48-S 004

文書情報処理に関する調査研究

昭和49年 3 月

JIPDEC

財団法人　日本情報処理開発センター

JIPDEC
48
S004

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なう必要があり、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４８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欧文字、数字またはカタカナで記述されていたために、とかくあいまいであるとされてきた文書情報を、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と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った、若松清司、上田陸奥夫、船崎武男、矢田光治、高森寛、杉本太郎、後藤榛男、野添篤毅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

目　　次

1. 総　　論 …… 1

1. 1　研究の目的 …… 1

1. 2　報告書の概要 …… 7

2. 日本語情報の分析 …… 9

2. 1　漢字の役割り …… 9

2. 2　漢字かな混り文出力の必要性 …… 14

2. 3　固有名詞処理に対する需要 …… 19

2. 4　公文書処理の考察 …… 23

2. 5　機械化の諸問題 …… 29

2. 6　適用分野 …… 31

2. 7　技術動向 …… 39

3. 日本語情報処理のためのカタカナ入力方式 …… 55

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 …… 55

3. 2　カタカナ入力方式へのアプローチの方法 …… 59

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 …… 62

4. 固有名詞処理のためのカタカナ入力方式 …… 83

4. 1　固有名詞の構成要素 …… 83

4. 2　固有名詞の変換テーブル …… 84

4. 3　固有名詞の漢字分析 …… 89

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、日本語情報をカナけん盤で入力し、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。そこでは、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目した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、変換率がたとえ低くとも、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太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、カナ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こ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、トップ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結論にもとづいて昭和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、一般の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な問題があることを究明した。このようにわが国においては、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験的考察を行ない、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（Ｃｏｍ－ｍｕｎｉｃａｔｉｏｎ）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」であると答える以外にないかもしれない。しかし、それではことがらや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。それにしても、日本語においては、情報という言葉が便利なためにか、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」にあたるといえよう。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の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あり、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している。また、伝達される内容は知識であれ何であれ、記号系列に変換され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には違いない。

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、フィーリングとか匂いなどでもよいということになる。もっとも、フィーリングや匂いであっても、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ために情報というものを論ずるときには、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と、②意思決定における判断の素材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、数式などのように論理的に厳密なものもあれば、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も共通理解が双方の間で明確になっ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は伝わらないか、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ぎ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communis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5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

る。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、互に向いあってい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、受け手の方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、知識の伝達という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、それが受け手になんらかの反応を惹き起し、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か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うたわれているが、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て、だれが、なにをするために、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ければ、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て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しかねない。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調査研究の成果として、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、各章でそれぞれつぎ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需要の高いといわれている固有名詞についてその構成を概述し、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 漢字の役割り

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た。

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ナルな日常業務の処理、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なか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、利用するかが各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、文書情報を扱う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）が行ないやすく、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ことなどがあげられる。

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いる。それは、従来の情報処理が、処理面の効率追求にあったためであり、この追求心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、漢字かな混り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なうための道具であるから、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り文による情報処理が当然要求されてくる。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が多いため、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り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、購買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、このようなかな文字（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な混り文で表現されており、生活環境のなかには、漢字かな混り文が、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、かな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、かな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）となったときに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、単語としての音をそなえている。

そのため、かな文字は、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。かな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恐れがある。

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。

このことは、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、出版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除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かな混り文が、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 漢字かな混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、人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。

⑤　ハードウェアを取扱う熟練技能者の教育訓練が十分にできないため、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、処理時間の速い、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語のけん盤を数千語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、表意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音文字に比べて

①　文字の字数が非常に多い。

②　字形が複雑である。

③　一定のルール（英字の場合のＡＢＣ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きない。

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

④ 読　み　方

⑤ 同字異体

などをどうするかが問題である。現在のところこれらに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、各メーカごとに独自のビット構成をとっており、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 漢字かな混り文出力の必要性

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、ＥＤＰ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り文出力の問題（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（タイトルや項目の説明など）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。漢字のもつ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り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り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、計画、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、その情報にもとず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

行為段階	行為要素
発生段階	（受ける） 受付・受理 受入・戻入・振替 受信 受領・収納 購入 検収
整理段階	（書　く） 記入・記帳・記録・消込 筆記・速記・転記 トレース・写書・浄書 タイプ・印刷 起案 （計算する） 計算・算出・集計 検算・精算・決算 （調べる） 調書・再書・下調 照合・対照・校正・読合せ 照査・精査・点検・診療 測量・測定 検査・試験・選考 検収 審査・審議
整理段階	分析・研究 （作　る） 作成 作票・作表 調整・調剤 （集める） 収集 編集 （分ける） 分配・配分・割当 仕訳・分類・分析 （整える） 整理・整頓・整備 準備 綴込・切抜・添付・挿入 移管 修繕 清掃 （話　す） 連絡
計画段階	（考える） 計画・企画 発案・立案・考案 見積 設計

行為段階	行為要素
計画段階	調整
	研修
	検討
	研究
	（話す）
	交渉・打合・折衝
	討議・合議・会議
	査定・算定
決定段階	（決める）
	決定・専決・決裁・採択
	承認・確認・認証
	評定・認定・算定・決算
	指定
	検印・捺印
	契約
実施段階	（伝える）
	命令・発令・指令・指示
	監督・視察
	立会・督促
	管理・統制

行為段階	行為要素
実施段階	司会
	（送る）
	送付・返送・発送・運搬
	（出す）
	発信・呼出
	払出・振替・支払・納付
	発行・発表
	提出・出願・届出・掲示
	派遣・出席・参加・出張
	（配る）
	交付・回付・回覧・供覧
	配付・配布・配給
	（知らす）
	通知・通牒・通告・通達・社達
	報告
	回答・応答
	申告・上申
	（整える）
	保存・保管・格納

がなされる。しかしこれらの各段階における行為は、ほとんど個人のなかだけで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。

このように受け取った情報を正しく理解し、それにもとずく意思決定を正確に他の人に伝達するために、浄書をしたり、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（漢字かな混り文）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にも漢字の問題が取り上げられ、漢字用の入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・購買・在庫などの台帳作成

⑩　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ＥＤＰ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ざわざ人手をかけて漢字に書き直したり、タイプ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。

また人名簿や見積書、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いけれども、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、大量事務処理の範疇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。

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５００万1,２２０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、あて名印刷機を用いたり、あるいは手書きによるものの比重が圧倒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。

表 2.2 漢字による固有名詞処理システムに対する需要の見通し(試案)

省庁名(順不同)	適用業務	使用する固有名詞の種類					需要のタイプ	業務内容の例示	備考
		地名	人名	企業名	民間事業所名	官公署名			
厚生省	厚生年金保険・健康保険・児童手当	○	○		○		⊙ (注)	保険料徴収事務，裁定事務，支給事務	
労働省	失業保険・労働者災害補償保険	○	○		○		⊙	同上	
警察庁	自動車運転免許管理	○	○				⊙	運転免許証の発給，違反通告，免許の停止・取消し通告	
運輸省	自動車登録	○	○	○	○	○	⊙	自動車登録証の発行，手数料徴収，重量税の徴収，リストの作成	
国税庁	源泉所得税徴収	○			○	○	△	受領通知の発行	
〃	申告所得税徴収	○	○	△			◎	申告用紙の配布，納入告知書の発行，督促通知，受領書の発行	
〃	法人税徴収	○		○			△	同上	
都道府県	事業税徴収	○	○		○		△	同上	
〃	自動車税徴収	○	○		○		⊙	同上	
市区町村	軽自動車税徴収	△	○		○		△	同上	
〃	住民税徴収(申告分)	△	○		○		◎	同上	
〃	源泉住民税徴収	○	○		○	○	△	納税額の通知，督促，受領通知	
〃	固定資産税徴収	○	○	○			○	同上	
〃	国民年金	△	○				△	同上	
〃	国民健康保険	△	○				△	同上	
〃	選挙公報	△	○				○	選挙人名簿作成，投票所入場券発行	

(注) 需要のタイプ欄の⊙は現行システムからの移行が比較的容易，◎は現行システムの拡充と相まって効果の発揮が期待されるもの，△は実施までにやや長期間を要するとみられるもの，を示す。

ユーザ	適用業務	使用する固有名詞の種類					需要のタイプ	業務内容の例示	備考
		地名	人名	企業名	民間事業所名	官公署名			
市区町村	住民福祉	△	○				○	米穀通帳、就学通知、児童手当、共済事業、各種表彰者の名簿作成、通知書の発行	
地方公営企業	水道事業、下水道事業	○	○		○	○	○	納入通知書および受領書の発行	
電電公社	料金徴収	○	○		○	○	⊙	同上	
NHK	同上	○	○		○		○	同上	
電力会社	同上	○	○		○	○	○	同上	
ガス会社	同上	○	○		○	○	○	同上	
郵政省	簡易生命保険	○	○				△	同上および所得控除証明書発行	
生命保険会社	生命保険	○	○				⊙	同上	
損害保険会社	火災保険・自動車保険・傷害保険	○	○		○		⊙	同上	
農業共済組合（共同または委託）	農業共済	○	○	○			△	名簿の作成、引受通知書、納入通知書、支払通知書等の発行	
各種共済組合	年金事業、医療給付	○	○		○	○	○	名簿、組合員証、年金証書、支給通知等の発行	
農業協同組合（共同または委託）	農業事業全般	△	○		△		△	組合員名簿、共済事業の受払、信用、経済等の各種事業の受払通知	
新聞・出版・放送企業	新聞・出版・放送事業	○	○	○	○	○	⊙	原稿の送受信、編集、版下作成、記事の保存、電光ニュース・字幕の編集	
大学等	学籍管理等	○	○				△	学籍簿、成績簿、授業料徴集、成績証明書発行等	
人企業	株式業務	○	○				◎	株主名簿の作成、総会組集、配当支払等	
各種企業	対顧客サービス	○	○	○	○	○	⊙	顧客名簿作成、ダイレクトメールあて名記入	
官公庁および情報企業	統計調査・世論調査・市場調査	○	○	○	○	△	○	母集団または個別的な標本の名簿の作成、訪問通知、調査票上の回答者欄の記入	
市電公社	電報	△	△	△	△	△	△		

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、にわかに脚光を浴びるようになってきた。そして、日本語の漢字単語、文法および構文の研究をはじめとして、漢字入出力装置の開発、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解明や、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究明したがここでは、各種の文書の中でも、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ともな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（請願法第２条、政治資金規正法第６条第１項等）。

訴訟法上では、物体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、布ぎれ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でき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体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）の集りをいい、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

になる。

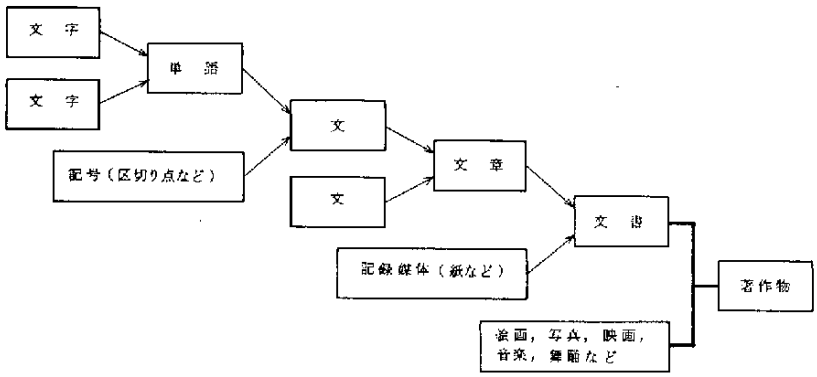


図２－１　　用語の相互関連

⑵　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、「私文書」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、同趣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＞に対する」（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１２月９日の次官会議申合せによる「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、当用漢字表（昭和２１年内閣告示第３２号）の「まえがき」に見受けられる。なお、これらに類似する用語としては、「公用文」というのがあり、現在でも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、ほぼ大同小異で、これらの考え方が、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かどうか。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」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、個人間に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

公文書と私文書とを対比すると、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、つぎ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 作成権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 階層構造をしている。

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っている。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、相互の関係の概略を図示すると、図2－2のとおりである。

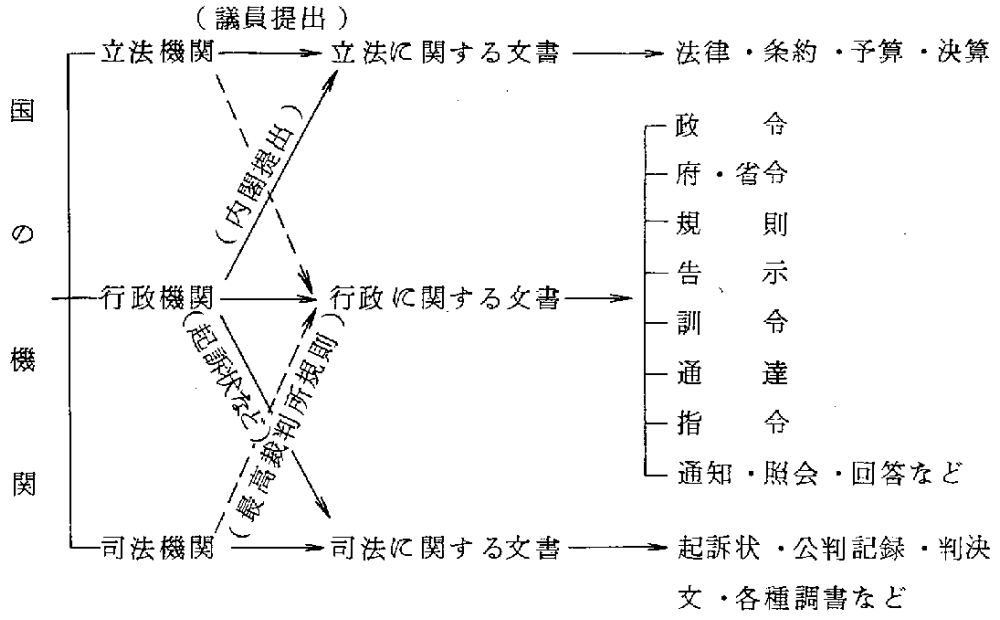


図２－２　国の機関と公文書の種類

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、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である。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（ひらかな、カタカナ）で発音どおりに、「かんじ」と書くな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回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て「かんじ」と打つなら、指先のタッチは４回で済む。このように日本語文書を清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を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。このかな文字を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かいると聞く。

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ずらさがある。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、まちまちの送りがな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、字の上手な人に清書してもらうか、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労力と時間を要する。清書するといっても、直筆の文字には必然的に個性的な特徴が混入し、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、文書がもたらす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凶といえば、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。字種が多いばかりでなく、一般の人にははっきり何種類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。送りがなに関するルールも不明確きわまりなく、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、分かち書きをしないことが多いからますますとまどう。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（漢字かな混り文出力）の適用分野として、人事・販売・購買・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・官庁提出用資料など限定された分野の情報や業務がある。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漢字が要求されるものや、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、いろいろな視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ぎらわしいもの。

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。

しかしその反面、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

③　漢字を処理するためのデータ・パンチやオペレーションのための要員の教育

など新しく発生する費用が、すでに導入されているコンピュータの費用に、ほとんど置き代ることなく加わって来る。日本語情報処理の問題は、適用面におけるメリットと、漢字を処理するために新しく発生する費用や要員の問題（一般に従来のパンチャやペレーションの作業に比べて複雑で高度の熟練を要するので、要員の確保や教育に問題がある）などデメリットな面とを十分検討したうえで、その導入を決定する必要がある。

以下に使用目的から見た適用可能分野の例を挙げる。

2.3.1　日本語文章の情報検索

今日のように企業を取りま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上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なわなければ、相当大きいむ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ＥＤＰ化したり、全社的な立場にた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を管理したり、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人達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 情報のアブスト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。

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、漢字けん盤にない文字や漢字出力装置に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、それらの業務を通し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・家族構成や考課など従業員各人のデータも人事検索システムを通して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人手によっておぎなったり、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
- 読みにくい
- 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人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、人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・営業で利用するダイレクト・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

など名種台帳類や伝票類の作成・発行が考えられる。

このようにコンピュータによる日本語情報処理は、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。

2.3.3 伝票の発行

現在、企業においては、販売・購買・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りをはたしている。

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ること、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できること

などが要求されている。

そのため、伝票処理をＥＤＰ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、人間の側にその譲歩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ＥＤＰ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の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き、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（漢字かな混り文）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性
- 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わ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

などの問題から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

１ビット	２種	７ビット	１２８種
２ビット	４種	８ビット	２５６種
３ビット	８種	９ビット	５１２種
４ビット	１６種	１０ビット	1,０２４種
５ビット	３２種	１１ビット	2,０４８種
６ビット	６４種	１２ビット	4,０９６種

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字分で漢字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、漢字の数がひじょうに多いといっても、実際にはそれほどめんどうな問題ではない。また、漢字のもつ有意性は、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いえば、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（日本語表現の漢字：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。この比較は専門のパンチャやオペレータがパンチした場合のことで、だれもが容易に入力操作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。

いっぽう、漢字かな混り文出力については、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、見通しはあかるい。しかしながら、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たものから、文章（日本語）を扱か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、身近にせ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ぎ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、②ＯＫＩ型けん盤さん孔機がある。

① けん盤キー……１９２個

シフトキー……　１２個

＜収容文字数＞　2,３４６字

※右手でけん盤キーを、左手でシフトキーを操作する。

② けん盤キー……６５０個

シフトキー……4段シ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、両足でシ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

表２－４　　漢字入力装置一覧表

メーカ	機　　種	文字種	コ　ー　ド	出　　力	概　　要
高千穂	T4112 漢字けん盤 ディスプレイ レコーダ	3,072		磁気 テープ	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
	T4012 漢字けん盤 レコーダ	3,072		カセット ・テープ	入力文字をカセットテープに書込む
	T4011 漢字けん盤 さん孔機	2,496	８単位２列	紙テープ	紙テープに漢字コードをさん孔する
沖電気	漢字けん盤 さん孔機	2,600	６単位２列 ８単位２列	紙テープ	
ＩＢＭ	漢字けん盤機構付２９さん孔機	3,600	２行 カード コード	カ　ー　ド	通常のカタカナ付カードさん孔機としても利用できる
富士通	FACOM 6801A 漢字入力 キーボード	3,840	８単位２列	紙テープ	オプションボードを付加すると、5376字になる。
谷村新興	SCK－200－D－54 漢字けん盤 さん孔機	2,688	６単位２列 ８単位２列	紙テープ	リーダの 接続可能
日　電	C－５１１０	2,145	８単位２列	紙テープ	邦文タイプライタの変形でワンタッチ方式である
東　レ	TORAY－ ８５１０	2,240	８単位２列	紙テープ	

千穂交易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した谷村新興製作所のＳＣ－４００型、和文タイプライタを利用した日本電気の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。

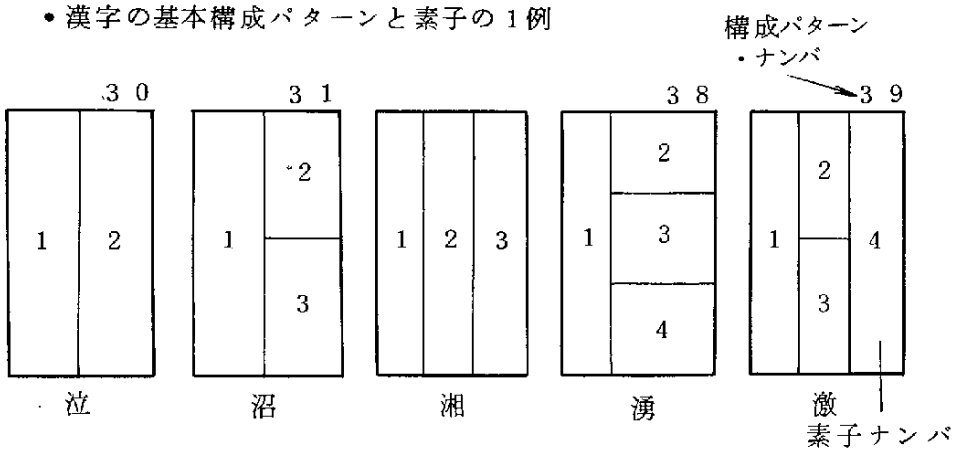
原データの作成については、原稿をみて漢字入力装置から入力するだけであるから問題はないが、問題なのは訂正処理を行なうための訂正データの作成である。英数字やカタカナ入力装置の場合にはベリィ（検孔）という方法があるが、漢字入力装置の場合は、能力的、コスト的にみてもなかなかむずかしい。そのための１つの方法としては、モニタ用紙に出力して印刷物校正の方法と同様に訂正することがある。ただしその場合には、部分訂正をするか、全データの打直しをするか、あるいは行送りはどうするかなどソフトウェアの処理が複雑になり、ただ単に追加、訂正、削除だけにとどまらない。いわゆるいくつかの編集業務を必要とする訂正処理の解決策として、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2,５００字に対

表２－５　漢字ディスプレイ一覧表

	沖　電　気	高　千　穂	日本電子産業	電　総　研	日　　電	東　　レ
文字表示形式	ドット	ドット	フライング・スポット	ドット	高解像度ブラウン管	ドット
最大表示字数	３２字×１６行（横）	６４字×２行（横）	32字×8行（横） 16字×32行（縦）	４０字×１０行		88字/行×66行 20字/行×15行
文　字　種	2,５７８	3,０７２	2,４９６	1,1４３～2,５００		
バッファ・メモリ	５１２字	１２８字	９６０字	８００字		
プ　リ　ン　タ	接　　続		接　　続			接　　続
そ　の　他	ライトペン使用（位置指定）	最大32台まで接続	種々のシステムとのインタフェイス	大型プロジェクト開発の一環	ライトペン使　用	

• 漢字の基本構成パターンと素子の１例



• 漢字のパターン入力方式の１例「湘」

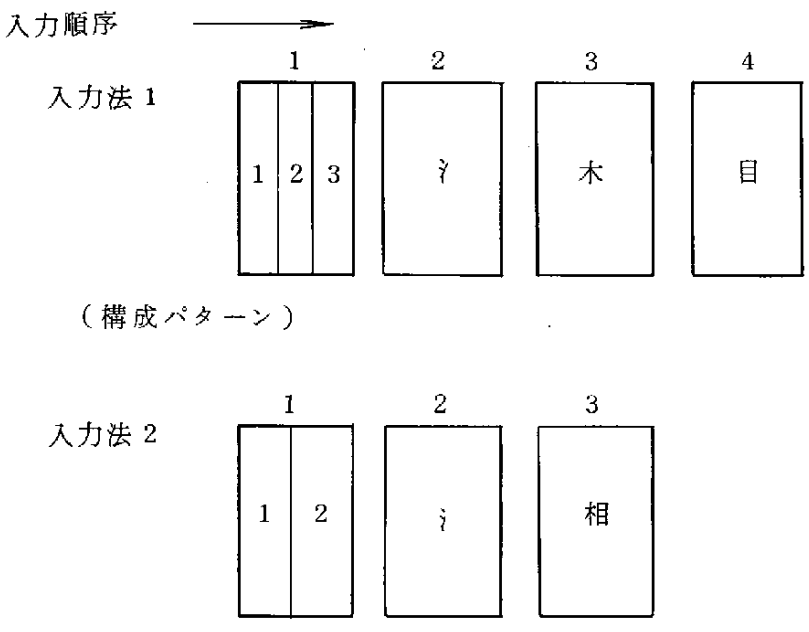


図２－３　漢字入力方法の例

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なう方法がある。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は漢字入力に関して苦心しており、い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ゲ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、文字品質などについ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述することにする。

(1) モニタ用の出力装置

モニタ用には、つぎ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、プリントすべき字種が入力装置の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、漢字かな混り文出力には欧文のように英語なら２６文字ロシア語なら３２文字と上限がはっき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、プリント機構は、やはり縦１５、横１８のピン・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、記憶装置がそれらに対応する文字パターンを探し出し、それをピン・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、これはピ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を応用したものであり文字を縦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

メーカ	日本電子産業	高千穂交易	富士通	沖電気	日本アイ・ビー・エム	日電	谷村新興	東芝
印字方式	静電記録方式	静電記録方式	静電記録方式	静電記録方式	ワイヤ・プリント	ワイヤ・プリント	活字打鍵	多針電極
用紙	静電記録紙	静電記録紙	静電記録紙	静電記録紙	通常の連続用紙	普通紙	普通紙	静電記録紙
印刷速度	12,000～24,000（字/分）	72,115(字/分)	600行/分	1,240行/分	300行/分 or 333行/分	100行/分	180行/分	107/600
文字パターン	フライングスポット	32×32ドット/分	15×18ドット/分	18×18ドット/分	18×22ドット/分	18×18ドット	活字ドラム	24×22ドット
文字種類	10,000	2,048～8,192	2,400 4,800 7,200	～2,600	～10,000	3,960	4,096	4,096
文字サイズ	2.2×2.2㎜～4.5×4.5㎜の8種	2.5×2.5㎜ 4.5×4.5㎜の4種	3.8×4.0㎜	6×6㎜	最大 4.6×5.8㎜	4.23×4.23㎜	12P	4.2×3.84㎜
コンピュータとのインタフェイス	オンライン可能	オンライン	FACOM230-25/35以上の機械とオンラインで稼動	ミニコン制御中・大型とのオンラインも可能	IBM360/30（32K）以上の機械とオンラインで稼動			
特徴	＊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 ＊種々のシステムとのインタフェイスが可能	＊ドットの数が多いため種々の文字を作成できるまた印字結果をオフセット印刷の原稿として使用	＊ＯＣＲ文字をプリントすることができるためターンアラウンドを可能にする	＊ＣＲＴディスプレイ上に表示してからプリントするため、簡単に訂正してプリントが可能	＊マトリックス文字発生プログラム・パッケージにより漢字の種類・字体などの変更ができる ＊完全にコンピュータのラインプリンタとして設計されている	＊多針電極による静電ドット記録方式であるため、普通紙を使用出来ること、スピードが速いことなど コストが安い	＊漢字タイプライタとオンラインになっている。モニタの役目をする	＊コンピュータとのオンライン可
基本構成	＊プリント制御装置 ＊文字発生装置 ＊プリンタ ＊入力機器(PT) JEM3100システムの中のプリンタ	＊磁気テープ装置 ＊フォント・メモリ（文字発生装置） ＊プリンタ	＊プリンタ制御装置 ＊リード・オンリー・メモリ ＊コンピュータ ＊プリンタ	＊サブ・コントロール ＊メイン・コントロール ＊ＣＲＴ ＊プリンタ ＊ミニコン	＊プリンタ ＊コンピュータ	＊プリンタ		

表２－７　　電算植字システム性能表

メ　ー　カ	写　　研	富　士　通	日本電子産業	R　C　A	I　B　M
印字方式	フラッシュランプ露光	フラッシュランプ露光	C　R　T	C　R　T	C　R　T
印字速度	９００／分	１２００／分	１０００００／分	４２００／分	13500(英文8ポ)
文字発生方式	字母光学式	字母光学式	フライングスポット	フライングスポット	フライングスポット
文　字　種	６４００字	１２７６２	２００００	４００００	外部記憶装置による
文字サイズ	９～18級　６種	６～14Ｐ　８種	４～２４Ｐ	４～９６Ｐ	４～１８Ｐ
特　　徴	文字変形たて よこ可 版下フィルム出力 10インチ×100フィート	〃 〃 ８インチ×100フィート	〃 モニタプリント可 〃 ２５０mm巾	〃	〃 9.44インチ巾
コンピュータ	ＨＩＴＡＣ―10	FACOM 270―20	ＨＩＴＡＣ―10	ＨＩＴＡＣ８４００	ＩＢＭ360―40
入力装置	紙テープ	紙テープ	紙テープ	紙テープ	紙テープ
出力装置	〃	〃	磁気テープ	気テープ	磁気テープ

(2) 電算植字用の出力装置

電算植字というのは、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。前述したモニタ用のプリンタは、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植字の場合は、出力がそのまま印刷物の版下となることから、つぎのような要求・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たり、ドット方式では、点が目に見えるようなものは不適当である。そのため、電算植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、

Ⓐ 光電式 ―電子式

Ⓑ 全電子方式

がある。

しかし電算植字は、雑誌、新聞など1回印刷して役目を終わるようなものには非効率であり、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は効果的である。

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

パ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。

○ノンブル柱処理

割付が決まった段階でページ、柱をつける。

(2) 棒　組　調　整

○ジャスティフィケーション処理

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。

○禁足、分離、禁止処理

行末、行頭に、カッコ類、感嘆符、疑問符、半音などが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、

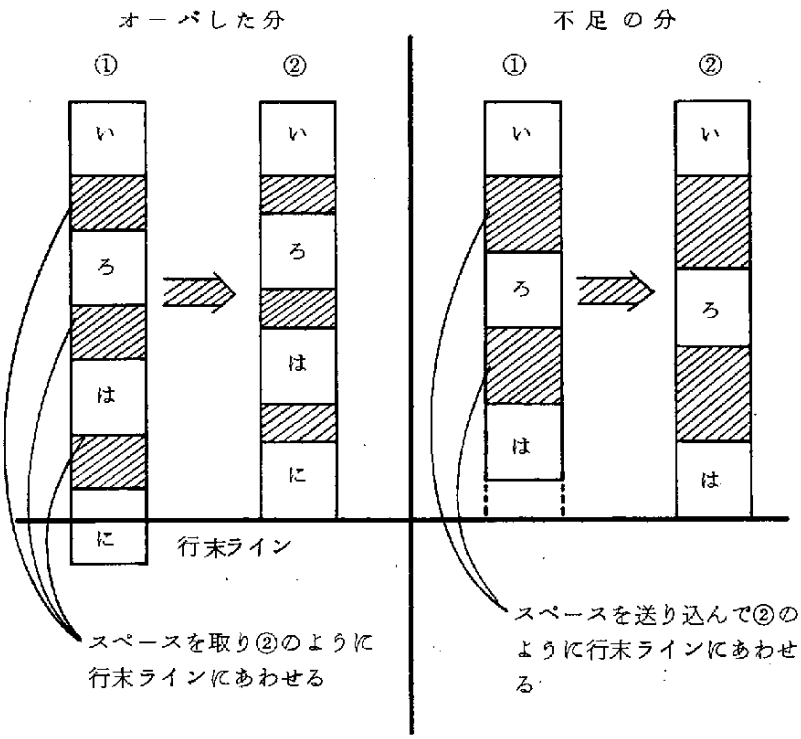


図２－４　行末処理の例

中央揃え、などが可能である。

○ハイフオネーション

英字の単語が２行以上にわたる場合、区切りにハイフオンを挿入する。

○そ　の　他

合成、数式などの処理を行なう。

(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・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か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

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増進され、その適用分野の急激な拡大がめ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きていると思われる。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、英語を自国語としない他の国々に対して、コンピュータによる自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、この意味からも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3,０００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

は大きな役割りをはたさねばならない。

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、取扱い上の人間の問題、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ずいた配列

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ずいた配列

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ぎ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から解放される。（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、

ⓐ 同音異義語の処理

ⓑ 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、ⓐの同音異義語の処理は入力装置を改良したり、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、ⓑ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、たとえばつぎ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

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き、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ぱんに要求される。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・訓読の取り扱いや送りがな、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、漢字植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文変換の処理は、きわめて困難な問題をかかえていることは前述してきたが、カタカナから漢字への変換を１００パーセント完ぺき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。たとえば、変換処理のすべてをコンピュータで片付けてしまおうとしないで、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。入力の段階で、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、分ち書きしてないカタカナ・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強いことを、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定するのはかなり難しい。言葉づかいがていねいかどうかなどで、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、もしそれがうまくいっても、「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断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、その負担の軽減ははか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、それを入力者がみて、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「徴候」は同音異義語とはいえない。どちらかに統一すべきであろうし、そのほか、「記念・紀念」、「出会う・出合う」なども、どちらかに整理しても、日本語はいっ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、貿易がしんてんするときは「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」とかかねばならない。そのほか、自然現象を「観照」するとかく、また、ほめてはげますときは「勧賞」するとかき、感心してほめるときは、「感賞」するとかく。しかし、すすめはげますときは「勧奨」とかくことになっている。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、「かんしょう」という単語には、「干渉」、「感傷」、「環礁」、「緩衝」、「管掌」、「癇性」、「冠省」、「簡捷」、「奸商」、「姦商」など合計１６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」、「謀る」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、まぎら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「かんしょう」、「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。「かんしょう」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、ひらがなどのファンクション・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式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ぎ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しながらエントリできる。
- 漢テレ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に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ぎ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に行なわせることによって、大形コンピュータの増設より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せている。

(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・パッケージの完備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) 端末は、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、入力データのチエックおよび編集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) 多くの文字を印書可能とする。

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。

○オフライン・システム

① カタカナ入力キー・ボード

付加機能 { ページ・ライタ / ディスプレイ（同音異義語表示用）

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 磁気テープ装置

⑥ 中形汎用コンピュータ・システム（メイン・コンピュータ）

○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ぎ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

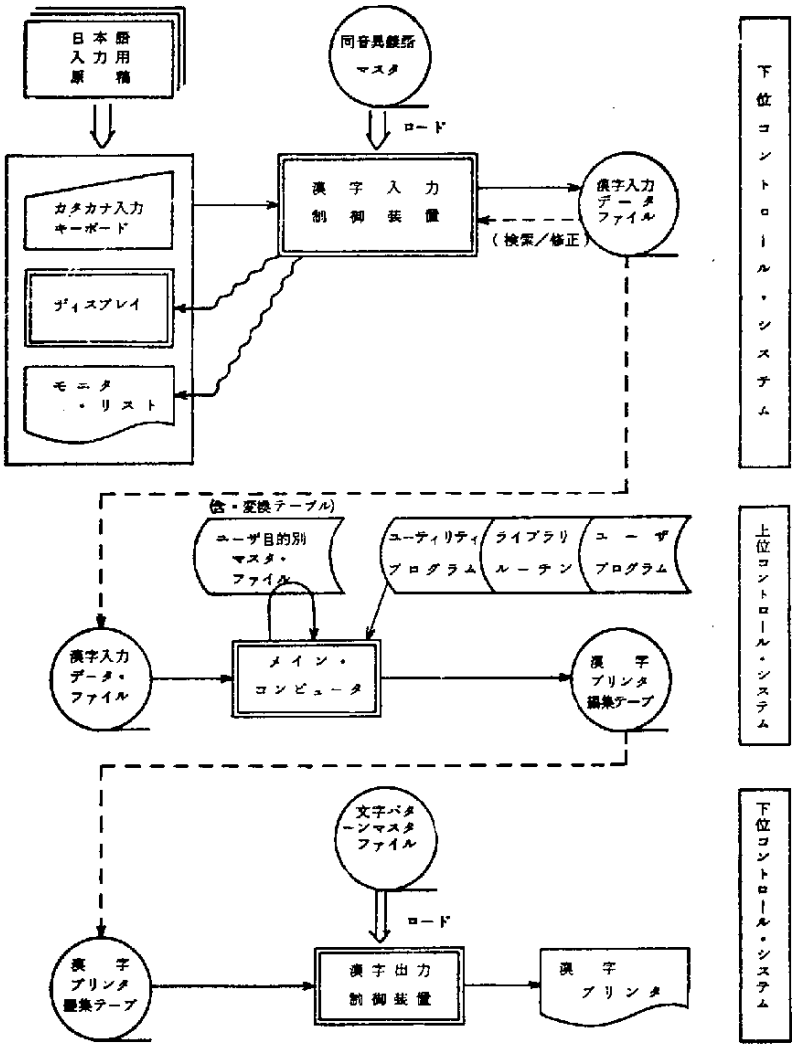


図 3－1　システムの構成－(1)オフライン処理

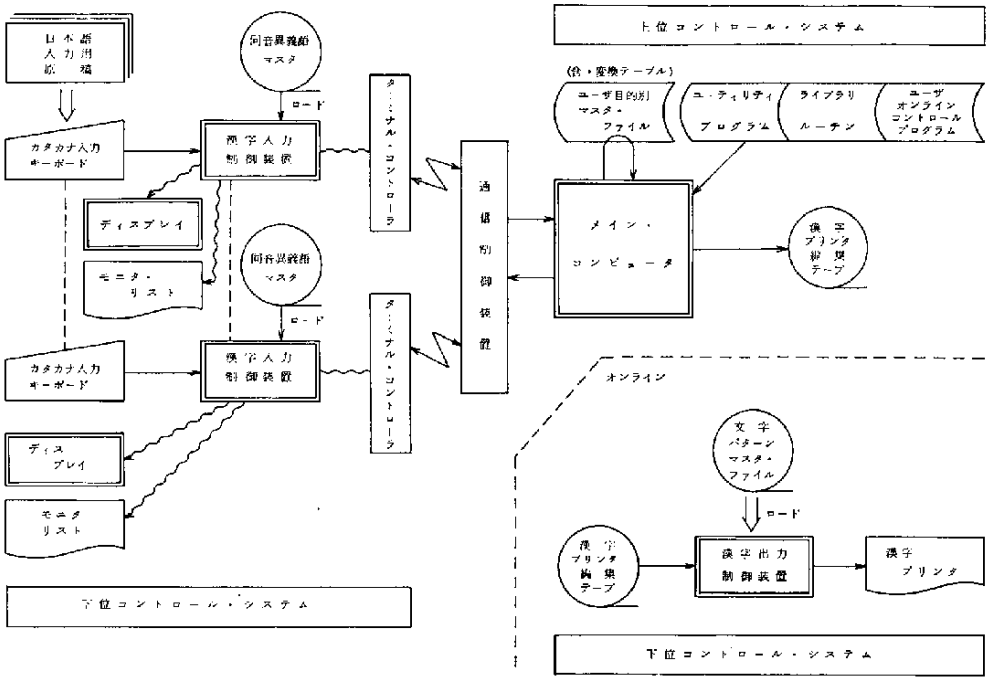


図３－２　システムの構成－(2)オンライン・インライン処理

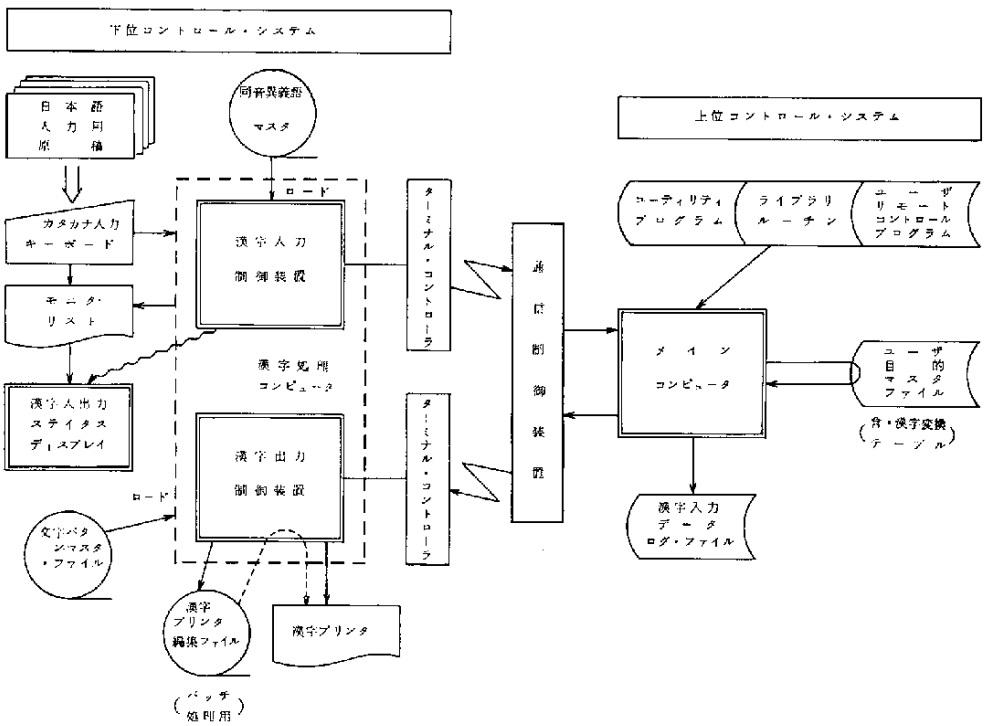


図３－３　システムの構成－(3)リモート処理

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。

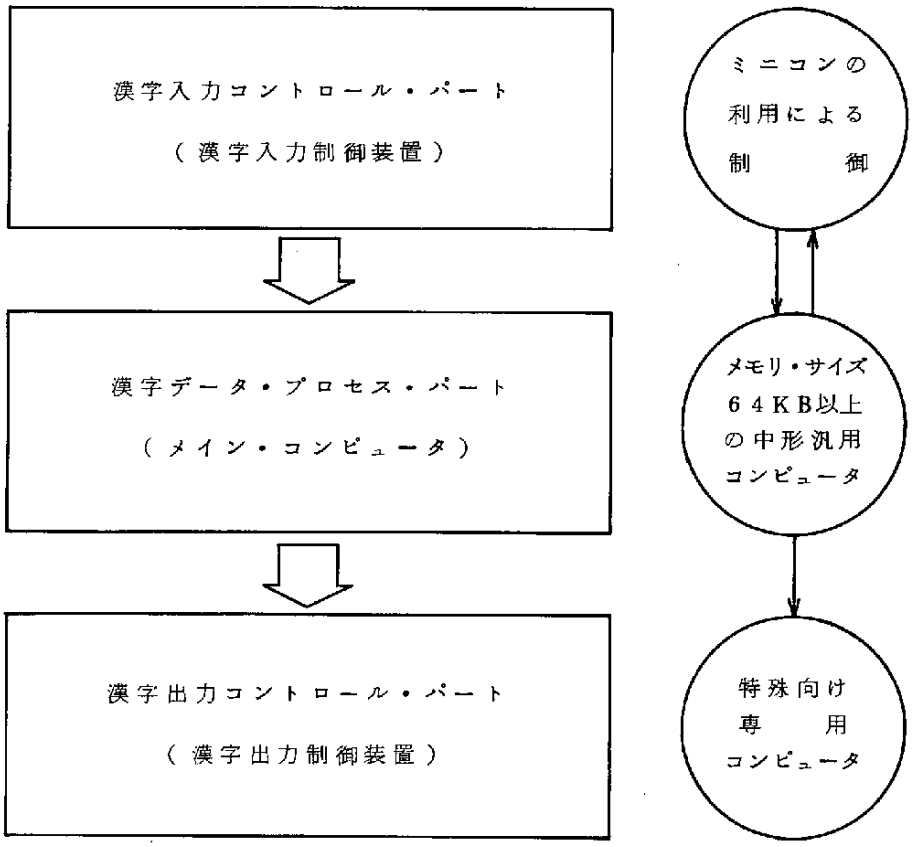


図３－４　システムの総合機能

以下に各パートのハードウェア仕様および構成を概述し、データ・プロセスの主な機能を、ソフトウェア・オリエンテッドに考察する。

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。

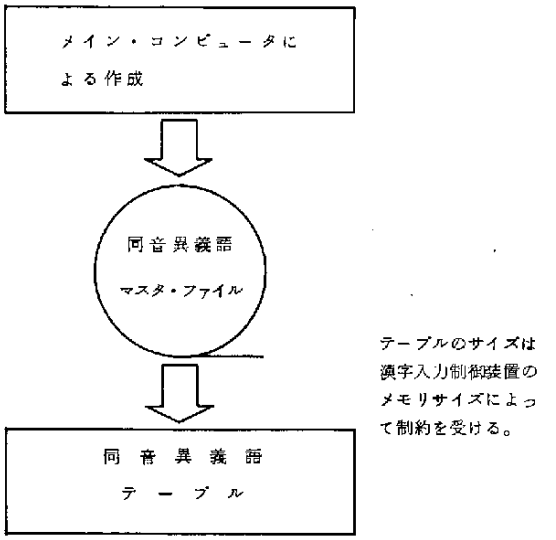


図３－５　同音異義語テーブルのイニシャライズ

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、入力イメージのチエックを行ない、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エ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・コンピュータにあま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、プログラマブル・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う。

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編集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・ファイルに属している漢字変換テーブルの記憶構造によって、変換処理効率が大巾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るので変換対象コード群のメインテナンスはすでに準備されているユーティ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 Identifire＝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ＦＩＤ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る。

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ＩＳＡＭ（ 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、つぎのような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なえばよい。

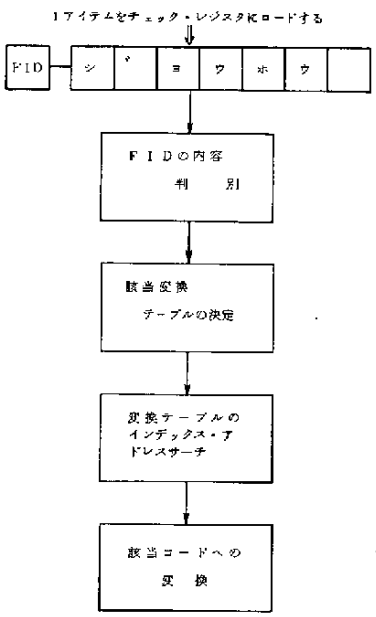


図３－６　非階層構造の変換テーブルによる変換

(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すためには、

- 同音異義語　　} 下位コンピュータ
- 専門用語
- 専門的一般語（使用頻度が多い一般語）　} 上位コンピュータ
- 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ぎのようにデータを分割しレベル化することができる。

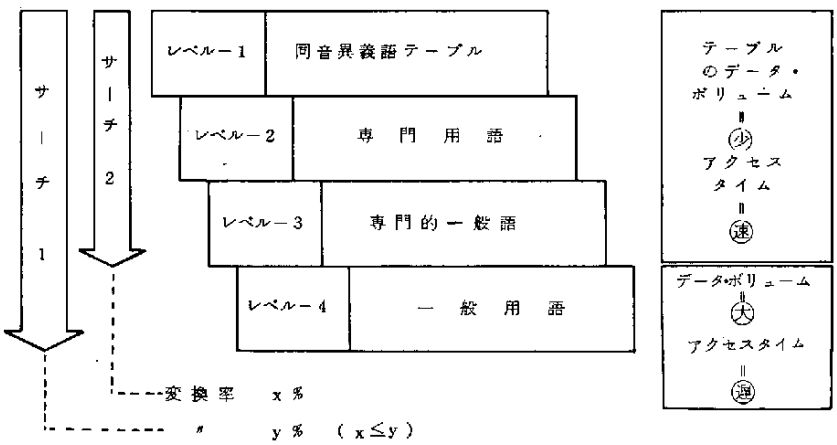


図3－7　階層構造の変換テーブルによる変換

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。

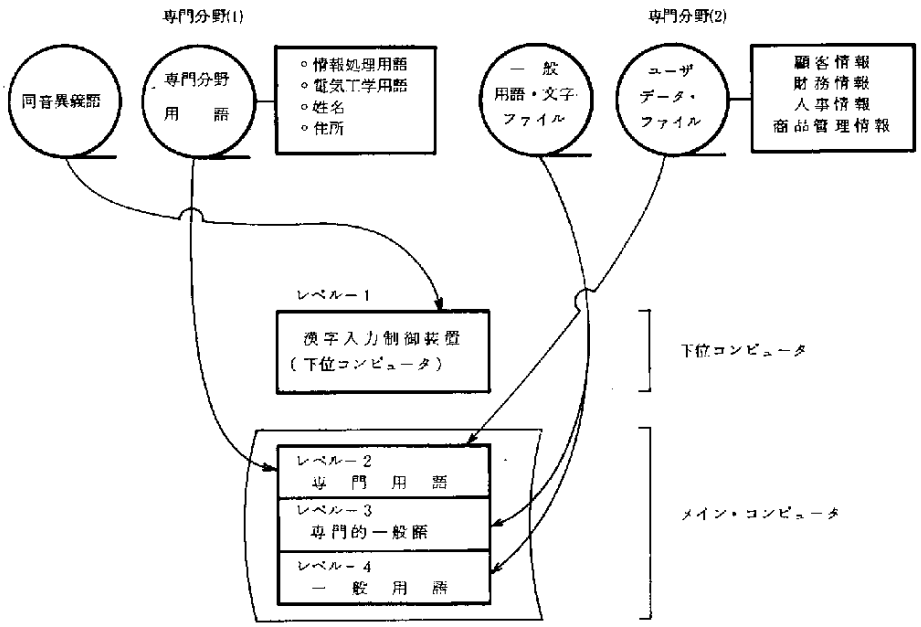


図３－８　階層化変換テーブルの構成

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イ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・ストーレッ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

計測しておき、

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ＦＩＤおよびカタカナＦＩＤをサーチしてから、

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・サーチと何等変らないが、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

階層		レベル－１	レベル－２	レベル－３	レベル－４
用語類別		同音異義語	専門用語	専門的一般語	一般用語
平均変換率	①	平均変換率　x％			
	②	平均変換率　y％			
対象変換テーブル		同音異義語ディクショナリ	専門用語ディクショナリ	セレクション（専用業務別）←	一般用語ディクショナリ
データ件数の１例		３００語～９００語	１００語～５００語	５００語～７００語	1,０００語～
記憶装置		下位コンピュータのメモリに常駐	〔ＩＳＡＭ〕 ○インデックス部　メインコンピュータのメモリ上 ○変換単語テーブル　高速DISC/DRUM	〔ＩＳＡＭ〕 ○インデックス部　高速DISC/DRUM ○変換単語テーブル　中速DISC/DRUM	大容量・低速DISK/DRUM
期待変換率		３０％	６０％	７０％	９０％
変換プロセスのステップ		1	2	3	4
使用コンピュータ		漢字入力制御装置	メイン・コンピュータ		

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。

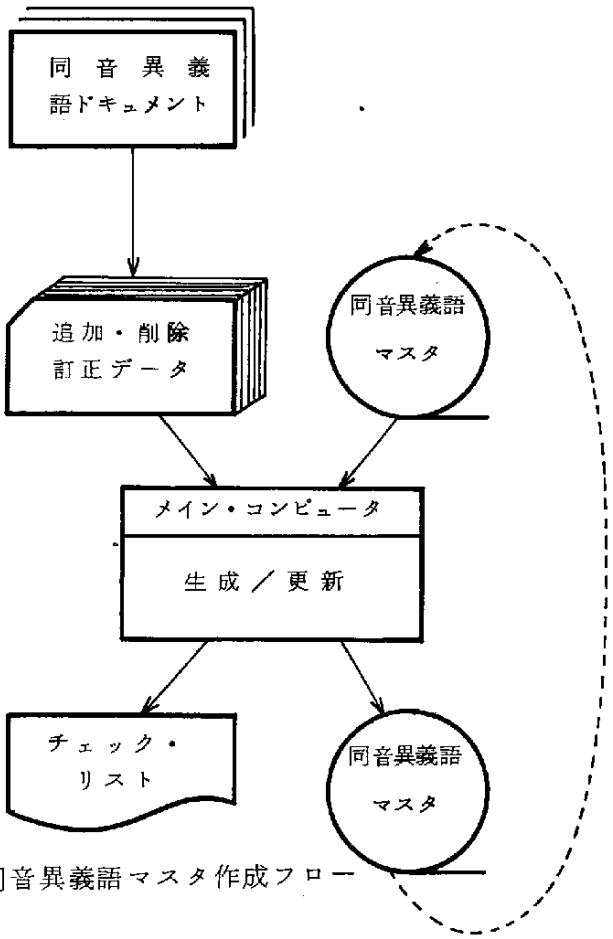


図3－9(a) 同音異義語マスタ作成フロー

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。

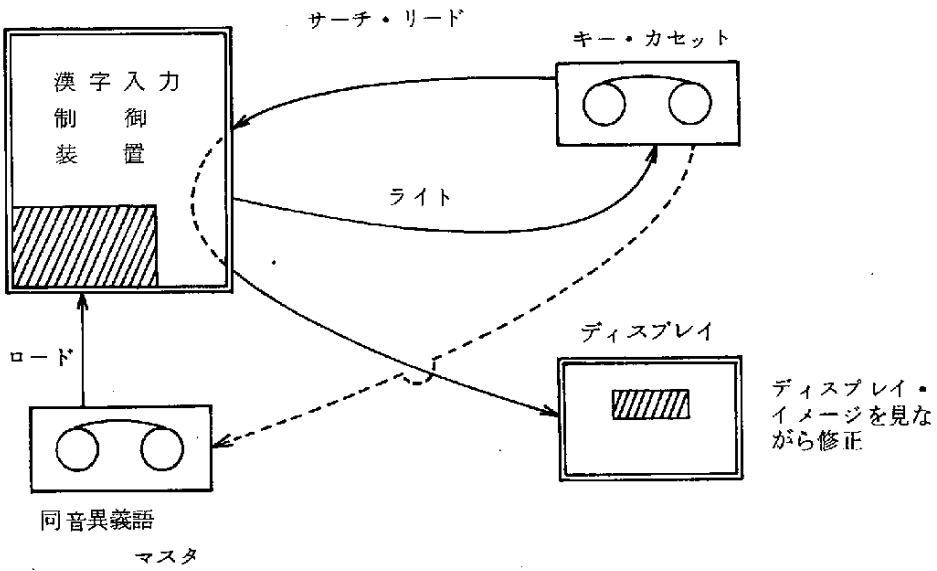


図３－９(b)　キー・カセットを利用した同音異義語マスタ作成

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。

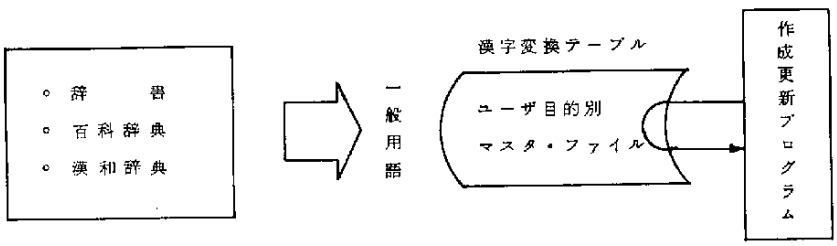


図３－１０

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。

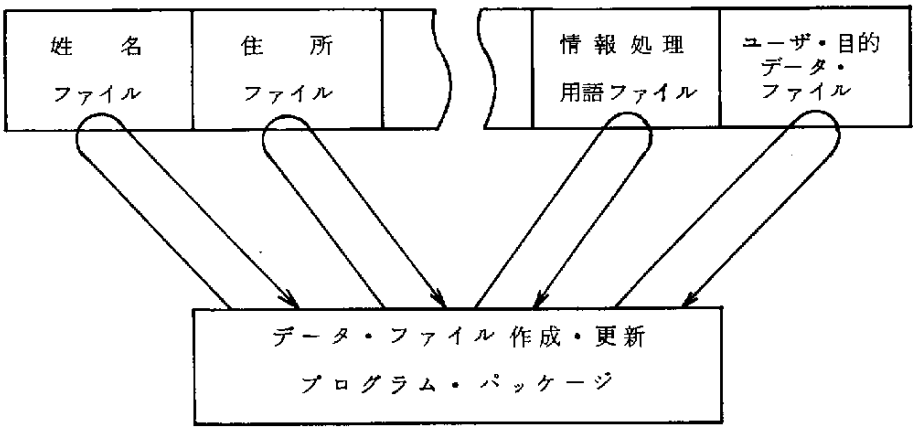


図３－１１

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。

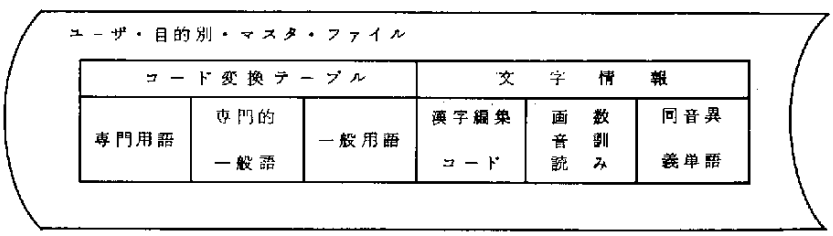


図３－１２

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・アドレッシングができること

• 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

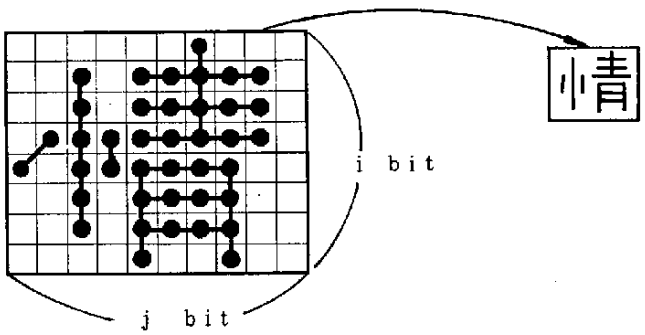
このファイルには、漢字プリンタで印書したい文字のパターンが内蔵されており、このファイルを「漢字出力制御装置」のメモリにローディングしておくことにより印書可能となる。

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Ｉｄｅｎｔｉｆｉｒｅ

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・レット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

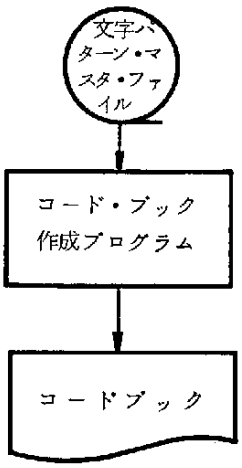


図３－１３

⑦ 漢字プリンタ編集テープの作成

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プを作成するには、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。

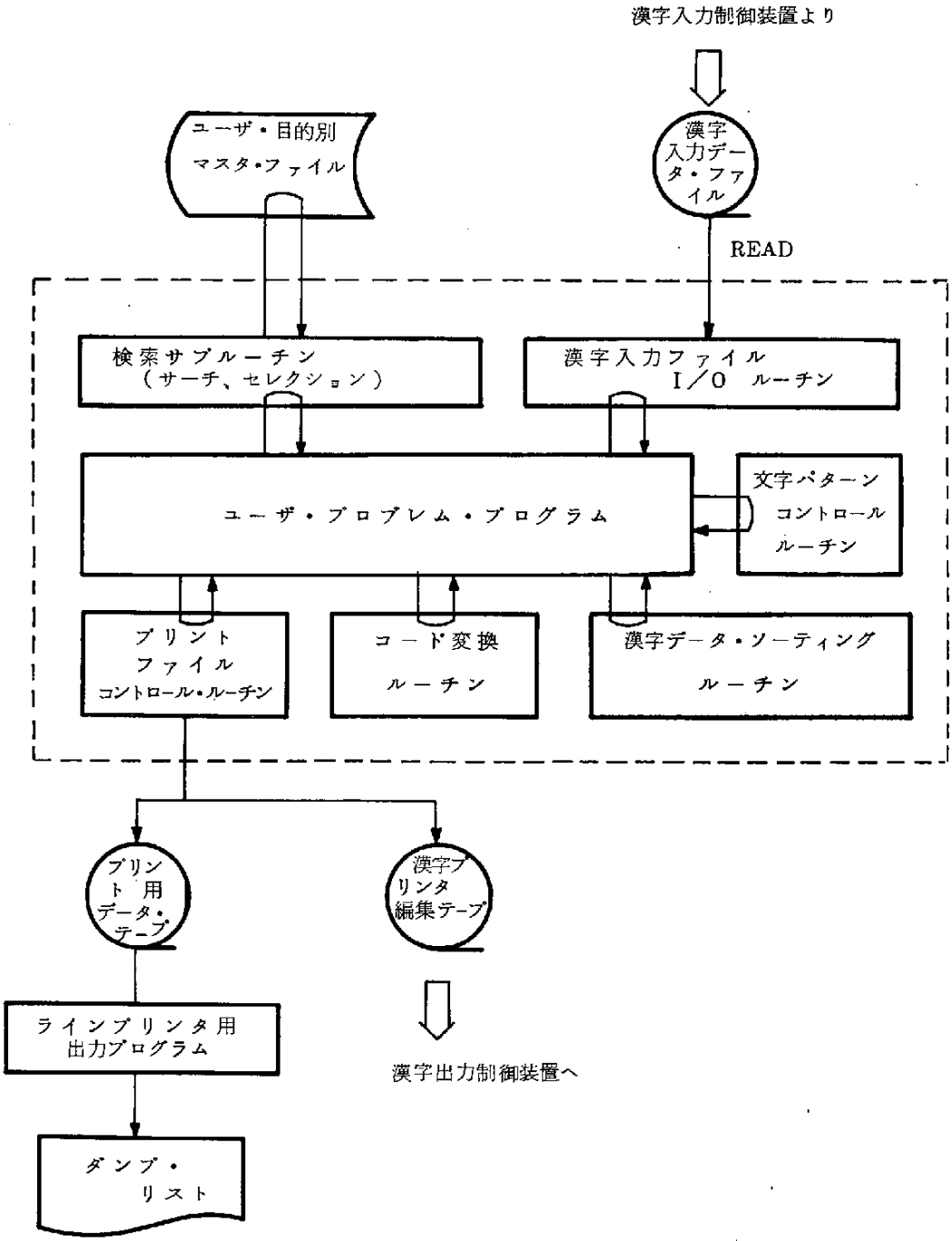


図３－１４　漢字プリンタ編集テープ作成プログラム構成図

(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プ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

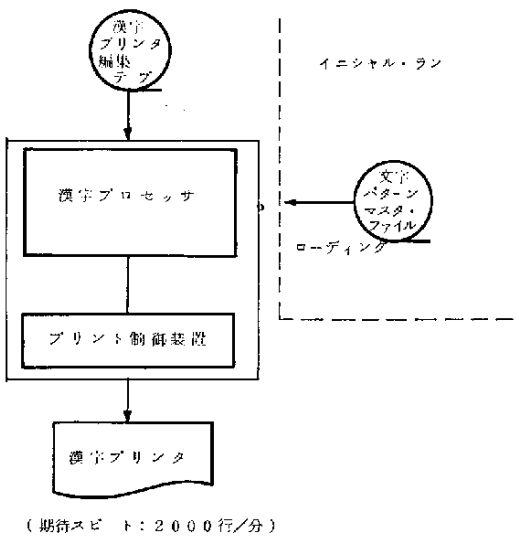


図３－１５ 漢字出力制御装置の構成

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

ベーシック・ソフトウェア	ユーティリティ・プログラム
• 漢字入出力ファイル・ハンドリング・ルーチン • プリント・ファイル・コントロール・ルーチン • 各種コード変換ルーチン • 漢字データ・ソーティング・ルーチン • 文字パターン・コントロール・ルーチン	• 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

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編成のうえで、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、相互の関連を大まかに示したものが図4－1である。

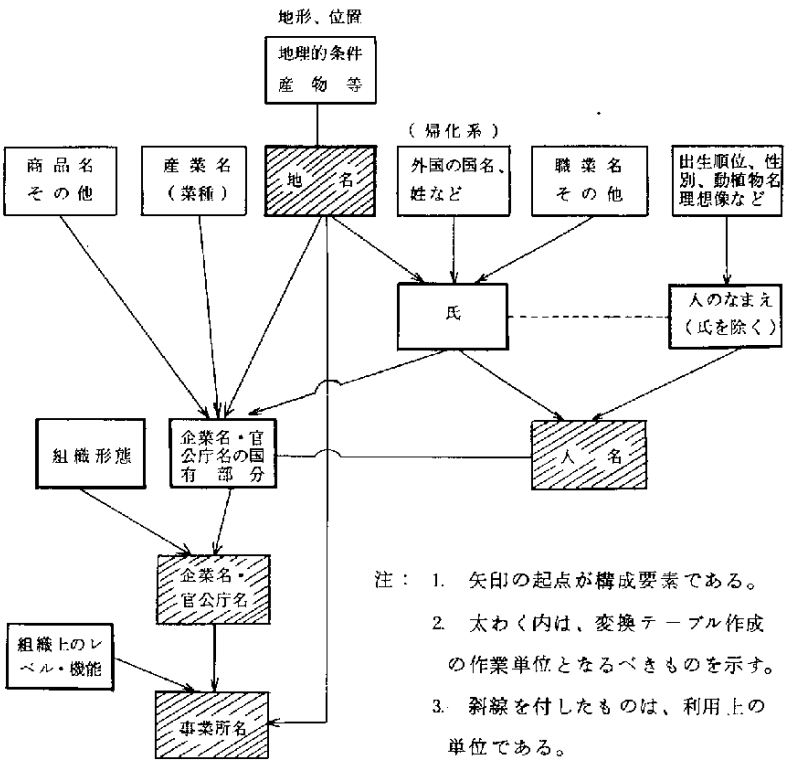


図4－1　固有名詞の構成要素

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、変換テーブルの編成単位を、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。仮りに、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、変換テーブルの編成、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であって、官公庁や企業で使われるような、組織、商品、機能などを示す単語は、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部内通達、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企業・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ぎのような諸点があげられる。

①　全国最高の人口集中地区で、各地発祥の人名が網羅的にみられる。

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有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象事業所については、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、事業処理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編集刊行し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いる点で、前記行政機構図に比べて、より網羅的であるといえる。

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。

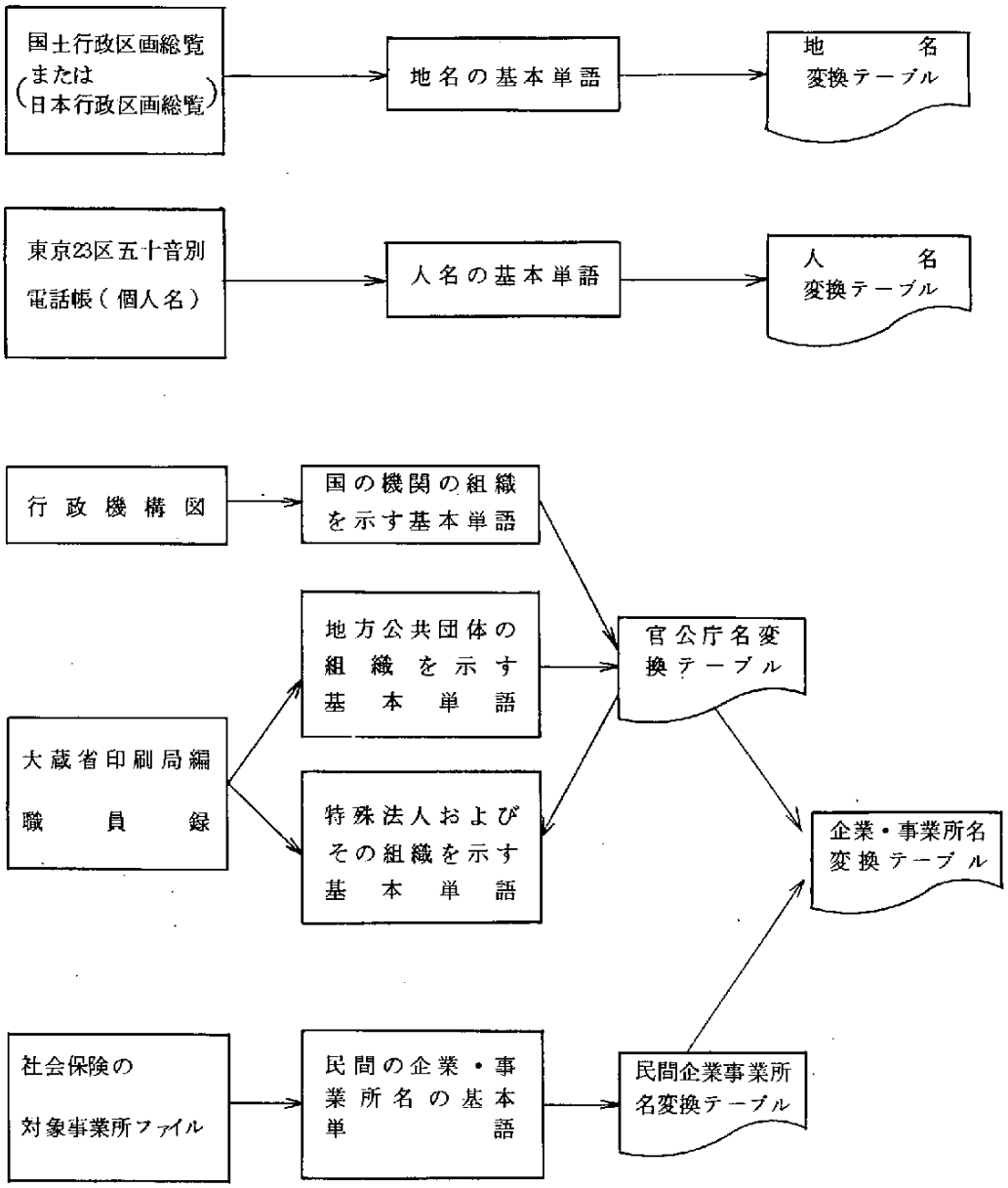


図４－２　固有名詞変換テーブル編成の手順　（案）

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¼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、

	計	民営
全国	5,308,892	5,130,579
東京23区	551,374	544,408

この調査の標本数は、

上巻	1,356
下巻	1,306
合計	2,662

したがって抽出率は、

約　0.5%　$\cdot\left(\frac{1}{200}\right)$

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、㈱、㈲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名を示すもの

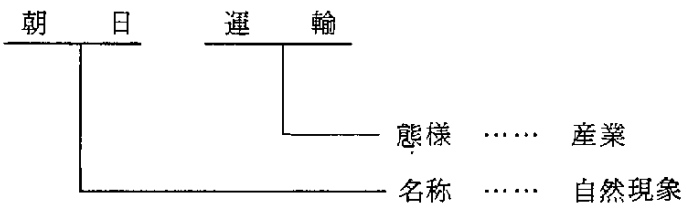
態様：事業所名のなかに事業所に関する状態、形態、様相などを示すもの

たとえば

商　品	……	鋼　管
産　業	……	運　輸
職　業	……	計　理
行　為	……	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）



(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）　　川島産業

② 地名（もしくは自然、理想、美的なもの） ＋ 態様

（例）　　日本鋼管

芙容産業

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

	使用語数	延べ使用語数	比率
氏名	826	1,248	1.5
態様	868	2,932	3.4
名称	969	1,499	1.5
計	2,663	5,679	2.1

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

	使用語数 前半／後半	延べ使用語数 前半／後半	比率
氏名	470／356	625／623	1.3／1.8
態様	566／302	1,468／1,464	2.6／4.8
名称	554／415	775／724	1.4／1.7
計	1,590／1,073	2,868／2,811	1.8／2.6

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。

表４－３　　五十音別熟語分析

	熟語								
	氏名			態様			名称		
	使用語数	重復使用語数	延べ使用語数	使用語数	重復使用語数	延べ使用語数	使用語数	重復使用語数	延べ使用語数
ア	53	28	81	8	8	16	45	23	68
イ	61	52	113	27	66	93	37	2	39
ウ	26	11	37	7	24	31	9	—	9
エ	7	3	10	13	28	41	22	8	30
オ	55	31	86	7	4	11	24	7	31
カ	54	31	85	72	125	197	46	22	68
キ	22	8	30	55	91	146	34	25	59
ク	25	5	30	9	8	17	8	—	8
ケ	3	—	3	52	81	133	11	3	14
コ	35	23	58	74	276	350	39	9	48
サ	35	26	61	14	103	117	54	22	76
シ	29	11	40	116	569	685	84	25	109
ス	20	20	40	6	10	16	24	2	26
セ	7	5	12	84	276	360	31	2	33
ソ	4	—	4	23	19	42	6	2	8
タ	67	46	113	16	12	28	69	28	97
チ	6	3	9	23	8	31	21	10	31
ツ	14	2	16	9	5	14	6	—	6
テ	4	1	5	31	100	131	11	1	12
ト	21	2	23	35	40	75	71	160	231
ナ	36	18	54	4	—	4	16	2	18

	熟語								
	氏名			態様			名称		
	使用語数	重復使用語数	延べ使用語数	使用語数	重復使用語数	延べ使用語数	使用語数	重復使用語数	延べ使用語数
ニ	14	1	15	4	—	4	37	82	119
ヌ	—	—	—	1	—	1	—	—	—
ネ	2	—	2	7	3	10	1	—	1
ノ	8	3	11	7	1	8	3	—	3
ハ	25	11	36	18	30	48	32	4	36
ヒ	19	4	23	19	32	51	21	5	26
フ	25	8	33	18	29	47	33	37	70
ヘ	2	—	2	3	—	3	4	4	8
ホ	17	3	20	22	44	66	14	—	14
マ	26	13	39	2	—	2	42	6	48
ミ	34	10	44	—	—	—	24	16	40
ム	8	2	10	2	2	4	5	2	7
メ	—	—	—	8	4	12	11	6	17
モ	13	3	16	8	11	19	5	1	6
ヤ	26	22	48	8	11	19	26	7	33
ユ	2	1	3	8	1	9	4	—	4
ヨ	17	9	26	15	28	43	6	—	6
ラ	—	—	—	—	—	—	3	—	3
リ	—	—	—	19	13	32	8	—	8
ル	—	—	—	—	—	—	—	—	—
レ	—	—	—	7	2	9	2	—	2

	熟語								
	氏名			態様			名称		
	使用語数	重復使用語数	延べ使用語数	使用語数	重復使用語数	延べ使用語数	使用語数	重復使用語数	延べ使用語数
ロ	—	—	—	6	—	6	6	1	7
ワ	4	6	10	2	—	2	14	6	20
計	826	422	1,248	868	2,064	2,932	969	530	1,499

総使用語数 5,679語

（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、態様に関する熟語に目立っており調査の後半で態様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様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様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。

表４－４　　事業所名の頻度調査

使用頻度(A)	使用語および数					延べ使用語数	
	氏名	態様	名称	計(B)	上位からの累計	上位からの延べ使用累計(C)	全対に対する割合(D)
１９７		（製作所）		1	1	１９７	3.5％
１９３		（商　店）		1	2	３９０	6.9
１２１			（東京）	1	3	５１１	9.0
１０８		（工　業）		1	4	６１９	１０.９
１０２		（商　会）		1	5	７２１	１２.７
８２		（商　事）		1	6	８０３	１４.１
８１		（産　業）		1	7	８８４	１５.６
６４			（日本）	1	8	９４８	１６.７
６１		（工業所）		1	9	１,００９	１７.８
４２		（事務所）		1	１０	１,０５１	１８.５
３２		（医　院）		1	１１	１,０８３	１９.１
３０		（研究所）		1	１２	１,１１３	１９.６
２９		（建設）（電機）		2	１４	１,１７１	２０.６
２８		（電　気）		1	１５	１,１９９	２１.１
２７		（営業所）		1	１６	１,２２６	２１.６
２４		（自動車）		1	１７	１,２５０	２２.０
２０		（協　会）	（富士）	2	１９	１,２９０	２２.７
１９		（印刷）（工場）（支店）		3	２２	１,３４７	２３.７
１８	（鈴木）			1	２３	１,３６５	２４.０
１７		（興　業）	（東洋）	2	２５	１,３９９	２４.６
１６		（機械）（工務店）（不動産）		3	２８	１,４４７	２５.５
１５							

使用頻度(A)	使用語および数					延べ使用語数	
	氏名	態様	名称	計 (B)	上位からの累計	上位からの延べ使用累計 (C)	全対に対する割合 (D)
14	1	2	1	4	32	1,503	26.5
13	2	2		4	36	1,555	27.4
12	1	4	1	6	42	1,627	28.7
11	3	7		10	52	1,737	30.6
10		4	1	5	57	1,787	31.5
9	2	6	1	9	66	1,868	32.9
8		13	2	15	81	1,988	35.0
7	4	10	3	17	98	2,107	37.1
6	6	11	5	22	120	2,239	39.4
5	10	31	5	46	166	2,469	43.5
4	11	29	15	55	221	2,689	47.4
3	44	44	36	124	345	3,061	53.9
2	96	127	77	300	645	3,661	64.5
1	645	555	818	2,018	2,663	5,679	100.0
	826	868	969	2,663		5,679	

（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）＝$\frac{(C)}{5679}$ × 100

2. 頻度16以上の語は語数の代りに実名を記入してある。

2,６６３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、

２８語まで	２５.５％
５２語まで	３０.６％
２２１語まで	４７.４％
６４５語まで	６４.５％

すなわち、使用語の1/6は、延べ使用語数の1/4を占める。事業所名は、平均2.1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）

表４－５

	頻　　度	使用事業所割合
所	５１８	１９.５％
商	３９３	１４.８
店	３９１	１４.７
業	３６０	１３.５
工	３５０	１３.１

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述する。

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、

東京都港区麻布飯倉町１－１－２

↓

東京都／港区／麻布飯倉町

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。（住居表示整備済区域および未整備区域を含む）つづいて漢字単語の基本単語の原則にしたがって

麻布飯倉町　→　麻布／飯倉町

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　／　狸穴町

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、「麻布飯倉町」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。

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ぎ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	荏原、江原
タマガワ	玉川、多摩川
トヤマチョウ	戸山町、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のか調べるために、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ぎ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、2,１４６という数になり、延べ字数4,４６１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

カスヤ	粕谷	カサイ	葛西
カミヤマチョウ	神山町	カスミ	加住
カミヤマ	神山	カタクラ	片倉
カナメ	要	カミカワ	上川
カスガチョウ	春日町	カワ	川
カスガ	春日	カワグチ	川口
カミヤチョウ	神谷町	カドタチョウ	門田町
カミヤ	神谷	カドタ	門田
カガ	加賀	カミイチブ	上壱分
カツシカク	葛飾区	カノヤ	叶谷
カツシカ	葛飾	カミト	神戸
カナ	金	カミツギ	神着
カマクラ	鎌倉	カナモリ	金森
カメアリ	亀有		

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述する。参考資料は「日本人の姓」（佐久間英、六藝書房）である。

調査にとりあげた姓の数は、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「巻島」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とおりまである。）をカタカナでカード・パンチした。

００００１　スズキ
００００２　サトウ
００００３　タナカ
　⋮
０００１５　イノウエ
０００１５　イウエ　　｝「井上」の２とおりの読み方
　⋮

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、堺、阪井、境、逆井
ショウジ…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

対応の読み方についてカタカナを漢字に直して調査してみると、

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せると思われる。

いっぽう、こうした頻度の高い姓をテーブル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。

表４－７　姓の読み方に対する同音異義語の例

アイザワ	相沢・会沢
アイダ	会田・相田・間
アイハラ	相原・粟飯原
アイバ	相場・饗庭
アオト	青砥・青戸
アカギ	赤木・赤城
アキタ	秋田・明田
アキバ	秋葉・秋庭・秋場
アキモト	秋元・秋本
アサイ	浅井・朝井
アサオカ	浅岡・朝岡
アサカ	浅香・安積
アサカワ	浅川・朝川
アサクラ	朝倉・浅倉
アサダ	浅田・朝田・麻田
アサノ	浅野・朝野・麻野
アサミ	浅見・浅海
アサヤマ	浅山・朝山
アダチ	足立・安達
アベ	阿部・安部・安倍
アライ	新井・荒井・新居
アラキ	荒木・新城・新木・安楽
アラタ	新田・荒田・新
アラタニ	新谷・荒谷
アラマキ	荒巻・荒牧
アラヤ	新谷・荒谷・新屋・新家
アラヤマ	新山・荒山
アンザイ	安西・安斎
アンドウ	安藤・安東
イイ	伊井・井伊
イカリ	猪狩・碇
イサカ	井坂・伊坂
イザキ	井崎・伊崎
イザワ	伊沢・井沢
イシイ	石井・石居
……	
（以下略）	

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、大学における学院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「公示」、「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16,286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われており、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2,３７７回であり、これは漢字熟語出現の総延べ数7,３９０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2.2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7,６９３回であり、これは漢字文字出現の総延べ数１６,２８６回のうちの４７.２％に相当する。このことは、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であり、グラフ化したのが図５－３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り、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逓減して行く。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３２％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個程度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、仮り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求される費用は、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％をねらうかで、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に表われて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逓減するということ、また、いっぽうでは変換処理の費用は少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文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用分野の登録単語数 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１３.４％の４００語にすぎなかった。登録単語数の割には、変換率が低いのは、登録単語の選択が電気工学用語集からかなり恣意的になされたためと思われる。ここで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

大　学　院	4月6日（土）午後1時	
大学第一・第二部		記　念　館
女子短期大学	4月9日（火）午前10時	
		青　学　講　堂
高　等　部	4月6日（土）午前9時30分	
		P．S．講　堂
中　等　部	4月5日（金）午前9時30分	
		青　学　講　堂
初　等　部	4月4日（木）午前10時	
		初等部礼拝堂
幼　稚　園	4月3日（水）午前10時	
		幼　稚　園

規　約

大学学生部規則一部改正

第2条　学生部長は大学長の下において、各学部長と連絡を保ちつつ、学生部を総括し、学生生活の福祉厚生ならびに助育に関する業務を処理する。

学生部副部長は、青山キャンパスに2名以内、世田谷キャンパスに1名をおく。学生部副部長は学生部長を補佐し、前項にかかげる関係業務を処理する。

学生部長および学生部副部長は教員の中から大学長が大学協議会に諮り、これを委嘱する。

学生部長および学生部副部長の任期は2年とする。

（昭和48年6月27日　理事会改正）

通　知

昭和48年度メキシコ政府留学生募集（文部省）

Ⅰ　専攻分野、募集人員

A．大学で開設中の講座を受講する留学生（30名）

スペイン語、メキシコおよび中南米の歴史、地理、経済、法律、文学、人類学、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なう留学生（10名）

指定された受入れ研究機関および専攻分野

Ⅱ　留学期間

昭和48年6月から10カ月間

Ⅲ　待　遇

往復渡航費、オアステペック休暇センター（10日間）滞在費、民宿（約9カ月間）経費負担、入学金・授業料免除

Ⅳ　応募資格

1．年令30才未満の者で、大学卒業または大学院在学中者、あるいはこれらを卒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能力を有し、単身渡航および滞在できる者

Ⅴ　応募手続

応募者は応募書類を大学庶務部庶務課を通じて、3月3日（土）までに文部省大学学術局留学生課あて提出すること

報　告

海外出張・帰国

寺崎　和郎（大学理工学部教授）

第14回国際クリーニング燃焼会議およびハワイ大学東西センター研究打合せのため7月12日から27日までアメリカへ出張した

金　　徳洲（大学理工学部教授）

学会出席および研究、講義のため7月20日から9月15日までアメリカ、イギリス、イタリアへ出張

嘉村　祐一（大学理工学部教授）

世界平和に対する大学の使命に関する各分野の話合いのため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

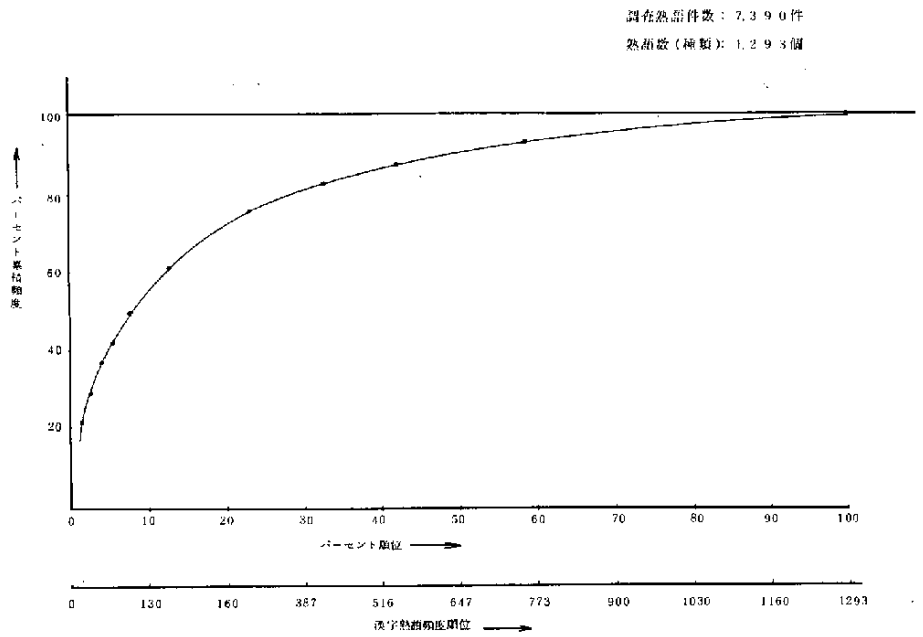


図５－２　大学内文書の漢字熟語：頻度順位とパーセント累積頻度

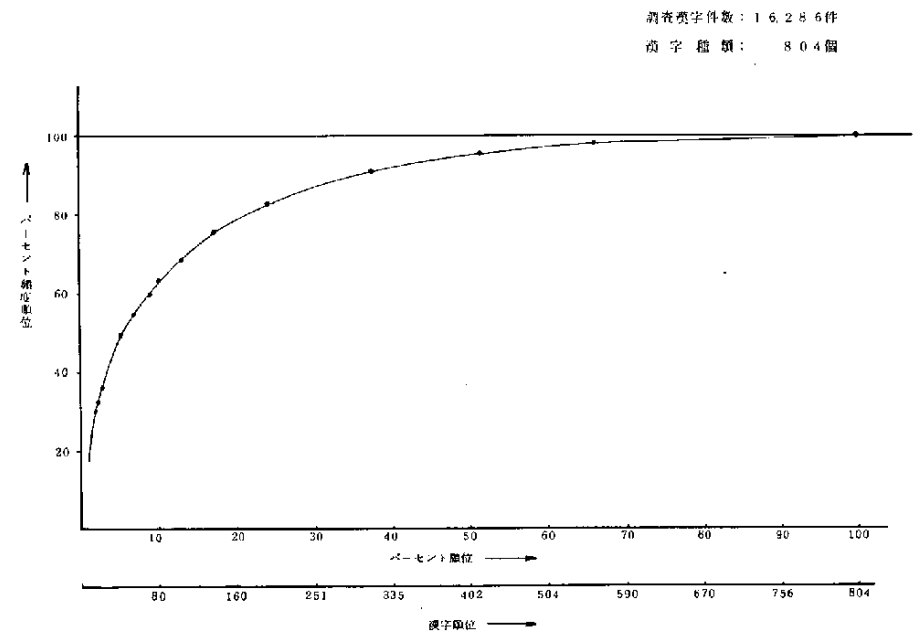


図５－３　大学内文書の漢字文字：頻度順位とパーセント累積頻度

表５－１　　大学内文書の漢字熟語の頻度と熟語数

頻　　度	熟　語　数	熟語数パーセント
343	1	.08
150～300	0	0
80～149	3	.2
50～ 79	11	.9
40～ 49	10	.8
30～ 39	17	1.3
20～ 29	33	2.6
10～ 19	106	8.2
6～ 9	117	9.0
3～ 5	246	19.0
2	215	16.6
1	534	41.3
	1,293	100

表５－２　　大学内文書の漢字文字の頻度と文字数

頻　　度	文　字　種	文字種パーセント
1327	1	.12
301～500	2	.25
201～300	6	.75
101～200	28	3.5
81～100	7	.9
61～ 80	24	3.0
41～ 60	26	3.2
21～ 40	78	9.7
11～ 20	99	12.3
6～ 10	104	12.9
3～ 5	156	19.4
2	97	12.1
1	176	21.9
	804	100

表５－３　　使用頻度の高い漢字熟語

順位	熟　語	使用頻度
1	大　学	343
2	昭　和	144
3	研　究	135
4	年　度	94
5	専　攻	79
6	募　集	78
7	出　張	74
8	学　部	71
9	午　前	59
10	教　授	57
11	高等部	54
12	午　後	54
13	応　募	52
14	日　本	51
15	研究科	49
16	経　済	48
17	手　続	48
18	入　学	46
19	女　子	45
20	幼稚園	42
21	教　育	41

順位	熟　語	使用頻度
21	理　工	41
23	学　校	40
23	分　野	40
25	初等部	39
25	中等部	39
27	支　給	38
27	短　期	38
29	提　出	37
29	文学部	37
29	礼　拝	37
32	庶務課	36
33	発　表	35
34	考　査	34
34	庶務課	34
36	文　学	33
37	学　院	32
38	教職員	31
38	経　営	31
38	文部省	31
31	青山学院	30
（頻　度　和		2,377）

表５－４　使用頻度の高い漢字文字

順位	文字	頻度	順位	文字	頻度
1	学	1327	21	日	139
2	部	427	22	員	132
3	大	396	23	事	132
4	研	263	24	生	132
5	究	248	25	記	130
6	教	234	26	者	128
7	会	225	27	午	124
8	務	210	28	人	120
9	年	206	29	国	119
10	科	179	30	本	117
11	出	173	31	経	117
12	山	172	32	理	111
13	文	165	33	院	107
14	期	159	34	専	106
15	授	156	35	書	105
16	募	155	36	給	104
17	等	146	37	応	103
18	和	145	38	費	102
19	昭	144	39	入	98
20	業	144	40	課	93

（４０位までの頻度和　7,593）

表５－５　大学内文書の漢字熟語の頻度順位と累積頻度

頻度順位	パーセント順位	累積頻度	累積パーセント
～ 8	0.006	1018	1.4
～ 20	0.015	1637	22.2
～ 34	0.026	2177	29.5
～ 55	0.043	2787	37.7
～ 71	0.055	3157	42.7
～ 98	0.076	3636	49.2
～ 168	0.128	4489	60.7
～ 298	0.230	5512	74.6
～ 425	0.329	6069	82.1
～ 544	0.421	6426	87.1
～ 759	0.587	6856	92.8
～1293	1.00	7390	100.0

表５－６　大学内文書の漢字文字の頻度順位と累積頻度

頻度順位	パーセント順位	累積頻度	累積パーセント
～ 18	0.022	4986	30.6
～ 20	0.025	5274	32.4
～ 24	0.030	5814	35.7
～ 46	0.057	8067	49.5
～ 57	0.071	8834	54.2
～ 71	0.088	9710	59.6
～ 84	0.104	10340	63.5
～ 106	0.132	11219	68.9
～ 140	0.174	12299	75.5
～ 193	0.240	13460	82.6
～ 304	0.378	14841	91.1
～ 414	0.515	15522	95.3
～ 531	0.660	15916	97.7
～ 804	1.00	16286	100.0

カタカナ入力漢字かな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て、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、とりあえず約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。

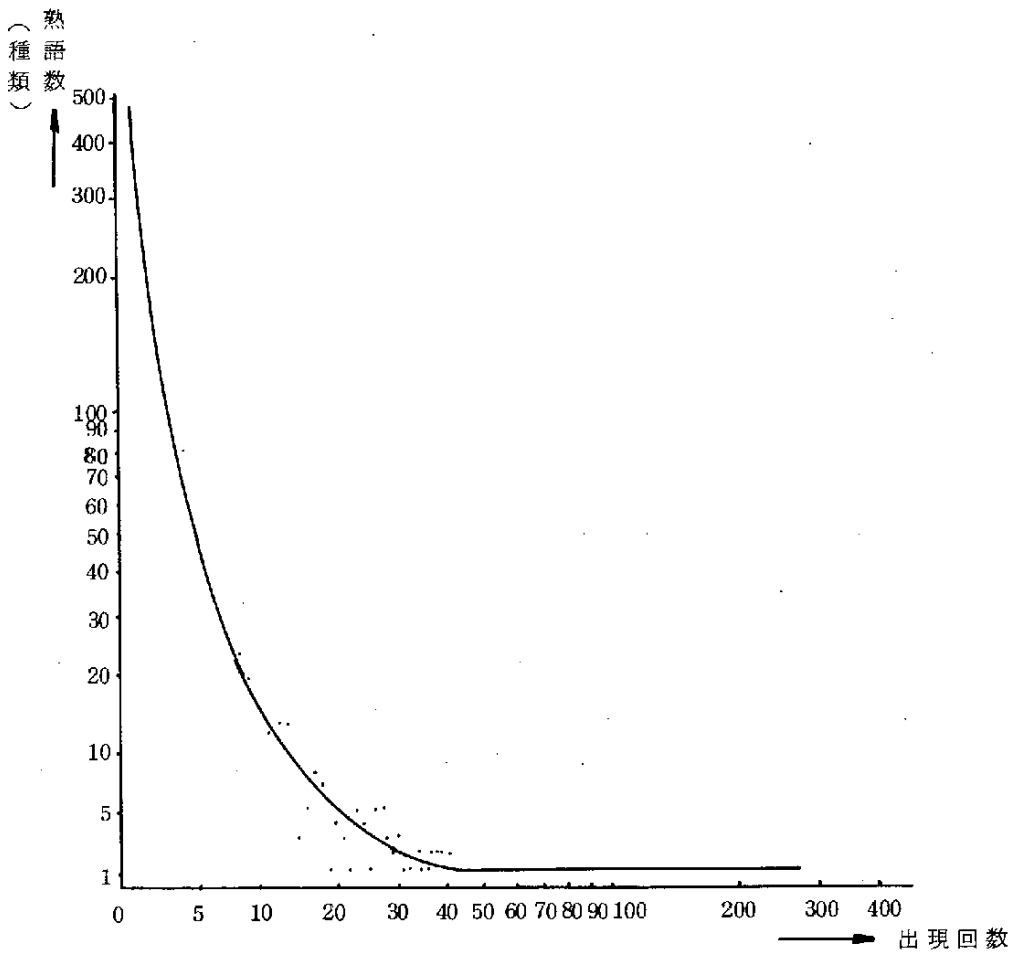


図５－４　出現頻度曲線（１６部分）

(2) 頻出語と使用熟語の分類

出現頻度のグラフを作成すると同時に、とくに頻度の高い熟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熟語を見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『学部に関する語郡』として、また、「庶務部」、「庶務課」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『時間を画するための語群』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、同じ語群内の熟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熟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『時間を画する語』それに『事務用語』などは、一般に年間を通して頻出し、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き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かな混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、いうまでもなく、その例を挙げるならば、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・コードおよびシソーラス委員会"が設けられ、医学用語、コード、シ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んで行なった。

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も多数扱うばかりでなく細分化、専門化も著しい。

日本語の造語の特性、これは医学用語ではとくに著しく現われ、そのためにか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、医学分野は未刊である。しかし、近年、学界の努力が実り、日本医学会から１９７４年中に「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５０,０００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）金原出版（約８０,０００語）
- 医学大辞典　　南山堂
- 内科用語集（１９６０）日本内科学会（約１４,８６０語）

また、英語の辞書として代表的なものは、

- Dorland's Illustrated Medical Dictionary（約５８,０００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語い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 Subjest 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くに当用漢字以外の漢字が他の分野に比べて数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

＊　桜井宣隆、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 No.11、２０９－２５５（１９６７）

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。

調査結果はつぎのとおりである。

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２０,２３４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）に含まれるものをみると、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％、全般で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・嚢・指・膚・頸

椎・喉・髄・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

これらの字種は一般編のリスト中でも上位を占めている。また、医学論文タイトルに特有な漢字としてつぎ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。

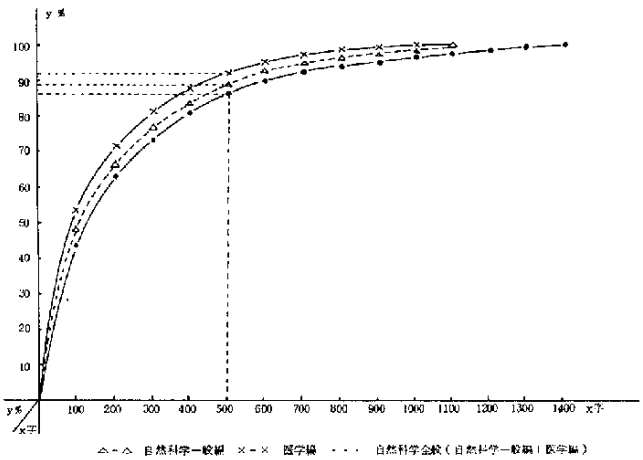


図５－５　頻度順位と全体に対する割合の関係図

表５－７　頻度順位と全体に対する割合の比較表

上位からの文字	自然科学一般編		医学編		自然科学全般（自然科学一般編＋医学編）	
	文字数	全体に対する割合	文字数	全体に対する割合	文字数	全体に対する割合
	字	%	字	%	字	%
1～100	4,177	48.17	6,176	53.49	9,286	45.89
1～200	5,722	65.99	8,222	71.22	12,905	63.77
1～300	6,643	76.61	9,388	81.32	15,044	74.35
1～400	7,268	83.81	10,139	87.82	16,521	81.64
1～500	7,706	88.87	10,614	91.94	17,562	86.79
1～600	8,021	92.50	10,945	94.81	18,312	90.50
1～700	8,255	94.85	11,161	96.68	18,852	93.16
1～800	8,365	96.47	11,305	97.92	19,234	95.05
1～900	8,465	97.62	11,405	98.79	19,530	96.52
1～1,000	8,565	98.77	11,505	99.66	19,730	97.50
1～1,100	8,665	99.93	11,544	100.00	19,897	98.33
1～1,200	8,671	100.00	—		19,997	98.82
1～1,300	—		—		20,097	99.32
1～1,400	—		—		20,197	99.81
1～1,500	—		—		20,234	100.00

表５－８　雑誌記事索引－自然科学編収録論文標題の漢字分布

		自然科学一般編	医学編	自然科学全般（自然科学一般編＋医学編）
件数（論文数）		1,032	1,095	2,127
総文字数		8,671	11,544	20,234
一論文あたりの平均漢字数		8.4	10.5	9.5
漢字の種類		1,106	1,039	1,437
内訳	当用漢字	988（89.3%）	863（83.0%）	1,176（81.5%）
	（全当用漢字に対する割合）	53.4%	46.6%	63.6%
	教育漢字	647	584	730
	教育漢字外漢字	328	271	428
	人名漢字	6	5	9
	補正漢字	7	3	9
	当用漢字外漢字	118（10.2%）	176（17.0%）	251（18.5%）

表５－９　一般編と医学編に出現する漢字の比較

No.	字種	度数	順位（一般）	順位（医学）	
74	癌 ○	63		28	○ } 医学分野に特有な字種
104	腎 ●	48		51	●
118	瘍 ○	45		58	
158	肝 ●	33		90	● 身体の部位を示す字種
202	鋼	26	78		
243	稲	22	99		
249	裁	21	102		
256	娩 ○	21		143	
261	船	20	105		
280	併	19		159	
294	投	18		166	
297	副	18		169	
310	畜	17	129		
342	材	15	152		
348	胆 ●	15		200	
352	歯 ●	15		203	
356	婦 ○	15		204	
361	澄 ○	14		206	
367	指 ●	14		211	
368	耳 ●	14		212	
374	膚 ●	14		218	
379	頭	13		225	
380	建	13	173		
384	社	13	177		
385	設	13	176		

No.	字種	度数	順位（一般）	順位（医学）	
390	糖	13		232	
396	給	12	190		
398	号	12	192		
404	準	12	194		
408	鋳	12	196		
410	膜 ○	12		244	
411	米	12	197		
429	繊	11	210		
437	椎 ●	11		260	
442	判	11		263	
447	林	11	219		
451	漁	10	222		
454	喉 ●	10		271	
461	掘	10	225		
466	随 ●	10		283	
471	鼻 ●	10		287	
476	慢 ○	10		288	
478	未	10		289	
480	安	9	238		
493	飼	9	251		
494	序	9		305	
497	難	9		311	
499	燃	9	257		
500	播	9	258		
			22種	27種	

表５－１０　医学編および一般編での頻度最上位の漢字

		一般編での順位		医学編での順位
1	性	5	関	8
2	症		用	11
3	血		研	5
4	究	4	究	4
5	研	5	性	1
6	例		電	
7	的	17	化	25
8	関	1	法	12
9	療		機	
10	験	18	工	
11	用	2	?	
12	法	8	水	
13	治		計	
14	経		物	
15	生		業	
16	病		分	
17	学	25	的	7
18	腫		験	10
19	児		子	
20	科		度	
21	及		合	
22	対		発	
23	検		気	
24	発		動	
25	化	7	学	17

表５－１１　雑誌記事索引医学編各漢字と頻度

○　当用漢字中の教育漢字
△　〃　教育漢字外漢字
∨　〃　人名漢字
×　〃　補正漢字

度数順位	文字		度数	全数比%/累計	度数順位	文字		度数	全数比%/累計
1	性	○	210	1.82	30	床	○	63	0.55
2	症	△	203	1.76	31	臨	○	63	0.55
3	血	○	177	1.53	32	体	○	61	0.52
4	究	○	174	1.51	33	子	○	60	0.52
5	研	○	174	1.51	34	外	○	58	0.50
6	例	○	170	1.47	35	分	○	58	0.50
7	的	○	164	1.42	36	影	△	57	0.49
8	関	○	132	1.14	37	剤	△	55	0.48
9	療	△	110	0.95	38	動	○	55	0.48
10	験	○	108	0.93	39	小	○	54	0.47
11	用	○	107	0.93	40	圧	○	53	0.46
12	法	○	101	0.87	41	使	○	53	0.4
13	治	○	94	0.81	42	細	○	52	0.45
14	経	○	91	0.79	43	酸	○	52	0.45
15	生	○	90	0.78	44	結	○	51	0.43
16	病	○	89	0.77	45	心	○	51	0.43
17	学	○	87	0.75	46	管	○	49	0.42
18	脳		76	0.66	47	機	○	49	0.42
19	児	○	76	0.66	48	神	○	49	0.42
20	科	○	75	0.65	49	内	○	49	0.42
21	及	△	73	0.63	50	白	○	49	0.42 368
22	対	○	72	0.62	51	腎		48	0.42
23	検	○	70	0.61	52	成	○	48	0.42
24	発	○	68	0.59	53	膜	△	47	0.41
25	化	○	67	0.58	54	胃	○	46	0.40
26	肺	△	66	0.57	55	実	○	46	0.40
27	患	∧	64	0.55	56	胞	△	46	0.40
28	痛		63	0.55	57	手	○	45	0.39
29	術	○	63	0.55	58	場		45	0.39

度数順位	文字		度数	全数比%/累計	度数順位	文字		度数	全数比%/累計
59	炎	△	43	0.37	95	害	○	30	0.26
60	診	△	43	0.37	96	期	○	30	0.26
61	高	○	42	0.36	97	尿	△	30	0.26
62	質	○	42	0.36	98	於		30	0.26
63	素	○	42	0.36	99	果	○	29	0.25
64	中	○	42	0.36	100	電	○	29	0.25 53.49
65	計	○	42	0.36	101	観	○	28	0.24
66	能	○	42	0.36	102	筋	△	28	
67	部	○	42	0.36	103	効	○	28	
68	理	○	42	0.36	104	作	○	28	
69	皮	○	41	0.36	105	査	○	28	0.24
70	骨	△	40	0.35	106	妊	△	28	0.24
71	後	○	40	0.35	107	時	○	27	0.23
72	疾	○	40	0.35	108	大	○	27	
73	核	△	39	0.34	109	物	○	27	
74	合	○	39	0.34	110	変	○	27	
75	特	○	39	0.34	111	方	○	27	
76	応	○	38	0.33	112	質	○	27	0.23
77	状	○	38	0.33	113	因	○	26	0.23
78	断	○	38	0.33	114	域	∧	26	
79	菌	△	37	0.32	115	障	△	26	
80	馨	∧	37	0.32	116	機	○	26	
81	人	○	37	0.32	117	組	○	26	
82	線	○	37	0.32	118	職	△	26	0.23
83	脳	○	37	0.32	119	下	○	25	0.22
84	液	○	36	0.31	120	会	○	25	
85	者	○	36	0.31	121	宮	○	25	
86	一	○	36	0.31	122	細	○	25	
87	脈	○	36	0.31	123	頭	○	25	
88	察	○	35	0.30	124	量	○	25	0.22
89	抗	△	34	0.29	125	過	○	24	0.21
90	肝	△	33	0.29	126	度	○	24	
91	新	○	33	0.29	127	不	○	24	
92	出	○	32	0.28	128	養	○	23	0.20
93	感	○	31	0.27	129	所	○	22	0.19
94	定	○	31	0.27	130	節	○	22	0.19

度数順位	文字	度数	全数比(%)累計
203	轟	15	
204	財 ○	15	0.13
205	院 ○	14	0.12
206	激	14	
207	最 △	14	
208	見 ○	14	
209	国 ○	14	
210	腔	14	
211	指 ○	14	
212	耳 ○	14	
213	娠	14	
214	前 ○	14	
215	脱 △	14	
216	天 ○	14	
217	図 ○	14	
218	膚 △	14	
219	剤 △	14	
220	論 ○	14	0.12
221	育 ○	13	0.11
222	英 ○	13	
223	活 ○	13	
224	計 ○	13	
225	頸	13	
226	門 ○	13	
227	考 ○	13	
228	試 ○	13	
229	種 ○	13	
230	石 ○	13	
231	注 ○	13	
232	構 △	13	
233	麻 △	13	
234	類 ○	13	0.11
235	維 △	12	0.10
236	柴 ○	12	
237	角 ○	12	
238	環 △	12	0.10
239	土 ○	12	0.10
240	再 ○	12	
241	自 ○	12	
242	臓 △	12	
243	胎 △	12	
244	臓	12	
245	毛 ○	12	0.10
246	医 ○	11	0.10
247	各 ○	11	
248	解 ○	11	
249	交 ○	11	
250	隆 △	11	
251	申 △	11	
252	三 △	11	
253	集 ○	11	
254	消 ○	11	
255	進 ○	11	
256	主 ○	11	
257	静 ○	11	
258	先 ○	11	
259	赤 ○	11	
260	維	11	
261	低 ○	11	
262	人 ○	11	
263	判 ○	11	
264	評 ○	11	
265	比 ○	11	
266	肥 ○	11	
267	文 ○	11	
268	保 ○	11	
269	離 △	11	0.10
270	運 ○	10	0.09
271	喉	10	
272	膵 △	10	
273	最 ○	10	
274	氏 ○	10	0.09

度数順位	文字	度数	全数比(%)累計
131	代 ○	22	0.19
132	腹 △	22	0.19
133	移 ○	21	0.18
134	異 ○	21	
135	開 ○	21	
136	行 ○	21	
137	傷 △	21	
138	射 △	21	
139	染 △	21	
140	清 ○	21	
141	造 ○	21	
142	題 ○	21	
143	娩	21	
144	問 ○	21	0.18
145	群 ○	20	0.17
146	謝 ○	20	
147	蛋	20	
148	年 ○	20	0.17
149	感 ○	19	0.16
150	基 ○	19	63.7
151	球 ○	19	
152	候 ○	19	
153	産 ○	19	
154	重 ○	19	
155	水 ○	19	
156	腸 ○	19	
157	毒 ○	19	
158	徴 ○	19	
159	併 △	19	0.16
160	気 ○	18	0.16
161	器 ○	18	
162	形 ○	18	
163	就 ○	18	
164	全 ○	18	
165	道 ○	18	
166	投 ○	18	0.6
167	乳 △	18	0.16
168	伴 △	18	
169	副 ○	18	
170	本 ○	18	
171	与 △	18	
172	力 ○	18	0.16
173	頭 △	17	0.15
174	系 ○	17	
175	型 ○	17	
176	酵 △	17	
177	色 ○	17	
178	上 ○	17	
179	精 ○	17	
180	腺	17	
181	変 △	17	
182	薬 ○	17	
183	輪 ○	17	
184	予 ○	17	
185	反 ○	17	0.15
186	価 ○	16	0.14
187	原 ○	16	
188	常 ○	16	
189	測 ○	16	
190	虫 ○	16	
191	二 ○	16	
192	眼 ○	15	0.13
193	急 ○	15	
194	近 ○	15	
195	鏡 ○	15	
196	現 ○	15	
197	構 ○	15	
198	視 ○	15	
199	切 ○	15	
200	胆 △	15	71.22
201	長 ○	15	
202	点 ○	15	

度数順位	文字	度数	全数比(%)累計
275	食 ○	10	0.09
276	椎 ○	10	
277	熟 △	10	
278	折 ○	10	
279	集 △	10	
280	談 ○	10	
281	調 ○	10	
282	値 △	10	
283	髄 △	10	
284	統 ○	10	
285	導 ○	10	
286	塔 △	10	
287	鼻 ○	10	
288	慢 △	10	
289	未 ○	10	
290	誘 △	10	
291	溶 △	10	
292	那 △	10	
293	流 ○	10	
294	立 ○	10	0.09
295	意 ○	9	0.08
296	音 ○	9	
297	義 ○	9	
298	眼 ○	9	
299	簡 ○	9	
300	処 ○	9	81.32
301	場 ○	9	
302	語 ○	9	
303	室 ○	9	
304	除 ○	9	
305	序 ○	9	
306	拿 ○	9	
307	遷 ○	9	
308	第 ○	9	
309	転 ○	9	
310	同 △	9	0.08
311	難 ○	9	0.08
312	放	9	
313	末 ○	9	
314	満 ○	9	
315	無 ○	9	
316	則 △	9	
317	連 ○	9	0.08
318	遺 ○	8	0.07
319	衛 ○	8	
320	疲 △	8	
321	温 ○	8	
322	隔 △	8	
323	較 △	8	
324	巨 △	8	
325	胸 ∧	8	
326	起 ○	8	
327	援 △	8	
328	際 ○	8	
329	死 ○	8	
330	周 ○	8	
331	制 ○	8	
332	脊 △	8	
333	栓	8	
334	総 ○	8	
335	団 ○	8	
336	地 ○	8	
337	痛 △	8	
338	伝 ○	8	
339	枚 △	8	
340	肉 ○	8	
341	班	8	
342	波 ○	8	
343	別 ○	8	
344	房 ∧	8	
345	味 ○	8	
346	門 ○	8	0.07
347	要 ○	8	0.07
348	利 ○	8	
349	烈 △	8	0.07
350	惠 ○	7	0.06 85.1
351	回 ○	7	
352	蓄	7	
353	局 ○	7	
354	業 ○	7	
355	健 ○	7	
356	係 ○	7	
357	呼 △	7	
358	硬 △	7	
359	策 ○	7	
360	十 ○	7	
361	殊 ∧	7	
362	絨	7	
363	循 △	7	
364	支 ○	7	
365	酔 △	7	
366	正 ○	7	
367	早 ○	7	
368	装 △	7	
369	損 ○	7	
370	統 ○	7	
371	暦 △	7	
372	択 △	7	
373	知 ○	7	
374	聴 △	7	
375	通 ○	7	
376	吐 △	7	
377	道 ○	7	
378	閉 △	7	
379	網 △	7	
380	幼 △	7	
381	様 ○	7	
382	燐	7	0.06
383	老 ○	7	0.06
384	路 ○	7	
385	和 ○	7	0.06
386	明	6	0.05
387	黄 ○	6	
388	開 ○	6	
389	鑑 △	6	
390	界 ○	6	
391	検 ∧	6	
392	興 ○	6	
393	狭 △	6	
394	混 ○	6	
395	広 ○	6	
396	訊	6	
397	左 ○	6	
398	侵 △	6	
399	止 ○	6	
400	鉄 △	6	
401	式 ○	6	
402	編 ∧	6	
403	重 △	6	
404	接 ○	6	
405	相 ○	6	
406	耐 △	6	
407	膵	6	
408	着 ○	6	
409	洞	6	
410	旧 ○	6	
411	溶 △	6	
412	膀	6	
413	右 ○	6	
414	硫 △	6	
415	々	6	0.05
416	団 ○	5	0.04
417	位 ○	5	
418	簡 △	5	0.04

度数順位	文字	度数	全数比(%)	累計
419	家 ○	5	0.04	
420	格 ○	5		
421	境 ○	5		
422	減 ○	5		
423	県 ○	5		
424	週 ○	5		
425	殺 ○	5		
426	災 ○	5		
427	照 ○	5		
428	少 ○	5		
429	施 △	5		
430	市 ○	5		
431	脂 △	5		
432	思 ○	5		
433	示 ○	5		
434	刺 △	5		
435	肢	5		
436	真 ○	5		
437	腹 △	5		
438	類	5		
439	析 △	5		
440	新 △	5		
441	寮	5		
442	存 ○	5		
443	像 ○	5		
444	単 ○	5		
445	植	5		
446	屋 ○	5		
447	皇 △	5		
448	結 △	5		
449	盤 △	5		
450	摩 △	5	89.9	
451	裂 ○	5		
452	被 △	5		
453	貴 ○	5		
454	浮 △	5	0.04	
455	覆 ○	5	0.04	
456	亡 △	5		
457	目 ○	5		
458	孟	5		
459	癒	5		
460	労 ○	5	0.04	
461	塩 ○	4	0.03	
462	喘	4		
463	横 ○	4		
464	我 ○	4		
465	加 ○	4		
466	覚 ○	4		
467	拡 ○	4		
468	顎	4		
469	寄 ○	4		
470	緩 △	4		
471	教 ○	4		
472	季 ○	4		
473	強 ○	4		
474	腱	4		
475		4		
476	誤 ○	4		
477	個 ○	4		
478	康 ○	4		
479	行	4		
480	宰	4		
481	鎖 △	4		
482	昭 ○	4		
483	住 ○	4		
484	女 ○	4		
485	身 ○	4		
486	事 ○	4		
487	写 ○	4		
488	終 ○	4		
489	昇 △	4		
490	来 ○	4	0.03	
491	針 △	4	0.03	
492	梢	4		
493	枢 △	4		
494	整 ○	4		
495	然 ○	4		
496	製 ○	4		
497	村 ○	4		
498	多 ○	4		
499	台 ○	4		
500	超 △	4	91.94	
501	復 △	4		
502	適 ○	4		
503	底 ○	4		
504	訓	4		
505	菊	4		
506	認 ○	4		
507	熱 ○	4		
508	農 ○	4		
509	濃 △	4		
510	薄 △	4		
511	配 ○	4	0.03	
512	板 ○	4	0.03	
513	標 ○	4		
514	疲 △	4		
515	伸	4		
516	弁 ○	4		
517	編	4		
518	母 ○	4		
519	防 ○	4		
520	報 ○	4		
521	補	4		
522	泡	4		
523	望 ○	4		
524	民 ○	4		
525	面 ○	4		
526	免 △	4		
527	腰 △	4		
528	抑 △	4		
529	冷 ○	4		
530	池	4		
531	頻	4	0.03	

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。この病気のグルーピング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4,７４９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字とそれほどかわりない。

＊　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

いっぽう、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１９.２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、ついで「性」「炎」「腫」の順である。このリストには第５位の「ン」以下「ル」「ー」（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・瘍・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

などが見うけられる。

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。

つぎに「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、病名を表わす接尾語（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。

表５－１２　医学大辞典の病名見出し語にみられた文字リスト

順位	文字	頻度	%	順位	文字	頻度	%
1	症	1,335	5.83	28	化	106	0.46
2	性	1,033	4.51	29	膿×	105	0.46
3	炎	551	2.41	30	胞	104	0.45
4	腫×	505	2.21	31	硬	102	0.45
5	ン	377	1.65	32	肺	101	0.44
6	ル	369	1.61	33	フ	101	0.44
7	病	354	1.55	34	細	100	0.43
8	一＊	326	1.42	35	状	98	0.43
9	血	312	1.36	36	ツ	95	0.42
10	候	240	1.05	37	脈	95	0.41
11	群	234	1.02	38	ウ	94	0.41
12	リ	233	1.02	39	心	93	0.41
13	膜	222	0.97	40	発	92	0.40
14	ア	201	0.88	41	内	91	0.40
15	ス	188	0.82	42	尿	88	0.38
16	骨	179	0.78	43	嚢	87	0.38
17	神	172	0.75	44	筋	86	0.38
18	皮	158	0.69	45	色	86	0.38
19	瘍×	152	0.66	46	疹	86	0.38
20	イ	149	0.65	47	節	86	0.38
21	経	143	0.62	48	癌×	85	0.37
22	髄	136	0.59	49	チ	84	0.37
23	毒	129	0.56	50	肉	81	0.35
24	結	124	0.54				
25	中	111	0.49	＊　長音			
26	工	110	0.48	×　当用漢字以外の漢字			
27	ク	106	0.46				

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、候、群、瘍、毒、膿

の９種がある。

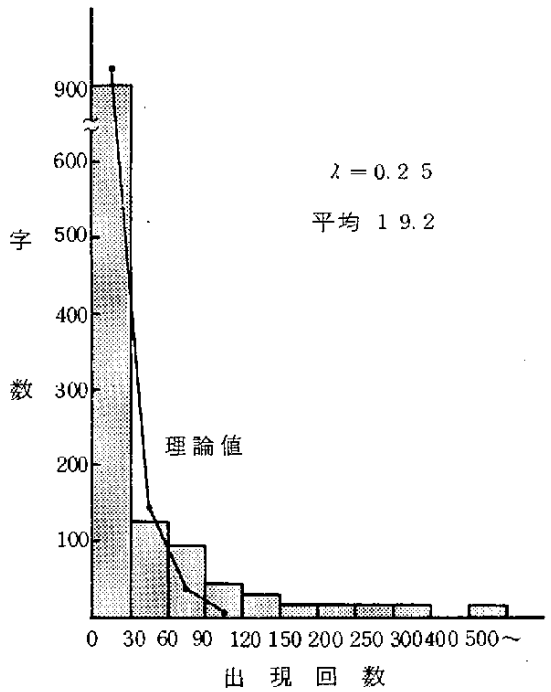


図５－６　　病名の出現頻度別度数分布（ポアソン分布）

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
- ○Ehlrich腹水癌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。

＊　斉藤　孝、医学用語の統計的調査（Ⅰ）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）

斉藤　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た用語。

たとえば、"甲状腺腫"、"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、展望"などの慣用的に附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

順位	単位語	頻度	%	順位	単位語	頻度	%
1	＊～　的	368	4.41	33	結　核	41	0.49
2	癌	286	3.43	34	＊～　剤	41	0.49
3	＊～　学	248	2.98	35	移　植	38	0.46
4	＊～　性	245	2.94	36	＊～　例	38	0.46
5	研　究	241	2.89	37	甲　状	37	0.44
6	組　織	232	2.78	38	ラッテ	37	0.44
7	細　胞	210	2.52	39	＊～　酸	36	0.43
8	病　理	182	2.18	40	転　移	34	0.41
9	腫　瘍	144	1.73	41	反　応	33	0.40
10	肝　(臓)	134	1.61	42	障　碍	32	0.38
11	＊～　症	125	1.50	43	白　血	32	0.38
12	肺　(臓)	92	1.10	44	染　色	32	0.38
13	化　学	71	0.85	45	臓　器	30	0.36
14	マウス	69	0.83	46	Virus	30	0.36
15	顕微鏡	65	0.78	47	淋　巴	29	0.35
16	脳	65	0.78	48	硬　変	27	0.32
17	電　子	65	0.78	49	動　物	26	0.31
18	＊～　腺	60	0.72	50	観　察	25	0.30
19	実　験	59	0.71	51	増　殖	24	0.29
20	肉　腫	58	0.70	52	酵　素	24	0.29
21	＊～　炎	57	0.68	53	＊～　像	23	0.28
22	変　化	55	0.66	54	病　変	23	0.28
23	＊～　系	51	0.61	55	＊～　法	23	0.28
24	形　態	50	0.60	56	中　毒	23	0.28
25	腎　(臓)	49	0.59	57	硬　化	23	0.28
26	動　脈	48	0.58	58	物　質	22	0.26
27	発　生	47	0.56	59	＊～　類	22	0.26
28	＊～　体	46	0.55	60	神　経	22	0.26
29	胃	43	0.52	61	所　見	22	0.26
30	培　養	42	0.50	62	家　兎	22	0.26
31	腹　水	41	0.49	63	多　糖	21	0.25
32	血　管	41	0.49	64	異　常	21	0.25

平均標題単位語数　　8語／標題

異なり語数　　1,104語

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

	単位語	特殊単位語*	総数
延べ語数	6,188	1,847	8,034
異なり語数	854	250	1,104
平均使用率**	7.3	7.4	7.3

＊　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 ＝ 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

Ｘ（語）	50	100	200	300	400	500	600	700	800	900	1,000
Ｙ（%）	43	58	70	76	83	86	89	92	95	97	98
延べ語数	3,456	4,661		6,671							

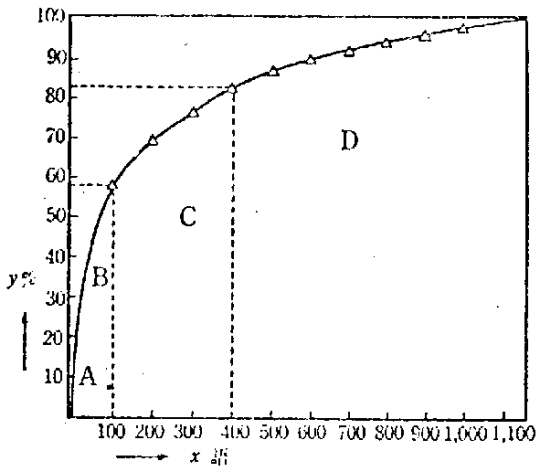


図５－７　単位語の異なり語数（Ｘ）
——累積使用率（Ｙ）曲線

(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異なり語数1,１０４語中、２５０語を占めているが、このうち接尾語を示す語として機能する補助語について評価をしてみる。

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瞭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 / 淋巴性肉腫 }　　{ 異種腫瘍細胞 / 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して、形態の描写に用いられる。

"花房状、ブドウ状、星状"のように「～状」は、物質を示す単位語との結合力かつ強く、

"Virus様、リウマチ様、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では、語なのか、句なのかの判断に苦しむ。

（例）　悪性型黒色棘細胞増殖病

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>

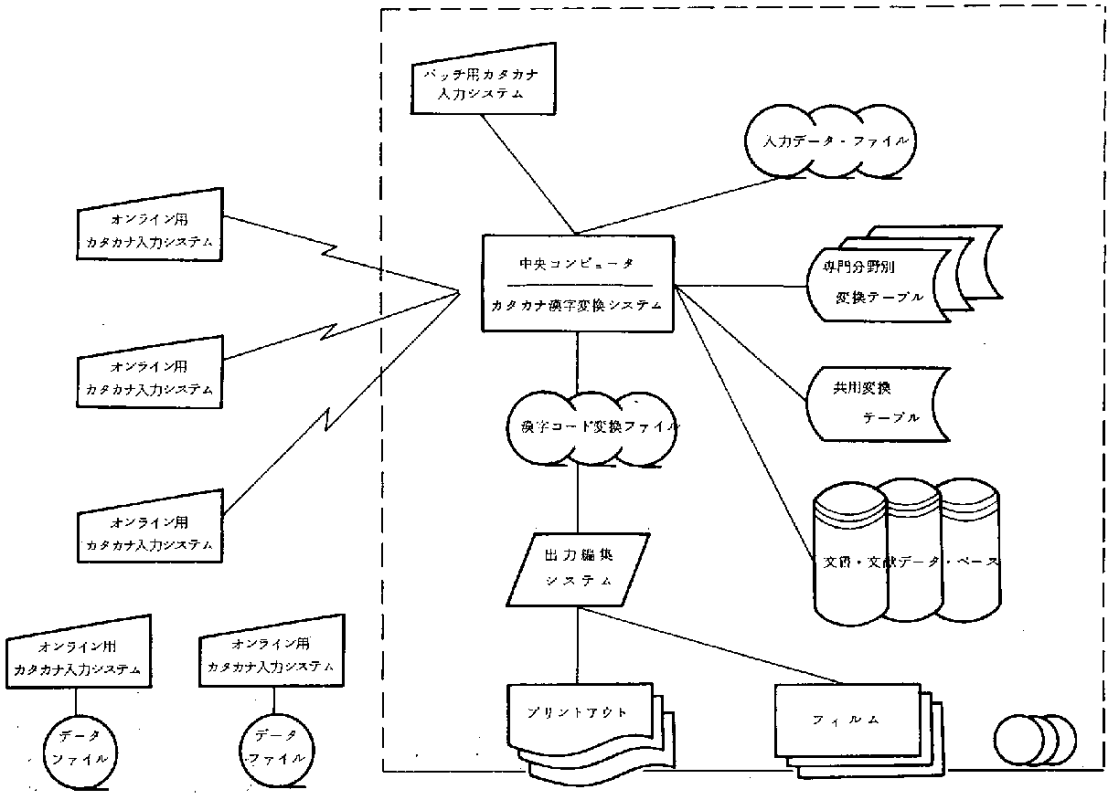
(2) カタカナ漢字変換システム……<処理>

(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に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・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。)



※　このデータファイルを中央コンピュータでバッチ処理する。

図６－１　カタカナ入力方式によるシステム構成の概念図

漢字かな混り文出力編集システムでは、漢字コードに変換したデータを、印刷や出版のために、禁則処理などを含めた編集処理を行なう。

システム構成の概念図は図6－1に示す。

6.2 システムの利点と効果

このカタカナ入力方式は、とくにつぎ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エ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れるので、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繋でき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って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、キーパンチャに過大の負担がかかっているのが現状である。

一般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ャにとっては、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も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、馴じみにくい記号や符号を操作者が覚えなければ

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字で「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おちる。

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みながら、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、言いかえれば、漢字けん盤入力方式のように熟練を要するパンチャを特別に教育する必要がないという利点がある。

(2) 蓄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ビット構成であって、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蓄積できるということは、漢字コードで蓄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、蓄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蓄積するとき、カタカナの入力データに、漢字出力のためのファンクション・コードを挿入して蓄積しておけば、検索情報を漢字かな混り文出力とカタカナ文出力といっ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伝送する場合も、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、高価な漢字かな混り文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エックしたり訂正(校正)するためには、漢字テレタイプからの出力テープを漢字出力装置にかけてモニタ用のゲ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原稿を見ながらパンチしても誤パンチがわからないまま見過したり、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ないながらチエ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。

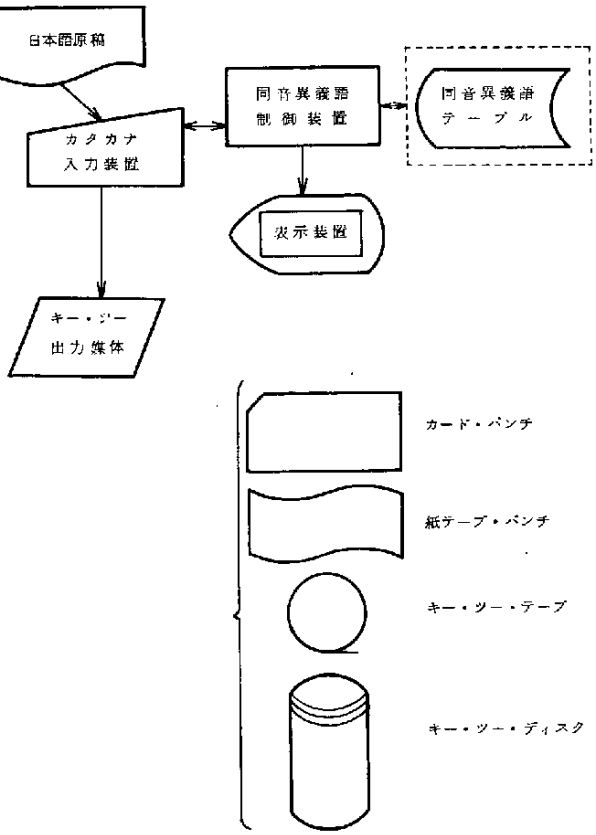


図６－２　カタカナ入力システムの構成

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けん盤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、実験的に漢字キー「＃」、カタカナキー「＊」を設定し、操作者が漢字・仮名を区分する。

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、同時にカタカナけん盤から入力操作を行なうが、この時、原文中（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

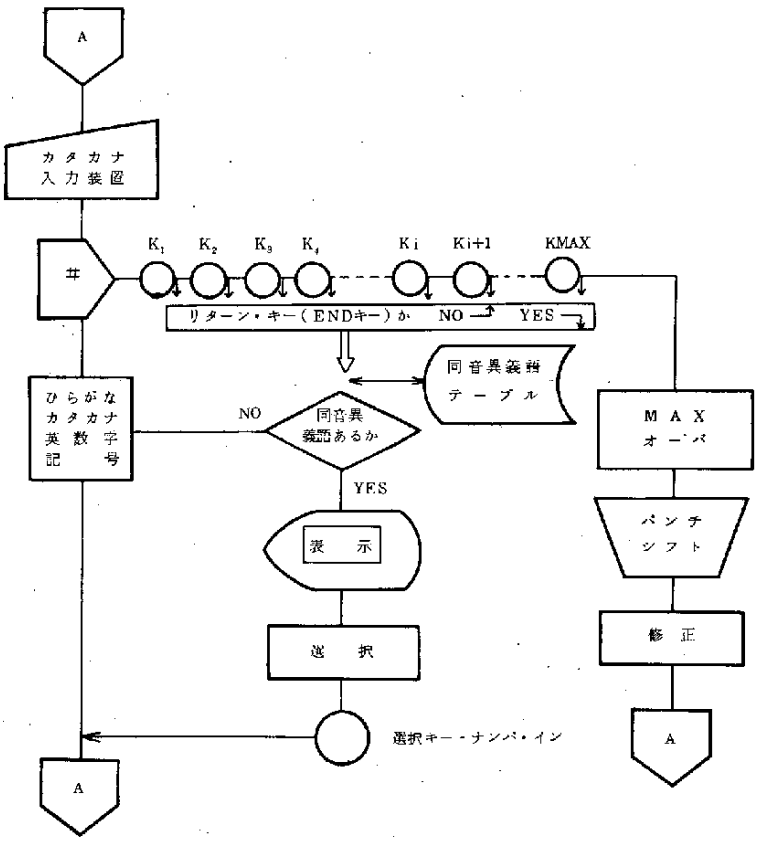
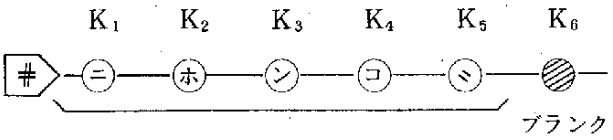


図６－３　カタカナ入力システム操作フロー

ョン・キー（この例では「＃」）を入力し、つづいて、その漢字の読みをカタカナ入力する。漢字は、単語（1～3字の熟語）単位に分割あるいはまとめて1区分とする。1区分のためのリターン・キー（漢字単語のENDキー）は、カタカナ以外ならばいずれのキーを入力してもシステムが自動的に判断する（ENDキーを別に付しても良い）。

（例）〔日本語についての……〕



単語1区分

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エ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エック・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、チエック・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、このアドレス指示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CRT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得策と思われる。

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。

キーボード操作　　（例）　〔情報処理システム……〕

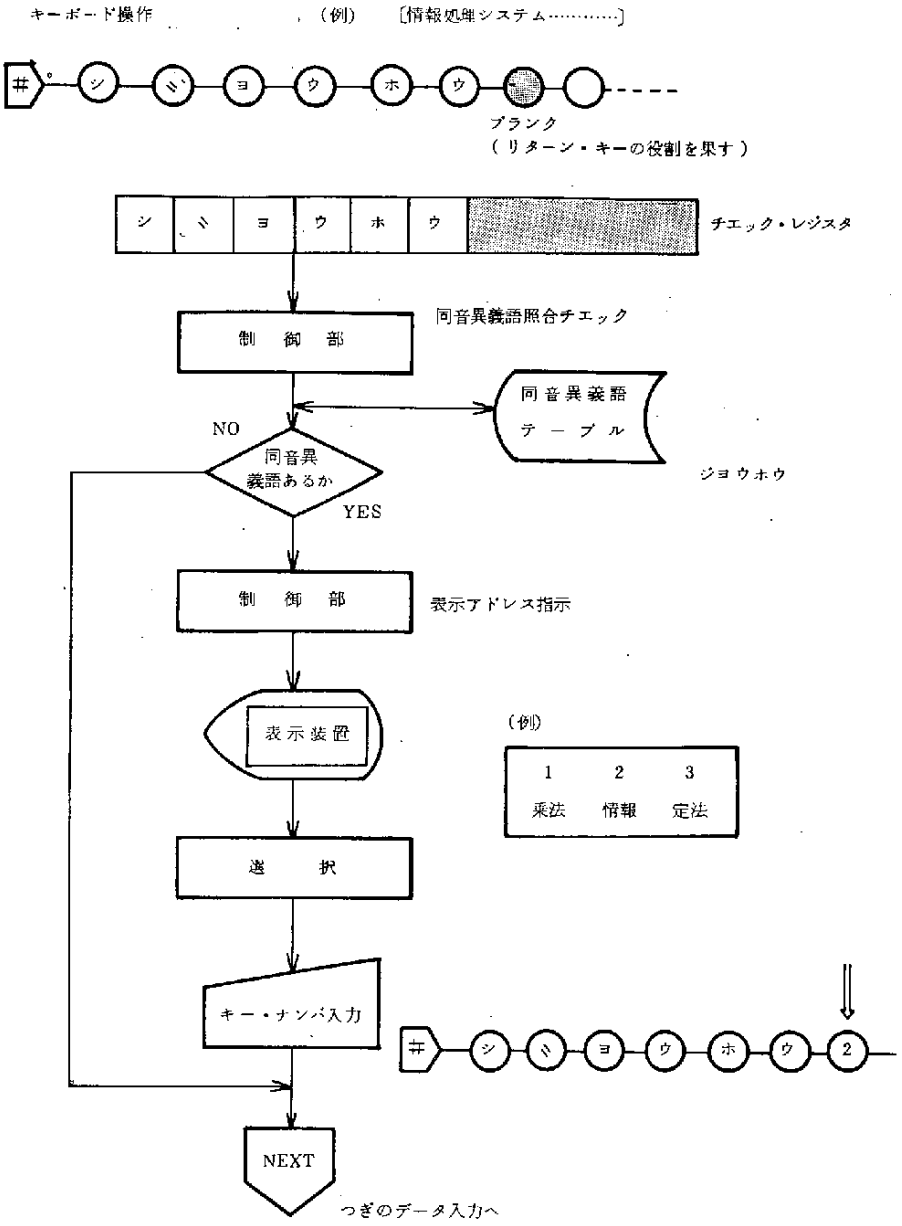


図６－４　同音異義語フロー

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かなければならないが、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章で述べるので、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。

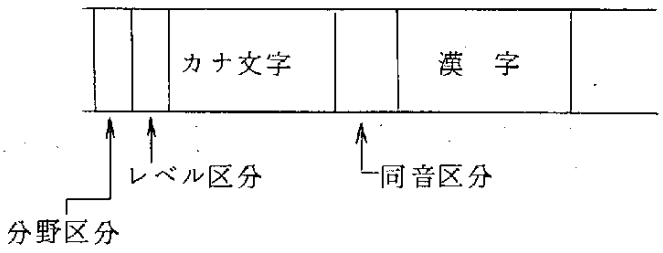
6.4.1　変換単語テーブルの格納処理

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なうために変換単語テーブル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リのチエックを行なう。

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

この最終ファイルは、つぎ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。



① 分野区分……分野別（政治、経済……など分野別変換が可能）

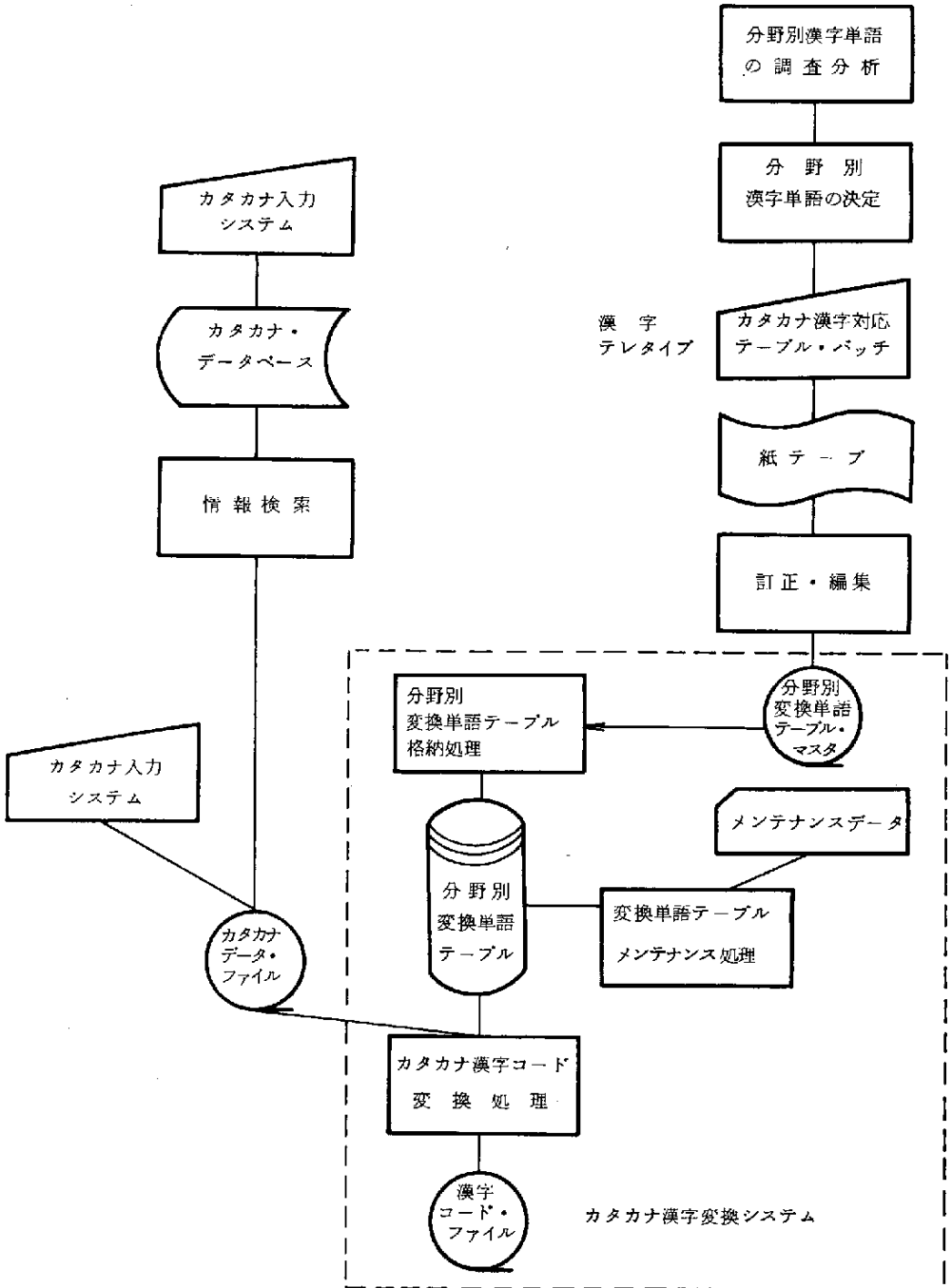

分野別漢字単語
の調査分析
分野別
漢字単語の決定
漢字
テレタイプ
カタカナ漢字対応
テーブル・パッチ
紙テープ
訂正・編集
カタカナ入力
システム
カタカナ・
データベース
情報検索
分野別
変換単語テーブル
格納処理
分野別
変換単語
テーブル・
マスタ
カタカナ入力
システム
メンテナンスデータ
分野別
変換単語
テーブル
変換単語テーブル
メンテナンス処理
カタカナ
データ・
ファイル
カタカナ漢字コード
変換処理
漢字
コード・
ファイル
カタカナ漢字変換システム

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

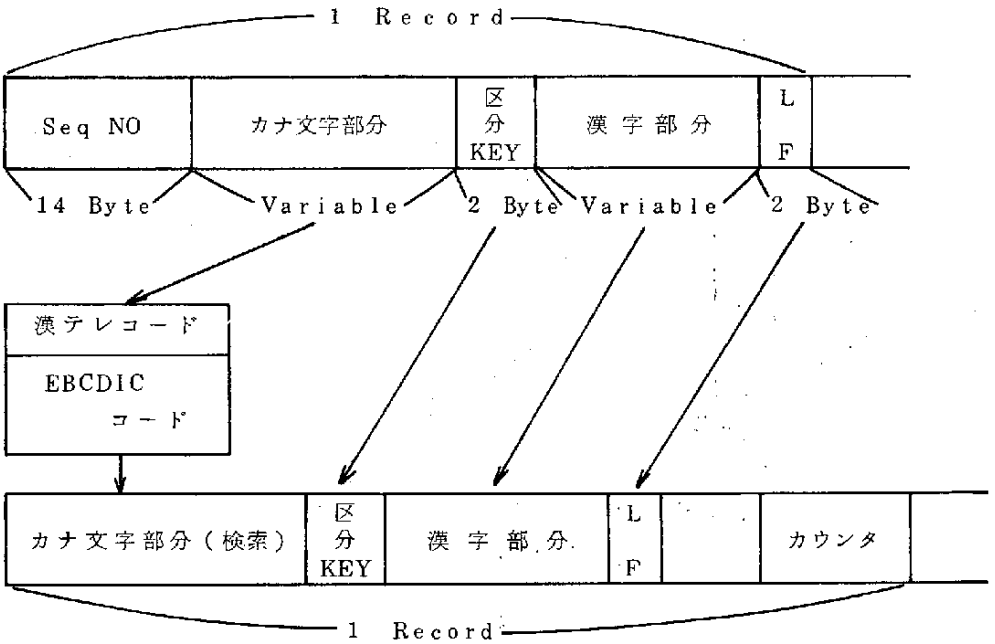
③ 同音区分……同音異義語の処理可能

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

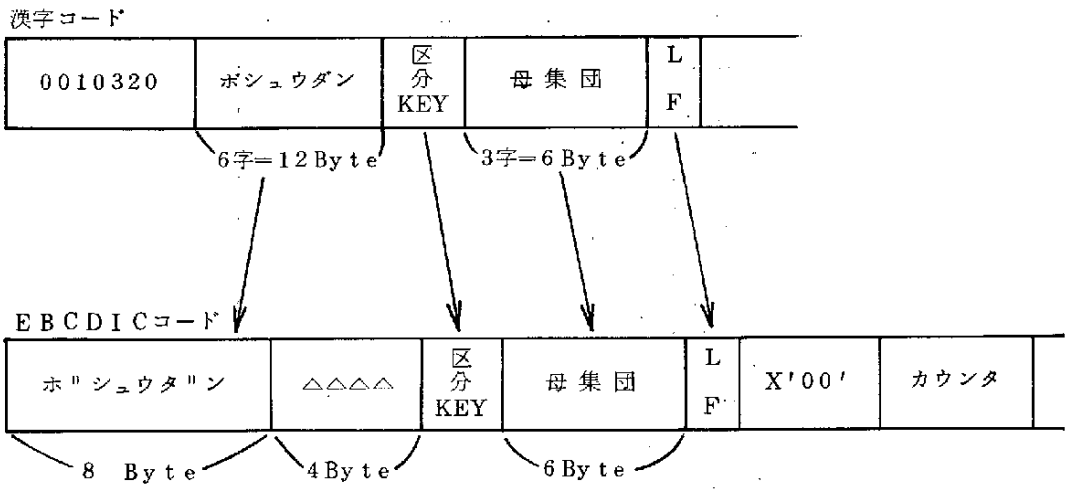
⑤ このファイルは、５０音順に配列され検索しやすい。

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。



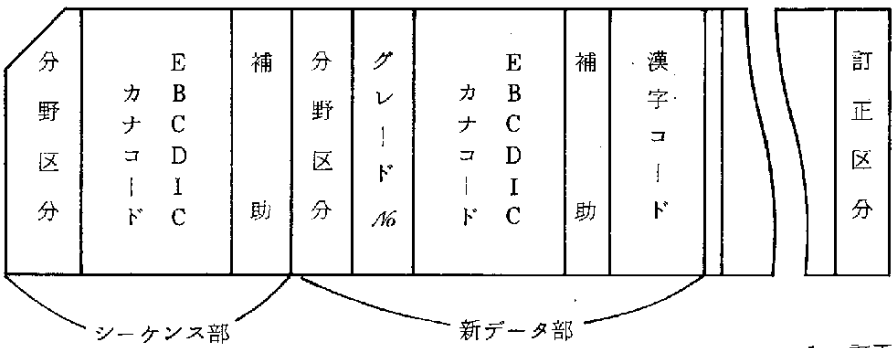
例



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、「赤字校正済プルーフ・リスト」以下の処理によってメンテナンスを行なう。手順に従って概述する。

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エ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パンチされる。



1. 訂正
2. 新規
3. 削除

。

図6－7　訂正カード

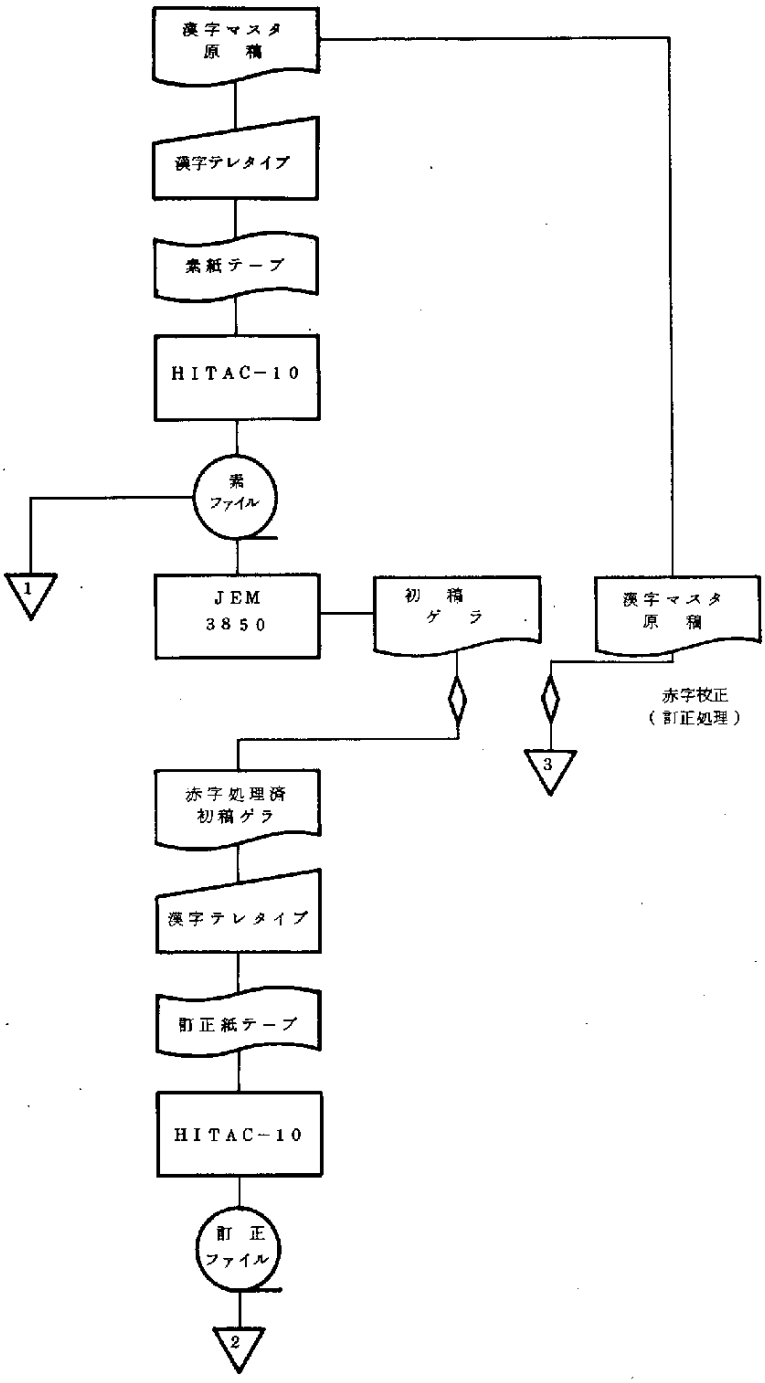
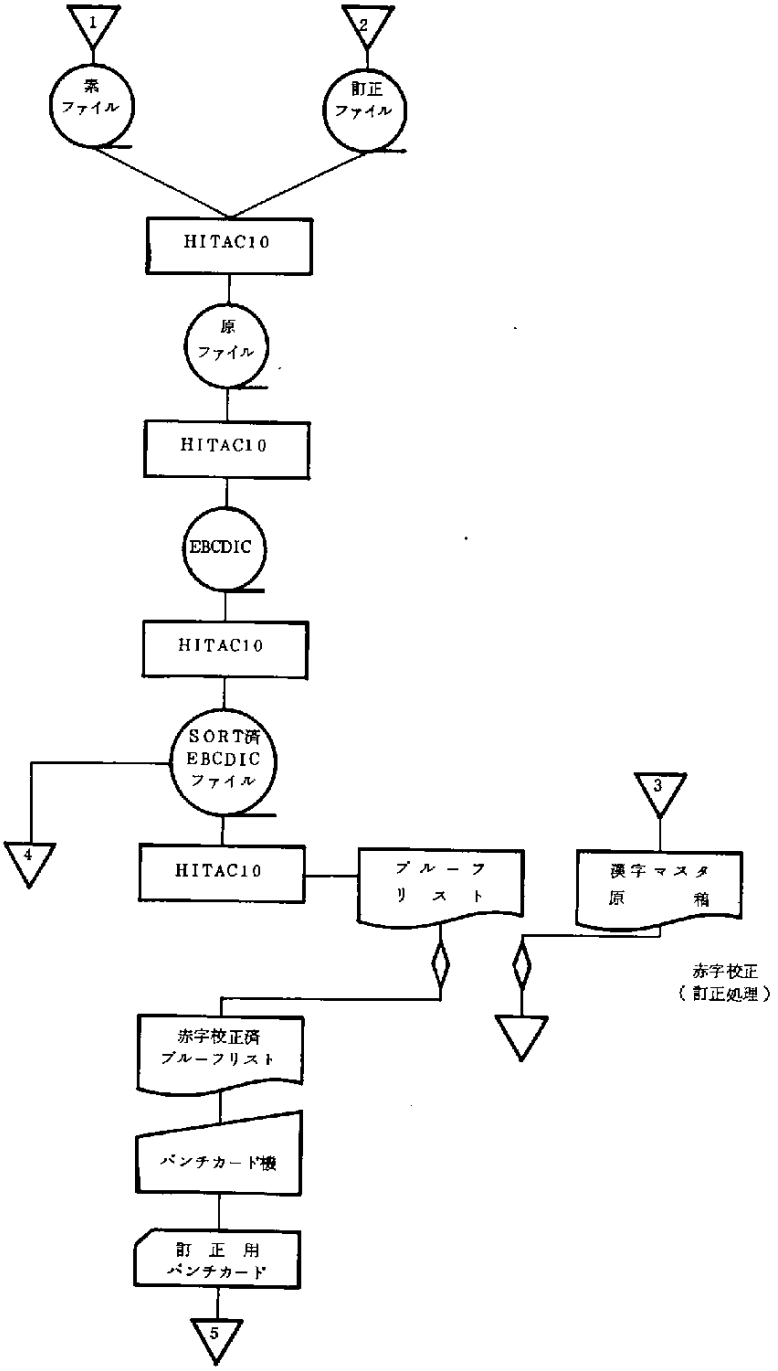
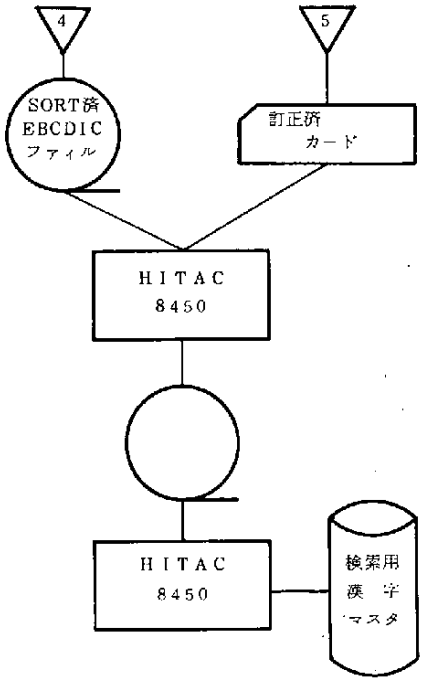

漢字マスタ
原　稿
漢字テレタイプ
素紙テープ
HITAC−10
素
ファイル
1
JEM
3850
初　稿
ゲ　ラ
漢字マスタ
原　稿
赤字校正
（訂正処理）
3
赤字処理済
初稿ゲラ
漢字テレタイプ
訂正紙テープ
HITAC−10
訂　正
ファイル
2

図６－６　漢字マスタ


1
素
ファイル
2
訂正
ファイル
HITAC10
原
ファイル
HITAC10
EBCDIC
HITAC10
SORT済
EBCDIC
ファイル
4
HITAC10
ブルーフ
リスト
3
漢字マスタ
原稿
赤字校正
（訂正処理）
赤字校正済
ブルーフリスト
パンチカード機
訂正用
パンチカード
5



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カタログに分解されて、ディスクにファイルされる。

(注) 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区切りが読みこまれ、その区切りの頭にあるファンクション・キー「＊」、「＃」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。

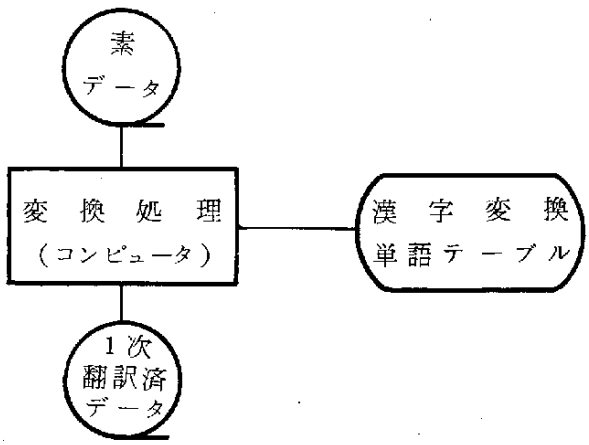


図６－８　コード変換処理

例をあげると

ex1

＃タカチホ　L2000 ヲ　＃ハツピョウ、＊ハードト……

→　高千穂L2000を発表、ハードと……

ex2

＃ツウサン　＃コウギョウ　＃カ　デ　＃カカク　＃ブ　ンリ…

→　通産工業課で価格分離

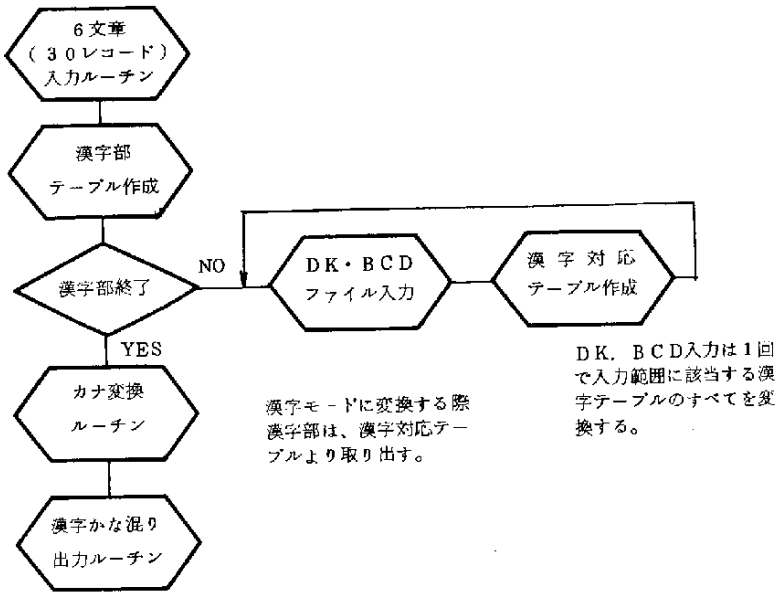


図６－９　コード変換処理フロー概図①

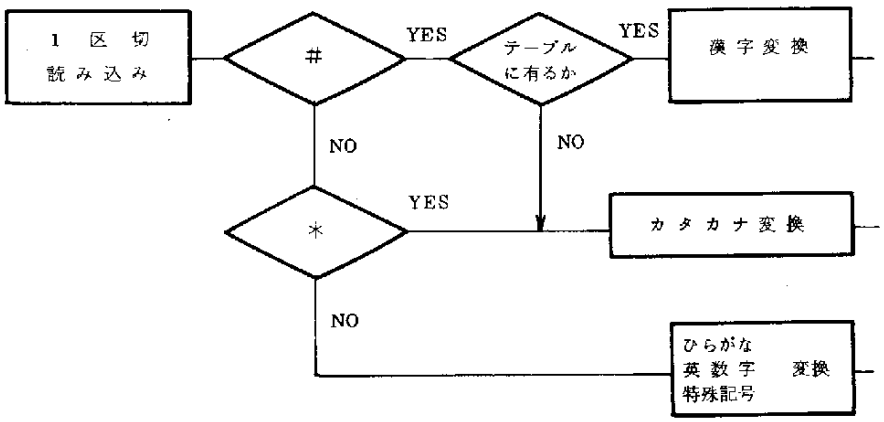


図６－１０　コード変換処理フロー概図②

(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。

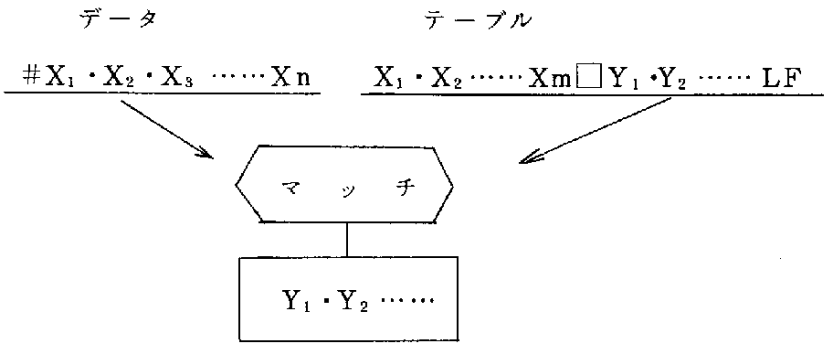


図６－１１　マッチング方法（この場合、検索方法はカタログ方式をとっている。）

(3) 検索におけるカタログ方法について

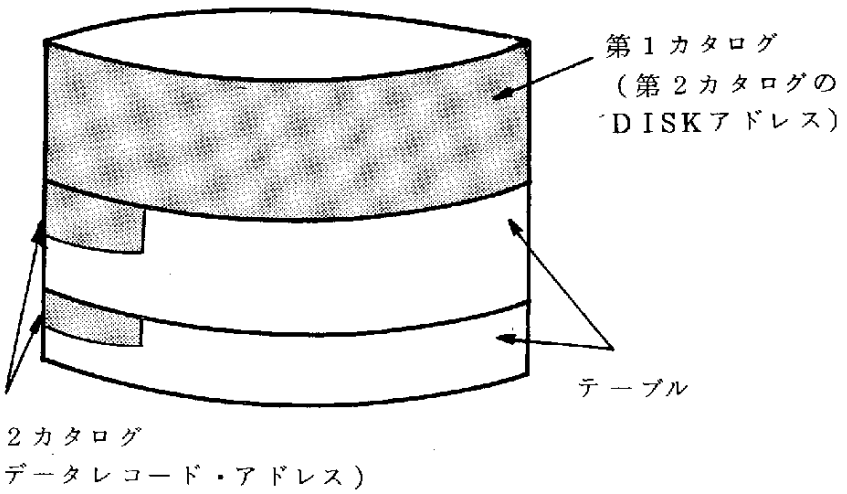


図６－１２　変換単語テーブルのカタログ方法

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。

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になっており、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。

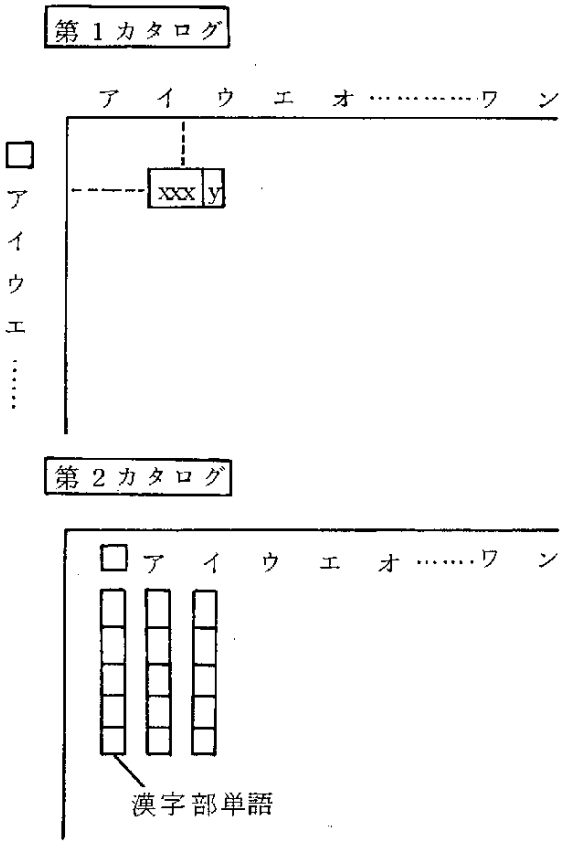


図6－13　カタログの内容

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上のアドレスを検出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。テーブルにないときは、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。

とくに「」、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……

ブランク

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

```
00130#コクサイ2#ツウカ1#シﾞヨウセイ カﾞ #アンテイ スルタメ ノ #セﾞンテイ#シﾞヨウケン ハ #キヨカﾞク ナ #カシﾞヨウ※トﾞル ノ # /
00140ショリ#モンダﾞイ オ, フクメテ, #キシﾞク#ツウカ1※トﾞル ノ #シンニン#カイフク ニ ナンラカ ノ ※メトﾞ オ, ツケルコトデアル, /
00150ダﾞカﾞ,コノヨウナ #カダﾞイ オ, #タン#キカン1 ニ #カイケツ スルコト ハ,#ショハン ノ #シﾞヨウセイ カラミテ キウメテ #コンナ /
00160ン ト イワザﾞル オ, エナイ.@
00170トスレハﾞ,#コクサイ2#ツウカ1#フ1#アンテイ トイウ#シﾞヨウタイ ハ カナリ #チヨウキ#カ1 シ,#トウセﾞン,#エン2 モ マタ #ケﾞ /
00180ンサﾞイ ノ #ヘントﾞウ#ソウハﾞセイ1 オ, #チヨウキ ニ #シﾞソﾞク セサﾞル オ, エナイタﾞロウ.@
00190#ヘントﾞウ#ソウハﾞセイ1 カﾞ #コテイ#セイ1 ニ クラベﾞルト,#ホﾞウエキ#シﾞツム#シﾞヨウ2,#シユシﾞユ ノ #シシヨウ オ, モ/ /
00200ツテイルコト ハ #シﾞシﾞツ デアル.@
00210トクニ #ニホン ノ ヨウニ,#カウセ#カンリ デ カンシﾞカﾞラメ ニ シハﾞラレテイル #シシﾞヨウ ニ オイテハ ソウデアル, シカシ,# /
00220コクミン#ケイサﾞイ ノ #カンテン カラスレハﾞ,#ヘントﾞウ#ソウハﾞセイ1 ニハ ツキﾞノヨウナ #リテン カﾞ アルコト モ ワスレテ ハ /
00230ナラナイ.@
00240ソノ #ダﾞイ1 1 ハ,#ケイサﾞイ#ウンエイ ニ オイテ,#コクサイ2#シユウシ#シﾞヨウ2 ノ #セイヤク カラ #カイホウ サレテ,#セイサ /
00250#センタク ノ #シﾞユウ2#トﾞ カﾞ ヒロカﾞルコト.@
00260#ダﾞイ1 2 ハ, #ユニユウ※インフレ オ, #シヤタﾞン シ,#コクナイ ノ #フﾞツカ#シﾞヨウシヨウ オ, #ヨクセイ スル#コウカ1 オ /
00270, モツコトデアル, #ニホン#ケイサﾞイ ハ コレマテ,#コクサイ2#シユウシ,#フﾞツカ,#フクシ トイウ 3#シﾞユウ2#ク ニ ナヤマサ /
00280レナカﾞラ,#ソウコﾞ ノ #トレートﾞ,オフ #カンケイ カラ #ユウコウ ナ #セイサク#シユダﾞン カﾞ 3スクミ ノ #シﾞヨウタイ ニ ア /
00290ルト サレテキタ.@
00300ダﾞカﾞ,#ヘントﾞウ#ソウハﾞセイ1 オ, トルコトニヨノツテ,#トウメン, #コクサイ2#シユウシ#シﾞヨウ2 ノ #テﾞカセ #アシ カセ カ /
00310ﾞ ハスﾞサレレハﾞ,#セイサク※センタク ノ #ハハﾞ ハ ソレタﾞケ ヒロカﾞル.@
00320シカモ,#コクナイ#フﾞツカ#シﾞヨウシヨウ ノ モノツトモ #ユウリヨク ナ #ヨウイン ト サレテイル #カイカﾞイ※インフレ ノ #シﾞヨウリ /
00330ク カﾞ #シヤダﾞン サレレハﾞ,#セイサク#トウキヨク ト シテハ,#コンコﾞ ノ #ケイサﾞイ#シセツ ノ #シﾞユウテン オ, アケﾞテ # /
00340コクミン#フクシ#コウシﾞヨウ ノ 1#テン ニ オクコト モ #カノウ ナ ハスﾞデアル.@
00350クエテ #イツハﾟン#シﾞツキﾞヨウ#カイ カラモ,#コンカイ ノ #ヘントﾞウ#ソウハﾞセイ1 ニハ,'※フロートスﾞし' シタトイウカ,#セﾞ /
00360ンカイ ホトﾞノ ツヨイ #ヒナン ハ キカレナクナノツタ.@
00370コウシタ #テン オ,#コウリヨ スルト, デキルタﾞケ #カウセ#シﾞユキユウ ニ ソクシタ #エン2 ノ #テキセイ#スイシﾞユン オ, ミツ /
00380ケル #ヒツヨウ#セイ2 カラモ, コノサイ ハ, シノツクリ #コシ オ, スエタ #エン2 ノ #チヨウキ#ヘントﾞウ#ソウハﾞセイ1 オ, /
00390#カクコﾞ シテ, ソノ #セﾞンテイ ニ タノツタ #ケイサﾞイ#ウンエイ カﾞ ノソマレルノデアル.@
00400タﾞ,#ヘントﾞウ#ソウハﾞセイ1 ノ #チヨウキ#カ1 ハ,#コクナイ ノ #フﾞツカ#シﾞヨウシヨウ,#インフレ オ, ハトﾞメスル,#ホシヨ /
00410ウ3 トハカナラスﾞシモ ナラナイ.@
00420ヨクシラレテイルトオリ, #サクネン#アキ カラ #ホンネン ニ カケテ #モウイ オ,フルノツテイル #オロシウリ#フﾞツカ ノ#キユウトウ#ヨウ /
00430イン ニハ, #ユニユウ#フﾞツカ#シﾞヨウシヨウ ノ ホカニモ,#フキヨウ※カルテル#トウ1 ノ #キヨウキユウ#セイケツ#テキ#コウイ,#ケイ /
00440キ1#カイフク ニトモナウ ※シﾞユヨウ ノ #カクダﾞイ,#フﾞモン#カン#シﾞユヨウ ノ ※シフト, イワユル #カシﾞヨウ#リユウトﾞウ#セイ /
00450ク,#チンキﾞン#コスト ノ #ソﾞウダﾞイ ナトﾞカﾞ アルカラデアル.@
00460ツマリ, イマノ #フﾞツカ#シﾞヨウシヨウ ハ, コレラノ #シヨ3#ヨウイン カﾞ #フクザﾞツ ニ カラミア/ツテ オコノツテイルワケテ, /
00470#カイカﾞイ※インフレ ノ #エイキヨウ カﾞ #ヘントﾞウ#ソウハﾞセイ1 デ #シヤダﾞン サレテモ, ナオ ネスﾞヨイ #シﾞヨウシヨウ#ヨ /
00480ウイン カﾞ ノコノツテイレハﾞ, #フﾞツカ#シﾞヨウシヨウ ト ※インフレ#アクセイ#カ1 ノ #キケン ハ サラナイ.@
00490#サイキン ノ#ロシウリ#フﾞツカ#シﾞヨウシヨウ ニハ,#コウエキ1#シﾞヨウケン ノ#キユウアク#カ1 ニ#タンテキ ニ シメサレテイルヨウニ /
00500,#モクサﾞイ,#タﾞイスﾞ,#ヨウモウ ナトﾞノ #ユニユウ#カカク ノ #ホﾞウトウ ニ ヨルトコロカﾞ キワメテオオキイ.@
00510シタカﾞノツテ,コノ#メン カラノ#コクナイ#フﾞツカ#ヨクセイ#コウカ1 ハ#コンコﾞ,#シﾞヨシﾞヨ ニ テﾞテクルモノト キタイサレルカﾞ,# /
00520トﾞウシﾞ1 ニ#フﾞツカ#アンテイ#カ1#タイサク トシテ,#セﾞンキ#シヨ3#ヨウイン オ, #シﾞヨキヨ ナイシ #カンワ スルヨウナ #シサ /
00530ク カﾞ #セツキヨク#テキ ニ トラレル #ヒツヨウ カﾞアル, ※インフレ ノ #コウシン ハ トリモナオサスﾞ #コクミン #フクシ ノ #ヒ /
00540テイ ニモ ツナカﾞルカラテﾞアリ, #ヘントﾞウ#ソウハﾞセイ1#カ ニ オイテモ,#インフレ#ヨクセイ カﾞ #イセﾞン,#シﾞユウタﾞイ ナ /
00550#セイサク#カタﾞイ トシテ ノコルコト ハ #ロン オ, マタナイ.@
00560ソノ#イミ デ #セイサク#トウキヨク カﾞ,#ヨキン#シﾞユンヒﾞ#リツ ノ サイキリアクﾞ ヤ #キﾟンコウ ノ#マトﾞクチ#キセイ1 ナト /
00570ﾟニヨノツテ,#キンユウ#セイサク ノ #キチヨウ#テンカン ニ ノリタﾞシタコト ハ #カシﾞヨウ#リユウトﾞウ#セイ2#キユウシユウ ノ タメニ /
00580 #サンセイ デキル.@
00590ダﾞカﾞ,#サクネン 10-12#カﾞツ ノ ※GNP#シﾞツシツ#セイチヨウ#リツ カﾞ #セﾞンネン#トﾞウキ#ヒ4 12.7% ニ ナノツタコ /
00600ト オ, ハシﾞメ,#ホンネン 1#カﾞツ マテ ノ#ケイキ1#ショ3#シヒヨウ カﾞ ツキナミ オオハハﾞ ナ #シﾞヨウシヨウ オ, シメシタ /
00610コト オ, モノツテ,#ケイキ1#カネツ ハ#ヒツシ ト #ハンタﾞン シ#コウテイ2#フﾞアイ ノ ヒキアケﾞ オ, #ケントウ シテイルトスレハ /
00620ﾟ,#モンダﾞイ カﾞ アル.@
00630アセﾞナラ,#ダﾞイ1 1 ニ,#サクネン#アキ カラ #ホンネン 1#カﾞツ ニ カケテ ノ#ケイキ1#シヒヨウ#オオハハﾞ#コウテン ハ,#セﾟ /
00640ンネン#トﾞウキ ノ #ケイキ1 カﾞトﾞンゾﾞコ #シﾞヨウタイ ニ アノツタコト ノ #ハンエイ デモアルカラデ,#ケイキ1 カﾞコノママ /
00650カネツ#シﾞヨウタイ ニ #トツニユウ スルカトﾞウカ ハ, モウシハﾞラク ノ #スイイ オ, ミナケレ ハﾞ ナラナイ.@
00660ダﾞイ1 2 ニ,#コンコﾞ ハ #ヘントﾞウ#ソウハﾞセイ1#イコウ ニ トモナウ ※デﾞフレ #コウカ カﾞ #ヨソウ サレルコトデアル, /
00670トクニ #オウシユウ#ツウカ1#シﾞヨウセイ ノ #リユウトﾞウ#カ1 ニ ヨノツテ,#エン2#ソウハﾞ ハ #トウシヨ#ヨソウ オ, #オオハハﾞ /
00680 ニ ウウマワル #エンタﾞカ#シﾞケﾞン カﾞ#ヒツシ ト ミラレルタﾞケニ, ソノ #エイキヨウ ハ #ユシユツ#カンレン#サンキﾞヨウ オ, /
00690#チユウシン ニ #ケイシ デキヌモノカﾞアロウ.@
00700コノヨウナ #シﾞヨウタイ#カ テ,#ホンカク#テキ ナ #キンユウ ヒキシメ,#リウ1#シﾞユヨウ#セイサク ニ ウノツタエレハﾞ,#カシﾞヨウ /
```

図 6 - 14(a) カタカナ入力データ例

00710#ﾘﾕｳﾄﾞｳ#ｾｲ2 ﾉ #ｷﾕｳｼﾕｳ ﾆﾊ #ﾕｳｺｳ ﾃﾞｱﾉﾂﾃﾓ,#ﾌﾞﾂｶ ﾊ ｿﾚﾎﾄﾞｻｶﾞﾗｽﾞ, ﾑｼﾛ #ｾｲｻﾝ ｶﾞ # /
00720ﾃｲｶ2 ｼﾃ,#ｵﾛｼｳﾘ#ﾌﾞﾂｶ#ｼﾞﾖｳｼﾖｳ ｵ, #ｼｹﾞｷ ｼｶﾈﾅｲﾄｵﾓﾜﾚﾙ.@
00730ｺﾉ #ｻｲ ﾊ,#ﾕｼﾕﾂ#ｶﾞﾀ2#ｻﾝｷﾞﾖｳ ﾉ #ﾅｲｼﾞﾕ#ﾃﾝｶﾝ ｵ, #ｿｸｼﾝ ｽﾙﾀﾒﾆ, ﾑｼﾛ #ｶﾉｳ ﾅ ｶｷﾞﾘ # /
00740ｺｸﾅｲ ﾉ #ﾌｸｼ#ｶﾞﾀ2#ｼﾞﾕﾖｳ ﾉ #ｶｸﾀﾞｲ ﾔ #ﾕﾆﾕｳ#ｿｸｼﾝ ﾆ #ｼﾞﾕｳﾃﾝ ｵ, ｵｲﾀ #ｾｲｻｸ ｵ, #ｼｺ /
00750ｳ2 ｽﾍﾞｷﾃﾞｱﾛｳ.@
00760#ｻﾝｷﾞﾖｳ#ﾃﾝｶﾝ ﾉ ﾀﾒﾉ ｲﾜﾕﾙ ﾏｴﾑｷ ﾉ #ﾁﾖｳｾｲ#ｴﾝｼﾞﾖ,#ﾛｳﾄﾞｳ#ﾘﾖｸ#ﾘﾕｳﾄﾞｳ#ｶ1 ﾉ ﾀﾒﾉ #ｼﾖ /
00770ｸｷﾞﾖｳ#ｸﾝﾚﾝ ﾅﾄﾞﾉ #ｼｻｸ ｶﾞ #ｾﾂｷﾖｸ#ﾃｷ ﾆ ﾄﾗﾚﾙ #ﾋﾂﾖｳ ｶﾞ ｱﾙｺﾄ ﾊ ｲｳﾏﾃﾞﾓﾅｲｶﾞ,#ｵｳﾍﾞｲ /
00780#ｼﾖｺｸ ﾆ ｸﾗﾍﾞﾃ #ﾃｲﾘﾂ ﾉ #ﾎｳｼﾞﾝ#ｾﾞｲ ﾉ #ｿﾞｳｾﾞｲ ﾔ #ｼﾖﾄｸ#ｹﾞﾝｾﾞｲ ﾅﾄﾞﾓ ｶﾝｶﾞｴﾗﾚﾃﾖｲ.
00790#ﾖｳ ﾊ,#ﾍﾝﾄﾞｳ#ｿｳﾊﾞｾｲ1#ｶ ﾆ ｵｲﾃ,#ﾆﾎﾝ#ｹｲｻﾞｲ ﾉ #ﾃｷｵｳ#ﾘﾖｸ,#ｾﾝｻﾞｲ#ﾉｳﾘﾖｸ ｵ, #ｼﾞﾕｳﾌﾞ /
00800ﾞﾝ ﾆ'ｲｶｼﾅｶﾞﾗ,#ｹｲｻﾞｲ#ﾀｲｼﾂ ｵ, ｺﾚﾏﾃﾞ ﾉ #ﾕｼﾕﾂ#ｶﾞﾀ2 ｶﾗ #ﾌｸｼ#ｶﾞﾀ2 ﾍﾄ #ｶｶﾝ ﾆ #ﾍﾝｶ /
00810ﾝ ｼﾃｲｸｺﾄ ｶﾞ #ｾｲｻｸ ﾉ #ｷﾎﾝ ﾃﾞｱﾙ.@
00820#ｹﾞﾝｲﾝ ﾊ ｻﾏｻﾞﾏﾀﾞｶﾞ,ｲﾏ #ｾｶｲ ﾉ #ｼﾖｸﾘﾖｳ#ｼﾞｼﾞﾖｳ ﾊ 2#ﾈﾝ ﾂﾂﾞｷ ﾉ #ﾌｻｸ ﾃﾞ #ｼﾞﾕｷﾕｳ /
00830ｶﾞ ﾋﾉﾂﾊﾟｸ ｼ, ﾜｶﾞ #ｸﾆ ﾉ #ﾉｳｾｲ ﾉ #ｷﾎﾝ ﾓ ｶﾝｶﾞｴﾅｵｽﾍﾞｷﾀﾞﾄﾕｳｺｴ ｶﾞ ｲﾁﾌﾞ ﾆ ｱｶﾞﾉﾂﾃｲ /
00840ﾙ.@
00850ｺﾉﾀﾋﾞ #ﾊﾂﾋﾟﾖｳ ｻﾚﾀ #ﾉｳｷﾞﾖｳ#ﾊｸｼﾖ ﾓ,#ﾉｵｿﾝ#ﾃｷ#ﾌｳﾄﾞ ﾉ #ﾎｾﾞﾝ ﾄﾄﾓﾆ,#ｾｶｲ ﾉ #ﾉｵｻﾝ#ﾌ /
00860ﾞﾂ#ｼｼﾞﾖｳ ﾉ #ﾌ1#ｱﾝﾃｲ#ｾｲ2 ﾆ #ﾀｲｵｳ ｼﾃ #ｼﾖｸﾘﾖｳ#ｼﾞｷﾕｳ#ﾄﾞ ﾉ ﾋｷｱｹﾞ ｵ, #ｷﾖｳﾁﾖｳ ｼﾃｲ /
00870ﾙ.@
00880ｼｶｼ, ｺﾚｶﾗ ﾉ #ﾉｳｾｲ ﾊ ﾀﾀﾞﾀﾝ ﾆ #ｼﾖｸﾘﾖｳ ﾉ #ｼﾞｷﾕｳ#ｼﾞｿｸ ｵ, ﾒｻﾞｽﾉﾃﾞﾊﾅｸ,#ﾃｷﾁ#ﾃｷｻﾝ /
00890ｵ, #ﾁﾕｳｼﾝ ﾄｼﾀ#ｺｳ1#ﾉｳﾘﾂ.#ﾃｲ1※ｺｽﾄ ﾉ#ｾｲｻﾝ1#ﾀｲｾｲ ｵ, #ｿｳｷﾕｳ ﾆ #ｶｸﾘﾂ ｽﾙﾉｶﾞ ﾓﾉﾂﾄﾓ /
00900 #ｼﾞﾕｳﾖｳ ﾅ #ｶﾀﾞｲ ﾃﾞｱﾙ.@
00910#ｾｶｲ#ｼﾖｸﾘﾖｳ#ﾉｳｷﾞﾖｳ#ｷｺｳ ﾉ #ｽｲﾃｲ ﾆ ﾖﾙﾄ,#ﾎﾝﾈﾝ ﾆ ｵｹﾙ #ｾｶｲ ﾉ #ｼﾖｸﾘﾖｳ#ｾｲｻﾝ1 ﾊ#ｷﾖ /
00920ｳｻｸ ﾃﾞｱﾉﾂﾀ #ｻｸﾈﾝ ﾄ ﾎﾎﾞｵﾅｼﾞｸﾞﾗｲ ﾉ ﾐｺﾐﾃﾞ,#ｹﾂｷﾖｸ 3#ﾈﾝ ﾂﾂﾞｷ ﾉ #ﾌｻｸ ｵ, #ﾖｿｳ ｼﾃｲ /
00930ﾙ.@
00940#ｺｸﾙｲ ﾉ #ﾌｻｸ ﾊ #ﾄﾞｳｼﾞ1 ﾆ#ｼﾘﾖｳ ﾉ #ｷﾖｳｷﾕｳ#ﾌﾞｿｸ ﾆﾓﾂﾅｶﾞﾘ,#ﾁｸｻﾝ#ﾌﾞﾂ ﾉ #ｾｲｻﾝ1 ﾆｵ /
00950ｵｷﾅ #ｼｼﾖｳ ｵ, ｵﾖﾎﾞｽ.@
00960#ﾉｵﾘﾝ#ｼﾖｳ ﾊ #ｻｸﾈﾝ 10#ｶﾞﾂ ﾆ "#ﾉｵｻﾝ#ﾌﾞﾂ#ｼﾞﾕｷﾖｳ ﾉ #ﾃﾝﾎﾞｳ ﾄ#ｾｲｻﾝ1#ﾓｸﾋﾖｳ ﾉ #ｼｱﾝ /
00970" ﾅﾙﾓﾉ ｵ, #ｺｳﾋﾖｳ ｼﾀ.@
00980ｺﾚﾊ,※FAO ﾉ "1980#ﾈﾝ ﾉ #ｼﾖｸﾘﾖｳ#ｼﾞﾕｷﾕｳ ﾉ #ﾃﾝﾎﾞｳ" ﾔ ※ﾛｰﾏ.ｸﾗﾌﾞ ﾉ "#ｾｲﾁﾖｳ ﾉ #ｹﾞ /
00990ﾝｶｲ" ﾅﾄﾞﾆ ｼﾒｻﾚﾃｲﾙ #ｼﾖｸﾘﾖｳ#ｼﾞｼﾞﾖｳ ﾆ ﾀｲｽﾙ #ﾋｶﾝ#ﾃｷ ﾅ ﾐｶﾀ ｵ, #ﾈﾝﾄｳ ﾆ ｵｲﾃ, ﾜｶﾞ /
01000#ｸﾆ ﾉ #ｼﾖｸﾘﾖｳ#ｼﾞｷﾕｳ#ﾘﾂ ｵ, ﾀｶﾒﾃｲｸｺﾄ ｵ, ﾈﾗｲﾄｼﾃｲﾙ.@
01010#ｺﾝｶｲ ﾉ #ﾉｳｷﾞﾖｳ#ﾊｸｼﾖ ﾃﾞﾓ #ﾐｷﾞ ﾉ ﾖｳﾅ #ﾓﾝﾀﾞｲ#ｲｼｷ ｵ, ﾓﾁﾅｶﾞﾗ #ﾉｳｻﾝ#ﾌﾞﾂ#ｼﾞﾕｷﾕｳ, /
01020#ﾉｳｷﾞﾖｳ#ｹｲｴｲ,#ﾉｳｷﾞﾖｳ#ｹｲｻﾞｲ ﾉ #ﾄﾞｳｺｳ ｵ, #ﾌﾞﾝｾｷ ｼﾃｲﾙ.@
01030#ﾉｳｷﾞﾖｳ#ｾｲｻﾝ1 ﾊ#ﾉｳｻﾝ#ﾌﾞﾂ#ﾕﾆﾕｳ ﾉ #ｿﾞｳﾀﾞｲ,#ｺﾒ ﾉ#ｾｲｻﾝ1#ﾁﾖｳｾｲ ﾔ #ﾌﾂｳ#ﾊﾀｹ#ｻｸﾓﾂ /
01040, #ｹﾞﾝｻﾝ ﾅﾄﾞﾆﾖﾉﾂﾃ 45,46#ﾈﾝ ﾄ #ｺｳﾀｲ#ｹｲｺｳ ｵ, ﾀﾄﾞﾉﾂﾀ.@
0105047#ﾈﾝ ﾉ #ﾉｳｷﾞﾖｳ #ｾｲｻﾝ1 ﾊ#ｼﾞﾔﾂｶﾝ ﾉ ﾓﾁﾅｵｼ ｶﾞ ｷﾀｲｻﾚﾙﾓﾉﾉ, ｲｾﾞﾝﾄｼﾃ #ﾃｲ1#ｽｲｼﾞﾕﾝ /
01060ﾆ ｱﾙ. #ﾉｳﾖｳ#ﾁ2 ﾉ #ﾘﾕｳﾄﾞｳ#ｶ1 ﾊ #ﾁｶ ﾉ #ｼﾞﾖｳｼﾖｳ ﾆ ﾖﾉﾂﾃ #ｿｶﾞｲ ｻﾚ,#ﾉｳﾁ#ｼﾖﾕｳ#ｹﾝ /
01070 ﾉ #ﾕｳｼﾖｳ#ｲﾃﾝ ﾊ #ｹﾞﾝﾀｲ ｵ, ﾂﾂﾞｹﾃｲﾙ.@
01080ﾏﾀ,#ｺﾒ ﾉ#ｾｲｻﾝ1#ﾁﾖｳｾｲ ﾆﾖﾙ #ｷﾕｳｺｳ#ﾃﾞﾝ ﾉ #ﾊﾝﾌﾞﾝ ﾁｶｸ ﾊ #ｺｳﾊｲ ｼﾀﾏﾏ ﾆ #ﾎｳﾁ ｻﾚ, ｿ /
01090 ｼﾃｲﾙ. ﾄｸﾆ #ﾉｳｷﾞﾖｳ ﾉ #ｼﾖｸｷﾞﾖｳ#ｼﾞﾝｺｳ ﾉ #ｹﾞﾝｼﾖｳ ｶﾞ ﾒﾀﾞﾉﾂﾃｲﾙ, #ﾉｳｷﾞﾖｳ ﾉ #ｼﾖｸｷ /
01100ﾞﾖｳ#ｼﾞﾝｺｳ ﾉ #ｹﾞﾝｼﾖｳ#ﾘﾂ ﾊ 40#ﾈﾝ ｶﾗ 44#ﾈﾝ#ﾍｲｷﾝ ﾃﾞﾊ 3.2% ﾃﾞｱﾉﾂﾀﾉｶﾞ,45#ﾈﾝ ﾆﾊ 5 /
01110.9%#ｹﾞﾝ,46#ﾈﾝ ﾆｲﾀﾉﾂﾃﾊ 9.6% ﾉ #ｵｵﾊﾊﾞ#ｹﾞﾝｼﾖｳ ﾄﾅﾉﾂﾀ.@
01120ｺﾉ #ｹﾞﾝｼﾖｳ#ｹｲｺｳ ﾊ #ｺｳﾚｲ#ｼﾖｸｷﾞﾖｳ#ｼﾔ1 ﾉ#ｲﾝﾀｲ,#ｺｳｼﾞﾖｳ2 ﾉ#ﾁﾎｳ#ﾌﾞﾝｻﾝ ﾆﾖﾙ #ﾃﾝｼﾖｸ /
01130 ﾅﾄﾞﾆﾖﾘ,#ｺﾝｺﾞ ﾄﾓﾂﾞｸﾓﾉ ﾄ ﾐﾗﾚﾙ. ｺﾚﾆ ﾄﾓﾅﾉﾂﾃ #ｾﾞﾝﾀｲ ﾉ #ﾉｳｶ#ｺｽｳ ﾓ ｼﾀﾞｲ ﾆ ﾍﾘﾂﾂｱ /
01140ﾙ. ﾄｺﾛﾃﾞ,#ｼﾖｳﾋ#ｼﾔ1 ｶﾞﾜｶﾗﾐﾀ#ｼﾖｸﾘﾖｳ#ﾋﾝ#ｶｶｸ ﾉ #ｼﾞﾖｳｼﾖｳ ﾊ ｶﾅﾘ #ｵｵﾊﾊﾞ ﾃﾞ ｱﾘ #ｶｹ /
01150ｲ ﾉ ｵｵｷﾅ #ｱﾂﾊﾟｸ#ｻﾞｲﾘﾖｳ ﾄ ﾅﾉﾂﾃｲﾙ. ｼｶｼ, #ﾉｳｻﾝ#ﾌﾞﾂ ﾉ#ｾｲｻﾝ1#ｼﾔ1#ｶｶｸ ﾊ,#ｼﾖｳﾋ#ｼﾔ /
011601#ｶｶｸ ﾖﾗｾﾞﾉﾂﾄ#ｼﾞﾖｳｼﾖｳ#ﾘﾂ ｶﾞ ﾋｸｲ.@
01170ｺﾚﾊ #ﾎｳｿｳ#ﾋ1,#ﾕｿｳ#ﾀﾞｲ,#ﾊﾝﾊﾞｲ#ﾃﾝｲﾝ ﾉ #ﾁﾝｷﾞﾝ #ｼﾞﾖｳｼﾖｳ ﾅﾄﾞ #ﾘﾕｳﾂｳ#ｹｲﾋ ﾆﾓ,#ﾓﾝﾀ /
01180ﾞｲ ｶﾞ ｵｵｸｱﾙｺﾄ ｵ, ｼﾒｼﾃｲﾙ. #ｼﾖｳﾋ#ｼﾔ1 ﾉ#ｼﾖｸﾘﾖｳ#ｼﾞﾕﾖｳ ﾊ #ｼﾖﾄｸ#ｽｲｼﾞﾕﾝ ﾉ #ｼﾞﾖｳｼﾖ /
01190ｳ ﾄﾄﾓﾆ #ﾋｼ ｵ, ｵﾉﾂﾃ #ｿﾞｳﾀﾞｲ ｼﾃｲﾙ.@
01200ｿﾉﾅｶﾃﾞ #ﾄｸﾁﾖｳ#ﾃｷ ﾅ ｺﾄﾊ, ﾃﾞﾝﾌﾟﾝ #ｼﾂ#ｼﾖｸﾘﾖｳ ﾉ #ﾜﾘｱｲ ｶﾞ ﾋｷﾂﾂﾞｲﾃ#ﾃｲｶ2 ｼ,#ﾄﾞｳﾌﾞ /
01210ﾂ#ｾｲ2 ﾀﾝﾊﾟｸ #ｼﾂ ﾉ #ﾋｼﾞﾕｳ ｶﾞ ﾀｶﾏﾉﾂﾃ ｲﾙｺﾄﾀﾞ.@
01220#ﾄﾞｳﾌﾞﾂ#ｾｲ2 ﾀﾝﾊﾟｸ ﾉ ﾅｶﾃﾞﾊ #ｳｼ #ﾌﾞﾀﾆｸ ﾅﾄﾞﾉ #ﾋﾖｳﾋ#ﾘﾖｳ1 ﾊ#ｿﾞｳﾀﾞｲ ｼﾃｲﾙ ｶﾞ #ｷﾞﾕ /
01230ｳﾆﾕｳ,#ﾆﾕｳｾｲ#ﾋﾝ,#ｹｲﾗﾝ ﾅﾄﾞﾆ ﾀｲｽﾙ #ｼﾞﾕﾖｳ ﾊ #ﾃｲﾀｲ ｷﾞﾐﾃﾞｱﾙ. ｻﾗﾆ,#ｼﾖｸ#ｾｲｶﾂ ﾉ #ｶﾝ /
01240ｲ#ｶ1 ﾆ ﾖﾉﾂﾃ #ｶｺｳ#ｼﾖｸﾘﾖｳ#ﾋﾝ ﾉ #ｼﾖｳﾋ ﾔ #ｶﾞｲｼﾖｸ ｶﾞ #ｷﾕｳｿﾞｳ ｼﾃｲﾙ #ﾃﾝ ﾆﾓ #ﾁﾕｳﾓｸ /
01250 ｼﾃｵｸ #ﾋﾂﾖｳ ｶﾞｱﾙ.@
01260#ｾﾞﾝｼﾞﾕﾂ ﾉ ﾖｳﾆ #ﾉｳｻﾝ#ﾌﾞﾂ ﾉ #ｾｲｻﾝ#ｼﾔ1#ｶｶｸ ﾊｺｺﾉﾄｺﾛ ｵﾁﾂｲﾀ ｳｺﾞｷ ｵ, ﾐｾﾃｲﾙ ｶﾞ,#ｺ /
01270ｸｻｲ2#ﾋｶｸ ｵ,ｼﾃｲﾙ ﾄ ﾜｶﾞ #ｸﾆ ﾉ #ﾊﾞｱｲ ﾊ ｲｾﾞﾝﾄｼﾃ #ｿｳﾄｳ ﾆ #ﾜﾘﾀﾞｶ ﾃﾞｱﾙ.ﾄﾆ,#ｴﾝ2 ｷ /
01280ﾘｱｹﾞ ﾆ ﾖﾉﾂﾃ ｺﾉ #ｶｸｻ ﾊ ﾏｽﾏｽ #ｶｸﾀﾞｲ ｼﾃｷﾃｲﾙ.@
01290ｼﾀｶﾞﾉﾂﾃ #ﾉｳｻﾝ#ﾌﾞﾂ ﾉ #ﾕﾆﾕｳ#ﾜﾘｱﾃ ﾉ #ｿﾞｳｶ, ｱﾙｲﾊ #ﾕﾆﾕｳ ﾉ#ｼﾞﾕｳ1#ｶ1 ﾊ ﾜｶﾞ #ｸﾆ ﾉ /

国際通貨じょうせいが安定するための前提条件はきょがくな過剰ドルのしより問題をふくめて，きじく通貨ドルのしんにん回復になんらかのメドをつけることである．だが，このようなかだいを　短期間にかいけつすることは，しよはんのじょうせいからみてきわめてこんなんといわざるを　えない。
とすれば，国際通貨不安定という常態はかなり長期化し，とうぜん，円もまた現在の変動相場制を長期に持続せざるを　えないだろう。
変動相場制が固定制にくらべると，貿易じつむ上，しゆじゆのししようを　もっていることはじじつである。とくににほんのように，為替管理でかんじがらめにしばられている市場においてはそうである。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なりてんがあることもわすれてはならない。
その第1は，経済運営において，国際収支上のせいやくからかいほうされて，政策選択の重度がひろがること。

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お，もつことである。にほん経済はこれまで，国際収支，物価，福祉という3重くになやまされながら，相互のトレード・オフ関係から有効な政策手段が3すくみの常態にあるとされてきた。
だが，変動相場制をとることによって，とうめん，国際収支上の手かせあしかせか　はずされれば，政策選択のはばはそれだけひろがる。
しかも，国内物価上昇のもっともゆうりよくな要因とされている海外インフレのじようりくがしやだんされれば，政策とうきよくとしては，今後の経済施設のじゆうてんを　あげて国民福祉こうじようの1点におくことも可能なはずである。
くわえて一般じつぎようかいからも，こんかいの変動相場制には，　フロートズレ　したというか，ぜんかいほどのつよいひなんはきかれなくなった。
こうした点をこうりよすると，できるだけ為替需給にそくした円の適性水準をみつける必要性からも，このさいは，じっくりこしを　すえた円の長期変動相場制をかくごして，その前提にたった経済運営がのぞまれるのである。
ただ，変動相場制の長期化は，国内の物価上昇，インフレをはどめする，保証とはかならずしもならない。
よくしられているとおり，きくねんあきからほんねんにかけてもういをふるっているおろしうり物価のきゆうとう要因には，輸入物価上昇のほかにも，不況カルテル党の供給制限的行為，景気回復にともなう需要の拡大，部門かん需要のシフト　いわゆる過剰流動性，資金コストのぞうだいなどがあるからである。
つまり，いまの物価上昇は，これらの諸要因がふくざつにからみあっておこっているわけで，海外インフレの影響が変動相場制でしやだんされても，なおねずよい上昇要因がのこっていれば，物価上昇とインフレあくせい化の危険はさらない。
さいきんのおろしうり物価上昇には，交易条件のきゆうあく化にたんてきにしめされているように，もくざい，だいず，ようもうなどの輸入価格のぼうとうによるところがきわめておおきい。
したがって，この面からの国内物価よくせい効果は今後，じよじよにでてくるものときたいされるが，同時に物価安定化たいさくとして，ぜんき諸要因をじよきよないしかんわするような施策が積極的にとられる必要がある。インフレのこうしんはとりもなおさず国民福祉のひていにもつながるからであり，変動相場制かにおいても，インフレよくせいがいぜん，じゆうだいな政策かだいとしてのこることはろんを　またない。
その意味で政策とうきよくが，預金準備率のさいきりあげや銀行のまどぐち寄生など　によって，金融政策のきちよう転換にのりだしたことは過剰流動性吸収のためにさんせいできる。
だが，きくねん10－12月のGNP実質成長率がぜんねん動機比12．7％になったことを　はじめ，ほんねん1月までの景気諸指標がつきなみおおはばな上昇をしめした
ことを　もって，景気かねつはひつしと判断し公定歩合のひきあげを　けんとうしているとすれば　，問題がある。

図6－14(b)　漢字かな混り文出力例

なぜなら，第1に，さくねんあきからほんねん1月にかけての景気指標おおはばこうてんは，ぜんねん動機の景気がどんぞこ常態にあったことのはんえいでもあるからで，景気がこのままかねつ常態にとつにゆう　するかどうかは，もうしばらくのすいいを　みなければならない。
第2に，今後は変動相場制いこうにともなうデフレこうかが予想されることである。とくにおうしゆう通貨じようせいの流動化によって，円相場はとうしよ予想をおおはばにうわまわるえんだか次元がひつしとみられるだけに，その影響は輸出かんれん産業を中心にけいしできぬものがあろう。
このような常態かで，ほんかく的な金融ひきしめ，総需要政策にうったえれば，過剰流動性の吸収には有効であっても，物価はそれほどさがらず，むしろせいさんが低下して，おろしうり物価上昇をしげ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ゆ転換をそくしんするために，むしろ可能なかぎり国内の福祉型需要の拡大や輸入そくしんにじゆうてんを　おいた政策を指向すべきであろう。
産業転換のためのいわゆるまえむきの調整援助，ろうどう力流動化のための職業訓練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，おうべい諸国にくらべて従来の法人税のぞうぜいや所得げんぜいなどもかんがえられてよい。
情は，変動相場制かにおいて，にほん経済のてきおう力，潜在能力をじゆうふニんにいかしながら，経済たいしつを　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年つづきのふさくで需給がひっぱくし，わがくにののうせいの基本もかんがえなおすべきだとゆうこえがいちぶにあがっている。
このたびはつぴようされた農業はくしよも，のおそん的ふうどのほぜんとともに，世界ののおさん物市場の不安定性に対応して食糧じきゆう度のひきあげを　きようちようしている。
しかし，これからののうせいはただたんに食糧のじきゆうじそくを　めざすのではなく，てきちてきさんを　中心とした高能率，低コストの生産体制をそうきゆうにかくりつするのがもっともじゆうようなかだいである。
世界食糧農業機構の推定によると，ほんねんにおける世界の食糧生産はきようさくであったさくねんとほぼおなじぐらいのみこみで，けつきよく3年つづきのふさくを　予想している。
こくるいのふさくは同時に資料の供給ぶそくにもつながり，ちくさん物の生産におおきなししようを　およぼす。
のおりんしようはさくねん10月に「のおさん物じゆきようの展望と生産もくひようのしあん」なるものを　こうひようした。
これは，FAOの「1980年の食糧需給の展望」やローマ・クラブの「成長の限界」などにしめされている食糧事情にたいする被官的なみかたを　ねんとうにおいて，わがくにの食糧じきゆう率をたかめていくことを　ねらいとしている。
こんかいの農業はくしよでもみぎのような問題意識をもちながらのうさん物需給，農業経営，農業経済の動向を分析している。
農業生産はのうさん物輸入のぞうだい，こめの生産調整や普通はたけさくもつ，　げんさんなどによって45，46年と交代傾向をたどった。
47年の農業生産はじやつかんのもちなおしがきたいされるものの，いぜんとして低水準にある。のうよう地の流動化はちかの上昇によって疎外され，農地所有けん　のゆうしよう移転はげんたいを　つづけている。
また，こめの生産調整によるきゆうこうでんのはんぷんちかくは後配したままにほうちされ，そしている。とくに農業の職業人口の現象がめだっている。農業の職業人口の現象率は40年から44年平均では3．2％であっ

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、文字種変換、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なう。

- 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

② 棒組調整処理

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、中央揃
- タブ、均等配分、ルビ
- 合成、ハイフォネーション
- そ の 他

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。

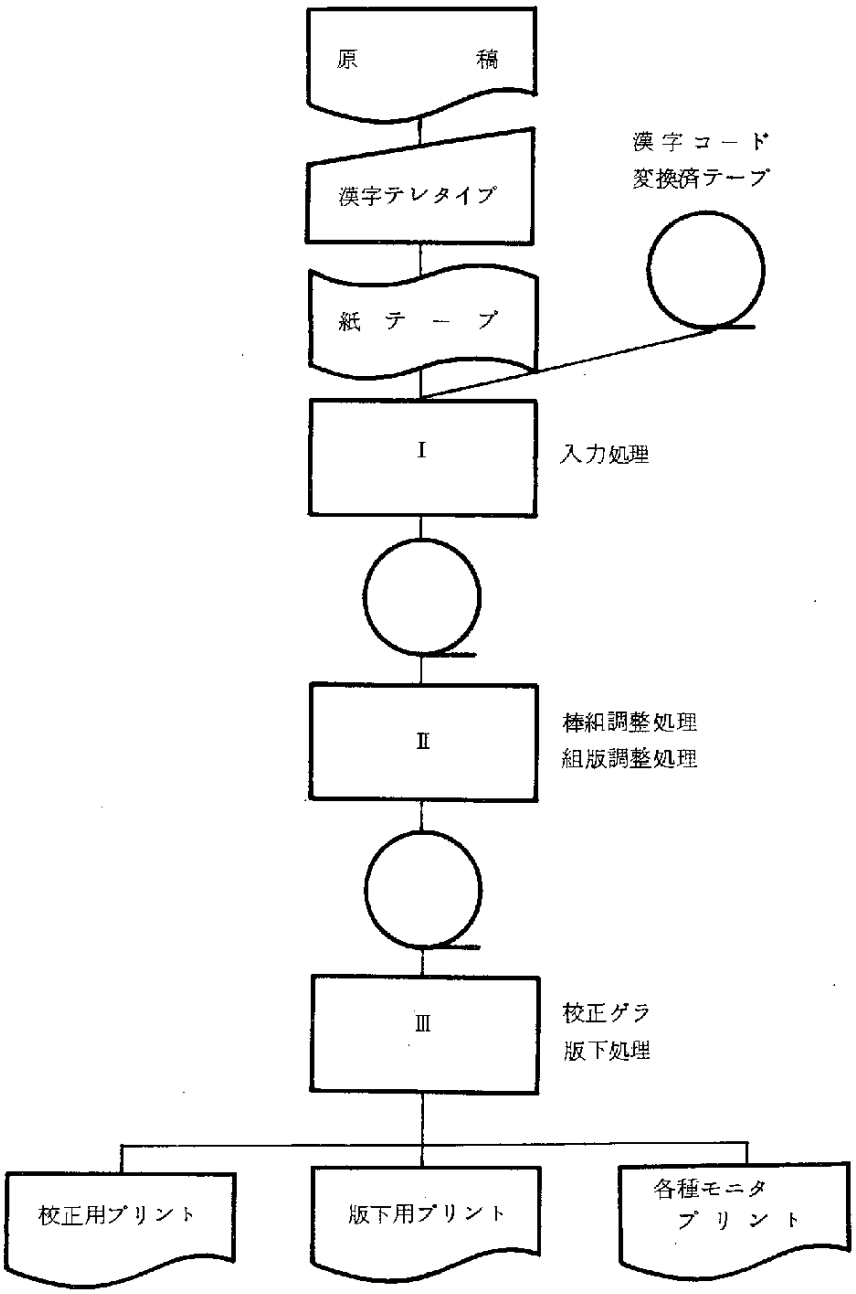


図６－１５　自動編集ソフトウェアの処理フロー

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。

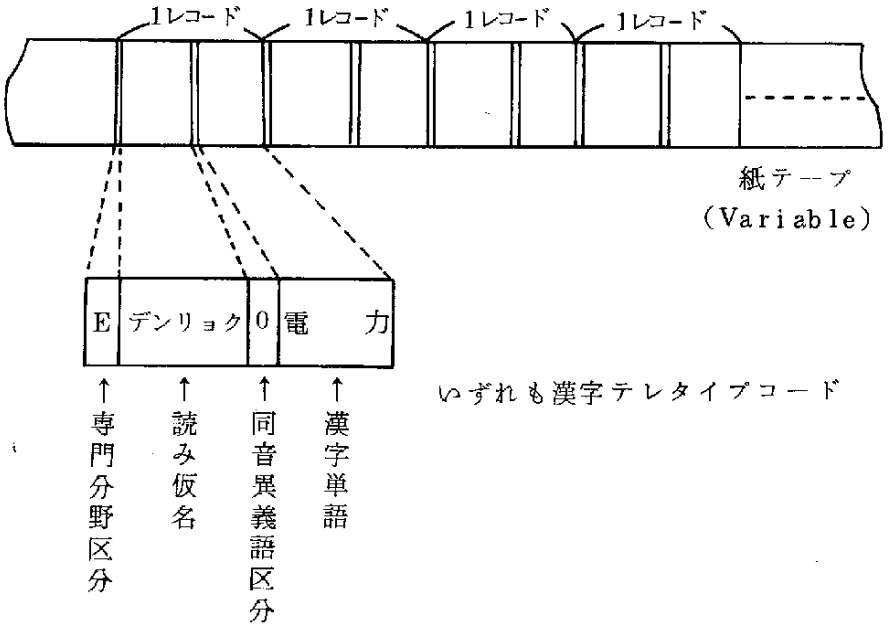
まず最初に専門分野における漢字および漢字単語の調査分析を行なう。学術書に関するような専門分野であれば、ハンドブック、用語辞典などの専門用語集を参考にしたり、専門書のキーワードや索引などから漢字単語を抽出することが手身近でよい。しかしながら、用語集などとしてまとまっていない分野においては、過去において出版された文書・文献のなかから漢字単語を抽出する方法がとられよう。この作業を進めるうえで、漢字単語の抽出とともに出現頻度をカウントする必要がある。これは、登録単語の規模をシステムにあわせて考慮したほうが当然ながら効率的なためであり、頻度の低い単語については、将来の必要性などを推測して判断しなければな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

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パンチ入力する。

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）



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 情報の洪水

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、ＯＥＣＤ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）、少くとも年率１０％で増大し続けていくとなると、１９８５年には、８００万件／年になるという。

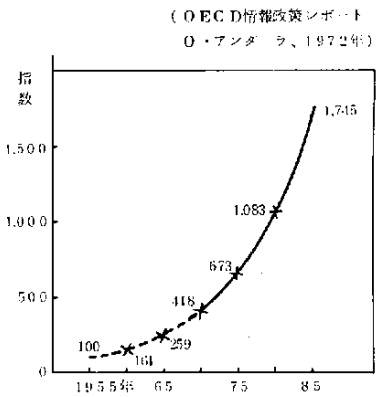


図７－１　科学技術情報のふえ方

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。

このような爆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、量的な問題をある意味で解決できたとみてもいいかもしれない。

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よって、コンピュータを使わせてもらっている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でしか通用しないのがほとんどである。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。

表７－１　ファイルの形態

業種＼ファイルの形態	ロールフィルム	ストリップフィルム	ジャケットフィルム	カードアパーチュア	マイクロフィッシュ	マイクロカード	計
水産	1	－	1	1	－	－	3
鉱業・石油	4	－	2	2	4	－	12
鉄鋼・非鉄	8	2	2	9	12	3	36
建設	15	－	9	10	3	－	37
食品	3	2	－	1	2	－	8
繊維・パルプ・紙	6	5	1	－	4		16
化学・ゴム・硝子・七石	34	4	16	7	14	－	75
機械（含金属製品）	42	9	18	33	16	－	118
電気機器	75	8	24	43	22	2	174
輸送機器	34	2	17	23	24	－	100
精密機器	23	7	5	7	9	3	54
その他製造	2	－	－	2	－	－	4
商業	9	－	7	－	5	－	21
金融・保険・不動産	29	1	1	3	4	1	39
運輸・倉庫	14	1	3	2	1	－	21
電力・ガス	8	1	4	6	4	－	23
サービス	18	4	7	3	6	－	38
通信・新聞・出版	24	1	3	4	6	－	38
官公庁	78	8	14	21	7	－	128
教育・宗教・文化施設・団体	116	19	8	7	24	5	179
計	543	74	142	184	167	14	1124

表７－２　被写体の種類

業種＼被写体	図面	図書文献	文書	カタログ	新聞	地図	帳票類	証券類	手形小切手	計
水産	3	－	－	－	－	－	－	－	－	3
鉱業・石油	6	4	1	－	1	－	－	－	－	12
鉄鋼・非鉄	18	12	9	4	－	－	1	1	－	45
建設	29	18	14	－	－	1	1	－	－	63
食品	5	2	4	－	－	－	－	－	－	11
繊維・パルプ・紙	－	14	3	－	－	－	－	－	－	17
化学・ゴム・硝子・七石	21	38	26	6	3	1	2	－	－	97
機械（含金属製品）	90	36	22	9	－	－	6	－	－	163
電気機器	110	64	40	17	3	－	6	4	－	224
輸送機器	69	37	31	14	－	2	9	－	1	163
精密機器	17	12	15	3	3	－	1	3	－	54
その他製造	1	－	1	－	－	2	－	－	－	4
商業	4	7	12	－	－	－	7	－	－	30
金融・保険・不動産	5	5	7	－	－	2	13	8	16	56
運輸・倉庫	10	2	10	2	1	－	1	－	－	26
電力・ガス	12	5	7	－	－	2	3	－	－	29
サービス	6	16	4	3	8	2	1	1	－	41
通信・新聞・出版	8	14	15	－	28	－	3	2	－	70
官公庁	57	29	44	1	13	12	15	－	－	171
教育・宗教・文化施設・団体	15	108	30	1	73	29	5	－	－	261
計	486	423	295	60	133	53	74	19	17	1560

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だりする方法である。

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ぎのような特徴がある。

- 記録が正確で信頼度が高い。
- 秘密の保護ができる。
- 情報を通常の資料の約$\frac{1}{100}$位に縮小できる。
- 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(30.5m)のフィルムをリールに巻いたもので、普通2,000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㎜、35㎜、105㎜などがある。16㎜巾のものは、通常一般文書、伝票、小切手、手形、証券などを記録する目的で使われている。35㎜巾のものは、学術文献、とくに新聞、図面などに多く用いられている。105㎜巾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) マイクロフイッシュ

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。マイクロフイッシュは、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てとり出すことができるようにしている。最近ではさらに、3,000コマ以上も収容できる超マイクロフイッシュも使われている。

(3) アパーチュアカード

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り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3,２００ページを標準の４×６インチのマイクロフイッシュに毎秒1,０００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。

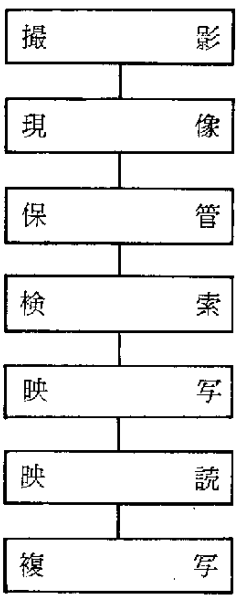


図７－２
マイクロフィルム作成手順

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、リーダ・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16㎜専用、16㎜と35㎜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）。

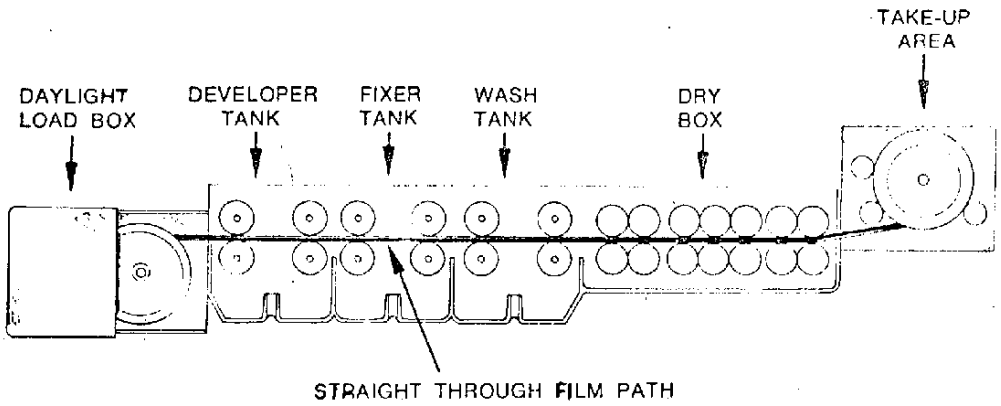


図７－３　マイクロフィルム現像処理機

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成するためのリーダ・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／㎜、周辺で２.５本／㎜

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ニング機構がスムーズであり、像が横になっていたり、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。

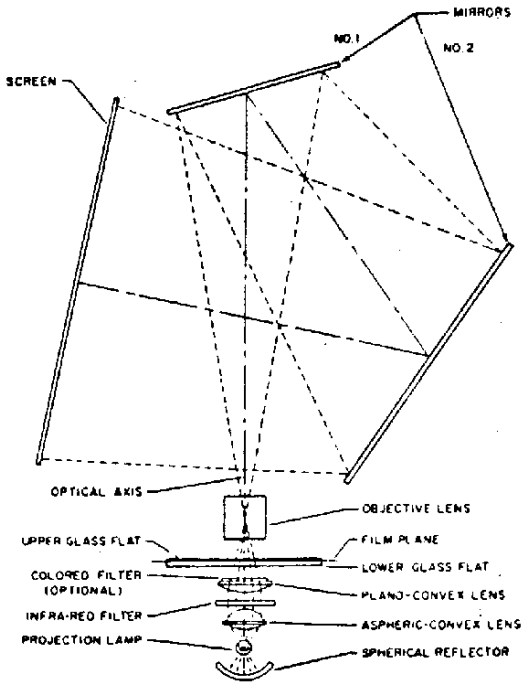


図7－4　マイクロ・リーダの構造の例

(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ァ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。

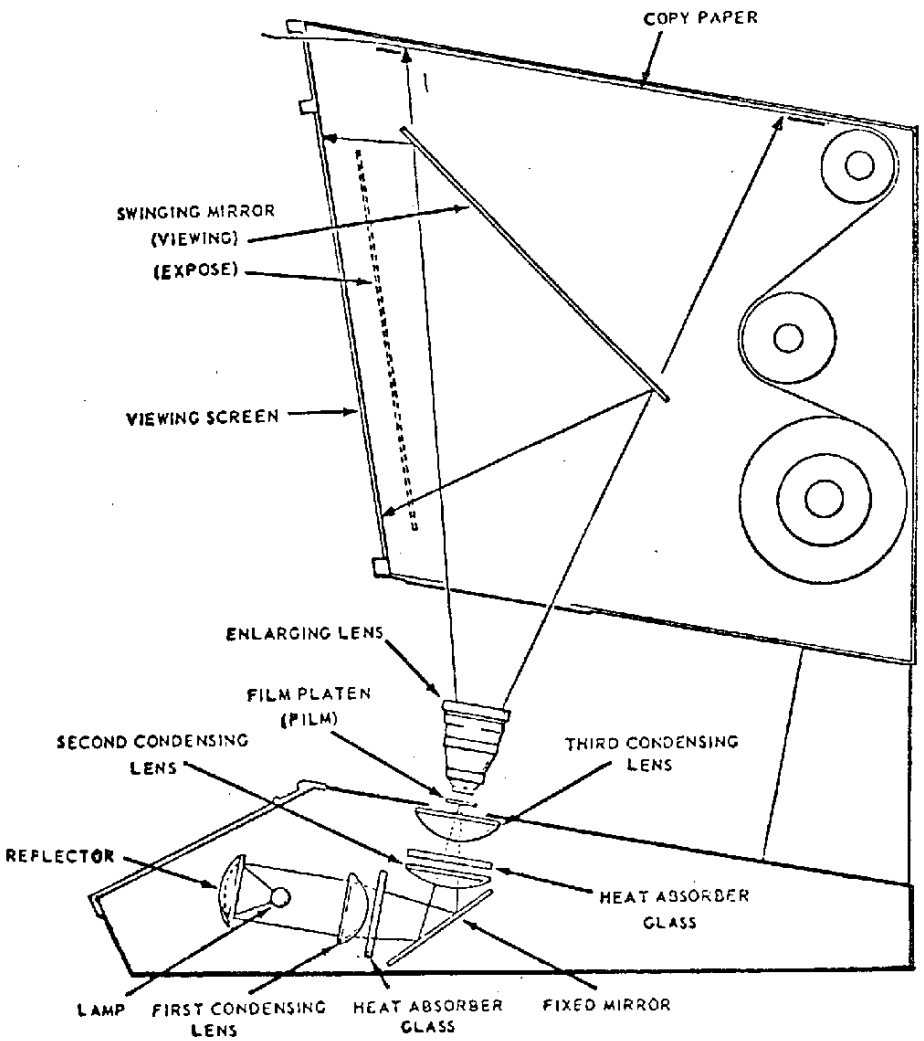


図７－５　リーダ・プリンタの構造の例

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ろいろある。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び出している手動検索方式がある。これに対して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なうことができる。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ておき、このマークをカウントして、コマを探す方法をビリップ方式と呼んでいる。自動検索装置つきの場合は、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

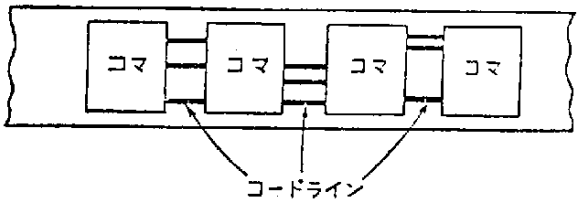


図７－６　コードラインフィルム

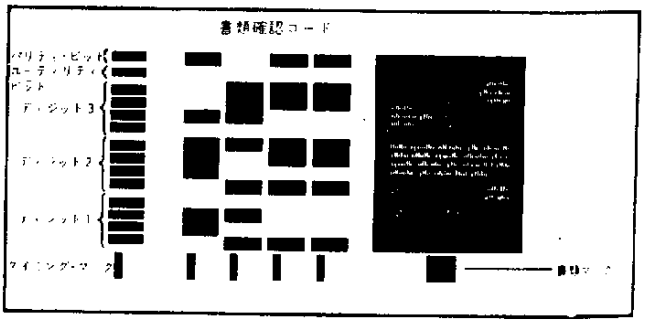


図７－７　ミラコードの画像・コードの配列

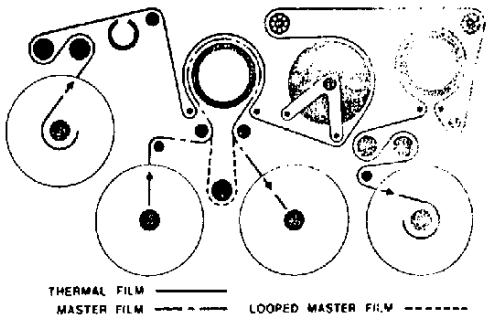


図７－８　ロールフィルム複写機

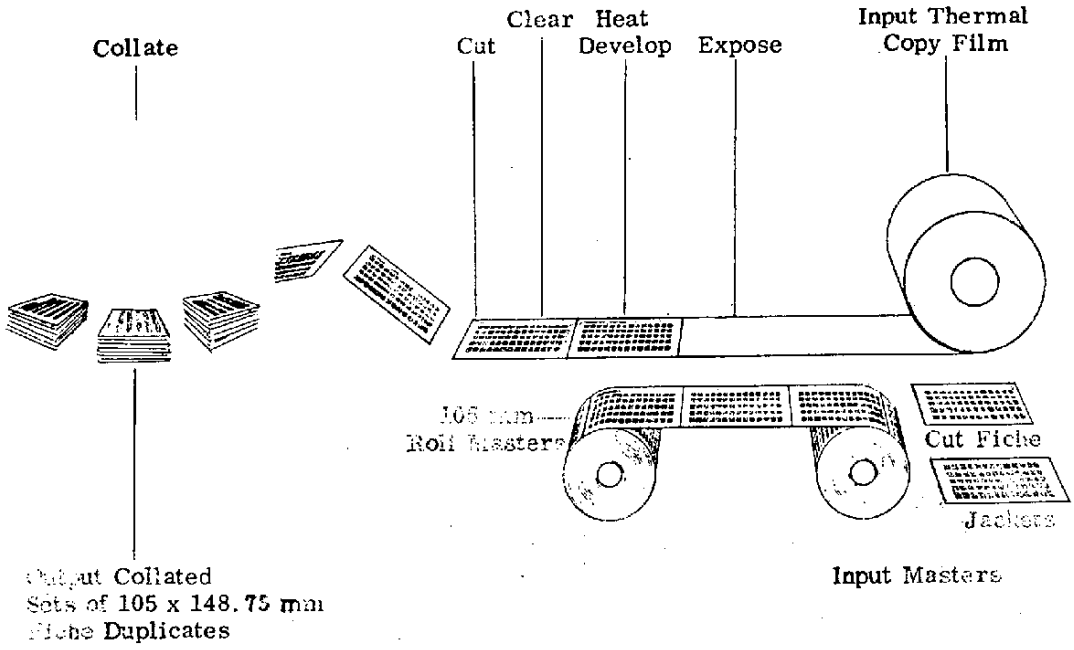


図 7－9　サーマルフィッシュ複写機

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

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。

⑵　ＣＯＭの利用状況

ＣＯＭの性能上から分類するとつぎの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。

表7－3　COM

メーカ（ディーラ）		（Information International社）極東貿易	（米国3M社）住友スリーエム	（カルコンプ）吉沢ビジネスマシン		（PERTEC社）三菱事務機械販売
型名		FR－80	エレクトロン・ビーム・レコーディングシステムシリーズF	1675	2100	PERTEC3700 マイクロフィルム・プリンター
磁気テープ	トラック数	7トラック，9トラック	7トラック，9トラック	7，9	7／9	7トラック，9トラックのいずれにも切換自由
	記録密度（bpi）	556，800，1600 bpi	200，556，800，1600	200～1600	200～1600	556，800，1600の3種
	キャラクタ・セット	BINARY，ASCH，BCD，EBCDIC	BCD，EBCDIC	BCD，EBCDIC，BCDIC ASCII他	BCD，EBCDIC，BCDIC ASCII他	EBCDIC，BCDIC，ASCII，その他各種有
	転送速度（KC）	60Kバイト	OFFUNE－60K	60	60	25，30，60の3種
文字	文字種類	128	英，数，カナ，記号	英数字・特殊記号・カタカナ・漢字・	英数字・特殊記号・カタカナ	64種（128種に拡張）
	文字発生方式		直接ストローク・ジェネレート	CRTによるドット方式	CRTによるドット方式	ドット・マトリックス方式（7×9）
	文字数		60 ～ 120	英数字・特殊記号：64，カタカナ：48	英数字・特殊記号：64，カタカナ：48	144キャラクタ／行
カメラソイルム	フィルムサイズ	35㎜，16㎜，105㎜，216㎜	16	16，35㎜，105㎜マイクロフィッシュ・105㎜フルフレーム	16㎜巻取・105㎜マイクロフィッシュ	16㎜または105㎜アンスプロケット・ロール
	フィルムの長さ（フィート）	106m	（16㎜）675'	200，400	200	200フィート（16㎜）100フィート（105㎜）
	フィルム記録速度	74mSEC／フレーム～1 SEC／フレーム	6万字／秒　2万行／分	10000～15000ライン／分	10000～15,000ライン／分	26000行／分
	縮小率	1／24～1／300	24／1，42－48／1	（35㎜）15：1（16㎜）21：1，24：1（105フィッシュ）42：1（105フルフレーム）10：1	24：1，42：1，48：1	24：1，42：1の2種
検索コード		IMAGE COUNT，MIRACODE	オドメータ，プリップ検索	ミラコード	イメージカウント	
チェック方式			パリティ方式			垂直方向，VRC，LRC
コンピュータインタフェイス		Burroughs，CDC，GE，Honeywell IBM360，370，NCR RCA Spectra，Univac	OFF	ON・OFF	ON・OFF	
接続機種				全機種	全機種	IBM他あらゆる機
ページサイズ（字×行）			132×64（66×32）	132～224×88	132×66	144×64，86
フォーム処理		プログラム制御	プロジェクタ方式，ストアド・プログラム	プロジェクタ	プロジェクタ	フォーム，オーバーレイによる
処理速度（ページ×分）			960ページ／分	120	180～310	250ページ～450ページ／分（実測）
モニタ		8″×10″モニタ	現像処理モニタ	ビジュアルモニタ	CRT直視	有
価格（レンタル）買取	円		¥75,500,000（112，210，310，410）¥78,250,000（113，210，310，410）	96万～4000万	3024万～1260万～	70万円
備考			熱現像方式（ドライシルバー）現像プロセス処理まで一貫オペレーション			

メーカ（ディーラ）		（Quantor社）高千穂交易		（CMS社）高千穂交易	丸紅エレクトロニクス	キヤノン販売	Eastman Kodac Company（クスダ事務機株式会社）		日商エレクトロニクス
型名		Quantor 105	Quantor 100	CMS 7000	UCC・COM300－2	J-COMシリーズ	コダックKOM-80マイクロフィルマ	コダックKOM-90マイクロフィルマ	SDI－4440
磁気テープ	トラック数	7／9チャンネル	7／9チャンネル	7／9チャンネル	7，9	7，9	9	7又は9	7又は9
	記録密度（bpi）	556，800，1600 BPI	556，800，1600 BPI	200，556，800，1600 BPI	556，800，1600	556，800 1600 BPI	556，800，1600	200，556，800，1600	200，556，800，1600
	キャラクタ・セット	BCD EBCDIC	BCD EBCDIC	EBCDIC，BCD，ASCII，その他いずれも可		ANY	EBCDIC	BCD EBCDIC	自由
	転送速度（KC）	60KC	60KC	300KC／秒	36	ON500，OFF60	60KC，90KC，120KC	120K 120KCまで	200IPS
文字	文字種類	1	1	16サイス（イタリック，ローマン，ライト，ボールドなど）	英，数，記号，カナ，漢字	英，数，記号，カナ，漢字	4種（標準，イタリック，標準太，イタリック太）	4種（標準，イタリック，標準太，イタリック太）	英，数字，カナ，特殊記号，年月日
	文字発生方式	ドット・マトリックス	ドット・マトリックス	ストローク・ジェネレート方式	ドット・ジェネレータ	フライング・スポット	ストローク方式	ストローク方式	キャラクトロン方式
	文字数	128種	128種	128個	128	96～192	121	121	128
カメラ・フィルム	フィルムサイズ	105㎜	16㎜	16㎜，35㎜，105㎜	16，35，105㎜	16．105㎜	16，35，82，55，105㎜	16 35，82，55，105㎜	16，35，70，105㎜
	フィルムの長さ（フィート）	100フィート	400フィート	ロール1000フィート，フィッシュ100フィート	200，1000	600，1200	200ft 500ft 1000ft	200ft 500ft 1000ft	1000ft
	フィルム記録速度	30,000キャラクタ／分	30,000キャラクタ／分	120KC／秒，50,000行／分	3ページ／秒		8.3㎜	8.3㎜	120Kc／秒
	縮小率	24×，27×，42×	275×	20×から50×まで任意	15～42倍	22:1，42:1	$\frac{1}{20}$ ～ $\frac{1}{50}$	$\frac{1}{20}$ ～ $\frac{1}{50}$	18，25，42，48
検索コード			イメージ・カウント	バー・コード，ミラー・コード，イメージ・コントロール	ハードウエア，ソフトウエア，タイトル／フィッシュ，イメージカウント，ミラー・ド，コードライン	イメージカウント	コードライン，イメージマーク，ミラコード	コードライン，イメージマーク，ミラコード	ミラコード，イメージカウント，コードライン
チェック方式				パリティチェック（ビット・チェック）			VRC	VRC	CRCチェック
コンピュータインタフェイス				可		ON・OFF	オフライン	オフライン	各種あり
接続機種				いずれも可		ほとんどの種類	IBM仕様のテープドライブユニット	IBM仕様のテープドライブユニット	IBM，NCR UNIVAC FACOM，HITAC，NEAC
ページサイズ（字×行）		90行／ページ，160字／行 64字／ページ，132字／行		256字／行，128行／ページ	136字×79行	232×64	132×64行	132×64行	132，136，140桁／行 64～88行／ページ
フォーム処理		1	1	1フィッシュ8種類まで可	フォームスライド	プロジェクタ方式	フラッシュによる重ね焼き	フラッシュによる重ね焼き	フォーム・スライド
処理速度（ページ×分）		200ページ／分	600ページ／分	ロール900ページ／分，フィッシュ270ページ／分	3ページ／秒	150～720	3～500ページ／分	3～500ページ／分	4ページ／分
モニタ			有	11インチ・モニター・テレビ			ヒューイング・スコープ，フォーマットディスプレイ	ビューイング・スコープ，フォーマットディスプレイ	ポラロイドカメラ ビューイング・スコープ
価格（レンタル）買取	円						4042万円	5,661万円	その他：モデルSDI4370 SDI4200(オンライン用) SDI4060(グラフィック用) SDI 1460
備考		自動現像機内蔵，自動カッタ内蔵	自動現像機内蔵	バッファメモリ（8K～32K）	モジュールの追加によりVCC・COM300－3種になりグラフィックCOMへのシステムアップ可能				

メーカ（ディーラ）		SEACO COMPUTER-DISPLAY社（青作オフィスメーション）		Gould社（住友商事）	日本電子産業
型名		SEACO421J	SEACO451グラフィックCOM	Beta COM70.0L	JEM-3100／PT-150
磁気テープ	トラック数	7又は9	7又は9	7または9	9
	記録密度(bpi)	200,556,800,1600	556,800,1600	556,800,1600	800 BPI
	キャラクタ・セット	ANY	,ANY(EBCDIC,BCD,ASCII その他何でも)		JEM漢字コード
	転送速度(KC)	30KC	36KC	36KCmax	288KC
文字	文字種類	英数字，カナ文字，記号等	英数字，カナ文字，記号	英数字，上下段ソフト，科学記号	英数字，カナ，漢字，OCR用文字
	文字発生方式	ストローク方式	ストローク方式	dot matrix	フライングスポット及びROM
	文字数	約110	ハードウエアキャラクタ約130文字，ソフトウエアキャラクタ256文字(種類は無限)	128	漢字2500／20,000英数字128字
カメラフィルム	フィルムサイズ	16,35,82,55,105㎜	16,35,105㎜(825㎜も可)	105,70,35,16㎜	16,35,82,55,105㎜
	フィルムの長さ(フィート)	200,400,600,1000	200,400,600	200ft	200ft
	フィルム記録速度	12,000行/分			MAX10,000字/秒
	縮小率	24×,42×	35㎜－12×，16㎜－24×	48:1,42:1,24:1,18:1,12:1	$\frac{1}{24}$
検索コード		ミラコード，ラインコード，イメージカウント	ANY(ソフトウエアによって作成)	Cut Mark, Image Count Mack	
チェック方式		VRC，LRC，CRC(自動修正機能)	VRC，LRC，CRC		
コンピュータインタフェイス		可	可	オフライン	各種Compオンライン又はMTオフライン
接続機種		ANY	ANY		132字×64行以内
ページサイズ(字×行)		140×64，160×80	ハードウエアキャラクタMAX160×80 ソフトウエアキャラクタANY	132×最大105行	可能
フォーム処理		フォームスライド	ハードウエア(フォームスライド)＝1，ソフトウエア＝無限	可	最大420ページ／分
処理速度(ページ×分)		150～250ページ/分	160ページ/分	最大370ページ／分	
モニタ		ポアサイト	モニタディスプレイ有		オプション
価格(レンタル)買取	円	2,800万～5,000万	約15,000万円		3,200万円(125万円)
備考					

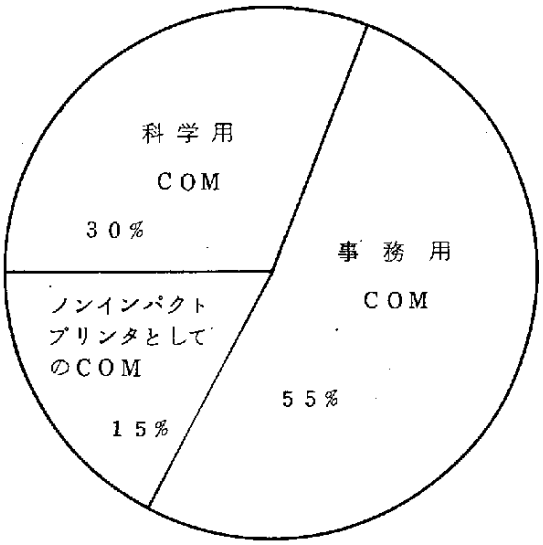


図 7－10

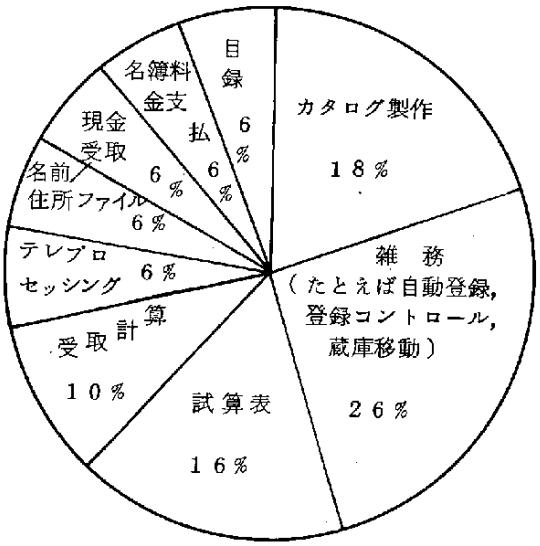
COMの使用例をあげると表 7－4 のようになっている（Auerbach 社の調査による）。

表 7－4

業務	応用範囲	分布
銀行業	要求払い計算	支店
自動車産業	部品リスト	供給人
航空	料金情報	代理人
図書館	書籍目録	閲覧室
特製鋼鉄	品目目録	倉庫
通信販売	カタログ情報	受注者
書籍出版	マイクロ出版	取引先
工学	科学用グラフ、工学用線図	エンジニア、技術者
金融および投資	チャート、グラフ、テーブル、一般印刷	従業員、依頼人、管理者
保険	供給人の委任状、供給人の適用システム、取引先一覧表、チェック・コピー検索、料金支払名簿、医者の記録、売人の記録	供給人 従業員 取締 支店
漫画映画および科学シーン	広告、技術、化学、科学的調査	テレビ局、映画館、研究所、大学
教育および訓練映画	教育と訓練	教室および講堂

ＣＯＭ使用別の割合は図７－１１のようになっている。

図７－１１



(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

キャラクトロンによって文字を発生させ、これをマイクロフィルムに写しと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種類位であり、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２５×３５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き、その所要速度は約8.3マイクロ秒（毎秒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太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。図

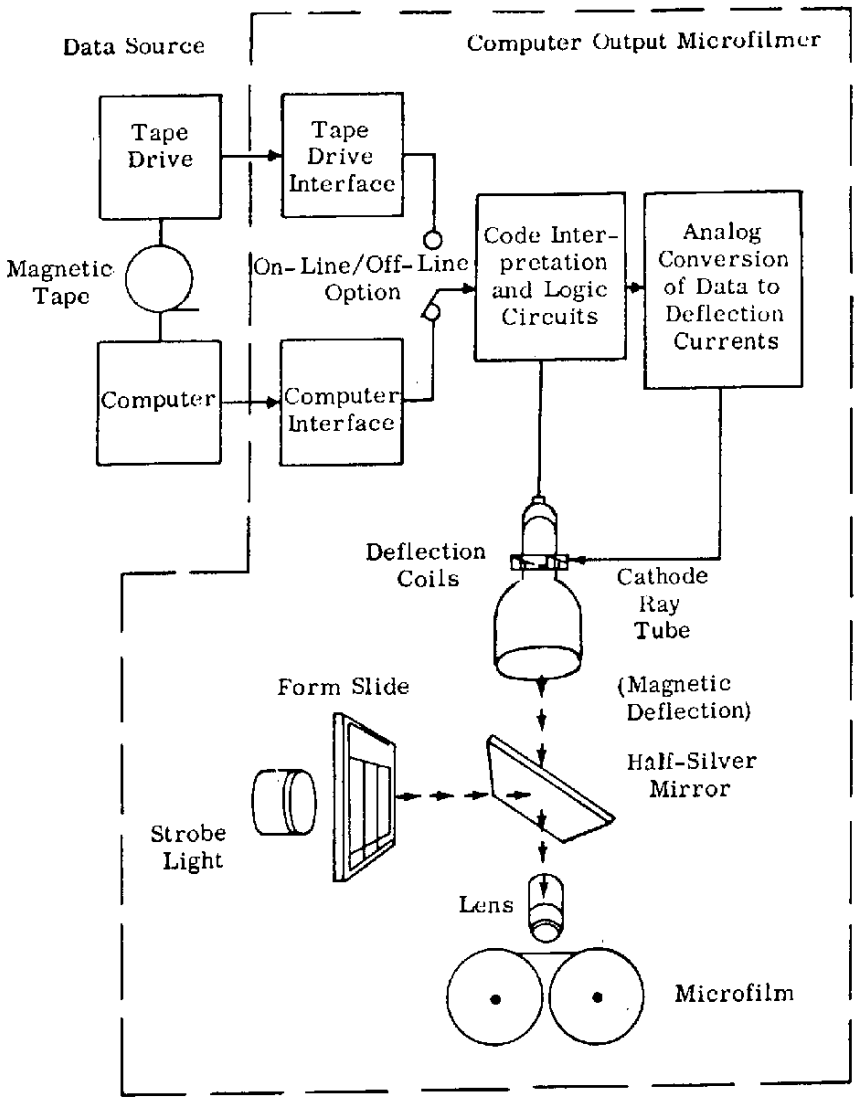


図７－１２　ＣＲＴによる記録

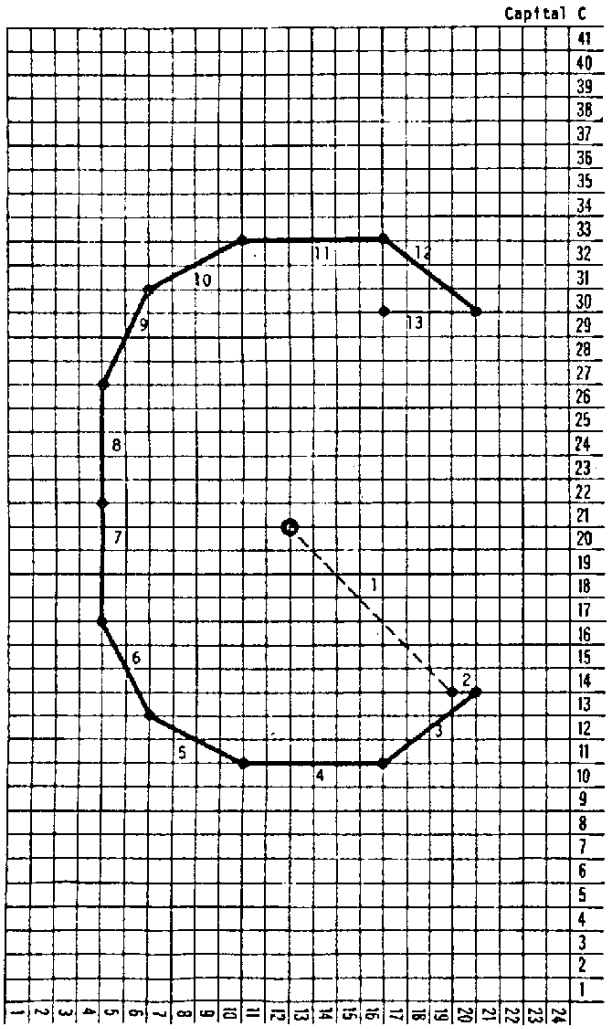


図７－１３　ＣＯＭ文字の一例

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

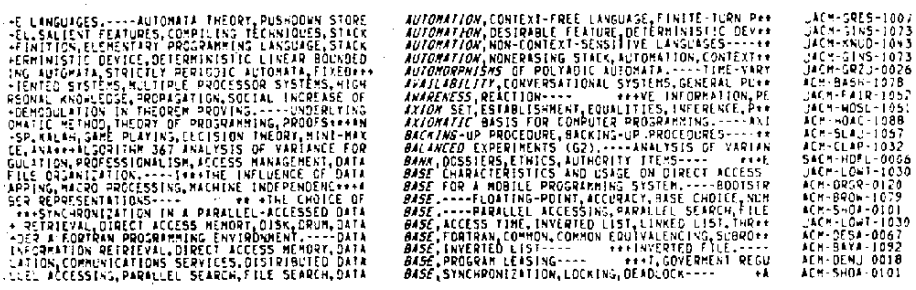


図7－14　イタリック体をキーワードに使用したCOMで撮影のKWICインデックス例

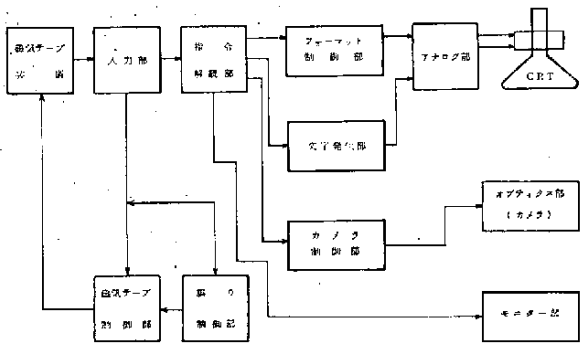


図7－15　COMのハードウェア・システム概要

ン形式と違ってマイクロフィルムに直接電子ビームが投射されるので、マイクロフィルムは真空装置の内部におかれることになる（図７－１７）。露出されたフィルムはプロセッサで熱現像される。

③　ＬＥＤによる記録

これはＬＥＤ（ Light Emitting Diode ）すなわち、発光ダイオードでビット信号を文字に表わすものである（図７－１８）。Memorex １６０３はＬＥＤとファイバ・オプチックスを組合わせたものである（図７－１９）。これに対してSeqnential Information System　Ｓ／ＣＯＭ－７０ではファイバ・オプチックスは用いていない（図７－２０）。

④　レーザによる記録

レーザによるマイクロフィルム上への記録はより一層の高密度の記録ができる。たとえばLaser Microfilming社のＰＤＲ－５システムは、８ミリフィルム上の３６チャネルに１インチ当たり5,０００ビットの割合で記録できる（図７－２１）。すなわち１イ

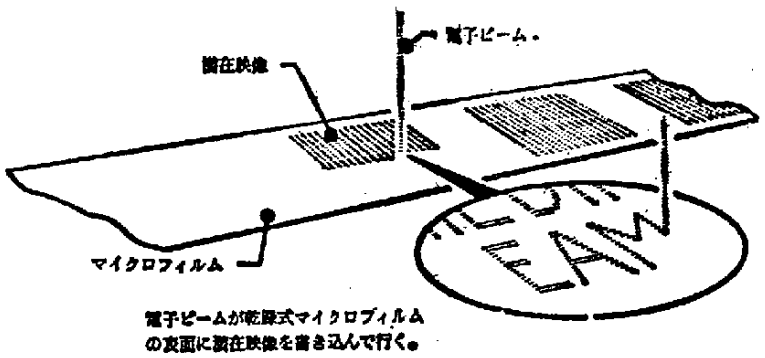


図７－１６　電子ビームによる記録

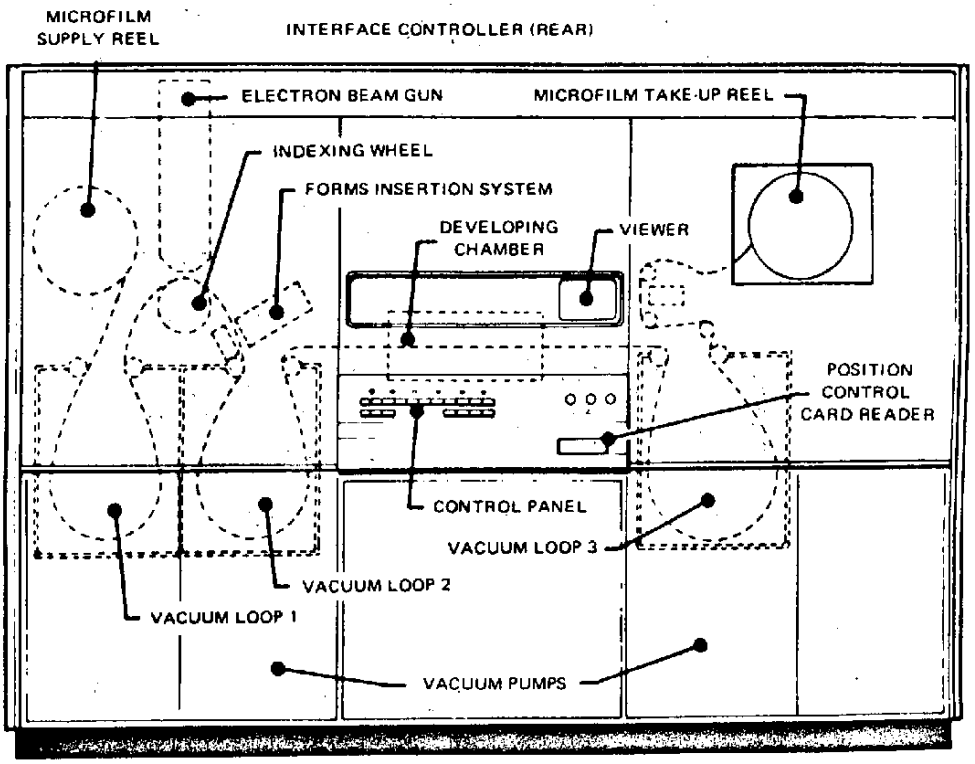


図７－１７　電子ビームによる記録

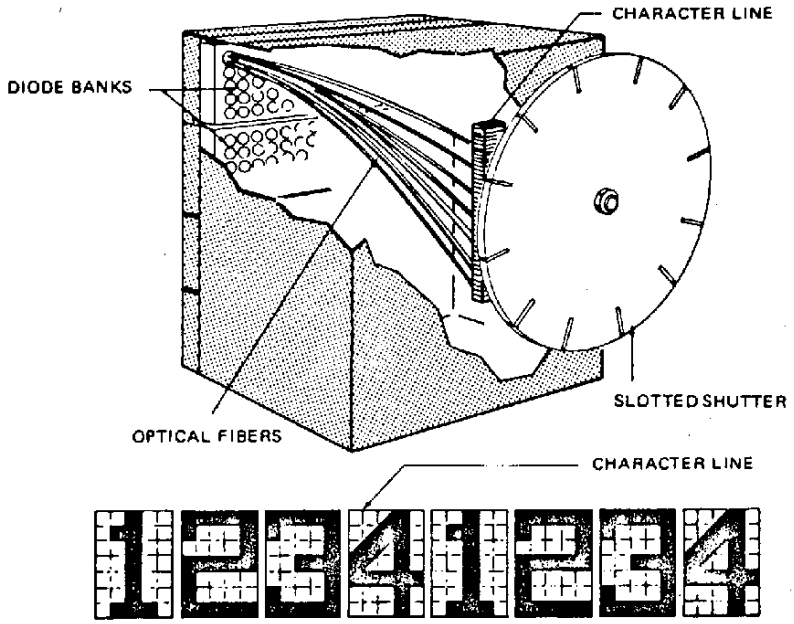


図７－１８　メモレックス文字発生器

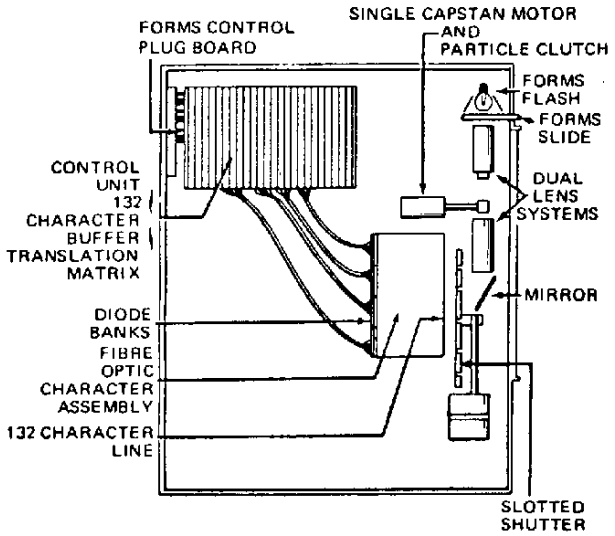


図７－１９　ファイバ・オプチックスによる記録

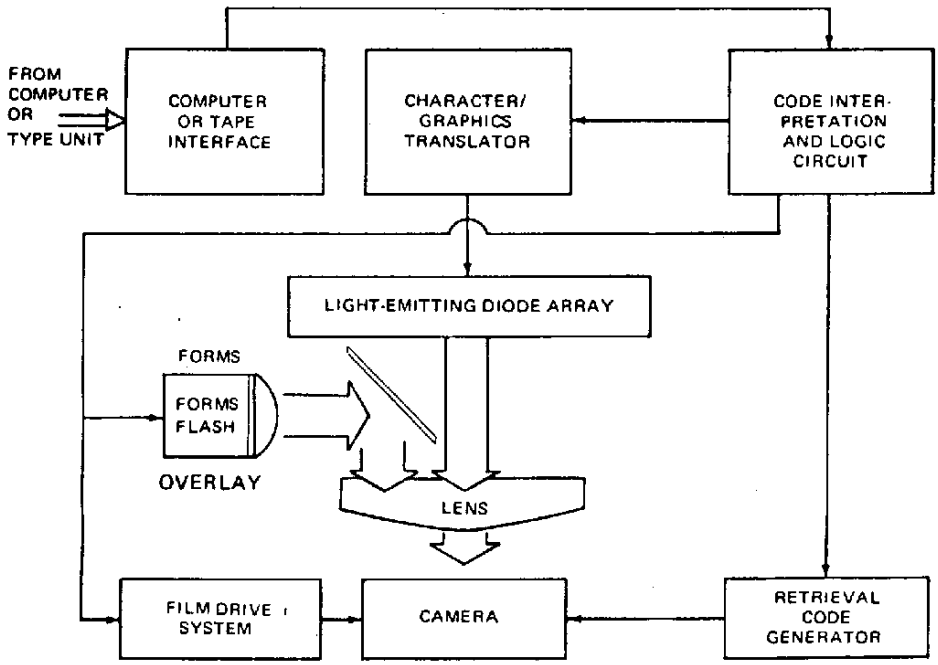


図７－２０　ＬＥＤによる記録

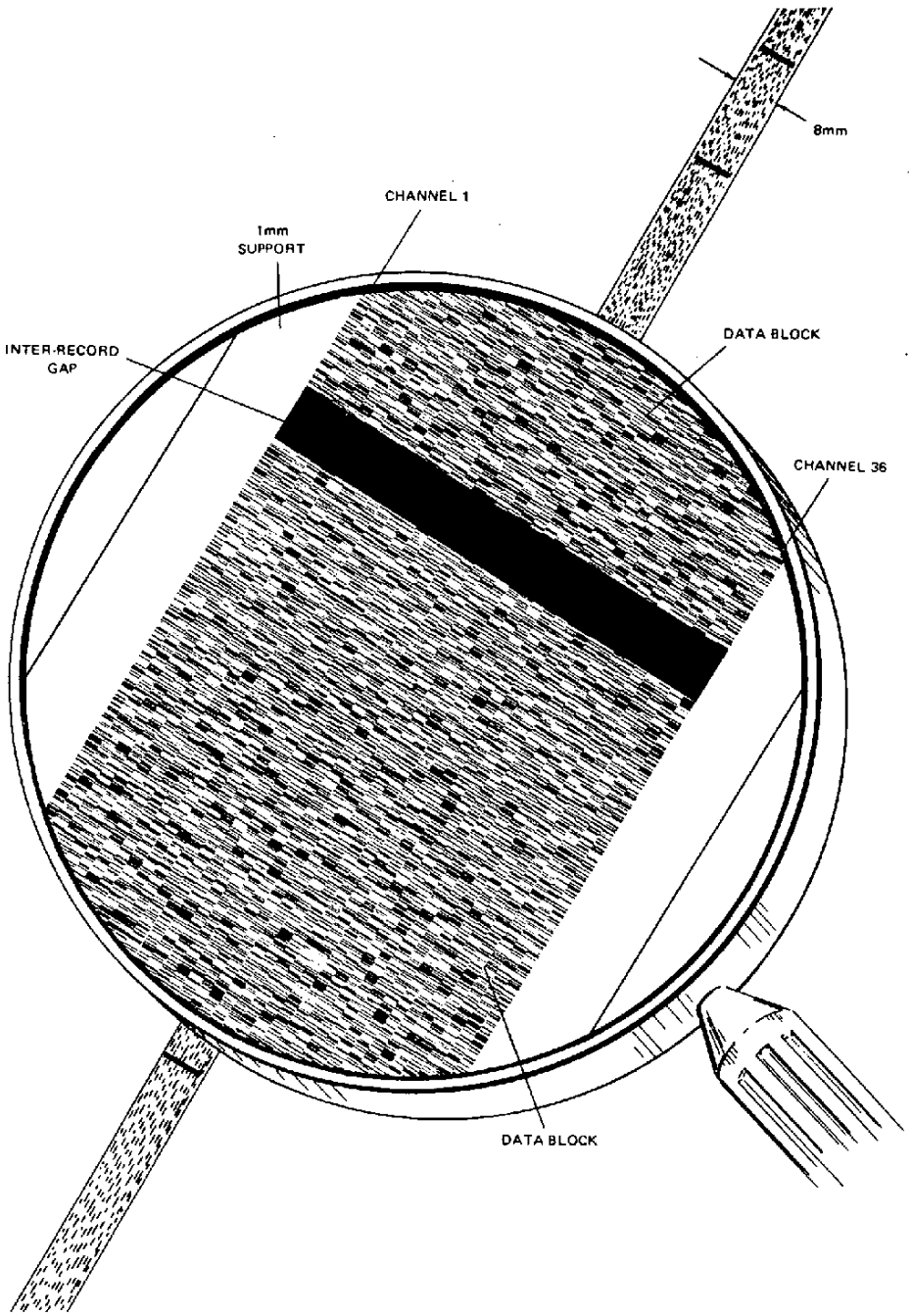


図 7－21　レーザによる記録の例

ンチ当り合計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にあたる。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ろ開発されている。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、アセンブラなどで使用でき、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。

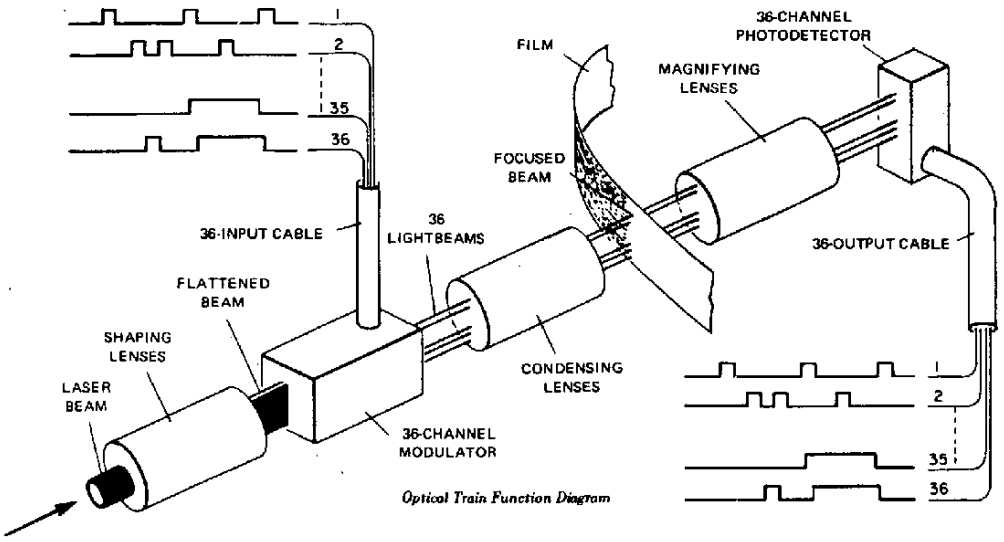


図７－２２　レーザ・ビームによる記録

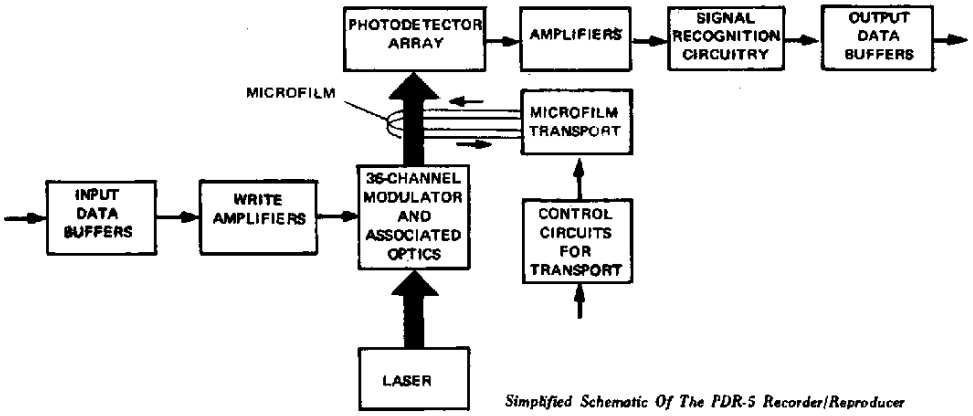


図７－２３　ＰＤＲ－５システムの概念図

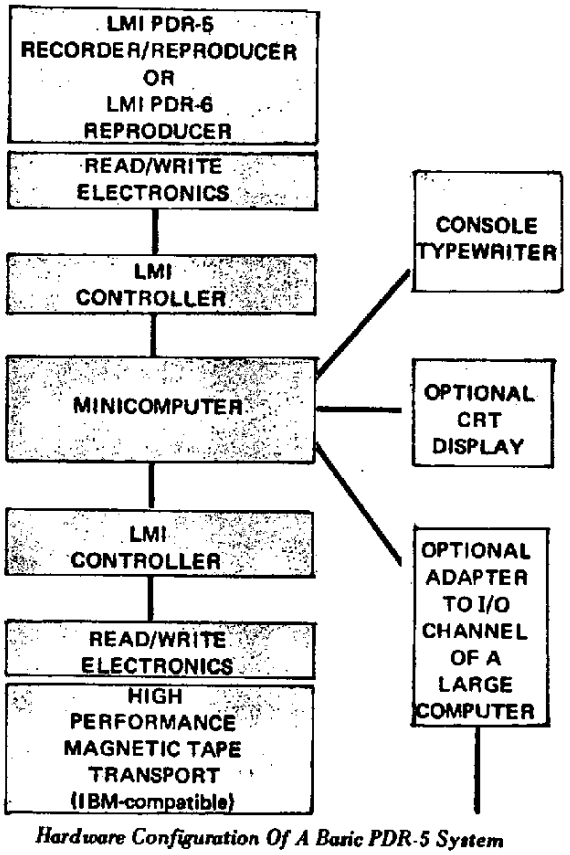
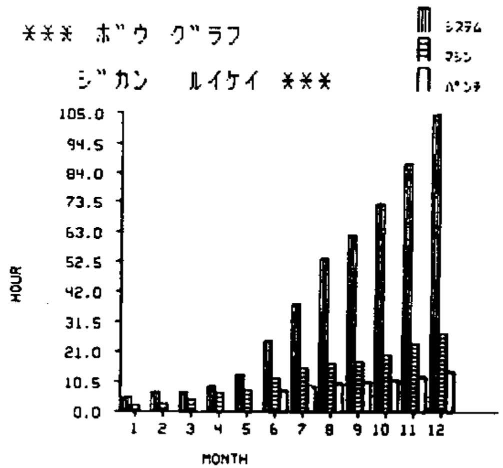


図７－２４　ＰＤＲ－５システムのハードウェア構成

***** ギョウム ベツ コスト ルイケイ ヒョウ (FY) ****

	1	2	3	4	5	6	7	8	9	10	11	12
*1 システム ジカン	5.0	2.0	0.0	2.0	4.0	12.0	13.0	16.0	8.0	11.0	15.0	17.0
	5.0	7.0	7.0	9.0	13.0	25.0	38.0	54.0	62.0	73.0	88.0	105.0
システムRジカン	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
*2 マシン ジカン	2.1	0.7	1.4	2.4	1.0	4.3	3.5	1.5	0.7	2.4	4.0	3.5
	2.1	2.7	4.2	6.6	7.6	11.9	15.4	16.9	17.6	20.0	24.0	27.5
リヨウA ジカン	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	1.4
	0.0	0.0	0.0	0.0	0.0	0.0	1.4	1.4	1.4	1.4	1.4	2.8
リヨウB ジカン	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
*3 パンチ ジカン	0.0	0.0	0.0	0.0	0.0	7.4	1.6	1.0	0.5	0.8	1.0	2.0
	0.0	0.0	0.0	0.0	0.0	7.4	9.0	10.0	10.5	11.3	12.3	14.3
ガイチュウヒ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
*4 ザイリョウヒ	3.0	2.0	5.0	4.0	2.0	7.0	4.0	3.0	2.0	5.0	6.0	4.0
	3.0	6.0	10.0	15.0	17.0	24.0	28.0	31.0	33.0	38.0	44.0	48.0
*5 トータル キンガク	40.0	11.0	20.0	41.0	22.0	140.0	91.0	75.0	50.0	100.0	110.0	95.0
	40.0	51.0	71.0	112.0	134.0	274.0	365.0	372.5	422.5	522.5	632.5	727.5



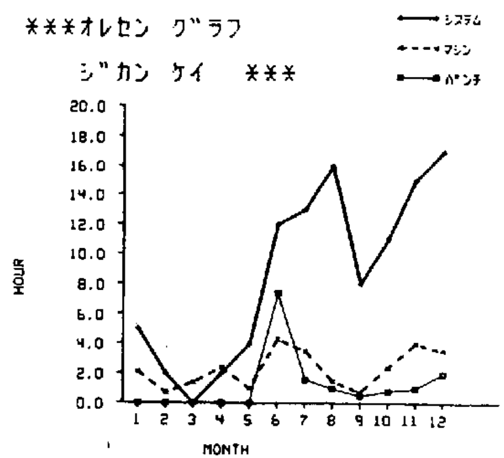


図7－25 PAGEによる出力

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ぎのようになる。

(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。

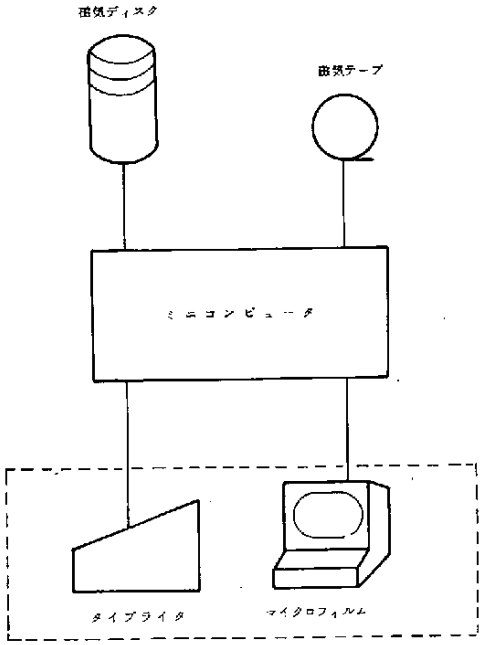


図７－２６　ミニコンピュータを利用する場合

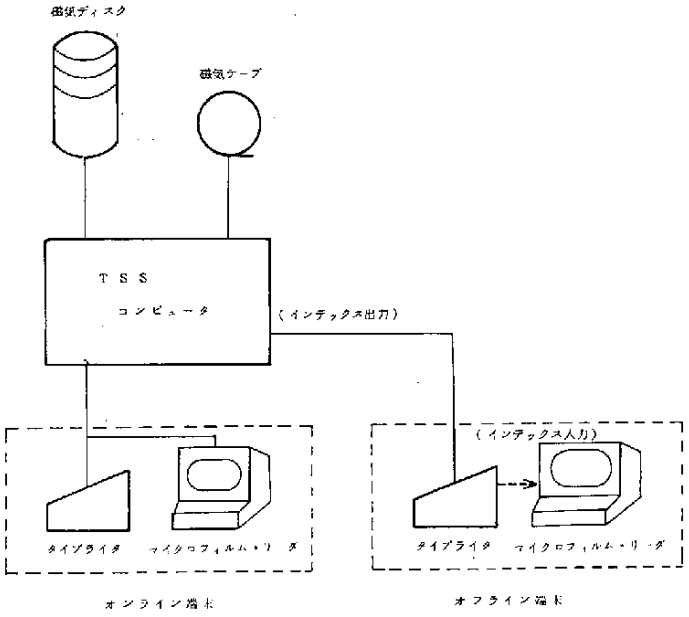


図７－２７　ＴＳＳ端末を利用する場合

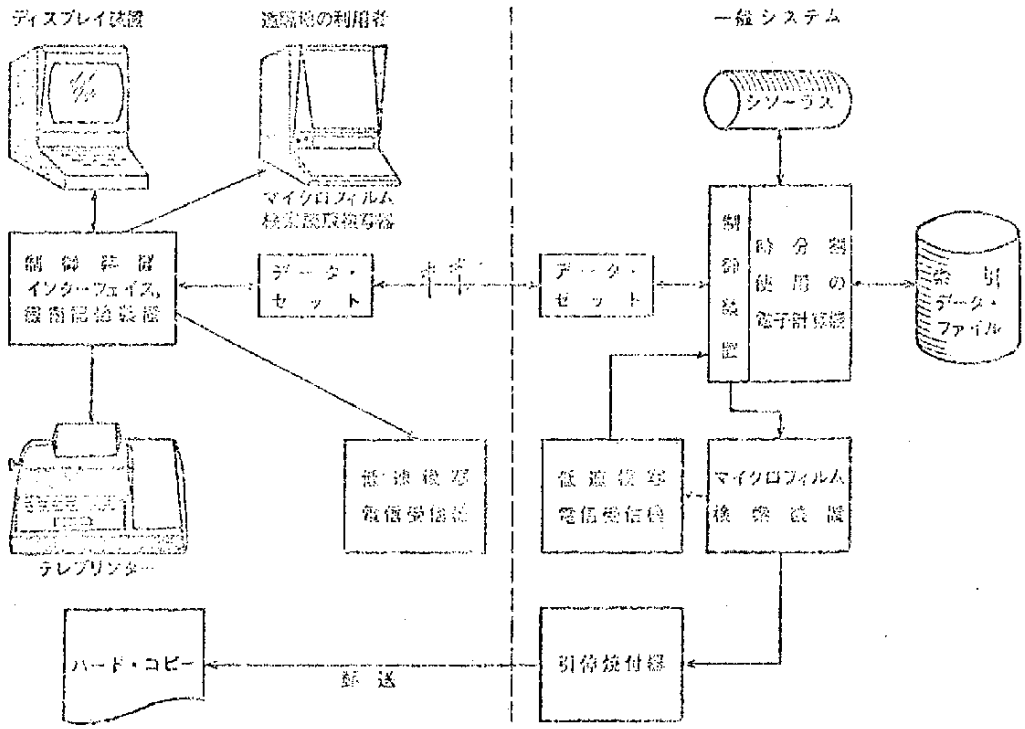


図７－２８　模写電信による場合

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、これを単にトップ・マネジント・レベルの情報システムとしてでなく、あらゆるレベルの管理者も利活用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ほどこ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。

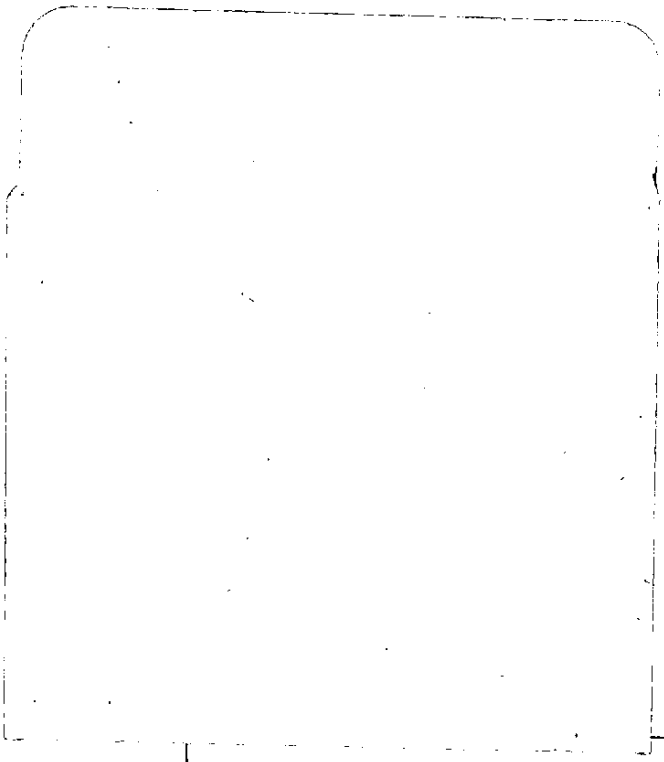
このように、数値や記号が情報として有効であったのは、情報システムが単一レベル、あるいはきわめて限られた範囲で利用されていたからであって、情報システムの利用が拡がるにつれて柔軟性のない数値や記号だけによる情報では、むしろ情報提供の相手に正確な意思を伝達できないばかりでなく、時によっては、情報提供の相手に重大なミスにつながる誤解を与えてしまう恐れも考えられる。

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、オペレーショナル・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、カタカナけん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、比較的簡単に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つは明るい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　嶺　社
東京都渋谷区代々木２丁目26番地
TEL（379）1816

48－S004