

## AI 研究および認知科学に対するジョン・サールの批判（＊）

服部裕幸(南山大学)

『アカデミア』人文・社会科学編（南山大学）  
第 77 号、2003, pp.81-103

はじめに

H.ドレイファスは早い時期から AI 研究に対して批判的であったが、彼に劣らず、J.R.サールもまた AI 研究に対して批判的な立場をとったことは今ではよく知られている。彼らの批判に対しては、当然のことながら、AI 研究を積極的に進める立場からの反論がなされたが、ここでは主としてサールの批判とそれをめぐる議論をとりあげ、サールの批判が AI 研究に対して果たした役割が何であったのか、あるいは、何でありえたのか、また何であるべきなのか、こうした点を見極めたいと思う。というのは、サールの議論は一見するとかなり明白であるかのように見え、サール自身もそれをほとんど「自明のこと」と考えているにもかかわらず、その後の議論の展開は、その議論がその見かけほどには明白でないことを示しているからである。

### 1 テューリング・テストと機能主義

コンピュータが人工知能たりうるかという問題に答えるためには、そもそもコンピュータに何ができればそれを人工知能とみなしうるのかという問題に答えなければならない。この問題に対して A.テューリングは、今日テューリング・テスト（1）と呼ばれる基準を提案している。（あるいは、通常はそのように理解されている。）そのアイディアは次のようなものである。壁を隔てて人間の質問者が端末を通じて人間もしくはコンピュータと質疑を行う。壁に隔てられているので、質問者は、自分の対話の相手が人間であるのかコンピュータであるのかは、端末を通じてのやりとりから判断するしかない。このとき、もしその質問者が、自分の対話の相手が人間であるのかコンピュータであるのか識別できなければ、そのときの対話の相手がコンピュータであった場合、そのコンピュータにも知能があるとしてよい。現実にはこのようなテストにパスしたコンピュータは存在しないが、それは今の場合問題ではない。というのは、われわれが問題にしているのは、コンピュータはテューリング・テストにパスできるのか、もし原理的にパスできないとすればその理由は何か、といったことだからである。

### 2 「中国語の部屋」の議論（CRA）

「中国語の部屋」の思考実験は今では誰もが知っていると思うが、以下の議論に

において重要な出発点となるので、再確認のため、労をいとわずこれについての簡単な記述を与えておこう。

サールは中国語をまったく解さない。彼には中国語と日本語の区別もつかない。それどころか、意味のある記号なのか、それともただの落書きなのかの区別さえつかない。そのような彼がある部屋に閉じこめられる。そしてまず、部屋の外から、中国語で書かれた文章と英語で書かれた規則集が彼に与えられる。さらに中国語の文章が与えられる。彼は規則集に従って、最後に与えられた中国語の文章に対応する適当な中国語の文章 といっても、サールにはそれが中国語の文章であることさえわからないのだが を作り、それを部屋の外へ出す。この規則集は非常によくできていて、部屋の外の人と部屋の中にいるサールとの間で会話がなされているかのように見える。すなわち、中国語で書かれた物語についての中国語による質問に対してサールが中国語で答えているかのように見えるのである。しかしながら、仮定によりサールは中国語を解さないので、もちろん部屋の外の人とサールの間で中国語による筆談がなされたわけではない。

さて、規則集はプログラムに相当し、サールはプログラムを実行したと解釈することができる。すると、ここには、入力、プログラムの実行、出力という点で中国語を解する人同士の間やりとりと同じであるにもかかわらず、実際には（少なくとも一人が）中国語を理解していない事例があることになる。

ここからサールは、コンピュータは中国語（の物語）を理解しているとは言えない、それはせいぜいのところ、中国語の理解のシミュレーションを与えているにすぎない、という結論を引き出すのである（Searle1980, Boden1990: 68-70）（2）。つまり、コンピュータは上の思考実験におけるサールのようなもので、見かけは中国語を理解しているように見えるかもしれないが、実際には中国語を何ら理解していないというわけである。当然のことながら、CRA が正しければ、チューリング・テストは有効ではないということにもなる。

### 3 当初の反論 「システム論」と「ロボット論」

この CRA に対していろいろな方面から反論が寄せられたが、そのうちの代表的かつ重要なものは「システム論」と「ロボット論」である（3）。

#### 「システム論」

部屋の中にいるサールが中国語を理解しないということは、まさに仮定により、その通りであるが、コンピュータに相当するのはサールではなくサールのいる部屋全体である。サールは、コンピュータの方で言えば、（比較的中心的な位置を占める）部品の一部にすぎない。個々の部品が中国語を理解しないというのは当然のことであり、したがって CRA は強い AI（4）に対する批判としては有効ではない。

### 「ロボット論」

この反論は、基本的には CRA を妥当な議論と認める。すなわち、コンピュータはそれだけでは中国語を理解しているとは言えない。しかし、感覚器官に相当するものを持ち、手足をそなえ、運動することができ、したがって、世界に対してみずから働きかけることができるようなシステム、すなわちロボットのようなものを考えてみよう。そして、それが中国語を入力として与えられ、出力として中国語を出し、あるいはしかるべき行動を起こし、その反応が中国語を解する人のそれと区別することができないとすれば、そのロボットが中国語を理解していないと言う根拠はないであろう。強い AI の研究者が「人工知能」と言う場合には、このようなシステムを考えているのである。

「システム論」については次に論じるので、ここでは「ロボット論」について簡単にコメントしておこう。サールは、ロボットを考えても事態は本質的には何も変わらないと主張する。テレビカメラやマイクを通じて外から中国語を入力し、規則集にしたがって中国語の音声を出したり、適当な操作をすることによって部屋の外にセットされたアームを動かしたり、あるいは部屋自体を移動させたりしたとしても、自分が何をしているのかを私はまったく理解していない、とサールは述べるのである (Boden1990 : 76-77)。このような批判に対する強い AI の研究者の側からの再反論は、結局は「システム論」と同じ種類のものになるように思われる。すなわち、中国語を理解する人と区別できないような仕方で反応するロボットを、それにもかかわらず実際には「中国語を理解していない」と述べる根拠は、強い AI の研究者によれば、何もないのである。ロボットの中のサールは中国語を解さないかもしれないが、それは、誰であれ、その人の腕の筋肉や心臓が、あるいは脳でさえも、中国語を解さないのと何ら変わりはない。中国語を解するのは全体として見た一人の人間なのである。

#### 4 第2ラウンド 人格の分裂 論

例えばピアノを弾ける（あるいは泳ぎが出来る）人が記憶を喪失したとしよう。この人は自分が誰であるのかを覚えていないし、ましてや自分がピアノを弾けること（あるいは泳ぎが出来ること）を覚えていない。しかし、ピアノを前にして自分がピアノを弾いて（あるいは、川に落ちたときに泳いで岸にたどり着いて）驚くということはあるようなことである。以上の例は技能知に関してであったが、同様のことが言語知においてありえないということは、ア・プリアリには言えないであろう。たしかに、言語知は技能知に似た側面はあるが、技能知と同一視することはできない。それゆえ、自分が中国語が話せるということを知らないにもかかわらず、中国語で話しかけられたときに（その話の内容がわからないにもかかわらず）ひとりでに中国語を話して、みずから驚く、というのはかなり異常な事態であろう。しかし、ありえないこ

とではない。多重人格者はこうした事例にかなり近いと考えられる。(ただし、多重人格者の場合、通常は、一つの人格が表に現れているときには、別の人格は表に現れてこないで、ここで考えている事態とも少し異なる。)

さて、「システム論」に対するサールの反論は、マニュアルやデータの入出力をすべて自分の頭の中に入れたとすれば、自分(サール)がその部屋と同じことになるが、それでもやはり、自分(サール)が中国語を理解していないことに変わりはない、というものであった(Boden1990:73)。すると、ここで起こっている事態はまさに、いま上で想定したようなものである。そうであるならば、異常な事態には違いないが、彼(サール)が中国語を理解していないとは言えないであろう。彼は自分が中国語を解することを知らないが、しかし、自らは中国語を解するのである。あるいは、サールの中に二つの人格が分裂して存在していると言ってもよからう。

## 5 CRAの射程(1) コネクショニズムも論駁されるのか

CRAの標的は当初は古典的AIであったように思われる。実際、サールの1980年の論文が書かれた当時は、AI研究の主流は古典的計算主義に基づくそれであった。(サールがシャंकのプログラムやヴィノグラードのSHRDLUやワイゼンバウムのELIZAに言及していたことを想起されたい。)実際、CRAの一つの定式化は

統語論は(それだけでは)意味論にとって十分ではない。

コンピュータ・プログラムはもっぱら形式的ないし統語論的な構造によって定義される。

心は意味論的内容をもつ。

故に、いかなるコンピュータ・プログラムもそれだけでは心にとって十分ではない。

というものであった(Searle1984:39)。したがって、仮にこの議論が妥当であるとしてもそこから直ちに、コネクショニズムによる人工知能研究が成功しないということとは帰結しない。なぜなら、コネクショニストのシステムでは、計算は、統語論的構造をもった表象に対して統語論的規則を逐次適用する、という仕方で行われるわけではないからである(服部2000)。

ところが、サールの実際の議論においては、「中国語の部屋」の中の人(サール)は中国語が理解できていない、ということがポイントになっている。すなわち、部屋の外にいる人には、中に中国語を解する人がいてその人と自分は会話をしている、と思えるのに、実は中の人には中国語が理解できていない。したがって、このシステムが中国語を理解していると言うことは誤りだ、と論じられるのである。元の例では、部屋の中の人には統語論的規則の適用を(自分が何を行っているかその意味もわからず

に) 逐次実行する、というように想定されているが、いま述べたように議論を理解するならば、その人は他の種類の規則を(自分が何を行っているかその意味もわからずに) 適用するということであってもかまわないように思われる。むしろ、重要なのは、中の人(サル)が中国語を理解していないこと、そして、外から見ると中に中国語を解する人がいるかのように見えることである。そうであるならば、CRAは古典的AIのみならず、コネクショニズムにも妥当する可能性がある。事実、後にサルはそのように主張することになる。次節では彼のその見解を見ることにしよう。

## 6 CRAの射程(2) 「中国語の部屋」から「中国語のジム」へ

「中国語の部屋」の中の人(サル)は、英語で書かれたマニュアルにしたがって部屋に入ってくる中国語の紙片を逐次的に処理して、結果としてその部屋はコンピュータ・プログラムを実行しているのと同じことになるように想定されていた(Boden1990)。ところで、先に述べたように、サルは当初は古典的計算主義に基づくAIを念頭に置いていたと考えられる。したがって、前節で見たように、古典的計算主義のパラダイムにとって代わるものとして喧伝されるコネクショニズムに基づくAIにはCRAは適用されない、とコネクショニストは考えるかもしれない。しかし、サルはこのコネクショニストの期待をうち砕くべく、次のように論ずる。

「中国語の部屋の代わりに、私[サル]が中国語のジム、つまり、英語のみを話す多くの人々がいるホールをもっていると想像しよう。この人々はコネクショニストの機構におけるノードとシナプスと同じ作業を遂行する...そして、その結果として、一人の人にルールブックにしたがって記号を操作させるのと同じ成果が得られる。このジムの中の誰一人として中国語を話しはしない...しかし、適当な調整をすれば、このシステムは中国語の質問に対して正しい答えを与えるであろう。」(Searle1990)

## 7 第3ラウンド 「システム論」批判はもう使えない

「中国語のジム」では部屋の中で多くの人々が忙しく動き回っているので、これに対して「システム論」による批判が加えられた場合には、もはやかつてのような反論は有効ではないように思われる。すでに見たように、「中国語の部屋」の場合、マニュアルやデータ入力やデータ出力を一人の人の頭の中に「押し込める」ことによって、全体としての「中国語の部屋」を誰か特定の人と同一視できることをサルは示そうとした。そしてそのことによって、彼は「システム論」が有効でないことを示そうとしたのである。ところが、今度の場合には、誰か一人の人の頭の中にマニュアルやデータ入力やデータ出力を「押し込める」ことはできない。「押し込める」のはせいぜい自分の遂行する作業に関する規則や入出力データにすぎない。したがって、その人が中国語を解さないという事実は、部屋全体が中国語を解さないということを何ら含意しないだろう。また、仮に一人の人の頭の中にマニュアルやデータ入力やデータ出

力を「押し込める」ことができた場合でも、その人は自分が関与しない作業のための規則を（意味もわからずに）記憶しただけであって、中国語の理解にとって不可欠な多くの作業を彼自身は遂行せず、彼以外の人々が遂行するという事情は変わらない。そして、「システム論」はその一連の作業全体が中国語を理解するということなのだ、と主張するのであるから、部屋全体が中国語を解すにもかかわらず、その人が中国語を解さないという事実は、何も不思議なことではない。というのは、その人（の行うこと）とその一群の人々（が行っている共同作業）が違うことはほとんど自明だからである。

8 CRAの射程(3) 人工知能研究や機能主義に対するものなのか、それとも行動主義に対するものなのか？

仮にCRAの標的には「中国語のジム」のようなものも含まれていたとしよう。するとそのときには、CRAの射程にはコネクショニズムも含まれることになる。この場合、CRAの構造は次のようなものと思われる。

（古典的なシステムであれコネクショニズムのシステムであれ）人工的なシステムの働き方は入出力の関係によって定義される。

心は入出力の関係によって定義されない。

故に、人工的なシステムは心をもたない。

この理解によれば、対話の相手がコンピュータであるのか人間であるのかを当の対話者が識別できなければ、コンピュータにも会話能力があるとしてよい、というのがチューリング・テストの通常解釈であるので、CRAの標的はチューリング・テスト（の有効性）ということになる。

「システム論」に対するサールの反論が正しければ、たしかにCRAは成功しているように見える。しかし、これでは「ロボット論」をブロックすることはできない。実際、サールを待つまでもなく、パトナムがかつて指摘したように、チューリング・テストは、その本来の形で理解するかぎり、妥当ではない。その一つの理由は、たとえば、コンピュータ 古典的なシステムであろうとコネクショニストのシステムであろうと が人間と区別できないような対話をすることができたとしても、それは指示をしないからである（5）。ところが、そのシステムが外部世界を認識できるような感覚器官に対応する機器と接続され、また外部世界に働きかけることができるような手足に対応する機器と接続されたとしたならば、その新しい拡大されたシステム というよりはロボット は、指示をすることができると考えられるからである。

しかし、すでに見たようにサールは「ロボット論」を有効とは認めない。（彼の議論が成功したかどうかは今は問わない。）したがって、サールのCRAは上のようなも

のではないかもしれない。少なくとも、そこで言及されるシステムはロボットのようなものも含むものとして理解されなければならない。すると、CRAの標的は行動主義ということになりそうである。

#### 9 CRAの射程(4) クオリア論との関係

「システム論」を批判する際、サールはデータの入出力作業や自分が遂行すべき作業を指示するマニュアルを(暗記するなどして)自分の頭の中に「押し込め」としても、その人(サール)が中国語を理解できないという事情は変わらないと主張していた。しかし、彼はとにかく、不思議なことに、中国語で話しかけられると、自分では知らないうちに、中国語で答えるのである。このことは、彼が中国語を理解しているという実感無しに中国語を理解する、という可能性を示唆する(6)。人格の分裂論はまさにこの点を追求したものであった。しかし、見方を変えれば、これはCRAの真の標的がクオリア不要論であったことを示唆する。つまり、「中国語を理解しているという実感」をクオリアになぞらえて考えるなら、「中国語の部屋」はクオリア無しの中国語理解者なのである。この場合には、CRAは次のようなものとして定式化できるであろう。

クオリアにとって入出力関係は十分ではない。

人工的なシステムの働きは入出力関係によって定義される。

心はクオリアをもつ。

故に、いかなる人工的システムもそれだけでは心にとって十分ではない。

今度の場合も、システムは、古典的であると接続主義のそれであるとを問わない。

#### 10 「統語論的性質は本来的性質ではない」という主張

Searle1992によれば、認知科学の背後にある基本的仮定は、脳は(一種の)コンピュータであり、心的過程は計算過程である、というものである(198)。(サールによれば)この二番目の仮定は、心にとって本質的なのはコンピュータ・プログラムだ、と言い換えられるので、古典的AI研究は認知科学の中心的位置を占めることになる。そしてCRAはもっぱらこの二番目の仮定を攻撃するものであった。CRAが登場して以来、その有効性にもっぱら議論が集中したために、この第一の仮定についての検討が表に出てくることはなかったが、サールの中ではもともと、心的過程は計算過程ではないというのは事柄のうちのたかだか半分にすぎなかった。彼によれば、第一の仮定も誤りなのである。Searle1992ではそのための論証がようやく前面に出てくる。

サールの分析では、第一の仮定の背後には次のような、より一般的な仮定がある。

すなわち、

- (1) 脳がデジタル・コンピュータであるという見解と競合する唯一の見解はある種の二元論である。
- (2) 脳過程が計算過程であるかどうかということは経験的な問題である。
- (3) コンピュータということで念頭に置くものは、抽象的なチューリング・マシンのようなものである。

以下では(2)と(3)について検討しよう。まず(3)から。これは基本的には機能主義者の仮定であり、実際のコンピュータは何らかの素材で作られているわけであるが、どのような材料で作られていても、そのようなことは問題ではなく、「計算論的に等価」(computationally equivalent)であれば同じコンピュータとみなしうる、というように表現されることもある。「多重実現可能性」(multiple realizability)という表現が用いられる場合もある。これに対してサールは二つの点で異議を唱える(208-9)。

まず第一に、この原理を認めると、どんなものでも何らかの計算をしているとみなしうるので、すべてがコンピュータになってしまうであろう。

第二に、そもそも、統語論は物理学に本来的な(intrinsic)性質ではない。

まず第一の論点から見ていこう。「緑の葉が本来的に光合成を行い、心臓が本来的に血液を送り出す」というような意味で、脳は本来的にデジタル・コンピュータであるかどうか」を彼は疑う。脳についてのどのような事実が脳をデジタル・コンピュータにするのか、と問うているときに、「すべてがデジタル・コンピュータなので、脳もデジタル・コンピュータである」というのは答えにならないとサールは主張する。(これに対して、葉や心臓の場合には、それらがもつ物理的特性の故に光合成をしたり、血液を送り出すと言える、と彼が考えていることは明白である。)

しかし、ここには誤解があるように思われる。機能主義者といえども、「すべてがデジタル・コンピュータなので、脳もデジタル・コンピュータである」と考えているわけではない。サールは計算過程と解釈できるということと、計算過程を実行(implement)ないし実現(realize)しているということとを区別していないようであるが、脳は、少なくとも表象を計算論的に処理しているのである。それはけっして、水が最短距離をとって下方へ流れていくのとは同じではないのである。たとえ、最短距離を計算して下方へ流れていくかのように見えたとしても。ここでは「表象」およびその「処理」ということがきわめて重要な役割を演ずるのであるが、サールはそのことを忘れて見えるように見える。もっとも、「表象」やその「処理」という概念自体、けっして自明ではなく、とりわけ古典的AIとコネクショニズムとの間の論争ではそれがキー・ポイントになる(7)のであるが、ここではそれに触れる必要はないだろう

う。

第二の論点は第一のそれに関連している。そしてまた、これは先にあげた(2)の仮定とも関連している。サールが「表象」のことを忘れたのは決して偶然ではない。というのは、彼にとっては、表象の処理とは結局のところ、少なくともそれが古典的AIで考えているような表象であるかぎり、統語論的処理を受けるものであるが、そのようなものは観察者に相対的な(observer relative)概念にすぎない。それは外部からの解釈に依存するのである。「計算とかアルゴリズムとかプログラムというような概念はシステムの本来的な物理的特徴の名前ではない。」(210) 彼によれば、ある過程を計算過程として特徴づけるということは、ある物理系の外部からの特徴づけにほかならない。そのような特徴づけは本質的に観察者に相対的なものである。これに対して、「質量」とか「引力」とか「分子」というような表現は本来的な特徴を名指している。「観察者がたとえ存在しなくなったとしても、世界には質量や引力や分子などは存在し続けるであろう。」ところが、「ピクニック日和」とか「バスタブ」とか「椅子」などのような表現は実在の本来的な特徴の名前ではない。適当な使用者や観察者が存在しなかったならば、ピクニック日和やバスタブや椅子などといった特徴は存在しないのである(211)。

しかしながら、この「本来的性質」、「観察者相対性」、「観察者」、「使用者」ということでサールが何を意味しているのかは実はあまり明らかではない。たとえば、「椅子」が観察者相対的概念の例としてあげられていた。椅子として使用する者ないしは椅子として使用されるものとして見る者がいなければ椅子などとは言えない(あるいは言ったとしても意味がない)ということなのであろうか。しかし、もし「として見る」ということを厳密に解するなら、いかなる性質も観察者相対的と言わざるを得ない。観察者がいなければ、何であれ、それについて何かを語ることは意味をなさないからである。しかし、サールはこの意味で観察者相対性を語っているわけではない。なぜなら、彼によれば、質量や引力などは観察者相対的ではないからである。では、「として見る」をどの程度ゆるく解すればいいのだろうか。心臓を、血液を送り出すものとして特徴づけることは、観察者相対的なのだろうか。サールは、ある箇所では、血液を送り出すという心臓の働きはその本来的性質であるということを示唆していた(208)。また、ある物理系を分子として特徴づけることも、その本来的性質についての話であると考えている(211)。その根拠を明白に述べてはいないが、それらのシステムの働きを適切に記述、説明するには観察者(生理学者や物理学者)はこれらをそのような系として捉えなければならないからではなからうか。ところが、もしそうであるとすると、脳の働きを適切に記述、説明するのに質量や引力などについて語るだけでは十分ではなく、神経回路や表象の処理というような語り方をすることが不可欠であるかもしれない。そうであれば、脳は情報処理システムであるとか、脳は統語論的処理を行っているというものは観察者相対的とは言えず、またそれ

らは脳の本来の性質ではないとも主張できなくなるであろう。

この問題と関連して、サールはいわゆる「小人問題」(homunculus problem)に言及している。心臓と脳をパラレルに語ることはできないとサールは考えているようであるが、その一つの根拠は、後者が「小人問題」に苦しむということである。「脳がデジタル・コンピュータであると考えれば、われわれは「では、そのユーザーは誰か?」という問題に直面する。」しかし、「小人」は本当に存在しなければならないのだろうか。サールは、高次のレベルでユーザーを想定しなければならないとすれば、低次のレベルでもそうである、と主張する(214)が、その論証は行っていない。また、「小人問題」は本当に心臓の場合には生じないのであるだろうか。血液を循環させるポンプとして使用するユーザーは誰なのかとなぜ問えないのだろうか。おそらく問うことは可能であろう。そしてその答えは明白である。その心臓をもつ動物なり人間なりがそのユーザーである、というのがそれである。同様に、脳をデジタル・コンピュータとして使用するユーザーは、当然のことながら、その脳を所有する動物なり人間であると答えることができるはずなのである。

#### 1.1 サールにおける還元主義的傾向とその問題点

統語論(的説明)が本来の性質(による説明)ではないとサールが考えるもう一つの根拠(と思われるもの)は、統語論は因果的効力を持たない、というものである。「実行されたプログラムは、そのプログラムを実行する媒体が因果的効力をもつということを別にすると、いかなる因果的効力ももたない。というのは、プログラムは実在しないからである...。」(215)おそらく、認知科学者はこのサールの主張に異議を唱えはしないであろう。しかし問題は、これが統語論的説明が本来の性質による認知過程の説明ではないとする根拠になるかという点である。サールのいわんとすることは、脳はコンピュータのように振舞い、それが認知を引き起こすと言われるが、「[計算を実行する]小人がいないので、コンピュータも脳もパターンを示しているだけである」(216)ということであろう。しかし、先に指摘した「小人問題」を取り去るならば、ここで問題となっているのは、統語論的説明は不可欠かということと、脳をコンピュータに見立てることは単なるコンピュータ・シミュレーション以上のものかということである。サールは、いずれに対しても否定的に答えるのであるが、最初の問いに対する彼の答えはある種の還元主義を前提しているように思われる。しかし、もしそうであるならば、心臓のもつ「血液を送り出す」という性質や「分子構造」なども本来の性質とは呼べなくなるであろう。二番目の問いに対する答えについては、サールはシミュレーションとそうでないものの区別が明白であるかのように考えているが、その区別はそれほど明白ではない。たとえば、飛行機の風洞実験はシミュレーションなのだろうか、それとも、「本物」(の縮小版)なのだろうか。以下においてこの二点についてもう少し詳しく述べよう。まず本節では第一点に

ついて考察する。

プログラムに言及することによってコンピュータの振舞いを説明することができるのと同じ意味において脳の振舞いを説明できるとする認知科学者の立場を批判するにあたって、サールは、「脳のいかなる機能についてであれ、それについてのこの種の説明は、それが神経生物学的レベルにおいてどのように働くのかということが実際に理解されているところでは、けっして受け入れられることはない」(217)と主張する。彼はレットヴィン他の論文を引用して、カエルが昆虫を捕える行動を説明するのにカエルの「虫検出器」のコンピュータ・シミュレーションを与えることはできるかもしれないが、誰もそのようなことはしないのであって、研究者らは神経生物学的レベルにおける説明を与えるのだと述べる。そしてまた、そのような神経生物学的レベルにおける説明はしかるべきプログラムのハードウェアでの実行にすぎないなどとは言わない、とも述べている(218)。要するに、サールによれば、ひとたびカエルの視覚システムが実際にどのように働くかということが理解されたならば、「計算のレベル」におけるカエルの昆虫認知の説明はまったく無関係(irrelevant)なのである。

しかしながら、カエルの視覚システムを研究している現場の生物学者が「神経生物学的レベルにおける説明はしかるべきプログラムのハードウェアでの実行にすぎないなどとは言わない」としても、そのことは必ずしも「神経生物学的レベルにおける説明はしかるべきプログラムのハードウェアでの実行にすぎない」ということを否定するものではない。また、仮にこの議論が正しく、したがって統語論が本来的性質でなく、観察者相対的な性質であるとすれば、まったく同じ議論によって、分子構造や血液を送り出すという心臓の機能なども本来的性質ではなくなってしまうように思われる。というのは、分子や心臓の振舞い説明するのに、物理学者の中には、究極的には原子やそれより小さな粒子の振舞いに言及しなければならぬと述べたり、それより上のレベルでの説明のもつ価値を低く見たり、あるいは分子構造の話や血液を送り出すというような語り方はヒューリスティックな価値を持つにすぎないと考えたりする者もいないわけではないからである。ところが、すでに見たように、サールは分子であるとか血液を送り出すという性質を本来的性質の例としてあげていたのである。

## 1.2 因果的効力について

次に、シミュレーションか複製かという問題に移ろう。ハリケーンの動きをシミュレートするプログラムができたからといって、ハリケーンの振舞いをそのプログラムによって因果的に説明できると主張する人はいないであろう。ところが、なぜか認知科学者は脳過程の説明の場合にはその種の主張をする。これはどうしたことなのだろうか、とサールはいぶかる(218)。サールの診断では、ここには、コンピュータがアルゴリズムを実行するというのと同じ意味において、脳が認知のための

アルゴリズムを実行しているかもしれない、という誤った仮定があるのである  
(219)。コンピュータがアルゴリズムを実行するというのがコンピュータの振舞いの因果的説明を与えるかに見えるのはなぜなのだろうか。少し長くなるが、サールの診断書を見てみよう。「[ テューリング・マシンを実行する ] 人間コンピュータは意識的にアルゴリズムの各ステップを経験する。したがって、その過程は因果的であると同時に論理的でもある。すなわち、そのアルゴリズムが入力記号から出力記号を導出するための一群の規則を提供するが故に、論理的であり、行為者がそのステップを経験するために意識的な努力をしているが故に、因果的である。機械のコンピュータの場合、全体としての作動しているシステムには外部の小人 [ 人間のユーザー ] が含まれている。そして、この小人がいるので、そのシステムは因果的であると同時に論理的でもある。すなわち、その小人がその機械の過程に対して解釈を与えるが故に、論理的であり、その機械のハードウェアがその過程を引き起こすが故に、因果的である。」(219-220) ところが、「脳の生の、盲目的な、非意識的な (brute, blind, nonconscious) 神経生理学的操作はこれらの条件を満たすことができない。脳コンピュータの場合には、人間コンピュータの場合に存在するような、アルゴリズムの意識的で、志向的な実行などというものはない。また、機械のコンピュータの場合に存在するような、アルゴリズムの非意識的な実行などもない。なぜなら、それを行うためには、物理的出来事に対して計算論的解釈を与える外の小人が必要だからである。脳の中にわれわれが見出しうるのはせいぜい、機械のコンピュータにおいて実行されるプログラムと形式的に類似した出来事のパターンにすぎない。しかし、そのようなパターンはそれだけではいかなる因果的効力ももたないのである。」(220)

この議論がどのような種類のものかということはいまだ明らかではないが、最後の文から示唆されるのは、脳過程の場合、そこに何らかの統語論的パターンが見出せたとしても、それを解釈するもの、すなわち「小人」が存在しないので、そのようなパターンはそれ自体としては因果的効力をもたない、という議論である。ところが、そのように解釈すると、脳過程に統語論的パターンを見出す研究者を解釈者として想定することが可能であり、したがって、脳とコンピュータを区別する本質的理由はなくなってしまう。おそらくこれはサールが意図していたことではなかろう。よりありそうな解釈は、何らかの統語論的パターンが因果的効力をもつのは、それが何らかの意識を持つ主体 それは通常の間人である場合もあれば、「小人」である場合もある

がその統語論的解釈に関与した場合である、というテーゼをサールが主張している、というものである。このテーゼはそれ自体として非常に興味深いものであるが、ここでは、サールはこのテーゼを擁護する議論を十分に展開しているようには見えない、ということを指摘するにとどめたい。

なお、このテーゼを擁護するために CRA をここで援用したらどうか、ということが考えられるが、これは見込みがないように思われる。CRA を当面の問題に適用できる

ようにアレンジしたものは次のようなものであろう。(ただし、次の定式化における「心」は、上の議論で登場した「意識」と置き換えることができるものとする。)

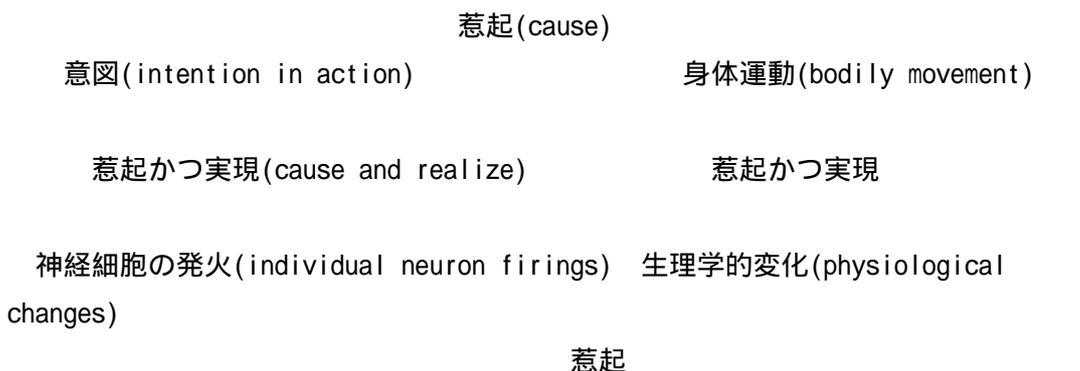
統語論は(それだけでは)因果的効力をもつのに十分ではない。

脳過程に見られるパターンはもっぱら形式的ないし統語論的な構造によって規定される。

心は因果的効力をもつ。

故に、脳過程に見られるパターンはそれだけでは心にとって十分ではない。

二番目の仮定については問題はないであろう。なぜなら、サールも認知科学者もこれは認めているからである。三番目の仮定は一般には議論の余地が大いにあるが、サールがこの仮定を受け入れるというのはそれほど不自然ではない。というのは、彼はある意味で「心身因果」を認めるからである(8)。たとえば、彼は次のような図式を受け入れている(Searle1983: 270)。



ここでは心的現象のうちの一つである意図が例に挙げられているが、これを意図に限定する必然性はなく、この図式を心的現象一般に拡張することは自然である。さて、問題なのは、一番目の仮定である。サールはこの仮定を受け入れるが、いまわれわれが論証しなければならないのはまさにこの仮定なので、ここの文脈に合うようにアレンジされた CRA を用いたのでは、論点先取になってしまうのである。

われわれは10節から12節にわたって長々とサールの議論を吟味してきたが、以上のことから、「統語論的性質は本来の性質ではない」というサールの主張は、かなり特殊な見解を採用しないかぎり受け入れがたい、と結論できるのではなからうか。

ところで、10から12節までの議論においては、われわれはもっぱら古典的AI研究での統語論を念頭に置いて考察を加えてきたが、接続主義についてはどうであろうか。接続主義で考えているような表象の処理の場合に、サールの元

の議論がそのまま妥当するかどうかは不明である。しかし、コネクショニストも、認知過程においては表象が使用され、処理過程において、古典的 AI での計算とは本質的に異なる仕方においてはああるが、その表象が変換されると考えているようなので、サールの主張のポイントが、表象の処理というのは本来的な性質ではない、ということであるならば、それはコネクショニズムにも適用されることになる。したがってまた、われわれの考察も、そのように理解されたサールの議論に適用できるはずである。

統語論を本来的な性質と見ないサールの立場は、ある意味では道具主義的に解釈されたデネットの立場に近いが、AI 研究や認知科学に対する二人の見解には大きな隔りがある。この違いが何に由来するのかということは興味深い問題であるが、これについては別の機会に改めて考えてみたい。

結びにかえて

CRA はサールの意図とは裏腹に、AI 研究（強い AI）の誤りを示すことに成功はしていないが、「意味とは何か」、「志向性とは何か」という問題は狭い意味での統語論的アプローチだけでは解決されないということを、CRA をめぐるその後の一連の議論は明らかにしたと言えるであろう。また、言語理解の場面ではしばしば「クオリアとは何か」、「それは言語理解にとっていかなる役割を果たすか」という問題が忘れ去られがちであるが、それは言語理解にとっても重要な役割を演じているかもしれないということを、CRA をめぐる議論は明らかにしたように思われる。さらにまた、サールの議論とそれに続く批判者たちの議論の展開は、サールの議論が古典的 AI 研究やコネクショニズムに基づく AI 研究や認知科学研究一般をも射程に入れたものであることを明らかにした。そして、彼の議論の説得力に問題はあるものの、それは心の科学に対するわれわれの安易な方法論に警鐘をならすものとして理解されなければならないであろう。最後に、サールは統語論的性質が本来的な性質であるかどうかという問題を提起したが、これは伝統的な心身問題に新たな方面から一石を投じたものと評価できるのではないだろうか。

注

(\*) 本稿のアイディアの骨子は、2000年 - 2002年にかけての科学研究費補助金（基盤研究（B）（1））による研究（課題番号 12410003）の報告書に掲載される予定の『『中国語の部屋』とその後』と題する論文においてその素描を与えられている。

(1) 彼自身は「イミテーション・ゲーム」という表現を用いていた。Turing1950を参照。なお、この論文は Boden1990 にも収められている。

(2) Searle1980 は Boden1990 にも収められている。本稿においては、この論文に言

及するときには後者における頁数を示すことにする。

(3)「システム論」と「ロボット論」についてのより詳しい検討は服部 1990 においてなされている。

(4) コンピュータが知能をもつ(あるいはより一般的に、心をもつ)とする立場は強い AI と呼ばれる。これに対して、コンピュータは知能や心をシミュレートできるとする立場は弱い AI と呼ばれる。サールは後者には反対しない。

(5) p トナム 1994、13 頁以下参照。

(6) Block2002、74 頁も参照。

(7) 服部 2003 を参照。

(8) ただし、言うまでもないことだが、サールは心身二元論者ではない。

## 文献

Block, N., 'Searle's Arguments against Cognitive Science', in Preston and Bishop 2002

Boden, M.A.(ed.), The Philosophy of Artificial Intelligence, OUP 1990

Dennett, D.C., The Intentional Stance, MIT 1987

Preston, J. and Bishop, M.(eds.), Views into the Chinese Room, Clarendon Press 2002

Haugeland, J., 'Syntax, Semantics, Physics', in Preston and Bishop 2002

服部裕幸、「心に関する機能主義の射程」、『愛知教育大学研究報告』、39、1990

服部裕幸、「コネクショニズムとは何か」、『科学哲学』、33-2、2000

服部裕幸、「『分散表象』は認知の説明にはたして役立つのか？」、戸田山和久他編、『心の科学と哲学』、昭和堂(近刊)

ホフスタッター&デネット編、『マインズ・アイ』、TBS ブリタニカ 1992

Putnam, H., Reason, Truth and History, CUP 1981

パトナム、『理性・真理・歴史』、法政大学出版局 1994

Searle, J.R., 'Minds, Brains and Programs', The Behavioral and Brain Sciences 3, 1980 (also in Boden 1990)

Searle, J.R., Intentionality, CUP 1983

Searle, J.R., Minds, Brains and Science, BBC 1984

サール、『心・脳・科学』、岩波書店 1993

Searle, J.R., 'Is the Brain's Mind a Computer Program?', Scientific American, Jan. 1990

Searle, J.R., The Rediscovery of the Mind, MIT 1992

Turing, A., 'Computing Machinery and Intelligence', Mind, Vol.LIX, No. 236, 1950