

## 新しい認知の理論としてのコネクショニズムの可能性

美濃 正（大阪市立大学）

『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.79-114

### 本章の問題

コネクショニズムが認知科学における新しいパラダイムとして提唱され多くの支持を集めるようになったとき、科学に関心をもつ一部の哲学者たちがコネクショニズムに関心を寄せたのは、おそらく、それが当時すでにいくつかの難題に直面していた古典的計算主義に代わる新しい認知の理論でありうるのではないか、あるいはより正確には、そのような理論の基礎をそれが提供しうるのではないか、という問題意識ないし期待を抱いたためであった。しかし、いささか奇妙なことに、この問題をめぐるその後の論議の過程をへて、コネクショニズムに共感を寄せる立場か逆に批判的な立場かという立場のいかに問わず、多くの哲学者が新しい認知の理論としてのコネクショニズムの可能性に対して否定的な見解をとるに到っているように思われる（このような見解の実例を次項で挙げる）。このような事態に立ち到ったのは、私の見るところでは、これまでのコネクショニズムに関する哲学者の間での論議において、次のようなきわめて基本的だが重要な問題についての検討と解明が十分にはなされてこなかったためである。つまり、そもそも認知の理論とはどのようなものであるのか（あるいは、あるべきなのか）、認知の理論としての古典的計算主義は正確にどのような主張を含む立場なのか、それにとって代わる新たな認知の理論がありうるとすればそれはどのような要件を満たさなければならないのか、そして、コネクショニズムはこれらの要件を満たすことができるようなものであるのか、といった問題である。そこで本節では、不十分ながら私なりに、これらの基本的な問題に対する適切な解答をするための見通しを与える作業に取り組んでみたい(1)。

### 最近の支配的見解

まず、私自身が多少関わった論議(2)のなかから、最近の支配的見解の実例を二つ挙げ てみよう。一方は（部分的に）親コネクショニズム的立場をとる論者の、他方は（おそらく）反コネクショニズム的立場に立つ論者の見解である。

このような見方のもとでは、信念や欲求のような、構文論的表象を含むと考えられる心的状態は脳の内部には存在せず、言語表現の形で脳の外部にのみ存在することになる。われわれはたとえば「地球はまるい」と声に出して言ったり、頭の中で言ったりしなくても、地球はまるいと信じていると考えているが、そのような信念は、それが構文論的表象を含むものと解されるかぎり、存在しない。そのような信念が存在するとすれば、それは「地球はまるい」という発話を行う傾向性としてのみ存在し、そしてこの傾向性ないしその物理的な基盤は構文論的構造をもたない。(3)

これに対して、「統語論的規則を適用して変換操作を行なっているかに見えるだけで、実際にはそのようなことは行なっていない」．．．ので、「古典的 AI モデルは真に認知活動を説明してはならず、むしろコネクショニストモデルこそが真の説明を与えているのだ」と応ずることは何の役にも立たないだろう。というのは、ここで言われている「真の説明」とは結局、より脳の働きに則した説明ということでしかないように思われるからである。ところが、問題は認知活動というのは．．．脳の活動そのものなのか、それとも抽象度のより高いレベルのものなのか、ということだからである。古典的 AI 研究者は後者であると考えるので、コネクショニズムは実現レベルの話をしているにすぎないと批判するのである。(4)

信原氏は、われわれの脳が一つのコネクショニスト・システムを成すと考えている。そしてそれゆえにこそ、少なくとも信念や欲求のような、構文論的構造をもつ種類の認知(状態)は脳の内部には存在しない、と主張するのである。信原氏の見解が、認知の理論としてのコネクショニズムの可能性を、少なくとも部分的に否定する見解であることは明白であろう。

これに対して服部氏の見解は、コネクショニズムは単に脳神経科学的レベルにおける脳活動のモデルを提供しうるにすぎない、というものである(ように思われる)。認知活動はより抽象度の高いレベルにおける脳活動の記述なのであるから、したがってこの見解からは、コネクショニズムが認知の理論、ましてや新しい認知の理論を提供しうる可能性はない、という結論が導き出されることになる。

以上のように、相対立する陣営に属する哲学者たちが、相異なった根拠にもとづいて、しかし認知の理論としてのコネクショニズムの可能性に対しては同じように否定的結論に到っていることが確認できる。このような事態になったことには、一つの共通原因が大きく働いているものと考えられる。それは、J.A.フォーダーと彼の仲間が提出したある論証である。

#### 認知能力の体系性からする論証

フォーダーらの論証は、認知能力の体系性という一般的事実から出発して、最終的には古典的計算主義こそ唯一可能な認知の理論だ(したがってコネクショニズムが新しい認知の理論となることはありえない)という結論を導き出そうとするものである。

まず、認知能力の体系性とは、何であろうか。フォーダーらの定義によれば、それは「意味論的に関連し合う心的状態の集まりがあって、．．．有機体はその集まりに属する一つの状態になることができるのは、それが[同じ集まりのうちの]それ以外の多くの状態になることができる場合にだけである」(5)という事態のことである。たとえば、あなたが「ジョンはその少女を愛している」ということを考えることができるのは、あなたが同時に「その少女はジョンを愛している」ということをも考えることができる場合にだけだろう。同様に、あなたが「冷たい水」と「熱いコーヒー」を欲することができるなら、あなたは「冷たいコーヒー」や「熱い水」を欲することもできるだろう。以上は、心的表象能力に関する体系性の事例であるが、同様に、表象の処理能力に関する体系性も認めることができる。たとえば、「P & Q & R」から「P」を推論できる人は、同時に「P & Q」から「P」を推論することもできるはずであろう。

このように認知能力は孤立した形においてではなく、意味論的に関連し合う他の多くの認知能力を相伴って成立する、という事態は、われわれ人間に代表される自然的な認知システム、つまり現に知られている人工的でない認知システムのすべてが共有する性質だと言うことができる。つまり、認知能力の体系性は、自然的な認知システムに関して成り立つ一つの法則性だと言える。なぜ、このような法則性が成り立つのだろうか、これがフォーダーらの問いかけである。

この問いかけに対する彼らの答えが彼らの論証の第二段階を形成する。それは、「心的表象には言語と同様に構文論的構造が備わっており、かつ、心的表象の処理はこの構文論的構造に従って、より正確に言えば、それに因果的に依存して行われるから」という答えである。よく知られたフォーダーの表現を借りて言うならば、要するに自然的認知システムには「思考の言語 language of thought」が備わっているがゆえに、その認知能力は体系性を本質的に示すのである。

後に明らかになるように、フォーダーらの論証の第二段階までに対しては私も同意する。しかし、彼らは論証をさらに先へと進め、次のようなディレンマを提出する。「もしコネクショニズムが認知能力の体系性を説明できないなら、それは認知の理論と呼ぶに値するものではないであろう。他方、もしそれがこの体系性を説明するために心的表象の構文論的な合成構造に則した（因果的に依存した）処理過程を要請することになるならば、それは結局、古典的認知モデルを実現するための一つの仕方を示しているにすぎず、それに代わる新たな認知モデルとは言えないことになるだろう。今述べた二つの可能性しかないのだから、いずれにせよ、コネクショニズムは古典的計算主義に代わる新たな認知の理論ではありえない。」(6)

このようにフォーダーらの論証は、認知能力の体系性を説明するには何が必要かという問題提起から始まって、最終的には認知の理論としてのコネクショニズムの否定に到っている。容易に見てとれるように、この最終的結論を導き出すさいに彼らが依拠している前提は、「構文論的構造に則した心的表象の処理はすべて古典的計算主義の考えるような処理にならざるをえない」ということである。(7) しかしながら、彼らの採用したこの前提は、一見したところそう見えるほどには説得性のある主張ではない。いずれにせよ、新しい認知の理論としてのコネクショニズムの可能性を探ろうとする本節のような試みにとっては、この前提はぜひとも突き崩さなければならないものである。したがって、以下の論述においては、フォーダーらの採ったこの前提の批判的検討が中心的部分を占めることになる。しかし、まずはそもそもいわゆる認知科学における認知の理論とはどのようなものである（べき）なのか、認知の理論としての古典的計算主義とはどのような主張であるのか、そして、なぜ古典的計算主義に代わる新しい認知の理論への待望が認知科学の内外に生じるに到ったのか、といった基本的な事柄の簡単な再確認から始めることにしよう。

## 認知の理論とは何か - - 認知記述の三つのレベル

いわゆる認知科学はさまざまな目的をもっている。たとえば、ポピュラーなもので言えば、われわれ人間に匹敵するような人工的な認知システム（ロボット）を作り出すことも、その目的の一つであろう。しかし、中心的な目的の一つはあくまでも、現に自然のうちに存在する認知システム、特に人間という認知システムのあり方を明らかにする理論を与えることにある。では、認知科学は、人間に代表される自然的認知システムを、どのような点から明らかにしようとしているのだろうか。ひとことでは、このようなシステムにおける認知の過程がどの

ようなメカニズムに従って進行するのか、を明らかにする理論の提示をめざしているのである。

この点についてもう少し詳しく述べれば、認知科学が提示しようとしている「認知の理論」は、一つの認知システム（における認知過程）を次の三つのレベルから解明することをめざしている(8)。

(1) そのシステムがある認知状態(9)から、次にどのような認知状態へと推移していくのかを、記述する最上位の抽象的レベル。

(2) このような認知状態間の推移が、より下位のどのような因果的プロセス（あるいは設計）に従って生じるのかを（数学的に）明らかにする中間的レベル。

(3) このような因果的プロセスが、具体的にどのような仕方で物理的に実現されているのかを明らかにする最も下位の物理的レベル。

認知科学における認知の理論は、与えられた認知システムにおける一つの認知過程を、三つの異なる観点から記述しようとするものである、と言ってもいいだろう。このことは、認知科学における認知の理論が一般に次のようないくつかの重要な前提にもとづいていることを示唆する。

#### 認知の理論の前提 - - 表象主義・機能主義・物理主義

まず抽象的トップレベルに関わることだが、認知科学における認知の理論の解明しようとするものは、言うまでもなく今日われわれが考えている意味での「認知」（のメカニズム）である。つまり、認知の理論は現行の認知概念を前提としている。では、現行の概念によれば、認知という事象の本質的特性は何であろうか。少なくとも大半の認知事象に共通する一つの本質的特性は、認知がそれ自身以外の何事か（特に外的世界で生じると考えられる何事か）に関わる、もしくはそれを表すという性質、つまり表象的性質であることに疑いの余地はないだろう。たとえば「雪は白い」という信念（という認知事象）は、明らかに<雪は白い>という事態に関わっている、もしくはそれを表している。（少なくとも大半の）認知事象は表象的性質をもつというこの考え、つまり表象主義が、したがって今問題になっている種類の認知の理論にとってははずすことのできない前提の一つになっているのである。(10)

第二の重要な前提は認知に関する広い意味での機能主義である。機能主義とは、一言で言えば、認知事象の本性のすべてもしくは少なくとも一部はその機能つまり因果的役割に存する、という考えである。たとえば「のどの渇きをいやしたい」という欲求（という認知事象）は、「ビールはのどの渇きをいやす」という信念が条件として与えられれば、「ビールを飲もう」という意図ないしそれに対応する行為を生じさせる強い傾向をもつだろう。この欲求は、ある与えられた条件（信念）から、別の認知状態（意図ないし行為）を導く機能を担っていると言ってもよい。このように、ある条件が与えられたとき次にどのような認知状態へと導く傾向があるか、というおのこの認知状態の因果的役割もしくは機能のうちこそ、機能主義はその認知状態の本性を見定める。つまり、そのような傾向もしくは機能をもたないものは、その種の認知状態とは認められない、とみなすのである。認知科学における認知の理論が特に中間的レベルにおいて解明をめざす認知状態間の推移のメカニズムとは、まさにこのような認知状態がもつ因果的傾向性同士の連関から主として成り立っているものであろう。したがって、認知の理論がこのようなメカニズムの解明に焦点を絞っているという事実は、それが機能主義を不可欠の前提としていることを強く示唆するのである。

最後の前提は、やはり広い意味での物理主義である。第三の物理的レベルの存

在がはっきりと示すように、認知の理論は、すべての認知事象が最終的には物理的に実現されると考えている。言い換えれば、認知事象はおのおのすべて何らかの物理的事象と同一の事象であるとみなしている。より具体的には、われわれ人間の場合なら認知事象はすべて何らかの脳神経科学的事象にほかならず、シリコンチップから成る認知システムの場合にはすべてそれらのチップの間で生じる何らかの電気化学的事象にほかならない、ということになる。したがって、認知科学における認知の理論は、認知事象がいかなる物理的事象からも独立に生じうることを認める、いわゆる心身二元論のような立場とは相入れない前提にもとづくのである。

古典的計算主義とはどのような認知の理論なのか

認知の科学における認知の理論一般に共通する特徴は以上のようなものであるが、この項では「古典的計算主義」と呼ばれている特定の認知理論がどのようなものであるのか、を簡潔に見ておこう。まず、この理論の基本的主張ないし前提を、先に示した認知記述の三つのレベルの区別に従って示しておこう(11)。

(1) 認知状態間の推移は、適切な物理的手段によって計算可能な関数(12)に従って生じる(トップレベル)。

(2) このような認知状態間の推移は、表象もしくは記号の処理(操作)に関わる、例外のない厳密な規則群に従って生じる。そして、これらの規則はコンピュータ・プログラム(一種のアルゴリズム)化可能なものである。このアルゴリズムが、認知状態間の推移を実現する因果的プロセス(の数学的表現)である(中間レベル)(13)。

(3) 認知状態間の推移を実現する、このようなアルゴリズムは、最終的には何らかの物理的プロセスによって顕在的もしくは明示的に実現される。つまり、アルゴリズムに現れるそれぞれの項に対応する何らかの現実的物理事象が認知システムの中に存在する(ボトムレベル)。

以上に加えて、古典的計算主義はさらに次のことをも主張ないし前提することを指摘しておくべきだろう。

(4) 少なくとも多くの認知事象は、統語論的構造(文法構造)を有する。

(2)で述べたような例外のない諸規則による表象の処理は、もちろん与えられた表象のもつ統語論的構造に従ってなされる。さもなければ(つまり表象の文法構造を無視して)、人間的認知システムが典型に示すような、認知状態間の意味論的に適切な(各認知状態の表す意味にかなった)推移がどうして可能であるのか、説明困難になるだろう。

古典的計算主義は認知という事象について結局どのような描像をもっているのか、きわめて単純化した形で要約しておこう。この描像によれば、認知活動というものはすべて何らかの問題解決活動である。したがって、認知状態間の推移は、与えられた問題の解決という定まったゴール(14)に向かうプロセスもしくはその一ステップだ、ということになる。そして、このプロセスは、認知システムにあらかじめ内蔵されている(例外のない)明示的諸規則を順次、適用することによって、所与の問題に関係する諸表象を処理ないし操作することをおして進められるのだ、とみなされている(15)。要するに、与えられた問題の解決という

定められたゴールに向かって、システムにあらかじめ備わっている定められた解決手順に従う表象処理（つまり計算）によって進行する過程、というのが古典的計算主義の思い描く認知像なのである。

#### 認知理論としての古典的計算主義への疑問

しかし、このような認知の描像に対しては、いくつかの本質的と思われる問題が指摘されてきた。この項においてそれらの問題について包括的に論じることとはとてもできないが、大きく分ければ、それらは一方では自然的（特に人間的）認知システムの能力の非完結性という事柄、他方では同じくこのようなシステムによる表象の処理に見出される（意味論的）関連性という事柄に関わっている。ここでは、後者の事柄に関わる一つの問題を取り上げ、すぐ後で簡潔に考察することにした。（16）

しかし、ごく直観的に考えても、古典的計算主義の認知描像には疑問が湧いてくるだろう。たとえば、コーヒーを飲みながら友人とおしゃべりをしているとしよう（17）。しゃべりながら私はカップを手にとってコーヒーを飲む。このとき明らかに、私の認知状態が推移して（つまり、どのようにしてか私はそのときコーヒー・カップに向かって手を伸ばす気になって、あるいはそうしようという意図が私の中にどのようにしてか生じて）、私の行動は起こったのである。しかしそのさい私の内部で、いつカップを手にとればよいのかについて明示的で例外のない規則に従っての計算過程が生じ、その結果として私はカップに手を伸ばしたのだ、などとすなおに考えられるだろうか。以上のような直観的考察をめぐらせただけでも、古典的計算主義の認知像が狭苦しすぎるものなのではないか、という疑いを禁じえなくなることだろう。少なくともわれわれ人間の思考（認知過程）は、もっと柔軟（ソフト）であるように思われるのである。

さて、古典的計算主義の認知描像に対してよく指摘されてきた重要な問題の一つは、いわゆる「フレーム問題」である。この問題の定式化はさまざまの仕方で可能であるが、たとえば次のような問題として考えることもできるだろう。認知システムが新たな情報を得たとき、そのシステムがすでに獲得している信念は適切な仕方で更新される必要がある。では、どの信念をどのように更新すればよいかを決定するための、実効性のある一般的方法はどのようなものだろうか。

通常のコンピュータでは、既得の情報はメモリー（記憶装置）に蓄えられている。しかし、それらの情報を一つ一つ、新たな情報に照らして更新すべきか否かチェックしていくという方法は、たとえ実行不可能ではないにせよ、人間的認知システムの実際のやり方とは考えにくい。そこで、古典的計算主義の内部で提起された方法は、更新の対象となりうる情報を、無関係な情報から区別して枠（フレーム）をはめる、という方法である（18）。つまり、新たな情報に関連性をもつ既得情報だけを、チェックの対象としてあらかじめ選び出しておくという方法である。

しかしながら、新たに獲得された任意の情報に関連性をもつ情報だけを選び出す、実効性のある一般的方法を見つけることは、人間のよう複雑で洗練された認知システムに関してはほとんど不可能であるように思える。なぜなら、与えられたコンテキストによって、任意の情報はほかの任意の情報に対して関連性をもちうるからである。たとえば、インドにおける現在の紅茶の価格は、太郎が午前8時半までに朝食をとるかどうかということに何の関連性ももたないように思える。しかし、もし太郎がインドの紅茶取引に多額の投資をしており、その価格が暴落したとすれば（そのようなコンテキストでは）、前者は後者に対して大いに

関連性をもつことだろう(19)。

このように「フレーム問題」は、古典的計算主義にとってはなかなか解決が困難な問題であるように思われる。そして、この立場にとって解決困難な問題はすでに述べたようにほかにも存在し、これらの問題はすべて人間の認知過程のソフトさに関わっていると考えられる。もちろん、これらの難題を解決することは古典的計算主義には不可能だ、ということが証明されたわけではない。しかし、この立場がこれらの問題の解決に成功しないままかなりの時間が過ぎているという事実は、少なくともそれにとって代わるべき新しい認知の理論を追求する強い動機を与えるものだ、と言うべきだろう。

#### 新たな認知の理論の前提 - - 「思考の言語」仮説

それでは、古典的計算主義にとって代わる新たな認知の理論は、どのような一般的前提にもとづいて構築されるべきであろうか。コネクショニズムと結びついた形で考えられる新しい認知の理論の具体的内実に入るまえに、この項ではこの問題について手短かに考えておきたい。

まず、上の「認知の理論の前提」という項で指摘した三つの主張（表象主義、機能主義、そして物理主義）は、そのまま前提として堅持されるべきだろう。なぜなら、強く言えば、これらの前提は、「認知科学」という企ての本質を構成するアイデアを表していると考えられるからである。

しかしこれらに加えて、次の二つの主張も前提として置かれるべきだろう。すなわち、

(a) 少なくとも多くの心的表象（認知事象）は統語論的構造をもつ。

および、

(b) 認知システムによるこれらの表象の処理は、表象のもつ統語論的構造に則して行われる。

(a)は先に指摘したように（本節 頁）、古典的計算主義の前提もしくは主張(4)にほかならない。また、そこで説明したように、この立場は(b)にもまたコミットしている。ところで、前提(a)と(b)は、思考に代表される心的表象（認知事象）が、文をはじめとする言語的表現と同様の構造をもつ対象であり、しかも、そのような言語的構造に従っての取り扱いをシステムの内部で受けるものである、と主張している。それゆえ、このような主張はしばしば「思考の言語」の仮説と呼ばれている(20)。したがって要するに、古典的計算主義にとって代わるべき新たな認知の理論は、認知科学における認知の理論一般の三つの共通前提は当然のこととして、さらにそれに加えて古典的計算主義を特徴づける前提の一つである「思考の言語」仮説をも、その理論的前提として引きつぐべきだと考えられるのである。なぜだろうか。

その根拠は、ホーガンとティーンソンによって説得的に述べられているので、その要点を確認しておくことにしよう(21)。

(ア) この世界において自然的認知システム、特に人間のようにきわめて複雑な認知システムが適応的に生存してゆくためには、それはそれを取り巻く環境のあり方をそのつど適切に把握する必要があるだろう。つまり、そのなかに存在するさまざまな持続的对象（個物）と、それらがもつ多様な性質や関係の複雑

な時間的变化とを追跡し続ける能力 をもっていなければならない。

(イ) また、環境の急な変化や自分自身の空間的移動のために、認知システムはしばしば 新しい環境に直面する。つまり、これまで出くわしたことのない新しい対象に遭遇し、それらがもつ未知の性質や関係を経験する。このような場合には、それらをそのようなものとして、すなわち新しい対象、あるいは対象の未知の性質や関係として適切に把握し追跡する能力が要求される。(22)。

(ウ) このように多数の(可能的には無数の)対象、およびその(可能的には無数の)性質ないし関係の変化を追跡しうるためには、少なくとも人間のように複雑な自然的認知システムは環境を表象する次のような方式をとっている、と考えるのがきわめて自然である。つまり、対象は対象として、その性質ないし関係は性質ないし関係として表象したうえで、それら種類を異にする表象を結びつけることによって、そのつど表象が必要な事態(ある対象がある性質ないし関係をもつ)の複合的な表象を作るという体系的な表象方式をとっている、と考えるのが自然である。(23)

(エ) さらに、複雑に変化する環境のなかで生き延びるためには、その環境の適切な表象を作るだけでは不十分である。認知システムはそれに加えて、そのような表象の(意味論的)内容に相応した行動をとってゆく必要がある。言い換えれば、おのおのの表象がその内容に相応したシステムの行動を引き起こすように、システム内部の因果的な構造化がなされていなければならない。たとえば、同一の対象に相異なる性質ないし関係を帰属させる諸表象は一般に、ある点(つまり同一の対象に関わる行動であるという点)では類似した、しかし別の点では(体系的に)相異なる諸行動を引き起こすように、システムができあがっている必要があるだろう。このようなことはしかし、それらの表象そのものがある点(つまり同一の対象への述定であるという点)では類似し別の点(相異なる性質・関係の述定であるという点)では相異なる構造をもつことによって互いに体系的に関係づけられているのでなければ、ほとんど実現不可能であろう。

(オ) 以上のことが意味するのは、人間のような認知システムのうちには自然言語と同様の統語論的機構が内蔵されていると考えることがほとんど避けがたい、ということである。つまり、要素的表象を規則に従った仕方で複合するという表象方式(<複合性>)をとることによって、必要に応じてさまざまな複合的表象(ときにはまったく新しい表象さえも)を自在に作り出す(<産出性>)とともに、作り出される諸表象どうしの(意味論的に)体系的な関係(<体系性>)を可能にする機構である。さらに、このような体系的表象方式に従って作り出される各表象がまさにその統語論的構造に則した処理を受け、表象の内容に相応した行動へと導くのだと考えることによってはじめに、認知システムの適応的生存の明確な説明が得られる。要するに一言で言えば、「思考の言語」仮説は、自然的認知システムのこの世界における生存を説明してくれるきわめて有力な仮説なのである。

#### フォーダー説の批判的検討(1) - - 古典的表象は不可避か

ホーガンとティーンソンの「追跡論証」はたいへん説得的であり、したがって「思考の言語」仮説は古典的計算主義に代わるべき新たな認知の理論にとっても必須の前提であると考えられる。しかし、まさにここで、本節のはじめのほうで紹介したフォーダーらによる「体系性からする論証」が、この新たな認知の理論の可能性に対する脅威となるように思われるのである。しかし、それは真の脅威

なのだろうか。この項では、この論証に含まれているフォーダーらの考え方の核心に対する批判を試みたい。

さて、彼らの考え方の骨組みは次のようなものである。

1. 自然的認知システムがもつ心的表象（ないしはその認知状態）は、もしそれがおよそ統語論的合成構造をもっているならば、すべて「古典的表象」である。つまり分かりやすく言い換えれば、心的表象は、その統語論的構成要素が全体表象の時空的部分を成すような構造になっているのである。(24)

2. したがって、心的表象（のトークン）においては、その統語論的構成要素（のトークン）がすべて（もとの表象の部分として）明示的・顕在的に実現されており、それゆえ各要素がその因果的効力を発揮することができる。だからこそ、もとの表象の統語論的構造に則した、さらにはまた意味論的内容に相応した表象の処理（言い換えれば、認知状態の推移）が可能となるのである。(25)

3. そして、表象の統語論的構造および意味論的内容に相応した表象処理が、これ以外の仕方でも可能であるとはとても考えられない。つまり、その統語論的構成要素が必ずしも全体表象のうちに顕在的・明示的に実現されない非古典的表象については、そのような表象処理が可能だとはまず考えられない。

フォーダーらの第一の主張は、統語論的構造をもつ心的表象はすべて上の意味での「古典的表象」だ、ということである。およそ統語論的合成構造をもつ表象が「古典的表象」でないことはありえないのだから、これはあまりにも明らかなことだ、と考えられてしまうかもしれない。しかしホーガンとティーンソンが指摘しているように(26)、非古典的な統語論的構成要素から成る表象（もしくは表象方式）というものは、じつはさほど珍しいものではなくむしろありふれたものである。問題点を明らかにするために、このことをまずはっきりさせておこう。たとえばごく単純な例を挙げれば、英語の"went"という表現（表象）は、フォーダーらの定義によれば、非古典的表象だと言わなければならない。なぜなら、それは明らかに単純表象ではなく、たとえば"laugh-ed"などと同様に動詞（の原形）と過去時制を表す統語論的要素との複合した表象とみなされるべきであるが、それにもかかわらずその統語論的構成要素（"go"と"ed"）はそのうちに顕在的・明示的には実現されていないからである。(27)つまり、英語という自然言語も、一見その反対に思えるかもしれないが、じつは厳密には「古典的表象」方式をとってはいないのである。そして註の(27)で触れたように、英語などよりもはるかに非古典性の程度の高い自然言語も多数、存在する。しかも、われわれは（少なくとも当該の言語を習得した人間は）、これらの言語に含まれる、非古典的な構成要素から成る諸表現（表象）を何の苦もなく適切に処理している。このような事実を照らすならば、およそ統語論的構造をもつ心的表象（思考の言語）はすべて古典的表象である、というフォーダーらの主張がそれ自体では何ら明白なものではないことが理解できるだろう。

#### フォーダー説の批判的検討(2) - - 何が真の問題か

じっさい、フォーダーら自身、この第一の主張がそれ自体で明白に成り立つ主張だ、と考えているわけではない。彼らがこの主張に固執する理由は別に存在する。彼らはたとえば次のように述べている。

複合的な活性ベクトル(28)の構成要素は、通常、「そこにある」わけではない。だから、もし複合的ベクトルのトークンの因果的帰結がその合成構造に応じたものになっているとしたら、それは奇跡と言うべきだろう。(29)

フォーダーらが非古典的表象(方式)の可能性を認めていることは、この引用文から明白であろう。したがって、彼らの議論のポイントはこの種の表象の不可能性を言い立てることにあるわけではない。この種の表象(方式)を採る認知システムが仮に存在したとしても、そのようなシステムは表象の統語論的合成構造に則した表象の処理を行うことがまずできないだろう、ということこそ彼らの真のポイントである。なぜなら、非古典的表象のうちにはその統語論的構成要素(のトークン)は顕在的には存在せず、したがって当然、存在しない構成要素はその因果的効力を発揮できないからだ、というのである。(30)

したがっていずれにせよ、非古典的表象(方式)が可能か否か、が真の問題ではないことは明らかであろう。問題はむしろ、古典的か非古典的かどちらの表象(方式)が採られるにせよ、表象処理のやり方が古典的計算主義の主張するようなものであるのか否か、なのである。では、表象の統語論的構造に則した(応じた)表象処理は古典的計算主義の主張するような仕方ではかまらず可能ではない、というフォーダーらの主張(上述の主張2および3)は真であり、また、そのための彼らの議論は適切なものなのであろうか。

ふたたびホーガンとティーンソンの指摘するように(31)、フォーダーらの議論は的をはずしていると言うべきである。統語論的合成構造に則した複合的表象の処理にとって肝心なのは、フォーダーらの言うように、その表象の統語論的構成要素(のトークン)が因果的効力を発揮するか否かである、のではない。そうではなく、その表象がある一定の統語論的構造をもつという事実によって当の表象自体がその構造に応じた因果的役割を果たしうるかどうか、ということこそが問題のはずである(32)。表象トークン自体がこのようにしてその統語論的構造に応じた因果的役割を担っていないとすれば、(じつはそれは本来の意味での「表象」トークンとは言いがたく、それゆえ)フォーダーらのように、ある物理的アイテムをその表象トークンの統語論的構成要素とみなそうとすることさえ、そもそも意味をなさないだろう。つまり、複合的表象の構成要素の因果的効力ということの有意味に言うためにさえ、じつはその表象自体が求められている因果的効力をすでにもっていることを認めなければならない。しかし、この肝心のことを認めるのに必要なのは、表象の古典性でも構成要素の因果的効力でもない。必要なのは、表象がともかく何らかの仕方ですべて一定の統語論的合成構造をもつこと、そしてこの構造に応じた因果的役割を当の表象に割り当てるようなシステムが形成されていることなのである。そして、非古典的表象もまた(古典的表象とは異なった仕方ですべて)統語論的構造をもつことは、すでに前節で確認済みのことである。

しかしさらに、表象処理において肝心のことは表象の構成要素そのものがではなく、表象がある統語論的合成構造をもつという情報が因果的効力を発揮することなのであるから、複合的表象の処理は「そこにある」表象の統語論的構成要素(のトークン)についての処理、つまり、あくまでも表象(記号)レベルの規則による表象処理(33)である必要もないことになる。つまり、(特に非古典的表象については)表象の統語論的構造に則した処理は、古典的計算主義の主張するような仕方で行われざるをえないというわけでもないことになるのである(34)。このようにフォーダーらの議論は、それが望む結論(古典的計算主義の主張する表象処理方式の不可避性)を打ち立てるには程遠いものに終わっている、と言わなければならない。しかし、表象処理の非古典的方式とはどのようなものなのか。

そのある程度具体的なイメージを示すことが、古典的計算主義にとって代わろうとする新たな認知理論にとっては重要な課題となるだろう。

#### 一つの経験的事例：RAAM ネットワーク

この課題について考える前に、表象の統語論的構造に則した処理が行われているとみなすことのできる、コネクショニスト・システムの一つの実例を簡単に見ておくことにしよう。それはポラックによって構築された「RAAM」と略称されるネットワーク(35)である。このネットワークを用いて彼が達成しようとした主要な目的は、固定した「幅」の表象（つまり、固定した数のユニットのセット）によって、さまざまな長さの回帰構造をもつ諸対象（じっさいには文の樹状構造）を表象させることであった。それがどのように達成されたのか、ネットワークの構造ならびに機能と彼の研究結果の概要を確認しよう。

- - 図 1 - -

図 1 にあるように、RAAM は三つの層から成る。入力層および出力層はおのの、16 個のユニットの集合体 3 つから成る。隠れ層は、16 個のユニットの集合体 1 つだけから成る。隠れ層の各ユニットは、入力層あるいは出力層の 48 個のユニットすべてと結合されている。さて入力層ではまず、与えられた文の述語部分、主語部分、目的語部分がこの順序で 3 つのユニット集合体のおのののに入力される（入力される各語にあらかじめ割り当てられている活性化パターンがユニット集合体のおのののに提示される）。次に、入力層とのユニット結合をとおして隠れ層において、この文の「圧縮表象」が形成される（つまり、入力層において 48 個のユニットによって表されていた文の各部分に関する情報と、文全体の構造に関する情報とが 16 個のユニットのうち圧縮されることになる）。図 1 に示されているのは回帰構造を含まない単文のケースであるが、回帰構造を含む場合には、まず最深の埋め込み文が入力され、その後その文の圧縮表象が入力層の第 3 のユニット集合体にコピーされるという仕方で次のステップの入力が行われる(36)。そして、こんどは複合文の圧縮表象が隠れ層に形成される。以上の過程が与えられた文の回帰構造の深度に応じて必要なだけ繰り返される。このようにして与えられた文のコード化のプロセスが完了した後、脱コード化（解読）のプロセスが始められる。こんどは、まずもっとも「外側」から出力層における解読が行われ、次にもっとも浅い埋め込み文の圧縮表象が隠れ層に戻されてその解読が行われる、という仕方で最深の埋め込み文に到るまで脱コード化のプロセスが続けられる。

RAAM の構造と機能は以上のようなものであるが、ポラックはこのネットワークを 13 個の文（その回帰構造の「最大深度」は 4 である）で「訓練」した。その結果、訓練キットの 13 個の文ばかりではなく、それ以外の 12 個の文についても、RAAM は与えられた文の回帰的樹状構造を正しく再構成することができるようになった、ということである。

この結果は次のような意義をもつものと理解することができる。イ)限られた範囲においてはであるが、RAAM ネットワークは（文の）表象の統語論的合成構造に応じた処理を行うシステムとみなすことが可能である。入力された（回帰構造をもつ）文の統語論的構造に一致した表象が、隠れ層の「圧縮表象」を介して、出力層において出力されるからである(37)。それにもかかわらず、ロ)与えられた文の圧縮表象がその統語論的構成要素を明示的に含んではないことは明白で

ある。なぜなら、隠れ層の各ユニットは入力層の3つのユニット集合体のすべてと結合しているからである。文の3つの構成要素に関する情報は隠れ層のすべてのユニットに完全に分散して受け継がれており、特定の構成要素の表象にだけ関わるユニットは存在しない。したがって圧縮表象は明らかに非古典的である。さらに、八)RAAMにおける表象の処理が古典的計算主義の処理方式に従っていることは、まずありそうもない。たしかに、個々のユニットの活性化の度合いは、ほかのユニットとの結合の重みづけに応じたあるアルゴリズムによって司られている。しかし、口)において確認したように、(特に隠れ層の)おのおののユニットそれ自体はいかなる特定の表象内容も担ってはいない。それゆえ、RAAMにおいては、表象レベルのアルゴリズムによる表象の処理が実行されている、とはきわめて考えにくいのである。

もちろん RAAM はきわめて単純な表象処理システムにすぎない。にもかかわらず、それは、非古典的表象の非古典的処理によってわれわれ人間のよう複雑このうえない認知システムが成立しているのかもしれない、という可能性を示唆してくれる人工的システムの好事例だと言うべきだろう。

表象の非古典的処理とはどのようなものか：打ち消し可能な因果的傾向性(38)

それでは、前々項の終わりに提起した問題にもどることにしよう。古典的計算主義が主張する方式によらない、しかも表象の統語論的構造に則した表象処理とは、どのようなものなのか。もちろん、それが表象レベルの規則に従っての計算によるものではないことは、すでに明らかになっている。しかし、表象の非古典的処理というものをより具体的・積極的に特徴づければ、どういうことになるのだろうか。

ホーガンとティーンソンが提案するのは、表象の「認知的力」というアイデアを文字どおりに受けとめるという考え方である(39)。この考え方によれば、自然的認知システムの各表象はその意味論的内容に応じた力を文字どおりにもっている。そして、このようなシステムにおける認知過程は次のようなものである。つまり、関係する諸表象がいわば一つのじょうごに一挙に投げ入れられ(40)、そこで相互作用させられる。そこから出てくる結果、つまりシステムの次の認知状態は、それらの表象の間の力関係によって自動的に決まる、というような過程である。それはちょうど、綱引きの結果が二組の引き手の間の力関係で決まるようなものである(41)。もちろん、引き手たちの力や綱の摩擦係数を数値化することによって、綱引きの経過と結果をコンピュータ・シミュレーションすることは容易だろう。しかし、現実の綱引きにおいては、シミュレーションにおけるような計算はいっさい行われず、ただ力の相互作用が生じているだけである。同様に、現実の認知過程もまた、諸表象間の力の及ぼし合いによるのであり、本来の意味での、つまり表象レベルでのいかなる計算も含まない、というのである。

表象の「認知的力」とはどのようなものなのか、より詳しく見てみよう。たとえば、眠れぬ夜を過ごしていた太郎は、ビールが飲みたくなった(42)。ビールは階下の台所の冷蔵庫にあるはずだった(と彼は信じていた)。そこで、太郎はベッドから起き上がり、階段を下りて台所まで行き、冷蔵庫を開けてビールを取り出した。もちろん、太郎のこの行為は、ビールを欲し、かつ、それは冷蔵庫の中にあると信じているという彼の認知状態(表象)のゆえに行われたのである。つまり、(すでに明らかになっていることだが)表象の認知的力とは、システムを別の認知状態(表象)へと向かわせる因果的傾向性にほかならない。

しかし、もしそのとき太郎が夜中に動きまわって家人の眠りを妨げるのは忍び

ないと思ったとすれば、彼はビールをあきらめたかもしれない。あるいは、太郎がダイエット中だったとすれば、彼の理性が欲望にまさることになったかもしれない。あるいはさらに、不審な物音を聞きつけたとすれば、何が起こったのか確かめようと台所ではなく音のしたほうへと向かったことだろう。つまり、さまざまの表象が太郎のビールへの欲求に打ち勝って彼を別の行為へと向かわせることがありうるのである。表象の認知的力は、ほかの表象（の因果的傾向性）によって打ち消し可能な因果的傾向性なのである。

しかも、ある表象（たとえば太郎のビールへの欲求）の因果的傾向性を打ち消すことのできる表象にはかぎりがないように思える。たとえばさらに、台所まで下りてビールをとってもどってくる間の寒さの思いが太郎を押しとどめることになったかもしれない。あるいはまた、台所に行けば太郎が恐れているある対象を避けることができないことが明らかになったとすれば、それもまた太郎にビールへの欲求を断念させることがありうることだろう。このように、仮に現実化したとすれば（たとえば）太郎のビールへの欲求に打ち勝つことになるであろう表象（認知状態）の数と種類にはかぎりがないのだとすれば、それは、表象同士の働き合いの過程は表象レベルの計算過程（アルゴリズム）にほかならない、という考えにとってはきわめて不利な事実であろう。なぜなら、それは、与えられた表象（たとえばビールへの欲求）を打ち消しうる表象であるための必要にして十分な条件をあらかじめ定めることは不可能だということを強く示唆するからである。しかし、その必要十分条件が特定されないかぎり、与えられた表象が打ち消される（例外的）場合のための、表象レベルのアルゴリズムは存在しえないのである。（43）

以上のように、ホーガンとティーンソンが提唱する認知の描像、つまり認知過程を表象のもつ打ち消し可能な因果的傾向性（認知的力）の間の相互作用と捉える描像は、（おそらく）古典的計算主義とは両立しえない描像である。つまり、彼らの描像によれば、われわれの認知過程は計算可能（ある適切な意味で）な関数として表されうるようなものではないのである（44）。他方、認知的力の相互作用としての認知過程という描像は、コネクショニズムとはきわめて相性がよい。コネクショニスト・ネットワークにおける各表象の（物理的）実現は、もしそれがいやしくも可能であるのなら、各ユニットの活性化にもとづく実現であることだろう。しかし、あるユニットの活性化は本来、（一定の重みづけをもつ）結合をとおして他のユニットの活性化に導くようなものである。したがって、ネットワークにおいて実現される表象があるなら、それらはすべてほかの多くのユニットの活性化、つまりはほかの表象を引き起こす因果的傾向性をもつだろうと考えられるのである。それゆえ、ネットワークにおいては、任意の複数の表象が同時に実現された場合には、それらは自動的に各々の因果的傾向性に応じて相互作用し合うことであろう。したがってまた、（太郎のビールへの欲求と、それが台所の冷蔵庫の中にあるという信念、そしてまた彼の欲求を打ち消しうる諸表象のように）互いに意味論的な関連性を有する諸表象が同時に実現された場合には、それらは自動的におのおのの意味論的内容に相応した仕方で相互作用し合うことであろう。しかも、ネットワークにおけるそのような表象間の相互作用がどのような結果に導くのか（ビールへの欲求が貫徹されるのか、それとも打ち消されるのか）もまた、そのつどの関係する諸表象間の認知的な力関係による自動的決定に委ねられるのであって、古典的計算主義におけるように表象レベルの例外なき規則のうちあらかじめ組み込まれている必要などはないのである（45）。

ネットワークにおいて表象はどのように物理的に実現されるのか（46）（47）

最後に、コネクショニスト・ネットワークにおける表象（認知状態）の（特に物理的）実現は、どのような仕方で果たされるのか、という問題について手短かに考察しておきたい。この問題に対する答えは、これまで述べてきたことのうちにある意味ではすでに含まれている。しかし、この問題に関する十分な理解が従来なされてきたとは言いがたいように思われる。たとえば、ネットワークにおいて実現される表象は（多数のユニットに）「分散」している、というようなことがしばしば言われている。しかし、「分散表象」、つまり一組のユニットの一定の活性化状態（活性化パターン）は、正確にどのような意味で表象の実現になりうるのだろうか。その点が十分明確になっていなかったために、たとえばフォーダーらにおけるように、「分散表象」は真の意味での表象ではない、というような議論さえ提出されることになったものと推測される。それゆえ、本項の表題に掲げた問題についてあらためて考察することはけっして無意味なことではないだろう。

さて、たしかにユニットの活性化パターン（より正確には、その物理的実現態）は、ある意味では、コネクショニスト・ネットワークにおいて表象を物理的に実現するものである。しかし、すでに述べたとおり、活性化パターン自身の内在的な物理的性質のうちには、それが実現している表象の統語論的構成要素に顕在的・明示的に対応するものは存在しない。つまり、活性化パターンによる表象の実現は内在的な実現ではない。むしろ、活性化パターンは、当の表象を傾向的に実現する。つまり、その表象がほかの表象との関わりにおいてもつ因果的傾向性が、活性化パターンの物理的傾向性によって実現されるのである。

さらに、活性化パターン（の物理的実現態）は、じつは単一の表象を実現するというよりは、むしろシステムの（そのときどきの）全体的認知状態（それは場合により、複数の単位表象の重ね合わされた状態でありうる）を実現するものである。そしてふたたび、その全体的認知状態が表象の複合体である場合に、それがいかなる表象の複合となっているかは、活性化パターンの物理的な内在的性質のうちには明示されない。したがってやはり、活性化パターンによる実現はここでもまた傾向的な実現である。つまり、システムのある全体的認知状態の因果的傾向性が、当の活性化パターンのもつ物理的傾向性によって実現されるのである。

では、ある活性化パターンの物理的傾向性がある表象の、もしくはある全体的認知状態のもつ因果的傾向性の物理的実現である、とどうして言うことができるのだろうか。その根拠は、最終的には次の点にあると考えるべきだろう。すなわち、問題となっている認知システムがもつ認知的潜在性(48)の総体と、このシステムを物理的に実現すべき物理的システム（ネットワーク）がもつ物理的潜在性(49)の総体とが構造的に対応し合う(50)、という点に存すると考えるべきだろう。言葉を換えて説明しよう。認知システムがその状態になりうるころの全体的認知状態は、その因果的傾向性のゆえに、ある全体的状態の次には別のある全体的状態へと、さらにこの第二の全体的状態は次に第三の全体的状態へと推移する傾向を有する等々といった相互間の推移関係によって、総体としていわば一つの地形を構成する。同様に、ネットワークの可能的な全体的物理状態もまた、その因果的傾向性にもとづく相互間の推移関係によって、総体としていわば一つの地形を構成するだろう。これら二つの総体的地形の一致対応（認知の総体的潜在性構造の物理的実現）こそが、個々の表象、もしくは個々の全体的認知状態の個々の全体的物理状態による実現（おのおのがこの地形上の同じ位置を占めることにもとづいて）を成り立たせる。けっして、古典的計算主義が言うように、個々の表象の（統語論的）構造が物理的状態の構造のうちには明示的に写しとられる

ことによって表象の物理的実現がいわば個別的に果たされるわけではない、と考えられるのである。

## 本節のまとめ

以上、本節においては、主としてホーガンとティーンソンのアイデアに従って、古典的計算主義にとって代わるべき新たな認知の理論の一つをきわめておおまかにスケッチしてきた。それによれば、古典的計算主義とは対照的に、(われわれ人間の)認知の描像は以下のようなになる(51)。

(1) 認知状態間の推移は、物理的手段によって計算可能な関数に従って生じるわけではなく、むしろ、おのおのの認知状態がもつ認知的力の競合と協力によって決定される(トップレベル)。

(2) 認知状態間の推移を実現する因果的プロセスは、表象(記号)レベルの例外なき規則に支配される過程として、言い換えればコンピュータ・プログラム化可能なアルゴリズムとして数学的に表現されうるようなものではない。それはむしろ、数学的には一つの動的システムとして表現されるようなものである(中間レベル)。

(3) このような動的システムの各状態の(したがってまた認知状態の)物理的実現は、脳神経系の物理的状态(もしくはプロセス)の現実的性質による顕在的・明示的な実現ではなく、むしろ、その潜在性による傾向的実現である(ボトムレベル)。

以上のような明確な対立点にもかかわらず、この新たな認知の理論は古典的計算主義と一つの重要な前提を共有している。それは言うまでもなく、「思考の言語仮説」という前提、つまり「脳神経系の物理的状态にもとづいて実現される(少なくとも多くの)認知状態は統語論的構造を有し、かつ、この構造に従って処理される」という前提である。

ホーガンとティーンソンが提示したこのような新しい(非古典的な)認知描像の価値はどこにあるのだろうか。一つには、本節において繰り返し述べてきたように、コネクショニスト・ネットワークのもっともパズリングな特徴(の少なくとも一つ)に対して、(別の点で大きな代償を払うことなく)納得のいく説明を与えるための道筋を示したことであろう。その特徴とは、ネットワークがある種の認知的タスクを古典的システムよりもはるかによく遂行しうるにもかかわらず、ネットワークのどの状態のうちにも、そのタスクに関連しているはずの表象に固有の構造(とくに統語論的構造)を目に見える形で(顕在的・明示的に)実現していると言えるような性質が見当たらない、ということであったと言えるだろう。そのために、一方ではコネクショニズムは古典的認知モデルを物理的に実現するための単なる代替手段にすぎないという消極的評価(居直った古典的計算主義)と、他方ではわれわれの常識的信念に反して認知のプロセスはじつは表象や統語論ぬきで行われるものなのではないのかという(いわば)哲学的冒険主義(さまざまの形の消去主義)とが生まれてきたのだと考えられる(52)。ホーガンとティーンソンは、一つには物理的システムによる表象の傾向的実現というアイデアと、いま一つには各表象がもつ打ち消し可能な因果的傾向性にもとづく表象同士の(計算によらない)相互作用をとおしての認知的推移(表象処理)というアイデア(53)によって、ネットワークにおける表象の実現がいかに原理的に可能であるのか、を示すことに成功した。そして、そのことによって、コネクショニズ

ムに関する上のようなきわどい哲学的評価がいずれも拙速にすぎる評価であることを示したのである。

しかしながら、ホーガンとティーンソンの新しい認知描像の真の価値は、むしろ次の点にあると考えるべきだろう。つまり、いま述べた作業を遂行するさいに、そもそも認知科学における認知の理論とはどのようなものであるべきであるのか、古典的計算主義とは正確にどのような特徴をもつ認知の理論であるのか、したがってまた、この理論にとって代わるべき新しい認知の理論はどのような要件を満たす理論でなければならないのか、といったきわめて基本的な問題に対して（おそらくはじめて）意識的に明確な答えを与えようとした点にある。これらの基本的問題に答えを与えることによってはじめて彼らは、コネクショニズムに関する上述のパズルを納得のいく仕方で解き、コネクショニスト・ネットワークという（物理的数学的）機構が認知の理論一般に対してもちうる本来の意義を正当に評価する可能性を開くことができたのだ、とすることができるだろう(54)。

むろん、ホーガンとティーンソンが示した認知描像は、古典的計算主義にとって代わりうる唯一の認知描像ではない。コネクショニズムと親和的な認知描像はほかにもありうるだろう。また逆に、彼らの認知描像は、コネクショニズム以外の物理的数学的機構とも親和的であるかもしれない。そして、この認知描像のもとで捉えられたものとしてのコネクショニズムが、われわれ人間という認知システムの正しい記述となっているのか、という問題はまだほとんど手つかずのままであると言うべきだろう。言うまでもなく、この問題に肯定的に答えることができるようになる前に、彼らの認知描像とコネクショニズムはさまざまなアプリアリな討究に耐え、経験的なテストをクリアする必要があるだろう。このように彼らの新たな認知描像に関して残されている問題は数多い。しかしそれにもかかわらず、この認知描像は、今後、認知科学およびそれと結びついた（いわば）認知哲学における真剣な討議の対象とするに十分に値する内容を提示していると考えられるのである。

## 註

(1) このように述べたが、本節における私の論述は実質的に、Horgan and Tienson(1996)に対する解説と注釈 - - しかも彼らの主要なアイデアに限定しての部分的で不完全な解説と注釈以上のものではない。この書において体系的に展開されているホーガンとティーンソンの立場は、十年に近い年月にわたって公刊され続けてきた、彼らの多数の共著論文をとおして発展させられ、練り上げられてきたものである。しかし、これらの共著論文で示されたアイデアの多くはこの書のなかに吸収総合されていると考えられるので、以下の参考文献にはこれらの論文は省略し、この書の名のみを挙げることにする。彼らの共著論文に関心のある向きはこの書の末尾にある文献表を参照されたい。

(2) この論議の発端となったのは、1999年11月の日本科学哲学会第32回大会におけるシンポジウム「コネクショニズムの哲学的意義」であった。このシンポジウムにおける議論の成果の一部は、同学会の学会誌『科学哲学』33巻2号に収められている。信原氏の立場は、その後さらに信原(2000b)として発展させられることになった。服部氏の考えは、本書所収の論文(第2章節)においてさらに詳しく展開されている。

(3) 信原(2000a), p.12.

- (4) 服部(2000), p.22.
- (5) Fodor and McLaughlin(1990), p.200. ただし、頁付けはリプリント版(Macdonald and Macdonald(1995))に従う。
- (6) Fodor and McLaughlin(1990), p.199を参照。ちなみに、先に挙げた信原氏の見解はこのディレンマの第一の角に凭れ、他方、服部氏の見解は第二の角に凭れることによって新たな認知の理論としてのコネクショニズムを攻撃している、と言えるだろう。
- (7) 構文論的構造に則した表象の処理は体系性の説明には不可欠だとされているのだから、したがって結局、彼らはコネクショニズムに限らず、古典的計算主義以外のいかなるモデルについても、それが新たな認知の理論である可能性を否定することになる。つまり彼らによれば、古典的計算主義だけが唯一可能な認知の理論であることになる。
- (8) 三つのレベルの区別は元来、Marr(1982), pp.24-25(邦訳 pp.26-28)において古典的な形で与えられた。ホーガンとティーンソンによるこの区別の一般化については、Horgan and Tienson(1996), pp.45-46を参照。
- (9) より正確に言えば、システムのある時点における全体的認知状態。
- (10) 逆に言えば、非表象主義(認知事象の表象性の否定)を帰結として伴うような特定の認知の理論は、いわば自滅的な理論だとみなされるべきだろう。そのような理論は出発点における前提であった現行の認知概念を放棄し、そのことによって「認知の理論」の名に値しないものになり果てるだろうからである。近年、「非表象主義」ということが認知科学およびその周辺で取りざたされるようになってきている。このレッテルを冠せられることもある人々の論稿をいくつか収めた論文集『ハイデガーと認知科学』(産業図書 2002年)が最近公刊された。しかし、この書物を十分に参照する暇はなかったので、そこに論稿が収められているいずれかの人々の立場が「自滅的な理論」であるのか否かについては、本節では論じることはいできない。
- (11) Horgan and Tienson(1996), pp.24-27を参照。
- (12) 単なる数学的計算可能性ではなく「適切な物理的手段による計算可能性」という制約は、認知科学の関心が(人間を含めて)物理的に実現可能な認知システムに限定されているということに由来する。ここで問題にされている「計算」は、当然、「古典的意味での計算」、つまり表象レベルの規則による計算という限定された意味である。Horgan and Tienson(1996), pp.25-27、さらにこの箇所につけられた註12~14も参照。
- (13) トップレベルの認知推移関数はこのようなアルゴリズムに従ってより具体的に計算される、と言ってもよい。
- (14) ただし、それはつねに「正しい」解決であるとはかぎらないし、一つの解決しかないというわけでもない。
- (15) このような表象処理は、すでに述べたように、表象の統語論的構造に従うものである。しかも、少なくとも最終的には、表象のもつ意味ではなく、もっぱらその形式的性質(形態)だけに依拠する処理だと考えられている。だからこそ、このような表象処理は純粋に物理的に実行可能だ、というわけである。
- (16) この項における論述は、Horgan and Tienson(1996), Ch.3に多くを負っている。また、Copeland(1993), Ch.5をも参照。
- (17) Horgan and Tienson(1996), pp.31-32を参照。
- (18) 「フレーム問題」という名はここに由来する。
- (19) Horgan and Tienson(1996), p.38を参照。
- (20) もはや周知のとおり、この仮説の古典的定式化はFodor(1975)において与

えられた。(21) Horgan and Tienson(1996), pp.81-86 を参照。

(22) ホーガンとティーンソンの議論は、(ア)および(イ)で述べられた事柄を出発点とするので、彼ら自身によって(思考の言語仮説のための)「追跡(能力からする)論証 tracking argument」と名づけられている(cf. Horgan and Tienson(1996), p.72; pp.81-86; et passim)。

(23) システムが表象することができなければならない、対象とその性質ないし関係との組み合わせは可能的に無限なのだから、それらの組み合わせすべての表象があらかじめシステムにセットされているとはきわめて考えにくい。また、システムはそれらの組み合わせのおののに対して単純な表象を作るのであって、新しい組み合わせの表象が必要になった場合にはいわばその場しのぎ的に新たな単純表象を用意するのだ、とも考えにくい。後者の点に関しては、以下の本文における考察をも参照されたい。

(24) より正確な定義は次のようなものである。「表象 E 1 が表象 E 2 の古典的構成要素であるのは、E 2 のトークンが実現されるときにはつねに E 1 のトークンも実現される場合にかぎる。」(Fodor and McLaughlin(1990), p.201.)この定義中の「トークン」とは、ある種類(タイプ)の表象の具体的・個別的な実現態のことである。

(25) フォーダーらはもちろん、表象レベルの例外のない明示的規則によって処理が行われると考えている。

(26) Horgan and Tienson(1996), pp.71-74 を参照。

(27) 同様に、"men"や"women"といった表現もまた非古典的表象である。名詞と複数を示す統語論的要素との複合した表現でありながら、それらの要素は明示化されていないからである。おそらく厳密に古典的な表象方式をとる言語は、論理言語に代表されるいくつかの人工言語にかぎられるだろう。

ところで、本節の草稿をある場所で発表したとき、"went"などは単なる例外としてかたづけることができるのではないか、という疑義が呈された。つまり、英語など自然言語の表象方式は基本的に古典的であって、少数の例外的ケースに対してはそれを扱うための特別の明示的処理規則を設ければ、古典的計算主義の表象処理方式を貫徹できるのではないか、という趣旨の反対論である。しかしながら、自然言語のうちには、非古典的表象(表現)がけっして「例外」とは言えないような言語が存在する。たとえばギリシア語やラテン語のような言語では、動詞の時制や数や態、名詞の格や数は、おののの統語論的構成要素の明示的複合によってではなく語形変化によって示されるのが、例外ではなく「規則」である。そして、思考の言語はこれらの自然言語よりもはるかに高い程度に非古典的な表象方式をとっているかもしれないのである。

また同じ機会に、"went"などは単に表層的表現にすぎず、その深層構造はいわゆる規則動詞と同じく、動詞の原形と過去時制を示す要素との明示的複合になっているはずだ、という批判も受けた。しかし、このような批判は見当ちがいと考えられる。なぜなら、古典的表象も非古典的表象も、もし(これはやや粗い言い方だが)それらの意味論的内容が同じであれば、(批判者のことばを借りれば)「深層」においては同じ統語論的合成構造を有するということは、いわば議論の前提になっているからである。そのうえで、たとえば言えば、"went"式の表象方式をとるかそれとも"go-ed"式の表象方式をとるか、が争点となっている。ふたたび批判者のことばを借りれば、まさに「表層」の表象方式のちがいこそが問題なのである。

(28) P. スモーレンスキーによって提案された、ネットワークにおける心的表象の表現方式(もしくは実現形式)のこと(cf. Smolensky(1988),(1991))。それ

によれば、複合的な心的表象は、その各構成要素を表す活性化ベクトル（これは、関連する各ユニットの活性化の度合いを一つの座標軸とする高次元ベクトル空間中の、一つのベクトルになる）に対して、テンソル積と重ね合わせという二種類の操作を加えることによって得られる一つのベクトルとして表現される。たとえば、"John loves Mary"という複合的表象は、

$[v(\text{John}) \times v(\text{subject})] + [v(\text{loves}) \times v(\text{verb})] + [v(\text{Mary}) \times v(\text{object})]$

という「複合的な活性化ベクトル」として表現される（"×"はテンソル積、"+"は重ね合わせを表す。また"v(subject)", "v(verb)"などは各構成要素が果たしている統語論的役割に対応するベクトルを示している）。「複合的な活性化ベクトル」はそれ自体、単一のベクトルになるので、そこには各構成要素のベクトル（あるいは、その果たす役割のベクトル）は明示されないことになる。

(29) Fodor and McLaughlin (1990), p.215.

(30) 上述のフォーダーらの主張の2と3（本節 頁）を参照。

(31) Horgan and Tienson (1996), pp.79-80 を参照。

(32) これは結局、表象のシステム（つまり認知システム）全体の因果的構造の問題である。つまりシステム全体が、表象から表象への推移がつねに各表象の統語論的合成構造に応じたものになるであろう、そういう因果的構造になっているかどうか、の問題である。言い換えれば、各表象は何らかのシステム全体と相対的にのみ、一定の統語論的構造をもつのであり、かつまた、それをもつという情報を担うのである。ここには認知の理論の第二の前提である機能主義の考え方（本節 頁参照）がひじょうによく現れていると言うべきだろう。さらに、本節の最後から二番目の項における考察をも参照されたい。

(33) 古典的計算主義の前提(2)（本節 頁）を参照。

(34) このポイントを、信原(2000a)における、「機能的合成性」概念にもとづくヴァン・ゲルダーのコネクシオニズム擁護論に対する駁論(pp.5-8)は捉え損なっている。そのため信原氏の駁論は、表象の統語論的構造に則した処理はすべて古典的計算主義の処理方式に一致せざるをえない、という誤った前提に立ってしまっている。

(35) RAAM についての記述は主として Bechtel and Abrahamsen (2002), pp.170-178 に依拠した。"RAAM"は、"recursive auto-associative memory"(回帰的自動連合記憶装置)の略称である。RAAM に関しては Pollack (1990)も参照。

(36) たとえば(Knew Pat(Loved John Mary))という複合文の場合には、(Loved John Mary)の「圧縮表象」(LJM)と表すことにする)が形成された後、第2段階の入力において入力層の第3のユニット集合体に(LJM)がコピーされ、残る2つのユニット集合体に'Knew'と'Pat'の入力が行われる。また(LJM)のような圧縮表象は、解読過程において生じる対応表象の正誤をチェックするための「標的」としても用いられる。

(37) より正確には、入力パターンと出力パターンとの、広汎な事例における一致というこの事実が、RAAM のユニット集合体の活性化パターンを統語論的構造をもつ表象とみなすことを可能にするのである。註(32)を参照。

(38) 「打ち消し可能な因果的傾向性」は"defeasible causal tendency"の訳である。傾向性一般に関しては、本書所収の柏端論文（第 章第 節）の 2.に有益な説明がある。

(39) たとえば Horgan and Tienson (1996), pp.95ff. を参照。

(40) たとえば Horgan and Tienson (1996), p.5 を参照。

(41) Horgan and Tienson (1996), p.105 を参照。

(42) この例は Horgan and Tienson (1996), pp.96-97 に挙げられている事例に若干

の脚色 を施したものである。

(43) この段落の論点に関しては、Horgan and Tienson(1996), pp.96-97; p.99 などを参照。

(44) 本節 頁、および、そこに付された註(12)を参照。われわれの認知過程がこのようにコンピュータ・プログラム化されうるものではないとしても、しかしそれは物理的システムによって実行可能なものではあるはずである。ホーガンとティーンソンによれば、その実行過程の適切な数学的表現は「動的システム dynamical system」である。動的システムとは、ごく簡単に言えば、高次元ベクトル空間のうちに描かれた一つの地形図である(ここでは、認知システムの全体的認知状態のおのおのがこの空間の一点として表される)。しかし、動的システムについては、紙幅の関係上、本節では立ち入って考察することはできない。Horgan and Tienson(1996), Chs.4 & 9 を参照。

(45) たとえば Horgan and Tienson(1996), p.99; p.153 などを参照。

(46) この問題に関する、本書所収柏端論文 2.における議論は、本項の議論と軌を一にするものであり、大いに有益で参考になる。

(47) この項の論述に関しては、Horgan and Tienson(1996), pp.146-153 を参照。また、本節 頁も参照。

(48) つまり、認知システムが仮にかくかくの全体的認知状態になったとしたら、次には しかじかの全体的認知状態に推移するだろう、という形の性質。注意すべきことは、認知システムが現実に前者の状態になる必要はなく、ただそうなりうるだけでよい、ということである。

(49) 同様に、ネットワークが仮にかくかくの全体的物理状態になったとしたら、次には しかじかの全体的物理状態に推移するだろう、という形の性質。やはり、ネットワークは前者の状態になりうる、というだけでよい。

(50) より正確には、前者と、後者のある部分構造とが同型対応する、という関係になる だろう。

(51) 本節 頁を参照。

(52) 本節 頁で示したフォードのディレンマを参照されたい。

(53) このアイデアの正味の内容は、すでに述べたように、数学的に一つの動的システム として表現されることになる。

(54) このような新たな非古典的認知描像が注目されるべき背景にはもちろん、すでに指摘したように、古典的計算主義がフレーム問題という一つの行きづまりに直面している(ように見える)という事情もある。そして、これもすでに示唆してきたように、コネクショニズムはこの新しい認知描像と結びつくことによって、フレーム問題を解決ないし解消するための道筋を開くことができるように思われる。たとえば、メモリーに蓄えられている諸表象を新たな情報に照らして適切に更新するための実行的で一般的な方法はいかなるものか、というフレーム問題の一形態に対しては次のような解答の方向を示せるだろう(cf. e.g., Horgan and Tienson(1996), pp.166-167)。もしコネクショニスト・システムを真正の認知システムと認めることができるのなら、ここでは、すでに獲得された表象がすべて明示的にメモリーに蓄えられる必要はない。つまり、そもそも変更すべき表象は何もないことも可能である。それらは必要に応じて再活性化されればよい。さらに、新たに得られた情報はユニット結合の重みづけの変更という形で実現されることも可能である。その場合、重みづけの部分的変更はシステム全体の活性化地形の変更をもたらさうだろう。つまり、既存のすべての表象は、再活性化のさい、この活性化地形全体の変更にともなって自動的に適切な変形を被ることが可能であろう。

## 参考文献

- Bechtel, W. and Abrahamsen, A. (2002), *Connectionism and the Mind: Parallel Processing, Dynamics, and Evolution in Networks*, 2nd edition, Blackwell.
- Copeland, J. (1993), *Artificial Intelligence: A Philosophical Introduction*, Blackwell.
- Fodor, J.A. (1975), *The Language of Thought*, Harvard University Press.
- Fodor, J.A. and McLaughlin, B.P. (1990), "Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work", *Cognition* 35, pp.183-204  
Reprinted in Macdonald and Macdonald (1995), pp.199-222.
- 服部裕幸 (2000), 「コネクショニズムとは何か?」 『科学哲学』 33 巻 2 号 日本科学哲学学会 pp.15-28.
- Horgan, T. and Tienson, J., (eds.) (1991), *Connectionism and the Philosophy of Mind*, Kluwer.
- Horgan, T. and Tienson, J. (1996), *Connectionism and the Philosophy of Psychology* MIT Press.
- Macdonald, C. and Macdonald, G. (1995), *Connectionism: Debates on Psychological Explanation Vol.2*, Blackwell.
- Marr, D. (1982), *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman and Co. (邦訳 乾敏郎、安藤広志訳 『ビジョン：視覚の計算理論と脳内表現』 産業図書 1987年)
- 信原幸弘 (2000a), 「コネクショニズムと消去主義」 『科学哲学』 33 巻 2 号 日本科学哲学学会 pp.1-14.
- 信原幸弘 (2000b), 『考える脳・考えない脳：心と知識の哲学』 講談社現代新書
- Pollack, J. (1990), "Recursive Distributed Representations", *Artificial Intelligence* 46, pp.77-105.
- Smolensky, P. (1988), "The Constituent Structure of Connectionist Mental States: A Reply to Fodor and Pylyshyn", *Southern Journal of Philosophy* 26, Supplement: Spindel Conference 1987: Connectionism and the Philosophy of Mind, pp.137-161. Reprinted in Horgan and Tienson (1991), pp.281-308.
- Smolensky, P. (1991), "Connectionism, Constituency, and the Language of Thought", in Loewer, B. and Rey, G. (eds.), *Meaning and Mind: Fodor and His Critics*, Blackwell. Reprinted in Macdonald and Macdonald (1995), pp.164-198.