

「分散表象」は認知の説明にはたして役立つのか？

服部裕幸（南山大学）

『心の科学と哲学-----コネクショニズムの可能性』戸田山・服部・柴田・美濃（編）  
（昭和堂）2003, pp.29-51

## 1 コネクショニズムと分散表象

古典的計算主義とコネクショニズムの関係について次のような議論がある。それはすなわち、統語論的規則の適用による変換操作に相当するものをコネクショニズムにおいても近似的に回復できることを示すことでコネクショニストのシステムでも体系性を説明できる、と主張することは、コネクショニズムを擁護するというよりはむしろ、コネクショニストのシステムは古典的計算主義の実現形態の一つでしかないというコネクショニズム批判を認めることになる、というものである(1)。しかし、このような議論には誤解が含まれている、という批判が出されている。というのは、コネクショニストは「古典的統語論の規則に従う変換操作の（近似的）対応物がコネクショニズムによって復元できるから、コネクショニズムは古典的統語論を頼りとして、『認知活動における産出性や体系性や合成性を説明できる』と主張しているわけではない。彼らのポイントはむしろ次のようなものだと考えるべきであろう。古典的統語論の規則の適用による表象の変換操作と近似してはいるが、それとははっきり異なる表象の変換操作の仕方があるかもしれない」(2)と考えるからである。しかしながら、この反論はポイントをとらえ損なっているように思われる。認知過程がある抽象的レベルで記述されなければならないということを認めるとすれば、その抽象的レベルで同じような統語論的構造とその変換システムが得られる。その抽象的レベルでの統語論的構造とその変換システムが複数存在してそれらが競合する場合、いずれかに、または第3のものに収束するかもしれない。かぎり、結局のところ、「コネクショニズムは実現レベルでの代案を出したに過ぎない」と古典的計算主義者が語るような事態しかそこにはないということに変わりはない。フォーダーが言わんとしたことはまさにこのことであつたはずである。

「古典的統語論の規則の適用による表象の変換操作と近似してはいるが、それとははっきり異なる表象の変換操作」が存在するかどうか、また、存在するとしてそれがどれほど「古典的統語論の規則の適用による表象の変換操作」と異なるかということとはたしかに経験的な問題であり、ア・プリオリに決定することはできない。ただ、古典的計算主義者は、認知活動における産出性や体系性や合成性を説明するには彼らが

認めているような規則のシステムが必要であろうと考えている。したがって、コネクショニストのシステムが近似的に復元するかもしれない変換操作のシステムもそれに類するものにならざるをえないだろうと予測するのである。(分散表象の非古典的構成要素がどのように変わるかということの規定する仕方は古典的統語論の規則の適用による表象の変換操作とは「はっきり異なる」が、これは問題となっている認知過程の抽象的レベルにおける話ではない。)要するに、古典的計算主義ではトップダウン的アプローチがとられ、(局所的表象への)規則の適用システムによる計算がなされるのに対して、コネクショニズムではボトムアップ的アプローチがとられ、(分散表象の)ネットワークによる変換処理計算がなされる。そして両者の相違は同じ事態を「上から見る」か「下から見る」かの違いでしかない、というのが真相ではなかろうか。

表象への変換操作はそれが局所的表象への規則の適用でないとしたら「因果的に無効」ではないか、という古典的計算主義者による批判に対するコネクショニストの反論の一つに、「弦楽四重奏を録音したテープ」の例を利用するものがある(3)。この反論の概略はこうである。そのテープには4つの楽器が発した音のそれぞれの表象の重ね合わせとしての1つの表象が記録されているだけである。個々の楽器の音の表象はテープ上にはどこにも見出せない。(それゆえ、それらは古典的な意味での構成要素表象ではない。)しかし、一つの楽器が別のものになれば、あるいはまた、別の音を出せば、テープ上の表象は別のものになる。この意味で構成要素は因果的な効力を持っている。これと同様に、コネクショニストのシステムでは分散表象の非古典的構成要素に何らかの規則が適用されるわけではないが、だからといってその構成要素が因果的な効力を持たないということにはならない、というわけである。さて、テープの例において個々の要素音が因果的な効力を持っていることは否定できないが、4つの楽器が発した音を重ね合わせる操作をあえて「表象の変換」過程ととらえる必要があるのだろうか。4つの音を別々に録音して、それらを重ね合わせて一つの四重奏の音の記録にすると考えても同じことである。それはおそらく(瞬間瞬間の)電気的信号の単純和(の連続)を作るにすぎない。それをあえて「表象の変換」と呼ぶことによって何が得られるのだろうか。同様に、分散表象の非古典的構成要素が因果的な効力を持つとしても、それを説明するにあたって「分散表象を変換する」というような表現を用いることにどれほどの意味があるのだろうか。これが問題となろう。以下ではこの問題をより一般的な観点から検討することにしたい。

## 2 ネットワーク・モデルは実際にはどのように働いているか

例として水雷探知システム(4)をとろう。このシステムでは、入力が入力層のユニットに入ると、各ユニットが活性化したりしなかったりする。すると、それらと結合した隠れ層のユニットが、それぞれの結合の重みづけに応じて活性化したりしなかつ

たりする。(その重みづけは学習過程で決まる。)そして、隠れ層の各ユニットが出力層のユニットに結合しているので、出力層のユニットがやはりそれぞれの結合の重みづけに応じて活性化したりしなかったりする。そして最終的に、入ってきた入力の水雷のそれであるか、岩のそれであるかが判別される。水雷であるのか岩であるのかの認知はこのようなメカニズムによってなされるわけである。いま仮に入力層のユニットが13個あるとすると、外部の対象の表象(と呼んでよいとすればの話であるが)はユニットの活性化の度合いを適当な数値で表現すれば、13次元ベクトルで表される。他方、出力層が2つのユニットからなるとすると、出力はとりあえずは2次元ベクトルで表現され、それが水雷あるいは岩という判断結果に対応する表象ということになる。したがって、ここでは与えられた13次元ベクトルをしかるべき2次元ベクトルへ変換する操作が遂行されたことになる。このベクトルはコネクショニストが言うところの分散表象を表現したものである。コネクショニズムの主張によれば、分散表象の処理は古典的計算主義におけるような規則に従った過程ではないので、離散的な操作の有限回の積み重ねによって記述することはできない。なんらかの記述が可能であるとすれば、それはたとえば微分方程式を用いてなされるだけかもしれない。

ここで問題なのは、そのような記述は時間の推移と共にシステムの状態がどのように変わっていくかということについての記述を与えるが、その変化が具体的にどのような仕方で起こるのかということについては何も語らないということである。システムの状態がどのように変わっていくかを具体的に述べるためには、おそらく個々のユニットが活性化しているかどうか、また結合の重みづけがどのようにになっているか、あるユニットに入ってくる入力の和が閾値を超えているかどうか、といったことを述べなければならないだろう。ところで、コネクショニストの言う分散表象がベクトル表示されるとすれば、個々のユニットがどのような状態になっているかを語るということは、ベクトルの成分がどのような値を取るかを語るということであり、結合の重みづけがどのようにになっているか、あるいは、あるユニットに入ってくる入力の和が閾値を超えているかどうか、といったことを語ることも、ベクトルの成分に関わる話であって、ベクトルそのものについての話ではない。これは、いうなれば、表象レベルの話ではないのである。

これに対して、古典的計算主義のシステムでは事情が異なる。いわゆる局所的表象が入力として与えられたならば、それに対して規則に支配された仕方である操作が適用され、次のステップに移行し、また別の操作が適用され、...といった具合に進み、(答えの出る種類のものであるならば)有限回のステップの後、出力が出されることになる。この出力がこのシステムによって変換された局所的表象ということになる。この入力と出力の間関係や内部状態の変化の有様は離散的時間(ステップの回数)を用いて関数として記述することができるかもしれないが、今度の場合には、変換の

途中の過程がどのようなものであるかということ、具体的に述べることができる。というのは、始めに与えられた局所的表象がどのような操作を適用され、どのように変形され、その後でシステムの内部状態がどのように変わったかということが追跡でき、どのステップについてであろうと、そのときの内部状態や、そのときにどの操作がどのような表象に対して適用されているかということ（原理的には）述べるからである。このようなことが可能であるのは、古典的計算主義のシステムにおいては、表象が局所的表象であり、それが複雑な統語論的構造をもっている場合には、その構成要素 これも通常はなんらかの表象である がどのような仕方と結合されているかということに即して操作が逐次的に適用されるからである。この局所的表象とその構成要素の関係は部分と全体の関係であり、コネクショニストのシステムにおけるベクトル表象とその成分の関係とは異なることに注意すべきである。したがって、古典的計算主義のシステムにおいては、局所的表象の変換過程の記述は一貫して表象レベルにおいてなされるのである。

このことは、古典的計算主義のシステムにおいては、表象について語ることがその処理がいかになされるかを語る際に本質的な役割を演ずる、ということの意味している。つまり、「表象を処理する」という語り方をすればはじめて変換過程を説明できるのである。これとは対照的に、コネクショニストのシステムにおいては、表象について語ることがその処理がいかになされるかを語る際に本質的な役割を演じてはいないように思われる。要するに、あえて「分散表象」と言ったところで、そのように語ることにあまり意味があるようには見えないのである。

### 3 表象とは何か

コネクショニズムに対するもっとも批判的な人物でさえコネクショニズムも古典的計算主義もともに表象主義の立場に立っているということを認めている(5)。しかし、前節の最後に挙げたような疑問が生ずる以上、この認識自体を疑ってみてもよいのではなからうか。そのためには、そもそも表象とは何なのかということ、あるいは少なくとも心的表象とは何なのかということの再吟味から始めなければならない。

XはYの表象(representation)である、あるいはXはYを表象している(represent)と言われる場合には、(a) XはYを表している、当然のことながら、(b) XはY自身ではない。ここで、(b)については問題はないと思われるが、(a)についてはさらに何ごとかを述べなければならない。XがYを表すとはいかなることだろうか。いろいろな場合が考えられる。

- (1) 黒雲はやがて雨が降るであろうということ（多くの場合）を表している。
- (2) 木の年輪はその木の樹齢を表す。
- (3) 扉にかかれた「WC」という文字はそこがトイレであることを表している。
- (4) 「テーブルの上にリンゴがある」という文はテーブルの上にリンゴがある

というこ とを表して（意味して）いる。

（５）富士山（らしき山）の絵は富士山を表している。

（６）富士山の写真も富士山を表している。

（７）心に浮かんだ富士山のイメージも（もしそのようなものがあるとすれば）富士山を 表している。

等々。

これらの事例に共通する特徴を見出すのは難しい。（３）や（４）や（５）については規約ないし慣習によってこのようなことが起こっている。これに対して、他の事例については、表象されているものと表象しているものの間に何らかの因果関係があることが決定的な役割を演じているように思われる。クリプキのような人ならば、前者も因果的図式の中に組み入れることができると主張するかもしれない。しかし、仮にそれが正しかったとしても、そのことは規約や慣習が自然化可能だということが示されるということの意味するにすぎないのであって、（３）（４）（５）と

（１）（２）（６）（７）が異なる種類のものであるということは否定しがた

い。逆に、ヴィトゲンシュタインならば、どの事例も結局は言語ゲームの中で初めて成立するにすぎないと主張するかもしれない。また、これらの事例に共通する特徴など存在せず、たかだか家族的類似性があるにすぎないとも主張するにちがいない。例を心的表象に限定してもこの事情は変わらないであろう。先に挙げた（７）は心的表象の場合の例であるが、以下もそうした例である。なお（７）（８）（９）の例においても先ほどと同じ問題が生じることを見てとることは容易であろう。

（８）これこれの視覚像は赤を表している（赤に対応する像である）。

（９）頭の中で行っている計算の中に現れる「たす」という語は加法を表して（意味して）いる。

さて、本論の主題は表象についての一般論を展開することではないので、目標をさらに限定して、「分散表象」と言われているものについて関わるかぎり、表象というものについて考えることにしよう。コネクショニズムのネットワーク・システムの典型的ないし代表的なものとしては、先に例として挙げた水雷探知システムの他に、文字認識システム、顔の判別システムなどがある。これらはいずれも古典的計算主義の立場では（どちらかと言えば）作ることが難しいシステムであった。（ちなみに言う、これらのシステムが扱う対象（認知対象）は統語論的構造を持っているとは通常は考えられていない。）これらのシステムに適用できそうな表象の概念は（６）や（８）に見られるそれではなからうか。というのは、そこでは隠れ層のユニット群の活性化のパターンが認知対象の表象である、とコネクショニストは主張したがっているからである。

（６）や（８）の例では、表象は表象される対象によって因果的に引き起こされている。（ただし、視覚像を心的なものと考えれば、（８）では心身間の因果関係

が想定されなければならない。)しかし、表象とそれを生起させた対象の間に因果関係があるか否かということによって表象関係の有無を説明することはできない。ボールが窓ガラスに当たってガラスが割れたとしても、割れたガラスがボール(が窓ガラスに当たったこと)を表象しているとは通塙へ言わない。その理由はおそらく、ボールがそもそも認知システムとは考えられないからであろう。一般には、認知システムの(ある部分の)状態とそれを生起させた対象の間に因果関係があれば、後者が前者によって表象されているということになる、というようなことはない。何らかの薬物によって部屋の壁一面を毛虫がはいずりまわっているような視覚像が生じたとしても、その視覚像はその薬物を表象しているわけではない。表象される対象と表象との間の因果関係は法則的でなければならない、という条件をつければこの問題を避けることはできるであろうが、いわゆる「選言問題」を回避することはできない。フォーダーはこの問題を解決するために非対称的法則性という概念を導入する(6)。他方、ミリカンは表象の「生産者」(すなわち原因)ではなくむしろ「消費者」(すなわち使用者)の観点を重視し、機能の概念に訴える(7)。ここでは彼らの提案 それらは現在提案されているもののなかでは代表的なものである の詳細には立ち入らないが、当面のわれわれの目的にとっては、ミリカンの提案は使えないだろう。なぜなら、彼女の提案では進化過程の概念が重要な役割を演ずるからである。ネットワーク・システムは人工物であり、通常の意味での進化の過程に位置を占めることはできない。ところが、そのようなシステムが表象を利用しているか否かをわれわれは問題にしているのである。そこで、とりあえずフォーダーの提案にしたがって考えてみよう。

細部にわたる議論は省くが、フォーダーにしたがえば結局はネットワーク・システムの「分散表象」と称するものは表象と言わざるを得ない。というのは、外的対象と隠れ層のユニット群の活性化のパターンとの間には、彼の言うような種類の因果関係が成立しているように見えるからである。そのかぎりでは、古典的計算主義もコネクショニズムとともに表象主義であるという彼の主張は正しい。しかし、フォーダーの提案にしたがうと、たとえばある種のバクテリアやゾウリムシでも外的対象を表象するということになることに注意すべきである。(この点はミリカンに従っても同じである。)

#### 4 認知活動の説明における表象の役割

しかしながら、人間の認知活動とバクテリアやゾウリムシの認知活動を同列に論ずることはわれわれの直観に反するように思われる。そして、その直観はある意味で健全である。両者の間にはたしかにかなり大きな相違がある。大雑把な言い方をすればそれは以下のような点である。バクテリアやゾウリムシの認知活動は行動に直結しているが、人間のそれは必ずしもそうではない。この点に関係するが、バクテリアやゾ

ウリムシは信念という形で表象を保持しつづけることはないように思われるのに対して、人間は信念という形で表象を保持しつづけることがある。バクテリアやゾウリムシは個体識別をしないとされる（し、またその必要性もないだろう）が、人間は時には個体識別を行なう（し、またその必要もある）。人間は自己を表象することができるが、バクテリアやゾウリムシはそのようなことはないように見える。また、おそらく以上の点に関係するであろうが、人間は表象を用いて推論を行うことがあるが、バクテリアやゾウリムシが推論を行なうということはないだろう。人間の認知活動とバクテリアやゾウリムシの認知活動の相違点はほかにも挙げることができるにちがいない。しかしまた、両者の相違は程度の相違であるかもしれない。たとえば、両者の中間に位置するような生命も存在するであろう。しかし、ここではこれ以上こうしたことを述べる必要はない。われわれが注目すべきなのは、以上のことから一つの興味深い点が浮かび上がってくるということである。そしてそれは、認知活動が体系性をもつかどうかということ、あるいはまた表象が統語論的構造をもっているかどうかという論点とも関連している。

認知活動にどれほどの体系性があるかということは経験的な問題であり、たとえば、その認知活動を行なう主体が対象をどのように認知し、その認知活動がどのように行動に影響を与えるか。その場合、ただちに行動に移ることもあれば、そうでないこともあるだろう。ということをごく簡単に説明することができるかにかかっているように思われる。たとえば、白血球が異物を自らのうちに取り込むメカニズムの記述を与えることは、（免疫学者は非常に難しいと言うかもしれないが）われわれの言語理解のメカニズムの記述を与えることと比較するならば、はるかに簡単であり、直接的に生化学的記述を試みることができるかもしれない。ここでは、フォーダーラが言うような体系性は必要ないだろう。（もちろん、このことは異物に出くわした白血球の振る舞いがランダムだということの意味してはいない。）

これに対して、たとえばわれわれの言語理解のメカニズムをいきなり生化学的レベルに降りて記述することができるとは到底思えない。そもそも、すでに母国語を習得した大人がその言語のある文を聞いたとき、その人は直ちに何らかの行動を引き起こすとはかぎらない。ひとまず記憶にとどめておいて、後になって必要なときにその情報を利用して、何らかの行動を引き起こすことになるかもしれない。また、その情報から、何か他の情報を引き出すかもしれない。さらにまた、（直後であろうと後になってであろうと）行動を引き起こす場合、そのようにして生起される行動は人によっても異なるであろう。また、その人は他の文を聞いたときとは違った対応をするであろうが、その文と先の文に共通の表現が現れれば、そのことに気づくであろう。また、その文が初めて耳にする文であったとしても（専門的な内容の文であったり、専門的な表現が使われたりしていなければ）それを理解することができる。このようなことを説明するためにわれわれは、その人は母国語に関する「言語能力」をもってい

る（母国語を知っている）と述べたり、他の「知識」を使って「推論」をすると述べたり、聞いたことを「記憶」にとどめておくと述べたり、聞いたことについてよく「考えた」上で「行動に移した」というような語り方をする。このような語り方をした場合、次に問題になるのは、そこで言われる「言語能力」とはどのようなものなのか（「言語知」の内容、中身は何か）そこで言及されている「知識」とは何であり、それはどのように整理された形で蓄えられているのか（それとも、そのような整理はまったくされていないのか）、どのような「推論」がなされるのか、その構造はどのようなものか、「記憶」は「思考」においてどのように利用されるのか、「考え」と「行動」はどのように結合されているのか、等々といったことである。

すると、たとえば言語能力について言えば、個別言語の文法を記述すること（だけではないにしても、少なくともそのこと）によって言語能力と言われるものの中身について（間接的に）語ることになる。というのも、知っていると言語がどのようなものであるかがわからなければその能力についても語りようがないと考えられるからである。そしてその際、現在の主流をなしていると思われる見解にしたえば、言語表現は統語論的構造をもっているものとして記述され、その文法を利用することによってわれわれは話し手の言わんとしたことを知るのである。ここに至って初めて体系的の話が前面に出てくる。すなわち、そのような文法は体系的に記述されるのである。ただし、これは経験的な問題であって、ア・プリオリに決まっているわけではない。われわれの言語の使用を適切に記述したり説明したりするには体系的記述が必要であるというにすぎない。犬に「お手」「お坐り」「伏せ」などという言葉を教え、その犬がそれらの言語を（ある程度？）理解できるようになったとしよう。しかし、その犬の言語能力について語るのに文法の体系的記述は必要ないだろう（し、そもそも文法記述さえ不要であるかもしれない）。また、個々の自然言語の文法記述は今日では（細部においては違いはあるものの）チョムスキー流の生成文法の形で与えられることが多いが、これも経験的な問題であって、ア・プリオリにそのように決まっているわけではない。ただ、チョムスキーアンが正しければ、われわれの言語の使用を適切に記述したり説明したりするのに適したよりよい代替案がないというだけのことである。

さて、ある言語の文法記述が体系的であるということから直ちに、その言語の使用者の言語能力 通常、これはその人の心の機能の一つと考えられている もまた体系的であるということは帰結しない。しかし、両者の間には、たとえラフな関係があるにすぎないとしても、ある程度の相関はあると考えるのは自然であろう。というのは、言語能力もまた体系的であると仮定しないと、体系的文法をわれわれが使いこなしているという事実を説明することが困難になるからである。かくして、心の中で思考の言語が体系的に処理されているという仮説が立てられるのである。いわゆる「思考の言語」仮説がこれである。たとえば、耳にした言葉に対応する表象、すなわち思



考の言語が心の中に存在し、心はそれを何らかの仕方で処理して言葉を理解するというわけである。そして、この処理の仕方が、古典的計算主義によれば、規則に従ったものなのである。むろん、そのように考えたからといって、ある言語の文法記述の体系性とその言語についての知（すなわち言語能力）の体系性が完全にパラレルな関係になっているということは帰結しない。何らかの程度においてそのパラレルリズムは破れているかもしれないのである。しかし、ここでは、それらの間にパラレルな関係が成立するか否かということは問題ではない。問題は、言語能力もまた体系的であるという仮定である。今のところその仮定なしで人間の言語能力を、そして結局は言語使用を、適切に説明したり予測したりすることが困難である二すれば、われわれはその仮定を受け入れざるをえないであろう。その意味でこの仮定の妥当性もまた経験的に判断されるべきものであって、ア・プリオリにその妥当性が証明されるわけではない。

前置きが長くなったが、ここで表象が認知活動の説明においていかなる役割を演じているかという問題に戻ろう。これまで述べてきたようなことが大筋において正しければ、人間の言語理解というような認知活動の説明には、統語論的な構造をもった心的表象（すなわち思考の言語）が必要とされ、また各々の心的表象は相互に体系的に関連しあっているという仮定が不可欠である、ということが（経験的に）主張できるであろう。ここでは表象の概念が説明の中で使われているが、それが単なる（心的）表象ではなく統語論的な構造を持った（心的）表象であるという点に注目すべきである。単なる（心的）表象ということであれば、（前節で見たように）ゾウリムシやバクテリアもその認知活動においてそれを利用していると言えるかもしれない。（通常の意味においては、ゾウリムシやバクテリアは心をもつとは考えにくいので、それらが利用する表象を「心的(mental)表象」と呼ぶのは問題があろう。単に「内的(internal)表象」と呼ぶにとどめるべきかもしれない。）しかし、それらの認知活動の説明は今見たような複雑な形をとらずにすむ（と考えられる）。したがってまた、それらが利用する表象が統語論的な構造をもつと仮定する必要もないし、表象の処理が規則に従ったものだと仮定する必要もない（と考えられる）。そして、後者のような説明には、実は、第2節において示唆したのと同様の疑問、すなわち、そこでの表象の概念の使用にどれほどの有効性があるのかという疑問が生じるのである。これに対して、人間の言語理解というような認知活動の説明に登場する心的表象の概念は、少なくとも先に述べたような仕方でその認知活動が説明されるかぎり、不可欠である。というのは、その表象を規則に従って処理する、という語り方をすることがここでは本質的だからである。

われわれは、表象の概念を使ったこれら二つのタイプの説明を区別すべきなのである。表象について論ずる際にわれわれは presentation という語と representation という語を使い分けるべきだ、という提案をグラッシュが最近行なっている(8)。彼の用

語法が適当であるかどうかはともかくとして、彼が言わんとしたことは基本的にはわれわれと同じことである。ゾウリムシやバクテリアが利用する表象は、彼の用語法にしたがえば presentation なのであり、それに対して、言語理解というような認知活動の説明に登場する表象は representation なのである。もしこの意味における representation に対してのみ「表象」という言葉を当てるとするならば、そして、分散表象が presentation とみなされるべきであるならば、コネクショニズムも古典的計算主義もともに表象主義であるというフォーダーの指摘は誤りということになる。そして、われわれが第2節で確認したことからは、分散表象は presentation であるということが示唆されるのである。

## 5 ワットの回転速度調節器(governor)とベクトル

分散表象を「表象」と呼ぶべきかどうかということ、それはグラッシュの意味での representation なのかどうかということ、そしてまた、最終的には、コネクショニズムが表象主義かどうかということ、は単なるネーミングの問題にすぎないと思われるかもしれない。しかし、そうではない。すでに明らかなように、表象(と称するもの)が認知活動の説明においてどれほど本質的な役割を演ずるかという点が問題なのである。そしてわれわれは、この点に関して、分散表象は認知活動の説明においてそれほど本質的な役割を演じていないということ、ベクトルの議論の検討を通じて、主張したいと思う。というのは、ファン・ゲルダーとベクトルの間で行われた論争において、ベクトルは最近、ワットの回転速度調節器(governor)によって一定の回転速度を保つエンジン・システムは(あるいは、そのようなシステムでさえも)表象を用いている、と主張している(9)からである。彼のこの見解はわれわれの見解とは(少なくとも見かけ上は)対極に位置するので、それを注意深く吟味する必要がある。

彼の議論を検討する前に、まずファン・ゲルダーの論点を簡単に見ておこう。彼は、認知システムについての伝統的アプローチ(古典的計算主義)に対する代替案としてダイナミカル・システムによるアプローチを提案する。すなわち、「認知システムはダイナミカル・システムであるのかもしれない」(10)と。ダイナミカル・システムとは何か。それは「その状態が数値的[に表されうるもの]であって、...その状態の変化の規則がそのような数値的[に表されうる]状態の系列を特定化するようなシステム」(11)のことである。(彼はベクトルで表示することもここで言う「数値的」に表すことの一つと考えている。)彼はA回転速度調節器を例にあげて二つのアプローチの特徴を説明している(12)。彼によれば、回転速度を一定に保つようなシステムを作り上げる場合、二つの考え方がありうる。一つは、(ステップ1)回転速度を測定し、(ステップ2)目標として設定された速度からの乖離があるかどうか調べ、(ステップ3)乖離がなければ、(ステップ1)に戻り、引き続き回転速度を測定する。もし乖離があれば、(ステップ4)スチームの圧力を測定し、(ステップ5)ど

れだけ圧力を変えればよいか計算し、(ステップ6) 必要なだけスチームパイプのバルブを開閉し、速度を上げ下げする。そして、(ステップ1)へ戻り、再び回転速度を測定する、というものである。これは基本的には古典的計算主義に基づいた設計思想である。これに対して、もう一つの考え方がある。それは、回転速度が上がると、回転軸にセットされた、金属球のついたアームが遠心力によって開き、アームに蝶番によって接続された梃子が押し上げられ、その結果、それに接続されたスチーム弁が少し閉まり、回転速度が下がる、というような仕組みをもった機械を作る、というものである。この設計思想においては、回転速度を測定するというような過程はこのシステムに存在しない。このシステムは純粹に力学的なメカニズムによって回転速度を一定に保つのである。このようなシステムはダイナミカル・システムの典型例である(13)。そして、ワットが実際につくった回転速度調節器はまさにこのタイプのものだったのである。

さて、ファン・ゲルダーによれば、これまで認知過程は第一のタイプのものという考え方に基づいて研究されてきたが、第二のタイプの考え方もあり、前者が行き詰まってきたかに見える現在、後者を真剣に考えるべきなのである。そして、コネクショニズムについてはファン・ゲルダーは次のように述べる。「コネクショニスト・モデルはすべてダイナミカル・システムである。...ネットワークが $n$ 個のニューラル・ユニットをもっているとする、いかなる与えられた時刻におけるそのシステムの状態も活性化値の $n$ 次元ベクトルにほかならない。そして、そのネットワークの振舞いは各ユニットの活性化値を更新する方程式によって決定される、そのようなベクトルの系列なのである。」(14)しかしながら、多くの標準的なコネクショニスト 彼らの立場を「古き良き時代のコネクショニズム」とファン・ゲルダーは呼んでいる。は古典的計算主義の思想とダイナミカル・システムの思想の間を目指しているように見える。というのは、彼らは基本的にはダイナミカル・システムであるものを用いながらも、それを、古典的計算主義が採用している認知システムについての概念と結合させようとしているからである。彼がここで念頭に置いているのは、コネクショニストのシステムも入力表象を出力表象へと変換するという点では古典的計算主義のシステムと同じであるが、使用する表象とその変換方法が古典的計算主義のシステムとは異なるのだ、すなわち、分散表象を用いるのであって、規則に従った変換作業を行なうのではない、というコネクショニストの主張である(15)。しかし、このような中間の道は彼によればありえない。「時の経過とともに、そのようなコネクショニスト・モデルは次第次第に、種類としては認知に関する計算論的な概念の[単なる]実現形態[にすぎないもの]に道を譲るか、あるいは、より完全にダイナミカルなモデルへと道を譲るかのいずれかになるとわれわれは期待すべきなのである。」(16)

ファン・ゲルダーがこの結論に与えた論拠はわれわれの前節での議論と基本的には同じである。すなわち、彼は「あるシステムが表象を含んでいるか否かを述べる信頼

できる方法は、そのシステムを記述するにあたって表象という用語を用いることが説明の上で何らかの有効性をもっているか否かを尋ねるというものである。」(17)とした上で、いくつかの理由を挙げて、コネクショニスト・モデルが表象を利用しているということを否定するのである。まず第一に、コネクショニスト・モデルをその一種として含むダイナミカル・システムを記述するときに、われわれは通常は表象の概念を用いない。たとえば、回転速度調節器の振舞いの記述は微分方程式によって与えられるが、そこに表象の概念は登場しない。第二に、回転速度調節器の例で言えば、アームの角度と回転速度の間には相関関係があるとしても、相関関係があるからといってそこに表象関係があるということにはならない。実際、ほとんどすべてのものが偶然であろうとなかろうと、他の何らかのものと相関関係をもつが、それらをすべて表象関係と呼ぶとすれば、表象の概念が空虚なものになってしまうであろう。第三に、そもそもアームの角度と回転速度の間に相関関係があるというような語り方をすることは不適切である。第四に、回転速度調節器は、アームの角度が上昇すると回転速度が下がり、そのことによって、度はアームの角度が減少し、回転速度が上がら始める、というような精妙なメカニズムになっている。すなわち、アームの角度の瞬間加速度はそのときの回転速度だけでなく、その時点でのアームの角度の変化率、その時点でのアームの角度などにも依存しているのである。したがって、アームの角度と回転速度の間に単純な表象関係を認めることができるほど単純な関係にはない。最後に、このシステムはいくつかのモジュールに分けることもできなければ、それらの中での表象の受け渡しもない。以上の理由は相互に関連し合っており、それぞれを切り離して考えることは妥当ではないが、ここでは便宜的に切り離して考えることにしよう。

第一、第三、および第五の理由はこれだけではそれほど説得力のあるものではない。現に記述に「表象」という言葉が使われていないということは、「表象」という言葉を用いての記述が不可能ということの意味しない。第二の理由について言えば、単に相関関係があるだけで表象関係を認めることはできないというのはその通りであるが、すでに第3節で見たように、多少制限された因果的相関関係の存在をもって表象関係があるとすることは、(われわれはそれに与しないが)それほど不合理ではないかもしれない。そしてその限りで、コネクショニスト・モデルは表象を用いていると言えるかもしれないのである。実際、フォーダーがそうであった。五つの中では第四の理由がもっとも重要であるように思われる。通常、XがYの表象であるというときにはYはXの表象ではない。すなわち、表象関係は対称的ではないのである。これに照らせば、第四の理由をもって、コネクショニスト・モデルは表象を用いていないと結論したくなるかもしれない。しかし、これは早計である。まず、第四の論点を理由として表象を用いていないと主張できそうなシステムは、回転速度調節器のような、ある意味では単純なシステムに限られるかもしれない。ゾウリムシのようなシス

テムについてはそのようなことは言えないかもしれないのである。さらに、回転速度調節器でさえも、厳密に言えば、回転速度（の変化）を表象と見て、それを処理してアームの角度を調節すると考えるときには、出力とみなされるアームの角度の変化は最初の、入力表象とみなされたときのアームの角度の変化と同じものではない。したがって、回転速度（の変化）とアームの角度（の変化）の間に表象関係を認めたとしても、それが同時に対称性ももっているということにはならないのである。たしかに、表象関係が時々刻々と連続的に入れ替わることにはなるが、そのこと自体は不合理なことではない。そのように考えることにいかなるメリットがあるのかという批判は可能である。そして、それはまさにわれわれの批判であるが、それはまた別の論点である。

このように、個々の理由はそれ単独ではさほど説得力をもつわけではないが、それらを全体として見たときには、それなりの力をもつように思われる。そしてその論点は、一言で述べるならば、コネクショニストのシステムを記述するにあたって「表象」という表現を用いることは説明の上で何の有効性ももっていないのではないか、ということになる。あるシステムが表象を含んでいるか否かに関するファン・ゲルダールの基準によれば、このことは、コネクショニストのシステムは表象を用いるシステムではないということの意味するのである。したがって、コネクショニストのシステムは分散表象を（規則に従うのとは異なる仕方）で処理するシステムである、とコネクショニストが語ったとしても、それは誤りか、もしくは空虚な主張なのである。

さて、ベクテルは以上のようなファン・ゲルダールの主張の論駁を試みている。ファン・ゲルダールのような論者 われわれもその中に含まれる は、回転速度調節器のようなシステムにまで表象の概念を適用すると、その概念が空虚になるか、あるいはその概念を用いた説明が興味あるものとはならなくなる、と考えているが、そのようなことにはならないとベクテルは主張する。このことを彼は、「これこそまさに、表象の概念が不可欠であると認知神経科学が考えるような種類のシステムに他ならない、ということを示すことによって」(18)論証しようとする。ここでの彼の議論は概略次のようなものである。

前提 1 脳の処理システムと回転速度調節器は同じ種類のシステムである。

前提 2 脳の処理システムの働きの説明には表象の概念が不可欠であり、その説明は空虚ではなく興味深いものである。

結論 故に、回転速度調節器の働きの説明には表象の概念が不可欠であり、その説明は空虚ではなく興味深いものである。

これは一種の類比論法であるが、その論法の持つ弱点は今問わないことにしよう。しかし、それを別にしてもこの議論には問題があるように思われる。以下、それを指

摘したい。

そもそも問題の発端は、コネクショニストのシステムは古典的計算主義のシステムの単なる実現形態の一つにすぎないのか、それともそうではないのか、ということであった。そして、この論争の過程で分散表象という概念が登場してきた。すなわち、コネクショニストのシステムには局所的な表象はないが、その代わりに分散表象がある、と提案されたのである。次いで、この分散表象がはたして「表象」と呼ぶに値するかという議論になってきた。脳の処理システムの働きが現在あるようなネットワーク・モデルで説明できるかどうかについては、多くの脳研究者は懐疑的であるが、将来においてはそれで説明ができるここでは仮定しよう。するとベクテルは、少なくともある種のネットワーク・モデルの働きの説明には表象の概念が不可欠であり、その説明は空虚ではなく興味深いものである、ということ的前提していることになる。議論の流れからして、ここでの表象は分散表象と考えなければならないだろう。しかし、この前提こそがまさに問題視されていたのではなかつたらうか。つまり、ネットワーク・モデルでは分散表象を処理しているのだという説明は、個々のユニットが活性化しているか否か、またユニット間の結合の重みづけの大きさはどのようになっているのか、という話以上の何かを付け加えているのか、ということが問題とされていたのではないのか。もしそうであるとすれば、ベクテルは、少なくとも議論の大きな流れの中では、論点先取の誤りを犯しているように見える。

ベクテルはまた、前提1が正しいことを論証しようとしているが、その企てにも問題がある。彼は、局所的表象であろうとなかろうと、脳の処理システムも回転速度調節器もともに表象システムである、ということを示すことによって前提1の正しさを示そうとしているように見える(19)。彼がその際に依拠する表象の概念はドレッキやミリカンのものである。ドレッキによる表象概念の分析は因果関係を利用するものであり、基本的にはすでに第3節においてわれわれが検討したものと同じである。ミリカンのそれはドレッキの表象概念の分析を、表象を「消費する者」という観点を強調することによって補完しようとするものである。彼女の分析は進化の概念に訴えるので当面の目的には使えない というのは、われわれは人工的なシステムをも問題にしているから ということを示すに述べたが、その点をあえて無視すれば、たしかに脳の処理システムも回転速度調節器もともに表象システムであると述べることはできるであろう。このことも第3節においてわれわれが確認したことから明らかである。しかし問題は、ドレッキやミリカンの表象概念の分析を利用して人間の認知過程の(いくつかの) 解明を行なうことがはたして適切か、ということであったはずである。ドレッキやミリカンの表象概念にしたがうかぎり、人間(の脳)やバクテリアや回転速度調節器もみな表象システムということになるであろう。このことをわれわれは否定しない。(ただし、その場合の表象はグラッシュの言う representationではなく、presentation にすぎない。) したがって、その意味ではたしかに前提1は論証された

かに見える。しかし、問題は、ドレツキやミリカンの意味で同じ種類のシステムであると主張しえたとしても、先の類比論法を妥当なものとするのにそれで十分か、ということである。われわれはそうではなかろう、と考える。ミリカン自身、彼女の意味では人間（の脳）もバクテリアもともに表象システムであるという点では同じであるが、そこには無視しえない違いもあるということを知っている(20)。そして、そのような違いこそが、ベクテルが行なったような類比論法の適用を不適切なものとするのである。表象の概念をバクテリアや回転速度調節器にまで適用したとき、それをを用いてのシステムの振舞いの説明が空虚になるであろうという論者の論点も、最終的にはそこにあるのである。したがって、ファン・ゲルダーやわれわれの主張を論駁しようというベクテルの試みは成功しているとは言えないように思う。

しかしながら、ベクテルを批判する論者やわれわれの議論も決定打を欠くものであることは認めなければならない。というのは、表象の概念を用いての認知の説明の効力ないし有効さという概念が明快さに欠けるからである。実際、ベクテルによると、脳神経学者は「脳の活動と末端の刺激との間の情報関係については、もしこの情報をさらなる処理過程において脳が利用していると仮定しなかったならば、それを見出すことにほとんど興味を示すことはなかったであろう。脳がこの情報を実際にどのように消費しているかを決定するためには、情報処理モデルを作り上げなければならない。そして、そのようなモデルにおいては、視覚処理過程の早い段階でコード化された情報が、視覚の通路の後の方の領域の活性化によってどのように利用されるかということが示されることになる。」(21)かくして彼は、脳神経学者は表象の概念に訴え、ト認知についての有効な説明を与えていると主張するのである。ただ、彼は問題の論文においては脳神経学者が表象の概念を用いて行なっている説明を実際に検討しているわけではない。事実として脳神経学者が表象の概念を用いて説明を行なっているとしても、そのことはその概念が有効に用いられているということ、あるいはその概念が不可欠であるということの意味しない。単に表象の概念を用いて説明を行なうだけであるならば、われわれは回転速度調節器にさえもそれが適用可能であることを否定はしない。ただわれわれは、それがせいぜい比喩的に使われるだけで、その働きの説明にとって表象概念は必要ないと主張するのである。脳神経学者による表象の概念を用いての説明についても同様であるというのがわれわれの予測である。

最後に次のことを指摘しておこう。古典的計算主義モデルにおいては、(局所的)表象は変換されたり、記憶されたりするが、いずれにしても同じタイプのものであり続けている。ところが、分散表象の場合にはそうでないように思われる。入力された「表象として機能するのは個々の神経細胞あるいはユニットの発火率ないし発火パターンである」(22)とされるが、仮に表象が記憶される。そして、表象の概念が有効に使われるとすればこのことは不可欠であるように思われる。とすると、コネクショニズムにおいてはそれは神経細胞あるいはユニットの発火率ないし発火パターンではな

く、むしろ神経細胞間あるいはユニット間の結合の重みづけ（のパターン）という形態をとるということである。この点もまた、「分散表象」は本当に表象と呼べるものであるのか、という疑念を抱かせるのである(23)。

#### 注

- (1) 服部 2000, p.22 参照。
- (2) 美濃 2001, p2.
- (3) Horgan and Tienson 1992, p.211. 美濃 2001 もこの例を引用している。
- (4) Churchland 1995, pp.79-83 参照。
- (5) Fodor and Pylyshyn 1988, p.9 参照。
- (6) Fodor 1987 の第 4 章を参照。
- (7) Millikan 1993 の第 4 章を参照。
- (8) Grush 2001
- (9) Bechtel 2001
- (10) van Gelder 1995, p.346.
- (11) Op.cit., p.368.
- (12) Op.cit., pp.347-350.
- (13) Op.cit., p.367.
- (14) Op.cit., p.370.
- (15) Op.cit., p.374.
- (16) Ibid.
- (17) Op.cit., p.352.
- (18) Bechtel 2001, p.334.
- (19) Op.cit., pp.334-337.
- (20) Millikan 1993, pp.97-101.
- (21) Bechtel 2001, p.340.
- (22) Op.cit., p.337.
- (23) 草稿段階において、美濃正、大沢秀介、柴田正良、戸田山和久、柏端達也、金子善彦の各氏から有益な批判や助言を頂いた。ここにお礼申し上げる。

#### 文献

- Bechtel, W. : 2001, 'Representations: from neural systems to cognitive systems', in Bechtel, W. et al. (eds.) 2001, pp.332-348.
- Bechtel, W. et al. (eds.): 2001, Philosophy and the Neurosciences, Blackwell.
- Churchland, P. : 1995, The Engine of Reason, the Seat of the Soul, MIT.
- Davis, M. (ed.) : 1992, Connectionism : Theory and Practice, Oxford.



- Fodor, J. A. : 1987, *Psychosemantics*, MIT.
- Fodor, J. A. and Pylyshyn Z. W. : 1988, 'Connectionism and cognitive architecture', *Cognition* 28, pp.3-71.
- Grush, R. : 2001, 'The architecture of representation', in Bechtel, W. et al. (eds.) 2001, pp.349-368.
- 服部裕幸 : 2000, 「コネクショニズムとは何か?」, *科学哲学* 33-2, pp.15-28.
- Horgan, T. and Tienson, J. : 1992, 'Structured representations in Connectionist systems?' in Davis, M. ( ed.) 1992, pp.195-228.
- Millikan, R. : 1993, *White Queen Psychology and Other Essays for Alice*, MIT.
- 美濃 正 : 2001, 「一昨年以来の論議の総括」, *名古屋哲学フォーラム*(2001.9.8., 於: 南山大学)での提題発表原稿
- van Gelder, T. : 1995, 'What might cognition be, if not computation', *Journal of Philosophy* 91, pp.345-381.