知識工学と人工知能 64－2
（1989. 5. 18）

自己認識論理のエージェント系への拡張とその階層的知識の表現への応用

外山勝彦　　　稲垣康善
名古屋大学工学部

理想的に合理的な２つのエージェントの間で十分な通信がある場合に対して，Moore の自己認識論理を自然に拡張し，各エージェントが，与えられた前提知識から拡張した自己認識推論によって得られる知識に関する論理，すなわち，２エージェント系の自己認識論理を形式化する．そのような知識は，安定性と依存性という２つの概念で特徴付けられることを示し，さらに，そのような知識と前提知識の関係を示す．また，この２エージェント系の自己認識論理を用いて，階層的な知識の表現を行い，そのような知識の下での推論を形式的に特徴付ける．これは，フレームモデルによる知識表現に対する論理的な意味付けを与えると考えられる．

Autoepistemic Logic for Two Agents and its Application to Representation of Hierarchical Knowledge

Katsuhiko TOYAMA　　　Yasuyoshi INAGAKI

School of Engineering, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-01 Japan

We propose autoepistemic logic for two agents, which is a natural extension of Moore's autoepistemic logic. This logic formalizes beliefs which two ideally rational agents with mutual communications can infer by the extended introspection from their initial beliefs. We show that these beliefs are characterized by two concepts, stability and groundness. We also investigate the relation between these beliefs and their initial beliefs. Furthermore, we apply this logic to represent hierarchical knowledge and to characterize the reasoning based on such knowledge. This representaion can be thought as a logical formalization of Minsky's frame model for knowledge representaion.

1. はじめに

高度な知識情報処理の実現のために，階層や構造のある知識やメタ知識を表現し，常識を用いた推論や，不完全な知識の下での推論を行うことができる知識処理システムが考えられている．そのような知識や推論の形式化として，Reiter のデフォルト論理（default logic）(2)，McDermott の nonmonotonic logic(3,4)，Moore の自己認識論理（autoepistemic logic）(1)，McCarthy のサカマスクリプション（極小限定）(circumscription)(7,8)などがあり，これらは非単調論理（nonmonotonic logic）と呼ばれている．このうち特に，自己認識論理は，知識システムにどのような知識が存在し，あるいは，欠如しているかについてのメタ知識を表現することができ，理想的に合理的な知的行為者であるエージェント（agent），たとえば，人間や計算機などが，あらかじめ与えられた前提知識から，自己認識推論（introspection）によって得られる知識の集合を特徴付けたものである．自己認識推論とは，エージェントが，ある知識を持っているとき，その知識を持っているというメタ知識を持つことであり，また，ある知識を持っていないとき，その知識を持っていないというメタ知識を持つことである．

ところで，いくつかの知識システムから構成される総合的なシステムにおける知識を解析する場合，それを構成する個々のシステムが持つ知識には，与えられた前提知識から推論によって得られる知識の他に，個々のシステムの間の通信によって得られる知識もあり，これら両方とも解析する必要がある．そのような知識について議論するためには，自己認識論理を複数のエージェントが存在する場合に拡張して形式化することが考えられる．

本論文では，理想的に合理的な２つのエージェントの間で十分な通信がある場合，あらかじめ与えられた前提知識から，各エージェントが自己認識推論によって得られる知識に関する論理を形式化する．

次に，この２エージェント系の自己認識論理を用いて，上位階層と下位階層の２層だけの場合の階層的な知識の表現と，その下での推論を形式的に議論する．従来，階層のある知識を扱う場合，フレームモデル(9)や非単調論理(1,2,3,4,7,8)などが用いられてきた．しかし，これらの知識表現モデルでは，表現された知識の意味が不明確であったり，その下での推論が不十分であったりする場合がある．本論文では，各階層を１つの知識システムと捉え，属性継承（inheritance）は，各システムが持つ知識の間の関係に関する知識と，上位階層であるシステムから下位階層であるシステムへの知識のメタ通信によって生じると捉える．

以下，本論文では，次の２章で Moore の自己認識論理(1)を概観する．３章では，２エージェント系の自己認識論理の構文と意味論について定義する．４章では，拡張した自己認識論理における理論の健全性と完全性の特徴付けを行い，エージェントが持つ知識と前提知識の関係を解析する．５章では，階層的な知識の表現とその下での推論について考察する．

2. 自己認識論理

自己認識論理(1)は，通常の命題論理式の他に，様相演算子Ｌを追加して得られる論理式を持ち，Ｌｐという論理式は，ｐという知識を持っているというメタ知識を表す．エージェントが持つ知識は論理式で表され，その集合を自己認識理論（autoepistemic theory），または，単に，理論という．

解釈（interpretation）とは，命題論理において通常用いられるような論理式への真理値の割当（assignment）であり，論理式集合Ｓのモデル（model）とは，Ｓの論理式がすべて真となる解釈である．論理式Ｌｐの意図された意味を反映する解釈は，理論Ｔに対し，Ｌｐが真であるのは，$p \in T$のとき，かつ，そのときに限るとする解釈であり，これをＴによる自己認識解釈（autoepistemic interpretation），または，単に，Ｔによる解釈という．すなわち，Ｔによる解釈において，Ｔは論理式Ｌｐを解釈するための指標である．論理式集合ＳのＴによる自己認識モデル（autoepistemic model），または，単に，Ｔによるモデルとは，Ｓの論理式がすべて真となるＴによる解釈である．

エージェントがあらかじめ与えられた前提知識から自己認識推論によって得られる知識の集合を特徴付けるとき，エージェントは理想的に合理的であるということを仮定する．すなわち，エージェントは合理的だから，与えられた前提がすべて真ならば，自己認識推論によって得られる知識はすべて真でなければならない．また，エージェントは理想的に合理的だから，自己認識推論によって得られる知識は，知識全体の中に含まれていなければならない．これらの概念をそれぞれ健全性，完全性といい，形式的には次のように定義する．

［定義２．１］（健全性）　理論Ｔが論理式集合Ａ（以下，前提（premise）という）に関して健全（sound）であるとは，ＡのＴによるすべてのモデルは，ＴのＴによるモデルであることをいう．　□

［定義２．２］（完全性）　理論Ｔが完全（complete）であるとは，ＴのＴによる任意のモデルにおいて真である論理式は，すべてＴに含まれることをいう．　□

［定義２．３］　理論Ｔが前提Ａの拡張（extension）であるとは，Ｔが次の条件(1)～(3)を満たしているときをいう：

(1)　$A \subseteq T$.

(2)　Ｔは完全である．

(3)　ＴはＡに関して健全である．

□

健全性と完全性を満たす理論を特徴付けるために，安定性と依存性という概念を定義する．

［定義２．４］（安定性）　理論Ｔは，次の条件(1)～(3)を満たすとき，安定（stable）であるという：

(1)　$p_1, \cdots, p_n \in T$, $p_1, \cdots, p_n \models p$ならば，$p \in T$.（$\models$は，命題論理における論理的帰結（logical consequence）を表す．）

(2)　$p \in T$ならば，$Lp \in T$.

(3)　$p \notin T$ならば，$\neg Lp \in T$.

□

［定理２．１］　理論Ｔが完全であるのは，Ｔが安定であるとき，かつ，そのときに限る．　□

ここで，理論Ｔと前提Ａに対して，次のような記法を定める：

$LT = \{Lp \mid p \in T\}$,

$\neg L\bar{T} = \{\neg Lp \mid p \notin T\}$,

$A' = A \cup LT \cup \neg L\bar{T}$.

［定義２．５］（依存性）　理論Ｔが前提Ａに依存（grounded）しているとは，$T \subseteq \{q \mid A' \models q\}$であることをいう．　□

［定理２．２］　理論Ｔが前提Ａに関して健全であるのは，ＴがＡに依存しているとき，かつ，そのときに限る．　□

［定理２．３］　前提Ａと理論Ｔに対して，ＴがＡの拡張であるのは，$T = \{q \mid A' \models q\}$であるとき，かつ，そのときに限る．　□

これらの定理により，理想的に合理的なエージェントが，あらかじめ与えられた前提知識から自己認識推論によって得られる知識の集合，すなわち，この前提知識の拡張は，安定性と依存性という２つの概念で捉えられ，前提知識とその拡張の関係が示された．

３．２エージェント系の自己認識論理

２エージェント系の自己認識論理は，各エージェントが，それぞれにあらかじめ与えられた前提知識から，拡張した自己認識推論によって得られる知識の集合を特徴付けるものである．この拡張した自己認識推論は，エージェント１（エージェント２）がある知識を持っているとき，その知識をエージェント１（エージェント２）が持っているというメタ知識をエージェント１，および，エージェント２が持つことであり，また，エージェント１（エージェント２）がある知識を持っていないとき，その知識をエージェント１（エージェント２）が持っていないというメタ知識をエージェント１，および，エージェント２が持つことである．

本章では，３．１節で，２エージェント系の自己認識論理の構文を定義し，３．２節で，その意味論を定義する．

３．１　構文

［定義３．１］　用いる記号は次の(1)～(3)である：

(1)　原子命題（atomic proposition）
　　：$P, Q, R, \cdots$.

(2)　論理結合子（logical connective）
　　：$\neg, \vee$.

(3)　様相演算子（modal operator）
　　：L_1, L_2.

□

［定義３．２］　論理式（formula）を次の(1)～(4)で定義する：

(1)　原子命題Ｐは論理式である．

(2)　ｐが論理式のとき，$\neg p$も論理式である．

(3)　p, qが論理式のとき，$p \vee q$も論理式である．

(4)　ｐが論理式のとき，$L_1 p, L_2 p$も論理式である．

□

$\wedge, \Rightarrow$などの論理結合子は，通常の意味の通りに用いる．なお，p, qなどの文字を論理式を表すために用いる．

$L_i p$（$i = 1, 2$）という形の論理式は，エージェントｉ（$i = 1, 2$）がｐという知識を持っているというメタ知識を表す．

原子命題と論理結合子$\neg, \vee$だけを用いてできる論理式（すなわち，L_1, L_2が出現しない論理式）を特に命題論理式（propositional formula）という．

［定義３．３］　自己認識理論（autoepistemic theory）（または，単に，理論）とは論理式の集合である．□

理論は，エージェントが持つ知識の集合である．ここでは，２エージェント系を扱うので，理論の組＜T_1，T_2＞を考え，これを単に，理論という．

３．２　意味論

理論の中の論理式は，エージェントの知識を表す．その真理値は，

(1)　対象世界における各命題の真偽，

(2)　エージェントが拡張した自己認識推論によって得られる知識の集合

の２つによって決定される．

［定義３．４］　解釈（interpretation）とは，次の条件(1)～(3)を満たす論理式への真理値の割当（assignment）Ｉである：任意の論理式ｐ，ｑに対して，

(1)　$I(p)=1$，または，$I(p)=0$のいずれか一方である．

(2)　$I(\neg p)=1$　iff　$I(p)=0$．

(3)　$I(p \vee q)=1$

iff　$I(p)=1$，または，$I(q)=1$．

□

［定義３．５］　論理式集合Ｓのモデル（model）とは，Ｓの論理式がすべて真となる解釈である．

□

解釈，モデルにおいては，$L_i p$の形の論理式は新しい原子命題として扱う．

$L_i p$の形の論理式の意図された意味は，エージェントｉが，ｐという知識を，その知識の集合である理論T_iの中に持っているということである．そこで，この意味を反映する解釈を定義する．

［定義３．６］　理論＜T_1，T_2＞による自己認識解釈（autoepistemic interpretaion）（または，単に，理論＜T_1，T_2＞による解釈）とは，次の条件を満たすような解釈である：

$I(L_i p)=1$　iff　$p \in T_i$．

（$i=1,2$）　□

［定義３．７］　論理式集合Ｓの理論＜T_1，T_2＞による自己認識モデル（autoepistemic model）（または，単に，理論＜T_1，T_2＞によるモデル）とは，Ｓの論理式がすべて真となる＜T_1，T_2＞による解釈である．□

理論＜T_1，T_2＞による自己認識解釈において，T_1，T_2は$L_i p$の形の論理式を解釈するために用いる指標である．理論は，原子命題への真理値の割当とは独立に，$L_i p$の形の論理式の真理値を決定する．

４．自己認識理論の健全性と完全性

本章では，２エージェント系の自己認識論理において，各エージェントが，それぞれにあらかじめ与えられた前提知識から，拡張した自己認識推論によって得られる知識の集合の特徴付けを行う．

２エージェント系の自己認識論理においても，各エージェントは理想的に合理的であるということを仮定する．すなわち，健全性と完全性を持つ理論の特徴付けを行う．２エージェント系における健全性は，各エージェントは合理的だから，１つのエージェントに与えられた前提がすべて真ならば，拡張した自己認識推論によって，そのエージェントが得られる知識は，すべて真でなければならないということである．また，２エージェント系における完全性とは，各エージェントは理想的に合理的だから，各エージェントが持つ知識に基づいて真となる知識は，各エージェントの知識全体の中に含まれていなければならないということである．これらの概念を次のように形式的に定義する．

［定義４．１］（健全性）　理論＜T_1，T_2＞が論理式集合の組＜A_1，A_2＞（以下，前提（premise）という）に関して健全（sound）であるとは，A_1の＜T_1，T_2＞によるすべてのモデルは，T_1の＜T_1，T_2＞によるモデルであり，かつ，A_2の＜T_1，T_2＞によるすべてのモデルは，T_2の＜T_1，T_2＞によるモデルであることをいう．□

［定義４．２］（完全性）　理論＜T_1，T_2＞が完全（complete）であるとは，T_1の＜T_1，T_2＞による任意のモデルにおいて真である論理式は，すべてT_1に含まれ，かつ，T_2の＜T_1，T_2＞による任意のモデルにおいて真である論理式は，すべてT_2に含まれることをいう．□

また，各エージェントが，前提＜A_1，A_2＞から拡張した自己認識推論によって得られる知識の集合T_1，T_2に対し，A_1，A_2は前提だから，それぞれ，T_1，T_2に含まれていなければならない．

そこで，特徴付けるべき知識の集合を，次のように定義する．

［定義４．３］　理論＜T_1，T_2＞が前提＜A_1，A_2＞の拡張（extension）であるとは，＜T_1，T_2＞が次の条件(1)～(3)を満たすときをいう：

(1)　$A_1 \subseteq T_1$，$A_2 \subseteq T_2$．

(2) $\langle T_1, T_2 \rangle$は完全である.

(3) $\langle T_1, T_2 \rangle$は$\langle A_1, A_2 \rangle$に関して健全である.

□

以下では，完全な理論，および，与えられた前提に関して健全な理論の特徴付けを考える.

［定義４．４］（安定性） 理論$\langle T_1, T_2 \rangle$は，次の条件(1)～(3)を満たすとき，安定 (stable) であるという.

(1) $p_1, \cdots, p_n \in T_i$, $p_1, \cdots, p_n \models p$ ならば，$p \in T_i$.

(2) $p \in T_i$ならば，$L_i p \in T_1$，かつ，$L_i p \in T_2$.

(3) $p \notin T_i$ならば，$\neg L_i p \in T_1$，かつ，$\neg L_i p \in T_2$.

$(i = 1, 2)$

□

理論が安定であるとは，直観的には次のような意味である．すなわち，いずれのエージェントも，自分が持つ知識から，それ以上の知識を拡張した自己認識推論によって得ることができないということである.

［命題４．１］ 理論$\langle T_1, T_2 \rangle$は安定であるとする．このとき，次の(1)～(4)が成り立つ.

(1) T_1が無矛盾であるとき，$L_i p \in T_1$ ならば，$p \in T_i$.

(2) T_2が無矛盾であるとき，$L_i p \in T_2$ ならば，$p \in T_i$.

(3) T_1が無矛盾であるとき，$\neg L_i p \in T_1$ ならば，$p \notin T_i$.

(4) T_2が無矛盾であるとき，$\neg L_i p \in T_2$ ならば，$p \notin T_i$.

$(i = 1, 2)$

□

［証明］

(1) $L_i p \in T_1$，かつ，$p \notin T_i$とする．定義４．４の(3)により，$\neg L_i p \in T_1$となり，T_1は矛盾し，仮定に反する.

(2) (1)と同様に示すことができる.

(3) $\neg L_i p \in T_1$，かつ，$p \in T_i$とする．定義４．４の(2)により，$L_i p \in T_1$となり，T_1は矛盾し，仮定に反する.

(4) (3)と同様に示すことができる.

□

命題４．１により，理論$\langle T_1, T_2 \rangle$が安定で，T_1, T_2が共に無矛盾ならば，

$$L_i p \in T_1 \ \text{iff}\ L_i p \in T_2 \ \text{iff}\ p \in T_i.$$

$$\neg L_i p \in T_1 \ \text{iff}\ \neg L_i p \in T_2 \ \text{iff}\ p \notin T_i.$$

$(i = 1, 2)$

であることが分かる．すなわち，理論$\langle T_1, T_2 \rangle$が安定で，T_1, T_2が共に無矛盾ならば，T_1のモデルは，すべてT_1の$\langle T_1, T_2 \rangle$によるモデルであり，かつ，T_2のモデルは，すべてT_2の$\langle T_1, T_2 \rangle$によるモデルである.

［定理４．２］ $\langle T_1, T_2 \rangle$を安定な理論とする．T_1の中のすべての命題論理式の$\langle T_1, T_2 \rangle$によるモデルは，T_1の$\langle T_1, T_2 \rangle$によるモデルであり，また，T_2の中のすべての命題論理式の$\langle T_1, T_2 \rangle$によるモデルは，T_2の$\langle T_1, T_2 \rangle$によるモデルである. □

［証明］

ＩをT_1の中のすべての命題論理式の$\langle T_1, T_2 \rangle$によるモデルとする．T_1の中の命題論理式は，Ｉにおいて真である.

T_1は無矛盾である．なぜならば，もしT_1が矛盾であるならば，T_1はＩにおいて真でない命題論理式を含まなければならないからである.

$p \in T_1$とする．$I(p) = 1$を示す.

ｐは，それと等価な次のような形に表すことができる:

$$p_1 \wedge \cdots \wedge p_s.$$

ただし，

$$\begin{aligned} p_i = p_{i,1} &\vee L_1 p_{i,2} \vee \cdots \vee L_1 p_{i,m} \\ &\vee L_2 p_{i,m+1} \vee \cdots \vee L_2 p_{i,n} \\ &\vee \neg L_1 p_{i,n+1} \vee \cdots \vee \neg L_1 p_{i,l} \\ &\vee \neg L_2 p_{i,l+1} \vee \cdots \vee \neg L_2 p_{i,k}, \end{aligned}$$

$p_{i,1}$は命題論理式.

$(i = 1, \cdots, s)$

$\langle T_1, T_2 \rangle$は安定だから，T_1は論理的帰結について閉じている．ゆえに，$p_1 \wedge \cdots \wedge p_s \in T_1$，$p_1 \wedge \cdots \wedge p_s \models p_i$ $(i = 1, \cdots, s)$ だから，$p_1, \cdots, p_s$は，すべてT_1に属する.

次に，$I(p_i) = 1$ $(i = 1, \cdots, s)$ を示す.

(1) $L_1 p_{i,2}, \cdots, L_1 p_{i,m}, L_2 p_{i,m+1}, \cdots, L_2 p_{i,n}, \neg L_1 p_{i,n+1}, \cdots, \neg L_1 p_{i,l}, \neg L_2 p_{i,l+1}, \cdots, \neg L_2 p_{i,k}$のうち，少なくとも１つが$T_1$に属する場合．$T_1$が無矛盾であることと，命題４．１，および，Ｉは$\langle T_1, T_2 \rangle$による解釈であることから，この中で$T_1$に属する論理式は，Ｉにおいて真である．したがって，$I(p_i) = 1$.

(2) $L_1 p_{i,2}, \cdots, L_1 p_{i,m}, L_2 p_{i,m+1}, \cdots, L_2 p_{i,n}, \neg L_1 p_{i,n+1}, \cdots, \neg L_1 p_{i,l}, \neg L_2 p_{i,l+1}, \cdots, \neg L_2 p_{i,k}$のいずれも$T_1$に属さない場合．定義４．４から，任意の論理式ｑに対して，$L_i q \in T_1$，または，$\neg L_i q \in T_1$のいずれか一方である．ゆえに，$\neg L_1 p_{i,2}, \cdots, \neg L_1 p_{i,m}, \neg L_2 p_{i,m+1}, \cdots, \neg L_2 p_{i,n}$,

$L_1p_{i,n+1},\cdots,L_1p_{i,l},L_2p_{i,l+1},\cdots,L_2p_{i,k}$のすべてが$T_1$に属する．$p_{i,1}$は$p_i$（$\in T_1$）とこれらの論理式からの論理的帰結だから，安定性により，$p_{i,1}\in T_1$．$p_{i,1}$は命題論理式だから，$I(p_{i,1})=1$．$p_{i,1}\models p_i$だから，$I(p_i)=1$．

(1)，(2)のいずれの場合にも，$I(p_i)=1$（$i=1,\cdots,s$）だから，$I(p)=1$．したがって，IはT_1の$\langle T_1,T_2\rangle$によるモデルである．

同様に，T_2についても証明することができる．

□

すなわち，安定な理論$\langle T_1,T_2\rangle$では，T_i（$i=1,2$）の論理式の真理値は，T_iの中の命題論理式の真理値にのみ依存して決まる．

［定理4．3］　理論$\langle T_1,T_2\rangle$が完全であるのは，$\langle T_1,T_2\rangle$が安定であるとき，かつ，そのときに限る．□

［証明］

（十分性）　$\langle T_1,T_2\rangle$は安定であるとする．T_1の$\langle T_1,T_2\rangle$による任意のモデルにおいて真である論理式は，すべてT_1に含まれることを示す．

$p\notin T_1$とする．pが偽であるT_1の$\langle T_1,T_2\rangle$によるモデルが存在することを示す．

pは，それと等価な次のような形に表すことができる：

$$p_1\wedge\cdots\wedge p_s.$$

ただし，p_i（$i=1,\cdots,s$）は定理4．2の証明と同様である．

$\langle T_1,T_2\rangle$は安定だから，T_1は論理的帰結について閉じている．ゆえに，$p_1,\cdots,p_s\models p$，かつ，$p\notin T_1$だから，$p_1,\cdots,p_s$のうち少なくとも1つは，T_1に属さない．これをp_iとする．

p_iは，$p_{i,1},L_1p_{i,2},\cdots,L_1p_{i,m},L_2p_{i,m+1},\cdots,L_2p_{i,n},\neg L_1p_{i,n+1},\cdots,\neg L_1p_{i,l},\neg L_2p_{i,l+1},\cdots,\neg L_2p_{i,k}$のうちのどの論理式からも論理的に帰結され，$p_i\notin T_1$だから，これらの論理式はいずれも$T_1$に属さない．

$p_{i,1}$は命題論理式であり，$p_{i,1}\notin T_1$だから，$p_{i,1}$はT_1の中の命題論理式からの論理的帰結ではない．ゆえに，$p_{i,1}$は偽で，かつ，T_1の中の命題論理式がすべて真である解釈が存在する．これを$\langle T_1,T_2\rangle$による解釈に拡張することができる．これをIとすると，$I(p_{i,1})=0$．

定理4．2により，IはT_1の$\langle T_1,T_2\rangle$によるモデルである．

$\langle T_1,T_2\rangle$は安定だから，$\langle T_1,T_2\rangle$による解釈において真であるL_ip，または，$\neg L_ip$の形の論理式は，すべてT_1に属する．$L_1p_{i,2},\cdots,L_1p_{i,m},L_2p_{i,m+1},\cdots,L_2p_{i,n},\neg L_1p_{i,n+1},\cdots,\neg L_1p_{i,l},\neg L_2p_{i,l+1},\cdots,\neg L_2p_{i,k}$はいずれも$T_1$に属さないから，これらの論理式は，いずれも$\langle T_1,T_2\rangle$による解釈において偽である．特に，いずれも$I$において偽である．

したがって，$I(p_i)=0$であり，さらに，$I(p)=0$．

同様に，T_2についても証明することができる．

したがって，$\langle T_1,T_2\rangle$は完全である．

（必要性）　$\langle T_1,T_2\rangle$は完全であるとする．

I_i（$i=1,2$）をT_iの$\langle T_1,T_2\rangle$による任意のモデルとする．

(1)　$p_1,\cdots,p_n\in T_i$，$p_1,\cdots,p_n\models q$とする．I_iはT_iの$\langle T_1,T_2\rangle$によるモデルだから，I_iにおいて$p_1,\cdots,p_n$はいずれも真である．ゆえに，$I_i(q)=1$．したがって，完全性により，$q\in T_i$．

(2)　$p\in T_i$とする．I_1,I_2は$\langle T_1,T_2\rangle$による解釈だから，I_1,I_2のいずれにおいてもL_ipは真である．したがって，完全性により，$L_ip\in T_1$，かつ，$L_ip\in T_2$．

(3)　$p\notin T_i$とする．I_1,I_2は$\langle T_1,T_2\rangle$による解釈だから，I_1,I_2のいずれにおいても$\neg L_ip$は真である．したがって，完全性により，$\neg L_ip\in T_1$，かつ，$\neg L_ip\in T_2$．

(1)～(3)により，$\langle T_1,T_2\rangle$は安定である．

□

ここで，次のような記法を定める：理論$\langle T_1,T_2\rangle$と前提$\langle A_1,A_2\rangle$に対して，

$$L_iT_i=\{L_ip\mid p\in T_i\},$$
$$\neg L_i\overline{T}_i=\{\neg L_ip\mid p\notin T_i\},\quad (i=1,2)$$
$$A_1{}'=A_1\cup L_1T_1\cup L_2T_2\cup\neg L_1\overline{T}_1\cup\neg L_2\overline{T}_2,$$
$$A_2{}'=A_2\cup L_2T_2\cup L_1T_1\cup\neg L_2\overline{T}_2\cup\neg L_1\overline{T}_1$$

とする．

$L_iT_i,\neg L_i\overline{T}_i$（$i=1,2$）は，エージェントが持つメタ知識に関する仮定であり，$A_1{}'$，$A_2{}'$は，これらとエージェントにあらかじめ与えられた前提を合わせたものである．

［補題4．4］　理論$\langle T_1,T_2\rangle$が完全で，前提$\langle A_1,A_2\rangle$に対し，$A_1\subseteq T_1$，かつ，$A_2\subseteq T_2$であるならば，

$$\{q\mid A_1{}'\models q\}\subseteq T_1,$$

かつ，

$$\{q\mid A_2{}'\models q\}\subseteq T_2$$

である．□

[証明]

IをT_1の$\langle T_1, T_2 \rangle$による任意のモデルとする．

(1)　$q \in A_1$とする．$A_1 \subseteq T_1$より，$I(q)=1$．

(2)　$q \in L_i T_i\ (i=1,2)$とする．qは$L_i p$という形の論理式であり，$p \in T_i$．Iは$\langle T_1, T_2 \rangle$による解釈だから，$I(q)=1$．

(3)　$q \in \neg L_i \overline{T}_i\ (i=1,2)$とする．$q$は$\neg L_i p$という形の論理式であり，$p \notin T_i$．$I$は$\langle T_1, T_2 \rangle$による解釈だから，$I(q)=1$．

(1)〜(3)により，A_1'の論理式はすべてIにおいて真であり，$A_1' \models q$となるqもIにおいて真である．したがって，完全性により，$q \in T_1$．

同様に，T_2についても証明することができる．□

定理４．３により，安定性は理論が完全であることを保証する．しかし，与えられた前提に関して理論が健全であるかどうかについては何も述べていない．これは，安定性の条件が，理論の中に含まれてはならない論理式について，何も規定していないからである．そこで，次に依存性という概念を定義し，健全性について考察する．

[定義４．５]（依存性）　理論$\langle T_1, T_2 \rangle$が前提$\langle A_1, A_2 \rangle$に依存 (grounded) しているとは，

$$T_1 \subseteq \{q \mid A_1' \models q\},$$

かつ，

$$T_2 \subseteq \{q \mid A_2' \models q\}$$

であることをいう．□

理論が前提に依存していることの直観的な意味は，各エージェントが持つ知識は，前提知識とメタ知識に関する仮定から帰結されるということである．依存性は，与えられた前提に関して理論が健全であることを保証する，次の定理４．５は，このことを示している．

[定理４．５]　理論$\langle T_1, T_2 \rangle$が前提$\langle A_1, A_2 \rangle$に関して健全であるのは，$\langle T_1, T_2 \rangle$が$\langle A_1, A_2 \rangle$に依存しているとき，かつ，そのときに限る．□

[証明]

（十分性）　$\langle T_1, T_2 \rangle$は$\langle A_1, A_2 \rangle$に依存しているとする．

IをA_1の$\langle T_1, T_2 \rangle$によるモデルとする．IがT_1の$\langle T_1, T_2 \rangle$によるモデルであることを示す．

(1)　$q \in A_1$とする．このとき，$I(q)=1$．

(2)　$q \in L_i T_i\ (i=1,2)$とする．qは$L_i p$という形の論理式であり，$p \in T_i$．Iは$\langle T_1, T_2 \rangle$による解釈だから，$I(q)=1$．

(3)　$q \in \neg L_i \overline{T}_i\ (i=1,2)$とする．$q$は$\neg L_i p$という形の論理式であり，$p \notin T_i$．$I$は$\langle T_1, T_2 \rangle$による解釈だから，$I(q)=1$．

(1)〜(3)により，A_1'の論理式はすべてIにおいて真であり，$\{q \mid A_1' \models q\}$の論理式もすべて$I$において真である．したがって，依存性により，$T_1$の論理式はすべて$I$において真である．

同様に，A_2の$\langle T_1, T_2 \rangle$によるモデルは，すべてT_2の$\langle T_1, T_2 \rangle$によるモデルであることを証明することができる．

したがって，$\langle T_1, T_2 \rangle$は$\langle A_1, A_2 \rangle$に関して健全である．

（必要性）　$\langle T_1, T_2 \rangle$は$\langle A_1, A_2 \rangle$に関して健全であるとする．

$p \in T_i\ (i=1,2)$ならば，$L_i p \in A_1'$だから，$L_i p$はA_1'のどのモデルにおいても真である．また，$p \notin T_i\ (i=1,2)$ならば，$\neg L_i p \in A_1'$だから，$L_i p$はA_1'のどのモデルにおいても偽である．

ゆえに，A_1'のモデルにおいて$L_i p$が真であるのは，$p \in T_i$であるとき，かつ，そのときに限る．すなわち，A_1'のモデルは，$\langle T_1, T_2 \rangle$による解釈である．また，A_1'のモデルにおいて，A_1の論理式はすべて真である．

健全性により，A_1の$\langle T_1, T_2 \rangle$によるモデルは，T_1の$\langle T_1, T_2 \rangle$によるモデルである．したがって，A_1'のモデルは，T_1の$\langle T_1, T_2 \rangle$によるモデルである．すなわち，T_1の論理式は，A_1'の論理式の論理的帰結である．

同様に，T_2についても証明することができる．

すなわち，$\langle T_1, T_2 \rangle$は$\langle A_1, A_2 \rangle$に依存している．

□

最後に，与えられた前提とその拡張の関係を示す．

[定理４．６]　前提$\langle A_1, A_2 \rangle$と理論$\langle T_1, T_2 \rangle$に対して，$\langle T_1, T_2 \rangle$が$\langle A_1, A_2 \rangle$の拡張であるのは，

$$T_1 = \{q \mid A_1' \models q\},$$

かつ，

$$T_2 = \{q \mid A_2' \models q\}$$

であるとき，かつ，そのときに限る．□

[証明]

（必要性）　$\langle T_1, T_2 \rangle$は$\langle A_1, A_2 \rangle$の拡張であるとする．補題４．４と，定理４．５により，

$$T_1 = \{q \mid A_1' \models q\},$$

かつ，

$T_2=\{q \mid A_2'\models q\}$

である．

（十分性）

$T_1=\{q \mid A_1'\models q\}$,

かつ,

$T_2=\{q \mid A_2'\models q\}$

であるとする．

このとき，$A_1\subseteq T_1$，かつ，$A_2\subseteq T_2$である．

また，定理4.5により，$\langle T_1,T_2\rangle$は$\langle A_1,A_2\rangle$に関して健全である．

さらに，次の(1)～(3)により，$\langle T_1,T_2\rangle$は安定である．ゆえに，定理4.3により，$\langle T_1,T_2\rangle$は完全である．

(1) $p_1,\cdots,p_n\in T_i$ $(i=1,2)$，$p_1,\cdots,p_n\models p$とする．$p_1,\cdots,p_n\in\{q \mid A_i'\models q\}$だから，$A_i'\models p$．すなわち，$p\in T_i$．

(2) $p\in T_i$とする．このとき，$L_i p\in A_1'$，かつ，$L_i p\in A_2'$．$A_1'\subseteq T_1$，かつ，$A_2'\subseteq T_2$だから，$L_i p\in T_1$，かつ，$L_i p\in T_2$．

(3) $p\notin T_i$とする．このとき，$\neg L_i p\in A_1'$，かつ，$\neg L_i p\in A_2'$．$A_1'\subseteq T_1$，かつ，$A_2'\subseteq T_2$だから，$\neg L_i p\in T_1$，かつ，$\neg L_i p\in T_2$．

したがって，$\langle T_1,T_2\rangle$は$\langle A_1,A_2\rangle$の拡張である．

□

5. 階層的な知識

本章では，本論文で拡張した自己認識論理を用いて，階層的な知識の表現とその下での推論について考察する．これは，フレームモデル[9]による知識表現に対し，論理的な意味付けを与えると考えられる．簡単のために，ここでは上位階層と下位階層の２層だけの場合について論じる．そのような知識の例として，

「鳥は飛ぶ」，
「鳥は翼を持つ」，
「ペンギンは飛ばない」，
「ペンギンは鳥である」

という知識を考える．

これを命題論理や１階述語論理で表現すると，次のようになる：

$T=\{\text{Is-Bird}\Rightarrow\text{Fly},\ \text{Is-Bird}\Rightarrow\text{Has-Wing},$
$\quad\text{Is-Penguin}\Rightarrow\neg\text{Fly},$
$\quad\text{Is-Penguin}\Rightarrow\text{Is-Bird}\}$

「ペンギンである」という知識があるとき，これを Is-Penguin と表現し，Tに加えると，$T\cup\{\text{Is-Penguin}\}$ は矛盾する．

そこで，非単調論理[1,2,3,4]ではペンギンは鳥の例外であると考えて，「鳥は通常飛ぶ」，「ペンギンは通常飛ばない」という表現をとる．

このうち，デフォルト論理[2]では，次のように表現される：

$D=\{\text{Is-Bird}:M\text{Fly}/\text{Fly},$
$\quad\text{Is-Bird}:M\text{Has-Wing}/\text{Has-Wing},$
$\quad\text{Is-Penguin}:M\neg\text{Fly}/\neg\text{Fly}\}$,

$W=\{\text{Is-Penguin}\Rightarrow\text{Is-Bird}\}$.

「ペンギンである」という知識 Is-Penguin があるとき，これをWに加えてできるデフォルト理論$\langle W\cup\{\text{Is-Penguin}\},D\rangle$の拡張は，次の$E_1,E_2$の２つである：

$E_1=Th(\{\text{Is-Penguin},\ \text{Is-Bird},\ \neg\text{Fly},\ \text{Has-Wing}\})$,

$E_2=Th(\{\text{Is-Penguin},\ \text{Is-Bird},\ \text{Fly},\ \text{Has-Wing}\})$.

また，nonmonotonic logic[3,4]や，自己認識論理[1]では，次のように表現される：

$A=\{\text{Is-Bird}\wedge M\text{Fly}\Rightarrow\text{Fly},$
$\quad\text{Is-Bird}\wedge M\text{Has-Wing}\Rightarrow\text{Has-Wing},$
$\quad\text{Is-Penguin}\wedge M\neg\text{Fly}\Rightarrow\neg\text{Fly},$
$\quad\text{Is-Penguin}\Rightarrow\text{Is-Bird}\}$.

この場合も，前提$A\cup\{\text{Is-Penguin}\}$ の拡張は，Fly を含むものと$\neg$Fly を含むものの２つである．

このように，従来の非単調論理では，「ペンギンは飛ばない」を含む拡張世界と，「ペンギンは飛ぶ」を含む拡張世界の２つの拡張世界が存在する．

ところで，フレームモデル[9]による知識表現では，図１に示すように，鳥とペンギンの２つのフレームを用意し，鳥フレームのスロットに

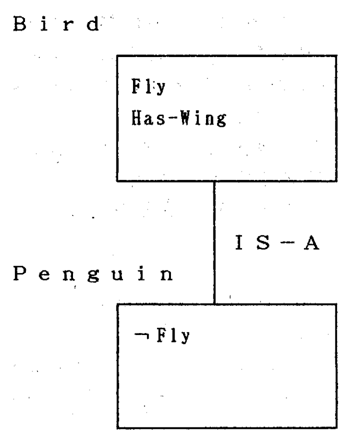

図１．　フレームモデルによる表現

「飛ぶ」,「翼を持つ」という属性を記述し，ペンギンフレームのスロットに「飛ばない」という属性を記述する．また，これら2つのフレームの間に属性継承を考える．これにより，「鳥は飛ぶ」,「ペンギンは飛ばない」,「ペンギンは翼を持つ」が推論され，「ペンギンは飛ぶ」は推論されない．

表現された知識の意味やその下での推論を形式的に議論するためには，適切な論理体系を構築することが望まれる．一方，フレームモデルのように，階層という概念を取り入れた方が，より適切な表現や推論ができる場合もある．しかし，フレームモデルにおいて表現された知識の形式的意味は明確でない．

ここでは，階層的な知識の表現を次のように考える．すなわち，各階層に与えられた知識は，対象世界の断面の表現であり，各階層は1つの知識システムである．そして，階層ごとにその知識を論理式で表現する．しかし，この表現は一般に不完全であり，階層間の属性継承は，上位階層の知識を利用して，下位階層の知識をより完全にすることと捉える．このためには，上位階層がある知識を持っているというメタ知識を下位階層が持つ必要がある．これは，上位階層から下位階層への，このメタ知識の通信が行われたとも考えられる．

以上のことを，各階層をエージェントと考えて，本論文で拡張した自己認識論理によって形式化する．

まず，そのようなエージェント1とエージェント2の間のメタ知識の通信は，2エージェント系に拡張した自己認識推論で捉えることができる．これは，既に，理論の安定性として特徴付けた．

さらに，エージェント1の持つ知識の利用によるエージェント2の持つ知識の完全化は，

$$L_1 p \wedge \neg L_2 \neg p \Rightarrow p$$

という形の表現の知識とすることができる．この知識は，属性継承は上位階層から下位階層へ行われ，上位階層が下位階層の知識を利用することはできないことから，下位階層に相当するエージェント2にあらかじめ与えておくのが自然である．これにより，エージェント1が知識 p を持つというメタ知識が通信されていて，p が無矛盾であるとき，すなわち，エージェント2が知識 $L_1 p$ を持つが，知識 $\neg p$ を持たないとき，p を推論することにより，エージェント2の知識に p を追加することができる．

前述の例は，次のように表現することができる（図2参照）：

$$A_1 = \{\text{Is-Bird},\ \text{Fly},\ \text{Has-Wing}\},$$
$$A_2 = \{\text{Is-Penguin},\ \neg\text{Fly},\ L_1 p \wedge \neg L_2 \neg p \Rightarrow p\}.$$

ただし，p は任意の論理式である．これは，一般に属性継承では，上位階層から下位階層へ，任意の知識が継承されることによる．

A_1, A_2 は，それぞれ，鳥フレーム，ペンギンフレームのスロットに属性としてあらかじめ与えられた知識の集合である．すなわち，鳥フレームのスロットに与えた属性である「飛ぶ」，「翼を持つ」を表す論理式 Fly, Has-Wing を A_1 に与え，ペンギンフレームのスロットに与えた属性である「飛ばない」を表す論理式 $\neg$Fly を A_2 に与えてある．さらに，A_1 には，「鳥である」を表す論理式 Is-Bird を与え，A_2 には，「ペンギンである」を表す論理式 Is-Penguin を与えてある．これは，各階層に与えられた知識は，対象世界の断面の表現であると捉えたことから，鳥フレームにおける断面では「鳥である」が成り立ち，ペンギンフレームにおける断面では「ペンギンである」が成り立つと考えられるということに基づく．

前提 $\langle A_1, A_2 \rangle$ の拡張 $\langle T_1, T_2 \rangle$ に対して，Fly, Has-Wing $\in T_1$，$\neg$Fly, Has-Wing $\in T_2$ であるが，Fly $\notin T_2$ である．T_1, T_2 はそれぞれ，鳥フレーム，ペンギンフレームにおいて成り立つ知識の集合である．すなわち，「鳥は飛ぶ」,「ペンギンは飛ばない」,「ペンギンは翼を持つ」が推論され，「ペンギンは飛ぶ」は推論されない．

ところで，本論文では，2つのエージェントの間に十分な通信があることを仮定し，これは

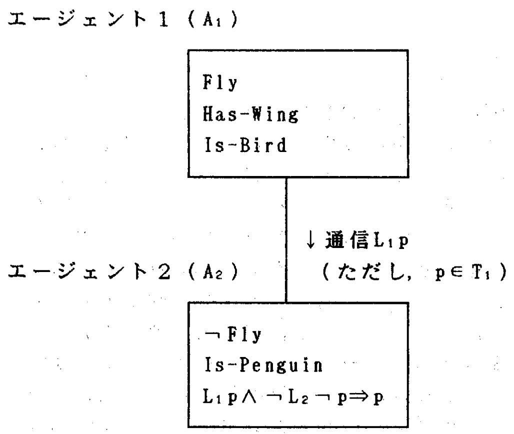

図2．　2エージェント系の自己認識論理による表現

形式的には，安定性として定義した．すなわち，上述の例では，エージェント１が知識ｐを持つ（持たない）とき，エージェント１が知識ｐを持つ（持たない）というメタ知識をエージェント２が持つが，同時に，エージェント２が知識ｑを持つ（持たない）とき，エージェント２が知識ｑを持つ（持たない）というメタ知識をエージェント１が持つことになる．これは，これらエージェントの間において，メタ知識の通信が双方向に行われるということを仮定していることになる．

ところが，本章で示したような階層的な知識の表現とその下での推論においては，一般に，下位階層は上位階層の知識を利用できるが，逆に，上位階層は下位階層の知識を利用することはできない．したがって，上位階層に相当するエージェントが，下位階層に相当するエージェントが持つ知識についてのメタ知識を持つことは必要でない．そこで，エージェントの間のメタ知識の通信や，各エージェントが持つメタ知識は，この通信の方向性を考慮して形式化することが望まれる．

しかし，知識システムの利用者が各階層に対して行う質問は，実際には，対象世界に関する質問が主であると考えられる．したがって，エージェントの間に双方向の通信を仮定し，他のエージェントの知識に関するどのようなメタ知識もエージェントは持つことができることを仮定していても，エージェントが持つ知識のうち，対象世界に関する知識（オブジェクト知識）だけに注目すれば，本論文で拡張した自己認識論理を用いて，階層的な知識を特徴付けることができる．

６．おわりに

理想的に合理的な２つのエージェントの間で十分な通信がある場合に対して，自己認識論理を自然に拡張し，各エージェントが与えられた前提知識から自己認識推論によって得られる知識の集合，すなわち，前提知識の拡張を安定性と依存性という２つの概念で特徴付けられることを示し，前提知識とその拡張の関係を示した．さらに，この拡張した自己認識論理を用いて，階層的な知識の表現を行い，そのような知識の下での推論を形式的に特徴付けた．これは，フレームモデルによる知識表現に対する論理的な意味付けを与えると考えられる．

本論文では，双方向通信を行う２エージェント系を扱った．エージェントが３つ以上の場合に対する自己認識論理の拡張や，エージェントの間の通信の方向性を考慮して，各エージェントが自己認識推論によって得られる知識を特徴付けることは，今後の課題である．

謝辞

日頃御指導下さる豊橋技術科学大学本多波雄学長，中京大学福村晃夫教授，並びに御討論下さる名古屋大学坂部俊樹助教授，平田富夫助教授をはじめ，研究室の皆様に感謝致します．

参考文献

(1) Moore, R.C.: Semantical Considerations on Nonmonotonic Logic, Artif. Intell., 25, 1, pp.75-94 (1985).

(2) Reiter, R.: A Logic for Default Reasoning, Artif. Intell., 13, 1/2, pp.81-132 (1980).

(3) McDermott, D., Doyle, J.: Non-Monotonic Logic I, Artif. Intell., 13, 1/2, pp.41-72 (1980).

(4) McDermott, D.: Nonmonotonic Logic II, JACM, 29, 1, pp.33-57 (1982).

(5) Konolige, K.: On the Relation between Default and Autoepistemic Logic, Artif. Intell., 35, 3, pp.343-382 (1988).

(6) Etherington, D.W.: Formalizing Non-monotonic Reasoning, Artif. Intell., 31, 1, pp.41-85 (1987).

(7) McCarthy, J.: Circumscription - A Form of Non-Monotonic Reasoning, Artif. Intell., 13, 1/2, pp.27-39 (1980).

(8) McCarthy, J.: Applications of Circumscription to Formalizing Common-Sense Knowledge, Artif. Intell., 28, 1, pp.89-116 (1986).

(9) Minsky, M.: A Framework for Representing Knowledge, Winston, P.H. (ed.), The Psychology of Computer Vision, McGraw-Hill (1975).

(10) 外山勝彦，稲垣康善：自己認識論理のエージェント系への拡張とその階層的知識の表現への応用，信学会情報基礎理論セミナー論文集，pp.123-124 (1989).

(11) 外山勝彦，稲垣康善：自己認識論理のエージェント系への拡張，情処学会全大論文集，pp.462-463 (1989).