

Title	合理的エージェント間のコミュニケーション
Author(s)	中山, 康雄
Citation	大阪大学人間科学部紀要. 1999, 25, p. 119-138
Version Type	VoR
URL	https://doi.org/10.18910/7022
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

合理的エージェント間のコミュニケーション

中山 康 雄

目 次

はじめに

1. 心的表示構造
2. 人工的エージェントの設計
3. 問題解決と行為
4. 発話による態度伝達
5. 原初的言語ゲーム
6. 会話の含意の分析

まとめ

合理的エージェント間のコミュニケーション

中山 康雄

はじめに

最近流行の複雑系の科学の考え方からすれば、社会組織は、個々のエージェントの活動の中から創発したものである。言語も個々のエージェントの問題解決や共同作業をより効率よく押し進めるために生まれてきたと考えられよう。生きること自体が、解くべき問題を常にエージェントに提出する。効率よい食物の獲得や安全な住居の確保は、生きることに不可欠である。

人間は、生まれてきた時、非力である。幼児は、他者の協力なしに生きることができない。幼児が、親に何がほしいのか、何が問題なのかを伝えることは、生きていくために極めて重要なことである。だから、幼児にとり言語を学ぶことは、生きていくための一つの手段となる。幼児は、まず、一語文を発するようになり、それから二語発話の段階にいたる。文法の面から見れば、幼児の発する語句は不完全と言えよう。しかし、コミュニケーションの観点から見れば、このような幼児の発話も十分目的を充足していると言える。

本稿は、合理的エージェント間のコミュニケーションの問題を問題解決のコンテキストの中で考察する試みである¹⁾。本稿では、合理的エージェントは、自ら目的を設定しその目的を達成するための実行可能な手段を探しこの手段を実現しようとする存在者として規定される。人間は、この意味で、合理的エージェントの一種とみなせるだろう。ここでは、議論を透明にするために、人間のようにふるまう人工的エージェントという合理的エージェントについて語ることにする。本稿の目的は人間の言語活動を説明することにあるが、人工的エージェントの活動を分析することにより、人間の言語活動の分析における新しい視点を提供することができると思うのである。

1. 心的表示構造

Davidson (1963) は、欲求のような賛成的態度と信念をあげることにより行為は説明されるとし、これを理由による行為の説明であるとする。「何故、ここを歩いているのか?」と聞かれて、我々は「家に帰るために」などと答える。このように答える時、我々は、家に帰りたいという欲求を持ち、今歩いていく方向に歩くことがこの欲求充足の手段で

あることを信じている。行為に合理的説明を与えるためには、行為者にある特定の信念や欲求や意図を帰属させることが必要になる²⁾。そして、このことは行為一般のみでなく、発話行為においても成り立つ。また、話者が多くの信念を自分と共有した合理的主体であると仮定することは、話者の発話行為の解釈を限定し、話者が何をこの発話の際に意図したかの解釈を容易にする。逆に、異文化から来た者にとっては、共有する信念の欠如のため余分な解釈の可能性まで吟味しなくてはならなくなり、原地語の発話の解釈が困難になる。我々が、他者の行為を予測できるのは、他者への合理性の帰属や他者の信念や欲求の推測が可能のためである。この前提なしに、我々は、他者を理解できない。

発話行為の説明を可能にするためには、信念、欲求、意図などの態度を持つものとして人間を捉える必要がある。そこで、人も含めたエージェントの心的状態を、{信念箱 (B-box), 欲求箱 (D-box), 意図箱 (I-box)} により表すことにする。ただし、B-box, D-box, I-box は、それぞれ D 論理式の集合とする³⁾。これを次のように図示する：

(a) {**B** : {...}, **D** : {...}, **I** : {...}}

以下、エージェントの心的状態の記述にあたり、議論に重要な点だけを取り出して描写することにする。

中山 (1998) などが論じているように、本来、信念状態は、単なる文集合や D 論理式の集合ではなく、それらの間の順序構造により記述されるべきである。また、欲求や意図に関してもこのことは当てはまる。しかし、議論の見通しをよくするため、本稿においてはこの単純化された心的状態の記述を用いた考察を展開することにする。

あるエージェント a が K ということについて \mathbf{X} という態度を取っているということを $[K] \in \mathbf{X}(a)$ で表す。括弧 [] は、引用符のような働きをし、 $[K]$ は D 論理式 K の名前だとする。そこで、 $[K] \in \mathbf{X}(a)$ の真理条件は次のように規定することができる：

(1) $[K] \in \mathbf{X}(a)$ の真理条件

M は現実世界のモデル、 \mathbf{X} は **B** か **D** か **I** かのいずれかである時、

$[K] \in \mathbf{X}(a)$ は M で真 \Leftrightarrow (a) $\{\mathbf{X} : \{K, \dots\}, \dots\} \Leftrightarrow K$ は a の \mathbf{X} -box の要素。

2. 人工的エージェントの設計

態度と行為の関係はどのようなものなのか？ 本稿では、この疑問に直接答える代わりに、人工的エージェントのモデルを提案し、このモデル上で、態度と行為の関係を説

明したい。モデルを用いることは、正確な記述とメカニズムを明らかにすることに役立つ。それは、人間における態度と行為の関係についての一つの見取り図 (picture) を提供する。それは、実際の人間の問題を直接には描かないが、どのようなことが物理的に可能なのかを見せてくれる。

アリストテレスは、理論的推論と並んで実践的推論を認めた。次の例で見るように、実践的推論では、欲求と信念から一つの行為が帰結すると考える⁴⁾：

私はのどが渴いている。

目の前の水はのどの渴きをいやすだろう。

目の前の水の入ったボトルに手を伸ばすという私の行為

欲求と信念から行為が帰結するとはいかなることを意味しているのか？ 意図や欲求というような心的な状態が身体運動という物理的事象をいかに引き起こすのか？ 意味論の範囲で考えるかぎりこれらのことは明らかでないように思われる。人工的エージェントを用いて、このようなエージェントに関してこれらの問題がいかに解かれるかを示すことは、実際の人間の問題を考えるための手助けになるだろう。

まず、はじめに生物が共通に持つ不断の活動という性質を認めたい：

(2) 不断に内的活動を続ける人工的エージェント

エージェントは、その内部で絶えず活動している。そして、エージェントは、絶えず行動の指針を求めている。意図や欲求などの賛成的態度は、エージェントが次にどのような行動を取るべきかを教える。賛成的態度は、ある行動を引き起こすというよりもむしろ、エージェントの活動にコントロールを与えている。

この視点から実践的推論の問題を考えると、実践的推論からの帰結は、エージェントが次に何をしたらいいかの指針であればよいということがわかる。エージェントが常に指針を求めてそれに従い行動するという構造を持っているとすれば、そのような指針が与えられればエージェントは行動を開始することになる。エージェントがその内部で絶えず何かをしているということは、並列の無限ループを持つプログラムを書くことで実現できる (図1参照)。

ここで、本稿での議論に重要となる人工的エージェントの基本的性質をまとめておく：

(3) 人工的エージェントの基本的性質

- a) 人工的エージェントは身体を持つ。
- b) 人工的エージェントは、命題的態度を持つことができる。

- c) 命題的態度には、「p を信じる」、「p を欲する」、「p を意図する」という三つの基本形がある。人工的エージェントの心的状態は、{信念箱 (B-box), 欲求箱 (D-box), 意図箱 (I-box)} により表される。
- d) B-box, D-box には、環境との相互作用で何かが書き込まれることがある。
- e) 人工的エージェントは外界を知覚でき、知覚情報を B-box に書き込む。
- f) 人工的エージェントは、基本的推論能力を持ち、B-box の内容の演繹的拡張を行うことができる。また、人工的エージェントは非単調推論を行うことができ、B-box の内容を無矛盾に保つことができる。
- g) B-box には、自分自身の能力についての評価が書き込まれており、どのような行動が自分に可能かについての改訂可能ナリストが存在する。
- h) 人工的エージェントは自己保存の欲求を持つ。
- i) 人工的エージェントは自己保存のために何が必要であるかのある程度の知識を B-box に持っている。
- j) [自己意識の原則] 人工的エージェント a は自分が何を信じ、何を欲し、何を意図しているかを知っている。即ち、 \mathbf{X} が \mathbf{B} か \mathbf{D} か \mathbf{I} の時、 $[\mathbf{K}] \in \mathbf{X}(a) \equiv ([[\mathbf{K}] \in \mathbf{X}(a)] \in \mathbf{B}(a))$ が成り立つ⁵⁾。

人工的エージェントの行動をコントロールする並列アルゴリズムの基本的特徴を次に示しておく。これは、図 1 に表されている並列のプロセスに相当するものである。

(4) 人工的エージェントの態度と行動を規定する並列アルゴリズムの特徴

まず、以下の規定で必要になる表記の規定をする：

Perform(ACT) は、ACT で表現された行為を遂行するプロセスを始動させる。

choice(Y) は、集合 Y から適切な一つの要素を取り出す関数とする。

ここで、Goal は話者が望む目的に相当する事態の成立を指すとする。また、 F は未来演算子で、 $F(\mathbf{K})$ は現在おかれているエージェントの状態から見て未来において \mathbf{K} が実現する時、真であるとする。

- a) B-box や D-box や I-box や行為遂行に関する操作は並列に進行する (図 1 参照)。
- b) B-box への書き込みや削除が外界との相互作用や推論により行われる。
WHILE alive(S) DO
 $\mathbf{B}(S) :=$ [外界との相互作用などからの影響による書き込みや削除]；
- c) D-box への書き込みや削除が外界との相互作用や B-box での推論の影響を受けて行われる。

WHILE alive(S) DO

Satisfied-Desire (S) := {K | ([K] ∈ D(S), [K] ∈ B(S)) が成り立つ}

D(S) := D(S) - Satisfied-Desire(S) ;

D(S) := [外界との相互作用や B(S) の変化などからの影響による書き込みや削除] ;

- d) I-box は望まれた目的の手段で実行可能と思われる自らの行為の集合として規定される。

WHILE alive(S) DO

I(S) := {Act(S) | (Act(S) は S が自分でなしうと思っている行為) & (([Act(S) → F(K₁)] ∈ B(S), ..., [K_{n-1} → F(K_n)] ∈ B(S), [K_n → F(Goal)] ∈ B(S), [Goal] ∈ D(S)) を満たす D 論理式 Goal が存在する)} ;

- e) 行為遂行 (I-box の内容が行為遂行に影響を与え、逆に、行為遂行が I-box に影響を与える) :

WHILE alive(S) DO

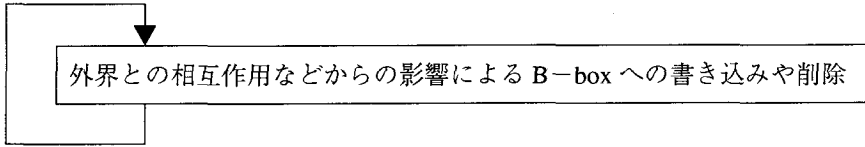
IF I(S) ≠ ∅ THEN ACT := choice (I(S)) ;

Perform (ACT) ;

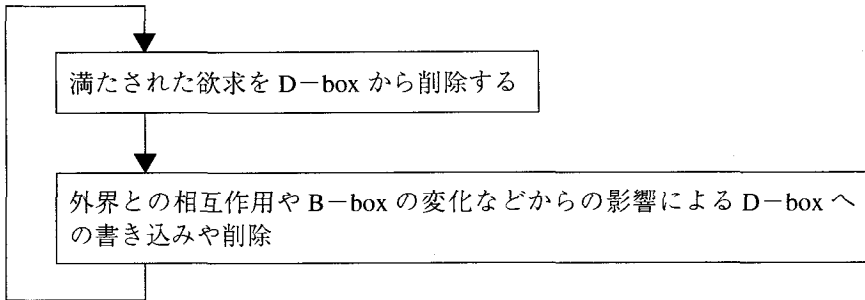
I(S) := I(S) - {ACT} ;

以下の4つの操作が並列して進行する

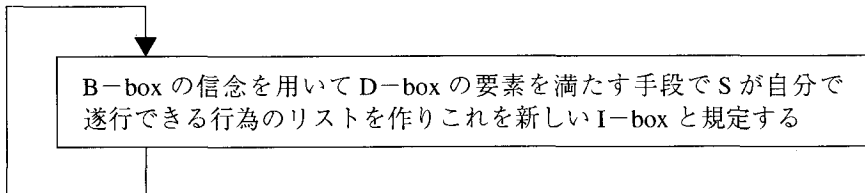
(4 b) B-box



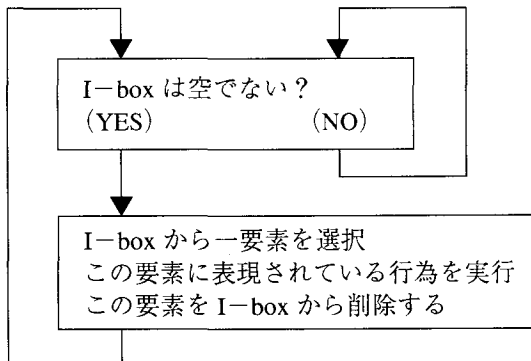
(4 c) D-box



(4 d) I-box



(4 e) I-box と行為遂行の相互作用



並列な無限ループを持ったアルゴリズム (図1)

3. 問題解決と行為

ある欲求を満たすためには、一般に、様々な道が可能である。ある欲求を満たすための方策を選択し、これを実行に移すことは、問題解決の一つでもある。自らの欲求を満たすために行為することに対応する意図の形成は、次のように表すことができる：

(5) 問題解決のための意図形成

(S) {**D** : {Goal, ...}, **B** : {Act(S) → F(K₁), ..., K_{n-1} → F(K_n), K_n → F(Goal), ...}, **I** : {...}} がある状態において成り立つなら、[Act(S)] ∈ **I**(S) が次の状態において成り立つ。

(S が Goal の実現を望んでおり、自分が行為 Act を実行することが Goal の実現をもたらすと考えているなら、S は行為 Act を実行することを意図する。この規定は (4d) に対応する。)

Goal を欲するが、即座に Goal を実現する方法が見あたらない時、考察により問題解決の方策を探し出すことが必要になってくる。そして、このような問題解決の一つの方策として他者との協力や他者への訴えかけがある。態度伝達の重要な側面は、聞き手の行為決定に影響をおよぼしうるところにある。ここに、話者にとっての新たな問題解決の可能性が生じる。自分ができないことを他者にやってもらったり、自分一人でできないことを他者との共同作業により解決したりする道が開けるのである。つまり、態度伝達がなされることで、その後の集団の振る舞いが影響されるのである。

(6) 問題解決のための他者への要求

X が態度を表し、**N** がエージェント名、**nx** がある項目名の時、(X.N.nx) という表記法で **nx** という項目の **N** というエージェントの **X**-box を表すことにする。

a) (S) {**D** : {Goal, ...}, **B** : {Act₁(H) → F(Goal), ...}, **I** : {...}}

(S は Goal が起こることを欲し、H が Act₁ という行為を遂行することによりこの Goal が達せられると S は思っている。)

b) (S) {**D** : (D.S. 6 a) ∪ {Act₁(H)}, **B** : (B.S. 6 a) ∪ {[Act₁(H)] ∈ **D**(S)} ∈ **B**(H) → F(Act₁(H))}, **I** : (I.S. 6 a)}

((6 a) で述べたことに加え、S は、H の Act₁ という行為の遂行を望んでおり、このことが H にわかれば、H は行為 Act₁ を遂行するだろうと思っている。?)

c) (S) {**D** : (D.S. 6 b), **B** : (B.S. 6 b) ∪ {Act₂(S) → F([Act₁(H)] ∈ **D**(S)) ∈ **B**(H)}, **I** : (I.S. 6 b) ∪ {Act₂(S)}}

((6 b) で述べたことに加え、S は、自分が Act₂ という発話行為を行うことにより、H の Act₁ という行為の遂行を自分が望んでいると H に知らせること

ができると思っており、またこの発話行為を実行することを意図している。)

ここで、Sの推論能力を認めると、上の心的状態から、自分がAct₂を実行することによりGoalが実現可能だとSが考えていることが帰結する。また、行為Act₂が発話行為であるなら、この行為遂行により指令型の発話行為が遂行されたことになる⁸⁾。

上の例で、「何故、あなたはAct₂を実行したのか」と後に問われたなら、Sは、「私はGoalが実現されることを欲したから」と答えることができる。というのは、もしSがGoalの実現を望まなかったなら、(6)で描写したようなSの態度変化は起こらず、SはAct₂を実行しなかっただろうからである((4)参照)。この時のSの行為には、合理的説明が可能であり、その意味で、Sは合理的に行為したとみなすことができる。

SがAct₂を実行することによりHに対しAct₁を要求した時、Hが実際にAct₁を実行するのは、次の協力の原則が成り立つ時である：

(7) HのSに対する協力の原則

[Act(H)] ∈ D(S) がHのB-boxにあるなら、Act(H)をHのI-boxに書き加える。即ち、[[Act(H)] ∈ D(S)] ∈ B(H)がある状態において成り立つなら、

[Act(H)] ∈ I(H) が次の状態において成り立つ。

(Sが自分にActという行為の遂行を望んでいるとHが思うなら、HはActという行為の遂行を意図するようになる。)

4. 発話による態度伝達

二人以上の合理的エージェントの間では、態度伝達が可能になる。話者Sは聞き手Hにある特定の自分の態度を伝えたいと思っている。つまり、Sは、彼がKで表される事象についてXという態度を持つということをHが知るようになることを意図する。ただし、ここで、XはBかDかIとする。この意図を実践に移すための一つの有効な方策は、SがKについてXという態度を持っているということをHが推論できるような文をSが発することである。これが、態度伝達の基本構図である⁹⁾：

(8) 発話による態度伝達の基本構図

a) (S){D : {[[K] ∈ X(S)] ∈ B(H), ...},

B : {Act(S) → F([[K] ∈ X(S)] ∈ B(H)), ...}, I : {Act(S), ...}}

(H){B : {...}, ...}

(SはKについてXという態度を持っているということをHに伝えたいと思っており、これを実現するためにはSはActを遂行すればよいと思っており、

また、Act を遂行するという意図も持っている。))

b) Act (S). (S は Act を遂行する。普通、Act は発話行為であるが、同様の目的が他のタイプの行為で実現できる場合もある。)

c) (S) {**B** : (B.S. 8 a) \cup {Act(S), $F([\mathbf{K}] \in \mathbf{X}(S)) \in \mathbf{B}(H)$ }, **D** : (D.S. 8 a),
I : (I.S. 8 a) - {Act(S)}}

(H) {**B** : (B.H. 8 a) \cup {Act(S)}, ...}

(S は、自分が伝えようとしたことがまもなく H に伝わるだろうと考えている。Act 遂行の意図は自らの行為により充足されたと S は思っているの、I-box から削除される ((4c), (4e) 参照)。

d) (H) {**B** : (B.H. 8 c) \cup { $[\mathbf{K}] \in \mathbf{X}(S)$ }, ...}

(S) {**B** : ((B.S. 8 c) \cup { $[\mathbf{K}] \in \mathbf{X}(S) \in \mathbf{B}(H)$ }) - $\{F([\mathbf{K}] \in \mathbf{X}(S)) \in \mathbf{B}(H)\}$,
D : (D.S. 8 c) - { $[\mathbf{K}] \in \mathbf{X}(S) \in \mathbf{B}(H)$ }, **I** : (I.S. 8 c)}

(S が K について \mathbf{X} という態度を持っているということを H は推論し、一方、S は、自分が伝えようとしたことが H に伝わったと考えている。当初の S の欲求は自らの行為により充足されたと S は思っているの、D-box から削除される。なお、以下の記述においては、(8c), (8d) に相当するステップは、統合して記述することにする。)

この時、S が K について \mathbf{X} という態度を本当に持っているなら、即ち、 $[\mathbf{K}] \in \mathbf{X}(S)$ が真であるなら、S の行為 Act は誠実 (sincere) であると言う。そして、Act が発話行為である時には、これを「誠実な発話行為」と呼ぶ。

(4) で提案されたアルゴリズムから、人工的エージェントについて次の原則が帰結する：

(9) 削除の原則：欲求や意図は充足されると消滅する。ただし、K についての欲求や意図が充足されるとは、K が真になるということである。

Austin (1962) や Searle (1969, 1979) が議論した言語行為 (speech act) の問題は、Nakayama (1998) で提案したように、合理的エージェント間の態度伝達の問題として捉えることができる。話者は、発話行為を遂行することにより、他者の行為をうながすことができ、当初の目的を達成できる。これが、次の「言語行為の成功」と呼べるプロセスである：

(10) 言語行為の成功 (指令型の発話の場合)

Act₁(S) は、ある S の発話行為の遂行を意味するとする。

a) (S) {**D** : {Act₂(H), $[[\text{Act}_2(H)] \in \mathbf{D}(S)] \in \mathbf{B}(H)$, ...},

$$\mathbf{B} : \{ \text{Act}_1(\text{S}) \rightarrow F([\text{Act}_2(\text{H})] \in \mathbf{D}(\text{S})) \in \mathbf{B}(\text{H}), [[\text{Act}_2(\text{H})] \in \mathbf{D}(\text{S})] \in \mathbf{B}(\text{H}) \rightarrow F(\text{Act}_2(\text{H})), \dots \}, \mathbf{I} : \{ \text{Act}_1(\text{S}), \dots \}$$

$$(\text{H}) \{ \mathbf{B} : \{ \dots \}, \mathbf{I} : \{ \dots \}, \dots \}$$

(SはHの行為 Act₂を欲しており、このことをHに伝えたいと思っており、これらを実現するためには Act₁を遂行すればよいと思っており、また、Act₁を遂行するという意図も持っている。)

b) Act₁(S). (Sは Act₁を遂行する。即ち、Sは Act₁で表された発話行為を遂行する。)

c) (S) { $\mathbf{B} : (\text{B.S. 10a}) \cup \{ \text{Act}_1(\text{S}), [[\text{Act}_2(\text{H})] \in \mathbf{D}(\text{S})) \in \mathbf{B}(\text{H}) \}$,

$$\mathbf{D} : (\text{D.S. 10a}) - \{ [[\text{Act}_2(\text{H})] \in \mathbf{D}(\text{S})) \in \mathbf{B}(\text{H}) \},$$

$$\mathbf{I} : (\text{I.S. 10a}) - \{ \text{Act}_1(\text{S}) \} \}$$

$$(\text{H}) \{ \mathbf{B} : (\text{B.H. 10a}) \cup \{ \text{Act}_1(\text{S}), [\text{Act}_2(\text{H})] \in \mathbf{D}(\text{S}) \}, \mathbf{I} : (\text{I.H. 10a}), \dots \}$$

(Sは、Act₁の遂行により自分が伝えようとしたことがHに伝わったと考えている。また、Hは、Sが Act₂(H)を欲しているということを推論している。)

d) (S) { $\mathbf{B} : (\text{B.S. 10c})$, $\mathbf{D} : (\text{D.S. 10c})$, $\mathbf{I} : (\text{I.S. 10c})$ }

$$(\text{H}) \{ \mathbf{B} : (\text{B.H. 10c})$$
, $\mathbf{I} : (\text{I.H. 10c}) \cup \{ \text{Act}_2(\text{H}) \}, \dots \}$

(Sは (10c) と同様の心的状態にある。Hが協力の原則 (7) を守っている場合には、Hは Act₂を遂行することを意図する。)

e) Act₂(H). (Hは Act₂を遂行する。)

f) (S) { $\mathbf{B} : (\text{B.S. 10d}) \cup \{ \text{Act}_2(\text{H}) \}$, $\mathbf{D} : (\text{D.S. 10d}) - \{ \text{Act}_2(\text{H}) \}$, $\mathbf{I} : (\text{I.S. 10d})$ }

$$(\text{H}) \{ \mathbf{B} : (\text{B.H. 10d}) \cup \{ \text{Act}_2(\text{H}) \}$$
, $\mathbf{I} : (\text{I.H. 10d}) - \{ \text{Act}_2(\text{H}) \}, \dots \}$

(Sの欲求はHの行為により満たされて消滅し、Hの行為遂行の意図も彼自身の行為遂行により充足され消滅する。)

例として、Act₁(S), Act₂(H) が次のことを意味する場合を考えてみよう：

Act₁(S) : Sが「醤油を取ってください」と言う。

Act₂(H) : Hが醤油をとる。

この例の場合、上の6段階のプロセスは、Sが「醤油を取ってください」と言い、このSの要請を理解したHが醤油を取るというシーンでのSとHの態度の変化を表していることになる。

(10b)の段階で発話行為が遂行される。(10b)の発話は、(10a)から(10c)への移行を引き起こすものならどんなものでもよい。それは、自分の知らない外国語でも幼児の一語発話でも間接的言語行為でもよいのである。ただ、SとHが共通の言語を持つなら、解釈の可能性がせばまり、この言語ゲームの遂行は容易になる。

5. 原初的言語ゲーム

Wittgenstein の原初的言語ゲームの例を考えてみよう。

「アウグスティヌスが与えているような記述のあてはまる、一つの言語を考えてみよう。その言語は、建築家 A とその助手 B との間の意思疎通に役立つのでなくてはならない。A は石材によって建築を行なう。石材には、台石 (Würfel)、柱石 (Säule)、石板 (Platte)、梁石 (Balken) がある。B は A に石材を渡さねばならないが、その順番は A がそれらを必要とする順番である。この目的のために、二人は、「台石 (Würfel)」、「柱石 (Säule)」、「石板 (Platte)」、「梁石 (Balken)」という語からなる一つの言語を使用する。A はこれらの語を叫ぶ。——B は、それらの叫びに応じて、もっていくよう教えられたとおりの石材を、もっていく。——これを完全に原初的な言語と考えよ。」(Wittgenstein (1953) § 2)

この原初的言語ゲームは、(6) と (10) の図式を用いて説明できる。このことは、原初的言語ゲームが問題解決を目的とした言語行為の成功したケースの一つとして解釈可能なことを示している。即ち、それは、必要な石材を B に取ってもらうために B に自分が何を欲しているかを語 Y を叫ぶことにより伝える試みと解することができる。ただし、Y は「Würfel」、「Säule」、「Platte」、「Balken」という語のいずれかとする。例えば、A が台石を B から得るために「Würfel」と叫ぶ言語ゲームの進行における A の心的状態の推移は次のように表せる：

(11) 言語ゲームの進行と A の心的状態の推移

Act₁(A) と Act₂(B) は次のことを表しているとする：

Act₁(A)：叫ぶ(A, “Würfel”) (A が「Würfel」と叫ぶことを表している。)

Act₂(B)：(台石(d₁), 手渡す(A, B, d₁)) (B が A にある台石を手渡すことを表している。)

a) (A) {D : {Act₂(B), ...},

B : {Act₁(A) → F ([[Act₂(B)] ∈ D(A)] ∈ B(B)), [[Act₂(B)] ∈ D(A)] ∈ B(B) → F(Act₂(B)), ...}, **I** : {Act₁(A), ...}

(B) {B : {...}, ...}

(A は B が台石を自分に手渡すことを欲し、自分が「Würfel」と叫ぶことにより、この欲求を B に伝えることができ、このことが台石を自分に手渡すという B の行為を引き起こすと考えている。また、A は「Würfel」と叫ぶことを意図している。この心的状態は (10a) に対応している。)

b) Act₁(A). (A が「Würfel」と叫ぶ。)

- c) (A){**B**: (B.A. 11a) ∪ {Act₁(A), [[Act₂(B)] ∈ **D**(A)] ∈ **B**(B)},
I: (I.A. 11a) - {Act₁(A)}, **D**: (D.A. 11a)}
 (B){**B**: (B.B. 11a) ∪ {Act₁(A), [Act₂(B)] ∈ **D**(A)}, ...}

(AはBが台石を自分に手渡すことを欲していることがBに伝わったと思っている。また、Bは、Aが自分が台石をAに手渡すことを欲していると思っている。この心的状態は(10c)に対応している。)

- d) BによるAct₂の遂行というAの望みを知ったBは、Act₂の遂行を意図し、Act₂を遂行する。つまり、BはAにある台石を手渡す。このBの行為遂行により、Aの(11a)の状態における欲求は満たされ、この言語ゲームは完了する。このプロセスは、(10d), (10e), (10f)のプロセスに対応している。

Aは、Bが台石をAに手渡すことを望んでおり、この望みのBへの伝達を通してこの望みが満たされると考えている((11a))。この態度伝達の望みが実現されるようAは適切な方策を選ぶことができる。例えば、AがBはドイツ語を理解できると考えれば、上の例のように「Würfel」と叫ぶだろう。また、Bがドイツ語を理解できないと思っても、何度かこのゲームを繰り返すうちにBは望まれた反応の仕方を習得するだろうと考えれば、やはり、Aは「Würfel」と叫ぶだろう。あるいは、それぞれの建築用石材を描いた4枚の絵をAは用意し、これをあげると同時にドイツ語の単語を叫ぶかもしれない。

このようなゲームを行うためには、相手が満足しているのかそれとも怒っているのかがだいたい推定できないといけなだろう。会話の前提にされているこのような理解は、言語習得以前にすでにある程度与えられていると考えられる。つまり、親と子供のコミュニケーションは言語習得以前からすでに始まっているといつてよいだろう。人間が犬などの動物に何かを命令し、その命令に動物を従わせることができるのも言語伝達の前に相手の欲求に関するある程度の理解が原初的な形で可能だからにほかならない¹⁰⁾。

6. 会話の含意の分析

会話の含意 (conversational implicature) の分析は、P. Griceにより着手され (Grice (1989) 参照)、Sperber and Wilson (1986) などにより発展させられてきた。本稿の態度伝達のプロセスを通してコミュニケーションを分析する仕方は、基本的に、GriceやDavid Lewisなどが提案し、SperberとWilsonが継承した「伝達の推論モデル」(inferential model) に沿っている。このモデルでは、伝達は、証拠の提示とその解釈からなる (Sperber and Wilson (1986) 邦訳 p. 3 参照)。我々の分析の特徴は、信念と意図の他に欲求という態度が大きな役割りを演じていることと、発話行為が行為の一形態として問題解決のコンテキストで分析されることにある。

関連性理論 (relevance theory) のような会話の含意の理論の強みは、字義的および非字義的語りにおけるコミュニケーションの問題を統一的な理論の中で論じることができるところにある。Sperber and Wilson (1986) の議論は、会話の分析に焦点が当てられているが、本稿の議論の関心は、その会話の背景に起こっている話者と聞き手の態度変化のプロセスとそれと連動した行為遂行の問題にある。Peter と Mary を二人の合理的エージェントとする時、発話行為における話者の心的状態は、以下のようにまとめることができる：

(12) 態度伝達のタイプ

以下、Act₁ (Peter) は K に関する Peter の発話行為とする。

a) 直接的に態度が表出される場合：

$$(Peter) \{ \mathbf{D} : \{ [K] \in \mathbf{B}(Mary), \dots \}, \mathbf{B} : \{ Act_1(Peter) \rightarrow F([K] \in \mathbf{B}(Mary)), \dots \}, \mathbf{I} : \{ Act_1(Peter), \dots \} \}$$

(これは、単純な問題解決 (5) に対応している。問題解決の Goal は、ここでは、Mary に K ということ信じさせることである。話者 Peter は、発話行為 Act₁ をこの目的を達成するための直接的手段と思っている。)

b) 間接的に態度が表出される場合：

Peter が Mary に向かって K₁ に関することを発話しつつ、K₂ と Mary が思うようになることを期待している。

$$(Peter) \{ \mathbf{D} : \{ [K_2] \in \mathbf{B}(Mary), \dots \}, \mathbf{B} : \{ Act_1(Peter) \rightarrow F([K_1] \in \mathbf{B}(Mary)), [K_1] \in \mathbf{B}(Mary) \rightarrow F([K_2] \in \mathbf{B}(Mary)), \dots \}, \mathbf{I} : \{ Act_1(Peter), \dots \} \}$$

(これは、問題解決の手段が部分手段を持つ場合である。話者 Peter は、Mary に K₁ を伝えることで K₂ も伝わると考えている。)

c) 聞き手の行為をうながしたい場合：

$$(Peter) \{ \mathbf{D} : \{ Act_2(Mary), \dots \}, \mathbf{B} : \{ Act_1(Peter) \rightarrow F([K_1] \in \mathbf{B}(Mary)), [K_1] \in \mathbf{B}(Mary) \rightarrow F(Act_2(Mary)), \dots \}, \mathbf{I} : \{ Act_1(Peter), \dots \} \}$$

(ここでは、話者 Peter は、Mary に K₁ を伝えることで彼女に行為 Act₂ をさせることをもくろんでいる。)

会話は、話者と聞き手が特定の心的態度を持ちながらそれぞれの問題解決を実行していくプロセスの一つと考えられよう。

(13) 会話における心的態度の推移 (Sperber and Wilson (1986) 邦訳 p. 67の例の変形)

$[\pm K] \in \mathbf{B}(S) \equiv ([K] \in \mathbf{B}(S) \vee [\neg K] \in \mathbf{B}(S))$ とおく。つまり、 $[\pm K] \in \mathbf{B}(S)$ は、 S が K について、それが真か偽かについて明確な信念を持っていることを表現している。また、Peter と Mary が締め切りの迫った仕事を今夜やりとげようとしているというのが与えられた状況とする。

K_1 : Mary drinks some coffee.

K_2 : Mary stays up.

Act₁ (Peter) : Peter says "Do you want some coffee?"

Act₂ (Mary) : Mary says "Coffee would keep me awake".

Act₃ (Peter) : Peter serves coffee for Mary.

a) (Peter) $\{\mathbf{D} : \{[\pm([K_1] \in \mathbf{D}(Mary))] \in \mathbf{B}(Peter),$
 $[[[\pm([K_1] \in \mathbf{D}(Mary))] \in \mathbf{B}(Peter)] \in \mathbf{D}(Peter)] \in \mathbf{B}(Mary), \dots\},$
 $\mathbf{B} : \{Act_1(Peter) \rightarrow F([\pm([K_1] \in \mathbf{D}(Mary))] \in \mathbf{B}(Peter)] \in \mathbf{D}(Peter)] \in \mathbf{B}(Mary), [K_2] \in \mathbf{D}(Mary), \dots\},$
 $\mathbf{I} : \{Act_1(Peter), \dots\}$

(Mary) $\{\mathbf{D} : \{K_2, \dots\}, \mathbf{B} : \{\dots\}, \mathbf{I} : \{\dots\}\}$

(Peter は、Mary がコーヒーを飲みたいのかどうか知りたいし、この自分の欲求を Mary に知ってもらいたい。また、彼は、「Do you want some coffee?」と Mary に向かって言うことで、この彼の欲求が Mary に伝わるとしており、この発話行為を意図している。そして、Peter は Mary が起きていたいのだと思っているが、実際、Mary はそう望んでいる。)

b) Act₁ (Peter). (Peter は、「Do you want some coffee?」と Mary に向かって言う。)

c) (Peter) $\{\mathbf{B} : (\mathbf{B}.Peter. 13a) \cup \{Act_1(Peter), [[[\pm([K_1] \in \mathbf{D}(Mary))] \in \mathbf{B}(Peter)] \in \mathbf{D}(Peter)] \in \mathbf{B}(Mary)\},$
 $\mathbf{D} : (\mathbf{D}.Peter. 13a) - \{[[[\pm([K_1] \in \mathbf{D}(Mary))] \in \mathbf{B}(Peter)] \in \mathbf{D}(Peter)] \in \mathbf{B}(Mary)\},$
 $\mathbf{I} : (\mathbf{I}.Peter. 13a) - \{Act_1(Peter)\}\}$

(Mary) $\{\mathbf{B} : (\mathbf{B}.Mary. 13a) \cup \{Act_1(Peter), [[[\pm([K_1] \in \mathbf{D}(Mary))] \in \mathbf{B}(Peter)] \in \mathbf{D}(Peter)] \in \mathbf{B}(Peter)\},$

$\mathbf{D} : (\mathbf{D}.Mary. 13a), \mathbf{I} : (\mathbf{I}.Mary. 13a)\}$

(Mary は、Peter が自分がコーヒーを飲みたいのかどうか知りたいのだと思うようになる。)

d) (Mary) $\{\mathbf{D} : (\mathbf{D}.Mary. 13c) \cup \{K_1\},$

$\mathbf{B} : (\mathbf{B}.Mary. 13c) \cup \{K_1 \rightarrow F(K_2), Act_2(Mary) \rightarrow F([\pm([K_1 \rightarrow F(K_2)]) \in \mathbf{B}(Mary)] \in \mathbf{B}(Peter)), [[K_1 \rightarrow F(K_2)] \in \mathbf{B}(Mary)] \in \mathbf{B}(Peter)\},$

$$\text{ter}) \rightarrow F([\![K_1] \in \mathbf{D}(\text{Mary})] \in \mathbf{B}(\text{Peter}))],$$

$$\mathbf{I} : (\text{I.Mary. 13c}) \cup \{\text{Act}_2(\text{Mary})\}$$

((13d)で述べたことに加え、Maryは、コーヒーは眼をさますと思うのでコーヒーを飲みたいと思い、この彼女の欲求が、「Coffee would keep me awake」とPeterに向かって言うことでPeterに伝わるとしており、そのように言うことを意図している。)

e) $\text{Act}_2(\text{Mary})$. (Maryは、「Coffee would keep me awake」とPeterに向かって言う。)

f) $(\text{Peter})\{\mathbf{B} : (\text{B.Peter. 13c}) \cup \{\text{Act}_2(\text{Mary}), [K_1 \rightarrow F(K_2)] \in \mathbf{B}(\text{Mary}), [K_1] \in \mathbf{D}(\text{Mary})\},$

$$\mathbf{D} : (\text{D.Peter. 13c}) - \{[\pm([K_1] \in \mathbf{D}(\text{Mary}))] \in \mathbf{B}(\text{Peter})\},$$

$$\mathbf{I} : (\text{I.Peter. 13c})\}$$

$(\text{Mary})\{\mathbf{D} : (\text{D.Mary. 13d}), \mathbf{B} : (\text{B.Mary. 13d}) \cup \{\text{Act}_2(\text{Mary})\},$

$$\mathbf{I} : (\text{I.Mary. 13d}) - \{\text{Act}_2(\text{Mary})\}$$

(Peterは、Maryが眼をさますためにコーヒーを飲みたいのだと思うようになる。)

PeterがMaryの欲求を満たすよう協力的行動を示す時には、このコミュニケーションに伴う心的状態の変化と行為の列に、その後の列を加えることができる：

g) $(\text{Peter})\{\mathbf{D} : (\text{D.Peter. 13f}), \mathbf{B} : (\text{B.Peter. 13f}) \cup \{\text{Act}_3(\text{Peter}) \rightarrow F(K_1)\},$

$$\mathbf{I} : (\text{I.Peter. 13f}) \cup \{\text{Act}_3(\text{Peter})\}$$

((13f)で述べたことに加え、Peterは、Maryにコーヒーを出せば彼女はコーヒーを飲むと思っている。また、彼はMaryにコーヒーを出すことを意図している。)

h) $\text{Act}_3(\text{Peter})$. (PeterはMaryにコーヒーを出す。)

i) $(\text{Peter})\{\mathbf{D} : (\text{D.Peter. 13g}), \mathbf{B} : (\text{B.Peter. 13g}) \cup \{\text{Act}_3(\text{Peter})\},$

$$\mathbf{I} : (\text{I.Peter. 13g}) - \{\text{Act}_3(\text{Peter})\}$$

$(\text{Mary})\{\mathbf{D} : (\text{D.Mary. 13f}), \mathbf{B} : (\text{B.Mary. 13f}) \cup \{\text{Act}_3(\text{Peter})\},$

$$\mathbf{I} : (\text{I.Mary. 13f}) \cup \{K_1\}$$

(Peterの行為により、彼のコーヒーを出すという意図は充足され消滅する。また、Maryはコーヒーを飲むことを意図している。)

j) K_1 . (Maryはコーヒーを飲む。)

k) $(\text{Mary})\{\mathbf{D} : (\text{D.Mary. 13i}) - \{K_1\}, \mathbf{B} : (\text{B.Mary. 13i}) \cup \{K_1\},$

$$\mathbf{I} : (\text{I.Mary. 13i}) - \{K_1\}$$

(Maryの行為によりコーヒーを飲むという彼女の意図と欲求は充足され消滅する。両者の欲求や意図が両者の行為遂行により充足されたため、この言語ゲームは完了する。)

Glis (1995) が指摘するように、Austin や Searle の言語行為の分析は、単文の発話行為の分析に限定されたものであるが、実際には、複数の文の発話により一つのまとまった言語行為が遂行される場合がほとんどである。また、ここで分析したように、コミュニケーションの参加者たちによる発話行為と行為は、普通、相互に関連しあったものである。我々が提案した態度記述による行為の分析は、共通の目的にそった複数のエージェントによる行為遂行の分析などにも共通の欲求の設定や協力の原則(7)を用いることにより拡張可能であろう。

まとめ

合理的エージェントは、態度を持ち、態度に依存して自らの行動をコントロールする主体である。合理性は、欲求と信念と行動の間に成り立つ関係と考えることができる。自らの行動を契機として自らの欲求を満たすことができると信じる時、その行動を実行することは合理的である。本稿は、発話行為をこのような合理的行為の文脈において分析しようとしたものである。その意味で、本稿は、言語行為論、Wittgenstein の言語ゲームの考え、会話の含意の理論、行為論にまたがる態度伝達の問題を総合的に論じている。

主体と時間の直示の問題、即ち、I (私) と now (今) をいかに扱うかという問題を本稿では議論しなかった。しかし、この問題は、態度の正確な記述には避けて通ることのできない問題である。また、行為論や言語行為論や心の哲学や信念文の意味論などに関して、このアプローチとの関係をさらに明らかにしなければならないが、これらの問題の検討は今後の課題としたい。

注

- 1) 同様の試みに Cohen らによる仕事がある (Cohen and Perrault (1979) 参照)。彼らは、プランに基づいた言語 (plan-based language) を用いて態度伝達を表現する。我々のアプローチが、動的意味論の言語を基本にした心的状態の記述を基に態度伝達を表現するのに対し、彼らは、前条件と効果と効果実現のための手段により構成される伝達行為という演算子を用いて態度伝達を表現する。我々のアプローチは、哲学的議論の道具として人工的エージェントを考へており、自律的ロボットの実際の制作を目標とはしていない。
- 2) Davidson (1963) は、意図は欲求と信念に還元消去できるという立場を取った。これに対し Bratman (1987) は、意図とは何であるか、行為と意図の関係は何であるかを分析し、意図は消去不可能であることを示そうとした。第2節の議論に見るように、本稿では、意図は消去不可能とされ、行為への移行に決定的な役割りを果たすと考えている。
- 3) D 論理式 (Discourse formula, D-formula) については、中山 (1996) 参照。D 論理式は、連言の論理結合詞を “,” で表し、スコールム定数という、その言語の語彙に含まれていない名前を用いるところに統語論的特徴がある。

- 4) この例は、Guttenplan (1994)、p.485からの変形である。
- 5) 精神分析の対象となる症例は、この自己意識の原則がこわれるところで表れると考えることもできる。特に、 $([K] \in \mathbf{D}(S), \neg([K] \in \mathbf{D}(S)] \in \mathbf{B}(S)))$ ということを作り立たせる K によって表されるある事態が存在する場合、即ち、 K ということを知っているが自分はそのことを知らないという事態が成り立つ場合が問題とされる。治療は、自己意識の回復により達せられると考えられよう。
- 6) 本稿では、時間と態度の関係を正確には記述しない。実際、心的状態は、その担い手としての主体と時間に依存し、態度の内容は、その態度がある主体により所有されている時間に関わっている。これらの問題を詳しく論じることは今後の課題として残された。
- 7) この時、 S は、Goal の実現を欲するがために、彼がその手段として有効だと信じる $\text{Act}_1(H)$ の実現を欲するのである。だから、 $\text{Act}_1(H)$ に関する S の欲求は、Goal に関する S の欲求を前提にしていると言えるだろう。
- 8) Nakayama (1998) 参照。Searle (1979) に従うと、指令型の発語内行為 (directives) は、要求するとか命令するとかのように、聞き手に何かをさせようとするという特徴を持つ。
- 9) 話者と聞き手が同様の言語理解を持つことは、発話行為の遂行による態度伝達の成功を助ける。語句の正しい意味の習得は、よりよいコミュニケーションの形成のための手段なのである。
- 10) Wittgenstein (1953) は、判断の一致は生活形式の一致を前提としていることを指摘している (§241 参照)。

参考文献

- Austin, J. L. (1962) *How to Do Things with Words*, Oxford UP.
- Bratman, M. E. (1987) *Intention, Plans, and Practical Reasoning*. (門脇・高橋訳『意図と行為』産業図書)
- Cohen, P. R. and Perrault, C. R. (1979) "Elements of a Plan-Based Theory of Speech Acts", *Cognitive Science*, 3 (3): pp. 177-212. Reprinted in: B. Webber and N. Nilson (eds.), *Readings in Artificial Intelligence*, Morgan Kaufman Pub., 1981, pp. 478-495.
- Davidson, D. (1963) "Actions, Reasons, and Causes," *Journal of Philosophy* 60, pp. 675-700.
- Glis, M. L. (1995) *Speech Acts and Conversational Interaction*, Cambridge UP.
- Grice, P. (1989) *Studies in the Way of Worlds*, Harvard UP.
- Guttenplan, S. (ed.) (1994) *A Companion to the Philosophy of Mind*, Blackwell.
- 中山康雄 (1996) 「動的意味論と有限クラス」『科学哲学』Vol. 29, 1996, pp. 139-153.
- Nakayama, Y. (1998) "Attitudes and Classification of Utterances," in: *Annals of the Japan Association for Philosophy of Science*, Vol. 9 No. 3, pp. 37-53.
- 中山康雄 (1998) 「認識の歴史性」『年報人間科学』Vol. 19, pp. 1-19.
- Searle, J. R. (1969) *Speech Acts—An Essay in the Philosophy of Language*, Cambridge UP.
- Searle, J. R. (1979) *Expression and Meaning—Studies in the Theory of Speech Acts*, Cambridge UP.
- Sperber, D and Wilson, D. (1986) *Relevance: Communication and Cognition*. Blackwell. (内田他訳『関連性理論—伝達と認知—』(1993) 研究社出版)
- Wittgenstein, L. (1953) *Philosophische Untersuchungen*. (藤本隆志訳『哲学探究』(1976) 大修館)

Communication between Rational Agents

Yasuo NAKAYAMA

Actions can be analyzed in the context of problem solving activities. Utterances are actions that are performed to convey speaker's attitudes. They are performed, because their performance contributes to solving some problems. In this paper, I try to clarify human speech activities by using a model of artificial rational agents.

An artificial agent is introduced as an agent that has a body and propositional attitudes like belief, desire, and intention. Its mental state is described as a set of B-box (a set of beliefs), D-box (a set of desires), and I-box (a set of intentions). These propositions are described by Discourse formulas that is formulas in a framework of dynamic semantics. This agent considers a desire as realizable when he believes that there is a performable action that will cause an event that satisfies it. In this case, the agent intends to perform this action. An agent can also try to solve a problem by conveying to his neighbor his desire for an action from his neighbor. This conveyance of desire can be realized by an utterance, where the speaker *S* believes that this utterance will convey his desire to the hearer *H* and that *H* will perform the action that *S* wants. In this way, utterances can be analyzed by describing speaker's attitudes.

In this paper, this approach is applied to the description of actions, speech acts, Wittgenstein's concept of primitive language games, and conversational implicatures.