

メタファーに基づく人工知能

月本 洋
(株) 東芝 研究開発センター

本稿では、メタファーに基づく人工知能 (MBAI) を提案する。従来の人工知能の多くは結合主義も含めて記号主義、計算主義であるが、これらは知能を人工的に実現するのに成功していない。人工知能の不可能性の主な論拠は、身体性、状況依存性、志向性である。しかしこれらをソフトウェアで実現するのは困難であるか不可能である。そこで身体性に基づく有意味性の形式、即ち経験の形式をソフトウェアで実現することを考える。この経験の形式とはカントの言う先天的な認識形式の拡張であり、経験の可能性の条件であるようなメタファーである。MBAI は基本領域、応用領域、投射から構成されるが、MBAI と従来の人工知能の最大の違いは、MBAI が基本領域を形式化することによって「意味」を扱える点であり、この点において、MBAI はソフトウェアで実現できる人工知能の限界である。

Metaphor Based Artificial Intelligence

Hiroshi Tsukimoto
Research & Development Center, Toshiba Corporation

This paper presents Metaphor Based Artificial Intelligence (MBAI). Artificial intelligence is based on symbolicism, which has not succeed in the artificial realization of intelligence. Impossibility of artificial intelligence mainly depends on embodiment, contex dependency, and intentionality. It is impossible or very difficult to realize these three items by software. So, we consider the realization of forms of experiences by software. The forms of experiences are the extensions of apriori forms of cognition (Kant) and are metaphors enabling experiences. MBAI consists of basic domains, application domains and projections. The biggest difference between MBAI and other AI systems is that MBAI can deal "meanings" thanks to the formalization of basic domains. In this point, MBAI is the best among AI by software.

1 はじめに

本論文の主旨は、「従来の人工知能（システム）を本質的に改良するにはメタファーが必要であり、そしてメタファーを用いて実現できる人工知能システムがソフトウェアで実現できる人工知能システムの限界である。」ということである。

従来の人工知能（システム）（以下、「システム」は省略して、「人工知能」とする。）は基本的に記号主義、計算主義と呼ばれる方法論に基づいているが、この方法では本当の人工知能は実現できないという批判があった [5], [6], [20]。「本当の人工知能」とは何かを定義するのは困難なので、ここでは自然知能、即ち人間の知能に似ているもの程度にしておく。その本当の人工知能は実現できないという批判の根拠は主に、身体性、状況依存性、志向性である。本稿では本当の人工知能を実現するために上記の身体性、状況依存性、志向性を実現するのは基本的に不可能であるか、もしくは非常に困難であるという立場に立ち、本当の人工知能が実現できなくても、ソフトウェアで実現できる最良の人工知能が実現できれば良いと言う立場に立つ。この「ソフトウェアで実現できる」とは、直観的に、現在の計算機上でソフトウェアを動かして実現できる、という程度の意味であり、ロボットを含んでいない。

そのソフトウェアで実現できる最良の人工知能を検討するために、有意義性が身体に依拠している事実に注目して、その有意義性、即ち経験の可能性の条件、即ち経験の形式に注目する。その形式は経験から得られるのではなく、逆に経験を可能にしているもの、もしくはそれを通してしか経験できないような枠組、もしくは経験の制約として機能する形式 [13] のことである。そしてその経験の形式とは、結論から言えば、カントの先天的な認識形式の拡張であり、経験を可能にしているようなメタファー、即ちそれなしでは経験自体が成立しないようなメタファーのことである [16], [12]。なお本稿では論理を特権視せず、論理もメタファーの基本領域の形式化の一つであると言う立場に立つ。

メタファーに基づく人工知能（以下 MBAI）を提示するのであるが、それは基本領域とそれが投射される応用領域から成り立っている。特長としては柔軟性、オントロジーの高度化、知識共有の高度化、意味の生成等がある。このようなシステムを構成するために必要な課題は、基本領域の列挙、基本領域の形式化、応用領域への投射の機構の解明、形式的表現方法の確立、形式的表現と自然言語との対応の明確化である。MBAI も記号処理システムであり、従来の AI との最大の差は、MBAI では基本領域を形式化して「意味」を扱えるようにしている点である。そして本論文では「メタファーを用いて実現できる人工知能システム MBAI がソフトウェアで実現できる人工知能の限界である」と主張するのであるが、これは基本領域の遡及不可能性に依拠する。

本論文の構成であるが、最初に 2 節で人工知能について述べる。従来の人工知能に簡単に触れ、そして人工知能不可能論について述べ、基本的にその不可能論に基づき人工的に知能を実現する方法を検討するが、実現が困難な故に、3 節で経験の形式をソフトウェアで実現することによる人工知能を提示する。4 節ではメタファーに基づく人工知能システム MBAI の構成と課題と特長等について述べる。最後に 5 節では従来の AI との違い、CYC、オントロジー、類推システム等との関連について簡単に触れる。

2 人工知能について

従来の人工知能は基本的に記号主義、計算主義に意識的であろうがなかろうが基づいていると言える。そして従来の人工知能は現在までのところ成功はしていない。その理由は種々あると思うが、一言で言えば知能を記号化し得ないと言うことである。人工知能の実現に関しては例えば [22] では 4 個の立場があると述べているが、ここでは以下の二つの立場を考える。一つは知能は記号化し得ないと言う立場であり、もう一つは記号化し得るけど現在までのところそれに成功していないと言う立場である。本論文の立場は「多分、知能を完全には記号化し得ないけれども、ある程度までは記号化できるであろう」と言う立場である。そしてどこまで記号化し得るかを検討することが重要であると考え。即ち知能の記号化の限界設定の問題が重要と考える。この問題を考える際にはドレイファス、サール等の人工知能不可能論者の意見を聞くのは有意義なことである。人工知能の実現不可能性の根拠は主に、身体性、状況依存性、志向性である。

2.1 従来の人工知能

エキスパートシステムは、例えば If Then 形式の規則を実装すれば、熟練者と同等の機能を有するシステムが構築できるという触れ込みで、一時ブームになったが、程なくその規則を獲得するのが困難であると言うことが分かり、その派生問題として知識獲得の問題が焦点となったが、特に有効な方法もなく、現在のところ失敗したと言う見方が一般的である。

また知識に重点を置いたシステムとして CYC がある。これは大量の知識を実装して人間の常識を実現しようと言う試みであり、現在も開発中である。このシステムは MBI と関連があり、後で再度簡単に触れる。

記号主義に対して結合主義（コネクショニズム）という立場があり、人間の脳の神経回路網をソフトウェアでモデル化することによって、知的な機能を実現しようとするものであるが、概念等をどう実現するのか議論をしている最中である。しばしば結合主義は記号主義に対峙するものとして取り扱われるが、これも基本的には広義の意味で記号主義である。

このような記号操作的な人工知能の可能性を肯定する立場は、計算主義、記号主義である。この立場は西洋の伝統的な哲学の中でも見出せる。代表例はギリシャのプラトン、近世ヨーロッパのデカルトである。

2.2 人工知能不可能論について

人工知能の可能性を否定する人は多いが代表的な論者としてはドレイファス [5]、[6]、サール [20]、ウィノグラード [24] が挙げられる。その主な根拠は、自然知能の身体性、状況依存性、志向性と言って良いであろう。身体性とは知能の主要部分が身体に依存していると言うことである。状況依存性とは知能が環境、状況に埋め込まれていて、例えば環境から分離することは不可能であるか、もしくは分離された部分はもはや知能というに値しないと言うことであり [29]、志向性とは知能の本質的部分である意識が志向性を有していることである。そして、記号主義、計算主義でこれらを実現するのは不可能であるということが人工知能不可能論者の論拠である。これらの議論の拠り所は知性における身体、状況等を強調したハイデッガー、メルロポンティ [18]、ウィットゲンシュタイン [25] 等である。

2.3 人工的に知能を実現するには

ここでは人工的に知能を実現するのに必要だといわれている志向性、状況依存性、身体性について簡単に述べる。

2.3.1 志向性について

志向性とは意識の志向性である [20]。意識の本質は自意識である。従って志向性が人工知能に必要なならば意識、自意識を実現せねばならない。ところで、この志向性を外部の観察者が想定するものであるという解釈 [4] もあるが、本稿ではその意見には賛同しない。意識、自意識、自己についての議論は種々ある。例えば意識とは他者のシミュレーションである (Humphrey) [10]。自己とは役割である (広松渉) [?]。他にもこのような議論があって、自己が、社会性、集団性と密接な関係があることは否定できない。更にもし意識が付随的なものならば、(完全な) ロボットに意識はいらぬ。もしそうでなくとも、ロボットにはロボットなりの「意識」があって良いのではないか。なにも人間と同じような意識が必要なわけではないであろう (広松渉) [32]。ところでわれわれの経験には常にこの自己というものが伴っていることもまた事実であり、何かを理解するとはこの自己というものの抜きには考えられない。この自己が人工的に知能を実現する時に必要であるのかないのかは、現在のところ筆者にはわからないが、ここでは理解と言う行為には自己は必要であるとしておく。

2.3.2 状況依存性について

記号主義、計算主義は、個々の記号の意味もしくは操作が全体から独立に定義でき、全体はその集まりである、という仮定を (暗黙に) 前提にしているが、これに対して、そのようなことが不可能ではないかという議論がある。即ち全体を考慮して初めて意味、操作が可能になるという議論である [19]。また知能においては、個々の表象の操作よりは、状況、背景などの原理的に表象不可能なものが重要な役割を果たしているという議論もある [24]。

更にアフォーダンス理論というのもあり、「アフォーダンスは動物にとっての環境の性質で、知覚者の主観が構成するものではなく、環境の中に実在する知覚者に価値のある情報である。」と主張する [29]。確かにそうかもしれないが、著者は正直なところ、こう言われてもどうしたら良いのか戸惑ってしまう。またこれに近いアプローチとしてブルックス [3] の試みもあるが、その可能性に疑問を持つ。

また状況理論 [1] というのもあるが、この方法で上記のような状況依存性を実現できるとは思えない。

2.3.3 身体性について

メルロ・ポンティが「受肉した主観」 [18] としての身体を強調したが、それ以降、人間の知性における身体的作用の重視は今世紀になって広く認められるところとなっている。身体が世界の意味の分節を行なうのであ

る。簡単な例は以下の通りである。我々はなぜ前後が理解できるのであろうか。それは前後に身体が非対象であり、基本的に一方向（前）しか見れないし、その方向に歩くからである。例えば全方位に対称な動物、即ち目の知覚器官が各方位に均等で、足等の歩行器官も各方位に均等であるような動物を考えてみよう。その動物は前後の意味を理解できないであろう。同様なことは上下にも言える。上下が意味を持つのは身体が重力を受けているからである。無重力状態で生息する全方位対称の動物に上下は無意味である。このような事柄はその他種々存在するが、空間的な身体が意味を分節するだけではなく、欲求や生理的事項の源としての身体も世界の意味の分節を行なっている [12]、[27]。

2.4 本稿の立場でのまとめ

本稿では、上記の、人間の身体性、状況依存性、志向性等に人間の知能が依存していることに基本的に同意するのであるが、人工知能を実現するために上記の身体性、状況依存性、志向性をソフトウェアで実現するのは基本的に不可能であるか、もしくは非常に困難であるという立場に立ち、本当の人工知能が実現できなくても、ソフトウェアで実現できる最良の人工知能が実現できれば良いという立場に立つ。即ちソフトウェアで実現できる最良のもので妥協すると言う立場である。それが本当の知能と違っていても、それは仕方のないことだとあきらめる。従って本稿ではソフトウェアで実現できる最良の人工知能について検討する。即ちソフトウェアでどこまで自然知能を模倣できるかを検討する。その結論は身体性等に基づく有意味性の形式、即ち経験の形式をソフトウェアで実現することではないかと思う。これについては次節で詳述する。なお上記の「ソフトウェアで実現できる」は、直観的には、ロボットを考慮していないということの意味する。即ちロボットによる「自己」の実現の可能性を否定していないからである。

3 経験の形式とメタファー

意味が理解できるソフトウェアは、前節の議論から自己を実現しなければならないので、諦める。そこで意味を理解できなくともその有意味性の形式を扱えるソフトウェアの実現を考える。前節で簡単に見たように、身体が世界の意味の分節を行なっている。この身体性等に基づく有意味性の形式をソフトウェアで実現するのである。有意味性の形式とはその形式からはずれば意味をなさない、即ち無意味になるような形式のことである。別の言い方をすれば、有意味な経験はかならずその形式を有しているような形式のことである。例えば、内外の形式である。内外、即ち包含は、我々人間が3次元世界の皮膚で区切られた容器であることに、その理解の基礎がある。この包含そのものをソフトウェアで理解することは不可能である。即ちある具体的なベン図を理解することはできない。なぜならばソフトウェア自体には3次元世界での容器性がないからであり、ソフトウェアは感覚、即ち感性を有していないからである。しかしその包含の有意味性の形式をソフトウェアが「理解」することは可能である。包含の有意味性の形式とは、詳細は後述するが、命題論理のことである。この形式、即ち命題論理を「理解」するとはその形式操作が行なえると言うことである。ソフトウェアはその形式操作を行なえるので、ソフトウェアは包含の形式を「理解」できる。ソフトウェアは包含自体は理解できないが、包含の形式は「理解」できるのである。

この有意味性の形式とは、我々人間の経験が有している形式のことでもあり、以降経験の形式と言う。この経験の形式はカントの先天的な認識形式の拡張であり、経験を可能にするようなメタファーのことである。経験の形式とは、我々人間が外界、対象、世界をどう理解し、経験しているかの枠組であり、これは、文化、言語に依存している形式と依存していない形式がある。後者は基本的に、人間に共通な身体に依拠していると思われる。

本節ではカントの認識論 [13]、その拡張である経験の形式即ちメタファー、レイコフ [16]、ジョンソン [12]、瀬戸賢一 [30]、[31] 等のメタファー理論、そしてメタファーと論理の関係の順に述べる。

3.1 先天的な認識の形式について - カントの理論 -

カントは純粹理性批判で経験を可能にしているような条件、形式を明確にしようとした。理性がどこまで語り得るのかを、即ち理性の限界を設定しようとする試みを行なおうとした。このような試みの動機は、ニュートン力学で決定論的に記述されては人間の主体性、自由などがなくなるので、カントとしては困って、理性を理論理性と実践理性にわけて、理論理性は現象界に、実践理性は物自体に関わるとし、人間の主体性を救おうとしたのである。このような動機はさておいて、理性の限界設定に関わる議論は、今日でも示唆に富むものであり、その有効性は否定できない [21]。カントの議論は、我々人間が知覚、認識している通りに世界は存在しない、ということであり、我々人間は種々の知覚、認識の形式を有しているものであり、それらの形式は経験によって得られるのではなく、先天的に有しているものであり、我々の知覚、認識はそれらの形式を通してしか可能ではないという

ことである。即ち、われわれはすべて経験から学ぶのではなく、その経験をそうとしか経験できないような先天的形式がわれわれに備わっていると言うことである。別の言い方をすれば外界の存在の形式を人間の認識の形式に転換したのである。カントは自分の理論を経験的実在論であるけれども超越論的観念論であると言っている。これは経験的に、即ち常識的に考えた場合の対象の存在性を否定はしなけれども、超越論的には、即ちメタ的にはその対象の観念性を主張するものである。

感性の形式としては時間と空間がある。3次元空間は普通に考えれば世界の存在の形式である。しかし少し考えてみれば、その3次元空間は我々人間の認識形式であるかも知れない。即ち、我々人間が世界をそう見ているだけで、「本当に」世界が3次元として存在していないかもしれない。ひょっとすると我々人間より高等な知性が存在しその知性にはこの世界は6次元に見えているかも知れないし、空間、次元とは全く異なる枠組で世界と接しているかも知れない。空間と言う形式が存在形式なのか認識形式なのかは議論のあるところであり、どちらが正しいかは決着がつかないかも知れない。

悟性の形式としてはいくつかのカテゴリーと規則がある。このカテゴリーは現在で言えば論理である。当時の論理学のレベルからしても、カントのこのカテゴリーの部分にはそれほど忠実である必要はないであろう [35]。この形式も存在形式なのか、認識形式なのかは議論のあるところである。

3.2 経験の形式、メタファー

ところでこのような形式は本当に先天的に人間に具備されたものなのであろうか。後天的に学習等で獲得した形式ではないのであろうか。これに関しても、多くの議論があり、後天的であると言うものも、先天的であると言うものもある。しかし発生心理学的な議論、もしくは哲学的な議論をひとまずおいて、発育終了後の普通の人間に話しを限定すれば上記の3次元空間みたいな経験の形式は「先天的」であると言って良いであろう。

本稿ではこれらの形式が存在形式か認識形式かは議論せず、経験の形式として理解しておく。例えば、空間と言う形式が、存在形式か認識形式かはわからなくても、我々人間の経験の形式であることは否定できない。即ち我々人間には世界は3次元に見えるのであり、我々人間は世界を3次元空間として経験、理解しているのである。また後天的であるか、先天的であるかも議論しない。そのような形式を発育終了後の通常の人間が有していれば良いとする。別の言い方をすれば、カントの先天的な認識形式を「先天的」な経験形式に拡張したことになる。そしてこの経験の形式とは以下で説明するメタファー、特に死んだメタファーである、即ちそれなしでは経験自体が成立しないようなメタファーである [16], [12]。ここで言う死んだメタファーとは「彼は一匹狼だ。」のような文学的なメタファーではないと言う意味である。たとえば「彼女の気持ちは私に伝わって来なかった。」は導管のメタファーである。「気持ち」が「彼女」と「私」をつなぐ導管の中を「伝わって来る」と表現していて、「気持ち」を導管の中を流れるものとしている。このようにメタファーで我々は表現するし、さらに言えばこのようなメタファーを用いずに表現することは困難もしくは不可能である。

次の文は [30] に載っている長尾真「人工知能と人間」の一部分である。

「人工知能研究は人間の知的活動がどのようなものであるかをコンピュータプログラムで模擬的に実現することによって、その内容を明らかにしようとするものである。これは公理論的立場、経験主義的立場のいずれにも共通しており、両者ともにこれを実現するのに記号を基礎におく。これには記号論的立場の記号という意味が含まれてはいるが、もっと単純にアナログに対するデジタルという意味合いが強い。・・・」

上記の文の太字はメタファーであり、具体的には次の通りである。[もの]: 存在のメタファー、[内容]: 容器のメタファー、[明らか]: 視覚のメタファー、[立場]: 空間のメタファー、[共通]: 空間のメタファー、[両者]: 擬人のメタファー、[基礎]: 建築のメタファー、[含まれ]: 空間のメタファー、[対する]: 空間のメタファー、[強い]: 力のメタファー

この例からも、メタファーは非常に多く使われていることがわかる。抽象的な表現は一般的にメタファーに基づく表現であり、身体性等に基づく（共同）主観的な表現であるとも言える。メタファー表現を可能にしているのは想像力である。想像力は通常は芸術的な想像力を想起しがちであるが、ここで働く想像力は経験に必須な想像力である [12], [14]。

3.3 メタファー理論について

ここでは、レイコフとジョンソンの「レトリックと人生」以降のレトリック理論について簡単に触れる [16], [17], [12], [30], [31]。

レイコフとジョンソンは「レトリックと人生」で日常言語の存在基盤にメタファーを見た。例えば「意味が伝わらない」は「意味」を導管の中を流れるものとして表現している。即ちここでは「導管のメタファー」を用い

ている。このような装飾としてではない日常のメタファー、思考や行動を条件付けている本質的・深層的認識への転換を彼らは強調した。

瀬戸は「レトリックと人生」を評価しながらも以下の点で批判する。

1. メタファーを偏重している。メタファー以外に、概念体系を構成すると考えられるものはメトミニー（換喩）、シネクドキ（提喩）等いくつかある。
2. 普遍性の観点が欠落している。メタファーは言語に依存しているから、文化相対主義である。しかし文化普遍的なメタファーもある。
3. メタファーの種類が未整理である。悟性、感性、空間、視覚等でメタファーを分類する必要がある。

1に関してはその通りであろうが、ここではメタファーだけを考える。2に関しては、文化相対的なメタファーもあれば文化普遍的なメタファーもある。文化相対的なメタファーの一例は、過去の経験であろう。

私は中国に行ったことがある。

I have been to China.

日本語では存在のメタファーであるが、英語では所有のメタファーである。また文化共通のメタファーは内外（包含）、上下等の基本的な空間メタファーであろう。前者の方は翻訳不可能性 [7], [15]に通ずるものであろう。後者の方は身体等人間に共通な部分に由来するメタファーである。

3.4 論理について

カント、ジョンソン等の理論と本稿の立場の差異の一つは論理に対する考え方である。本論文では論理をもっとも基本的なものと思わない。したがって厳密性を放棄する。純粹理性批判のカテゴリー表は今の言葉でいえば論理であるが、カントはこれをその他の図式とは別ものと扱っている。ジョンソンでもその辺りは不明瞭である。本稿では論理はメタファーの形式化であるという立場を鮮明に取り、論理の特別視はやめる。即ち、メタファーの形式化になりえないような論理は、現実世界で理解可能なモデルを持たないものであるから、単なる記号操作体系とみなすことにする。

では具体的に命題論理は何であるのか。命題論理は空間メタファーの中で最も基本的な包含のメタファーの形式化である。ベン図は命題論理を理解する時に補助的に使われる図であるが、単なる補助手段でなく、基本的なものであることを主張したい。なぜならば、我々は命題論理の計算規則なしで、このベン図だけで包含を理解できる。しかし、ベン図なしで、命題論理の計算規則だけで包含を理解することは、非常に困難であるか不可能である。したがって、命題論理よりもベン図の方が我々人間の包含の理解にとって、より基本的なものなのであると言って良いと思う。

命題論理と言っても、古典、非古典いくつか存在する。古典命題論理はベン図の線の部分を独立したものとみなさない包含の形式化であり、直観主義命題論理はその線の部分を独立したものとみなした包含の形式化である。その他の非古典命題論理のいくつかはこのような解釈ができるかも知れないし、また別のいくつかはこのような解釈ができないかも知れないが、それは重要な問題ではない。なぜならば非古典命題論理は古典命題論理の規則を部分的に除去して作られたものであり、包含等のメタファーの基本領域と対応が取れるように、その規則の除去はなされていないからである。したがって、包含等のメタファーの基本領域と対応がとれないような命題論理は、記号論理であることには間違いないが、現実世界での理解可能なモデルを持たないので、単なる記号操作体系にしか過ぎないと言って良いと思う。

次に、論理と自然言語の関係について簡単に触れる。通常、雑な言い方をすれば、論理は自然言語の形式化であると論われている。しかし、上記のような考え方をすれば、論理は自然言語の下でのごくメタファーの形式化であることになる。論理と自然言語の間には乖離がある。論理和、論理積、否定、論理的含意等の論理接続詞が自然言語の接続詞と対応がとれないと言う問題が存在している。例えば最も典型的なのが論理的含意であり、論理的含意の違和を除去するために適切論理と言う論理も研究されている。論理的含意以外の論理接続詞にも違和がある。例えば論理積は日本語の「かつ」と対応するが、英語の“AND”とは対応しない。英語の“AND”は日本語の「かつ」と「と」に対応し、「と」は論理積とは対応しない。しかし上記のように、命題論理を包含のメタファーの形式化であると考えれば、論理は自然言語の直接的な形式化ではなく、論理的接続詞が自然言語の接続詞と対応がとれていなくとも良いことになる。即ち自然言語と論理が対応がとなくて良いわけである。

さて、次に述語である。ところで「AはBに含まれている」を $In(A, B)$ と2項述語 In を用いて書くこともできる。以降、簡単のためにこの記法を述語記法と言う。ところで包含は、現在の命題論理の論理和、論理積、

否定を用いなくても In のような述語を用いて述語記法で書ける。しかしこれを述語記法とせず特殊な記号としたのは、これらが最も基本的な空間メタファーである包含のメタファーの形式化だからであろう。そこで以下で、現在の命題論理と述語記法について簡単に比較する。 $In(A, B)$ とソフトウェアで実装しても、計算機にはその包含の意味はわからない。しかし、ブール代数でも、LK でも良いが、命題論理の形式操作をソフトウェアで実装すれば、計算機は少なくとも、包含の意味は分からなくとも、包含の形式操作はできるのである。ここで「意味とは差異である [33]」を思いだしてもらいたい。 $In(A, B)$ と書いただけでは、他の表現とは差異はない。 $In(A, B)$ ではなく、 $Out(A, B)$ と書いてもスペルの差異を除けば同じである。人間はこの In, Out という英語表現から内、外を読みとれ、現実世界の内、外を想起することができるが、計算機にはそれができない。計算機にわかるのは単にスペルの違いだけであるが、我々が計算機に分かってもらいたい差異はスペルの差異ではない。しかし、例えば $In(A, B) \wedge Out(A, B) = \perp$ と書けば、 In, Out に他の記号操作とは違うなんらかの関係を与えたことになる。勿論 $\wedge, =, \perp$ はこの式で同時に定義されることになる。包含に関するこのような関係をきちんと与えたのが命題論理である。この命題論理の形式操作が他の形式操作と違う点が命題論理が包含の形式的な「意味」を有すると言うことである。

述語記法で表現可能な包含が命題論理として形式化できるのであるから、他の基本領域の形式化を行なうことも可能ではないかと考えられる。これは現在の述語論理で述語表現されている述語を「命題論理化」することを意味する。このような「命題論理化」により、計算機はより良くその述語の「意味」を理解できるのである。

すべてが基本領域の形式化であるとするならば、述語表現 ($P(a) : a$ は P) 自体もなんらかのメタファーであることになるが、これは時間メタファーであると言え、カントが考察した超越論的統覚、根源的統覚がその基本であると考えられるが、詳述は別稿で行なう。

4 メタファーに基づく人工知能システム MBAI

4.1 構成

本稿では MBAI を提示するのであるが、それは基本領域と投射と応用領域から成り立っている。基本領域とは例えば、以下の通りである。空間で言えば内外 (包含)、上下、前後、等であり、その他には道、周期、容器、充満、建築物、導管等がある。応用領域とは上記の基本領域が投射される領域であり、例えば、経済、教育、政治等の領域がある。

4.2 課題

このようなシステムを構成するために必要な課題は以下の通りである。

基本領域の列挙 : 例えば内外 (包含)、上下、前後等のいくつかの空間に関する領域、道、周期、容器、充満、建築物、導管等の領域が基本領域であると考えられる。レイコフ [16]、ジョンソン [12]、瀬戸 [31] 等がいくつかの領域を挙げているが、それぞれ異なる。整理、列挙せねばならない。

基本領域の形式化 : 各基本領域を形式化する必要がある。例えば、包含のメタファーの形式化は (古典、直観主義等の) 命題論理であり、建築のメタファーの形式化はいわゆるオントロジーの一種になるであろう。例えば空間メタファーで包含の次に基本的なメタファーであると考えられる上下に関する形式化は、上下を導入したベン図、即ちある方向に非対称なベン図の包含に関する、命題論理と同様の形式的な体系になるであろう。建築のメタファーの形式化としては、設計、基礎、主構造 (床、壁、屋根)、副構造 (柱、窓、ドアなど) が考えられる。これはいわゆるオントロジーである。したがって、論理とオントロジーとはメタファーの基本領域の形式化としては同一視される。この基本領域の形式化は一意に決まらない。包含の形式化として古典命題論理と直観主義命題論理の二つがあるのがその例である。

応用領域への (組合せ的) 投射 (推論) の機構の解明 : 例えば建築 \rightarrow 理論、戦争 \rightarrow 議論、金 \rightarrow 時間等の投射がある。また組合せ的投射としては建築、戦争 \rightarrow 理論、議論等がある。この投射が頻繁に使われているのが死んだメタファーであり、あまり使われないのが生きたメタファーである。そして頻繁過ぎて固定化されたような投射は死んだメタファーというよりは、基本領域に戻る事が語源を知るようになるようなものであろう。さらに初めて使われるような投射は文学的なメタファーなのであろう。

形式的表現方法の確立と自然言語との対応の明確化 : 例えば、「この文章の意味が伝わる」は導管メタファーであるが、これのメタファーレベルでの表現の一例は「動き (導管 in 文)」となろう。なお、一つの自然

言語表現に対応するメタファー表現は一つではなく複数存在する。逆に一つのメタファー表現に対応する自然言語表現も複数存在する。

4.3 基本領域の遡及不可能性

メタファーの応用領域の表現は基本領域に遡及することができるが、メタファーの基本領域はそれ以上他のある領域によって表現することはできない。この基本領域を理解するには言語の外に出なければならぬ。即ち感性レベルでの理解を必要とする。これはカントの言う図式に相当する。人間は個々のイメージ（表象）を抱くことはできても、一般的なイメージ（表象）を抱くことはできない [2]、[9]。これに対し、言語レベルでは一般的な観念（犬、三角形等）しか存在しない。イメージ（表象）は感性の対象であり、言語は悟性の対象である。この感性と悟性の間には飛躍が存在するが、この両者に関わっていて、これらをつなぐものがカントの言う図式であり、この図式を処理しているのが人間の想像力である [13],[14]。

ソフトウェアは身体、感性、想像力等を有していないがゆえに、ソフトウェアが意味自体を理解することは、不可能である。したがって、ソフトウェアではメタファーの基本領域は遡及不可能なのである。このように考えれば、MBAI は、表現を基本領域に還元して、その基本領域が言語レベルで遡及不可能であることを考えれば、意味を扱う点で、ソフトウェアで実現できる人工知能の限界であることが分かる。

そしてこの基本領域の形式操作を通じて、ソフトウェアが「意味を理解」できるのである。なぜならば各基本領域の形式操作が異なるからである。即ち意味とは差異である [33]。なおこれは記号とパターンの統合の問題 [26],[28] や、記号接地問題 (Symbol Grounding Problem) [8] と密接に関連する。

4.4 特長

このようなシステムの特長は、柔軟性、オントロジーの高度化、知識共有の高度化、意味の生成等であろう。

柔軟性に関しては、自然言語の表層的な表現に捕われずに、「意味」に基づいた表現を行なうので、従来の人工知能より柔軟な処理が可能である。別の言い方をすればメタファーに基づく理解とは、文を状況単位で理解しているとも言える [16]。

また、オントロジーの高度化に関しては、やはり「意味」に基づいた表現をしているので、自然言語に基づくオントロジーよりは高度なものを構成することができる。しかし、メタファーを用いて表現するのは主に抽象的なものである。したがって、具体的な存在物に関してのオントロジーは他の手法のものと同じである。

そして同様の理由で、知識共有の高度化も可能である。知識共有の例で言えば翻訳がそうであろう。翻訳は異種の言語間の知識共有であり、それは基本的には身体等を共有していること、即ち身体等に基づく基本メタファーを共有しているからこそ可能なのである。

そして最後に、「意味」の生成であるが、我々の理解が、複数の基本領域を組み合わせる応用領域に投射することによってなされているのなら、逆にある複数の基本領域を組み合わせる応用領域に投射することによってわれわれに理解可能なことを、即ち有「意味」なことを生成することができると考えられる。別の言い方をすれば、基本領域を適当に組み合わせる応用領域に投射することによって得られる表現は、基本的に人間に理解可能であり、その表現に対応する慣習的な自然言語表現が存在すれば、それはいわゆる死んだメタファーであり、慣習的な自然言語表現が存在しなければ、それは生きたメタファーになる。例えば日本語では「時間がある。」と存在のメタファーで表現するが、これは死んだメタファーである。これを「時間を持つ。」と所有のメタファーで表現すると生きたメタファーになる。但し英語では、この限りではない。

5 他のシステムとの比較

5.1 従来の AI システムとの比較

さて、メタファーに基づく人工知能（以下 MBAI）と言っても、結局記号処理システムであることには変わりはないのである。そうすると一体どこが従来の人工知能と違うのであろうか。結論から言うと、メタファーの基本領域を形式化してそれを応用領域に投射しているところが、従来の人工知能との最大の違いである。MBAI も、記号処理システムなので、記号の意味は理解はできない。しかしその記号を理解可能にする形式を実装しているので、理解していても、基本的に、記号を操作するレベルでは理解している人間とはほぼ同じことが可能なのである。もっとも基本的な例は包含のメタファーであろう。この場合に MBAI には包含は分からない。しかし包含を理解している人間が包含を理解していない人間に具体的な図を用いずに記号操作でのみで教える場面

で行なう記号操作、即ちその一つの形式化である古典命題論理を記号的に操作する限りにおいては、その体系がブール代数であろうが、自然演繹体系であろうが、MBAIは同等の能力を有するのである。したがって従来の人工知能も古典命題論理を実装しているのならば、包含のメタファーを理解しているのである。したがって包含のメタファーに関してはMBAIと従来の人工知能の能力は同じである。しかしそれ以外では異なるのである。MBAIでは包含以外の種々のメタファーを実装する。空間のメタファーで言えば、例えば上下、前後等のメタファーであり、他の基本領域では、容器のメタファー、道のメタファー、建築のメタファー等々である。例えば建築のメタファーの形式化としては、設計、基礎、主構造(床、壁、屋根)、副構造(柱、窓、ドアなど)の関係が記述されることになる。

5.2 その他のシステムとの比較

CYC : 常識の実現を目指してLenat等がCYCの開発を行なっている。MBAIも目指すところは似ているが、MBAIには述語論理がないのに対し、CYCは認識レベルでは古典一階述語論理を用いている。ここが最大の違いである。

オントロジー : 大規模知識ベース、知識再利用、情報共有等でオントロジーが注目されている[34]。しかし、オントロジーの構築は難しい。人間は対象のオントロジーを持ってはいないのでないか、状況に応じてその具体的な対象のある側面に注目して、いくつかの基本領域のオントロジーを組み合わせているのではないか、と思われる。例えば、「考え」にはさまざまな(基本)領域が投射される。その領域は例えばたべもの、植物、商品、資源、お金、ファッションである[16]。したがって「考え」のオントロジーを構築することは不可能であり、いくつかの他の領域のオントロジーを状況に応じて投射していると考えべきであろう。

類推システム : 従来の類推システムは、基本領域、応用領域の個別の具体的な構造にはあまり関心を示さなかった。また類推を用いて何か新しい関係を発見する、学習することに関心を示していた。MBAIでは類推を用いて発見、学習を行なうことを主主題としてはいない[11][23]。

6 おわりに

本論文ではメタファーに基づく人工知能システムMBAIについて述べたが、記述がまだ曖昧な点が多いが、今後は細部をつめるとともに、基本領域の列挙、基本領域の形式化、応用領域への(組合せ的)投射(推論)の機構の解明、形式的表現方法の確立と自然言語との対応の明確化等の課題を研究して行く予定である。

参考文献

- [1] J. Barwise and J. Perry, *Situations and Attitudes*, MIT Press, 1983. (状況と態度、土屋等訳、産業図書、1992)
- [2] G. Berkeley: *A Treatise concerning the Principles of Human Knowledge*, 1710. (人知原理論、大槻訳、岩波書店、1929)
- [3] R.A. Brooks: Intelligence without Representation, in *MIT Teche Report*, 1988.(表象なしの知能、柴田訳、現代思想, Vol.18.3, pp. 85-105, 1990)
- [4] D. C. Dennett: *The Intentional Stance*, MIT Press, 1987. (志向姿勢の哲学、若島他訳、白揚社、1996)
- [5] H. Dreyfus: *What Computers Can't Do: A Critique of Artificial Reason*, Harper & Row, 1972. (コンピューターには何ができないか、黒崎他訳、産業図書、1992)
- [6] H. Dreyfus and S. Dreyfus: *Mind Over Machine*, John Brockman Associates, 1986. (純粋人工知能批判、椋田訳、アスキー出版、1987)
- [7] P. Feyerabend, *Against Method* New Left Books, 1975. (方法への挑戦、村上訳、新曜社、1981)
- [8] S. Harnad: The Symbol Grounding Problem, *Physica D*, Vol.42, pp.335-346,1990.

- [9] D. Hume: *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, Oxford University Press, 1888. (人性論、大槻訳、岩波文庫、1948)
- [10] N. Humphrey: *The Inner Eye*, Faber and Faber, 1986. (内なる目、垂水訳、紀伊国屋書店、1993)
- [11] B. Indurkha: *Metaphor and Cognition*, Kluwer Academic, 1992.
- [12] M. Johnson: *The Body in the Mind*, The University of Chicago Press, 1987 (心の中の身体、菅野他訳、紀伊国屋書店、1991)
- [13] I. Kant: *Kritik der Reinen Vernunft*, 1787. (純粹理性批判、篠田訳、岩波書店、1961)
- [14] I. Kant: *Kritik der Urteilskraft*, 1790. (判断力批判、篠田訳、岩波書店、1964)
- [15] T. Kuhn: *The Structure of Scientific Revolutions, 2d ed.*, University of Chicago Press, 1970. (科学革命の構造、中山訳、みすず書房、1971)
- [16] G. Lakoff and M. Johnson: *Metaphors We Live By*, University of Chicago Press, 1980. (レトリックと人生、渡部他訳、大修館書店、1986)
- [17] G. Lakoff: *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, University of Chicago Press, 1987, case study 2. (認知意味論、池上他訳、紀伊国屋書店、1993)
- [18] M. Merleau-Ponty: *La phénoménologie de la perception*, 1945. (知覚の現象学、竹内他訳、みすず書房、1967)
- [19] W. V. O. Quine: *From a Logical Point of View*, Harper Torchbooks, 1961. (論理的観点から、中山他訳、岩波書店、1975)
- [20] J. Searle: *Intentionality*, Cambridge University Press, 1983.
- [21] P.F. Strawson: *The Bounds of Sense*, Methuen, 1966.(意味の限界、熊谷他訳、勁草書房、1987)
- [22] S. Torrance: *The Mind and the Machine*, Ellis Horwood, 1984. (AIと哲学、村上監訳、産業図書、1985)
- [23] E.C. Way: *Knowledge Representation and Metaphor*, Kluwer Academic, 1991.
- [24] T. Winograd and F. Flores: *Understanding Computers and Cognition*, Ablex Publishing, 1986. (コンピュータと認知を理解する、平賀訳、産業図書、1987)
- [25] L. Wittgenstein: *Philosophische Untersuchungen* 1953. (藤本訳、哲学探究(ウイットゲンシュタイン全集8)大修館書店、1968)
- [26] 麻生英樹: 柔らかなシステムにおける記号とバタンの統合をめぐる、RWC情報統合ワークショップ'95、pp.157-166, 1995.
- [27] 尼ヶ崎 彬: ことばと身体、勁草書房、1990.
- [28] 伊庭 幸人: 基礎的問題から見た情報統合、人工知能学会誌、Vol.11, No.2, pp.193-200, 1996.
- [29] 佐々木 正人: アフォーダンス - 新しい認知の理論、岩波書店、1994.
- [30] 瀬戸 賢一: メタファー思考、講談社、1995.
- [31] 瀬戸 賢一: 空間のレトリック、海鳴社、1995.
- [32] 広松 渉: 哲学の越境、勁草書房、1992.
- [33] 丸山 圭三郎: ソシユールの思想、岩波書店、1981.
- [34] 溝口理一郎: 形式と内容 - 内容志向人工知能研究の勧め -、人工知能学会誌、Vol.11, No.1, pp.50-59, 1996.
- [35] 山下 正男: 論理学史、岩波書店、1983.