

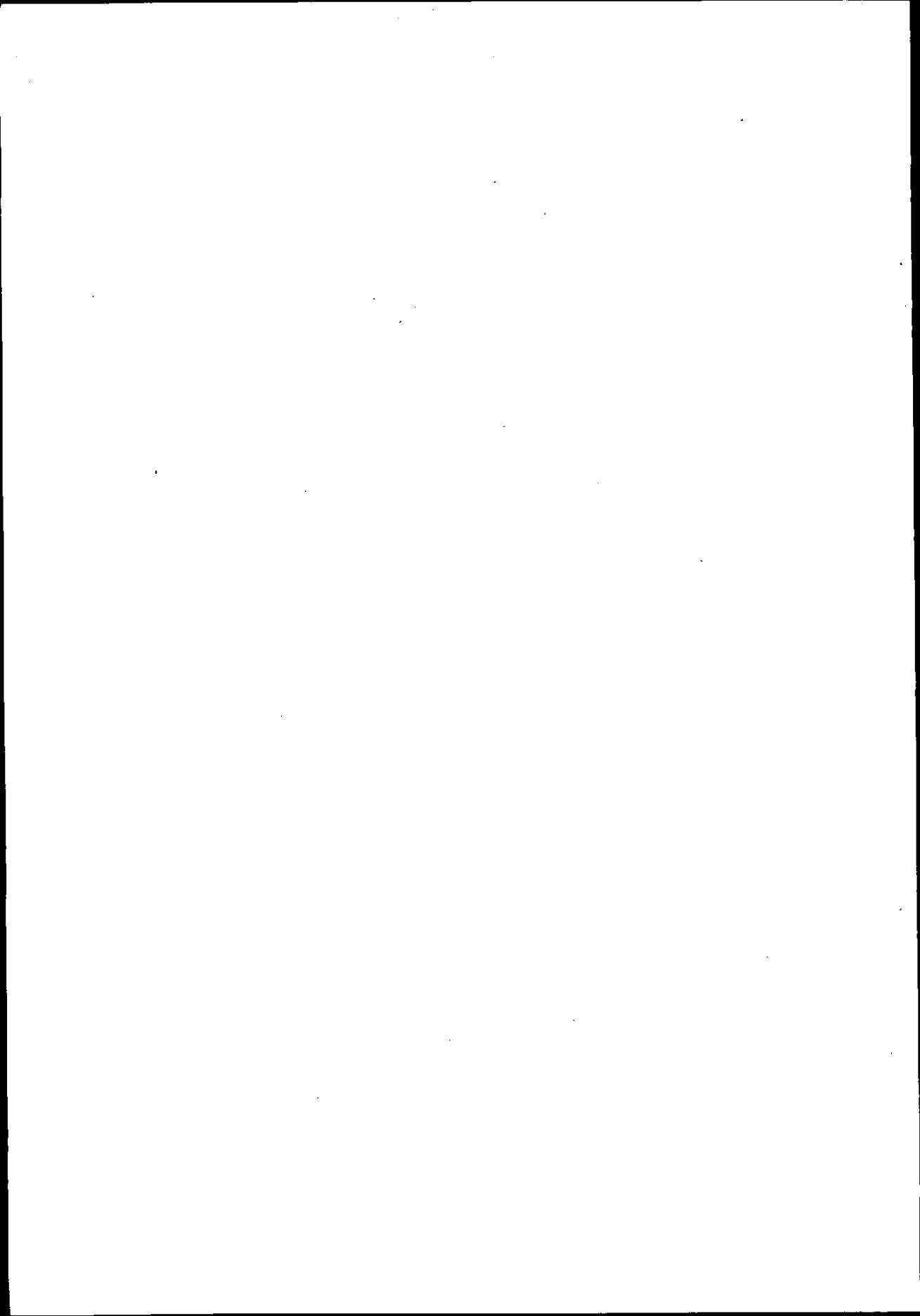
第2回(昭和46年度)研究奨励金 受賞研究の概要

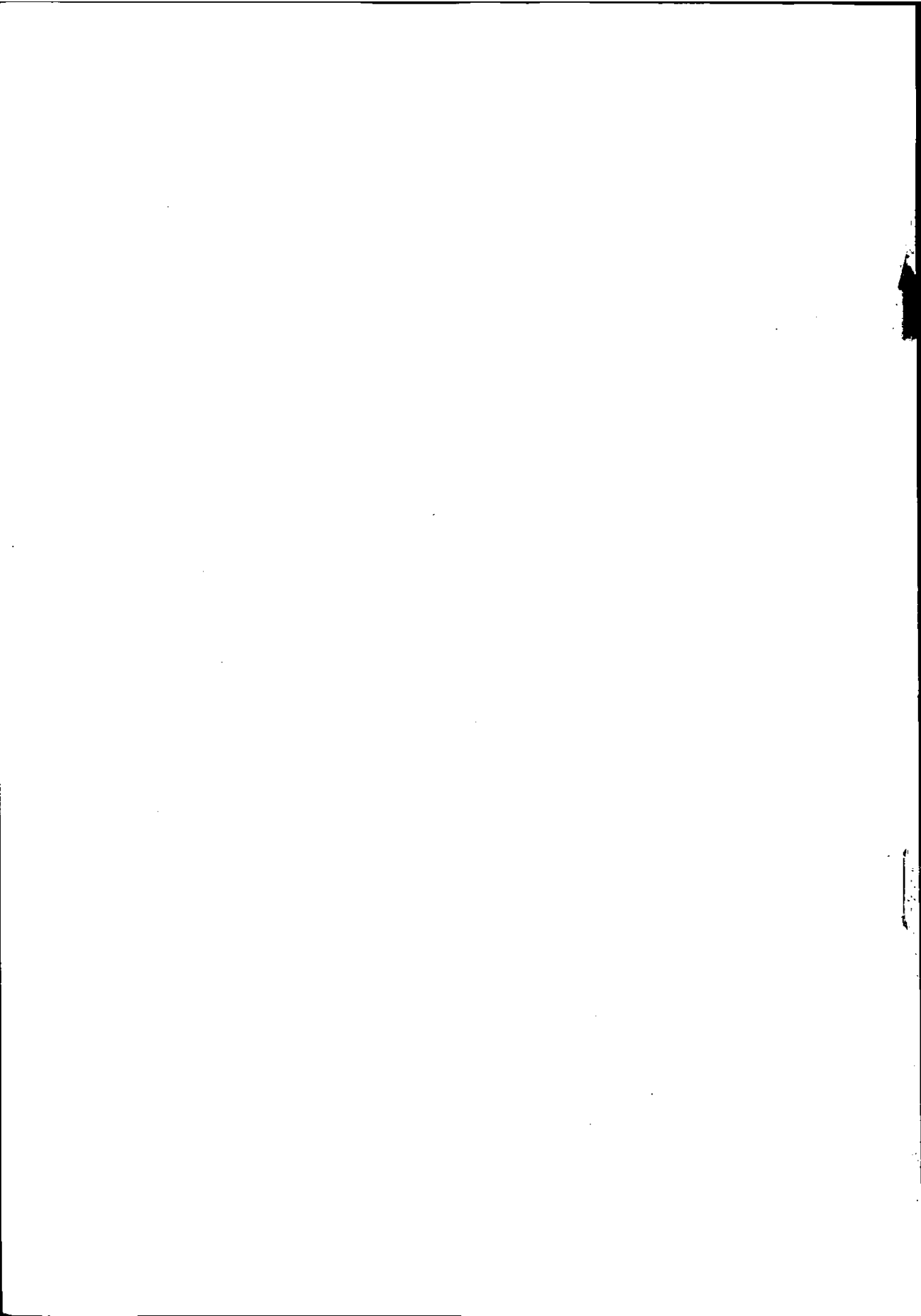
昭和47年 3月

財団法人

日本経営情報開発協会







この研究奨励金制度は、本協会に対する民間の寄附金の一部をもって基金を創設し、その果実を

(1) 電子計算機の高度利用に寄与する技術と知識に関する研究

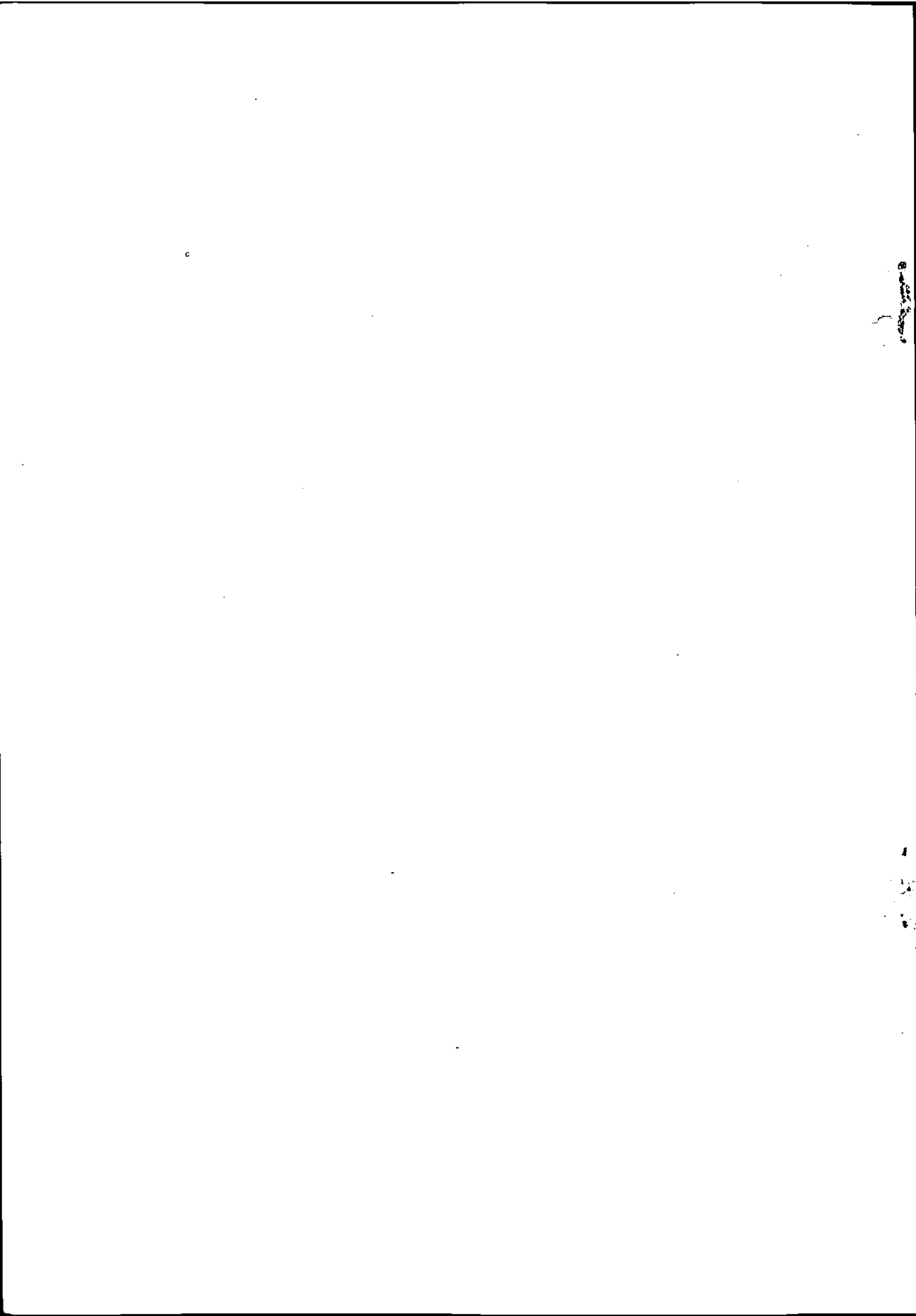
(2) 電子計算機の適用分野の拡大開発に寄与する技術と知識に関する研究

のうち優れたものに対し、研究奨励金として交付することを目的としている。

公募は、年1回行なうことになっており、本協会の依頼した大学、研究機関、関連団体および本協会役員等の推薦人(機関)を通して推薦を受けることとしている。

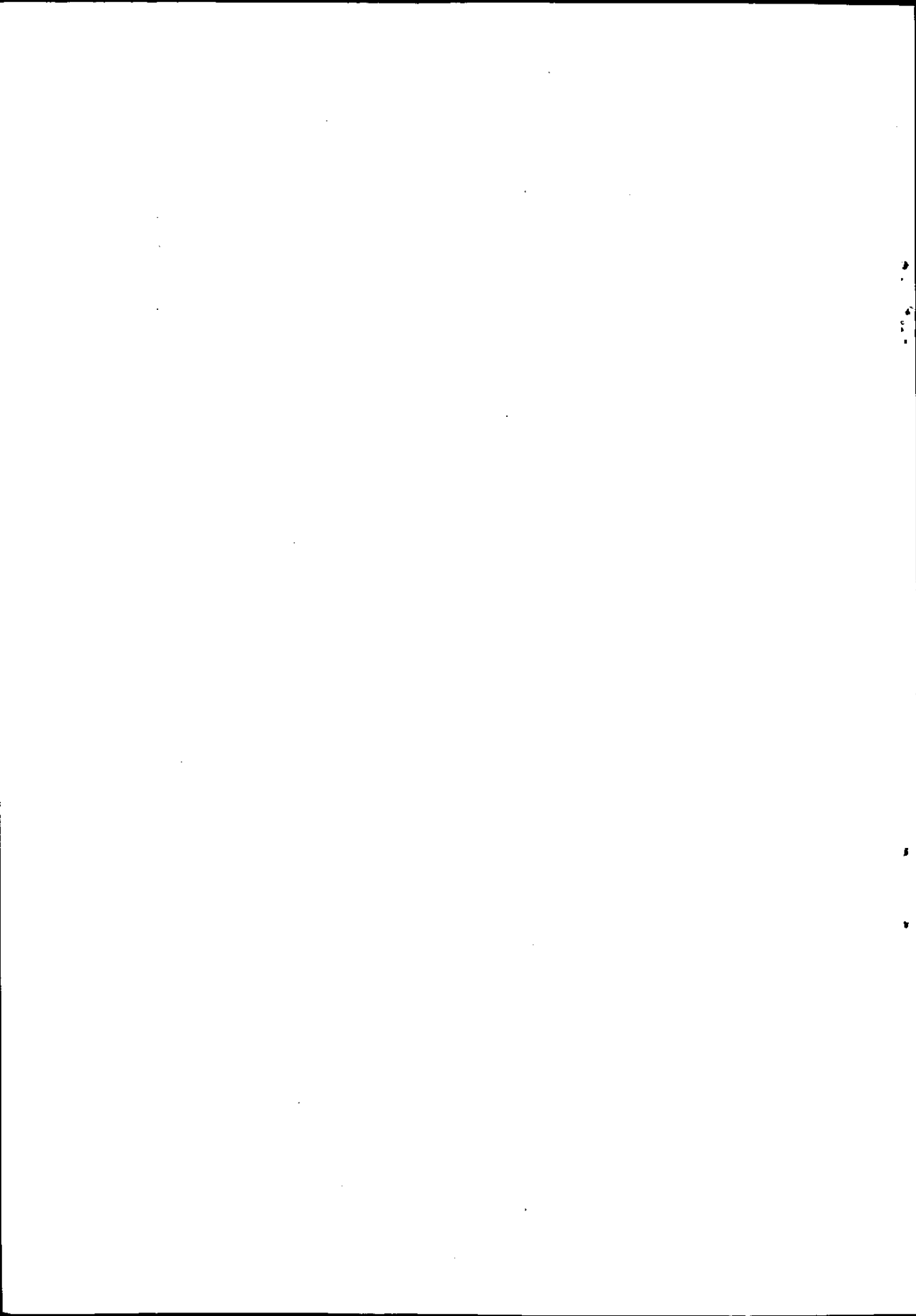
審査は研究奨励金審査委員会において厳正公平に行なうことになっている。

今回はその第2回として昭和46年9月に公募を行なった。



目 次

I	推薦および審査のあらまし	1
	○ 推薦状況	1
	○ 選考方針および選考の経緯	2
	○ 選考の結果	2
	○ 研究発表会	4
II	授与式における平田副会長の挨拶	17
III	受賞研究の概要	9
1.	電子計算機の音声入出力方式の研究	東北大学教授 城戸健一 他5名
		11
2.	法律文に対する論理的推論を伴う質問回答システムの一方式	東京女子大学教授 水谷静夫 東京芝浦電気(株) 斎藤 孝 他1名
		27
3.	胸壁面上における心電図電位分布の計算機映画によるカラーアニメーションの開発と実用化	北海道大学教授 吉本千禎 他2名
		53
4.	染色体のオンライン解析システム	日本アイ・ビー・エム(株) 主任研究員 飯坂譲二 東京大学助教授 小泉 明 他2名
		61
5.	記号処理言語の開発とその応用	九州大学教授 藤野精一 他6名
		79
6.	自然語科学情報のデータパターン及びインタレスト・パターン処理の基礎研究	東京大学教授 藤原鎮男 他1名
		91
7.	実時間計算機による思考過程分析用実験システムの開発	大阪大学教授 鈴木良次 他4名
		111
8.	常駐FORTRAN コンパイラの開発	東京大学大学院生 武市正人 他1名
		121



I 推薦および審査のあらまし

○ 推 薦 状 況

本協会が依頼した推薦人(253人)から64件の推薦があった。機関別の推薦状況は次の通りである。

大 学	51件
会 社	9
その他	4
計	64件

なお、研究テーマの種類は次の通りである。

医療に関するもの	15件
コンピュータ教育に関するもの	7
ソフトウェアに関するもの	6
パターン認識に関するもの	5
数学・ORに関するもの	5
そ の 他	26
計	64件

○選考方針および選考の経緯

- (1) 独創的な研究であり、近い将来利用に供せられる可能性の大きいもの、各種の専門分野にまたがる共同研究等を優先的に考慮した。
- (2) 選考は審査委員会（委員長 堀越楨三氏、副委員長 山内二郎氏）において厳正に行なわれた。

○選考の結果

選考の結果次の通り決定した。

件 名	担 当 者	奨励金交付額 千円
電子計算機の音声入出力方式の研究	東北大学教授 城戸健一 他 5 名	2,000
法律文に対する論理的推論を伴う 質問回答システムの一方式	東京女子大学教授 水谷静夫 東京芝浦電気㈱ 斎藤 孝 他 1 名	2,000
胸壁面上における心電図電位分布 の計算機映画によるカラーアニメーションの開発と実用化	北海道大学教授 吉本千禎 他 2 名	2,000
染色体のオンライン解析システム	日本アイ・ビー・エム㈱ 主任研究員 飯坂譲二 東京大学助教授 小泉 明 他 2 名	2,000
記号処理言語の開発とその応用	九州大学教授 藤野精一 他 6 名	1,000

自然語科学情報のデータパターン	東京大学教授	1,000
及びインタレスト・パターン処理 の基礎研究	藤原鎮男 他1名	
実時間計算機による思考過程分析	大阪大学教授	1,000
用実験システムの開発	鈴木良次 他4名	
常駐FORTRAN コンパイラの開 発	東京大学大学院生 武市正人 他1名	1,000
計 8 件		千円 12,000

○研究発表会の開催

今回の交付対象となった研究は、いろいろの専門領域に分れているが、電子計算機の高度利用という点では強く結びついている。

一般にこのような優れた研究はそれぞれの専門領域で発表されているが、他の専門領域の研究者に広く知らされる手段が欠けており、しかも、最近の学術研究は異種の専門領域にまたがるものや、各専門の境界領域に位置するものが多くなっている。そこで本協会は研究奨励金の交付を機会に、これら各領域の研究を一堂において発表し、他部門の研究者に広く周知していただき、電子計算機を媒介として知識や技術の交流を促進することを企画した。

第2回研究発表会は次の要領で開催する。

日 時 昭和47年3月16日(木)～17日(金)

(9:00～17:00)

場 所 霞が関ビル1階 ブラザホール

日時および講演題目

3月16日(木)

9:30～10:30	常駐FORTRAN コンパイラの開発	東京大学大学院生 武市正人 他
10:30～12:00	自然語科学情報のデータパターン及びインタレスト・パターン処理の基礎研究	東京大学教授 藤原鎮男 他
13:00～15:00	記号処理言語の開発とその応用	九州大学教授 藤野精一 他

15:00~17:00 染色体のオンライン解 日本アイ・ビー・エム(株)
析システム 主任研究員
飯坂譲二
東京大学助教授
小泉 明 他

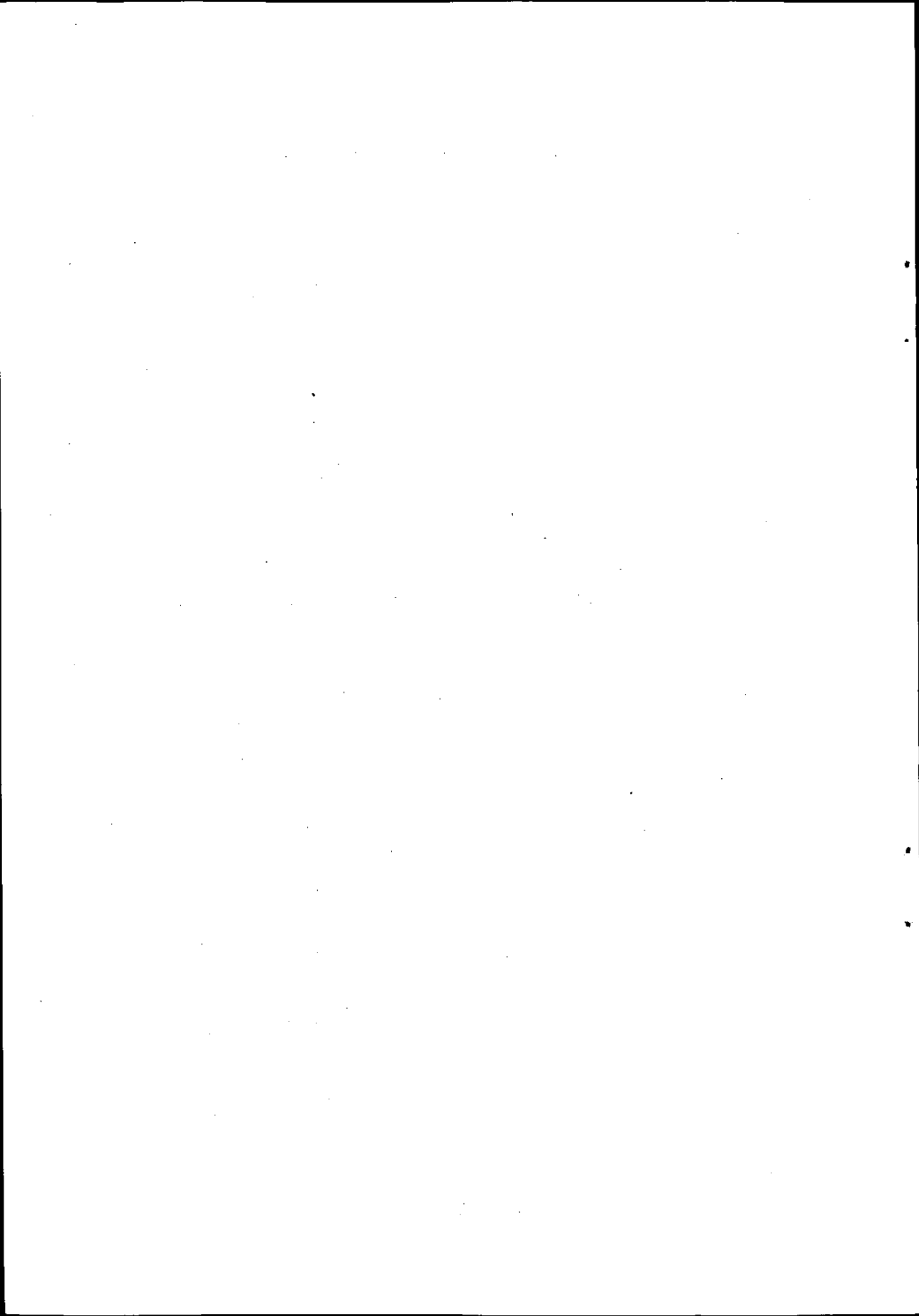
3月17日(金)

9:00~10:30 法律文に対する論理的 東京女子大学教授
推論を伴う質問回答シ ステムの一方式 水谷静夫
東京芝浦電気(株)
斎藤 孝 他

10:30~12:00 実時間計算機による思 大阪大学教授
考過程分析用実験シス テムの開発 鈴木良次 他

13:00~15:00 胸壁面上における心電 北海道大学教授
図電位分布の計算機映 吉本千禎 他
画によるカラーアニメーションの開発と実用
化

15:00~17:00 電子計算機の音声入出 東北大学教授
力方式の研究 城戸健一 他



Ⅱ 授与式における平田副会長の挨拶

このたび第2回の研究奨励金をお受けになる榮譽を担われた皆様に心からお祝い申し上げます。

本協会は、コンピュータの高度利用、適用分野の拡大開発をすすめることにより、わが国の社会・経済の発展に寄与することを目的としており、研究奨励金はそのための重要な事業の一つであります。

コンピュータの利用や開発に関する研究は非常に広範囲に亘り、かつ奥が深いものであります。殊に、これからの研究は1分野、1学究の殻の中に閉じこもってでは、達成できなくなるのではないかと思います。種々の科学に跨り、色々の分野の人々が集まって研究を推しすすめることが必要であります。それだけに費用も膨大なものになります。

アメリカにおけるコンピュータを基礎とした科学や産業が非常にすすんでおりますのは、このような研究体制をとり、かつ、模大な費用を注ぎこんでいるからに外なりません。

本協会の研究奨励金制度は発足後日が浅く、まだ金額的にも小さいものではあります、これが誘い水となって、わが国のコンピュータ関係の研究が活発に進められる一助ともなれば幸いです。

ると思っております。

今回、皆様方に差し上げる奨励金も誠に僅かでありまして、到底皆様方の研究費をカバーするには足りないでしょうが、これを種子に優秀な成果を上げられることを期待してやみません。

授与式にあたって、一言お祝を申し上げる次第であります。

昭和47年2月16日

山内審査副委員長の挨拶

第2回研究奨励金を受賞されました皆様方に心からお祝いを申し上げます。

実は、あと2カ月ほどで、私の研究者歴は50年になりますが、その間いろいろな形態の研究をして参りましたし、可成りの金額の研究費も使って参りました。

今の日本で問題となりますことは、戦後のGHQの通達により、会計制度が変わった結果研究活動に非常な支障をきたすようになったということです。

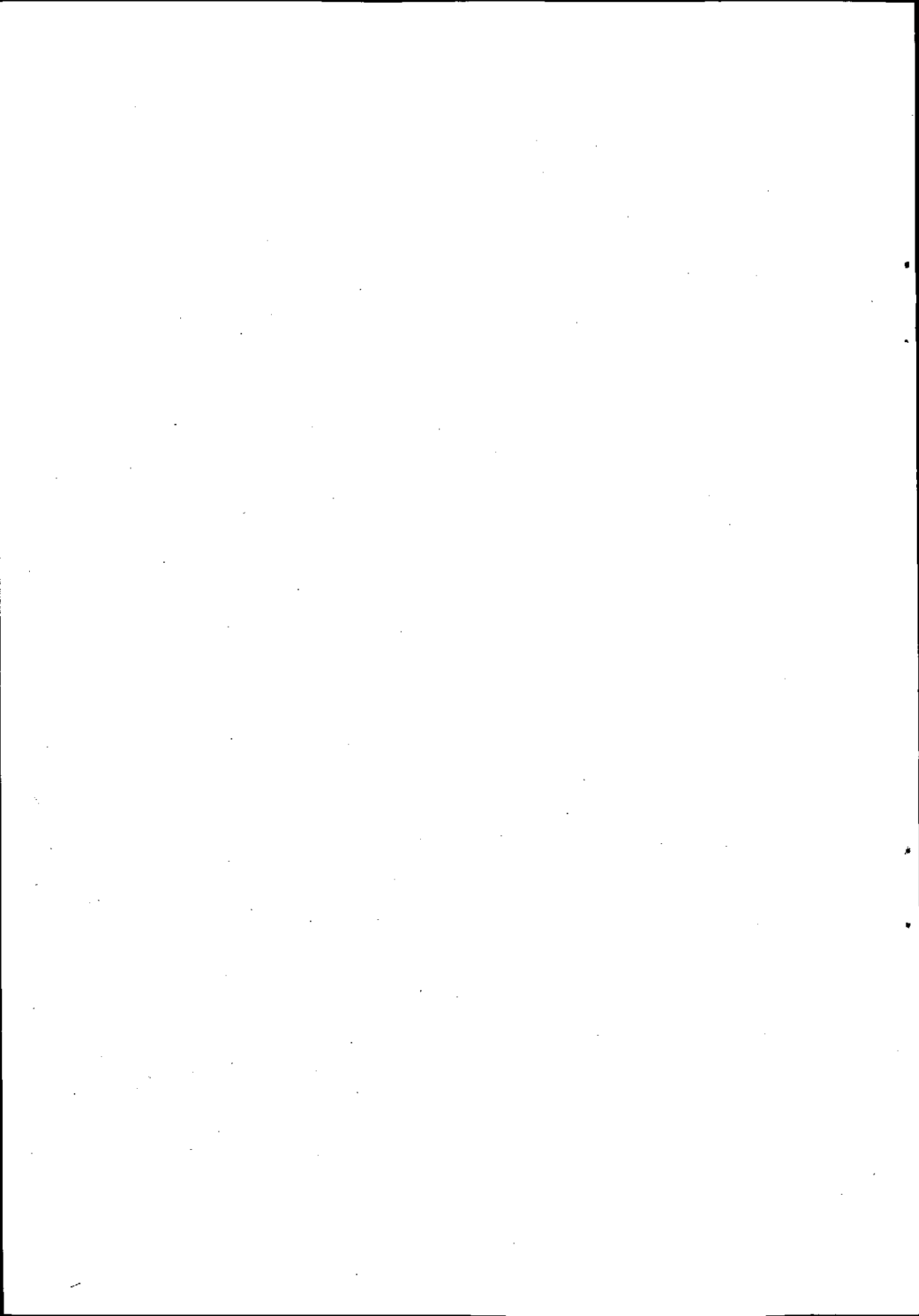
戦前および戦中までは、非常に重要な研究には、相当の金額が、案に使えました。

人も十分に雇うことができました。また、ポケット・マネーを出さなくても、会計検査院は研究費を大幅に認めてくれました。

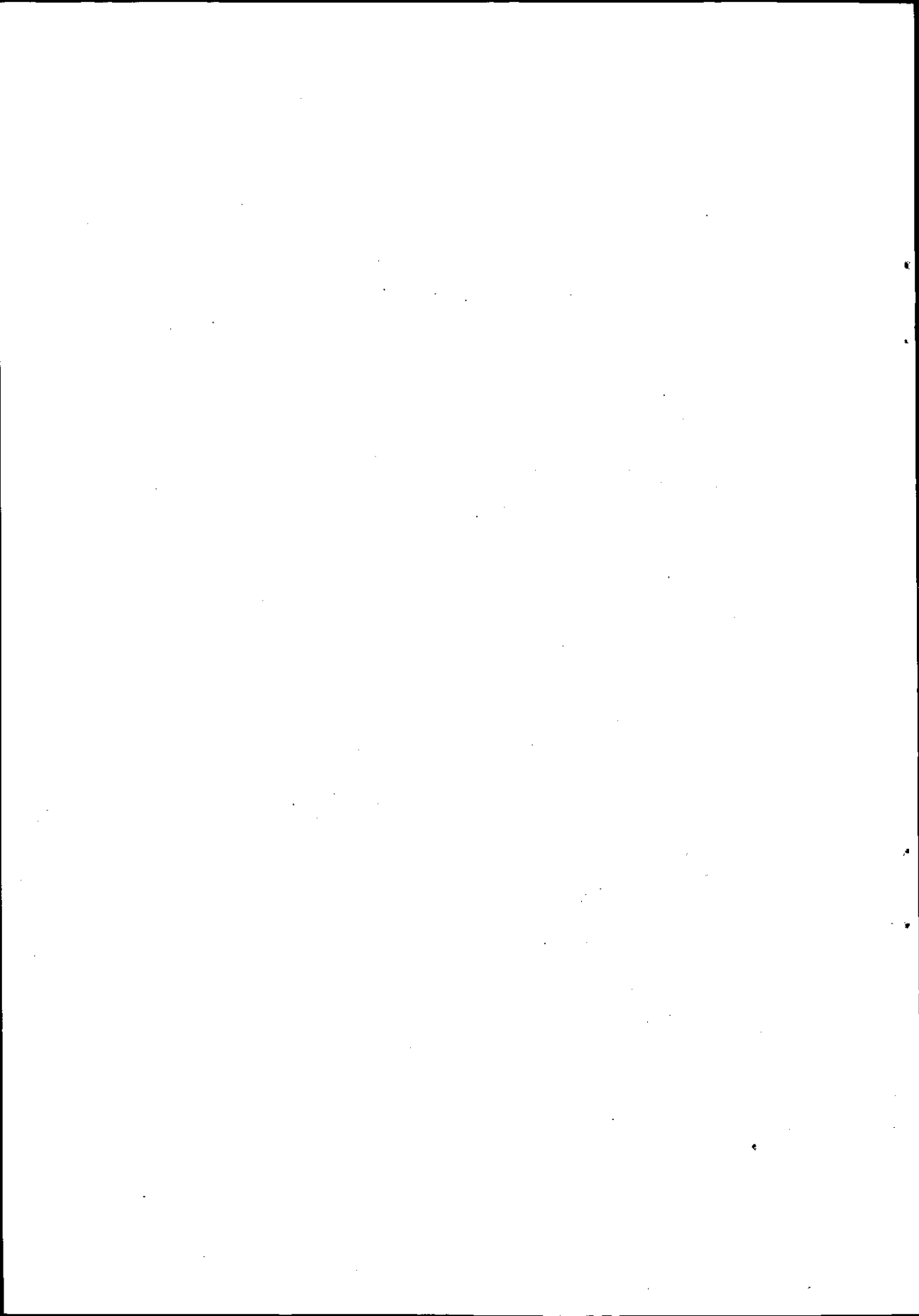
それが戦後になりますと、GHQの通達によって、研究費が非常に窮屈になり、研究者がポケット・マネーを出さないと十分な研究をやっていけないようになってしまいました。日本経営情報開発協会の研究奨励金は、そういう意味におきまして、非常に具合のよい研究の潤滑油になるものと思います。他所の奨励金に比べますと、金額がやや少いかもかもしれませんが、研究者の立場からいたしますと、本当に何とも言われぬ有難いものだと思います。どうか十分に有効に使っていただきたいと思います。

日本人は共同研究が、非常に不得手ですが、今日ここに受賞されました方々の研究は、全部が共同研究となっております。

皆様方が研究をうまくやられますことは、今後のわが国の共同研究を一步前進させる手本となりますので、どうか立派な成果をあげられますことをお祈りして、お祝いの言葉と致します。



Ⅲ 受賞研究の概要



電子計算機の音声入出力方式の研究

東北大学電気通信研究所教授 城戸 健一

” 助教授 鈴木 久喜

” 助教授 比企 静雄

” 助手 粕谷 英樹

” 助手 金森 吉成

” 助手 板橋 秀一

電子計算機の入出力媒体として音声を用いることの必要性和重要性とは、電子計算機の大型化・高性能化にともない、ますます増大してきている。それは、電子計算機の用途が、数値情報処理にとどまらず、数値化されない情報の処理の分野にひろがっていることにも由来するものである。しかし、音声研究の現状は、その要求に応えるには未だ不十分であって、電子計算機の音声出力が特別に簡単な用途（たとえば、数値計算の結果の回答、駅における案内のアナウンスのようにきまり文句で語りが少ないもの）に使われはじめたにすぎず、音声による情報の入力の実用化は、まだかなり先のことになるものと考えられる。そして、音声出力を一般的なものとするためにも、音声による情報入力のための自動認識のためにも、人間における音声の生成と認識の過程に基礎をおいた本質的な研究の推進が要求されているわけである。

1. 音声研究と実用化の現況

音声の研究の歴史は古いが、近代的な研究が始まったのは第2次大戦後であり、電子計算機の出現は、研究に実用的な目標を与えるという面と、研究のために有力な手段を提供するという面との両面で、研究に拍車をかけることになった。最初、音声の合成と自動認識とは、別の課題として扱われてきたが、どちらも音声の本質にさかのぼらなければ研究の進展が望めないということから、それらは、目的を多少異にする同じ研究でなければならないことが認識されるようになった。

現在実用化されている音声合成方式は、録音編集方式とよばれるもので、音声をこま切れにしたものを録音しておき、必要な順序でつぎ合わせるものであるから、一見、音声の本質にまでさかのぼっての知識はいらないかのように見える。しかし、音声は、活字を並べるようにつぎ合わせることができず、この方式では、つぎはぎの単位を短文または熟語の程度に長くする必要がある。もし、一音節または二音節の単位でつぎはぎをするならば、ピッチ、強さ、ホルマント等の時間的变化を伴った多数の音節を用意して、それを適当につないで行かない限り、何を話しているのかも全くわからなくなるのである。そこで、この方法は大量の記憶容量を必要とし、メッセージの種類を豊富にするには不適當である。そのためには、音調の制御の困難性を克服しなければならず、制御法則を見出すためには、人間における音声発声の心理的、生理的過程までも解明しなければならないことになる。録音編集方式よりは本質論的な音声合成方式として、口道の伝達関数を電子回路または計算機シミュレーションで模擬し、その極ど零とを制御するいわゆるターミナルアナ

ログ合成方式や、口道の幾何学的形状を同様に模擬するいわゆる口道（ボーカルトラクト）アナログ合成方式が研究されているが、そのために制御法則が多少簡単にはなるにしても、心理的・生理的過程の解明が必要なことはかわらない。

音声自動認識に関するものとしては、単母音の性質はすでによく調べられ、第1、第2および第3ホルマント、あるいは、第1、第2ホルマントと基本波周波数がわかれば、単母音はほとんど完全に認識されることがわかっている。したがって、ホルマント周波数と基本波周波数抽出の技術が確立されさえすれば音声自動認識は可能となるかのような印象を受けるが、連続音声の中ではそれらの時間的変化が大きく、その変化を記述する法則（これは、音声合成のための法則と同じである）を作ることの困難、個人性による変動の除去の困難、それらの経時的変動補正の困難等のために、末だに連続音声の自動認識は未完成の問題である。数字音声、あるいは、比較的少数個の限定単語音声の自動認識が非常に高い率で実行できたという報告もあるが、それらは、個々の発声者が短時間の間に同じ発声をくりかえし、前の発声から抽出したパターンを標準パターンとして後での発声を認識するという方式をとっていて、音声の本質が必ずしも解明できなくても、現在の段階の知識をうまく利用すればここまで行き得るという例を示したものである。しかし、これでは、認識しようとする単語数が非常に多くなったり、連続した文章になったり、また、不特定多数の発声者の発声を扱うようにすることは困難であり、そのためには、より本質論的な方法が必要であると考えられる。

2. 本研究の思想的背景

人間における音声の生成と認識の過程をモデル化すると、図1のように表わされる。

すなわち、思考と記憶によって作られた概念は言語的表象として発声器官を制御するための神経指令を作るが、神経指令は直接的には弁別

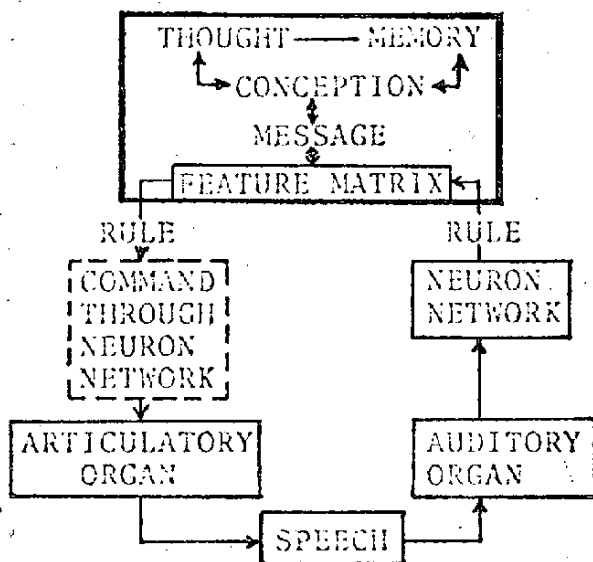


図1. Model of speech processing in man

るものと考えられる。神経指令は発声器官、すなわち、肺、声門、口道等を動かして音声を発声させる。発声された音声は、聴器に入り、蝸牛で神経回路網を通るパルスに変換され、神経回路網で処理されて大脳の聴覚中枢に至る。聴覚中枢ではこれを言語中枢に送り、言語中枢ではこれを発声過程におけると同様の弁別の特徴として受けとる。発声の際には、発声のために作られた弁別の特徴の時系列が聴器を通しての入力によって作られたものと一致するか否かで、発声の制御を行なう。聴取のみの過程でも、認識のためには弁別の特徴を一度形成することが必要で、その弁別の特徴の時系列が言語的表象と化して記憶内容もしくはそれと思考とによって

作られる概念によって作られる言語的表象と一致するか否かがたしかめられ、一致したときはじめて、音声認識されたことになる。

著者等は、このような人間における音声の生成と認識の過程を解明し、その模擬をすることによって、音声の合成と認識の方法を確立しようと考えている。一見、遠回りのように見えても、このような本質論に基いた方法でなければ、着実な成果を上げることはむずかしいと信じているのである。

3. 生理的パラメータによる音声の合成

発声器官としての口道は、それを動かす多くの筋肉によってその形を作られ、目的に応じた音声を作っているのであるが、その音響的性質は、口道の断面積の口道中心線にそっての形で規定される。口道断面積パターンの時間的变化は、口道を形作る「あご」「舌」その他の運動によって起きるので、それらが動くときの生理的・力学的拘束を受けていて、その拘束条件に反するような変化はできない。したがって、口道を変化させる筋肉を動かす神経指令の段階を合成器の入力とし、筋肉や骨などによる制約の加わった口道断面積の時間的变化にもとづく連続音声の合成を行なえば、比較的少ない情報量の指令で、自然な音声を合成できるものと考えられる。

このような観点にたって口道の運動に関係する筋肉をしらべると、主要なものは14個あり、それらと口道の形との関連も生理学的研究の成果を利用してつけることができることが明らかになったので、NEAC-3200を使った計算機シミュレーションによって、母音の合成を行な

っている。現在、単母音および2音節の連続母音の合成に成功し、制御法則作成のための研究を進めている。

口道を音響的に駆動するものは声帯音であり、音声の自然さ、個人性等は、音韻性に関係する口道の形よりは、声帯音に大きく依存する。特に、アクセント、パターンは声門の動きの生理学的観測によって明らかになることを発見し、それによって、発声時のアクセント、ピッチの変化を詳細に調べる方法を確認した。この結果は、前記の音声合成装置の声帯波音源の制御のために重要なものである。

4. 音声の自動認識

4.1 方法論

従来、音声の自動認識は、音声波の短時間スペクトル、帯域3波器による分析出力、ホルマント周波数等に基づく類別という形でのみ進められてきた感があるが、著者等は、第2節に述べたような思想的背景のもとに、図2のような認識システムを考えている。すなわち、音声をまず周波数分析をしてその出力から必要な特徴量を抽出する。周波数分析としては、人間の蝸牛における周波数分析に近いことが望ましいと考えられるが、蝸牛から大脳に至る間の神経回路網での処理について神経生理学的に明らかにされたものをみると、低次のニューロンでは分析出力の周波数選択度を向上させるだけのことが行なわれ、高次のニューロンで周波数の上昇、下降、振幅の増加、減少などに対応する信号が表われているようであることと、蝸牛の分解能($Q \approx 3$)では後の処理に便利でないため、 $Q = 3.9$ と $Q = 6.0$ の単峰性3波器を作って実験した結果から

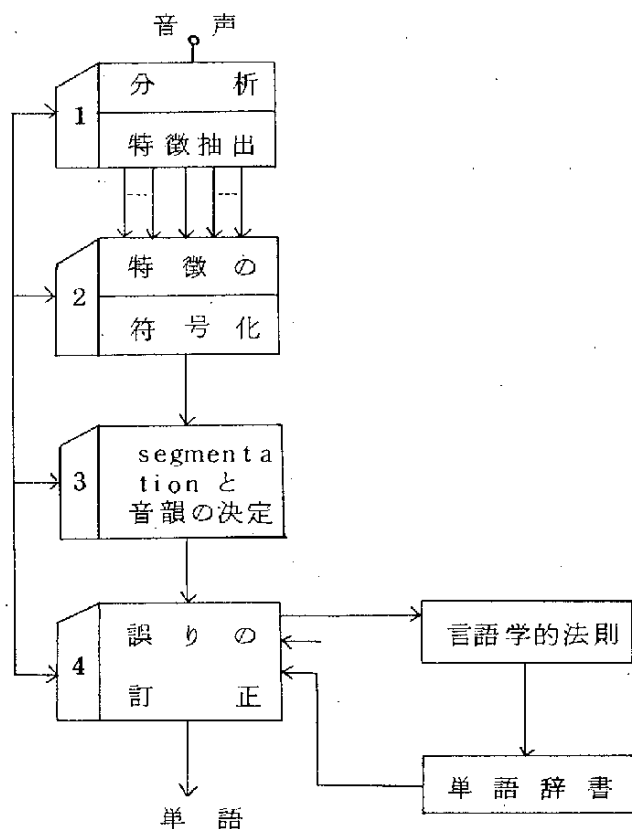


図2. 単語音声自動認識のシステム図

$Q=6$ の単峰性3波器を周波数分析に用いることとした。(これは、研究の進展によって変更されることもあり得る)このような分析器の出力を利用して、音韻ごとの区分け(セグメンテーション)と分類(クラシフィケーション)が行なわれる。その結果は、ほぼ、音素記号系列のようなものになるであろう。このようにして作られた音素記号系列は、必ずしも正しい認識出力になるとは限らない。それは、連続音声中には、子音、母音の脱落、ホルマント周波数、基本波周波数あるいは強さの極端な移動(変化)があるのが普通で、単純な処理方法でそれによる誤り

をすべて解決することは不可能だからである。そこで、最後の段階として、言語学的情報にもとづく誤りの訂正を行なう。これは、図1のモデルでは、中枢において行なわれていることである。音声のもつ言語的性質について調べた結果では、音素の結合規則は非常にゆるやかで、それだけでは誤り訂正に効果を上げることが困難なことがわかったので、意味情報としての、辞書との対照を主に行なうこととした。辞書との対照の効果は非常に大きく、たとえば、無声破裂音(p・t・k)相互間、あるいは鼻子音(m・n・ng)相互間の類別が全くできない場合(音響的性質だけで類別することは一般に困難である)それぞれ37.5%あるいは、21.6%の誤り率になるものが、辞書を用いて訂正すれば、それぞれ6.3%あるいは3.4%に改善される等の結果が得られている。

辞書との照合は、このような音素記号のレベルでなくても、弁別の特徴のレベルでも行なうことができる。

音声認識を一つの方式だけで行なうと、その方式に特有の誤認識をする可能性がある。そのような場合には、幾つかの認識系を同時に動作させ、それらの結果から適当な選択、または誤り訂正を行なうことが有効であると考えられる。著者等は、図3のように、このような重複処理を行なうことが必要であると考えている。

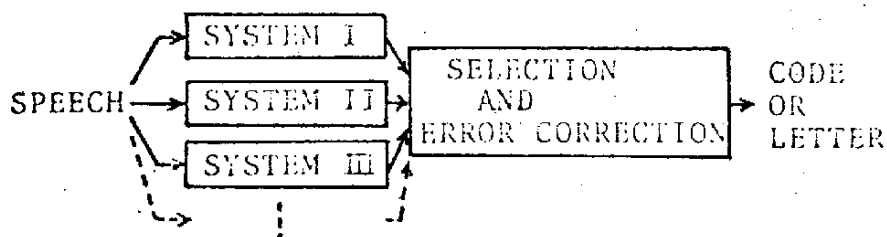


図 3. Recognition scheme

4.2 $Q=6$ の単峰性 3 波器による分析と認識

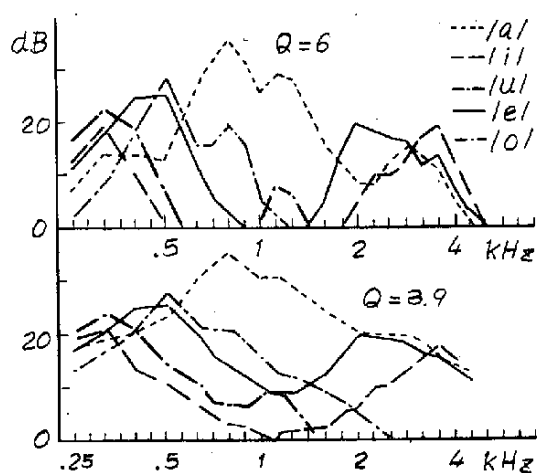


図 4
Spectra of five Japanese vowels analyzed by the Single tuned filters.

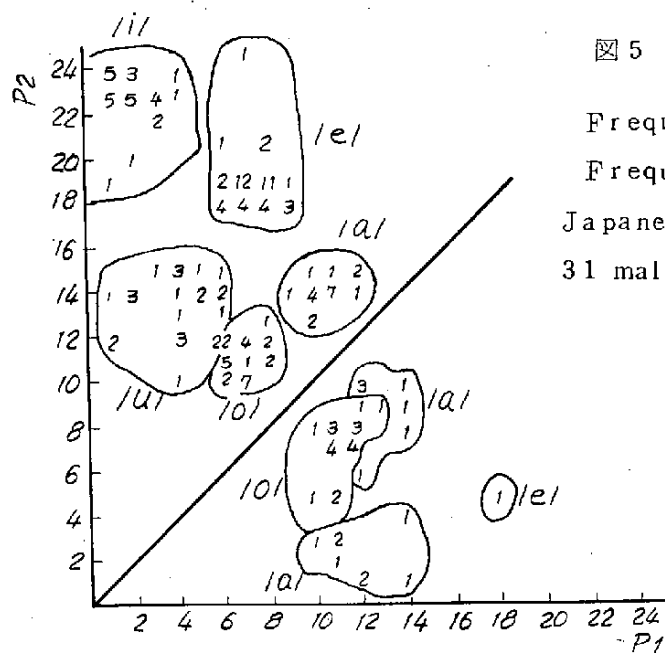


図 5
Frequencies of P1 vs.
Frequencies of P2 for five Japanese vowels uttered by 31 male adults.

図2の第1段階の分析器として、どのような3波器を使うべきであるかを検討する一段階として、 $Q=6$ および $Q=3.9$ の単峰性3波器によって単母音の分析を行なった。3波器の中心周波数は、 250Hz から 6300Hz まで $1/2$ オクターブ等間隔29チャンネルである。その分析出力の2例を図4に示す。この図だけからはどちらがよいという積極的な結果は得られないが、他の多くの分析の結果も参考にして、さしあたり、 $Q=6$ の3波器をえらぶこととした。

特徴量として、この分析出力から何をとり出すべきかが問題である。著者等は、大まかな形としてのスペクトルモーメントと、ニューロン回路網でのスペクトルの尖鋭化や動きの表現に直観的に対応づけやすいスペクトルのピーク（ローカルピーク）との両方についての処理を試みたが、ここでは、ローカルピークの処理についてだけ述べる。

31人の成人男子の発声した日本語5母音の分布を、大きい順に第1、第2のピーク（それぞれ、P1およびP2とする）のチャンネル番号によって図示したものが、図5である。この図で 45° の斜線より下に来るのは周波数の高いチャンネルに生じるピークの方が低周波数のピークより大きい場合であって、ローカルピークの大小の関係は余り重要でないから、これを周波数の低い方から順に並べなおすと、第1、第2ホルマントパターンと類似のものとなる。しかし、/a/の第2ピークが非常に低い周波数にくるのは、基本波の低次の高調波が独立して第2ピークと観測され、第1、第2ホルマント周波数が近接したためにその付近に第1ピークが1個だけ検出されたことによるものである。したがって、この場合だけは、/a/のP1-P2パターンは、周波数の低いP2をすてて、P1とP2が一致して 45° の斜線上にあるものとしなければな

らない。このように、P1-P2パターンによってほとんど完全に単母音の識別ができることは、図から明らかであろう。

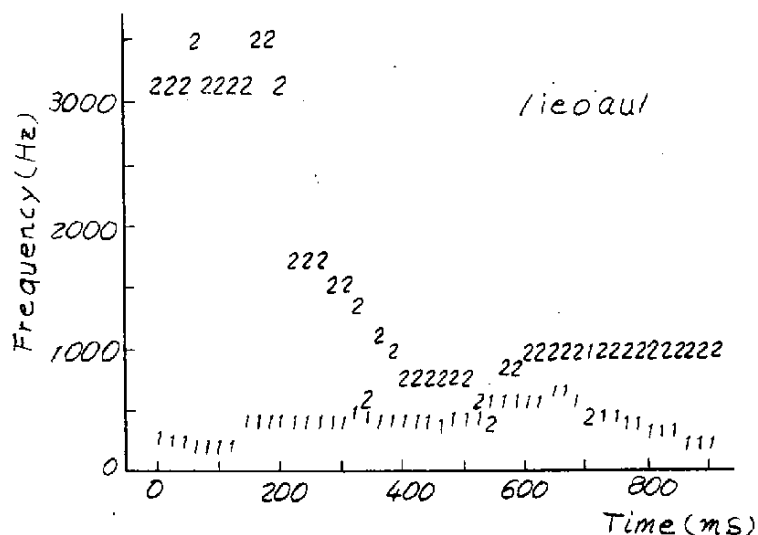
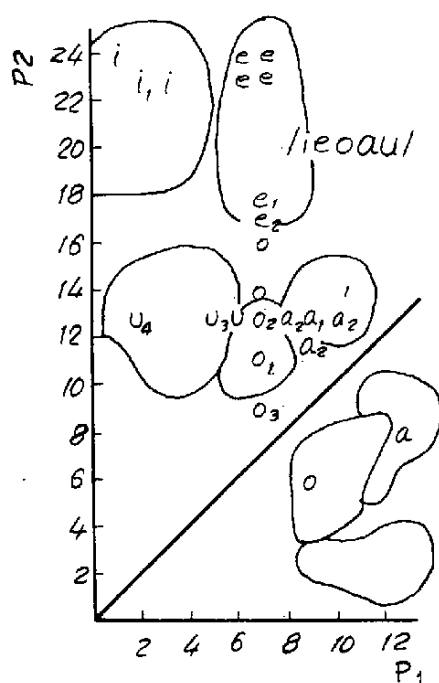


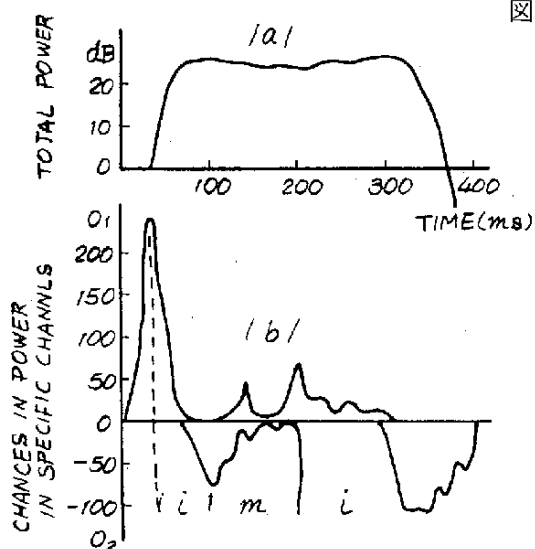
図6 Transitions of P1 and P2 in the vowel continuation /ieoau/

次に、母音連続の分析を行なった。/ieaou/ の分析例を、図6および7に示す。図6は横軸を時間にとって、P1およびP2の周波数の時間的变化を示したもので、図7は、それをP1-P2平面上にプロットしたものである。図7には、図5と同じ母音領域を示す。連続して発声したために、前図の領域外に検出される場合が出ている。しかし、これらは別に作った規則によって何れかの母音として認識されている。その結果を、i, e, a, o, u の文字で示す。下添えの数字は、生起回数である。

同じ分析出力を用いて、連続音声のセグメンテーションを行なうこともできる。これは、一つは、たとえば10 ms, ごとの分析出力を音素



⊠ 7 Frequencies of P1 vs. Frequencies of P2 for the vowel continuation/ieoau/. Subscripts indicate the occurrence frequencies.



⊠ 8 Segmentation of an utterance/imi/

記号系列に変換した後で、同種の記号群ごとに分ける方法であり、この場合、継続時間の短いものを消去するか否かには、言語学的処理も必要となる。別の一方法は、分析出力の時間的变化に注目する方法で、その結果の1例を図8に示す。/imi/という発音の場合には、音声のエネルギーの時間的变化はほとんどないので、それに基づくセグメンテーションはできない(破裂子音では、それだけでできる場合が多い)が、分析出力を適当に処理することにより、このように行なうことができる。

5. おわりに

電子計算機の音声入出力を可能とすることを目指し、かつ、人間における音声の生成と認識に立脚点をおいて行なっている著者等の研究を、その背景と思想的根拠に基いて概説することを試みた。しかし、そのすべてを述べるとほり大な紙数を要するので、音声合成に関する研究の方法論と、自動認識に $Q=6$ の単峰性3波器を用いて行なっている最近の研究の一部を述べるに止めた。本文では簡単に触れただけであるが、自動認識における単語辞書の利用のような、言語学的情報の利用は非常に重要である。

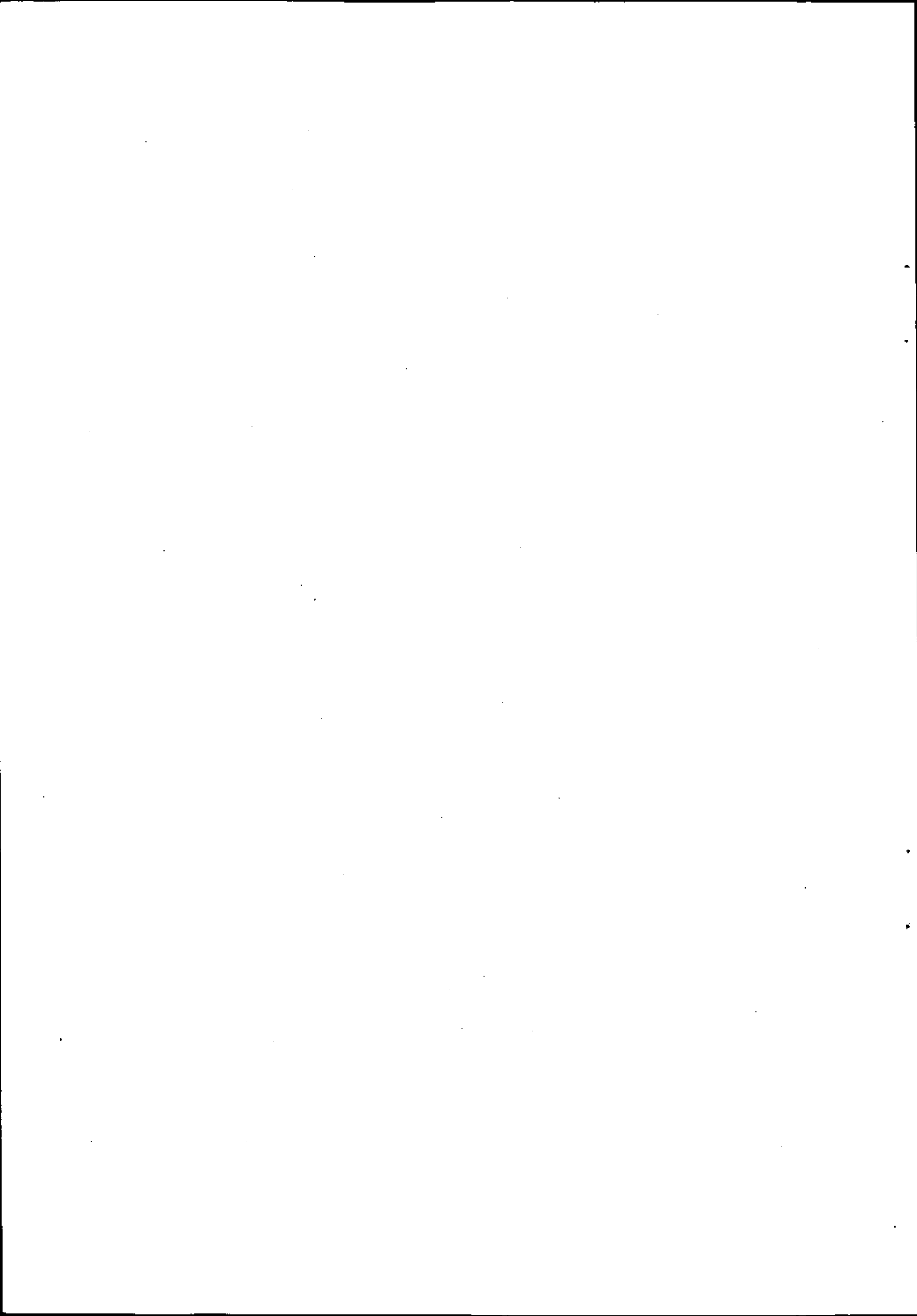
文

献

1. 粕谷, 鈴木, 城戸: 年令, 性別による日本語5母音のピッチ周波数とホルマント周波数の変化, 日本音響学会誌, 24巻6号(1968)
2. 城戸, 粕谷, 板橋: 音声自動認識系のモデルについて, 電子通信学

- 会電音研資料 E A - 6 8 - 2 4 (1 9 6 9 - 0 2)
3. 野口, 城戸: Q の低い単峰 3 波器による母音の分析, 電子通信学会電音研資料 E A - 6 8 - 2 5 (1 9 6 9 - 0 2)
 4. 板橋, 鈴木, 城戸: 単語中の幾つかの子音の辞書による識別, 電子通信学会論文誌 5 4 - C, No. 1 (1 9 7 1)
 5. 板橋, 鈴木, 城戸: 日本語の高頻度単語の音形上の幾つかの性質, 電子通信学会論文誌 5 4 - C, No. 5 (1 9 7 1)
 6. 板橋, 牧野, 城戸: 単語音声のセグメンテーションに関する一考察, 電子通信学会電音研資料 E A 6 9 - 2 1 (1 9 6 9 - 1 1)
 7. 板橋, 牧野, 城戸: 辞書と音形規則を利用した単語音声の認識, 東北大学通研シンポジウム論文集 8 (1 9 7 1) II-9-1~16
 8. 板橋, 城戸: 辞書と音形規則を利用した単語音声の認識, 日本音響学会誌, 2 7 巻 9 号 (1 9 7 1)
 9. 比企: 生理的パラメータによる音声合成のための調音機構のモデル, 日本音響学会誌, 2 7 巻 9 号 (1 9 7 1)
 10. 北脇, 後藤, 城戸: 音声分析用フィルター群についての一考察, 電子通信学会電音研資料 E A 7 0 - 2 4 (1 9 7 1)
 11. 垣田, 比企: 候頭の制御状態の観測の一方法, 日本音響学会講演論文集 1 - 1 - 9 (1 9 7 1 - 5)
 12. S. SATO, Y. KAKITA, S. HIKI: Model of Accentuation of Loanwords in Japanese, 日本音響学会誌, 2 8 巻 1 号 (1 9 7 2)
 13. 白淵, 粕谷, 城戸: スペクトルモーメントを用いた単語音声の音素群への区分化, 日本音響学会講演論文集 3 - 1 - 1 7 (1 9 7 1 - 1 1)

14. 若原, 鈴木, 城戸, スペクトルの local Peak による単語音声の
音響的特徴による記述, 日本音響学会講演論文集 2-1-18
(1971-11)



法律文に対する論理的推論を伴う
質問回答システムの一方式

東京女子大学教授 水谷 静夫
東京芝浦電気 榊 孝
電算機システム技術部

〃

大木 令子

〃

1. 目的と課題

〔目的〕 この研究は、蓄積情報ファイルから質問文に関連して捜し出されたデータに基づき、論理的推論によって回答を導く、事柄検索〔fact retrieval〕の一手法の開発を目的とする。

計算機による情報検索の汎用システムとして現在実用段階に達しているものは、文献検索の性格のものであると言っても大過あるまい。文献検索は所望情報の提供という観点からすれば間接的な方式である。従って当然、直接に所望情報を引き出す事柄検索が要求されることになる。特急券発売など特殊目的に限定されたシステムでは事柄検索が既に実行されているわけであるが、こゝに意図するのは、更に広く、情報ファイルの中身を入れ替えるだけで同じ探索原理によって種々の分野のデータが処理できるような、柔軟性を蔵するシステムである。無論これは最終目的であって、今ただちにそれを目指すことはできない。そこで、後述のような課題を立て、直接にはそれについて研究する。

以上の趣旨からしてこの研究では、原理上は、対象をきびしく限定する必要がなく、かつ将来の発展性が見込まれる — という条件を満たす

方法を求めることになる。事柄検索システムとして良いものがなかなか現われないのは、どんな方法が有効かの見定めがつけにくいからである。これを裏から言えば、文献検索がそれなりの実用段階に達したのは、その手法の原理的な単純さによる。そこでは、情報記述用（および質問用）言語として keyword 等いわば名詞的な語を主軸とする素朴な syntax に従う言語を使い、検索性論理としてはわずかに命題論理の範囲にとどまっている。これでは言語としての記述力が乏しく、論理の点でも射程が小さいから、この線での補強を図っても事柄検索に成功する見込みは期し難い。一方、いわゆる自然言語そのまま（あるいはこれを少々加工したもの）による方式は、それができれば最も望ましいけれども、自然言語自体の本来の性質からして、また自然言語研究の現在の水準からして、はなはだ困難である。それゆえわれわれは厳密な論理の裏づけを持ちながら記述力において柔軟な形式言語を求める方向を選ぶ。そのような形式言語として、後述のとおり、ゲーデルの公理的集合論に採用されたのと同種の言語を使うことにする。

問題はどのように事柄検索システムに使う言語と論理という根本的な所にまで立返るわけであるが、これをただ抽象的に考えているだけでは伸展が乏しい。そこで次のような課題を設定した。

〔課題〕 事柄検索のうち、処理法の性質がかなりはっきりしておりかつ用途も相当に広いと思われる、法令検索による質問回答システムを考える。具体的には特許法を対象として、「かくかくの事が適法か」

「かくかくの権利／義務があるか」「かくかくの事が適法であるためにはどんな条件が課せられているか」というたぐいの質問に対し、特許法から演繹できるときには直接にその答（すなわち内容）を出し、演繹が

(このシステムでは)できないときにはその旨を告げる、モデル的なシステムを作る。

なぜ試みとして法令を選んだかは§ 2.4 に述べるとして、次の注釈を加えておこう。法律関係の事柄検索には少なくとも下記三種のタイプがあると思われる：

一、われわれの課題のように、いわば一般的な、適法性に関して答えるもの — 模型化して言えば、法律の条文は概して「 $\mathcal{A}$ なら $\mathcal{B}$ 」の形をとっている。そこで質問「 $\mathcal{C}$ なら $\mathcal{B}$ か」に対して然りか否かを、また質問「どんな条件の下で $\mathcal{B}$ が適法か」に対して、適法な $\mathcal{A}$ が規定されていればその $\mathcal{A}$ を、そうでなければその旨を、一般的な問題として答える種類のものである。

二、たとえば遺産相続の場合の相続人間での分配額を答えるというように、特定の規定「 $\mathcal{A}$ なら $\mathcal{B}$ 」に関し $\mathcal{A}$ のパラメタを与えて $\mathcal{B}$ の値を答えるもの — 計量法学的な処理とも言えよう。

三、「 $\mathcal{A}$ なら $\mathcal{B}$ 」についての一般的な判断でなく、この適用の仕方すなわち個々の具体的ケースがその規定に言う $\mathcal{A}$ たり得るか否かの判断に関する、先例を求めるもの — これが判例探しであるとも言えよう。

われわれがこの研究で直接に念頭に置いているのは、無論、上記の一の仕事である。

2. 課題に対する方略

2.1 法律文記号化の意義

法律に関する質問回答システムの試作に当り、なぜ法律の原文をゲージル言語のような一般にはなじみの薄い形式言語表現に翻訳するのか、

すなわちわざわざ記号化の手間をかけるのかについて、こゝで述べておきたい。それはこのシステムの考え方と深くかかわる事柄だからである。

一般論を後廻しとして、まず簡単な例示をしよう。今、未成年の a 君がたばこを吸ったとする。この行為は、未成年者喫煙禁止法第一条に照らして明らかに違法である。「明らかに」と言ったが、どんな判断手続で明らかなのであろうか。この判断を形成する過程を表わすために、次のような仮の記号化を試みる（「仮の」とは、あとで用いる記号化法そのままでなく、説明を見易くする目的で簡略にしてある意である）。

「未成年者 a が喫煙した」ことを、

$a \in M \wedge a \in K$ と表わす。「 a が未成年者の集合に属しかつ喫煙するという集合にも属する」意である。

「満二十年ニ至ラサル者ハ煙草ヲ喫スルコトヲ得ス」を、

$\forall x (x \in M \supset \langle x, K \rangle \in J \text{ 禁制})$ と表わす。「すべての日本国民 x について、もし x が未成年なら、その x を喫煙と組合せて考えた行為 $\langle x, K \rangle$ が法により禁ぜられたものの集合に属する」意である。

更にわれわれは次の事を法適用の公理としている：

$\forall i \forall x (x \in I \wedge \langle x, i \rangle \in J \text{ 禁制} \supset \phi(x, i) = \text{否})$ すなわち、もし人 x が i をし、かつ $\langle x, i \rangle$ が法で禁ぜられていれば、 $\langle x, i \rangle$ という法律行為の適法性は否だ（つまり違法だ）—— という事が、だれのどんなカテゴリーの行為についても成り立つ。

以上に基づいて、先の判断形成の道筋は次のような証明図に写し取れる。（こゝではゲンツェンのNK流のものの略式表示による。）

$$\begin{array}{c}
\frac{a \in M \wedge a \in K}{a \in K} \quad \frac{\frac{a \in M \wedge a \in K \quad \forall x (x \in M \supset \langle x, K \rangle \in J \text{ 禁制})}{a \in M \quad a \in M \supset \langle a, K \rangle \in J \text{ 禁制}}}{\langle a, K \rangle \in J \text{ 禁制}} \quad \frac{\forall x \forall t (x \in t \wedge \langle x, t \rangle \in J \text{ 禁制} \supset \phi(t, K) = \text{否})}{\forall x (x \in K \wedge \langle x, K \rangle \in J \text{ 禁制} \supset \phi(t, K) = \text{否})} \\
\frac{a \in K \wedge \langle a, K \rangle \in J \text{ 禁制}}{\phi(a, K) = \text{否}}
\end{array}$$

これでa君の喫煙行為の違法性が、事実と条文と法適用の公理と形式論理の手續とだけで示せたわけである。先に「明らかに」と言ったのは、われわれが法の運用に関してかような演繹図式を当然の事と感じているからであった。それを形式的にきちんとした形に書き上げると厄介な証明図になるにしても、証明図を書き上げるのと同等の判断形成を当然の事と感ずるのは、この図式がa君のその喫煙行為に一回だけ使えるというようなものでなく、多くの類例について広く使えるようなものだからである。

ここで特に注意すべきは、ひとたび条文の記号化が適切に行なわれ（かつ、法適用の公理も適切に記号化され）たなら、事実認定に関する部分を除いては、処理は形式論理の内部の操作に帰着してしまうという点である。この限りにおいてわれわれは問題を文字どおり「機械的に」扱うことができる。われわれはしかし、当面、適法性の判定までを計算機にさせようとは思わない。情報検索としては、たとえば上図の「 $a \in K \wedge \langle a, K \rangle \in J \text{ 禁制}$ 」を変形した「 $a \in K$ という前提の下で $\langle a, K \rangle \in J \text{ 禁制}$ 」を導く所までを扱おうとする。ここから先の判断は法律家に委ねればよからう。問題はむしろ今述べたような回答をどのようにして導くか

にある。

上記のような形式的操作を施して回答を導くためには、自然言語のままの表現、すなわち法律の原文では、扱いにくい事が多い。操作の観点からすれば法律の文言はなお十分には形式化されていず、また同じ形の表現で異なる事を意味する場合もあるからである。たとえば前記の「コトヲ得ス」は未成年者喫煙禁止法では不作為の義務を表わしているが、一般には権利の否認にも使える。口語の（最近の）法律では前者は「してはならない」後者は「することができない」で表わすことになっているが、「しなくてはならない」が作為義務よりは弱い訓示的規定にも使われるというような事はある。ここに述べたのはわずかに問題の一端であって、文言だけからでは紛らわしい場合が多い（特に、人間と違って良識による補強が期待できない計算機を相手にするとき、かかる紛らわしさは重大な問題となる。）それゆえなまの条文ではなくそれを形式言語で記号化した表現を使うのであるが、記号化は既にある解釈を予定している。法の解釈にはまたそれなりの問題を生ずるが、ここでの解釈はいわゆる文理解釈にとどめるものとする。

2.2 解法の原理

以上の態度でもう一度ふり返って見れば、（制定）法の体系そのものは、現に形式化されているか否か、また実際に矛盾をはらむ部分があるか否かにかかわらず、建前としては無矛盾な公理系であるという性格を有する。これは単独法を取った場合にも同様である。それゆえわれわれの課題を実行するための一つの方針は次のようなものである。

質問が「条件 q を前提として、この前提の下では Q か」という形で発せられるものとしよう。ここに q は命題式の有限列か空列かであり、 Q

は一個の命題式であるとする。また n を 1 以上の (有限の) 整数として
「 $P_1, \dots, P_n$ 」が条文ファイルに蓄積してある条文記号式の適宜の列、
「 X 」が最終的には $P_1 \wedge \dots \wedge P_n$ と両立し得る形をとるに至る命題変項
か空式かであるとして、

- (1) q 自身が矛盾をはらんでいると確かめられれば、その旨を答えて
検索を打切る。
- (2) q が整合的なら (あるいは、もっと弱く、矛盾の存在がつきとめ
られなければ)、下記の甲) か乙) かに帰着させられるか否かの処
理を行なう:

甲) $P_1, \dots, P_n, q, X \vdash Q$

処理後の X が空式 (か全く形式的な条件を示す命題式か) な
ら、「然り」と答え、実質的な条件を示す命題式なら「 X の
下では然り」と答える。

乙) $P_1, \dots, P_n, q, X \vdash \neg Q$

同様に、処理後の X の中身に応じて、「否」または「 X の下
では否」と答える

- (3) 上記のどちらとも判定できなければ「このシステムでは不確定」
と答える。

質問が、適法に Q を行なうための条件を求めるものである場合も、こ
の変種として扱う。質問のこの二大別はいわゆる yes-no question (わ
れわれは「T 型」と称する)、wh- question (われわれは「W 型」と
称する) に対応する。T 型質問は証明問題であり、W 型質問は解の存在
を証明した上でその値 (数値とは限らない) を見出だす計算問題である。
われわれは W 型質問の処理を T 型質問の処理の応用として行なおうとし

ている。

さてT型質問の処理は、いわゆる「定理の自動証明」と同じタイプの問題である。ところで一階述語論理で既に決定手続がない事が知られているから、われわれは当然、ある種の制限を課した上でのアルゴリズムの発見を志すことになる。このシステムはいわゆる heuristic なものではない。前記甲)か乙)かに帰着させようとする手続をあらかじめシステムに与えておいて、それにのっとった推断をさせるわけである。しかもそれを連系操作 (string manipulation) で行なおうとする法律の条文は概して

$$\forall x \forall y \dots \forall z (A_1 xy \wedge \dots \wedge A_m xz \supset Cxy \dots z)$$

$$\forall x \forall y \dots \forall z (A_1 x \supset (B_1 xy \wedge \dots \wedge B_n xz \equiv Cxy \dots z))$$

のような形に記号化できる。従って、推論図の $\forall$ とり、 $\supset$ とり (modus ponens) と (しばり) 変項の改名とを主流とする変形操作を根気よく繰り返すことによって、質問に答えられる場合がかなり多い。われわれのシステムはこのようなアイディアに立つものである。

2.3 ゲーデル流の形式言語の採用

法律には事実問題のほかに権利問題がからんでくる。前者はいわゆる物理言語でも記述できようが、後者は一種の様相判断に関係するので、通常の物理言語では扱い切れない。様相論理として知られる体系も、通例の解釈では法的世界に適用できない。たとえば L5 の名で有名なルイスの体系には、「 $\Box p \supset p$ 」「 $p \supset \Diamond p$ 」が公理式として置かれ、「 p が必然的なら p だ」「現に p なら p は可能だ」のように解せられている。これを法的世界に流用するとき、「 $\Box$ 」を義務の様相、「 $\Diamond$ 」を権利的様相に引き当てるのが割合に自然な解釈と思われるが、納税の義務があ

っても脱税者が現に存在する事とか、現に殺人者がいるからとて殺人の権利があるとは言えない事とかからして、前記のような解釈のし直しはあまり役に立たない。一方また次のような事もある。特許法第三十九条4項には

…ときは、出願人の協議により定めた一の出願人のみが特許又は実用新案を受けることができる。協議が成立せず、又は協議をするこ
とができないときは、…

の文言があるが、下線を附した二つの「できる」の性格は異なる。前者が法的な可能態（こゝでは権利の是認）を言っているのに対し、後者は事実としての可能態を言っている。（体系L5の「◇」は後者を表わす道具だと解せられる。）このように、同じ可能態にしても法的なものと事実的なものとの使い分けが必要である以上、安易に既成の様相論理体系を借用するわけにはいかない。

こゝでわれわれが着目したのがゲーデルの集合論言語である。すなわち、適宜の様相常項を設け、それに関してある種の公理を課すれば、様相論理における推論と同等の事ができるはずである。これがゲーデル流の形式言語を採用する第一の理由である。

次に、NKのような一階述語論理を規準として考えるとき、法律文では一階の述語に翻訳されるものを変項的に扱った表現が多い。（§ 2.1の例で言えば、「 $\langle a, _K \rangle \in J$ 禁制」の「 K 」がそれであり、更に法適用の公理と仮称した式の「 i 」のような使い方がその典型である。）これらについては高階の述語論理に訴えればよいが、高階述語論理の操作は技法的に大変面倒である。この点をゲーデルの集合論言語は巧みに切り抜けている。これがゲーデル流の形式言語を採用する第二の理由であ

る。

ところで法律文の表現には、他の条文の参照や読み替え規定などに見られるとおり、法律文自体を対象とした表現つまりメタ表現が含まれる。対象表現とメタ表現とを同列に扱うと色々なパラドクスの生ずる事が知られているが、この危険を回避するためにもゲーデルの集合論言語は有効であり、特に、角括弧〔corner〕と呼ばれる対記号「」を導入することによって、この種の危険は完全に防止できると思われる。

以上の理由によりわれわれはこのシステムでのデータ記述用言語として、角括弧を導入したゲーデル言語を使う。

2.4 われわれの課題のために法令を選んだ理由

つけたりとなるが、§ 1の課題を実行するに当って法令を対象とした理由を述べておこう。一言で済ませば、法律文の検索が例として割合に扱いよいと思われたからである。しかもそこには、従来のQAシステムや自動証明の研究でほとんど手がついていない問題もあるからである。これらの事には§ 2.1 - 2.3で暗に言及したつもりであるが、ここにもう一度まとめて述べる。それは次のとおりである。

- (1) 原文の表現が（技術文献や新聞記事などに比して、格段）整っているので、日本語として扱いよい。
- (2) 建前として公理論化が可能なのはである。従って形式的操作の成功する見込みが大きい。
- (3) 条または項を単位に考えたとき、かなり限られた文型で済むと期待できる。
- (4) 様相的表現が本質的にきいてくる点、通例の一階述語論理では律し切れない点で、従来のこの類の研究とは異なる新たな問題を含む。

(5) 見なし規定・読み替え規定などメタ表現もデータにする必要がある点で、言語理論としても興味深い。

(6) 法令が量的にきわめて多くなっている現在の社会では、法律文の質問回答システムの開発にそれなりの価値があると思われる——要望の多い判例捜しにしても、かかるシステムを含まなければなるまいし、また法令間の整合性の検討などへの応用も考えられる。

最後に、特許法を選んだ理由は、この法律が割合に最近の立法にかかり内容も技術に関する事なので、比較的是っきりした扱いができようと考えたことである。

3. システムの概観

このシステムにおける処理の大筋は図1に示すような五段階を踏むものである。

第一段階：質問の確認——これはシステムに発せられた質問文を一文ずつ読む段階である。（大型機の利用を予定しないので、まとめては読まない。）読まれた質問文はアナライザに渡される。こゝで構文的チェック（と一部の意味論的チェックと）をし、作りのよい連系である場合にはその構文を反映する系図様のリストが添えられる。作りの悪い連系である場合には、悪い点を指摘してこの質問の処理を打切る。（当面は会話型処理を考えていない。）

なお質問文は、日本語によるのではなく、やはり形式言語によってつづる事を前提にしている。将来は日本語がそのまま使えるようにすべきであろう。また質問文連系のほかにリスト・ジェネレータにより系図仕立てにするのは、第四段階での処理に備えての事である。

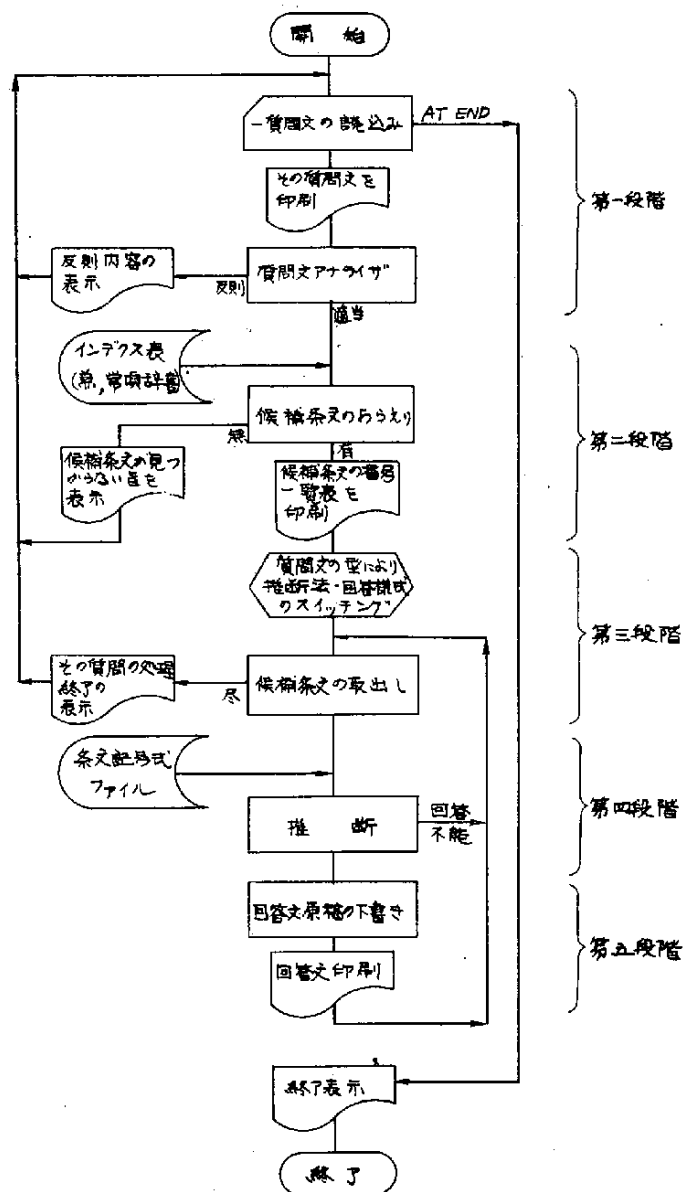


図1 システムの概要

第二段階：探索範囲の決定 — 純理論的にはこの段階は必須ではない。

しかしファイル中の条文をすべて見るのは能率的でないから、この段階をはさむ。ここではキーワード合せの手法で、回答が導けそうな候補条文（の集合）を決定する。質問文の大半の常項がキーとして使われる。ある部分連糸がそのような常項である場合には、アナライザが構文解析過程でその見分けができるように、常項連糸のシンタクスが決めてある。候補条文の決定にはインデクス表を利用する。インデクス表の各項目は（原則として）個々の常項に対応し、常項の形が見出しとなっており、それに構文論的情報、時として意味論的情報（この一部で類義常項や上位概念を指す常項の指定ができる）、更にその常項が現われている条文の番号一覧が添えてある。この条文番号一覧をデータとして候補条文のあらえりを行なう。従ってこのシステムでは文献検索も補助手段に使っていることになる。

第三段階：探索の指図 — 質問文の型に応じた処理法のスイッチングおよび未処理データ（候補条文）の管理をする。（参照条文へのつなぎなどもこの部分に受持たせる予定である。）

ついでながら質問文の型について述べておこう。先にT型とW型との区別を述べたが、質問文の前提列の部分が空列か否かで甲・乙の二種を分ち、従って都合四種の型に分けられる。型の違いによって第四段階・第五段階の処理手続が異なる。次の事に注意しておくのも、むだではあるまい、T乙型の「 $q \rightarrow Q$ 」は「前提 q の下で Q か」に対応するが、実際問題として q は条件の一部分しか提示しまい。（前提条件が枚挙できるほどの知識を有すれば、質問を出すまでもなくその事について十分に知っているであろう。）従ってシステムはこの型の質問を「指定した

前提 q に、もし必要なら追加条件 X を追加して、どんな X があれば適法に Q と言えるか」の意と解釈する。無論 X は命題変項であるから、 W 型質問とは区別される。 W 型は「いつ」とか「だれに」とかのように項の値を求める質問である。 W 型質問にあっても追加条件を補うための命題変項 X が附加される。 X の値の決め方については § 4 で例示する。さて T 乙型に対するシステムの解釈法と同様に、 T 甲型質問「 $? \rightarrow Q$ 」についてもシステムは「どんな条件 X を追加すれば、適法に Q と言えるか」の意と解釈する。 W 型でもこの方針に変わりはない。

第四段階：推断による処理 — この段階がこの研究の中心的な部分である。質問文と候補条文との構文を利用し、構文を写し取っている系図式リストの走査をしつつ、証明図の構成と同等の効果をあげる処理をする。その大方針は、適宜回の $\forall$ とり、 $\exists$ とりを繰返して Q か $\neg Q$ かが導けるか否かを調べる事である。その他の論理記号に関する操作も、手順化してプログラムに織り込んである。特に存在量化命題式の処理はメタ定理を参照して行なうようになっている。かような方法は heuristic ではないが、法律文には一定のパターンがあり、その適用法も概して限られた図式に当てはまるので、推断法の型を幾つか決めてそれに従うのが能率的だと思われる。

現在のところ質問文の q や Q の部分に許される構文にはかなりの制限があるが、それでも下記の程度の質問を記号化し得るほどの表現力は与えてある：

- どんな場合に特許出願の審査を請求することが許されているか
- 自分のした特許出願にみずから異議を申立てることができるか
- 飲食物に関する発明について特許が受けられるか

- ・未成年者でも特許手続を行なうことができるか
- ・二人以上による共同発明については、その中の一人が単独で特許出願をすることはできないか
- ・審査官の資格は何によって決められているか
- ・特許出願に対し異議がある場合、いつだれに異議申立をすることができるか

そしてまた、この程度の質問に答えられることが当面の目標でもある。なお条文記号式のレコードには、その条と関係の深い条への参照がつけてあり、その限りにおいて二条以上を使う推断をする道がついているが、今回はかゝる推断の実行は保留した。（参照条文の所在だけは回答文の中に入れる。）

第五段階：回答文作成 — この部分については特別の説明はいるまい。回答文の体裁は§ 4 に例示する。なおこのシステムはCOBOLでプログラミングしている。

4. 質問文処理例 — 推断法を中心に

このシステムがどのように働くかの大体の感じを理解してもらうために、T型・W型の質問各一例について推断法の部分を中心とした例解を示す。その前にシステムが扱う形式言語GLについても、簡単ながら述べておく必要を覚える。

4.1 システムが扱う言語GL

われわれのシステムが処理対象とする形式言語は、ゲーデル流の集合論言語族に属する一言語である。詳細は別にマニュアルを用意し、それに述べてある。ここでは後掲の例解に必要な範囲で述べよう。なお読み

易さを考えて、計算機にフィードする形そのままではない。(実は下記の形がそのまま使える方が便利であるが、今日の計算機の character set はそれほど豊かでないので、やむなく記法を変えた。そのためプログラムの負担が増す事も、ついでに言うておく。)

GLの末端語彙〔terminal vocabulary〕は、純理論的には、本来の述語を表わす記号「 ϵ 」, 変項記号, 縛り変項記号, 論理記号「 $\neg$ 」「 $\wedge$ 」「 $\vee$ 」「 $\supset$ 」「 $\forall$ 」「 $\exists$ 」および若干の補助記号から成る。更に定義によって末端語を増すことができる。しかしそうした文法論議よりも、ここで大切なのは次のような事である。

まず基本命題式として $\langle \textcircled{1} _ \textcircled{2} _ \dots _ \textcircled{n} \rangle \epsilon \alpha$ の形のものを使う。この「 α 」は常項と呼ばれるものであり、常項は特定集合(一般的には class)の名である。たとえば $\langle a _ h _ d \rangle \epsilon$ 特許願は、そうした基本命題式の一つである。これは日本語の表現の「事物 a に関し(法)人 h が時 d において(日本政府に)特許出願をする」が表わす内容に対応する各常項に対しその元である順序 n つ組 $\langle \tau_1 _ \tau_2 _ \dots _ \tau_n \rangle$ の n が指定されていて、たとえば常項「特許願」に対しては、有体物/無体物をさす項、人/法人/人の集まりをさす項、時をさす項の三つがこの順に並べられた順序三つ組でなければならない。こういう規定は常項辞書に書き込まれている。なお「 $\langle \tau \rangle$ 」と「 τ 」とを同一視する。従って「 $\langle \tau \rangle \epsilon \alpha$ 」は単に「 $\tau \epsilon d$ 」と書かれる。また「 $_$ 」は空白記号である。

さて、前述のように「 a に関し h が d において特許出願する」は「 $\langle a _ h _ d \rangle _$ 特許願」であるが、これを体言化した「 a に関し h が d において特許出願すること」は、「 $\langle \langle a _ h _ d \rangle _$ 特許願 $\rangle$ 」のように入子になった順序対で表わす。これは命題式ではなく、変項の一種である。

従ってこれ自体の真偽は問われない。真偽が問われるのは

$$\langle h \sqsubset \langle a \sqsubset h \sqsubset d \rangle \sqsubset \text{特許願} \rangle \in J \text{ 可能}$$

のような命題式になった場合である。この命題式は「 h にとって、 a に関し h が d において特許出願することが、法的に可能である」意を表わす。そしてまたこれと同じ内容を、ゲーデル集合論の取極めにのっとり、別の形により、

$$k = \langle \langle a \sqsubset h \sqsubset d \rangle \sqsubset \text{特許願} \rangle \Delta \langle h \sqsubset k \rangle \quad J \text{ 可能}$$

と書くこともできる。($\langle \tau_1 \sqsubset \tau_2 \sqsubset \dots \sqsubset \tau_m \sqsubset \tau_n \rangle$ は $\langle \tau_1 \sqsubset \langle \tau_2 \sqsubset \dots \sqsubset \langle \tau_m \sqsubset \tau_n \rangle \dots \rangle \rangle$ と同義だと定義されている。)

かような記法は、慣れないと大変むずかしいように感ぜられるが、われわれの実験結果ではそれほど難解な記法ではない。自然言語による表現の内容について十分な知識を有する者にとっては、原表現をパラフレーズするのと同じくらいの時間でGLによる記号化が行なえることが、確かめられている。

われわれのGLでは、表現自体を対象とする表現、つまりメタ表現について角括弧「 $\sqsubset$ 」を利用し、これを対象化することもできる。これは法律文のいわゆる読み替え規定などの記号化に有用である。後に角括弧を利用した例を示す。

この§ 4.1は端折った解説なので分りにくいかも知れないが、実際にはさほどむずかしいわざではない。その事は、§§ 4.2 - 4.3の実例を見れば明らかであろう。

4.2 T乙型質問処理の一例

下記の例は乙型に限らずT型質問に対する処理の一つの典型である。

自然言語で表わした質問内容：特許出願に対して異議申立ができるか

この質問をパラフレーズすれば「 a に関し h が時 d に特許出願するという法律行為を k で表わすとする。この k について人 i が異議を申立てることが法的に可能か」という事になろう。かかるパラフレージングの下に、この質問内容を GL で表わすわけであるが、異議申立に関する基本命題式の文型は（常項辞書により）

$\langle \textcircled{1} \textcircled{2} \textcircled{3} \textcircled{4} \textcircled{5} \rangle \in \text{異議申立}$ で「物事②に対する①を内容とする異議を③が④に時⑤において申立てる」意を表わすと約束してある。また法的権利の保証について

$\langle \textcircled{1} \textcircled{2} \textcircled{3} \rangle \in J$ 可能で「（法律上の）人①にとり、②を③として実現すること、つまり②に関する行為③が、法的に可能だ」の意を表わす約束もある。これらによつて、

記号化した質問文（の一つ）： $k = \langle \langle a \text{---} h \text{---} d \rangle \text{---} \text{特許願} \rangle \text{---} ? \rightarrow \text{---} \langle i \text{---} b \text{---} k \text{---} i \text{---} j \text{---} e \rangle \text{---} \text{異議申立} \rangle \in J$ 可能 と記号化される。

参考までに言えば、これと同じ内容を甲型にして

$? \rightarrow \text{---} \langle i \text{---} b \text{---} \langle \langle a \text{---} h \text{---} d \rangle \text{---} \text{特許願} \rangle \text{---} i \text{---} j \text{---} e \rangle \text{---} \text{異議申立} \rangle \in J$ 可能と表わすこともできるが、前者の形に翻訳した方が処理し易い。（なお cf. § 4.3）

さて上記の質問には特許法第五十五条1項

出願公告があったときは、何人も、その日から二月以内に、特許庁長官に特許異議の申立をすることができる。

で答えられるわけであるが、その

条文記号式： $\forall x \forall p \forall s \forall o \forall i \forall u \forall q \forall y ($

$\langle x \text{---} p \text{---} s \rangle \in \text{特許願} \wedge o = \langle x \text{---} p \text{---} s \rangle \text{---} \text{特許願} \wedge \langle o \text{---} i \rangle \in$

出願公告 $\wedge i \in \text{日}$

$\supset (t \leq u \wedge u \leq t+2 - \text{月} A (\text{本条 2 項 } [x]))$ 指示物 $\subset y$

$\equiv \langle q _ \langle y _ o _ q _ \text{特許庁長官} _ u _ \rangle _ \text{異議申立} \rangle \in J(\text{可能})$

但し計算機内での表示では、長さの節約と処理の簡便化とのため、存在量子が先立っていない前置普遍量子子は略記する。(上記で点線を下に引いた所が省かれることになる。)また上記で本条 2 項に言及した部分は、多様な質問に応ずることに備えて補った部分である。記法上の解説を追加すれば、

$\langle \textcircled{1} _ \textcircled{2} \rangle \in \text{出願公告} \dots \textcircled{1}$ の出願公告が時 $\textcircled{2}$ において行なわれる。

$\textcircled{1} \in \text{日} \dots \textcircled{1}$ は日単位で扱われる時だ。

$\alpha' \beta$ の形の表現には次のような定義がある；ここに $\exists!$ を $\mathcal{A}$ はであるクラス がただ一つだけ存在する意：

$(\exists! x (\langle x _ \alpha \rangle \in \beta)) \langle \alpha' \beta _ \beta \rangle \in \beta \wedge (\neg \exists! x (\langle x _ d \rangle \in \beta) \supset \alpha' \beta = \phi)$

さて質問文中の常項「特許願」「異議申立」「J可能」が候補文あらえりのためのキーとなる。五十五条 1 項はこれらをすべて用いているから、当然、候補条文となる。以下に、この条文に対する推断法の概要を述べる。

まず質問文の系図と条文の系図とは図 2、図 3 に示すとおりである。図 2 のアイウエオ、図 3 の数字は、それぞれ質問文連系、条文連系において現われた末端語の、その出現順に従って、それに対応するリストセルを示す符牒である。(原文のすべての末端語がリストセルになるわけではない。)ところで質問文は「 $q \rightarrow Q$ 」の形、条文の骨組は「 $\mathcal{A} \supset (\mathcal{B} \equiv \mathcal{C})$ 」の形をしている。しかも結果的には条文が「 $\mathcal{A} \wedge q' \wedge \mathcal{A}_2 \supset (\mathcal{B} \equiv Q)$ 」と解釈できるので、この事に基づいて回答するわけであるが、その事は次のようにして確かめられる。

最初にQ部と条文帰結部とが同じ基本命題式の型を有するか否かを調べる。これは二つの系図のリスト走査と連系合せとによって実行でき、しかもその間に条文が「 $\exists x(B \equiv C)$ 」の構文型である事も分る。Q部はセル「テ」に始まる部分系図〔テ〕であり、条文帰結部Cはセル「52」に始まる部分系図である。これら部分系図〔テ〕〔52〕は構造を同じくする。しかも質問側常項の所在セル「ツ」「ト」は条文側の「51」「53」がさす常項と同一である。また本来の述語も「テ」と「52」とに見るとおり一致する。そこで、末端連系の一致しない部分についての対応をつけ、次のような改名表を作ってみる。

質問側 i b k j e

 ↑ ↑ ↑ || ↑

条文側 q y o 長官 u

次に、条文の条件部である $\exists x, B$ (の系図)を走査して、質問q部(の系図)と整合するか否かを調べる。そのためには、部分系図〔1〕を条文側の部分系図〔5〕,〔9〕,〔19〕,〔22〕,〔25〕,〔30〕,〔40〕を順次突き合わせてゆけばよい。〔22〕までの所で〔1〕の否定と見合うものが現われればこの条では回答が出せない。〔1〕と見合うものがあれば、その部分は追加条件とならない。残りが回答文での追加条件Xとなる。この例では〔1〕と〔9〕とが次の対応関係を介して見合う：

質問側 k a h d

 ↑ ↑ ↑ ↑

条文側 o x p s

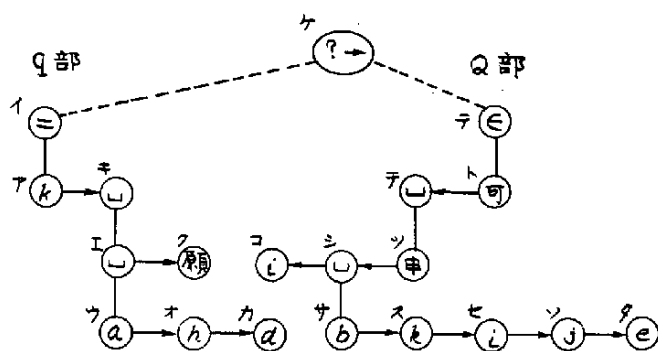


図 2 質問文の系図

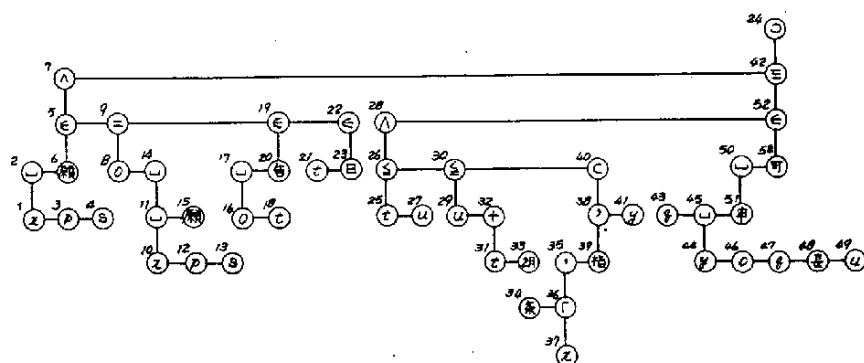


図 3 第五十五条 1 項の系図

しかもこの対応関係は先の改名表と整合する。それゆえ最終的な改名表は

質問側	<u>i</u>	<u>b</u>	<u>k</u>	<u>j</u>	<u>e</u>	<u>a</u>	<u>h</u>	<u>d</u>
改名表	↑	↑	↑		↑	↑	↑	↑
条文側	<u>q</u>	<u>y</u>	<u>o</u>	長官	<u>u</u>	<u>x</u>	<u>p</u>	<u>s</u>

となる。条文側の縛り変項「 q 」等を質問側の対応する変項「 i 」等にし書き改める推断は、推論図 $\forall$ とりによって保証されている。したがって、

五十五条1項, $j = \text{長官} \vdash \forall t (\mathcal{U}_1 \wedge q \wedge \mathcal{U}_2 t \supset (\mathcal{B}_t \equiv Q))$

が言える；こゝに「 $'$ 」をつけたのは条文の縛り変項を改名表に従って変項にした表現である。更にこの結果に基づいて

五十五条1項, $q, j = \text{長官}, \mathcal{U}_1 \wedge \exists t (\mathcal{U}_2 t \wedge \mathcal{B}_t) \vdash Q$

五十五条1項, $q, j = \text{長官}, \mathcal{U}_1 \wedge \exists t (\mathcal{U}_2 t \wedge \neg \mathcal{B}_t) \vdash Q$

が共に言える。しかもこの証明図をその都度構成する必要はない。なぜなら、条文の構文型と改名表の整合性から一般的に導ける事だからである。

結局、回答文は次のとおりである；なお「 t 」のような縛り変項は、特に断わってなければ普遍量化を受けていると約束する：

* 特許法 55:1

$\langle a, h, d \rangle \in \text{特許願}$ $\wedge$ $\langle k, t \rangle \in \text{出願公告}$ $\wedge t \in \mathbb{N}$ $\wedge j = \text{特許庁長官}$ として、
もし $t \leq e \leq t+2$ -月 $\wedge$ (本条2項「 x 」) 指示物 $c \in b$ なら J 可能、
そうでなければ J 可能でない。

参照 * 月数計算) 3:1, 3:2 * 申立手続) 55:2,

* 出願公告) 51:1, 51:3, 51:4 * 申立理由) 49

この参照の所は条文記号式ファイルのその条(項)につけてある情報を

4.3 W甲型質問処理の一例

記号化した質問文 (の一つ) : $\rightarrow \llcorner \langle h _ 審査官 \rangle ' 資格 _ ? \rangle \in$ 規定

ここに、常項辞書によって次の基本命題式の型が取極めてある：

$\langle ①, ②, ③ \rangle \in \text{資格} \cdots \cdots ①$ が、 $②$ が $③$ であるための資格の総体だ。

＜①_②＞E規定……①を法（の条項）②で規定する。

この質問には特許法第四十七条２項「審査官の資格は、政令で定める。」
によって答えられるが、その

条文記号式： $\langle o_1 _ p _ 審査官 \rangle \in 資格 \supset \exists o (\langle o_1 _ o \rangle \in 規定 \wedge o \in 政令)$

質問文・条文の系図はそれぞれ図4、図5のとおりである。

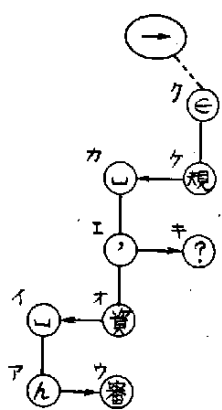


図4 質問文の系図

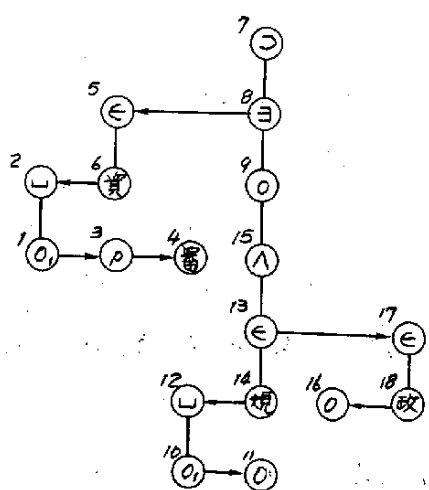


図5 第四十七条2項の系図

§ 4.2 の場合と同様に質問Q部を条文帰結部と突合せることになるが、この例では帰結部が連言でしかも存在量化式なので少々手がかかる。条文側系図をセル「7」からはいつて「15」まで（直系子孫をたどりつつ）調べることにより、上記の構文情報が得られる。そこでQの系図を順次部分系図〔13〕、〔15〕と突合わせる仕事に進む。

〔ク〕と〔13〕との系図の構造は完全には同じでないが、セル「エ」からの子孫がなければ同じである。そういう場合には≪深さの異なる系図の整合性≫検査ルーティンに入れる。この例では、条文側の「 o_1 」が質問側の「 $\langle h_審査官 \rangle$ ’資格」と整合的な対応をすればよい。その事は次のような≪処理≫ルーティンに調べさせる：帰結部で今のような問題が起こったとき、すなわち「 ξ 」が「 $\rho'\alpha$ 」に対応すると想定されたとき、もし条件部の方に「 $\langle \xi_p \rangle \leq \alpha$ 」または「 $\xi = \rho'\alpha$ 」を指定する表現があれば、想定した対応が正しいものとする。（そうでなければ依然この想定を未確認のままで処理を続けて最後に整合性を調べる。無論、矛盾が発見されれば処理を打切る。）

さて、部分系図〔17〕も突合せ範囲にある事を記憶した上で、まず〔ク〕と〔13〕との照合を始める。上述のような構造のくい違いが見出されるから、システムが新たな変項「 k_{01} 」を発生させ

質問側	$k_{01} = \langle h_審査官 \rangle$ ’資格	?
	↑	↓
条文側	o_1	o

という想定をする。そこで条文条件部を調べると、部分系図〔5〕の所で「 $\langle h_審査官 \rangle$ ’資格」と「 o_1 」との対応の想定が正しい事が分り、全体としての改名表は次のとおりになる：

		$k_{01} = \langle h_{\text{審査官}} \rangle$	資格	?
改名表	↑	↑		↓
	o_1	p		o

それゆえ「?」が対応する「 o 」についての規定のうち、条文で質問文との照応が取れずに残っているものを取り出す。それは部分系図〔17〕である。これが存在量化を受けた帰結部にあるので、回答文は次のようになる：

* 特許法 47:2
 " ? " を " o " として, $o \in$ 政令である o がある。
 参照 特許法施行令 12

胸壁面上における心電図電位分布の計算
機映画におけるカラーアニメーションの
開発と実用化

北海道大学応用電気研究所教授 吉本 千禎

“ “ 助手 高谷 邦夫

“ 大型計算センター 助手 中村 鎮雄

1. はじめに

心臓の活動の生理学的情報を非観血的に得る手段としての心電図は心臓病の診断において欠せないものである。標準に誘導とよばれる上下肢からの6波形と胸部からの6波形から心筋の機能変化を見出すことには多くの経験の累積が必要である。これは立体的に複雑な形態を有する心筋に発生する活動電位が位置情報の極端に欠如した形で、単に1本の時間軸上のグラフとして表現されることに起因する。そこで心電図の診断では単に12本の心電図を形態的に比較するか、又は異った位置から眺めたことを意味する12本の心電図を頼りに電気双極子ベクトルの方向、大きさを想定して頭の中で心筋上に広がる活動電位の分布を組み立てるかの方法がとられる。しかし標準に誘導からベクトルの動きを正確に把握すること自体が容易ではなく、たとえベクトル心電図によってベクトルの軌跡が与えられても点から面上に広がる活動電位の変化を再現して考えるのには相当の経験が必要である。

心電図診断を形態的な比較分類に頼る方法は集団検診などを主な目的

とする電子計算機を利用した自動診断装置として発達し実用化の段階に至っている。しかし心臓病患者の精密な診断の際に行なわれる心筋上の活動電位を推定する思考過程を容易にする技術的アプローチは未発達である。

心電図の胸壁面上の電位分布は心筋の活動電位の投影であるから、胸壁面上の電位分布を直接観測する手段があれば、診断の大きな部分をしめる立体的心臓のイメージに重なる心電図の把握という作業は非常に容易になる。胸壁面上の電位分布は気圧配置図と同じように等圧線で表現するのが一般的であり理解も容易であるので、等電位曲線を胸壁上に描き心筋の部分的な現象も表現しようというのがこの研究の着眼点である。

天候の変化もその日一日の気圧配置図では予測し難く、何枚かの時間の異なる図を見る必要があるように、心電図の診断にも電位分布図を1心拍にわたって観測する必要がある。特に心電図の場合心臓が収縮する時期において急激な電気現象の変化を生じるので、その変化に追従する電位分布図の枚数もほう大なものになる。この大量の画面を能率的に眺める方法が講じられなければやはり診断に対する貢献は少ない。この弱点を補う手段としてアニメーションが考えられ、約1秒位の1心拍を数分位に引伸して映画化してやると位置情報、時間情報ともに含まれた総合的な心電図情報を得ることができる。

ここで心電図の波形に立返ってみると、診断の際に注意してみる正電位か負電位かということに等面位線図が直感的に訴える手だてがないのに気がつく。この点を解決する方法として効果的なのは色分けすることであり、カラーディスプレイによって、アニメーションの場合に考える基準を与える零電位線の表現も可能になる。比較的容易にできるフィル

ター同期型の順次式カラーディスプレイによって心電図電位分布のカラーアニメーションを可能にしたのである。

2. 胸壁面上の心電図電位分布処理システム

胸壁面上の心電図電位分布に関する処理システムは現在大きく分けて三つの独立した機能を有する部分から成り立っている。それは

- 1) 心電図の計測および前処理
- 2) 測定点間の補間および等電位線図の作図
- 3) 等電位線図のカラー表示

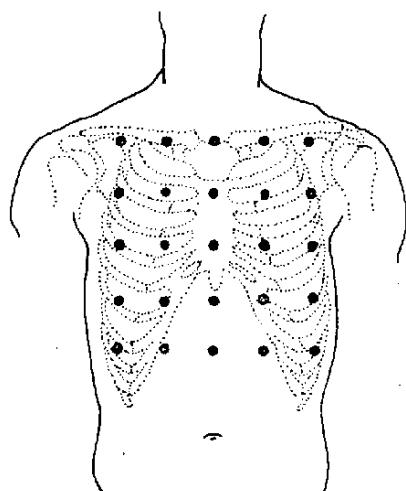
の各部分である。これらの処理はアナログデータを取り扱ってかつ図形表示用のグラフィックディスプレイを備えた医学用電子計算機 TOSBAC-3000/M によって行っている。

2.1 心電図の計測および前処理

心電図などのアナログ情報を取り扱う場合には原始データを得るのにハードウェア、ソフトウェア両面の特殊な手続きが必要なのは言うまでもないが、実際に標準的な数値処理に至る前に前処理と呼ばれる雑音除去、キャリブレーション等の操作が必要である。

等電位線図を作図するためのデータは第1図に「・」で示した電極位置から得られる心電図である。鎖骨下窩を一边とする正方形上にとられた25点の電極位置からの心電図は胸壁面上の電位分布を乱さないために一点ずつ順番に第II誘導とよばれる手足からの心電図を時間基準にして測定する。通常10心拍位を雑音除去のために加算平均してデータとしている。このデータは次に零電位に基線を合わせ、基準電位によって

キャリブレーションを行ない
さらにアニメーションを行なっ
たときに各コマの間が滑らか
に変化するように加重移動平
均をかける。次に測定点の順
に並んでいるデータを、一定
時間間隔内の25測定点のデ
ータから成る1レコードに作
りかえる分類作業を行なう。



第1図 胸壁面上の電極位置

心電図を25の測定点から
とるために要する時間は約

20分位であり、患者には大きな負担はかからない。

2.2 測定点間の補間および等電位線図の作図

等電位線図を描くためには胸壁面上の任意の点における電位が与えられなくてはならないが、測定値として既知の電位は25点しかないので、補間計算を行って測定点間の電位を知る必要がある。局所的な電位分布も忠実に再現する必要を満たすために、実際の値と近似値の偏差の絶対値が大きくなならないチェビシェフ近似を用いて補間した。

この補間計算は等電位線図の作図と一体になって1つのプログラムを構成している。胸壁面上の縦方向、横方向とも両端の電極位置間の距離を80等分し、 81×81 点の補間値を計算させた。今作図するものは、あらかじめ与えられた電位の等電位線であるから、補間計算で得たある場所の電位がその電位の近傍にあるかどうかを調べてある巾の中に入れば画面上でその点を光らせばよいわけである。このシステムでは正負そ

れそれに10本までの等電位線を求めることができるようになっている。画像の点数は約3000点まで許され、カラー表示のために正の等電位線と負の等電位線は分けられている。零電位は正負に重複して入っている。作図された胸壁面上の等電位線図は磁気テープに格納され、一連の画像の作図が完了した段階で、カラー表示用のプログラムに渡される。

2.3 等電位線図のカラー表示

この第3のプログラムによって64枚1組の形で磁気テープに記録されている画像を順次表示していくが、正の電位か負の電位かの区別がないと波形の逆転などの情報を見落したり、零電位の移動により心筋の興奮の部分的なおくれなどを適確につかめないで、正、零、負の電位をそれぞれ赤、白、青というように色分けする表現法を行っている。この原理は、赤、緑、青の三原色のフィルターをはめ込んだ円板を高速で回転し、赤のフィルターがディスプレイ前面に来たときに赤のパターンを表示するという具合に、フィルターとパターンの切替の同期をとる順次方式である。この場合には原色をそのまま利用しているが2色混合により、識別可能な18色の色を出すことができる。グラフィックディスプレイの表示管は赤、緑、青の三原色のスペクトラムを平均して出せるブラウン管を使用した。等電位線図の下に時間の基準として用いた第Ⅱ誘導の心電図を緑で表示し、それに現在の図がどの時刻のものであるかを示すインデキータを付けて理解しやすくしてある。

3. 結果および今後の問題

この研究の最終的な目標である電位分布心電図の臨床応用という見地

から眺めると、現在の2次記憶装置である磁気テープの速度、大きさの制限により計算機映画というよりむしろ計算機スライドの程度までしか到達できなかったために、充分満足いくものにはなっていない。約500枚の画像を記憶するためには1Mバイトの容量が必要であり1秒間に10枚の画像を表示するとしてもバイト単位で40KHzの転送速度が必要とされる。そこで、表示された画像を一度16mm映画にとり、それによって分布心電図の臨床応用の可能性を検討した。第2図に正常な人の



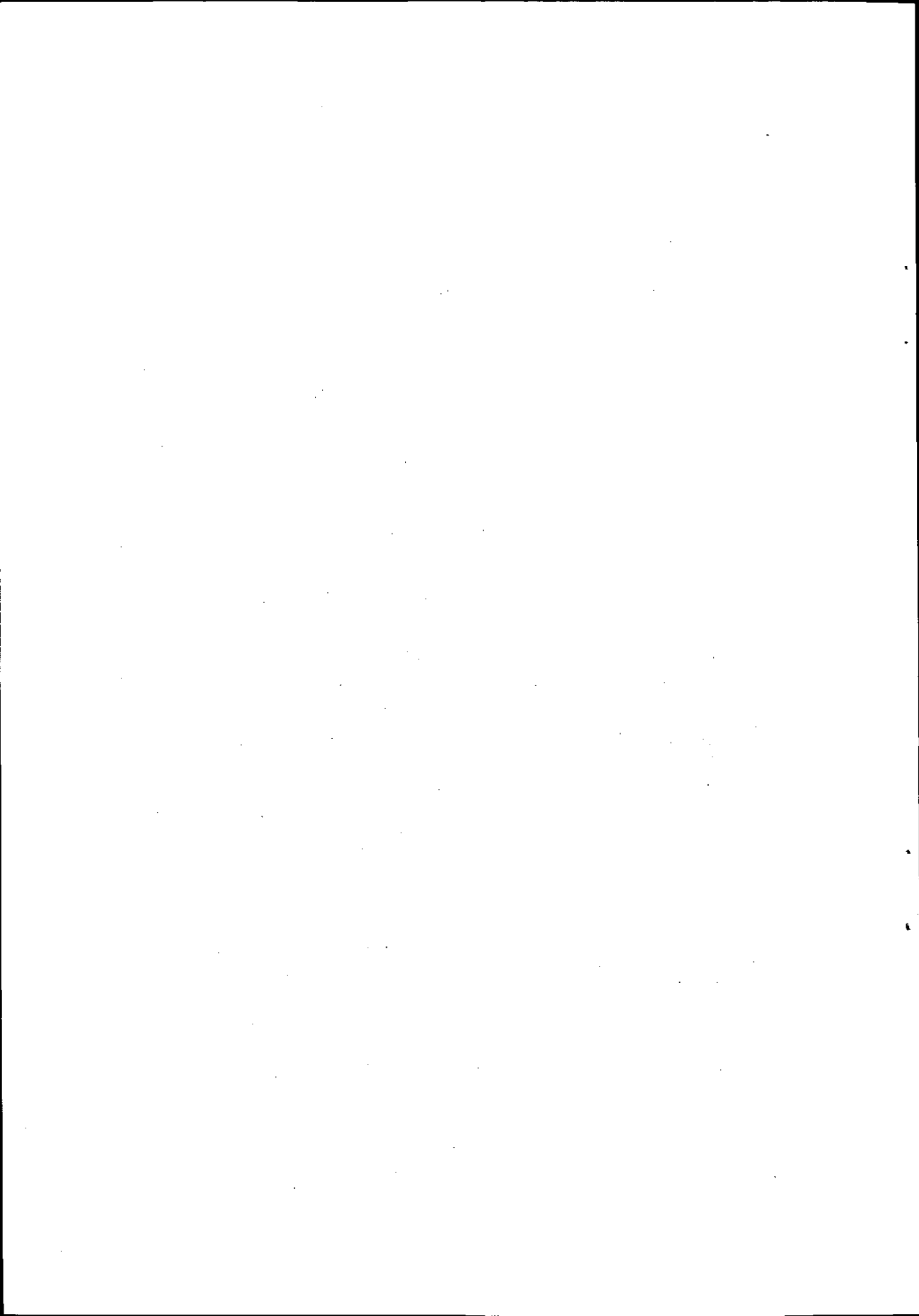
第2図 正常な電位分布心電図

分布心電図を示した。等電位線はP波に対して0.05mVおき、QRS群に対しては0.5mV おき、T波に対しては0.2mV おきにとってある。零電位は正常な場合には全体をとおして、左上から右下へ分布し、大巾な変化は見られない。P波に関しては振巾が小さく不明な点が多いがこの方法によりはじめて興味深いパターンの変化を観察できた。QRS群では急激なパターンが変化であるが、その過程に含まれる局所的な現象も明瞭に把握できる。零電位は若干時計方向に回転し復元する。拡張期に対応するT波のゆっくりした過程は静かな池に広がる波紋のようなきれいな電位分布を見せてくれる。

今後の課題として残されていることはフィルムに依らない映画化であるが、高密度・高速のディスク装置などにより解決できる予定であり、電子計算機映画により電位分布心電図を見て心疾患の診断を下すための診断体系の整理の方がむしろ大きな問題といえよう。

文 献

- 1) 吉本千禎：心電図の等電位線図的表現法およびその応用，日新医学 41. 11, 580-585 (1954)
- 2) YOSHIMOTO, C. TAKAYA, K. et al: Computer Animation of ECG Field Distributions, 8th ICMBE, 16-3, (1969)
- 3) 吉本千禎，高谷邦夫：体表面電位分布心電図のカラーアニメーション——T-3000/M による——，医用電子と生体工学，18, (1) 94-98, (1970)



染色体のオンライン解析システム

日本アイ・ビー・エム(株)主任研究員 飯坂 譲二

東京大学医学部助教授 小泉 明

東京大学医学部附属病院企画室 開原 成允

1. はじめに

人間や生物の固有性、遺伝に密接な関係のある染色体の研究は色々な方面で積極的に行なわれている。医用における形態学的な染色体解析の研究の意義として、臨床的には現在先天性疾患に伴う疾病の診断、判定に非常に有力な武器とされている点がある。(図1参照)例えば、染色体解析の結果G-21に属するものが対にならず、3個存在する場合にダウン氏症候群であることの判定がなされる。臨床的な判定のために染色体の解析が行なわれる以外に、最近では、染色体が細胞分裂に密接に関係していることから腫瘍のような分裂の激しい細胞にあつて、癌性、非癌性の判定に利用しようという試みもあり、染色体の異常観察によって癌の早期発見への手がかりとする研究もある。また、現在の様に人間をとりまく環境の変化が激しく、これまでの人間に与えられていた環境とは別の急速に変化している環境問題を研究する上での重要な研究手段となっている。薬物、放射線、公害等の人間の遺伝的影響、ビールス等の染色体への影響、染色体異常による自然流産の発生等、染色体解析の研究は重要な意義をもっている。(図3)

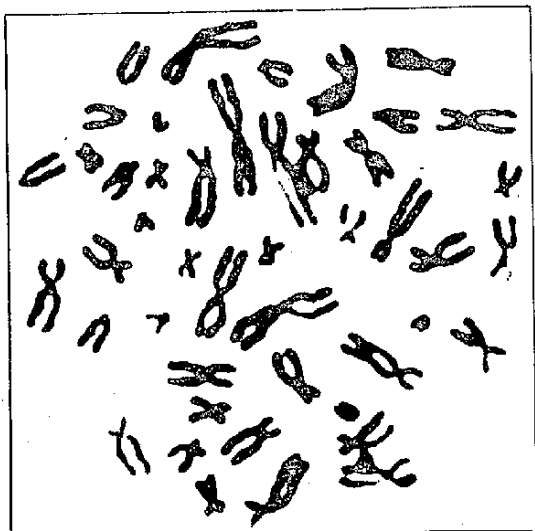


図1 染色体の顕微鏡写真

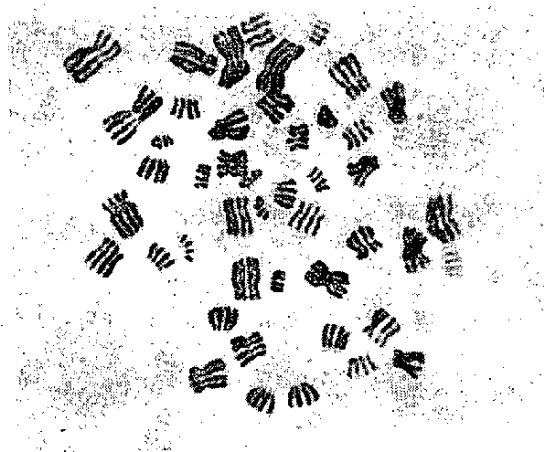


図3 放射線による染色体の異常

図 2 染色異常と臨床例

細胞遺伝学的な判定	典型的な臨床上的状態
46XX	女性
46XY	男性
G-トリソミー	ダウン氏症候群
18トリソミー	E・症候群(多発性先天性奇形)
13トリソミー	Di 症候群(多発性先天性奇形)
XXX	三重X症候群
XXY	クラインフェルター氏症候群
XYY	XYY 症候群
XO	ターナ氏症候群
D-D 転座	D ₁ -症候群, D-D 転座保有
D-G 転座	ダウン氏症候群(遺伝性, 自発性)
	D-G 転座保有
G-G 転座	G-G ダウン氏症候群(遺伝性, 自発性)
	G-G 転座保有
他の転座	多くの多発性奇形, 精薄に関連(?)
B の短腕部分欠損	クリ・ド・ジャ氏症候群
18 の短腕部分欠損	先天性奇形, 精薄の報告あり
18 の長腕部分欠損	
D の短腕部分欠損	
D の長腕部分欠損	網膜芽腫, 先天性奇形
G の短腕部分欠損	(正常ともいわれる)
G の長腕部分欠損	先天性奇形, 膿性, 脊髄白血病
Y の長腕部分欠損	(不妊症の症例あり)
他の欠損	(含まれている染色体による)
等方法染色体	
環状染色体	
逆位	親の保有状態が多発性流産, 先天性奇形に関連
モザイク	上記の組合せ

2. コンピュータによる染色体の解析

コンピュータによる染色体の自動解析で広く行なわれているプロセスは次のようなものである。(図4参照)

- (1) 分裂中期の標本の顕微鏡写真を撮る。又は顕微鏡に直接TVカメラを設置し、イメージを電気信号に変える。
- (2) 写真にとられた染色体のコンピュータ処理の可能なデジタル量に変換し、デジタルイメージを得る。

写真を読みとるために比較的広く利用されている装置には、フライング・スポット方式のディジタイザーがある。コンピュータから指示されたCRT(ブラウン管)上の位置X, Yにスポットを出し、フィルムを通過した光量をPM(光電倍增管)で受けてフィルム上X, Yに対応する点の濃度で測定する。フィルムの読みとり精度は、一画面 $256 \times 256 \sim 1024 \times 1024$ 程度が広く利用されており、濃度も8~256段階にわたっている。この装置は未露光フィルムを装填すればコンピュータによる処理結果を画像としてフィルムに撮影できる。

- (3) 顕微鏡下の染色体像は図1のようなものであるがそれもデジタルイメージをもとに1つ1つの染色体に分離するのが次の過程である。対象の分離は他の医用情報に比して比較的容易ではある。図6のように画面全体のレベル分布から明らかなように染色体の濃度と他の部分の濃度が異なり、バックグラウンドの補正によって容易に黒、白のみで中間調を含まない二値画像が得られ、輪郭抽出のコンピュータ処理を行なわせ染色体の形を浮かび上がらせることができる。染色体の1つ1つを分離するには、コンピュータで行なわせるのにきわめて困難な

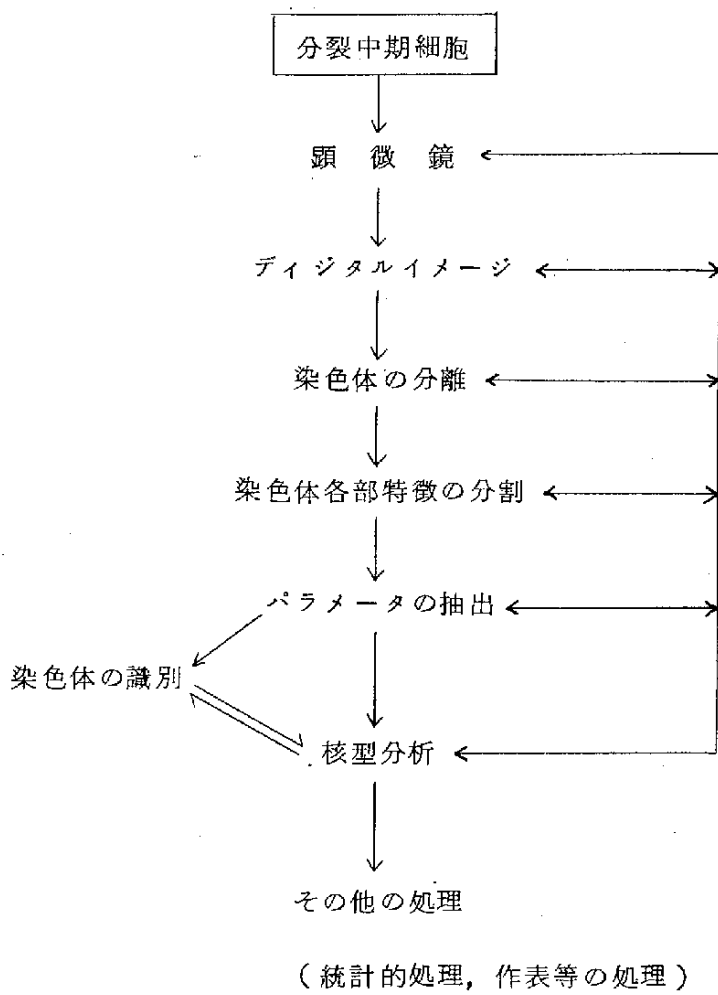


図4 コンピュータに依る自動解析の過程

面がある。それは標本を作成する上で押しつぶし法というものがよく使用されているが、互いに個々の染色体が分離している像を必ずしも得られるとは限らない。

たとえば重なりあい、あるいは腕の折れまがり等の像が混っている標本をえられる場合の方が多い。従って1つには染色体の個々を分離処理をコンピュータで行なわせるためには多くの処理が必要である。

(4) 染色体各部特徴の分離とパラメタの抽出

1つ1つの染色体のデータが分離されれば1つ1つの染色体の各特徴についての測定がおこなわれる。たとえば腕の部分の長さあるいは着視点の測定等がおこなわれ、これを染色体像の特徴づけようとしてデータをうけいれる。

個々の染色体の測定としては

1) ディメンション

(1) 全 長

(2) 面 積

(3) 周囲の長さ

(4) 各 arm の長さ

(5) Arm 比 (短腕と長腕の長さの比)

(6) セントロメリック・インデックス (短腕の長さの全長に対する百分比)

2) 濃 度

(1) 全光学濃度

(2) 光学濃度分布

等が主に測定される。

(5) 染色体の識別

(4)の染色体の特徴の分離と密接に関連した染色体の識別がある。染色体には大小、種々の形があり、常染色、性染色体等の識別が必要である。

(6) 核型分析

染色体はその特徴によって、A～Gの22対の常染色体と性染色体X, Yに分類される。

染色体はデーパー規約によって定量的に表現するために次の表現によって扱われる。

(a) 常染色体およびX染色体の長さの総和で各染色体の長さを割ったものを1000倍する。(図5の横軸, 但し%で表わしている)

(b) 長腕と短腕の比(図5の縦軸)

(c) 着糸の長さ, 染色体のセントロメアの位置を相対値で表わす。

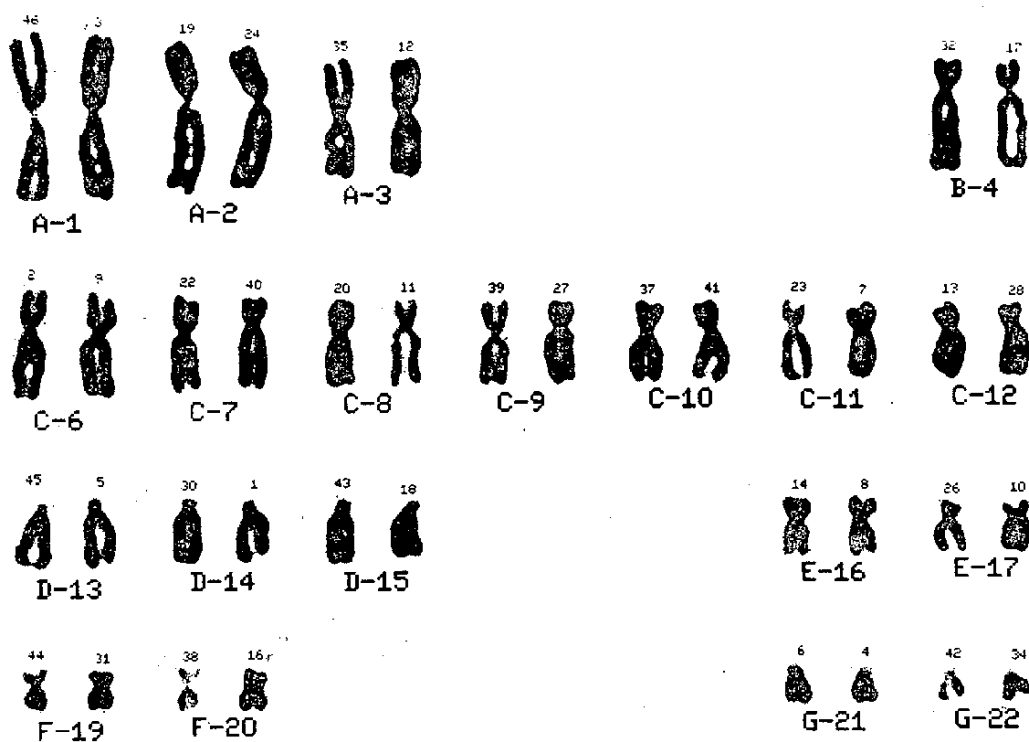
短腕の長さ/染色体の全長

このようなデータを求めた上で各染色体の分類が行なわれる。分類は正常染色体の分布の確率密度の数学モデルをもとに分類を行なう方法がある。すなわち、腕長Lと腕比CLの平面でi番目の正常染色体の分布の確率密度関数を

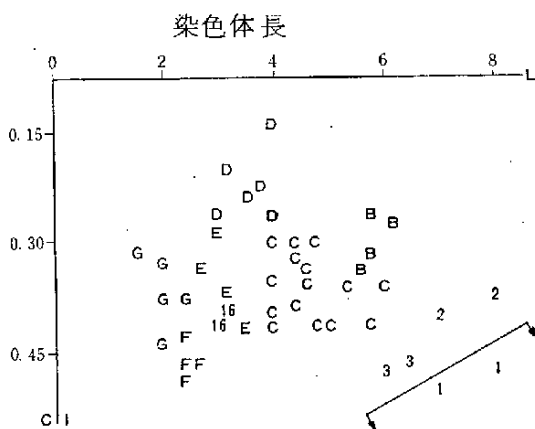
$$P_i(L, CL) = \frac{1}{2\pi} \frac{1}{\sqrt{SL_i, SCL_i}} \times \exp \left\{ -\frac{1}{2} \left\{ \frac{(L - \langle Li \rangle)^2}{SL_i} + \frac{(CL - \langle CL_i \rangle)^2}{SCL_i} \right\} \right\}$$

但し SL_i, SCL_i : i番目の分散

$\langle Li \rangle : \langle CL_i \rangle$: 期待値



(a) 核型分析の写真



(b) 分類の基準

図 5 核 型 分 析

に対し実測した染色体（腕長 L ，腕比 CL ）が i 番目の正常染色体である確率 $P_i(C', CL')$ を最大にする i を決定する。このような計算はコンピュータにとっては容易に行なえるものである。

このような核型分析の結果は図 5-a のような写真の型式でアウトプットされる。

3. 染色体のオンライン解析

従来の人手による染色体の解析方法によると 1 日当り 5 セル程度の処理能力である。前述したように染色体解析の要求度が高まるにつれその自動化をはかる必要がある。一方これまでこの種の研究には染色体を一旦、写真に撮影した上でその写真からディジタイザーで得られたデータをもとにして解析を行なう方法が多く採用されていた。しかし一般には写真のみをベースに、後の処理を全面的にソフトウェアの問題に転化するには問題が多い。図 4 のように、場合によってはオリジナルの染色体のイメージに立ちもどる必要もある。すなわち、オンライン解析システムによって

- ① 写真撮影のステップを染色体解析からのぞく、又これによって撮影系の雑音が入る機会をなくす。
- ② 自動化と実用化の 2 点をのぞむための低価格化
- ③ 従来の写真が人間視覚に適した染色を示す染料を用いているが、更に染色体解析に適した染色法の有無の検討
- ④ 波長の異なった光源の利用や光学系による位相差像等、同一対象の多種の像の利用によってコンピュータ処理を容易にする。
- ⑤ 閉じた系の採用によって多くの標本のうち、細胞分裂中期の固定

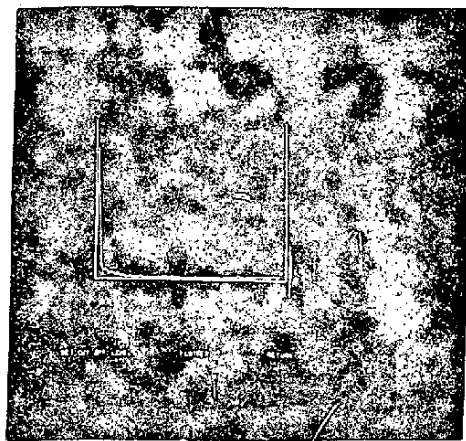


図6 染色体像の濃度レベルの分布

とその選択が可能となる。

以上のようなことが期待できる。

そこで、現在 I T V カメラを利用した顕微鏡のオンライン化とその処理システムを研究している。

A . ハードウェア

システムの構成は、図 7 に示してある。一般に I T V カメラは、濃度に対する解像度はよくないが、染色体像は図 6 に示すように二値的な像に近いのでビデオコンヤプランビュン等の撮影管を利用できる。顕微鏡像の走査システムとして光源をスポットにして走査する方式もあるが、染色体は細胞の核内に存在しているので細胞よりはるかに小さく顕微鏡の倍率も大きなものが必要であり従って光源も強力なものが必要である。スポット方式ではこの点光量が不足するので、I T V を採用する。視野の精度は通常のテレビ操作方式を利用する。すなわち、染色体の場合、一画面の精度を通常のテレビ方式の 525 本で十分と考えられる。一般に、T V のスキャニングの走査のスピードとコンピュータの入力の転送速度との間にずれがあるので、走査制御回路によってそのコンピュータが指令した位置のみ出力をサンプリングする方式をとる。サンプルされた信号は増幅器によって増幅され、A D 変換されたあとコンピュータ・インターフェイスを通してシステムに入る。使用しているシステムは 32 K 語であるので、一度には記憶装置に入らないためコアがいっぱいになったところへ磁気ディスクへ書き出す。1130 のデータ・チャネルを利用して、外部同期によってコアの中にサンプルされたデータはよみこまれる。コアの中のテーブルがいっぱいになったとき、自動的にテーブルを切り換えるテーブルス

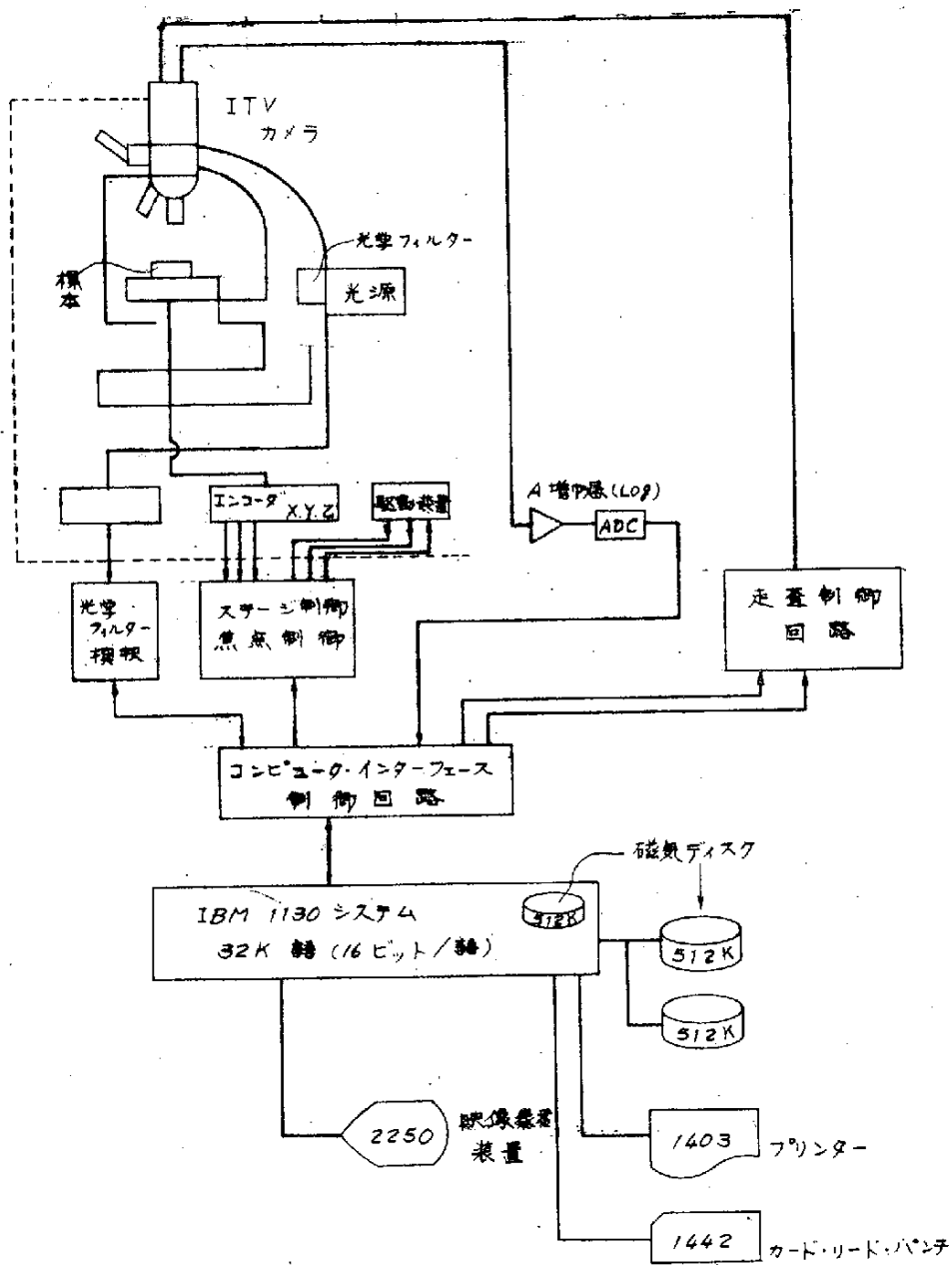


図7 染色体のオンライン解析システム

イッチングをコンピュータ・インターフェイスで行なわせる。いったんディスクに入った一画面のデータは磁気ディスクにストアされ、のちの解析処理に引き渡される。解析処理の結果、必要に応じてさらに他の像あるいは特別な像の拡大像を得たい場合にはコンピュータからの指示によって倍率レンズを交換し、ステージを制御して、必要なデータを取りこむ方式にする。よみこまれたデータはX、Yの二次元的なデータであるが必要に応じてZ方向の制御を稼働にする。これは重なりあった染色体の分離等への効果があることを期待している。光源は現在のところ連続的に波長をかえる方式ではなく、光学フィルターの選択によって必要な入射光の色を選択して現在使われているフィルターをコンピュータはディジタル量で読み取っている。コンピュータの指定した位置X、Yのデータを取りこむためにはXレジスター及びYレジスターにある座標を各々DA変換し、同期信号発生数を計数したものととの比較をとってサンプリングを行なう方式が簡単なものであろう。

B. ソフトウェア

一般に染色体像が比較的情報処理の問題として取り上げられたのは像そのものが比較的黑あるいは白の二値的なイメージに近いということがある。これは図のように一般的濃度レベルのヒストグラムをとるとわかるが二値的であるといっても、バックグラウンドにはいくらかのドリフト等のノイズがある。従って得られた像の全体について何らかの補正をほどこす。つまり前置処理を行なって、その結果1つ1つの鮮明化された染色体像とともに解析をおこなう。まず必要な前置処理というのは、その像によって明確に規定されるものであるが、理論的にははっきりしていない面があるのでその前置処理の種類及びパナメ

タについては選択的に行なう必要がある。このような為には前置処理が行ないやすいということが必要条件である。従って対話形式のシステムによって画像処理のテクニックがコンピュータプログラムとして登録しておき、それを自由にユーザーが取り出すことができること。

1つ1つの染色体の像を解析するわけであるが、それを解析するにあたって必要な測定量というものは色々なものがあり、その必要に応じて測定結果をファイルする機能及び測定されたデータの整理というものは長腕比のようなもの以外のデータの整理の仕方の発見的な意味については対話形式が良い。特に重なり合った染色体の自動的な分離は、いくつかの試みはあるが困難であり、人間の弁別能力が優れているので人間の介入が便利である。このような事情から全自動染色体解析システムに至る前に対話型式の染色体解析システムが有効と考え現在研究中である。

このシステムは次のような目的と、それを果たすための機能を保持している(図8)。

1) 染色体の処理について

- デジタルイメージのファイリング・バイナリーパターンへの前処理
- 染色体の輪郭の抽出
- 染色体の整列、分類
- オーバーラップした染色体のライトペンによる分離
- カリオタイプ解析と統計

2) マン・マシン・コミュニケーションとして

- 処理画面のモニター表示

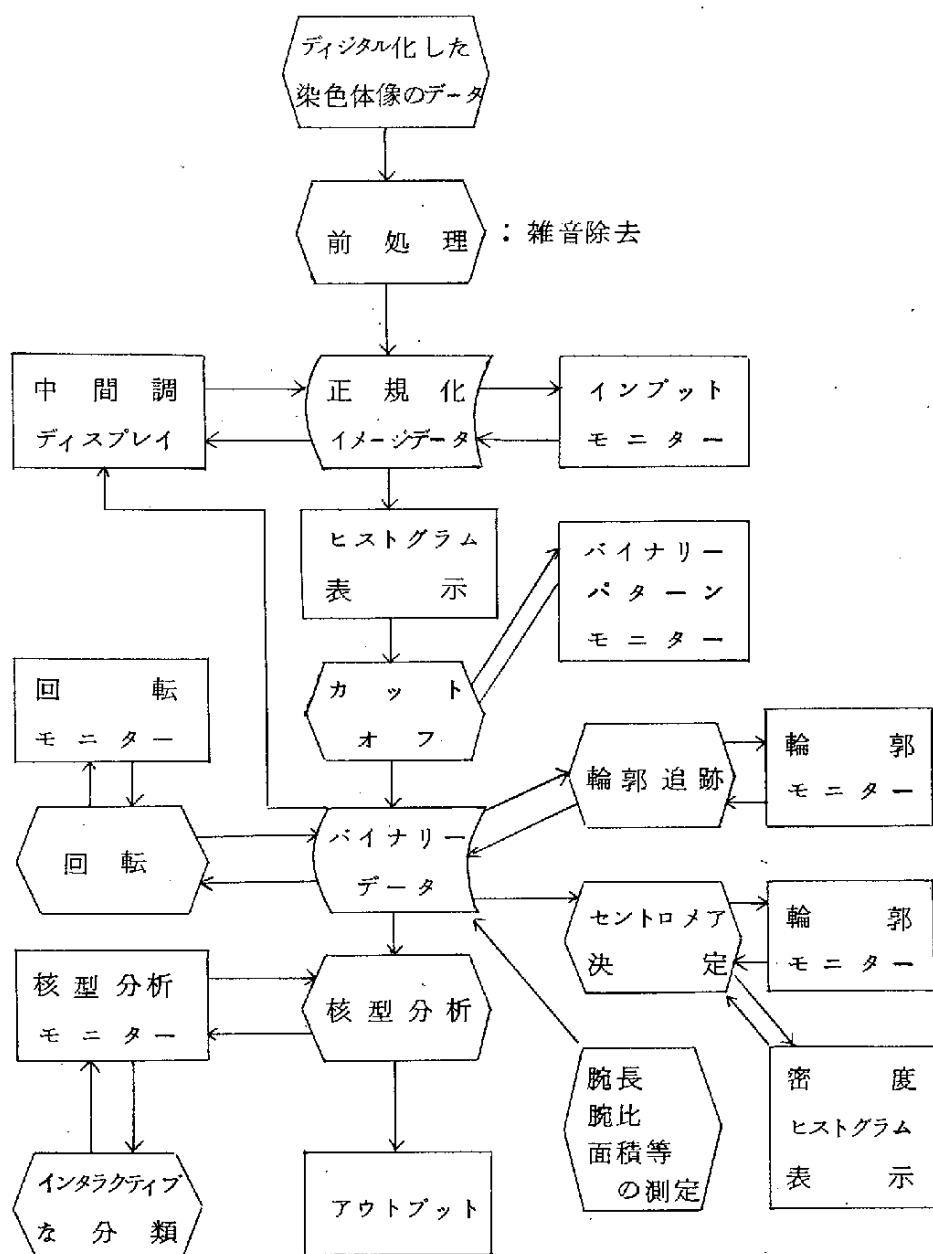


図8 染色体解析のソフトウェアシステム

○ 処理結果のグラフィック・モニター・ライトペンによる処理の選択 (図 9)

○ タイプライターからのコマンド入力

3) オペレーション・コントロールについて

○ 映像表示装置を利用した染色体解析の全過程のコントロール

○ 染色体解析の追加, 修正

○ 染色体処理言語 (コマンドの追加修正, イメージ・データのフェイリング)

○ 染色体の整列分類におけるライトペンによる修正

○ カリオタイピングのライトペンによる再分類

処理の流れ及びシステム構成は図 8 に示してある。画面全体に拡がっている染色体は 1 つ 1 つ分離され、磁気ディスク上にファイルされる。勝手な方向に並んでいる染色体は最少面積の外接四角形の方角によって整理する。この方向を定めるのに固有楕円の主軸によって求める例もある。

4. おわりに

以上簡単に現在研究中のオンライン、染色体解析システムについて紹介した。染色体の形態学的な分類のみで遺伝的な問題が解決するわけではなく、しかも、その形態学自体が染色性、あるいは吸光性、あるいは立体的な情報、その他総合的な見地から自動的な分離がなされなければならないし、まだ多くの未知の問題が残っている。このようなシステムが最終的には実用化というものをめざしてはいるものの、その前の、段階として我々のシステムが有効に働くことを期待している。

さらに、その染色体の解析をすすめる上で色々な情報を総合的にインテグレートするシステムというものが必要となるであろう。その1つの例として図10にえがいてみた。さらに染色体の解析の有効性が普及するにあたっては標本の収集方法、あるいはその解析結果のへんそう方法、あるいは解析するセンターの設置等、多くの問題があるがこの我々の研究がこの方面への足がかりとなれば幸である。

情報処理の立場からは画像情報の処理という問題は今後多くのコンピュータ利用の上で重要な問題であり、かつ染色体像の解析を通じて貢献することもできることを期待している。

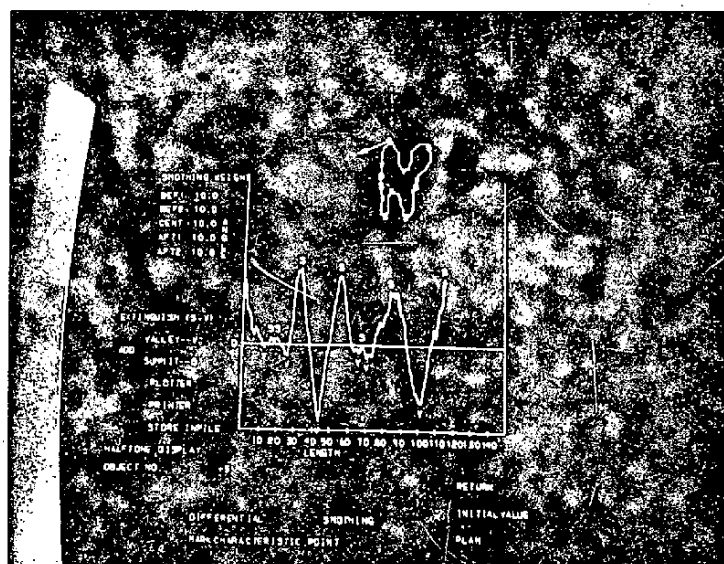


図9 マン・マシン方式による染色体の識別

コンピュータの判定した染色体像とその特徴点を確認修正する。

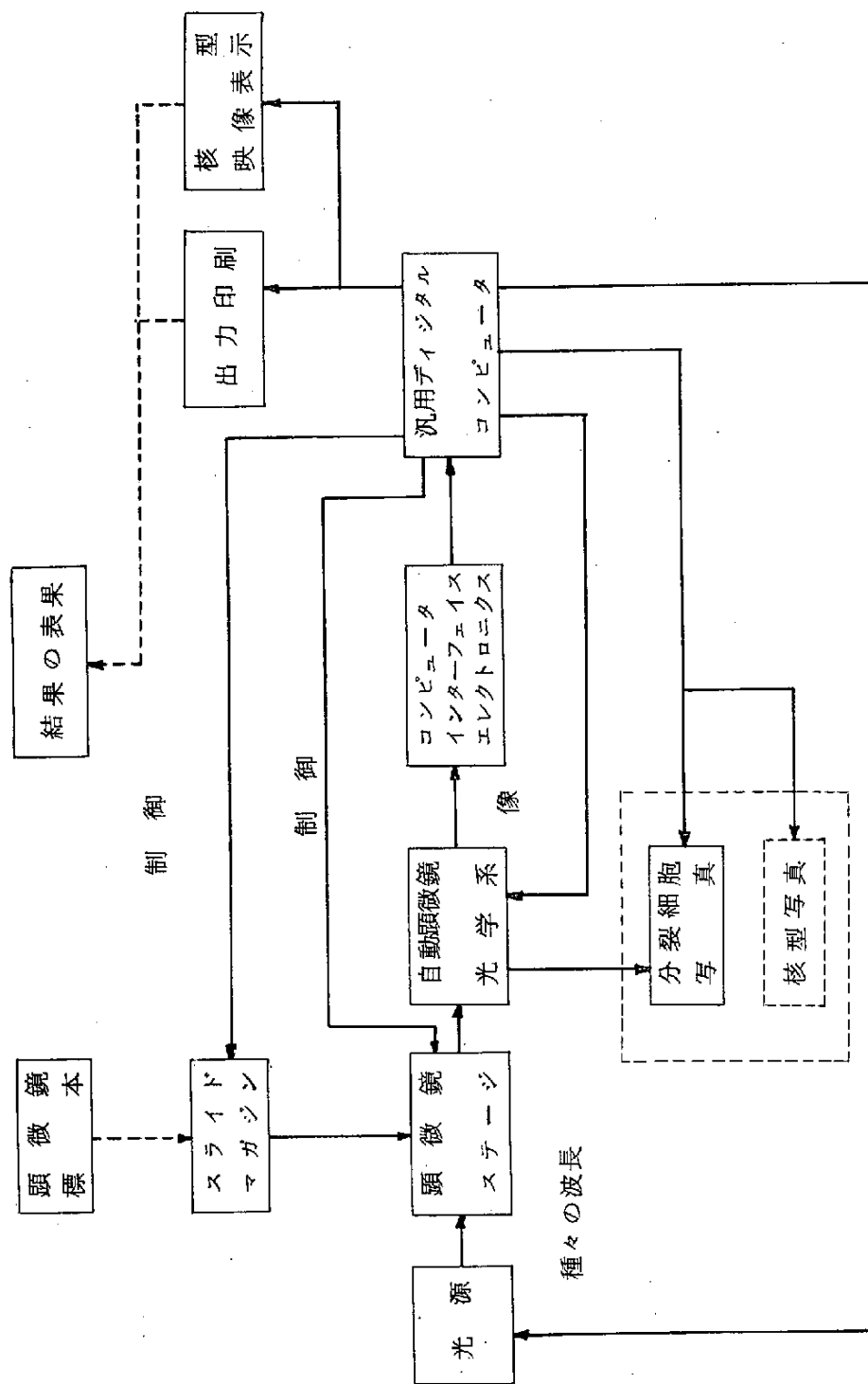


図 10 総合的な染色体の解析

記号処理言語の開発とその応用

九州大学理学部 教授	藤野 精一
佐賀大学教育学部助教授	飯盛 末夫
九州大学理学部 講師	有川 節夫
鹿児島大学理学部 助手	富樫 昭
九州大学理学部 助手	新島 耕一
九州大学大型計算機 センター助手	武藤 直彦
九州大学理学部 助手	棚次 奎介

1. 記号処理

記号処理の代表的な例に代数式の微分という問題がある。この例を通じて、最初に、記号処理の概要を説明しておこう。

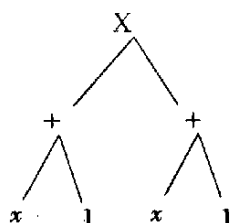
$(x+1)^2$ の x に関する導関数は $2x+2$ である。このような式の微分をコンピュータで実行するには、どのようにすればよいであろうか

? 考えなければならない問題は

1° コンピュータの外部における式の記号化
と

2° 記号化された式のコンピュータ内部における処理方法
とである。1° はコンピュータに与えられる情報の表現の問題であり、2° は表現された情報の処理方式を考えることに相当する。上の問題について、このことを具体的に考えてみよう。 $(x+1)^2$ の演算の順序は第1図に示すように木構造表現を利用して表わすことができる。

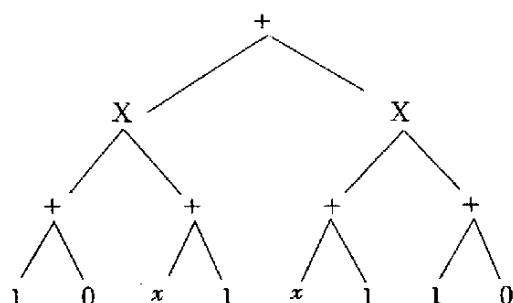
第1図の意味は図から自明であろう。もし、コンピュータの中にこの図に相当する表現を入れることができれば木の根に相当する部分の X の



第 1 図

記号の部分に乗算に関する微分法則を使用する。そのためには、その下の部分の+に関する微分処理を必要とすることになる。上のXから下がって帰納的に微分法則を使用して第2図に示すような木構造表現あるいはそれに相当する他の表現を得ることができれば、第1図の表現をコンピュータに入れたのと逆の

方向で第2図の表現をコンピュータの外にとりだすことができ、
 $1 \times (x+1) + (x+1) \times 1$, すなわち $2x+2$ と解釈することができる。



第 2 図

この際 $(x+1)^2$ という記号列は第1図に対する慣用化された表現であるが、これ以外に第1図を適確に表現する記号列は構成されないものだろうか？ たとえば、

$$X + x \ 1 + x \ 1 \quad (1)$$

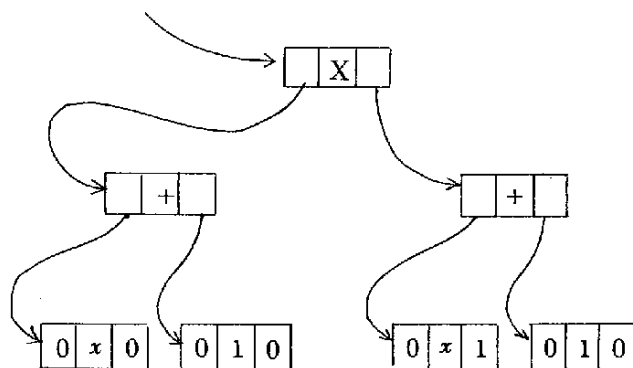
あるいは、カッコを付して

$$(X(+x \ 1)(+x \ 1)) \quad (2)$$

と表現してみてもどうであろうか？ これは、よく知られているように、いわゆるポーランド記法で表現した $(x+1)^2$ である。コンピュータで $(x+1)^2$ の計算を実行する場合であれば、逆ポーランド記法

$$x \ 1 + x \ 1 + X \quad (3)$$

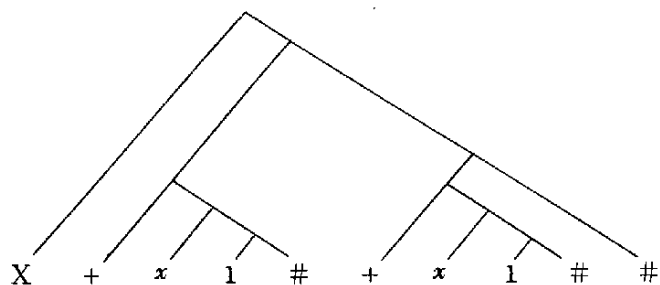
の方が有用であるが、記号列を左から右へ読みこんでいくとき、第1図に相当する第3図のような関連状況を記憶の中に格納するには(1)または(2)の方が適切である。



第3図

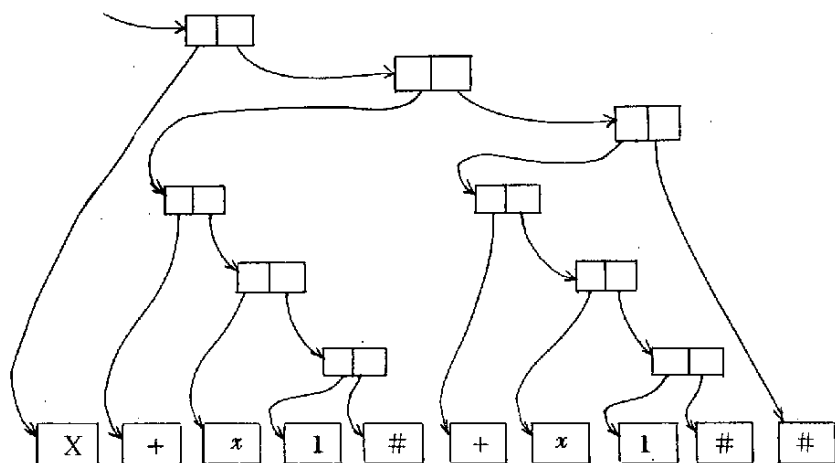
このような図では1つのセル（場合によっては3個のセルを1つにまとめる）を3個の部分にわけて第1の部分では木構造の左の部分、第2の部分（中央）では木構造の各節の上ののっかっている記号、第3の部分（右）では木構造の右の部分の表示を示している。

ところで、 $(x+1)^2$ は第4図のようにも表現される。



第 4 図

これは最下欄に一線にならんだ記号列を左からならべて書けばわかるように、(1)の形となっている。ただし、#という各演算段階の終末につけた記号を除去して考えてのことである。これは第5図のように記憶の中に格納することができる。



第 5 図

第3図を処理して微分の結果を得る方式と、第5図を処理して得る方式とは若干の相違がでてくるのは当然であろう。

以上のことから、記号処理の問題では

1° 与えられた情報とその情報の構造（連関状況）をどのように表現すれば、コンピュータの中で適確な表示を得るか？

ということと、

2° このような記憶内での表現をもとにして処理をする方式を表現するにはどのような構造の言語を使用すればよいか？

という2点から記号処理用の言語を作成する必要がある。

いままでに知られている記号処理言語の代表的なものをあげれば

IPL V, ⁶, LISP 1.5, SNOBOL 4

等が考えられる。もちろん他にも重要な言語もあるが、ここでは上の言語を代表としてとっておこう。IPL V は記号処理としては最初に考えられたという点で古典的なものとしての代表であり、⁶は機械に近い表現、SNOBOL 4 は人間側に近い言語、これに対してLISP はこの中間をいく特色ある言語という意味で代表と考えられるのである。

これらの言語は、いうまでもなく機械固有の言語ではないから、言語の仕様の他に、その使用のためには処理システムの作成を必要とする。記号処理言語を考える上の出発点として、われわれはLISP を代表的なものとしてとり、処理システムの作成という点に焦点を絞った。いわば追試の段階から研究を開始したのである。

2. 自動的なM-S変換

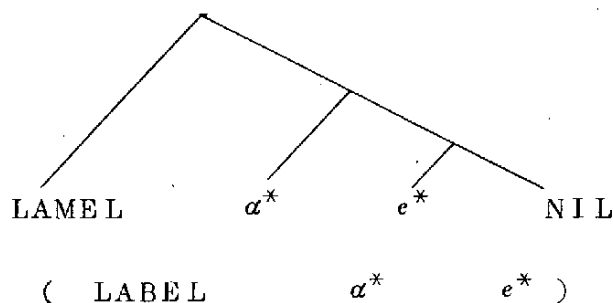
LISP 1.5 はJ. McCarthy によって開発された言語である。このマニュアルは一読では読み通せないという欠点をもっている。最初の数頁はがまんできるが、段々と壁にぶつかる代物である。それはどういう点かということ、情報構造をS式で、処理方式をM式で表現し、これを別々

に区別している間はまあまあといった感じであるが、M式で表現したプログラムをコンピュータに入れる際になって問題が生ずるのである。このM式をS式に変換して入れるからである。S式というのは(2)のようなリストを巢作りにしたような構造が主になった表現であるので、この形の式を読みこんで第5図のような関連状況を記憶内に作成するのは簡単なプログラムを用意することによって達成する。ところがM式をコンピュータに入れて使用するにはこのままの形では不適當な個所が相当でてくる。そこで、与えられた各M式に対して、1つのS式を対応させる規則を作成し、この規則を帰納的に使用してS式を作ることにした。こうすることによってデータを入れるのと同じプログラムで、M式に対応するS式が記憶の中に格納され、このS式化された本来のM式を使用してデータのS式を処理させればM式を入れる特別のプログラムを必要としないということになった。

M式を同等なS式に変換する規則は、次のように表現される(くわしくは〔1〕参照)

1° M式が関数の場合:

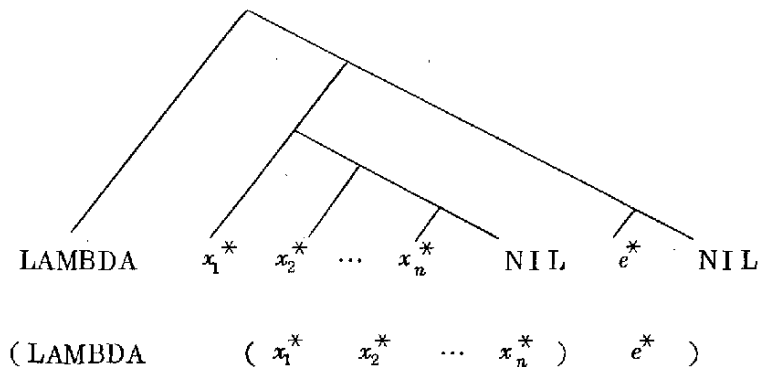
(i) label [α ; e]



第 6 図

ここに α^* , e^* はそれぞれ α , e の S 式への変換形である。以下
 $*$ 印はこの意味に使用する。また NIL は前節の # に相当する
 LISP 特有の終止記号列である。

(iii) $\lambda \left[(x_1 ; x_2 ; \dots ; x_n) ; e \right]$



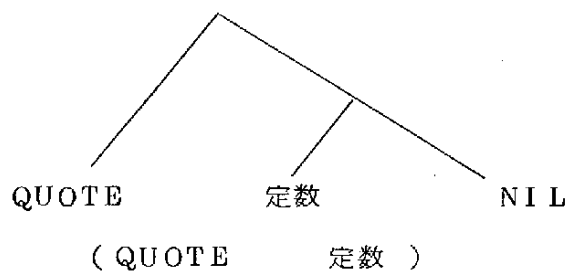
第 7 図

(iii) 関数の名前

小文字を大文字に変更。

2° M 式が形式の場合；

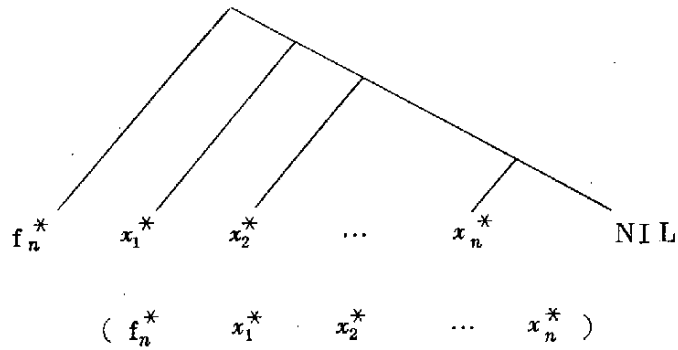
(i) 定 数



第 8 図

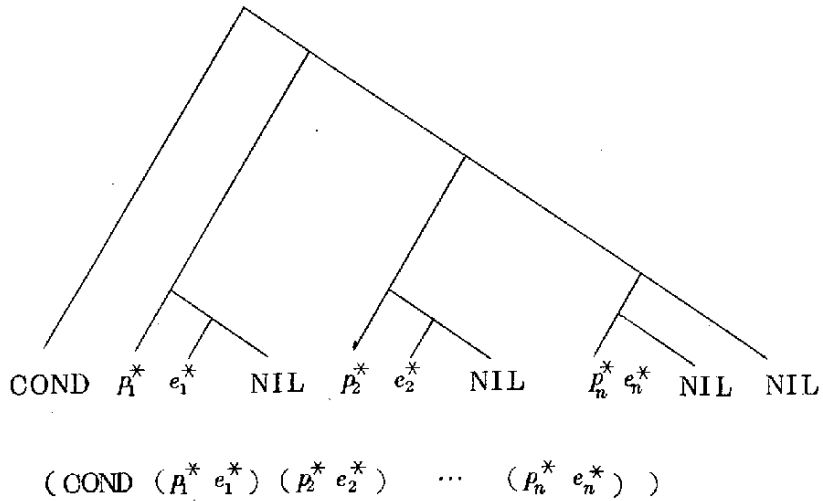
(ロ) 変数は関数の名前の場合と同じ

(ハ) $f_n [x_1 ; x_2 ; \dots ; x_n]$



第 9 図

(ニ) $[p_1 \rightarrow e_1 ; p_2 \rightarrow e_2 ; \dots ; p_n \rightarrow e_n]$



第 10 図

このような木構造表現の使用はLISPの処理関数の処理方式を理解するのにずいぶん有益であった。たとえば car, cdr, cons 等は図の上で処理

が明示される。

この変換された形をつねに使用することによってマニュアルの説明が難解さを増すことがわかったので、われわれはこのM-S変換を自動的に実行するプログラムを作成することを思いついた。このようにすればM式を直接使用できて理解しやすくなる。このプログラムの作成は最初は簡単のように見えたが、結局言語の構造解析の一般的手法をここでもとらざるを得なかった。それは"transition diagram"の方法とよばれるもので、スタックを利用して現在処理している段階がどの段階であるかを記憶して、その段階の処理に移る。その処理が終了すればスタックの第1番目の内容をポップ・アップして第2番目に格納してある処理段階標示をみてもとに帰る。これをくり返すことにより、最終の結果を得ることに成功した。それが〔2〕である。使用したコンピュータはOKITAC 5090-Cである。この自動的なM-S変換プログラムの使用により、入力としてはM式も使用できることになり、LISPが一段と利用しやすくなる。

〔3〕ではM-S変換が実質的には2位の構文則向き翻訳模型であることが示された。これによって〔2〕における自動プログラムの正当性が裏づけられる。

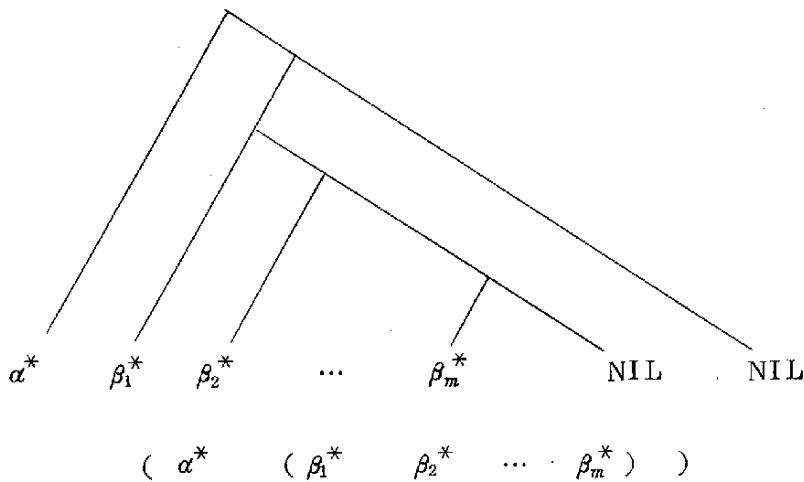
3. 言語解析における優先順位行列の自動作成

LISPを言語解析に応用したのが〔4, 5〕である。優先順位行列についてのくわしい解説は紙数の関係で〔1〕にゆずることにするが、優先順位行列の方法というのは単純句構造言語の構文解析の上昇型解析の一手法としてR.W.Floydによって導入され、N.Wirth-H.Weberによって

一般化された方法である。〔4〕では LISP が Wirth-Weber 型の優先順位行列をコンピュータによって作成できることを実際に示した。方法は、この場合の入力情報としての単純句構造文法の表現をたくみに利用することにある。たとえば、生成規則

$$\alpha \rightarrow \beta_1 \beta_2 \cdots \beta_m$$

を第 11 図のように表現することから出発する。



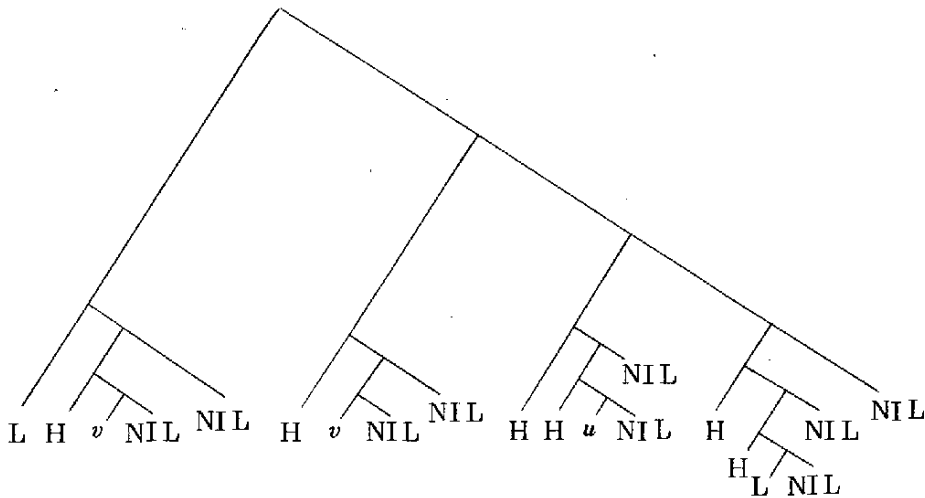
第 11 図

有限個の生成規則よりなる文法の場合はこれを 1 つの単位としてリストを構成する。たとえば文法が次のようになっているとする：

$$\left\{ \begin{array}{l} L \rightarrow Hv \\ H \rightarrow v \\ H \rightarrow Hu \\ H \rightarrow HL \end{array} \right.$$

(ここに u , v は基礎記号)

このときの S 式の木構造表現は次のようになる。



第 12 図

これをデータとして処理関数を作成していくのである。〔5〕では実際にこの問題に適用して成功した結果を示した。

4. 今後の研究

いままでにあげていなかった記号処理言語の 1 つに SPRINT 2 とよばれる面白い言語がある。これは構造がいたって簡単で、機械言語に近い形をしている上に、スタック 2 個を自由に駆使できるようになっている。われわれはこの言語を土台にして必要な改良を加え、他の言語の処理プログラムを作成する一助にしたいと考えている。実際、簡単な FORTRAN, 制限した LISP の場合にはシステム作成がうまくいっているようである。情報構造の表現も自由性があるので、かなり期待がもてそう

である。さらに、オートマドン論や、人工言語論への応用を考察してコンピュータ利用の各理論への有効性をたしかめたい。

文 献

- [1] 藤野精一： リスト処理と言語解析，（朝倉書店・プログラミング演習シリーズ3）（1970） 192頁
- [2] 藤野精一，富樫 昭： M-S translation in a list processing language LISP， 九大理学部紀要 24 №2（1970）， 274-290
- [3] 藤野精一，富樫 昭： Syntax-directed translation scheme of M-S translation in a list processing language LISP， 九大理学部紀要 25 №1（1971）， 6-9
- [4] 飯盛末夫，藤野精一： List processing for finding Wirth-Weber type precedence relations between symbols， 九大理学部紀要 25， №1（1971）， 152-166
- [5] 飯盛末夫： A remark on list processing for finding Wirth-Weber type precedence relations between symbols， 九大理学部紀要 25， №1（1971）

自然語科学情報のデータパターン及び インタレスト・パターン処理の基礎研究

東京大学理学部教授 藤原 鎮男

助手 山本 毅雄

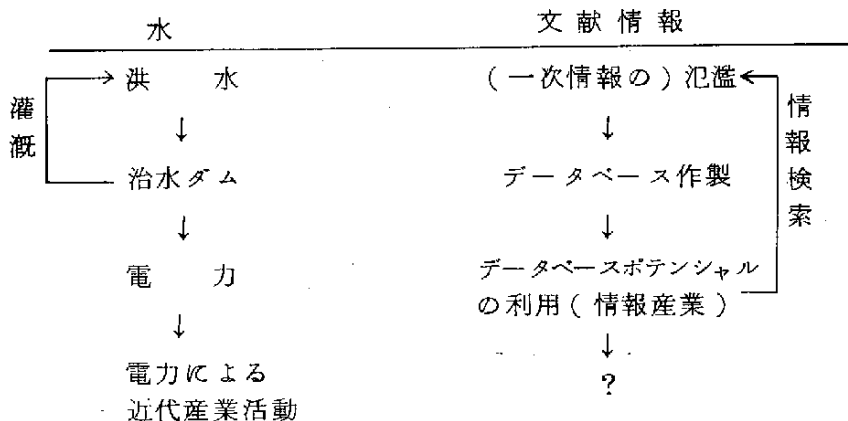
1. 化学情報検索の概況とその意義

1.1 情報エネルギー

情報の氾濫とその処理法の最近の進歩とは、ともに世上喧しくのべられるが、その認識はやゝ表面的のきらいがあるようである。情報氾濫の社会的意義をまず考え将来への影響を述べ、それによって我々が本研究を志した動機を示したい。

同じく氾濫という言葉が用いられるので水と対比して情報の社会的意義を考えよう。第1表は水が治水ダムの建設によってそのエネルギーが

第 1 表



蓄わえられ、電力にそれが変換され、その結果近代工業社会がつくられることを示している。ダムに蓄積された水は一部そのままの形で灌漑用水として利用されることも重要である。

さて情報の場合、水の各段階との対比は第1表より明らかであるが、注意すべきことは、情報のもつポテンシャルエネルギーについて世人の認識が低く、ダムに相当する情報データベースがつくられたあと、そのエネルギーの利用すなわち、電力および近代工業に相当する社会活動がまだ存在しないことが深く認識されていないということである。いふなれば情報検索は灌漑用水的利用の範囲を出ないのであって、その意味ではきわめて初歩的段階にあるといわねばならない。

水力の電力への変換に対応する情報エネルギーの変換が将来どのような形態で起るか、それを考えるのが最も重要な情報科学の課題である。その一つの方向は、計算機による情報処理が、部門部門に狭く閉じこめられていた情報データを部門を超えた場に解放し、いわば専門の解放を行うところにある。それはまだ情報データの形態にとどまっているが、自由なエネルギーとして部門を超えた利用が始められたわけであり、部門別の形態を異にし、局所に存在が制限されていたエネルギーが変換され、流動し始めたといってよい。それは水力の電力への変換に相当する事態の始まりといってよいかもしれない。

そうであるとする、この場合重要なことは異なる部門の情報エネルギーを利用し易い形式で利用しうるようにすることであり、自然語情報を本研究でとり上げる理由の一つはここにある。

1.2 学問の学問、科学の科学の誕生

従来、専門家とはある領域の知識の占有者であるとされた。しかし文

文献情報に関する限り機械検索の発展により専門知識といえども門外漢が自由に利用出来るようになった。それは専門分野の解放である。その結果、狭い個々の専門の上に立つ総合的な学問が成立する。比較文学などは従来もあったが、より理想的な共同研究、統計的研究法、システム研究法、科学政策の決定法などというものが生れうるわけであり、それは従来の専門別の学問技術と異質の方法論のテクノロジーで、画期的な新分野である。

しかし、そのようなものの必要性は漠然とすべての人が感じていることであり筆者のみの発言ではないであろう。

本研究において筆者らが考えるのはそれを実際に進める基本技術である。

専門の解放といって専門別に大量に蓄積された情報を無差別に出したのでは、これは乱雑に放置された情報と結局は大差がない。専門家を尊重するのは、一つの分野に関してその人がもつ情報に価値を認めるだけでなく、その人がそれを蓄積する過程に個々の情報について評価をすることに価値を認めるのである。

機械検索ではこの評価が欠けており実際には別に設定された基準に依存せねばならない。たとえば文献情報は一次情報として a)学会誌、国際学術誌、速報誌、d)協会誌、大学紀要、研究所報告、c)技術報告、社内報告、および c)学位論文、などに分類されるが、学術上の正しさからいえば a)、d)がよく、b)、c)は低く評価されるので機械検索の場合もこの分類は一つの基準となりうる。しかしながら文献情報の評価は必ずしも学術上のレベルや完全さにあるのではなく、着想やある場合にはある化合物をとり上げたかどうかということだけで大きな価値を与えうる場合がある。すなわち、機械検索が新しい学問の学問をつくるのに寄与す

るには、インプット・アウトプットを含めてデータベースの内容に評価を含まねばならないが現状はこの点で実績はないといえる。

評価はその操作内容からみれば比較である。評価なり比較なりを文献情報について行なう場合通常とられる方法は結局は人間の思考によっている。すなわち、今迄のべた評価ないし比較の作業はひっきりん人手により、検索者の思考、指示により行なわれるものといってよい。しかし乍ら、筆者がここでいいたいのは、機械検索の終局的な目標としては、文献情報内容に立ち入って、そのもつ文脈、構成、用語の機械的解析法を樹立し、それによって「文献」の本体の解析を行なうところに真の意味の検索があり、その作業をへて明らかにされた文献情報の特性を集録、分類するとき、それは始めて現在存在するところの専門分野から全く独立に解放された知識の分類となり、それは我々が現在全く知る所のない機械検索によって創造されたが生れるということである。現在これは全くの目標にすぎないが本研究の指向するところである。

1.3 化学文献情報の量と種類

これについては最近筆者が別に記したので御参照を仰ぎたい。(藤原鎮男, 学術月報24巻11号, 22頁 昭47)

以上の基本的背景のもとに我々の行なった研究の概要を以下にのべる。

東大理学部化学教室 藤原鎮男, 同情報科学研究施設 国井利泰,
東大大型計算機センター 山本毅雄, お茶の水女子大学 細矢治夫,
東大大型計算機センター 故池田親民, 牛丸 守, 仲野憲一, 熊井友子

2. 実際の作業

2.1 自然語情報処理の問題点

自然語は、その表現の柔軟性と内容の豊かさによって、科学情報のようにたえず変化・発展する膨大な量の情報を表現するのに適している。これを十分に利用・評価するためには、次のような点の解決が必要である。

- a 自然語の特質をつかんだ適切な検索法の開発・単純な一致・不一致による検索だけでなく、種々のロジックに従う検索、文脈に含まれる情報の検索などができるようになっていなければならない。
- b 自然語のあいまい性・可塑性に対処するため、人間の判断をシステムにとりいれてゆく方法の確立。そのためには、システムと人間の対話や、システムを介する人間と人間の情報交換が十分におこなわれなければならない。また、作業の結果に対する人間の判断が、自然な形でシステムにフィードバックされ、システムが全体として賢くなつてゆかなければならない。
- c 可変長の大量データの処理法の確立。特に、長大な文字連系を含むファイルの生成と管理、大量のデータとオンライン・システムの関係のつけ方などは大きな問題である。

上の各点にわれわれなりの解決を呈示すべく、

- (1) 統辞論的近接度の考えを利用した自然語（英語）の内容解析法の開発
- (2) 自然語科学情報蓄積・検索・配布システム TSIR-1 開発をおこなった。

2.2 統辞論的近接度

たとえば、英語の文章中で

- a COBALT, COPPER AND ZINC COMPLEXES OF
- b TRIAMMINE COMPLEXES OF COBALT ARE
- c COMPLEX SPECTRA WERE OBSERVED COBALT
SOLUTION.....

という3種の表現があるとき、連糸“COMPLEX”はa, bでは“錯化合物” cでは“複雑な”を意味する。このように文脈によって意味の異なりうる場合、これを機械検索で一般的に解析することは非常に複雑な大問題であり、大量の自然語データに対してこれを可能にする方法は現在存在しないといってよい。

ところで、これらの文章の中での COMPLEX, COBALT という二種の連糸の関係に話を限り、これらのデータが“コバルトの錯化合物”に関する文章であるか否かの解析は可能であろうか。われわれは、次のように、“統辞論的近接度”という概念を導入することによって、上の解析が可能なることを示した。¹⁾

まず、検索質問により連糸M(質問語、上の例でいえば COMPLEX)が与えられ、検索システムから打ち切り長(自然数) n_{prox} とふたつの辞書すなわち前置終止記号辞書と後続終止記号辞書が与えられているとする。ここで前置(後続)終止記号辞書は、後の例のように、一般にはことなる何個か(零個を含む)の連糸の表である。これらの連糸を、前置(後続)終止記号という。

定義：連糸Mの前方(後方)近傍とは、次のすべてを満足する連糸である。

1. この連糸はMの前方(後方)に連続している。
2. 前置終止記号(後統終止記号)を含まない。
3. 長さは n_{prox} 以下である。
4. データの始まり(終り)とMの間にあるか、さもなければ長さ n_{prox} であるか、さもなければ前置終止記号ではじまる(後統終止記号で終る)。

定義：連糸N(例えば上記の例の COBALT) が少なくとも連糸Mの前方近傍又は後方近傍に含まれるとき、NはMと統辞論的に近いという。

このような意味でのMとNの近さ(統辞論的近接度)は、当然、M、 n_{prox} 、前置および後統終止記号辞書の与え方によって種々の場合がある。たとえば、Mの前方近傍と後方近傍は、¹⁾

- a Mと同じ語の残り部分
- b Mのとなりの語
- c Mから一定距離以内にある部分
- d Mの形容詞句

などのひとつにはほぼ一致させることが可能である。dのためには、

d_{prox} を十分長くとり、前置終止記号辞書として通常の終止符、関係代名詞、助動詞、動詞、前置詞のリスト、後統終止記号辞書としては終止符、関係代名詞、助動詞、動詞のリストを与えればよい。

このように、統辞論的近接度の考え方によって、いろいろな意味での「近さ」を統一的に定義することができ、したがって広範囲の文脈解析を同一のアルゴリズムを用いて(単にデータだけの変更で)おこなうことができる。また、上のようにデータとして与える辞書に多少不完全さがあっても、結果の信頼性にはそれほどの影響を与えないことがありう

ることに注意したい。たとえばdの場合、辞書から動詞のリストを除いても、助動詞を多用するデータに対しては結果に影響が少なく、しかも検索の手間をはぶく効果は大きい。

2.3 TSIR-1の開発^{2)~12)}

2.3.1 全体の流れ

TSIR-1は、2.1でのべた目的を実現するため、東大大型計算機センターのHITAC 5020 TSS(2次元アドレス付)をベースとし、この機能を種々の点で拡張した自然語情報の蓄積・検索・配布システムである。

このシステムの特徴は、

- ① 大規模データを扱いながら、TSSと強く結びついていること。特に利用者とシステムとの交渉はすべてTSSを通じておこなわれ、人間がファイル構造や検索技術の非本質的な細部にわずらわされることなくシステムと対話できること。
- ② CAS (Chemical Abstracts Service)をはじめとする大量文献データベースと、これを検索してえられるサブファイル、各利用者が会話形式で生成するローカルなデータ、あるいは紙テープを介してえられる測定値情報などを同一ファイル構造に変換してシステムにとりこみ、ソフトウェアの共用可能性と、データ源の弁別可能性を保ったこと。
- ③ 検索質問を重視し、これをひとつの知的財産としてオンライン・ファイルにたくわえ、システムとの会話による生成・更新が常時可能であるようにしたこと。これによって、同一グループの利用者間での共有、他グループとの間の情報交換あるいは機密保持がすっきりした形で可

能となり、また検索結果の評価を検索質問の更新の形でシステムにフィードバックすることが自然におこなえるようになった。

④ 検索プロセスは、表を利用したフレキシブルなものとし、2.2の方法を含む種々の検索アルゴリズムが可能であること。また、システムがバッチ的におこなう大量検索においては、各利用者のファイルにアクセスして検索質問をあつめ、これを編集して能率のよい検索をおこなっていること。

⑤ 検索結果を原データベースのサブファイルとして、さらにくわしい検索や将来の利用のためのデータベースとする道をひらいたこと。

などである。Fig 1. に全体の流れを示す。

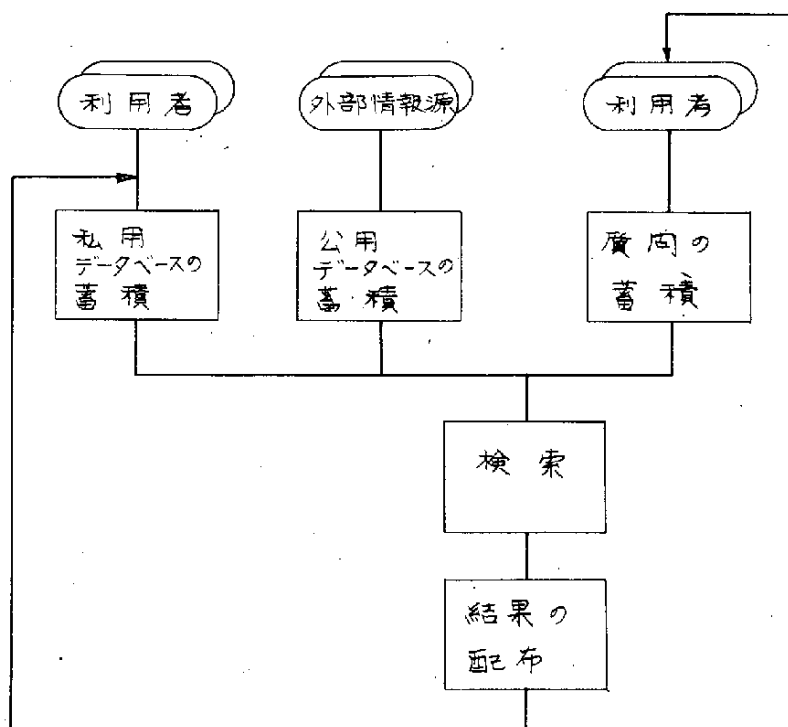


Fig 1. 全体の流れ

2.3.2 私用データベースの蓄積

このシステムの利用者(科学者)は、共同研究者とファイルを共有するグループを作り、このグループ単位でシステムに登録することができる。これらの利用者は、手持ちの文献データをTSS端末から自分のグループのファイルに入れ、私用データベースのレコードを生成することができる。この生成(Fig 2, Fig 3.), 更新, 印刷(Fig 4.) あるいはKWIC 索引生成などを, 利用者になるべく負担をかけない形で会話形式によってすすめるためのTSS ライブラリプログラムを作った。^{3), 4), 8)}

一方, 質問者の指定によって, 公用大規模ファイルに対してシステムが定期的におこなう現況速報検索の結果を, 公用ファイルのサブファイルの形で私用データベースへ配布する機能を計画した。このために, システムが各利用者のファイルに次々とアクセスし, 情報を配布してゆけるようなオンラインファイル管理システムの開発¹¹⁾が必要となった。

このように, ローカルに生成したレコードと, 大規模データベースからのレコードが同一ファイルに入り, 同一のソフトウェアで扱いうるためには, これらの共用可能性・弁別可能性を保たなければならない。そのためにはファイル構造・ソフトウェアの両面にわたって種々の注意が必要であった。^{4), 5)}

2.3.3 公用データベースの蓄積^{5), 9)}

システムは, いくつかの外部情報源(現在はChemical Abstracts Condensatesのみ)から磁気テープデータをうけ入れ, これをシステムで使える形に変換・蓄積する。Chemical Abstracts Condensates テープを, 2回テープ-テープ変換して, トラック数, 密度, 文字コード, ブロック構造等の変換をおこない, TSIR-1の公用データベース

\$RUN(GENERATE-STF)

INPUT RECORD FORMAT
HOW MANY AUTHORS?

... 3

KEYWORDS, 1 OR 0?

... 0

ARTICLE TITLE, 1 OR 0?

... 1

JOURNAL NAME, 1 OR 0?

... 1

CODEN, 1 OR 0?

... 0

VOLUME NUMBER, 1 OR 0?

... 1

PAGE, 1 OR 0?

... 1

YEAR, 1 OR 0?

... 1

ANY OTHER CATEGORY, 1 OR 0?

... 1

INPUT IDNO IN 4 DIGITS

... 4096

INPUT IDMOD IN 2 DIGITS

... 01

ANY OTHER CATEGORY, 1 OR 0?

... 0

Fig 2. 会話形式によるデータベース生成 (1)

```

INPUT AUTHOR NAMES, ONE BY ONE
... FUJIWAKA S

... AOYAGI K

... MIYAMAE T

INPUT TITLE. INPUT ==SIGN AT THE END
... NUCLEIDIC MASS MEASUREMENT BY ION CYCLOTRON RESONANCE=

INPUT JOURNALNAME
... BULL. CHEM. SOC. JAPAN

INPUT VOLUME NUMBER
... 43

INPUT PAGE
... 561

INPUT YEAR
... 1970

INPUT DATA ELEMENT, IDNO= 4096---SIGN AT THE END
... A NEW METHOD FOR MEASURING THE NUCLEAR MASS BY ION CYCLOTRON

... RESONANCE SPECTROSCOPY IS DESCRIBED. THE METHOD IS APPLIED

... TO THE DETERMINATION OF THE NUCLEAR MASS OF ARGON-40.

... THE ACCURACY OF THE METHOD AND POSSIBLE SOURCES OF ERROR ARE

... DISCUSSED.=

INPUT ABSTRACT NUMBER IN 4 CHARACTERS
... 0027

LINES LEFT= 38
INPUT 1 IF NEXT DATA INPUT
INPUT 0 IF INPUT END AND FILE REGISTRATION WANTED

... 0

INPUT FILENAME IN 4 DIGITS
... BCSP
FILENAME= FILEBCSP

CPUTIME 005.8SEC

```

Fig 3. 会話形式によるデータベースの生成 (2)


```

$RUNC(PRINT=STF)
INPUT STF TSSC FILE NAME

... FILEHCSP
INPUT SERIES NUMBER IN 4 DIGITS

... 0027
00597001 0000A FUJIWAHA S
00597002 0000R ANYAOI K
00597003 00009 MIYAMAF I
00597001 00035 NUCLEIDIC MASS MEASUREMENT BY ION CYCLOTRON RESONANC
F
00597001 00016 BULL. CHEM. SOC. JAPAN
00597201 00002 43
00617301 00003 561
00597001 00006 000070
10007001 000F7 A NEW METHOD FOR MEASURING THE NUCLEAR MASS BY ION C
YCLOTRON RESONANCE SPECTROSCOPY IS DESCRIBED. THE M
ETHOD IS APPLIED TO THE DETERMINATION OF THE NUCLEAR
MASS OF ARGON-40. THE ACCURACY OF THE METHOD AND P
OSSIBLE SOURCES OF ERROR ARE DISCUSSED.

00127001 0000R XX000027
00547001 00009 XX0000027
INPUT 1 IF NEXT SERIES NO INPUT
INPUT 0 IF INPUT END
... 0
CPU TIME 002.75FC

```

Fig 4. 会話形式によるデータベースの印刷

とした。

2.3.4 質問の蓄積¹¹⁾

利用者は、TSS端末から会話形式または紙テープ入力により質問を提出し、自分のグループのファイル内に蓄積する。各利用者は、自分の質問の更新や、自分のグループ内の他の利用者の質問を参考にして新しい質問を作成することを、常時必要な時にやれることになる。

一般性のある質問は公用ファイル中にライブラリデータとして登録すれば、各利用者がこれを読み出して使うことが可能になる。

2.3.5 検 索¹⁰⁾

検索作業はつぎの3種類にわけられる。

- ① 大規模現況速報検索、外部データベースからシステムに新しいデー

タが入ってくるとき、これを材料にする。検索質問（2次質問）は多数の利用者の質問（1次質問）をまとめて編集したものであり、検索対象は磁気テープに入っている。質問者は検索の出力方法として、ラインプリンタへの文献リスト、質問者のファイル中へのサブファイル作成、質問者の磁気テープファイルへのサブファイル作成のどれかを選択することができる。

② 小規模遡及検索・サブファイルを所有する質問者が必要に応じてオンラインで実行する。2次質問はこの利用者（グループ）の1次質問を何題かまとめたものである。結果の出力方法として、質問者のファイル中へのサブファイル作成、質問者の端末への文献リストのどちらかを選択できる。

③ 大規模遡及検索・大規模データベースに対する遡及検索要求が一定限度以上に集まった時に実行する。検索対象、出力方法などはその都度決定する。

Fig 5. に、②の結果の文献リストの例を示す。

検索の内容は、多数のライブラリプログラムとしてシステムに登録しており、作業の種類と質問データの内容によってこれらのうちの必要なものが動的にリンク、実行される。その種類は大別して次のようになる。

① 項目の存否の検索・あるレコード中に、求めるデータ項目（たとえば文献の発行年）が存在するか否かの検索。

② 質問語の存否の検索：ある項目のデータ要素中に、求める質問語が存在するか否かの検索。これはさらにつぎの3種類にわかれる。

a 語（W）検索・データ要素が構造を持たない単語である場合（たとえば人名、誌名など）、質問語が、データ要素内のあるきまった

CHARGE-TRANSFER INTERACTION AND CHEMICAL REACTION. 2. INTERACTION OF S-TRIAMINOBENZENE WITH SEVERAL ELECTRON ACCEPTORS

T.YAMAOKA S.NAGAKURA
BULL.CHEM.SOC.JAP. VOL43, 355 (1970) CODEN=BCSJAB

CRYSTAL AND MOLECULAR STRUCTURE OF DI-MUE-CHLORO-BIS(PAI-1-ETHOXY CARBONYL-2-CHDROXYALLYL) DIPALLADIUM(2)

K.ODA N.YASUOKA T.UEKI N.KASAI M.KAKUDO
BULL.CHEM.SOC.JAP. VOL43, 362 (1970) CODEN=BCSJAB

STERIC STRUCTURE OF L-PROLINE OLIGOPEPTIDES. 4. PROTON MAGNETIC RESONANCE STUDIES OF L-PROLINE OLIGOPEPTIDES

H.OKADAYASHI T.ISEMURA
BULL.CHEM.SOC.JAP. VOL43, 359 (1970) CODEN=BCSJAB

NUCLEAR MAGNETIC RESONANCE IN PARAMAGNETIC SOLUTION. PROTON MAGNETIC RESONANCE STUDIES OF SOME SUBSTITUTED ANILINE DERIVATIVES IN THE PRESENCE OF PARAMAGNETIC NI(2) AND CO(2) ACETYLACETONATES

T.YONEZAWA I.MORISHIMA Y.AKANA K.FUKUTA
BULL.CHEM.SOC.JAP. VOL43, 379 (1970) CODEN=BCSJAB

ELECTROCHEMICAL PROPERTIES OF .ALPHA.-FERRIC OXIDE IN SULFURIC ACID SOLUTIONS

H.YANEYAMA H.TAMURA
BULL.CHEM.SOC.JAP. VOL43, 350 (1970) CODEN=BCSJAB

CHROMIUM(3) COMPLEXES WITH IMINODIACETIC ACID OR L-ASPARTIC ACID

H.MIZUOCHI S.SHIRAKATA E.KYUNO R.TSUCHIYA
BULL.CHEM.SOC.JAP. VOL43, 397 (1970) CODEN=BCSJAB

ALTERNATING COPOLYMERIZATION OF .BETA.-CYANOACROLEIN WITH .ALPHA.-METHYLSTYRENE

I.TAKEMURA H.SUMITOMO
BULL.CHEM.SOC.JAP. VOL43, 334 (1970) CODEN=BCSJAB

部分と一致するか否かの検索。

b 句 (P) 検索・データ要素が複合語、句、簡単な文など(たとえば論文標題など)、質問語がデータ要素内の任意の部分と一致するか否かの検索。

c 文 (S) 検索・データ要素が複雑な文あるいは長い文章のとき、ふたつの質問語の間の統辞論的近接度をしらべる検索。これについては現在予備的な試みの段階である。

以上のような質問語別の内容検索の結果を総合し、その間のロジックを判定してはじめて、ある文献レコードがある検索質問に適合するか否かが決まる。これを「質問者別処理」とよぶ。内容検索、質問者別処理を含めた検索の主要部分の流れ図を Fig. 6. に示す。

2.3.6 結果の配布

結果の配布は、テレタイプ端末あるいはラインプリンタに対する文献リスト出力、KWIC 索引出力のほか、磁気テープあるいはディスクファイルの形式で原ファイルのサブファイルを配布することを特色とする。このためには、システムが各質問者のファイルに次々とアクセスして結果の書き込みをすることが必要である。

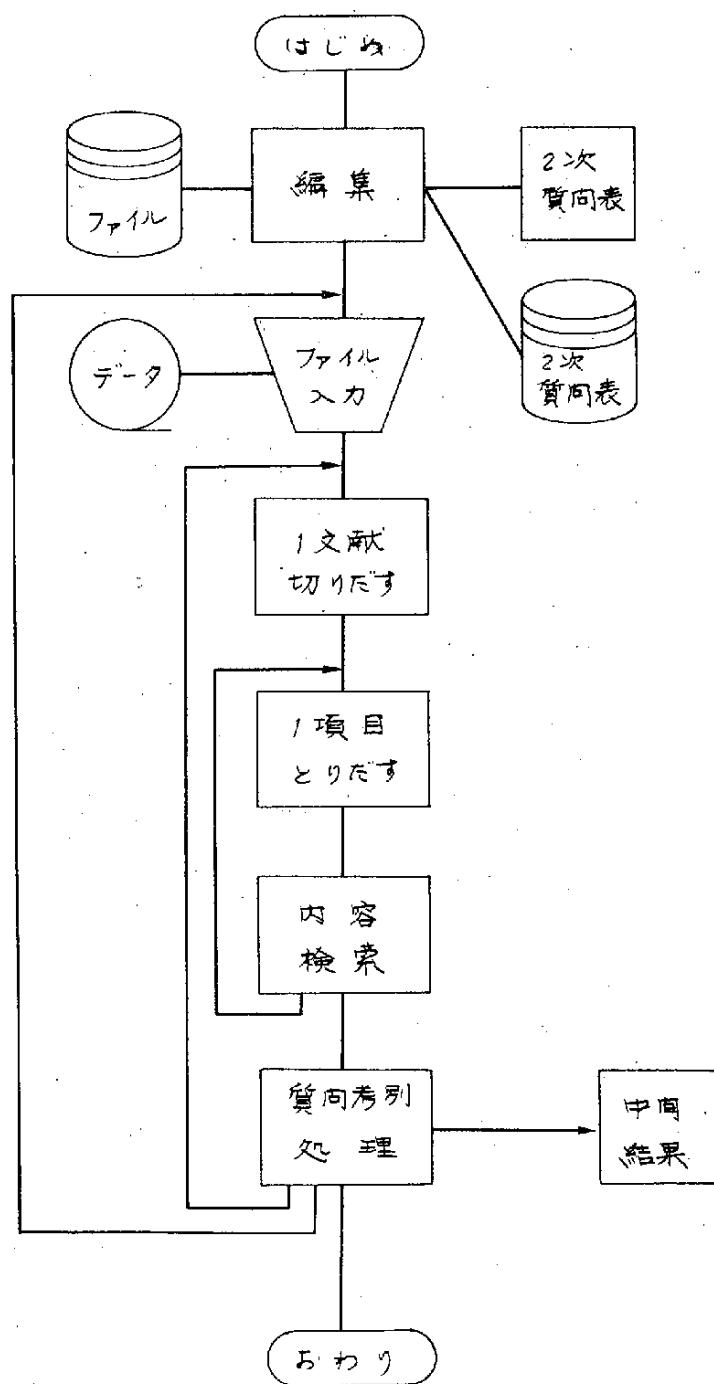


Fig 6. 検索の主要部分

文献および講演

- 1) Syntactical Proximity : Partial Syntactical Analysis of Natural Language Data Records
T.Yamamoto, S.Fujiwara, J.Chem.Doc., 11, P.(1971)
- 2) 科学情報蓄積, 検索および配布システムの研究
山本, 梅沢, 池田, 仲野, 牛丸, 済賀, 染谷, 国井, 高橋
東大大型計算機センター年報, 1, 3 (1971)
- 3) 自然語文献情報データベースの会話形式による生成と利用
東大大型計算機センター年報, 1, 50 (1971)
- 4) Todai Scientific Information Retrieval (TSIR-1) System. I.Generation, Updating and Listing of a Scientific Literature Data Base by Conversational Input
T.Yamamoto, T.Kumai, K.Nakano, C.Ikeda, T.L.Kunii, H.Takahasi, S.Fujiwara, J.Chem.Doc., 11, P.(1971)
- 5) Todai Scientific Information Retrieval (TSIR-1) System. II.Generation of a Scientific Literature Data Base in a Center-Oriented Format by a Tape-to-Tape Conversion of CAS SDF Data Base
T.Yamamoto, M.Ushimaru, T.L.Kunii, H.Takahasi, S.Fujiwara, J.Chem.Doc., 印刷中
- 6) オンライン情報検索の研究——分析化学誌の例
山本, 白川, 藤原,
分析化学会第20年会発表(1971.10.東京)
科学情報オンライン検索システム TSIR-1
情報処理学会第12回大会発表(1971.12.東京)

7) I. システムの全体像

国井, 山本, 藤原, 高橋

8) II. データベースの対話入力による生成

熊井, 山本, 池田, 仲野, 牛丸, 染谷, 国井, 藤原, 高橋

9) III. データベースのテーブル変換による生成

牛丸, 山本, 国井, 藤原, 高橋

10) III. 検索プロセス

山本, 白川, 熊井, 池田, 国井, 藤原, 高橋

11) V. 質問入力とディスク管理

池田, 山本, 国井, 藤原, 高橋

12) VI. オンラインテーブル I O C S

仲野, 池田, 山本, 国井, 藤原, 高橋

実時間計算機による思考過程分析用 実験システムの開発

大阪大学基礎工学部教授	鈴木 良次
〃 〃 講師	佐藤 俊輔
〃 〃 助手	末田 統
〃 〃 助手	小林 欣吾
大阪大学文学部 教授	田中 正吾

1. はじめに

学習や教育のプロセスに計算機を活用しようとする気運は近年ますます高まっているが、一方では機械による教育に反撥する動きもある。これは、人間が人間を教育するということが一体どういう意味をもっているのかが明らかにされていないことに大きな原因があると考えられる。教育が個人の才能を最大限に引き出すためのものである以上、機械による画一化された教育は好ましくないとする意見と、マスプロ化をよぎなくされている現状では、個別教育の良さを復活し、保っていくためにも計算機は導入されるべきだという意見とがきかれる。いずれにしても、人間は一体いかにして学習や発見を行なっていくのか、その特性はいかなるものであるのかという心理学で問われている問題を明らかにしていくことが必要であろう。特にこれまで十分に研究が行なわれていなかった人間と人間の、また人間と機械の間の相互作用を分析することが大切である。この報告は、その第一歩として、思考過程の分析をめざし、心理学実験の能率化をはかる実時間計算機実験システムについて述べるも

のである。

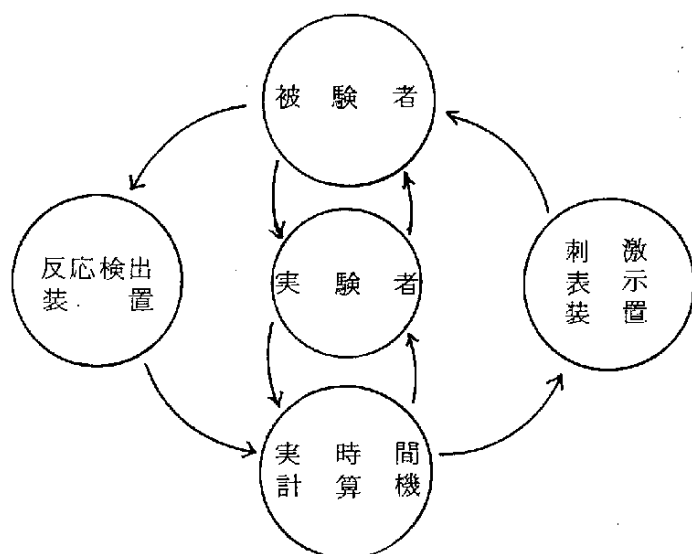
2. 心理学実験のためのトータルシステム

心理学実験は基本的には、実験者によって提示された刺激に対し、被験者がどのような反応を示したかを分析することから構成されているといえる。普通には、提示される刺激のパターンやタイミング、順序はあらかじめきめられていて、1シリーズの実験の終わったところで被験者の反応のデータが分析され、次のシリーズの実験がきめられる。しかし、実験によっては、一つ一つの反応をみながら、次に提示する刺激の性質を決定しうることを望ましい。場合によっては、正しい反応をひき出すために実験者から被験者への「強化」の情報を送りたいこともある。このように、心理学実験においては、実験者と被験者との間に存在するフィードバックループがしばしば重要な役割を果たすことがある。このフィードバックループを通じて、心理学実験が「適応的 (adaptive)」になりうるものであり、実験の能率向上にきわめて有益である。

ところで、このように実験を「適応的」にするためには、1.多種多様の刺激パターンを短時間のうちに順序をきめて提示しうること、2.被験者の反応を速やかに分析して次の刺激を決定しうることなど即応性と融通性のある実験システムが要求される。そして、このような条件をみたすものとして、実時間計算機が導入されることになる。

すでに述べたことではあるが、心理学実験は、実験者から被験者への刺激の提示、被験者からの反応の検出、実験者による反応の分析の三つの段階からなる。したがって心理学実験においてコンピュータに期待される働きも、(1)刺激パターンの準備と提示、(2)反応データの収集、(3)デ

ータの処理、解析となる。もちろん、現在のコンピュータが、実験者に完全にとって代ることはできないので、コンピュータを中心とした実験システムの中に実験者が割込んでやりとりできるチャンネルが用意されている必要がある。これらの関係をまとめると第1図の如くなる。



第1図 実時間計算機による心理学実験システム

図にも示してあるように、実験システムは、(1)刺激表示装置、(2)反応検出装置、(3)コンピュータの三つに分けられる。すでに明らかな如く、コンピュータはここでは反応データの解析、刺激のコントロールを行なうものである。

心理学実験においては、刺激としては視覚パターンが最も多く利用され、次いで音響パターン、時には動物実験では電気ショックなども利用される。視覚パターンとしては、文字、文章、数字、記号、シンボル、図形、略画、写真の如く、抽象化の進んだものから全くの具体的な事物

のパターンまで用いられている。それぞれによって適当な表示装置が考えられるが、これまでは、フィルムプロジェクターがよく活用されている。これならば、色彩も含めて多様なパターンの提示が可能であるが、パターンの選択に即応性をもたせようとする機構に特殊な工夫が必要となる。抽象化されたパターンの場合にはむしろブラウン管表示装置が便利である。パターンの修正が即座に行なえたり、新たなパターンを発生させることも容易であるから。さらに、実時間での解析結果の表示の行なえることは実験者が実験プログラムに割込むためにも都合がよい。

反応検出にもいろいろの方法が考えられる。たとえば、皮膚電気反射はウソ発見器として利用されているが、普通には、応答選択器のボタンを押して反応を知らせる。コンピュータと結びついたものとしてはタイプライター、ライトペンが便利である。音声による反応は限られた範囲のものならばよいが、一般には音声言語の計算機による認識の問題がからんでくるので容易には利用できない。

コンピュータの役割は大まかにはすでに述べたが、その中のデータの収集、刺激装置のコントロールを円滑に行なうには周辺装置との結合に融通性がなければならない。まず心理学実験ではしばしば反応に要した時間の測定が行なわれるので、そのためにはコンピュータでコントロールしうるクロックを備えておくことが望まれる。次に、反応がアナログ量としてあたえられることもあり得るので、アナログ入力端子が必要であり、データ解析の結果、外部にアナログ信号を送ることも考えられるので、DA変換器も必要となる。コンピュータはこのような情報のやり取りのみでなく、データの分析と実験計画の決定もある程度は行なうものである。刺激パターンの時間的・空間的配列をコントロールしたり、

実験パラメータの変更・修正，問題のむずかしさの程度を被験者に合うように調整する働きなども，コンピュータに期待されているものの例である。コンピュータを使ったパターン発生が有効に利用されたものとして，Julesz による奥行知覚のプロセスの研究がある。これは，コンピュータがランダムなパターンを発生しやすいということによっており，心理学実験でしばしば論じられるランダム性も，コンピュータによって解決される部分が多い。

以上心理学実験のためのトータルシステムの満たすべき条件の概要を述べたが，もちろん，システムはハードウェアだけで動くものではなく，ソフトウェアの開発が現在では，より重要な課題となっている。このことは，計算機による教育の場合も同じで，ハードウェアとしてはほとんど同じ実験システムを用いることができる。

3. 心理学実験のいくつかの例

ソフトウェアシステムとしてどのようなものを用意したらよいかという問題に答えるためには，いくつかの具体例をとりあげ，この経験が集積される必要があると考えている。

以下には，われわれのとりあげている研究課題のあらましを述べる。

(1) 図形の知覚実験

図形パターンの特徴抽出過程を明らかにするためのもので，図形のどのような特徴が知覚の手がかりとなるか，図形をどこまで単純化(略画)しても同定に差し支えないかなど図形の幾何学的特徴と知覚能力との関係を明らかにするための実験システムを開発している。実験者があらかじめ用意した無意味図形をCRTに表示し，被験者はこの図形を限られ

た数の点と直線で近似していく。近似はライトペンを使って行なわれ、略図は直ちに磁気テープに格納される。一つの原因について多数の被験者による略図がつくられるわけであるが、まず、原因のどのような部分に多くの人の注意が集中するかを統計的に処理する。また、統計処理と別の分析としては、略図から原因をあてさせる実験を多数の被験者について行ない、略図の「質の良さ」を評価する。これらの結果を合わせて、図形パターンの特徴抽出のプロセスを分析しようとするものである。

(2) 振動感覚に関する心理実験

二次元に分布した振動刺激は視覚パターンの代りとなってよく利用される。たとえば、盲人の文字読取装置はその一例である。ここでは、振動感覚を通しての学習の可能性をできるだけ追ってみようとするもので、刺激パターンの発生、被験者の反応の記録と解析をオンラインで行なうシステムの開発を行なっている。

(3) 記憶過程の解析

人間の記憶過程がどのようなものかは、計算機の新しい方式の研究に関連して非常に興味をもてる問題ではあるが、あまりわかっていないものの一つである。ここでは、文字、図形、記号をテストパターンとして、一定時間内での記銘と再生の実験を行なう。そして、どのようなパターンが記憶されやすいか、どのような学習法が記憶の能率をたかめるか、再生されたパターンがどのように変形しているか、再生の際の記憶内容の走査方式はどのようなものかなどを調べるためのシステムの開発を行なっている。

(4) 人間の規則発見過程の分析

Bruner らによって確立された帰納的推論の概念は、まだその基本的

論理構造が解明されたとはいえない。われわれはイルーシスゲームと呼ばれる規則発見のゲームを用い、人間における概念形成のプロセスに迫ることを試みている。イルーシスゲームとは、与えられたパターン列から次のパターンを推論し、かつ、パターン列への働きかけを行なうことによって自らの推論形式を変形させていくゲームである。この変形規則がどのような論理構造の上に行っているかが研究の対象であり、人間の思考過程をたえず追跡しうるシステムが必要とされる。

(5) 小グループ内の相互作用の分析

人間が人間を教育するとはどういうことか。個別教育がよいのか集団教育がよいのかなど、教育の現場に計算機を持ち込むまでに検討しなければならない重要な問題がいくつかある。このことははじめにも指摘した通りである。ここでは、これらの問題に関連して、G.Pask らが行なった小人数の情報交換システムにおける相互作用の働きについての分析の問題をとりあげる。まず、人間というものを非定常なものと考え、まわりとの情報交換によって自分の内部状態を変えていく自己組織系としてとらえる。問題は、このような人間が何人か、ある限られた環境の中におかれたいわゆる「社会」システムを考えたとき、この「社会」が有効な活動を行ない、安定なシステムとして存在するための条件は何かということである。Pask らは、3人の被験者が制限された環境条件の中で、帰納的推論の能力を高めていく実験を試みている。

小グループの人間同志をむすびつける情報のチャンネルをコントロールすることによって、グループ内部の結合状態、相互作用を変化させて、グループの全体としての活動がどのように変わるかを研究し、このことによって、個人にとっての情報とグループにとっての情報とがどのように

関連しているかを調べることは、教育システムのあり方を考えるうえにも役立つことが期待される。

計算機を用いた心理学実験はここに述べた以外にもすでに多くの試みがある。たとえば人間が判断を下すときの思考過程をペーズの定理を基礎にしたモデルで分析しようとする研究がある。(A. Rapoport, 1964) また、心理学の分野で定量的な研究が非常にむずかしいとされている概念形成の問題に対し、コンピュータで制御される実験が計画されたことがある。(Weinberg, 1965) さらに、心理物理実験や条件反射形成の実験は、コンピュータの導入が容易であり、早くからいくつかの研究が報告されている。

4. あ と が き

教育を計算機の助けを得て行なおうとする理由の一つは、はじめにも述べたように、経済性を度外視すれば、個別教育の良さがとりもどせるということである。現在のようにクラス単位の教育では、生徒のもっている初期能力、学習の速さ、学習のタイプのちがいを無視せざるをえないが、個別的指導が可能になると、うまく学習させる確率が増すことが期待できる。すなわち学習の経過や成績をみて、次にあたえる問題の程度や内容を適切に選択することができるからである。しかし、これは、教育のプログラムがきちんとしている場合にだけいえることであって、そうでないときにどれほどの期待がおけるかはまだ未知数である。特に、計算機が教師の代りをしたときに、相互作用の性質がどのように変わってしまうものか、集団教育あるいは小人数の複数教育を個別教育にしてしまうときに失なわれるものは何かなどが問題となる。教育機械がすでに

多くの現場で利用されているが、問題はそこに組み込むプログラムであり、そのための明確な指導原理の確立が望まれているのであるが、現状では、教師の経験と直観に依存する部分が多い。そのような意味で、人間の思考過程の特性、小人数グループでの相互作用の役割などが明らかにされることが望まれる。

現在、われわれの一人、田中は、初等幾何学の定理の証明を生徒がどのような手順で見つけていくかを分析しうるソフトウェアを開発している。これは、証明にいたる道すじがいくつかあって、しかも、きちんとわかっている場合である。次には、予期しない手を生徒が考え出した場合にも対処しうるようなソフトウェアを開発し、人間の思考過程にまつわるあいまいさを少しずつでもなくしていくように努力している。

参 考 文 献

1. Pask, G. (1962): Interaction between a group of subjects and an adaptive automaton to produce a self-organizing system for decision-making.
Self-Organizing Systems. M.C.Youits et al. Editors, Washington, Spartan Press (p.p. 283-311)
2. Pask, G., (1964): Man as a system that needs to learn. Advances in Cybernetics, vol. 1, S.Beer, Editor: New York, Academic Press.
3. Rapoport, A. (1964): Sequential decision-making in a computer controlled task. J.Math, Psychol., p.351-374.

4. Weinberg, G.M.(1965): Experiments in problem solving, Unpublished doctoral dissertation, Univ. of Michigan.
5. Uttal, W.(1967): Real-time computers, Technique and applications in the psychological sciences, New York, Harper & Row, Publishers.

常駐 FORTRAN コンパイラの開発

東京大学大学院 武 市 正 人

〃 〃 米 沢 明 憲

当研究は、昭和45年秋に、日本経営情報開発協会のもとで公募された「教育実習用FORTRAN コンパイラ開発コンクール」への応募作品に関するものである。

コンクール実施にいたる経過、その目的等については、すでに報告されている（〔1〕，〔2〕）ので、詳細は割愛する。

従来、ミニコンピュータ（1語16ビット、主記憶4096語程度）のソフトウェアとして、メーカーから提供されるFORTRAN コンパイラは、そのほとんどが、目的プログラムを一度外部記憶（多くの場合は紙テープ）に出し、コンパイラの翻訳を行なう部分に格納する（あるいはコンパイラの実行処理部を格納する）ことによって実行に至るものであった。このような、いわゆる2パス方式は、比較的小さいプログラムを大量に処理しようとする教育実習には適切でない。

FORTRAN のソースプログラムの入力から翻訳、実行までを1パスで行ない、しかも次のプログラムを処理する際に、コンパイラを再び収納する必要のない「常駐コンパイラ」は、教育実習には欠かせないものである。また、このようなコンパイラは、ミニコンをパーソナル・コンピュータとして簡単な計算に用いるときにも、取扱いが容易であり、有用なものである。

ここで開発したコンパイラは、松下通信工業のMACC 7/S 4K 語用のものである。

MACC 7/S には、メーカー提供の常駐コンパイラが備わっている。
(ミニコン各機種にメーカーから提供されているFORTRANの比較が
〔3〕にある)。しかし、このコンパイラは、機能が小さく、使用上も不
便な点があるように思われた。コンパイラ開発コンクールに応募するに
際して、手元にあったこの機種を用いたが、これらの点を考慮して、全
く新しく、機能面では、ほぼJIS-3000を使用上扱いやすいように
留意して作製した。

詳しい説明書等の用意はある(〔4〕)が、ここでは、その概略を記
すことにする。

なお、機器構成は、コンクール応募要項にあるように、本体及び入力
用紙テープ読取装置(PTR)、入出力用タイプライタである。

1. 言語の機能について

JIS規格水準3000 FORTRAN とほぼ同等の機能を有する。

文字型データの入出力及び移動が行なえるように工夫されている。

また、従来のミニコン用FORTRAN(例えば松下通信提供のコンパ
イラ)に備わっていなかった点は次のような機能である。

配列は1次元及び2次元を扱う。

関数副プログラム、サブルーチン副プログラムがある。COMMONの
機能がなく、プログラム単位間のデータの結合はすべて引数による。

常駐の基本外部関数として、EXP, ALOG, SQRT, SIN, COS,
ATANが備わっている。

また、データの入力紙テープまたは、タイプライタ打鍵により行な
うので、自由書式入力である。出力は、F, E, Iの各変換と、文字出

力のためのA変換ができるがすべて固定書式である。

さらに、目的プログラムの最適化などは行なわないので、多くの箇所に一般の算術式を書くことが許される。例えば、出力並び、配列要素の添字式、DO文の制御パラメータ等。

2. コンパイラの構成と処理できるプログラムの大きさについて

コンパイラは、翻訳部(トランスレータ)と実行部(エクゼキュータ)から成る。翻訳部は、FORTRANプログラムを逐次的に、殆んど逆ポーランド記法に近い形で中間言語をつくり出す。(上で述べた目的プログラムに相当する)。実行はインタプリティブに行なわれるので直接機械語に翻訳する方法に比べて多少時間がかかるが、コンパイラ本体の大きさ、目的プログラムの大きさの点で利点がある。

コンパイラ常駐部分は、基本外部関数を含めて約3200語である。したがって、ユーザーの使用できる領域(目的プログラム及びデータの領域)は800余語である。

例えば、15元連立一次方程式を解くプログラムは処理可能である。とくに大きい配列を用いないプログラムのときは、FORTRANの文100個程度から成るプログラムは処理できる。

文番号の数、DOの入れ子の深さ等はほとんど制約とならない。

3. 使用する際の扱い易さについて

紙テープによる入力为主になることから、ソースプログラム、データともに自由書式である。入力はタイプライタからも可能である。

ソースプログラムの入力機器をプログラム途中で変更することができ

るので、ディバグ等が能率よく行なえる。

目的プログラムの再実行が可能である。

誤りは、翻訳時、実行時とも、誤りの生じた場所とともに3桁数字で種類が表示される。誤りの種類は表を引いてしらべる。翻訳時20種、実行時8種の誤りが検出される。

コンパイラ開発コンクルの対象ではないが、ユーティリティルーチンがいくつか用意されている。一つはあとの例で示す、プログラムを目的プログラムの形で格納できるようにしたプログラム編集印刷のルーチンである。また、約200語を占める基本外部関数のうち必要でないものを除いて、ユーザーの領域を拡大するためのルーチンもある。

4. 例

このプログラムは、紙テープに穿孔されたFORTRANのソースプログラムをデータとして、それを編集（行ごとに形式を整える）し、印刷するためのものである。プログラム単位ごとにページが改められ、1ページ50行の印刷が行なわれる。

下の結果は、まず、プログラムは、各行が左端に詰めて紙テープに穿孔されていたものをコンパイルし、このプログラムをデータとして与え、印刷したものである。

各種計算の例、及び、コンクル審査の際のテストプログラムは、〔2〕に紹介されている。

```

2 :
55 WRITE(3,120)A;
   IF(A-S) 51,52,51;
   LINE END;
52 IF(LINE-54) 53,53,54;
53 LINE=LINE+1;
   WRITE(3,150);
150 FORMAT(/);
   GO TO 2;
   PAGE END;
54 CALL PCLR(LINE);
   PAGE=PAGE+1;
   GO TO 1;
   END;

```

```

3 :
   SUBROUTINE INT(C,D,P,S,A0,A9,B,PAGE);
   READ(1,100) C,D,P,S,A0,A9,B;
100 FORMAT(/7A1);
   WRITE(3,110);
110 FORMAT(/7HDATE : );
   READ(2,120);
120 FORMAT(6H      );
   WRITE(3,130);
130 FORMAT(/7HPAGE : );
   READ(2,140)PAGE;
140 FORMAT(17);
   WRITE(3,150);
150 FORMAT(/1H-/////31H*** FORTRAN SOURCE PROGRAM*** );
   WRITE(3,120);
   CALL PCLR(5);
   RETURN;
   END;

```

```

4 :
   SUBROUTINE PCLR(X);
   DO 1 I=X,60;
1   WRITE(3,100);
100 FORMAT(/);
   WRITE(3,110);
110 FORMAT(/1H-);
   RETURN;
   END;

```

S@:09 (紙テープ上のデータ)

```

1 :
    FORTRAN SOURCE PROGRAM PRINT;
    DIMENSION X(10);
    C=' , D=8, P=0, S=SEMICOLON, B=BLANK, AO=0, A9=9;
    PAGE INITIALIZED;
    CALL INT(C,D,P,S,AO,A9,B,PAGE);
1  CONTINUE;
    PAGE BEGINNING;
    WRITE(3,100) PAGE;
100 FORMAT(///17,2H : ///);
    LINE=5;
    2 CONTINUE;
    LINE BEGINNING;
    READ(1,110) A;
110 FORMAT(/A1);
    I=0;
    22 IF(A-A0) 3,20,21;
    21 IF(A-A9)20,20,3;
    20 I=I+1;
    X(I)=A;
    READ(1,120)A;
120 FORMAT(A1);
    GO TO 22;
    3 IF(I) 5,5,4;
    STATEMENT NUMBER;
    4 DO 13 J=1, 7-1;
    13 WRITE(3,130):
130 FORMAT(IX);
    DO 23 J=1,I;
    23 WRITE(3,120) X(J);
    IF(A-B)6,11,6;
    5 IF(A-C) 6,7,6;
    COMMENT LINE;
    7 WRITE(3,140);
140 FORMAT(1X,1H',6X);
    11 READ(1,110) A;
    GO TO 55;
    6 IF(A-D)8,9,8;
    CONTROL S;
    9 WRITE(3,120) A;
    GO TO 54;
    8 IF(A-P)50,10,50;
    CONTROL O;
    10 WRITE(3,120) A;
    CALL PCLR(LINE);
    STOP;
    STATEMENT BODY;
    50 WRITE(3,160);
160 FORMAT(8X);
    GO TO 55;
    51 READ(1,120) A;

```


引用文献

- (1) ミニコンピュータにおけるFORTRAN コンパイラ開発コントロール報告書，日本経営情報開発協会，昭和46年12月3日。
- (2) 森口繁一，「教育実習用ミニコンピュータのコンパイラ開発コントロールについて」，第13回プログラミングシンポジウム報告集，1972年1月。
- (3) 安楽省吾，「ミニフォートラン」，bit 臨時増刊ミニコン，1971年12月，共立出版。
- (4) 武市正人，MACC 7/S FORTRAN（文法，解説，使用法）1971年9月。

請求 番号		経 46-38		登録 番号			
著者名 日本経営情報開発協会							
書名 第2回(昭和46年度)研究奨励金 受賞研究の概要							
所属	帯出者氏名	貸出日	返却 予定日	返却日			

Japan Computer Usage Development Institute

財団
法人 日本経営情報開発協会

東京都千代田区霞ヶ関 3 丁目 2 - 5 霞ヶ関ビル 30 階 TEL 581 - 6 4 0 1 (代)