

対話参加者の心的記述とその動的変化に関する枠組

村田 剛志 米澤 明憲
東京大学理学部情報科学科

協調的な対話システムを構築していく上で、対話の相手についての信念記述の枠組を持つことが必要である。その信念記述の中には自己についての信念が含まれているので入れ子の構造になる。本論文では対話参加者の信念やルールなどの心的状態がどのように表現され、そしてその記述が対話中にどのように変化していくかを記述する枠組を与える。様相演算子を用いて信念などの構成要素を記述し、それらに対する操作を定義する。これによって発話による命題や意図の伝達などの現象を、信念記述の動的変化という形で明確に表現することができる。

A Framework for Describing Mental State and its Dynamic Changes within Dialog Participants

Tsuyoshi Murata Akinori Yonezawa

Department of Information Science, Faculty of Science,
The University of Tokyo

7-3-1 Hongou, Bunkyo-Ku, Tokyo, 113, Japan

In order to make a dialog system cooperative, it needs to have a framework of describing beliefs of the partner who converses with it. Because such descriptions contain beliefs of the system itself, the entire belief description structure is nested. This paper proposes a framework in which we can describe both the participant's mental state and its dynamic changes during the dialog. Modal operators are employed to express the mental states and manipulations on the modal operators are defined. With this framework, we can clearly express various phenomena of conveying propositions and intentions in terms of dynamic changes of belief descriptions.

1 はじめに

自然言語を用いた協調的な対話システムを構築するためには、単に入力された文を字義通り解釈するのでは不十分である。ユーザーの意図を理解し、陥りやすい誤りを認識してそれをカバーしていくような機構が必要である。Allen と Perrault の研究 [2] で論じられているように、例えば駅員が客に列車の時間について尋ねられたら、単に聞かれたことだけに答えるのではなく、客の意図(この場合はその列車に乗りたいか、その列車に乗ってくる人を出迎えたいなど)を理解し、その障害となるものを取り除くような答をする(例えばその電車が発車するホームを教えるなど)のがより協調的な対話であると言える。

対話によってシステムがユーザーから得られる情報が単調に増加していくようなときは複雑な推論はそれほど必要ではないが、対話がうまく進行していかないとき、ユーザーが発話するに至った推論過程をシステムが推論して、それに対して適切に対処する(例えば、システムがユーザーの目標だと思っていたものを修正するか、ユーザーが誤った信念を持っていることを教えてあげるとか)ことが望ましい。

そのためには、対話の相手についての何らかの信念記述を持つことが必要である。それによって、相手の推論機構をこちら側でシミュレートし、それに対して適切な対応をすることが可能になる。

相手の信念に対して推論を行うものとしては Cohen と Perrault の研究 [5] がある。これは INFORM や REQUEST など、いくつかの発話行為 [8] の適用可能条件とその結果起こる信念の変化を定義し、それを組み合わせることで発話のプランを組み立てるものである。しかし、発話行為の適用可能条件と、その結果起こる変化を組み合わせることで起こるべき発話を探すやり方は機械的でやや不自然であり、対話が進行するにつれて心的状態が変化する過程を記述しにくい。また、対話の流れがスムーズにいかないときの対処が述べられていない。

このような問題を解決するには、信念集合の構造について考える必要がある。ある対話参加者の全ての信念の中には、その人から見て対話相手が信じている信念の集合が含まれている。ところがその相手の信念の集合の中には相手から見て自分が信じている信念の集合が含まれているはずなので全体として入れ子の構造になる。対話中にその入れ子の構造の中で、信念がどのように変化するかについて考察することは、相手の推論過程を推論した上で対応するなど、協調的な対話を実現する上で必要である。

本研究では、対話中における対話参加者の信念の動的变化を記述することを目標とする。発話によって入れ子の構造の中の信念がどのように変化するかについて考察し、発話と、入れ子構造内の信念に対する操作とを対応づける。それによって発話からその発話をするに至った推論過程を考察することができ、さらには発話生成の

ためのベースとすることができる。

2 心的記述の枠組の概観

1節で述べたように、本研究では入れ子の構造の中の信念の変化について考察する。(図1参照)この構造の中に、信念や、今ある信念から新たな信念を導出するためのルールなど、次の節で述べるような要素が入る。対話が進むにつれて、この構造の中の要素は動的に変化する。信念が入れ子のより内側に伝わったり、外側に伝わったりする。新たな信念が従来の信念と矛盾することもある。本研究では構成要素を入れ子の内側から外側へ、外側から内側へ移動させる操作を、対話参加者の発話行為に対応づける。すなわち、発話が行なわれたことによって、心的記述がどのように変化するかについて考察する。我々の枠組は現在の段階では、すでに行われた対話に対してその際の心的記述の変化を説明するだけだが、発話をするための他の条件などを考慮することによって、発話の生成、つまり心的記述がどのような条件を満たすときに発話が起こるかについて議論するための土台になると考えられる。

Searle [8] は命令、依頼、約束など多くの種類の発話行為について検討しているが、本研究においては基本的な発話行為である主張と質問について検討することにする。

本研究では、構成要素として信念やルールに加えて、HOLE と呼ぶものを新たに導入した。信念は何らかの情報を表すのに対し、HOLE は情報が欠けていることを表す。これにより、主張と質問の二つの発話行為を統一的に扱うことが出来る。

枠組全体は対話参加者の一方からの主観的な世界の記述を表す。そもそも対話においては相手の実際の信念を参照できない。相手の発話から相手の信念を推論するのであり、得られた信念は観察した側からの主観的な記述にならざるを得ない。従って、事実とは異なる信念を持つこともあり得る。事実と異なることがわかったらその信念を取り除き、更にその信念から導かれた信念を取り除くなどの真理維持 (truth maintenance) [6] が必要になる。

信念その他、これから述べる要素は層 (layer) の中に存在する。入れ子構造の一番外側の layer は自分の信念の集合である。様相演算子 BEL を用いれば (BEL I P) が真となるような P の集合である。

その内側に相手の信念であると自分が信じることの集合がある。つまり (BEL I (BEL YOU P)) が真であるような P の集合である。(図 1)

以下、同様に自分が信じていると相手が信じていると自分が信じていることなどが定義される。これらは理論的には無限に続くのであるが、推論の際はその中で必要な有限個の layer を用いる。以下便宜的に、layer に外側から 0, 1, 2, ... と番号をつけることにする。

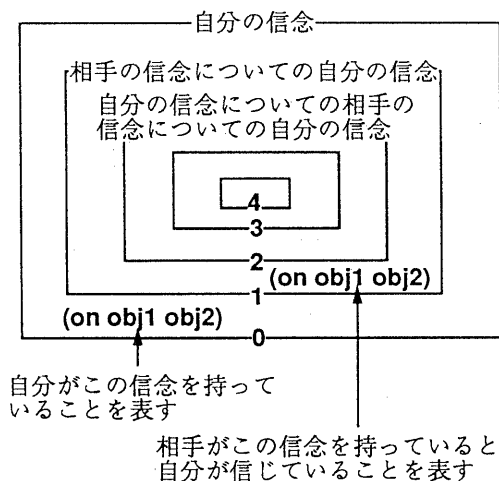


図 1: 入れ子の中の信念

3 入れ子構造の枠組における要素

前の節で信念ルールを要素とすることは述べたが、その他にも対話を表現するために必要な要素がある。この節では我々の枠組で導入される要素を定義する。

信念: 何かを信じていることを表す信念の記述は次のようになる。(predicate arg1 arg2 ...) で、述語が predicate で引数が arg1, arg2, ... であるような信念を表す。図 1にあるように、この信念が入れ子の中のどこの layer に属するかによって、誰の信念であるかが異なってくる。(例 (on object1 object2))

BELREF: 信念の具体的な内容でなく信念の存在を記述するために BELREF を導入する。(BELREF #arg agent (predicate arg1 ... #arg ...)) は、agent が信念 (predicate arg1 ... #arg ...) を、引数 #arg の値も含んだ完全な形で持っていることを表す。(例 (BELREF #x agent (at train1 #x)) は agent が信念 (at train1 #x) を引数 #x の値も含んだ完全な形で持っていることを表す。) Allen[1] の KNOWREF では (A KNOWREF the x:P(x)) に相当する。(もちろんここでは一方の対話者からの主観的な記述なので KNOWREF ではなく BELREF とする。)

BELIF: 信念の肯定または否定の存在を記述するために BELIF を導入する。(BELIF agent (predicate arg1 ...)) で、agent が信念 (predicate arg1 ...) の真偽についての信念を持っていることを表す。(例 (BELIF agent (at A office)) は、agent が (at A office) の真偽についての信念、すなわち (at A office) または (not (at A of-

fice)) を持っていることを表す。)

以下では、今まで述べた構成要素 (信念、BELREF、BELIF) を basic-element と表すことにする。

HOLE: 信念は何かを信じているという記述であるのに対して、信念が欠けているという記述をするために HOLE を導入する。HOLE は目標状態を表す WANT から作られ、信念の欠如であるその HOLE を消去するために推論が進められる。(HOLE #arg agent (predicate arg1 ... #arg ...)) で、agent が信念 (predicate arg1 ... #arg ...) の引数の一部 #arg を欠いていて、しかもその値が対話の目標を達成する上で必要であることを明示的に表す。(例 (HOLE #x agent (at agent1 #x)) は、agent が信念 (at agent1 #x) の引数の一部 #x を欠いている、つまり信念の中の agent1 の居場所という値が得られないことを表す。)

信念の中の引数の一部が欠けている場合と、信念の全体が欠けている場合に分けて、HOLE の記述を次のように定義する。

- layer n にある agent の信念の中に、信念 n: (predicate arg1 ... arg ...) を完全な形で持っておらず、引数 arg が得られないことを、n: (HOLE #arg agent (predicate arg1 ... #arg ...)) と表す。
- layer n にある agent の信念の中に、信念 n: (predicate arg1 ...) が存在しないことを n: (HOLE #ALL agent # (predicate arg1 ...)) と表す。(引数は全て bind されているものとする)

なお、すでにある信念集合から推論を経て導けるような信念の内容であっても、必要とする信念が明示的に存在しなければ HOLE は作られ、その後で HOLE を消去するための推論がなされる。4.3節で HOLE に対する操作について詳しく述べる。

ルール: A→B の形式のルールを記述する。(RULE basic-element1 basic-element2) で、basic-element1 が存在するときに basic-element2 が存在するというルールを表す。(例 (RULE (and (have A B) (have B C)) (have A C)))

意図: 意図は、対話参加者の目標状態を表す。(WANT agent basic-element) は、agent が basic-element を持つことを目標状態とすることを表す。(例 (WANT agent1 (at agent1 Tokyo)) は、agent1 が (at agent1 Tokyo) という信念を持つことを目標とすることを表す。)

これらの要素は先に述べた入れ子構造のいずれかの layer に属している。従って、その layer の番号を要素の前につけて表すことにする。例えば 0:(on A B) は、(on A B) という信念が入れ子の一番外側の layer にあること (つまり自分が "A が B の上にある" という信念を持つ

ていること)を表す。

4 要素に対する操作

この節では3節で定義された要素に対する操作について述べる。

4.1 発話に対応する操作

入れ子の心的構造の中に、3節で述べた要素が存在する。それぞれの要素は自分の属する layer 内の他の要素との関係で変化したり、新たな要素を作り出したりする。例えば、情報の欠如を表す HOLE と、その欠如している情報を表す信念の両方があるときに HOLE が消去されたり、信念とルールから新たな信念を導出したりするのがそれにあたる。また各要素は自分の属する layer だけでなく、他の layer の要素との関係でも変化する。例えば、相手が知らないことを教えるという発話行為は、我々の枠組では自分の layer にある信念を使って隣接する相手の layer 内の HOLE を消去するという操作に対応する。

本研究では発話行為を、入れ子の心的構造の構成要素に対する操作として捉える。この節では主張、質問の2つの発話行為に対応した構成要素に対する操作を定義する。発話をするとは、この枠組では layer n にある要素を layer n+2 に書き込む操作に対応する。例えば A と B が対話をしていて A の信念が layer 0 に属すると仮定する。そのとき layer 2 は“A の信念だと B が信じている”と A が信じている要素の集合を表す。A が B に発話した結果、layer 0 の要素が layer 2 に書き込まれる。

一般に異なる layer の構成要素に対する操作は次のようになる。以下の説明では“自分”とは入れ子の心的記述の枠組を持つ側の対話参加者を表し、“相手”とはそうでない側を表す。

let-me-know 操作: 質問という発話行為は、相手に対して“自分はあることを知らない”ということを知らせることである。それが起こるためには、自分がそのことに関する信念を持たず、相手がそのことに関する信念を持っていると信じている必要がある。この発話行為に対応するのは、“layer n に HOLE があり、layer n+1 に BELREF または BELIF があるとき、layer n と同じ HOLE を layer n+2 に作る”という操作である。この操作の記述は次のようになる。

• n: (HOLE #arg agent1 (predicate ... #arg ...)) と n+1: (BELREF #arg agent2 (predicate ... #arg ...)) が存在するとき n+2: (HOLE #arg agent1 (predicate ... #arg ...)) を書き込む。(図 2)

• n: (HOLE #ALL agent1 #((predicate ...))) と n+1: (BELIF agent2 (predicate ...)) が存在するとき n+2:

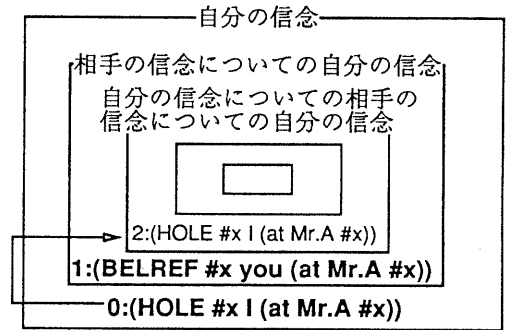


図 2: 信念の操作 1

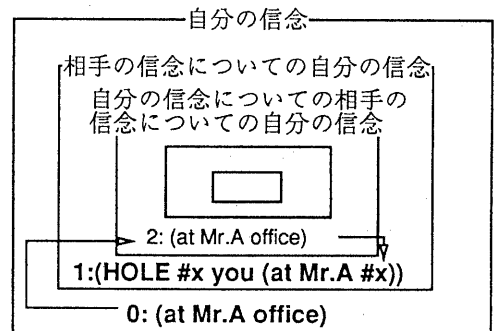


図 3: 信念の操作 2

(HOLE #ALL agent1 #((predicate ...)) を書き込む。

I-inform-you 操作: 相手が知らないことを知らせるという発話が起こるためには、自分がそのことに関する信念を持っていて、相手がそれに関する信念を持っていないと信じている必要がある。この発話に対応するのは、“layer n に信念があり、layer n+1 にその信念に関する HOLE があるとき layer n の信念を layer n+2 に作る”という操作である。この操作の記述は次のようになる。

• n: (predicate arg1 ...) と n+1: (HOLE #arg agent (predicate ... #arg ...)) が存在するとき n+2: (predicate arg1 ...) を書き込む。(図 3)

I-correct-you 操作: 相手の持つ、自分についての誤った信念を解消するための発話が起こるためには、自分がある信念を持っていて、それとは矛盾する信念を自分が持っている相手と誤って信じていると自分が信じている必要がある。例えば相手の誤った信念を消去するために、自分が A と信じていると発話するときには、自分

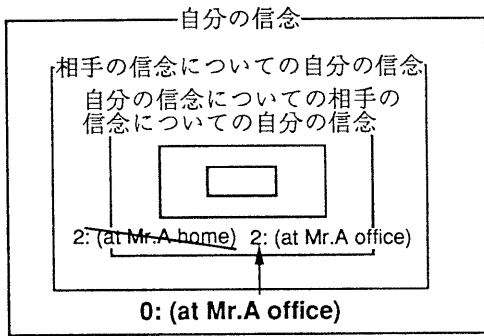


図 4: 信念の操作 3

は A と信じているのに、こちらが (not A) と信じていると相手が信じていると自分が信じていることになる。この発話を我々の枠組で説明すると 2 つ離れた layer (いわば自分と、自分の中の相手の中の自分) は自分という同じ対象に関する記述であるから、必ずしも一致している必要はないが、矛盾しているものであってはならないから発話が起ったということになる。

この発話に対応するのは、“layer n に信念があり、layer n+2 にそれと矛盾する信念があるとき、layer n の信念を layer n+2 に作る” という操作である。この操作の記述は次のようになる。

• n: (predicate ...) と n+2: (not (predicate ...)) が存在するとき n+2: (not (predicate ...)) を消して、n+2: (predicate ...) を付け加える。(図 4)

本研究では Cohen[5]の研究では INFORM として定義されていたものを I-inform-you 操作と I-correct-you 操作の 2 つに分けて定義した。この 2 つの操作は、ある layer の信念を 2 つ内側の layer に書き込むという点では共通しているが、前者はその書き込んだ結果が、次に述べる I-follow-you 操作のために発話された側の信念の layer にまで及ぶが、後者は発話した側の信念の layer に留まるため、両者の発話の効果は異なる。従って本研究では両者を分けて定義した。

I-follow-you 操作: この操作は発話とは結びつかないが、先の I-inform-you 操作と結びつく操作なので、ここで定義する。相手がある信念を持っていると信じていて、自分はその信念を欠いているとき、相手の信念を自分の信念とすることがある。例えば相手が“会議は 9 時に始まる”と信じていて、自分は会議の時刻についての信念を持っていないとき、自分も“会議は 9 時に始まる”と信じる場合などがそうである。つまり、対話中に相手が述べた信念を自分の信念とすることである。

これに対応するのは、“自分の layer 内の HOLE で

欠けている内容を相手が隣接する layer に信念として持っている信じるとき、相手と同じ信念が自分の layer に作られる” という操作である。この操作の記述は次のようになる。

• n:(HOLE #arg agent (predicate ... #arg ...)) と n+1: (predicate arg1 ...) が存在するとき n:(HOLE #arg agent (predicate ... #arg ...)) を消去して n:(predicate arg1 ...) を付け加える。

4.2 WANT に対する操作

今まで述べた信念、ルール、HOLE、BELREF、BELIF などの要素と、対話参加者の意図を表す WANT とを結び付けるための操作を考える。

意図に関する推論としては Cohen と Levesque の研究 [4] など様々な研究があるが、ここでは WANT によって目標となった信念の HOLE が作られるというもっとも単純な操作のみにとどめることにする。この WANT が出発点になって目標状態の HOLE が作られ、そしてその HOLE を消去するための推論が行われる。

make-hole 操作: layer n に n:(WANT agent (predicate ...)) という意図が存在するとき、n:(HOLE #ALL agent #((predicate ...)) という HOLE を作る。

4.3 HOLE に対する操作

本研究の枠組においては、情報の欠如を表す HOLE をどのように扱うかが重要な問題になる。なぜならまず目標状態を表す WANT から HOLE が作られ、そしてその HOLE を消去するために推論や操作が行われるからである。

HOLE に対して考えられる操作規則としては 3 つある。まず同一 layer 内の信念との操作、次に他の layer の要素との操作、そして同一 layer 内の RULE との操作である。

同一 layer 内の HOLE と信念に対する操作

HOLE の定義から明らかであるが、同一の layer 内に信念と、その信念に関する情報の欠如を表す HOLE が共存することはあり得ないので、その場合は HOLE を消去する必要がある。

delete-hole 操作: agent の layer n の中に n: (predicate arg1 ...) と n: (HOLE #x agent (predicate ... #x ...)) の両方が存在するときには HOLE を消去する。

delete-hole-all 操作: agent の layer n の中に n: (predicate arg1 ...) と、n: (HOLE #ALL agent #((predicate ...)) が存在するときには HOLE の方を消去する。

異なる layer 間の HOLE と信念に対する操作

これは 4.1 節の“発話に対応する操作”の let-me-know 操作と I-inform-you 操作と I-follow-you 操作であり、すでに述べた。

同一 layer 内の HOLE と RULE に対する操作

例えば、“会議がある日なら社長は出社している”というルールがあり、“社長が出社している”という信念が欠けているなら、会議がある日だという信念も欠けていると見なすことが出来る。すなわち、 $A \rightarrow B$ というルールがあり、かつ信念 B の情報の欠如があるとき、これから A の情報の欠如が導くのは妥当である。このような推論の目的は、信念 B を得るために信念 A の HOLE を作って、その HOLE を消去させるような操作をする(つまり信念 A を得るための操作をする)ことによって、abductive な推論を行うことである。この操作の記述は次のようになる。

hole-propagation 操作: layer n に $n:(\text{RULE} (\text{and} (\text{pred}_1 \dots) \dots (\text{pred}_m \dots))) (\text{predicate } \text{arg}_1 \dots \text{arg}_k \dots))$ と、 $n:(\text{HOLE } \# \text{arg agent } (\text{predicate } \text{arg}_1 \dots \# \text{arg} \dots))$ ($\# \text{arg}$ は k 番目の引数とする)が存在するならば、 $i=1 \dots m$ について、 $(\text{pred}_i \dots)$ の引数に arg_k が含まれていればそれを $\# \text{arg}$ に置換して $n:(\text{HOLE } \# \text{arg agent } (\text{pred}_i \dots \# \text{arg} \dots))$ をつくる。そうでなければ $n:(\text{HOLE } \# \text{ALL agent } \# (\text{pred}_i \dots))$ を付け加える。

4.4 その他の操作規則

今まで述べてきたような、発話と直接結び付く操作の他に、対話を我々の枠組で説明するために必要な操作をこの節で定義する。

まず、BELREF に関する HOLE の中の引数を変更する操作を考える。これは、自分の BELREF に関する HOLE の代わりに、相手の BELREF に関する HOLE を作る操作である。その相手の BELREF に関する HOLE が消去できたら、そのときは相手の layer には BELREF があり、自分の layer には BELREF に関する HOLE があるので、改めて相手に質問をする、つまり let-me-know 操作をすることによって自分が求める信念を得ることが出来る。

change-agent 操作: $n:(\text{HOLE } \# \text{ALL agent1 } \# (\text{BELREF } \# \text{arg agent1 } (\text{predicate } \dots \# \text{arg} \dots)))$ という HOLE が存在するときに、 $n:(\text{HOLE } \# \text{ALL agent1 } \# (\text{BELREF } \# \text{arg agent2 } (\text{predicate } \dots \# \text{arg} \dots)))$ という HOLE を書き込む。

また対話の際、両者の間に行き違いが生じることがある。その原因としては、

- 両者が異なった basic-element に基づいて推論を行なっている。
- 両者が異なったルールに基づいて推論を行なっている。

などが考えられる。このような対話の行き違いとその修復をこの枠組において表現したい。そのために次のような操作を考える。

対話相手が異なったルールを持つ場合については、Pollack[7]の論文の中で述べられているように、この心的記述を持つ側の信じているルールの条件部の一部を欠いたルールを対話相手が信じている可能性がある。したがって、対話相手が異なったルールを持っている場合は、条件部の一部を欠落させたルールを相手を持つと仮定して推論をすることにする。この操作の記述は次のようになる。

wrong-rule 操作: 対話相手が異なったルールを持つ場合、心的記述を持つ側が信じている layer n のルール $n:(\text{RULE} (\text{and} (\text{pred}_1 \dots) \dots (\text{pred}_k \dots))) (\text{predicate } \dots))$ から、 $n < m$ である layer m に $m:(\text{RULE} (\text{and} (\text{pred}_1 \dots) \dots (\text{pred}_{i-1} \dots) (\text{pred}_{i+1} \dots) \dots) (\text{predicate } \dots))$ ($1 \leq i \leq k$) を新たに作って推論を進める。

相手が自分とは異なるルールを持つということをもどのように検出するかは大きな問題である。この枠組では、対話相手が異なるルールを持つことを明示的に述べた場合か、それまで述べた操作規則だけからでは HOLE を消去出来ない場合にこの操作をするものとする。

5 具体的な例

今まで述べた信念記述の枠組の有効性を示すために、具体例を示す。対話相手についての心的記述や、対話相手の推論過程についての推論が必要な例として、次のような間接発話行為を含む対話を我々の枠組を用いて説明してみよう。

A: 時計を持っていますか?
 B: 時間は分かりません。この時計止まっているんです。

A が時間を尋ねようとして上のような発話をして、B がそれに対して分からないと答える対話である。

この対話を A の側から考えるとする。最初に次のような信念及びルールがあると仮定する。ここで (timeis t) は時刻が t であるという信念を表す。

0 : (WANT A (BELREF #t A
 (timeis #t))) (1)

0 : (RULE (have X watch)
 (BELREF #t X (timeis #t))) (2)

1 : (BELIF B (have B watch)) (3)

まずここで (1) に make-hole 操作を適用して、

$$0 : (\text{HOLE} \# \text{ALL} A \# (\text{BELREF} \# t A (\text{timeis} \# t))) \quad (4)$$

このあと (2) と (4) から、hole-propagation 操作を適用して、

$$0 : (\text{HOLE} \# \text{ALL} A \# (\text{have} A \text{ watch})) \quad (5)$$

しかしこの HOLE は消去できない。そこで change-agent 操作を用いて、(4) の (BELREF ... A ...) を (BELREF ... B ...) にする。

$$0 : (\text{HOLE} \# \text{ALL} A \# (\text{BELREF} \# t B (\text{timeis} \# t))) \quad (6)$$

これは、この HOLE が消去されれば B に BELREF があることが分かり、B に対して改めて時間を尋ねることと時間を知ることが出来るので有効である。

そして (2) とこの HOLE から、hole-propagation 操作を使う。

$$0 : (\text{HOLE} \# \text{ALL} A \# (\text{have} B \text{ watch})) \quad (7)$$

B についての BELIF(3) があるので、let-me-know 操作でこの HOLE を layer 2 に書き込む。これが A の “時計を持っていますか?” という発話に対応する。

$$2 : (\text{HOLE} \# \text{ALL} A \# (\text{have} B \text{ watch})) \quad (8)$$

そして I-follow-you 操作で述べた規則により layer 2 にあるこの HOLE が layer 1 に書き込まれる。

$$1 : (\text{HOLE} \# \text{ALL} A \# (\text{have} B \text{ watch})) \quad (9)$$

これに対する B の推論過程は B の発話を聞くまで分からない。B の発話を聞いた段階で、A は B のこの発話に至る推論過程について推論することが出来る。それは、B の “時間はわかりません。この時計止まっているんです。” という発話は B の layer 1 の信念

$$1 : (\text{and} (\text{have} B \text{ watch})(\text{not} (\text{working} \text{ watch}))) \quad (10)$$

$$1 : (\text{not} (\text{BELREF} \# t B (\text{timeis} \# t))) \quad (11)$$

を layer 3 に書き込む操作に対応するから、この操作が起こるためには次のような推論過程があったことが分かる。時計が動いていないことを B が伝えているのはそれが時間がわからないことの理由になっているからであることが A の側から語用論的に推論できる。このことから時計が止まっていないことが時間を知るための条件になっていることがわかる。(10) と (11) の因果関係から、B は次のようなルールを持っていることが A の側から推論できる。

$$1 : (\text{RULE} (\text{and} (\text{have} X \text{ watch})(\text{working} \text{ watch})) (\text{BELREF} \# t X (\text{timeis} \# t))) \quad (12)$$

そして、B は A が let-me-know 操作をするための前提条件から、layer 3 に BELIF があるためと判断する。

$$3 : (\text{BELIF} B (\text{have} B \text{ watch})) \quad (13)$$

ここで B の (9) への対処は 2 通りある。B は A が (9) の HOLE を持っていて単純に質問したと考えるか、A が他の HOLE から hole-propagation 操作をして (9) を導いて間接発話行為を行なったと考えるかである。前者であれば、B は A の HOLE を消去するために I-inform-you 操作をする。すなわち “はい、持っています。” と答える。しかしここではそうではないので B は A が hole-propagation 操作を行なったと考えて、その操作に必要なルールを探す。B の持っているルールの中には適切なものが見つからないので、B は A が異なったルールを持っていると見なす。そして B は wrong-rule 操作、つまり B の持つルールの条件部の一部を欠いたルールを作る操作を行なって、B は、時計を持ってさえいれば時間が分かるという誤ったルールを A が信じているという仮説を立てることが出来る。

$$2 : (\text{RULE} (\text{have} X \text{ watch})(\text{BELREF} \# t X (\text{timeis} \# t))) \quad (14)$$

$$3 : (\text{RULE} (\text{have} X \text{ watch})(\text{BELREF} \# t X (\text{timeis} \# t))) \quad (15)$$

そしてこのルール (15) と BELIF(13) から、

$$3 : (\text{BELIF} B (\text{BELREF} \# t B (\text{timeis} \# t))) \quad (16)$$

を導出することができる。

従って、B の発話は、layer 1 と 3 の違いを解消するための発話であることが A の側から説明できる。つまり我々の枠組では I-correct-you 操作に対応し、(11) と (16) の違いに対し、(11) を layer 3 に書き込み、(15) を打ち消すためにこのルールでなく別のルール (12) を信じていることを layer 3 に書き込む。以上が B の発話 “時間はわかりません。この時計止まっているんです。” の発話によって起こる信念記述の変化である。

上の例では A が B の推論過程を推論したが、それが A の次の発話に生かされていない。次に A が B の推論過程を推論した上で発話した例について考える。

A: 時計を持っていますか?

B: 時間はわかりません。この時計止まっているんです。

A: 時計を持っているかどうかを聞いたんです。

この例では A が B の推論過程を推論して、B が A の発話を間接発話行為と見なしたことに對し、A が自分の発話が単純な質問であったことを B に伝えている。

この対話を A の側から考えるとす。最初に各 layer には次のような信念及びルールがあると仮定する。

0 : (WANT A (have B watch)) (17)

0 : (RULE (and (have X watch) (working watch)) (BELREF #t X (timeis #t))) (18)

1 : (BELIF B (have B watch)) (19)

まず make-hole 操作から A に次の HOLE が生じる。

0 : (HOLE #ALL A # (have B watch)) (20)

これと BELIF(19) から、A は let-me-know 操作を行う。これが A の “時計を持っていますか?” という発話に対応する。

2 : (HOLE #ALL A # (have B watch)) (21)

このあと A は HOLE を消去するような B の発話 (例えば “はい、持っています” など) を期待する。しかし、予想と異なる発話 (“時間はわかりません。この時計止まっています”) を得たので、B の推論過程について推論することになる。この部分の B の推論過程に関する推論は先に述べた例と同様で、B は

3 : (BELIF B (BELREF #t B (timeis #t))) (22)

を導き、layer1 と layer 3 の違いを解消するための発話をして、それが I-correct-you 操作に対応することが A の側から説明できる。

これに対して A は、自分は間接発話行為ではなく、単純に時計の有無を尋ねるために let-me-know 操作をしたことを知らせる。すなわち

0 : (HOLE #ALL A # (have B watch)) (23)

を layer 2 に書き込む操作をする。これが最後の A の発話 (“時計を持っているかどうかを聞いたんです”) に対応する。

この例は単純な場合のものであるが、この枠組を用いることによって他の対話参加者の心的記述が必要な場合の対話を説明できる。特に対話中に相手が予想と異なる発話をしたときに、相手の推論過程について推論することによって、相手がシステムと異なる信念や異なるルールを持っていることを検出する作業の助けになる。また、発話のための他の条件を考慮することで、すでに行なわれた対話の説明だけでなく、対話システムにおける発話生成機構のベースとして利用できる。

6 まとめ

入れ子の構造における信念を扱ったものとしては、Ballim と Wilks の研究 [3] があるが、そこでは対話相手や第三者の持つデフォルトの信念集合をそれ以外の信念

集合から生成することが目標であり、本研究の主題とは異なっている。

Cohen[5] らの論文では、INFORM や REQUEST などのオペレーターを組合せることで発話行為のプランニングを行なっている。しかし、発話することは入れ子の心的記述の 2 つ内側の layer に要素を書き込むことだという、より本質的な規則が Cohen のオペレーターの適用条件からは見えてこない。また、対話中の信念の動的変化を記述しにくく、対話の流れがスムーズにいかない場合に対処するためには多くのオペレータを定義する必要がある。

本研究では対話システムにおける対話参加者の心的状態の動的な変化を記述する枠組を与えた。入れ子の心的構造の中で信念やルールなどの要素を導入し、それらに対する操作を発話行為と関係づけて定義した。本研究においては情報の欠如を表す要素 HOLE を新たに導入することによって、質問と主張を同一の形式で表現することができた。対話中の発話を、layer の構成要素に対する操作として捉え直すことで、対話中の信念の動的な変化の記述や、相手の推論過程を推論する際の記述が容易になる。

謝辞

この研究を進めるにあたって様々なアドバイスを頂いた東京大学理学部情報科学科米澤研究室の皆さん、特に自然言語グループの皆さんに感謝いたします。

参考文献

- [1] Allen, J.F., Recognizing Intentions From Natural Language Utterances. In M. Brady and R.C. Berwick, eds., *Computational models of discourse*. Cambridge, MA: MIT Press, 1983.
- [2] Allen, J.F. and Perrault, C.R., *Analyzing Intention in Utterances*. Artificial Intelligence, 15:143-178, 1980.
- [3] Ballim, A. and Wilks, Y. *Artificial Believers*, Lawrence Erlbaum Associates, 1991.
- [4] Cohen, P.R. and Levesque, H.J., Persistence, Intention, and Commitment, In P.R. Cohen, J. Morgan, and M. E. Pollack (Eds.) *Intentions in COMMUNICATION*, The MIT Press, 1990.
- [5] Cohen, P.R. and Perrault, C.R., *Elements of a Plan-Based Theory of Speech Acts*, Cognitive Science 3(3), 177-212, 1979.
- [6] Doyle, J., *A truth maintenance system*, Artificial Intelligence, 12:231-272, 1979.
- [7] Pollack, M.E., Plans as Complex Mental Attitudes, In P. R. Cohen, J. Morgan, and M. E. Pollack (Eds.) *Intentions in COMMUNICATION*, The MIT Press, 1990.
- [8] Searle, J.R. *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press, 1969.