くはCOBOLよりもFORTRANの方が便利であるので，FORTRANでの言語処理プログラムの開発が望まれる。しかしこの場合，単語の各文字を独立に扱うことが必要となるのでENCODE文，DECODE文が必要であり，これにはFORTRAN JIS-7000のレベルが必要である。

文字としては，ローマ字書きがプログラム上便利であるが，漢字かなまじり文の符号についてプログラムを作ることはそれほど困難ではない。ただし漢字かなまじり文の入力をカードで行なう方式がまだできていないので問題があろう。

他の言語処理プログラミング言語としてはSNOBOL，LISPなどがあるが，現在これらは一部の研究者の間でしか用いられていない。

この種のシステムで必要な本質的な機能は，辞書を能率良く参照できる機能と，樹枝状構造の変換が自由に速くできる機能とである。言語処理をもっと容易にするためには，この種の言語が漢字かなまじり文に対して用いられるように改良され，多くの計算機で用いられるようになる必要があろう。

7.5 応 用 の 将 来

(1) 新 聞，出 版

現在すでに新聞，出版のシステム全体がかなり電子計算機に依存した形となってきており，将来ますます自動化されていくことは明らかであろう。新聞原稿の編集システムの自動化の最大の問題は，紙面の割り付けを実物大で行なうことである。これは現在できていないため，記事を紙面に割り付け

ても見出しや見出しの活字の大きさなどに関しては実感をもってCRT上で決定することができない。したがって少なくとも見出しについては漢字そのものが見出しの場所に表示される表示CRTが必要である。しかもこれらの漢字は字体の指定もできなければならない。したがって非常に大きなサイズのディスプレイで十分な分解能をもつ必要がある。おそらく10,000×15,000の分解能があればよいだろうから，これを作ることは可能と考えられる。文字のディスプレイのもととなる文字は，かなり多く用意しなければならないところに問題があろう。

次に問題となるのはこのような表示装置が最も有効に使えるソフトウェアを整備することであり，これは使用経験にもとづいて徐々に改良されていくだろう。写真やグラフも，実際にモニタCRT上に出して編集できるかどうかは将来のハードウェアの発達に待つ必要があるが，その可能性は十分ある。

出版における問題は写植を自動的に行なうことであり，電子計算機と結合された写植システムは最近非常に発達してきた。今後はその速度を上げることと，エディティング機能を改良することである。今後はコストも徐々にさがってくるだろうから，ミニコンピュータと結合されたシステムは広く用いられる可能性がある。

(2) 放　　送

放送においてはテレビジョンへの文字の出力ディスプレイが現在最も大切なものであろうが，そのほかに放送電波を利用したデータ伝送があり，各家庭でファクシミリのようなものでコピーをとるというシステムも考えられる。テレビジョンへの文字出力における問題は，入力を迅速に行なうこと，文字の配列をうまく指示できること，さらに図面との合成ができることが必要である。これは現在の技術で実現可能なことであるから，そのうち利用されることになろう。現在あまり用いられていないのは，質の良い文字を簡単に出せる安いシステムがないからである。

最近の新しい分野としてはCATVがある。これはケーブルによって情報を伝達するもので電波などのように法的制限が強くないという利点がある。このCATV網は単にTVだけでなく，その他種々の情報を送ることができ，また計算機との結合によって使用者からの積極的な働きかけにより情報をひき出すことも可能となるだろう。

(3) 情　報　検　索

アメリカでは種々の分野においてドキュメント，データが整理され，オンラインで情報検索ができるようになっている。これに対して日本ではこの種のデータ蓄積，情報サービスを行なっているところはほとんどない。問題は技術にあるというよりは，そういった情報をほんとうに利用したいという要求が少ないこと，そういった情報サービスの仕事に大きな価値，重要性を積極的に認めないこと，そしてなによりも情報を積極的に流通させ率直に受入れようという態度のない日本の風土に根本的な原因があるように思われる。さらに蓄積された情報の正確さや，それがつねに更新されているかどうかということに対する疑問，さらに種々の統計に対する不信感なども存在する。

英語，ローマ字による情報の蓄積，検索に比べて，漢字かなまじり文の場合には，入力／出力ともコストが高くつくということもあまり普及しない原因であるかもしれない。しかし限られた分野では

着実にのびていっているから，これから数年先には相当な発展が望めよう。情報検索が用いられる分野としては，情報サービスが中心で次のようなものが代表的である。

株　価　情　報

予約：催物，ホテル，航空機，レンタカー

レジャー情報，海外観光情報

科学技術刊行物情報

気　象　情　報

特　許　情　報

医　療　情　報

時　事　情　報

判　例　情　報

商品，人材，不動産情報

一　般　知　識

(4) CAI(注1)，QA(注2)システム

計算機による教育は最近研究が盛んになってきているが，もっと自然言語によって質問応答が行なえるようになる必要があろう。すなわち，スライド・プロジェクタに出る問題を見て，YESかNOをボタンで押したり，単に数値の答をするだけでなく，生徒の方から積極的に質問を発することができるようになることである。そのためには文章が正しく計算機で分析され，理解され，蓄積されているデータから適切な演繹が行なわれることが必要となる。その結果得られる答を適切な文章で出力しなければならない。これらは，これからの問題であるが数年先にはある程度使えるものができるであろう。

(5) 機　械　翻　訳

機械翻訳は依然として困難である。技術論文に限って，しかも専門分野をかなり狭くとれば，ある程度の翻訳は可能であろう。しかしそのためには辞書作成，文法の確立など多くのことがなされねばならない。したがってしっかりした専門のグループが形成されて研究を進める必要があろう。最近アメリカその他で，意味を積極的に利用した分析が試みられつつあるが，分析のアルゴリズムの確立とともに辞書をどのように構成すべきかについて研究すべきであろう。限られた分野では，案外早く実用されることになるかもしれない。

（注1）CAI：Computer Asisted InstructionあるいはComputer Aided Instructionの略

（注2）QA：Question and Answerの略

参 考 文 献

参 考 文 献

漢字情報処理一般

(雑誌特集号)

(1) データ通信，特集／文字とマン・マシン・システム，1971. 11，情報研究出版会

花島，データ通信に適した漢字システム

安田，漢字情報処理システムの現状と将来

野沢，T4100漢字情報システムのハードウェア，ソフトウェア

飛田，漢字情報処理システム

斉藤，江川，日本語の文字機械処理上の問題点，とくに漢字をめぐって

(2) FUJITSU，漢字情報処理特集，1970，Vol. 21，No. 7，富士通研究所，

高須，新聞製作工程電算化の現状

石原，朝日新聞社における漢電システムについて

古田，凸版印刷における組版の機械化

倉持，印刷工程の省力化システム

秋草，柿沼他，漢字処理とソフトウェア

塚谷，漢字処理とハードウェア

篠田，秋草，全自動写真植字システム

塚谷，石川，FACOM6531C全自動写真植字装置

(3) ビジネス・コミュニケーション，特集漢字処理システム，Vol. 9，No. 10，1972，企画センター

木沢，漢字処理システムの現状と問題点

清正，行政における漢字処理システム

学研他，実践化ユーザが語る漢字処理システムの利点と問題点

各社の漢字処理システム

(4) 情報処理，漢字情報処理特集，Vol. 10，No. 5，1969，情報処理学会

植村，漢字かなまじり文KWIC索引

長谷川，高速漢字プリンタ

坂井，長尾，寺井，部分パターンによる漢字の合成

斉藤，永井，山田，大迫，漢字の機械への入力

釜江，小野，テレビ電話を用いた文字表示のシミュレーション

高橋，広田，計算機による日本語の処理

石綿，機械処理の対象としての漢字

長井，漢字入出力装置の動向と技術的問題

大場，JICSTにおける日本語抄録の電算機処理
安田，わが国の新聞社における漢字情報処理
林原，矢野他，最近の日刊紙における文字使用状態について
木沢，漢字のネモニックコード化への試み
穂鷹，漢字のInputの一方式
多田，漢字の性質について
高沢，$x-y$プロッタで漢字を書かせる一つの試み
（5） 印刷雑誌，漢字情報処理システム小特集，Vol. 55, No. 3, 1972, 印刷学会出版部
塚谷，FACOMシステム
佐藤，IBMシステム
野沢，T4100システム
伊達，JEM－3100／4100システム
（6） FUJITSU,Vol. 23, No. 6, 1972, 富士通研究所
田中，磯上，森本，甲田，菊地，伊藤，FACOM6305A／B手書数字読取OCR
石井，萩原，菅野，木村，土田，柿沼，FACOM6521A漢字プリンタ
井上，佐藤，林，尾形，黒田，FACOM6221R$_{II}$文字ディスプレイ装置
（7） エレクトロニクス，特集漢字情報処理システム，Vol. 17, No. 1, 1972, オーム社
長尾，漢字情報処理システムを考える
石綿，漢字まじりの日本語処理
石井，FACOM漢字処理システム
上田，IBM漢字情報処理システム
渡辺，漢字入力表示装置
金田他，T4100漢字情報処理システム
（8） 特集日本語，数理科学，Vol. 7, No. 11, 1969, ダイヤモンド社
（その他）
（9） 高橋，日本語の機械処理，東洋経済新報社，昭45
（10） 高橋，研究日本語情報処理(I)～(V)，事務と経営，1971. 5～1971. 10.
（11） 新谷，早瀬，星沢，普及期に入る漢字情報処理システム（1）～（3）—連載中—，電子と経営，1973.2－1973.4，電波新聞社
（12） 長谷川，漢字処理システムの概念と問題点，Computer Report, Vol. 12, No. 8, 1972，日本経営科学研究所
（13） 市場総合開発研究所企画調査部，わが国における漢字情報処理システム市場——その現状と展望，市場総合開発研究所，昭47
（14） 坂井，コンピュータ利用における日本語処理——その現状と将来，電子と経営，No. 14, 1972, 電波新聞社

日本語，言語

(1) Chomsky. N, Aspects of the theory of Syntax, MIT Press, 1965

(2) Bach. E, An introduction to transformational grammers, Holt, Rinehart and Winston, Inc, 1964

(3) D. S. Worth, A Deep Index of Derivational Morphology, RM-5143-PR, 1966. RAND, Corp.

(4) J. Robinson, PARSE, A System of Automatic Syntactic Analysis of English Text, RM-4654-PR, 1965, RAND Corp.

(5) H. Kacera, W. H. Francis, Computational Analysis of Present-day American English, Brown Univercity Press, 1967

(6) 国立国語研究所，分類語彙表，秀英出版，昭39

(7) 国立国語研究所，電子計算機による新聞の語彙調査(Ⅰ)，(Ⅱ)，国立国語研究所報告37，38，秀英出版，昭45

(8) 国立国語研究所，同音語の研究，国立国語研究所報告20，秀英出版，昭36

(9) 五十嵐，山田，ドイツ語文生成のための変換方法，情報処理，Vol. 7, No. 2, 1966，情報処理学会

(10) 水谷，松原，坪井，同字異訓熟語集，計量国語学，No. 58, 1971，計量国語学会

(11) 山田，五十嵐，変換文法理論による日本文生成，情報処理，Vol. 7, No. 5, 1966，情報処理学会

(12) 本社使用活字頻度調査，毎日新聞社東京工務局，昭28

(13) 国立国語研究所，現代雑誌九十種の用語用字(1)，(2)，(3)，国立国語研究所報告21，22，25，秀英出版，昭37

(14) 国立国語研究所，現代新聞の漢字調査（中間報告），国立国語研究所資料，No. 8，秀英出版，昭46

(15) 石綿，名詞＋助詞「に」＋動詞の表現構造，情報処理学会 Computational Linguistics 研究委員会資料

(16) S. Ceccato, Linguistic Analysis and Programming for Mechanical Translation, Gardon & Brench, 1961

(17) 坂井，日本語のゆくえ，データ通信，1969. 4，情報研究出版会

(18) C. E. Osgood et al, The Measurement of Meaning, Univ. Illinois Press, 1957

(19) 雑誌記事索引自然科学編における漢字の頻度調査，国立国会図書館，1967

(20) 漢字コード委員会，標準コード用漢字表（試案），情報処理学会，1971

(21) 国立国語研究所，計算機による国語研究V.国語研究報告，秀英出版，昭48

(22) 国立国語研究所，現代語の語彙調査，婦人雑誌の用語，国語研究所報告4，1953

(23) Schachter, Kernel and non-Kernel sentences in transformational grammer, Proceeding of the ninth Informational Congress of Linguists, Monton, 1964

(24) 江川，漢字かな混り文の「自動単位分割」に関する研究，計量国語学，1967-43144計量国語学会

(25) 野村，新聞使用漢字の試行的分析，電子計算機による国語研究Ⅱ，国語研究所報告No. 34，秀英出版，昭44

(26) 山田，P-S理論による日本語文法，国語学，No. 63, 1965, 国語学会

(27) 国立国語研究所，類義語の研究，国語研究所報告28，秀英出版，昭40

(28) 田中，堀江，新聞語い調査データによる同音短単位表と同形短単位表の作成，LDP（月報別冊），No. 6, 1970

(29) T. Sakai, M. Nagao, Sentena Generation by Semantic Concordance, International Conference on Computational Linguistics 1965

(30) N. Chomsky, Syntactic Structure, Mouton & Co, The Hague, 1962

(31) J. J. Kantz, J. A. Fodor, The Structure of a Semantic Theory, Language. 39, 1963

(32) C. E. Osgood, et al, The Measurement of Meaning, Univ. Illinois. Press, 1957

(33) 林　，当用漢字字体表の問題，国語シリーズ53，光風出版，1963

(34) 長尾，Context-Free Phrase Structure Languageの構造分析，昭和38年度情報処理学会全国大会，昭38.12

(35) 坂井，長尾，文章の生成　—句構造と変形構造—，昭和39年度情報処理学会全国大会，昭39.12

(36) 杉田，機械翻訳からみた日本語，昭和46年電気四学会連合大会，1971. 10

言語情報処理

(1) 杉田，A Study on Mechanical Translation from English into Japanese, 京都大学博士論文集，昭43

(2) H. Nisimura, A Trouble with Computer Arrangements of natural Words, Electrotechnical Laboratory MITI, Tokyo,

(3) Language and Machines, Computers in Translation and Linguistics, National Academy of Sciences, 1966

(4) 坂井，杉田，渡辺，電子計算機による和文英訳，情報処理，Vol. 10, No. 6, 1969, 情報処理学会

(5) 寺井，電子計算機による漢字の合成，京都大学修士論文集，昭44. 3.

(6) 石綿，言語の意味と言語情報処理，電子計算機による国語研究，国立国語研究所報告31，秀英出版，昭43

(7) 坂井利之，「翻訳するコンピュータ」，講談社，昭和44．9
(8) H. Eggers u. a, Elektronische Syntaxanalyse der deutschen Gegenwants sprache, Max Niemayer, 1969.
(9) 田町，言語の機械翻訳(Ⅲ)，電子通信学会誌，Vol. 46, No. 11, 1963, 電子通信学会
(10) 栗原，言語の機械翻訳(Ⅰ)，電子通信学会誌，Vol. 46, No. 11, 1963, 電子通信学会
(11) 栗原他，仮名文の漢字混り文への変換について，九大工学集報39(4)，PP. 659－664, 1967, 九州大学
(12) 坂井，杉田，機械による英文和訳，電子通信学会誌，Vol. 49, No. 2, 1966, 電子通信学会
(13) 植村，機械翻訳（和文英訳）の一方法について，京都大学修士論文集，昭41
(14) 西村，機械翻訳の一模型，情報処理，Vol. 8, No. 2, 1967, 情報処理学会
(15) 和田，機械翻訳の現状と将来，情報処理，Vol. 6, No. 3, 1965, 情報処理学会
(16) 坂井他，計算機による和文英訳，電子通信学会専門委員会資料，昭44－01
(17) 松下，かなやローマ字の入力をいかに漢字かなまじり文に変換するか，ソフトウェア科学，1972. 3, 産報
(18) 相沢，計算機によるカナ・漢字変換，情報処理における日本語（漢字）の取扱い，PP. 1－11, 日本経営科学協会セミナーテキスト，
(19) KING et al, Machine Translation of Chinese, Scientific American, Vol. 208, No. 6, 1963,
(20) A. D. Booth, Machine Translation, North－Holland, 1967
(21) 坂井，長尾，最近の言語処理研究について，情報処理，Vol. 10,. No. 1, 1969情報処理学会
(22) M. Nagao, Engineering Approach to Language—Key : Computational Linguistics machine—IR－System, 計数言語学調査セミナー
(23) M. Nagao, Engineering Approach to Language Problems, U. S－Japan Survey Seminar on Computational Linguistics, Hawaii, 1968
(24) 蓼沼，坂本，玉川，英文分析のためのKWIC，情報処理学会機械翻訳研究委員会資料，1967
(25) 木村，構文解析自動化の研究Ⅱ，国語研究所報告34，秀英出版，昭44
(26) 石綿，構文解析自動化の研究Ⅰ，電子計算機による国語研究Ⅰ，国語研究所報告31，秀英出版，昭43
(27) 野村，漢字かなまじり文の文字連続，電子計算機による国語研究Ⅳ，国語研究所報告46，秀英出版，昭47
(28) T. Winograd, Understanding Natural Language, Edingburgh University Press, 1972
(29) 坂井，言語情報処理，ことばの宇宙，Vol. 2, No. 2, 1967
(30) 坂井，長尾他，機械翻訳の１つの試み，昭和38年度情報処理学会全国大会，昭38.12
(31) 坂井，長尾，電子計算機による文章の生成，電気通信学会雑誌，Vol. 49, No. 1, 1966

(32) 坂井，長尾，言語処理研究の動向，電気学会雑誌，Vol. 89－6, No. 969, 1969

(33) 坂井，情報処理における自然言語の問題点,電子通信学会誌，Vol. 53, No. 10, 1970

入　　力

(1) 吹抜他，あいまいさ種別のフィードバックを有するカナ鍵盤漢字入力装置，電子計算機研究会資料，No. EC72－5, 1972－5, 電子通信学会

(2) 森田他，文献速報自動印刷システムのための入出力字種の選定，第5回情報科学技術研究集会論文集，PP. 7－16, 1968

(3) 西村，電試漢テレ仕様案，機械翻訳研究委員会資料，1965, 情報処理学会

(4) 平田，寺尾，符号化文字の読み取り装置の試作，簡単な自動パンチ，情報処理，Vol. 10, No. 4, 1969, 情報処理学会

(5) 高木，品名の電子計算機への直接入力について，情報処理，Vol. 9, No. 2, 1968, 情報処理学会

(6) 小沢，インプット研究，日本経営出版会，PP. 467－496,

(7) 黒崎，カナけん盤漢字表示方式，昭和43年度情報処理学会大会予稿集

(8) JICST電子計算機室，漢字インプットデータのディスプレイ装置を用いた校正実験　総合報告書

(9) 斎藤他，漢字の機械への入力，情報処理，Vol. 10, No. 4, 1969, PP. 194－198情報処理学会

(10) 寺井，中田，オンライン手書き片仮名文字認識，情報処理，Vol. 13, No. 3, 1972, 情報処理学会

(11) G. F. Groner他, On Line computer classification of handwritten chinese character as a translation aid,

(12) 新方式の漢字入力／表示装置，情報管理，Vol. 12, No. 9, 1971, 日本科学技術情報センター

(13) 佐伯，ソクタイプによる漢字処理法，情報処理，Vol. 11, No. 10, 1970, 情報処理学会

(14) 富田，野口，大泉，岩元，手書き片仮名文字の認識，電子通信学会誌，Vol. 50, NO. 4, 1967, 電子通信学会

(15) R. Casey, G. Nagy, Recognition of Printed Chinese Characters, IEEE Trans, EC－15. 1, 1966. 2.

(16) 保原，和文入力に関する一考察，情報処理，Vol. 11, No. 2, 1970, PP. 108－111, 情報処理学会

(17) 大倉，C－5110漢字入力装置を通して見た日本語における常用漢字，1972. 3, NEACニュース

(18) 情報処理学会，Preprints for Seminar on Input/Output Systems for Japanese and

Chinese Characters, U. S－Japan Committee on Scientific Corporation, 1971
(19) 池田，鈴木，清野，タブレット式図形入力装置と蓄積管表示装置をもつ図形入出力端末の構成と制御について，情報処理学会関西支部プログラミング言語研究会，1970. 23
(20) 石田，入出力機器の現状と動向，情報処理，Vol. 12, No. 1, 1971, 情報処理学会

出　　力

(1) S. W. Levin, A Very High Speed Electro－Optical Mechanical Phototypesetting Machine, IEEE TRANS, EWS－ 11, No. 2, 1968. 8, PP. 31－55
(2) J. C. Miller, A Simple Display for Character and Graphics, IEEE TRANS, C－17, No. 5, 1968. 5, PP. 470－475, Computer Design Publishing Corp.
(3) R. H. Stoty, A New Display Terminal, Computer Design, 1968. 4, PP. 80－86, Computer Design Publishing Corp.
(4) M. E. Stevens, et al, Automatic Typographic－Quality Typesetting Techniques, A State of the Art Review, NBS, Monograph 99, 1967. 7
(5) M. Yoshimura, T. Iijima, A Method for the design of the Font (An example of the Font of KATAKANA for the line printer)
(6) Y. Mita et al, A high speed microfilmprinter for Chinese Character, The 1970 SID Symposium, PP. 96－97
(7) D. M. Aredom, An overview of the computer output microfilm field Proc, FJCC 1969, PP. 613－623,
(8) 斎藤，キャラクタ・ディスプレイ，エレクトロニクス，昭44. 6，オーム社
(9) 三田，コンピュータ・アウトプット・マイクロフィルム，情報処理，Vol. 12, No. 12, 1971, 情報処理学会
(10) 鈴木，池田，清野，コンパクトな文字発生方式について―ひらがなの発生―，昭和45年度第11回大会講演予稿集，PP. 345－346, 情報処理学会
(11) 鈴木，池田，清野，コンパクトな文字発生方式について，情報処理，Vol. 11, No. 9, 1970, PP. 517－522, 情報処理学会
(12) 近藤，COMとその周辺機器，産報，
(13) Auerbach, Computer Output Microfilm, Auerbach Technology Evaluation Service, Auerbach INFO Inc,
(14) J. K. Koeneman, Computer microfilm－A Cost Cutting Solution to the EDP Output bottleneck Proc, FJCC, 1969, PP. 629－635
(15) ドット式漢字ディスプレイ装置，昭和43年度電気四学会資料
(16) ドット式漢字ディスプレイ装置，昭和43年度情報処理学会大会予稿集，情報処理学会
(17) 電子式漢字プリンタの研究進む，データ通信に漢字まじり文を，通研月報，電電公社

(18) 岩井，電子計算機出力としての漢字プリンタ，情報処理学会第9回大会講演予稿集，昭43，情報処理学会

(19) 井上，エレクトロニクス・イメージング・システム，テレビジョン，Vol. 21, No. 11, 昭42，テレビジョン学会

(20) 宮崎他，ファイバー光学記録管によるTV信号の記録，録画研究委員会，

(21) 村上，画像ディスプレイとコンピュータシステム，ビジネスコミュニケーション，Vol. 9, No. 5, 1972, 企画センター

(22) BETA INSTRUMENT Co.，Hard-Copy CRT Printer, Computer Design, 1968. 8, Computer Design Publishing Corp

(23) IBM2285映像複写装置，エレクトロニクス，昭44.2，オーム社

(24) C. A. STEINBERG, Information Storage & Retrieval SYSTEM, INFO DISPLAY, 1970. 2,

(25) W. F. Heagerty, Ideographic composing machine,

(26) 小川，漢字テレプリンタの現状と問題点，第9回情報処理学会大会予稿集，1968，情報処理学会

(27) 斎藤，漢字ディスプレイ装置の開発，昭和43年情報処理学会講演予稿集，P. 111, 1968，情報処理学会

(28) 林原，斎藤他，漢字ディスプレイ装置，沖電気時報，Vol. 36, No. 2, 昭46

(29) 滝沢，漢テレ・全自動モノタイプ方式の使用符号について，印刷雑誌，Vol. 49, No. 9, PP. 8-15, 1966, 印刷学会出版部

(30) 中摩，後藤他，漢字の陰極線管表示方式，情報処理，Vol. 9, No. 6, 1968, 情報処理学会

(31) 漢字情報処理システムT4100, 情報管理，Vol. 14, No. 2, 1971，日本科学技術情報センター

(32) 高速度の電気－光学－機械式写真植字機，海外技術ハイライト，Vol. 2, No. 10, PP. 58-59,

(33) 新谷他，高速漢字プリンタ装置，エレクトロニクス，昭43.11,PP. 92-97, オーム社

(34) 石田，小型ディスプレイ装置とその応用，情報処理，Vol. 10, No. 3, 1969, 情報処理学会

(35) P. Dickson, Latest word in printing spell new electronics market, Electronics, 1967. 5. 29, PP. 137-141

(36) 竹松，マイクロメーションシステム，Computer Report, Vol. 8, No. 11, PP. 17-16 日本経営科学研究所

(37) H-7830文字表示装置システム，日立評論，Vol. 51, No. 2, 日立製作所

(38) 吉村，飯島，文字字体の一設計法，情報処理，Vol. 11, No. 3, 1970, 情報処理学会

(39) ANPARI, Newsprint in TSS Age, Special Reportof ANPARI, 1968. 8,

(40) 木村，市川他，音声応答装置，情報処理，Vol. 12, No. 7, 1971, 情報処理学会

(41) New high speed microfilm printer used by navy's NORC during missile simulation, Control(Eng), 1968. 12. PP. 35－36,

(42) G. Weiss, Photo－quality CRT displays find new jobs, Electronics, 1970. 6. 8, McGraw－Hill PUBLICATION

(43) 渋谷，原，プロッターによる漢字書き，情報処理，Vol. 9, No. 3, 1968, 情報処理学会

(44) 釜江他，試作ビデオ応答装置について，研究実用化報告，Vol. 19, No. 1, 電電公社電気通信研究所

(45) RANK PRECISION INDUSTRIES LTD, THE XERONIC HIGH SPEED COMPUTER PRINTER, THE XERONIC PRINTER, 1962. 8,

(46) 霜田他，静電記録紙の実用化，FUJITSU, Vol. 20, No. 3, 1969, 富士通研究所

(47) 坂井，堂下他，周辺制御装置と入出力用XYレコーダ，電子通信学会電子計算機研究会資料，昭和43. 7

(48) D. Ophir et al, The Brookhaven Raster Display, Comm of the ACM, 1968. 6, PP. 415－416

(49) Edward K. Yasaki, The Computer & Newsprint, Datamation, 1963. 3

(50) 釜江他，テレビ電話を用いた文字表示のシミュレーション，情報処理，Vol. 9, No. 5, 1968, 情報処理学会

(51) 釜江，小野，テレビ電話のデータディスプレイへの応用，通研月報，Vol. 18, No. 3, 1969, 電電公社

(52) ラインプリンタによるカタカナ印字字体規格(原案)，JEIDA－5, 44－C－187, 電子工業振興協会

(53) 日本経営科学研究所，特集COMシステム,Computer Report, 昭45. 9，日本経営科学研究所

(54) E. F. de HAAN他, The Plumbicon, a new television camera tube, Pyhilips Technical Review, Vol. 25, No. 617

(55) 内藤，禰津，ビジコンによる漢字の発生,情報処理，Vol. 10, No. 6, 1969, 情報処理学会

(56) 吉田，木内，ビジコンの緩速走査のターゲット電流，第4回テレビジョン学会全国大会予稿集，昭43

(57) 堀，辻，ビジコン容量性残像の計算，第4回テレビジョン学会全国大会予稿集，昭43

(58) 漢字プリンタ，電子と経営，1973. 2, 電波新聞社

(59) 木村他，音声応答装置，情報処理，Vol. 12, No. 7, 1972, 情報処理学会

(60) 板倉，音声応答システムの現状と課題，Computer Report, Vol. 12, No. 8, 1972

(61) ディスプレイとその応用特集，電子通信学会誌，Vol. 53, No. 4, 1970, 電子通信学会

(62) 高橋，IBM漢字情報処理システム，ビジネスコミュニケーション，Vol. 9, No. 10, 1972, PP. 58－62, 企画センター

(63) 和田，漢字などの入出力装置について，情報管理，Vol. 14, No. 1, 1971，日本科学技術情

報センター

(64) H. Wada, A traial for Kanji Display on Video Scope, Key. ; computational linguistics, display Kanji, 計数言語学調査セミナー

パターン認識

(1) 坂井，長尾他，部分パターンによる漢字の記述，電子通信学会オートマトン・インフォメーション理論研究会資料，昭44－1

(2) 宮田，大泉，仮名文字パターン認識の考案，プログラミングシンポジウム，1966－1

(3) 岩元，富田他，結合行列を用いた手書き片仮名文字の識別，東北大学電通談話会記録，1966－06

(4) R. J. Spinrad, Machine recognition of hand R. J. Spinrad printing, INFORMATION AND CONTROL, 8. 2, 1965－04

(5) 坂井，文字の識別，電子通信学会誌，Vol. 46, No. 11, 昭38，電子通信学会

(6) 飯島，森，人間の識別能力に迫るOCR, ASPET/71, 日経エレクトロニクス，1972. 5. 22, 日経マグロヒル社

(7) G. L. Fischer, Optical character recognition, SPARTAN, 1962

(8) 森，増子他，OCRのための東芝形書体数字，昭和41年電気四学会連合大会予稿集

(9) 飯島，パターン認識について，工業技術院電気試験所創立75周年記念講演会予稿集

(10) 桜井，淵，パターン情報処理の研究について，情報処理，Vol. 13, No. 4, 1972, 情報処理学会

(11) K. Nakata Others, Problems in Chinese Character Recognition, First U. S. A－JAPAN Computer Conference, 1972, PP. 296－301, 情報処理学会

(12) E. C. Greanis, P. F. Meagher, The recognition of handwritten numerals by contour analysis, IBM J. 1963. 1

(13) 富田，野口他，手書き片仮名文字の認識について，オートマトンと自動制御研究会資料，1966. 6, 電子通信学会

(14) 中田，音声の認識と発生，情報処理，Vol. 13, No. 4, 1972, 情報処理学会

(15) 寺井，中田，オンライン手書き片仮名文字認識，情報処理，Vol. 13, No. 3, 1972, 情報処理学会

アプリケーション

(1) 不破，西尾，文献速報自動作成システムについて，情報管理，Vol. 10, No. 4, 1967, PP. 1－27, 日本科学技術情報センター

(2) 高崎，電子計算機による選挙報道，エレクトロニクス協議会会報，1962. 10, エレクトロニクス協議会

(3) 鈴木，安田，電子計算機による新聞記事の作成，リアルタイムシンポジウム予稿集，1965，鉄道サイバネティックス協議会
(4) 安田，電子計算機による新聞編集，日本新聞学会秋季大会予稿集，1965. 10,
(5) 林，JICSTにおける日本語の言語情報処理，データ通信，情報研究出版会
(6) 桜井，情報検索用機械，情報処理，Vol. 7, No. 6, 1966, 情報処理学会
(7) 情報センターの自動編集機とその周辺，印刷雑誌，Vol. 51, No. 6, 1968, PP. 2－5，印刷学会出版部
(8) 鈴木，辻，安田他，NEAC－2206によるオリンピック記事報道過程の自動化の実験，NEAC－Journal, 1965. 5, 日本電気
(9) 細野，井上，木村，音声応答システムの旅客サービスへの応用，第7回鉄道におけるサイバネティクス利用国内シンポジウム予稿集，1970. 12
(10) 安田，新聞編集における電子計算機，総合ジャーナリズム研究，1966. 1
(11) 安田，新聞編集の自動化とコンピュータ，数理科学，1966. 5, ダイヤモンド社
(12) 安田，わが国の新聞社における漢字処理，情報処理，Vol. 10, No. 5, 1969, PP. 340－347, 情報処理学会
(13) 石田，ANNECSへの挑戦，新聞印刷技術，No. 59, 1972, PP. 2－4
(14) 吉田，NELSONの概要，新聞印刷技術，No. 59, 1972, PP. 5－8
(15) 畔上，これからの新聞製作システム，印刷界，1972. 4, PP. 54－57
(16) 石田，自動新聞編集製作システムANNECSの全容，日経エレクトロニクス，1972. 5. 8, PP. 68－79, 日経マグロウヒル社
(17) 藤田，コンピュータによる新聞製作システム，情報処理，Vol. 13, No. 6, 1972, PP. 402－406, 情報処理学会
(18) 馬渡，斎藤，電算植字小事典，日本印刷技術協会，昭46
(19) 印刷学会出版部，コンピュータ植字，印刷学会，昭47
(20) 高橋，電子計算機植字の現状，情報管理，Vol. 11, No. 5, 1968, PP. 245－254，日本科学技術情報センター
(21) 臨時増刊電子植字，印刷雑誌，Vol. 55, No. 4, 1972, 印刷学会出版部
(22) 特集電算写植システムの動向，印刷雑誌，Vol. 55, No. 4, 1972, 印刷学会出版部
(23) 後藤，日本語を入出力するコンピュータ・システム，学習コンピュータ，学習研究社
(24) 坂井，情報処理技術と新聞の将来，新聞研究，No. 210, 1969: 1

そ の 他

(1) 石橋，言語から見たコンピュータと人間のインターフェイス，オートメーション，Vol. 15, No. 7,
(2) 高橋，情報検索システムにみるハードウェアの役割，エレクトロニクス，昭43.10, PP. 47－

51, オーム社

(3) 情報処理学会規格委員会，漢字の符号化・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，昭46.3、PP. 82－97, 情報処理学会

(4) 坂本，Automatic Language Processing System－KWIC index file－, CL研究委員会資料，1969. 5, 情報処理学会

(5) 植村，データ自動圧縮編集を持つKWIC索引システム，情報処理，Vol. 9, No. 6, 1968, 情報処理学会

(6) 田中，電子計算機によるワードリスト作成上の一問題，国立国語研究報告31，1968, PP. 115－132

(7) 菊地，自動索引法と自動分類法展望Ⅰ，情報処理，Vol. 8, No. 13, 1965, 情報処理学会

(8) 日本情報処理開発センター，記事情報検索のためのデータマネジメント，46－S006，日本情報処理開発センター

(9) 田中，ことば，コンピュータ，漢字データのプログラム的処理，ソフトウェア科学，1970. 9, 産報

(10) 橋本，伊藤他，日本語キーワード索引システム，第3回ドキュメンテーション研究集会発表論文集，PP. 167－176, 1966, 日本科学技術情報センター

(11) 植村，日本語KWIC索引の試み，第5回情報科学技術研究会発表論文集，PP. 143－152, 1968, 日本科学技術情報センター

(12) 石綿，新聞用語調査の用例印字プログラム,COBOL KWIC，電子計算機による国語研究Ⅲ，国語研究所，昭46

(13) 坂井，コンピュータ時代の言語問題，特集言語時代，ことばの宇宙，昭42.5

(14) 坂井，これからの言語と情報，Energy, Vol. 4, No. 2, 1967. 4

(15) 坂井，杉田，電子計算機による文章の図形化，CL委員会資料，1969. 9 9

本報告書の作成に当っては，電算写植システムを利用しています。この電算写植システムの利用については，ペティ株式会社，および株式会社デンショクのご協力を戴きました。

禁無断転載

昭和48年3月発行

発行所　財団法人　日本情報処理開発センター
東京都港区芝公園3丁目5番8号
機械振興会館内
TEL (434) 8211 (代表)

印刷所　株式会社　デンショク
東京都千代田区神田小川町3－28
TEL (291) 5175 (代)

47－R 001