ISSN 0582－4478

京都大学附属図書館報

# 静脩

1985年12月

The Kyoto University Library Bulletin

号　外

## 電子計算機の進歩と学術情報

附属図書館長　西　原　　宏

### 電子計算機の進歩と学術情報

スライド　1

電子計算機の進歩と学術情報

電子計算機と在来技術の結合による新情報技術：

1．電子計算機と電気通信の結合によるデータ通信（INS，VAN）
2．電子計算機と機械工程の結合によるロボット技術
3．電子計算機と事務作業の結合によるオフィス・オートメーション（OA）

題はいかめしいが難しいことではなく気楽なお話をさせて頂く。

電子計算機（通称コンピュータ）が急速に進歩して，社会に強い影響を与えている。それを在来技術と計算機との結合による新情報技術という観点からまとめたのがスライド1の3項目である。

1．計算機の進歩に伴って on-line 化が進んだ（この場合 on-line 化とは通信線に接続すること）。それによって遠隔利用及び対話的（会話的）利用ができるようになった。始めは専用回線が用いられたが，国際回線を含む公衆回線に接続されるようになり，利用の範囲が地理的にも格段に拡がり，地球規模の情報ネットワークが実現した。現在おびただしい数の情報ネットワークが出来て利用されている。

電々公社（85・4からはNNT㈱）の INS とか，他の企業による VAN は連日のように新聞紙面その他に姿を見せている。

2．ロボットのする仕事は，手作業という人間らしい仕事で，機械化（電算化）が却って難しく，化学工業やセメント工業などのいわゆるオートメーションに比べて実現が遅れた。マンガに出て来るようなロボットは実在しない。

3．事務処理という人間の手仕事を対象にしているうえに自然言語（われわれの場合には主として日本語でかな漢字まじり文という独特の表記法で書かれる）の処理・取扱いを含んでおり，その難しさも奥行きの深いものである。

さて，これらの新技術における計算機の役割は，いわゆる数値の計算ではなく，むしろ情報の処理である。具体的に計算機が仕事をするのは記号の処理，記号列の処理である。言語の計算機処理もこれに帰する。

電子工学が進歩発達すると何でも出来るようになるかというとそうではなく，その時々の限界があり，その限界は言語とか文化とかに対する人間の理解洞察の深さによっており，却ってそれらを探究する学問を促すように思われる。

便利で本当に役に立つ学術情報ネットワークシステムを実現するには，差当り通信回線や機器設

本稿は，1985年2月15日（金）に開催した講演会（近畿地区国公立大学図書館協議会主催）の概要を収録したものである。

備の性能容量が不十分であるという点で整備が必要であるが，それだけではなく，上に述べたような観点と努力が大切であると思う。

いまひとつの事情としては，わが国では情報需要が，米国などの場合に比べて大変弱いといわれている点である。学術情報については高い経費を支出して情報を入手するという習慣がないというか，とに角しっかり育っていないということである。一方，情報の入手（あるいはそのための投資）が大きな利益を生むような場合には，事情が全く異なり，情報サービス業は大繁昌ということになるのである。

ここで情報についてしばらく考えて見ることにしよう。

## ロスチャイルドと情報に関する逸話

スライド　2

Rothschild 家
父：Meyer Amschel Rothschild (Frankfurt am Mein)
第一子：Amschel Meyer Rothschild (Frankfurt am Mein)
第二子：Solomon Rothschild (Wien)
第三子：Nathan Meyer Rothschild (London)
第四子：Karl Rothschild (Napoli)
第五子：James Rothschild (Paris)

スライド　3

岩波西洋人名辞典（増補版）　p.1729
ロスチャイルド
Nathan Meyer Rothschild (1777.9.16-1836.7.28)
…ウォーターローの会戦には，伝書鳩を用いていちはやくナポレオンの敗戦を知り，ロンドン市場で証券投機を行い巨富を得た。…

情報とは何か，ということを考えるとき，私のいつも想い出す，古い逸話がある。私が小学生であった頃の——今から50年位前の——古い記憶である。

19世紀の世界の金融界に重きをなしたロスチャイルド(Rothschild)家は，情報というものの重要性に早くから着眼していたという。

ナポレオンがウォーターローでウエリントン将軍の引きいる連合軍と会戦したとき(1815-6-18)，その勝敗に関する情報をいち早く入手してそれを活用して大をなしたという話である。私の記憶では，ナポレオンが勝ったという虚報を，伝書鳩を使ってわざとロンドンに伝え，自分だけがナポレオンの敗北を知っていて，大いに利を得たというのである。この逸話の根拠を京大附属図書館の資料で調べて見た結果をお話してみよう。

先ず，当のロスチャイルド家の英雄たちであるが，スライド2のように Frankfurt am Mein の Meyer Amschel Rothschild (1743—1812) は5人の息子があり，Amschel Meyer Rothschild は Frankfurt に，Solomon は Wien に，Nathan Meyer は London に，Karl は Napoli に，そして James は Paris に店をもたせた。

岩波西洋人名辞典によると，スライド3のように，第3子ネイサン（ロンドン）は，だれもナポレオンが勝ちウェリントンが敗けると思っていたときに，人の知らない，そして確率の低いナポレオン敗北の情報を伝書鳩を使って入手し，それを活用して巨富を得たとある。

記憶とは違って，人を欺くことはしていない。New Century Cyclopedia of Names, New York (1954) にも同様の事が出ている。

## Dictionary of National Biography

スライド　4

Dictionary of National Biography vol.XVII, MacMillan Co., New York 1909, pp.306-309
Rothschild は大陸に急使や伝書鳩による通信網を持っていた。
代理人 Roworth が，Waterloo の戦で連合軍の勝つたことを速報したオランダの新聞を，第一報としてロンドンにはこんだ。Nathan は直ちに英国政府に報せたが，大臣たちは信用しなかった。
ダウニング街は，後に別の情報源から確認した。

スライド　5

Rothschild自身がWaterlooからの報せをもたらし，この情報が一般に知れわたる前に，市場操作に十分な期間これを独占したというのは，作り話である。
(Nathan Rothschild は，英国の国会議員にもなった

ところが，Dictionary of National Biography にはスライド4及び5のように以上の話とは違う事の経緯が具体的に示されている。ロスチャイルドは早くから情報の重要性に着目しており，大陸に a staff of courier（急使，早飛脚に相当か？）を置いていたのみならず伝書鳩による通信組織を作っていた。

ロスチャイルドの代理人の Roworth という人がウォーターローの戦の結果のしらせを Ostend

人で，世を欺くような事はしなかったと思われる。平素からの，情報に対する組織的な努力が巨富をもたらしたのであろう。この逸話は，情報というものの特質や属性をよく伝えている。）

(Oostende と同じか？）で待っていた。そして，連合軍の勝利を速報したオランダの新聞を手に入れ，急いで英仏海峡を渡り，ロンドンに第一報をもたらした。鳩ではなく人が運んで，６月20日の朝到着した。

こうしてロスチャイルドは誰よりも早く情報を得たのであるがスライドのようにここからが前の話と違う。

ところが，このスライドを作ってから後に，現在発売中の中央公論1985年３月号を見たところ，奥本大三郎氏（横浜国立大学助教授 仏文学）が〝高貴なる畸型〟―ロスチャイルド家の人びとと博物学　という文を書いている。そのなかでネイサンの市場操作・心理操作が具体的に書いてある。伝書鳩で偽の報らせは送ってはいないが，株を売りにまわって，ネイサンがウェリントンの敗北を知ってのうえのことだと思わせたというのである。

ひよっとしたら，英国政府に報らせたが信用されないのを見て市場操作をやったのかも知れないと思えて来る。しかし，これ以上は本格的な調査をやらないとわからないだろう。後日に譲るとして先に進もう。

## 情報の経済理論

スライド　6

情報の経済理論

1．情報には機密性の有無があり，機密性のあるものは，経済的価値をもち得る。
2．確率の低い情報ほど，与える情報量が多い。
3．情報の購入者は微少またはゼロの費用で複製できる。情報は人に譲渡しても依然として手もとに残っており，情報の取引は不可逆である。（A→B→A→C≒A→C）
4．情報の有用さは他人の保有に影響される。与える影響の大きい情報ほど拡散し易く，拡散すると価値が下る。
（野口悠紀雄：情報の経済理論）

ネイサンに関する事実がどのようなものであったにせよ，この逸話は情報についてつぎのようなことを物語っていると見てよいだろう。

1．人の知らない情報は知っている人に大きな利益をもたらす
2．起り難いようなことが起った という情報（人々がなかなか信じないような情報）は価値が高い。

ということである。

このように情報には大へんな経済的価値が伴うことがあるので，情報の経済学が関心をもたれている。

野口氏のこの表題の著書にはロスチャイルドの逸話が事実に反するということが明記してある。そして，先の逸話から学んだことが，スライド６に示したように経済理論の裏付けをもってまとめられている。

情報とは何か？　経済学において無定義概念とされている。定義されていないのではあるが，その性質や価値などの属性について理解したり考えたりすることができるのである。

スライド６の１，２については先に述べた通りである。重要なのは，３．コピー，複製のことである。情報の特質の一つは微少またはゼロの費用でコピーできることである。そのときコピーされる内容が情報であるという。このことが情報の取引きに

$$A \to B \to A \to C \neq A \to C$$

という不可逆性を生み出す。* そして，現在の社会に版権・著作権についての深刻な問題をもたらしている。

４．は独占というロスチャイルドの逸話の主題である。

全体として情報はまだまだつかみどころのない存在である。

コピーと同時に記憶の内容も情報と言ってもよいだろう。

＊ 不等記号≠の左辺は，Ａが情報をＢに伝え，それをＡが一たん取り返して改めてＣに伝えたことを表わす。また右辺はＡが始めから直接Ｃに情報を伝えたことを表わしている。この式は，左辺と右辺が等価でないことを意味している。

## 情　報　量

スライド　7

情報の量
2進数による位置の表現
（4桁の例）


桁数を増す程，位置は正確に表わされる。
位置の確率分布が一様なら，すべての4桁の2進数（通信文）は等しい出現確率を持つ。確率分布が一様でなく，例えば図の左半分に集中していると，1で始まる数は0で始まるものより確率が大きい。
確率の小さい（稀な）通信文の方が確率の大きい（ありふれた）ものより情報の量が多いと考えて，それを $1/p$ （$p$ は確率）の2を底とする対数で表わす（単位ビット）。

ここで情報の量の評価について一寸考えておくことにしよう。

図の直線（線分）の上の任意の点の位置を2進数で表わす。先ず線分を2等分する点をとり，問題の点が左にあれば1，右にあれば0とする。図の場合は1である。

つぎに左側の2等分点をとり，同様に0か1かをきめる。以下これを繰返す。$n$ 回繰返すと $n$ 桁の2進数が得られ，この数は線分を $2^n$ に等分した寸法精度で点の位置を表わしている。桁数をふやすほど位置は正確に表わされる。

図では4桁の2進数 1100 が太線の部分に問題の点があることを表わしている。全体を16等分して16桁の2進数に各1/16の区間を対応させ，点のある区間の桁を1，他を0とすることによって，0001000000000000のようにして，同じ精度で点の位置が与えられるが，この場合には16桁も使っていることに注意してほしい。その2進数には1のある桁は1つしかない。あとは0である。このように0と1で表わす場合には図の方法（20の扉の方法）が最小の桁数の表現を与える。

点の位置の分布が偏っているとどうなるか。例えば図の左半分に点が偏っていると，最上位の桁が1である数の出現確率が0のものより高い。

出現確率の低いほど与える情報量が多いと考えるのが自然である（ロスチャイルドの逸話）。そこで出現確率 $p_i$ の通信文の与える情報量を $\log_2 \frac{1}{p_i}$ で表わす。単位はビット（bit）である。この表現は情報量について自然な理解認識を与える。

先に挙げた2進数が線分上の点の位置について与える情報量は，一様分布なら，各桁1ビットずつとなる。$n$ 桁なら $n$ ビットである。

情報や言語の確率統計から見た側面は大変重要である。しかし一方，その内容に立入った取扱いが増々重要になって来ている。

## 電子計算機の進歩

スライド　8

電子計算機の進歩と世代
第1世代（真空管）：
　ENIAC（Pennsylvania 大学，1946年，外部プログラム方式）
　EDVAC（同，1947年，Neumann 式，プログラム内蔵方式）
　EDSAC（Cambridge 大学，1959年，オペレーティング・システム及び標準サブルティン）
第2世代（半導体，トランジスター）：
　磁気コアメモリ，コンパイラー言語，小型化，高速化）

ここで計算機の進歩について，スライド8を見ながら古い想い出話をはじめよう。

私が京大電気を卒業したのが終戦直後の昭和20年9月であった。米国から週刊誌などが入って来るようになった。1946年（昭和21年）ペンシルバニア大学で ENIAC という真空管式の電子計算機が作られ艦砲の弾道曲線の計算をして，実際の艦砲射撃より短い時間で結果がわかるという記事が TIME か NEWSWEEK に出ていたという遠い記憶がある。そのとき深い感銘をうけ，これから電子計算機の時代が来るという予感がしたという記憶がある。京大図書館で調べて見ると NEWSWEEK の1946-2-18号に出ていた。弾道曲線のことは出ていない。この計算機は外部プログラム方式で演算・計算の内容は配線で表現する。1947-3-17号の NEWSWEEK には EDVAC が出ている。記事の内容にも写真にも記憶が残っていた。この計算機はフォン　ノイマンによるプログラム内蔵式で，計算の順序を指示するプログラムをあらかじめ読み込んでおき，その手順・命令を順次読み出しながら計算を進める。今日の計算機も基本的にはこのやり方である。これによって計算機はハードウェアとソフトウェアから成り立っているという考え方が一般的になった。

つぎのEDSACでは，計算だけでなく計算機の動作・状態などをもソフトウェアで制御するオペレーティングシステム（OS）と，多数の利用者プログラムに共通にあらわれる成分をサブルティンとして用意しておく方式がとられ，その後の計算機は皆この方式にならっている。

真空管式の欠点は真空管の寿命である。ENIACでは18000本使われた。トランジスタが出来て実用性が急激に向上したのである。真空管式の計算機ではフジフィルムの人が76というラジオ用の3極管を使い，手作りでレンズ設計用の計算機を作って実用にしたということがあった。

しかし当時（それから10年後でも）私の先生であった方々は皆，日本製の計算機が実用になる見込は全くないという意見であった。当時は一般にそう考えられていた。

## 第3世代から第5世代まで

スライド　9

第3世代(集積回路，IC)：
オペレーティング・システム（OS），ファミリー・シリーズ，TSS会話型処理，ソフトウェアの重要性，毎秒百万回の演算（MIPS）
第3.5世代（大規模集積回路，LSI）：
IC メモリ，大容量メモリ，データベース会話型処理の実用化
第4世代（超LSI）：
100 MIPS，スーパー・コンピュータ
第5世代(?)：
並列処理，非 Neumann 式，人工知能，知的インターフェース，知識ベース

スライド9について説明する。

第3世代：集積回路が用いられ，会話型処理が行われ速度が著しく速くなった。ソフトウェアの重要性が増し，ソフトウェア産業が興ることになった。

第3.5世代：あらかじめ世代の名前がきめてしまってあったのでこんな名前になった。LSI が使われ，データベースが普及した。データベースが発達したのは，記憶装置が安くなったことによる。

第4世代：超 LSI が用いられる。スーパーコンピューターと呼ばれる。逐次計算では速度の限界がある。そこで並列にすすめることの出来る演算は並列に進める。今日の実用機でスーパーコンピュータと呼ばれているものもある。

第5世代：これはまだ実現しておらず開発目標にすぎない。これについてはまた後でふれる。非ノイマン方式と言うがノイマン自身は並列方式を提案したということである。

## 最近における電子計算機の顕著な進歩

スライド　10

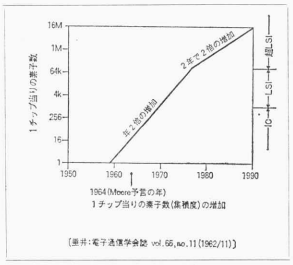
電子計算機の最近の進歩
1．回路の小型化と高密度化→集積度の増大
IC（5.6㎜×5.6㎜のチップに数百個の素子)→LSI（1000個以上の素子）→超 LSI（10万個以上の素子，回路の間隙 1 $\mu$m 程度）
集積度は，メモリ素子は1年に2倍，演算素子は2年に2－3倍の割合で増加
2．素子の価格の低下
メモリ素子は年に40%，論理素子は年に25%
3．システムの大型化，マイクロ・エレクトロニクスの発達

最近の進歩をスライド10のようにまとめておく。

要点は回路の小型化と低廉化である。一方システムは大型化と小型化の両分極を生じている。

## 集積度の増加

スライド　11



[垂井：電子通信学会誌 vol.65, no.11(1982/11)]

小型化は，具体的には，集積度の増大によって達成されているので，これをスライド11で見ておこう。

IC，LSI，超 LSI と進むにつれて，1つのチップ上の素子の数がどんどん増加した。どんなに急速に伸びたか，図に示されている通りである。この図は引用の文献を参考にした略図である。

1961年に Moore が年に2倍の増加という予言をした。この予言の通りに経過して来たのは驚くべきことである。内訳は回路の細密化，回路の改良とチップの寸法を大きくしたことである。

超 LSI の時代になってからは2年で2倍のペ

ースになるようである。このように Moore の予言通りに進行した集積度の増大と素子当りの価格の低下が計算機とその応用の技術の急速な進展をもたらした。

## 将来の素子

スライド 12

LSI，超LSI のつぎは？
計算機を構成する基本素子は，２つの平衡状態をもち，一方から他へ移るとき，オン・オフのはたらきをする。
遺伝子工学的に合成した蛋白質分子枠のうえにスイッチの役をする化合物（ポリフィリン）を，規則的に配置した分子デバイスが提案されている。

このような素子を使ってどのように計算機を構成するかという問題は全くこれからのことであろう。

計算機が進歩すると，だんだんその構造まで人間に似て来るという気もする。

## 第５世代の計算機とはどんなものか

スライド 13

第５世代の計算機
その概念は1970年代に遡る。
開発研究の野心的性格を反映して，構成が複雑である。
主な機能：問題解決と推論，知識ベース，知的インターフェイス，
適用分野：音声認識・応答システム，パタン認識，自然言語の処理
関連分野：認知科学，知識工学，人工知能，エキスパート・システム

早くから（多分1970年代の前半から）その概念はあった。

今日その主な機能として期待され，あるいは開発の目標となっているのは，スライド13のように従来の算術的演算と人工言語（プログラム言語）を基礎とする計算機とは違って，推論，知識の蓄積と活用，音声や自然言語による人間（非専門的人間）との交信などが考えられている。

第５世代計算機の研究開発は直接計算機に関係しない他の分野の研究の影響を受け，相互にかかわり，互に他の発達をたすける関係にある。

人間の知識の情報量は従来の計算機とは比較にならないほど多く，そのことを考えただけでも第５世代の計算機がいつになったら出来るのか，見当がつかない。

先に見たように，計算機には，マイクロ・エレクトロニスの爆発的な発達進歩と，第５世代計算機のように緩かな進歩とがまざっている。どの分野にもこのようなことはあるだろう。

## プログラム言語

スライド 14

プログラム用人工言語
多数のプログラム用人工言語が作られ，研究されてきた。
主なものを挙げると，
機械語，アセンブラ語，FORTRAN，ALGOL，PASCAL，BASIC，COBOL，APL，LISP，SMALLTALK，PROLOG
などである。

FORTRAN に代表されるプログラム言語は，その内容が指令的(instruction)で，プログラムによって制御の流れを示すという性格をもっている。その意味で手続き的言語（これは判り難い表現）という。

Lisp などの新しい言語には，これと違って，制御の流れというよりは関係を記述するという性格をもつものが多い。その意味で非手続き的で記述的な言語と呼ばれる（これも判り難い表現）。

Lisp は１種類の実行文すなわち関数の呼び出ししかもっていない。1950年代の終りごろ，当時 MIT にいた John McCarthy によって作られた。

記号論理学を基礎とした新しい言語である Prolog は1972年フランスのマルセイユ大学で作られ，後に英国のエジンバラ大学で新しい処理系が作成された。このように最近はヨーロッパでこの方面の研究が盛であるということである。

日本経済新聞昭和60年２月10日(日)朝刊につぎのような記事（抜粋）が載っていた。

国際経済経営会議レポート②
1984-11-20～22（東京）
1984 International Conference on Economics of Management
京大から坂井・長尾両教授が出席，長尾氏が司会
MIT から Marvin Minsky 教授が出席。
Minsky 氏は日本の５世代計算機の計画が論理型の Prolog を採用したのは誤りと主張。
Minsky 氏は Frame 理論で有名で，Frame と

いう知識の単位を基礎にして人間の知的能力を説明しようとしている。

## 情報検索とインデクシング

スライド 15

情報検索システム

A．検索しようとする情報の種類
　A－1．事実検索：データや情報の検索
　A－2．文献検索：所要の情報を含む文献
B．検索の方法の種類
　B－1．書誌的事項や項目による検索
　B－2．キーワードによる検索
　B－3．意味内容による検索

図書館は大量の情報（現状では主として言語で書かれ表現された情報）をあつめ，これを便利に提供する機関であるという風に見える。

収書に加えてスライド15に示したような索引を作りこれを検索する方法便宜を備えるという事が大切である。

一方では browse（自由閲覧，自由検索）ということを忘れてはいけないだろう*。検索が機械化（電算化）されると，それが出来なくなるという風な心配があるようで，この点は恐らく説明不足にも原因があるだろう。

＊ 名前を覚えなくても情報をとり出せるシステムを browser という。

スライド 16

文献検索・
インデクシング言語

情報検索は索引の作成（整理）と捜索から成る。
資料整理のための意味を表現する道具がインデクシング言語であり，分類法によるものと用語法によるものとがある。
用語法には，統制付き用語による方法（件名標目による方法，シソーラス法）と文脈付き用語法(KWIC, KWOC, KWEST (-IC, -OC) PRECIS など）がある。
分類法と用語法の合体も考えられる。
中村幸雄　情報管理
Vol. 27 No.5 1984-8

文字列を取り扱うときの基本的な問題に整列（sorting）がある。整列とは，与えられた要素の集合を，ある定義された順序にならべることをいう。

自然言語処理のための整列の応用として KWIC は重要である。これは文の集合（文章またはテキスト）が与えられたとき，文中に現われるすべての単語をとり出し，これをアルファベット順に整列させ，その単語の含まれる文を文脈として付加して出力するものである。Luhn が初めて作った。ひとつの文のなかに$m$個の単語があると，この文は$m$行の KWIC 表を占める。すなわち KWIC 表に$m$回現われる。

各単語がそれぞれ窓のKWを示す位置に来るように文の位置をずらして出力する。各単語をKWを示す窓に示し，そのつぎにその単語の表われる文を出力すると KWOC となる。

図書館の，例えばインデクシング関係の資料や，人工知能関係の資料等もそれを専門としない人にとっては難読難解である。整理課長の話では KWIC は図書館職員なら誰でも知っているということであるがこれを「文脈付き見出語」であると書かれては，読んでも仲々わからない。

図書館の職員の人たちには自分で興味をもつ問題について個人的に深く勉強して相当のレベルに到達している人が多いが，図書館全体が組織としてそのレベルの仕事をするためには，例えば，図書館全体で，関係の課で，あるいは担当の係りで図書館の仕事に活かすという観点から勉強・研修を継続的に行うことが大切ではないかという気がする。

それに加えて，図書館のかかわる仕事の内容(例えばインデクシング)を解り易く説明するという努力も必要ではないか。例えば入門用の資料を作成することが必要ではないか。

解り易いかどうかはその人のバックグラウンドにもよるであろう。

## 自動インデクシングと言語情報処理

スライド 17

自動インデクシング

Ⅰ．表現形式に基づく検索用ファイルの自動作成
(1)法令の参照関係　(2)文献の引用関係
Ⅱ．日本語の文構造解析による索引の自動作成
(1)文節構成語の認定とキーワードの自動抽出　(2)構造解析によるキーワードのロール付与　(3)確認・修正のための会話型校

検索のためのインデクスを作成することは，図書館にとって重要であるがまた骨の折れる仕事である。これを自動化するという考え方がある。その内容はスライド17のようなものである。

正
Ⅲ．不要語の除去によるキーワード自動抽出
(1)かな漢字まじり文の自動分かち書き (2)自立語の認定 (3)不要語の除去
長尾真 日本語情報処理 第9章

自動インデクシングは，必然的に言語処理を伴う。かな漢字まじり文で表現される日本語の処理には固有の難しさがある。勿論或程度の手がかりはある。例えば重要な語は大てい漢字かカタカナで書かれており，ひらがな書きの部分は相対的に重要度が低いという経験的事実があるそうである。

スライド 18

日本語情報処理

日本語情報処理あるいは広くこれからの情報システムを考えるには，計算機のハードウェアとソフトウェアに加えて，言語情報そのもの(知識ウェア)が必要である。真に有用なシステムは豊な知識をそなえ，それを適切に提供できるものである。これからの目標は，このようなシステムの構築である。
長尾真：自然言語処理と機械翻訳

言語の計算機処理にはよく文脈自由型文法が用いられるが，それは文の意味に立ち入らず，文の構造を句構造の規則（書き換え規則）を中心として，文の全体から細部に向って規定して行く文法である。これに対して文脈規定型文法もあり，いずれも専門外の者にとっては難解である。

物語を，その内容についての質問に答えるという意味で理解するためには尨大な常識が必要であることが指摘されている。長尾氏は，知識を計算機に与える方法として，学習機能の重要性を強調している。この関係の研究はその長い歴史にもかかわらずまだ殆ど進んでいないということである。計算機が（あるいは人間が）ある文をよんで理解するというのはどういうことなのだろうか。理解のあかしとして，読んだ文に関する質問に応答できることや，つぎにあらわれる文について予測できることなどが挙げられるだろう。

スライド 19

プロダクション・システム

人間がもっている知識をどのようにコンピュータにのせて表現するかという問題に対して，プロダクション・システム〔2〕という普遍的に使えるような形式を作り出した…〔1〕
〔1〕長尾真 ぇれきてる 1984—14
〔2〕A. Barr & E.A. Feigenbaum 編，田中幸吉訳
人工知能ハンドブック 第一巻，共立出版 (1983)

スライド19に示したように，人間がもっている知識をどのようにコンピュータにのせて表現するかという問題に対してプロダクション・システムという普遍的に使えるような形式が見出されているということである。

ただしこの方法はある種のトピックスについてはすばらしい能力を発揮するが，全体としては人間よりはるかに低いレベルにあるとされている。

## プロダクション・システム

スライド 20

プロダクション・システムの構成

プロダクション・システムは次の3部分から成っている。
1．ルール・ベース：プロダクションの規則
IF X (条件部, LHS)
THEN Y (実行部, RHS)
2．文脈：文脈リスト (CL) の短期メモリ
3．インタプリタ：システムの動作の制御（次に何をするかを定める。）
田中幸吉訳 人工知能ハンドブック

ここでプロダクション・システムについて，ごく簡単な説明をしておくことにしよう。

プロダクション・システムはある種の問題解決のモデルになり得るとされているが，そこでは人間の知識はルール・ベースの形で与えられている。

スライド 21

プロダクション・システムの例題

ルール・ベース
P1. IF On-CL green THEN Put-On-CL produce.
P2. IF On-CL in box THEN Put-On-CL cake.
P3. IF On-CL refrigerated THEN Put-On-CL perishable.
P4. IF On-CL perishable AND On-CL 7kg THEN Put-On-CL turkey.
P5. IF On-CL 7kg AND On-CL produce THEN Put-On-CL watermeron.
実行：

スライド21に示した最も簡単な例題を考えて見よう。

ルール・ベースはつぎのように構成されている。

P1. もし文脈リスト（CL）に緑色が含まれていれば作物をCLに入れる(CLの先頭に追加する。以下同様)。
P2. もし包装がCLに含まれていれ

CL＝（green, 7kg）→
（produce, 〃）→
（watermeron, 〃）

ば菓子をCL'に入れる。

P3. もし冷蔵がCLに含まれていればくさるをCLに入れる。

P4. もしくさる及び7kgがCLに含まれていれば七面鳥をCLに入れる。

P5. もし7kg及び作物がCLに含まれていれば西瓜をCLに入れる。

最初に緑色と7kgがCLに含まれる状態で，適用可能なルールのうち最も番号の若いものを，ただし重複記入を避けて実行すると，最終的には西瓜という答がCLの先頭に得られる。われわれの知識はルールベースの形で，答を導くのに役立つ訳である。

## 日本語について

スライド 22

手書き文字（漢字）
の機械読み取り
1．文字パターンの輪郭をしらべて，候補文字を選ぶ。
2．手書き文字のくせを消去し，各候補文字の標準パターンとの類似度を計算し，検定する。
3．複数の候補が残っているときは，前後の文字を見て単語として，その意味から決定する。
例えば，地名ファイルを参照して確定する。
ゑれきてる 1984—12

われわれが学術情報の問題を考える場合に直面する問題は「日本語の機械処理」である。欧米語を主な対象として進められている研究の成果が必らずしも直ちに適用できない。

先ず，日本語文字の機械読取りはどんなことをするのかを見よう。印刷文字の読み取りは，例えばつぎのように進められるだろう。

イ．文字パターンの走査：印刷された文字を白黒のパターンとして $m \times n$ の目に分けて走査すると0（白），1（黒）の列が出来る。

ロ．前処理・正規化：活字の型，大きさ，字体などによる個性を消去して字画等を検出する。

ハ．候補文字の選択：計算機のメモリにパターンとその文字コード（符号）との対照表の形で入れてある字引のなかから，先に読み取った文字パターンと比較すべき候補文字を選択する。これは全数比較が実行不可能であるために行うのである。*

ニ．照合識別・認識：候補文字と照合して最も一致のよいものを識別する。一般には完全な一致が得られなかったり，複数が残ったりするだろう。

ホ．後処理：必要なら前後の文字との続き具合などによる後処理が行われるだろう。

ヘ．最後に文字コードがきまって読み取り作業を終る。

手書きの場合はずっと面倒になる。手書き文字によるワードプロセッサー入力は商品化されたものがあるが，文字を強く限定しないと無理であろう。ただし手書き入力の場合には筆順などの寄与もある。

トの1，文字線をたどって字画を検出する。

トの2．特徴を検出し，その特徴を使って候補文字を選択し，候補文字と比較しつつ字画を検出する。

チ．そのうえで印刷文字の処理・読み取りと同様の手順に移行する。

こういうことになっているが，詳細はよくわからない。昔あったローマ字論やカナ文字論（当然分かち書き）を採用していれば，この困難は避けられたのであるが，われわれは，日本語ワードプロセッサーの実用化などもあって，かな漢字まじり文を将来の日本語表記法として，改めて選んだことになろう。**

* JIS漢字に限定するとJIS第一水準2,965字。康熙字典は49,188字

**高田宏は“エッセーの書き方”という本のなかで“かな漢字まじり文は，世界の言語表記のなかでとびぬけて優れている。”と書いている。

スライド 23

鈴木・大田：問題提起
「分かち書き」
滋賀大国文，第21号，
p.47—57（1983）
日本文を分かち書きする場合には，単語ごとに切るか，文節ごとに切るかが大きなもんだいで，…
…分かち書きについてのルール，範例の未整理さが大

スライド23によって分かち書きの問題にふれる。

日本語について身近かな問題の一つに分かち書きの問題がある。

幼児のための本は

きく災いして，点訳活動が継続できずに挫折してしまう人が多い。…
…単純明快なルールの確立は等閑に付されるべきではないと思う。…

分かち書きになっているが，学年進行と共に分かち書きがなくなって行く。これは膠着語の特長なのであろうか。

スライド 24

Terry Winograd:
サイエンス 11,1984

自然言語の計算機処理の研究の将来について，3つの方向を示唆している。
1．話し言葉の研究にもっと重点が置かれるようになるだろう。
2．言語の使い方についての制約をさらに慎重に，理論的に加えるようになるだろう。
3．自然言語と形式言語を組み合わせたようなシステムを開発することである。

Terry Winograd*: Computer Software for Working with Language, Scientific American 11, 1984 日本版
に自然言語の計算機処理の研究の将来について，進むべき3つの方向が示されている。それらをスライド24にまとめて示した。

3は自然科学の論文等では昔からやっていることのように思われる。

Winograd は自然言語の計算機処理に対して大いに貢献した。** Winograd 以後のシステムの特色は，ある文の解析から得られる解釈の結果から，多分生じているであろう事態や状況についての推測を行い，次に現われる文をその推測のもとに解釈しようという考え方である。最も有力な推測が当らないときには，他の可能性を探すのである。

* Stanford 大学

**長尾真：自然言語の理解，情報処理 vol. 19, No.10, 1978-10

## 古い歴史

スライド 25

新しいアイデアにも
古い歴史が

1．プロダクション・システムは最初に，Postによって提案された．… AIにおける現在のシステムは，Post の提案した形式とは全く変わったものになっている。

電子計算機や通信手段の進歩に基づく学術情報のオンライン化とか人工知能(AI)というような，新しい技術やその背後にある基礎研究について調べているうちに，新しいアイデア・着想と思われるものにも長い(古い)歴史があるという場合（もの）があるのに気がついた。

Emil L. Post: Formal reductions of the general combinational decision problem, American Journal of Mathematics Vol. LXV 1943 pp. 197-215
2．KWIC についても Luhn (1959) に100年くらい先行する研究があると言う。

例えば，人間が知識を活かして行う問題解決の一つのモデルと考えられるプロダクション システム(PS) の最初の提案は，1943年(昭和18年) の Post の論文であるということである。もっともこの論文の内容と今日の AI における PS とはすっかり様替りしている様であるが，キッカケはここにあるという認識である。

この論文は記号論理学の分野に属する研究である。この論文に関連して（連想で）ずっと若い頃記号論理学の講義をきいたことを思い出した。大学卒業後間もない頃 Hilbert-Ackermann の「記号論理学の基礎」という本の日本語訳を出された伊藤誠氏（広大後に九大）のこの本に基づく講義を文学部の木造旧館の講義室できいたという記憶である。

附属図書館のカード目録で調べたところ，この訳本のカード（著者名）が7枚あって，一番古いのが1954年で謄写版となっている。

1957年には大阪教育図書出版となっている。私もこの謄写版のものをもっていて，これで講義をきいたことを急に想い出したが，とうとう私の蔵書の中には見付からなかった。残っていたとしてももうボロボロだろう。

もう一つは KWIC である(スライド25参照)。

## むすび

自然言語はまことに不思議なものである。こんなに難しいことを考えたり表現したりすることのできる言語がどのようにして，進化して来たのか，若し言語能力が，人間というハードウェアを働かせるソフトウェアであるとすると，不特定の人々の間で，長い歴史のなかでそれが作り上げられたということになり，実に不思議な気がする。

自然言語の機械処理を研究している人達は，

Lisp とか Prolog というような，深い研究を経て作り上げられたプログラム用人工言語を使ってプログラムを書いて，それを以て自然言語の処理をする訳だが，それだけではどうしてもうまく行かないということである。

さて，様々な言語による表現の奥に，言語に依存しない，共通の意味の世界があって，その世界における或る意味が，それぞれの言語によって表現されるという風に，表現の世界と意味の世界に言語の世界を分けることができるかと言うと，どうもそうではないらしいのである。

谷崎潤一郎の文に〝最初に思想があって，然る後に言葉が見出されるという順序ではなく，まず言葉があって，然る後にその言葉に当て嵌るように思想をまとめるということもある〟というのがある。これは谷崎が自分の経験から言っているので，味わい深い。

そのような次第で，機械ほん訳の過程においては，例えば

日本語の文→日本語の文構造（表層構造）→日本語の意味（深層構造）→A語の深層構造（意味）→A語の表層構造→A語の文

のように処理されるのである。これは可逆的であろうか？　恐らくそうではないだろう。長尾氏は〝機械ほん訳システムは複雑なシステムであるから，簡単に全体をとらえることは困難である。〟と言っている。この言葉は，われわれが日常使っている言語の不思議さを表わしているように私は思う。

日本語についての深い関心と理解をもって図書館用の日本語の標準化と，その機械処理を考えてもよいのではないだろうか。

京都大学附属図書館報「静脩」号外　1985年12月10日発行・編集：静脩編集委員会（責任者附属図書館事務部長）
発行：京都大学附属図書館・京都市左京区吉田本町・電　大代751—2111(内線)2611～2646