

# 国立国語研究所学術情報リポジトリ

## A linguistic analysis for artificial intelligence : Basic considerations

メタデータ	言語: jpn 出版者: 公開日: 2017-03-31 キーワード (Ja): キーワード (En): 作成者: 田中, 卓史, TANAKA, Takushi メールアドレス: 所属:
URL	<a href="https://doi.org/10.15084/00001055">https://doi.org/10.15084/00001055</a>

# 人工知能のための言語分析

——基礎的考察——

田 中 卓 史

## 1. はじめに

電子計算機を利用した情報処理が盛んになり、列車座席の予約、銀行の窓口業務等、日常生活とも関わり合いを持つようになってきている。この傾向は、将来にわたりますます進むものと予想されるが、その際、人と機械とのインタフェイスが重要になってくる。人が機械と情報を交換する際の理想の形は、機械の方を人間に近づけること、即ち、機械が本来持つ能力に加えて、人間の言葉を理解し、人間型の思考を行なう能力を持たせることであろう。そのためには、まず、我々自身の言語に関する情報処理の機構を明らかにすることから始めねばならない。

会話において、話者の頭脳内情報は、途中、言語の形をとって聴者の頭脳内へと移動する。読書においても、情報の移動に関して同様のことが言える。これらの場合に、頭脳を情報処理機械と見なすと、言語は頭脳内情報の外部表現として考えることができる。外部表現としての国語は、長年月を経て、内部情報を能率よく表わすことができるように進化し、発達して現在の形をなしているのであろう。とすると、国語には、頭脳内の情報の持つ構造が反映されているはずであり、国語を手がかりとして、逆に、頭脳内の情報処理の仕組みを解明できると考えている。国語の分析を通して見出される構造や法則は、頭脳模型の側面を描くものとしてとらえることが重要である。

## 2. 研究の方法

頭脳内での情報処理機構を明らかにする方法として、解剖学的、生理学的、あるいは心理学的方法等があろうが、情報処理システムとしての大局的把握を

行なうためには、言語を手がかりにするのが良いと考えている。しかし、この情報処理システムは、言語から直接的に調べる方法があるというわけではない。

我々に行なえるのは、計算機に代表される自動機械の延長として頭脳模型を描くこと、頭脳模型の仮説に対し、国語の分析から裏づけるデータを得ること、さらに、この模型をソフトウェアシステムとして計算機上に実現し、人と類似の言語動作を行なわせることである。なお、これらは独立して行なえるものではなく、部分的な仮説とそれに基づく分析作業の繰返しにより、模型の側面が少しずつ明らかになり、次第に詳細な構造が定まるものと考えている。

ここ数年来、人工知能研究の分野において、言語情報処理に関する研究が盛んになり、限られた範囲内ではあるが、言語の理解、知識に基づく推論、質問応答過程等の研究が試みられている。これらは分析的な立場よりもむしろ、人と類似の言語動作を行なわせることを目的として、知識の表現や処理過程に工夫がなされているが、頭脳模型を描く上で参考になることが多い。

現在、情報処理システムとしてのごく粗い頭脳模型として、一種のデータベースシステムを考えている。データベースシステムはデータベース（データの基地）とそれを操作し、利用するアルゴリズムから成る。データベースとは、特定の個人が生まれて以来、現在に至るまでに獲得してきた莫大な情報のファイルであり、近似的に個人の概念体系と呼ぶことができよう。一方、アルゴリズムとは、生得的に持っていたと仮定する概念体系を作り上げる能力であり、また、概念体系を内部情報源として利用し、言語理解、推論、思考等を行なう能力であって、処理系と呼ぶことにしている。

研究は、データベースとして見た概念体系を明らかにする段階と、アルゴリズムとしての処理系を明らかにする段階に大きく分けて行なう。最初から、人間と同規模の模型を考えることは困難であるから、概念体系としては比較的閉じた世界を例にとり、また、処理系としては人間の理性的側面に限定し、論理的思考を模倣させることを目標に研究を進める。

### 3. 概念のデータベース

#### 3-1 概念について

情報処理システムとして見た頭脳内に、概念体系に相当するデータベースの存在を仮定した。このデータベースの構成要素として情報の単位を想定し、仮に、概念と呼ぶことにする。この概念の外部表現として、語が対応して存在すると考える。語の意味がこのデータベースから生ずるものとすれば、語を意味的側面から分類・分析することは、間接的にこのデータベースを分析することを意味する。

頭脳内に形成された概念のデータベースは森羅万象にわたる莫大なものと想像され、分析に当たり多くの困難が予想される。それゆえ、分析作業は日常一般の分野と多数の専門分野の概念体系に分けて進めるのが良いと考えている<sup>2)</sup>。これまでに専門分野として、比較的閉じた世界と見なせる電子回路の概念体系を例にとり分析を進めている。分析の第一段階として、電子回路がどれだけの概念から構成されているか、また、それらはどのように分類されるかについて明らかにするため、電子回路の教科書に現われた文をカナデータとしてカードファイル化し、そこで用いられる用語の意味分類を行なっている。

専門分野を対象としても、分析の過程で用いる方法や、分析の結果として得られるデータベースの構造は、分野によらない不変的なものを含むはずであり、より大規模な分野に対する分析の見通しを得ることができると考えている。

#### 3-2 概念の分類

語の意味分類を行なう際の問題点は、分類の基準となる意味的要素をあらかじめ客観的に定め難いことである。特定の用途のための分類と異なり、概念体系を把握するための手段としての意味分類を行なうのであるから、分類された結果よりもむしろ分類の基準を見つけること自体が重要である。

今、分類を行なうための意味的基準を‘見方’<sup>2)</sup>と呼ぶことにする。見方（把握の仕方）を分析対象となる母集団に作用させ、基準に合う語を抽出する。どのような見方を設定するかは、積極的に構造を見出そうとする直観に基づいて定める。見方を作用させ、語がうまく分離・抽出されれば適切な見方の設定で

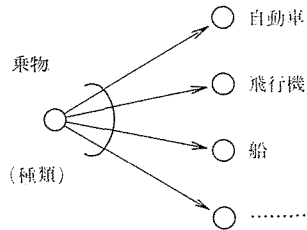
あり、語が全く抽出されなければまずい見方の設定である。また、いかなる見方でも分離されない語が残る場合は、いまだ見方（把握の仕方）が不十分であると考えることができる。なお、意味分類は同義語でない限り、語の意味的特徴をとらえ一語一分類となるまで行なえようが、見方による分類は概念体系の大局的把握を目的としているから、見方に深く内容が入るような詳細な分類は行なわない。

分野によらない基本的な見方となるものに‘もの’、‘こと’、‘抽象的ものごと’、‘とき’、‘ところ’、‘さま’等がある。専門分野に応じてこれらを細分する見方を設定できる<sup>9)</sup>。

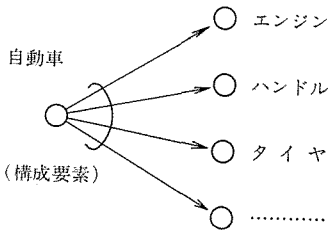
### 3-3 概念の連結構造

語は見方によりグループ化されるだけでなく、相互に意味的な関係を持っている。例えば、良く知られている上位・下位語の関係（図1）のほか、全体と部分あるいは構成要素との関係（図2）、概念とその属性（内包）を表す語との関係（図3）等がある。

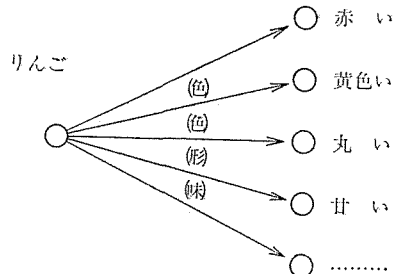
これらの語相互の意味的關係は、対応する概念間に存在する関係から生ずるものとする。即ち、頭脳内におけるデータベースには、語に対応する情報の単位‘概念’が存在して、それらの間の意味的關係が陽に蓄えられていることにする。なお、これら概念間の関係は、概念の一側面に着目して得られる関係と見なせるものが多いので、射影関係<sup>1)</sup>と呼ぶことにしている。



(図1) 上位・下位語の関係



(図2) 全体・部分の関係



(図3) 概念・属性の関係

概念間の二項関係は概念を節点、関係を枝とする有向グラフで表わされ、網構造をなす。この網構造は連想の構造を表わすものとして見なすことができる。

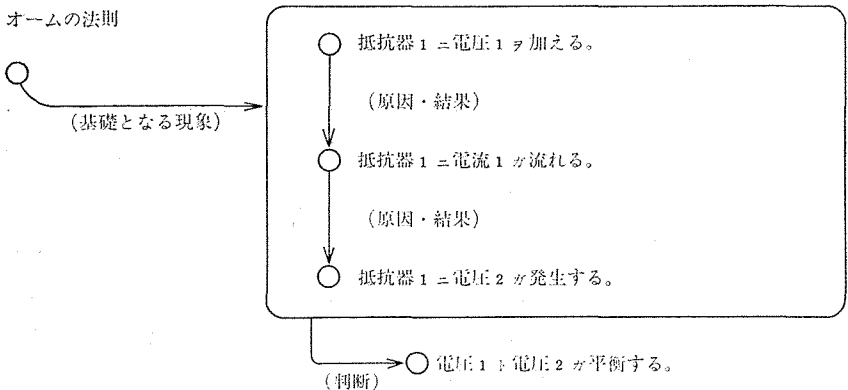
このような概念のなす構造を、一般に、概念の連結構造<sup>2)</sup>と呼んでいる。概念の連結構造における一つ概念は、他の概念との関係において相対的に定義されていると見なせる。

なお、概念の射影関係を求める作業において、用いる関係は概念の全般にわたり通用する基本的なものに限定すべきである。この射影関係は、後に処理系を考える際、アルゴリズム中に組み込まれる要素となるので、基本的関係という制限の中で語相互の意味をとらえることが重要である。

### 3-4 知識の蓄積

語間の意味的關係を陽に与えることにより‘自動車はエンジンとハンドルとブレーキと……とから構成されている’とか‘りんごの味は甘く、形は丸く、色は赤い’等ごく浅い知識が蓄えられたと見なせる。それでは、より深い知識はどのような形に蓄えられているのであろうか。

今、概念と呼んだ情報の単位を語から句や単文へと拡張することを考える。(この情報の単位に対し、概念という言葉の使用はもはや妥当でなくなるかも知れない。) その際、単位となる句や文の構成規則(分解規則)をシステムに



(図 4)

与える必要がある。(この規則を意味的基本構文<sup>2)</sup>と呼ぶことにしている。)

語の意味分類において、この概念を表すものとして分類された語は、文の形をとって完結した情報となる。文で表わされる概念間の関係は、語の場合にくらべ、はるかに知識と呼ぶにふさわしい内容を記述することができる。

情報の単位を文へと拡張すると、図4に表われた行為記述文と現象記述文間の原因・結果の関係、現象記述文相互の原因・結果の関係、現象記述文と判断記述文の関係の他、状況(状態)記述文とよく行なう行為記述文、行為記述文と目的、理由記述文の関係等の意味的關係、文を命題あるいは述語として見なして成り立つ論理的關係等を考えることができる。また、文の構成要素としての語の指示関係も重要になる。(図4において、抵抗器は各文とも特定の同一物を指示していることが必要であり、電圧に関しても同様のことが言える。)

しかし、語間の意味的關係が単なる網構造と見なせたのにくらべ、句や文はそれ自体も構造を持ち、その構成要素としての語に指示関係が必要になること、また、文間の単なる二項関係としては内容がうまく切り取れず、多項間の意味的關係を考慮せねばならないことから、網構造の拡張を考えるだけでは不十分であり、新たに知識を乗せる構造(わく組)を定めることが必要である。

今、知識の構造を定めるため、再び電子回路を例にとり蓄えるべき情報を考えると、あたかも演劇の台本ようになる。特定の回路に関する知識は、使用目的や機能等、射影関係により表わされる内包的、要約的なものの他に、より実質的なものとして、まず、回路現象の舞台となるべき回路の場が設定され、その中には、大小道具に相当する具象物(回路素子、部分回路)が存在し、登場人物に相当する各部の電圧、電流、信号が設定され、せりふあるいはト書として、電気現象、回路の状態、回路を操作する行為等が記述される。これらの記述の間には、原因・結果の関係を始め、種々の意味的關係が与えられる。

これらの多項間の関係は無制限に広がるものではなく、台本内で閉じており、概念のデータベース内においては局所化されて存在することになる。このような演劇の台本に相当する情報の単位をモデル<sup>10)</sup>と呼ぶことにしている。

(物理現象が微分方程式で表わされ数学モデルと見なせるように、自然現象、人の行為、様相、判断等の集合体からなる知識の言語モデルと考えたことによ

る。)

モデルと呼ぶ情報の単位を処理系から扱えるようにするには、モデルの構造（わく組）を定めこの中にデータを納めること、納められるデータ（文）もまた一定の構造の下に構成されていることが必要である。モデル内で用いる文は、構造に制限を設けないと、一般の文と同規模となり、この中から情報を分離し利用するアルゴリズムが複雑になりすぎる。また、制限が強すぎると情報がうまく乗らなくなる。できるだけモデルの構造に情報を吸収させ、要素となる文は、構造の単純な基本的なものとする必要がある。これらの構造をどう定めるかについては、実際に多くのモデルを構成してみる必要がある。

#### 4. 処理系の構成

##### 4-1 処理の単位

頭脳内に形成された概念体系を、処理手続（アルゴリズム）から独立したデータベースであると仮定した。このデータベースは語の意味分析を通して明らかにできると考えたのであるが、一方、言語理解、言語生成、推論・思考過程等を実行する処理系はどのようにして明らかにするのが良いのであろうか。概念体系にくらべ、処理系を明らかにするための分析的、直接的方法はさらに無いように思える。

処理系を構成するアルゴリズムに関して一般的に言えることは、頭脳模型において作り付けとなるため、データの内容に関して一般性のある、汎用的なものとなるべきことである。即ち、アルゴリズムからデータの内容に依存する部分を排除すべきであり、アルゴリズム中で用いて良い情報とは、データベースの構造（わく組）の情報、文の構造の情報等、データの内容に依存しないものである。このことは、逆に構造の情報を取り扱う処理単位が定義できることを意味する。データベースに関する処理単位とは、例えば次のようなものである。

##### (i) 概念の連結構造に関しては、

データとなる語の上位概念を求める、下位概念を求める、属性を求める、部分を求める、全体を求める、関連するモデルを求める、……がある。



(ii) モデルと呼ぶ概念（の集合体）に関しては、

モデルを構成する場を求める、場で起こることを求める、ことの主役を演ずるものを求める、……があり、さらにモデル内のことに関して、原因を求める、結果を求める、理由を求める、……があり、これらはさらに、現象を求める、行為を求める、様相を求める、……に細分できる。

(iii) モデル中の文に関しては、

主体を求める、対象を求める、場所を求める、手段を求める、時を求める、……があり、モデル中で用いる文の格構造と関連する。

以上はデータベースの検索を中心としたものである。いずれも概念の構造が明らかになるにつれ、おのずと定まってくる。これらの処理単位に対応し、設定された情報からデータベースを生成する操作の単位も定義することができる。

この他に、概念データベースの処理とは直接関係しないが、入力情報（文章）の操作、文法情報に基づく構文解析、出力情報の文としての生成等、システム構成に必要となる処理単位も定義せねばならない。これらの処理に関して、文法情報のテーブル等、概念のデータベースとは別に小規模のデータベースが必要になるう。

#### 4-2 処理単位の統合

処理の単位は言語理解あるいは推論・思考過程等、一連の処理過程を実行するために統合されることが必要である。では、統合された処理過程とは、どのようなものが存在するのであろうか。

これに対する手がかりは、再び語の意味分類から得ることができる。意味分類において‘精神行為を表すこと’として分類された語（考える、理解する、計算する、解析する、説明する、……）は巨視的ではあるが、処理系の制御状態を表わすものとして考えることができよう。

今、計算機とのアナロジーで考えるならば、マクロな命令が基本的な命令へと展開されるように、また、基本的な命令がマイクロ操作の列へと解読実行されるように、精神行為を表わす文は処理単位の組み合わせとして解読実行されねばならない。即ち、システムは、精神行為を表わす文が処理系の制御状態と

してセットされた時、処理単位を次々に起動するための命令デコーダを持たばなねなければならない。それでは、どのように処理単位を組み合わせると、それぞれの精神行為が実行できたことになるのであろうか。

精神行為の例として‘入力文を理解する’に着目しよう。処理過程を定めるためには、入力文を理解したという状態が、内部情報のどのようになった状態を指すかについて定めることが必要である。

理解の第1段階は、単独の文としての構文解析が行なわれた状態である。この段階においても、名詞句の決定、係り受けの決定等に文法情報の他、概念データベース（語間の意味的關係）の情報が必要になる。

さらに進んだ段階とはどのような状態であろうか。入力文の文法的構造が判明した後、疑問文、命令文であればシステムの制御状態へ直接影響を与えるであろうし、また疑問文、命令文を問わず文全般にわたり入力文の内容に関連して概念のデータベースを参照し、過去に蓄えられた情報との比較を行なうことが必要であろう。入力文に関連する情報は、入力文中の語をキーとしてその語に付随するモデル、あるいは、その語から概念の連結（語の意味的關係）をたどり関連語に付随するモデルから得られる。これらのモデル中に類似の文（名詞、名詞句から連体修飾を除き、上位語に置き換え、また述部から助動詞等を除く操作をして意味的な基本構造が一致する文）が見つければ、そのモデルをひな型として、モデル中の対象を入力文に則したものに置き換え、新しく類推モデルを生成できよう。

入力文が類推モデルにおける多項關係の要素として見なせることは、入力文が言外に多くの情報を提供したことを意味する。命令文であれば命令を実行するため、また疑問文であれば疑問に答えるためにはまず何をなすべきかという、より具体的行動の情報も得られようし、文章（文の列）中の文であれば、聞き手にも共通の常識として述べられることのない行間の情報を補うことができよう。

理解の最終段階は、入力文章の情報が概念のデータベースにおけるモデルと同じ構造にとらえられる段階であると考えている。同じ構造に変換された情報は、新しいモデルとして概念データベース中に組み込むことができる。

## 5. システムの構成

### 5-1 模型の拡張と制御系

精神行為を表わす文が処理系の制御状態に対応すると考えたが、これは模型システムにおいて、語句や文が単に処理対象としてのみ取り扱われるのではなく、システムの動作とも関わり合うことを意味する。より一般的に語句や文と模型システムとの関わり合いを調べるためには、より人間に近い模型を想定しておくことが必要である。

今、実現の可能性は別にして、計算機の入出力装置の代わりに人間の手足等の行動装置、目耳等の感覚装置を実物そっくりに作り、接続したシステムを考える。各装置のきめ細かな制御はそれぞれに付随する制御装置が行なうことにする。

一方、中央処理装置には多数の処理単位に対応するプログラム（処理系）とこれらを順次起動する制御プログラムが存在する。ここで制御プログラムの機能を拡張し、行動装置・感覚装置への制御指令も出すことにする。

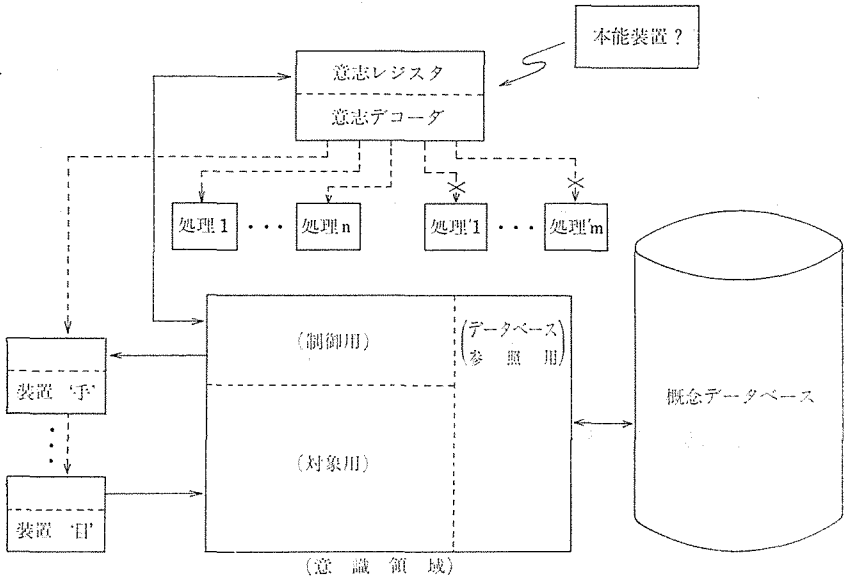
中央処理装置とこれらの装置との間の情報の交換は言語の形を用いることにする。行動装置であれば、中央処理装置から手足の行為を表わす文が送られてくる。

行動装置に付随する制御装置はこの文を解析し、行動の実行に必要な、装置の細部にわたる信号の列を生成する。但し、行動装置の制御装置が解析できる文（文というよりパラメータとしての語列と呼ぶ方が適当であろうが）の構造は限られたもので、用いられる語も装置の基本動作と直接関係するものに限定される。（例） 足：歩く、走る、けとばす、右、左、前、後、ゆっくり、……

なお、手足の行為は多くの場合、対象を把握するための視覚装置からの信号を必要としている。この種の信号は言語レベルではなく、制御信号のレベルで戻す必要がある。

感覚装置においても、発生する情報を言語の形で中央処理装置へ送ることにする。特に、視覚装置で発生する文には、目に見えるすべての具象物の名詞を始めとしてその動作や状態、関係等を記述する多くの語が含まれる。

行動装置が内部情報を外界への作用に変換したのに対して、感覚装置は外界



(図5)

の現象や状態等を内部情報へと変換する。これまでに述べた模型システムを図式的に画くならば(図5)のようになる。

図5において、制御プログラムの状態を表わすレジスタを考え‘意志レジスタ’と呼んでいる。また、各処理単位を実行するプログラムの作業領域を考え‘意識領域’と呼ぶことにしている。実線は言語情報の流れを示し、点線は制御指令の流れを示す。処理‘1’～処理‘m’は必ずしも意志の制御下で動作していると見なせないプログラムの存在を示す。(例)意識領域中に生成されたモデルのデータベースへの新規登録(学習)、データベース中の不用モデルの消去(忘却)

### 5-2 システムの動作

意志レジスタに(精神)行為の文がセットされると、意志デコーダ(プログラム)は、行為を実行すべく各処理単位プログラム、及び周辺装置に順次起動する指令を与える。デコーダは精神行為を表わす文のすべてを解読できるわけではない。(デコーダが莫大になりすぎる。)解読できる精神行為は限定すべきで、一般に行為を表わす文は、一度、意識領域にセットした上で概念データバ

ースを参照し、類推モデルを用いてより基本的な行為、あるいは状況に適した行為に分解した後、意志レジスタにセットされることにする。意志は解読可能なものに取り換えられるため、意志レジスタにスタック機能を持たせ、古い意志を保存することも必要になろう。意志レジスタに行為の文がセットされると、動作が開始されるが、意志レジスタのセット自体はいかなる情報に基づき、どのプログラムが行なうのが適当であろうか。

- (i) 意志の制御から離れたプログラムが外部からの命令文を無条件に終止形になおし意志レジスタにセットする。
- (ii) 本能装置を作り意志の根源となる情報を作り出す。

### 5-3 システムと語との関係

すべての語は、概念のデータベース中において、データとして使われることにより、相対的に陰な形で定義されていると見なせるが、システムから直接的に定義される語も多数存在する。それらの一部を次に示す。

- (i) 制御プログラムによる定義……考える、理解する、解析する
- (ii) 処理単位による定義……原因、結果、目的、理由、種類
- (iii) 行動装置による定義……手、足、つかむ、なげる、右、左
- (iv) 感覚装置による定義……りんご、自動車、流れる、甘い

これらは、語の意味分類においてシステム的な見方を設定することにより明らかにすることができる。

## 6. おわりに

頭脳模型を設定することにより、国語の意味的側面の分析を通し得られる種々の結果を、システム的な見地から位置づけることを意図している。現在の模型で位置づけられない結果が出れば、その時点で模型を修正すれば良いと考えている。

現在の模型システムは次の特徴がある。

- (i) 制御装置（プログラム）、行動装置は言語情報をシステムの動作や行動に変換する。
- (ii) 感覚装置は外界の現象や状況を言語情報に変換する。

(iii) (i), (ii)により機械の動作(行動)と外界の現象・状況が言語情報の形で対象化され統一的に取り扱える, さらに, ‘状況—行為—現象’等の多項関係をモデルとしてデータベースに登録, 参照できる。

これらは, システムが単にプログラムの, あらかじめ定まった道の上を条件を選択しながら進むというのではなく, 自己の動作を一度対象化して取り扱い, 過去の知識や経験を参考にし, 起こるべき動作結果を予測した上で, 状況に適した動作や行為を決定できるシステムであることを意味する。

この模型を実現させるためには, 分野を限定した上でそこで用いられる語句や文の詳細な分析が必要になる。現在, 電子回路の分野を例として分析を進めているが, 概念の表現には言語だけではどうしても不十分であり, 図形(回路図)と言語との関係を明らかにすることが必要である。

なお, ここで述べた事柄は, 生前, 御指導, 御鞭撻を賜った九州大学工学部教授故栗原俊彦先生の思想に負う所が多い。深甚の謝意を表し, 御冥福を祈念する。

(52. 9. 30)

#### 参 考 文 献

- 1) 栗原・吉田・藤田: 単語の意味分類—モデル論的考察一, 九大工学集報, 42.3, p. 247 (1967—06)
- 2) 栗原俊彦: 自然言語の機械処理, 情報処理, 14.4, p. 267 (1973—04)
- 3) 岡田・田町: 自然語および図形解釈のための単純事象概念の分析および分類, 電子通信学会論文誌, 56—D.9, p. 523 (1973—09)
- 4) 長尾・辻井・田中: 意味および文脈情報を用いた日本語文の解析—名詞, 単文の処理, 情報処理, 17.1, p. 10 (1976—01)
- 5) 古川康一: 人工知能とデータベース, 情報処理, 17.10, p. 927 (1976—10)
- 6) 田中穂積: 談話の理解とメモリモデル, 電子技術総合研究所彙報, 40.7 (1976)
- 7) 長尾・辻井: 自然言語処理プログラム, 情報処理, 18.1, p. 63 (1977—01)
- 8) 雨宮・島津他: 図形世界を話題とした質問解答システム, 情報処理, 18.8, p. 799 (1977—01)
- 9) 栗原・田中: 電子回路に関する定性的な思考の言語によるシミュレーション(第1報), 44.4, p. 574 (1971—08)
- 10) 栗原・田中・前田: 同(第2報), 44.5, p. 645 (1971—10)
- 11) 栗原・田中・前田: 同(第3報), 45.1, p. 116 (1972—01)