

情報検索論のモデル論的展開

Applying a Model Theory to Information Retrieval

鈴木志元
Yukimoto Suzuki

Résumé

Most of the theories dealing with information retrieval are based on the concepts of set and Boolean algebra. 'Boolean algebra' is a general term of several algebraic systems which correspond one-to-one to various logical systems. In the present theory of information retrieval, however, the term 'Boolean algebra' is meant to be only an ordinary Boolean algebra, and this restricts the logic of information retrieval to the propositional logic.

Generally, syntactic structures of some language systems are thought to be some algebraic systems. So, if we can formulate information retrieval by means of the different types of Boolean algebras from the one used in the existing IR systems, this will lead to the other formulation in the syntax of query languages.

In this paper, we adopt a model theory to formulate the semantics of information retrieval, since it is the most fundamental explanatory frame for communication and is expected to be one of the fundamental tools in library and information science.

- I. はじめに
- II. モデル論的研究の歴史的展望
 - A. モデル論 : Tarski
 - B. モデル論的意味論 : Montague
 - C. プログラム意味論 : Scott
 - D. データベース論 : Minker, Reiter
 - E. 情報検索論 : Marek, Lipski

鈴木志元；東京大学教育学研究科博士課程，東京都文京区本郷7-3-1
Yukimoto Suzuki; Graduate School of Education, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo.
1988年9月30日受付

III. モデル論的情報検索論

A. モデル論

1. 真理対応説
2. 公理的方法
3. モデル論

B. モデル論的意味論

C. 代数的定式化

D. 索引とモデル論

IV. おわりに

注・引用文献

I. はじめに

現在稼働している情報検索システムに用いられている方法は、ほとんどが2値ブール論理に基づいたものである。この方法はいくつかの欠点が指摘されており、情報検索のプロセスを人間の思考にあつたものにしよとする試みが数多くなされてきた。ファジー論理の利用もそのような試みの1つである。

通常の論理体系とは異なる論理を用いて情報検索システムを拡張しようとする試みは、今のところファジー論理の利用以外には見あたらない。しかし、これまでに研究がなされてきた論理体系は、命題論理やファジー論理の他にもいくつかある。これらはすべて論理を人間の思考に近づけようとして編み出された体系である。情報検索を考える際に特定の論理体系にのみ依存しては、論理学が我々に与えてくれる成果の大きな部分をみすみす見逃すことになる。

情報検索を様々な論理体系のもとで考えようとするとき、情報検索を根本的に見直す必要がある。基本に戻って考えれば、一般に情報システムとは情報源と利用者とを結ぶコミュニケーションシステムであると見なすことができよう。そしてコミュニケーションという状況に関する理論を最も論理的に展開しているのがモデル論である。本論文では情報検索の基本的な枠組みとしてモデル論を用いる。

モデル論は数学的真理を定義しようとして開発された理論であり、いわば数学的認識論とでも言いうる理論である。最近、言語学やコンピュータ科学へのモデル論の応用が盛んに行われており、その有用性が認められてきた。情報検索をモデル論的に展開してみようという試みも過去にいくつかなされており、データベース研究者の

関心と呼んだ。しかし現在の所、図書館・情報学の研究者の話題となるまでには至っていない。

本論文は、モデル論という考え方の基礎を紹介し、その枠組みの中で情報検索がどの様に捉えられるかについて述べる。本論文の主たる目的は、図書館情報学を展開する枠組みの1つとしてのモデル論の有効性を示すことである。その1つの応用例として、情報検索論をモデル論的に展開した。

II. モデル論的研究の歴史的展望

この章ではモデル論の発生とその波及についての歴史的概略を述べる。モデル論の理論的解説は第三章で行う。

理論というものは突然生まれるわけではなく、それらしきものは昔からあると考えるのが通例であろうけれど、モデル論を最初に定式化したのは Alfred Tarski である、と見なすのが一応の定説となっている。本章でも Tarski を中心に説明を行う。モデル論の他分野への影響についても、Tarski とのつながりをみていくと流れを捉えやすい。

A. モデル論：Tarski¹⁾

A. Tarski は《真なる文》を定義しようとする試み、ある文の述べていることが実際に成り立っているならその文は真、成立していなければ偽、とした。述べていることが実際に成立しているとは、例えば、「雪は白い」という文を考えるとき、次の式の右辺と左辺が同値になるということである。

「雪は白い」は真である \Leftrightarrow 雪は白い
つまり一般的にいうなら、文の名前をTとしたとき、

Tは真である \Leftrightarrow T
を、真理概念の満たすべき基準とするのである。

しかし、この様に定めると矛盾が生じてくる。自己言及による矛盾であり、一般に「ウソつきパラドックス」という名で知られている。例えば、

(*) 「文 (*) は真ではない」

という文を上基準に当てはめると、文の名前は (*) であるから、

文 (*) は真である \Leftrightarrow 文 (*) は真でないとなる。これは矛盾である。つまり、ある言語体系に上のような真理基準を満たす述語「真である」を加えると矛盾が生ずるのである。これは言い替えば、真理という概念がメタ概念であるということ、即ち、文の真偽は言語に関する知識のみで決定することはできない、ということである。更に言うなら、文の真偽を論ずるには、その文によって言及されている世界についての知識を備えて置く必要がある、ということの意味しているのである。「その文によって言及されている世界」というものを明示し、表現と世界との対応を明確に示さなければ、文の真偽を論ずることはできない、という命題がモデル論の主張なのである。したがってモデル論的に言えば、人間によって認識され抽象化された対象世界と言語表現との体系的な対応づけが与えられて初めてその言語の意味論を論ずることができる、といえる。

E. モデル論的意味論：Montague

モデル論の考えを自然言語の意味論に応用したのが R. Montague である。彼の作った意味論は、モデル論の自然言語という分野への言い替えであるから、当然のことながら、今問題にしている世界の完全な記述を前提としている。このことが度々言語学者の攻撃的になった。しかし、モデル論という説明装置の明快さは捨てがたく、論理学者と言語学者の共同作業によりいくつかの改良が加えられ、モデル論的意味論として意味論の中に確固たる地位を占めるにいたっている。

いまモデル論的意味論という言葉を用いたが、モデル論が文の真偽を論ずることから発生したように、『ある文の意味を知るということは、その文が真であるためには世界がどのようであればならぬかを知ることである』²⁾ という具合に真理という側面を強調したときには、真理条件の意味論という言葉が用いられることもある。また、可能世界意味論という言い方がなされることもある。Dowty は可能世界意味論を次の様に説明している。

『ある文に対して真理値を与える場合、その文が真であ

るためには「世界はいかようであらねばならないか」に関して言及することになる。したがって、文の意味は単にそうである世界のみならず、そうであるかもしれない、あるいはそうであったかもしれない世界すなわち、他の可能世界 (possible world) 一にも依存する。(中略) 可能世界という概念の直観的な意味合は、ものがいかようであるか、あるいはいかようであるかもしれないかを、意味的に、関与する限りきめ細かく完全に規定するということである。ある特定の世界すべての可能世界の中の1つは、ある文の真理値に影響を及ぼすことのできるあらゆるもの、すなわち、文がそれについて述べることのできるあらゆるものを含むことになる。』²⁾ (下線は鈴木による)

この文は Montague 意味論の性格をきわめて簡潔に表現している。「真理値に影響を及ぼすことのできるあらゆるもの」を、「関与する限りきめ細かく完全に規定する」ことができ、初めて意味論を論ずることができる、ということが Montague 意味論の基本的前提なのである。これはモデル論における主張 (真理概念はメタ概念であって、真理を規定するには言語の知識の他に世界に関する知識も必要である) に対応するものである。

しかし先にも述べたように、「あらゆるもの」を「完全に規定する」という条件はいかにも厳しく、言語学者の非難がこの点に集中された。Montague による1970年の提唱以降のモデル論的意味論の十数年の歩みは、これらの欠点を解消する試みの歴史であったともいえる。このような改良の1つとして最近注目を集めているのが、J. Barwise による状況意味論である。人間が世界の完全な記述を行えるわけがない、という非難に応じて部分的情報を扱おうと試みている。状況意味論は、『同じ言語表現が異なった状況のもとで用いられるという言語の特質が、従来の言語哲学で無視されてきたと考え、表現の解釈及び真理値決定における状況を重要視する。意味は状況間の関係であると考える』³⁾ 理論なのである。

C. プログラム意味論：Scott

モデル論の応用として有名なものにプログラム意味論がある。プログラム意味論は、プログラムの検証などに数学的な基礎を与えるものとして、盛んに研究が行われているものである。中島によれば、プログラム意味論がめざしているのは次のようなことである。『プログラムの正しさを証明という数学的議論によって確立することが検証である。プログラムの正しさとはプログラムがそ

の作成の意図のとおり働くことである。証明が可能になるためには、作成者の意図（すなわち仕様）やプログラムの意味が数学的議論の対象となりうるものでなくてはならない⁹⁾。コンピュータ科学も意図の伝達という根本的な問題で悪戦苦闘しているようだが、モデル論の基本的性格を考えれば、この問題に対する有力な概念装置としてモデル論の枠組みを借りようとするのは、いわば当然の成行きであったように思う。

通常プログラム意味論と言われているものには、操作的意味論、公理の意味論、表示の意味論がある。この中でモデル論を応用して理論を展開しているのは表示の意味論である。これは論理学者 D. Scott が1970年初頭に定式化した意味論であり、数学的に最も明快なものとして、現在では（多少とも理論的な）コンピュータ科学の常識になりつつある。

表示の意味論における意味論の展開も、自然言語に対するモデル論の意味論の展開の仕方と基本的にはまったく同じである。まずプログラム表現からなる統語論の領域を定め、次にプログラムの意味が直観的に理解できるような豊かな意味領域を定め、これら2つの領域の間の対応を規定すればよい。ただ統語領域が完備半順序集合になりその上の連続写像を扱わねばならないので、位相的議論が必要になる、という面倒さはある。

D. データベース論：Minker, Reiter

1960年代の初頭から質問応答システム等において、論理学の応用はなされていた。これらの研究は、定理の証明をへて論理プログラミングとして現代につながっている。データベース論のなかで論理学を意識したものとしては、Codd による関係型データベースシステムの研究があげられよう。関係型に関連して述語論理に基づいた問い合わせ言語の開発等が行われた。

データベース論の中でモデル論をとりあげたのは Minker と Gallaire それに Nicolas である。彼らは『データベース：理論と解釈』¹⁰⁾と題する論文を発表しているが、ここでいう理論とは形式的体系(第III章を参照)のことである。彼らはデータベースに蓄積されている情報の体系を形式的体系として捉え、ある関係の外延的データ（関係型データベースならタプルの集合）を解釈としてデータベースを論じている。

モデル論的にデータベースを扱おうとする研究者のなかでも特に注目に値すると思われるのが R. Reiter である¹¹⁾。彼はデータベースが規定するモデルを閉世界と

して捉え、それを更に非単調論理として定式化したデフォルト論理を展開している。閉世界意味論について簡単に説明しよう。

データベースはある世界の記述であるが、もとより完全な記述は望むべくもない。記述の対象を1つの部屋に限ってみても、その部屋に存在しないものをすべて列挙することは不可能である。そこで、「存在する」という積極的な言明がないときには「存在しない」と解釈しよう、という考えに基づく意味論を閉世界意味論という。演繹のデータベースにおいて、ある命題を証明することができなければ、その命題の否定が成り立つと見なすのである。あるいはこのデータベースはある公理を入れられているだけ（即ち、情報が不完全）であって、その公理があれば問題となっている命題は成り立つのかもしれない。しかし、たとえそうであったとしても、今はこのシステムを1つの閉じた完全な世界とみなして推論を行おう、とするのが閉世界意味論である。

閉世界意味論をさらに展開していったのがデフォルト論理であり、この他にもデータベースの意味論を扱う論理が非単調論理という名のもとに研究されている。

E. 情報検索論：Marek, Lipski

情報検索をモデル論に定式化しようとする試みは、図書館情報学とは別の文脈でなされた。コンピュータの理論的研究の一環としてなされたもので、主としてポーランドの学者によって研究が行われている。論理学におけるポーランド学派の伝統に則っているのであろうか。Tarski もポーランド学派の代表的人物の1人である。

最初にモデル論的定式化を試みたのは Marek と Pawlak である¹²⁾。彼らは情報検索システムを形式的体系として定式化し、情報検索論を数学的に展開しようとした。彼らの情報検索システムとは次の4つの要素の組である。

X：対称集合（文献項目の集合）

A：ディスクリプタの集合

I：属性の場合

U：索引関数 AからXのべき集合への写像

Uは次の条件を満たす。

1) $i \in I, a \in A, b \in A, v \neq w$ なら

$$U(v) \cap U(w) = \emptyset$$

例えば、 i が色という属性で、 a が赤というディスクリプタ、 b が黒というディスクリプタであるとすると、このとき、赤と索引されているものの集合

とは重ならない。

$$2) \quad U \cup (a) = X$$

各文献項目には各属性について必ず1つのディスクリプタが索引付られている。

このシステムで認められている問い合わせは、このシステムを形式的体系とみたときの項と式(第三章参照)である。項はある条件を満たす対象の集合を求めるときに用いる問い合わせであり、式はある条件を満たす対象の間にある関係が成り立つかどうかを尋ねる問い合わせである。

項の意味値を次のように定義される。

- 1) $\langle a \rangle = U(a)$
- 2) $\langle \sim t \rangle = X - \langle t \rangle$
- 3) $\langle s \cdot t \rangle = \langle s \rangle \cap \langle t \rangle$
- 4) $\langle s + t \rangle = \langle s \rangle \cup \langle t \rangle$
- 5) $\langle s \rightarrow t \rangle = (X - \langle s \rangle) \cup \langle t \rangle$

彼らの定式化は2値ブール代数の範囲内であり、通常の意味の域をでいていない。しかし、彼らの目指したものは情報検索のモデル論的な定式化であり、その目的は果たしているといえる。

彼らの後を受け継いだのが、同じくポーランド人の Lipski⁹⁾であった。彼は、Marek と Pawlak による情報検索のモデルを拡張し、不完全な情報をもつシステムに対する問い合わせをも表現できる言語を作ろうとして、様相論理の体系を採用した。この言語では様相論理の必然性演算子と可能性演算子に対応するものとして、‘surely’や‘possibly’といった表現を使うことができる。

命題様相論理とは、命題論理に必然性を表す論理記号が付け加わったものである。だから、統語論としては、文法的に認められる式の範囲がこの論理記号のぶん広がった、という変化が起こるだけである。意味論としては、Kripki による可能世界モデルが用いられている。命題論理はある定まった時点に於ける命題の真偽を問題にしているが、可能世界モデルを用いれば時間の経過をも視野にいれた体系の論理を扱える。ここでいう可能世界とは、各時刻における世界を表しているともみなすことができるのである。また、各可能世界を情報の集合と考え、世界と世界との関係を情報の集合の包含関係と解釈することもできる。(時間との関連でいうなら、後世の持つ情報量の方が大きい、という解釈とみなせる)

Lipski は可能世界を情報の集合とみなした解釈にしたがって、不完全な情報をもつデータベースの意味論を

構築する試みを行った。不完全なデータベースに情報を加えていって次第に完全なシステムにすることは可能であろう。そのプロセスに様相記号の解釈を対応させたのである。このとき、 $\square B$ (即ち、「Bであることが必然である」)は「我々の現在の知識のあらゆる可能な拡張において命題Bが真である」と解釈される。

問い合わせに対する意味値を次のように定義することができる。

- 1) $\langle (i.A) \rangle = (\text{属性 } i \text{ に関して } A \text{ と索引づけられている文献の集合})$
- 2) $\langle 0 \rangle = \phi, \langle 1 \rangle = X$
- 3) $\langle -t \rangle = X - \langle t \rangle$
- 4) $\langle t + s \rangle = \langle t \rangle \cup \langle s \rangle$
- 5) $\langle t \cdot s \rangle = \langle t \rangle \cap \langle s \rangle$
- 6) $\langle t \rightarrow s \rangle = \langle X - \langle t \rangle \rangle \cup \langle s \rangle$
- 7) $\langle \square t \rangle = (\text{現在のシステムのあらゆる拡張における } t \text{ の値に含まれる文献の集合})$
 $= \cap \langle t \rangle$

1)から7)は項に対する値である。式に対する値も同様に定義される。必然演算子 \square を解釈するときは、システムのあらゆる拡張を考え、可能演算子を解釈するときは、条件を満たすあるシステムの存在が言えればよい。

$$\langle \langle t \rangle \rangle = (\text{ある拡張されたシステムにおいて } x \in \langle t \rangle \text{ となる文献 } x \text{ の集合}) \\ = U \langle t \rangle$$

彼のこの定式化は位相ブール代数を念頭においたものである。論理に対応した代数を用いて情報検索システム(ないし、その検索言語の意味論)を定式化する試みを行ったのはこれが最初であろう。しかし、残念ながら具体的な計算方法を規定するまでにはいたらなかった。

III. モデル論的情報検索

情報検索一般を理論的に定式化するには、情報検索その根底から見直す必要がある。本論文ではこの試みに最適な装置として、数学基礎論におけるモデル論を採用する。モデル論は数学的真理を定義する試みから発生した理論であり、検索式の真偽を論ずる枠組みとして最適なものであろう。また、言語学やコンピュータ科学への応用も広がってきており、情報検索をモデル論的に定式化することによって、それらの方面から情報検索へアプローチする可能性も開けてくることが期待される。

この章ではモデル論発生の前提となった「真理」という概念から公理的方法、モデル論やモデル論の意味論、

そしてその代数的定式化までを解説する。

A. モデル論

1. 真理対応説

真理という概念をどのように捉えるかについては昔から多くの説がある。その中で最もポピュラーなものが真理対応説であろう。それは次のような主張を行うものである。

『真理とは信念と実在する世界の特徴すなわち事実との対応である』⁹⁾

この定義で第一に問題となるのは、事実とは何か、という点であろう。これについては多くの説があるけれど、ここでは神野に従って、事実を次のようなものとする。

『事実とは、世界という時間空間的連続体からわれわれの認識によって切り取られたものである。それに対し、事物、性質、出来事は、これらこそがまさに世界という時間・空間的連続体をなしているものである。(中略) 知性をもつ存在は、存在を見つめ存在を解釈することができる。すなわち、世界を解釈して事実の世界となすことができる。(中略) 事実とはわれわれが自らの関心と知識に応じて世界から選び出したものである』⁹⁾

ここに引用した事実のとらえ方は、我々がデータベースにおける情報を考えるときに参考となるものである。データベースを構成する情報とは世界における生の事物や出来事ではない。蓄積されている情報は、索引者によって切り取られ解釈された世界の断片である。つまり、データベースを構成している情報とは、我々がここで定義した事実なのである。

また、真理対応説において事実と対置させられたのは信念であった。信念という言葉は、世界観という言葉で言い替えてみよう。一般に情報要求とは世界観をより完全な形にしようとする要求をあらわしたものであるとみなすことができる。とすれば、真理対応説における信念に情報検索論における情報要求(問い合わせ)を対置させることができよう。

2. 公理的方法¹⁰⁾

数学的理論は、ある概念を規定する定義と命題に対する証明とから構成されている。一般にある概念を定義するには別の概念が必要であり、ある命題を証明するには、その前提(証明に用いられる事実)となる別の命題が必要である。これらの概念のつながりと命題のつながりを遡っていけば、最も基本的と思われる概念と命題に

いきつく。最も基本的と思われる概念と命題をもとに厳密な定義と証明とを与えることによって数学的理論を体系的に構築しようとする方法が公理的方法といわれているものである。理論構築に必要な他のすべての概念がそれを基にして定義されるだろうと考えられる基本概念を無定義概念といい、それに基づけば理論中の命題がすべて証明できるであろうと考えられる基本的な命題を公理という。無定義概念と公理を仮定すれば、他の命題はすべて論理的に導かれる。

公理を定めるときには、当然その意味を考え、理論にとって本質的と考えられるものを選んだのであるが、一旦選んでしまえばそれらが本来持っていた意味を無視していくことができる。このように公理が何を指し示しているかという詮索を抜きにして公理を純粋に仮定と考え、ある命題が公理から証明(推論)できるか否かだけを問題にすれば、経験世界における私的な判断から切り離され、完全に形式化された問題として処理することが可能となる(一般に、システムという名で機械化されているのは、この形式的な部分である)。このように、人間が現在認識できる範囲の中から基本的な事実と思われるものを公理として選び、その公理をもとに論理という人間の武器によって切り崩せるところを切り崩そう、とする方法を公理的方法といっている。(あくまでも、現在認識できる範囲、である。将来において、公理を修正する必要は当然起こるだろう。それにともなって、システムの更新が行われる)

例えば、自然数論に於ける公理とは自然数のすべてをいい尽くしている「一応」考えられる命題の集まりである。したがって、全能の神が作った自然数論を本来の自然数論と考えるなら、公理から展開される自然数論は人間の作った自然数論であり、本来の自然数論のある種の近似像でしかない。つまり公理的方法とは、『仮に自然数とはそのようなものであるとしておいて、ともかくその性質を調べてみよう』¹⁰⁾という暫定的な方法なのである。

こうしてみると先に述べた真理論は、いわば公理的な真理論であったことがわかる。事実というものを突き詰めて考えていけばどうしても認識という壁に突き当たらざるを得ない。この壁に挑戦するのも方法の1つであろうが、我々は認識論の泥沼に入り込むことを避け、事実とは我々が世界から切り取るもの(我々が今認識しているもの)だと規定し、その規定によって定まる範囲の中で真理というものを考えようとしたのであった。これ

は、情報システム一般を考えるとときにおさえておくべき基本的事実である。つまり、システムに蓄えられた情報は、ある角度から切り取られた近似的な情報なのである。この事実を無視して情報の意味などを論じるのは無意味であろう。知識ベースの議論において、知識という言葉ではなく信念という言葉が多く用いられるのは（例えば、Doyle の TMS のように）、この事情による。

公理的方法に則って数学的真理を厳密に論じようとするのが、次に述べるモデル論である。

3. モデル論¹⁰⁾

数学や論理学においては、自然言語の曖昧さを避けるため、記号を用いる。記号で理論を言い表せるためには、記号の使い方についての明確な規定がなければならない。記号の意味を考えずその形式的な構造を研究する立場を統語論 (syntax) といい、記号に意味を対応させその内容を考える立場を意味論 (semantics) という。つまり、数学のような形式的理論を展開するには3つの道具建てが必要なのである。

1つは、記号である。普通、数学的理論は個体記号、関数記号、述語記号、論理記号等からなっている。理論展開に用いる前にこれらを明確に規定しておかなければならない。

2つめは、統語論である。これは3つの部分からなる。1) 論理式を定める規則：この理論で用いる式(命題)はどのような形のものであるのか。論理記号が1つだけでは式にはなれない。なお数学的理論では、文に相当する式のほかに、名詞句に相当する項も規定するのが普通である。2) 公理：理論展開のもとになる論理式はなにか。3) 推論規則：論理式から論理式を導く規則。理論展開のルールである。

統語論においてある論理式が成り立つとは、公理としてある式を選び、さらに推論規則を定めたとき、その論理式が公理から推論規則を用いて導ける場合のことである。このようにして導かれた論理式を定理と呼んでいる。公理の選び方によって（即ち、理論によって）1つの論理式が成り立ったり成り立たなかったりする。ユークリッド幾何学と非ユークリッド幾何学を思い起こしてほしい。

ここまでは意味を考えていない。単なる図形としての記号とその記号の組合せ方を規定しただけである。しかし、これだけの道具で、形だけではあるが1つの体系ができあがる。この体系を形式的体系と呼んでいる。

さて、3つめは意味論である。形式的体系に存在して

いるのは単なる記号例（論理式）であった。そこでは式の意味はまったく考慮にいれられていない。記号にある解釈が与えられて初めて意味を問題とすることができ

る。形式的に意味解釈を導入する前に文の真偽について少し考えてみよう。文の真偽は、文の構成要素である原子文の意味づけと言語以外の事実（世界）とから決定される。例えば、「ジョンは犬である」という文を理解するには、ジョンや犬が何を指すのかという意味づけが必要であるばかりでなく、この発話がどのような状況（世界）で行われたのかという情報も必要である¹¹⁾。

「ジョン」がQ氏の飼犬の名前であり、「犬」が動物の犬をさしているのなら、ある現実の世界において上の文は真となる。あるいは、「ジョン」がある小説の主人公であり、「犬」が警察の手先を指しているのなら、上の文は小説という架空の世界において意味をもつ。

つまり意味論には次の種類のものを記述する道具が必要となるのである。

1. 世界：少なくとも1つ可能な世界がある。これは記号の集合に対応し、その直観的な意味を我々に納得させるだけの豊かさをもった対象の集合であればなんでもよい。
2. 意味割当(解釈)：文を構成している記号に世界における対象を対応させる関数。
 - 1) Dは対象領域と呼ばれる空でない集合である。
 - 2) 個体記号 c に対し、 c の解釈と呼ばれるDの元 $I(c)$ がある。
 - 3) 関数記号 f に対し、 f の解釈と呼ばれるD上の関数 $I(f)$ がある。
 - 4) 述語記号 P に対し、 P の解釈と呼ばれるD上の述語 $I(P)$ がある。

いま考えている形式的体系の公理がある構造においてすべて真である（真という意味をもつ）とき、この構造をその形式的体系のモデルであるという。このモデルにおいて、公理から導かれる定理もすべて真となる。論理学における完全性定理によって、定理が証明されること（これは統語論でのみならず）と、その定理(命題)があるモデルにおいて真であること（これは意味論でのみならず）とは同値となり、真の命題から別の真な命題を導出する操作を形式的体系(統語論)における規則で代行するこ

とができる。この事実が次に述べるモデル論的意味論の鍵となるものである。

B. モデル論的意味論

第II章で述べたように、モデル論的意味論はモデル論の考えを自然言語の意味論に応用したものである。前節で述べた完全性定理によって公理が真となる世界（モデル）では公理から推論規則を用いて導かれた式（定理）が真となった。つまり定理を証明するプロセス（統語操作）においては意味をいちいち考えることはせず規則に則って機械的に定理が導かれた。いま、証明のプロセスの陰に対応する意味操作があって、定理の意味（真）が公理の意味（真）からこの意味（操）作をへて得られると考えれば、モデル論とモデル論的意味論との対応は明らかである。つまり個々の統語論の単位から文が構成されるのとまったく同じようにして個々の意味論の単位から文全体の意味が構成されていくと考えるのである。これがモデル論的意味論の中心となる考えであり一般に構成性原理（または、Frege の原理）と呼ばれているものである。

例として次のような言語 L_0 ¹²⁾ を考え、どの様に意味を定めるのかをみよう。言語 L_0 の統語論は次のものからなっている。

・ L_0 の基本表現

名 辞 a, t, m

1項述語 $G, B,$

2項述語 F

・ 統語規則

1. δ が1項述語で、 a が名辞ならば、 $\delta(a)$ は文である。
2. γ が2項述語で、 α と β が名辞ならば、 $\gamma(\alpha, \beta)$ は文である。
3. ϕ が文ならば、 ϕ は文である。
4. ϕ と ψ が文ならば、 $[\phi \cap \psi]$ は文である。
5. ϕ と ψ が文ならば、 $[\phi \cup \psi]$ は文である。
6. ϕ と ψ が文ならば、 $[\phi \rightarrow \psi]$ は文である。
7. ϕ と ψ が文ならば、 $[\phi \leftrightarrow \psi]$ は文である。

1と2は原子文(その文の一部に他の文を含まない文)を形成する規則であり、3以降は、1つないしそれ以上の文から別の1つの文を形成する規則である。

統語規則は文の集合を帰納的につくりだす。たとえば、規則1によって1項述語 B と名辞 m から原子文 $B(m)$ がつくられる。また、規則2によって2項述語 F と名辞 $a,$

t から原子文 $F(a, t)$ がつくられる。さらに規則3によって文 $B(m)$ と文 $F(a, t)$ から文 $B(m) \cap F(b, t)$ がつくられる。このようにして、文が無限につくられていく。

次に言語の意味論である。われわれの意味論は各基本表現に意味値を与えるところから出発する。意味規則の役割は、基本要素の意味値にもとづいてより大きな構成要素の意味値を決定することである。

たとえば、名辞 a, t, m は各々「安倍晋太郎」、「竹下登」、「宮沢喜一」なる個体をその意味値としてとるものとしよう。また、1項述語は、その意味値としてその述語が真となる個体の集合をとるものとする。たとえば、 B の意味値を頭のはげた個体の集合とし、 G の意味値を眼鏡をかけた個体の集合としよう。すると、文 $G(a)$ は安倍晋太郎が眼鏡をかけているときに真となる。2項述語の意味値は個体の対の集合である。 F の意味値として1番目の個体が2番目の個体を友人であると考えているような対の集合としよう。たとえば、安倍晋太郎が竹下登を友人であると考えているのなら、すなわち F の意味値の中に対 (安倍晋太郎, 竹下登) が含まれているのなら、 $F(a, t)$ は真となる。このようにモデル論的意味論においては意味論を集合論的に展開できるのである。

表現 a の意味値を $\langle a \rangle$ と書くことにすると L_0 の基本表現の意味値と意味規則が次のように書ける。

・ 基本表現

$\langle a \rangle$ = 安倍晋太郎

$\langle t \rangle$ = 竹下登

$\langle m \rangle$ = 宮沢喜一

$\langle G \rangle$ = 眼鏡をかけている個体の集合 = $\{\langle a \rangle\}$

$\langle B \rangle$ = 頭のはげている個体の集合 = $\{\langle m \rangle\}$

$\langle F \rangle$ = 1番目の個体が2番目の個体を友人であると考えている対の集合 = $\{(\langle a \rangle, \langle t \rangle), (\langle m \rangle, \langle a \rangle)\}$

・ 意味規則

1. δ が1項述語で、 c が名辞ならば、 $\delta(c)$ が真となる必要十分条件は $\langle c \rangle \in \langle \delta \rangle$ 。
2. γ が2項述語で、 α と β が名辞ならば、 $\gamma(\alpha, \beta)$ が真となる必要十分条件は $\langle \alpha \rangle, \langle \beta \rangle \in \langle \gamma \rangle$ 。
3. ϕ が文ならば、 $\neg \phi$ が真となる必要十分条件は ϕ が真でないこと。
4. ϕ と ψ が文ならば、 $[\phi \cap \psi]$ が真となる必要十分条件は ϕ と ψ が共に真であること。
5. ϕ と ψ が文ならば、 $[\phi \cup \psi]$ が真となる必要十分条件は ϕ または ψ が真であること。

6. $\phi\psi$ とが文ならば, $[\phi\rightarrow\psi]$ が真となる必要十分条件は ϕ が偽か, あるいは ϕ が真であること。
7. ϕ と ψ が文ならば, $[\phi\leftrightarrow\psi]$ が真となる必要十分条件は ϕ と ψ がともに真か, あるいはともに偽であること。

これらの意味規則と基本表現に対する意味値の割当が与えられれば, すべての文の意味(真理値)を決定することができる。

例えば, 文 $G(m)$ は, 規則1により $\langle m \rangle \in \langle G \rangle$ のとき真となる。しかし, いま定めたモデルによると $\langle G \rangle = \{ \langle a \rangle \}$ であり $\langle m \rangle$ は $\langle G \rangle$ に含まれていないから文 $G(m)$ は偽となる。

1項述語の意味として, その述語が真となる個体の集合をとった。情報検索という状況において1項述語に相当するものはある属性であり, その属性を表現するディスクリプタである, と考えられる。1項述語「G」の意味値が「安倍晋太郎」であったように, 1項述語「参考業務」の意味値を{参考業務について述べている文献} = {文献1, 文献3, ……}と定めることができるのである。1項述語「参考業務」を満たす個体は文献1と文献3と……という具合に問題となる世界(モデル)を関与する限りきめ細かく完全に規定してはじめて意味論を展開することができる。この規定を行うのが索引者の仕事である。

C. 代数的定式化

前節で述べたことを代数的に定式化することができる。代数とは数学的構造の記述であり, 統語論とは言語学的構造の記述である, という考えれば, このような定式化を試みようとするのは, 構造主義の名を出すまでもなく, ごく自然な発想であろう。

ある集合が, その上で定義された演算に関して閉じている(この集合の元に演算を行った結果が再びこの集合の元となる)とき, この体系を代数系(あるいは単に代数)という。例えば, 有理数の全体は通常の加法と乗法, それに逆元を対応させる演算3に1/8を対応させる)とによって体(field)という代数になる。(加法や乗法のように2つの元にある1つの元を対応させる演算子を2項演算子, 逆元演算子のように1つの元にある1つの元を対応させる演算子を1項演算子という)これらの演算子によって閉じていることは容易にわかる。整数の集合は加法と乗法に関して閉じているものの逆元演算子に関しては閉じていない。つまり, 整数の全体は体とい

う構造をなしていない。(環という構造になっている)

言語 L_0 の統語論を考えたとき, 集合としては文の集合をとる。1項述語は1つの名辞を引数とする1項演算子であり, この演算の結果は文である(統語規則1)。2項述語は2つの名辞を引数とする2項演算子であり, この演算の結果もまた文である(統語規則2)。例えば, 1項演算子 G をこのシステムのように解釈するなら $G(a)$ は「安倍晋太郎は眼鏡をかけている」と解釈される文になる。 G, B, F のような演算子は言語 L_0 に固有な演算子であるけれど, どのような言語体系においても用いられる一般的な演算子と見なされているのが論理記号 $\neg, \cap, \cup, \rightarrow, \leftrightarrow$ である。 \neg は1項演算子, 残りはすべて2項演算子である。これらの演算子に関しても文の集合は閉じている, ということを述べているのが統語規則の3から7である。即ち, 統語操作という演算に関して文の集合は閉じており, これをある代数構造と見なすことができるのである。この代数系を一般に統語代数と呼んでいる。

言語 L_0 の意味規則は統語規則に対応するように規定されている。したがって, 意味値の集合にも文の集合と類似の構造が入り代数となる。これを意味代数と呼んでいる。

言語 L_0 は説明のためのおもちゃ言語であるから演算子(述語)の種類や数は多くない。より通常の言語体系ではさまざまな種類の演算子(述語)が用いられるであろう。そのような言語体系に対応する代数系は多ソート代数(ソートは品詞に対応すると考えてよい)という名で一般に論じられている¹³⁾。

ここに挙げた例から統語論を代数的に論ずることがいかに自然なことであるかおわかりいただけたことと思う。我々の目的の1つは, 検索言語の統語論をより自由かつ明確に規定することにある。現在使用されているブールの検索言語の統語論は通常のブール代数をなしているが, これは1つの単純な例にすぎない。我々はもっと表現力に富む検索言語を作りたいのである。新しい言語の定式化は統語論をある代数として捉えることによって容易に行うことができる。というのは意味論もその代数に類似な(準同型な)代数として定式化できるからである。即ち, ある代数を核として情報検索の統語論と意味論とを平行に論ずることができる。例えば, ブールの検索言語の拡張としてブール演算子以外の演算子をも述語として含んでいるような検索言語を考えてみよう。その言語の統語論をブール代数と似た同じようなある代数と

して規定することができるだろう。

古典的命題論理は通常のブール代数として表すことができ、ブール代数は集合演算として表現することができた。そして、このつながりが情報検索の基礎をなす理論として用いられてきた。しかし論理体系は命題論理ばかりではない。人間の思考をより詳細に表現できる論理体系をめざして古典的命題論理の拡張や新しい論理体系の構築が行われてきた。その中には古典的述語論理もあれば直観的命題論理もある。更に命題様相論理もある。そしてこれらはそれぞれ完備ブール代数、擬ブール代数、位相ブール代数として表すことができるのである¹⁴⁾。これらの代数を統語論とするシステムが作れたなら、そのシステムの索検言語は現在使われている命題論理を越えた表現力を持つであろう。

いま通常のブール代数とそれに類するものだけをとりあげた。集合演算との対応を考えたからである。統語論として一般的な代数でいいのなら様々な述語を工夫して表現力豊かなシステムを作ることも可能であろうが、それに対応する意味操作として文献集合を操作しなければならないことを考えると、今のところブール代数の系列に絞らざるを得ない。

D. 索引とモデル論

これまでは情報検索における検索にのみ焦点をあわせて論じてきた。ここでは図書館情報学のモデル論的考察の1例として、索引をモデル論的に考えてみよう。

一般に文の意味は、その文の構成要素である語彙の意味を扱う語彙意味論と、文を形作っている文法体系に対応してこれらの語彙の意味から文全体の意味がどのように構成されるかを扱う構成意味論とによって決定される。モデル論的意味論として上に紹介したのは、構成意味論のほうである。これは明確に体系だてて論ずることができる。語の同義、類義等を扱うのが語彙意味論であり、通常、意味論と言ったときに、こちらを想定されることが多い。語彙意味論は、いまだに体系的分析の対象とはなっておらず、可能なのは、直観的ないし統計的な分析のみである。

語の意味を研究する方法として普通挙げられるのは、Carnap の意味公準による方法と、語彙素と呼ばれる比較的少数の一般の意味要素の組合せをもとにして語彙構造を記述しようとする方法である。意味公準は数学的理論における無定義用語や公理に相当するものであり、語彙素はその言語の語彙要素間の意味関係を記述するため

に設定された理論上の要素である。これらは、あくまでも、記述するための道具であって、記述者の意志で選ばれるものである。

データベース全体の規模で意味を問題にしようとするとき、1つの文献に含まれる意味内容を大幅に抽象し数語のディスクリプタとして表さなければ、大量のデータに対して対扱のしようがない。この作業をへて文献情報はデータベースに蓄積される。意味内容の抽象という作業を行うのが索引者（あるいは統計的処理機）である。1つの文献は1つの統語単位であるとみなせる。即ち、1つの文献についてさまざまな側面から分析を行うけれど、それらの分析データの全体が1つにまとまってはじめて意味あるものとなる。そこで言語になぞらえて言えば、文献は1つの語彙であると考えられる。索引者はシソーラスという基本語彙素集に則ってその語彙（文献）の意味を分解し、それを語彙素（ディスクリプタ）の集合（あるいはストリング）として表現するのである。つまり、情報蓄積検索システムというコミュニケーション・システムを意味論に対応づけて考えようとするとき、個々の文献に適切なディスクリプタを付与する索引づけ（indexing）というプロセスに対しては、語彙意味論が、索引づけによって付与されたディスクリプタにもとづいて要求された文献集合を構成する検索というプロセスに対しては構成意味論が対応するとみなせる。さらにモデル論的にいえば、索引づけとは1つの統語単位（文献）に対してシソーラスによって定まっている世界（対象領域）の元を割り当てる操作であり、この操作によって1つの構造（モデル）が定まるのだ、ということが出来る。このモデルが1つのデータベースを形成する。意味割当は索引者による文献の認識の表現であり、十分に分析的な語彙意味論が存在しない以上、最終的には索引者による意志決定に依存するものである。

Carnap は意味公準の決め方について次のように述べている。

『ある体系の創作者が、「独身」と「既婚」という性質を指示するためにそれぞれ述語“B”と“M”とを望むとしよう。これらの性質が両立不可能と言うことを、それ故公準Pを立てなければならないということをしてどのようにして彼は知るのだろうか。これは認識問題ではなく、意志決定に関する問題である。（中略）「黒い」と「カラスのぬれ羽色」という語に対応させるのに述語“BI”と“R”を彼が望むとしよう。「黒い」の意味はまったく明確であるのに、「カラスのぬれ羽色」の意味は

日常言語ではかなり曖昧である。「カラスのぬれ羽色」がいつもあるいは大抵の場合「黒い」を合意しているかどうかを見いだすために、普通の用法の内省や統計的調査に基づいて熱心に研究しても彼に有利な事は何もない。彼の体系の述語“*R*”と“*BI*”とが、前者の“*R*”が後者の“*BI*”を論理的に合意するような具合に、使用されることを彼が望んでいるかどうかを決めることがむしろ彼の仕事である。もしそのように望んでいるならば、彼はその体系に公準

$$“(x) (BI(x) \supset R(x))”$$

を加えなければならないし、そうでないならば加える必要はない¹⁵⁾、

認識問題ではないと彼が言うのは、事実の認識という問題に関わることが意味論の本来ではない、という意味であろう。あるシステムの意味論にとって重要なのは、それがコミュニケーションの道具として機能するということであって、認識論を論ずることにあるのではない。システムの意味論に対するこのような姿勢は、事実の認識に対して我々がとった公理的方法と同じものである。究極の認識などと言うものが望むべくもない以上、システム設計者はどこかで手を打たなければならない。そして、その決定を意味公準として明確に言明しなければならない、と Carnap は言っているのである。また、統計的調査が役に立たぬとは、統計的事実が意味を決定するわけではない、という意味である。もちろん何らかの参考になるだろうが、あくまでも決定するのはシステム設計者の役割なのである。

これまで述べたことをもとに、情報蓄積検索システムを全体的に見ると次のようになる。索引づけによって文献とディスクリプタとの間にある対応がつけられた。1度、対応がつけられたなら、文献とディスクリプタのどちらを意味値と考へてもさしつかえない。文献の意味値としてディスクリプタの集合をとるのが索引づけのプロセスであり、ディスクリプタの意味値として文献集合をとるのが検索プロセスである。より正確にいえば、情報検索における問い合わせ（検索式）はいくつかのディスクリプタの組合せからなっているので、それぞれのディスクリプタの意味から問い合わせ全体の意味を構成する作業が必要である。この作業は、検索式に含まれている統語演算子に対応する集合演算子を意味演算子として、ディスクリプタの意味である文献集合から検索式全体の意味を構成するものである。1つ1つのディスクリプタの意味値である文献集合から問い合わせ全体の意味値を

作っていく過程はまさに構成の名に値する。

情報検索とは統制語彙とそれにもとづく索引という意味解釈関数（言語表現と可能世界との対応を規定するもの）に従って問い合わせ（検索式＝ディスクリプタから構成される文）を解釈し、意味値として文献集合を提示する意味解釈システムなのである。

IV. おわりに

情報検索は、集合とブール代数の概念を中心にして理論構成が行われてきた。ブール代数とは、かなり広い内容を含んだ代数系の総称である。この代数系の特徴は、そこに含まれている代数の各々が、それぞれある論理体系に対応しているというところにある。通常のブール代数には古典的命題論理、位相ブール代数には様相論理、という具合にである。

ブール代数はこのような多様性と論理の意味をもっているにもかかわらず、これまで情報検索において用いられてきたのは通常のブール代数のみであった。そして、これまでの情報検索システムの柔軟性のなさはこのブール代数に対応している命題論理のそれであった。

一方、モデル論的意味論からの示唆によって、ある言語体系をモデル論的に定式化するなら統語論と意味論を共に類似な代数として捉えられることがわかった。それなら、情報検索をモデル論的に定式化すればその統語論と意味論をある代数として統一的に論じることができよう。そしてこのように代数的な取扱いを行うことによって、情報検索の統語論と意味論をブール代数に対応している様々な論理体系のもとで展開する可能性が開けてくるのである。

本論文はこれらの理論に必要な概念を説明し、その基礎的な部分の展開を行った。基本的な枠組みはモデル論である。これは意味論を展開するときの最も論理的な概念装置であり、意図や意味の伝達を扱う場面への適用が注目されているものである。この枠組みの中で図書館情報学を考へてみようではないかという提案を行うことが本論文の目的の1つであった。情報検索論を様々な論理体系のもとで展開しようとしたときに必然的に登場する代数的定式化はモデル論的考察の1つの例として述べたものである。

論理学ないしは代数的に情報検索論を展開してみても、それがすぐにあるシステムの実現に結び付くとは限らない。実現するとしてもそれにはかなりの時間を要すると考へるのが妥当であろう。にもかかわらず、あえて

理論的展開に固執した。「理論は我々が考える。実現についてはコンピュータ屋さんが考えてくれ」という立場があってもよかろうと考えたからである。むしろこの方が図書館情報学の本来の姿に近いと思うのだがいかなるものであろう。

本論文は新しい成果を展開したものではない。新しい枠組みを解説し、先行研究を紹介することによってその有効性の一端を示したにすぎない。このような研究は始まったばかりであって展開はこれからである。ここに述べたことはその出発点の明示であり、いわば研究プログラムの提示であった。このプログラムにそっての研究が次の仕事になる。

- 1) Tarski, A. 真理の意味論的観点と意味論の基礎。In: 坂本百大編。現代哲学基本論文集Ⅱ。勁草書房。1987。p. 51-120.
- 2) Dowty, D. R, Peters, S. and Wall, R. E. モンテギュー意味論入門。井口省吾等訳。三修社。1987.
- 3) 石川彰。モデル論の意味論を用いた言語学。In: 太田朗, Felix Lobo 編。海外言語学情報 第2巻。p. 3-15。
状況意味論の原典は次の図書であるが、数学的に整理されているとは思えない。
Barwise, J. and Perry, J. *Situations and Attitudes*. MIT Press. 1983.
- 4) 中島玲二。数理情報学入門スコット・プログラム理論。朝倉書店, 1982。p. 11.
- 5) Gallaire, H., Minker, J. and Nicolas, J. M. ed. *Logic and Data Bases*. Prentice Hall Press. 1978。
が最初のものである。引用した彼らの論文はこれに含まれている。この方面に関しては次の論文集がある。
Gallaire, H., Minker, J. and Nicolas, J. M. ed. *Advances in Data Base Theory*, vol. 1. Prentice Hall Press. 1981.
Gallaire, H., Minker, J. and Nicolas, J. M. ed. *Advances in Data Base Theory*. vol. 2. Prentice Hall Press, 1984.
- 6) Reiter, R. On Closed World Databases. In: Gallaire, H. et. *Logic and Data Bases*, 1978, p. 55-75.
Reiter, R. 'A Logic for Default Reasoning.' *Artificial Intelligent*, vol. 13, 1980. p. 81-132.
- 7) Marek, W. and Pawlak, Z. "Information Storage and Retrieval Systems: Mathematical Foundations," *Theoretical Computer Science*, vol. 1, p. 331-354 (1976).
- 8) Lipski, W. Jr. "On Semantic Issues Connected with Incomplete Information Databases," *ACM Transactions on Database Systems*, vol. 4, no. 3, 1979. p. 262-296.
Lipski, W. Jr. "On Databases with Incomplete Information," *Journal of the Association for Computing Machinery*, vol. 28, no. 1, January 1981. p. 41-70.
- 9) 神野慧一郎。真理論の系譜。In: 新岩波哲学講座。経験 言語 認識。1986。p.281-312.
- 10) 公理的方法に関する記述は、次の図書を参考にした。
福山克。数理論理学。培風館。1980。
梅沢敏郎。記号論理学。筑摩書房。1970.
- 11) この例は、次の図書からとった。
神野慧一郎, 内井愨七。論理学：モデル論と歴史的背景。ミネルヴァ書房。1976.
- 12) この例は、2)にあった例を修正したものである。
- 13) 多ソート代数を用いてモンテギュー意味論を代数的に論じているのは Janssen, T. M. V. *Foundations and Applications of Montague Grammar Part 1: Philosophy, Framework, Computer Science*. Amsterdam, Center for Mathematics and Computer Science, 1986.
- 14) 論理を代数的に定式化している文献は多くない。代表的なものとして Rasiowa, H. and Sikorski, R. *The Mathematics of Metamathematics*. Polish Scientific Publishers, 1963。
日本語の文献は次のものしか見あたらない。
松本和夫。数理論理学。共立出版。1980。
松本和夫。情報数学1。森北出版。1980.
- 15) Carnap. R. 意味と必然性。永井成男訳。紀伊國屋書店。1973。p. 276.