

# 心は(どんな)コンピュータなのか 古典的計算主義 vs. コネクショニズム

戸田山和久(名古屋大学)

『シリーズ心の哲学 ロボット篇』所収  
信原幸弘編、勁草書房(2004, 7 出版予定)

## 一 計算主義とは何か、それは何を問題にしているのか

本章の主題は「計算主義(computationalism)」だ。計算主義の主張を一言で言えば、「心は計算機だ」ということになる。しかし、これはひどく曖昧で、ほとんど無内容すれすれではないだろうか。というのも、計算機って何なの?ということがどうしようもなく曖昧だからだ[1]。したがって、本当に意味のある問いは、「心は計算機なのか」ではなくて、「心が計算機であるとしたら、それはどんな計算機なのか」である。

計算主義は、心の入出力ではなく、内部メカニズムに目を向けさせる。歴史的に見れば、これが計算主義のテーゼが果たした最も重要な役割だった。二〇世紀前半には「科学的」心理学と言えは行動主義を意味した。内観に頼っていた一九世紀心理学に対抗するために、行動主義は心の中をブラックボックスと見なして研究対象から除外し、外部から観察可能な刺激と行動だけに視野を限定した。こうした行動主義の圧制に対して、計算主義は、心という計算機がどのように情報処理をしているかを研究しよう、心の中を研究しても立派な科学研究なんだ、と提案する。これが認知革命(cognitive revolution)と言われる動きだ。人工知能研究も認知科学もこの革命が生み出した。心は計算機だというテーゼは、この革命の核心をなすスローガンとなったのである[2]。

(1) 認知関数と表象: これからの議論に必要ないくつかの概念を導入しておこう。言うまでもなく、すべての計算機が心つまり認知システムであるわけではない。認知システムを特徴づけるのは、「何を計算しているか」という点だ。しかし、この表現は要注意である。算盤でかけ算をしている場合を考えてみよう。このとき、算盤プラス人間という計算機が計算しているのは、かけ算という関数である。そして、計算プロセスが進行するときに操作されているものは算盤の玉だ。このように計算については、どのような関数が計算されているか、何が操作されて計算過程が進んでいるか、の二つの観点が重要であることがわかる。このどちらもが「計算機が計算しているもの」と言いうる。認知システムが計算する関数は、認知関数(cognitive function)と呼ばれる。認知過程は、こうした認知関数の計算として捉え直される。そして認知システムは、この認知関数を表象(representation)を操作することによって計算する点で他の計算システムから区別される。

(2) アーキテクチャ[3]: Java, Lisp, C++といった高級プログラミング言語を例にとろう。こうした言語は異なるアーキテクチャをもつと言われる。同じ関数を計算する場合でも、これらの言語は、統語論、許される基本的操作(かけ算が基本的な関数として入っているかどうかといったこと)、操作される対象、メモリなどのリソースの使い方がそれぞれ異なっているために、異なるアーキテクチャだ。特に、基本的操作・関数の定義域と値域が表象であるようなアーキテクチャを認知アーキテクチャと言う(Fodor & Pylyshyn, 1988: 10)。

どのように関数が計算されるのかは、システムのアーキテクチャに依存する。そうすると、計算主義者の問うべき問題は次のようなものになる。人間の心が計算機だとしよう。そのとき、心のもつ認知アーキテクチャはどのようなものなのか? [4]。これに答えるためには、計算過程で操作されている表象とはいかなるものか、そして表象がどのように操作されて計算が進行するのか

を明らかにしなければならない。いまのところ、この問いにどう答えるかについてのコンセンサスがあるわけではない。しかし、有力な研究プログラムが二つ存在し、それぞれ、古典的計算主義(記号主義)、コネクショニズムと呼ばれている。

さてそこで、本章の目的を次の点に置こう。(1)古典的計算主義とコネクショニズムがそれぞれどのような特質をもった認知研究の枠組みであるのかを明らかにする。(2)古典的計算主義による批判からコネクショニズムをできるかぎり擁護する。(3)その擁護のための議論を通じて、両者の違いがどこにあるのかをさらに明確にするとともに、この種の対立に決着をつけるためには何をしなければならないかを明らかにする。

## 二 古典的計算主義の認知モデル

まずは古典的計算主義が認知をどのようにモデル化してきたかを整理しておこう。今後、本章では古典的計算主義をたんに「古典主義」と呼ぶことにする。

(1)心は形式システムである:心は形式システム(formal system)として理解される。形式システムとは、離散的なアイテムをその形式的な性質に言及した規則に従って操作することで計算を進めるシステムのことだ。たとえば形式論理学の体系は形式システムの一例と言える。そこでは、離散的な記号からなる論理式を、その形に言及した規則、たとえば「A B なる形をした式からは A を取り出してよい」といった推論規則にしたがって操作することで計算が行われる。チューリングマシンも形式システムだし、フォン・ノイマン型コンピュータ(市販されているほとんどすべてのコンピュータ)も、チューリングマシンとは異なるアーキテクチャをもつ形式システムだ。また、さまざまなプログラミング言語もやはり、それぞれアーキテクチャを異にする形式システムを定めている。古典主義では、心もこうした形式システムの一つだと考える[5]。

心のアーキテクチャがどのようなものであるかははっきりしない。それがフォン・ノイマン型だと考える人はいまではほとんどいないだろう。とは言え、どちらも形式システムである。この点で心のアーキテクチャをフォン・ノイマン型コンピュータで実装することは少なくとも原理的にはできそうに思われる。こうして人工知能のプロジェクトに正当化が与えられる。

(2)統語論的構造を持つ表象:古典主義は、心という形式システムが操作している離散的なアイテム=表象について、さらに次のような2つの要件を課す[6]。

[S1:表象の合成性(compositionality)](i)心的計算の対象になる表象は統語論(syntax)を持つ。つまり、複合的な表象は、原子的な表象から統語論的な規則にしたがってできあがっている(統語論的合成性)。(ii)そして、複合的表象の内容は、構成要素となっている原子的表象の内容とその複合的表象のもつ統語論的構造とが決まれば、ただ1つに定まる。つまり、表象の意味論も、その統語論的合成性に依存する形で合成性をもつ(意味論的合成性)。

[S2:表象処理の統語論感受性(syntactical sensitivity)]心的計算過程における表象への操作がどのようなものになるかは、その表象がもつ統語論的構造に依存して決まる。

(3)思考の言語仮説:これらの要件を満たす典型例は言語に他ならない。そこで、古典主義は次のような基本的仮定を置く(Fodor, 1975)。心の中には言語のような表象がある。これをメンタル語(Mentalese)あるいは思考の言語(language of thought)と呼ぼう。あらかじめ明示的に与えられたアルゴリズムに従って、思考の言語で書かれた文をその統語論的構造に応じて変形する。この操作によって、心的計算(思考などの認知過程)が進んでいく。

(4)統語論的エンジンとしての心:表象は統語論と意味論をもつ。S2 は、にもかかわらず、心的計算にかかわりをもつのは統語論だけだと述べている。言いかえれば、表象のもつ統語論的・形式的性質(それがどういう「形」をしているか)と、意味論的性質(それが何を意味しているか)のうち、前者だけが計算過程において因果的効力をもつということだ。心は表象の統語論的性質によって駆動するエンジンなのである。

### 三 認知研究の3つのレベル

計算主義はシステムがどのような認知関数を計算しているかだけに関心をもつのではない。むしろ主要な関心は、どのようなアーキテクチャによってその関数を計算しているかにある。(Marr, 1982: 24-25)は認知過程を記述する際の3つのレベルを次のように区別している。

- (i) 計算論的レベル(computational level) どのような認知関数が計算されているか
- (ii) アルゴリズムレベル(algorithmic level) その認知関数がどのように計算されているか
- (iii) 実装レベル(implementational level) 上記のアルゴリズムの各ステップをなす操作が、より下位のメカニズムによってどのように実行されるか

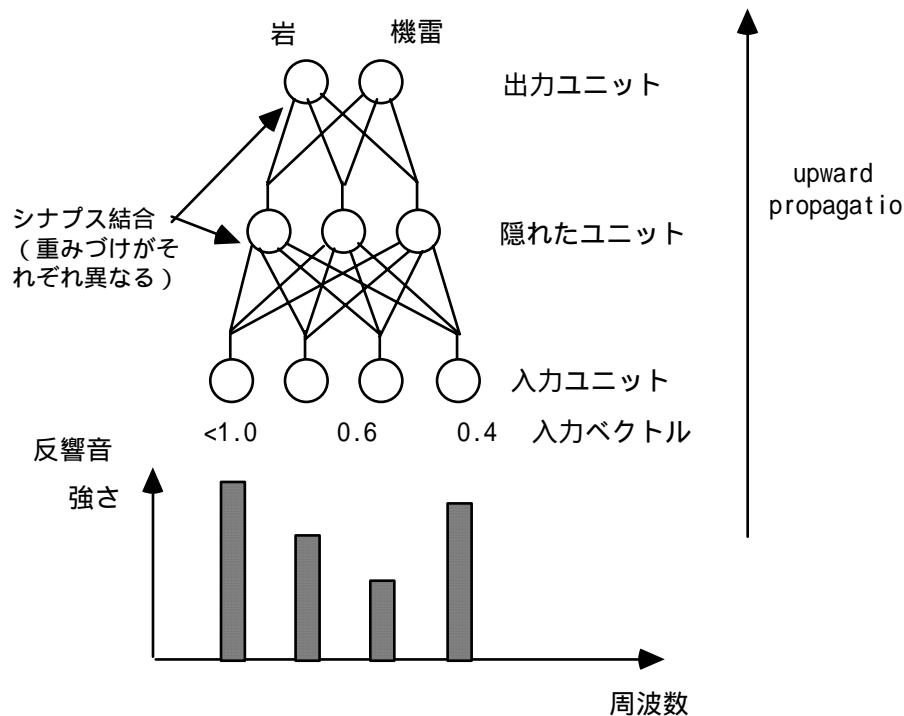
いくつか注意すべきことがある。まず第一に、これら3つのレベルはお互いに他のレベルを制約するけれども、完全に決定はしない。ある関数を計算できるアルゴリズムは何通りも考えられ、計算論的レベルからの制約ではアルゴリズムは一つに決まらない。また、提案されたアルゴリズムがとても下位のメカニズムによっては実装できないようなものであるなら、その提案の信憑性は疑わしくなる。このように、各レベルは比較的独立に研究できるのだけれど、その際には他のレベルからの制約が手がかりになるという関係がある。第二に、この階層は相対的である。というのは、あるアルゴリズムの各ステップを実装しているメカニズムじたいが、なんらかの関数の計算であり、その計算にも別のアルゴリズムがあって...ということがいくらでもありうるからだ。したがって第三に、ここで言う「実装」を物理的「実現(realization)」と混同してはいけない。そもそも実装という語は、PASCAL で書かれたプログラムが、アセンブラ言語で実装され、アセンブラ言語が物理的コンピュータの機械語で実装される、というような使い方をされる。決して、アルゴリズムとその物理的実現の関係に限定されたものではなく、異なるレベルのアーキテクチャ間(つまり抽象的マシン間)の関係一般に用いられる概念である。本章では、「実装」と「実現」とを区別して使っていく。

アルゴリズムレベルの探求は、上下のレベルからの制約を受ける。しかし、古典主義者は上からの制約をより重視するようだ。彼らの研究は次のように進む。人間はしかじかの認知関数を計算している。その関数を計算するにはこんなアルゴリズムが適切なはずだ(こうして提案されるモデルはしばしば「規範的モデル」と言われる)。しかし実験などを通して知られる現実の人間の行動は、規範的モデルからの逸脱やバイアスを含んでいる。それを手がかりに、はじめ提案したアルゴリズムを修正し、より現実的なモデルに近づけていく...

一方、最終的に心のアーキテクチャは脳において実現されているという事実も無視できない。だから、神経回路の働きという下からの制約をより重視してアーキテクチャレベルでの研究をすすめることも有意義な方向だろう。コネクショニズムはこうした発想に基づく認知研究のプログラムである。

### 四 コネクショニズムの認知モデル[7]

(1) コネクショニスト・ネットワーク(以下 CN)の構造:いきなり典型例を見てしまおう(図1)。CNは多数のユニット(円で示す)が互いに結合された構造をもっている。ユニットはニューロンに、ユニット間の結合はシナプスに相当する。一見して、CNのアーキテクチャはおなじみのパソコンのそれとは異なることがわかる。まず CPU が存在しないし、メモリも、メモリに内蔵されたプログラムもない。



ユニット間の結合のそれぞれには、ある特定の重み  $w_i$  (正負の実数値) が割り当てられている。各ユニットは、まず、自分に結合している他のユニットのそれぞれから入力された信号の強さ  $x_i$  に、各結合に割り当てられた重み  $w_i$  を乗じたものの総和をとる(この値をそのユニットの活性化値と言う)。次に、活性化値に応じた強さの信号を他のユニットに向けて出力する[8]。

このようなユニットを多数結合させたネットワークとして認知システムのアーキテクチャを考えようというのがコネクショニズムだ。もちろんユニットをどのように結合するかについては選択の余地がある。図1は階層型ネットワークと呼ばれる。他方、全結合型ネットワークと呼ばれるものは、各ユニットが他のすべてのユニットと結合をもつ。

(2) CN は何をするか: 次に、CN 全体の振舞いを見よう。図1は、ソナーからの反射音を用いて岩と機雷を識別するシステムだ(もちろん単純化してある)。最下層にある4つのユニット(入力ユニット)は、それぞれ4つの周波数帯(400 450 ヘルツ、450 500 ヘルツ、...)での反射音の強さに対応させてある。各入力ユニットは自分が担当する周波数帯の音の強さに応じて異なる度合いで活性化される。4つの入力ユニットがそれぞれどれくらいの度合いで活性化されているかは、例えば  $\langle 0.4, 0.3, 0.9, 0.7 \rangle$  のような4元ベクトルで表現できる。これを入力ベクトルと言う。入力ベクトルは CN の外部にある情報をコード化したものに相当する。

中間層は「隠れ層(hidden layer)」と呼ばれる。入力ユニットの活性化値は、それぞれの入力ユニットから出力される信号を介して隠れ層へと伝播してゆく。隠れ層にあるそれぞれのユニットは、入力層の4つのユニットすべてから信号を受け取り、4つの信号のそれぞれの強さに重みを乗じたものの総和に対応する度合いで活性化される。こうして、隠れ層にある3つのユニット全体の活性化の様子についても、同様に活性化ベクトル(3元ベクトル)が考えられる。これを「隠れ層の活性化ベクトル」と言う。

次に、隠れ層と最上層(出力層)の間でも同様の伝播が起こる[9]。この CN は、入力層から出力層に向けて一方向に信号が伝達されるため、フィードフォワード・ネットワークと言われる。ようするに、CN は全体として、入力ベクトルを出力ベクトルに変換するベクトル変換装置である。

(3) 学習と訓練: この CN に計算させたいのは、反射音をコード化したベクトルを入力すると、

岩か機雷かについての判定結果を出力するような認知関数だ。もちろんこれは適当に結合重みを割り当てていたのでは実現できない。そこで必要になるのが学習(訓練)である。CNの学習とはようするに重みづけの調節だ。各結合の重みをうまく調節して、岩の反射音を入力すると出力ベクトルが $\langle 1, 0 \rangle$ になり、機雷の反射音を入力すると出力ベクトルが $\langle 0, 1 \rangle$ になるようにすること。これが次の目標だ。

CNの学習には教師ありと教師なしの二種類がある。教師あり学習では、いくつかの入力データとそれに対する意図された出力値(教師データ)との組(これを訓練セットと呼ぶ)を用いて、出力が教師データに近づくように重みを調節する。教師なし学習では、そうした教師データを使わずに学習を行う。これは、正解があらかじめわからない場合、たとえば入力データ間にある隠れた相関を発見させたいようなときに有効だ。

それぞれの学習に応じて、重みを自動調節するための規則が提案されている。これを学習則と呼ぶ。たとえば、教師なし学習ではヘップ則やコホネンのアルゴリズムが有名だ。コネクショニズムの先駆的研究は古典主義と同じくらい古いが[10]、ローゼンブラットの2層ネットワーク(パーセプトロン)では排他的選言を計算できないことを(Minsky & Pappert, 1969)が証明して以来、コネクショニズム研究は一時期下火になってしまった。それが復活し、今日のように盛んに研究されるようになったきっかけは、ラメルハートが誤差逆伝播法(backpropagation)という教師あり学習則を発表すると同時に、3層以上のネットワークでは排他的選言を計算できることを示したことだった(1986年)。

(4)CNが行う処理の特質:学習を終えたCNはいったいどれほどのことができるようになっていのだろうか。そして、CNの能力と処理の方法にはどのような特質があるのだろうか。

(i)優れた一般化能力:多くの実験結果は、ネットワークが訓練セットにない新奇な入力についても正しい出力を行う能力を獲得することを示している。印象的な実例は、写真を入力として与え、顔であるかどうか、性別、誰の顔であるかを判断して出力するCNを用いた実験である(柴田、二〇〇一:第5章参照)。学習後にまったく新しい事物や人物の写真を入力しても、顔かどうかに関しては100%、性別に関しては81%の正解率を示した。

こうした結果が示唆するのは、CNが入力と出力を一対一に対応させた表を検索するような仕方(table lookupと言われる)で処理を行っているのではないらしいということだ。このCNは、顔とは何か、また男女の顔の違いとは何かについての一般的知識を獲得して、それを新しい事例に一般化していると言いたくなる。このことを指して、CNは高い一般化能力をもつと言う。

(ii)劣化した入力に対する耐性:先の実験ではさらに、訓練セットにあった写真の顔をそれぞれ5分の1だけ隠し、それを正しく見分けられるかどうかを調べたところ、正解率にほとんど変化がないことが確かめられた。一般にうまく訓練されたCNは、ノイズや欠損によって劣化した入力に対しても出力まで辿り着き、しかも劣化に柔軟に対応して正解に到達できることがわかっている。

(iii)全体論的な処理:顔認識ネットワークにおいては、どの写真の処理にも隠れ層のユニットすべてが関与し、しかも、このユニットは輪郭担当、このユニットは髭の有無担当...といった具合に、個々のユニットが顔の要素的特徴の識別を専門に行っているのではない。処理はネットワーク全体に高度に分散していて、特定の作業を専門に担当するモジュールはない。

(iv)非アルゴリズム的な処理:CNに明示的なアルゴリズムがあるとしても、シナプスから入ってくる信号の強さと結合重みを掛け合わせて、それを足し合わせて...という個々のユニットが従う規則しかない。顔認識や岩・機雷識別という処理(=認知関数の計算)全体を支配する明示的なアルゴリズムは存在しない。

(v)CNの得意技:CNは顔認識のようなパターン認識に非常に優れている。その他、多数の(必ず満たされる必要がなく、いくつかを満たせばよいという意味で)柔らかな制約を満足する解を見いだすような問題[11](best-match problem)や、巡回セールスマン問題のように独立に

解くことのできるサブ問題に分割できないような問題 (connected problem) を近似的に解くこと、あるいは頭を動かしても眼が対象を追いつけるようにするという運動制御に秀でている。逆にこれらの問題の解決は、古典主義のアプローチでは非常に難しい。

(5) 分散表象: 岩・機雷識別システムを転用して、人々の恋愛模様を記憶させてみよう。簡単のために、入力ユニットの活性化値は0か1のいずれかとする。ロス、チャンドラー、レイチェル、モニカの4人をそれぞれ、00、01、10、11とコード化する。そして「ロスはレイチェルが好きだ」という文を、 $\langle 0,0,1,0 \rangle$ という入力ベクトル、「チャンドラーはチャンドラー(じしん)が好きだ」を $\langle 0,1,0,1 \rangle$ で...という具合にコード化する。こうすると、16通りの文を入力することができる。この文の中には真な文と偽な文がある。そこで、真偽の判定を出力させることにする。入力した文が真のときは $\langle 1,0 \rangle$ が出力され、偽のときは $\langle 0,1 \rangle$ が出力されるように教師あり学習を行うと、このCNは16個のどの文を入力しても、その真偽を答えてくれるようになる。このシステムは、誰が誰を好きかについての情報を蓄えていると言える。

古典主義は表象を思考の言語で書かれた記号列と考える。たとえば「ロスはレイチェルが好きだ」という思考は、ロス、レイチェル、愛しているという関係という3つの原子的表象がある形式で結合したものと捉えられる。コネクショニズムではこれとは対照的に、表象は隠れ層の活性化ベクトルとして存在する。3つの隠れユニット上で、「ロスはレイチェルが好きだ」はベクトル $\langle 0.5, 0.7, 0.3 \rangle$ により表象され、「チャンドラーはチャンドラーが好きだ」は $\langle 0.3, 0.2, 0.9 \rangle$ により表象されるという具合だ。こうした独特な表象のあり方が分散表象と呼ばれる。この場合、どの文も、利用可能な3つのユニットのすべてを用いて表象されている。そして、異なる文を表象するために同じユニット群が用いられる。したがって、特定の文を表象するための専用のユニットといったものはない。また、どのユニットがロスを表象しているのかという問いも意味をなさない。分散表象はどこからどこまでのユニットがロスの表象で、どこからどこまでがレイチェルの表象を担っているか、といったことに意味がないような仕方、3つの隠れユニットにまたがっている。

## 五 古典主義者 vs. コネクショニスト

### 五-一 古典主義者がコネクショニズムを批判する～認知の体系性による議論

コネクショニズムも古典主義も、認知を表象に対する計算として捉えているという限りで、表象主義の一種だと言える。両者が食い違うのは、表象とその処理をどのようなものと考えているからだ。すでに見たように、思考の言語と分散表象は、同じ表象と呼ばれるにしてもずいぶん異なる。古典主義者の眼には、分散表象はどうしてもナサケナイ表象で、とても認知の説明に役立ちそうにないものと映るらしい(服部、二〇〇〇、二〇〇三)。こうして古典主義者は、分散表象に依拠するコネクショニズムは認知理論を与えることはできない、と論じることになる。

ジェリー・フォードらは、人間の認知について成り立っているいくつかの事実を根拠にして、コネクショニズムは古典主義と正面から対立するようなライバルではありえず、認知理論を提供できるのは古典主義しかありえないという議論を展開した[12]。その事実とは、認知能力の生産性と体系性と呼ばれてきたものである。それを(Aydede, 1997: 78)に従って整理し直して提示しよう。なぜこのような再整理が必要なのは後になって明らかになる。

(RI) 形成における規則性 (formative regularities) 表象を生み出して保持する場面で見られる規則性

(I-i) 生産性 (productivity) 人間はいくらでもたくさんの表象を産出し心に抱くことができる。たとえば、Ross believes that Joey loves Monica と思うことのできる人は、Chandler believes that Ross believes that Joey loves Monica、Rachel believes that Chandler believes that Ross believes that Joey loves Monica...も思うことができる。

(I-ii) 形成における体系性 (formative systematicity) 人間の認知状態は、ある状態になることができるためには、その状態と関連しあったそれ以外の多くの状態にもなれるのでなくてはならない、といった性質をもつ。たとえば、人が Joey loves Monica と思うことができるのは、その人が Monica loves Joey と思うこともできる場合に限られる。

(RII) 変形における規則性 (transformational regularities) 表象を変形し処理する場面で見られる規則性

(II-i) 推論における体系性 (inferential systematicity) [13] 「P かつ Q かつ R」から「P」を推論できる人は、「P かつ Q」から「P」を推論することもできる。どちらか一方の推論だけできる人、というのはヘンだ。このように、人間の認知は、ある処理ができるためには、関連しあったそれ以外の多くの処理もできるのでなくてはならない、といった性質をもつ。

これらはすべて、人間の認知に見られる特異な現象だ。まともな認知理論であれば、この現象を説明する責任がある。とりあえず、この現象を一括して「認知の規則性」と呼ぼう。

さて、ここからフォーダーらは次のように議論をすすめる。

【議論 1】思考の言語を仮定すれば、認知の規則性を説明することができる。したがって、古典主義は認知の規則性を説明できる。

【議論 2】これに対し、コネクショニズムは、規則性の説明に不可欠な道具立て、つまり統語論的構造に従って処理される表象をもたないから、それを説明することができそうにない。規則性を説明できないならば、コネクショニズムは認知理論の名に値しない。

【議論 3】コネクショニズムがどうか規則性を説明できたとしても、その際に統語論的構造をもち、その構造に従って処理される表象を用いてしまったならば、それは古典主義的アーキテクチャを実装する一つのやり方を与えたにすぎないから、コネクショニズムはやはり認知理論の名に値しない。

以上の議論は、本来は認知の規則性全般から出発して行うべき議論だと思うのだけれど、体系性だけがクローズアップされて論争が進んでしまったので「体系性からの議論」と呼ばれている。議論2と議論3はあわせてディレンマを形成している。ここから引き出される結論は、いずれにせよコネクショニズムは認知理論にならない、というものだ。こうしてコネクショニズムを斥け、規則性を説明できる古典主義が晴れて王座に坐ることになる。しかし、そんなにうまくいくのだろうか。

## 五・二 コネクショニストからどんな反論ができるか

フォーダーらの批判をどのようにして斥けたらよいだろう。一つのやり方は、そもそも前提になっている、人間の認知は規則性を示しているという「事実」を拒否することだ[14]。そうすれば説明すべき現象じたいが消えてなくなってしまうわけだから、議論はスタートできない。確かに、人間の認知はそんなに生産性があり、体系的だろうかと疑ってみることはできる。いくらでも深い回帰節をもった文を理解できるということは明らかにウソであるし、他のかけ算はできるのになぜか  $4 \times 7$  だけはできないという人がいたっておかしくない。さらに、発達心理学的な研究から子どもの認知では体系性が見られない場合があることがわかっている。なのにフォーダーらは、認知の規則性はたんに人間という生物において偶然に見られる事実ではなく自然法則だ、つまり認知とよべるような能力なら本質的・必然的に規則性を示さなくてはならないと考えている。これは明らかにいきすぎだろう。フォーダーらは、日常言語に見られる体系性や生産性を心的処理へと読み込んでいのではないかと疑われても仕方がない(美濃、二〇〇〇: 五一頁)。

とは言うものの、少なくとも大人の通常の認知では、ある程度の規則性が広範囲にわたって見られることも事実としては認めなければならないだろう。認知理論を目指すなら、こうしたある程度の規則性が事実として存在することを説明できるにこしたことはない。以下では、フォーダーらの挑戦を門前払いするのではなく正面から受けとめた場合、コネクショニストに何ができるかを見てみよう。

認知の規則性を示すような CN を組み立てる、あるいは既存の CN からそうしたものを探し出すことができれば、少なくとも議論 2 への有力な反例になるだろう。これはこれでやりがいのある課題だ。というのも、コネクショニズムはパターン認識には強いけれど、いかにも思考の言語が関与していそうな論理的推論や言語処理には弱い研究枠組みだったからだ[15]。

### 五・三 古典主義モデルの実装としての CN

最も安直なやり方は、古典主義者が与えるアルゴリズムと表象をトップダウンに CN に実装してしまうことだ。このような研究にもそれなりの意義がある。その結果えられた CN は、劣化した入力への耐性などの望ましい性質を保持していると期待できるからだ。しかし、この方針は議論 3 の餌食になるように思われる。つまり、この場合、コネクショニズムは古典的アーキテクチャを実装する一つのやり方にすぎず、それゆえ認知理論の名に値しないとされてしまう。

この議論には飛躍がある。古典的アーキテクチャの実装であることからただちに認知理論ではないということが帰結するわけではない。帰結するように思ってしまう背景には、「3レベル主義」とでも呼ぶべき誤解がある[16]。それは、認知関数 / アルゴリズム / 実装という「認知研究の3つのレベル」を、階層は3つありそれですべてだと固定的に捉えてしまうことに由来する。そうすると、最下層の「実装」は「物理的実現」と混同される。この場合、もしコネクショニズムがアルゴリズムレベルにある古典主義の実装にすぎないということになると、コネクショニズムは、古典主義的アーキテクチャがどのようにして脳によって物理的に実現されているかについての仮説という位置づけを自動的に与えられてしまう。

しかし、すでに指摘したように、「3つのレベル」は相対的な位置関係を示すものとして捉えるべきだ。したがって認知研究は、最上階に認知関数、最下階に物理的実現のレベルがあり、中間にそれぞれ上位のレベルを「実装」という関係で結ばれたいくつもの階層をもつものとして考えることができる。この中間の階層はすべて情報処理モデルだと言ってよい。そして、所与の認知現象を説明するためにどの階層が本質的関連性をもつかということは、アプリアリには決まらず、あとから経験的に見いだされるべきことがらである。かりに、古典主義モデルの実装であるとしても、そのことだけではコネクショニズムから認知モデルの地位を奪うことはできない[17]。

### 五・四 RAAM とチャルマーズの実験

とは言うものの、古典主義モデルの実装を与えるにすぎないなら、コネクショニズムが認知の規則性に古典主義者と別種の説明を与えることは期待できず、彼らの説明は古典主義的説明に寄生したものになってしまう。そこで、もう少し独自性を追求してみよう。コネクショニストも古典主義のライバルになるような説明を与えられることを示し、議論 2 を正面突破する作戦だ。そのためには、もう少し本来のコネクショニズムらしいネットワークの構築が必要になる。(van Gelder, 1990) が挙げているいくつかの候補をから、ポラックの RAAM (Recursive Auto-Associative Memory) を紹介しよう[18]。RAAM はその名の通り、一種のメモリの働きをする3層ネットワークである。その目的は、ある一定数のユニットからなる隠れ層の中に、(LOVES JOEY MONICA)、(BELIEVES ROSS (LOVES JOEY MONICA))、(BELIEVES CHANDLER (BELIEVES ROSS (LOVES JOEY MONICA)))...といった回帰節をもった文を分散表象として記憶させておき、出力層にもとの文を再現させようというものだ。入力層は 48 個のユニットからなるが、これらはそれぞれ 16 個のユニットからなるグループ(A、B、C としよう)に分かれている。



出力層も入力層とまったく同じつくりになっている。隠れ層は 16 個のユニットからなり、これらのユニットはそれぞれ入力・出力層のすべてのユニットに結合している。

文の入力 コード化 隠れ層での記憶 解読 もとの文の出力、という一連の過程は次のようにして行われる。まずは、(LOVES JOEY MONICA)を入力したとしよう。LOVES をグループ A に、JOEY を B に、MONICA を C に入力する。それぞれの単語にはあらかじめ 16 個のユニットのどれを活性化させ(活性化値は 1)、どれをさせないかの異なる組み合わせを割り振っておく。たとえば、LOVES は 16 個のユニットの 8 番目と 9 番目を活性化させることによって表す、といった具合だ。こうして、グループ A、B、C に属するユニットが、それぞれのグループに入力された語に応じた仕方で活性化する。この信号は、隠れ層に伝播する。隠れ層では 16 個のユニットがそれぞれ 0 と 1 の間の何らかの値で活性化する。こうして、隠れ層の中に最初の文が圧縮された形で分散的に表象されることになる。どのユニットが主語担当、どのユニットが動詞担当ということはなく、JOEY も MONICA も 16 個のユニット全体に分散して保持されている。あとは、この逆の過程を辿って、出力層の A グループに LOVES、B に JOEY、C に MONICA が、それぞれ 16 元活性化ベクトルとしてもとどおり出力される(学習がうまくいってればの話)。

さて、問題は回帰節をもった文だ。こうした複合文に対してはユニット数もグループ数も足りないように思われる。これをクリアするための工夫が再帰的構造を付加することだった。隠れ層の 16 個のユニットからそれぞれ入力層 C グループの 16 個のユニットに 1対1 に結合を張っておく。こうすると、次のようなことができるようになる。まず、第一段階では先ほどと同様に、LOVES を A、JOEY を B、MONICA を C に入力する。そうすると隠れ層には(LOVES JOEY MONICA)の圧縮表象が生じる。この隠れ層の活性化ベクトルを、再帰的な結合を使って、グループ C の入力ユニットに「コピー」してやる。そうすると次の段階で、A に BELIEVES を、B に ROSS を、そして C に(LOVES JOEY MONICA)の圧縮表象を入力することができる。そうすると、隠れ層に(BELIEVES ROSS (LOVES JOEY MONICA))の圧縮表象を記憶させることができたことになるだろう。解読は、この逆を行えばよい。出力層の A に BELIEVES、B に ROSS、C に(LOVES JOEY MONICA)の圧縮表象が出力されてくるから、C の活性化ベクトルをもういちど隠れ層に戻してやれば、出力層に LOVES、JOEY、MONICA がリカバーできる。

ポラックは 13 個の文からなる訓練セットで RAAM を学習させ、その結果、訓練セットに属さない 12 個の文についても、その複合構造を記憶し正しく再現することができた。もちろん限界はある。一つには、出力された活性化パターンが単語に相当するものか、それとも何らかの文の圧縮表象なのかを判定する手続きが必要になること。もう一つは、4層まで入り組んだ文は何とか処理できるが、それ以上の「深い」回帰節をもった文は処理できなくなってしまうということである。

RAAM が認知の規則性をそれぞれどの程度示すことに成功しているかを見てみよう。まず、生産性 (I-i) についてはどうか。訓練セットに含まれない文も正しく再現されたということは、RAAM がかなりの一般化能力を持っていることを示すと同時に、隠れ層の圧縮表象は何らかの仕方で、入力文の統語論的構造についての情報を保持していることを示している。このことは、訓練セットにはなかった新しい構文の文も正しく記憶できたという事実にも現れている。しかし、4層以上の回帰節をもった文はうまく記憶できなくなるといことは、RAAM が示す生産性は完全なものではないことを示している。しかし、事実として人間が示している生産性も不完全なものであるのだから、このことは致命的な欠点にはならないだろう。形成における体系性 (I-ii) はどうか。訓練セットの語彙を使って 16 個の単文をつることができる。これらの単文のうち訓練セットにあらかじめ含まれていたのは 4 つだけだった。にもかかわらず、RAAM は残りの 12 個の文もうまく一般化して記憶・再生することができた。やはりある程度の体系性が示されたと言ってよいだろう[19]。

しかし、評価を行えるのはここまでだ。RAAM はメモリである。それは、回帰的な統語論的構造をもった文を圧縮表象の形で保持するだけで、その圧縮表象に対してさらに操作を加えて

認知タスクを実行するようなシステムではない。したがって、変形における規則性(RII)をCNは示しうるかという問いには、そもそもRAAMは無関係だ。さらに、単なるメモリでは古典的なシンボル構造を実装したにすぎないと言われてしまう(Chalmers, 1990: 9)し、フォーダーらの批判に答えたことにはならない[20]。

(2)チャルマーズの実験:(Chalmers, 1990)は、RAAMの圧縮表象をさらに統語論感受性をもつような仕方で処理することができ、処理の場面で見られる規則性を説明できることを示した。思考の言語のような構造を失っている圧縮分散表象を解読の過程ぬきでじかに処理しても、統語論的な変形ができるなら、この種の処理は古典的モデルを超えたものとなるはずだ。チャルマーズが選んだ変形は、能動態から受動態への書き換えである。具体的には、(Ross Love Rachel)のような構造を(Rachel (Is Love nil) (By Ross nil))に書き換える[21]。チャルマーズは変換ネットワーク(transformation network, TN)という3層ネットワーク(入力層は16ユニット、隠れ層は実験に応じて13~29ユニット、出力層16ユニット)をつくり、RAAMに接続した。タスクは次のように進む。まず、(Ross Love Rachel)をRAAMに入力し圧縮表象に変換する。その圧縮表象がそのままTNに入力され、TNでの処理を経て、TNの出力層から16元ベクトルが出力される。これを再びRAAMの隠れ層に直接に入力して解読すると、(Rachel (Is Love nil) (By Ross nil))が得られる。...もちろん、こうなるように学習させるわけだ。訓練セットには、40個の能動文・受動文のペアが使われた。その結果、最初の学習後には65%の正解率だったが、さらに20ペアを訓練セットに追加して学習させ直したところ、訓練セットにない20個のペアについて100%の正解率となり、かなりの一般化能力を発揮した。

この実験によってフォーダーらの議論2が反駁されたと考えてよいだろうか。しかし、実験からどんな帰結を読みとるかについてはもう少し慎重さが必要だ。RAAMプラスTNというネットワークが、規則性を呈している(exhibit)ことは、ちょっとおまけして認めよう。しかし、議論2が主張していたのは、コネクションニズムは認知の規則性を説明(explain)できない、ということだった。この2つは同じことではない。「なんだか知らないがうまくいった」ではなく、コネクションニストには、圧縮分散表象には統語論的構造がないように見えるのに、どのようにしてそれに対する統語論的操作ができたのかについて何らかの説明が求められる。このためには、ライバルの古典主義はどのような根拠に基づいて認知の規則性を説明していたのかをおさらいする必要がある。

#### 五一五 統語論的構造をもつ表象と思考の言語はイコールではない

古典主義は思考の言語を措定する。思考の言語を統語論的構造にしたがって操作することが認知なのであれば、たしかに認知の規則性は簡単に説明がつく。ここでは議論をわかりやすくするために、フレイグ星人という架空の存在を想定してみよう。フレイグ星人の思考の言語は述語論理の言語であり、思考を支配するアルゴリズムは述語論理の形成規則と演繹規則になっている。そうすると、形成における体系性は次のように説明できるように思われる。「きみ(フレイグ星人)がRabという表象をもてるとしよう。ところで心的表象は合成性を持つのだ。つまり、複合的な表象Rabは、原子的な表象Rとaとbから、「大文字の後ろにはしかじかの個数の小文字をつなげてよい」という形成規則に従って組み上げられたものだ。だから同じ形成規則に従ってきみの心は表象Rbaも形成することができてしまう。だから、きみがRabという表象をもてるのだったら、表象Rbaももてることになる。」同様に、生産性も形成規則の再帰性に訴えれば説明がつく。

推論における体系性はどうか。「PかつQかつR」は思考の言語では、(P&(Q&R))という形をした複合表象として表象されている。これは次のような統語論的構造をもっている。つまり全体として( & )で、 のところがさらに( & )という形をしているという構造だ。ところで、表象への操作がどのようなものになるかは、表象がもつ統語論的構造だけで決まる。推論には「( & )を に変形せよ」という規則がある。(P&(Q&R))はいま述べたような統語論的構造をもつ

ているので、きみはこの規則をそれに適用して P を引き出したわけだ。だったら、きみは(P & Q) から P を引き出すこともできるはずだろう？なぜなら、(P & Q)も( & )という形をしているからさ。だから、「P かつ Q かつ R」から P への推論はできるけど、「P かつ Q」から「P」の推論はできないなどということはありません。」

以上の「説明」を分析するとわかることは、規則性の説明が本質的に依拠しているのは、表象が述語論理の言語であるという仮定ではなく、S1 と S2、つまり表象の合成性と処理の統語論感受性だということである。形成における規則性を説明するには S1 で十分だが、変形における規則性も説明しようとするとき S2 に訴える必要が出てくるという違いはあるが[22]。というわけで次が確認できたこととしよう。

(1) S1 は RI の説明に十分であり、S1 + S2 は RII の説明に十分である。

次も言えるかどうかは不透明だ。

(2) S1 は RI の説明に不可欠であり、S1 + S2 は RII の説明に不可欠である。

しかし、いまのところ、認知の規則性を説明すべきだとするなら、S1 と S2 を置くことが最有力な説明方法だということは言えそうである。そうすると、思考の言語仮説と S1・S2 との関係が気になってくる。なぜなら、S1 + S2 が思考の言語仮説を含意するなら、認知の規則性の最良の説明は思考の言語仮説を含意する、ということになるからだ。

ところが、S1、S2 を満たす表象は、フォーダーらが想定している思考の言語に限られない。こんどはゲーデル星人に登場してもらおう。ゲーデル星人はフレイゲ星人と同じくらい論理的推論が得意なのだが、一つ違いがある。彼らの心的表象は論理式ではなく、何とそのゲーデル数なのだ[23]。ゲーデル数とは、特定の形式言語の記号とそれからなる列(原子式、論理式、証明、あるいはまったくでたらめな記号列)のそれぞれについて、一対一に対応づけられた自然数であり、次の条件を満たすものである。(i) 記号列が与えられたときにそのゲーデル数を求める実効的な手続きがある。(ii) 任意の自然数が与えられたときにそれがゲーデル数であるか、そしてどんな記号列のゲーデル数を判定する実効的な手続きがある。したがって、ゲーデル数を見れば、それがどんな統語論的構造をもった論理式に対応するものなのかが一意的にリカバーできる。

このようにして、形式言語内部での統語論的操作(記号列の変形)は、対応するゲーデル数を用いた数値計算に置き換えることができる。フレイゲ星人が(P & Q)を P に変形して推論をしているとき、ゲーデル星人はたとえば 15762546 という数をしかるべき帰納的関数に代入して、8 という答えを出したりしている[24]。外から見ると、どちらも「P かつ Q なんだけど」と言われて「じゃ、P だ」と答えているのだが、心の中で起きているはずいぶん異なっている。

ゲーデル星人の頭の中を覗いてみると、そこには「&」に対応する原子的表象もなければ推論規則もない。あるのは、自然数の表象と推論規則とは似ても似つかない計算規則だ。にもかかわらず、ゲーデル星人はフレイゲ星人とまったく同じだけの推論ができるし、フレイゲ星人と同等の認知の規則性を示している。フォーダーらに言わせれば、これはまったくの奇跡ということになるだろう。しかし、これは奇跡でも何でもない。すでに指摘したように、ゲーデル数は(i) (ii)の性質を満たすことによって、統語論的構造に関する情報をコード化して保持している。そして、ゲーデル数に対する計算もその情報に依存しそれを利用する形で進んでいく。

そこで次のように考えよう。フォーダーらが想定した思考の言語は、合成性と統語論感受性(S1 と S2)を満たす一つの、おそらく最も直観にかなったやり方だが、唯一のやり方ではない。要請 S1 を次のように言いかえてみよう。「複合的表象は統語論的な構造の情報をもっている」[25]。そして S2 は、「心的計算における表象への操作は、その表象がもつ統語論的構造についての情報に依存して決まる」と言いかえる。このように、ちょっとだけ「表象が統語論をもっている」ということの意味をゆるめてやれば、ゲーデル数とそれに対する計算としてとらえた表象と心的処理も、合成性と統語論的感受性もち、それゆえ S1 と S2 を満たしていると考えられる。こうしてゲーデル星人の示す規則性も、S1 と S2 に訴えて同様に説明することが可能になる。

思考の言語とゲーデル数はどちらも S1 と S2 を満たしているが、その仕方はずいぶん違う。複合表象(P&Q)からは P を引き出せるが R は引き出せない、同じように、ゲーデル数 15762546 から 8 は引き出せるが 546 は引き出せない。だから或る意味で、思考の言語とゲーデル数のどちらも、抽象的なレベルでは同じ構成要素と統語論的構造をもっていると言える。しかし、思考の言語では、複合表象(P&Q)のトークンには、統語論的構成要素(、P、&、Q、)のトークンがすべて空間的構成要素として顕在的に含まれている。つまり、複合表象があれば、構成要素もそこにある。そして、複合表象の統語論的構造は、こうした構成要素の配列によってこれも顕在的に示されている。複合表象がこのような合成性をもつ場合、それは連鎖的合成性(concatenative compositionality)をもつと言う。合成性の要請を満たそうとすれば、構成要素をそのまま空間的に並べていけばよいのは道理だ。

ゲーデル数では事情が異なる。15762546 から引き出すことのできる統語論的構成要素 8 は、複合表象の空間的構成要素、つまり部分としては含まれていない。では、それでもなぜ 8 はゲーデル数 15762546 の統語論的構成要素だと言えるのだろうか。それはすでに述べたように、実効的な手続きがあって、ゲーデル数から一意的に構成要素と統語論的構造をいつでもリカバーすることができるし、そのようにして陰伏的にゲーデル数に「潜んでいる」統語論的構造が、ゲーデル数がどのように計算されるかに因果的にかかわりをもつからである。ゲーデル数は連鎖的合成性をもたない。つまり、ゲーデル数の表現はトークン化されたときに、文字通りその構成要素を含むわけではない。しかし、S1、S2 を満たす程度にはやはり統語論的構造をもち、統語論感受性をもった処理を受けつけるのである。(van Gelder, 1990: 368-375)は、こうした表象の統語論的合成性を、機能的合成性(functional compositionality)と呼んでいる[26]。

#### 五-六 フォーダーの議論は成り立たない

もうおわかりだろう。RAAM の分散圧縮表象は機能的合成性をもつ表象のもう一つの事例なのである。それゆえに、統語論感受性をもった処理が可能となり、そのことが認知の規則性を説明する。フォーダーらは、統語論感受性をもった処理には、表象の統語論的構成要素が空間的にトークン化されていて、それぞれが因果的効力をもつことが不可欠だと考え、連鎖的合成性をもつ表象 = 思考の言語の想定を正当化できると考えた。しかし、統語論的感受性をもつ処理に必要なのは、(美濃、二〇〇三:九五頁)の巧みな言い方を借りれば、「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じたいがその構造に応じた因果的役割を果たすことだ。

(信原、二〇〇〇)は、機能的合成性の概念に訴えてフォーダーらの批判からコネクショニズムを擁護しようとする議論を次のように評価している。S1・S2 を満たす分散表象は統語論的構造をもつと言ってよい。古典主義的表象と分散表象の違いは、我々の目から見て構成要素が顕在化しているかどうかにはすぎない。表象がどう処理されるかという観点から見れば、どちらも表象のもつ統語論的構造に基づいて処理されるわけだから、処理の観点からは明示的な構成要素をもつと言ってよい。したがって、コネクショニストモデルは古典主義的モデルにほかならないということになる。こうしてコネクショニズムを擁護しようとする試みは失敗に終わる。

この議論は「したがって」の前後で大きな飛躍がある。処理の観点から見ればどちらも明示的な構成要素をもつと言えるということにかりに認めたとしても、そこからどちらのモデルも同じだという結論は導かれない。「処理の観点」から見て重要なことは、何を処理しているかだけではない。どのように処理しているかもそれ以上に重要だ。そして、この点でまったく異なる提案をするために、古典主義とコネクショニズムは依然として実質的な違いをもつアーキテクチャなのである。その違いは次の点にある。古典主義では、時空的にトークン化している構成要素そのものが因果的効力をもつ。そしてそれが因果的効力を持つのは、構成要素(=表象)に言及する規則がそのシステムに実装されており、その規則に従って構成要素が処理されるからだ。つまり古典主義的モデルでは、どうしてしかじかの処理がなされたのかの説明に、これこれの構

成要素が「そこにあった」ことと、表象レベルでの規則が顔を出す。

これに対し、CN で因果的効力を持つのは、分散表象が特定の統語論的構造をもつという情報である。これが因果的効力を持つのは、そのように学習がなされたからであって、表象とその統語論的構造に言及する規則を CN に実装してあるからではない。CN で実行されているのは、個々のユニットが受け取った信号の強さに重みづけを乗じ、総和をとって...という計算を指示するアルゴリズムである。これは表象をどうこうするというアルゴリズムではない。CN は全体として何らかのアルゴリズムにしたがった計算をしているように「見える」かもしれないが、アルゴリズムによる表象の処理が実際になされているわけではない[27]。

## 五・七 最後の逆襲？

本節の最後に片づけておかなければならない問題がある。フォーダーらはスモレンスキのネットワークについて次のような論評をしているからだ。

コネクショニストに対して認知の体系性(規則性)がもたらす難問は、CN のアーキテクチャのもとでも規則性が可能であることを示すことではない。そうではなく、規則性がいかにして必然的であるか、つまり認知能力の規則性が自然法則であるのはいかにしてかを示すことなのである(Fodor & McLaughlin, 1990:348-349, McLaughlin, 1993)。そして、さらにフォーダーらは次のように主張する。ネットワークを「Rab を表象できる Rba も表象できる」が成り立つようにワイヤリングすることができることは認めよう。しかし、CN ではそれとまったく同様に、「Rab を表象できる Rzq も表象できる」が成り立つようにワイヤリングすることだってできてしまう。一方、古典主義的モデルでは、Rab を表象できるための条件が満たされたなら必然的に Rba を表象できるための条件も満たされてしまう。つまり、古典主義では認知の規則性が本質的に組み込まれており、規則性が必然化される(necessitate)のに対し、コネクショニズムではそれについて中立的であり、そうもできるしそうでないこともできるという具合に高々偶然に規則性をもつにすぎない。したがって、古典主義では規則性を説明できるのに対してコネクショニズムでは説明を与えることはできない。

こうした議論については、次の三点を指摘して反論することができる。

(1) 古典主義的なアーキテクチャ(思考の言語への統語論的操作)だけからは規則性を法則的な帰結として導くことはできない。それができるとしても、ある特定のアルゴリズムにしたがった処理を背景にしたときのみである。だとするなら、どのようにもワイヤリングできるのだからコネクショニズム的アーキテクチャは規則性を法則的に必然化しないという具合に、コネクショニストにだけアーキテクチャだけからの必然化を要求するのはフェアではない。CN のアーキテクチャ + 特定の重みづけ、というペアによれば、規則性が必然化される。これは、古典主義的アーキテクチャ + 特定のアルゴリズムによってはじめて規則性が必然化されるのと同様である(Hadley, 1997a)。

(2) じつは、古典主義なら認知の体系性を説明できるという主張がそもそもかなり怪しい。(i) Rab を表象できる人は、R、a、b を表象できる。(ii) 人間が複合的対象を構成する処理過程は、これらの原子的表象にアクセスできる。これらを認めたとしても、そこから、その人が Rba を表象できるということは、もちろん可能にはなるが必然化されるわけではない。R、a、b を Rab にまとめあげる過程の存在から、それを Rba にまとめあげる過程の存在は出てこないのである(Matthews, 1997)。(Aizawa, 1997a)はこの論点をさらに進め、表象の(連鎖的)合成性だけからは、可能な表象間の体系的な関連についてのいかなることも帰結しないと、極端な例として一つの文しか許さない合成的な統語論を想像してみることを促している。古典主義者もコネクショニストも認知の規則性を説明できないということは、フォーダーらの説明の基準が厳しすぎることを示唆している。

(3) 一方で(Aizawa, 1997a, 1997b)は、コネクショニストの「説明」を、周転円というアドホックな仮定を置いて惑星の動きを説明するプトレマイオスの宇宙論になぞらえる。フォーダーらの批

判のポイントは、こうした思惟経済に反する恣意的な仮定(特定の重みづけがそれに相当する)の存在にあるのだ、というのがアイザワの理解である。しかし私は、おかしいのはアイザワ、フォーダーらの科学的説明についての理解の方だと思う。彼らは、被説明項が必然化される根拠を、それが理論言明から恣意的仮定を置かず演繹されることに見いだそうとしているようである。こうした説明観は工学や生物学、医学におけるシミュレーションやモデル化による説明にそもそも当てはまらない(Matthews, 1997)。訂正されるべきは、彼らの説明観の方と思われる[28]。

## 六 コネクショニズムと思考の言語

### 六一一 思考の言語仮説とは何だったのか

これまでにわかったことを整理しよう。まず、次のことが指摘できる。S1 と S2 からは、表象の連鎖的合成性つまりフォーダーらが考える意味での「思考の言語仮説」は必ずしも帰結しない。したがって、フォーダーの思考の言語仮説は、(S1) 表象の合成性、(S2) 処理の統語論感受性に加え、それらがどのような仕方で実装されているかについての特定のコミットメント(連鎖的合成性)を含む立場だということになる。つまりフォーダーらの立場じたい2つのレベルにまたがった主張であり、実装レベルでの選択を含んでいる[29]。

表象の合成性と処理の統語論感受性を満たすもう一つのやり方(機能的合成性をもつ表象)がある。コネクショニズムの分散表象はその候補だ。この発見により、「思考の言語仮説」を S1 と S2 を満たす立場全般を指すものと考え、古典主義とコネクショニズムはその対立するサブカテゴリーと位置づける論者が現れてきている[30]。しかし、これはあまりよい分類法とは言えない。一つには、コネクショニストには認知理論であるために S1、S2 にコミットしなければならない理由がないからである(後述)。もう一つには、この名称は、論理的推論などの認知過程が、言語的ではないが統語論的構造をもつような表象によって担われている可能性を見えにくくするからだ。たとえば(大沢、二〇〇三)は、シンボル表象(言語)に対比した意味でのイメージ表象(典型的には図)を操作することによっても論理的推論が可能であることを示し、それがコネクショニズムでの表象の有力候補であることを主張している。イメージ表象も統語論をもち、それゆえ S1・S2 を満たしうるのだが、これをも思考の「言語」と呼んでしまうと、古典主義的な表象との重要な違いが隠されてしまう。

これは言葉遣いの問題だと言ってしまえばそれまでだ。しかし、何らかの仕方で古典主義とコネクショニズムを区別しながら包括することのできる用語があった方がよいだろう。そこで、妥協的だがとりあえず次のように整理しておくことにしよう。

(i) 強い思考の言語仮説(古典主義) = (S1) + (S2) + (CC) 連鎖的合成性

(ii) 弱い思考の言語仮説 = (S1) + (S2) + (FC) 機能的合成性

これまでにわかったことを一言で言えば、コネクショニズムは弱い思考の言語仮説と両立可能であり、それゆえに認知の規則性を説明できるということである。

### 六一二 弱い思考の言語仮説とコネクショニズムの関係

コネクショニズムの分散表象は機能的合成性をもち、したがって統語論感受性のある処理を受けつけることが可能である。このことにより、コネクショニストはフォーダーらの議論 2 を斥けることができる。ここまでは私は正しいと思う。しかし、以上のことと、コネクショニストは弱い思考の言語仮説を受け入れなければならないということとはまったく別である。本節ではこのことを示し、弱い思考の言語仮説をコネクショニズムの基本仮説として採用すること、ましてや(美濃、二〇〇三)のように「認知理論」であるための要件としてア priori に要請することはナンセンスだと主張したい。そのためにまず2つの点を指摘しておこう。

(1) 機能的合成性はタスク相対的である:まず、「機能的合成性をもつ表象」という概念は文脈依存的であることに注意しよう。RAAM の圧縮表象が機能的合成性をもつとされるとき、どのような論拠に基づいてそう主張されるのかを振り返ってみると、このことはよくわかる。

RAAM 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見なすことが可能である。入力された(再帰構造をもつ)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。(美濃、二〇〇三:九八頁、強調は原著者による)

圧縮表象が機能的合成性・統語論をもつ(S1)のは、入力された外部表象(英語の文)の統語論的構造をそれにコード化して保存し、いつでも解読して、入力と同じ統語論的構造をもった外部表象を再現できるからに他ならない。またチャルマーズの実験で、圧縮表象が陰伏的にもっている統語論的構造に即した処理がなされている(S2)と言われるのも、最終的に対応する受け身文という外部表象がリカバーできたからである。つまり、この圧縮表象は、もともと統語論的な構造をもった入出力に対して統語論的変形というタスクを行うためのシステムで用いられているからこそ、それに寄生する形で統語論的構造をもつ。

岩・機雷識別システムの分散表象と比べてみるとよくわかる。この場合の分散表象は表象であることは疑いないが、統語論的構造をもっているとは思えない。それはなぜか。この場合の入力は反射音のスペクトルだった。これに統語論的構造がないからである。

いくつか注意が必要だ。まず第一に、だからといって機能的合成性という概念は無効だとか空虚だと主張しているわけではない。統語論的タスクを行うCNの表象は機能的合成性を現にもっている。そうでないCNに現れる表象は機能的合成性を現にもたない。言いたいことは、分散表象が合成性をもつかどうかは、タスクに相対的に決まるということだ。あらかじめ明示的な統語論的構造をもつ外部表象の操作というタスクを行う限りにおいて、内部表象も機能的に合成性もちうる。

(2) 機能的合成性と連鎖的合成性の重要な差異:第二に、次のような反応が考えられる。タスク相対性は機能的合成性に限ったことではない。古典的な表象が連鎖的合成性をもつかどうか、あるいはそもそも表象かどうかということすらタスクと相対的ではないか。これはその通り。何らかの物理的対象があったときに、それだけを取り出して連鎖的合成性・統語論をもつかどうか問うのはナンセンスだ。英字ビスケットというお菓子がある。これを袋に詰めている菓子屋の店員は、或る意味で文字に操作を加えているかもしれないが、それは統語論的操作ではない。しかし、このビスケットを並べてスクラブルをプレイしたりメッセージを残そうとしている人は統語論的操作を加えている。このとき、ビスケットの列は連鎖的合成性をもつ。いずれにせよ、あるアイテムが統語論・合成性をもつかどうかは、どのような処理にそれが埋め込まれているかによって決まる。

表象がもつ統語論的構造の同一性基準を考えてみよう。連鎖的合成性をもつ表象の場合でも、形式言語のように最初から統語論的規則を明文化してあれば話は別だが、そうでない場合とあえずは機能的に同定するしかない。たとえば、アイテム「~P」の統語論的構造、たとえば原子的表象なのか複合表象なのかは、それだけ見ても決まらない。その「~P」が処理を受けると、「~」をはずされて「P」だけになったり、逆に「~ ~P」になったりといったこと、「P」と「~P」ではそのあと受ける処理に違いが出てくること...。こうしたことによって「~P」の統語論的構造は決まる。ここまでは、連鎖的合成性と機能的合成性にいかなる違いもない。

しかし、両者にはやはり大きな違いがある。連鎖的表象では、いったんこのようにして機能的に統語論的構造の同一性基準(機能的基準)が取り出されたら、その基準は容易に、表象の空間的配置や形態による基準(形態的基準)に転写でき、後者を前者の代用品として用いるこ

とができる。つまり、「~」は一項文演算子、「P」は原子文と決めれば、あとは同一システムのどこに出てきても、同じ形をしたアイテムは同じく原子文だし、同じ配列になっている複合的アイテムは同じ統語論的構造をもつと言えるようになる。なぜなら、連鎖的表象ではその機能がアイテムの形と配置によって果たされるからだ。

一方、機能的合成性しかもたない分散表象は、機能的基準以外のもう一つの基準をもてず、あくまでもネットワーク内で果たす役割によってしかその構造を同定できない。回帰的ネットワークではこの事情はもっと深刻である。なぜならこうした CN では、隠れ層の活性化パターンは1ステップ前にどのような処理がなされたかにも依存する。そうすると、隠れ層にまったく同じ活性化パターンを再現しても、それが同じ機能を果たすとは限らない。分散表象は、それがもつ機能的合成性を、その表象自体のもつ別の観察可能な性質に転写することがおそくできない。

(3) 弱い思考の言語仮説はあまり意味がない: 以上の考察は何を含蓄するだろうか。連鎖的合成性をもつ古典的表象は、統語論的構造の同一性基準を二重にもつことができる。強い思考の言語仮説が認知理論の基本テーゼの身分をもちうるのは、この二重性による。どういうことか。思考の言語仮説は、たんに人間が自然言語を処理しているときには心の中にも言語のような表象があるという仮説ではない。一見言語を使っているようには思えない知覚や顔認知などを含む認知処理のすべてが、言語的構造をもった表象を操作することで行われているという仮説でなければならない(そうでなければ、認知理論の基本的仮定とは言えないだろう)。思考の言語がこうした普遍的な理論装置の役割を果たすためには、そもそも思考の言語が「そこにある」のかどうか、思考の言語の文がどのような統語論的構造をもっているのかが、その果たしている機能、それが利用されているタスクといったんは独立に同定できなくてはならない。そうでないと、言語的なタスクでないような認知過程も思考の言語に担われている、という言い方に実質的な意味を与えることができない。連鎖的合成性をもつ古典的表象の場合は、形態的基準を合わせもつことによってそれが可能になる。明らかに統語論的なタスクに現れ機能的な意味で統語論的構造をもっている表象と、形態的な基準において同一の表象が、別のとくに統語論的には見えないタスクにおいて同様の処理を受けている、という事態が少なくとも概念的に可能なので、人間の認知はみんな思考の言語の統語論的処理だという主張に意味がある。

これに対し、高々機能的合成性しかもたない分散表象においては、それが統語論的構造をもった表象なのかどうか、どんな統語論的構造をもっているのかが、つねにどのようなタスクの中でその表象が処理されているかに概念的に依存してしまい、二重の同一性基準をもつことができない。このために、「統語論的タスクに見えないタスクにおいても、じつは統語論的構造をもった表象に対する統語論感受性をもつ処理がなされているのだ」という仮説(弱い思考の言語仮説)に実質的な意味を与えることができない。分散表象が(弱い意味での)思考の言語であるのは、入出力がすでに古典的表象であるときに限られる。ゲーデル数にしても、RAAMの圧縮表象にしても、それが機能的合成性をもつのは、連鎖的合成性をもつ論理式や自然言語の文との間に実効的で一意的な変換手続き(前者ではゲーデル数づけの規則、後者ではRAAM そのものがそれを計算している)が存在することに依存している。少なくとも、コネクショニストはそのような例しか提示できていないのである。というわけで、もしコネクショニストが弱い思考の言語仮説をあらゆる認知処理に当てはまる基本的仮説として採用したいなら、入出力の連鎖的合成性に依存しない形で、分散表象が機能的合成性をもつのはいかなるときかについての基準を提案しなければならない。しかし、それは機能的合成性の概念の性格からしておそらく不可能だろう。

結論はこうである。ある種のタスクにおいて分散表象が機能的合成性をもちうるという概念は空虚ではないし、それによってコネクショニズムは認知の規則性を説明できないというフォーダーラの批判を斥けることもできる。しかし、機能的合成性をもつ表象の統語論的処理によって



すべての認知過程を理解すべきだという弱い思考の言語仮説を採用すべきだという根拠はないし、採用しようにもこの「仮説」はこのままでは無内容である[31]。

### 六・三 思考の言語仮説にさよなら

この結論が正しいなら、コネクショニストは S1・S2 に依拠せず、つまり思考の言語を想定せずに、認知の規則性を説明する方向を探るべきだということになる。最後に、そうした方向を簡単にスケッチして本章を終えることにしよう。そのためには、いくつかの断片的なアイデアを撚り合わせる必要がある。

(1) 言語は表象ではないかもしれない: (Cummins, 1996) は表象の本性についての大胆な提案を行い、自然言語についての我々の考え方をがらっと変えることを要求している。彼の「表象のピクチャー理論」(Picture Theory of Representation: PTR) では、地図のような図的表象を心的表象の典型例と考える。地図が実際の地形の表象として使えるのは、その地図と実際の地形が構造として類似しているからである。もう少し詳しく言えば、地図という構造(表象構造 representing structure)に含まれる対象、たとえば線や四角形などと、実際の地形(内容構造 content structure)に含まれる対象、たとえば道路や建物と一対一の対応があり、表象構造に含まれる対象同士に  $n$  項関係が成り立っている(たとえば 火 の右にある)ときには、それぞれに対応する内容構造側の対象にも対応する  $n$  項関係が成り立っている(消防署はコンビニの東にある)。こうした関係が2つの構造のあいだに成り立っているとき、2つの構造は同型(homomorphic)であると言う。カミンズは、すべての自然な表象(心的表象はその一例である)の基礎をなすものはこうした構造間の同型性(類似)であると考え(Cummins, 1996: 96-111) [32]。

ところが、自然言語の語や文といったコンベンショナルな「表象」の場合は、それらとそれらが意味する構造との間に同型性が成り立たないことは明らかだ。したがって、PTR が正しいなら、言語記号は表象ではないか、少なくとも表象の典型例ではないことになる。

自然言語が表象でないとしたら、それはいったい何なのだろう。(Cummins, 1996: 135-140) は次のように続ける。自然言語が思考伝達の道具であることは認めよう。しかし、これが可能なのは、自然言語の記号トークンが思考を表象するからではない。コミュニケーションのためにはそうした表象性は必要ない。語や文が私の思考を誰かに伝えるのは、その語や文が類似の思考をコード化している図的表象を相手の脳の中に引き起こすからである。自然言語の記号はトリガー(trigger)であって内容をもたない。したがって、自然言語がコミュニケーションの道具であることは、それが表象であることを意味しない[33]。

(2) 思考のファシリテーターとしての自然言語: 思考の言語仮説を拒否するということは、思考の媒体(media)として統語論的表象を拒否することだ。カミンズの提案がもし正しければ、思考の言語を拒否するより強い動機が与えられる。言語が表象でないということと、認知は表象に対する操作であるということからは、言語は認知の媒体にはなれない、ということが帰結するからである。

ここで生じる問題は、言語のように構造化した媒体なしでどうやって高度な抽象的思考が可能なのか、というものだ。これは次のように言いかえてもよい。言語は抽象的思考にとって不可欠に思われる。そして、内観からも、たしかに自然言語は意識的な思考の多くに伴う。抽象的思考において言語が媒体としての働きをしているのでなかったなら、いったいどのような働きをしていると言うのか。一つの答えは、言語は思考のファシリテータの役割を果たすというものだ。手書きの文字のような外的言語記号は、勝手に消えたりしないでそこにとどまり続けてくれる。そのおかげで、我々は紙の上の文字列を外部記憶装置として用いることができ、心の中でパターン認識だけを行うことによって解けないような問題も解けることになる。(Rumelhart et al., 1986: 45) はそれを筆算を例にとって説明している。3桁の数同士のかけ算のような複雑な問題解決は、とてもコネクショニズム的なパターン認識だけで行っているとは思えないし、少なくとも

私は心の中だけでこの計算をすることができない。しかし、紙の上の外的記号の助けを借りれば簡単に計算できてしまう。外的な言語は思考の補助手段として役立つ。

(3) 自然言語の内面化: ヴィゴツキーは、子どもが自然言語をコミュニケーションの道具として身につけたのち、それを内面化して認知の道具にも使い始めることを指摘している (Vigotsky, 1962)。子どもは複雑な問題を解く場合に、最初は、「こうするでしょ。そして、ここをこうするでしょ...」と自分に語りかけながら実行する (egocentric speech)、その後、内語 (internal speech) を頼りに問題を解くようになる。自然言語は内面化されることによって、思考・認知の道具となる。

(4) 手続き的知識の集積としての思考: 以上から、言語を要求するよう見える高度な思考が我々に可能であるという事実を、次のような仕方で説明する道が拓けてくる。

(i) 高度で抽象的な思考は、自然言語の外的記号を道具として使ってはじめて可能になる。人間の「高度な」認知能力の本体は、ほんらい外的な統語論的構造をもったアイテム (自然言語) を相手に作業を行うための手続き的知識にある。したがって、認知システムに言語のような構造をもった表象を実装しておく必要はなく、そうした外的記号を扱う手続き的知識だけを実装すればよい。

(ii) しかし、我々はいつも紙と鉛筆を使って考えるわけではない。心の中だけでもそうした複雑で抽象的な思考ができるように見えるのは、発達の過程で、脳が統語論的構造をもった外的自然言語を内面化するからだ [34]。こうして、複雑な思考・推論の能力は自然言語の内面化に由来するが、それはつねに不完全である。我々は依然として、非常に複雑なことから外的記号の助けを借りないと考えることができない。

(iii) 自然言語が内面化されるということは、それが思考の表象的媒体つまり思考の言語になるということではない [35]。心の中で言語が果たしている役割は外で果たしていたものと同じである。つまり、思考のファシリテータであるか、脳の各部分をつなぐコミュニケーションの道具、つまり、脳の一部が他の部分にしかるべき (分散) 表象を生じさせ認知活動を引き起こすためのトリガーである。こうして内面化された自然言語が強力な認知のための道具になる。

(iv) じつは人間の認知が示す規則性はフォーダーらが考えているほど普遍的ではない。とくに、人間のあらゆる認知タスク (知覚、パターン認識) も規則性を示すという主張はほとんど説得力がない [36]。規則性を示す認知タスクはほとんど自然言語使用にかかわるものばかりである。こうした認知タスクにおいて規則性が見られるのは、自然言語が統語論的構造をもっているからに他ならない。自然言語は思考の言語の外化したものだから統語論的構造をもつのではなく、それじしんで構造をもつ。古典主義者や弱い思考の言語を認めるコネクショニストは、規則性を説明するためには、言語と同じような統語論的構造を顕在的であれ明示的であれ思考の媒体である内部表象がもたねばならないと考えた。ここでわかったことは、それ以外の第三の道が可能だということである。

このような立場を (O'Brien & Opie, 2002) は「ラディカルなコネクショニズム」と呼んでいる [37]。彼らはそのスローガンを次のように表現している。我々は言語を使って (with language) 考えるのであり、言語において (in language) 考えるのではない (O'Brien & Opie, 2002: 315)。規則性の説明に思考の言語は必要不可欠ではない。

#### 六・四 おわりに

哲学者はつい、思考の言語仮説は認知理論であるための必須条件であるとか、認知理論であるからにはかくかくしかじかのレベルに定位しなければならないとか、それは実装レベルのことであって認知理論ではない、という具合に、ほんらい経験的仮説であるはずのものを、ア・プリオリな要請へと「昇格」させたがる。しかし、このような昇格にはあまり根拠がない。このことを私は本章を通じて示してきたつもりだ。哲学者が行うべきことは他にもある。たとえばラディカルなコネクショニズムは、まださまざまな着想の寄せ集めにすぎない非常に弱い経験的仮説だ。これをより首尾一貫したものにし、その弱点を明らかにし、今後取り組むべき問題を明示する手

伝いをする事、これも哲学者にとってやりがいのある課題ではないだろうか。

## 注

[1] これは言うまでもなく「計算」という概念じたいが捉えどころのないものだからだ。もちろん、計算可能な関数という概念は数学的に明確な定義が提案され受け入れられている(チャーチ・チューリングのテーゼ)。しかし、これは計算とは何かの明確化とはあまり関係がない。計算概念の捉えにくさについては(戸田山、二〇〇二)あるいは(Smith, 2002)を参照。

[2] 思考は計算だという考え方じたいは少なくともホッブスやライプニッツまでさかのぼることができる。したがって、認知的革命は計算主義の誕生ではなく、復活と位置づけるべきなのだろう。本章では、計算主義の歴史を扱うことはできない。関心のある読者は(Scheutz, 2002: 1-21)を見ていただきたい。

[3] ここでの「アーキテクチャ」概念の説明は、(Aydede, 1997: 63)を参考にした。

[4] というわけで、本章では認知が表象に対する計算であることを前提したうえで、それがどのようなアーキテクチャをもつのかということの問題にしていくことにする。しかし、この前提は自明ではない。力学系アプローチをとる人々の中には、そもそも認知は計算ではなく、したがって表象も必要ないと主張する人々もいる(van Gelder, 1995)。私は(美濃、二〇〇三、一〇七頁)のように、非表象主義の認知理論がただちに「自滅的な理論」だとア prioriに言えるとは思わない。非表象主義の是非をめぐる議論は、(服部、二〇〇三)(金子、二〇〇三)あるいは本書の第 章に譲ることにする。

[5] 形式システムとしての心の特徴は、すでに述べたように、そこで操作される離散的アイテムが表象であり、したがって内容をもつという点に求められる。そこで、心は「解釈された形式システム」と言われたりもする(Haugeland, 1985)。

[6] (Fodor & Pylyshyn, 1988: 13)では、この2つの要件が古典的アーキテクチャを定義する特質だと考えられている。

[7] 紙幅の制限のため、ここでのおさらいは本当に簡単なものになってしまう。本章のメインの議論は、読者に接続ニズムにかかわる基礎的事項の理解を要求する。というわけで、初めて接続ニズムに接する読者は、あらかじめ(柴田、二〇〇一、第5章)あるいは(戸田山ほか、二〇〇三、序章)などによって、接続ニズムについてのさらに詳細な紹介を読んでいただくとありがたい。

[8] 現実の神経細胞では、活性化値がある閾値を超えたときにだけ一定の強さの信号を出力し、閾値を超えないときには信号を出さないといった性質(全か無かの法則)が成り立っている。この関数  $f$  は非連続な階段関数になるはずだが、こうした非連続関数は数学的に扱いにくいので、階段関数をなめらかにつないだ S 字型の関数(シグモイド関数)で代用することが多い。

[9] ここでは岩と機雷の識別だけが問題だから、出力ユニットは2つでよい。だから出力ベクトルは2元ベクトルになる。

[10] コネクショニズムの先駆としてよく指摘されるのはマカロックとピッツによるニューロンの数理モデルの提案である (McCulloch & Pitts, 1943)。近年、すでに学生時代のチューリングが unorganized machine という名のもとに、コネクショニズムを先取りするアイデアを抱いていたということが指摘されている (Copeland & Proudfoot, 1996, 1999; Teuscher, 2002)。だとすると、チューリングは古典主義のルーツであると同時にコネクショニズムのルーツでもあることになる。

[11] (Bechtel, 1986)によれば、科学的推論はこれまでに考えられてきたような演繹的・帰納的推論ではなく、外的表象や物理的装置、実験材料などとの相互作用における柔らかな制約を充足することである。どんなアプローチをとっても、通常はいずれかの制約には違反せざるをえないからである。したがって最良の解決は制約のいくつかに反するものになる。こうして、柔軟な制約をうまく扱える CN はこの点で科学哲学へ貢献しうる、とベクテルは考える。

[12] これから紹介する古典主義者によるコネクショニズム批判は、(Fodor & Pylyshyn, 1988) (Foder & McLaughlin, 1990, 1996) (McLaughlin, 1993)で展開されたものである。論争の常として、論点や概念の理解に食い違いや混乱が生じたり、何が問題だったのかが論争の過程を通じて後からはっきりするといったことがこの論争にも見られる。ここでは紙幅の制限のために、論争そのものの経過を通時的に追いかけることはできない。したがって、これから展開されるのは現在の時点から合理的に再構成された論争である。また、上記の批判論文すべてにフォーダーが深くかかわっているので、少々乱暴だがこの論争における古典主義陣営の人々を一括して「フォーダーら」と呼んでしまう。フォーダーらとその批判の標的になったスモレンスキのやりとりのほとんどは (MacDonald & MacDonald, 1995) に再録されている。

[13] これはフォーダーらの言い方では、推論の統一性 (coherence of inference) と呼ばれていた。

[14] (信原、二〇〇〇a) (信原、二〇〇三) (Nilrasson & van Gelder, 1994) (Waskan & Bechtel, 1997) がこの方向で議論を進めている。信原に批判的な (美濃、二〇〇〇) も、認知が偶然の事実としてではなく本質的に体系的だとする限りでフォーダーらの立場を批判している。

[15] たとえば、コネクショニズムの難問の一つに変項束縛問題 (variable binding problem) というものがある。Ross gave Rachel the ring. から Rachel owns the ring. を推論するには、 $xyz[GIVE(x,y,z) \text{ OWN}(y, z)]$  のような規則の変項  $y$  と  $z$  に Rachel と the ring を代入し  $GIVE(Ross,Rachel,the ring) \text{ OWN}(Rachel, the ring)$  を導かなければならない。しかし、量子化子は変項  $y$  と  $z$  のそれぞれ2カ所の現れを束縛しているため、二カ所の  $y$  の両方に Rachel を代入し、二カ所の  $z$  の両方に the ring を代入する必要がある。これは思考の言語を処理して推論をすすめていく古典主義者には簡単なことだ。なぜなら、二カ所の  $y$  は同じ「形」をした記号だからである。しかし、こうした記号の同一性に頼ってことを進めていけないコネクショニストにとっては、二カ所の  $y$  に同じものを代入するという、こんな単純なことが CN や脳のようなネットワークにやらせようとするとひどく難しい問題になってしまうのである (Bechtel & Abrahamsen, 2002: 165)。自然言語処理の分野におけるコネクショニズムの現状についてのサーヴェイとしては (Christiansen & Chater, 1999) が良い。

[16] これは (Lycan, 1987: chap.4) にある two-levelism 批判から学んだ論点である。

[17] もちろん、この評価は「認知モデル」あるいは「認知理論」というものにどのような要件を課すかに依存している。(美濃、二〇〇三)は、認知理論に表象主義、物理主義、機能主義、思考の言語というかなり古典主義よりの要件を課した上でもなお、コネクショニズムがその要件を満たしうるとことを示している。その証明したいは妥当だ。しかし、認知理論の要件をこれほど厳しいものにする必要があるのかについては疑問が残る。特に、思考の言語仮説を受け入れることが認知の理論であるために必要だという論点は明らかに根拠を欠き、いきすぎだと思う。これについては後で触れる。

[18] RAAM についての詳細は、もちろん (Pollack, 1990) を見ればよいが、(Bechtel & Abrahamsen, 2002: 171-175) (美濃、二〇〇三: 九六頁～九九頁)にもわかりやすい説明がある。もう一つの例はスモレンスキーのテンソル積表象 (Smolensky, 1988: 10) である。実のところ、フォーダーらと論争したコネクショニスト側の当事者はスモレンスキーなのだが、彼の CN は RAAM と異なり、本章で例示した3層フィードフォワードネットワークとはかなり違ったアーキテクチャをもっているため、例として扱いにくい。しかし、スモレンスキーとフォーダーらの論争の本質的なポイントは、事例に依存しないので、ここでは RAAM を事例とさせてもらうことにする。スモレンスキーのテンソル積表象を事例として使いながら、フォーダーらの【議論 2】に反駁を試みた論文としては (Guarini, 1996) を参照。

[19] こうした実験を評価する際にもうひとつ忘れてはならないポイントは、実験の際に用いられた単純化されたタスクをそのまま、現実に人間が行っているサイズのタスクに「スケール・アップ」することができるか、という問題だ。哲学者はこの論点を見落としがちである。というのも、哲学者は「CN が認知の規則性を示すことが原理的に可能か」という具合に問いを立てるので、現実離れした簡単なモデルでも哲学的議論のためには十分だと考えてしまうからだ。しかし、どちらの研究枠組みが現実の人間の認知理論として有望かといったことを考えている局面では、程度の問題を無視するこうした態度は危険かもしれない。(Haselagar & van Rappard, 1988) は RAAM を題材の一つに取りあげて、こうした CN の現実の認知へのスケール・アップの可能性を疑問視し、コネクショニズムは古典主義が直面しているフレーム問題を解決・回避できるという楽観論に異議を唱えている。しかし、本章ではこの問題に深入りできない。今後の課題としておこう。

[20] フォーダーらが分散表象では認知の規則性を説明できないとするいちばんの理由は次のものだ。統語論的構造をもった分散表象がかりにあったとしても、その構成要素は分散し重なり合ってしまうので、離ればなれに「そこにある」わけではない。そんな分散複合表象が、その構成要素と統語論的構造に因果的に依存した仕方で処理されるとしたら、それは奇跡と言うべきだ (Fodor & McLaughlin, 1990: 215)。つまり、フォーダーらの批判の核心は、分散表象は S1 を満たすことができるかもしれないが、S2 は決して満たせないという点にある。

[21] nil というのは空所を表すダミー記号である。

[22] だから、二種類の規則性をわざわざ区別して導入したのである。

[23] ゲーデル数をこの論点の説明に使うというアイデアは (Aydede, 1997: 73)、(van Gelder, 1990: 361-363) から借用した。

[24] ゲーデル数を割り当てる方法はいろいろありうるが、どれによったとしても、ごく短い論理式のゲーデル数ですら非常に桁数の多い数になってしまう。だから、ここにある数字じたいは

あくまでも説明のために適当に選んだインチキなものであるから注意してほしい。

[25] これはきちんと言うと、(i)一意的に構成要素をリカバーできる、(ii)その統語論的構造を一意的に同定できる、ということである。

[26] 機能的合成性のさらに詳しい説明は(信原、二〇〇〇: 六～八頁)参照。また、この言葉を使っているわけではないが、(美濃、二〇〇三: 九四～九六頁)も機能的合成性を擁護している。また(柏端、二〇〇三)は形而上学的観点から機能的合成性をもつ表象の実在性を説得的に証明している。

[27] 信原の議論についてのこの批判は、(美濃、二〇〇三: 一一一頁)によるものを私なりに言いかえたものにすぎない。信原はかなりラディカルなコネクショニストである。ここでの信原の議論は、「弱い思考の言語仮説をもつようなヤワなコネクショニズムは、オレの目から見れば古典主義と五十歩百歩だ。そんなヤワなコネクショニズムは嫌いだ」と言っているにすぎないように思われる。後で明らかになるように、私じしんは信原の目指しているよりラディカルなコネクショニズムにシンパシーをもっているのだが、この議論の説得性に限っては美濃の評価の方が正しいと思う。

[28] 以上の論点はもちろん、心理学における説明とは何かという問題に関係してくる。これは重要な話題だがここでは深入りすることができない。(Cummins, 1983)を参照。心理学的に興味深い説明は実現(instantiation)による説明ではなく機能的分析でなければならないとするカミンズの見解は、(Matthews, 1997)で検討され斥けられている。

[29] もちろんこのようにレベルをまたいだコミットメントが含まれることをもって認知理論として不純だという具合に批判するつもりはない。そのような批判は認知理論とは何かについての3レベル主義誤解の産物であり、完全に的はずしている。認知の理論は中間のいくつものレベルにまたがるさまざまな理論的選択の複合体と考えるべきだ。

[30] (美濃、二〇〇三)(Horgan & Tienson, 1996)(Aydede, 1997)を見よ。

[31] (Horgan & Tienson, 1996: 81-86)は、コネクショニストもここで言う弱い思考の言語仮説を採用すべきだと主張し、その根拠として追跡論証(tracking argument)と呼ばれる議論を展開している。さらに(美濃、二〇〇三)は追跡論証を「たいへん説得的」として評価し、(弱い)思考の言語仮説を認知の理論の必須の要件としている。追跡論証とは大雑把に言えば、次のような議論である。複雑きわまりない世界でうまく生きていける生物は、外界のデータを「何が・どうした」という具合に対象と性質・関係(主語と述語)に分けて表象し、それぞれを追跡することができなくてはならないはずだ。

これは、別途検討しなければならない興味深い議論だが、最低限のことだけ述べておく。(i)対象と性質をそれぞれ追跡することのできる表象過程があるということからは、構成要素として主語と述語を(陰伏的に)もつ表象を使って認知がなされるということとはただちには出てこない。(ii)この議論は、主語と述語に分けて表象するのが能率的だという直観に依拠している。これは脳の表象能力を過小評価している(Garson, 1997: 347-348)。脳が分散表象を用いているなら脳のコード化能力は莫大なものと考えられる。少なくとも我々の直観はこのような議論をするには役立たない。(iii)以上を譲ったとしても、せいぜい追跡論証から言えることは、主語述語に分節した表象を使って思考する生物は生存において有利だということどまりであり、思考がそのような表象を媒体としている(つまりそのような表象への統語論的操作が思考である)という

ことではない。(iv)さらに、百歩譲って思考の言語仮説が認知理論の有力で生産性の高い仮説であると認めたとしても、それを美濃のように認知理論であるための要件とすることは倒錯していると言わざるをえない。追跡論証の詳細については(美濃、二〇〇三:九一～九二頁)を参照。

[32] この表象の捉え方について注意すべきことを述べておこう。(1)まず、構造 R(地図)が構造 C(地形)を表象するのは、R に含まれる各対象が C の何らかの対象を表象しているからではない。むしろじたいは逆で、R と C の間に構造的な同型性が成り立っているとき、R は C を表象していると言い、そのことから派生的に、R の各対象が C の対象を表象していると言われるのである。(2)PTR は、カミンズ自身が(Cummins, 1989)で展開した概念役割意味論の失敗を踏まえて提案されたものである。概念役割意味論では、構造 R の表象内容が R がどのような役割を心的処理において果たしているかによって決まる。これに対し、PTR では構造 R の表象内容は R が心的処理でどのように用いられるかとは独立に、同型性によって決まるという点で、表象内容は表象に内在的(intrinsic)である。(3)PTR では、表象するという関係は2つの構造間の対応であるので、構造をもたない表象はありえないが、その構造が統語論的構造であることはもちろん要求されない。

[33] トリガーとしての言語という考え方をうまく説明している例を挙げておこう(Cummins, 1996: 70)。「牛の乳を搾ってこい」と言われたとする。あなたは、バケツをもって納屋に行き、一連の作業を行う。ここで「牛」が果たしている役割は、ただあなたの一連の行動の引き金を引くことだけである。バケツをもって納屋に行くためには、なるほど、バケツの表象と納屋の表象、納屋の扉や錠の表象が必要だろうが、牛の表象は必要ない。「牛」はトリガーであればよく、牛を表象している必要はない。しかし、牛に近づいて行って乳を搾るときには、牛の表象が必要だろうと言われるかもしれない。それはその通りだが、この作業のためにはるかに有効なのは、牛の図的表象であることは明らかだ。牛はどんな形をしていてどのあたりに乳房がついているのかについての情報を引き出せるからである。いずれにせよ、自然言語の「牛」が牛の表象として役に立つ場面はここには存在しない。

[34] もちろん、「内面化」とは何かを明らかにし、そのメカニズムを説明する仮説を立てる作業が残っている。(Rumelhart et al., 1986: 38-48)では、外的記号環境のメンタルモデルを構成して、それまで外的記号に対して行われていた手続きがそのメンタルモデルに対してなされるようになる、という仮説が提案されている。

[35] 信原は、内面化された自然言語をしきりに脳の外に置こうとする。たとえば、内語は音量ゼロの発話であり脳の出力なのだから脳の外にある(信原、二〇〇〇a: 一一頁)(信原、二〇〇〇b: 二〇一～二〇二頁)とか、内語による思考は外的表象の操作による思考のふりをするのであり、外的表象は脳の外にあるのだから、そうした思考は脳の外で行われている(信原、二〇〇〇b: 二〇三頁)、といった具合である。これはどう鼻屑目に見ても説得力を欠いており、(美濃、二〇〇〇: 五二～五五頁)によって適切に批判されている。おそらく信原は、内面化された自然言語の処理を脳の内部に置いてしまうと、そのタスクに関して CN は思考の言語の実装を与えるものにすぎなくなり、脳はコネクショニズムと古典主義のハイブリッドシステムになってしまうのを恐れているように思われる。まず私は、脳がどのようなアーキテクチャなのかは経験的に見いだされる問題だから、一部が古典主義的アーキテクチャになっていたってよいではないか、なぜこんな無理をしてまで脳全体をコネクショニスト・アーキテクチャだと言い張らねばならないのかと疑問に思う。さらに、内面化された自然言語を脳の中に置いたとしても、ただちに思考の言語を復活させ脳をハイブリッドシステムにしてしまうことにはならないと思われる。な

ぜなら第一に、内面化された自然言語がメンタルモデルやあるいは脳の他の部位に対するトリガーであるかぎり、それはまったく思考の言語仮説が想定するような思考の言語ではないからだし、第二に、我々は内面化された自然言語の使用を止めていつでも外部記号を使った思考に切り替えることができるからである(暗算しようとしてうまくいかないのを筆算に切り替える)。

[36] (信原、二〇〇〇a:八～一〇頁)は知覚には体系性(本章で言う規則性)が欠けていると論じている。

[37] この立場にたつ論者は他に(信原、二〇〇〇a)(信原、二〇〇〇b)(Churchland, 1995)がいる。また、(Bechtel & Abrahamsen, 2002: 178-194)も参照。

## 参考文献

- Aizawa, K., 1997a, "Explaining Systematicity", *Mind and Language* 12: 115-136.
- Aizawa, K., 1997b, "Exhibiting versus Explaining Systematicity: A Reply to Hadley and Hayward", *Minds and Machines* 7: 39-55.
- Aydede, M., 1997, "Language of Thought: The Connectionist Contribution", *Minds and Machines* 7: 57-101
- Bechtel, W., 1996, "What should a connectionist philosophy of science look like?", in McCauley, R. (ed.), 1996, *The Churchlands and their Critics*, Oxford: Blackwell:121-144
- Bechtel, W. & Abrahamsen, A. 2002, *Connectionism and the Mind*, 2nd ed. Oxford: Blackwell.
- Chalmers, D., 1990, "Syntactic Transformations on Distributed Representations", *Connection Science* 2: 53-62
- Christiansen, M. H. & Chater, N., 1999, "Connectionist Natural Language Processing: The State of the Art", *Cognitive Science* 23: 417-437
- Churchland, P. A., 1995, *The Engine of Reason, the Seat of the Soul*, Cambridge MA: The MIT Press
- Copeland, B. J., & Proudfoot, D., 1996, "On Alan Turing's Anticipation of Connectionism", *Synthese* 108: 361-377
- Cummins, R., 1985, *The Nature of Psychological Explanation*, Cambridge MA: The MIT Press
- Cummins, R., 1989, *Meaning and Mental Representation*, Cambridge MA: The MIT Press
- Cummins, R., 1996, *Representations, Targets, and Attitudes*, Cambridge MA: The MIT Press
- Fodor, J. A., 1975, *The Language of Thought*, Harvard University Press
- Fodor, J. A. & Pylyshyn, Z. W., 1988, "Connectionism and Cognitive Architecture: A Critical Analysis", *Cognition* 28: 3-71
- Fodor, J. A. & McLaughlin, B. P., 1990, "Connectionism and the Problem of Systematicity: Why Smolensky's Solution doesn't Work", *Cognition* 35: 183-204.
- Fodor, J. A. & McLaughlin, B. P., 1996, "Connectionism and the Problem of Systematicity (continued): Why Smolensky's Solution Still doesn't Work", *Cognition* 62: 109-119.
- Garson, J. W., 1994, "Cognition without Classical Architecture", *Synthese* 100: 291-305.
- Garson, J. W., 1997, "Syntax in a Dynamic Brain", *Synthese* 110: 343-355.
- Guarini, M., 1996, "Tensor Products and Split-Level Architecture: Foundational Issues in the Classicism-Connectionism Debate", *Philosophy of Science* 63 Issue Supplement: S239-S247.
- Hadley, R. F., 1997a, "Cognition, Systematicity and Nomic Necessity", *Mind and Language* 12: 137-153.
- Hadley, R. F., 1997b, "Explaining Systematicity: A Reply to Kenneth Aizawa", *Minds and Machines* 7: 571-579.
- Haselager, W. F. G. & van Rappard, J. F. H., 1998, "Connectionism, Systematicity, and the Frame Problem", *Minds and Machines* 8: 161-179
- 服部裕幸、二〇〇〇「コネクショニズムとは何か?」、『科学哲学』33-2: 15-28
- 服部裕幸、二〇〇三「分散表象は認知の説明にはたして役立つのか?」、(戸田山ほか、二〇〇三、二九～五一頁)所収
- Haugeland, J., 1985, *Artificial Intelligence: The Very Idea*, Cambridge MA: The MIT Press
- Horgan, T. & Tienson, J., 1996, *Connectionism and the Philosophy of Psychology*, Cambridge MA: The MIT Press
- 金子善彦、二〇〇三「表象なしのコネクショニズム」とその行方」、(戸田山ほか、二〇〇三、五二～七八頁)所収
- 柏端達也、二〇〇三「コネクショニズムは素朴心理学に対して何か言えるのだろうか」、(戸田山ほか、二〇〇三、一四四～一七七頁)所収
- Lycan, W., 1997, *Consciousness*, Cambridge MA: The MIT Press
- MacDonald, C. & MacDonald, G., 1995, *Connectionism: Debates on Psychological Explanation*, Oxford: Blackwell
- Marr, D., 1982, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W. H. Freeman & Co.
- Matthews, R. J., 1997, "Can Connectionists Explain Systematicity?", *Mind and Language* 12: 154-177.
- McLaughlin B. P., "The Connectionism/Classicism Battle to Win Souls", *Philosophical Studies* 71: 163-190
- McLaughlin B. P., "Systematicity, Conceptual Truth, and Evolution", in C. Hookway & D. Peterson (eds.), *Philosophy and*



- Cognitive Science, Cambridge: Cambridge University Press: 217-234
- McLaughlin B. P., 2000, "Computationalism, Connectionism, and the Philosophy of Mind", in (eds.), *A Companion to the Philosophy of Mind*, Oxford: Blackwell: 135-151
- 美濃正、二〇〇〇「思考のありか 信原説への疑問」、『科学哲学』33-2: 43-56
- 美濃正、二〇〇三「新しい認知の理論としてのコネクショニズムの可能性」、(戸田山ほか、二〇〇三、七九～一一四頁)所収
- Minsky, M. A., & Papert, S., 1969, *Perceptrons*, Cambridge MA: The MIT Press
- Niklasson, L. F. & van Gelder, T., 1994, "On being Systematically Connectionist", *Mind and Language* 9: 288-302
- 信原幸弘、二〇〇〇a「コネクショニズムと消去主義」、『科学哲学』33-2: 1-14
- 信原幸弘、二〇〇〇b「考える脳・考えない脳:心と知識の哲学」講談社現代新書
- 信原幸弘、二〇〇三「コネクショニズムと日常的推論」、(戸田山ほか、二〇〇三、二二六～二四五頁)所収
- O'Brien, G. & Opie, J., 2002, "Radical Connectionism: Thinking with (not in) Language", *Language and Communication* 22: 313-329
- 大沢秀介、二〇〇三「ビクトリアリズムと像的論理」、(戸田山ほか、二〇〇三、二四六～二七二頁)所収
- Pollack, J. B., 1990, "Recursive Distributed Representations", *Artificial Intelligence* 46: 77-105
- Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E., 1986, "Schemata and Sequential Thought Processes in PDP Models", in McClelland, J. L., Rumelhart, D.E., and the PDP Research Group, *Parallel Distributed Processing*, vol.2, Cambridge MA: The MIT Press: chap. 14
- Scheutz, M. (ed.), 2002, *Computationalism: New Directions*, Cambridge MA: The MIT Press.
- 柴田正良、二〇〇一「ロボットの心:7つの哲学物語」講談社現代新書
- Smith, B. C., 2002, "The Foundation of Computing", in (Scheutz, 2002: 23-58)
- Smolensky, P., 1988, "On the Proper Treatment of Connectionism", *Behavioral and Brain Sciences* 11: 1-74
- Smolensky, P., 1990, "Tensor Product Variable Binding and the Representation of Symbolic Structures in Connectionist Systems", *Artificial Intelligence* 46: 159-216
- Smolensky, P., 1991, "Connectionism, Constituency and the Language of Thought", in Rey, G. (ed.), 1996, *Meaning in Mind: Fodor and his Critics*, Oxford: Blackwell:
- Teuscher, C., 2002, *Turing's Connectionism*, Berlin: Springer Verlag.
- 戸田山和久、二〇〇二「局所的意味論と現象としての計算:人工生命と現象としての計算をめぐる天使と悪魔の対話」、『Computer Today』no.111: 40-45
- 戸田山和久ほか編、二〇〇三「心の科学と哲学:コネクショニズムの可能性」昭和堂
- van Gelder, T., 1990, "Compositionality: A Connectionist Variation on a Classical Theme", *Cognitive Science* 14: 355-384
- van Gelder, T., 1995, "What might Cognition be, if not Computation?", *Journal of Philosophy* 101: 345-381
- Vygotsky, L. S., 1962, *Thought and Language*, Cambridge MA: The MIT Press
- Waskan, J. & Bechtel, W., 1997, "Directions in Connectionist Research: Tractable Computations without Syntactically Structured Representations", *Metaphilosophy* 28: 31-62