

## 人文科学とコンピュータ

杉田 繁 治

国立民族学博物館

大阪府吹田市千里万博公園10-1

情報処理学会において「人文科学とコンピュータ」の研究会が誕生すべき必要性和意義について述べる。そして人文科学の分野における研究の特徴と、具体的にどのような事柄にコンピュータが利用されるのか、また利用者はどのような要求を持っているのかを明確にし、人文科学の分野における知的生産の技術としての情報処理の特徴と考慮しなければならない点を指摘する。

さらに人文科学の分野でコンピュータをもっと活用できるシステムにするにはどうすればよいか、非専門家にも使いやすいフレンドリシステムとは何か、コンピュータシステムに今後期待されている問題は何かについて述べる。

## COMPUTER AND HUMANITIES

Sugita, Shigeharu

Expo Park, Senri, Suita, Osaka 565, Japan

The necessity and importance of establishment of a new research group "Computer and Humanities" in the information processing society in Japan are described.

Several characteristic features concerning computer use in humanity studies are compared with other use such as in engineering field. An important concept 'user friendly system' are discussed. And also some problems which must be taken into consideration when computer systems are developed as tools for humanity use are listed

## 1 「人文科学とコンピュータ」研究会の誕生

「人文科学とコンピュータ」という研究会が情報処理学会において誕生した。それがどうしたのだといわれるかもしれない。しかしこれは大きな変化の現れなのである。

学会ができてからほぼ30年になり、会員も3万人近い大学会ではあるが、今迄はどちらかという工学色が強かった。情報処理というよりもコンピュータ技術を推進することが中心であった。ハードウェアにしろソフトウェアにしろそれを作る側からの問題提起が多かったのではなかろうか。

コンピュータで何をやりたいかではなく、現在あるコンピュータでは何が出来るかという発想が強いものであった。言語の翻訳や文字認識、音声認識などパターン認識の問題も、人工知能としてのコンピュータの開発をねらっていた。

その発想を変えてユーザの論理からコンピュータに関わるシステムを考えようという動きが出てきたのである。つまり従来の方法がトップダウン的であったのに対してボトムアップ的アプローチへと変化したのである。この変化を生み出したのはTSS（タイムシェアリング）技術、広域ネットワークシステム、そしてマイクロコンピュータの出現であろう。これらによって従来コンピュータ専門家のみが扱っていたのが、利用者層の拡大によって非専門家も参加できるようになったのである。

### <コンピュータの大衆化>

米国においては1960年代の後半からその傾向が見えているが、日本では1970年代の後半からであろう。約10年の差がある。しかし最近10年の間にワープロやパソコンが急激に普及しだして一般大衆にとってもコンピュータがみじかに感じられるようになってきたのである。

そこで利用者からの要求や不満があちこちから出るようになった。一般社会でもコンピュータに対する関心が深まり、それだけ意識も高まってきた。欧米ではタイプライタに親しみがあるからキーボードが自然にコンピュータにつながっていくが、日本では漢字など多数の文字がありそれをどうコンピュータに入力するかで大きな壁があった。しかしかな漢字変換の方法によって漢字なども容易にコンピュータに入力できるということになって爆発的に広まってきたのである。

コンピュータの利用層の大衆化によって、今迄にはなかった問題が発生した。それはコンピュータに関する利用方法について十分な知識を利用者が持っていることが期待できなくなったということである。しかし一般の人に対しては一から使用方法を勉強しなければ使えないというのでは普及はおぼつかない。エンドユーザにとっては自分のしたいことの方に注意をすればよいということではなければありがたいがない。テレビは単にチャンネルを押せばよく、デンプも番号を押せばよいというようになっている。

### <利用者を中心にした文明>

そこでいかにして一般大衆あるいは専門外の人にとっても使いやすいシステムにするかということが非常に重要な問題になってきたのである。このようなことは必ずしもコンピュータだけに限らない。自動車やコピー機、家電製品など不特定多数の人が関わる機器については要求される機能である。またエレクトロニクス関係のみならず政治、経済における制度や、都市の構造、建造物、また本や雑誌などすべてについて利用者にとって分りやすく、安全であり、精神的に落ち着いた気持ちで対応できることが求められている。このよ

うなものの蓄積、それが文明というものである。したがってフレンドリを追及することはより高度な文明を指向することにもなる。

## 2 人文系からの発想

コンピュータの利用として日本では文科系は遅れていた。それは日本の人文科学がどちらかという論理実証的というよりは思弁的な方法を用いることが主流であったからであろう。しかし最近コンピュータの一般大衆への普及につれて、人文科学の世界においてもコンピュータを使い出す人が増えてきた。大型コンピュータではなくパーソナルコンピュータは個人主義の指向が強い文科系のひとにとってなじみやすいものとして受とられているのではなかろうか。

このように広いユーザ層の出現によって情報処理の分野も新しい変化に対応しなければならなくなってきたのである。しかしコンピュータのように様々な分野に活用されるものに対していちいちその分野特有の問題のみを対象にして議論していても、それはアカデミックな学会活動としてはどうかという面もあろう。確かに個別の問題としては意味が薄い、かなり大きな集団に共通している問題があるとすればそれを明確にするのは意義のあることである。

情報処理学会を狭い意味に考えるのではなく、例えばそれは総合百貨店であると考えることができよう。いろいろな対象をカバーしていることが望ましいのである。特に情報処理というものは一般論ではなく実際に情報が発生し、それを処理する現場特有の問題を抜きにしては意味がない場合が多い。そういう現場感覚を持ちながら、しかし学問的に面白い問題を発掘していくことが重要であろう。

人文系はどちらかというコンピュータや数理に弱いという気持の人が多い集団である。その分野の人がどのような目でコンピュータを見ているか、またどのような問題を抱えているかを知ることは作る側にとって有用な情報となるであろう。またそれによって新しいシステムが作り出されるとしたらユーザにとっても大変ありがたいことである。このような両者を結ぶチャンネルが今迄制度として用意されていなかったのが不思議なくらいである。やっこのたびそれが実現する方向に一歩 進んだのである。

## 3 人文科学におけるコンピュータ利用の特徴

コンピュータ利用の点からいって、文科系とか理科系とかの区別はあまり意味がないかもしれない。むしろコンピュータに関する経験とかによるところが大きいのであろう。これからの若い人々が 1970年代からは大分 進歩するかもしれない。しかしこれまでは文学、社会学、歴史学、地理学、民族学などにおいても積極的に使われた例はあまりない。たかだか統計処理を中心とする問題や、文献検索、テキストの入力およびインデックスの作成などに応用されているにすぎない。

それは人文系の研究者は数学や機械に弱く、コンピュータに対しても拒絶反応を起すからだとわれていた。しかしそれは利用者側の問題ではなく、コンピュータシステムの開発者側に問題があったのではなかろうか。今までコンピュータによる情報処理の中心をなしていた人々は、ハードウエアやシステムソフトウエアの専門家ではあっても、必ずしも具体的な情報処理を必要としているユーザではなかった。道具を作る立場の人が主で、で

きた道具の誤りがないかどうかを試す程度には使用するが、実際には使用していない人が多い。そのような人にはプログラミングや操作手順などはなんでもないことであろうが、エンドユーザからすればあまりにも多い強制的な約束事に対する反発がある。

### ＜コマンドレベルの整備＞

ものには階層ということがある。あるいはユニットといえばよいかもしれない。そのユニットは一定の機能をもっているが、その内部がどのようなになっているのか、そのユニット以下のレベルに対しては知らなくてもよいというようになっていると分かりやすい。テレビがどのような原理で出来ているのか知らなくてもブラックボックスとしてTVがある。電子レンジも時間の設定の仕方のみ知っていればよい。ワープロ専用機は電子手帳などと同様ある目的にとって必要な機能のみを分かりやすくしたキーで表現しているので普及している。

ところがコンピュータでわれわれが日常的に行なっているような知的生産の作業、写真を整理したり、AV資料を見たり聞いたりしていることをコンピュータでやろうとしたら一から開発していかねばならないのが現状である。コンピュータに対する一般的なことはOSの発達によって随分楽になった。OSのよしあしはどのような機能をユニット化しているかによる。CPUの命令体系にしてもRISC（縮小命令コンピュータ）ということが出ている。あまり細かいコマンドではかえって使いづらい。かといって大まかなコマンドでは小まわりがきかない。

知的生産のためのコンピュータとしてどのようなコマンドを用意すればよいか、応用的な事柄に関することはこれからの課題である。現在はまだ分かりやすいコマンドレベルでイメージや音響などが手軽に使えるようにはなっていない。パソコンレベルでのマルチメディアシステムの開発が待たれているところである。

### ＜マン・マシンシステムの必要性＞

今ひとつ人文系でコンピュータ利用があまり活発でなかった理由をあげるとすれば、その研究内容が実際問題として、コンピュータで処理できないような性質のものが多いということではなかろうか。しかしそれは問題全体としてすべてをコンピュータ化できないというだけで、部分部分には利用可能なのであるが、人と機械がうまく共存して問題を解決していくようなシステムの開発が出来ていなかったのである。

とくに民族学のような総合的な分野では、その対象として取り扱う情報およびその媒体の種類は、極めて多岐にわたっている。数値や文字情報のみならず、写真、スライド、フィルム、音楽、音声、物、など、しかもそれらから新しい情報を引だし命題を構築するのであり、情報処理は最終目的ではない。おのずとそこに工学的情報処理の開発目的と、人文系での利用者に対する情報処理技術には異なった観点が出てくる。

### ＜コンピュータ利用の比較＞

工学系と人文系におけるコンピュータ利用の特徴を対比させると次のようなことがある。

情報処理	v a	知的生産
解析	v s	統合
大量少種類	v s	少量多種類
汎用	v s	個別
多重利用	v s	単発利用

技術系においては実験などから大量に集められたデータに対して統計処理などを施してある一つの量に集約させる、あたかも大量のゴミを処理するようなやり方に対して、人文系では観察された極く少ない情報から人間生活の深層構造を明らかにしようとする知的活動が中心であるという違いがある。工学系がデータの解析を行なうのに対して、人文系では様々な種類のデータを総合的に考えることが重要であるという違いがある。実験系ではデータを変えて同じプログラムを何回も使うのに対して、知的生産においてはある性質を検証すればそのプログラムはもはや用はないという、使い捨てに近い場合が多いという違いもある。

#### 4 フレンドリシステム

人文科学研究の分野におけるコンピュータ利用については、人工知能（AI）というよりはいかにして研究者の持っている知識を引き出すか、あるいはばらばらな情報を統合させて新たな知識を生み出すか、そのための道具の開発が重要である。コンピュータが出現するまでの長い間、人は手作業でいろいろなことをやってきた。文献を読み、カードを作成し、計算やグラフ化を行ない、写真をファイルしたり、そのようにして仮説を頭のなかに構築したりしてきたのである。このプロセスをコンピュータを活用して行なうことが出来ないかというのが一番の問題である。

文科系の研究者はどちらかといえば機械類を嫌うタイプの人が多い。恐怖感を与えず安心して自然なやり方で使えるようなコンピュータシステムの開発が望まれているのである。機械が勝手に判断して何かをやるようなタイプのAIよりも、何かやりたいことが自然な操作ですぐ出来るようなフレンドリシステムが求められているのである。

その内容は必ずしも高級なことがらではない。しかし大量のデータから必要なものを捜すということは単純ではあっても実際に役に立つものである。特に画像や音響データなどの場合、単に所在情報だけではなくその内容そのものが映像や音響として目や耳で聞くことが出来れば知的生産にとって大いに役立つ。文字情報だけではなく映像、音響などが総合的に関係付けられたマルチメディアシステムの構築が重要となるのである。

#### 5 コンピュータに望むもの 〈こんなものいらない〉

コンピュータが日本に導入されてからもう数十年になるが未だに当初からの形式に引きずられており、新しい展開ができていないように思われる。それは専門家だけのものであったからで、これからは急激な変化を示すに違いない。一寸気が付くことだけでも次のようなことがある。

- 面倒なキーボード
- 役に立たないマニュアル
- 不自然な形式での対話
- 一覧性を殺す小さな表示部
- 遅い画像情報処理

コンピュータの知識のない人はいくらそのマニュアルを読んでもどうしたらよいか分らないものである。すでに他のコンピュータを使ったことのある人でも他社の物だと分らな

い所が多い。本当に素人にも分るマニュアルというのはほとんどない。誰かに説明してもらおうと分るのにマニュアルだけでは分らないというのはどこかに欠点があるのであろう。本当はマニュアルなどなしでも使えるようになっていなければいけないのである。

### <何が欠けているのか>

現在のコンピュータシステムは満足できるものになっているであろうか。ゲームやワープロにおいては多くの人がある程度満足しているが、いざ実際に今迄人間がたやすくやっていたことをコンピュータの下で自動化しようとするとうまくできない問題があまりに多いのに気づくであろう。

コンピュータは先ず数値や文字を扱う道具として発達してきた。しかしわれわれの日常生活では画像や音響情報が普通である。にもかかわらず現在のコンピュータは画像や音響の扱いに向いていない。文字や数値の延長線上から考えているかであろう。むしろ画像から発想して新しいシステムを開発すべきであろう。

コンピュータの出現した1950年頃から既にパターン認識や学習などが研究されたが、結局まだたいしたことはできていない。最近再びAIということが話題になっているが、いかにハードウェアが進歩したとはいえ人間の知的活動の機構についての知識はまだあまり進展していない。

せまい意味での問題にしばってみても現在のコンピュータには欠けているものが多い。  
例えば

文字、画像、音響の自然な入出力

関連情報の同時表示

五感との連動

運動感覚と知覚感覚との連動

予備知識なしで使えるシステム

など。これらを解決しなければフレンドリとはいえない。

## 6 人文系コンピュータの環境整備

<マルチメディアシステム> 人文系では数値化することがむづかしいデータが多い。写真や、考古発掘資料、美術作品、音楽、音声などそのまま対象として扱うことが普通である。現在のコンピュータではそのままでは扱えない場合があり、ましてやパソコンレベルではできない場合が多いので、やむなく文字・記号で代理させてやっているのが現状である。しかし人文系の研究には知的生産の道具として、直感的なヒラメキを刺激するような情報を提供するシステムが必要である。この点からも本来人文系のコンピュータシステムはマルチメディアシステムであるべきなのである。

<利用の実態調査> 個人の利用はその存在がなかなかつかめない。成果が発表されれば表面に出ることもあるが、通常積極的に公表されることがない。今自分がどのようなデータを入力し研究しようとしているかを他人に知られるのを嫌がることもあろう。あるいはデータに誤りがあるかもしれないのでよくよく吟味してからにしようと考えているかもしれない。もしそうでなくても、現在の状況ではどこに申し出ればよいのか、その受け皿がない。大学の大型センターの利用者ならばその機会はある。しかし文科系の個人研究者はパソコンでやっているケースが多く、大きなプロジェクトなら

いざ知らず、ほとんど人に知られることがない。このたび誕生した研究会をコミュニケーションの場として活用されるよう運営しなければならない。

**< データアクセスの問題 >** 利用者が各種のパソコンを持っていると、その通信方式、特にファイル転送や文字種、コードが問題となる。オンライン的に結合されていない場合では、フロッピーディスクのフォーマットの違いによる互換性がとれないケースがある。またフロッピーも各種のサイズ、密度があり、媒体自体の互換性がとれないから、オンライン的にデータ交換を行なう方が容易になる。はたして個人レベルでのデータフォーマットの標準規格が必要か否か、検討をしなければならないであろう。

データフォーマットについても同様のことが言えよう。単純なレコード長やブロック長はあっても、その中に使われている種々の特殊記号などの説明がなおざりにされている場合が多い。コンピュータのデータとしてはリストの形式であろうと、磁気テープやフロッピーの形態であろうと、その媒体の中にデータに関する記述が含まれていることが望ましい。画像データの場合には標準的なヘッダーフォーマットを定めているが、テキストデータに関しても、標準規格が必要か否か、検討をしなければならないであろう。

**< 多種文字の入出力 >** かな漢字変換によって漢字の入力は可能になったが、ローマ字や漢字だけではなく世界に多数ある諸言語の文字が容易に入力できる環境を作っておく必要がある。アラビア、ハングル、タイ、デバガナリー、チベット、ビルマ、などなど。人文科学のコンピュータ応用では、テキストの入力がかなり重要な問題になる。いろいろな文字が簡便に入力できるとまた新しい利用が生まれることになる。

**< コード体系 >** 世界の言語をコンピュータで扱えるようにするためには統一なコードが必要になる。その字種を全てカバーしようとするれば少なくとも3バイトは必要である。しかし通常は同時に多種類の言語を混在させて使用することは稀であろう。そこで3バイトコードのようなものを作るよりは、言語切換えソフトコードを設定して、各々の言語はそれだけで独立しているようにしておくのがよいのではないかと思う。

コード体系で問題となることは、ASCII, JIS, シフトJIS, EBCDICなどコンピュータにより異なるシステムがあることと、外字の問題である。例えばJISの第一水準、第二水準までは、その字種の選択にいろいろ問題があるとしても一応共通基準となっており、各種の装置がサポートしている。しかしそれ以外の字種が全く統一が取れていない。国文学や漢籍などでは異体字や偽字なども含めて約1万5千字から3万字が必要といわれている。それぞれの機関で必要に応じてJIS以外の文字パターンを作っているのが現状である。しかもそのコードの割当方法がそれぞれ勝手であるから、データ交換のための互換性が取れない。このようなまだ認知されない状態のものに関して、「人文科学とコンピュータ研究会」において検討していかねばならない問題であろう。

**< 自然な入力 >** 従来コンピュータへの情報入力はほとんどがキーボードからであった。しかし人文系のデータはアナログ的なものが多いのでそれらを直接入力できるようになっていないと使いづらい。CCDカメラによる画像の入力や、音響情報の入力も利用者は単に資料をセットするだけでいけるようなシステムが開発されなければ

ばならない。その扱いが容易であればまた新しい応用が考えだされるものなのである。必ずしも機械による自動認識でなくてもよい。録音テープやVTR、映画などの資料から、研究者が情報を選択しながら入力する時にやりやすい装置の開発が必要である。

美術館などで光ディスクを用いて収蔵品やカタログの作品を蓄積し検索している所まで来ている。テレビカメラから直接ビデオディスクに入力する簡便な装置が利用できるようになってきたからである。これを更に安価にしかつ画像の質を良くすることに勉める必要がある。ただデータとして画像が入ればよいというのではなく、写真と同じ位の質を持っていないと満足できない。

＜シソーラス＞ コンピュータを使って検索をしたり、言語を扱う場合に一番問題になることは、同類語の扱いである。人間であれば必ずしも同一の表現でなくても、意味的に似た単語であれば一致したと判断する。しかしコンピュータにはそのような常識が与えられておらず、普通の場合勺子定規な一対一の完全一致でしか検索しないから、もれてしまう情報が多くなる。これを救うにはシソーラスというものを導入することが最低必要である。シソーラスは一種の語彙集であるが、同義語関連語などを定義した辞書である。とくに民族学で使用する用語を中心としたシソーラスを開発する必要がある。これは今はやりの人工知能やエキスパートシステムの問題とも継がるもので、コンピュータにインテリジェンスをもたせる第一歩である。

ワープロの普及はかな漢字変換辞書の出現によって可能になった。人文系のコンピュータが普及するには自然言語の扱いが容易になることであり、その為にはシソーラスが用意されることが第1歩である。一般語彙、ある分野毎の辞書を作成することが是非必要である。

＜著作権の問題＞ 今までコンピュータの分野での著作権の問題は、プログラムに関するものが多かった。しかし人文系の分野ではテキストを扱う場合が多いので、コピーとか、一般の著作における引用、放送やカラオケ音楽におけると同様の問題が生じつつある。コンピュータにテキストデータを入力して、個人的に利用している分には何も問題はない。データベースとして他の人にも利用させるようになると著作権や出版権などの問題が出てくる。フロッピーで簡単にデータがやりとりされると、安易に考えがちである。

もとのテキストと同様の形式でなく、それを加工した場合でもやはり問題になるであろう。例えばテキスト全文を入力して、それを使ってKWIC索引を作ったとしよう。それは形式は異なるが、元の文章を切り貼りしたようなもので、いわば行単位の引用と同じであり、著作権に抵触することになろう。しかしKWICの形式に加工したこと自体には一種の編集権のようなことが発生すると考えられよう。いずれにしても原著者や出版社に対して何等かの許可を得ないでデータベース化し、それを流布させること、あるいはそれを公表することは道義に反するであろう。個人レベルのデータベースであって、研究者仲間だけに個人的に無償で伝達されるようなデータであると、ややもすると安易に考えがちであるが、十分注意しなければならない。

今後百科事典や辞書などがCD-ROMなどで普及してくるが、さまざまな問題が生じることが予想される。