

Article

アーティクル

深層脳内情報処理から学ぶもの —機能脳科学の観点から—

Learning from Deep-Level Cognitive Processing — A Suggestion from Functional Brain Science —

苔米地 英人
Hideto Tomabechi

コグニティブリサーチラボラトリーズ株式会社
Cognitive Research Laboratories, Inc.
tomabechi@crl.co.jp, http://www.crl.co.jp

1. 深層脳機能と人工知能

人工知能研究は、1970年代頃から本格的に始まり、特に70年代後半から80年代の、いわゆるアメリカのAIブーム時に飛躍的に発展した。しかし90年代に入ってから、人工知能の基本的なパラダイムにおけるめざましいブレークスルー的学説は、あまり出現していないというのが実感である。

実際90年代に入ってから、人工知能学者が、分子生物学や機能脳科学などにその専門分野を移行したケースは少なくなかった。人工知能の分野はもはや、学説的なブレークスルーが期待できないと思われたからであろう。

しかし分子生物学の分野において、計算機による大量な情報処理を行う必要が発生しつつあり、計算機科学への学者の関心が再び高まっている。また機能MRIについても、MRI画像をデジタルデータ処理した後に、さまざまなアルゴリズムによって再処理が行われるようになり、そこで計算機科学者の力が要求されている。人工知能学会に、分子生物学や機能脳科学の研究をする人たちが増えてきたのも、最近の需要による結果であろう。

もともと人工知能は、認知科学と表裏一体に進歩してきた。認知科学はその定義上、認知をファンクショナルな集合体として捉えるファンクショナルリズムをドグマとしている。そのファンクション自体が、もともと記号的なものであり、そのシンボルを操作できるシステムをつくり上げることが、1970年代から80年代半ばまでの人工知能研究の傾向であった。

1980年代半ば以降は、記号的処理が限界に達し、ニューラルネットワーク

などの統計的な手法の重要性がにわかに高まり、シンボリックなAIだけでなくサブシンボリックなAIも、人工知能研究の主要なテーマとなった。

サブシンボリックな研究が進行するにつれ、電気信号や化学反応のレベルから知能を解明したい学者たちが、MRI研究の学者たちと結びつき、いわゆるファンクショナルMRI (fMRI) を利用した知能研究が行われ始めたのが、1990年代初頭である。

1991年11月発行「サイエンス」誌の表紙に、ハーバード大学医学部グループのfMRIによる脳の活性化した視覚野の画像が掲載されたことも、まだ記憶に新しい。

その後の90年代は特に、fMRIを利用した脳機能の局所的研究や、その地図作り(ブレインマッピング)が、3テスラレベルの強いマグネットを使って世界中で盛んに行われ、局所レベルでは詳細な部分まで解明が進んだ。また研究に伴い、いくつかの不思議な現象も確認された。

例えば、瞑想や催眠中に強い光などをイメージすると、視覚野が活性化するという報告がなされている。これは、1980年代に提唱されたジェリー・フォード(J. A. Fodor)らによるモジュラリティ仮説の発想[Fodor 85]からは、説明し難い現象である。

モジュラリティ的な発想では、脳機能の各部位はモジュールに局所化され、それは情報を下位から上位のモジュールに上げていくといった関係であり、上位から下位へのフィードバック関係はあっても、上位レベルにおける情報が下位レベルの情報を生成するという現象は起こり得ない。

fMRIによる発見までは、例えば催眠や瞑想中の幻覚や光のイメージは、第46野などにおける、ワーキングメモリを利用した意識化された活動が、海馬などを利用した記憶の引き起こしと結びつけられて発生する現象であろうと考えられてきた。

つまり、V1からV5にあたるような初期認識野の役割は、あくまでも視神経からの入力信号を、色、形、動きなどから認識し、その情報を前頭葉といった脳内の各機能野に伝達することである。これらの各機能部位の並列的な処理の最終的な統合場所である、第46野などの部位で起こる現象が、逆にごく初期の認識や信号を生成するということは、当初のモジュラリティ的な枠組みでは考えられなかった。

例えば、ビデオカメラで画像を受け取り、それをデッキに録画し、モニターに再生するという枠組みにおいて、ビデオカメラのレンズが眼にあたり、レンズの後ろにあるCCDが視覚野に相当し、ビデオデッキが海馬を利用した記憶メカニズムに相当し、モニターが第46野に代表されるワーキングメモリと仮定する。

これまで、催眠や瞑想で認識されるような光や幻覚は、この回路における何らかの誤作動で誤った信号が合成され、モニターに映ったものと考えられていた。もしくはワーキングメモリを利用した意識的な活動により、存在していない光を、デッキからの情報などを利用してモニターに生成する、つまり海馬などを利用して、前頭連合野で認識するというメカニズムで説明されていた。

しかし瞑想中、幻の光を見ている状

態で視覚野が活性化しているならば、モニターに誤作動で存在していない光が映っている状況に例えるなら、レンズのすぐ後ろにある CCD で、信号が発火しているということになる。つまり、脳内の前頭連合野の抽象化された情報が、初期認識野の神経の活性化を促しているのである。

また運動野との関係においても、行動を起こそうとした場合、前頭葉が活性化する前に、運動野が活性化するという現象も確認されている。[澤口 00]によれば、猿が「自由意思」で実際に手を動かす 1～2 秒前から活動が高まるニューロンが、補足運動野で多く発見されたという。これも動作を命令する内省的な意識の発生以前に、すでに運動そのものが開始していることになる。

また、運動をイメージするだけで、運動野が活性化するという報告もなされている。これは、抽象度の高い情報空間における情報処理が、物理的レベルに近い抽象度の低い情報空間における情報処理に影響を与えていることを示している。

生物学の世界でも最近、治ろうとする強力な意思が、統計的に有意なレベルでガン細胞の活性化を押さえたのではないかという報告があった。この事例も、脳内の前頭連合野において、治りたいという抽象的な思考活動が、物理世界に近い分子レベルに情報伝達されたことによって引き起こされた現象であるといえないだろうか。

人工知能学の情報処理は、70～80年代においては、抽象化された記号処理レベルで行われており、いわゆる内部表現と呼ばれる心的モデル、認知モデルも、記号の情報のみを扱っていた。しかし現在、人間の情報処理を模倣しようとするならば、遺伝子レベルから極めて抽象度の高い情報空間まで、統合的に全方向の情報伝達がなされるモデルをベースにしなければならない状況になっている。

一時、分子生物学や機能脳科学に流れた人工知能学者たちが、最近になって戻っている事実も、こういった傾向と無関係ではない。また、学習、記憶、

概念、言語といった各分野での研究が、認知科学と人工知能の学問領域の中で多くのモデルとなっており、そのいくつかは、解明された脳内情報処理のパターンによって説明されている。一方で、仮想空間の臨場感など、従来の学説では説明の難しいモデルがいくつか出現しているのも事実である。

2. 物理空間と情報空間に共通する プリンシプルの存在

この章では、人工知能という学問分野にとらわれず、物理空間と情報空間の両方に対して、ファンクショナルなアプローチを試みたい。

近代以降、物理空間におけるさまざまな理論が発見されたおかげで、科学技術は大きな発展を遂げた。量子力学に基づいて電子レンジがつけられ、ニュートン力学によって月までアポロが安全に飛行し、最近では、相対論誤差を利用するような宇宙開発も進んでいる。ただしこういった文明の機器をつくりだした理論の方程式は、物理空間に限定された定理を説明しているに過ぎない。

私は物理空間だけでなく、情報空間にも共通に通用するプリンシプルがあるのではないかと考えている。

a. 唯識論という情報空間のプリンシプル

例えば、仏教哲学の古くからある概念に、唯識がある。[中村 98]によると、唯識説は、インドでマイトレーヤ(弥勒菩薩)が説いた思想が源流となっており、それを三蔵法師が勉強し、中国に持ち帰って『成唯識論』に集約した。この論は後に法相宗という宗派を中心に日本に伝わっている。唯識とは、万物すべての存在の実体は「空」であるが、対象を分別し、知る「識」の力でそれらがあるものとされているという教えである。識には、六識すなわち眼、耳、鼻、舌、身、意があり、心のさらに奥に自我執着心と自我意識である、末那識、阿頼耶識が存在する。

唯識論は、唯心論とは異なっている。唯心論では、心に事物が映っていないければその存在はないものとされるが、唯識論では事物が心に映っていないと

も、実体は空であるがそれらは存在しているにとらえられている。例として、富士山を考えてみる。唯心論的発想では、富士山を全く知らない人にとって、富士山の存在はないと同じである。しかし唯識論では、たまたまその人の識の範囲にないだけで、富士山は、その実体は空であるが存在しているとされる。

唯識論は、ある個人の認識について考察していない。唯識的意味での抽象度が究極まで高まった状態を指す空は、どちらかという人工知能学者や現代分析哲学者が用いる[]の記号に似ている。つまり完全に自由な変数である。すべての情報は[]との単一化に成功するので、あらゆる情報と無矛盾であるということから、潜在的にすべての情報を含有しているともいえる。[]は、唯識論における空と同様であると考えることができるだろう。

機能脳科学の立場から考えると、唯識論は、物理空間のプリンシプルがすべて情報空間のプリンシプルから発生していると解釈している。私なりに考えた現代的な唯識論の解釈では、例えば波動方程式といった物理世界の定理を記述する記号を、情報空間の何らかのプリンシプルに準じたものであるとしている。ある物理現象は、ある現象段階のプリンシプルに基づいており、現象の抽象度が上がれば上がるほど、それを記述する記号の情報空間のプリンシプルは高階なものとなり、プリンシプルは階層的に対象と対応しているとしている。つまり、抽象度の階層的集合体として宇宙空間をとらえている。いってみれば唯識論では、宇宙空間にははじめに情報空間があり、物理空間はその情報空間のルール上、たまたま発現している状態に過ぎないとみているのである。

機能脳科学自体の学問領域においては、物理空間、情報空間という区別にとらわれず、情報の抽象度という見地からプリンシプルを探っていく。すなわち抽象度が低ければ物理世界に近く、高ければ情報世界に近いということである。最も抽象度が低いのは、現実的な物理存在そのものであり、非常に高

度な抽象空間は、複雑な思考空間などが例としてあげられる。つまり、物理空間と情報空間との差は、単に抽象度の差に過ぎないとみている。

これは実のところ、唯識の発想に近い。唯識の場合も抽象度は、色即是空、空即是色といわれるように、物理世界の物と情報が単純に表裏一体として同じものであるとされている。

法相宗から現代仏教までの伝統的な唯識の解釈では、物理世界を心に映った情報世界に過ぎないとしてきたことから、現在でも学者たちは唯識を唯心論的なものとしてとらえている。しかしさらに現代的な機能脳科学的な視点から唯識をとらえると、抽象度が0の物理空間から高次の抽象空間までの全範囲における表現論的なプリンシプルとすることができる。そのことを応用して解釈すると、あらゆる抽象度に存在する情報状態において、何らかのプリンシプルが存在しているという見解が、もともと唯識の発想の中にあるといえる。

古くから東洋にある法相や古来学僧が考えた唯心的な解釈を超えて、源流の唯識論を現代的な機能科学的立場から見直したとき、情報空間と物理空間両方に通用するプリンシプルがあってもよいととらえることができるのである。

b. チョムスキーの宇宙視野のプリンシプル

言語学者チョムスキーは、言語情報に関して、情報空間のプリンシプルの存在をユニバーサルグラマーという学説を通じて主張した。彼は、GB 理論 (Government and binding theory: 統率束縛理論) [Chomsky 82], Barriers (障壁理論) [Chomsky 86], PP (Principles-and-parameters approach: 原理とパラメータ) [Chomsky 95], などに基づいたユニバーサルグラマーという、宇宙共通のプリンシプルに基づいた文法があると考えた。その文法に対するパラメータのチューニングが、たまたま日本語、あるいは英語に過ぎないのである。さらにその説を応用的に解釈すると、人類が地球上で進化に失敗し、唸り声や嬌声しか発すること

ができないとしても、本来言語現象そのものはあるとみなすことができる。

チョムスキーはユニバーサルグラマーのプリンシプルの前提として、言語クラスの階層性をチョムスキーヒエラルキーとして定義した。後にその考え方は、情報問題の複雑性に関する理論であるコンプレキシティ論 (計算量の複雑性理論) に発展していく。加えてチョムスキーは、言語現象そのものも抽象度でヒエラルキー化することができることを発見している。

彼のいうクラス分けとは、言語をその表現能力に応じて分類したもので、Regular language, Context-free language, Context-sensitive language, Recursively enumerable language というヒエラルキーをつくり、さらにその言語の文法と、実際記述するための計算機のクラスをも定義した。例えば Context-free language は、Context-free grammar (文脈自由文法, 2型文法) クラスの文法をもち、Nondeterministic pushdown automaton (非決定的プッシュダウンオートマトン) が、その言語を解釈するマシンとされている。ちなみに Recursively enumerable language は、無制限文法 (0型文法) クラスとなり、Turing machine が解釈のマシンとして必要になる。

こういった情報のクラス分けはその後、1970年代後半から80年代にかけてさかんになったコンプレキシティ論にまで発展した。これは物理空間における熱力学の第一法則と同じように、明らかに情報宇宙に必ず通用する一つのプリンシプルである。情報世界にも階層性があることを指摘した点で、斬新な説だといえる。

チョムスキーにある面では影響を受け、ある面では独自に発展したコンプレキシティ論は、もともと情報空間だけの理論としてつくられている。しかし計算機あるいはチップが物理世界のルールで動いているので、計算量が複雑になればなるほど、コンピュータが長時間稼動しなくてはならないといったように、情報空間のプリンシプルが物理空間にも影響を与えるケースが証明できる。これは離散数理やコンプレ

キシティ論のなかで、注目すべき点であろう。もともと物理世界だけの話を前提として作られ、その世界だけで完結している物理方程式とは、本質的に違っているのである。

言語現象において、最初からチョムスキーは完全に情報空間の現象だけを扱っており、そこからチョムスキーヒエラルキーを考え出した。それが後のコンプレキシティ論に強い影響を与え、分析に計算機を用いることで情報空間の制約が物理空間にも通用することが、明確にされたわけだ。相対性理論で物理世界の常識を覆したアインシュタインより少し先を行ったのがチョムスキーだと言ったら、ほめすぎだろうか。

これと似たものに、音階や協和音がある。純粋に数学的なバランスが物理的周波数と対応し、我々が脳の聴覚野からそれを受け取ると心地よくなったり、不快感を覚えたりする。音楽も情報と物理空間にまたがるプリンシプルであり、ある意味でベートーベンやモーツァルトもチョムスキーレベルの人物であったとみることもできる。

c. 人工知能分野におけるプリンシプルの探索

しかしながら人工知能という学問分野において、いくら素晴らしいプリンシプルを発見しても、最後は記号によって記述するしか、それを説明する手立てはない。宇宙共通のプリンシプルはけっして記号化できないはずはないが、その記号は非常にメタフィジカルな性質を帯びているに違いない。例えば、物理学の波動方程式などを考えてみると、もともとその式は記号世界の原理ではなく、物理世界の原理を非常に抽象化し、記号化したものである。過去の人工知能の学問分野においては、そこまで抽象化した記号は使われておらず、わかりやすい関数が扱われているに過ぎない。そういった意味で、人工知能分野における高度に抽象化された記号が、そろそろ出現してもよいのではないかと思っている。それには記号化するシステムから、構築しなくてはいけないのかもしれない。

生物においてこの原理は、恒常性維持機能 (ホメオスタシス機能)、あるいは

は恒常性を維持させるような、自我と環境のフィードバック機能として働いていると考える。このことは、人工知能や深層心理を考察するうえで、どのような関係があるのか。

人工知能はその定義上、ファンクショナルなサイエンスである。ファンクショナルであるためには、すべての理論を記号化しなくては意味がない。人工知能学においては、人間の知能情報を抽出し、記号化し、その記述すなわちファンクションを裏付けるマシンを作り上げることが研究命題であろう。

理論の記号化が前提なので、人工知能研究の立場から深層心理を考えたとき、それは無秩序な、完全に分散化されたような非記号を、記号化していく作業であるといえる。フロイトのいうトラウマのような深層心理に埋まっているものをとらえようとするとき、その抽象度が高く言葉や論理となって表出しているものであれば記号化は進めやすく、単に信号状態に近い、曖昧な、抽象度が低いものであれば、作業は困難である。

深層心理だけでなく、脳内情報処理の解析についても同じことがいえる。脳内情報を化学的レベル、あるいは物理的な電気信号としてとらえたとき、抽象度を下げればフォークのレベルにまで考察が及ぶ。どのレベルまで抽象度を下げるかにもよるが、同じ物理情報のレベルでも、モルキュラーな分子生物レベルで分析するのか、電気レベルで解析するのかによって、記述の抽象度における差異が生じる。しかしそこからさらに情報を付加していけば、脳が作り出す心という現象にまで到達する。

まとめるなら、一番抽象度の低い世界は、波動方程式あるいは十次元のヒモ理論のレベル、抽象度が上がると、物理から化学といった分子生物レベル、あるいは電気信号レベルの世界に、さらにもう一段階上がると、無秩序な一般的に無意識と呼ばれる非記号的な信号状態の情報世界、あるいは準記号的なイメージ世界に、そしてさらに抽象度が上がると、やっと記号となるのである。同じ記号でも「山」といった抽

象度の低い記号から「唯識」、「美」、「愛」などの抽象度の高い記号がある。究極的には「空」や「無」といった概念に至るのだろう。

脳の物理的な解析の手法は現在、ブレインマッピングが主流である。しかし私はそのやり方について、あまり成果を期待していない。物理的な意味での局所的観察が意味をなさないというわけではなく、マッピングしても人間にしか通用しないものであれば、そこから、あらゆる宇宙世界共通の原理を導き出すのは難しいと考えるからである。人間と同じ脳のつくりが10万年前の猿にあったとしても、地球以外の星の生物と比較して脳の言語野が左右違っていたら、たまたまこの地球上で生命が発生し、その進化の結果ブレインマッピングどおりの脳になったにすぎず、それが宇宙の定理とはいえない。

人工知能学的視点から深層意識を考えたとき、脳内現象と全身に散らばる神経細胞や心といった対象を違う抽象度で認識すれば、その抽象度によって対象を表す表現は違ってくる。脳あるいは心は記述の抽象度によってさまざまな表現方法があるわけで、脳と心はどこで切り離されるかという議論には、ほとんど意味がない。過去の人工知能は、シンボリックな領域に限られ、記号化ということにこだわりすぎていた感がある。人工知能よりもニューラルネットワークが一時復権したのは、準記号レベルでかなり研究成果を上げた結果であると考えられる。

今、心の分析という側面だけで人間の認知が理解できている学者はいない。トータルに人間の脳に迫るには準記号、非記号のレベルまで解明する必要があるだろう。心の概念は、記述の抽象度の差でしかないというのは、現代的な機能脳科学を経験した後の人工知能学者の本音ではないか。そのうえで精神空間と物理空間を統一するような原理があれば、大変おもしろいことになる。

結論的に、あらゆることは抽象度の差異によってその性質が決定されると私は考えている。それを前提としたう

えで、その抽象度のレベルの差異を超えた、すべての抽象度に通用するような宇宙の原理があるのではないだろうか。物理空間にも精神世界にも当てはまる原理ならば、それによって両世界を統一することも可能である。また片方の世界から片方に影響を与えることもできる。念じたらものが動いたという超常現象とされていた世界の話も、物理空間が精神世界に感応することもあり得ると仮定すれば、不思議な話ではないだろう。逆に、物理的作用が精神に影響を与えることも考えられる。

言葉を交わして意思が伝わるという現象も、よく分析すれば不可解である。もともと純粹に情動的な心に音声は振動として影響を与えることで、複雑な意思を疎通させることができる。まさに物理世界と情報世界が互いに密接に共鳴している例だといえよう。

3. グラウンディング問題<心と体のつながるところ>

グラウンディング問題は、1980年代の半ば頃から、ニューラルネットワークと記号処理の関わりを中心に研究していた、元イェール大学のAIプロジェクトメンバーで、当時UCLAの助教授であったマイケル・ダイア(Dyer, M. G.) [Dyer 88, Dyer 90]によって主に提唱されていた問題である。

一般にグラウンディング問題とは、人工知能の分野では、なんらかの記号がニューラルネットワークのような準記号的、もしくは非記号的な情報状態において、どのように表現されているかという問題であった。伝統的にストロングAIと呼ばれている人工知能は、すべて記号のみで人間の認知を解明し、記述することを前提としているので、そこにグラウンディング問題が生じる余地はない。ストロングAIにおいては、ファンクショナリズムが純粹に信じられ、記号処理を行う対象が人間の脳であっても、算盤、ノイマン型コンピュータであっても全く関係がないので、記号状態がハードウェア上でどのような情報状態であるかということは、問題

とならない。

一方現代的な人工知能の考え方では、あくまでその知能処理を行うハードウェアの性能がその記号処理レベルまで強い影響を与えると考えられているので、必然的にグラウンディング問題が問われることになる。

例えばニューラルネットを考えてみよう。ここでいうニューラルネットとは、伝統的なバックプロパゲーションネットワーク (BPNET) のことだが、リカレント型のニューラルネットに、時間ごとに順番に単語レベルで文章を入力し、訓練の終わったネットワークをクラスタ分析してみると、文法の雲のようなものが出現している。その雲は、名詞、動詞、壊れやすいもの、壊れにくいものといったような、単語間の構造をもった依存関係がネットワーク上に生まれたもので、そのことはエルマン (Elman, J. L.) の研究 [Elman 92] などで当時知られていた。

その後ニューラルネットのクラスタ分析を進めるうちに、個々のニューロンが特定の文法的素性を認識する役割を果たすことなどが示されてきた。さらに脳神経科学が認知科学の強い影響を受け、解剖学的研究から機能科学的研究へその学問領域が波及したとき、1991年11月の「サイエンス」誌の表紙を飾ったことでも知られるハーバード大学での機能MRI (fMRI) の研究成果により、脳内での情報処理のグラウンディング問題も注目されるようになった。

現在では、グラウンディング問題といえ、人間の脳内で記号処理がどのように行われているかという問題を指すことが多いだろう。もちろん人工神経回路網における同様な問題も、人類の認知解明の足がかりになるべきものだと考えている。

グラウンディング問題の現在の機能脳科学における解明方法は、fMRI (核磁気共鳴断層撮影法) や PET (Positron Emission Tomography: 陽電子放射断層撮影法) を使った脳内機能活性時の局所マッピング研究に留まっており、実際脳内の神経レベルでの活性化状態を直接的にすばやく十分な解像度で解

析できる段階には、まだほど遠い。このような技術の開発は、読者である若手研究者にぜひ挑戦してもらいたい研究テーマである。ハーバードグループのfMRIを超えるような、脳内状態のグラウンディング問題にまで示唆を与えるような解明ツールができることを期待したい。

グラウンディング問題とは、つまるところ人間の心と体の繋がる場所はどこか、またどのように繋がっているかの研究である。このような説明の仕方をする、伝統的なデカルト的二元論の立場の延長線上ととられがちであるが、あくまでこのようなテーマのモチベーションとなっているベースは、脳内の情報処理と、認知上の情報処理が分離された世界に存在する実体という伝統的な立場でなく、あくまでも脳内情報処理と認知処理は一体であり、その差異は、我々科学者の記述における抽象度の差異に過ぎないという発想である。2章でも詳しく述べているが、物理状態を低い抽象度とし、情報状態を高い抽象度としたとき、ある事象は、それを表す記述方法に応じて、連続的な抽象度における、あるレベルのフィルタを通じて語られているにすぎない。この場合グラウンディング問題の本質を一度突き詰めて考えてみる必要がある。

つまりグラウンディング問題そのものも、実は我々が精神と脳内情報処理の解明をどのような抽象度で記述するかによって、問題そのものが定義されることになるのである。BP ネットを始めとする人工神経回路網の研究の伝統から、我々はたまたまグラウンディング問題をネットワークの統計的遷移状態に記号化されるまでの、グルーピングされた情報状態における問題としているが、実際は、グラウンディング問題を異なる抽象度でとらえることも可能である。例えば分子生物レベルでとらえることもできれば、極めて抽象度を上げて形而上的なレベルでとらえることもできる。

このように見ていくと、グラウンディング問題とは、我々の存在そのものを二つ以上の異なる抽象度で複合的に

眺めるための窓であり、また我々が慣れた記述の手法、例えば論理式で不慣れな情報状態を記述するための道具でもある。まだまだこの問題をテーマとしている研究者は少ないが、私は一つの人工知能学におけるブレイクスルーの重要な窓であると考えている。

4. グラウンディング問題の双方向性

これまで、機能脳科学、分子生物学、グラウンディング問題と、それにまつわるストロング AI (記号的 AI) や、ニューラルネットワークを利用した人工知能研究の一連の流れを説明してきたが、ここでグラウンディング問題を実際の生命検証の問題としてとらえたとき、興味深い事実が発見されていることを紹介したい。それは、グラウンディングは双方向的であるということである。

1章でも触れたが、ジェリー・フォード (Fodor, J. A.) が説いたモジュラリティ仮説などの従来の学説では、脳内の情報処理は完全に下位から上位へのヒエラルキーにモジュール化されており、情報は下位から上位への一方向のみに伝えられ、上位から下位への影響は、極めて少ないフィードバック現象のみであると考えられていた。しかしこれに矛盾する現象が、fMRI による研究でいくつか発見されているのである。

代表的な例としては1章でも述べたが、変性意識化された催眠状態で何らかの強烈な光のイメージを思い浮かべる (幻視する) と、人間の後頭部にある視覚野で神経回路網の活性化が確認できることが報告されている。この結果は、モジュラリティ仮説と矛盾する。幻視について改めて双方向性の観点から説明すると、モジュラリティ的発想では、視覚野はビデオカメラの CCD のような初期認識野であり、そこからの情報は前頭連合野へ一方向的に発信されるだけで、前頭連合野での認識が視覚野に影響を与えることはないと言われていた。しかし、変性意識状態が第46野、あるいはその他の脳のどの部位で起こっているかはいまだ解明されていないが、言語的誘導などによって言

語野や何らかのワーキングメモリ上の情報処理が、認知に深い影響を与えていることは間違いない。

幻視による視覚野での発火という本質的にモジュラリティ仮説と矛盾する現象は、より高位な脳の部位から、初期認識野に対して強い影響を与えられているということを物語っている。これ以外にも、高位な認知活動が、物理状態の抽象度に近い生態レベルにまで強い影響を与えていることが知られている。

一般に我々のホメオスタシスフィードバック [苫米地 94] が、思考やイメージから強い影響を受けていることは、有史以前から人類にとっては当然のことであった。例えば恐ろしいことを考えると、心臓がドキドキするというのは、まさにグラウンディングの双方向性を表す現象の一つであろう。

ここでは、本稿の記述の便宜上あくまでも精神的な認知空間と生態的な情報空間を別個のものとしてとらえる伝統的な立場の説明に従って述べているので、グラウンディングに双方向性があることが矛盾しているように思われるかもしれない。しかし2章でも述べたように、認知的な情報空間と、生態的な情報空間は本来同じもので、その違いは抽象度の差異にすぎないという考え方に照らし合わせれば、このようにグラウンディングに双方向性があることは、想像に難くない。このことは、ダマジオ (Damasio, A. R.) も、脳の視覚野における双方向性を著書の中で図式化し、説明している [Damasio 94]。

抽象度の差異という見地からグラウンディングの双方向性について述べてきたわけだが、機能脳科学の分野に留まらず、双方向性はデカルト二元論における情報空間と物理空間や、言語学における意味と構造についても当てはまる現象である。フッサール、モンテグューにおける伝統的文法理論では、個々の単語を積み重ねた構造が単調的に文全体の意味をつくり上げるという二元論的な考え方がベースとされていた。しかし、一つの現象を異なる抽象度から解釈する現代認知科学的な見地

から眺めると、文全体の意味を理解しないと、個々の単語の品詞を決められないという逆説的な現象が確認されている。

心理学の分野においても、ゲシュタルト現象において双方向性が確認されている。あるものについて、全体の認識が同時もしくは先になされていないと、個別要素の認識ができないというものである。これは、一方向的な部分からの情報が全体に伝達され事象が認知されるというモジュラリティ的発想では、説明できない。ただ私は、モジュラリティ的内部構造の存在を否定しているわけではなく、もともとモジュラリティ的な階層そのものが実は、さらに抽象度の高い空間におけるヒエラルキーを反映しているのではないかと考える。

これまでのモジュラリティ的な考え方がまだ根強く、AIや機能脳科学に影響を与えているので、私はあえてここで、これを問題にしたのである。

解剖学的脳科学にファンクショナルリズムを導入した脳科学を機能脳科学と呼ぶのと同様、ファンクショナルリズムを導入した分子生物学を、仮に機能分子生物学と呼んでみる。将来的に分子生物学がこのような機能分子生物学へ拡大していけば、同じようなグラウンディング問題が、また繰り返し研究、あるいは討論のテーマとなることが予想される。脳科学の分野で発見されてきたように、例えば思考や思想といった抽象度の高い情報が、神経回路網の物理状態に強い影響を及ぼしているのと同様に、思考思想といった情報状態が分子生物レベルの情報状態に強い影響を与えているという発見が、多く報告されることになるだろう。

例えば、強く治ろうとする意思がウイルスを消滅させ、信心深い人が祈りでガンを克服するといったことが、奇跡ではなく、グラウンディング問題の窓を通して、生体にとっては当然の現象であるといった説明がなされることも考えられなくはない。逆に、そのようなメカニズムが解明されれば、大胆に言えば人類は意思の力で自らを自由に一代で進化させることが可能に

なるかもしれない。

もしかすると、進化は単なる統計的淘汰の結果ではなく、生物の生きようとする強い意思が生体に分子生物レベルで影響を与えた結果引き起こされているのかもしれない。それ以外のはるかに包括的な進化論が、今後生み出されることも考えられる。いずれにしろ、我々が機能脳科学と情報処理を通して学んできた成果をベースに類推すれば、分子生物レベルでも、このような精神状態と分子生物状態の深い関わりが解明される可能性が高い。なぜなら、もともと我々の立場では、分子生物状態も同じ人間=生命存在という抽象的な記述の一つに過ぎず、生命現象は全抽象度の世界で同時に起きていることは間違いないからである。

5. 内部表現における抽象度の階層性

心象を表す用語に、内部表現 (internal representation) と呼ばれるものがある。人工知能の学問領域では、1960年代のキリアン (Quillian, M. R.) による研究 [Quillian 69] の時代から、そういった内省的な意識を、記号の論理式や意味ネットワークといったプログラムやデータ構造で表現している。シャंक (Shank, R. C.) らのスクリプト、MOP理論や、現在一般的な動的オブジェクトによるクラスヒエラルキーなどは、そういった内部表現の一例である。

もちろん記号化されたものがすべてというわけではなく、内部表現はあらゆる抽象度で表現することができ、心理学では自我、生物学では生体、宗教的には魂と呼ばれているものが、それにあたる。

a. 自我、脳内機能と内部表現

例えば、自我について考えてみる。我々個人が自分を内省的に意識するとき、あるいは外界を意識するとき、その意識そのものが内部表現であるといえる。自分とは何かという命題を与えられたとき、ある人は思考したり、ある人は物理的に握っているボールペンの感触を確認したりする。このように、人は遺伝的に、あるいは受けた教育水準や経験に従って、それぞれ内省的な

認識に最適な抽象度を、無意識のうちを選んで、好きなボールペンを選ぶにしても、その人の内部表現の抽象度に応じて決められるのである。例えば物理世界、すなわち本能的な世界に近い内部表現をもつ人は直感的に色を選び、抽象度が高く記号的な内部表現をもつ人は、論理的に考慮し、仕事の内容に最適な機能を備えたペンを選ぶであろう。

人間はもともと保持している内部表現の抽象度の範囲内において、変化する環境に応じて、自分にとって最も外側の世界を認識しやすい抽象度を選択している。個々人を形成する自我そのものは、あらゆる抽象度で流動的に表現されている。また個性という概念も、観察者が他人を見て感じる内部表現の抽象度の違いによって生み出された副産物にすぎない。認識する側と認識される側の関係において特定の抽象度が決められ、個性とされるものが内部表現に記述されるだけなのである。だから個性とは本来、その瞬間ごとに変化する、観察者によっても左右される、流動的なものであるといえよう。

つきつめれば、自我や個性というところらえ方も、言葉の記述の差異に過ぎない。外部から判断された個性を自我が認知し、自分自身の内部モデルをつかって自我がそれに合わせていくという循環が生じる場合もあるので、自我と個性とは切り離すことができない概念であるともいえる。極論するなら、我々が通常永続的な固有存在としての自我をとらえているものは、存在しない。自我とは、神経細胞レベルにおいて、外の世界を認識し自力でシンボルを操作できる範囲内で選んだ抽象度における、自分自身のある時点でのモデルに過ぎない。

自我=内部表現の様相の傾向は、人類の進化に応じ変化している。言葉をもたなかった原始人は、動物的な本能が強いので、自我は現代人よりはるかに物理世界に近い抽象度である。現代において哲学者は、非常にシンボリックなロジックで物事を認識するので、その内部表現の抽象度は高階である。

人工知能は、こういった内部表現を

科学者が最も操作しやすい抽象度の記号で表した姿である。シンボリックな内部表現をもつロボットができれば、定義上は自我が存在するといえよう。近い将来ペットロボットなどが、電池が切れようとするとき恐怖心が走るといった反応を示すようになるかもしれない。

内部表現そのものは、脳の中で一体どのように表現されているのか。それはゲノム上で、人間の器官や機能がどのように表現されているかが解明されていないように、本質的に未知の領域である。視覚野がどの位置にあるかなど、局所化されている事実は判明しているが、器官の各機能が実際どのようにニューラルネットワーク上にエンコードされているかはわかっていない。それについてはさまざまな解釈があるが、一つの例として、[Tomabechi 91]において、脳内の脳波はFM放送におけるベースフリクエンシイのようなもので、我々はその上に情報を周波数変調して記述することで情報処理しており、脳内全体が一つのホログラフィックな空間であると説いたことがある。その仮説が、FMNN (Frequency Modulation Neural Network) である。

こういった機能脳科学と分子生物学におけるゲノムとの関係において、あるいは脳機能の位置を把握する局所化研究と、解剖学的なメカニズムの解明過程において、機能がどうエンコードされているのか、その仕組みを抽象的な記述によって解読することが、今後の我々の研究命題ではないだろうか。

内部表現には分散性がある。例えば、ある人物像を思い浮かべようとしたとき、同じ視覚情報でも、色、動き、明るさといった条件の記憶は、脳のそれぞれ違う部位に電気的表現配列として保存されており、その情報が、連合野などにある収束ゾーン [養老 99] に集まり、イメージが再現される。記憶がしまわれる部位とは、長期記憶は側頭葉、手順記憶は被殻、無意識の記憶は扁桃核、本能的記憶は尾状核などがあげられる。過去の記憶イメージは正確に再現されるわけではない [中島 00]。分散された情報が一度に起動しなければ、

例えばある人物について顔はわかるが、声が思い出せないという状況に陥る。つまりそれは、当時 V1 から V5 に伝達された視覚野からの情報は保存状態が良く、収束ゾーンによって円滑に起動され、聴覚野からの情報は、記憶の保存がなされなかったか、うまく起動しなかったことによって生じたのである。この場合、声は仮のイメージで代用される。

この脳内情報処理を 2 章で説明した抽象度の階層性を見地からとらえるなら、記憶などの内部表現は、抽象度が下がると音声やセル画、シナリオが書かれた脚本のように分散し、抽象度が上がると一つの映画のような画像として再現されるのである。

b. デプログラミングにおける内部表現の操作

個人のパーソナリティ、あるいは個性や性格は、内部表現において、ある特定の抽象度で記述されている認知パターンであるといえる。「洗脳原論 [苦米地 00a]」にも記したデプログラミングの手法は、最適な抽象度に標準を合わせ、内部表現=自我の操作、あるいは書換えを行っていく作業である。

最適な抽象度とは、デプログラマーにとって最も操作しやすい抽象度のことで、言語的操作と視覚的操作が両方同時に行えるような、言語的抽象度よりも少し低く、生体的抽象度よりは少し高いレベルを指す。ただしその範囲内においても、被洗脳者のパーソナリティによってデプログラマーは、さらに細かい段階の抽象度とその性質を選択する必要がある。ある人が論理的で思索家ならば、その内部表現の記述の抽象度は高く、本能的で身体能力に優れ直観型の人物ならば抽象度は低い。絵画にたとえるなら、前者はカディンスキーのような点や線だけの現代美術、後者はミレーのような、自然主義的風景のような内部表現をもっているといえ、わかりやすいだろうか。

ここでお断りしておくが、デプログラミングという手法は反社会的なテロ組織のメンバーについて行われるべきもので、けっしてみだりに個人に対し濫用されるべきではない。私の紹介する例

は、この鉄則に準じた状況において行われたものであることを明記しておく。

Lisp のプログラムを作るとき内部表現を最も書きやすい抽象度の記号やラムダ式で表そうとするのと同じように、デプログラミング中は、被洗脳者の言葉や動作を注意深く観察し、最もグラウンディングに働きかけやすい抽象レベルがどこか、探索する。それを決定したなら、そのレベルからの働きかけを、さまざまな手法で徹底的に行う。

デプログラミングの際の抽象度を定めるポイントをあげれば、ミュージシャンのような音に敏感な人には声からアプローチし、画家といった視覚が発達した人には、ジェスチャーが有効であろう。また信仰している宗教の流派によっても、あるものは身体的であったり、あるものは論理的であったりするので、やはりそれぞれの教義の特徴に合わせて、アプローチの方法を選択する。

抽象レベルを絞って接近すると、相手の意識がより変性しやすく、ホメオスタシスフィードバックが起りやすくなり、接近のきっかけとなったものがグラウンディングの窓となっていく。

信者へアプローチする初期の段階で、働きかける抽象レベルのセグメントを間違えると、デプログラムに失敗する確率は高まる。例えば音楽好きで、聴覚野がグラウンディングと密接に結びついている人物に、いくらパズルのように論理的な哲学ディベートを仕掛けても、全く興味を示さない。かえってそれをしたことで警戒心が高まり、私を見てもグラウンディングは、なんら機能しなくなってしまう。最悪の場合、私の声に対し不快感や嫌悪感を抱く、マイナスのホメオスタシスフィードバック [苦米地 94] が作動することさえあるだろう。

だから、デプログラミングは初対面のときが最も大事なのである。細かい言葉や動作から、相手がどのレベルの内部表現をもった人物か、なるべく早く見極めなくてはならない。そのため私の場合は、相手が話したことを全部記憶し、空間に場所を決めて一つ一つその内容をイメージ的に貼り付け、

整理し、作戦を立てる。

デプログラマーとしての私が最も操作しやすい抽象度は、記号的抽象世界より少し低い、画像イメージ世界のレベルである。そのレベルは言葉で介入できる世界なので、記号での操作は可能であるが、厳密には記号レベルより少し身体感覚に近い。抽象度が決まれば、人工知能研究の場合と同じように、内部表現を観察し、そこから全体像を描きやすくなる。

描く方法としては、紙に書きながらコミュニケーションを進めるケースもあるが、実際は目の前の空間に受信したイメージをジェスチャーで示していく。[苦米地 00a] で紹介したケースでは、空中に現実世界と教義世界の二つの円を描き、それを一つにしてみせることで、現実世界に対する信者の臨場感を回復させる試みをした。

Lisp のプログラムを書くとき、最初の 1 行から最後の 2 万行目までの構成をあらかじめビジュアルに細部までイメージしながらキーボードをたたくように、被洗脳者が宗教的抽象世界でなく現実世界に生きているという実感が湧くよう、計画的に内部表現を操作し、現実世界におけるホメオスタシスフィードバック機能の稼働を促進していく。

ホメオスタシスフィードバック (恒常性維持機能) とは、卑近な例で説明するなら、寒いときは毛穴を調整して体を温め、暑ければ発汗して体温を下げるといった、無意識レベルでの生体と環境のフィードバック関係を言い表したものであるが、生体 [苦米地 93] とは、全抽象度における生体を指す。物理的意味での生体だけでなく、記号的抽象空間における生体も、その範疇に入る。例えば瞑想による仮想体験のようなイメージ世界の記憶の情報や、論理的な考え方のパターンといったものも、生体としてとらえる。その生体と、外側の環境情報が合わさり、ホメオスタシスフィードバック現象が生まれるわけである。このことを人工知能的観点からとらえるなら、内部表現と環境とのフィードバック関係ともいえよう。

この現象を利用してカルトは信者に薬物などを投与し、例えば教祖のマン

トラを耳で聞くと快感が走るよう、グラウンディングのセッティングをする。

ホメオスタシス機能には、興味深い現象がある。それは内部表現を書き換えると、認知パターンの変化によって外界の感じ方が変わり、その異なった環境からの刺激のフィードバックによって、さらに内部表現が変わるという機能である。

また、一つの抽象度で内部表現の操作を行うと、別の抽象レベルで整合的な動きが起こる。ある抽象度で書換えが成功すれば、その書き換えられたデータはあらゆる抽象度に反映される。催眠現象は、その好例である。

例えば、右手が軽いという情報を内部表現に書き込むと、右手が軽くなった気分になる。そして実際写経をしていた右手が上がったなら、字が書けなくなるので物理世界に影響が出る。さらに右手が下ろせない状況が続けば、それは生体の身体感覚だけでなく、記憶から、抽象度の高い信仰心、宗教観にも影響を及ぼし、内部表現を書き換えていく。

これは 2 章に述べた情報世界と物理世界に通用するプリンシプルの存在を解明する、鍵となるような現象ともいえるのではないか。

またこの論文において、これまでグラウンディング問題を人工知能研究者の目からとらえてきたが、主体である自我にとってのグラウンディングの窓という視点から考えれば、複合的な抽象度を統一する窓であるともいえる。どこに窓があるかは個人差があり、ミュージシャンなら音が、画家なら色がその窓となる。例えば、音は単に物理的な要素に過ぎないが、それは音楽が好きな人の感情や思考、イメージにまで大きな影響を及ぼす。音に敏感な人が悲しい曲を聞けば、コミカルな漫画を読んでも、曲にまつわる悲しい思い出が一緒にしてまぶたの裏によみがえり、自然に涙が出てくる。そしてそのことは、高い抽象度の思考にも影響を与える。

デプログラマーはそういった、できるだけ多くの抽象度に対して働きかけることができるグラウンディングの窓

を、短時間で見つけることが肝要である。できるなら二つ以上の抽象レベルから書換えを行っていけば、成功率は高まる。

「グラウンディングの窓」は、[苦米地 00a] で述べた、エリクソン派の用語である「アンカー」と同義語である。アンカーは簡潔に説明するなら、記憶や教義といった抽象情報と、具体的な行為、あるいは感情が密接に結びついた状態を指す。絵や音楽といったトリガーに対するアンカーの作動状況を細かく観察していけば、その人のグラウンディングパターンがより正確に把握できる。

抽象度という問題とは別に、グラウンディングの窓、すなわちアンカーという見地から [苦米地 00a] を読み直すと、認知科学に詳しい人なら、また違った楽しみ方ができるに違いない。ここでは、オウムのイニシエーションなどについて述べているが、今や洗脳は、薬物をいっさい使用することなく、言葉による催眠などで容易に行える時代になった。オウムは「CIA 洗脳実験室 [苦米地 00b]」に描かれているような、1950年代から60年代に開発された、薬物を使用する行動療法的な、サイキックドライビングなどといった古い技術を採用していたに過ぎない。

現在では、洗脳という形式にとらわれず、さらに直接的な内部表現の書換えができる技術が開発されている一方、人工知能や脳科学の研究成果によって内部表現の分析が進み、知識と経験があれば、抽象度を合わせるだけで、パーソナリティを復元する書換作業ができるようになった。まさに洗脳者とプログラマーとのいたちごっこが、続いているわけである。こうした現場においてこそ、物理空間と情報空間におけるプリンシプルの発見が急がれるべきではないか。

c. 内部表現におけるファイアウォールの構築

この項では、純粋な状態での内部表現を悪影響からブロックする方法について考えていきたい。詳しくは将来出版する予定の「続・洗脳原論(仮称)」に記すつもりなので、ここではその基

本となる原理だけ公開する。

一般の人が洗脳されないための防御法は、ウイルスからパソコンをプロテクトする方法に似ている。具体的には、ファイアウォールのようなものを認知の中に築くことにより、悪意の第三者による介入的な内部表現の操作を防御する。このファイアウォールは外部からの情報の侵入を阻止するだけでなく、ウイルスのような、内面に巣くう情報的で破壊的な活動をする存在から自我を防御する役割も果たす。

例えば、我々の認知の解明に将来人工知能学者が成功したとすると、身体が有機物であるか無機物であるかに関係なく、一つの巨大な情報処理システムが完成することになる。高度な情報処理システムであれば、必ずいつかはハッカーの手によって、それを破壊もしくは制御不能にするウイルスが発明されることは、間違いないだろう。

ウイルスは、どのように認知から認知へと感染していくのか。まず考えられるのは、現在の典型的なコンピュータウイルスのように、直接的な情報のやりとりによって自我に感染するタイプである。あるいは遺伝子に貼り付き、親から子へと鎖のように感染していく潜伏期間の長いウイルスも開発される可能性がある。

当然そこまで認知の解明が進めば、人工知能だけでなく、これらと同じような有機物でできた我々人間の脳内情報処理を破壊するウイルスを、人間の手で発明することもできる。それが分子生物学的に合成されたものなのか、単に情報のみの実体なのか、私にも予想がつかない。

しかしそこまで高度な人工知能が出現しなくても、認知へ介入するウイルスをつくることは、実はたやすいことなのである。Windows 2000 は、天才級のエンジニアが多数含まれた、何百人、何千人の専門家の手によって開発された。そのレベルの技術を可能にする会社は、世界でマイクロソフトや、IBM、SUN など数社しかない。ところが Windows 2000 を壊すプログラムは、たった一人の高校生によっても書けてしまう。ソフトを作成する能力が

なくても、主要機能のメカニズムがある程度わかれば、セキュリティが万全な全く新しいプログラムを作成するより、破壊ウイルスの開発ははるかに容易なのである。

この事実から、心のファイアウォールを構築させることができる人物は、プログラマー、あるいは洗脳者以上の能力が要求される。ただ緻密な理論を構築するだけでなく、すべてのあらゆるアタックの方法について、防御メカニズムを用意しなくてはならない。一方、破壊的ウイルスを作るだけなら、それほど難しいことはないはずである。

心の破壊ウイルスは、ごく最近出現し、現代に蔓延しているものだけでなく、過去ある天才によって作成され、場合によっては全人類がすでに犯されているものもあるだろう。あるいは生得的な遺伝ではなく、情報的な伝達で、世界中に飛散しているかもしれない。

科学技術が加速度的に進歩する未来において、構築された情報世界のプリンシプルが悪意の第三者に利用され、生体の内部表現をさらにたやすく破壊するウイルスが新たに生み出される危険は、十分考慮されるべきである。

それでは具体的に、こういった破壊的ウイルスから自我を防御するには、どうしたらよいか。一つはそういった外部からの侵入を、阻止するという方法である。もう一つは、いったん入ったウイルスを、ワクチンとしての機能を備えたファイアウォールが攻撃し、消滅させるという方法である。

侵入を防ぐには、例えばオウム級の洗脳法であれば、認知側が不用意に第三者の介入によって変性意識状態に導かれない方法で対抗する。または変性意識状態に陥っても、内部表現の書換えが行われようという瞬間に、それを意識化させる方法を身につけておけばよい。それ以外にもいくつかの方法があり、私が開発したテクニックもあるが、時期をみて別の機会に記したい。

一方、侵入したウイルスに対するワクチン的な防御法であるが、これはいわゆる情報処理における免疫システムのことである。あらかじめ何らかの情報を脳内に埋め込んでおき、脳のホメ

オスタシスを破壊しようとする情報処理があると、それに対し攻撃を加え殲滅するのだが、これはまだ研究段階で、成功例はない。もしも成果があれば、いずれかの学会で公表したい。

あえて今考えている方法を述べるなら、人の認知には絶対譲れないプリンシプルが、幼少期からの教育によって埋め込まれている。例えば人を殺してはいけない、盗んではいけないといった価値観である。もしカルトからの洗脳を受け、この禁忌のプリンシプルに接触したとき、禁忌プリンシプルがワクチンとして作用し、カルトのゆがんだ価値観を駆逐する可能性はある。しかし禁忌のプリンシプルは戦争などが発生した場合、容易に壊されてしまうので、そのプリンシプルを我々の手で確固たるものにできるかという課題が残されている。

心のファイアウォールについて知りたい人は、洗脳に関する私の著作を読破されたい。洗脳を知ることで、似たような状況下に置かれたとき、警戒心を喚起することができる。しかしあくまでこの予防法はアラートシステムの段階に留まっており、これだけではファイアウォールまでの効果は期待できない。

カルトは宗教だけではない昨今、外部から自己の純粋な内部表現への介入が行われたとき、自らの選択で、心のファイアウォールを使って拒絶したり、書き換えられた個所を修復したり、書き換えようとする力を消滅させるという能動的なワクチンのアタックをあらかじめ自分の意識下に埋め込んでおくことは、重要な健康管理の一環である。

まちがえてはいけないのは、心のファイアウォールやワクチンは、個人の選択で内部表現にインストールされるものであり、全員の自我に強制的に構築されるものではないということだ。

6. バイオインフォマティクスからコグニティブバイオインフォマティクスへ

認知科学の成功と神経科学におけるfMRIの誕生により、脳科学にファン

クショナルリズムの考え方が持ち込まれ、機能脳科学が誕生したのと同様に、今後分子生物学にもその流れが押し寄せ、functional molecular biology、もしくはfunctional bio-science、functional genetic engineeringといった研究が盛んになることが予想される。

もちろん現在のいわゆるバイオインフォマティクスは、そのための土壌を十分我々に用意してくれている。ただし気をつけなければいけないのは、現在のバイオインフォマティクスは、機能脳科学と比較してみると、人工神経回路網の研究が進められた1980年代頃の方法に似通っていることである。脳科学において、分子生物学の代表的な素材であるゲノムのような、情報状態が凍結された神経の活性化状態のスライスというものは、人工回路網のクラスタとしてしか存在しておらず、学問のためであっても活性状態を保ったままの人間の脳を研究において用いることは不可能である。それに比べゲノムは、そのものの情報状態をエンコードしているので、分子生物学研究の方がはるかに機能研究を行いやすいはずだ。

もちろんこのことは、一つ一つのゲノムが単体で有意な情報状態を保っているという前提で話をしている。現在におけるバイオインフォマティクスの研究手法は、基本的にゲノム情報を高速に統計処理ならびに比較処理し、有意な情報ユニットを探し出すという手法である。これが成果をあげており、少なくともある特定個人のゲノム情報の分類化が終了し、3万~4万の遺伝子数の存在が確認されたと報告されている。もちろんこれからは、すべての人間対人間のゲノム比較、すなわち30億対30億のベクトルポジションの比較へとスケールが拡大するため、ケタ違いに大変な作業が待っているのだが。

重要なのは、これらが成功したとしても、まだマシンのOSのバイナリーが分類解析された状態にすぎず、これを終えてからようやく、バイナリーからOSを推測するという作業が始められるのである。WindowsやLinuxで試しても、おそらく我々の現在もって

いるツールでは、不可能とも思える作業が待っていることになる。それはゲノム上にエンコードされた抽象空間に存在する人間の認知能力を含めたOSを解析する膨大な作業である。

これこそ分子生物レベルでの表現の抽象度から、人間の生体機能レベルにおける情報処理のメカニズムを、グラウンディングの窓を通して推測する作業にほかならない。ここからがfunctional molecular biologyすなわち機能分子生物学が最も必要とされる領域である。そのような研究手法はfunctional bio-informaticsもしくは、cognitive bio-informaticsと呼ばばいいだろうか。

ファンクショナルリズムの考え方なくしては、例えばある特定の病気を遺伝子治療で治せたとき、こうすれば治るといふ臨床的な手法は発見できても、なぜ治るといふ基礎医学的な解明をすることは難しい。

つまり科学者には、機能分子生物学レベルで、まだまだ大きな仕事待ちを受けているのである。言い換えれば、ある個人のゲノムが解析されたことにより、これから先は人工知能を経験した計算機科学者の活躍が、分子生物学においても大いに期待できるといえる。

7. 次世代人工知能の可能性

次世代人工知能に期待される未来像は、大まかにいえば二つ考えられる。一つは人間の拡張機能として進化していくタイプ、もう一つは、人間に代わる生命を新たに創造することによって生み出されるタイプである。

前者については、私がハイパーセルフ化と呼んでいるタイプで、エージェント機能をさらに高度化したものである。エージェント機能とは、本人の代理人となる秘書、もしくは執事を情報空間(サイバースペース)に置き、それに対し自然言語やジェスチャーなどで何らかの指示を与えると、その知的なエージェントが本人になり代わり、あらゆる作業を行ってくれるというものである。

しかしこの本来の意味でのエージェ

ントタイプは、開発に時間がかかるであろう。なぜなら人間そのものの認知がまだ解明されておらず、さまざまな行為における人の意図を正確に理解できない現実において、場合によっては人の一生以上もの容量の仕事を手工知能に期待するのであるから、十分な試行実験を重ねなければ実現は困難である。

一方でハイパーセルフの考え方は、あくまでも人間の脳に対してなんらかのインタフェースを用意し、高度な抽象的思考活動は人間に任せ、簡単な思考活動を機械に任せようという発想である。例えば自転車や自動車、飛行機の発明は、人間の物理的な機能を拡大するという産業革命の究極的な理想の姿を体現した物であった。その後20世紀に入って発明された計算機は、人間の機能を物理空間でなく情報空間で拡大させることができるという点で、原始時代における火の発見に匹敵するような発明品であり、それは画期的な進歩を社会生活にもたらした。

現代社会における会計ソフトやワードプロセッサは、ある程度知的ではあるが、あまりクリエイティブでない仕事を自動化し、処理してくれている。これをもっと人間にとって自然なインタフェースでつなぎ、我々の意識活動に直結させることが実現すれば、まさにハイパーセルフのアーキテクチャーとなる。できれば何らかの方法で、脳と計算機を物理的につなぎ合わせることができれば、難しい抽象思考は人間が行い、多くの生活時間を消費しているあまり知的でない情報活動は計算機に任せるといったことも可能になる。このように人間の機能の一部を計算機が担うことが、次世代人工知能の進む一つの方向であろう。

その一方、人間という存在にまったくとらわれず、無機的な情報処理主体をベースにし、たんぱく質でできた我々人間の分子生物レベルまで機能と構造を模倣したうえで、全く異なる物性により新たに人工的な知能を産出することも、次世代人工知能のもう一つの可能性として考えられる。少なくとも現在のエンジニアリングのレベルで

は、我々が扱いやすい情報処理の素材はシリコンであり、銅線である。こういった我々が工学的に処理しやすい素材を利用したうえで、全く新しい人工知能を生み出すこともできる。ただし単純に記号レベルでの情報処理を模倣する過去のストロングAIとは異なり、脳神経回路網のレベルのみならず分子生物のレベルまでも模倣することを試みるのである。

おそらくあえてこの方向を意識しなくても、現在人間の外観と物理的機能の模倣に留まっているヒト型ロボットは、いずれ神経回路網や分子生物レベルまで踏み込んだ研究によって、より人間と融合性の高いものに仕上がっていくことが予想される。その結果、外見や声、会話の内容も人間そっくりであるが、物性は人間と全く異なる人工的な素材で合成された人工知能ロボットが生み出されるであろう。その後、この無機的な生命体ならではの進化を遂げたり、固有の感情も持つかもしれない。こういった方向性も次世代人工知能に期待される一つの可能性である。

どちらの方向性を選ぶかは、我々科学者一人一人の生き方と哲学の問題であろう。

◇ 参 考 文 献 ◇

[Chomsky 82] Noam Chomsky: Some Concepts and Consequences of the Theory of Government and Binding, The MIT Press Books (1982)
 [Chomsky 86] Noam Chomsky: Barriers, The MIT Press Books (1986)
 [Chomsky 95] Noam Chomsky: The Minimalist Program, The MIT Press Books (1995)
 [Damasio 94] アントニオ・R. ダマジオ原著, 田中三彦 訳: 生存する脳—心と脳と身体の神秘, 講談社, pp. 161 (2000)
 [Dyer 88] Michael G. Dyer: Symbolic neuroengineering for natural language processing, a multilevel research approach/UCLA Collections (1988)
 [Dyer 90] Michael G. Dyer: Symbol processing, connectionism, and artificial neural networks for high-level cognitive tasks, natural language research at the UCLA Artificial Intelligence Laboratory, UCLA Collections (1990)
 [Elman 92] Jeffrey L. Elman: Grammatical Structure and Distributed Representations, in Davis, S., ed., Connectionism: Theory and Practice, Vol. 3 in the series Vancouver Studies in Cognitive Science, Oxford

University Press (1992)
 [Fodor 85] Jerry A. Fodor 原著, 伊藤笏康, 信原幸弘 訳: 精神のモジュール形式 人工知能と心の哲学, 産業図書 (1985)
 [中島 00] 中島節夫: ウソの記憶と真実の記憶, 河出書房新社 (2000)
 [中村 98] 中村 元: 大乘仏典・完成期の大乘仏典, 東京書籍, pp. 157-188 (1998)
 [Quillian 69] M. R. Quillian: The teachable language comprehender, BBN Scientific Report 10 (1969)
 [澤口 00] 澤口俊之: わがままな脳, 筑摩書房, pp. 80-81 (2000)
 [Tomabechi 91] Hideto Tomabechi and Hiroaki Kitano: Beyond PDP: the Frequency Modulation Neural Network Architecture, In Proceedings of Eleventh International Joint Conference on Artificial Intelligence (1991)
 [苫米地 93] 苫米地英人: サイバーVR マルチモーダルシステム—自然言語・仮想現実・生体情報の融合 (1993), 徳島大学工学部研究報告第39号 (1994) 収録
 [苫米地 94] 苫米地英人: サイバーホメオスタシス仮説—マルチモーダリティの臨場感パラダイム, 統合とマルチモーダリティ・ワークショップ論文, 日本ソフトウェア科学会「言語と知能」研究会 (1994)
 [苫米地 00a] 苫米地英人: 洗脳原論, 春秋社 (2000)
 [苫米地 00b] 苫米地英人 訳, ハービー・ワインスタイン 原著: CIA 洗脳実験室, デジタルハリウッド出版局 (2000)
 [養老 99] 養老孟司 監修, リタ・カーター 原著, 藤井留美 訳: 脳と心の地形図—思考・感情・意識の深遠に向かつて, 原書房, pp. 196, 238 (1999)

2000年12月8日 受理

— プロフィール —



苫米地 英人
 1983年上智大学外国語学部英語学科卒業(言語学)。1993年カーネギー・メロン大学博士(Ph.D.計算言語学)。1996年通商産業省情報処理振興審議会専門委員着任。1998年コグニティブリサーチラボラトリーズ株式会社代表取締役社長に就任。現在、マルチメディア振興協会法務委員会委員、国際交流委員会委員兼任、文部省領域研究ゲノム情報科学高度化委員会委員。ACM, IEEE, 米国人工知能学会, 米国認知科学会など各会員。情報処理学会, 電子情報通信学会, 言語処理学会など各会員。