「フレーム問題」の解消
――人工知能研究への一提言――

羽地　亮

0　はじめに

「フレーム問題（frame problem）」は、1969年にマッカーシーとヘイズ（McCarthy and Hayes 1969）によって提唱されて以来、人工知能（AI）研究において長らく難問とされてきた。しかも、人間の精神をコンピュータになぞらえる昨今の論調において、この問題は、単なる形式論理上の問題や工学上の問題にとどまらず、AI 研究によってはじめて明るみに出された「根の深い新たな認識論的問題」（Dennett 1984, p.148）とみなされる。すなわち、コンピュータにとってのみならず、人間にとってもまた「フレーム問題」は問題であるというのである。

これに対して、私は次のことを主張したい。①「フレーム問題」は現在の AI 研究に特有のアポリアである。そして、それがアポリアである限り、AI は実現しないだろう。②人間の精神に適用された「フレーム問題」は疑似問題である。人間にとって「フレーム問題」は存在しない。③なぜ人間にとって「フレーム問題」が存在しないのかを考察することによって、人間の認識や概念の構造についての知見が得られ、ひいてはこれが AI 研究の方向性に関する示唆を与える。

1　「フレーム問題」とはどういう問題か

1・1　R2D2への道

周知のとおり、「フレーム問題」の正確な定義は、今のところ存在しない。この問題が何を指しているかについて、研究者の間のコンセンサスは存在しない。しかしながら、あちこちで引用されている話なので気が引けるけれども、デネットがこしらえた次の物語は（Dennett 1984, p.147f.）、「フレーム問題」を差し当たって直観的に理解することに役立つだろう。

R1と名付けられた一台のロボットがあった。ある日、R1の予備バッテリーをしまってある部屋に時限爆弾が仕掛けられ、それはまもなく爆発するようにセットされていた。部屋には一台のワゴンがあり、バッテリーはその上にある。R1はバッテリー救出作戦を立てた。すなわち、PULLOUT（WAGON, ROOM）という行動を行えば、バッテリーを部屋から持ち出すことができると考えた。R1はただちにこれを実行した。ところが、不幸なことに爆弾もまたワゴンの上にあった。R1は爆弾がワゴンの上にあることを知っていたが、ワゴンを引っぱり出すことが、バッテリーと一緒に爆弾も持ち出すことになるということに気が付かなかった。自分が計画した行動のこの明白な帰結を見落としていたR1は、部屋の外で爆発してしまった。

技術者たちは考えた。ロボットは自分の行動の帰結として、自分の意図したものだけではなく、副産物についての帰結も認識できなければならない、ロボットは周囲の状況の記述を用いて自分の行動を計画するから、そのような記述から副産物についての帰結を演繹（deduce）させればよい、と。こうしたわけで、R1D1（robot-deducer）がつくられた。R1D1は R1と同じ苦境にたたされた。R1D1も、PULLOUT（WAGON, ROOM）を考えついた。それからR1D1は、設計されたとおり、この行動の帰結を考え始めた。R1D1は、ワゴンを部屋から引っぱり出しても部屋の壁の色は変わらないということを演繹し、ワゴンを引けば車輪が回転するだろうという帰結の証明にとりかかった。そのとき爆弾は爆発した。

技術者たちは考えた。われわれはロボットに、関係のある（relevant）帰結と関係のない（irrelevant）帰結との区別を教えてやり、関係のないものは無視するようにさせなければならない、と。こうしたわけで、R2D1（robot-relevant-deducer：分別のある演繹ロボット）がつくられた。R2D1も例の苦境にたたされた。すると、驚いたことに、このロボットは、部屋に入ろうともせず、じっとうずくまって考えていた。設計者たちは「何かしろ」と叫んだ。R2D1は「してますよ」と答えた。「私は、無関係な帰結を探し出してそれを無視するのに忙しいんです。そんな帰結が何千とあるんです。私は、関係のない帰結を見つけると、すぐそれを無視しなければならないもののリストにのせて、……」また爆発してしまった。

これらのロボットはみな「フレーム問題」に苦しんでいる。事態を的確に判断し、すばやくそれに対処できるようなR2D2ロボットをつくるためには、設計者たちは「フレーム問題」を解かなければならない。

１・２　「一般化フレーム問題」

汎用のプログラム可能なデジタルコンピュータは、1941年ドイツのコンラート・ツーゼによって開発され(1)、以後、急速な進歩を遂げた。コンピュータは、チェスやチェッカーのようなゲームや、論理学の定理の証明のように、世界が閉じているという仮説をとることができる場合は、大きな成功をおさめた。こうした仮説の下では、行為に先立って、その行為に直接は関係しない有限の周囲の状態（行為を囲む「枠（frame)」）について完全な目録をこしらえ、それを書き換えることによって次の動作に利用することができる。しかし、現実の世界のような開いた世界の場合は、行為を囲むフレームについての目録は、ほとんど無限なものとなり、それを完成させることはできない。こうしてコンピュータはフレーム問題に直面する。

マッカーシーとヘイズによる「フレーム問題」のオリジナルな規定では（McCarthy and Hayes 1969, pp.477-490)、電話を持っている人間 P が、電話帳で人間 Q の電話番号を調べて、電話をかけ会話するという状況が設定されている。こうした行為を古典的な形式論理で記述しようとすると、電話を持っている人間 P は、電話帳で電話番号を探した後でも、まだ電話を持っている、というような人間にとっては自明な条件をいくつも加えざるを得ない。こうした条件の記述の量は、膨大なものになってしまって手に負えなくなる。すなわち、一般的には、ある行為を形式論理で記述するとき、その行為のフレーム（その行為によって変化しない事柄）が変化しないということを効率的に記述するにはどうすればよいか、というのがオリジナルな「フレーム問題」である。

しかしながら、オリジナルな「フレーム問題」の定義は、狭すぎて知識表現の研究に有意義でない、というのがデネット（Dennett 1984）や松原（松原 1990）の主張である。特に松原は、オリジナルな定義を拡張して「一般化フレーム問題」を提唱している。デネットは、狭い方の問題に対する「解決」が真の困難を広い方の問題に押しつける可能性を示唆している。私としても、デネットや松原に賛同して「フレーム問題」をより広くとりたい（しかし、私は、デネットや松原とは違って、「フレーム問題」を人間にまで適用することは妥当でないと考える)。ここでは、松原の「一般化フレーム問題」を取り上げることにする。

松原によれば、「フレーム問題」は最初は形式論理の中の問題として議論が始まったので、マッカーシーら論理主義の AI 研究者は、非単調論理のような形式論理の拡張を「フレーム問題」にからめて議論する傾向がある（この場合の論理主義とは、計算機上では形式論理だけで知識を表現すべきであり、そうすることによって原理的に計算機は人間のような高度の知性がもてるはずである、とする立場である)。この傾向の原因は、論理

主義者が「フレーム問題」の定義を形式論理における記述の量を減らす問題に狭く限定しているためである。しかし、情報に対しては空間とともに時間も関係している。空間に対応するのが記述の量とすれば、時間に対応するのが処理の量である。知識表現の効率を考えるときには記述の問題と処理の問題とを込みにする必要があり、形式論理からのアプローチは、記述の量の処理の量への転嫁であったのである。

例えば、非単調論理において、様相記述の記号を M とすれば、「M（電灯がついている)」は、「電灯がついている」の命題が否定されていない限り「電灯がついている」と考えてよい、という様相命題である。この「M（電灯がついている)」という記述が存在すれば、もしも「電灯がついている」ことが偽と判明していなければ、「電灯がついている」を真と仮定して推論を進める（いかに状態が変化しようとも、電灯に関係のある行為がなされない限り、いちいち「電灯がついている」と断る必要はない)。後で「電灯がついている」が偽と判明したならば、この命題を真と仮定したことによって得られた定理はすべて取り消すことになる。しかし、「電灯がついている」という命題がある状態で真か偽かを決定するためには、その命題に関する記述を探して過去に遡って処理を行う必要があるので、それだけ余計に処理の量を要する。また、ある仮定が否定された際に、その仮定に基づいて生成された定理をたどって否定していく処理の手間も小さくない。「フレーム問題」に関しては、非単調論理は単に記述の量を推論の量に転嫁しているに過ぎない。

これに対して、松原の「一般化フレーム問題」は、記述の量を減らすとともに、処理の量を減らすことを新たな問題として考える。「一般化フレーム問題」の本質は、膨大な情報を、記述するにしろ処理するにしろ、いかに扱うかということである。記述の量と処理の量とが計算論的にトレードオフの関係にあることは自明であり、その一方だけの増減を議論するのは、知識表現の効率を考える上で無意味である。「一般化フレーム問題」は、情報処理の主体が膨大な情報のうちの一部しか参照することができないことから生じるのである。

以上、松原の「一般化フレーム問題」を概観した。しかしながら、松原が「解決」するべきであるとしたこの問題を、私は疑似問題として「解消」するべきであるとみなす。なぜそうみなせるのかが次の問題である。

2　なぜ「フレーム問題」は疑似問題なのか

松原が言うように、「フレーム問題」は、情報処理の主体が膨大な情報のうちの一部し

か参照できないことから生じるようにみえる。しかし、松原の議論には、ある哲学的な前提が存在し、しかも松原自身がそれをはっきり認めている（松原 1990, p.236）。それは「表象主義」といわれるものである。私がここで「表象主義」というのは、行為に対して認識が先行し、認識とは世界を抽象的に表象する記号を操作することである、という考え方である。世界からの「入力」を表象・記号に変換しこれを操作することによって次の行為を企図する（「出力」する）というのは、認識や行為の主体がコンピュータであれ人間であれ、ごく自然な考え方である。この自然な考え方は、AI研究者にとっては、人間がコンピュータと連続的な存在であり、したがって、人間もコンピュータも「フレーム問題」を抱えている、という主張を結果する。

これに対して、私は、AI研究者が「表象主義」の立場にたつ限り、AIの実現にとって「フレーム問題」は解決不可能な問題として立ちはだかる、と考える(2)。なぜなら、「表象主義」は行為と認識との分離を前提としており、この認識の側にあらゆる場合に記述の量も処理の量も爆発させずに情報処理を行わせるというのは、有限な情報処理能力しかもたない主体にとって不可能なことだからである。さらに私は、「表象主義」の立場にたつのではない、人間の知性の把握の仕方が存在する、と考える。そして、私はこのことを根拠にして、人間にとって「フレーム問題」は存在しない、と考える。以下にこれらを明らかにする議論を提示する。

２・１　メタ推論規則・サーカムスクリプション

前節でみたように、論理主義者は様々な非単調論理を考案して「フレーム問題」の解決をはかったが、いずれも失敗に終わっている。通常の（単調な）形式論理では、ある公理系がより包括的な公理系に拡張された場合、前者の公理系によって証明されるすべての定理は、後者の公理系によって証明される定理の中に含まれる。論理の単調性とは、公理が増加すれば定理も増加するような単調な関係のことをいう。ところが、非単調論理の場合、公理が増加すると、証明できる定理は減少する。以下にみるのは、マッカーシーが考案したサーカムスクリプション（circumscription）という一種の非単調論理である（McCarthy 1980, 1986）。この論理の特長は、世界の中の差し当たり非関与的な事柄を無視する最も端的な定義を含むことである。大澤が、マッカーシーの叙述をうまくまとめているので、これを利用する（大澤 1990, pp.271-276）。なぜすでに不首尾に終わるとわかっている議論を紹介するかといえば、サーカムスクリプションの限界を見通すことによって、「フレーム問題」を生み出している真の源泉が露になるからである。

サーカムスクリプションは、特殊なメタ推論規則をともなった二階述語論理である。

その中には「対象Oが、事実Aのもとで性質Pをもつことが推論されたとき、性質Pをもつ対象はOに限る」と解釈できるような推論規則が定義されている。この推論規則は、性質Pをもつものを Oだけに囲い込み、O以外の他の対象も性質Pをもつ可能性があるにもかかわらず、それらについては無視することを定めている。

定義：論理式 A(P) における述語 P のサーカムスクリプションとは、次のような文図式である。

$$A(\Phi) \wedge \forall x\,(\Phi(x) \supset P(x)) \supset \forall x\,(P(x) \supset \Phi(x)) \qquad (1)$$

$\Phi(x)$ は任意の論理式。

ここでは一つの変数 x のみの場合を書いたが、一般的には n 個の変数の組で考えるべきである。A(P)――すなわち述語 P(x) を含む論理式――が定理であるとき、上記の文も定理になる。

この推論規則は、次のことを意味する。論理式Φで表現できるような性質や関係を満たす対象 x が、述語 P が表現するような性質をもつことがわかっているとき（つまり、$\Phi(x) \supset P(x)$ のとき)、P を満たす x はすべてΦをも満足するとみなすことによって、ΦとPがまったく合致していると考えるということを、この規則は含意する。$A(\Phi)$ は、Pによって満足されている条件が、Φによっても満足するという仮定を表現する。

例：ブロック世界において、次の文Aが成り立っているとする。

$$\text{isblock a} \wedge \text{isblock b} \wedge \text{isblock c} \qquad (2)$$

すなわち、a も b も c もブロックであると主張されている。この isblock という述語のサーカムスクリプションは、定義によって次のように書き表すことができる。

$$\Phi(a) \wedge \Phi(b) \wedge \Phi(c) \wedge \forall x\,(\Phi(x) \supset \text{isblock x}) \supset \forall x\,(\text{isblock x} \supset \Phi(x)) \quad (3)$$

ここで、

$$\Phi(x) \equiv (x = a \vee x = b \vee x = c)$$

とおけば、サーカムスクリプションの図式 (3) の前件は当然真になるから、

$$\forall x\,(\text{isblock x} \supset (x = a \vee x = b \vee x = c)) \qquad (4)$$

と結論することができる。これはブロックが、aかbかcに限るということを意味している。つまり、ブロックであるということが、最初の文 A（(2)のこと）において主張

されていた対象のみに限定され、他の対象がブロックである可能性を無視していることになる。

ところで、ある公理系のモデルとは、その公理系のすべての公理が満足するような、各変数および論理式への真理値の割当法のことである。モデルとは、いわば、その公理系によって記述されている世界のことである。また、公理系の中の述語 P に関する極小モデルとは、公理系の公理を満足し、また P が真となるようなモデルの中で、真である論理式の範囲が最も小さくなるもののことである。極小モデルとは、公理系が記述していると解釈できるような（複数の）世界の中で、最も小さいものである。このとき、次の定理が成り立つ。

定理：公理系 A の中の述語 P についての任意のサーカムスクリプションは、P についての任意の極小モデルにおいて真である。

証明：M を P についての A の極小モデルとする。P′ を (1) の前件のΦの代わりのものとする。前件の右半分より、P は P′ を拡張したものである。もしここで (1) の後件が満たされなかったならば、P′ は P の真部分であることになる。この場合には、P 以外のすべての述語をそのままにし、P を P′ に置き換えれば、M の真部分モデル M′ を得ることができる。これは、M の極小性の仮定に反する。

これらを考慮した上で、次のような記法を導入すると、サーカムスクリプションの直観的にわかりやすい別の形式的表現を得ることができる。すなわち、$\forall x\,(U(x) \supset V(x))$ を、$U \leqq V$ と表しておく。これは、U のモデルが V のモデルの部分集合である、ということである。これを使えば、述語 P についてのサーカムスクリプションは、次のように書き表すことができる。

$$\text{Circum}\,[N(P) \wedge (U \leqq P) : P] \equiv N(U) \wedge (U = P) \qquad (5)$$

N(P) は、P が肯定形としては出現しない論理式。

Circum [.... : P] は「P についてのサーカムスクリプションは…」と読む。

(5) の基本的含意は (1) と同じだが、(5) の方が単純で、サーカムスクリプションという操作がいかなるものであるかをわかりやすく示している。これは、$U(x) \supset P(x)$ のとき——U が P のモデルであるとき——、P の定義として U を採用してしまうことによって、P を最小化する操作である。このとき、U 以外にも P という性質をもつ対象が存在するかもしれないのに、そのことは全く無視される。

しかしながら、サーカムスクリプションと名付けられたこの推論を行う手順が機械的

な仕方で定義できたとしても、たいてい、その手順にしたがった推論の量は膨大になる。つまり、サーカムスクリプションは、意に反して、無視の操作を非効率的にしか代行しない。サーカムスクリプションによる推論は、ちょうどR2D1と同じように、多くの時間をかけて、無関係な部分を無視するのである。

２・２　サーカムスクリプションの限界と「フレーム問題」の真の在処

ところで、実はサーカムスクリプションの計算可能性は一般的には保証されていない。サーカムスクリプションは、対象となる論理式 A(P) が、次のような意味で、述語 P に関して孤立している場合に、計算可能なのである。

定義：論理式 A(P) が P に関して孤立しているということは、それが次のような形をとっている場合である。

$$N(P) \wedge (U \leqq P) \tag{6}$$

ここで N(P) は肯定形の P が出現しない論理式であり、U は述語 P を含まない述語の組である。なお、P が単一の述語であるとは限らない。一般的には P は m 個の述語の組であると考えるべきである。

P に関してサーカムスクリプションをほどこした場合には、P のモデル U がそのまま P の定義として採用されることになる。(6) は次のことを意味している。A(P) が P に関して孤立しているということは、U が P を含んではならないことから端的に示されるように、A(P) の中で P が P 自身を用いて自己参照的循環的に定義されていないということに他ならない。このような自己参照的循環的定義が排除されているとき、サーカムスクリプションの計算可能性が与えられる。

自己参照的に定義される述語にサーカムスクリプションが適用できないのはなぜだろうか。サーカムスクリプションとは、その述語についての極小モデルを指定することである。ところが、述語の定義が自己参照的であるとき、極小モデルが得られない。なぜならば、P を定義し、P に取って代わられるべき述語 U の中に再び P が見出されるため、その P に関して同じ置き換えが反復されなくてはならないからである。このような反復は、P が P を含む U によって自己参照的に定義されている以上は、決して終わらない。つまり、P のどんなモデルに対しても、それよりさらに小さいモデルを見つけることができるため、極小モデルには絶対到達しない。それゆえ、このような P に対しては、サーカムスクリプションが計算できないことになる。

しかしながら、サーカムスクリプションが、無視する操作を表現することによって「フ

レーム問題」の解決を企図しつつも無益な試みに終わったことを鑑みれば、サーカムスクリプションがそれに対して禁止されているような操作の内にこそ、「フレーム問題」の問題性を捉える手がかりが隠されてはいないだろうか。「フレーム問題」とは、例えば「以下同様の仕方で」とか「他の事情が等しければ」という句によって表されるような操作を効率的に記述できないという問題である。「以下同様の仕方で」とは、行為を、したがってそれを表現する述語を、循環的に定義する方法の一種である。サーカムスクリプションは、自己参照的循環的定義を、推論の対象から排除してしまったが、むしろ、自己参照性・循環性は概念を定義したり説明したりするときの基本的な形態である。集合論のような数学の基礎的な部分を参照してみれば、循環を含んだ仕方で集合を定義することがしばしば行われる。より一般的には（羽地 1997)、われわれの日常の概念を説明するとき、当の概念の必要十分条件を与えることは普通不可能である。なぜなら、言語を説明する事象が当の言語に依存する結果として、説明と説明されるものとは区別できなくなるからである。日常使われる言語・概念の説明ということは、自己参照的構造、循環的構造をなしているのである。

したがって、われわれは、概念に包含される具体的事象を順次挙げていくことでしか説明を行うことができない。例えば、何かを「知っている」とはどういうことか説明するとき、「知っている」という言葉の使用が記述される。このとき説明はそのまま説明されるものとなる。そしてその説明には、原理的には終点がない。概念の内包や外延といったものはなく、ただ、概念に包含される具体的事象が、多様な類似点と相違点との重なり合いにおいて、無限に連なっているばかりである。その無限の連なりが「以下同様の仕方で」のような句によって代表されるのである。このような曖昧な記述を、その曖昧さを保存したままでコンピュータにプログラムすることが極めて困難であることは、容易に見て取れる。コンピュータにとっての「フレーム問題」は、われわれが日常的に用いている概念の自己参照的循環的構造に存する。

２・３　認識は行為である

さらに問を進めよう。なぜわれわれ人間の概念は、自己参照的循環的構造をもっているのか。例の「表象主義」にしたがえば、概念や認識は、われわれの行為を導く規則のような役割を果たさなければならない。しかし、「入力」の後の情報処理の結果がこのように曖昧なものであるならば、情報処理の主体はどのような「出力」を行うべきか困惑することになろう。

こうした問題に答えるための鍵は、知的システムにおけるある事実を考察することで

手に入るだろう。佐々木が紹介しているヘルドとハインの古典的な実験は、認識が行為に他ならないことを劇的に明らかにしている（佐々木 1987, pp.17-20)。誕生時から暗闇の中で母ネコとともに育てられた同腹の子ネコ五対が、歩けるようになるのを待って、特殊な状況下で「見る」体験を与えられる。その際、対にされたネコの片方は、装置内のどこへでも自ら移動できる。ところがもう片方のネコは、もう一方のネコの動きを忠実に同時に再現するゴンドラに乗せられて、無理矢理移動させられる。自由に移動するネコ（行為しているネコ）もゴンドラのネコ（移動はするが行為はしていないネコ）も、平等に一日三時間、十日間にわたり、こうした条件下での「見る」体験を与えられる。

ネコたちが見る能力を獲得したかどうかは、見ることと密接な関連をもつ行動を基準にして観察された。第一に、実験者がネコを両手でもち、ゆっくりと机の上に下ろすときに起こる、足で机の面の近づきを予測するような着地姿勢の有無、第二に、実際にはずっと同じガラスの板に覆われているが、見えとしては途中から床がなくなり落ちてしまいそうになっている「視覚的崖」を回避できるかどうか、第三に、突然目の前に現れる人の手に対して目を閉じる「瞬目反射」の有無、これら三種類の行動が基準となった。いずれの行動も動物の生存にとっては不可欠であり、最も基本的な視覚行動である。

観察の結果、すべての自由に移動するネコには、これら三種類の視覚行動が現れ、彼らが見る能力を獲得したことを示した。ところが、すべてのゴンドラのネコは視覚行動の兆候すら示さなかった。彼らに与えられなかったのは、自ら引き起こした身体の動きが見えの変化の原因になるという経験であった。確かにゴンドラの上でネコの足は動いていた。しかしその動きは、動物の身体の動きが本来もっている、動きが見えの変化と密接に対応しているという性質を失っていた。したがって、彼らにとって、「見る」体験は、視覚の機能の獲得には繋がらなかったのである。

こうした事実が示しているのは、ネコにおいては、見ることが行為することを含意するということである。「表象主義」は行為と認識との分離を前提としていたが、動物の実験結果では、行為と認識とは緊密な一体性を有していることになる。

それでは、人間ではどうか。佐々木の報告は、またしても興味深い（佐々木 1987, pp.27-29)。人間の眼球は、人間がものを見ているとき、常に活発に動いている。その動きの中には、肉眼ではほとんど捉えることのできない非常に速い周期の揺れである「眼震」、痙攣的な左右のふるえとして現れる「眼振」、ゆるやかな「漂流」、視点を次々に変えていく際に現れる「飛越」などがある。そこで、鏡の付いたコンタクトレンズを用いて眼球の微動と対象の像の動揺とを同調させることで、眼球の微細な動きによる効果を相殺し、「静止網膜像」をこしらえてみる。このような動かない像は、もし視覚が対象を

受動的に写し取ることによって成立する認識であるならば、理想的な光学的映像である。しかしながら、このような揺れない網膜像は奇妙な知覚をもたらす。

静止網膜像を与えられた知覚者が見たものは、意味ある部分（例えば対象が顔ならば、鼻や首や頭のような部分）が明滅するように消え去りそしてまた現れる、対象の像の不安定な変転であった。網膜像の静止は正常な見えを崩壊させるのである。

外界の安定した見えを得るためには、まず、「眼震」や「眼振」といった眼球の細かな動きが重要であり、さらにこのような微動を基層として、眼はさらに大きく動いている。対象に向かう眼球は、まず一点を凝視（停留）し、次々と起こる「飛越」運動によって、停留点を移動させていく。その動きは年齢によって大きく異なる。三歳児の停留点の動きは対象の中心に集中し広がりをもたない。一方、六歳児の停留点の動きの量は三歳児の四倍であり、その軌跡はまるで指で図形の輪郭をなぞるようである。そして、眼球の動きの発達差は、図形の見えの成立と密接に対応している。六歳児にとっては、先に見た図形と同じ図形を、後から与えられたいくつかの図形の中から正しく選択することは容易である。しかし、このような簡単な図形の再認に多くの三歳児は失敗する。

眼球の運動は、対象のある場所に移動して対象の輪郭や表面を「なぞる」行為の代わりをしているのである。視覚という認識は、対象を受動的に写し取ることによっては成立しない。人間においても、見ることは行為することに他ならないのである。

以上の心理学上の事実から、次のように考えることができる。われわれは、認識から生じた「表象」に導かれて行為するのではない。そうではなく、認識は行為に他ならないのである。この行為は、心に浮かぶ何か規範的なものに導かれて生じるのではない。誇張して言えば、それは端的に生じるのである。このことは、われわれ人間の自然的事実である。先に言及したように、「知っている」という概念を自己参照的循環的な仕方でしか説明できないのは、この概念が行為だからである。例えば、「手を挙げる」行為は、明示的に定義することはできず、強いて説明するなら、このような手の挙げ方があり、あのような手の挙げ方があり、以下同様、と自己参照的循環的に語るしかない。より厳密に言えば、ある行為がその行為として同定されるのは、行為がなされる状況に依存してのことであり、その状況がどのような状況か説明するためには、またも当の行為に言及せざるを得ない。すなわち、行為の説明にとって、自己参照的循環的状況依存性ということは、本質的なことなのである。

「フレーム問題」が現在のAI研究に特有のアポリアであることも、今や説明がつく。AI研究の主流は「表象主義」に依拠しており、行為と認識とを分離して考える。そして、認識の局面だけを取り上げてそれを形式論理で完全に明示化してプログラムに書こうと

するが、行為としてのわれわれの概念の自己参照的循環的状況依存性がAIを悩ませるのである。また、一方の行為の局面の処理には永遠にたどりつけない。AIの実現のための第一歩は、「表象主義」を捨てることであろう。

２・４　概念は身体的である

それでは、概念の自己参照的循環的状況依存性ということに直面しているわれわれ人間は、なぜ「フレーム問題」に悩まないのだろうか。「以下同様の仕方で」によって、なぜわれわれはお互いに了解できるのだろうか。解答の第一段階は、それはAIと違ってわれわれ人間においては、認識と行為とが一体となっているからである、というものであろう。しかし、これでは完全な解答にはなっていない。われわれがある概念を説明するときに、原理的には終点がないはずの説明をどこで切り上げればよいのだろうか。換言すれば、概念の説明と説明でないものとをどうやって区別しているのか。認識と行為との一体化の議論は、確かに、われわれの概念の自己参照的循環的状況依存構造の由来を明らかにするが、こうした概念がいかにしてわれわれにおいて運用されているのかについての知見を教えてはくれないのである。

残念ながら、私はここで完全な答案を提出できないことを告白しなければならない。しかしながら、解答への手がかりは確かにある。レイコフの認知言語学は、人間の概念一般が、人間の心のはたらきのみによって生み出されてきたのではなく、人間の身体性と深く関わっていることを、豊富な研究事例の引用によって明らかにしている（Lakoff 1987)。

例えば、人間の色彩概念は、人間の生理学的な条件に大きく制約されている。目と脳の神経経路の研究によれば、色彩概念に関わる６タイプの細胞がある。それらは、色相を決定する４つの細胞と明度を決定する２つの細胞である。色相を決定する４つの細胞は２組の対になっていて、一方の対は青と黄の知覚に関わり、もう一方の対は赤と緑の知覚に関わる。それぞれの細胞の興奮の度合いの様々な複雑な組み合わせによって、様々な色彩が知覚される。基本的な色彩語は言語によって様々であるが、ある言語に基本的な色彩語が２つしかない場合は、それらは黒と白である。基本的な色彩語が３つのときは、黒と白と赤である。４つのときは、その４番目は黄か青か緑のいずれかである。言語が概念体系を決定するというようなウォーフ流の考え方は、必ずしも真ではない。色彩概念は身体化されている。

また、古典的なカテゴリー理論では、カテゴリーを決定する属性はすべての成員により等しく共有されており、カテゴリー内において特別の地位をもつ成員は存在しないは

ずである。しかし、最近の研究の示すところでは、カテゴリーの成員の中には、そのカテゴリーに極めてふさわしい中心的な成員から、他のカテゴリーにも同時に分類されてしまいそうな疑わしい成員まで、多くの段階がある。カテゴリーの中で最も中心的な成員は「プロトタイプ」と呼ばれる。「鳥」のカテゴリーの中では、コマドリはニワトリ、ペンギン、ダチョウに比べて「プロトタイプ的な」成員である。「椅子」のカテゴリーの中では、机用の椅子は揺り椅子、床屋用の椅子、電気椅子に比べてプロトタイプ的である。そして、どの成員がプロトタイプ的であるかを決める基準は、ゲシュタルト知覚、心的イメージを引き起こす力、学習・記憶・使用の容易さ等々の身体的なものである。

さらに、古典的なカテゴリー理論では、例えばカテゴリー間の次のような階層構造（タクソノミー）がつくられる。論理的にはそれぞれのレベルは対等であり、どれかが特権的な地位を与えられるというわけではない。

始発点（植物、動物）
生活形（木、藪、鳥、魚）
中間形（広葉樹、針葉樹）
属（オーク、カエデ）
種（サトウカエデ、ホワイトオーク）
変種（葉に切れ込みのあるウルシ）

しかし、ツェルタル語を話す人々の言語の研究を通じて分かったのは、これらのレベルの中に優先的に関わりの対象として選ばれるレベルがあるということである。ジャングルの中で原話者に目に付いた植物の名前を言ってもらうと、原話者は種のレベルではなく属のレベルの名前を言う傾向があった。（話者は種の区別をする能力があり、その名称も知っていた。）原話者はまた生活形のレベルや中間形レベルの名前を言うことはなかった。この属のレベルが「基本レベル」と呼ばれるものである。属のレベルが基本レベルである理由は、またしても、それが最も容易に知覚し、意見が一致し、学び、覚え、名付けることができるレベルであるという、身体的な条件である。

なお、基本レベルは文化によってある程度異なる。例えば、都会の文化において、人は木というカテゴリーを基本レベルとして扱うかもしれない。また樹木の専門家は特異なカテゴリーを基本レベルとするかもしれない。

これらの研究はまだ暫定的な部分を残しており、われわれの概念の運用のメカニズムを完全に明らかにするほど、完成されてはいない。そして、これらの研究は、われわれの概念の運用に対してよりも、むしろ、われわれの概念の習得過程に対して焦点を合わ

せる傾向がある。しかしながら、明らかにわれわれは、概念を用いるとき、われわれの文化に根ざした根拠をもつとともに、われわれの身体的条件に根ざした根拠をももっている。すなわち、われわれの身体は、概念の説明と説明でないものとを自ずから区別している。「以下同様の仕方で」という了解の仕方は、われわれの精神にとっては曖昧なものとしかみなされないが、われわれの身体にとっては、それはほとんど明晰判明な了解の仕方なのである。したがって、概念は身体的であるという斬新な洞察は、さらに掘り下げて研究される価値をもつと思われる。この身体性の洞察においてこそ、人間の知能が満たしていて機械の知能が欠いているものを、最も明示的に解明できる見通しがある。

3　結語

断っておくが、私は、ドレイファスのように（Dreyfus 1992）、人工知能の研究を否定したいわけではない。ここで私が主張したのは、「表象主義」に基づく研究は、まったく実りがない、「フレーム問題」を解くことはできない、ということだけである。したがって、並列分散処理（parallel distributed processing）のようなモデルについては、ある程度の共感すらもっている。しかし、ここでこの新しいAIのモデルについて論じる余裕はもはやない。

ところで、ここまでの議論において、「フレーム問題」の疑似問題性を追求することによって、人間の認識や概念について、一定の知見が得られた。私はこの知見がAI研究においても軽視できない論点を含んでいるのではないかと思う。これをまとめることによって、AI研究への提言としたい。

(1) AI研究が人間の知性をシミュレートしようとするとき、「表象主義」ではない、知性の把握の仕方がある。それは、人間の知性が身体と密接不可分に結びついた仕方で情報を処理しているということである。人間の身体性において、関与的な情報のみを取り上げ、非関与的な情報を自然に無視する仕組みが必ず備えられている。認知に関する諸科学の進展に伴って、やがてこの仕組みは全貌を現してくるはずである。こうした科学と手を携えて、身体性をシミュレートすることは、非常に困難ではあろうけれども、不可能ではない。

(2) 動物であれ人間であれ、知的システムが行う行為は、頭の中の表象・規則・命令にしたがってなされるのではない。もちろん、頭の中の規則や命令にしたがって行為するという場合がまったくないとは言えない。しかし、行為は原理的には規則や命令を必要とはしない。正しくは、何らかの状況が行為を生じさせると言うべきである。そして、

言うまでもなく、この行為を生じさせる状況には、やはり身体的条件ということが重要なモメントとして含まれる。AI を備え、閉じた実験室ではなく開いた現実の世界で行為するロボットをつくるにおいても、諸科学と手を携えて、身体性の理解を深めなければならない。

(3) 本論文では、示唆するにとどまっているが、AI に思考させたり行為させたりしようというときに、身体性だけを考慮するわけにはいかない。いかなる知的システムも文化や社会を生み、その文化的社会的制約の中で活動している。この文化的社会的制約が、私がこれまで強調してきた身体性を条件づけている場合さえ十分に想定可能である。文化や社会を生まない知性というのは、ほとんど矛盾した概念ではなかろうか。AI 研究者にとって予想すらしない提言であろうが、私は、知性に対する文化的社会的な可能性の制約を明らかにすることが、AI 研究にとっても重要なことであると考える。したがって、AI 研究者は、人文諸科学、社会諸科学の知見にも耳を傾けるべきである。

注

(1) 一般には世界初のデジタルコンピュータは、アメリカの ENIAC (1945年完成) であるとみなされている。確かに、Zuse のコンピュータが継電器を使用しているのに対して、ENIAC は真空管を使用した電子式であり、より進歩していると言えるかもしれない。しかし、このように、現代のコンピュータの特長をどれだけ備えているかを判断の基準にするなら、イギリスで開発されたプログラム内蔵型 (ノイマン型) の Manchester Mark I (1948年完成、のちプログラム内蔵型に改良) が世界初のデジタルコンピュータということになるだろう (ENIAC はプログラム内蔵型ではない)。いずれにしても、ENIAC が世界初のコンピュータであるという説には根拠がないと思われる。

(2) 松原らは、表象主義をとらないときでさえフレーム問題からは逃れられないと主張する (松原、橋田 1990)。これに反論するためにはもう一つ論文が必要である。

文献

Dennett, D. C. (1984): "Cognitive Wheels : The Frame Problem of AI", in Boden, A. B. ed., *The Philosophy of Artificial Intelligence*, Oxford University Press, 1990, 147-170.

Dreyfus, H. L. (1992): *What Computers Still Can't Do: A Critique of Artificial Reason*, MIT Press

羽地亮 (1997):「家族的類似性について」(神戸大学哲学懇話会『愛知』第12、13合併号, 84-92頁)

Lakoff, G. (1987): *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, The University of Chicago Press

松原仁 (1990):「一般化フレーム問題の提唱」(J・マッカーシー、P・J・ヘイズ、松原仁『人工知能になぜ哲学が必要か』, 哲学書房, 175-245頁)

松原仁、橋田浩一 (1990):「表象なしのロボットもフレーム問題に悩む」(『現代思想』1990年7月号, vol.18, no.7, 160-167頁)

McCarthy, J. and Hayes, P. J. (1969): "Some Philosophical Problems from the Stand-point of Artificial Intelligence", in Meltzer, B. and Michie, D. ed., *Machine Intelligence* 4, Edinburgh University Press, 463-502.

McCarthy, J. (1980): "Circumscription : A form of non-monotonic reasoning", *Artificial Intelligence*, vol.13, 27-39.

McCarthy, J. (1986): "Applications for circumscription to formalizing commonsense knowledge," *Artificial Intelligence*, vol.28, 89-116

大澤真幸 (1990):「知性の条件とロボットのジレンマ　フレーム問題再考」(『現代思想』1990年3-4月号, vol.18, no.3-4, 3月号140-159頁, 4月号270-288頁)

佐々木正人 (1987):『からだ：認識の原点』, 東京大学出版会

(文学研究科研修員、龍谷大学非常勤講師)