特　集　「エージェントの基礎と応用」

エージェントのメンタルモデル

Models of Mental States of Agents

片桐　恭弘*
Yasuhiro Katagiri

*　ATR 知能映像通信研究所
ATR Media Integration & Communications Research Laboratories.

1995年5月8日　受理

Keywords: rational agents, knowldge and action, intention, logical formulation.

1. はじめに

環境に関する知識に基づき，達成すべき目標を目指して行為を遂行することは，知的主体の備える基本的特徴の一つである．このような能力を備えたエージェントは合理的エージェントと呼ばれる．本稿では，合理的エージェントの設計の基礎となる論理モデルについて述べる．はじめに，知識・信念という心的状態と行為との関係について様相論理を用いたモデル化を取り上げる．次に，合理性概念の中核となる意図の扱いについて触れる．さらに，これらのモデルの応用の実例として言語コミュニケーションにおける言語行為の扱いを紹介する．最後に，最近の状況依存性に関する話題にも簡単に触れる．

2. 知識と行為の論理モデル

2・1 知識と信念

知識や信念などの心的状態を形式的にモデル化するために，本来必然性・可能性を形式化するために開発された様相論理(modal logic)が利用できることに最初に気づいたのは，哲学者のHintikkaであった[Hintikka 62]．述語論理と比較したときの様相論理の特徴は，可能世界を意味論的対象として導入した点にある．可能世界の哲学的な位置づけに関してはさまざまな議論がある．しかし，我々の目標のためには，可能世界は世界のとり得る状態のさまざまな可能性と考えればよいだろう．例えば，ゲームの探索木中の各節点は，ゲームの進行に伴ってとり得るゲーム状態のさまざまな可能性を表す．それらの可能なゲーム状態はみなそれぞれ独立な可能世界と考えることができる．可能世界のすべてで真なる命題が必然的に真な命題である．

可能世界の概念を利用すると，エージェントの知識・信念に関する命題の真偽を次のように定めることができる．知識と信念はほぼ同様に定義されるので，以下ではまず知識を例として述べる．可能世界のなかには，エージェントの知っていることと整合的なものと整合的でないものとがある．エージェントの知識と整合的な可能世界をエージェントにとっての知識代替世界(epistemic alternative)と呼ぶ．すると，エージェントが知っていることはエージェントの知識代替世界のすべてで真である．成り立つかどうか知らないことは，知識代替世界のうちのある世界では真だが別の世界では偽となっている．形式的には，可能世界の集合 W と W 上の二項関係 $\mathcal{K}_a$ を用いて，ある世界 $w \in W$ でエージェント a がある命題の成立を知っているとは，以下のように定義される．

$$w \models \mathrm{KNOW}_a(p)$$

iff

$$\forall w' \mathcal{K}_a(w, w') \supset w' \models p$$

$\mathcal{K}_a$ は到達可能性関係と呼ばれ，$\mathcal{K}_a(w, w')$ は，w' が w にいるエージェント a にとって知識代替世界となっていることを示す．

この定義は，必然性に関する様相論理で標準的なKripke意味論と同型である．したがって，健全かつ完全な公理化が可能であることが知られている．ただし，知識や信念は各エージェントごとに固有の様相演算子 KNOW_a によって表現される．また，到達可能性関係

$\mathcal{K}_a$もエージェントごとに定義される．具体的な公理系は到達可能性関係にどのような性質を要求するかに依存するが，以下に代表的な公理化の一つを示す．これは$\mathcal{K}_a$として最も強い同値関係を仮定した場合の公理系で，様相論理のS5と呼ばれる体系に相当する．

(1) 公　理

(分　配)　$\mathrm{KNOW}_a(p)\wedge\mathrm{KNOW}_a(p\supset q)\supset\mathrm{KNOW}_a(q)$

(知　識)　$\mathrm{KNOW}_a(p)\supset p$

(肯定内省)　$\mathrm{KNOW}_a(p)\supset\mathrm{KNOW}_a(\mathrm{KNOW}_a(p))$

(否定内省)　$\neg\mathrm{KNOW}_a(p)\supset\mathrm{KNOW}_a(\neg\mathrm{KNOW}_a(p))$

(2) 推論規則

(Modus ponens)　$\vdash p$かつ$\vdash p\supset q$ならば$\vdash q$

(Necessitation)　$\vdash p$ならば$\vdash\mathrm{KNOW}_a(p)$

分配公理は，エージェントが知識に基づいて推論可能なことを示す．知識公理は知識の内容は正しいという直観に相当する．肯定および否定内省公理は，自分自身の知識の有無自体に関する知識を保証する．信念に関しては，信念の内容がつねに正しいと限らないため，知識公理の代わりに下の公理を用いることが多い．これは，信念の内容は無矛盾であることを示している．

(信　念)　$\mathrm{BEL}(\neg p)\supset\neg\mathrm{BEL}(p)$

これらの公理化が知識や信念に関する我々の直観を完全に反映しているというわけではないし，そもそも知識に関して万人に共通で安定した直観があるという保証もない．このような論理的定式化は，知識や信念の関与する現象を形式的にモデル化するための手段となるだけでなく，知識や信念に関する我々の直観を明示的な形で表現し吟味する手段としても重要な役割を果たす．直観との食違いの一部分は公理の選択によって吸収可能であるが，論理的形式化をとる限り避けられない問題もある．分配公理は可能世界による知識・信念の定式化と不可分であるが，これを認めることはエージェントに論理的帰結に関する無制限の推論能力を認めることに相当する．これは論理的全知の問題として知られている．

2・2　共有知識・共有信念

複数のエージェントが協力して行動するときには，行動に関する合意あるいは情報の共有が必要である．言語によるコミュニケーションによって通常もたらされるのも話し手と聞き手の間での情報の共有である．このような情報共有は単に複数のエージェントが同じ知識・信念を抱いているという状態とは異なっている．明日10時から会合を持つという場合，単に参加者各人が会合が明日10時からだという信念を持つだけでは合意とはいえない．当然お互いに他の参加者も会合に関して同じ情報を持っていることを信じている必要がある．さらにそれだけでも不十分で，任意の深さの信念の埋込みが必要なことは，埋込みのどこかのレベルで誰かが会合に関する情報を持っていない場合には合意とは言い難いことからわかる．このような情報の共有状態は，共有知識あるいは共有信念と呼ばれる．

共有知識・共有信念の重要性を示す現象として，コンウェイパラドックスがある．これは以下のようなものである．「太郎と次郎が初対面だとする．2人とも京大の卒業生だが相手のことを知らない．自分の出身校は当然知っているので，この場合2人とも「どちらか一方は京大出身である」ことを知っている．そこへ花子がやってきて2人に相手の出身校がわかるかと尋ねる．何回尋ねても当然2人はわからないと答えるだろう．次に同じ状況で，まず花子が2人に「どちらか一方は京大出身ですよ」と告げ，それから2人に相手の出身校がわかるか尋ねるとしてみよう．今度は，最初は2人ともわからないと答えるが，その答えを聞けば太郎は次のように推論することができる．もし次郎が京大出身でなければ花子の言葉を聞いて太郎が京大出身であることを知るだろう．それなのに次郎がわからないと答えたのは次郎自身が京大出身だからにほかならない．次郎も同様に推論することが可能だから，花子の2度目の問いかけに対しては，2人とも相手が京大出身であると答えることができる*．ところが，花子の発言自体はすでに太郎も次郎も知っていたことを述べたのに過ぎない．したがって，発言は何も新しい情報を追加しないはずである．それにもかかわらず二つのシナリオはこのように明白に違った結果をもたらす．なぜだろうか？」

このパラドックスは，「どちらか一方は京大出身である」ということを2人とも知っているというだけの状態と，その情報が共有知識となっている状態との違いを如実に示している．

様相論理を用いたモデル化では，グループGの構成員の持つpに関する共有知識$\mathrm{MK}_G(p)$は以下の式の最大不動点として与えられる．

$$\mathrm{MK}_G(p)\equiv E_G(p\wedge\mathrm{MK}_G(p))$$

ただし，

$$E_G(p)\overset{\mathrm{def}}{\equiv}\bigwedge_{a\in G}\mathrm{KNOW}_a(p)$$

* もちろん2人が出身校にふさわしく知的であればだが．

また，共有知識成立の条件は可能世界を用いると以下のように記述される．

$$w \models \mathrm{MK}_G(p)$$

iff

$$\forall w' \mathcal{K}^*(w, w') \supset w' \models p$$

ただし，$\mathcal{K}^*$ は $\mathcal{K}_a$ の総和の推移閉包であり，以下のように定義される．

$$\mathcal{K}^*(w, w') \overset{\text{def}}{=} \begin{cases} \exists a \mathcal{K}_a(w, w') & \text{or} \\ \exists a, w'' \mathcal{K}_a(w, w'') & \\ \quad \wedge \mathcal{K}^*(w'', w') & \end{cases}$$

2・3 知識と行為

電話番号を知っていればその相手と電話で連絡をとることができる．このように知識は行為に影響を与える．また，電話帳を調べることによって電話番号の知識を得ることができる．すなわち，行為は結果として新しい知識をもたらす．そして新たに獲得された知識はさらに次の行為に反映される．このように，知的エージェントの行動では知識と行為とは密接に結びついている．Moore[Moore 85]は，**2・2節**で述べた知識・信念の形式化に行為の形式化を加えて知識と行為の相互関係のモデル化を与えている．基本的アイディアは，行為の実行による状態変化を二つの可能世界間の関係と捉え，知識・信念の場合と同様に行為に関しても様相論理による形式化を与えるというものである．

エージェント a が行為 ACT を実行するというイベントを項 $\mathrm{DO}(a, \mathrm{ACT})$ によって表す．さらに，その行為実行の結果 p が得られるという命題を $\mathrm{RES}(\mathrm{DO}(a, \mathrm{ACT}), p)$ と表す．RES は行為実行に関わる様相演算子である．行為実行のイベントによる可能世界間の遷移関係は，当然誰がどの行為を実行したかというイベントに依存する．そこでこの遷移関係をイベント e を新たに引数として $\mathcal{R}(e, w, w')$ によって表す．w でイベント e に相当する行為の実行が不可能な場合には，$\mathcal{R}(e, w, w')$ を満たす w' が存在しない．また，行為実行の結果は決定的と想定する．すなわち，e, w を固定したとき，$\mathcal{R}(e, w, w')$ を満足する w' はたかだか一つである．すると，命題 $\mathrm{RES}(\mathrm{DO}(a, \mathrm{ACT}), p)$ の真偽は以下のように定義される．

$$w \models \mathrm{RES}(\mathrm{DO}(a, \mathrm{ACT}), p)$$

iff

$$\exists w' \mathcal{R}(\mathrm{DO}(a, \mathrm{ACT}), w, w') \wedge w' \models p$$

さらにこの定義を利用すると，エージェントの能力の概念を間接的に規定することができる．エージェント a が行為 ACT によって p を実現する能力を持っていることを $\mathrm{CAN}(a, \mathrm{ACT}, p)$ と表すと

$$\forall a(\exists X(\mathrm{KNOW}_a((X = \mathrm{ACT}) \wedge \mathrm{RES}(\mathrm{DO}(a, \mathrm{ACT}), p))) \supset \mathrm{CAN}(a, \mathrm{ACT}, p))$$

が成立する．この定義は，能力に関する以下の性質を形式化したものである．

- エージェントは実行すべき行為を知っている．
- エージェントは行為が実行可能なことを知っている．
- エージェントは行為実行の結果 p がもたらされることを知っている．

これらを用いて，電話番号に関する知識が連絡をとる行為を可能にすることは次のように捉えられる．

$$w \models \exists x \mathrm{KNOW}_a(x = \mathrm{PN} \wedge \mathrm{RES}(\mathrm{DO}(a, \mathrm{DIAL}(\mathrm{PN})), \mathrm{TALKTO}(\mathrm{B})))$$

が成り立てば，上の式から

$$w \models \mathrm{CAN}(a, \mathrm{DO}(a, \mathrm{DIAL}(\mathrm{PN})), \mathrm{TALKTO}(\mathrm{B}))$$

が成り立つ．このとき，$\mathcal{K}_a(w, w')$ が成り立つすべての世界 w' で PN は同じ電話番号 x になる．これがエージェント a が電話番号を知っていることに相当する．そして w' から行為 $\mathrm{DO}(a, \mathrm{DIAL}(\mathrm{PN}))$ によって遷移する先の世界 w'' では，必ず TALKTO(B)が成立する．

同様に，電話帳を引くことによって電話番号を知ることができるのは次のように捉えられる．

$$w \models \exists x \mathrm{KNOW}_a(x = \mathrm{LOOKUP} \wedge \mathrm{RES}(\mathrm{DO}(a, \mathrm{LOOKUP}), \exists z \mathrm{KNOW}_a(z = \mathrm{PN})))$$

が成り立てば

$$w \models \mathrm{CAN}(a, \mathrm{DO}(a, \mathrm{LOOKUP}), \exists z \mathrm{KNOW}_a(z = \mathrm{PN}))$$

が成り立つ．このとき，$\mathcal{K}_a(w, w')$ が成り立つすべての世界 w' で LOOKUP は同じ行為に対応する．これがエージェント a が電話帳を引くという行為がどういうものかを知っていることに相当する．さらに w' から行為 $\mathrm{DO}(a, \mathrm{LOOKUP})$ によって遷移する先の世界 w'' では，$\mathcal{K}_a(w'', w''')$ が成り立つすべての世界 w''' で PN が同じ番号 x になる．これが行為実行の結果エージェントが電話番号を知ることに相当する．

このように様相論理の枠組みは，知識と行為との密接な相互依存関係を数学的に厳密な形で表現し取り扱うための方法を提供してくれる．これらは，エージェントの仕様記述，およびエージェントによる行為の結果に関する推論や計画立案推論の基礎となる．

3. 合理的エージェント

知的エージェントは自身の生存をはじめとしてさまざまな目標を持っており，それらの目標の実現のために，合理的に行為を選択し実行する．すなわち，その行為を実行すれば目標が達成できることを知っている(信じている)ような行為，あるいは少なくともその行為が目標の実現に最適であると知っている(信じている)ような行為を選んで実行する．そのような特徴を持った行為選択を行うエージェントは合理的エージェントと呼ばれる．合理的エージェントは，行為選択のために現在の世界の状態に関する情報，行為による世界の状態変化に関する情報を備えていなければならない．それらの情報に基づくエージェントの行為選択の推論は，熟考(deliberation)あるいは計画立案(planning)などと呼ばれる．初めに最適な行為を推論し，その後にそれを実行するという手順が可能ならば理想的であるが，現実には変化する環境，限られた情報，限られた計算資源によってそのような推論と行為との逐次的な分離は不可能である．エージェントの意図は，変化する環境，限られた情報・資源のもとでエージェントの将来の行為を規定し，エージェントの行動に一貫性を与える機能を果たしている．以下ではまず，単独エージェントの意図と行為に関する論理モデルについて述べ，さらに，複数のエージェントが共同で行う行為とその背後の意図の取扱いについて述べる．

3・1 意図と行為

我々の日常生活において，意図は我々の行動に密着している．例えば，我々は学会での発表を事前に決断し，申込み，原稿を用意し，発表準備をして当日発表を行う．これらいずれの行為も意図的に遂行される．さらに人と人との協調においても，コミュニケーション，交渉，妥協などの手段を用いることによって，我々は意図の伝達・共有を行っている．このように，目標を持ち，その実現に向けて行動選択を行う合理的エージェントにおいては，意図は行為の選択に重要な役割を果たしている．

行為の背後の意図では，行為遂行の時点で存在する行為意図(intention in action)と，エージェントの将来の行為を規定する将来に向けられた意図(future-directed intention)とを区別することができる．前者が，例えば，手をすべらせて皿を落として割るのと，うっぷんを晴らすためにわざと皿を投げて割るのとの差をもたらす意図であるのに対して，後者は，明日東京へ出かけて会合に参加するつもりだという場合の意図である．

(1) 意図の機能

Bratman [Bratman 87]は将来の行為に向けられた意図の機能として以下の3点をあげている．

- 計画立案の入力として機能する．
- 計画立案の範囲を制約する．
- 行為を引き起こす原因となる．

学会発表をすると決定した時点では具体的な発表準備まで完了している必要はない．しかし，ひとたび発表するという意図を持ったならば，我々は実際の発表へ向けて準備を進める．発表という意図はその後の具体化のための計画立案に対して入力として機能する．同様に，ひとたび発表すると決断したならば，例えば学会の当日に友人と旅行に出かけるというように発表と競合するような予定を組んだりはしない．すなわち，発表の意図はその後の計画立案の範囲に制約を与えている．そして最終的に発表の意図は，実際の発表という行為を引き起こす原因となる．このように，意図には一定の範囲の行為の遂行あるいは状態の達成を持続的に追求するという性質がある．意図の持つこのような特性はcommitmentとも呼ばれる．

(2) 意図と副作用

意図の持つ性質を論理的に記述する際に問題となる性質に，副作用に関する意図の非存在がある．エージェントがpという状態の達成を意図しており，pの達成は副作用としてqの実現を伴うことをエージェントがたとえ知っていたとしても，エージェントはqの実現を意図しているとは言い難い場合がある．例えば，サッカー選手が重病にかかって足を切断しないと助からない状態になったとしよう．医師は彼の足を切断することによって彼の命を助けることを意図する．しかし，医師は同時にそれは彼からサッカーを奪うことになることを知っている．そのような場合に，手術を決断した医師がサッカー選手からサッカーを奪うことを意図しているとは直観的に言い難い．

(3) 意図的行為に関する信念

くしゃみをする，手をすべらせてコップを落とすというように，偶然的に起こる行為は意図的行為とは通常呼ばれない．したがって，意図的行為は満たすべき条件として，行為遂行の前にどの行為を遂行するかあらかじめエージェント自身が決めており，エージェントはその行為を遂行すると信じているという趣旨の条件が必要と考えられる．一方，意図の機能で述べたように，行為遂行の開始の時点で目標達成に至る行為の細部に至るまで完全に確定していることはまれであ

り，またその必要は通常ない．ところが，幸福になりたい，金持ちになりたいなど，単に漠然とした目標はあるが，それを実現するための具体的な方策をなんら持ちあわせないという場合，それらは単なる願望であって意図とは言い難い．意図的行為のためには行為遂行の前にエージェントは行為に関する少なくとも部分的な計画を持っており，それに従って行動することを信じているという趣旨の条件が妥当である．

意図的行為に関する信念の問題の複雑性を示す例として，以下のような例が知られている[Chisholm 66, Searle 83]．太郎は叔父の殺害を意図して拳銃を持って叔父の家へ車で向かった．途中太郎は叔父を殺すという考えに興奮して運転を誤り歩道に乗り上げて歩行者をひき殺してしまった．その被害者はたまたま太郎の叔父であった．この場合，太郎の殺害の意図が叔父の死の原因となってはいるが，太郎のやったことは意図的な殺害行為とは言い難い．

(4) 意図の論理モデル

Cohen と Levesque[Cohen 90a]は，単一エージェントの持つ意図の論理的定式化のために持続目標(persistent goal)の概念を提案している．単一のエージェント a の持つ p を達成するという将来に向けられた意図の中心部分は，信念，知識，目標を用いて以下のように定義される持続目標に帰着される．

$$\text{P-GOAL}_a(p) \stackrel{\text{def}}{=} \text{BEL}_a(\neg p) \wedge \text{GOAL}_a(\Diamond p) \wedge \text{KNOW}_a(\text{UNTIL}(\text{BEL}_a(p) \vee \text{BEL}_a(\Box \neg p), \text{GOAL}_a(\Diamond p)))$$

ここで，KNOW_a，BEL_a，GOAL_a はそれぞれ知識，信念，目標の認識様相を，$\Box$，$\Diamond$，UNTIL はそれぞれ always, eventually, until の時間様相を表す．持続目標は将来に向けられた意図の次のような性質を捉えている．

- 将来に向けられた意図は，現在成り立っていないと思っていることを将来達成することを目標とする(第1，2項)．
- 目標が達成された，あるいは達成不可能であるということがわかるまで目標達成を目指し続ける(第3項)．

持続目標を用いて，行為 ACT を実行するという意図は以下のように定義される．

$$\text{INT}_a(\text{ACT}) \stackrel{\text{def}}{=} \exists X\, \text{KNOW}_a(X = \text{ACT}) \wedge \text{P-GOAL}_a\ (\text{DONE}\ (\text{UNTIL}(\text{DONE}(\text{DO}(a, \text{ACT})), \text{BEL}_a(\text{DOING}\ (\text{DO}(a, \text{ACT}))))?; \text{DO}(a, \text{ACT})))$$

ここで，DONE, DOING はイベントから命題を構成する様相演算子であり，DONE(e), DOING(e) は，それぞれイベント e がちょうど起こったところ，e が現在進行中だという命題を表す．また，p? は命題 p の成立をテストするイベントであり，$e_1; e_2$ はイベント e_1 に引き続いてイベント e_2 が起こるという複合イベントを表す．

上の意図の定式化は，持続目標で示した意図の特徴のほかに意図的行為の持つ次の性質を捉えている．

- 意図的行為ではエージェントはどの行為を遂行するか知っている．
- 意図的行為は偶然に遂行されるのではなく，事前にエージェントは当該行為の遂行を信じている．

意図のこの定式化は，具体的な計画立案の過程，および計画の部分性には立ち入っていない．しかし，意図の持つ直観的性質のかなりの部分を論理的に明示的に表現している．例えば，上の定義のうち，持続目標の2番目の持続的に目標達成を目指し続けるという性質が，単純な目標と異なり，意図が commitment の性質を持つことを保証している．また，この論理的定式化が副作用に関する意図の非存在を満足することは $\text{P-GOAL}_a(p)$ と $\text{KNOW}_a(p \supset q)$ から $\text{P-GOAL}_a(q)$ が帰結しないことによって示される．

3・2 共同意図と共同行為

ピアノで連弾をする，連れ立ってどこかへ出かける，対話を行うなど，複数のエージェントが共同して一つの目標達成を目指すときに，共同意図が行為の背後に想定される．Searle[Searle 90]によれば，共同意図は，目標達成を目指した個々人の意図の単なる総和以上のものである．これは，例えば2人がたまたまピアノのそれぞれのパートを同時に弾いただけではピアノ連弾にならないことに表れている．

複数エージェントによる共同行為の背後の各エージェントの持つ心的状態のモデルとして Cohen と Levesque[Cohen 91, Levesque 90]は共同意図の論理モデルを提案している．共同意図を捉えるために共同持続目標(joint persistent goal)の概念を導入する．これは，前述の持続目標を複数エージェントに拡張したもので，共有信念，共有知識，共有目標を用いて以下のように定義される．

$$JPG_G(p) \overset{def}{=}$$
$$B_G(\neg p) \wedge MG_G(\Diamond p) \wedge$$
$$MK_G(UNTIL(MB_G(p) \vee MB_G(\Box \neg p), WMG_G(p)))$$

ここで，MB，MK はそれぞれ，共有信念，共有知識の認識様相を表す．認識様相 MG，WMG はそれぞれ，共有目標，弱共有目標と呼ばれ，以下のように定義される．

$$MG_G(p) \overset{def}{=} MB_G(\underset{x \in G}{\wedge} GOAL_x(p))$$

$$WMG_G(p) \overset{def}{=} MB_G(\underset{x \in G}{\wedge} WG_{x,G}(p))$$

ただし，

$$WG_{x,G}(p) \overset{def}{=}$$
$$[\neg BEL_x(p) \wedge GOAL_x(\Diamond p)]$$
$$\vee [BEL_x(p) \wedge GOAL_x(\Diamond MB_G(p)]$$
$$\vee [BEL_x(\Box \neg p) \wedge GOAL_x(\Diamond MB_G(\Box \neg p))]$$

単独エージェントの意図と同様に，共同行為の背後の共同意図は，共同持続目標を用いて次のように定義される．

$$JI_G(ACT) \overset{def}{=}$$
$$\exists X(MK_G(X = ACT) \wedge$$
$$JPG_G$$
$$(DONE$$
$$(UNTIL(DONE(DO$$
$$(G, ACT)),$$
$$MB_G(DOING\ (DO$$
$$(G, ACT)))) ?;$$
$$DO(G, ACT)))$$

この共同持続目標の定義は共同意図の次のような性質を捉えている．

- 共同意図は，現在成り立っていないとグループ構成員全員に共通に信じられていることを将来達成することを目標とする(第1，2項)．
- 目標が達成された，あるいは達成不可能であるということがグループの構成員全員の共通了解となるまで目標達成を目指し続ける(第3項)．
- 目標達成に関して個人的に知り得た情報はグループ全員での共有を目指す(弱共有目標)．

意図の定義と共同意図の定義との比較からわかるように，共同意図は，概略，意図のうちの信念，知識，目標の認識様相を対応する共有版に置き換えることによって得られる．単純な置換えとの相違点は，弱共有目標の利用にある．共同行為では，目標が達成されたことをエージェントの誰かが個人的に知り得たときにはそれを構成員すべてに知らせる必要がある．そうでないと，すでに目標が達成されたことを知らずにむだな行為を続けるエージェントが生じる可能性がある．例えば，誰かの落とした財布を複数の人が共同で探すというような場合，誰かが財布を見つけたならば，その発見の事実を皆に知らせることによって初めて共同探索は終了する．発見事実の周知がなければ，多くのエージェントは発見の事実を知らずにむだに探索を続けることになってしまうが，それでは共通目標を目指した共同行為とは言い難い．目標が達成不可能な場合についても同様である．例えば，複数の人が連れ立ってある目標地へ向かって移動しているときに，誰かが何らかの事情で目標地まで行けなくなった場合，その人は勝手にグループから離れるのでなく，その前にグループ構成員全員に共同目標達成に関わる変更を伝える．弱共有目標は共同行為の持つこのような性質を捉えるために導入された．

4. 言 語 行 為

真理条件意味論に代表されるように言語の意味の考察が主に言語表現の持つ内容に着目するのに対して，言語行為の考え方はコミュニケーションにおいて言語の果たす機能に着目する．ここでは，合理的エージェントのメンタルモデルの応用として言語行為の論理モデルについて述べる．

4・1 言語行為の考え方

Austin[Austin 62]は，裁判官による判決宣告，進水式での船の命名，オリンピックの開会宣言など発話によって世界の状態に実際的な変化をもたらす遂行的発話(performative uttrance)の分析を通じて，言語の持つ，相手に何かを伝達する，警告する，依頼する，命令する，約束するなど言語コミュニケーションにおいて何らかの効果をもたらす機能に着目した．このように言語を何らかの効果をもたらす行為の遂行と捉えるのが言語行為(speech act)の考え方である．

Searle[Searle 69]は言語行為を大きく，主張型(assertive)，要求型(directive)，約束型(commissive)，表出型(expressive)，宣言型(declaration)の5種類に分類した．

言語行為では，内容の真偽より発話が所期の効果を生み出すために満たすべき条件が問題となる．例えば，判決の宣告は担当の裁判官以外が行っても効力を持たないし，担当の裁判官であっても自宅で判決文を読み上げたのでは判決の宣告とはならない．言語行為が効力を持つための条件は適切性条件(felicity condition)と呼ばれる．Searle[Searle 69]は言語行為の適切性条

件として以下の五条件をあげている．

(1) **正常入出力条件**

正常な言語行為の遂行に必要な基本条件．話し手・聞き手は言語が理解できる，発話を聞き取る・読み取ることができるなど．

(2) **命題内容条件**

発話の命題内容に関する条件．例えば依頼の場合，命題内容は聞き手の将来の行為でなければならない．

(3) **準備条件**

言語行為の本質ではないが，言語行為の成立に必要な条件．例えば依頼の場合，聞き手は依頼内容の行為の遂行が可能，話し手は聞き手が行為を遂行可能と信じている，何もしない場合に聞き手が依頼行為を遂行するかどうかは話し手・聞き手どちらにも不明な必要がある．

(4) **誠実性条件**

言語行為の誠実な遂行のために必要な条件．誠実な依頼のためには，聞き手が依頼された行為を遂行することを話し手が実際に望んでいる必要がある．

(5) **本質条件**

言語行為を成立させるために本質的な条件．依頼の言語行為が成立するためには，発話が聞き手に依頼内容の行為を行わせようとする試みである必要がある．

4・2 言語行為の論理モデル

言語行為は，言語コミュニケーションを構成する基本単位として提唱されたが，Cohen と Levesque [Cohen 90b]は合理的エージェントの概念を用いることによって言語行為は合理的エージェントによる行為の一種と捉えられ，言語行為のさまざまな条件は合理性によって自動的に導かれることを示した．中心的なアイディアは以下のものである．

・言語行為は聞き手の心的態度を含む世界の状態に一定の変化を引き起こすことを目指した話し手による意図的な試みである．

(1) **意図的な試み**

まず，エージェント a による意図的な試みは複合的行為として次のように定義される．

$$\begin{aligned}&\mathrm{ATTEMPT}(a, e, p, q)\overset{\mathrm{def}}{=}\\&\quad[\mathrm{BEL}_a(\neg p)\\&\quad\wedge\mathrm{GOAL}_a(\mathrm{HAPPEN}(e\,;p?))\\&\quad\wedge\mathrm{INT}_a(e\,;q?)]?\,;e\end{aligned}$$

ただし，HAPPEN は DONE，DOINGと同様イベントから命題を得る様相演算子で，HAPPEN(e) はイベント e がこれから起こるところだという命題を表す．エージェント a は行為 e によって現在成立していない状態 p の成立を目標とする．p は単なる目標なので，e の結果実際に実現されないこともあり得る．その場合でもエージェント a は e によって少なくとも q は実現することを意図している．

(2) **依頼の言語行為**

意図的な試みを用いて依頼の言語行為は複合的行為として以下のように定義される．

$$\begin{aligned}&\mathrm{REQUEST}(S, H, e, \mathrm{ACT})\overset{\mathrm{def}}{=}\\&\quad\mathrm{ATTEMPT}\\&\quad(S, e, \phi, \mathrm{MB}_{\{S,H\}}(\mathrm{GOAL}_S(\phi)))\end{aligned}$$

ただし，

$$\begin{aligned}&\phi=\mathrm{INT}_H(\mathrm{ACT})\\&\quad\wedge\Diamond\mathrm{DONE}(\mathrm{DO}(H, \mathrm{ACT}))\end{aligned}$$

ϕ は，話し手 S は，行為 e(典型的には命令文の発話)によって聞き手 H が依頼内容の行為 ACT を意図し，将来それを実際に実行することを表す．依頼は，ϕ の実現を目標とし，同時に最低限その目標を公にすることを意図した話し手による意図的な試みである．

このように依頼言語行為を定義すると，一定の条件を満足する文脈のもとでの命令文の発話が依頼の言語行為となることが，命令文の機能および合理的エージェントの協力的振舞いの性質から論理的に導かれることを示すことができる．

(3) **依頼行為の適切性条件**

依頼の言語行為に関する適切性条件は論理モデルのなかに以下のように位置づけられる．

(a) **正常入出力条件**

命令文の発話イベントが正常に起こるという条件．

(b) **命題内容条件**

命令文の内容は聞き手の行うべき行為 ACT でなければならない．

(c) **準備条件**

・聞き手による依頼行為 ACT の実行可能性およびそれに関する信念の条件には，GOAL の定義から話し手・聞き手とも聞き手が ACT をけっして実行しないわけではないと信じていることが導けることが対応する．

・何もしない場合に聞き手が依頼行為 ACT を実行するかどうかわからないという条件は，意図的な試みの定義から，聞き手が ACT を意図し，将来実行する (ϕ) と話し手がすでに信じている場合には，依頼の言語行為とはならないことに相当する．

(d) **誠実性条件**

REQUESTの定義から，話し手は聞き手がACTに対する意図を持ちそれを実行すること(ϕ)を目標としていることが保証される．

(e) **本質条件**

REQUESTの定義は意図的試みATTEMPTによっている．

合理的エージェントの概念に基づく言語行為の論理モデルの特徴は以下のようにまとめられる．

- 言語コミュニケーションを行うエージェントに関して，実際の実装レベルに立ち入る前の抽象レベルでの論理的仕様を与えている．
- それによって言語行為およびそれに関与する諸概念に論理的に明確な定式化を与えている．
- 言語行為自体を基本的単位ととる必要はなく，意図的な試みという概念を利用することによって，言語行為を合理的エージェントのとる行為という一般的な枠組みのなかで記述している．

5. 状況依存性のモデル

これまで述べた様相論理に基づく合理的エージェントの論理モデルは，エージェントを外側から眺める理論家の立場からの定式化である．理論家はエージェントとそれを取り巻く環境の双方のすべてを考慮に入れることができる．それに対してエージェント自体は自分を取り巻く環境に関する部分的な情報しか得ることはできない．しかもエージェントの心的状態・行為のどちらも環境に依存している．例えば，握りしめたこぶしを降りおろすという同じ手の動きでも，ドアを前にして行えばノックという行為になるが，ナイフを握りしめた状態で太郎のすぐ前で行えば，太郎を刺すという行為になる．このように行為においてはエージェントの直接の制御下にある動作(movement)と，その動作が一定の条件を満たす環境下で起こったことによってもたらされる結果(achievement)とを区別する必要がある[Israel 93]．

知識に関する様相論理的定式化は知識をエージェントの内部状態とエージェントを取り巻く環境状態との相関関係と捉えている．一方，知識には行為を引き起こす原因としての機能もある．エージェントがどの動作を実行するかは当然エージェントの内部状態にのみ依存する．前者の意味での知識は理論家から見た知識であって，環境に関する不完全な情報しか持たないエージェントには，この意味での知識は利用不可能である．後者の内部状態のみがエージェントにとって利用可能であるが，内部状態が現実にどのような知識に対応するかはエージェントを取り巻く環境に依存し，エージェントの知り得る範囲を超えている[Katagiri 94]．このような，エージェントの心的状態・行為双方の環境依存性を取り入れた論理モデルの構築には知識に基づく分散システムの概念[Halpern 90]が有効であると期待される．

6. おわりに

環境に関する知識に基づき，達成すべき目標を目指して行為を遂行する合理的エージェントの設計の基礎となる論理モデル研究の現状について概観した．この研究分野は我々の常識的・哲学的直観と論理的厳密性とがせめぎあう分野である．様相論理は主に合理的エージェントの外部仕様を与える役割を果たしたが，今後は分散システムなどの手法を利用してエージェント設計を目指す方向の研究の進展が望まれる．

◇参考文献◇

[Austin 62] Austin, J. L.: *How to Do Things with Words,* Oxford University Press (1962).

[Bratman 87] Bratman, M. E.: *Intention, Plans, and Practical Reason,* Harvard University Press, Cambridge, Mass. (1987).

[Chisholm 66] Chisholm, R. M.: Freedom and action, Lehrer, K. (ed.), *Freedom and Determinism,* Random House (1966).

[Cohen 90a] Cohen, P. R. and Levesque, H. J.: Persistence, intention and commitment, Cohen, P. R., Morgan, J. and Pollack, M. E. (eds.), *Intentions in communication,* chapter 3, pp. 33-70, MIT Press (1990).

[Cohen 90b] Cohen, P. R. and Levesque, H. J.: Rational interaction as the basis for communication, Cohen, P. R., Morgan, J. and Pollack, M. E. (eds.), *Intentions in Communication,* MIT Press (1990).

[Cohen 91] Cohen, P. R. and Levesque, H. J.: Confirmations and joint action, *Proc. 12th Int. Joint Conf. on Artificial Intelligence,* pp. 951-957 (1991).

[Halpern 90] Halpern, J. Y. and Moses, Y.: Knowledge and common-knowledge in a distributed environment, *J. ACM,* Vol. 37, No. 3, pp. 549-587 (1990).

[Hintikka 62] Hintikka, J.: *Knowledge and Belief,* Cornell University Press (1962).

[Israel 93] Israel, D., Perry, J. and Tutiya, S.: Executions, motivations and accomplishments, *Philosophical Review,* Vol. 102, No. 4, pp. 515-540 (1993).

[Katagiri 94] Katagiri, Y.: A distributed system model

for actions of situated agents, *Conf. Information-oriented Approaches to Logic, Language and Computation* (1994).
[Levesque 90] Levesque, H. J., Cohen, P. R. and Nunes, J.: On acting together, *AAAI-90,* pp. 94-99, Morgan Kaufmann (1990).
[Moore 85] Moore, R. C.: A formal theory of knowledge and action, Hobbs, J. R. and Moore, R. C. (eds.), *Formal theories of the commonsense world,* chapter 9, pp. 319-358, Ablex Publishing Corporation (1985).
[Searle 69] Searle, J. R.: *Speech Acts,* Cambridge University Press (1969).
[Searle 83] Searle, J. R.: *Intentionality: An Essay in the Philosophy of Mind,* Cambridge University Press (1983).
[Searle 90] Searle, J. R.: Collective intentions and actions, Cohen, P R. Morgan, J. and Polack, M. E. (eds.), *Intentions in Communication,* chapter 19, pp. 401-415, MIT Press (1990).

著 者 紹 介

片桐 恭弘(正会員)

1981年東京大学大学院工学系研究科情報工学専門課程修了. 工学博士. 自然言語処理, 機械翻訳, 対話理解を研究. 現在, ATR知能映像通信研究所第四研究室長. 情報処理学会, 日本認知科学会, AAAI, IEEE各会員.