文書情報処理に関する調査研究

昭和49年3月

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48年度に実施した「文書情報処理に関する調査研究」の成果をとりまとめたものであります。
序

経済の成長と情報化社会の進展にともなって著しい変貌を遂げつつある産業界において、経営者が常に安定した企業経営を維持していくためには、経営のための関連情報を十分に活用して適確な意思決定を行う必要があり、このような情報を必要に応じて経営者に提供する、いわゆる経営情報システムの開発が、内外の企業、研究機関等において活発に行なわれております。

当財団においても、昭和43年度以来、コンピュータを有効に利用して合理的な経営情報システムを確立するという観点から、あらゆるレベルの管理者に理解しやすい情報を提供するための情報処理システムについて調査・研究を実施してきました。昭和44年度には、その一環として、従来わが国ではコンピュータの出力が数文字、数字またはカタカナで記述されていたために、とくにおいまいであるとされてきた文書情報、漢字かな混り文で提供できるようにした日本語情報処理システムの基本構成を検討し、カタカナけん盤に同音異義語の判別処理機能を組合せて考案した、カタカナ入力漢字かな混り文出力方式の日本語情報処理システムについて考察を加えました。

その結果として、わが国においては、日本語文の合理的な入出力方式を確立することがもっとも重要な課題であるという認識を深めるとともに、日本語情報処理システムの構築にあたっては、漢字けん盤による入力方式のそれに比較してデータ入力の操作が簡便で、しかも理解しやすい出力情報を得ることができる、このカタカナ入力漢字かな混り文出力方式を広く採用することが望ましいとの結論を得ました。

ここに、今年度の研究実施にご尽力ならびにご支援を賜わたった、若松清司、上田隆実、船崎武男、矢田光治、高森寛、杉本太郎、後藤根男、野添義穂の各氏および関係各位に心より感謝の意を表しますとともに、この報告書が各方
面に活用されわが国情報処理産業の発展の一助として寄与できれば幸いに存じます。

昭和49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
# 文書情報処理に関する調査研究

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>目</td>
<td>次</td>
</tr>
<tr>
<td>1. 総 論</td>
<td>1</td>
</tr>
<tr>
<td>1.1 研究の目的</td>
<td>1</td>
</tr>
<tr>
<td>1.2 報告書の概要</td>
<td>7</td>
</tr>
<tr>
<td>2. 日本語情報の分析</td>
<td>9</td>
</tr>
<tr>
<td>2.1 漢字の役割</td>
<td>9</td>
</tr>
<tr>
<td>2.2 漢字かな混り文出力の必要性</td>
<td>14</td>
</tr>
<tr>
<td>2.3 固有名詞処理に対する需要</td>
<td>19</td>
</tr>
<tr>
<td>2.4 公文書処理の考察</td>
<td>23</td>
</tr>
<tr>
<td>2.5 機械化の諸問題</td>
<td>29</td>
</tr>
<tr>
<td>2.6 適用分野</td>
<td>31</td>
</tr>
<tr>
<td>2.7 技術動向</td>
<td>39</td>
</tr>
<tr>
<td>3. 日本語情報処理のためのカタカナ入力方式</td>
<td>55</td>
</tr>
<tr>
<td>3.1 カタカナ入力漢字かな混り文出力システムの必要性</td>
<td>55</td>
</tr>
<tr>
<td>3.2 カタカナ入力方式へのアプローチの方法</td>
<td>59</td>
</tr>
<tr>
<td>3.3 カタカナ入力方式による日本語情報処理の実験的考察</td>
<td>62</td>
</tr>
<tr>
<td>4. 固有名詞処理のためのカタカナ入力方式</td>
<td>83</td>
</tr>
<tr>
<td>4.1 固有名詞の構成要素</td>
<td>83</td>
</tr>
<tr>
<td>4.2 固有名詞の変換テーブル</td>
<td>84</td>
</tr>
<tr>
<td>4.3 固有名詞の漢字分析</td>
<td>89</td>
</tr>
</tbody>
</table>
5. 分野別の漢字および漢字単語の分析 .......................... 105
   5.1 大学内文書に使われる漢字・熟語調査 ...................... 105
   5.2 医学分野における用語調査 ............................ 118

6. カタカナ入力漢字かな混り文出力システムの実験 ............... 143
   6.1 システムの概要 ......................................... 143
   6.2 システムの利点と効果 .................................. 145
   6.3 カタカナ入力システム .................................. 148
   6.4 カタカナ漢字変換システム ................................. 152
   6.5 漢字かな混り文出力編集処理 ............................. 163
   6.6 変換単語テーブルの決定 ................................. 170

7. 文書情報とマイクロフィルム .................................. 173
   7.1 情報量の拡大と多様化 ..................................... 173
   7.2 マイクロフィルムのあらまし .............................. 176
   7.3 マイクロフィルム用機器 .................................. 178
   7.4 マイクロフィルムによる文書情報検索システム ................ 202

8. 今後の課題 ..................................................... 205
1. 総論
1. 総 論

1.1 研究の目的

昭和47年度の事業報告書「階層構造の情報処理システム」において、日本語情報処理システムにおける新しい処理方式をとりあげ、情報の価値を高めるための工夫の1つとして当財団で考案したカタカナ漢字変換の処理方式を紹介した。考案した方式では、日本語情報をカタカナに変換し、カナコードで情報の蓄積と伝送を行ない、漢字かな混り文で出力するようになっている。それでは、効率の尺度として単にカタカナから漢字への変換率を採用せずに、いわゆる出力情報を利用する者の理解度を高めることに着眼したカタカナ漢字変換システムの有用性を強調した。情報の価値は、必ずしも変換率の高低によって決まるのではなく、変換率が低くて低くとも、専門分野の知識が深ければ、漢字に変換できなかった用語を大字のかな文字で出力するという工夫だけで、理解度の高い日本語情報を利用者に提供できることを例示するとともに、人間の理解力に応じた出力情報を提供できるようにするためには、カタ漢字変換テーブルを階層化する必要があることも指摘した。当財団で考案した方式でも、時間をかけさえすれば変換率を改善して専門知識の浅い人達に理解度の高い出力情報を提供できるとはいうまでもない。しかしながら迅速に情報提供を行なうためには、トッパから現場までの各レベルごとに漢字で出力しないと誤解を招く恐れのある重要な用語を十分に吟味して、変換テーブルの充実をはかる必要がある。そこで同音異義語の判別回数を少くするための手段として、漢字述語の区切りを原則として2つにした変換単語テーブルを作成し、変換テーブルのなかにどのような漢字を最小限いくつ単独に用意しておけば、あらゆるレベルの管理者に理解度の高い出力情報を提供できるかを解明して、有効な日本語情報処理システムを実現するための手掛りを示した。
このような昨年度事業の経緯にもとづいて昭和48年度の事業では、現在各機関において開発されている漢字けん盤入力方式による日本語情報処理システムを調査・検討した結果、一般的漢字けん盤入力方式では規模が大きく、操作も複雑で人間工学的問題があることを究明した。このようなわが国においては、日本語文の合理的な入出力方式を確立することがもっとも重要な課題であるということを認識した。このような技術動向を背景として当財団では、日本語文入力方式の実証的考察を行ない、カタカナけん盤と同音異義語の判断機能を組合せた日本語文入力方式による文書情報処理システムの開発を進めめた。すなわち、この研究ではカタカナけん盤によって入力した情報を漢字かな混りの日本語情報に文書化して出力する、いわゆるカタカナ入力・漢字かな混り文出力方式の実験システムを検討し、日本語情報処理システムを合理的に運用するためのソフトウェア技術を究明した。また、文書情報の蓄積と検索を効率的に運用するためのコンピュータ・マイクロフィルミング方式に関する調査研究も行なった。

なお、わが国では、"情報"という言葉の概念が不明確なために、"文書情報"とか、"文字情報"とか、"日本語情報"といった形で"情報"という言葉がきわめてあいまいに使われている。

そこで、"情報(Information)とコミュニケーション(Communication)の定義づけ"について述べることにする。

"情報"という概念は、かなり抽象度の高いもので、だれでもが納得のいく厳密な定義は難しいようである。しいて問われれば、辞書に書いてあるように、それは「ことがらの知らせ」、「事情の知らせ」であると考える以外にないかもしれない。しかし、それではながら、事情の知らせとは何のことかという問いを提起するだけで、あまりよい解答にならない。それにして、日本語においては、情報という言葉が便利なためか、いささか乱用気味といえる。便利な言葉といったわけは、とくに、他の言葉
と容易に結びついて新しい熟語を形成することが可能だからである。たとえば、文字情報、文書情報、言語情報、パターン情報、音声情報、日本語情報など□□情報という具合にどんな言葉を前に持ってきてもううまく結びついてしまう。そのほか、情報化、情報化社会、情報処理というように別の言葉を後につけて熟語を形成するケースも多い。情報処理という言葉は英語の「Data Processing」の訳語であり、情報化社会はダニエル・ベルのいった「The Post Industrial Society」にあたるといえる。このように英語では、別にInformationという言葉が使われていない熟語でも、日本語に訳されると情報という言葉が使われることが多い。日本語の情報という言葉はそれだけ広い概念をもっているのである。言語情報とか日本語情報に対応する英語の訳語といっても、Language InformationとかJapanese Informationなどという熟語は通用しないだろう。

情報（Information）と類似的の言葉にコミュニケーション（Communication）という言葉があるが、その概念についても多くの人々によっていろいろな定義がなされているのは、日本語においても英語においても同じでおり、コミュニケーションの定義づけにはかなりの混乱がみられる。

ここでは、情報とコミュニケーションを対比しながら、2つの概念を整理してみよう。

まず、情報の概念であるが、一般に情報というときには送り手(Sender)と受け手(Receiver)を想定し、伝達されるものは知識であるとされている。また、伝達される内容は知識であれば何であれ、記号系列に変換されていることを前提にしている。たしかに、手振り足振りで何かを知らせても、その手振り足振りは記号であることは違いならない。

いっぽう、ほかの定義によれば、意思決定とか問題解決しようとする主体にとって、判断の助けとなるものはなんでも情報であるというものもあ
る。この定義では送り手がいなくても、情報は存在する。情報は記号系列に変換された知識でなくてもよく、フィーリングとか何かどのでもよいということになる。もっとも、フィーリングや何かあっても、それらが意思決定や問題解決に役立つとなると、それは立派な知識であるといい出す人も出てくるかもしれない。

情報に関するこれらの定義のうちで、どれが妥当であるかについてはここではあまり論じないことにする。

企業活動など人間の組織活動における情報伝達や情報システムのあり方を究明する必要に情報というものを論ずるときは、①送り手と受け手の両方がいて知識が伝達される状況での情報、②意思決定における判断の素でとしての情報とを区別して考える方が混乱がなくてよい。さらに、前者の状況における情報は「コミュニケーション」と区別して情報を考えなければならないし、後者の状況における情報は「データ」と区別して考えるべきである。

一般に、情報という言葉には、送り手と受け手の間、とくに、人間と人間の間に、なんらかの知識、すなわち意味内容が伝達される状況が想定されている。しかも、送り手が伝えたたい知識や意味内容は記号系列に変換され、また、受け手がその記号系列を知識や意味内容に変換して、はじめて、理解されることになっている。記号とは前もって意味が与えられているシンボルである。言葉、文字、数字、図形などが普通に使われる記号であり、数式のように論理的に厳密なものであれば、自然言語のように論理的にはルーズなものもある。しかし、大切なことは、送り手と受け手の間で正しく意味内容が交換されるように、すなわち、情報というもののが成立するためには送り手と受け手の間にかなり共通の理解、あるいは知識や合意が存在していなければならないことである。少なくとも、記号の意味については共通知識が必要であり、たとえば、ある数字がどこかの会社の電話番号なのか、自分の会社の売上げ高なのか、株価なのか、ということについては共通理解が双方の間で明確にされていなければならない。伝達された
数字や、コンピュータが打ち出した数字がどういう意味をもち、それらがどのように活用されるかについても共通理解がなければならない。ある地域に住む世帯の平均家計支出がいくらであるとメッセージが送られても、それが送り手から受け手への知識の伝達となるには、送り手と受け手の双方に家計支出とか平均値という概念についての共通知識があるから理解できるのである。しかも、このメッセージが時間的にいつ伝達されるかについても、双方間に合意があることが多い。合意は暗黙の合意であることもあるが、これが無視されるとこのメッセージは情報としての意味が失われることもある。情報は知識の伝達であるという意味での情報が成立するには、まず、その伝達される知識を記号系列に変換する送り手と、それをさらに変換して意味内容をくみとる受け手の双方に共通の認識とか知識の枠組みが必要である。それら共通のものがなければ、送り手の伝えようとする意味内容が伝わらないか、伝わっても送り手が伝えようとしたものとは異った意味内容が受け手に伝えられる可能性が大きくなくなる。

組織だった人間活動とかシステムにおいて構成メンバ間の情報交換が正
確かつ効果的に行なわれるためには、それら構成メンバが事前につきのような多くのものを共有していなければならない。その共有されるべきものは、認識や知識だけにとどまらず、目的、期待、予想、関心などをはじめ、ときには、経験や習慣、価値観や世界観まで含まれる。このように、組織やグループ内の構成メンバが共有するもの、共通性を促進しようとする努力・行為はコミュニケーションと呼ばれるものである。コミュニケーション（communication）という言葉はラテン語のcommunicareからきたもので、これは共通（common）という意味である。

正午を知らせるベルの音や一日の就業時間の終了を告げるベルの音が情報としての意味をもつのは、事前に、正午と夕方の時にはベルがなるという合意、習慣があったからであり、ベルの音が人々のどんなアクションにつながるかについてもすでにコミュニケーションが確立しているからであ
ル。コミュニケーションをともなわない状況における単なる情報発生の例は、大学における大衆団交であろう。ここでは、互に向かいあっていたながら情報の送り手の意味したメッセージは受け手に伝わらないか、ときには、全く反対の意味内容に受けとられることがある。このようにコミュニケーションが貧弱な状況においては、送り手がある意味や知識を記号系列、たとえば言語にして送っても、受け手のそれは送り手が意図した意味や知識とはかなり異ったものを受けとるのが普通である。そこでは、知識の伝達といろ意味での情報は流れていない。しかし、このように送り手の意図した意味内容ではないものを受け手が受け取った場合でも、それが受け手にならかな反応を惹き起し、その意思決定やアクションに影響を及ぼすなら、その受け手は情報を受けたことになる。このような場合は、そのメッセージの送り手者が意図した意味や知識は伝わらないのであるから、送り手は単にデータを送ったにすぎない。その場合には、データを受け手なりの知識、観点、偏見で評価して、なんらかの意味づけをしたのである。

最近は、全国情報ネットワーク化と経営情報システム化（MIS）などコンピュータを活用しての情報システムの整備が図られているが、それらの情報システムが、利用する人間集団や組織にとって役立つものとなるかどうかは、結局のところ、組織構成メンバ間のコミュニケーションが充実しているかどうかにかかっている。このことはあまり強調されていないようである。効果的な管理情報システムを構築するための前提条件とし、だれが、なぜをするために、どんな情報を、いつ、必要とするかについて共通理解・合意がなければならない。そこから得られた情報がどんな応用のされ方をするのかについて、共通認識が必要である。組織内のどの管理者がどんな意思決定に、どんな情報を必要とするかがわからないならば、その意思決定者に有益な情報を提供することはできない。有益で効果的な情報システムは豊かなコミュニケーションを前提として成立するのである。

これまで考えられてきた、コンピュータを主役とする情報システムでは
その入力も出力も記号系列であり、その中での情報操作も主として論理操作である。これまでの情報システムは記号、数字、論理だけで機能するきわめて非人間的な組織といえよう。そのようなシステムからどんなに多量の出力が打ち出されてもコミュニケーションの貧しい環境においてはその出力は利用者にとって情報というよりもむしろ単なるデータにすぎない。管理者や意思決定者がそれぞれ異なる認識、概念、価値、目的などで情報システムからのデータに思い思いの意味づけを行なったとしたら、その情報システムをコミュニケーションに役立たることは不可能である。むしろ、コミュニケーションを一層混乱させる事態をひき起こさなければならない。したがって、企業や組織における情報システム化の進展は必ずしもそのままではコミュニケーションの充実には結びつかない。情報システム化の進展のためには、コミュニケーションの問題を真剣にとりあげる必要がある。

1.2 報告書の概要

当財団では、昭和43年度以来経営情報システムに関する研究の成果として、43-S004「経営情報システムの理論とサブシステム」、44-S003「中堅企業のMIS構造」、45-S004「機械工業の生産情報システムにおける意思決定機構の解析」、45-S001「経営予測のためのデータ・マネジメント」、46-S006「記事情報検索のためのデータ・マネジメント」、46-S007「機械工業における制御情報システム」、47-S001「階層構造の情報処理システム」といった一連の報告書、および47-R001「日本語情報処理の技術動向調査報告書」をとりまとめている。

今年度の報告書は8つの章からなっており、各章でそれぞれつきのような内容をとりまとめている。

第1章「総論」では、研究の目的を記述し、「情報」という言葉の背景を究明して、文書情報、日本語情報、文字情報などの定義づけを明らかにし
ている。

第2章「日本語情報の分析」では、日本語情報処理システムを開発することの意義や必要性について述べるとともに、日本語情報処理の適用可能な分野、および現状の技術動向について述べている。

第3章「日本語情報処理のためのカタカナ入力方式」では、日本語情報処理を行なううえでもっとも重要な課題となっている入力方式について考察するとともに、カタカナ入力方式の利点について解説している。

第4章「固有名詞処理のためのカタカナ入力方式」では、日本語情報のなかでももっとも重要かつ高いといわれている固有名詞についてその構成を概説し、固有名詞用の変換テーブルのあり方、および固有名詞である事業所名・人名・地名についての漢字単語を分析している。

第5章「分野別の漢字および漢字単語の分析」では、専門分野別の変換テーブルを作成するための準備として、大学内文書と医学文書に関する漢字および漢字単語の調査結果を考察している。

第6章「カタカナ入力漢字かな混り文出力システムの実験」では、当財団が考案した、カタカナ入力方式による日本語情報処理システムに関する実験結果を概説している。

第7章「文書情報とマイクロフィルム」では、情報蓄積の1つの方法としてマイクロフィルムをとりあげ、文書情報の検索システムへの適応性を述べるとともに、マイクロフィルムの技術動向を紹介している。

第8章「今後の課題」では、当財団が考案したカタカナ入力方式による日本語情報処理システムの展望について述べている。
2. 日本語情報の分析
2. 日本語情報の分析

2.1 漢字の役割

コンピュータをはじめとして、その周辺装置や通信回線、および端末機器の発達や進歩は、あらゆる分野において情報処理のEDP化を急速に拡大してきている。

このことは、コンピュータがソロバン代りの演算処理を目的として利用されてから、情報処理システムとしてのオペレーションな日常業務の処理、管理資料の作成、OR的手法などを利用したいろいろな統計資料の作成、経営計画に関する戦略的な資料の作成など幅広く利用されるようになってきている事実をみても理解できよう。

しかしコンピュータを利用したアプリケーション・システムの発展・拡大の過程をふりかえると、その基礎をなしているのは、各種の数値情報であると思われる。存在する大量のデータの中から必要な情報をいかに効率よく選択、分類、処理（演算）して、利用するかが企業にとっても個人にとっても重要な問題であり、コンピュータを中心としたEDPシステム発展に大きな影響を及ぼすものの1つであるといえる。コンピュータ処理の中心となっているのが数値計算であるが、これは情報の量全体や質からみるとほんの一部にすぎない。そのため、今後コンピュータを利用して処理する分野をさらに広げていくためには、文書情報を利用する処理に期待が大きい。このため文書情報の処理を必要とする分野の研究がさらに一段と進められることは、もはや疑う余地のない所である。

このことは、コンピュータが米国で発明されたものであるということと、数字や英字のように文字の数や順序関係がはっきりしている場合には、人間用の情報からコンピュータ用の情報への変換（コード化）が行ないうえすと、数値計算を中心とした処理において、処理遂行の際に、数字や英字、
特殊記号が使用できれば十分活用することができるようがげられる。

このため給与計算、販売管理、在庫管理、生産管理、財務管理、会計管
理など、いろいろな分野の事務処理において、情報の処理面では効率が良
くても、出力（印刷）の面では、コンピュータの処理結果を表示する文字
に対して不満が出てくる。それは、従来の情報処理が、処理面の効率追求
にあたってであり、この追求が今日の進歩をもたらしてはきたけれども、
欧米と日本との比較では問題にならない文字表現の問題が、最近とく
に大きな問題となってきている。

これは、漢字かな混入文を扱う日本人であるということと、コンピュー
タの利用分野が広がってきたこと、コンピュータに人間が使われるのでな
く、コンピュータはあくまでも人間が情報処理を行うための道具である
か、それを使用する上でのいろいろな制約は、人間の側からよりも道具
としてのコンピュータ側から人間が満足できるような形に接近してくるべ
きだという理由で起こってきた。

これは日本人として当然の要求であり、従来の情報処理に対する新しい
見方であるということができる。我々が日常の生活のなかで用いたり、企
業活動をする場合に使用しているほとんどの情報が日本語であることか
らも漢字かな混入文による情報処理が当然要求される。しかし、わが国
へコンピュータが導入された当初は、コンピュータは、数字、英字、特殊
記号しか使用できないものという考え方多いため、日本語の文章を英字
やローマ字で処理を行ない、その後ようやく漢字かな混入文による表現に
対する要求が強くなって、英字と同じ表音文字であるカタカナによる表現
が可能となった。

これはコンピュータによる情報処理の出力表示の方法としては、非常に
大きな前進であったということができる。

従来、手書きやゴム印などを使用して処理されていた給与明細書の人名
や、販売、購入関係の台帳に使われる顧客や下請業者の名称、および住所、
製品名や規格寸法などの処理や官庁関係の届出資料の文字情報の１部（カナ文字の使用が許されている範囲のもの）について、それぞれカタカナを利用し、事務の省力化がなされるようになった。とくに、企業においては、このほかにも、販売・購買・在庫・生産などの分野で使用されている各種の伝票類など、カタカナの利用される範囲はかなり広い。またコンピュータのハードウェア、ソフトウェア、および端末機器や通信技術の進歩によって、情報サービス（文献検索や出版物のサービス、ダイレクトメールなど）の分野においてもＥＤＰ化が進み、カタカナの使用が多くなって来た。

しかし、これまでコンピュータの出力（印刷）に対する不満が解消されたわけではない。日常使用している日本語の文章は、このようなかたかた文字（ひらがな、カタカナ）ばかりではなく、新聞、雑誌などの文書はすべて漢字かかな混じり文で表現されており、生活環境のなかには、漢字かかな混じり文が、感覚的にも感情的にも一般的な表現方法として受けとられ、かたかた文字だけの表現に対しては、とくに幼い子供でないかぎり異和感を覚える。

また、かたかた文字は表音文字であるから個々の文字には一定の音だけがあって意味はなく、いくつかの文字を組み合わせて言葉（単語）となったときはじめて意味が表現されるのに対して、漢字のような表意文字は、個々の文字に固有の意味があり、文字そのものが、単語としての音をとなえている。

そのため、かたかた文字は、漢字のような表意文字とちがって文字の意味や感情や情緒など文字としてのニュアンスが伝わらないので、日本語としての異和感が生じるものと思われる。かたかた文字だけの表現方法を用いた場合には、同音異義語などの取り扱いができないため、誤った情報が伝達されたり、意味を取り違えるなどの問題が生じる懸念がある。

そこで最近になって、コンピュータの出力情報として漢字の必要性が叫ばれ、漢字処理のシステムや漢字処理のための入出力装置がコンピュータ
メーカーなどで数多く開発されてきている。このように我々が、コンピュータに対して持っていた“処理面”の要求のほかに、より使いやすいという“満足面”の要求が強まってきていることと、従来の数値計算中心の情報処理から文章情報を中心とした情報処理の分野まで、その適用面が広がってきたためであるといえる。

こうした企業活動における重要な機能である情報処理の分野において、文書情報とくに漢字による表現が重要視されるようになったのは、
① コンピュータによる情報処理分野の拡張
② 文書情報処理の省力化
③ 文書情報処理の迅速化
④ 文書情報処理の正確性
といった面からも理解できる。

これは、各企業とも十分認識しているとは思うが、コンピュータ・メーカや報道、出品など限られた業種を除いてあまり普及しておらず、ここ2〜3年前頃から製造業など一般企業においても徐々に利用されるようになってきた。

しかし、前述したように企業活動のほとんどどの情報は、日本語であり、日本語を処理するうえで切り離すことのできない漢字を混じり文が、その必要性を認められながらも普及しなかった理由としては、
① 漢字を混じり文を処理するのに適当なハードウェアやソフトウェアがなかった。
② ハードウェアがあっても、高価であるため、入手でそれを処理する労働力と比較してあまり効果がない。
③ ハードウェアの制約（文字数や出力表示方式など）が多く、特定の用途以外に利用できない。
④ ハードウェアの取扱いが非常に専門的で、高度の技能的熟練度が要求される。

— 12 —
⑤ ハードウェアを取扱う熟練技能者の教育訓練が十分できないため、
なかなか補充するのがむつかしい。

があげられよう。そのため、やむなく英数字の組合せやカタカナを使用するか、それが使用できない箇所については、従来通り人間側に負担をさせてきた。

いっぽう、最近になってハードウェアの進歩により、処理時間の速さ、
大容量の活字が扱える（4,000〜6,000字）漢字入出力装置が開発されたことや数百万の記録を数千年の文字に展開するような制約（書体、
文字の大きさ、横書き縦書きなどの制約）が比較的自由に選択できるハードウェアも開発され、入手できるようになったことが、最近の漢字入出力装置の普及につながっている。

だが漢字入出力装置が高価であるという問題は、まだ解決されていない。
これは漢字情報を直接入力することに対する問題があり、ハードウェアとしての
漢字入出力装置が高価になる原因になっていると思われる。

また、表現文字としての漢字の宿命であるといえるが、英字やかな文字
のような表現文字に比べて
① 文字の字数が非常に多い。
② より長くなるのである。
③ 一定のルール（英字の場合の A B C 順、かな文字の場合の五十音順）
で分類したり、配列したりすることがでなければならない。
④ 文字の読み方が、一定していない。

などがあげられ、これら漢字の特徴が、コンピュータで情報を処理する上
では短所となっている。

このため漢字をコンピュータで処理するためのコード化（ビットの構成）
を考える場合、
① 取り扱う字数
② 分類方法
③ 配列
（1）読み方

（2）同文字

などをどうするかが問題である。現在のところこれらは絶対的なルールがないため、ハードウェア・メーカ独自の考え方で処理されている。そのため、各メーカーごとに独自のビット構成をとっている。ハードウェアの互換性がなく、しかも日常使用されている漢字をカバーするためには、どうしても大がかりで高価なハードウェアになってしまうという事情がある。

2.2 漢字かな混り出力の必要性

今日、多くの企業において、経営計画、販売、購買、生産、人事、会計、財務、技術、研究、設備保全などいろいろな分野において、そこで発生・伝達・処理される情報の処理手段としてコンピュータが広く使用されている。

そして、これらいろいろの分野における情報は、

① バッチ処理
② リモート・バッチ処理
③ リアル・タイム処理
④ タイム・シェアリング処理
などによって処理されている。

またこれらの情報は、

1）数値情報
2）文章情報
3）図形情報

に分類されるが、いずれに対しても業務の目的、性質、ハードウェアの制約などにあわせて使用されているのが現状であろうと思われる。

情報処理の中心になってきたのは数値情報であるが、企業活動における情報処理の手段として、E・D・P化が進んでオフラインからオンライン、計
算処理から情報検索や図形処理へとその適用範囲が広がってくるにしたがって、従来あきらめていた漢字かな混文出力の問題（日本語による出力）にもメスが入れられ、かな文字や漢字の入出力装置が、作られるようになった。

このことは、情報を処理する内容が数値計算を中心としたものの場合には、取り扱う文字も数字を中心としたものであり、その補助的な役割として（タイトルや項目的説明など）英字やカタカナをわずかに使用すればよかったが、情報処理の適用範囲が広がり、情報検索やタイム・シェアリング・システムが取り扱われるようになったり、漢字の入出力装置が開発されてくると、日常使っている文字を自由に使用して、情報を処理したいという要求はますます強いものになってくる。日常生活を支えている情報の基礎は漢字にある。漢字のもと特長をぬきにしては、わが国の情報を取り扱うことは考えられない。従来、情報処理の効率化が強く要求され、重要視されていたため情報処理における漢字の問題は、印字効率の低下を招くなどの理由により軽視されていたが、コンピュータによる情報処理の適用分野が広くなってくると、漢字の問題を軽視することはできなくなり、コンピュータによる情報処理に漢字の問題が取り入れられてきた。わが国の"情報"の中心が漢字を主体とした漢字かな混文文であることを考えてみてもその重要性は理解できるし、今後ますますその役割の重要性が理解されるであろう。

わが国では、身のまわりの情報が、発生、整理、計算、決定、実施の段階でそれぞれいかに情報が処理され、その処理を行なう場合の行為としてどんなものがあり、その行為がEDP化の対称として、どのような適用業務が考えられるであろうか。その行為を理解するために、各段階における行為の要素を表2－1に示す。

このように情報を処理する場合、受け取った情報を正しく理解し、その情報にもとづいて計画をたてたり、決定をしたり、実施をするという行為
表2-1 人の行為要素

<table>
<thead>
<tr>
<th>行為段階</th>
<th>行為要素</th>
<th>行為段階</th>
<th>行為要素</th>
</tr>
</thead>
<tbody>
<tr>
<td>発生段階</td>
<td>(受ける)</td>
<td>（作る）</td>
<td>分析・研究</td>
</tr>
<tr>
<td></td>
<td>受付・受理</td>
<td>作成</td>
<td></td>
</tr>
<tr>
<td></td>
<td>受入・戻入・振替</td>
<td>作票・作表</td>
<td></td>
</tr>
<tr>
<td></td>
<td>受領・回収</td>
<td>調整・調剤</td>
<td></td>
</tr>
<tr>
<td></td>
<td>購入</td>
<td>(集める)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>検収</td>
<td>収集</td>
<td></td>
</tr>
<tr>
<td>理段階</td>
<td>(書く)</td>
<td>編集</td>
<td></td>
</tr>
<tr>
<td></td>
<td>記入・記帳・記録・消込</td>
<td>分配・配分・割当</td>
<td></td>
</tr>
<tr>
<td></td>
<td>筆記・速記・転記</td>
<td>仕訳・分類・分析</td>
<td></td>
</tr>
<tr>
<td></td>
<td>トレース・写書・浄書</td>
<td>(整える)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>タイプ・印刷</td>
<td>整理・整頓・整備</td>
<td></td>
</tr>
<tr>
<td></td>
<td>起案</td>
<td>準備</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(計算する)</td>
<td>練込・切抜・添付・挿入</td>
<td></td>
</tr>
<tr>
<td>理段階</td>
<td>計算・算出・集計</td>
<td>移管</td>
<td></td>
</tr>
<tr>
<td></td>
<td>検算・精算・決算</td>
<td>修繕</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(調べる)</td>
<td>清掃</td>
<td></td>
</tr>
<tr>
<td></td>
<td>調書・再書・下調</td>
<td>(話す)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>照合・対照・校正・読合せ</td>
<td>連絡</td>
<td></td>
</tr>
<tr>
<td></td>
<td>照査・精査・点検・診断</td>
<td>(考える)</td>
<td></td>
</tr>
<tr>
<td>計画段階</td>
<td>測量・測定</td>
<td>計画・企画</td>
<td></td>
</tr>
<tr>
<td></td>
<td>検査・試験・選考</td>
<td>発案・立案・考案</td>
<td></td>
</tr>
<tr>
<td></td>
<td>検収</td>
<td>見積</td>
<td></td>
</tr>
<tr>
<td></td>
<td>審査・審議</td>
<td>設計</td>
<td></td>
</tr>
<tr>
<td>行 為</td>
<td>行 為 要 素</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>決 段</td>
<td>調整</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>研修</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>検討</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>研究</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>交流・打合・折衝</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>討議・合議・会議</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>査定・算定</td>
<td></td>
<td></td>
</tr>
<tr>
<td>（決める）</td>
<td>決定・専決・決裁・採択</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>承認・確認・確認</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>評定・認定・算定・決算</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>指定</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>検印・捺印</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>契約</td>
<td></td>
<td></td>
</tr>
<tr>
<td>実施</td>
<td>命令・発令・指令・指示</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>監督・観察</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>立会・督促</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>管理・統制</td>
<td></td>
<td></td>
</tr>
<tr>
<td>（伝える）</td>
<td>司会</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>（送る）</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>送付・返送・発送・運搬</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>（出す）</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>発信・呼出</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>払出・振替・支払・納付</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>発行・発表</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>提出・出願・届出・掲示</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>派遣・出席・参加・出席</td>
<td></td>
<td></td>
</tr>
<tr>
<td>（配る）</td>
<td>交付・回付・回覧・供覧</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>配付・配布・配給</td>
<td></td>
<td></td>
</tr>
<tr>
<td>（知らす）</td>
<td>通知・通牒・通告・通達・通達</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>社連</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>報告</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>回答・応答</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>申告・上申</td>
<td></td>
<td></td>
</tr>
<tr>
<td>（整える）</td>
<td>保存・保管・格納</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
がなされる。しかしこれらの段階における行為は、ほとんど個人のなかだけに処理されるのでなく、その個人に関連するまわりの人達との間で情報の伝達という行為がなされる。このため受けとった情報を正しく理解するために、書いたり、計算するという行為に始まって、調べたり、分類したり、作表するというような情報整理の行為が行なわれる。さらにこの整理された情報をもとに計画し、決定し、実施する。

このように受信取った情報を正しく理解し、それにもとづく意思決定を正確に他の人々に伝達するために、検書をしたり、タイプしたりしているし、資料を作成したり、検討、調整、決定、指示、発表、報告も行なっている。そしてこの一連の情報処理に用いられているほとんどすべての情報は、日本語（漢字が混じり文）を中心とした情報で行なわれている。このためわが国においては、情報を正しく理解して意思決定を行ない、誤りなく正確に情報を伝達するには、漢字の問題を無視して情報を処理することができないことは言うまでもない。

そのため、コンピュータを利用した情報処理の取扱いの中でも漢字の問題を取り上げられ、漢字用の人出力装置が開発されるようになった事は当然の問題である。このコンピュータを利用した漢字処理の問題は、従来人手によって処理していた、

① 文書情報の検索
② 人名簿の処理
③ ダイレクト・メール
④ 証券、株式の処理
⑤ 文書管理
⑥ 見積作成
⑦ マニュアル作成
⑧ 官庁提出用資料の作成
⑨ 営業・経理・購入・在庫などの台帳作成

—18—
各種帳票類の発行
などEDP化の新しい適用分野が開発できよう。
またこれら情報処理における漢字出力の問題は、従来コンピュータの出
力として使用されてきたローマ字やカタカナなどに比べて、情報を受けと
る側から見て読みやすくしかも理解しやすいばかりでなく、従来すでにED
P化されている他のシステム（人事管理、営業管理、会計処理、購買管
理、在庫管理、生産管理などの情報処理システム）と一元的に連続したもの
として処理できるメリットが得られる。
このことは、コンピュータで処理された出力レポートをわずかざ人手を
かけて漢字に書き直したり、タイピングで打ち直す必要もなくなる。このよう
に一度コンピュータで処理した情報を読みにくいとか読み誤りの可能性が
あるために人手によって再加工するといった問題は、人名、会社名、住所、
製品名、部品名、規格、工程名など比較的固定化されているながらも、繰り
返し性の強いものが多く見受けられ、多大な労力と時間を費すものが多い、
このような情報の再加工の問題は、漢字出力が可能になればなくなるし、
漢字処理の問題は、すでにEDP化されているシステムに組込むことによ
り解決することができる。
また人名簿や履歴書、マニュアル作成作業のように漢字を必要とするニ
ーズは大きくけれども、それを処理するための労力や時間が問題となるも
のは、コンピュータによる処理を考えず、最初から人手や印刷によって行
なわれていることが多い。

2.3 固有名詞処理に対する需要
わが国における文書の情報処理（それは当然漢字処理の問題に繋がるも
のである）のうち、大量事務処理の範囲に属するものを取り上げてみると、
その筆頭になるべきものは、なんといっても各種の名簿の作成と郵便物な
いしはこれに類する往復文書などのあて名の印刷であろう。

— 19 —
参考までに郵政省の昭和46年度郵政統計年報（郵便編）について、内国通常郵便（小包は含まれない）の引受件数をみると、昭和31年度の5,000万1,220通に対し、昭和46年度は1,175万9,449通と、約15年間に2倍以上の伸びを示している。また、昭和46年度における内国郵便（通常郵便）の料金納付形態別件数の割合をみると、切手貼付のものが56％、料金別納および料金後納によるものが44％で後者の比率が予想外に高い。もちろん、料金別納などの場合でも、あえて名印刷機を用いたり、あるいは手書きによるもののが ASAP的に高いとは思われるが、コンピュータの普及、人件費の上昇などにより、コンピュータからの出力によるものの件数が相当上昇しつつあるものと推定される。

ところで、名簿およびあて名を構成する主な要素は、(1)個人、法人または各種の機関の名称と、(2)それらの住所（または所在地・本籍など）を示す地名であることはいうまでもない。そこで、現行の各種事務処理システムおよびコンピュータの利用状況からみて、地名、人名、企業名および事業所名（官公署名を含む。以下「固有名詞」と総称する）の漢字による出力に対する利用の可能性およびその程度を、アプリケーション別に推測したものが、表2-2である。同表自体、未だ試案の域を出ないものであるが、これらの固有名詞、とくに地名および人名の処理に対する需要がきわめて強いことは、疑う余地がない。
<table>
<thead>
<tr>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
</tr>
</thead>
<tbody>
<tr>
<td>表 2.2 集合による医療名の系統処理システムに対する需要の見通し（試案）</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>サイドバー</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
</tr>
</thead>
<tbody>
<tr>
<td>表 2.2 集合による医療名の系統処理システムに対する需要の見通し（試案）</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>部</td>
<td>屋</td>
<td>用</td>
<td>途</td>
<td>用す</td>
<td>する</td>
<td>被</td>
<td>各</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.4 公文書処理の考察

わが国における情報処理の分野では、近頃、文字情報の処理あるいは文書情報の処理といういわゆる日本語の情報処理が、いわゆる情報処理の分野として、漢字入力装置の開発、漢字の字種や、コードおよびキーボード上の配列の統一などの動きも、逐次本格化しつつある。

しかしながら、日本語情報処理システムの開発にあたって、もっとも重要な問題点の1つである文書そのものの性質の説明や、形態的な分類についての研究は、情報処理システムとの関連づけという点に着目する限りにおいては、未着手の状態といえる。

第1章「総論」で、「情報」という言葉の背景を完明したがここでは、各分野の文書の中でも、情報処理の対象としてもっとも適切とみられる公文書を素材として、カタカナ入力方式を採用した日本語情報処理システムの開発にもとめよう諸問題をとりあげ、これを通じて、日本語情報処理の全般にわたる問題解決の端緒を得ることを狙って考察したものである。

2.4.1 公文書の定義

公文書も、文書の一種であることは、いうまでもない。したがって、公文書の性質を明らかにするためには、順序として、まず文書とはいかなるものであるかを考察する必要があるろう。

(1) 文書とは

「文書」という言葉の定義については、いくつかの説明が行なわれているが、つぎに代表的な例を2つほど掲げる。

(1) 文書とは、文字その他の記号の組合せによって、思想的な意味を表現している有形物をいい、その記号は、電信符号や略号でもよく、有形物には、木、石、金属をも含むが、写真、音盤、楽譜のようなものは、文書ではない（末川博編、日本評論社刊、新版法学辞典）。

(2) もっとも普通の用法としては、書面の意味に用いられ、口頭によ
る表示に対するものである（請願第2条、政治資金規正法第6条第1項等）。

訴訟法上では、物語に各人が知覚できる記号によって思想を表示したもの、または紙片、帳簿、布告などに文字その他の符号によって思想をあらわしたもの、と解されている。後者の意味における文書で証拠とすることがあるものを書証という（民事訴訟法第311条〜第331条）。

刑法上では、事実の証明または権利義務を内容とする意識の表示で発言できる符号または象形によって物語に記載されたものをいうものとされ、このうち、発音符号によるものを狭義の文書といい、象形によるものを図画という（刑法第155条、第156条、第158条および第159条）。公職選挙法第142条の文書図画もこの意味である。（佐藤達夫ほか2名編、学陽書房刊、法令用語辞典）

いっぽう文書の内容を構成するものは、「文章」である。「文章」は、思想、感情、事実などを示す1つまたは2つ以上のつながりをもつ文（センテンス）の集まりといい、一般的には、書き言葉であり、文が連続した散文形式のものを指すことが多い。文章が、紙などの記録媒体上に、文字その他の記号の組合せによって表現されたときに、それは文書となる。文は単語および区切り点などの記号で構成され、単語は文字で構成される。

文章は、その執筆の目的からみて、つぎのように分類することができる。
① 書簡文………特定の相手と交信するための文章
② 記事文………事実を記録し、または報道するための文章
③ 解説文………知識を秩序だてて解説するための文章
④ 文芸作品・論説文………読者の思考や行動を促すための文章

上記のような各種の用語の相互関連を図示すると、図2−1のよう
(2) 公文書とは

「公文書」という言葉は、元来、刑法および民事訴訟法において、「公文書」に対する言葉として用いられてきたものであるが、これにも大別して、2 とおりの説明が行なわれている。すなわち、刑法第 155 条第 1 項で「公務所マタハ公務員ノ作ル可キ文書」と規定しているのは、組織体としての官公署と、裁判官、検察官、公証人、普通地方公共団体の監査委員などのように単独で権限を行使する公務員をも区別する趣旨から述べられたものと思われ、法律関係の辞典などでは、同類旨の説明が一般的に行なわれている。これに対し、官公署名による文書であっても、その作成のための書記的作業は、自然人である公務員が行なうのであるから、その作成者を「公務員」であるとし、
「公務員がその職務上作成した文書をいい、それ以外の＜私文書＞に対す」（平凡社、世界大百科辞典）と説明している例もみられる。

行政の分野では、公文書に相当する用語として、「公文」ないしは「交換公文」という言葉が用いられることが多く、昭和21年2月9日の次官会議記録によると「官庁の用語・用字をやさしくする件」においても、「公文用語の手引」という表現が用いられている。また、「公用文書」という表現もあり、当用漢字表（昭和21年内閣告示第32号）の「またがき」に見受けられる。なお、これらに類似する用語としては、「公文文」というのがあり、現在で比較的頻繁に用いられているが、これは、「官公庁用の文章」に対相当するものといえよう。

以上の例に見られるように、司法の分野では、あらゆる文書を、公文書と私文書の2つに区分しており、行政の分野でも、多少の表現の差異が認められる場合があるにしても、大同小異で、これらの考え方、他の分野にも、広く波及しているように見受けられる。

いっぽう、事務管理の立場では、「ドキュメント」、「帳票」、「ファイル」など、文書を意味する言葉が盛んに用いられているが、公文書と私文書というような分類は、特別には問題にされていないようである。

しかしながら、日本語情報処理の立場からは、文書の分類に対する上記のような認識だけで十分といえるであろうか。公文書でないものは、すべて私文書として一括してさしつかえないだろうか。これらの点については、この際改めて検討の必要があると思われるが、結論から先に言えば、あらゆる文書を、まず「業務用文書」と「私文書」に大別し、前者をさらに、「公文書」、「公文」と「事業用文書」に区分するのがよいのではなかろうか。その理由の要点だけをあげれば、第1には、官公庁以外の企業、団体などが業務上作成する文書を「私文書」として、個人が他人と交際上交換する手紙の種類と同一視するのは、明らか
かに不合理と考えられることであり、第2には、それらの事業用文書の内容および形態が、個人間で交換される書簡などよりも、むしろ公文書に近似していることである。

2.4.2 公文書の特徴

公文書と私文書を対比すると、いくつかの相違点が見受けられるが、それらのうち、公文書の特徴としてあげられるのは、つきのようなものであろう。
① 内容に強制力を伴うことが多い。
② 普遍性および公開性が高い。
③ 文書としての寿命が長い。
④ 冗長度が極めて乏しい。
⑤ 形態が統一化されている。
⑥ 取扱手続が制度化されている。
⑦ 作成者者が限定されている。
⑧ 解釈上の共通的な原則が確立されている。
⑨ 階層構造をもっている。

上記のような公文書の特徴は、情報処理システム構成の基本仕様に対し、大きな影響を与えるものであるので、改めて細目にわたる検討を進める必要がある。

2.4.3 公文書の種類と作成機関

わが国には、立法、行政、司法の3つの機関があり、それらの機関の性質によって、作成される公文書の種類も異なる。各機関の種類と、そこで作成される公文書の種類について、相互の関係の略歴を図示すると、図2-2のとおりである。
図 2-2 国の機関と文書の種類

注：点線は、当該機関の内部管理（司法行政など）に関するもので、衆議院規則、最高裁判所規則などがあるもので、これに該当する。

地方公共団体の場合は、図 2-2 のうち、司法に関する部分、政令および府・省令の部分を除いて、国の機関の場合に準ずる。

2.4.4 その他の問題点

公文書を対象とする情報処理システムの設計にあたっては、上記のほか、つぎのような諸問題についても、さらに具体的な検討を進める必要があるろう。

① 公文書についての内容・性質による分類の確立
② 基本単語の抽出のための基礎となる文献の選定
③ 単語表の編成区分の設定
④ 法令名の略語、共通的な接尾語、同音異義語などの取扱い
2.5 機械化の諸問題

日本語文書を書く方の問題としては、まず、書くことに非常に時間がかかることがある。それは漢字の文字パターンがきわめて複雑だからである。たとえば、「漢字」という単語も、表音文字のかな文字（ひらがな、カタカナ）で発音どおりに、「かんじ」と書くながらきわめて簡単であるが、漢字で書けば、「漢」が14画、「字」が5画であるように、合計19画の紙面へのタッチが要求される。もし、一般的なかな文字タイプライタを使っただけに、打つことができる4回で済む。このように日本語文書を書書するのにかかる労力と時間は大変なものであり企業組織、あるいは社会全体としても負担は大きい。これは主として、表意文字である漢字を使うこととかな文字を使うこととの比較であるが、日本語を書くための機械化、すなわち、日本語のためのタイプライタの普及もさまたげられている。和文タイプライタでは、熟練者でないかぎり手で書くよりも操作速度は遅いし、しかもそれを使いこなすには職業訓練が必要である。したがって、和文タイプライタは単なる清書機械にすぎず、日本語文書を書く作業を機械化するという意味でのタイプライタではない。書く作業を機械化する手段としては、言語の表記法に、何らかの表音文字体系採用しなければならない。ある企業では、社内文書にかな文字を採用しているケースがある。このような文字採用して生ずるメリットは、かな文字タイプライタを活用できるので書くことの機械化が可能であるということである。著作家のなかにも、かな文字タイプライタで原稿を書く人が何人もいると聞く。

日本語文書を読むことに関する問題としては、まず、書くことが機械化できない場合手書きの文章では文字の読みやすさがある。これは書く人それぞれが大きな文字、小さな文字、各人勝手な略字、誤字、まちまちの送りが手で書くということが主な原因である。急いで文字を書くときなどは、自分で書く文字の大きさなどコントロールできないものである。そこで読
みやすくするには、字の上手な人に書かせてもあらか、和文タイプライタで打つしかない。しかしながらどちらも非常な努力と時間を要する。書くということも、直筆の文字には必然的に個性的な特徴が入るし、規格化された文字を書くことはできない。組織内の管理情報としては、文書がもっとも内容に書く人の個性があらわれるのはともかく、媒体としての文字は非個性的な規格化されたものが望ましい。媒体まで、個性化されているようでは、読む側にとってはむしろ雑音（ノイズ）に感じる。それよりも、文字は情報伝達の正確性を第1に期すものであり明確で読みやすい記号でなければならない。

日本語は、なんといっても、読むことよりも書くことに問題があり、その表記法のトラブルの元因といえば、表意文字である漢字を使うことにある。まず、第1に、日本語は正書法が確立していないという意味で、世界でも類のない言語といわれる。たとえば、「わかる」、「分る」、「解る」、「判る」、「分かる」、「解かる」、「判かる」など各個人によって書き方がまちまちである。しかも文字種が非常に多いこと、そして、それら多種の文字を識別するために、文字パターンが複雑となって、二重の負担となっている。字種が多ければかりでなく、一般の人にははっきり何かということもあまりよくわからない。このため、和文タイプライタの普及も遅れている。送りがなるに関するルールも不明確さまりなく、しかも、ときどき修正されたりしている。これなどは、日本語の表記法において、表音体系であるかな文字と表意体系の漢字とが混然と結びついていることから生じているものである。しかも、分から書きをしないことが多いからですますますとまどう。表音体系といっても、カタカナ、ひらがな、のほかにNHKなどというようにPRの外来語には英字もつかう。このような複雑な表記法をもつ言語は、他の文明語には類がない。これは日本語情報処理システムの技術開発にとって大きな負担である。
2.6 適用分野

一般企業における日本語情報処理（漢字を含む文書作成）の適用分野として、人事・販売・購入・経理といった特殊な適用分野や、特許情報・文献情報・役員会資料・庁舎提出用資料など限定された範囲の情報処理が適用される。しかしながらこうした一部の分野には、適用出来るが、それ以外の分野や業務には適用できないという性質のものではなく、コンピュータによる情報処理の出力結果として構成されるものや、ローマ字やかな文字に比べて漢字の方がより効果的であるというものを取り扱う分野や業務に対してはすべて適用することが出来る。しかしここでは、いろいろな観点から見て利用価値が大きいと想像される分野の例をいくつか取りあげることにした。

日本語情報処理システムの適用分野は、
① 文献検索のように大量のデータを迅速に読んだり、照合したり、検索する必要のあるもの。
② 顧客サービスの向上につながるもの。
③ 氏名、住所など固有名詞の多いもの。
④ 読み誤りをなくし、正確性を重視するもの。
⑤ 同一データ（製品名・工程名・規格など）を繰り返し使用するもの。
⑥ ローマ字やかな文字ではまずらわしいもの。
⑦ 現在EDP化されているもので、出力レポートを漢字になおすために手書きによる再加工を行なっているもの。
⑧ 漢字でなければならないため人手によって処理されているもの。

など、その処理の内容が省力化、正確性、迅速性、サービス向上につながる業務について適用できる。

しかしこの反面、日本語情報処理システムの導入は、
① 漢字を処理するためのハードウェアの導入
② 日本語情報処理システムの開発
２．３．１ 日本語文章の情報検索

今日のように企業を取り巻く環境がつぎつぎと変化したり、技術革新の進歩がはげしい時代に、企業が外部から取り残されることなく、環境の変化に順応し、時代とともに歩んでいくためには、企業をとりまく内外の情報をいかに手にコントロールしていくかが重要な問題である。この企業環境の激しい変化に対応するため企業活動としての情報管理をどうするかという事が重要視されている。

今日の情報の増加は、爆発的な増え方をしており、集中的な情報管理を行うなければ、相当大いにただな労力や時間や費用を費すことになるため、どうしても E D P 化したり、金社的な立場だった情報管理システムを作成する必要性が出てくる。

このためつぎに示すようないろいろの情報源から必要な情報を見つけ出し、加工・分析したうえで、整理・記録しておき、利用者から要求のあったとき、必要な情報を提供したり、あるいは変化する情報をつねに新しいものに更新したり、不要になった情報を廃棄するための情報管理したり、より広い活用をはかるための PR を行なったりするための情報管理システムが要求される。

１．各種論文や雑誌などの文献類
② 官公庁刊行物
③ 業界や学会の研究報告、統計資料、調査資料類
④ 社内の研究報告、統計資料、調査報告書類
⑤ 内外の特許情報
⑥ アンケート結果
⑦ その他新聞の切抜資料や各種印刷物
⑧ 講習会資料

また、このような情報管理システムのうちでも、その大部分を占める文章に関する情報検索システムの開発にあたっては、非常に多い情報源のなかから
1）必要な情報を見つけ出すむずかしさ
2）分析・加工するむずかしさ
3）分類・整理するむずかしさ
などの困難な問題があるため、図書館学の技術や特殊なドキュメンテーションの技術を必要とする。そのため、多くの企業においては、専門のセクションを設け、専門の担当者のもとで、ＥＤＰセクションの協力を得ることによって開発を行なったり、システムの運営を行なうようにしている。

またこのほか、社内一般のユーザや専門分野の専門に対するサービスとして、文献のコンテンツ・サービスや抄録サービスを行なっている。

日本語文章の情報検索のなかで主要な分野を占める文献検索システムや特許情報管理システムにおいては、
① 収集した情報を取捨選択したうえ、固有の分類によって配列した収集資料目録の作成
② 情報のアBSTラクトを示す抄録の作成
③ 新しく収集した情報の内容を簡単にまとめた解説案内の作成などの作業が発生する。

－33－
そのため、これらの作業に日本語情報処理を適用することは、従来のローマ字やかな文字に比べて、
1）利用者が読みやすい
2）文脈がはっきりする
3）学術名や技術用語など、まぎらわしいものがはっきりする
4）読み誤りをなくし、正確度を向上できる
5）国内・海外の情報の区分がはっきりする
などのメリットが得られる。

いっぽう、ローマ字やかな文字に比べて、文字の絶対数が多いため、決定的な分類、配列ができないので、漢字けん盤がない文字や漢字出力装置のない文字が発生した場合の処理についても考慮しておく必要がある。

2.3.2 人事管理（人名簿の作成）

企業の規模や業種、あるいは企業の方針や性格によって企業組織や構成は異なるが、一般的な人事の業務内容と範囲は、下記に示すようなものがあり、それらの業務を通じて多くの統計資料や従業員1人1人に関する資料が作成されている。

① 人事政策に関するもの
② 管理組織に関するもの
③ 人事管理のための各種規定に関するもの
④ 給与制度に関するもの
⑤ 採用・異動・定員・配置など人事計画に関するもの
⑥ 福利厚生に関するもの
⑦ 安全衛生に関するもの
⑧ 教育訓練に関するもの
⑨ 労働組合に関するもの
⑩ その他、人間関係に関するもの
これらの業務のうち給与計算や各種の人事統計作成作業は、すでにＥＤＰ化されている企業が多いし、個人の住所・学歴・職歴・所属・家族構成や授業などで従業員各人のデータも人事検索システムを通じてEDP化されている。

しかしこれら個人個人のデータのうち姓名・所属・住所・学歴・職歴など本来漢字で表現したい項目であってもコンピュータの入出力装置や端末装置に英字、数字、カタカナ、一部の特殊記号しか取り扱えないというハードウェアの制約から限定されているため、あらかじめＥＤＰシステムのなかからそれらの項目を除いておいて、あとでそれらの項目を手によせておきなったり、あるいはコード化をしたり、カタカナを使用したりしている現状である。このため出力結果を見る場合、その表示方法に慣れていないと

・理解しにくい
・コード変換が要求される
・同音異義語の取扱いが不明
・分かち書き（文・節・区・語の区切り）の問題
・読めにくい
・誤りやすい
などの不便さが生じる。

このためコンピュータの出力情報を手によって人事資料の作成作業を行なったり、漢字項目を加えるための再加工をしたりしなければならない問題などが生じる。

これらローマ字やカタカナによる処理や、手による処理から生じる不便さを解消するためにも、あるいはまた事務能率を向上させるためにも漢字をＥＤＰシステムに取り入れることは、大きなメリットである。

このほか日本語情報処理の可能なＥＤＰシステム・ファイルとして、姓名、住所、学歴、職歴、所属、特技、専門技術、家族構成、家庭状況、
考課など個人別名簿を作成しておけばよい。そうすることによって従来
人手で処理していた。

・人事統計や人事管理資料における漢字処理
・従業員の住所録や社内電話番号簿の作成
・辞令や身分証明書などの作成
・国税や地方税の納税者台帳など各種台帳の作成
・健康診断票や労災資料作成
などの名簿・台帳・帳票類の作成や幅広いコンピュータの活用、および
ＥＤＰシステムと台帳システムとの断絶や処理の重複がなくなるため、
コンピュータを利用した管理システムの一元化ができる。そのため
① 省力化
② 正確化
③ 迅速化
④ 円滑化
がはかれた人事管理効率をあげるとともに企業全体の生産性の向上に寄与
することができるなどのメリットが得られる。
また人事管理の問題と同じようにして
・販売・購買などに利用する顧客や下請業者の業者テーブルの作成
・購買・在庫・会計で利用する単価簿の作成
・販売・営業で利用するディレクトメールの宛名
・見積書や注文書の発行
・生産管理で利用する標準作業工程書や標準作業時間テーブルの作成
・機械台帳や固定資産台帳の作成
・商品カタログの作成
など各種台帳類や伝票類の作成・発行が考えられる。
このようなコンピュータによる情報処理は、ＥＤＰシステムの
今後の新しい適用分野の開発や活用に大きく貢献が期待される。

- 36 -
2.3.3 伝票の発行

現在、企業においては、販売・購入・生産・会計など企業活動のあらゆる分野でいろいろな種類の伝票がそれぞれの目的で使用されており、これらの伝票は情報伝達の手段として重要な役割をはたしている。

そのためこれらの各種の伝票は、情報処理を含む事務管理制度上から要求される機能に、もっともよく適合するように設計され、処理されて、運用されるようなものでなければならない。このため、企業活動の目的を達成するうえで、伝票本来の目的に適合するような固有機能を備えていこと、記入や取り扱いが誤りなく行なわれるように設計されていることなどが必要である。このほかに伝票は、処理を容易にするために、

・書きやすいこと
・読みやすいこと
・取り扱いやすいこと
・伝票処理を含むシステム全体のコストが安くてできること
などが要求されている。

そのため、伝票処理を E D P 化する場合、それらの処理を含む事務管理システムの目的・機能・運用を十分検討したうえで、それにもっとも適した形に伝票処理システムを設計するようにしている。

しかし現在のコンピュータで取り扱える文字の制約から英字や数字、およびカタカナと一部の特殊記号しか使用できないので、表示面を除いた処理面など、システムの他の面の問題についてはいろいろと配慮がなされていても文字表示の問題については、人間の側にその課題を要求している面が多い。

そのため書きやすさ、読みやすさ、取り扱いやすさといった、本来伝票に課せられるべき要求が人間の側に課せられる結果となっている。

現在開発されている多くの伝票発行を含む E D P システムにおいては、これら漢字を扱う処理が出来ないため、
① コード化したり、カタカナあるいはローマ字などを使用して処理を行なっている。

② 伝票の目的や用途に応じて、最初からタイトルや項目的説明など日本語で表示すべき性質のものを事前に印刷したホーム・シートを作成しておく、コンピュータの出力には、なるべく漢字を含まないもので
すむようにし、どうしても漢字を必要とする部分だけ、不便でも英字
やカタカナを使用する。

③ コンピュータの出力に漢字が使用できないため、全体のシステムの中から伝票発行の部分だけ取り出し、ＥＤＰ以外の方法で処理する。
などの方法がとられている。

このように売上げ伝票・注文伝票・作業指示伝票などいろいろな伝票
がコンピュータによって、われわれが日常使用している日本語（漢字か
を混在文）として表現できるならば、
・コード化によるシステムの複雑
・英語や日本語のローマ字表現、カタカナ表現による読みにくさ、理
解のしにくさ
・ホーム・シートによるオペレーションのわずらしさや処理効率の
低下
・伝票発行を他の手段に頼ることで生じる他の処理との断絶
などの問題が解消される結果となり、この意味からも伝票発行の分野
におけるコンピュータによる日本語情報処理のメリットは大きいし、さら
に現在ＥＤＰ化されていない分野、いわゆる人手を介して処理されて
いる日本語文を主体とした伝票発行の分野にも適用する事が可能となる
ため、そのメリットはさらに大きなものになると思われる。

2．3．4 その他の適用分野

このほか、業務として日本語文を取り扱っている分野については、す
べて適用できる可能性をもっているが、コンピュータによる日本語情報
処理は、ハードウェアおよびソフトウェアのコストと、その情報の価値やタイミングなどのバランスを十分検討しなければならない問題がある。変更処理のないただ1回だけの情報処理の問題までEDP化することはかえってデメリットの大きなものとなる。

そのためどんな情報をコンピュータで日本語情報処理の問題として取り上げるかということは、企業の性質や情報の性質によって異なるので、その情報を扱う人達で十分検討を行なったうえで決定する必要がある。他社で文献検索や人事管理に漢字のEDP処理を行なって大きなメリットがあったからといって、自分の会社でもうまくいくとは限らない。

しかしコンピュータの適用分野の領域に日本語情報としての文章情報処理の分野が開拓されてきたことは、注目すべきことであり、非常に大きな意義をもっている。

そこでこれまで述べてきた業務以外にどんな適用分野が考えられるか、その代表的な適用項目をあげておく。

(1) 見積書の作成
(2) ダイレクト・メールや自動振込通知などの業務
(3) 株式などの通知、案内書の作成
(4) 各種マニュアル作成業務
(5) 全社的なニュースや移動通知速報の作成

2.7 技術動向

繰り返し述べるようにコンピュータで日本語情報を取り扱う際にそれに付随して、いろいろな問題が生じている。まず、漢字の数が欧米語にくらべてはるかに多いということがあげられる。実際にコンピュータの必要メモリ（文字のビット構成）はどれくらいか計算してみると、表2-3のようになる。
表2-3

<table>
<thead>
<tr>
<th>1ビット</th>
<th>2種</th>
<th>7ビット</th>
<th>128種</th>
</tr>
</thead>
<tbody>
<tr>
<td>2ビット</td>
<td>4種</td>
<td>8ビット</td>
<td>256種</td>
</tr>
<tr>
<td>3ビット</td>
<td>8種</td>
<td>9ビット</td>
<td>512種</td>
</tr>
<tr>
<td>4ビット</td>
<td>16種</td>
<td>10ビット</td>
<td>1,024種</td>
</tr>
<tr>
<td>5ビット</td>
<td>32種</td>
<td>11ビット</td>
<td>2,048種</td>
</tr>
<tr>
<td>6ビット</td>
<td>64種</td>
<td>12ビット</td>
<td>4,096種</td>
</tr>
</tbody>
</table>

このように日常よく使われている2,500種の漢字を表現するには11ビットから12ビットであれば十分である。一般にコンピュータ処理では6ビットまたは8ビットで1文字を表現しているので、従来の2字で漢字1字を構成すればよい。単にビット構成の可能性からいえば、漢字の数がじょうずに多いといっても、実際にはそれほどめんどような問題ではない。また、漢字の名前有性は、かな文字やローマ字にはなく、日本人にとって情報量がもっとも多い文字であり、漢字1文字に16ビットを使用しても、漢字かな混り文出力の理解度からすれば、多大な効果が得られる。

出力に比べて問題となるのは入力のスピードである。現在の漢字テレタイプライタは、1分間に70～90字程度可能であるが、カナ・タイプライタに換算した場合140～180字/分に対応する（日本語表記の漢字：カナはおよそ1:2）。ところが、通常のカナ・タイプライタでは、1分間に200字くらいであるから、現状では、かな文字入力の操作のほうが速い。この比較は専門のバンチャやオペレータがバンチした場合のことで、だれもが容易に入力操作ができるというわけではない。欧米のようにタイプライタで日常文書を打っているところでは、現在の英字主体のコンピュータを使いこなすことに対して、あまり抵抗感はないと思われるが、日本のように入力やタイプ（和文タイプライタなどで）は専門家にまかせる習慣
にあっては大きな問題である。

素人でも取扱いが簡単な漢字タイプライタもいくつか開発されているが、操作速度に難点があるため実用的には問題がある。

以上のように、漢字かつ出力力については、1分間に7,000万文字前後のスピードをもつプリンタも開発されているといわれており、見通しはあるが。しかししながら、コンピュータの利用も、これまでのような英数字を中心にしてしたものから、文章（日本語）を扱おうとするときに漢字の処理は欠くことができないものであり、身近に近まっている問題である。

2.7.1 入力装置

日本語の文章で一般的に使われる漢字の数は、およそ2,500字といわれるが、アプリケーションにによっては、5,000～10,000字も必要とする場合がある。この多くの字種を扱う入力をいかに手がわよくできるかが漢字処理のポイントとなっている。現在ももっとも多く使用されているのが、漢字けん盤式のさん孔機である。この形式の入力装置は、漢字の字種に比例してキーが多くなり入力操作の速度は遅く、1分間に60字～80字といわれている。装置自体の規模もかなり大形なものである。代表的なものとしては、①新興型けん盤さん孔機、②O K I型けん盤さん孔機がある。

① けん盤キー…………192個
ソフトキー…………12個
＜収容文字数＞ 2,346字
※右手でけん盤キーを、左手でソフトキーを操作する。

② けん盤キー…………650個
ソフトキー…………4段ソフト足踏ペダル方式
＜収容文字数＞ 2,600字
※両手でけん盤キーを、両足でソフトキーを操作する。

最近では、漢字ディスプレイ装置およびカセット・テープを備えた高
<table>
<thead>
<tr>
<th>メーカ</th>
<th>機 種</th>
<th>文字種</th>
<th>コード</th>
<th>出 力</th>
<th>概 要</th>
</tr>
</thead>
<tbody>
<tr>
<td>高千穂</td>
<td>T4112</td>
<td>漢字用盤</td>
<td>3,072</td>
<td>磁気テープ</td>
<td>入力文字をディスプレイ64文字ブロックでテープに書き込む</td>
</tr>
<tr>
<td></td>
<td>T4012</td>
<td>漢字用盤</td>
<td>3,072</td>
<td>カセット・テープ</td>
<td>入力文字をカセット・テープに書込み</td>
</tr>
<tr>
<td></td>
<td>T4011</td>
<td>漢字用盤</td>
<td>2,496</td>
<td>8単位2列</td>
<td>紙テープに漢字コードをさん孔する</td>
</tr>
<tr>
<td>沖電気</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2,600</td>
<td>6単位2列</td>
<td>紙テープ</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8単位2列</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I B M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>漢字用扇形</td>
<td>3,600</td>
<td>2行カードコード</td>
<td>カード</td>
<td>通常のカタカナ付カードさん孔機としても利用できる</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>富士通</td>
<td>FACOM 600A</td>
<td>漢字入力キーボード</td>
<td>3,840</td>
<td>8単位2列</td>
<td>紙テープ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>谷村新興</td>
<td>SCK-200-D</td>
<td>漢字用扇形</td>
<td>2,688</td>
<td>6単位2列</td>
<td>紙テープ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8単位2列</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>日電</td>
<td>C-5110</td>
<td></td>
<td>2,145</td>
<td>8単位2列</td>
<td>紙テープ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>東レ</td>
<td>TORAY-8510</td>
<td></td>
<td>2,240</td>
<td>8単位2列</td>
<td>紙テープ</td>
</tr>
</tbody>
</table>

- 42 -
千種交易KKのT4型漢字入力装置、通常のカタカナ・キーパンチと併用できるIBM29型の入力装置、また従来のカナ・タイプライタを利用する谷村新興製作所のSC-400型、和文タイプライタを利用する日本電気のC-5110型など新しいタイプの入力装置が開発されている。

その他、日立製作所が開発した手書き文字認識装置は、楷書体の教育漢字881字とカナ文字を含んだ合計952字をライトペンを使用して、手書きで入力ができる。また、三菱信託銀行がアメリカのCS社より導入した漢字読取り用のOCRなど、最近では、いろいろな方式の入力装置が開発されてきている。

日本語文章の入力処理は、大きくわけて、
① 原データの作成
② 訂正（校正）データの作成
の2つの目的がある。

原データの作成については、原稿をみて漢字入力装置から入力するだけであるから問題はないとしながら、問題は訂正処理を行なうための訂正データの作成である。英数字やカタカナ入力装置の場合はペリィ（検孔）という方法があるが、漢字入力装置の場合は、能力的、コスト的にみてなかなかむずかしい。そのための1つの方法としては、モニタ用紙に出力して印刷物校正の方法と同様に訂正することができる。ただしその場合には、部分訂正をするか、全データの打直しをするか、あるいは行送りはどうするかなどソフトウェアの処理が複雑になり、ただ単に追加、調正、削除だけにとどまらない。いわゆるいくつかの編集業務を必要とする訂正処理の解決策として、入力装置に付随した漢字ディスプレイ装置が開発されている。しかしながら、漢字ディスプレイ装置は高価なうえに漢字処理機構との接続が複雑であり、コンピュータとオンラインでしか処理できないなどの欠点があるといわれている。

以上各種の入力方法について述べたが、一般に扱う約2500字に対
<table>
<thead>
<tr>
<th>文字表示形式</th>
<th>沖電気</th>
<th>高千穂</th>
<th>日本電子産業</th>
<th>電総研</th>
<th>日電</th>
<th>東レ</th>
</tr>
</thead>
<tbody>
<tr>
<td>ドット</td>
<td>ドット</td>
<td>フライング・スポット</td>
<td>ドット</td>
<td>高解像度ブラウン管</td>
<td>ドット</td>
<td></td>
</tr>
<tr>
<td>最大表示字数</td>
<td>32字×16行（横）</td>
<td>64字×2行（横）</td>
<td>32字×8行（横）</td>
<td>16字×32行（縦）</td>
<td>40字×10行</td>
<td>88字/行×66行</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20字/行×15行</td>
</tr>
<tr>
<td>文字種</td>
<td>2,578</td>
<td>3,072</td>
<td>2,496</td>
<td>1,143</td>
<td>2,500</td>
<td></td>
</tr>
<tr>
<td>バッファ・メモリ</td>
<td>512字</td>
<td>128字</td>
<td>960字</td>
<td>800字</td>
<td></td>
<td></td>
</tr>
<tr>
<td>プリンタ</td>
<td>接続</td>
<td>接続</td>
<td>接続</td>
<td>接続</td>
<td></td>
<td></td>
</tr>
<tr>
<td>その他</td>
<td>ライトペン使用（位置指定）</td>
<td>最大32台まで</td>
<td>種々のシステムとのインタフェイス</td>
<td>大型プロジェクト開発の一環</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>接続</td>
<td>使用</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
・漢字の基本構成パターンと素子の1例

構成パターン
・ナンパ

素子ナンパ

・漢字のパターン入力方式の1例「湘」

入力順序

入力法1

1
2
3
4

（構成パターン）

入力法2

1
2
3

（構成パターン）

図2-3 漢字入力方法の例
して、出版などの分野ではさらに多くの文字を扱わなければならず、漢字けん盤上に配列できない文字の入力方法を考える必要がある。

その解決策として考えられることは、漢字けん盤を使わずに、日本語文の読み方（発音）どおりにカタカナで入力し、コンピュータでカタカナから漢字へのコード変換処理を行う方法がある。また、従来の漢字けん盤のキーを増したり、あるいは、図2-3に示すような文字構成のパターンを構成エレメントによって組み合せる方法もある。

いずれにしても各メーカーは漢字入力に関して苦心しており、まだ最適な入力装置は開発されていない。

2.7.2 出力装置

機械式の漢字出力装置は、これまでにも各種開発されてきたが、大きく分けて
① ラインプリンタのような機能をもったもので主としてモニタ出力として使用するもの。
② 印刷の版下作成（版組みの合理化）のための電算植字として利用するもの
③ ①、②の両用のもの

などであろう。モニタ出力というのは、文字品質や文字の大小などにはあまり効果を考えずに単なる校正用のグラ出力のようなものである。これに対して、②の場合は、出力がそのまま印刷物の原版となるため、文字品質などによって良質の出力が要求される。以下にそれぞれの目的・用途によって、どのような出力装置が開発されているか概説することにする。

(1) モニタ用の出力装置

モニタ用には、つきのような種類がある。

① 機械方式
② 光学方式
(3) ドット方式
(4) ワイヤ・プリント方式
(5) 全電子方式
漢字プリンタの場合には、プリンターを何種類か入力装置の2,500種以上必要であり、しかも固有名詞など特殊な文字が、いつどんな文字が必要になるか予想することができない。そのため、漢字的な混り
支が出力には欧文のように英語なら26文字ロシア語なら32文字と上限がきりりしている言語をプリントする場合と比較にならないほど
の難しさがあるといえる。以下に各方式の出力装置を紹介する。点の
集合で表示するドット方式の代表的なものはFACOM6501である。
その基本的な機構は、それぞれの文字が縦15、横18のドット
・マトリックスに分解され、その文字パターンが、記憶装置に格納さ
れている。いっぽう、プリンター機構は、やはり縦15、横18のビン
・マトリックスの電極で構成されている。このドット方式は1つの文
字信号が送られてくると、記憶装置がそれに対応する文字パターン
を探し出し、それをビン・マトリックスの電極に伝えて、感光紙に感
光させる方法である。
IBMのワイヤ・プリント方式では、文字を縦18、横22のドッ
ト・マトリックスに分解して、オンラインのコンピュータによってデ
ィスクに記憶される。出力形態はFACOM6501が感光紙に感光
させるのに対して、これはビンについているインクによって通常のコ
ンピュータで使用している用紙に出力される。
高千穂交易KKのＴ４１００漢字情報処理システムの場合は、フォン
ト・メモリとCRTを応用したもので、文字を縦32×横32のドッ
ト・マトリックスに分解して、フォントメモリ上に記憶する。出力は
ファイバ・チューブに映像されている文字を電子記録紙に印字する方
式である。その他にも、各メーカーでこの方式は多く採用されている。
### 表2-6 漢字プリンタ性能表

<table>
<thead>
<tr>
<th>メーカ</th>
<th>高田電子産業</th>
<th>高田電子産業</th>
<th>高田電子産業</th>
<th>高田電子産業</th>
<th>高田電子産業</th>
<th>高田電子産業</th>
<th>高田電子産業</th>
<th>高田電子産業</th>
<th>高田電子産業</th>
<th>高田電子産業</th>
<th>高田電子産業</th>
<th>高田電子産業</th>
<th>高田電子産業</th>
<th>高田電子産業</th>
</tr>
</thead>
<tbody>
<tr>
<td>印字方式</td>
<td>静電レコード方式</td>
<td>静電レコード方式</td>
<td>静電レコード方式</td>
<td>静電レコード方式</td>
<td>静電レコード方式</td>
<td>静電レコード方式</td>
<td>静電レコード方式</td>
<td>静電レコード方式</td>
<td>静電レコード方式</td>
<td>静電レコード方式</td>
<td>静電レコード方式</td>
<td>静電レコード方式</td>
<td>静電レコード方式</td>
<td>静電レコード方式</td>
</tr>
<tr>
<td>用紙</td>
<td>静電記録紙</td>
<td>静電記録紙</td>
<td>静電記録紙</td>
<td>静電記録紙</td>
<td>静電記録紙</td>
<td>静電記録紙</td>
<td>静電記録紙</td>
<td>静電記録紙</td>
<td>静電記録紙</td>
<td>静電記録紙</td>
<td>静電記録紙</td>
<td>静電記録紙</td>
<td>静電記録紙</td>
<td>静電記録紙</td>
</tr>
<tr>
<td>表面速度</td>
<td>12,000〜24,000 (字/分)</td>
<td>72,115(字/分)</td>
<td>600行/分</td>
<td>1,240行/分</td>
<td>通常の連続用紙</td>
<td>100行/分</td>
<td>180行/分</td>
<td>107/600</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>文字パターン</td>
<td>フライングスポット</td>
<td>32×32ドット/分</td>
<td>15×18ドット/分</td>
<td>18×18ドット/分</td>
<td>18×22ドット/分</td>
<td>18×18ドット</td>
<td>活字 ドラム</td>
<td>24×22ドット</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>文字種類</td>
<td>1,000</td>
<td>2,048〜8,192</td>
<td>2,400</td>
<td>4,800</td>
<td>7,200</td>
<td>〜2,600</td>
<td>〜1,000</td>
<td>3,960</td>
<td>4,996</td>
<td>4,996</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>文字サイズ</td>
<td>2.2×2.2mm〜4.5×4.5mmの8種</td>
<td>2.5×2.5mm</td>
<td>3.8×4.0mm</td>
<td>6×6mm</td>
<td>最大</td>
<td>4.6×5.8mm</td>
<td>4.23×4.23mm</td>
<td>12P</td>
<td>4.2×3.84mm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>コンピュータとのインタフェイス</td>
<td>オンライン可能</td>
<td>オンライン</td>
<td>FACOM230〜25/35</td>
<td>IBM360/30</td>
<td>IBM360/30</td>
<td>IBM360/30</td>
<td>IBM360/30</td>
<td>IBM360/30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>特徴</td>
<td>＊インプットに磁気テープ、紙テープのほかにコンピュータから直接入力可能</td>
<td>＊どちらの文書を印刷することができるため、オフィス、オフィスの原稿として使用できる</td>
<td>＊OCR文字をDBMにより表示してからプリンタに送るため、単にDBMで印字する必要がある</td>
<td>＊CRTディスプレイ上に表示してからプリンタに送るため、単にCRTで印字する必要がある</td>
<td>＊オフセット印刷の原稿として使用できる</td>
<td>＊オフセット印刷の原稿として使用できる</td>
<td>＊オフセット印刷の原稿として使用できる</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>基本構成</td>
<td>＊プリンタ制御装置</td>
<td>＊文字発生装置</td>
<td>＊プリンタ</td>
<td>＊文字発生装置</td>
<td>＊プリンタ</td>
<td>＊文字発生装置</td>
<td>＊プリンタ</td>
<td>＊文字発生装置</td>
<td>＊プリンタ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>製品名</td>
<td>JEM3100</td>
<td>JEM3100</td>
<td>TYP3100</td>
<td>TYP3100</td>
<td>TYP3100</td>
<td>TYP3100</td>
<td>TYP3100</td>
<td>TYP3100</td>
<td>TYP3100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 48

- 48
表 2-7 電算機字システム性能表

<table>
<thead>
<tr>
<th>メーカ</th>
<th>写研</th>
<th>富士通</th>
<th>日本電子産業</th>
<th>R</th>
<th>C</th>
<th>A</th>
<th>I</th>
<th>B</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>印字方式</td>
<td>フラッシュランプ発光</td>
<td>フラッシュランプ発光</td>
<td>CRT</td>
<td>CRT</td>
<td>CRT</td>
<td>CRT</td>
<td>CRT</td>
<td>CRT</td>
<td></td>
</tr>
<tr>
<td>印字速度</td>
<td>900/分</td>
<td>1200/分</td>
<td>10000/分</td>
<td>4200/分</td>
<td>13500 (英字8ポ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>文字発生方式</td>
<td>字母光学式</td>
<td>字母光学式</td>
<td>フライングスポット</td>
<td>フライングスポット</td>
<td>フライングスポット</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>文字種</td>
<td>6400字</td>
<td>12762</td>
<td>20000</td>
<td>40000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>文字サイズ</td>
<td>9〜18級</td>
<td>6〜14P</td>
<td>4〜24P</td>
<td>4〜96P</td>
<td>4〜18P</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>特徴</td>
<td>文字変形させてよこ可</td>
<td>＂</td>
<td>＂</td>
<td>＂</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>貼下フィルム</td>
<td>貼下フィルム</td>
<td>モニタプリン可</td>
<td>モニタプリン可</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>出力</td>
<td>10インチ×100フィート</td>
<td>8インチ×100フィート</td>
<td>250mm巾</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>コンピュータ</td>
<td>HITAC-10</td>
<td>FACOM 270-20</td>
<td>HITAC-10</td>
<td>HITAC-8400</td>
<td>IBM360-40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>入力装置</td>
<td>紙テープ</td>
<td>紙テープ</td>
<td>紙テープ</td>
<td>紙テープ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>出力装置</td>
<td>紙テープ</td>
<td>磁気テープ</td>
<td>空気テープ</td>
<td>磁気テープ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(2) 電算機用の出力装置

電算機字というのは、漢字プリンタで出力したものをそのまま出版物の版下として使用する印字方式をいう。前述した文字用のプリンタは、文字の収容数やスピードおよび出力用紙についての制限はあったが、電算機字の場合は、出力がそのまま印刷物の版下となることから、つきのような要求・制限事項がある。

① 文字の大小、縦組・横組の自由

文字の大きさが自由に変えられること（2〜20ミリくらい必要）。文字の並びが縦にも横にも変えられること。

② 文字の種類

文字の書体が明朝体だけでなく多くの種類が用意されていること。
※日本語→明朝体、ゴシック体、教科書体など
※英語→ロマン体、ゴシック体、イタリック体、袋文字など

③ 文字の品質

何万、何十万もの印刷物の原版となるため、文字の品質は非常に高いものが要求される。通常のラインプリンタのように字がおどっていったり、ドット方式では、点が目に入るもののは不適当である。そのため、電算機字用としてはドット方式よりも内部に文字パターンをもっているもののほうがよく、その方式としては、

A 光電式→電子式
B 全電子方式

がある。

しかし電算機字は、雑誌、新聞など1回印刷して役目を终えるようなものは非効率であり、資料性のあるもの、部分修正を加えて何度も印字が必要になるものに向いている。たとえば、同じ内容のものを体裁を変えて出したいとき、いわゆる同一の小説をポケット版、文庫本、単行本などで出版するようなときに効果的である。

— 50 —
2.7.3 ソフトウェア

一般に、ソフトウェアは、システム・ソフトウェアとアプリケーション・ソフトウェアに分類できる。日本語情報を処理する場合、システム・ソフトウェアとしては、つぎのような処理があげられる。

(1) 入力
○ ファンクションコード処理
入力したデータのファンクションにしたがって文字の大小、書体の選択、組方向、形状などを処理する。
○ スペース処理
バーレンとか句読点などの前後に体裁をととのえるための適当なスペースをつける処理。
○ キャラクタ・ピッチ処理
漢字や数字、特殊文字に対してピッチ付けをし、文字のバランスをとる。
○ ページ割付処理
1段の字数、行数などを考慮して、1ページあたりの割付をする。
○ ノンプル柱処理
割付が決まった段階でページ、柱をつける。

(2) 棟 張 調 整
○ カスタマイゼーション処理
行の処理であり、一定の間隔内に文字をそろえるためのスペースを送りこんだり、取り除いたり、平体をかけたりする処理。
○ 構足、分離、禁止処理
行末、行頭に、カッコ類、感嘆符、疑問符、半音などきた場合の特別な処理。
○ ルビ処理
漢字の横あるいは上にふりがなをつける。この場合に、均等配分、
中央揃え、などが可能である。

○ハイフォネーション

英字の単語が2行以上にわたる場合、区切りにハイフォンを挿入する。

○その他の

合成、数式などの処理を行う。

(3) 図版処理

図版、写真、見出しなどの版下を入れる場所にスペースをとっておく処理。

以上のような基本的な文字、文書処理を基本として各種のアプリケーション
ジョン・システムが生じてくる。アプリケーション・システムは、ユーザ
によって各々開発されるもので、
① 入出力のハード機能
② ユーザ独自のアプリケーション
が考慮されなければならない。しかし、いずれにしても、上記のシステ
ム・ソフトウェアの全部あるいはその一部の機能を使用しなければ、
日本語情報の処理はできない。
現在、開発されているおもなアプリケーション・ソフトウェアとしては、
○科学技術文献速報編集システム
○百科辞典など書籍編集システム
○特許情報検索システム
○図書館における図書文献検索編集システム
○新聞自動編集システム
○その他の
などがある。
3. 日本語情報処理のためのカタカナ入力方式
3. 日本語情報処理のためのカタカナ入力方式

3.1 カタカナ入力漢字かな混り文出力システムの必要性

現代のように企業環境の変化が激しく、それにともなうおよびただしい情報量の中から必要な情報を適格に効率よく選択することは、コンピュータを中心とした情報処理システムをぬきにしては論ぜられないようになってきている。このため多くの企業においてそれぞれの企業活動の分野でEDPシステム化が行なわれており、それにともなってハードウェアやソフトウェア技術の発展が著しく、その適用分野の拡大がみざましいものとなった。

このようなEDPシステムの適用分野の拡大が、日本語文章を扱う情報処理システムの開発に対する強い要望となっていると考えられる。日常使用している日本語文のほとんどは漢字かな混り文により表現されている事からしても漢字かな混り文を中心とした日本語情報処理の問題が、日本人にとっていかに重要であるかということが理解されると思う。また日本語情報処理に関する問題は、わが国特有の問題であり、他のアプリケーション・プログラムや情報処理システムの問題とは異なって欧米などでは発生しない問題である。しかし日本語情報処理システムを開発することは、英語を国語としない他の国々に対して、コンピュータによる国語の処理に対する1つの処理方式を提示することになり、この意味からは重要な問題であるといえる。

今日、日本語の辞書に載っている漢字は5万字以上といわれ、そのうち新聞や雑誌など通常使われている漢字は、約3000字といわれている。このように日本語の問題は、文字の種類の多いこと、文字構成の複雑なことなどが原因となってコンピュータによる日本語情報処理の問題を困難なものにしていったが、EDPシステムの発展は、漢字の処理を可能にすると
ともに日本語による文章情報処理の効率化がなされるようになってきた。
しかし漢字処理は、
• 文字の字数が非常に多い
• 字形が複雑である
• 一定のルールで分類したり、配列したりすることができない
• 文字の読み方が一定していない
• 字形が時代によって変化する
などの問題を解決しなければならない必要があり、これらはいずれも情報処理における入出力の問題として集約することができる。
このため現在開発され、発表されているほとんどの日本語情報処理システムは、従来のコンピュータの入出力装置とは別に漢字を扱える専用の入出力装置をもち、それによって処理がなされている。そのため従来のハードウェア・コストに加えて漢字処理のための入出力装置が必要になり、それだけハードウェア・コストがプラスされることになる。
しかしそれでも現在使用されている漢字をすべて取り扱うことは、
① システムの複雑化
② 効率の低下
③ 高価なシステム
などの理由により不合理である。
本来、漢字を中心とした日本語情報処理の問題は、コンピュータの出力情報と人間とのインタフェースの改善にあり、人間の側にたった情報伝達の効率化や情報サービスの向上にあるといえる。このためコンピュータによる日本語情報処理の恩恵をもっとも受けるのは、人間の側にあり、コンピュータの側にとっては、あまりありがたい問題といえない。
しかし人間側の情報処理の効率向上という問題は、情報の伝達や意思決定を行なうのが人間である以上、人間系を含めた情報処理システムにおける効率向上やシステム活用につながる改善など、システム全体の効率化に
は大きな役割をはたされなければならない。

そのため日本語情報処理のためのシステム効率の向上と漢字入出力装置
導入によるコスト・アップとのバランスを考えなければならないが、この
場合ハードウェアのコストだけでなく、取扱い上の問題、データの
修正の問題、システムの互換性の問題などハードウェアの導入にともなう
いろいろな問題を考慮する必要がある。

現在、発表されている漢字入力装置は、タイプライタのように漢字キー
ボードを用いる方法のものが多く使用されている。この場合、コンピュー
タで処理するためには、従来の入力装置と同様にコンピュータが処理でき
るように文字をコード化しなければならない。しかし漢字には英字やカタ
カナのように一定の配列や順序がないため、漢字入出力装置を製作してい
るメーカーが、それぞれのコード体系を作成し、他社の製品との互換性は考
えられていない。

そのため、漢字入力装置の漢字キーボードの配列は、オペレータの操作
の容易性を考え、
① 文字の使用頻度による配列
② 部首や画数など文字の形にもとづいた配列
③ 音・訓索引（五十音順）や英字のアルファベット順など文字の読み方
にもとづいた配列
などが使用されている。

この他、文字のパターンを分解して、それを組合せて処理したり、漢字入
力用のOMRなどで処理するものもある。

このように漢字のコード化に関しては、いろいろな問題が混在しており、
日本語情報処理をより複雑なものとしている面がある。こうした背景のも
とにこの文書情報処理の研究では“カタカナで入力して漢字で出力する”
方式を採用することにした。この方式によってつきのようなメリットが得
られる。
１ 従来のカナ・タイプライタ（入力装置）があれば、特別な漢字入力装置を必要としない。

２ 漢字のコード化に対する複雑な問題を解消する。（一定のルールで処理できる）

３ オペレータの操作が容易である。

４ 漢字コードのような大きなビット構成をとらないので、処理効率が向上する。

５ 従来のものとシステムの互換性が得られる。

６ データ伝送などの場合も特別な（漢字用の）通信制御装置を必要としない。

７ 漢字出力装置のない場合あるいは漢字出力を必要としない場合にはカタカナで出力できるので、システムの使用範囲が拡張できる。

８ 変換テーブル（カタカナ→漢字）の構成を変えれば英語→漢字などの翻訳作業にも適用できるというソフトウェア・ロジックとしての拡張性を持っている。

９ 変更処理が容易

しかし、その反面、

⑩ 同音異義語の処理

⑪ 変換テーブル作成のむずかしさ

など問題になることもあるが、⑪の同音異義語の処理は人手装置を改良して、変換テーブルの構成を工夫したりすることによって解消されるであろうし、⑪の変換テーブルの作成も専門分野別にするなどして解決されるよう。

このようなカタカナ漢字変換システムを企業における文章情報処理に適用する場合には、たとえばつきのような利用方法がある。

① 個人や特定の顧客、下請業者など、そのデータの件数や名称、住所などがはっきりしており、しかもいろいろな利用のされ方や繰り返し性の

— 58 —
非常に高いものについては、人事ファイルや顧客ファイル、業者ファイルにカタカナ・データを入力しておく、それと同時に同音異義語の判別処理を行なった中間ファイルを作成しておくことによって、コンピュータから出力するたびごとに同音異義語の判別処理をしなくてすむので有効である。

(2) 文献検索や特許情報検索など各種の情報を入力する場合には、同音異義語の判別処理が頻繁に要求される。その場合、これら同音異義語について、変換テーブルを専門分野別（機械・電気・化学……）経営・人事・生産など）・発生頻度別の階層レベル付けなどの工夫を行なっておくことで同音異義語判別処理の減少がはかれる。

この他、日本語情報処理における音・文字の取り扱いや送りが、字体など複雑な問題があるが、これは表音文字と表意文字との対応ずけがないこと、漢字の標準的なコード体系がないこと、などの理由によるものであろう。

また今後ますます漢字を中心とした文章情報処理システムの重要性が強まってくると思われるので、文字の大きさ、ピッチ、書体、行処理などについても自由に簡単に扱うことができて、安価な漢字プリンタ、漢字ＣＯＭ、漢字検字装置の開発が要望される。

3.2 カタカナ入力方式へのアプローチの方法

カタカナ入力漢字かな混り変換の処理は、きわめて困難な問題をかかえているとは前提してきたが、カタカナから漢字への変換を１００パーセント完璧なシステムを開発しようとする試みは、現実的でないと思われる。ただそれでも、変換処理のすべてをコンピュータで片付けてしまうとしないで、人間が簡単に分担できるところは人間が処理した方が効率がよい。人入の段階で、分ち書きして入力するということはそれほどやっかいなことではないし、どの単語を漢字に変換してほしいかを指定することも
決して大きな負担になることではない。それを、分ち書きしないカタカナ・データを入力して、そのなかから、漢字変換すべきことばを見つけることまでコンピュータで処理しようとすれば、それだけでも大変な作業であり、処理時間も長くなってしまう。

漢字変換処理の主要な問題である同音異義語の処理にしても、現在開発が試みられているいくつかのシステムでは、前後の文脈や状況判断から同音語中の特定の漢字単語をえらぶ処理を完成しようとしているようである。しかし、そのような処理技術がいかに高度化されても、やはり限度があるし、情報処理量からいっても現実的でなくなる可能性が高いことを、ここで強く指摘したい。たとえば、「あなた」という語を漢字変換するとき、 「貴方」、「貴女」、「貴男」のどれをえらぶかを処理するとき、まず、その前後の単語からコンピュータが判断するのはかなり難しい。言葉がいていないかどうかなどで、文書の受け手の性別が判定されなければ、正解は困難であり、もしがくめつくっても、「貴方」か「貴男」かをえらぶのはさらに難しいことは言うまでもない。また、すべての同音異義語の判定に文書の受け手の性別などは、判断する必要がなく、なかにはその単語が名詞か動詞を判定すればよいものもあるであろう。同音異義語の判定のために、前後の文脈から状況判断するといっても、その状況判断のポイントが、1つ1つの単語について異なる。

同音異義語の処理は、すべてをコンピュータ処理しようとせず、人間が少し介入するだけで、その負担の軽減はかなり知れないものがある。したがって、たとえば、ある単語を漢字に変換するよう入力指示したとき、もし同音異義語がいくつかあれば、使用頻度の高いものをコンピュータからCRTなどに表示して、それを入力者にみて、適当なものを指定するというようなマン・マシン対話形式のシステムを開発する試みは、当面はきわめて現実的なアプローチであるといえる。

いっぽう、言語処理の負担を少しでも軽減する方向へと、言語の表記法
を整理することの必要性も強調されるべきであろう。たとえば、「兆候」
と「微候」は同音異義語とはいえないので、どちらに統一すべきであろうし、
そのほか、「記念・紀念」、「出会い・出合う」なども、どちらに整理
しても、日本語はいつこうに貧しくなったりはしない。このような例はまだ
まだたくさんある。「文部省刊行物表記の基準」では、「じゅうぶん」
と書くほうがよいが、「十分」と書いてもよいということになっている。
しかし、「日本国憲法」で「充分」とかれていることからすると、法律
文書では「充分」と書くことになっているらしい。また日本新聞協会編の
「新聞用語集」によれば、「十分」と書くことになっている。
また、文化が「進展」するとかくが、貿易がしんでんするときは「伸展」
と書くのが正しい。美術品を「鑑賞」するとかくが、桜の花をかんしょう
するのは、「観賞」とかねばならない。そのほか、自然現象を「観照」
するととかく、また、ほとんどはますときは「勘賞」するとかく、感心して
ほめるときは、「感賞」するとかく。しかし、すすめはですときは「勘
契」とかうことになっている。これら意味の違いは、きわめて微妙であり、
はたして同音異義語といえるであろうか。これらのほかに、「かんしよう」
という単語には、「干渉」、「感傷」、「環礁」、「緩衝」、「管掌」、
「塩性」、「冠省」、「簡捷」、「奸商」、「姦商」など合計166箇の熟
語が辞書にはのっている。「きしゃ」という言語には、6種類の別字の熟
語があるし、「きこう」などは24種類もの熟語を数えることができる。
名詞だけでなく、動詞にも同様の問題が多く、「はかる」については「計
る」、「量る」、「測る」、「図る」、「説る」などがあり、前後の文脈
から文書を書く人が使いたい漢字はどれであるかを判定するコンピュータ
処理の可能性にも限界があることは明らかである。むしろ、コンピュータ
処理だけでなく、われわれ日本語を使うものにとっても、ましてわしく誤
用の多い上述のような漢字の使いわけは整理しなければならない問題であ
る。
上記の「かんしゅう」、「きこう」など、同音異字熟語が多いといっても、それらの熟語の使用頻度を考えると大きな差がある。「かんしゅう」では、「鑑賞」、「干渉」、「感傷」などが頻度が高い。これら高頻度の熟語を使ってうまく変換出力できれば、95％ぐらいのケースでも正しい変換ができたとしてもよいであろう。100パーセントの変換率の達成を目標とするか、95％を目標とするかで変換処理システムの規模や複雑さ、および労力などすべて違ってくる。

3.3 カタカナ入力方式による日本語情報処理の実験的考察

3.3.1 システムの概要

前節までに日本語情報処理の全般的な考察を述べるとともに、いろいろな利点や効果をもったカタカナ入力方式の意義や必要性などについて述べてきたが、ここでは実験的にカタカナ入力方式による日本語情報処理システムを設計し、システムの拡張性や可能性についての考察を述べることにする。なお、この概念にもとづいた実験システムについては第6章で述べることにする。

このシステムは、オフライン日本語情報処理機能を初期のステップとして考察し、さらにオンライン処理からリモート処理にまで拡張することを目的にして設計している。このシステムのねらいは、日本語情報を処理するうえで、もっとも困難な問題となっているデータ・エントリーに工夫がなされており、専用オペレータを必要としない簡易入力処理に注目される。すなわち、カナ・タイプライタと同じキーボードに漢字、ひらがななどのフントショーション・キーを付加しただけで、カタカナ→漢字・ひらがなに変換可能な機能が完備されている。なお、漢字出力装置はすでに開発されているドット方式の漢字プリンタを想定して考察した。

3.3.2 システムの特長

(1) ハードウェア・システムの拡張性が配慮されている。単に日本語
文章を入出力するだけでなく、さらには、オンライン処理からリモート処理までの総合情報処理システムの思想にもとづいて設計されてい
る。
(2) 出力に関しては、特殊な用紙を必要としない方を採用しオペレーション・コストの低減をはかる。
(3) 単語単位にコード変換することによりつけのメリットがあげられる。
・単語ハンドリングによって、入力オペレータが文脈を理解ながらエントリできる。
・漢字に比べてキーボードの数が少ないため入力オペレーション時間が短縮される。
・変換テーブルの記憶構造は、階層（レベル）化方式を採用することでより漢字コード検索および変換処理の効率向上に寄与できる。
(4) システム構成を上位システムと下位システムとに分化した階層システムをとることによりつけのようなメリットがある。
・プロセス・タイムのスピード・アップ
・システム設計の容易性・拡張性
・業務を上位・下位のレベルに分割し下位の作業を小形コンピューターを行わせることによって、大形コンピュータの増設よりもはるかに経済的である。
・入力時点での速い応答を求められる作業要素は、下位レベルのコンピュータ側で処理されている。
(5) できる限り多種類の文字を入力可能にする。
データ・ファイル作成・更新プログラム・パッケージの完成によりユーザの目的に適したマスタ・ファイルをあらかじめ準備しておくことによって、地名、人名、一般用語などの単語群を、ファイルの容量限度にまで拡張することができる。
(6) 婚束は、インテリジェント・ターミナル・システムの概念を採用す
る。

それにより、入力コードを内部コードに変換するだけでなく、入力データのチェックおよび発送機能も端末側が兼ね備える。

また、さらには、内部コードへ変換した際に同音異義語テーブルによって、基本単語への変換も行なうので、上位のメイン・コンピュータへのロードの軽減に寄与する。

(7) 多くの文字を印刷可能とする。

3.3.3 システムの構成

このシステムは、下記の各機器によって構成され、そのシステム構成は図3-1から図3-3に示すとおりである。

1. オフライン・システム
   ① カタカナ入力キー・ボード
       ページ・ライタ
       付加機能
   ② 漢字入力制御装置
   ③ 漢字プリンタ
   ④ 漢字出力制御装置
   ⑤ 印刷装置
   ⑥ 中形漢用コンピュータ・システム（メイン・コンピュータ）

2. オンライン・システムおよびリモート・システム

   上記機器のほかにつきの機器が必要である。
   ⑦ 通信制御装置
   ⑧ ターミナル・コントローラ
   ⑨ その他、通信回線用コントローラ
図3-1 システムの構成-(1)オフライン処理
図 3-2 システムの構成—(2)オンライン・インライン処理

図 3-3 システムの構成—(3)リモート処理
3.3.4 総合的機能の説明

システムの総合的機能は、図3－4に示す3つの部門に大きく分類することができる。

![システムの総合的機能図]

3.3.5 各パートの機能

(1) 漢字入力コントロール・パート

① 同音異義語テーブルへのローディング

このパートのもっともねらいとするところは、同音異義語の変換機能にある。

同音異義語マスタ・ファイルはあらかじめ、上位のメイン・コン
ピュータにより作成・更新されているので、同ファイルを漢字入力制御装置のメモリにロードする。（イニシャル・ラン）

したがって、日本語のデータをエントリする以前に必ず、このイニシエイションを行なっておかなければならない。

図 3-5 同音異義語テーブルのイニシャライズ

② データ入力のプロセス

カタカナ入力のキーボードは、漢字入力制御装置に接続され、付加機能として、ディスプレイおよびページ・ライタがあげられる。

また、データの入力の他に、ベリファイ、サーチなどのオペレーションを行なうこともできる。

データ・エントリのプロセスは、キーボードから入力されたデータ（１単語）に対し、入力イメージのチェックを行ない、内部コードに変換する。このコードをもとに１単語分のコードを同音異義語テーブルから探し出し、ディスプレイ上に表示する。オペレータがディスプレイ上をチェックしたうえで選択キー番号をタイプインす
ると、漢字入力データ・ファイルに1アイテム（1単語）書き込まれ同時にモニタ・プリントされてから、つぎのアイテムの入力へ移る。

③ 漢字入力装置のインテリジェント・ターミナル化

メイン・コンピュータにあたり負担をかけないように、入力コード変換→入力コードチェック→データの編集→簡単な処理（検索・更新）→出力データの編集などの処理を端末側で行なう。

また、制御装置は、プログラムブル・バッファ方式を採用し、編集方法などの処理方式が変ったとしても最適なバッファ・デザインが可能である。

② 漢字データ・プロセス・パート

このパートは、メイン・コンピュータの処理を示しておりシステムの中軸をなすものである。

すなわち、このパートのシステム構成は、漢字入力および漢字出力制御装置に対し、上位のコントロール・システムとして位置づけられる。

メイン・コンピュータの主な役割はつぎに示すとおりである。

・各種コードの変換処理（漢字単語・ゴジックカタカナ・ひらがな・英数字・記号）
・同音異義語マスタの作成更新
・一般用語ファイルおよび専門用語ファイルの作成
・ユーザ目的別・マスタ・ファイルの作成・更新
・出力用文字パターン・マスタの作成・更新
・出力用文字パターン・コード・ブックの作成
・漢字プリンタ編集テープの作成

① 各種コードの変換処理（漢字単語・ゴジックカタカナ・ひらがな・英数字・記号）
漢字入力データ・ファイルを1ブロックずつ、バッファ・エリアに移し、さらに1単語単位（1アイテム）に分解したうえでワーク・エリアへ1アイテムずつセットする。

そして、それら各アイテムが、漢字単語コード、ゴジック・カタカナコード、ひらがなコード、英数字・記号コードのいずれかに変換を行なる。

同音異義語に関しては、すでに下位のコントロール・システム側で、処理されているので、変換の必要はなく、漢字コード用に一意的に構築規則どおりに編集しなおすだけでよい。

ここでは、ユーザ目的別マスタ・ファイルに残している漢字変換テーブルの記憶構造によって、変換処理効率が大いにダウンしたりアップしたりする。

まずこのシステムにおけるコード変換技術は、つぎのように特徴づけられる。

a. 単語単位で変換を行なうので、漢字を1字ずつコード変換するとよりもはるかに効果的である。

b. 同音異義語は、すでに漢字入力コントロール・パートで変換されているため、コードのサーチがより少ない回数ですむ。

c. 変換テーブルはユーザ目的別、マスタ・ファイル上に登録されているので変換対象コード群のメンテナンスはすでに準備されているユーザリティやライブラリを利用することで容易に行なうことができる。

d. 変換テーブルの記憶装置は、ディスクおよびドラムをベースとしたランダム・アクセス記憶装置を使用し、大量のコード群の蓄積や、速いアクセス・タイムの要求に対して、十分に応じられる。

e. 変換対象データ（1単語）の構造がシンプルでありコード種別の判別機能（Function Identifier=FID）を持っている。
すなわち、キーボード操作によって入力対象の 1 アイテム（1 単語）が、漢字単語、ゴジック・カタカナ、ひらがな、英数字記号のうちどれかを判別する F I D を単語先頭に持つことによって、変換処理を簡単にしていく。

変換テーブルは、前述したように、記憶構造のとり方によって変換スピードが異なる。

すなわち、一般には I S A M ( Index Sequential Access Method ) 構造を採用すればよく、あとは変換テーブルにおけるコードの種類、ボリュームに応じて、つきのような 2 とおりの方法を使用する。

(i) 非階層化構造の変換テーブル

変換テーブル自体が 1 次元構造であるため、1 アイテム（1 単語）ずつ、または、英数字やひらがなの場合は、1 文字ずつサーチを行えばよい。

図 3 ～ 6 非階層構造の変換テーブルによる変換

— 71 —
(ii) 階層構造の変換テーブル

初期のシステムでは適用分野の目的にあった変換テーブルを
準備すればよく、わざわざ変換テーブルを複雑化することはな
い。

しかしながらユーザのニーズをより以上に満たすためには、

- 同音異義語
- 専門用語
- 専門的一般語（使用頻度が多い一般語）
- 一般用語

といった分類が必要になる。

この4つのカテゴライズはつきのようにデータを分割しレベ
ル化することができる。

\[
\text{レベル} \begin{array}{|c|c|}
\hline
1 & 同音異義語テーブル \\
2 & 専門用語 \\
3 & 専門的一般語 \\
4 & 一般用語 \\
\hline
\end{array}
\]

\[
\text{変換率} \begin{array}{|c|c|}
\hline
x \% & y \% (x \leq y) \\
\hline
\end{array}
\]

図 3-7 階層構造の変換テーブルによる変換

したがってこれら変換テーブルが、専門分野別にモジュール
化されてユーザ目的別マスタ・ファイルに登録されることにな
る。

- 72 -
図 3-8 階層化変換テーブルの構成

レベル 3 とレベル 4 との分類はユーザの適用業務によって異なるため、システムの稼動が進むにつれ逐次メンテナンスを怠りなく行なう必要がある。

ただし、レベル 4 に関しては、マス・ストレージの容量をかなり必要とするため、あらかじめ
○ レベル 1 ～レベル 3 の平均変換率と
○ レベル 1 ～レベル 4 の平均変換率とを計測しておく。

ユーザ側の期待変換率と平均変換率とを合わせて使い分ける必要がある。
すなわち、漢字 F I D およびカタカナ F I D をサーチしてから、

- 73 -
ワーク・エリアにセットするまでは、非階層化構造のテーブル・サーチと同等変らないが、その後、レベル1～レベル3またはレベル4までのサーチがシステムに付加されることになる。

その際、レベル4をサーチする場合、アクセス回数が増すことと、一般用語自体が大量であることが予想されるため、運用目的に応じた使い方をすべきである。

したがって多少の変換率を下げたとしても、文章の意味を十分に理解することができるうえにコンピュータのランニング・タイムが少なくて済むことになる。

表3-1 階層化変換テーブルの特徴

<table>
<thead>
<tr>
<th>論層</th>
<th>レベル1</th>
<th>レベル2</th>
<th>レベル3</th>
<th>レベル4</th>
</tr>
</thead>
<tbody>
<tr>
<td>用語類別</td>
<td>同音異義語</td>
<td>専門用語</td>
<td>専門的-一般用語</td>
<td>一般用語</td>
</tr>
<tr>
<td>平均変換率</td>
<td>平均変換率 x%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>対象変換テーブル</td>
<td>同音異義語</td>
<td>専門用語</td>
<td>セレクション</td>
<td>一般用語</td>
</tr>
<tr>
<td>データ件数の比例</td>
<td>300語</td>
<td>1000語</td>
<td>5000語</td>
<td>10000語</td>
</tr>
<tr>
<td>記憶装置</td>
<td>下位コンピュータのメモリ用</td>
<td>インデックス部</td>
<td>インデックス部</td>
<td>DISK/DRUM</td>
</tr>
<tr>
<td></td>
<td>送受信端末</td>
<td>DISC/DRUM</td>
<td>DISC/DRUM</td>
<td></td>
</tr>
<tr>
<td>期待変換率</td>
<td>30%</td>
<td>60%</td>
<td>70%</td>
<td>90%</td>
</tr>
<tr>
<td>電磁プロセスのステップ</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>使用コンピュータ</td>
<td>設定入力</td>
<td>メイン・コンピュータ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(2) 同音異義語マスタの作成・更新

漢字入力制御装置のメモリにロードするために、あらかじめメイン・コンピュータ側で、同音異義語マスタを作成および更新しておかなければならない。

この場合図 3-9(a)のように、同音異義語マスタの媒体は、磁気テープであり更新処理はメイン・コンピュータで行なうことを前提としているが、むしろ漢字入力制御装置にキー・カセットなどのリード・ライト処理装置が付加されていれば、同音異義語の更新作業は、下位コンピュータ側で行なうことも可能である。
図 3-9 (b) キー・カセットを利用した同音異義語マスタ作成

③ 一般用語ファイル、専門用語ファイルの作成・更新

一般用語ファイルは、あらかじめシステム側から提供されるファイルであり主に図 3-10 のような資源から情報を得る。

図 3-10

専門分野用語マスタ・ファイルは、ユーザのニーズによって、たとえば図 3-11 のようなデータ群から形成される。
④ ユーザ目的別・マスタ・ファイルの作成・更新（コード変換テーブル）

前述の一般用語・専門分野用語ファイルからユーザ目的別、マスタ・ファイルをセレクションする。

すなわちこのファイルが漢字コード変換テーブルとなり、専門用語、専門的一般語、その他の一般用語としての階層構造ファイルを構成するものである。

図 3-11

図 3-12 のファイル群は蓄積形データベースであるため、記憶装置の条件として

・大容量格納可能であること
ランダマイズ・アドレスリングができること
ハードウェアのアクセス・タイムはあまり高速を必要としない
ISAM形ファイル構造が可能であること
が必要となる。

5. 出力用・文字パターン・マスタの作成・更新

このファイルには、漢字プリンタで印字したい文字のパターンが
内蔵されており、このファイルを「漢字出力制御装置」のメモリに
ローディングしておくことにより印字可能となる。

文字パターンは、漢字出力制御装置のメモリにローディングする
ために必要となるアドレスとそのパターンが1組にまとめられてお
り、最初のブロックには、漢字出力制御装置のメモリにローディン
グすることを指示するステートメントが入っている。

文字パターン:
① 文字Identifiare
② 印刷に必要なマトリックス・パターン

6. 出力用・文字パターン・コードブックの作成

この処理は、漢字入力および出力にともなってシステムの管理者
が、出力文字パターンを保守するために必要となる。コードブック
としてはつぎの種類のブック・レッドを作成することができる。
• 音訓インデックス
• 画引インデックス
• 一般用語パターン・インデックス
• 各種専門用語パターン・インデックス

図 3-13

⑦ 漢字プリンタ編集テーブルの作成

漢字コード変換テーブルが内蔵されているユーザ・目的別マスタ
・ファイルを参照しながら漢字入力データ・ファイルをセレクト、
コード変換、分類などの処理を行なって、「漢字プリンタ編集テー
プ」を作成する。

単純な漢字プリンタ編集テーブルを作成するには、ユーティリティ
・プログラムを利用すればよいが、特別の計算、分類、編集などを
行なうには、ライブラリ・ルーチンを呼び出して、ユーザがプログ
ラミングを行なうことも可能である。
図 3-14 漢字プリンタ編集テープ作成プログラム構成図
(3) 漢字出力コントロール・パート

① 主な機能

漢字出力制御装置は、あらかじめ、メイン・コンピュータで作成される文字パターン・マスタ・ファイルを漢字プロセッサのメモリ部にローディングしておかなければならない。 （イニシャル・ランニング）これにより、漢字プリンタ編集テーブルからデータをリードして、すでに漢字プロセッサに記憶されている文字パターン・テーブルにしたがって、漢字プリンタにデータを送り込む。

漢字プリンタは、漢字プロセッサから送られてくる、文字パターン、プリント位置情報、および文字のタテ・ヨコのサイズなどの指定によりプリント・アクションが行なわれる。

なお、ここに設計した漢字プリンタは、すでに他の機関で開発されているものからこのシステムに必要な機能を採り入れるなどして設計したものである。

② 漢字出力制御装置の構成

![図3-15 漢字出力制御装置の構成](image-url)
文字パターン・マスタ・ファイルは、メイン・コンピュータ側で
作成・更新される。

3.3.6 日本語情報処理におけるソフトウェア体系

このシステムには、上位のメイン・コンピュータを中軸として下位の
入出力制御装置との間で、円滑なプロセスを行なうために表 3-2 のよ
うなソフトウェアを用意する。

表 3-2

<table>
<thead>
<tr>
<th>ベーシック・ソフトウェア</th>
<th>ユーティリティ・プログラム</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 漢字入出力ファイル・ハンドリング・ルーチン</td>
<td>• 同音異義語マスタの作成・更新</td>
</tr>
<tr>
<td>• プリント・ファイル・コントロール・ルーチン</td>
<td>• 一般用語ファイルおよび専門用語ファイルの作成</td>
</tr>
<tr>
<td>• 各種コード変換ルーチン</td>
<td>• ユーザ目的別・マスタ・ファイルの作成・更新</td>
</tr>
<tr>
<td>• 漢字データ・ソーティング・ルーチン</td>
<td>• 出力用文字パターン・マスタの作成・更新</td>
</tr>
<tr>
<td>• 文字パターン・コントロール・ルーチン</td>
<td>• 出力用文字パターン・コードブックの作成</td>
</tr>
<tr>
<td></td>
<td>• 漢字プリンタ・編集テーブルの作成</td>
</tr>
</tbody>
</table>
4. 固有名詞処理のためのカタカナ入力方式
4. 固有名詞処理のためのカタカナ入力方式

4.1. 固有名詞の構成要素

地名、人名をはじめとする各種の固有名詞のうち、ある種のものは、それらの生成の過程で、相互に転用あるいは部分的に引用された結果、同音・同義の文字が、随所に共通的にみられる。したがって、それら共通的な要素と、特定の種類の固有名詞だけに特有の要素とを区分することは、効果的なカタカナ漢字変換テーブルの構成のうえで、重要な課題である。

そのような意味で、各種の固有名詞の生成の過程をたどり、相互の関連を大きく示したものが図4-1である。

注：1. 矢印の起点が構成要素である。
2. 太めの線は、変換テーブル作成の作業単位となるべきものを示す。
3. 斜線を付したものは、利用上の単位である。

図4-1 固有名詞の構成要素
4.2 固有名詞の変換テーブル

4.2.1 変換テーブルの構成

カタカナ漢字変換システムにおいて、変換テーブルの構成単位を、どの程度の大きさに決めるか、ということは、きわめて重要かつ難解な問題である。仮にコンピュータの内部メモリが無限大であると想定しても、あらゆる分野の、あらゆる種類の基本単語を、1つのテーブルに納めることは、変換テーブルの構成、入力およびメインテナンス、同音異義語の処理などの問題があり、さらには検索に要するサイクルタイム自体も無視できないであろう。

それでは、固有名詞に限定した場合はどうであろうか。前述のように地名、人名、企業名などは、相互に関連し合っているので、全体をまとめて1つのテーブルにすることも、それなりの理由はあると思われる。しかししながら、たとえば市町村の住民関係事務における公職選挙の投票所入場券や、児童の就学通知書の発行などのように、おもに個人を対象とする事務においては、人名および比較的小範囲の地名があれば十分あって、公庁や企業で使われるような、組織、商品、機能などを示す言語は、ほとんど不要である。

したがって、固有名詞については、将来実用化が進んだ場合は、改めて統合を検討するにしても、当面はつぎの3種類のテーブルに分割して作成するのがよいと思われる。

① 地名変換テーブル
② 人名変換テーブル
③ 企業・事業所名（官公庁名を含む）変換テーブル

上記3種類の変換テーブルの使い分けについては、つぎのような組合せが考えられる。
(1) 対住民行政、人事管理など、おもに個人を対象とする事務……地名変換テーブルと人名変換テーブル。
(2) 官公庁の組織管理や都内通達、社会保険料の徴収など、おもに官公庁、企業、事業所などを対象とする事務……地名変換テーブルと事業所名変換テーブル。

(3) 株主名簿の管理、ダイレクトメールの発送など、個人と法人との双方を対象とする事務……前記3種類の変換テーブルの全部（注：この場合でも、あらかじめ名簿を2種類に分けて処理することも考えられる）

4.2.2 変換テーブルの基礎資料

(1) 地名変換テーブルの基礎資料

地名を構成する基本単語の抽出には、以下のような資料の利用が考えられる。これらの資料は、いずれも地名の読み仮名がつけられているので、その意味でも利用価値が高い。

① 都道府県コードおよび市区町村コード

これには、工業標準化法に基づく日本工業規格（JISコード）、行政管理庁が設定した設計基準としてのコード、および自治省が設定した地方公共団体コードという3種類のものがある。

② 町字コード調査表

自治省が、全国統一的な町字コード（別名「住所コード」）を設定するために、昭和46年度以来、全国の市区町村に依頼して作成している資料である。市街地については町・丁（住居表示の街区など）、周辺部および農山漁村については、大字・字またはこれに準ずる通称の区域（一般的には、区、集落、部落などと呼ばれる範囲の区域で、郵便物のあて先の住所にも用いられるもの）を単位として、コード付けをしたもので、住居表示に関する法律（昭和37年法律第119号）に基づく住居表示が行なわれている区域については、これによる表示に合わせて、従来の町名または字名による表示を用いる二重表示が行なわれている。
コードの構成は、都道府県２けた、市区町村３けたの数字（上位のコードは、都道府県コードおよび市区町村コードに同じ）の下に、町字を示す４けたの数字と、１けたの検査数字を加え、計１０けたの数字によって構成されている。

③ 国土行政区画総覧・日本行政区画総覧
コード利用の便利さという点を別にすれば、上記2種類の資料の内容を一応網羅し、かつ市販されているので、変換テーブルの作成には、最も手頃な資料といえよう。

(2) 人名変換テーブルの基礎資料
人名の変換テーブル作成の基礎資料としては、種々のものが考えられるが、東京23区五十音別電話帳（個人名）3冊が、最も適切なものと思われる。その理由としてはつきのような諸点があげられる。
① 全国の高人口集中地区で、各地発行の人名を網羅的にみられる。
② 個人だけを対象に、五十音順に編成されているので、変換テーブルに組み替える手間が少なくて済む。
③ 手近なところにあるので、随時、利用することができる。

(3) 企業・事業所名変換テーブルの基礎資料
企業名および事業所名（官公庁およびその内部組織の名称を含み、以下同じ）の変換テーブル作成の基礎資料としては、つぎに掲げる4種類ものが考えられる。
① 東京23区五十音別電話帳（企業名）
この資料では、企業名を五十音順に配列するほか、官公庁およびこれに準ずる法人も別掲されているので、上記(2)に述べた人名の場合と同様の利点がある。しかしながら、電話の加入着者名義単位であるため、企業および官公庁の内部組織の名称、東京都区部以外における特定の官公庁名簿については、必ずしも網羅されていないという問題がある。
② 社会保険の対象事業所ファイル
厚生省（厚生年金保険、健康保険および児童手当用）と、労働省（失業保険および労働者災害補償保険用）とがある。いずれも東京都内の対面事務所については、コンピュータによりカタカナを用いて処理されているので、カタカナ漢字変換テーブルの作成はもとより、事業用システムそのものに、漢字を導入することが、技術的、コスト的および時間的にみて、比較的容易と思われる。

3 行政機構図

行政機構図には国の行政機関のほかに、国会および裁判所も含まれており、国の機関全部が対象の範囲になっている。

4 職員録

大蔵省印刷局が毎年発行行っている上下二分冊からなる資料である。国の機関を示す組織図においては、上記行政機構図に比べて出先機関にかかわるものが若干簡略化されているが、この点については、人名欄に掲げる職名によってより詳細に読み取れるものである。地方公共団体を含めて点で、前記行政機構図に比べて、より網羅的であるといえばよい。

4.2.3 変換テーブルの編成

固有名詞に係る基本単語の抽出については、上に述べたように、各種の基礎資料が考えられるが、各資料の特性および作業の難易を考慮して、図4-2に掲げるような組合せおよび作業手順によって、変換テーブルを編成するのが合理的であると思われる。
図 4-2 固有名詞変換テーブル編成の手順（案）
4.3 固有名詞の漢字分析
4.3.1 事業所名の分析

(1) 東京23区五十音順電話帳からの事業所名調査
職業別電話帳、上巻、下巻の収録数は、

上巻 約 505 千件
下巻 約 478 千件

と推定できる。このうち個人名だけのものは約半ほどと推定されるが、個人名が事業主である、すなわち個人事業所などの例もかなりあると思われる。この多くは事業所とは別の電話番号、たとえば社長宅、専務宅という例もかなりある。

約100万件近い電話番号の所有者が、即事業所とは、必ずしも断言できないが、個人名を除いては、ほとんど事業所名の形態をなしているようである。

逆に、電話帳に記載されていない事業所もかなりの数にのぼるものと思われるが、この点については推測の域を出ることはできない。

統計局の47年事業所統計調査による全国の事業所数および東京23区内の事業所数は、

<table>
<thead>
<tr>
<th></th>
<th>計</th>
<th>民営</th>
</tr>
</thead>
<tbody>
<tr>
<td>全 国</td>
<td>5,308,892</td>
<td>5,130,579</td>
</tr>
<tr>
<td>東京23区</td>
<td>551,374</td>
<td>544,408</td>
</tr>
</tbody>
</table>

この調査の標本数は、

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>上 巻</td>
<td>1,356</td>
</tr>
<tr>
<td>下 巻</td>
<td>1,306</td>
</tr>
<tr>
<td>合 計</td>
<td>2,662</td>
</tr>
</tbody>
</table>

したがって抽出率は、

約 0.5% $\left(\frac{1}{200}\right)$

である。
(2) 標本抽出の方法と漢字名詞の分析

① 偶数ページの第2列の最上段の事業所

② 事業所名を熟語に分解し、職、係などは除いて、

五十音順別

氏名、態様および名称別

に記録、度数をとる。

ここで、氏名、態様および名称をあらわす熟語をつぎのように定義する。

氏名：事業所名のなかに氏または氏を示すもの

態様：事業所名のなかに事業所に関する状態、形態、様相などを示すもの

たとえば

商品 事業
産業 事業
職業 事業
行為 事業

製作所

名称：事業所名のなかに、地名、屋号（動植物名、氏名略、旧地名、理想像、自然現象、美的表象などを含む）その他を示すもの

（例）

朝日 運輸

態様 事業
名称 自然現象

(3) 調査の結果

事業所名は、およそつぎのような構成になっている。

① 姓（もしくは名）＋態様
（例） 川崎産業
② 地名（もしくは自然、理想、美的なもの） ＋ 態様
（例） 日本鋼管
美容産業
協和銀行
③ 屋号（地名、氏名略など）
（例） 備前屋

三 越

事業所名は、総数 2,662 に対して、使用語数は延べ 5,679 語である。このことは、事業所名は上記の構成にもとづいて、平均 2.1 の単語または熟語で成り立っていることを示している。

表 4-1

<table>
<thead>
<tr>
<th></th>
<th>使用語数</th>
<th>延べ使用語数</th>
<th>比 率</th>
</tr>
</thead>
<tbody>
<tr>
<td>氏名</td>
<td>826</td>
<td>1,248</td>
<td>1.5</td>
</tr>
<tr>
<td>態様</td>
<td>868</td>
<td>2,932</td>
<td>3.4</td>
</tr>
<tr>
<td>名称</td>
<td>969</td>
<td>1,499</td>
<td>1.5</td>
</tr>
<tr>
<td>計</td>
<td>2,663</td>
<td>5,679</td>
<td>2.1</td>
</tr>
</tbody>
</table>

また、単語または熟語を氏名、態様または名称と区分すると、氏名および名称の重複使用度は態様のそれよりも低い。
調査を前半および後半に分けた結果は表 4-2 のとおりである。
表4－2

<table>
<thead>
<tr>
<th></th>
<th>使用語数</th>
<th>延べ使用語数</th>
<th>比 率</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>前半</td>
<td>後半</td>
<td>前半</td>
</tr>
<tr>
<td>氏名</td>
<td>470 / 356</td>
<td>625 / 623</td>
<td>1.3</td>
</tr>
<tr>
<td>態様</td>
<td>566 / 302</td>
<td>1,468 / 1,464</td>
<td>2.6</td>
</tr>
<tr>
<td>名称</td>
<td>554 / 415</td>
<td>775 / 724</td>
<td>1.4</td>
</tr>
<tr>
<td>計</td>
<td>1,590 / 1,073</td>
<td>2,868 / 2,811</td>
<td>1.8</td>
</tr>
</tbody>
</table>

使用語数すなわち新規に出現する単語または熟語は後半の方が少ないのは当然であるが、とくに態様に関する単語または熟語は少ない。これは態様に関する語の頻度が次第に高くなることを示している。表4－3は五十音順ごとの使用語数、重複使用語数および延べ使用語数を各熟語ごとに表に示してある。
<table>
<thead>
<tr>
<th></th>
<th>熟語</th>
<th>氏名</th>
<th>憧</th>
<th>様</th>
<th>名称</th>
</tr>
</thead>
<tbody>
<tr>
<td>使用</td>
<td>重複</td>
<td>延べ</td>
<td>使用</td>
<td>重複</td>
<td>延べ</td>
</tr>
<tr>
<td>語</td>
<td>語数</td>
<td>語数</td>
<td>語数</td>
<td>語数</td>
<td>語数</td>
</tr>
<tr>
<td>ア</td>
<td>53</td>
<td>28</td>
<td>81</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>イ</td>
<td>61</td>
<td>52</td>
<td>113</td>
<td>27</td>
<td>66</td>
</tr>
<tr>
<td>ウ</td>
<td>26</td>
<td>11</td>
<td>37</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>エ</td>
<td>7</td>
<td>3</td>
<td>10</td>
<td>13</td>
<td>28</td>
</tr>
<tr>
<td>オ</td>
<td>55</td>
<td>31</td>
<td>86</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>カ</td>
<td>54</td>
<td>31</td>
<td>85</td>
<td>72</td>
<td>125</td>
</tr>
<tr>
<td>キ</td>
<td>22</td>
<td>8</td>
<td>30</td>
<td>55</td>
<td>91</td>
</tr>
<tr>
<td>ク</td>
<td>25</td>
<td>5</td>
<td>30</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>ケ</td>
<td>3</td>
<td>-</td>
<td>3</td>
<td>52</td>
<td>81</td>
</tr>
<tr>
<td>コ</td>
<td>35</td>
<td>23</td>
<td>58</td>
<td>74</td>
<td>2</td>
</tr>
<tr>
<td>サ</td>
<td>35</td>
<td>26</td>
<td>61</td>
<td>14</td>
<td>103</td>
</tr>
<tr>
<td>シ</td>
<td>29</td>
<td>11</td>
<td>40</td>
<td>116</td>
<td>56</td>
</tr>
<tr>
<td>ス</td>
<td>20</td>
<td>20</td>
<td>40</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>セ</td>
<td>7</td>
<td>5</td>
<td>12</td>
<td>84</td>
<td>2</td>
</tr>
<tr>
<td>ソ</td>
<td>4</td>
<td>-</td>
<td>4</td>
<td>23</td>
<td>19</td>
</tr>
<tr>
<td>タ</td>
<td>67</td>
<td>46</td>
<td>113</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>チ</td>
<td>6</td>
<td>3</td>
<td>9</td>
<td>23</td>
<td>8</td>
</tr>
<tr>
<td>ツ</td>
<td>14</td>
<td>2</td>
<td>16</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>テ</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>ト</td>
<td>21</td>
<td>2</td>
<td>23</td>
<td>35</td>
<td>4</td>
</tr>
<tr>
<td>ナ</td>
<td>36</td>
<td>18</td>
<td>54</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>熟</td>
<td>語</td>
<td></td>
<td>熟</td>
<td>語</td>
</tr>
<tr>
<td>--------</td>
<td>----</td>
<td>----</td>
<td>--------</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td></td>
<td>氏</td>
<td>名</td>
<td>頻</td>
<td>様</td>
<td>名</td>
</tr>
<tr>
<td></td>
<td>使</td>
<td>用</td>
<td>重複使</td>
<td>延べ使</td>
<td>使</td>
</tr>
<tr>
<td></td>
<td>用数</td>
<td>用数</td>
<td>用数</td>
<td>用数</td>
<td>用数</td>
</tr>
<tr>
<td>は</td>
<td>14</td>
<td>1</td>
<td>15</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>ヌ</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>ネ</td>
<td>2</td>
<td>-</td>
<td>2</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>ノ</td>
<td>8</td>
<td>3</td>
<td>11</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>ハ</td>
<td>25</td>
<td>11</td>
<td>36</td>
<td>18</td>
<td>30</td>
</tr>
<tr>
<td>ヒ</td>
<td>19</td>
<td>4</td>
<td>23</td>
<td>19</td>
<td>32</td>
</tr>
<tr>
<td>フ</td>
<td>25</td>
<td>8</td>
<td>33</td>
<td>18</td>
<td>29</td>
</tr>
<tr>
<td>ヘ</td>
<td>2</td>
<td>-</td>
<td>2</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>ホ</td>
<td>17</td>
<td>3</td>
<td>20</td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>マ</td>
<td>26</td>
<td>13</td>
<td>39</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>ミ</td>
<td>34</td>
<td>10</td>
<td>44</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ム</td>
<td>8</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>メ</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>モ</td>
<td>13</td>
<td>3</td>
<td>16</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>ヤ</td>
<td>26</td>
<td>22</td>
<td>48</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>ユ</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>ヨ</td>
<td>17</td>
<td>9</td>
<td>26</td>
<td>15</td>
<td>28</td>
</tr>
<tr>
<td>ラ</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>リ</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td>ル</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>レ</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

- 94 -
<table>
<thead>
<tr>
<th></th>
<th>熟語</th>
<th></th>
<th>態様</th>
<th></th>
<th>名称</th>
</tr>
</thead>
<tbody>
<tr>
<td>使用語数</td>
<td>重複使用語数</td>
<td>延べ使用語数</td>
<td>使用語数</td>
<td>重複使用語数</td>
<td>延べ使用語数</td>
</tr>
<tr>
<td>ロ</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>ワ</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>計</td>
<td>826</td>
<td>422</td>
<td>1,248</td>
<td>868</td>
<td>2,064</td>
</tr>
<tr>
<td></td>
<td>2,932</td>
<td>969</td>
<td>530</td>
<td>1,499</td>
<td></td>
</tr>
</tbody>
</table>

(注)
(1) 重複使用語数：2回以上使用した語の延べ使用語数
(2) (使用語数)+(重複使用語数)=(延べ使用語数)

表4-3をみると、部分的に重複使用語数の非常に多いところがある。
この傾向は、態度に関する熟語に目立っており調査の後半で態度に関する熟語の重複頻度が増加している。したがって、調査件数を増加しても態度に関する新しい熟語の出現増加率は減少して、重複使用の頻度が増加するであろう。すなわち、事業所名の構成要素のなかの態度に関する熟語は反復使用される率が多い。それは、表4-4をみるとさらによくわかる。
<table>
<thead>
<tr>
<th>使用頻度（A）</th>
<th>氏名</th>
<th>態様</th>
<th>名称</th>
<th>1</th>
<th>1</th>
<th>197</th>
<th>3.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>193</td>
<td></td>
<td></td>
<td>商店</td>
<td>1</td>
<td>2</td>
<td>390</td>
<td>6.9</td>
</tr>
<tr>
<td>121</td>
<td></td>
<td></td>
<td>東京</td>
<td>1</td>
<td>3</td>
<td>511</td>
<td>9.0</td>
</tr>
<tr>
<td>108</td>
<td></td>
<td></td>
<td>工業</td>
<td>1</td>
<td>4</td>
<td>619</td>
<td>10.9</td>
</tr>
<tr>
<td>102</td>
<td></td>
<td></td>
<td>会社</td>
<td>1</td>
<td>5</td>
<td>721</td>
<td>12.7</td>
</tr>
<tr>
<td>82</td>
<td></td>
<td></td>
<td>事業</td>
<td>1</td>
<td>6</td>
<td>803</td>
<td>14.1</td>
</tr>
<tr>
<td>81</td>
<td></td>
<td></td>
<td>業務</td>
<td>1</td>
<td>7</td>
<td>884</td>
<td>15.6</td>
</tr>
<tr>
<td>64</td>
<td></td>
<td></td>
<td>（日本）</td>
<td>1</td>
<td>8</td>
<td>948</td>
<td>16.7</td>
</tr>
<tr>
<td>61</td>
<td></td>
<td></td>
<td>工業所</td>
<td>1</td>
<td>9</td>
<td>1009</td>
<td>17.8</td>
</tr>
<tr>
<td>46</td>
<td></td>
<td></td>
<td>事務所</td>
<td>1</td>
<td>10</td>
<td>1051</td>
<td>18.5</td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
<td>医院</td>
<td>1</td>
<td>11</td>
<td>1083</td>
<td>19.1</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td>研究所</td>
<td>1</td>
<td>12</td>
<td>1113</td>
<td>19.6</td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
<td>建設（電機）</td>
<td>2</td>
<td>14</td>
<td>1171</td>
<td>20.6</td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
<td>電気</td>
<td>1</td>
<td>15</td>
<td>1199</td>
<td>21.1</td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td>営業所</td>
<td>1</td>
<td>16</td>
<td>1226</td>
<td>21.6</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td>自動車</td>
<td>1</td>
<td>17</td>
<td>1250</td>
<td>22.0</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td>協会（富士）</td>
<td>2</td>
<td>19</td>
<td>1290</td>
<td>22.7</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td>印刷（工場）</td>
<td>3</td>
<td>22</td>
<td>1347</td>
<td>23.7</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td>（支店）</td>
<td>1</td>
<td>23</td>
<td>1365</td>
<td>24.0</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td>興業（東洋）</td>
<td>2</td>
<td>25</td>
<td>1399</td>
<td>24.6</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td>機械（工務店）</td>
<td>3</td>
<td>28</td>
<td>1447</td>
<td>25.5</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td>不動産</td>
<td>1</td>
<td>23</td>
<td>1365</td>
<td>24.0</td>
</tr>
</tbody>
</table>

表4-4 事業所名の頻度調査
<table>
<thead>
<tr>
<th>使用頻度</th>
<th>使用語および数</th>
<th>延べ使用語数</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>氏名 態様 名称</td>
<td>上位から延べ使用語数</td>
</tr>
<tr>
<td></td>
<td></td>
<td>累計 (B)</td>
</tr>
<tr>
<td>14</td>
<td>1 2 1 4 3 2</td>
<td>1,503</td>
</tr>
<tr>
<td>13</td>
<td>2 2 1 6 4 2</td>
<td>1,555</td>
</tr>
<tr>
<td>12</td>
<td>1 4 1 6 4 2</td>
<td>1,627</td>
</tr>
<tr>
<td>11</td>
<td>3 7 1 10 5 2</td>
<td>1,737</td>
</tr>
<tr>
<td>10</td>
<td>4 1 5 5 7 1</td>
<td>1,787</td>
</tr>
<tr>
<td>9</td>
<td>2 6 1 9 6 6</td>
<td>1,868</td>
</tr>
<tr>
<td>8</td>
<td>1 3 2 1 5 8 1</td>
<td>1,988</td>
</tr>
<tr>
<td>7</td>
<td>4 10 3 17 9 8</td>
<td>2,107</td>
</tr>
<tr>
<td>6</td>
<td>6 11 5 2 2 1 2 0</td>
<td>2,239</td>
</tr>
<tr>
<td>5</td>
<td>1 3 1 4 6 1 6 6</td>
<td>2,469</td>
</tr>
<tr>
<td>4</td>
<td>1 2 9 1 5 5 2 2 1</td>
<td>2,689</td>
</tr>
<tr>
<td>3</td>
<td>4 4 3 6 1 2 4 3 4 5</td>
<td>3,061</td>
</tr>
<tr>
<td>2</td>
<td>9 6 1 2 7 7 3 0 0 6 4 5</td>
<td>3,661</td>
</tr>
<tr>
<td>1</td>
<td>6 4 5 5 5 8 1 8 2 0 1 8 2 6 6 3 5 6 7 9 1 0 0 0</td>
<td>5,679</td>
</tr>
<tr>
<td></td>
<td>8 2 6 8 6 8 9 6 9 2 6 6 3 5 6 7 9</td>
<td></td>
</tr>
</tbody>
</table>

（注）1. 上位からの延べ使用累計（C）= Σ（使用頻度（A）+ 語数（B））
全体に対する割合（D）= \frac{(C)}{5679} \times 100
2. 頻度16以上の語は語数の代わりに実名を記入してある。
2663語のうち、使用頻度の多い順に延べ使用語数に占める割合は、

<table>
<thead>
<tr>
<th>語数</th>
<th>頻度</th>
</tr>
</thead>
<tbody>
<tr>
<td>28語まで</td>
<td>25.5%</td>
</tr>
<tr>
<td>52語まで</td>
<td>30.6%</td>
</tr>
<tr>
<td>221語まで</td>
<td>47.4%</td>
</tr>
<tr>
<td>645語まで</td>
<td>64.5%</td>
</tr>
</tbody>
</table>

すなわち、使用語の2/3は、延べ使用語数の1/4を占める。事業所名は、
平均21の単語または熟語から成り立っているので全事業所の半数は
28語程度の語のいずれかを使用していることになる。

また、「製作所」あるいは「商店」という単語を使用している事業
所の推計は、それぞれ、200×200＝40,000事業所と考えら
れる。

使用頻度の多い単語を構成する文字は当然使用頻度が高いが、同じ
文字が種々の単語を構成するような文字の頻度は相当に高くなる。
たとえば、

「製作所」の「製」

よりも

「商店」「商会」「商事」の「商」

「工業」「産業」「工業所」「営業所」の「業」

の文字の方がはるかに頻度が多い。（表4-5参照）

<table>
<thead>
<tr>
<th>表4-5</th>
<th>頻度</th>
<th>使用事業所割合</th>
</tr>
</thead>
<tbody>
<tr>
<td>所</td>
<td>518</td>
<td>19.5%</td>
</tr>
<tr>
<td>商</td>
<td>393</td>
<td>14.8%</td>
</tr>
<tr>
<td>店</td>
<td>391</td>
<td>14.7%</td>
</tr>
<tr>
<td>業</td>
<td>360</td>
<td>13.5%</td>
</tr>
<tr>
<td>工</td>
<td>350</td>
<td>13.1%</td>
</tr>
</tbody>
</table>
4.3.2 東京都の住所名の分析

東京都内の（23区および都下の市郡すべてを含む）住所名を、漢字単語の単位に分割するとともに整理集計した結果を概説する。

まず、東京都内の住所を区、市、郡、町、村などの区分に分割し、たとえば、

東京都港区麻布mdat町 1－1－2
↓
東京都/港区/麻布mdat町

のような作業を行なった。このときの数は約1,504件であった。（住所表示整備基準および未整備区域を含む）ついて漢字単語の基本単語の原則にしたがって

麻布mdat町 → 麻布//dat町
市谷加賀町 → 市谷/加賀町
西大久保 → 西/大久保

のように分割した。これは、「麻布」、「市谷」、「西」をそれぞれ接頭語とみなしたためである。

○「麻布」はこの他に

麻布 / 十番
麻布 / 睦穴町
などにも使われるし、「西」は、

西 / 神田
西 / 落合
西 / 蒲田

のように接頭語として数多く（高頻度）使われるので別に登録したほうが効果的である。しかも、「麻布mdat町」のように字数の多い漢字単語の登録によって、記憶装置へのテーブル格納が効率悪くなる原因になる。

－99－
こうした作業を漢字単語について頻度をカウントしながら集計し、カタカナ——漢字対応表を表4－6のように作成した。その結果を分析してみると、その概数は、

総単語数 約1,745件
同音異義語数 約30件

である。同音異義語の例をいくつかあげてみるとつきのようなものがある。

読み仮名  同音異義語
アオイ  葵、青井
アサヒチョウ  朝日町、旭町
イズミチョウ  和泉町、泉町
エバラ  荆原、江原
タマガワ  玉川、多摩川
トヤマチョウ  ユ山町、富山町

（以下略）

いっぽう、東京都内の住所名にどのような漢字がどれくらい使われているか調べるために、前述の作業に引きつづき、漢字、1字1字についての分析を行なった。

延べ字数 約4,461字
字種 約543字

（当用漢字432字、他111字）

使用頻度の高い漢字はつきのようなものである（頻度30以上）
町 (612)  田 (130)  東 (90)
大 (80)    谷 (80)    西 (76)
本 (75)    川 (74)    小 (62)
山 (62)    南 (61)    野 (60)
上 (58)    神 (54)    中 (53)
橋 (48)    木 (48)    北 (46)
台 (45)    原 (45)    日 (44)
井 (44)    市 (43)    下 (40)
新 (40)    島 (38)    千 (38)

これら頻度 30 以上の漢字 27 種で、2146 という数になり、延べ
字数 4461 字の約 48% を占める割合になる。いわゆる東京都内の住
所名を扱うときに、2 度に 1 度はこれら 27 種の漢字のいずれかを扱っ
ていることになる。

表 4-6 カタカナ — 漢字対応表の例

<table>
<thead>
<tr>
<th>カスヤ</th>
<th>畑谷</th>
<th>カサイ</th>
<th>葛西</th>
</tr>
</thead>
<tbody>
<tr>
<td>カミヤマチョウ</td>
<td>神山町</td>
<td>カスミ</td>
<td>加住</td>
</tr>
<tr>
<td>カミヤマ</td>
<td>神山</td>
<td>カタクラ</td>
<td>片倉</td>
</tr>
<tr>
<td>カナメ</td>
<td>要</td>
<td>カミカワ</td>
<td>上川</td>
</tr>
<tr>
<td>カスガチョウ</td>
<td>春日町</td>
<td>カワ</td>
<td>川</td>
</tr>
<tr>
<td>カスガ</td>
<td>春日</td>
<td>カワグチ</td>
<td>川口</td>
</tr>
<tr>
<td>カミヤチョウ</td>
<td>神谷町</td>
<td>カドタチョウ</td>
<td>門田町</td>
</tr>
<tr>
<td>カミヤ</td>
<td>神谷</td>
<td>カドタ</td>
<td>門田</td>
</tr>
<tr>
<td>カガ</td>
<td>加賀</td>
<td>カミイチブ</td>
<td>上条分</td>
</tr>
<tr>
<td>カツシカク</td>
<td>萩飾区</td>
<td>カノヤ</td>
<td>叶谷</td>
</tr>
<tr>
<td>カツシカ</td>
<td>萩飾</td>
<td>カミト</td>
<td>神戸</td>
</tr>
<tr>
<td>カナ</td>
<td>金</td>
<td>カミツギ</td>
<td>神着</td>
</tr>
<tr>
<td>カマクラ</td>
<td>鎌倉</td>
<td>カナモリ</td>
<td>金森</td>
</tr>
<tr>
<td>カメアリ</td>
<td>亀</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

-101-
4.3.3 日本人の姓の分析

カタカナ漢字変換処理における固有名詞処理のうちで、もっとも問題とされている日本人の姓について行なった調査分析の結果を概説する。
参考資料は「日本人の姓」（佐久間英、六藝書房）である。

調査によりあげた姓の数は、当書籍に掲げてある一覧表のなかから頻度の高い上位 4,000 である。第 1 位の「鈴木」、第 2 位の「佐藤」、第 3 位「田中」………以下第 4,000 位「矢島」というように。
これらの姓の順位 1, 2, ……4,000 をキーにして、姓の読み方（代表的な読み方がとりあげてあるが、1〜6 とおりである。）をカタカナでカード・バンチした。

<table>
<thead>
<tr>
<th>00001</th>
<th>スズキ</th>
</tr>
</thead>
<tbody>
<tr>
<td>00002</td>
<td>サトウ</td>
</tr>
<tr>
<td>00003</td>
<td>タナカ</td>
</tr>
<tr>
<td>……</td>
<td>……</td>
</tr>
<tr>
<td>00015</td>
<td>イノウエ</td>
</tr>
<tr>
<td>00015</td>
<td>イウエ</td>
</tr>
</tbody>
</table>

姓の読み方が何とおりか複数になるものもあるために、データ件数は、4,000 から 5,155 に増加した。このカードを磁気テープに入力し、カタカナ・ソートを行なってから、同音異義語の対応表を作成した例が表 4-7 である。読み方の数 4,312 のうち約 600 が複数対応であることがわかった。また、対応のもっとも多いのは、

サカイ……酒井、坂井、鞍、阪井、鶴、逆井
シオウジ……庄司、庄子、東海林、荘司、正司、小路
ハタノ……波多野、畑野、秦野、波田野、幡野、羽田野

の 6 対応であった。
これに対して 2 対応の読み方は約 480 (80 %) もあることから、2

102
対応の読み方についてカタカナを漢字に直して調べてみると、

アサイ 浅井、朝井
アサオカ 浅岡、朝岡
アサカワ 浅川、朝川
アサクラ 浅倉、朝倉
アサヤマ 浅山、朝山

という例のように、頭1字「浅」と「朝」が違うといったものも数多くある。こうした特徴のある2対応については、何らかの工夫によって同音異義語の減少が果されると思われる。

いっぽう、こうした頻度の高い姓をテープに持たせることによって、ある種の姓に関する対応の割合はどの程度になるか。これはカタカナから漢字への変換処理に役立つ調査である。参考例として実験してみたのは、「電気学会名簿」（昭和41年）であるが、約15,000人の会員から1,000件のサンプリングをしてみると、その対応数は約70%であった。いわゆる15,000人のうち10,500人分は、この4,000の姓でたりるということである。

−103−
表 4-7 姓名の読み方に対する同音異義語の例

<table>
<thead>
<tr>
<th>アイザワ</th>
<th>相沢・会沢</th>
<th>アラキ</th>
<th>荒木・新城・新木・安楽</th>
</tr>
</thead>
<tbody>
<tr>
<td>アイダ</td>
<td>会田・相田・間</td>
<td>アラタ</td>
<td>新田・荒田・新</td>
</tr>
<tr>
<td>アイハラ</td>
<td>相原・粟原原</td>
<td>アラタニ</td>
<td>新谷・荒谷</td>
</tr>
<tr>
<td>アイバ</td>
<td>相場・要庭</td>
<td>アラマキ</td>
<td>荒巻・荒牧</td>
</tr>
<tr>
<td>アオト</td>
<td>青砥・青戸</td>
<td>アラヤ</td>
<td>新谷・荒谷・新屋・新家</td>
</tr>
<tr>
<td>アカギ</td>
<td>赤木・赤城</td>
<td>アラヤマ</td>
<td>新山・荒山</td>
</tr>
<tr>
<td>アキタ</td>
<td>秋田・明田</td>
<td>アンザイ</td>
<td>安西・安東</td>
</tr>
<tr>
<td>アキバ</td>
<td>秋葉・秋庭・秋場</td>
<td>アンドウ</td>
<td>安藤・安東</td>
</tr>
<tr>
<td>アキモト</td>
<td>秋元・秋本</td>
<td>イイ</td>
<td>伊井・井伊</td>
</tr>
<tr>
<td>アサイ</td>
<td>浅井・朝井</td>
<td>イカリ</td>
<td>猪狩・猿</td>
</tr>
<tr>
<td>アサオカ</td>
<td>浅岡・朝岡</td>
<td>イサカ</td>
<td>井坂・伊坂</td>
</tr>
<tr>
<td>アサカ</td>
<td>浅香・安積</td>
<td>イザキ</td>
<td>井崎・伊崎</td>
</tr>
<tr>
<td>アサカラ</td>
<td>浅川・朝川</td>
<td>イザワ</td>
<td>伊沢・井沢</td>
</tr>
<tr>
<td>アサクラ</td>
<td>朝倉・浅倉</td>
<td>イシイ</td>
<td>石井・石居</td>
</tr>
<tr>
<td>アサダ</td>
<td>浅田・朝田・麻田</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>アサノ</td>
<td>浅野・朝野・麻野</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>アサミ</td>
<td>浅見・浅海</td>
<td>(以下略)</td>
<td></td>
</tr>
<tr>
<td>アサヤマ</td>
<td>浅山・朝山</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>アダチ</td>
<td>足立・安達</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>アベ</td>
<td>阿部・安部・安倍</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>アライ</td>
<td>新井・荒井・新居</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
5. 分野別の漢字および漢字単語の分析
5. 分野別の漢字および漢字単語の分析

5.1. 大学内文書に使われる漢字・熟語調査

教育・研究機関内部の文書に使われる漢字の文字と熟語の種類および出現頻度などを調べる目的で、大学における学校広報をサンプルにとって調査を行なった。これは毎月2回にわたって、教職員に配布されるもので、学校からの通知事項、職員の住所や身分の移動、結婚、出産、改名などのニュースを載せているものである。この広報のうちから約1年間分合計、22部を標本として選び、とくに、そのなかの「公告」、「学事」、「通知」、「報告」、「規約」の欄に現われた漢字をすべて調査した。（サンプルを図5-1に示す）

調査の結果を集計すると、漢字熟語の出現延べ数は7,390回であったから、広報1部あたりにすると約350個（延べ数）の熟語が使われていることになる。また、漢字の文字としては、延べ総数1,6286回であり1部あたり約780個の漢字文字が使われていることがわかる。

各文字や熟語は、それぞれ異なる出現頻度で現れており、それら熟語や文字が何個あったかを整理したものが、表5-1、表5-2である。この表5-1からもわかるように、全部で1,293種類の漢字熟語が現われた。このうちもっとも頻度の高い熟語は「大学」という熟語であり、その出現頻度は343であった。頻度が30回以上の熟語は41個ありそれらを表5-3に示す。1,293種類の熟語中、頻度が1～2回のものは、表5-1からもわかるように749種あり、全体の熟語の58％近くになっている。

文字種の方は、以外に少なく、合計804種の文字が用いられている。このうち「学」の文字がもっとも多く用いられており頻度は1,327回であった。また、表5-2からもわかるように804種の文字のうち、頻度が
1〜2回の文字が273種あり、全体の32.5%に相当する。

表5−3にみられるように、使用頻度の高い41個の熟語の頻度の和は2377回であり、これは漢字熟語出現の総延べ数7390回のうちの、32.2%に相当する。すなわち、調査した約1年分の広報22部に使われた漢字熟語の32.2%は、表5−3に示した41個の熟語のどれかであったことになる。このように、頻度順位で何位までの熟語で、何パーセントの累積頻度になるかを示したのが表5−5である。また、表5−5の関係を図5−2にグラフで示す。

漢字文字についていえば、表5−4の使用頻度の高い40個の頻度の和は7693回であり、これは漢字文字出現の総延べ数16286回のうちの47.2%に相当する。このことから、この文書に現われる漢字文字は、およそ2回に1回は表5−4に示した40個の文字のどれかであったということである。これらの関係をさらに明らかにするため、頻度順位で何位までの文字で何パーセントの累積頻度になるかを示したのが表5−6である。

表5−5、表5−6および図5−2、図5−3は、カタカナ入力漢字かな混り文変換処理のための変換テーブルに登録しておくべき熟語の個数すなわち変換テーブルの規模、とそこから生ずるメリットとしての変換能力に関連して、きわめて重要なデータを提供している。

一般論として、変換テーブルが大きいほど、すなわち、そこに登録されている漢字文字や熟語が多いほど、変換能力は高まる。しかし、その反面変換テーブルが大きいほど、そこから特定の語を探索する処理は複雑となる、変換処理時間は増大する。この変換処理時間が増大する割合は、変換テーブルの大きさ、Xに比例して増大するというよりも、Xのべき乗に比例して、あるいは指数関数的に増大していくと考えられる。いっぽう、図5−2、図5−3に見られるように、登録単語数を増加していっても、それから得られる変換率、すなわち探している単語が変換テーブルにある
確率の増加分はきわめて急速に連続して行く。このことは、どの言語にも
いえる共通の特性として、言語はきわめて使用頻度の高い単語と、非常に
稀にしか使われない単語など使用頻度が極端に異なる単語から成り立っ
ているという事実から生じてくる。この傾向は、経済とか電気工学などと
特定の分野で使われる単語に限定すると、ますます顕著になってくる。

われわれの調査した文書においても、表5－3に示した41個の使用頻
度の高い熟語を変換テーブルに登録しておけば、この文書に出現する、お
よそ3.2%の漢字熟語を変換できることがわかる。使用文字の種類にいた
っては、表5－4の40個の文字だけで、この文書に現われる約半分をカ
バーする。さらに、図5－2からは516個の熟語を登録しておけば
85%の熟語漢字変換率を達成できることがわかる。この変換率を約10%
引き上げて95%にするには、約1000個の熟語が登録されていなけ
ればならず、変換率85%のときの約2倍の登録単語数が要求される。い
ま、仮に変換処理に必要な労力や時間は登録単語数の2乗に比例すると
仮定すると、変換処理にかかる負担は4倍になる。すなわち、変換率を85%
から95%に10%だけ改良するのに要する費用は、変換率が85%
のときの費用の約4倍になるということになる。

同様のことが、漢字文字種と変換に関する図5－3についてもいえる。
わずか200個程度の文字が登録されていれば、この文書の85%の文字
がカバーできる。しかし、もし95%の文字をカバーしようとすれば、
400個近くの文字が登録されていなければならない。

以上の観察から、かなり一般的な結論として、カタカナ入力、漢字かな
混り文変換処理の開発についての基本的前提は100%変換率を目標とし
ないということである。それは、言語のもつ特性からして実現は不可能で
ある。また、85%程度の変換率をねらう、95%を考えうかで、技術開
発のストラテジーのうえでも、かなり根本的な相異がでてくるのではない
だろうか。いずれにしても、どの程度の変換能力を達成していたかについて
の明確な認識が必要である。その際考慮されるべき基本要因は、図5-2や図5-3に表われるように変換テーブルの規模を拡大してもそれが生み出す変換能力の改善は急速に遅減するということ。また、いっばくでは変換処理の費用は少なくとも変換テーブルの規模のベキ乗に比例するであろうと思われることである。変換能力とそれに要する開発費用および処理費用とのトレード・オフを考慮した適正規模というもののが存在するという認識が重要である。

これまでに行なってきたカタカナ入力漢字かな混り交出力方式による日本語情報処理システムの実験によれば電気工学分野の登録単語数2964個の変換テーブルの規模で変換処理を行なった実験結果によれば、漢字変換要求数1,496件程度の処理量に対して、変換率は71.3％であった。しかも、変換テーブルの登録単語から使用された単語はわずか1,340ある400語にすぎなかった。登録単語数の割には、変換率が高いのは、登録単語の選択が電気工学用語集からなり姿意的になされたためと思われる。ここでの調査した文書の場合は、変換率が71％程度を達成目標とするなら、図5-2から、登録単語が160個程度の変換テーブルで十分であることがわかる。これらは登録単語の選択にあたっては、対象文書の分野における単語の使用頻度の調査がいかに大切であるかを指摘している。
### 学事

昭和49年度入学式日程

<table>
<thead>
<tr>
<th>学  院</th>
<th>日  期</th>
<th>时  间</th>
</tr>
</thead>
<tbody>
<tr>
<td>大学院</td>
<td>4月6日（土）午後1時</td>
<td></td>
</tr>
<tr>
<td>大学第一、第二部</td>
<td>同上時</td>
<td></td>
</tr>
<tr>
<td>他学部</td>
<td>4月9日（火）午前10時</td>
<td></td>
</tr>
<tr>
<td>健学部</td>
<td>同上時</td>
<td></td>
</tr>
<tr>
<td>科 学</td>
<td>4月6日（土）午前9時30分</td>
<td></td>
</tr>
<tr>
<td>人文学</td>
<td>同上時</td>
<td></td>
</tr>
<tr>
<td>技術</td>
<td>4月5日（金）午前9時30分</td>
<td></td>
</tr>
<tr>
<td>理学部</td>
<td>同上時</td>
<td></td>
</tr>
<tr>
<td>共同</td>
<td>4月4日（木）午前10時</td>
<td></td>
</tr>
<tr>
<td>外部</td>
<td>4月3日（水）午前10時</td>
<td></td>
</tr>
</tbody>
</table>

### 規 約

大学学生規則一部改正

第２条 学生部長は大学院の下において、各学部長と連絡を保ちつつ、学生部を経営し、学生生活の維持整然ならびに調理に関する事務を処理する。

学生部長は、青谷キャンパスに2名以内、他学部キャンパスに1名をおく。学生部長は学生部長を担当し、関係にかかわる関係部を知悉する。

学生部長および学部部長は教員の中から大学院が大学院会に諮り、これに委任する。

学生部長および学生部部長の任期は2年とする。

（昭和48年6月27日 理事会教）

### 通 知

昭和49年度メキシコ政府留学生募集（文部省）

I. 招考对象、募集人員

A. 大学で開設中の講座を受講する留学生（30名）

メキシコ語、メキシコおよび中南米の歴史、地理、経済、歴史、文学、人文学、考古学等

B. 研究機関等において特定のテーマによる専門的

研究を行う留学生（10名）

II. 留学期間

昭和49年5月から10月間

III. 賞  与

往復航空費、オフスチック体験センター（10日間）費用負担、関税（約9千円）超額負担、入学金、授業料免除

IV. 応募資格

1. 年収30万円未満の者で、大学卒業または大学院

在学生で、あるいはこれらのことを授業または修了

している者

2. 心身ともに健全な日本人で、スペイン語の基礎

的な能力を有し、自身の能力および将来できる

V. 応募手続

応募者は応募書類を大学院当局に提出し、

3月3日（土）までに文部省大学学術部留学生課

あて提出すること

### 報 告

海外出張・帰国

寺崎 和彦（大学理工学部教授）

第4回国際クリリング観分研究会およびハワイ大学

東南センター研究討合会のため7月12日から27日まで

アメリカへ出張した

金 徳兼（大学理工学部教授）

学会出席および研究、滞在のため7月23日から9月

15日までアメリカ、イギリス、イタリアへ出張

岩村 幸一（大学理工学部教授）

世界平和に対する大学の役割に関する各分野の協合

いのため7月21日から29日まで韓国（ソウル市）へ出

張した

図5 - 1 調査文書の例

- 109 -
図5-2 大学内文書の漢字熟語：頻度順位とパーセント累積頻度

図5-3 大学内文書の漢字文字：頻度順位とパーセント累積頻度

-110-
表5-1 大学内文書の漢字熟語の頻度と熟語数

<table>
<thead>
<tr>
<th>頻度</th>
<th>熟語数</th>
<th>熟語数パーセント</th>
</tr>
</thead>
<tbody>
<tr>
<td>343</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>150〜300</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>80〜149</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>50〜79</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>40〜49</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>30〜39</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>20〜29</td>
<td>33</td>
<td>2.6</td>
</tr>
<tr>
<td>10〜19</td>
<td>106</td>
<td>8.2</td>
</tr>
<tr>
<td>6〜9</td>
<td>117</td>
<td>9.0</td>
</tr>
<tr>
<td>3〜5</td>
<td>246</td>
<td>19.0</td>
</tr>
<tr>
<td>2</td>
<td>215</td>
<td>16.6</td>
</tr>
<tr>
<td>1</td>
<td>534</td>
<td>41.3</td>
</tr>
<tr>
<td></td>
<td>1,293</td>
<td>100</td>
</tr>
</tbody>
</table>

表5-2 大学内文書の漢字文字の頻度と文字数

<table>
<thead>
<tr>
<th>頻度</th>
<th>文字種</th>
<th>文字種パーセント</th>
</tr>
</thead>
<tbody>
<tr>
<td>1327</td>
<td>1</td>
<td>1.2</td>
</tr>
<tr>
<td>301〜500</td>
<td>2</td>
<td>2.5</td>
</tr>
<tr>
<td>201〜300</td>
<td>6</td>
<td>7.5</td>
</tr>
<tr>
<td>101〜200</td>
<td>28</td>
<td>3.5</td>
</tr>
<tr>
<td>81〜100</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>61〜80</td>
<td>24</td>
<td>3.0</td>
</tr>
<tr>
<td>41〜60</td>
<td>26</td>
<td>3.2</td>
</tr>
<tr>
<td>21〜40</td>
<td>78</td>
<td>9.7</td>
</tr>
<tr>
<td>11〜20</td>
<td>99</td>
<td>1.23</td>
</tr>
<tr>
<td>6〜10</td>
<td>104</td>
<td>1.29</td>
</tr>
<tr>
<td>3〜5</td>
<td>156</td>
<td>1.94</td>
</tr>
<tr>
<td>2</td>
<td>97</td>
<td>1.21</td>
</tr>
<tr>
<td>1</td>
<td>176</td>
<td>2.19</td>
</tr>
<tr>
<td></td>
<td>804</td>
<td>100</td>
</tr>
</tbody>
</table>
表5-3 使用頻度の高い漢字熟語

<table>
<thead>
<tr>
<th>順位</th>
<th>熟語</th>
<th>使用頻度</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>大学</td>
<td>343</td>
</tr>
<tr>
<td>2</td>
<td>昭和</td>
<td>144</td>
</tr>
<tr>
<td>3</td>
<td>研究</td>
<td>135</td>
</tr>
<tr>
<td>4</td>
<td>年度</td>
<td>94</td>
</tr>
<tr>
<td>5</td>
<td>専攻</td>
<td>79</td>
</tr>
<tr>
<td>6</td>
<td>募集</td>
<td>78</td>
</tr>
<tr>
<td>7</td>
<td>出張</td>
<td>74</td>
</tr>
<tr>
<td>8</td>
<td>学部</td>
<td>71</td>
</tr>
<tr>
<td>9</td>
<td>午前</td>
<td>59</td>
</tr>
<tr>
<td>10</td>
<td>教授</td>
<td>57</td>
</tr>
<tr>
<td>11</td>
<td>高等部</td>
<td>54</td>
</tr>
<tr>
<td>12</td>
<td>午後</td>
<td>54</td>
</tr>
<tr>
<td>13</td>
<td>応募</td>
<td>52</td>
</tr>
<tr>
<td>14</td>
<td>日本</td>
<td>51</td>
</tr>
<tr>
<td>15</td>
<td>研究科</td>
<td>49</td>
</tr>
<tr>
<td>16</td>
<td>経済</td>
<td>48</td>
</tr>
<tr>
<td>17</td>
<td>手続</td>
<td>48</td>
</tr>
<tr>
<td>18</td>
<td>入学</td>
<td>46</td>
</tr>
<tr>
<td>19</td>
<td>女子</td>
<td>45</td>
</tr>
<tr>
<td>20</td>
<td>幼稚園</td>
<td>42</td>
</tr>
<tr>
<td>21</td>
<td>教育</td>
<td>41</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>順位</th>
<th>熟語</th>
<th>使用頻度</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>理工</td>
<td>41</td>
</tr>
<tr>
<td>23</td>
<td>学校</td>
<td>40</td>
</tr>
<tr>
<td>23</td>
<td>分野</td>
<td>40</td>
</tr>
<tr>
<td>25</td>
<td>初等部</td>
<td>39</td>
</tr>
<tr>
<td>25</td>
<td>中等部</td>
<td>39</td>
</tr>
<tr>
<td>27</td>
<td>支給</td>
<td>38</td>
</tr>
<tr>
<td>27</td>
<td>短期</td>
<td>38</td>
</tr>
<tr>
<td>29</td>
<td>提出</td>
<td>37</td>
</tr>
<tr>
<td>29</td>
<td>文學部</td>
<td>37</td>
</tr>
<tr>
<td>29</td>
<td>礼拜</td>
<td>37</td>
</tr>
<tr>
<td>32</td>
<td>庶務課</td>
<td>36</td>
</tr>
<tr>
<td>33</td>
<td>発表</td>
<td>35</td>
</tr>
<tr>
<td>34</td>
<td>考査</td>
<td>34</td>
</tr>
<tr>
<td>34</td>
<td>庶務課</td>
<td>34</td>
</tr>
<tr>
<td>36</td>
<td>文学</td>
<td>33</td>
</tr>
<tr>
<td>37</td>
<td>学院</td>
<td>32</td>
</tr>
<tr>
<td>38</td>
<td>教職員</td>
<td>31</td>
</tr>
<tr>
<td>38</td>
<td>経営</td>
<td>31</td>
</tr>
<tr>
<td>38</td>
<td>文部省</td>
<td>31</td>
</tr>
<tr>
<td>31</td>
<td>青山学院</td>
<td>30</td>
</tr>
</tbody>
</table>

（頻度 和 2377）
表 5-4 使用頻度の高い漢字文字

<table>
<thead>
<tr>
<th>順位</th>
<th>文字</th>
<th>頻度</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>学</td>
<td>1327</td>
</tr>
<tr>
<td>2</td>
<td>部</td>
<td>427</td>
</tr>
<tr>
<td>3</td>
<td>大</td>
<td>396</td>
</tr>
<tr>
<td>4</td>
<td>研</td>
<td>263</td>
</tr>
<tr>
<td>5</td>
<td>究</td>
<td>248</td>
</tr>
<tr>
<td>6</td>
<td>教</td>
<td>234</td>
</tr>
<tr>
<td>7</td>
<td>会</td>
<td>225</td>
</tr>
<tr>
<td>8</td>
<td>務</td>
<td>210</td>
</tr>
<tr>
<td>9</td>
<td>年</td>
<td>206</td>
</tr>
<tr>
<td>10</td>
<td>科</td>
<td>179</td>
</tr>
<tr>
<td>11</td>
<td>出</td>
<td>173</td>
</tr>
<tr>
<td>12</td>
<td>山</td>
<td>172</td>
</tr>
<tr>
<td>13</td>
<td>文</td>
<td>165</td>
</tr>
<tr>
<td>14</td>
<td>期</td>
<td>159</td>
</tr>
<tr>
<td>15</td>
<td>授</td>
<td>156</td>
</tr>
<tr>
<td>16</td>
<td>勲</td>
<td>155</td>
</tr>
<tr>
<td>17</td>
<td>等</td>
<td>146</td>
</tr>
<tr>
<td>18</td>
<td>和</td>
<td>145</td>
</tr>
<tr>
<td>19</td>
<td>昭</td>
<td>144</td>
</tr>
<tr>
<td>20</td>
<td>業</td>
<td>144</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>順位</th>
<th>文字</th>
<th>頻度</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>日</td>
<td>139</td>
</tr>
<tr>
<td>22</td>
<td>員</td>
<td>132</td>
</tr>
<tr>
<td>23</td>
<td>事</td>
<td>132</td>
</tr>
<tr>
<td>24</td>
<td>生</td>
<td>132</td>
</tr>
<tr>
<td>25</td>
<td>記</td>
<td>130</td>
</tr>
<tr>
<td>26</td>
<td>者</td>
<td>128</td>
</tr>
<tr>
<td>27</td>
<td>午</td>
<td>124</td>
</tr>
<tr>
<td>28</td>
<td>人</td>
<td>120</td>
</tr>
<tr>
<td>29</td>
<td>国</td>
<td>119</td>
</tr>
<tr>
<td>30</td>
<td>本</td>
<td>117</td>
</tr>
<tr>
<td>31</td>
<td>經</td>
<td>117</td>
</tr>
<tr>
<td>32</td>
<td>理</td>
<td>111</td>
</tr>
<tr>
<td>33</td>
<td>院</td>
<td>107</td>
</tr>
<tr>
<td>34</td>
<td>専</td>
<td>106</td>
</tr>
<tr>
<td>35</td>
<td>書</td>
<td>105</td>
</tr>
<tr>
<td>36</td>
<td>給</td>
<td>104</td>
</tr>
<tr>
<td>37</td>
<td>応</td>
<td>103</td>
</tr>
<tr>
<td>38</td>
<td>費</td>
<td>102</td>
</tr>
<tr>
<td>39</td>
<td>人</td>
<td>98</td>
</tr>
<tr>
<td>40</td>
<td>課</td>
<td>93</td>
</tr>
</tbody>
</table>

（40位までの頻度和 7593）
表5-5 大学内文書の漢字熟語の頻度順位と累積頻度

<table>
<thead>
<tr>
<th>頻度順位</th>
<th>パーセント累積</th>
<th>累積頻度</th>
<th>累積パーセント</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.006</td>
<td>1018</td>
<td>1.4</td>
</tr>
<tr>
<td>20</td>
<td>0.015</td>
<td>1637</td>
<td>2.2</td>
</tr>
<tr>
<td>34</td>
<td>0.026</td>
<td>2177</td>
<td>2.9</td>
</tr>
<tr>
<td>55</td>
<td>0.043</td>
<td>2787</td>
<td>3.7</td>
</tr>
<tr>
<td>71</td>
<td>0.055</td>
<td>3157</td>
<td>4.2</td>
</tr>
<tr>
<td>98</td>
<td>0.076</td>
<td>3636</td>
<td>4.9</td>
</tr>
<tr>
<td>168</td>
<td>0.128</td>
<td>4489</td>
<td>6.0</td>
</tr>
<tr>
<td>298</td>
<td>0.230</td>
<td>5512</td>
<td>7.4</td>
</tr>
<tr>
<td>425</td>
<td>0.329</td>
<td>6069</td>
<td>8.2</td>
</tr>
<tr>
<td>544</td>
<td>0.421</td>
<td>6426</td>
<td>8.7</td>
</tr>
<tr>
<td>759</td>
<td>0.587</td>
<td>6856</td>
<td>9.2</td>
</tr>
<tr>
<td>1293</td>
<td>1.00</td>
<td>7390</td>
<td>10.0</td>
</tr>
</tbody>
</table>

表5-6 大学内文書の漢字文字の頻度順位と累積頻度

<table>
<thead>
<tr>
<th>頻度順位</th>
<th>パーセント累積</th>
<th>累積頻度</th>
<th>累積パーセント</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>0.022</td>
<td>4986</td>
<td>3.0</td>
</tr>
<tr>
<td>20</td>
<td>0.025</td>
<td>5274</td>
<td>3.2</td>
</tr>
<tr>
<td>24</td>
<td>0.030</td>
<td>5814</td>
<td>3.5</td>
</tr>
<tr>
<td>46</td>
<td>0.057</td>
<td>8067</td>
<td>4.9</td>
</tr>
<tr>
<td>57</td>
<td>0.071</td>
<td>8834</td>
<td>5.4</td>
</tr>
<tr>
<td>71</td>
<td>0.088</td>
<td>9710</td>
<td>5.9</td>
</tr>
<tr>
<td>84</td>
<td>0.104</td>
<td>10340</td>
<td>6.3</td>
</tr>
<tr>
<td>106</td>
<td>0.132</td>
<td>11219</td>
<td>6.9</td>
</tr>
<tr>
<td>140</td>
<td>0.174</td>
<td>12299</td>
<td>7.5</td>
</tr>
<tr>
<td>193</td>
<td>0.240</td>
<td>13460</td>
<td>8.2</td>
</tr>
<tr>
<td>304</td>
<td>0.378</td>
<td>14841</td>
<td>9.1</td>
</tr>
<tr>
<td>414</td>
<td>0.515</td>
<td>15522</td>
<td>9.5</td>
</tr>
<tr>
<td>531</td>
<td>0.660</td>
<td>15916</td>
<td>9.7</td>
</tr>
<tr>
<td>804</td>
<td>1.00</td>
<td>16286</td>
<td>10.0</td>
</tr>
</tbody>
</table>

-114-
カタカナ入力漢字から混り文出力システムの開発にあたって、その変換
率を高めるためには、入力するデータの分野によって専用の変換用テーブルを準備し、各テーブルには出現頻度の高い熟語から登録すべきことを述べてきた。

以下に、大学内文書用の変換用テーブルを作成する際の基礎調査の種類
とデータ分析についての注意事項を学院広報に使用された漢字を例にとっ
te、具体的に説明することにする。

(1) 基礎調査用サンプル数の基準
まず、大学内の文書に使われている漢字・熟語の特性を知るために、
どの程度の基礎データが必要であるかが問題となる。そこで、とりあげ
ず約1年分の学院広報に出現した漢字熟語の頻度調査を行ない、その結果
をグラフにしてみた。

広報16部（出現熟語数5,807個）についての頻度曲線を図5－4
に示す。つぎに、広報6部を加えて22部（出現熟語総数7,295個）
について同様に曲線を描いてみると6部分のデータ1,488件を追加し
たにもかかわらずほとんど同じ特性のカーブを持っていることがわかっ
tた。そして、追加した1,488件のうち6部の中に初めて現われた熟語は
147件であり、残りの1,341件は先に作成された16部について
の変換テーブルに既に登録されていた。
図 5-4 出現頻度曲線（16 部分）
(2) 頻出語と使用頻度の分類

出現頻度のグラフを作成すると同時に、ときに頻度の高い語をまとめたものが表 5-3 であるが、これら上位にランクされている語をみると、使用用途に共通性が見出され、いくつかに分類することができる。たとえば、「青山学院」、「大学」、「短期大学」に使われる「短期」、「高等部」、「中等部」、「初等部」、「幼稚園」などの語は「学院全体の機構に関する語群」として一括でき、さらに「経済」、「経営」、「理工」、「文学」、「文学部」、「教育」などの語は「学校に関する語群」として、また、「勤務部」、「勤務課」などは「事務組織に関する語群」のようにまとめることができるであろうし、「昭和」、「年度」、「午前」、「午後」などといった語は「時間を示すための語群」としてまとめられる。このほかにも、「事務用語群」「経済用語群」といった分類によって、ほとんど、すべての語を分類することができるはずである。そして、これらのいくつかの語群は全体として「使用範囲のちがいによる分類」ということでまとめておくことができよう。

1年間の学院広報に出現した漢字および漢字単語を整理してみると、この使用範囲のちがいによって分類された語は年間を通じて、ほぼ均等に頻出しており、また、同じ語群内の語はどれも同じ位の出現頻度をもっているという共通性がある。

(3) 出現漢字の季節変動性

前項の分類方法とは別に、学院広報の発行時期によって語の出現頻度に非常に大きな変化があるものと、そうでないものを分類してみるとことができる。前述したように「学校の機構に関する語」や「時間を示す語」それに「事務用語」などは、一般に一年を通して頻出し、発行時期には、ほとんど左右されていない。これに対して、「募集」、「応募」、「手続」、「考査」、「発表」、「入学」といった語は2月〜4月のい
わゆる入試期の前後の広報に頻出するが他の月には希にしか現われていない。また、表5－3には掲載されていないが「卒業」という語は学年末にしか使われない。このように大学内の広報に使用される漢字熟語は学校全体の行事に依存する傾向が見出される。

このように時期によって使用される漢字熟語が異なるという『出現漢字熟語の季節変動性』が、大学内の広報（文書も含めて）における漢字の最大的な特徴であるといえる。

5.2 医学分野における用語調査

近年、コンピュータによる医学・医療情報処理の発展は、めざましいものである。医学分野においても、従来からコンピュータ処理などで扱われてきた文字は英数字が主体であったが、日本語の文書情報、文献情報の処理には漢字のかなに混じり文を用いることが要求されるようになってきた。コンピュータ処理の前提としては、入力情報の標準化が、最重要の問題であることは、いままでもなく、その例を挙げれば、厚生省では昭和48年度より“医療情報システム検討委員会”の中に“医学用語・コードおよびソーラス委員会”が設けられ、医学用語、コード、ソーラスについての基本的検討が続けられている。

このような動きの中で、医学用語、とくに日本語の医学用語の調査、検討はこれまであまり広くは行なわれてはいなかった。

この節では、上記のことを踏まえて、医学用語集の現状、医学用字、用語に関するこれまでの調査・研究の結果を述べることにする。調査の対象として学術論文の標題や学術用語集を選び行なった。

5.2.1 わが国の医学用語の現状

医学の領域はかなり広く、わが国有数の医学辞典である「南山堂医学大辞典」には、つぎのような学問分野の用語が含まれていることが記載されている。
化学、物理学、生物学、動物学、植物学、遺伝学、心理学、農学、薬学

このように、医学の分野で扱う用語は他の分野の用語よりも多様かつ多岐でなく細分化、専門化も著しい。

日本語の造語の特性、これら医学用語では特に著しく現われ、そのために公認された用語集がなかなか作成しにくい。文部省発行の学術用語集は各学問分野ごとに刊行されているが、医学分野は未利である。

しかし、近年、学界的努力が実り、日本医学会から1974年に「医学用語集」が発行される予定といわれる。これは日本語医学用語標準化の基本的ツールになるものであると期待される。

・医学用語集（1974年）日本医学会（収録用語約5,000語、外国語→日本語）

このほか、これまでにわが国で編集された、医学分野の辞書、用語集としては、下記のものがある。

・医学大字典（1958年）金原出版（約8,000語）

・医学大辞典 南山堂

・内科用語集（1960年）日本内科学会（約14,800語）

また、英語の辞書として代表的なものは、

・Dorland's Illustrated Medical Dictionary

（約5,800語）

が挙げられる。

このように医学全領域をカバーするには数万語の用語が必要になる。しかし、これらはシソーラスのような、ある分類整理された言葉ではなくどちらかと言えば“自然語”である。

現在、日本語の医学用語に関するシソーラスは存在しないが、米国の医学文献検索システムMEDLARSで用いられているシソーラスMeSH（Medical Subject Headings）では、
8,983語（1974年版）
7,590語（1970年版）
5,977語（1964年版）
に分類・整理された用語が用意されている。
医学・医療分野での文書情報としては、学術文献ばかりではなく、病歴（カルテ）や、検査データなど、“生”の情報も存在するが、ここでは省略する。

5.2.2 医学論文の標題に用いられる漢字分析

(1) 一般科学技術論文との対比

医学分野において用いられる漢字は一般的に種類が多くともに当用漢字以外の漢字が他の分野に比べて多くあることが言われている。また、一般の学術分野、とくに科学技術分野で用いられる漢字と医学分野で用いられる漢字とは、それぞれどのような特徴があるであろうか。

これらの問題について、論文のテキストではなく、論文の内容をもっとも簡潔に表わしている論文標題（タイトル）に注目して、考察してみる。調査のサンプルとしては、国立国会図書館で発行している「雑誌記事索引－自然科学編」を取り上げた。

(2) 「雑誌記事索引」（国立国会図書館）における漢字頻度調査

「雑誌記事索引」誌は国立国会図書館から刊行されている、国内でもっとも刊行歴の長いインデックス誌であり、人文科学編と自然科学編（現在は科学技術編）に分けられている。自然科学編は一般編と医学編（薬学を含む）から成っている。この調査は「雑誌記事索引－自然科学編」における漢字の頻度調査、図書館研究シリーズ


-120-
然科学編」の第14巻8号（1963年8月発行）を対象に選び、ここに収録された雑誌論文の標題に用いられている漢字の頻度を調べたものである。
「雑誌記事索引－自然科学編」に収録されている雑誌は、一般編－252種、医学編＝129種（うち欧文誌20種）であり、サンプルに取りあげた論文数は一般編＝1,032件、医学編＝1,095件であり、合計2,127件の論文が調査対象となった。

調査結果はつきのとおりである。
1) 総文字数（延べ数）は一般編で8,671字、医学編では11,544字
合計2,0234字であった。
2) 一般論文標題あたりの平均漢字数は一般編8.4字、医学編で10.5字である。これによって医学論文のタイトルが、他の科学技術分野より漢字が多く使われていることがわかる。
3) 漢字の種類としては、当用漢字（1,850字）に含まれるものを見ると、一般編の方が医学編よりも多く89.3％が当用漢字で、医学編は83％であった。また全当用漢字数のうち50％前後（一般編53.4％、医学編46.6％）が使用されていた。
4) 一般編、医学編の共通使用漢字は708字で漢字の種類1,437字の49.2％であった。
5) 漢字の使用頻度順の字数と使用全数比累計関係を示すと図5-5表5-7のようにになる。この図・表から100字で全体の51％を占めており、200字で68％、300字で79％、400字で85％、500字で全体の90％前後（一般編で88.87％、医学編で91.94％、全編で86.79％）を占めていることがわかる。いわゆる医学編の方が一般編より、同一字数においてカバーできる割合が大きいことが示されている。

ここで、図5-5、表5-8から医学分野の用語が一般の科学技術
分野に比較してどのような特徴があるかについて考察してみる。
まず、医学編で用いられている漢字の種類は1,039種で一般編の1,106種に比べて約10%少ない。いっぽう前述したが、1論文標
題あたりの平均漢字数は、医学編では10.5字に対して一般編では
8.4字で、医学分野の論文タイトルに漢字が多いという特徴がみられ
る。しかも漢字の種類は少ないが、医学特有の字種が用いられている
こともわかる。

つぎに、一般・医学両編全般についての頻度表で第500位までの
字種をみると、一般編あるいは医学編のいっぽうにしか出現しない漢
字をリストすると表5－9のようになる。両編に共通して出現しない
漢字は49種あり、医学編のみが27種、一般編のみが22種となっ
ている。医学編のみの27種の漢字のうち医学分野に特有と思われる
字種（表5－9のうち○、●印のもの）が17種あげられる。またこ
れらの漢字のうち、身体の部位を表わす漢字（解剖学用字）として、
腎・肝・胆・腫・指・膚・頭
椎・喉・随・鼻
などがある。

いっぽう、医学編および一般編の論文標題の漢字字種の出現頻度リ
スト（表5－10）の上位をそれぞれ見るとつぎののようなことが理解
できる。

○科学技術論文のタイトルに特有な字種が両編に限らず多くみられ
る。たとえば性、究、研、的、関、法、学、などの漢字は、医学
分野に限らず、あらゆる分野で論文タイトルの漢字として用いら
れている。

（例）～に関する研究

～法について

～学的研究

—122一
これらの学種は一般編のリスト中でも上位を占めている。また、医学薬物タイトルに特有な漢字としてつきのようなものがあげられる。

症・例・療・験・用・治・経

（例） ～症の治験例
～症の治療の経験

医学編に出現する漢字の頻度リスト（500位まで）は表5-11に示すとおりである。ここに、当用漢字、人名漢字、補正漢字とそれ以外の漢字が区別されている。当用・人名・補正漢字以外の漢字に入るものは

1位から500位まで 41種（8.2%）
501位から1039位まで135種（約13%）

で、頻度数の低い漢字に圧倒的に非当用漢字が多く、とくに頻度
1の漢字（順位745位以降）については
295種のうち81種（約28%）

が、非当用漢字である。

図5-5 頻度順位と全体に対する割合の関係図
表5-7 類度順位と全体に対する割合の比較表

<table>
<thead>
<tr>
<th>上位</th>
<th>自然科学一般編</th>
<th>医学編</th>
<th>自然科学一般編</th>
<th>医学編</th>
</tr>
</thead>
<tbody>
<tr>
<td>からの文字</td>
<td>文字数</td>
<td>全体に対する割合</td>
<td>文字数</td>
<td>全体に対する割合</td>
</tr>
<tr>
<td>1～100</td>
<td>4,177</td>
<td>4.117</td>
<td>6,176</td>
<td>5.439</td>
</tr>
<tr>
<td>1～200</td>
<td>5,722</td>
<td>6.599</td>
<td>8,222</td>
<td>7.122</td>
</tr>
<tr>
<td>1～300</td>
<td>6,643</td>
<td>7.561</td>
<td>9,388</td>
<td>8.132</td>
</tr>
<tr>
<td>1～400</td>
<td>7,268</td>
<td>8.331</td>
<td>10,139</td>
<td>8.782</td>
</tr>
<tr>
<td>1～500</td>
<td>7,706</td>
<td>8.887</td>
<td>10,614</td>
<td>9.194</td>
</tr>
<tr>
<td>1～600</td>
<td>8,021</td>
<td>9.250</td>
<td>10,445</td>
<td>9.481</td>
</tr>
<tr>
<td>1～700</td>
<td>8,255</td>
<td>9.485</td>
<td>11,161</td>
<td>9.668</td>
</tr>
<tr>
<td>1～800</td>
<td>8,365</td>
<td>9.647</td>
<td>11,305</td>
<td>9.792</td>
</tr>
<tr>
<td>1～900</td>
<td>8,465</td>
<td>9.762</td>
<td>11,405</td>
<td>9.879</td>
</tr>
<tr>
<td>1～1,000</td>
<td>8,565</td>
<td>9.877</td>
<td>11,505</td>
<td>9.966</td>
</tr>
<tr>
<td>1～1,100</td>
<td>8,665</td>
<td>9.993</td>
<td>11,544</td>
<td>10.000</td>
</tr>
<tr>
<td>1～1,200</td>
<td>8,671</td>
<td>10.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1～1,300</td>
<td>-</td>
<td>-</td>
<td>20,097</td>
<td>9.932</td>
</tr>
<tr>
<td>1～1,400</td>
<td>-</td>
<td>-</td>
<td>20,197</td>
<td>9.981</td>
</tr>
<tr>
<td>1～1,500</td>
<td>-</td>
<td>-</td>
<td>20,234</td>
<td>10.000</td>
</tr>
</tbody>
</table>

表5-8 雑誌記事索引一自然科学編収録論文標題の漢字分布

<table>
<thead>
<tr>
<th>項目</th>
<th>自然科学一般編</th>
<th>医学編</th>
<th>自然科学一般編</th>
<th>医学編</th>
</tr>
</thead>
<tbody>
<tr>
<td>件数（論文数）</td>
<td>1,032</td>
<td>1,095</td>
<td>2,127</td>
<td></td>
</tr>
<tr>
<td>総文字数</td>
<td>8,671</td>
<td>11,544</td>
<td>20,234</td>
<td></td>
</tr>
<tr>
<td>一論文あたりの平均漢文字数</td>
<td>8.4</td>
<td>10.5</td>
<td>9.5</td>
<td></td>
</tr>
</tbody>
</table>

漢字の種類

<table>
<thead>
<tr>
<th>内容</th>
<th>当用漢字</th>
<th>全当用漢字に対する割合</th>
<th>教育漢字</th>
<th>人名漢字</th>
<th>綴正漢字</th>
</tr>
</thead>
<tbody>
<tr>
<td>988 (89.3%)</td>
<td>863 (83.0%)</td>
<td>1,176 (81.5%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53.7%</td>
<td>5.6%</td>
<td>6.8%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>647</td>
<td>271</td>
<td>428</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>328</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>当用漢字外漢字</td>
<td>118 (10.2%)</td>
<td>176 (17.0%)</td>
<td>251 (18.5%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

-124-
### 表 5-9 一般編と医学編に出現する漢字の比較

<table>
<thead>
<tr>
<th>番号</th>
<th>字</th>
<th>番号</th>
<th>字</th>
</tr>
</thead>
<tbody>
<tr>
<td>74</td>
<td>鼻</td>
<td>63</td>
<td>28</td>
</tr>
<tr>
<td>104</td>
<td>咽</td>
<td>48</td>
<td>51</td>
</tr>
<tr>
<td>118</td>
<td>鼻</td>
<td>45</td>
<td>58</td>
</tr>
<tr>
<td>158</td>
<td>腸</td>
<td>33</td>
<td>90</td>
</tr>
<tr>
<td>202</td>
<td>心</td>
<td>26</td>
<td>78</td>
</tr>
<tr>
<td>243</td>
<td>腸</td>
<td>22</td>
<td>58</td>
</tr>
<tr>
<td>249</td>
<td>腸</td>
<td>21</td>
<td>102</td>
</tr>
<tr>
<td>256</td>
<td>鼻</td>
<td>21</td>
<td>143</td>
</tr>
<tr>
<td>261</td>
<td>腸</td>
<td>20</td>
<td>105</td>
</tr>
<tr>
<td>280</td>
<td>鼻</td>
<td>19</td>
<td>159</td>
</tr>
<tr>
<td>294</td>
<td>腸</td>
<td>18</td>
<td>166</td>
</tr>
<tr>
<td>297</td>
<td>腸</td>
<td>18</td>
<td>169</td>
</tr>
<tr>
<td>310</td>
<td>器</td>
<td>17</td>
<td>129</td>
</tr>
<tr>
<td>342</td>
<td>材</td>
<td>15</td>
<td>152</td>
</tr>
<tr>
<td>348</td>
<td>鼻</td>
<td>15</td>
<td>200</td>
</tr>
<tr>
<td>352</td>
<td>鼻</td>
<td>15</td>
<td>203</td>
</tr>
<tr>
<td>356</td>
<td>咽</td>
<td>15</td>
<td>204</td>
</tr>
<tr>
<td>361</td>
<td>咽</td>
<td>14</td>
<td>206</td>
</tr>
<tr>
<td>367</td>
<td>咽</td>
<td>14</td>
<td>211</td>
</tr>
<tr>
<td>368</td>
<td>腸</td>
<td>14</td>
<td>212</td>
</tr>
<tr>
<td>374</td>
<td>腸</td>
<td>14</td>
<td>218</td>
</tr>
<tr>
<td>379</td>
<td>鼻</td>
<td>13</td>
<td>225</td>
</tr>
<tr>
<td>386</td>
<td>鼻</td>
<td>13</td>
<td>173</td>
</tr>
<tr>
<td>384</td>
<td>鼻</td>
<td>13</td>
<td>177</td>
</tr>
<tr>
<td>385</td>
<td>鼻</td>
<td>13</td>
<td>176</td>
</tr>
<tr>
<td>398</td>
<td>鼻</td>
<td>13</td>
<td>232</td>
</tr>
<tr>
<td>396</td>
<td>鼻</td>
<td>13</td>
<td>232</td>
</tr>
<tr>
<td>398</td>
<td>鼻</td>
<td>13</td>
<td>232</td>
</tr>
<tr>
<td>404</td>
<td>腸</td>
<td>12</td>
<td>194</td>
</tr>
<tr>
<td>408</td>
<td>腸</td>
<td>12</td>
<td>196</td>
</tr>
<tr>
<td>416</td>
<td>鼻</td>
<td>12</td>
<td>244</td>
</tr>
<tr>
<td>411</td>
<td>鼻</td>
<td>12</td>
<td>197</td>
</tr>
<tr>
<td>428</td>
<td>鼻</td>
<td>11</td>
<td>210</td>
</tr>
<tr>
<td>437</td>
<td>椎</td>
<td>11</td>
<td>260</td>
</tr>
<tr>
<td>442</td>
<td>鼻</td>
<td>11</td>
<td>263</td>
</tr>
<tr>
<td>447</td>
<td>鼻</td>
<td>11</td>
<td>219</td>
</tr>
<tr>
<td>451</td>
<td>鼻</td>
<td>10</td>
<td>222</td>
</tr>
<tr>
<td>454</td>
<td>椎</td>
<td>10</td>
<td>271</td>
</tr>
<tr>
<td>461</td>
<td>椎</td>
<td>10</td>
<td>225</td>
</tr>
<tr>
<td>466</td>
<td>椎</td>
<td>10</td>
<td>283</td>
</tr>
<tr>
<td>471</td>
<td>鼻</td>
<td>10</td>
<td>287</td>
</tr>
<tr>
<td>476</td>
<td>椎</td>
<td>10</td>
<td>288</td>
</tr>
<tr>
<td>478</td>
<td>鼻</td>
<td>10</td>
<td>289</td>
</tr>
<tr>
<td>480</td>
<td>安</td>
<td>9</td>
<td>238</td>
</tr>
<tr>
<td>493</td>
<td>鼻</td>
<td>9</td>
<td>251</td>
</tr>
<tr>
<td>494</td>
<td>鼻</td>
<td>9</td>
<td>335</td>
</tr>
<tr>
<td>497</td>
<td>鼻</td>
<td>9</td>
<td>311</td>
</tr>
<tr>
<td>499</td>
<td>鼻</td>
<td>9</td>
<td>257</td>
</tr>
<tr>
<td>500</td>
<td>鼻</td>
<td>9</td>
<td>258</td>
</tr>
</tbody>
</table>
表 5-10 医学編および一般編での頻度最上位の漢字

<table>
<thead>
<tr>
<th></th>
<th>一般編での順位</th>
<th>医学編での順位</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>性</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>症</td>
<td>用</td>
</tr>
<tr>
<td>3</td>
<td>血</td>
<td>研</td>
</tr>
<tr>
<td>4</td>
<td>究</td>
<td>究</td>
</tr>
<tr>
<td>5</td>
<td>研</td>
<td>性</td>
</tr>
<tr>
<td>6</td>
<td>例</td>
<td>電</td>
</tr>
<tr>
<td>7</td>
<td>的</td>
<td>化</td>
</tr>
<tr>
<td>8</td>
<td>関</td>
<td>法</td>
</tr>
<tr>
<td>9</td>
<td>療</td>
<td>電</td>
</tr>
<tr>
<td>10</td>
<td>騙</td>
<td>18</td>
</tr>
<tr>
<td>11</td>
<td>用</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>法</td>
<td>水</td>
</tr>
<tr>
<td>13</td>
<td>治</td>
<td>等</td>
</tr>
<tr>
<td>14</td>
<td>経</td>
<td>物</td>
</tr>
<tr>
<td>15</td>
<td>生</td>
<td>業</td>
</tr>
<tr>
<td>16</td>
<td>病</td>
<td>分</td>
</tr>
<tr>
<td>17</td>
<td>学</td>
<td>25</td>
</tr>
<tr>
<td>18</td>
<td>腫</td>
<td>騙</td>
</tr>
<tr>
<td>19</td>
<td>児</td>
<td>子</td>
</tr>
<tr>
<td>20</td>
<td>科</td>
<td>皮</td>
</tr>
<tr>
<td>21</td>
<td>及</td>
<td>合</td>
</tr>
<tr>
<td>22</td>
<td>対</td>
<td>発</td>
</tr>
<tr>
<td>23</td>
<td>検</td>
<td>気</td>
</tr>
<tr>
<td>24</td>
<td>発</td>
<td>動</td>
</tr>
<tr>
<td>25</td>
<td>化</td>
<td>7</td>
</tr>
</tbody>
</table>
### 表 5-11 雑誌記事案引医学論各漢字と頻度

<p>| 文字 | 頻数 | 全文比 | | 文字 | 頻数 | 全文比 |
|------|------|--------||------|------|--------|
| 5.9 | 丸 | 4.3 | 0.37 | 9.5 | 木 | 3.0 | 0.26 |
| 6.0 | 條 | 4.3 | 0.37 | 9.6 | 電 | 3.0 | 0.26 |
| 6.1 | 高 | 4.2 | 0.36 | 9.7 | 電 | 3.0 | 0.26 |
| 6.2 | 資 | 4.2 | 0.36 | 9.8 | 原 | 3.0 | 0.26 |
| 6.3 | 電 | 4.2 | 0.36 | 9.9 | 電 | 3.0 | 0.26 |
| 6.4 | 中 | 4.2 | 0.36 | 1.0 | 電 | 2.9 | 0.25 |
| 6.5 | 笑 | 4.2 | 0.36 | 1.1 | 一 | 0.1 | 0.24 |
| 6.6 | 電 | 4.2 | 0.36 | 1.2 | 電 | 2.8 | 0.22 |
| 6.7 | 電 | 4.2 | 0.36 | 1.3 | 電 | 2.8 | 0.22 |
| 6.8 | 電 | 4.2 | 0.36 | 1.4 | 電 | 2.8 | 0.22 |
| 6.9 | 電 | 4.1 | 0.36 | 1.5 | 電 | 2.8 | 0.22 |
| 7.0 | 電 | 4.0 | 0.35 | 1.6 | 電 | 2.8 | 0.22 |
| 7.1 | 電 | 4.0 | 0.35 | 1.7 | 電 | 2.8 | 0.22 |
| 7.2 | 電 | 4.0 | 0.35 | 1.8 | 電 | 2.8 | 0.22 |
| 7.3 | 電 | 3.9 | 0.34 | 1.9 | 電 | 2.7 | 0.22 |
| 7.4 | 電 | 3.9 | 0.34 | 2.0 | 電 | 2.7 | 0.22 |
| 7.5 | 電 | 3.9 | 0.34 | 2.1 | 電 | 2.7 | 0.22 |
| 7.6 | 電 | 3.8 | 0.33 | 2.2 | 電 | 2.7 | 0.22 |
| 7.7 | 電 | 3.8 | 0.33 | 2.3 | 電 | 2.7 | 0.22 |
| 7.8 | 電 | 3.8 | 0.33 | 2.4 | 電 | 2.7 | 0.22 |
| 7.9 | 電 | 3.7 | 0.32 | 2.5 | 電 | 2.7 | 0.22 |
| 8.0 | 電 | 3.7 | 0.32 | 2.6 | 電 | 2.7 | 0.22 |
| 8.1 | 電 | 3.7 | 0.32 | 2.7 | 電 | 2.7 | 0.22 |
| 8.2 | 電 | 3.7 | 0.32 | 2.8 | 電 | 2.7 | 0.22 |
| 8.3 | 電 | 3.7 | 0.32 | 2.9 | 電 | 2.7 | 0.22 |
| 8.4 | 電 | 3.6 | 0.31 | 3.0 | 電 | 2.7 | 0.22 |
| 8.5 | 電 | 3.6 | 0.31 | 3.1 | 電 | 2.7 | 0.22 |
| 8.6 | 電 | 3.6 | 0.31 | 3.2 | 電 | 2.7 | 0.22 |
| 8.7 | 電 | 3.6 | 0.31 | 3.3 | 電 | 2.7 | 0.22 |
| 8.8 | 電 | 3.5 | 0.30 | 3.4 | 電 | 2.7 | 0.22 |
| 8.9 | 電 | 3.4 | 0.29 | 3.5 | 電 | 2.7 | 0.22 |
| 8.10 | 電 | 3.3 | 0.29 | 3.6 | 電 | 2.7 | 0.22 |
| 8.11 | 電 | 3.3 | 0.29 | 3.7 | 電 | 2.7 | 0.22 |
| 8.12 | 電 | 3.3 | 0.29 | 3.8 | 電 | 2.7 | 0.22 |
| 8.13 | 電 | 3.3 | 0.29 | 3.9 | 電 | 2.7 | 0.22 |
| 8.14 | 電 | 3.3 | 0.29 | 4.0 | 電 | 2.7 | 0.22 |</p>
<table>
<thead>
<tr>
<th>序号</th>
<th>文字</th>
<th>度数</th>
<th>全数比例%</th>
<th>序号</th>
<th>文字</th>
<th>度数</th>
<th>全数比例%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>代</td>
<td>22</td>
<td>0.19</td>
<td>167</td>
<td>孔</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>2</td>
<td>移</td>
<td>22</td>
<td>0.19</td>
<td>168</td>
<td>伴</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>3</td>
<td>田</td>
<td>21</td>
<td>0.18</td>
<td>169</td>
<td>人</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>4</td>
<td>疑</td>
<td>21</td>
<td>0.18</td>
<td>170</td>
<td>本</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>5</td>
<td>行</td>
<td>21</td>
<td>0.18</td>
<td>171</td>
<td>与</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>6</td>
<td>专</td>
<td>21</td>
<td>0.18</td>
<td>172</td>
<td>力</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>7</td>
<td>米</td>
<td>21</td>
<td>0.18</td>
<td>173</td>
<td>兆</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>8</td>
<td>修</td>
<td>21</td>
<td>0.18</td>
<td>174</td>
<td>勾</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>9</td>
<td>造</td>
<td>21</td>
<td>0.18</td>
<td>175</td>
<td>载</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>10</td>
<td>录</td>
<td>21</td>
<td>0.18</td>
<td>176</td>
<td>海</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>11</td>
<td>金</td>
<td>21</td>
<td>0.18</td>
<td>177</td>
<td>上</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>12</td>
<td>畜</td>
<td>21</td>
<td>0.18</td>
<td>178</td>
<td>周</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>13</td>
<td>道</td>
<td>21</td>
<td>0.18</td>
<td>179</td>
<td>伸</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>14</td>
<td>派</td>
<td>21</td>
<td>0.18</td>
<td>180</td>
<td>蜜</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>15</td>
<td>给</td>
<td>21</td>
<td>0.18</td>
<td>181</td>
<td>朝</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>16</td>
<td>援</td>
<td>21</td>
<td>0.18</td>
<td>182</td>
<td>堆</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>17</td>
<td>邮</td>
<td>21</td>
<td>0.18</td>
<td>183</td>
<td>封</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>18</td>
<td>年</td>
<td>20</td>
<td>0.17</td>
<td>184</td>
<td>予</td>
<td>17</td>
<td>0.15</td>
</tr>
<tr>
<td>19</td>
<td>适</td>
<td>19</td>
<td>0.16</td>
<td>185</td>
<td>独</td>
<td>17</td>
<td>0.15</td>
</tr>
<tr>
<td>20</td>
<td>基</td>
<td>19</td>
<td>0.16</td>
<td>186</td>
<td>通</td>
<td>17</td>
<td>0.15</td>
</tr>
<tr>
<td>21</td>
<td>圆</td>
<td>19</td>
<td>0.16</td>
<td>187</td>
<td>优</td>
<td>17</td>
<td>0.15</td>
</tr>
<tr>
<td>22</td>
<td>厨</td>
<td>19</td>
<td>0.16</td>
<td>188</td>
<td>阿</td>
<td>16</td>
<td>0.14</td>
</tr>
<tr>
<td>23</td>
<td>流</td>
<td>19</td>
<td>0.16</td>
<td>189</td>
<td>蝴</td>
<td>16</td>
<td>0.14</td>
</tr>
<tr>
<td>24</td>
<td>画</td>
<td>19</td>
<td>0.16</td>
<td>190</td>
<td>虫</td>
<td>16</td>
<td>0.14</td>
</tr>
<tr>
<td>25</td>
<td>水</td>
<td>19</td>
<td>0.16</td>
<td>191</td>
<td>电</td>
<td>15</td>
<td>0.13</td>
</tr>
<tr>
<td>26</td>
<td>洗</td>
<td>19</td>
<td>0.16</td>
<td>192</td>
<td>电</td>
<td>15</td>
<td>0.13</td>
</tr>
<tr>
<td>27</td>
<td>母</td>
<td>19</td>
<td>0.16</td>
<td>193</td>
<td>电</td>
<td>15</td>
<td>0.13</td>
</tr>
<tr>
<td>28</td>
<td>橙</td>
<td>19</td>
<td>0.16</td>
<td>194</td>
<td>电</td>
<td>15</td>
<td>0.13</td>
</tr>
<tr>
<td>29</td>
<td>梁</td>
<td>19</td>
<td>0.16</td>
<td>195</td>
<td>电</td>
<td>15</td>
<td>0.13</td>
</tr>
<tr>
<td>30</td>
<td>阳</td>
<td>19</td>
<td>0.16</td>
<td>196</td>
<td>电</td>
<td>15</td>
<td>0.13</td>
</tr>
<tr>
<td>31</td>
<td>安</td>
<td>19</td>
<td>0.16</td>
<td>197</td>
<td>电</td>
<td>15</td>
<td>0.13</td>
</tr>
<tr>
<td>32</td>
<td>沙</td>
<td>18</td>
<td>0.15</td>
<td>198</td>
<td>电</td>
<td>15</td>
<td>0.13</td>
</tr>
<tr>
<td>33</td>
<td>田</td>
<td>18</td>
<td>0.15</td>
<td>199</td>
<td>电</td>
<td>15</td>
<td>0.13</td>
</tr>
<tr>
<td>34</td>
<td>家</td>
<td>18</td>
<td>0.15</td>
<td>200</td>
<td>电</td>
<td>15</td>
<td>0.13</td>
</tr>
<tr>
<td>35</td>
<td>电</td>
<td>18</td>
<td>0.15</td>
<td>201</td>
<td>电</td>
<td>15</td>
<td>0.13</td>
</tr>
<tr>
<td>36</td>
<td>均</td>
<td>0</td>
<td>0</td>
<td>202</td>
<td>电</td>
<td>15</td>
<td>0.13</td>
</tr>
<tr>
<td>序数</td>
<td>文字</td>
<td>用数</td>
<td>余数</td>
<td>序数</td>
<td>文字</td>
<td>用数</td>
<td>余数</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>275</td>
<td>金</td>
<td>10</td>
<td>009</td>
<td>311</td>
<td>業</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>276</td>
<td>茶</td>
<td>10</td>
<td>009</td>
<td>312</td>
<td>僧</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>277</td>
<td>课</td>
<td>10</td>
<td>009</td>
<td>313</td>
<td>甲</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>278</td>
<td>矫</td>
<td>10</td>
<td>009</td>
<td>314</td>
<td>演</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>279</td>
<td>奧</td>
<td>10</td>
<td>009</td>
<td>315</td>
<td>維</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>280</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>316</td>
<td>涼</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>281</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>317</td>
<td>達</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>282</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>318</td>
<td>潤</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>283</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>319</td>
<td>源</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>284</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>320</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>285</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>321</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>286</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>322</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>287</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>323</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>288</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>324</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>289</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>325</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>290</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>326</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>291</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>327</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>292</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>328</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>293</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>329</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>294</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>330</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>295</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>331</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>296</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>332</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>297</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>333</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>298</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>334</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>299</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>335</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>300</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>336</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>301</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>337</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>302</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>338</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>303</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>339</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>304</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>340</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>305</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>341</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>306</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>342</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>307</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>343</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>308</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>344</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>309</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>345</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
<tr>
<td>310</td>
<td>創</td>
<td>10</td>
<td>009</td>
<td>346</td>
<td>創</td>
<td>9</td>
<td>008</td>
</tr>
</tbody>
</table>
5.2.3 病名に用いられる漢字（医学辞典における調査）

医学では、従来から人、生体の正常な状態についてよりも、異常な状態（病気・疾病）について論じられることが、圧倒的に多かった。この病気のグループや分類には病気の名称、すなわち病名（disease name）が採用されてきた。もちろん医学用語、解剖学用語、手術式名、検査用語など多くのカテゴリーの用語を含まれるが、この病名は、その中でも最も医学用語の特色、特質を持ったものであろう。

前述したように、日本の医学用語は、いまだ標準化あるいは統一化されたものがないが、各種の医学辞典、用語集が出版されている。ここではわが国の代表的な医学辞典である「医学大辞典」（南山堂；改訂10版1964年）の見出し語について検討する。*

(1) 調査方法

この医学大辞典は見出し語に日本語を用い、これらの用語に解説を加えたものである。見出し語としては、医学はもちろんのこと、医学薬学に関連深い分野の用語を収録している。

調査方法は、辞典の見出し語に出現する“病名”を抽出したがその総病名数は4749件であった。なお、これらの見出し語には、漢字ばかりではなく、カタカナ、ひらがなも含まれている。

(2) 調査結果および考察

医学大辞典の見出し語にみられた病名に関する文字種の出現頻度（第30位まで）は、表5－12のようである。なお、総数4749件の病名のなかには、同一の病名で名称が2個以上あるものも含まれている。文字の種類、すなわち異なり字数は1,195字であった。
この数字は、前項の「雑誌記事索引」での調査結果、1,039字とそれほど変わらない。

* 参考資料：秋山房雄「成人保健学」Ⅲ、保健学的接近と言語情報 P112－P137、南山堂1973
いっぱい、延べ字数は22,886字で、異なり字数1,195種に対し、それぞれ平均1.92回出現したことになる。
出現頻度の高い文字を見ると、最も高いのは「症」で1,355回、ついて「性」「炎」「臓」の順である。このリストには第5位の「ン」以下「ル」「〜」（長音）などのカタカナが、第30位までに9種ある。これは病名に人名あるいは外来語が採用されるケースが多いためであろう。
当用漢字以外の漢字（1位〜50位までに）は
腫・癒・膿・癌
があった。
接尾語と考えられる字種としては、
症・性・炎・腫・病・群・化
などが見つけるられる。
出現頻度の分布曲線は図5－6に示すようにポアソン分布があてはまる。
つきに「雑誌記事索引」の調査と比較してみると（表5－10参照）両調査結果の第30位までに入っている字種としては、
症（辞典第2位、索引第2位）、性（2,1）、腫（4,18）
病（7,16）、血（9,3）、経（21,14）、
化（28,25）
がある。
この辞典の調査は病名に使われている漢字であったが、「雑誌記事索引」の調査では論文タイトルに使われている漢字であるから、その字種および頻度順位に異なりがあるのが当然である。しかしながら、病名を表す接尾語（字）である“症”は共に1位、2位を占めておりまた医学用語で単位語を合成する際に用いられる“性”が2位、1位となっているのは特徴的である。

－132－
表5-12 医学大辞典の表名見出し語にみられた文字リスト

<table>
<thead>
<tr>
<th>順位</th>
<th>文字</th>
<th>頻度</th>
<th>%</th>
<th>順位</th>
<th>文字</th>
<th>頻度</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>症</td>
<td>1,335</td>
<td>5.83</td>
<td>28</td>
<td>化</td>
<td>1,06</td>
<td>0.46</td>
</tr>
<tr>
<td>2</td>
<td>息</td>
<td>1,033</td>
<td>4.51</td>
<td>29</td>
<td>漣</td>
<td>1,05</td>
<td>0.46</td>
</tr>
<tr>
<td>3</td>
<td>炎</td>
<td>551</td>
<td>2.41</td>
<td>30</td>
<td>腹</td>
<td>1,04</td>
<td>0.45</td>
</tr>
<tr>
<td>4</td>
<td>腫</td>
<td>505</td>
<td>2.21</td>
<td>31</td>
<td>硬</td>
<td>1,02</td>
<td>0.45</td>
</tr>
<tr>
<td>5</td>
<td>サ</td>
<td>377</td>
<td>1.65</td>
<td>32</td>
<td>肺</td>
<td>1,01</td>
<td>0.44</td>
</tr>
<tr>
<td>6</td>
<td>ル</td>
<td>369</td>
<td>1.61</td>
<td>33</td>
<td>フ</td>
<td>1,01</td>
<td>0.44</td>
</tr>
<tr>
<td>7</td>
<td>病</td>
<td>354</td>
<td>1.55</td>
<td>34</td>
<td>細</td>
<td>1,00</td>
<td>0.43</td>
</tr>
<tr>
<td>8</td>
<td>一</td>
<td>326</td>
<td>1.42</td>
<td>35</td>
<td>状</td>
<td>98</td>
<td>0.43</td>
</tr>
<tr>
<td>9</td>
<td>血</td>
<td>312</td>
<td>1.36</td>
<td>36</td>
<td>ツ</td>
<td>95</td>
<td>0.42</td>
</tr>
<tr>
<td>10</td>
<td>候</td>
<td>240</td>
<td>1.05</td>
<td>37</td>
<td>脈</td>
<td>95</td>
<td>0.41</td>
</tr>
<tr>
<td>11</td>
<td>群</td>
<td>234</td>
<td>1.02</td>
<td>38</td>
<td>ウ</td>
<td>94</td>
<td>0.41</td>
</tr>
<tr>
<td>12</td>
<td>リ</td>
<td>233</td>
<td>1.02</td>
<td>39</td>
<td>心</td>
<td>93</td>
<td>0.41</td>
</tr>
<tr>
<td>13</td>
<td>膜</td>
<td>222</td>
<td>0.97</td>
<td>40</td>
<td>発</td>
<td>92</td>
<td>0.40</td>
</tr>
<tr>
<td>14</td>
<td>ア</td>
<td>201</td>
<td>0.88</td>
<td>41</td>
<td>内</td>
<td>91</td>
<td>0.40</td>
</tr>
<tr>
<td>15</td>
<td>ス</td>
<td>188</td>
<td>0.82</td>
<td>42</td>
<td>尿</td>
<td>88</td>
<td>0.38</td>
</tr>
<tr>
<td>16</td>
<td>骨</td>
<td>179</td>
<td>0.78</td>
<td>43</td>
<td>袋</td>
<td>87</td>
<td>0.38</td>
</tr>
<tr>
<td>17</td>
<td>神</td>
<td>172</td>
<td>0.75</td>
<td>44</td>
<td>筋</td>
<td>86</td>
<td>0.38</td>
</tr>
<tr>
<td>18</td>
<td>皮</td>
<td>158</td>
<td>0.69</td>
<td>45</td>
<td>色</td>
<td>86</td>
<td>0.38</td>
</tr>
<tr>
<td>19</td>
<td>炎</td>
<td>152</td>
<td>0.66</td>
<td>46</td>
<td>疾</td>
<td>86</td>
<td>0.38</td>
</tr>
<tr>
<td>20</td>
<td>イ</td>
<td>149</td>
<td>0.65</td>
<td>47</td>
<td>筋</td>
<td>86</td>
<td>0.38</td>
</tr>
<tr>
<td>21</td>
<td>經</td>
<td>143</td>
<td>0.62</td>
<td>48</td>
<td>疫</td>
<td>85</td>
<td>0.37</td>
</tr>
<tr>
<td>22</td>
<td>随</td>
<td>136</td>
<td>0.59</td>
<td>49</td>
<td>チ</td>
<td>84</td>
<td>0.37</td>
</tr>
<tr>
<td>23</td>
<td>毒</td>
<td>129</td>
<td>0.56</td>
<td>50</td>
<td>肉</td>
<td>81</td>
<td>0.35</td>
</tr>
<tr>
<td>24</td>
<td>結</td>
<td>124</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>中</td>
<td>111</td>
<td>0.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>工</td>
<td>110</td>
<td>0.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>ク</td>
<td>106</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 長音
× 当用漢字以外の漢字

－133－
病名のうちで、身体の部位（解剖用語）を示す字種として、たとえば、

膜、骨、神（経）、皮、（神）経、髄、胞、肺、脈、
心、筋
など 11 種があげられる。
また、病名を表わす接尾語（字）としては

症、炎、腫、病、役、群、傷、毒、濁
の 9 種がある。

図 5-6 病名の出現頻度別度数分布（ポアソン分布）
5.2.4 病理学領域における医学用語（単位語による分析）

ここでは医学分野の1つの領域である「病理学」に関する論文標題を対象に調査を行った。論文標題に用いられている医学用語を単位語に分け、これらの種類と出現頻度について検討した結果を述べる。

(1) 調査対象

病理学における論文標題として、わが国最大の医学抄録誌である「医学中央雑誌」に掲載されている論文を引用した。「医学中央雑誌」は医学分野の抄録誌として1908年より刊行され、国内で発表される医学および関連領域の原著論文、学会発表を網羅的に収録している。

調査対象としては、この雑誌の第172〜182巻（昭和37年度）の分類件名「病理解剖学」に収録されている論文、1,837件のうち、任意に1,000件の標題をサンプルとして選定した。

(2) 用語の抽出

用語抽出は、論文標題中の、助詞、助動詞、接頭語、などの非重要語を除外して、2字以上に結合した状態で行なった。たとえば、

○細胞分裂の調節機構に関する研究
○腫瘍移植性に対する純系動物の病理学的検討
○Ehrlieh 腹水瘤のマウス皮下移植結節から抽出した Oncotrephin と Antitrephin

の標題では、下線部の文字や語を除いた語が、すべて調査対象となっている。

※ 杉藤 孝、医学用語の統計的調査（1）病理学領域における論文標題を例として、きたさと5(2)35-95（'66）
杉藤 孝 索引作業のための自然語処理の研究 —— 医学用語の計量的調査、Library Science 5, 51-72（'67）
この例からもわかるように、医学用語の特徴としては、数語が結合して用語を形成しているものが多いことである。このように2字以上の文字が結合したものを合成語とした。

用語を、最小の意味単位、あるいは認定単位に区別するためにつぎのような規則を設けることにする。

○合成語：つぎに述べる単位語と単位語、または単位語と特殊単位語の結合によっての用語。
たとえば、“甲狀腺腫”、“十二指腸虫”、“病理学的研究”など。

○単位語：漢字2字の集合で、1つの特定の概念を表わすことができる語。
例外として“所見、研究、症例、検討、展覧”などの慣用的付加する用語を常用語という。

○特殊単位語：
漢字1字でも用語として認められる語。
たとえば、“胃”、“肝”など。

また、補助語、特殊単位語のうち接尾語の機能を果す漢字1字の語。例えば、“～的、～様、～型、～状、～体、～病、学”など。

用語の抽出は上記の基準にしたがって行なわれるが、合成語を単位語にいかに分割するかは、非常に難しい。一応、2字を単位語の原則として、3字以上の合成語は、できるだけ2つ以上の単位語に分けるようにする。

(3) 調査 結果

論文の標題1,000件について、単位語および特殊単位語の出現頻度を調査した結果を表5-13に示す。

単位語延べ語数 8,037語
表 5-13 単位語の生起頻度

<table>
<thead>
<tr>
<th>順位</th>
<th>単位語</th>
<th>頻度</th>
<th>%</th>
<th>順位</th>
<th>単位語</th>
<th>頻度</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>*～</td>
<td>368</td>
<td>4.41</td>
<td>33</td>
<td>結核</td>
<td>41</td>
<td>0.49</td>
</tr>
<tr>
<td>2</td>
<td>症</td>
<td>286</td>
<td>3.43</td>
<td>34</td>
<td>*～</td>
<td>41</td>
<td>0.49</td>
</tr>
<tr>
<td>3</td>
<td>*～</td>
<td>248</td>
<td>2.98</td>
<td>35</td>
<td>移植</td>
<td>38</td>
<td>0.46</td>
</tr>
<tr>
<td>4</td>
<td>*～</td>
<td>245</td>
<td>2.94</td>
<td>36</td>
<td>*～</td>
<td>38</td>
<td>0.46</td>
</tr>
<tr>
<td>5</td>
<td>研究</td>
<td>241</td>
<td>2.89</td>
<td>37</td>
<td>甲狀</td>
<td>37</td>
<td>0.44</td>
</tr>
<tr>
<td>6</td>
<td>組織</td>
<td>232</td>
<td>2.78</td>
<td>38</td>
<td>ラッテ</td>
<td>37</td>
<td>0.44</td>
</tr>
<tr>
<td>7</td>
<td>細胞</td>
<td>210</td>
<td>2.52</td>
<td>39</td>
<td>*～</td>
<td>36</td>
<td>0.43</td>
</tr>
<tr>
<td>8</td>
<td>病理</td>
<td>182</td>
<td>2.18</td>
<td>40</td>
<td>転移</td>
<td>34</td>
<td>0.41</td>
</tr>
<tr>
<td>9</td>
<td>腫瘍</td>
<td>144</td>
<td>1.73</td>
<td>41</td>
<td>反応</td>
<td>33</td>
<td>0.40</td>
</tr>
<tr>
<td>10</td>
<td>肝(鱗)</td>
<td>134</td>
<td>1.61</td>
<td>42</td>
<td>障害</td>
<td>32</td>
<td>0.38</td>
</tr>
<tr>
<td>11</td>
<td>*～</td>
<td>125</td>
<td>1.50</td>
<td>43</td>
<td>白血</td>
<td>32</td>
<td>0.38</td>
</tr>
<tr>
<td>12</td>
<td>肺(鱗)</td>
<td>92</td>
<td>1.10</td>
<td>44</td>
<td>染色</td>
<td>32</td>
<td>0.38</td>
</tr>
<tr>
<td>13</td>
<td>化学</td>
<td>71</td>
<td>0.85</td>
<td>45</td>
<td>臓器</td>
<td>30</td>
<td>0.36</td>
</tr>
<tr>
<td>14</td>
<td>マウス</td>
<td>69</td>
<td>0.83</td>
<td>46</td>
<td>Virus</td>
<td>30</td>
<td>0.36</td>
</tr>
<tr>
<td>15</td>
<td>銀微鏡</td>
<td>65</td>
<td>0.78</td>
<td>47</td>
<td>淋巴</td>
<td>29</td>
<td>0.35</td>
</tr>
<tr>
<td>16</td>
<td>腦</td>
<td>65</td>
<td>0.78</td>
<td>48</td>
<td>硬変</td>
<td>27</td>
<td>0.32</td>
</tr>
<tr>
<td>17</td>
<td>電子</td>
<td>65</td>
<td>0.78</td>
<td>49</td>
<td>動物</td>
<td>26</td>
<td>0.31</td>
</tr>
<tr>
<td>18</td>
<td>*～</td>
<td>60</td>
<td>0.72</td>
<td>50</td>
<td>観察</td>
<td>25</td>
<td>0.30</td>
</tr>
<tr>
<td>19</td>
<td>実験</td>
<td>59</td>
<td>0.71</td>
<td>51</td>
<td>増殖</td>
<td>24</td>
<td>0.29</td>
</tr>
<tr>
<td>20</td>
<td>肉腫</td>
<td>58</td>
<td>0.70</td>
<td>52</td>
<td>酢素</td>
<td>24</td>
<td>0.29</td>
</tr>
<tr>
<td>21</td>
<td>*～</td>
<td>57</td>
<td>0.68</td>
<td>53</td>
<td>*～</td>
<td>23</td>
<td>0.28</td>
</tr>
<tr>
<td>22</td>
<td>変化</td>
<td>55</td>
<td>0.66</td>
<td>54</td>
<td>病変</td>
<td>23</td>
<td>0.28</td>
</tr>
<tr>
<td>23</td>
<td>*～</td>
<td>51</td>
<td>0.61</td>
<td>55</td>
<td>*～</td>
<td>23</td>
<td>0.28</td>
</tr>
<tr>
<td>24</td>
<td>形態</td>
<td>50</td>
<td>0.60</td>
<td>56</td>
<td>中毒</td>
<td>23</td>
<td>0.28</td>
</tr>
<tr>
<td>25</td>
<td>腎(鱗)</td>
<td>49</td>
<td>0.59</td>
<td>57</td>
<td>硬化</td>
<td>23</td>
<td>0.28</td>
</tr>
<tr>
<td>26</td>
<td>動脈</td>
<td>48</td>
<td>0.58</td>
<td>58</td>
<td>物質</td>
<td>22</td>
<td>0.26</td>
</tr>
<tr>
<td>27</td>
<td>発生</td>
<td>47</td>
<td>0.56</td>
<td>59</td>
<td>*～</td>
<td>22</td>
<td>0.26</td>
</tr>
<tr>
<td>28</td>
<td>*～</td>
<td>46</td>
<td>0.55</td>
<td>60</td>
<td>神経</td>
<td>22</td>
<td>0.26</td>
</tr>
<tr>
<td>29</td>
<td>胃</td>
<td>43</td>
<td>0.52</td>
<td>61</td>
<td>所見</td>
<td>22</td>
<td>0.26</td>
</tr>
<tr>
<td>30</td>
<td>培養</td>
<td>42</td>
<td>0.50</td>
<td>62</td>
<td>家兎</td>
<td>22</td>
<td>0.26</td>
</tr>
<tr>
<td>31</td>
<td>腹水</td>
<td>41</td>
<td>0.49</td>
<td>63</td>
<td>多糖</td>
<td>21</td>
<td>0.25</td>
</tr>
<tr>
<td>32</td>
<td>血管</td>
<td>41</td>
<td>0.49</td>
<td>64</td>
<td>異常</td>
<td>21</td>
<td>0.25</td>
</tr>
</tbody>
</table>
この調査でのサンプルは病理学領域であるため、出現した単位語は病理学特有の語がかなり多くみられる。

いっぽう、前項で定義した補助語としては、

的、学、性、症、腺、炎、系、体、剤、例、酸、像、法、類

がある。

特殊単位語としては

癌、肝、肺、脳、腎、胃

などがあるが、大部分は臓器を示す用語である。

頻度順位第1位から第10位までの単位語によって、全体の約4分の1、27.47％を占めている。そのうち、「～的、～学、～性」などの接尾語の和は全部の10.33％である。

表5－13では上位64位までの単位語をリストしたが、サンプル1,000論文から抽出された異なり語数1,104語のうち、延べ単位語、特殊単位語の内訳は表5－14に示すとおりである。

表5－14

<table>
<thead>
<tr>
<th></th>
<th>単  位  語</th>
<th>特 殊 単  位  語 *</th>
<th>総  数</th>
</tr>
</thead>
<tbody>
<tr>
<td>延べ語数</td>
<td>6,188</td>
<td>1,847</td>
<td>8,034</td>
</tr>
<tr>
<td>異なり語数</td>
<td>854</td>
<td>250</td>
<td>1,104</td>
</tr>
<tr>
<td>平均使用率*</td>
<td>7.3</td>
<td>7.4</td>
<td>7.3</td>
</tr>
</tbody>
</table>

* 特殊単位語 (補助語37を含む)

(外国語213)

** 各異なり語の平均使用率 = \frac{延べ語数}{異なり語数}
異なり語数 (X) とその使用率 (Y) の累積パーセントは表 5-15 のようになり、これを図に表わすと図 5-7 のようになる。

表 5-15

<table>
<thead>
<tr>
<th>X (語)</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
<th>1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y (%)</td>
<td>43</td>
<td>58</td>
<td>70</td>
<td>76</td>
<td>83</td>
<td>86</td>
<td>89</td>
<td>92</td>
<td>95</td>
<td>97</td>
<td>98</td>
</tr>
<tr>
<td>延べ数</td>
<td>3,456</td>
<td>4,661</td>
<td>6,671</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

図 5-7 単位語の異なり語数 (X)
—— 累積使用率 (Y) 曲線
(4) 特殊単位語（補助語）の分析

特殊単位語および補助語が総当たり語数 1,104 語中、250語を含めているが、このような接尾語を示す語として機能する補助語について評価をもつめる。

第 4-1 表における補助語は、合成語中で機能的な働きを示す語と考えられる。例として、「～的」、「～様」、「～状」、「～化」について検討する。

a）～的

もっとも出現頻度の高い単位語である。「雑誌記事索引」では第7位であった。この語の意味の働きは不明確であることが多いが、「らしい」という性質あるいは、「ある傾向」を意味している。「～的」はいかなる単位語とも結びつきやすく形容詞化の機能を果たしている。

（例）

・電子顕微鏡的細胞組織

・実験的肉芽組織形成

b）～性

性質または属性を表わし、大部分は単位語に結合しているが、つぎの場合などは、両者の意味の差は明確ではない。

淋巴肉腫　異種腫瘍細胞
淋巴性肉腫　異種性腫瘍細胞

c）～様、～状

いずれも様子、状態を示す機能を果たし、形態の描写に用いられる。

～花房状、～葉状、～囊状～のように「～状」は、物質を示す単位語との結合力かさ強く、

～Virus様、～ウマ様、～腫瘍様～のように、「～様」は病名、現象を示す単位語と結合する場合が多い。
d) ～ 化

作用や変化を示す。

"四塩化、石灰化、繊維化、安定化、実験化、活性化、癌化、異型化" などのようにいかなる単位語とも容易に結合する。

以上述べてきたように、医学用語の特長として、これらの補助語の機能を自由に使用して（ある意味では乱用して）、無制限に近い造語の傾向がある。これを模式化すると

"p化 q状 r性 s 的"

のように表現できる。つぎの例のようなものは、語なのか、句なのかの判断に苦しむ。

(例) 悪性型黑色棘細胞増殖症

直腸周囲組織異所寄生

電子顕微鏡的細胞組織病理学的研究

このような例は、例外ではなく常に論文標題に現われるのである。
6. カタカナ入力漢字かな混り文出力システムの実験
6. カタカナ入力漢字かな混り文出力システムの実験

6.1 システムの概要

このシステムは、日本語の文章を、カタカナけん盤から同音異義語の判断を含む入力処理を行わない、コンピュータでカタカナ漢字変換処理を施して漢字かな混り文に出力するシステム全体を示している、いわゆるカタカナ入力方式によるトータル・システムである。

このシステムを大別すると、
(1) カタカナ入力システム……………………＜入力＞
(2) カタカナ漢字変換システム……………………＜処理＞
(3) 漢字かな混り文出力編集システム…………＜出力＞
の3つのサブシステムに分類できる。

カタカナ入力システムは、単なるカタカナ・パンチによる入力ではなく、カタカナ漢字変換でもっとも問題となる同音異義語の判別処理を、原始データ（原文・原稿）にもっとも近い入力時にて人間が判断を行わないながら入力する方式を採用した。日本語文をカタカナにするときの分かち書きの単位は、漢字2字の単語を原則にするとともに、漢字区分のファンクショングキーを漢字単語の頭に付加することが条件となっている。詳細は6.3節に述べる。

カタカナ漢字変換システムでは、カタカナ入力システムで作成したカタカナ・データ、いわゆるコンピュータ処理で一般的なEBCDICコード、あるいはJISコードなどのデータから、漢字出力装置で扱う漢字プリンタ・コードへの変換処理を行なう。なお、このシステムでは、専門分野別による処理方式を採用することによって漢字の字種の減少と同音異義語の減少をはかっているため、変換テーブルは専門分野別に構成される。

（テーブル構成については6.6節で述べる。）
図 6-1 カタカナ入力方式によるシステム構成の概念図
6.2 システムの利点と効果

このカタカナ入力方式は、とくにつきのような利点や効果がある。

① 入力操作が容易である。
② 情報の蓄積と伝送に既存のコード体系をそのまま適用できるので合理的である。
③ データ入力者の手元にモニタ出力をプリントアウトできるのでチェックや校正が容易である。
④ カタカナ入力システムの専門化やモジュール化がはかるので、高価な出力編集システムに多数のカタカナ入力システムを容易に連接できる経済的な日本語情報処理システムの構成が可能である。

以下にそれぞれの利点について概説する。

(1) 入力操作

情報量が増大するとそれにともないコンピュータで処理する内容がますます複雑になるのはもちろんのことデータ入力のための作業が多くなり、キーパンチに過大の負担がかかっているのが現状である。

一般的の日本語情報処理システムを考察してみると、データ入力の操作を行なう漢字テレタイプのパンチにとっては、まさに非人間的な労力が課せられているといえる。日本語には、漢字という複雑でしかも類似した文字パターンが多いという特徴があるので漢字配列のけん盤を操作するには気苦労があるうえ、誤操作（ミス・パンチ）の恐れもある。

このような入力操作の難点を解消するためにこれまでに漢字を記号化や符号化して入力する方法も考えられてはいるが、いずれも漢字の字数に対応する数だけ、既に述べたく記号や符号を操作者が覚えないならば

—145—
ならないという難点がある。英字 26 文字の組み合せでカナ文字を表現できるローマ字入力方式でも、日本語の文章を入力できるけれども、カナ文字表現と比較してみると、「ＫＡ」の 2 文字を「カ」、「ＫＩ」「キ」というように、ローマ字入力方式はデータ入力の操作回数という面などで効率がある。

そこで、この研究では、カタカナ入力方式を採用し、操作者が日本語の文章を読むながら、いわゆる自然語で漢字かな混り文出力のためのデータ入力を可能にしている。50 個程度のけん盤でデータ入力ができるので、だれでも容易に入力システムを操作できるという利点、言い換えれば、漢字けん盤入力方式のように熟練を要するパンチを特別に教育する必要がないという利点がある。

(2) 番積と伝送のためのコード体系

カタカナ入力方式では、コード体系が、英字、数字、記号を含めて、一般的な 8 ビット構成なので、既存のコード体系をそのまま利用できるという利点がある。

漢字コードは、字種が多いために一般に 12 ～ 16 ビット構成であったが、将来漢字コードの標準化がなされたとしてもビット構成の低減は望めない。

カタカナ入力システムから漢字かな混り文出力システムの間で、既存のコード体系がそのまま使用できるということは、現用のコンピュータ・システムで採用している、ハードウェアやソフトウェアを改変することなく利用できるので実用的である。8 ビット構成のカナコードで日本語の文章を番積できるということは、漢字コードで番積するよりもデータベースの管理面で効率がよい。とくに 12 ～ 16 ビット構成の漢字コードでは、6 ビット×2 とか、8 ビット×2 というように 2 バイトで 1 文字を表現しているので、番積データの管理やメンテナンスなどに複雑なソフトウェアが必要である。
カタカナ入力方式で日本語の文章を書くとき、カタカナの入力データに、漢字出力のための機能を組み込み、漢字出力のためのエクステンションコードを挿入して書く。おき、検索情報を漢字から混じり出力力などで対応した具合に必要に応じて2とおりの形式で出力できるという利点がある。

また、通信回線などを使って日本語の文章を送信する場合は、漢字コードで行なうよりもカタカナ・コードで行なうほうが効率がよい。データ伝送の正確さや速さという面ばかりでなく、高価な漢字から混じり出力の出力編集システムを共用する場合を考えても、漢字テレタイプなどは採用せずに、既存のカタカナ、英字、数字、記号を備えた端末装置だけで日本語情報処理システムを構成できるという利点がある。

(3) 入力情報のチェックと訂正

漢字テレタイプから入力した情報をチェックしたり訂正（校正）するためには、漢字テレタイプからの出力データを漢字出力装置にかけてモニタ用のガラ刷りをプリントアウトする必要がある。

このように漢字出力装置がデータ入力者の手元になければデータ入力の操作者が無理を見ながらパッチしても誤パッチがわからないまま見過ごしたり、また訂正の場合にも漢字出力装置を使って訂正処理を行なわなければならないので出力情報の完成に手間も時間もかかるという欠点がある。

これに対してカタカナ入力システムを利用すると、カタカナの入力データをそのまま汎用的なプリンタに打ち出すことができるのでデータ入力者が操作を行いないながらチェックも訂正も簡単にできるという利点がある。このカタカナ入力システムでは、同音異義語の処理をシステムの選択指示にしたがってデータ入力者が直接実行するようになっている。

(4) システムのモジュール化

このカタカナ入力システムは、カタカナけん盤、同音異義語表示装置および同音異義語を判別するための制御装置といった機器で構成される
が、同音異義語の表示装置については、入力システムの適用分野を専門
別にモジュール化することによって簡略化することが可能である。
このような業務別、専門分野別に専門化したカタカナ入力システムを
多くの場所に設置し、これらの入力システムをオンライン処理であろう
と、オフライン処理であろうと、中央に設置した汎用の漢字かな混り文
出力編集システムに連繋してやれば日本語情報処理のネットワーク・シ
ステムを形成することができる。（図6-1参照）

6.3 カタカナ入力システム
カタカナ入力システムの構成を図6-2に示す。

図6-2 カタカナ入力システムの構成
この入力システムの大きな特徴は、同音異義語の表示方法にある。操作方法は一般のカタカナ入力の操作と同じであるが、文書中の漢字・仮名を区別するため、実際的に漢字キー「#」、カタカナキー「*」を設定し、操作者が漢字・仮名を区別する。

このカタカナ入力システムにおける同音異義語制御部、および表示装置、テーブルの関連を図 6-3 の操作フローに示し、その概要を述べる。

まず、カタカナ入力装置の操作者（データ作成者）は、日本語の原稿を読みながら、同時にカタカナ入力の操作を行うが、この時、原稿中（原稿）に漢字が現われたならば、漢字区分のために漢字ファンクション

図 6-3 カタカナ入力システム操作フロー
リターン・キーの信号が同音異義語制御部に入力されると、「#」からリターン・キーまで（上例では「ニホンゴ」）の読みに対して同音異義語があるかないかのチェックがその制御部で行なわれる。（ファンクションキーの信号が入力されると、あとに続く漢字の読み仮名の入力からはチェック・レジスタに記憶がはじまり、リターン・キーまで続く。）

制御部では、チェック・レジスタに記憶された読みを、同音異義語テーブルと照合し、同音異義語がある時は、制御部をとおして表示装置に信号が送られる。表示装置への信号とは、つまり同音異義語テーブルに対応する表示アドレスのことで、このアドレス指定は制御部にてコントロールされる。なお、この制御部は、ミニコンピュータ以下の規模・機能で可能である。

表示装置には、高価な装置としてC R Tディスプレイなどが考えられるが、実用化のための安価な表示方法としてスライド表示的なものや、マイクロフィルムの応用が見られる。

また、分野別にモジュール化した変換テーブルと、表示テーブルを用意しておき、必要に応じてモジュールごとに交換すれば、システムの利用価値はさらに高まる。
図6-4 同音異義語フロー

-151-
6.4 カタカナ漢字変換システム

ここでは、カタカナ入力システムで作成されたカタカナ・データを漢字コードに変換する処理を行なう。そのためには、あらかじめ漢字単語に関するカタカナ・漢字対応の変換単語テーブルを作成しておくなければならず、変換テーブルに登録する漢字単語の決定については6.6節および3章で述べる。ここでは変換テーブルの格納処理と変換単語テーブルのメンテナンス処理、およびコード変換処理について述べる。（図6-5プロセス・フローのうち枠内）なお変換テーブルは、いずれも分野別に単語単位で構成されるので、以下、変換単語テーブルと呼ぶことにする。

6.4.1 変換単語テーブルの格納処理

この格納処理は、カタカナから漢字かな混り文への変換を行うため、変換単語テープを作成する作業である。

変換単語テーブルに登録決定された原稿は、頻度順に入力され、漢字テレタイプを使って紙テープ（2バイト/1字）に穿孔し、コンピュータにより50音順にSORTし、さらに同音異義語の区分と単語のダブルのチェックを行なう。

このデータは、磁気テープに記録されるファイルとし、訂正処理を行うなかったのち、最終ファイルとして完成される。

この最終ファイルは、つきのようなフォーマットであり、分野、レベル、同音それぞれの区分と、カナ文字、漢字で構成されている。

<table>
<thead>
<tr>
<th>分野区分</th>
<th>カナ文字</th>
<th>漢字</th>
<th>同音区分</th>
</tr>
</thead>
</table>

① 分野区分…………分野別（政治、経済……など分野別変換が可能）
図 6-5 プロセス・フロー
② レベル区分………………1～3区分まであり、検索効率の比率を変更可能
③ 同音区分………………同音異義語の処理可能
④ カナ・漢字の対応文字数は、原則として1漢字2カナ文字とし、最大6漢字の対応テーブル保持可能
⑤ このファイルは、50音順に配列され検索しやすい。
　　 ここで素ファイルから漢字マスタ（最終ファイル）への処理概要を掲げる。
　　 素ファイルから最終マスタを作りあげる処理は、以下のとおりである。

![図解](image_url)
例

漢字コード

<table>
<thead>
<tr>
<th>ポシュウダン</th>
<th>区分</th>
<th>母集団</th>
<th>L</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>0010320</td>
<td>KEY</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6.4.2 変換単語テーブルのメンテナンス処理

図6-6のうち、「赤字校正済ブルーフ・リスト」以下の処理によってメンテナンスを行なう。手順に従って概説する。

① EBCDICマスタ（SORT済EBCDICファイル）での校正を行うためマスタ原稿と再チェックを前段階の処理で行なったが、再度、この段階でデータ・ミスが発見された場合、グレード部、漢字部のデータが修正カードとしてバンチされる。

図6-7 訂正カード

1. 訂正
2. 新規
3. 削除
図 6-6 漢字マスタ
② 訂正カードによりEBCDICファイルの内容を修正する。KEYは、EBCDICカナ部・グレード部よりなる。
③ EBCDICファイルをディスクにカタログ・ファイルをする。このディスク・ファイルは漢字かな混り文出力に必要な漢字マスタ・テーブル（変換単語テーブル）となる。

漢字マスタ・テーブルは、第1カタログ、第2カタログに分類されて、ディスクにファイルされる。

(4) 第1カタログ、第2カタログについての概略はつぎの6.4.3項を参照。

6.4.3 コード変換処理

(1) データの変換

変換方法は、まず文章の1区切りが読まれ、その区切りの頭にあるファンクションキーチ「＊」、「＋」により、それぞれの処理方
法が違ってくる。
「♯」は、変換単語テーブルのサーチを行なって、漢字変換処理をするための記号である。もし、テーブルに該当する単語がなければ漢字単語はカタカナのアンチ文字に変換される。
「＊」は、外来語であるカナ文字コードをテーブル・サーチして変換するための記号である。
「♯」、「＊」以外のものは、ひら仮名、英数字、特殊記号のコードをサーチして変換する。
このように、カナ文字データの変換とは、カナ文字（EBCDIC）コードを漢字コードに変換することを意味するものである。

![図 6-8 コード変換処理](image)

例をあげると

ex1

♯タカチホ L2000 ハ #ハツビョウ ハハード ト………
→ 高子陸 L2000 を発表、ハードと………

ex2

♯ツウサン #コウギョウ #カ デ #カカク #ブ シリ…
→ 通産工業課で価格分離
図 6-9 コード変換処理フロー概図①

図 6-10 コード変換処理フロー概図②
変換単語テーブルの検索方法は、データの区切り部分と、テーブルのカナ文字部分をコンピュータ内部でカタログ方式によりマッチングさせ、一致したときに、テーブルの漢字部分を取り出しファイルする方法である。

データ              テーブル
                           #X₁・X₂・X₃・…・Xₙ          X₁・X₂・…・Xₘ□Y₁・Y₂・…・LF
                           ↓                        ↓
             マッチ
                           ↓                        ↓
                           Y₁・Y₂・…

図 6-11 マッチング方法（この場合、検索方法はカタログ方式をとっている。）

（3）検索におけるカタログ方法について

第1カタログ
（第2カタログのDISKアドレス）

第2カタログ
（データレコード・アドレス）

図 6-12 変換単語テーブルのカタログ方法

このシステムにおける大きな特徴にカタログ方式による検索があげられる。

-161-
まず、変換すべきデータに対して、第１カタログで頭２ケタのカナ部を検出する。検出したポインタ（y）が第２カタログのアドレスとなっているが、さらにデータの頭３ケタ目のカナで該当する漢字部の単語をとり出す。それらの単語のどれをとるかは、同音異義語区分で選択させる。

変換される１ブロックのカナ部が４ケタ以上の場合は、すべて同音異義語とみなす。

この処理の特徴を述べると、

① カナ入力……EBCDICコード体系で、カタカナ、ひらがな、漢字区分キーにより分けられており、１回に入力するデータは、6文章ずつ処理される。

② DISKマスタ……磁気テープマスタ内で50音順のものを、DISKに第１カタログ、第２カタログをもち、ダイレクト（直接コード演算方式）にカナ対応、漢字レコードの入っているDISK上のアドレスを検査出来るファイリングになっている。

③ 検索……②で述べたようなDISKマスタを使用し、カナから漢字へダイレクトに変換する。テーブルがないときは、前述したよ
うにアンチのカタカナに変換される。

④ 変換……③で述べた漢字部を除いたカタカナ、ひらがな、特殊記号コードを漢字コードに変換する。
	ると「」、幼、促音についても処理される。

⑤ 出力……この場合、全データが漢字コードになっており、つぎに述べる汎用編集処理の入力となる。

(4) 入力データの１文章は、最大５レコード構成であるが、最大３０レコード（６文章）分のエリアを確保し、#～の漢字部をピックアップする。その後コア内に漢字部の対応テーブル（最大３０レコード×２０語）を作成する。

(5) 同音異義語は、#で区分されるカナ文字最後のそのつぎの数字で区分される。

ex. #コウセイ①□ノ□#ホウホウ……

プランク

数字で区分

6.5 漢字かな混り文出力編集処理

漢字かな混り文のデータ（素ファイル）が作られると、電算植字システムによって出力編集の処理が行なわれる。なおこの電算植字システムは学習研究社で開発したシステムを利用している。

とくに、出力に使用される汎用編集ソフトウェアはつぎのようなもので構成される。

(1) 自動編集ソフトウェア

自動編集ソフトウェアとは、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であり、HITAC-10の磁気テープ・ベース用に開発されている。

このソフトウェアは、漢字情報をコンピュータをもって正確かつ迅速に
図 6-14(a) カマナ入カデータ例
国際通貨じょうせいが安定するための前提条件はきわめてはっきりした通貨政策のしりょう問題をふくめて、きじゅう通貨ドルのしんにん回復にかんならのメソをつけることであろう。だが、このように考えたうえで、短期期にかいていくことは、もはやほんのじょうせいからみてきわめてこんんだきわざをえない。とすれば、国際通貨不安定という態勢はかなり長期化し、どうせ、むしも現在の変動相場制を長期に持続せざるをえないとだろう。

変動相場制が固定制にくらべると、貿易じつむ、じゆじゅのししようをもっていることはじつである。

とくににほんのうおおだの為替管理でかんじがまったように、相場の市場におけるまではそうである。しかし、国民経済のかんてんからすれば、変動相場制にはつぎのようにしとめてあることわすれてはならない。

その第1は、経済平衡において、国際収支のせいないかひょうして、政策選択の重がひろがること。

第2は、輸入インフレをしやだんし、国内の物価上昇をよくせいする効果で、もっともである。にほん経済はこれで、国際収支、物価、福祉という3者にかすめられながら、相互のトレード・オフ関係から有効な政策手段が3つみの常態にあると言われた。

だが、変動相場制をとることによって、とうれん、国際収支上の手かせあしかか、はずされれば、政策選択のほばれだけひろがる。

しかも、国内物価上昇のもっともゆうりょうを変えるといわれる海外インフレのようくくかげやだんさんれば、政策とうきょうとしては、今後の経済状況のじゅうてんを、あけて国民経済こうじょうの1点にひかえることも可能ならばある。

くわえて、つぎじゅこうかいからも、こんかいの変動相場制には、フロートスレしたといったが、せんかいほどのつよいものはかきえなくなった。

こうした点をこうりすると、できるだけ為替信にとくしてカレンの適性水準をみつける必要性からも、このさいは、じっくりを、すえの円の長期変動相場制をかくこととして、その前提にたった経済新機をそられるのである。

ただ、変動相場制の長期化は、国内の物価上昇、インフレをはきものる。保証とはかならずしもならない。

よくしられれているとおり、きくねあきからほんねにかけているもう一人をめるいるおししれ物価ゆうとう要因には、輸入物価上昇のほかに、不動カルテルの供給制限の行為、景気回復にもなる需要の拡大、部門間需要のシフト、いわゆる通貨流動性、資金コストのどうだいなどを考えられる。

つまり、いまの物価上昇は、これらの諸要因がふくいつこからみあっておこっているわけである。変動相場制の影響をみるため、そうせないうえの必要性から、のほしな家族経済の危険はさらに。

さいきのにおししれ物価上昇と、交易条件のきゅううあかくにたんてきにかかせられているように、もう、だいすくようまの輸入価格のふうえによることがわかめておき。

したがって、この面からの国内物価じょせいは今後、じょじょにでてくるものときたいされるが、同時に物価安定化にきくよう、きんじは経済をじょきょうのよういかんするような施策が積極的にとられる必要がある。

インフレをこうしんは、とくなくおおう国民福祉のうえにおかえるものであり、変動相場制かにおいても、インフレをこうせいがいせん、じゅうでな政策がきたいとしてのこることはらんをまたない。

その意味で政策どうきょうが、預金準備率のきあいあげや銀行のまどくえきなどによって、金融政策のきちょう転換にのりだしたこととは通貨流動性助成のためにきたいできる。

だが、きくねあき10〜12のGDIの実質成長率がせんねん動態比12.7％になったことを、はじめ、ばんねん1月までの景気誘致がつきためおおばはな上昇をしめた

ことを、もって、景気からはついてと判断し、公定歩合のききあげを、けんとうしているとすれば、問題があらう。

図6-14(b) 漢字をかみ混じり文字例

－166－
なぜなら、第1に、悪気ねんあきからはねん1月にかけての景気指標おおはばってんは、ぜんねん動機の景気がどうずことき常態にあったことのばんえいでもあるからで、景気がこのままかだね常態にとっては도ううかばは、こうしなくもにゆうなるべきない。

第2に、従来は変動相場制にこことともなって実際の景気をうかげることである。とくにおうしゅう通貨じょうせいの流動化によって、円通場はとうもろも予想をおおはばにうか文字えんがた次元がひつしとみられるだけ、その影響は輸出のねん産業を中心にかきついたものがある。

このような景気で、ばんかくの動機はきょうしんで、経済指標にうったえらばつ、通貨流動化の動きに対すはあつって、物価はそれほどさがらず、むしろせきさんが低下して、おろしごり物価上昇をしけきしかねないとおもわれる。

このさいは、輸出型産業のないじゅうそれと共しづするため、むしろ可能なければ国内の福祉型需要の拡大や輸入そししんじゆうてんを おいた政策を指向すべきであろう。

産業構造のためのいわゆるまえきの調整援助、うちずけ力流動化のための政策指針などの施策が積極的にとられる必要があることはいうまでもないが、おうしゅう経済にうるべての政策のこうすけいを果たすためにもかながれべよい。

第3は、変動相場制において、におはな経済のたきょうは力、潜在性をじゅうふんにいかしながら、経済たいしゅを これまでの輸出型から福祉型へとかかんに変換していくことが政策の基本である。

原因はさまざまだが、いま世界の食糧事情は2年つづきのばくくで展開がひっかかず、かがくにののうせいの基
本もがんがえなすべきだということがいちじるしいことである。

このたびとつづももれた農家はしあやしい、ねんたるふうとのがんとともに、世界ののあたとなった市場の不安定性に対する食糧じゅうおうのひきあいを きょううちようしている。

しかし、これからのおうせいはただただに食糧のじゅうおうを ひとめずのではなく、てきちてきさんが 中心とする高能率、低コストの生産体制をじゅうおうにかくうつにするのがもっともじゅうおうであらす。

世界食糧農業機関の推奨によると、ねんたるにおける世界の食糧事情はしあやしくあたたかたくんとほおなほ
けがれたみこめで、けつきよも3年つづきのあくを 予想している。

こくるいのじゅつは同時に出る供給もふくももががり、ちくさんがの生産におおかんししようを おぼす。

おわりとせはつきねん一月に ものいさよじゅいようの延長と生産もくひょうのしあや を こひょうした。

これには、FAOの「1981年の食糧機関の展望」やローマ・クラブの「成長の限界」などにとらわれている食
糧事情にたいする技術的なみかたを ねんとうにいて、たがくにの食糧じゅうおうをかかめていくことを お
らしいとしている。

こんかいの農家はくしょいでもみじよめたる問題意識をもちながらのうさんが生卵、農業経営、農業経済の動向を 
分析している。

農業生産はのがさんの輸入のぞうた、このの生産調整や普通はたけきもつ、げさんなどによって45． 46年と交差傾向をたどった。

47年の農業生産はじやつかのかのもなおしけがきたきされるものの、いぜんとして低水準にある。のようじる地
の流動化はちかの上昇によって絶えられ、農地所有者 げんのゆうしゅう移はげたいをつづけている。

また、このの生産調整によるきゅうくうてんのはんふらなば後配したままはうちされ、そのまわり。とく 
に農業の生産人口の現象がでまった。農業の生産人口の現象数は40年から44年平均では3．2％であっ

－167－
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。

① 入力処理

漢字テレタイプから入力された紙テープのデータに対して、文字頭変換、文字の大小、および縦・横のファンクション処理、英数字、特殊文字などのピッチ付加処理を行う。

- ファンクション・コード処理
- スペース・コード処理
- キャラクタ・ピッチ処理
- ジャスティフィケーション処理
- その他の

② 検査調整処理

入力処理を経た文字データは、1行単位をもって、字詰、禁則、ルビ、ハイフォネーションなどの処理がなされる。

- 禁則、分離
- 行末、行頭、中央換
- タブ、等配分、ルビ
- 合成、ハイフォネーション
- その他の

③ 組版処理

ここでの処理は、文章全体の編集処理であり、図版、見出し、写真、ページ割りなどの処理が行なわれる。

④ 校正、版下作成処理

校正用、版下用プリント、また各種モニタや台帳などのプリント処理を行う。

-168-
図 6-15 自動編集ソフトウェアの処理フロー
6.6 変換単語テーブルの決定

日本語の文章をカタカナ入力方式によって処理する場合、コンピュータ処理以前に行なうこの変換単語テーブルの決定はもっとも重要な作業である。

カタカナから漢字への変換処理において、その変換率はシステムの影響ではなく、変換テーブルに登録する漢字および漢字単語いかんによって決まってしまう。とくにカタカナから漢字への変換で問題となる同音異義語の処理に関連するため、十分かつつ慎重な判断・決定が要求される。

ここに設計したシステムでは、同音異義語の判別について検討した結果、専門分野別構成とともに単語（熟語）単位の漢字、いわゆる漢字単語を基本にした変換テーブルがもっとも効果的であると判断し、変換単語テーブルの決定を行なっている。

以下に変換単語テーブル格納までの手順にそって概説する。

まず最初に専門分野における漢字および漢字単語の調査分析を行なう。学術書に関するような専門分野であれば、ハンドブック、用語辞典などの専門用語集を参考にしたり、専門書のキーワードや索引などから漢字単語を抽出することが手近でよい。しかしながら、用語集などとしてまとまっていない分野においては、過去において出版された文書・文献のなかから漢字単語を抽出する方法がとられよう。この作業を進めるうえで、漢字単語の抽出とともに出現頻度をカウントする必要がある。これは、登録単語の規模をシステムにあわせて考慮したほうが当然ながら効率的なためであり、頻度の低い単語については、将来の必要性などを推測して判断したければならない。

こうした作業・検討の結果、専門分野別に漢字単語を決定するが、このとき1分野に限らず共通性のある単語に対しては一般語共用テーブルとして別に登録する必要性も生じる。とくに固有名詞に関しては共通性が高いので別に固有名詞テーブルとして登録する必要がある。
変換単語テーブルに登録する漢字単語が決定したら、つぎに漢字単語とその読み仮名（カタカナ）を対応させて、漢字テレタイプで紙テープにバ
ンチ入力する。
このうち漢字単語部を除いてEBCDICコードに変換するとともに、固定フォーマットに編集して、分野別変換単語テーブル、マスタの磁気テ
ープを作成する。
（例）

紙テープ

（Variable）

いずれも漢字テレタイプコード
7. 文書情報とマイクロフィルム
7. 文書情報とマイクロフィルム

7.1 情報量の拡大と多様化

7.1.1 情報の流動

情報化社会の進展にともない、情報過剰とか、情報格差とか、情報公害などの言葉が使われるようになってきた。経済企画庁の数年前の情報量の推計によると、昭和40年から昭和60年までの20年間に情報量は約90倍に伸びるという試算がなされた。これは新聞、雑誌、テレビなど世の中に出されるすべての情報量を文字数に換算したものであるが、
昭和60年の1年間の1人当たりに換算すると、約43億字、普通の書籍にして約10万冊という勘定になるという。

科学技術情報についての予測は、OECDDのアンダーラ教授によってなされたが、それによると、1970年の1年間に約200万件の論文が発表され（6,000〜7,000件/日）、少なくとも年率10％で増大し続けていくとると、1985年には、800万件/年になるという。

![グラフ](image)

図 7-1 科学技術情報の増加

図 7-1は1955年を100とした場合の1985年までの情報量の増加予測を示す。
このような多発的に増大する情報の中から自分に必要な情報を適切に、しかも必要な時に手に入れるようにすることは、そう簡単ではなくなっ
tてきている。そこでコンピュータによる情報検索が開発され、実用的に
も用いられている。

7.1.2 情報形態のいろいろ

コンピュータによる情報の整理は、大量のデータを処理することの問
題につきるので、最短の問題をある意味で解決できたとみてもいいかもしれませんが。

しかし、現在のコンピュータを利用する場合には人間が通常使っている文字や図などを取り扱うというより、コンピュータで扱い易い範囲に
合わせている。すなわち人間がコンピュータに歩みをした、コンピュー
タを使わせてもらうという見方の方が正しいような状況にある。
たとえば、日本人は、通常漢字やひらがなを用いているにもかかわらず、
現在のコンピュータでは英字またはカナ文字しか通用しないのがほとんど
とんどである。それも手書きで簡単にやりとりすることは困難で、キーパ
ンチによる手段をとっている。したがって、文書情報の場合でも、従来
の形式でのやりとりをコンピュータで行なうことは難しい。

7.1.3 文書情報のマイクロフィルム化

マイクロフィルムに文書情報を入れることは、通常の文書の形式で扱
えるし、また図や写真や色彩なども表わせる可能性がある。このような
文書情報のマイクロフィルム化は一般に広く採用され、その装置も各種
のものが出回っている（表 7－1）、（表 7－2）。しかしこれらのマ
イクロフィルムへの文書情報は、日常われわれが使っている形式になっ
ていて読むことは簡単であってもその必要な情報を引き出してくれるのは
記号または番号で行っている。これは機械向きであって人間向きではな
い。これを人間向きの通常の人間の言葉で書くことは難しいことである
が、ここにコンピュータを組み合わせることによって一部は解決する。

—174—
### 表 7-1 ファイルの形態

<table>
<thead>
<tr>
<th>業種</th>
<th>ファイルの形態</th>
<th>ロファイル数</th>
<th>ファイル数</th>
<th>アカウント数</th>
<th>メーカー数</th>
<th>計</th>
</tr>
</thead>
<tbody>
<tr>
<td>水産</td>
<td>魚類</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>化石・石油</td>
<td>石油</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>鉄鋼・非鉄</td>
<td>鉄鋼</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>36</td>
</tr>
<tr>
<td>食品</td>
<td>食品</td>
<td>15</td>
<td>9</td>
<td>10</td>
<td>3</td>
<td>37</td>
</tr>
<tr>
<td>機械・バブル・紙</td>
<td>紙</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>化学・ガス・鉱子・土石</td>
<td>鉱子</td>
<td>34</td>
<td>4</td>
<td>16</td>
<td>7</td>
<td>75</td>
</tr>
<tr>
<td>金属製品・金物製品</td>
<td>金物</td>
<td>42</td>
<td>9</td>
<td>18</td>
<td>18</td>
<td>118</td>
</tr>
<tr>
<td>電気機器</td>
<td>電気</td>
<td>75</td>
<td>8</td>
<td>24</td>
<td>22</td>
<td>174</td>
</tr>
<tr>
<td>輸送機器</td>
<td>輸送</td>
<td>34</td>
<td>2</td>
<td>17</td>
<td>23</td>
<td>100</td>
</tr>
<tr>
<td>機械工具機械</td>
<td>機械</td>
<td>23</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>34</td>
</tr>
<tr>
<td>その他</td>
<td>その他</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>織物</td>
<td>織物</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>金融保険不動産</td>
<td>金融</td>
<td>29</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>運輸・倉庫</td>
<td>運輸</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>電力・ガス</td>
<td>電力</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>23</td>
</tr>
<tr>
<td>サービス</td>
<td>サービス</td>
<td>18</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>38</td>
</tr>
<tr>
<td>通信新報出版</td>
<td>通信</td>
<td>24</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td>医療</td>
<td>医療</td>
<td>78</td>
<td>8</td>
<td>14</td>
<td>29</td>
<td>128</td>
</tr>
<tr>
<td>教育・文化・病院・団体</td>
<td>教育</td>
<td>115</td>
<td>19</td>
<td>8</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>計</td>
<td>計</td>
<td>543</td>
<td>74</td>
<td>142</td>
<td>184</td>
<td>167</td>
</tr>
</tbody>
</table>

### 表 7-2 被写体の種類

<table>
<thead>
<tr>
<th>業種</th>
<th>被写体</th>
<th>構造</th>
<th>組合</th>
<th>建築</th>
<th>住宅</th>
<th>産業</th>
<th>計</th>
</tr>
</thead>
<tbody>
<tr>
<td>水産</td>
<td>水産</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>化学・ガス・鉱子・土石</td>
<td>化学</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>12</td>
</tr>
<tr>
<td>鉄鋼・非鉄</td>
<td>鉄鋼</td>
<td>18</td>
<td>12</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>食品</td>
<td>食品</td>
<td>29</td>
<td>18</td>
<td>14</td>
<td>-</td>
<td>1</td>
<td>63</td>
</tr>
<tr>
<td>機械・メーカー</td>
<td>機械</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>金融保険不動産</td>
<td>金融</td>
<td>-</td>
<td>14</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>17</td>
</tr>
<tr>
<td>化学・ガス・鉱子・土石</td>
<td>化学</td>
<td>21</td>
<td>38</td>
<td>23</td>
<td>6</td>
<td>3</td>
<td>97</td>
</tr>
<tr>
<td>機械製品</td>
<td>機械</td>
<td>90</td>
<td>36</td>
<td>22</td>
<td>9</td>
<td>-</td>
<td>153</td>
</tr>
<tr>
<td>電気機器</td>
<td>電気</td>
<td>110</td>
<td>64</td>
<td>40</td>
<td>17</td>
<td>3</td>
<td>224</td>
</tr>
<tr>
<td>輸送機器</td>
<td>輸送</td>
<td>69</td>
<td>37</td>
<td>31</td>
<td>14</td>
<td>-</td>
<td>182</td>
</tr>
<tr>
<td>その他</td>
<td>その他</td>
<td>17</td>
<td>12</td>
<td>15</td>
<td>3</td>
<td>3</td>
<td>54</td>
</tr>
<tr>
<td>組合</td>
<td>組合</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>計</td>
<td>計</td>
<td>486</td>
<td>423</td>
<td>296</td>
<td>66</td>
<td>133</td>
<td>53</td>
</tr>
</tbody>
</table>
すなわちキーワードは現在のコンピュータで使用できる文字を採用し、実際のソースの文書データは、日常使っている文字を書いたり読んでる方法である。

7.2 マイクロフィルムのあらまし
7.2.1 マイクロフィルムの特徴
マイクロフィルムは、資料を縮小撮影したフィルムで、その縮小したフィルムを必要に応じて取り出し、拡大して読み取る方法で利用している。
マイクロフィルムにはつきのような特徴がある。
• 記録が正確で透明度が低い。
• 秘密の保護ができる。
• 情報を通常の資料の約 \frac{1}{100} で縮小できる。
• 再生はいろいろな形に簡単にできる。
• 保管寿命が長い。
7.2.2 マイクロフィルムの種類
マイクロフィルムは、大きくわけるとロール状のものとカード状のものがある。
(1) ロールフィルム
これは 100 フィート（30.5 m）のフィルムをリールに巻いたもので、普通 2000 コマの写真を入れることができる。ロールフィルムの巾には 16 \text{mm}、35 \text{mm}、105 \text{mm} などがある。16 \text{mm} 帯のものは、通常一般文書、伝票、小切手、手形、証券などに記録する目的で使われている。35 \text{mm} 帯のものは、学術文献、ときに新聞、図面などに多く用いられている。105 \text{mm} 帯のものは大形図面や地図を、より精度高く詳細に記録し、ハードコピーで作る目的で利用している。
(2) マイクロフィッシュ

これは数10コマを縦横に並べたカード状のもので300コマ程度収容できる。マイクロフィッシュは、論文、文献、文書などを分類してファイルし、図書の整理カードのように肉眼で読めるような表題をつけと、取り出すことができるようになっている。最近ではさらに、3000コマ以上も収容できる超マイクロフィッシュも使われている。

(3) アーチチュアカード

ロールフィルムを1コマまたは数コマにカットし、紙カードの窓をあけてある部分にはつけて使用するもので、分類や検索に便利になっている。

(4) PCMI

PCMI（Photochromic Microimage）は、NCR社のもので3200ページを標準の4×6インチのマイクロフィッシュに毎秒1000ページの速度で入れられるスーパマイクロフィルムである。

7.2.3 マイクロフィルムの作成と利用

マイクロフィルムを作成する手順は図7-2のようになっている。

撮影
現像
保管
検索

写
読

図7-2
マイクロフィルム作成手順

－177－
これらの流れの中では、撮影カメラ、フィルム自動現像機、引き伸ばし機、保管容器、マイクロフィルム・リーダーや、リーダー・プリンタなどのマイクロフィルム用機器が、それぞれ必要に応じて用いられている。

7.3 マイクロフィルム用機器

7.3.1 マイクロフィルム撮影カメラ

撮影カメラとしては、平床式カメラ、マイクロカード・カメラ、ロータリカメラなどが用いられている。

(1) 平床式カメラ

これはテーブル状の複写台の上に被写体を置き、撮影縮小率に応じて上下動させ、上部から撮影する装置である。平床式カメラは設計図面、帳簿、一般文書、地図、新聞などあらゆる分野に利用できる。

(2) ロータリカメラ

これは被写体とフィルムを互いに撮影に応じたスピードで走らせ、一部ずつ撮影するカメラである。これはシート状被写体を連続的に高速度で撮影できるため小切手や手形、株券など証券類に利用されている。

(3) マイクロカード・カメラ

これはシート状フィルムの中に数10コマを定められた順序で撮影するものである。写したフィルムは16mm専用、16mmと35mm兼用のフィルム自動現像機で現像する。

7.3.2 現像処理機

撮影されたフィルムは、一般の写真フィルムより非常に長尺なのでマイクロフィルム用の自動連続の現像処理機が使われている。しかも現像に要する暗室は不用であり、高品質のフィルムの現像ができるので能率的である（図7-3）
7.3.3 再現装置

マイクロフィルムはそのままでは読みとれないので、これを読むための再現装置としては、マイクロフィルムをスクリーン拡大投影して読むためのマイクロ・リーダと、マイクロフィルムからハードコピーを作るためのリーダ・プリンタがある。

(1) マイクロ・リーダ（図7-4）

マイクロ・リーダには、透過式スクリーンと反射式スクリーンがある。マイクロ・リーダは普通の部屋で容易に読めることが重要であるがスクリーン上の像の明るさは、光源や投影レンズやフィルムなどを同じ条件にして比較した場合、透過式スクリーンの方がはるかに明るく見える。また読む姿勢が楽にとれるようにするための自由度は透過式の方が大きいので、販売されているリーダの約90％は透過式であるといわれている。

リーダにとって以上の他に大切なことは、スクリーン上の文字が適当な大きさになるように倍率を考慮することである。なお像の解像度はJIS規格ではスクリーン上の中心で4本/mm、周辺で2.5本/mm
以上としている。また1コマづつずらしていくときのスキャンニング機構がスムーズであり、像が横になってしまったり、上下逆になっていたら簡単に操作できるような回転機構も大切である。

図7-4 マイクロ・リーダの構造の例

(2) リーダ・プリンタ（図7-5）
リーダ・プリンタには、銀塩安定化方式、静電式のエレクトロフォックス方式および電解方式がある。現在主に使われているプリンタ機構によると、スクリーン上の像をプリント作成するのに、押ボタンを押して10～20秒でハード・コピーが得られるものが多い。
ENLARGING LENS

SECOND CONDENSING LENS

REFLECTOR

LAMP
FIRST CONDENSING LENS
HEAT ABSORBER GLASS
FIXED MIRROR

SWINGING MIRROR (VIEWING)
(EXPOSE)

VIEWING SCREEN

COPY PAPER

FILM PLATEN (FILM)

THIRD CONDENSING LENS

HEAT ABSORBER GLASS

図 7-5 リーダ・プリンタの構造の例
7.3.4 マイクロフィルム検索装置

マイクロフィルム検索の方法は、フィルムの種類によっているいろである。マイクロフィッシュの場合は、主題、著者名、資料番号などの見出しを指示することによって検索し、さらに各コマの内容をスクリーンに写すなどして必要な情報を選んで出している手動検索方式がある。これに対してもマイクロフィッシュおよびその各コマにアドレスをつけておいて、そのアドレスをキーから打ち込むことによって検索を行う自動検索がある。また数十枚のマイクロフィッシュをカートリッジに入れておいて、各マイクロフィッシュに凸をつけておき、これをハンマでたたいて必要なマイクロフィッシュを選び出す方法もある。

ロールフィルムの場合は、1つの資料のグループのはじめに、内容の見出しや撮影手順などをグループの切れ目ごとにターゲットを入れ、またある一定のコマ数ごととか、一定のフィルムの長さごとに資料の像と反対、たとえばネガフィルムの場合、白ぬきにするなどにしたフラッシュを入れておいて、このフラッシュとターゲットとコマ番号を組み合せることによって検索を行うことができる。これをフラッシュ・ターゲット方式とよんでいる。

またコマとコマの間にラインを写し出しておいて、これをマイクロ・リーダにかけて高速に送ると、このラインは連続した線に見える。この線の位置を検索の項目ごとにつけしておくと、リーダのスクリーンの目盛りから、現在の情報が何であるかをみつける手がかりとなる。これをコードライン方式とよんでいる。（図7-6）

ロールフィルムのコマの下部にあらかじめ小さなマークを写し出しておき、これのマークをカウントして、コマを探す方法をピリップ方式と呼んでいる。自動検索装置つくりの場合は、番号をキーで指定することによって、その番号がカウントされその場所の像がスクリーン上に写し出される。

ロールフィルムに白と黒で表現した2進法コードを資料といっしょに
図 7-6 コードラインフィルム

図 7-7 ミラコードの画像・コードの配列

図 7-8 ロールフィルム複写機
図 7-9 サーマルフィッシュ複写機

写しておくと自動検索するのに便利で、これをミラコード方式と呼んでいる。（図 7-7）

7.3.5 複写機

マイクロフィルムは、その情報を必要な場所ですぐ使用できるように、同一のフィルムをコピーしておくことが多く、これには図 7-8のようなロールフィルムの複写機や図 7-9のようなマイクロフィッシュ複写機が使われている。

7.3.6 COM

(1) COM とは

COM（Computer Output Microfilming）とは、コンピュータの出力を磁気テープに出力し、このテープの内容をラインプリンター
に出すとスピードが遅いので、マイクロフィルムに記録する方法を指している。コンピュータと直接して使うことのできるものもあり、これらをオンラインCOMと呼んでいるが、一般にはオフラインで使っている場合が多い。また文字以外に図を描くことのできるものもあり、これをグラフィックCOMと呼び主として科学用COMとして使っている。

COMに対してCIM（Computer Input Microfilming）はマイクロフィルムに記録されたデータをOCRで読み取れるようにしたものである。

現在市販されているCOMの一覧表を表7−3に示す。

(2) COMの利用状況

COMの性能上から分類すると次の3つに分けることができる。

① ノンインパクト・プリンタとしてのCOM
　ラインプリンタの代わりに使えるようにしたもので、1行132文字で64行からなる標準の出力を打出すようになっている。

② 事務用COM
　線、円、グラフ、棒などの簡単なグラフを描いて見出しをつけることのできるプリンタとして使えるようにしたものである。

③ 科学用COM
　精密な図が描け、XYプロッタの代わりとしても使えるようにしたものである。
　これらの用途の割合は図7−10のようになっている。
<table>
<thead>
<tr>
<th>メーカー (Information International社)</th>
<th>品番</th>
<th>機種</th>
<th>カルゴン (PERTECH)</th>
<th>三菱電動機製品数</th>
</tr>
</thead>
<tbody>
<tr>
<td>ディーゼル</td>
<td>FR-80</td>
<td>167</td>
<td>2100</td>
<td>PERTECH 3700</td>
</tr>
<tr>
<td>トランク交</td>
<td>1トランク、2トランク</td>
<td>7</td>
<td>7 *</td>
<td>7トランク交、2トランク交</td>
</tr>
<tr>
<td>線間距離 (kg)</td>
<td>550,700,1600</td>
<td>200,1500</td>
<td>200〜1600</td>
<td>550,800,1600の3種</td>
</tr>
<tr>
<td>キャリプレーキー</td>
<td>2インパ,ACDC,</td>
<td>BCD,EBCDIC</td>
<td>RCD,EBCDIC,BCDIC,</td>
<td>EBCDIC,BCDIC,</td>
</tr>
<tr>
<td>ブレーキ速度 (KG)</td>
<td>120</td>
<td>60〜120</td>
<td>60</td>
<td>25〜50、60の3種</td>
</tr>
<tr>
<td>文字種類</td>
<td>128</td>
<td>128</td>
<td>128</td>
<td>64種（128幅に優先）</td>
</tr>
<tr>
<td>文字発生方式</td>
<td>直接ストローカ・ジェネレータ</td>
<td>CRTによるドット方式</td>
<td>CRTによるドット方式</td>
<td>CRTによるドット方式</td>
</tr>
<tr>
<td>文字数</td>
<td>60〜120</td>
<td>60〜120</td>
<td>60〜120</td>
<td>60〜120</td>
</tr>
<tr>
<td>ティルトサイズ</td>
<td>35m,105m,105m,216mm</td>
<td>15〜35m(50mmピクセル)・105m(35フレーム)</td>
<td>105m(35フレーム)</td>
<td>15m(75m)〜15m(25m)</td>
</tr>
<tr>
<td>フォームサイズ (フィールド)</td>
<td>105mm</td>
<td>200、400</td>
<td>200</td>
<td>200〜400、100</td>
</tr>
<tr>
<td>スピード (フレーム)</td>
<td>74SEC/Hフレーム</td>
<td>6、5〜7、5秒</td>
<td>5〜7、5秒</td>
<td>5〜7、5秒</td>
</tr>
<tr>
<td>小数</td>
<td>1/21〜1/290</td>
<td>24/1,142〜1/4</td>
<td>1/21〜1/290</td>
<td>24/1,142〜1/4</td>
</tr>
<tr>
<td>検出コード</td>
<td>IMAGE COUNT, MIRACODE</td>
<td>オフ</td>
<td>オフ</td>
<td>オフ</td>
</tr>
<tr>
<td>チューナ方式</td>
<td>バイテビ方式</td>
<td>バイテビ方式</td>
<td>バイテビ方式</td>
<td>バイテビ方式</td>
</tr>
<tr>
<td>コンピュータインタフェース</td>
<td>Burroughs CR-GF Honeywell 120,120,210,310,410</td>
<td>OFF</td>
<td>OFF</td>
<td>OFF</td>
</tr>
<tr>
<td>明度調整</td>
<td>全範囲</td>
<td>全範囲</td>
<td>全範囲</td>
<td>(全画面あらゆる明</td>
</tr>
<tr>
<td>ベーンサイズ (字幅)</td>
<td>132×64(16x12)</td>
<td>132〜242×88</td>
<td>132〜242×88</td>
<td>132〜242×88</td>
</tr>
<tr>
<td>フィールド高</td>
<td>ブラックフィールド</td>
<td>ブラックフィールド</td>
<td>ブラックフィールド</td>
<td>ブラックフィールド</td>
</tr>
<tr>
<td>機種速度 (ページ/分)</td>
<td>500ページ/分</td>
<td>120</td>
<td>180〜410</td>
<td>20ページ〜6ページ/分</td>
</tr>
<tr>
<td>メーカー名</td>
<td>ヒューズ・バーチャル</td>
<td>ヒューズ・バーチャル</td>
<td>ヒューズ・バーチャル</td>
<td>ヒューズ・バーチャル</td>
</tr>
<tr>
<td>価格 (レシート)</td>
<td>1,825万円(31,210,310,410)</td>
<td>301万〜720万円</td>
<td>301万〜720万円</td>
<td>301万〜720万円</td>
</tr>
<tr>
<td>メーカ</td>
<td>(Quantor酬)</td>
<td>(C ME酬)</td>
<td>東京エレクトロニクス</td>
<td>キャノン（株）</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-----------</td>
<td>----------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>型名</td>
<td>Quantor 105</td>
<td>Quarter 100</td>
<td>CMS 7004</td>
<td>UCC・COM 300-2</td>
</tr>
<tr>
<td>品目</td>
<td>SDI 4370</td>
<td>SDI 7400</td>
<td>SDI-4480</td>
<td>SDI-4480</td>
</tr>
<tr>
<td>質量</td>
<td>79GHz</td>
<td>56GHz</td>
<td>56GHz</td>
<td>56GHz</td>
</tr>
<tr>
<td>足高高</td>
<td>80GHz</td>
<td>70GHz</td>
<td>70GHz</td>
<td>70GHz</td>
</tr>
<tr>
<td>バンド</td>
<td>BCD EBCDIC</td>
<td>BCD EBCDIC</td>
<td>EBCDIC</td>
<td>EBCDIC</td>
</tr>
<tr>
<td>転送速度</td>
<td>9KC</td>
<td>9KC</td>
<td>9KC</td>
<td>9KC</td>
</tr>
<tr>
<td>文字長</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>文字効率</td>
<td>128</td>
<td>128</td>
<td>128</td>
<td>128</td>
</tr>
<tr>
<td>寸法</td>
<td>1.2m</td>
<td>1.2m</td>
<td>1.2m</td>
<td>1.2m</td>
</tr>
<tr>
<td>重い</td>
<td>15-22k</td>
<td>15-22k</td>
<td>15-22k</td>
<td>15-22k</td>
</tr>
<tr>
<td>丸の型</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
</tr>
<tr>
<td>半径</td>
<td>0.3m</td>
<td>0.3m</td>
<td>0.3m</td>
<td>0.3m</td>
</tr>
<tr>
<td>軽い</td>
<td>1.2m</td>
<td>1.2m</td>
<td>1.2m</td>
<td>1.2m</td>
</tr>
<tr>
<td>重い</td>
<td>15-22k</td>
<td>15-22k</td>
<td>15-22k</td>
<td>15-22k</td>
</tr>
<tr>
<td>丸の型</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
</tr>
<tr>
<td>半径</td>
<td>0.3m</td>
<td>0.3m</td>
<td>0.3m</td>
<td>0.3m</td>
</tr>
<tr>
<td>軽い</td>
<td>1.2m</td>
<td>1.2m</td>
<td>1.2m</td>
<td>1.2m</td>
</tr>
<tr>
<td>重い</td>
<td>15-22k</td>
<td>15-22k</td>
<td>15-22k</td>
<td>15-22k</td>
</tr>
<tr>
<td>丸の型</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
</tr>
<tr>
<td>半径</td>
<td>0.3m</td>
<td>0.3m</td>
<td>0.3m</td>
<td>0.3m</td>
</tr>
<tr>
<td>軽い</td>
<td>1.2m</td>
<td>1.2m</td>
<td>1.2m</td>
<td>1.2m</td>
</tr>
<tr>
<td>重い</td>
<td>15-22k</td>
<td>15-22k</td>
<td>15-22k</td>
<td>15-22k</td>
</tr>
<tr>
<td>丸の型</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
</tr>
<tr>
<td>半径</td>
<td>0.3m</td>
<td>0.3m</td>
<td>0.3m</td>
<td>0.3m</td>
</tr>
<tr>
<td>軽い</td>
<td>1.2m</td>
<td>1.2m</td>
<td>1.2m</td>
<td>1.2m</td>
</tr>
<tr>
<td>重い</td>
<td>15-22k</td>
<td>15-22k</td>
<td>15-22k</td>
<td>15-22k</td>
</tr>
<tr>
<td>丸の型</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
<td>20K, 50K, 100K</td>
</tr>
<tr>
<td>半径</td>
<td>0.3m</td>
<td>0.3m</td>
<td>0.3m</td>
<td>0.3m</td>
</tr>
</tbody>
</table>
| 品名 | SEACO COMPUTER-DISPLAY | Gaudio（作成者） | 日本電子産業
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>型番</td>
<td>SEACO414J</td>
<td>Beta COM T 0.0 L</td>
<td>JEM-3000/PT-150</td>
</tr>
<tr>
<td>線幅</td>
<td>5.512</td>
<td>5.512</td>
<td>9</td>
</tr>
<tr>
<td>線間駆動(点)</td>
<td>200,550,800,1500</td>
<td>550,800,1500</td>
<td>800 BP1</td>
</tr>
<tr>
<td>フォントセット</td>
<td>ANY(EBFDC1.BCD,ASCII1</td>
<td>ANY(EBFDC1.BCD,ASCII1</td>
<td>JEW輸入コード</td>
</tr>
<tr>
<td>驗速(RC)</td>
<td>30 KC</td>
<td>30 KC</td>
<td>265 KC</td>
</tr>
<tr>
<td>文字幅</td>
<td>真名,小文字,数字,等号</td>
<td>真名,小文字,数字,等号</td>
<td>真名,小文字,数字,等号</td>
</tr>
<tr>
<td>文字密度</td>
<td>n×4</td>
<td>n×4</td>
<td>125</td>
</tr>
<tr>
<td>スクリーンサイズ</td>
<td>56.55,89.55,105.0mm</td>
<td>56.55,89.55,105.0mm</td>
<td>105.70,35.16mm</td>
</tr>
<tr>
<td>フィルム厚さ</td>
<td>200,400,600,1090</td>
<td>200,400,600,1090</td>
<td>200 ft</td>
</tr>
<tr>
<td>フィルム配列</td>
<td>120.00/57.25/10mm</td>
<td>120.00/57.25/10mm</td>
<td>180 ft</td>
</tr>
<tr>
<td>基本単位</td>
<td>24.422</td>
<td>24.422</td>
<td>48.813</td>
</tr>
<tr>
<td>検索コード</td>
<td>VHC,LRC,CRC(自動製造記号)</td>
<td>VHC,LRC,CRC</td>
<td>Cut Mark, Image Mask</td>
</tr>
<tr>
<td>テザック方式</td>
<td>VHC,LRC,CRC(自動製造記号)</td>
<td>VHC,LRC,CRC</td>
<td>VHC,LRC,CRC</td>
</tr>
<tr>
<td>ソンジマーク</td>
<td>可</td>
<td>可</td>
<td>オプション</td>
</tr>
<tr>
<td>組積位置</td>
<td>ANY</td>
<td>ANY</td>
<td>組積位置</td>
</tr>
<tr>
<td>ベージャンクス(字数)</td>
<td>140×64,160×86</td>
<td>132×最大16 16行</td>
<td>132×最大16行</td>
</tr>
<tr>
<td>フォーム設定</td>
<td>ANY</td>
<td>ANY</td>
<td></td>
</tr>
<tr>
<td>記録速度(ページ/分)</td>
<td>15〜25ページ/分</td>
<td>15〜25ページ/分</td>
<td>最大20ページ/分</td>
</tr>
<tr>
<td>優先度</td>
<td>16位シンタテイ</td>
<td>16位シンタテイ</td>
<td>オプション</td>
</tr>
<tr>
<td>販売価格</td>
<td>5280円〜6200円</td>
<td>約15000円</td>
<td>312万円〜425万円</td>
</tr>
<tr>
<td>備考</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
COMの使用例をあげると表7-4のようなになっている（Auerbach社の調査による）。

### 表7-4

<table>
<thead>
<tr>
<th>項</th>
<th>業務範囲</th>
<th>分類</th>
</tr>
</thead>
<tbody>
<tr>
<td>銀行業</td>
<td>儲蓄払い等</td>
<td>受理</td>
</tr>
<tr>
<td>自動倉庫業</td>
<td>焼き出しスト</td>
<td>受取</td>
</tr>
<tr>
<td>宿題</td>
<td>自動倉庫</td>
<td>クリスマスリボン</td>
</tr>
<tr>
<td>金融機関</td>
<td>代理店</td>
<td>受取</td>
</tr>
<tr>
<td>関連機器</td>
<td>品質保証</td>
<td>介助</td>
</tr>
<tr>
<td>通信関連</td>
<td>マイクロフォード</td>
<td>受取</td>
</tr>
<tr>
<td>医療機器</td>
<td>エンジニア</td>
<td>技術者</td>
</tr>
<tr>
<td>広報</td>
<td>チャート、グラフ、テーブル、一般表</td>
<td>従業員、役職人、管理</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>項</th>
<th>業務範囲</th>
<th>分類</th>
</tr>
</thead>
<tbody>
<tr>
<td>銀行業</td>
<td>儲蓄払い等</td>
<td>受理</td>
</tr>
<tr>
<td>自動倉庫業</td>
<td>焼き出しスト</td>
<td>受取</td>
</tr>
<tr>
<td>宿題</td>
<td>自動倉庫</td>
<td>クリスマスリボン</td>
</tr>
<tr>
<td>金融機関</td>
<td>代理店</td>
<td>受取</td>
</tr>
<tr>
<td>関連機器</td>
<td>品質保証</td>
<td>介助</td>
</tr>
<tr>
<td>通信関連</td>
<td>マイクロフォード</td>
<td>受取</td>
</tr>
<tr>
<td>医療機器</td>
<td>エンジニア</td>
<td>技術者</td>
</tr>
<tr>
<td>広報</td>
<td>チャート、グラフ、テーブル、一般表</td>
<td>従業員、役職人、管理</td>
</tr>
</tbody>
</table>

- 189 -
(3) COMの構成

COMで文字を発生する方法として、CRT、電子ビーム、LED（Light Emitting Diode）、レーザなどが使われている。

① CRTによる方法

キャラクターによって文字を発生させ、これをマイクロフレームに写し取る方法をとっている（図7-12）。事務用COMの使用できる文字数は120種類であり、その文字の形成方法としてストローグ・ジェネレータによる場合が図7-13である。これはCRTの1文字分の区画内の25x35の格子からなる仮想マトリックス上に磁気テープから送られてきた電気信号に相当する文字が電子ビームのストロークによって描かれていく、その所要速度は約8.3マイクロ秒（每秒12万字）という速度である。ストローク方式の場合、特殊文字を使用したいときは、ジェネレータ回路の入れかえによって入れかえることができ、漢字プリンタとして使うこともできる。またストローク方式では電子ビームの大きさを電気的に変化することによって、標準文字に対する太文字を書いたり、マトリックスのY軸をずらすことでイタリック体を描くこともできる。
Data Source

Tape Drive

Magnetic Tape

Computer

Tape Drive Interface

On-Line/Off-Line Option

Computer Interface

Code Interpretation and Logic Circuits

Analog Conversion of Data to Deflection Currents

Deflection Coils

Cathode Ray Tube

(Magnetic Deflection)

Half-Silver Mirror

Form Slide

Strobe Light

Lens

Microfilm

図 7-12 CRT による記録
図 7-13 COM文字の一例
7-14はKWICインデックスの作成をイタリックス体を入れて見易くした例である。

図 7-15 はCRTによるCOMのハードウェアの構成例を示したものである。

② EBRによる方法

EBR（Electron Beam Recording）は電子ビームをドライシルバ・フィルムに露出して記録するもので、通常の銀塩の湿式のフィルムは使用しない（図 7-16）。この場合CRTのスクリー
電子ビームが乾燥式マイクロフィルムと相まって使用されるかのように記述されている。図7-16に示される。
図7-17 電子ビームによる記録
図7-18 メモレックス文字発生器

図7-19 ファイバ・オプチックスによる記録
FROM COMPUTER OR TYPE UNIT

COMPUTER OR TAPE INTERFACE

CHARACTER/GRAPHICS TRANSLATOR

CODE INTERPRETATION AND LOGIC CIRCUIT

LIGHT-EMITTING DIODE ARRAY

FORMS
FORMS FLASH
OVERLAY

FILM DRIVE: SYSTEM

CAMERA

RETRIEVAL CODE GENERATOR
図 7-21 レーザによる記録の例

−198−
インチ当たり合計180,000ビットである。これに対して8トラックの磁気テープは、1インチ当たり1,600ビットの場合は14,400ビットであるからこれはレーザの記録の12分の1にあたり。なお図7-22はレーザ・ビームによる記録の原則を示し、図7-23はPDR-5システムの概念図を示し、図7-24はPDR-5システムのハードウェア構成を示している。

(4) COM用ソフトウェア

COMによる出力は、単なる数字の便利さだけでなく見やすさ、使いやすさの点からグラフに表わすことが要求されてきている。そのためのソフトウェアがいろいる開発されている。たとえば、Canon社のソフトウェア・パッケージ“PAGE”は事務用グラフィックスの分野で使用されるもので、ラインプリンタとXYプロッタの両方の機能をもたせている。これはCOBOL、FORTRAN、アセンプラなどで使用でき、折れ線グラフ、棒グラフ、円グラフ、財務パターンなど描くことができる。出力例を図7-25に示す。

図7-22 レーザ・ビームによる記録
Simplified Schematic Of The PDR-5 Recorder/Reproducer

Hardware Configuration Of A Basic PDR-5 System
7.4 マイクロフィルムによる文書情報検索システム

マイクロフィルムとコンピュータを組み合せたマイクロフィルム検索システムはいろいろの構成があり、主なものをあげるとつきのようになる。

(1) ミニコンピュータを利用する場合

ミニコンピュータを検索装置の専用機として用いるもので図 7-26 のようになる。

(2) TSS端末を利用する場合（図 7-27）

TSS 端末からキーワードなどの必要な情報を入力して、TSS コンピュータに接続されているファイルを検索し、合致したデータからマイクロフィルムの番号をタイプライターに印字するものである。この番号に合ったマイクロフィルムを手動またはマイクロフィルムのキーをたたいて自動検索を行なう。

さらにマイクロフィルムを TSS 端末にオンラインで接続して行なう方法もある。

(3) 模写電信を利用してする場合

利用者はオンライン・ディスプレイで質問し、このシステムからは利用者と会話しながら必要な情報を探していく。必要な情報が得られたらマイクロフィルムのアドレスをたよりに原文を見る。遠隔地にマイクロフィルムがない場合には模写電信で送るようにする（図 7-28）。

(4) ビデオモニタを利用する場合

端末にはデータ・ファイルやマイクロフィルムをもたず、必要な情報はすべて中央のコンピュータから送ってもらうシステムである。このシステムのファイル管理はすべて中央で行なうので、端末での運用管理は楽になる。
図7-26 ミニコンピュータを利用する場合

図7-27 TSS端末を利用する場合
図7-28 検索電信による場合
8 今後の課題
8 今後の課題

これまで一般に考えられてきた情報処理システムでは、所定の理論式や経験則のロジックにしたがって導出した数値や記号を、所要のフォーマットに整理整頓して作成したテーブルやグラフそのものが情報であるという概念をよりどころとしており、そこにコンピュータを活用することが情報処理技術の基本であると規定しているように思われる。

しかしながら、経営情報システムの形成という観点に立って、これを単純にトップ・マネジメント・レベルの情報システムとしてではなく、あらゆるレベルの管理者も利用可能できるような情報システムに拡張するためには、単純な数値や記号による情報、すなわち定形的な数値情報や記号情報のほかに、各レベルの管理者が数値情報や記号情報に対して自己の判断に基づく処理をとくとして文書の形式に加工した情報をやりとりできるようにする必要がある。

このように、数値や記号が情報として有効であったのは、情報システムが単一レベルであるためきわめて限られた範囲で利用されていたからであって、情報システムの利用が拡がるにつれて柔軟性のない数値や記号だけによる情報では、むしろ情報提供の相手に正確な意思を伝達できないばかりでなく、時によっては、情報提供の相手に重大なミスにつながる誤解を与えてしまう恐れもある。

少なくとも、経営情報システムにおいて、トップ・マネジメントから、オペレーション・レベルの管理者にいたる、あらゆるレベルの管理者間で情報を伝達する過程においては、管理者が適宜に自己の考え方や判断を文書の形式で情報システムに入力できるようにすることが必要であり、いわゆる文書情報処理に関する研究・開発の重要性は今後ますます高まる傾向にある。

48年度に進めた文書情報処理に関する研究事業では、カタカナ化盤によって情報を入力し、漢字かな混り文で出力するための基本的なシステム設計を
考察したが、比較的簡単な操作できる入力サブシステムを開発することにより、日本語情報処理システムの普及にともなって予想されるデータ・エントリの要員不足を解消するという、この研究事業の目的の1つは明確に見通しを得たと結論することができよう。

今後の課題としては、文字による情報の伝達や記憶に大きな利点をもつ、この日本語情報処理方式を発展させて、官公庁における通達文書、出版社における初稿、生産企業における製品マニュアル、病院におけるカルテといった文書情報の作成に利活用できるようなトータル・システムを実現するために、それぞれ専門分野別のカタカナ・漢字単語テーブルを整備・拡充し、入力システムに使用する周辺機器を開発することである。
禁 無 断 転 載
昭和49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 有限会社 秀 崎 社
東京都渋谷区代々木2丁目26番地
TEL（379）1816

48-S004