

Infant Agents 間の対話による概念獲得と対話戦略が与える影響

田口 亮, 山本 航, 桂田 浩一, 新田 恒雄

豊橋技術科学大学 大学院工学研究科

〒441-8580 豊橋市天伯町雲雀ヶ丘 1-1

E-mail: (taguchi, yamato, katurada)@vox.tutkie.tut.ac.jp, nitta@tutkie.tut.ac.jp

あらまし 本報告では Infant Agent (IA) と人間, および IAs 間の対話による概念獲得について述べる. 実験の第一段階では, 3名のMother (人間) が, 各々のIA に対してスクリーン上のオブジェクトを指示しながら, オブジェクトの特徴(色・形・位置, 操作)を発話することで概念を教える. 第二段階には, 各IAが他のIAと対話を通して概念を共有化する. IAs 同士の対話学習においては, IAの質問の仕方や答え方といった対話戦略の良否が重要になる. そこで教示IAを特定した教師あり対話学習と, 教示IAを特定しない教師なしの対話学習の二つのケースについて, それぞれ対話戦略の比較実験を行った. 概念獲得に至る対話数を指標とする実験結果から, 対話戦略の立て方の重要性が示された.

キーワード 概念獲得, Infant Agent, 対話戦略, 相互学習

Effect of Dialogue Strategies on Concept Acquisition Made with Interactions among Infant Agents

Ryo Taguchi Wataru Yamamoto Koichi Katurada and Tsuneo Nitta

Graduate School of Engineering, Toyohashi University of Technology

E-mail: (taguchi, yamato, katurada)@vox.tutkie.tut.ac.jp, nitta@tutkie.tut.ac.jp

Abstract: This paper described concept acquisition through the interaction between an infant agent and a mother (human) and the interaction among infant agents (IAs). At the first stage, three mothers teach concepts to each IA by pointing an object on a screen and speaking an attribute of the object. The attribute includes color, figure, location, and operation to an object. At the second stage, IAs interact each other and share the concepts. Effective interaction needs sophisticated dialogue strategies, such as how to ask and how to answer. We investigated two types of experiments, the first concept learning assigns a supervisor to an IA and the second one does not assign a supervisor. The experimental results, that are evaluated using a number of dialogue turns for each strategy, showed the importance of dialogue strategies.

Keyword Concept Acquisition, Infant Agent, Interaction Strategies, Mutual Learning

1. はじめに

マルチモーダル対話 (MMI) システムでは, フロントエンドの知覚処理以上に, 背後の知能処理を与える機構の役割が重要である. コンピュータとの対話では, 人間が対話で利用している知能処理と同様の処理機構を, コンピュータが獲得して初めて円滑な対話が成立する. 当面, コンピュータの聴覚モダリティと視覚モダリティを上位の知能レベルに関連付けながら, 適切な応答を返す一連の処理が必要である. こうした処理は, 対話シナリオを記述する言語 [1] を用いて代行する方法が当面実用的であろう. しかし, 人間が記述するルールの上でしか動作

しない方法には限界がある. 一方, テキストや音声 (ディクテーション) を用いて, 人間が予め記述した画像に対するアノテーションを利用する方法が試みられている. また, 一步進めて画像 (中の個々のオブジェクト) と対応する音声 (言語) を関連付けながら概念 (画像と音声との関係) を獲得する機構の研究も始められている [2]-[7]. こうした研究は, コンピュータに概念と言語を獲得させることを可能にする.

本研究は, 複数のエージェントが相互にマルチモーダル対話を行う実験を通して, 知覚情報から高次知能情報処理機構を獲得することを可能にする方法論の確立を目

指している。本報告ではその第一歩として、複数の教示者（人間）が育てたコンピュータ内の幼児エージェント（以下 IAs: Infant Agents と呼ぶ）を連携させ、概念獲得を効率良く行う方法を検討する。まず2章で概念獲得と文法獲得について述べる。3章では人間による教示、4章では教示役 IA による教示を通じた概念獲得について述べる。続く5章では前章の課題を受け、新たな戦略を追加し対話戦略を比較する。6章では教示 IA を特定しない対話学習について述べた後、最後の7章で実験から得た知見をまとめる。

2. 概念と文法の獲得 [8]

IAs は、知識を持たない状態から学習を始め、同一空間内にある画像オブジェクトの特徴を視覚情報として、また、人間や他の IAs が発する音声の特徴を聴覚情報として受け取った後、両者の対応関係を概念として獲得する。同時に、概念を獲得した IAs は、概念がどのような時系列として現れたかを確率的に表現し、文法を獲得する。図1にIAによる概念と文法の獲得の流れを示す。

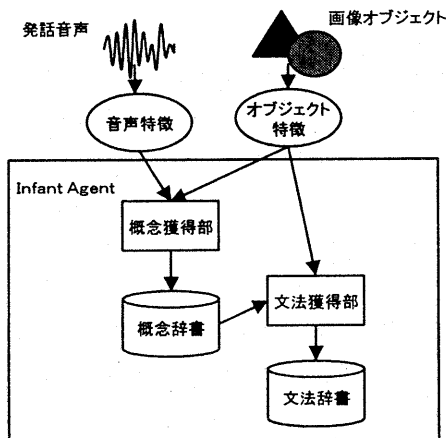


図1: IAによる概念と文法の獲得の流れ

2.1. 視覚情報

IAs は視覚と聴覚情報を刺激として受け取り、学習を進める。視覚情報は、オブジェクトの属性とオブジェクトに対する操作（移動）を含むオブジェクト特徴として与えられる。聴覚情報は発話から得られる音声特徴として与えられる。教示者である人間（Mother）は、視覚・聴覚情報を対して IAs に与える。以下では、視覚情報と聴覚情報の与え方について説明する。

2.1.1. 視覚情報の与え方

今回は、あらかじめコンピュータ内に画像オブジェクトを配置した世界を構成し、色や形等の属性を与えると共に、空間座標等もあらかじめ与えておく。すなわち、画像オブジェクトは特徴抽出が正確に行われ、仮想空間

内に配置されたと仮定する。

(1) 画像オブジェクトの操作

図2に示す画像オブジェクトを画面上に配置した実験システムを製作した。画面中央には IA が表示され、Mother は IA の周囲のオブジェクトの指示と移動を行いながら、IA に視覚情報を教示する。

(2) オブジェクト特徴

今回は、画像オブジェクト及び操作に対して形状（丸、三角、四角）、色（赤、青、白）、位置（上、下、左、右）、動作（移動）の計11個の属性与えた。各属性は、それぞれ一つのベクトル（以下オブジェクトベクトル）として表わされている。オブジェクトベクトルを連結したベクトルは、個々のオブジェクトの特徴（オブジェクト特徴）を表し、IA はこれを視覚情報として取り込む。

(3) オブジェクトの操作履歴の保存

Mother がシステム内のオブジェクトを操作すると、オブジェクト特徴が得られる。Mother が、オブジェクトを繰り返し操作する度に、オブジェクト特徴はファイルに順に記録される。操作履歴は、学習の際、1行づつ記録され IA へ順に入力される。

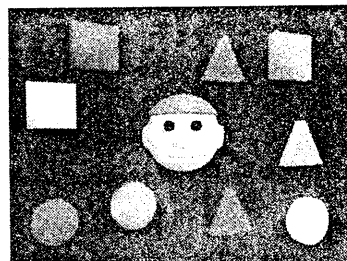


図2: オブジェクト操作システム

2.1.2. 聴覚情報の与え方

Mother による教示音声はオブジェクトの操作と同時に録音される（サンプリング周波数 16[kHz]、量子化ビット数 16[bit]）。録音音声は、1発話ごとに切り出された後、特徴量に変換される。特徴量としては、標準的な MFCC パラメータセット（MFCC(12)+ Δ Cep(12)+ Δ Δ Cep(12)+ Δ P+ Δ Δ P）を使用した。これを聴覚情報とする。

2.2. 概念と文法の獲得

IAs は視覚情報（オブジェクト特徴）と聴覚情報（単語音声特徴）の対応関係を概念として学習する。聴覚情報はオブジェクトの特徴を単語音声として与えるだけではなく、簡単な文の発話として与えることができる。また、概念を獲得した IAs は、入力音声に含まれる属性の並びを文法として獲得する。

2.2.1. 概念の獲得

IAs は知識を持たない状態から学習を始める。IA は、教示された情報（聴覚情報、視覚情報）を、図3に示す

形式に変換する。この図の2行目以下は、音声特徴（概念呼称）を形成するオブジェクト特徴（属性ごとのベクトルで表現）を示している。図4は、教示情報の例を示す。オブジェクトベクトルに対応する部分が1となっている。

音声特徴ファイル		
指示	移動	-
赤丸	青	白
左	四角	三角
上	右	-
	下	-

図3：教示情報

file1		
1	0	-
1	0	0
0	0	1
0	1	-
1	0	-

図4：教示情報の例

IAs は、最初の教示があると、図4に示した教示情報を履歴として登録する。これを概念辞書と呼ぶ。概念辞書は、複数の音声特徴を辞書項目としたオブジェクト特徴の記録となっている。以降は教示があると、入力聴覚情報の音声特徴と、既に登録されている音声特徴との間で、フレーム同期 DP マッチングを行う。今回は、入力音声に含まれる単語数を最大3語までとし、1~3単語に対してフレーム同期 DP 法を適用して距離を求め、得られた三つの距離から最適単語数を判定する。次に、単語（概念呼称）ごとにこの距離を求めた後、距離が一定の閾値より小さい場合は一致音声とし、大きい場合は未知音声とし、以下の処理を行う。

まず音声特徴が一致した場合は、図5の例に示すように、対応する概念辞書のオブジェクト特徴に、新しい教示から得たオブジェクト特徴が加えられる。ここで教示音声から1単語と判断されると、辞書項目の音声特徴に教示音声特徴が追加され、登録パターン数が増える。一致するものがない場合は、1単語と判断されたときのみ、新しい概念として教示情報が概念辞書に新規登録される。

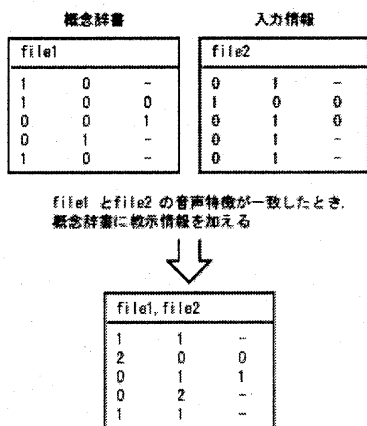


図5：概念辞書への追加

オブジェクト特徴が以下の条件を全て満たした場合、音声特徴とそのオブジェクト特徴の対応を、概念として確定する。

- (1) 学習回数が3回以上である。
- (2) ある属性内で、同じ特徴が現れた頻度が90%以上である。
- (3) (2)を満たす属性が一つだけである。

2.2.2. 文法の獲得

IAs はオブジェクトの属性を品詞に見立て、教示音声の各品詞間の接続を求めた後、接続確率を 2-gram に置き換え確率正規文法を獲得する。

3. 人間が教示して概念を獲得した場合

3.1. 実験条件

学習は、単一の概念呼称（1単語音声）を全ての概念についてランダムな順序で教示し、これを5回繰り返すことから始める。以降は、1発話内の概念数を徐々に増やしながら、ランダムな順序で発声・教示した。教示文は、全てあらかじめ与えられた文法（位置、色、形状、動作の順）に従っている。

学習実験は、3名のMotherが各々IA-A、IA-B、IA-Cを教示する形で進められ、各128組のデータ（実世界オブジェクトの属性ベクトルと音声特徴の組）を得た。これらの学習データを使用して、概念獲得および文法獲得をIA別に行った。

3.2. 実験結果と考察

全てのオブジェクトの属性ベクトルについて、頻度の高い音声特徴を調べたところ、全ての概念関係を正しく獲得していた。また、文法獲得についても、最終的に得られた 2-gram を観察した結果、3例とも、位置、色、形状、動作の順となり、学習の際の発話語順と一致した。図6は、一回の発話で教示した概念数（単語数）を横軸に、単語マッチング時の正解率を縦軸に取ったものである。教示音声を誤認識することはなかったが、未知語として扱われた例が見られた。教示者により差はあるが、単語数が多いほどリジェクトが増え、正解率が低下する傾向がある。

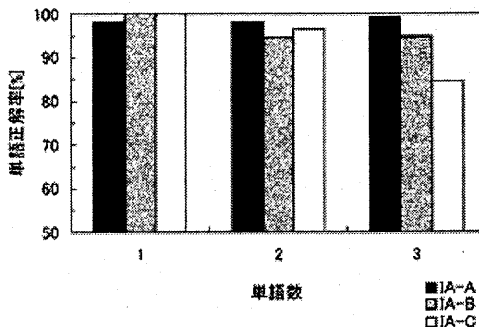


図6：単語正解率

4. IAs 間の対話学習：教示 IA を特定した場合

IAs がネットワーク上に分散して存在する状況を考える。個々の IA は自律的に他の IAs と対話しながら、未獲得の概念を得ることができる（相互学習）。以下、異なる学習過程を持つ IAs 同士が、対話によって概念を獲得する場面を想定する。なお 4 章と 5 章では、質問者側の IA（質問 IA）と、教示者側の IA（教示 IA）を分け、一方向の学習を考える。質問 IA はオブジェクトを示し、それに対する教示 IA の応答（オブジェクト操作+音声発話）を通して未獲得の概念を学習する。

4.1. 対話戦略

対話を協調的に進めることの重要性が指摘されている（Grice の公準（質、量、関係、様態）[9]）。対話による概念獲得を円滑に進めるためには対話戦略が不可欠である。そのためここでは Grice の公準を考慮して二つの対話戦略（質問戦略と教示戦略）を導入する。

質問戦略：質問 IA は、教示 IA から得た音声で、概念辞書に存在しない場合、新概念として辞書登録した後、教示 IA に再発話を求める。また、複数単語を含む音声の一部が未知語の場合は、その部分の再発話を求め、再発話で得られた音声を概念登録した後、複数単語を含む音声で学習を行う。（なお、質問を音声で伝える部分は、現在未実装である）

教示戦略：教示 IA は、質問 IA に対して音声で応答する。この時、1 発話に含まれる単語数を学習が進むにつれ徐々に増加させる。ただし、質問 IA から、再発話を求められた時は、発話の単語数を減らして応答する（1 単語の場合は、概念辞書内の別の音声を使用して応答する）。一方、質問 IA から、一部分の再発話が求められた時は、対応部分の発話を行う。

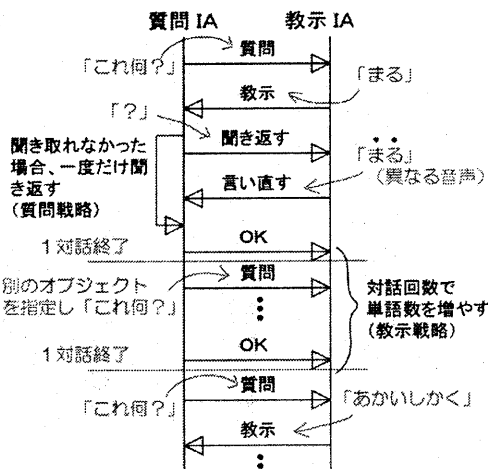


図 7：教示戦略と質問戦略を用いた対話学習

両戦略を同時に用いた場合の対話学習の流れを図 7 に示す。

4.2. 実験条件

2 章に説明した概念獲得後の IA-A, IA-B, IA-C を順に教示 IA として、相互学習を行う。実験は以下の二つに分けて行った。

(1) 対話戦略の違いの影響：

対話戦略の使用を以下の五種類に分けて比較した。

- 戦略なし（教示の単語数はランダム）
- 戦略なし（1 単語発話のみで教示）
- 教示戦略のみ（10 回ごとに単語数を増加）
- 質問戦略のみ
- 対話戦略あり（質問と教示の二つの戦略を使用。ただし、同じオブジェクトに対する発話を 2 往復までに制限した）

なお、質問 IA には初期知識を与えていない。

(2) 初期知識の違いの影響：(1) のうち、対話戦略ありの場合について、初期知識の影響を調べる。但し、異なる話者の音声を DP マッチングする場合、現行の認識方式では対応できないため、今回は教示 IA, 質問 IA は同じ話者が発話したものを使用する。なお、実験では教示 IA が初期に持つ概念数を、3 と 6 の二通り（ランダムに選択）とした。

(1) および (2) の実験を、全ての IA (IA-A, IA-B, IA-C) について行い、平均を求めた。

4.3. 実験結果

対話戦略の違いが IA の概念獲得過程に及ぼす影響を図 8 に示す。IAs が初期概念を持たない場合、対話戦略なしで概念を得ることは不可能である。教示者が一方的に 1 単語発話のみで教示すれば、戦略なしでも概念を与えることはできるが、この場合、文法は獲得できないことになる。教示戦略だけを導入した場合は、獲得できない概念および文法が生じた。

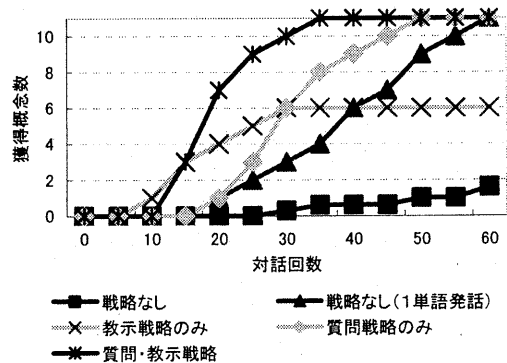


図 8：概念獲得過程（対話戦略の違い）

次に、質問戦略のみを適用すると全ての概念と文法を獲得できた。さらに、教示戦略を含む対話戦略を用いると、学習効率も良く、文法も全て獲得することができた。

初期知識の違いが IA の概念獲得過程に及ぼす影響を図 9 に示す。教示戦略の単語数を増加させる条件は「教示 10 回ごと」であり、初期知識に依存しない。そのため、質問 IA が初期知識を持つ場合、1 語発話による教示が行われる対話の初期段階では、既知の概念しか教示されない可能性が高く、概念獲得が効率良く進まない。その結果、図に示すように初期知識の量が多くなるにつれ、学習効率が低下する。これは 1.1 程度程度の少ない概念数の場合なら弊害は少ないが、概念数が増えた場合や、不完全な IA 同士の相互学習を行う場合には、概念獲得を遅らせる大きな障害となる。

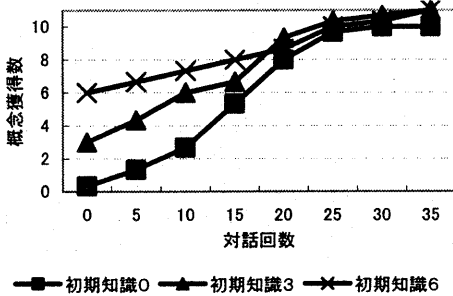


図 9：概念獲得過程：初期知識の違い

5. 質問者の初期知識を戦略に組み入れた場合

4 章での課題を踏まえ、ここでは以下に述べる新たな戦略を導入する。

質問 IA は常にオブジェクト操作全体について質問する。しかし実際は、その操作に関する概念のいくつかは既知であり、それ以外の概念を教示されなければ、質問 IA の概念獲得は進まない。質問 IA の知識量が増えるに従い、教示が無駄になるケースが多くなるからである。教示 IA の立場から見ると、質問された情報に Grice の量の公準に対応する工夫がないため、無駄な教示は避けられない。

5.1. 質問 IA の戦略

量の公準を満たすためには、質問 IA の知りたいことを質問に付加する必要がある。これには「この形は何？」という質問方法が考えられるが、そのためには「形」という上位概念を知っていなければならない。そこで逆に、質問 IA は知っている事を発話し、教示 IA はその発話を踏まえて教示することを考える。そうすれば上位概念がなくても、ある程度、量の公準を満たす事ができる。

質問 IA が、既知の概念を発話した場合の対話の流れを図 10 に示す。質問 IA は質問時に、対象のオブジェ

クト操作に関する既知の概念を 1～3 語で発話する。教示 IA はその発話を評価し、質問 IA の発話よりも 1 語多く（最大 3 語）発話する。このとき質問の内容に応じて教示内容を変えることも考えられるが、今回は教示 IA の発話内容をランダムに決定した。以下、これまでの教示戦略を旧戦略、ここで述べた戦略を新戦略と呼ぶ。

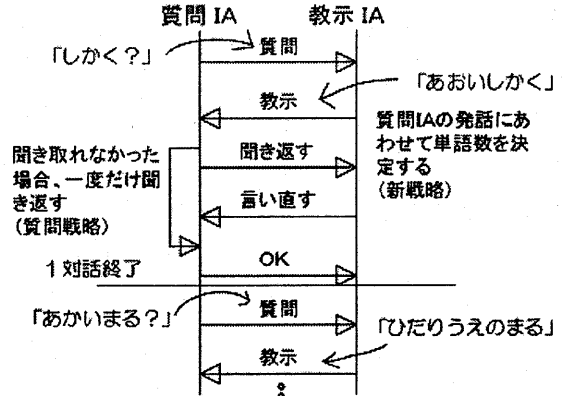


図 10：質問 IA の戦略

5.2. 戦略の比較実験

4.2 で説明した条件の下で、旧戦略と新戦略の学習効率比較を行った。その結果を図 11 に示す。旧戦略は初期概念が増えるに従って獲得効率が悪くなるのに対して、新戦略では獲得概念数が多いほど、効率が良くなっていることが分かる。新戦略は相手の学習進捗を発話から間接的に知ることができるため、早い段階から 2～3 語の発話を行うことができる。それに対して旧戦略は、単語数を増やす閾値（対話回数）によってその効果が大きく変わる。今回使用した 10 回という閾値は、堅実な戦略となっていることから、初期知識が少ない場合に効率が良い反面、多い場合は効率が悪くなった。

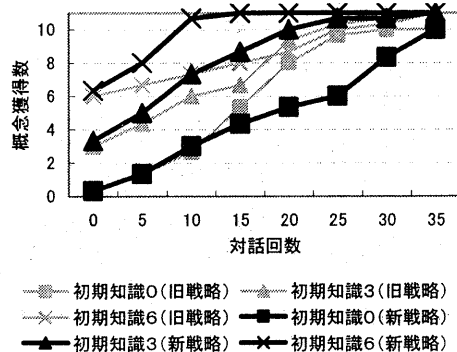


図 11：新戦略と旧戦略の比較（教示者は特定）

6. IAs 間の対話学習：教示 IA を特定しない場合

これまでの完全な知識を持った IA が一方的に教示する対話学習の場合を扱ってきた。以下では学習の参加者が相互に発話しあい、お互いが同時に概念を獲得できる場合、言い換えると教示 IA を特定しない場合を考える。

6.1. 実験条件

前述した実験条件のうち、一方的であった質問-教示を、一対話ごとに交代するように変更して教示 IA を特定しない対話学習を行った。

IA-A (獲得概念数 8) と IA-B (獲得概念数 6) を用意する。これまでの実験と同様に、IA-A, IA-B は同じ話者により教示された音声を使用した。使用概念は前章と同じ 11 個である。初期の共有概念数は 3 とした。

実験はまず、IA-A と IA-B による対話学習の実験を行い、旧戦略と新戦略の比較を行った。次に、質問 IA の発話を利用して教示 IA が概念を獲得できるよう改造し、同様の実験を行った。

6.2. 実験結果

実験の結果を図 1 2 と 1 3 に示す。ここでは横軸をターン (発話の一往復) 数としている。

共有概念が 3 個という条件は、5 章の実験の初期概念数 3 個に対応する。この実験でも新戦略は 5 章の結果と同等の効率が得られた。また、質問時の発話で教示も概念獲得を行うようにした場合、質問-教示という役割は無くなり、お互いが知っていることを発話し合うことで、1 つの対象に対し同時に概念を共有化することができる。そのため、図に示すように、旧戦略と比べかなり効率が良くなった。

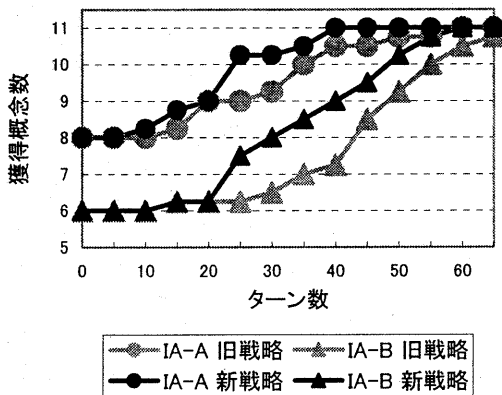


図 1 2 : 新戦略と旧戦略の比較 (教示者は特定せず)

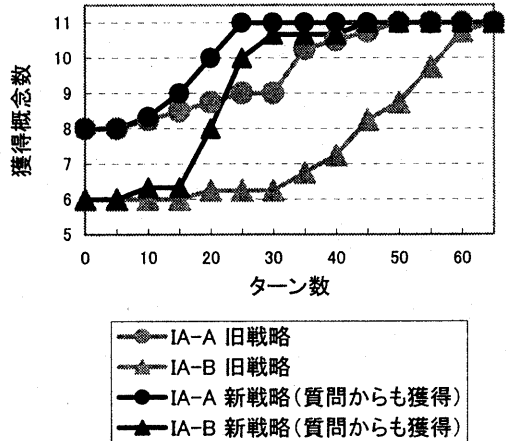


図 1 3 : 新戦略 (質問からも獲得) と旧戦略の比較

7. まとめ

本報告では、Infant Agents が得た概念を共有化する過程に注目し、対話戦略の重要性を明らかにした。今後は、概念の拡張 (上位概念等) と共に、より多数の IAs を用いた際の振る舞いを評価したい。

文 献

- [1] 小林 聡, 中村 有作, 桂田 浩一, 山田 博文, 新田 恒雄: "マルチモーダル対話記述言語 XISL の提案" 情報処理学会研究報告 2001-SLP-37, pp.43-48 (2001).
- [2] A.L.Gorin, S.E.Levinson and A.Sanker, "An Experiment in Spoken Language Acquisition", IEEE Trans on Speech and Audio Processing, Vol. 2, No.1, PART II, pp224-239, 1994
- [3] 中川聖一, 升方幹雄, "視聴覚情報の統合化に基づく概念と文法の獲得システム", 人工知能学会, Vol.10, No4, pp.619-627, 1995
- [4] 赤穂昭太郎, 速水悟, 長谷川修, 吉村隆, 麻生英樹, "EM 法を用いた複数情報源からの概念獲得", 電子情報通信学会論文誌, Vol.J80-A, pp.1546-1553, 1997
- [5] D.Roy, "Integration of Speech and vision using mutual information", In Proc. Of ICASSP, Vol.4, pp2369-2372, 2000
- [6] 金景柱, 岩橋直人, "知覚情報の統合に基づく言語音声単位の獲得アルゴリズム", 信学技報, TL200-21, pp.9-16, 2000
- [7] 黒田和史, 荒井秀一, 石川和雄, "画像・音声メディアからの概念獲得における対話を用いた学習の効率性評価", 信学技報, AI2001-45, pp.87-94, 2001
- [8] 新田恒雄, 越坂禅, 桂田浩一, "Infant Agents 間での対話による概念知識獲得", 人工知能学会全国大会, 2002 1A1-07
- [9] 石崎雅人, 伝康晴, "言語と計算 3 談話と対話", 東京大学出版会, pp.13-31, 2001