

幼児エージェントにおけるバイアスの形成と言語の構造化

篠原 修二[†] 田口 亮[†] 橋本 敬^{††}
桂田 浩一[†] 新田 恒雄[†]

近年、人間の幼児は、形状類似バイアス、相互排他性バイアスなど、学習バイアスと呼ばれる様々な制約を利用して、効率良く言葉の学習を行っていると考えられるようになってきた。バイアスが言語学習に有効に作用するためには、幼児が学習すべき言語と幼児の持つバイアスが整合的でなければならない。本研究では、経験を通して言葉を学習する幼児エージェント (Infant Agent: IA) の単純なモデルを構築し (1) IA が何らかのバイアスを持つ場合、世代交代を繰り返すことによって彼らの話す言語がどのように変化するのか (2) 教示言語に応じて、IA はどのようなバイアスを形成するのか、という 2 つの問題について分析した。本稿では、まず IA 間でコミュニケーションが成立するためには、IA が「A ならば B」から「B ならば A」を推論する傾向性 (対称性バイアス) を持つことが不可欠であることを示す。次に、対称性バイアスを実現するメカニズムを提案し、それを IA に実装する。この IA を用いて数値実験を行った結果、第 1 に、この IA は世代交代を繰り返すにつれて、自らの学習能力に適した形に言語を構造化すること、第 2 に、構造化された言語を学んでいく過程の中で、様々な学習バイアスを自律的に形成し、効率的な語彙学習を達成することが示された。この結果は、幼児の発達過程において観察される様々な学習バイアスが、対称性バイアスという単一の原理から導出できることを示唆する。

Emergence of Learning Biases and Structuring of Language Caused by Infant Agent with Symmetry Bias

SHUJI SHINOHARA,[†] RYO TAGUCHI,[†] TAKASHI HASHIMOTO,^{††}
KOUICHI KATSURADA[†] and TSUNEO NITTA[†]

Recently, it has been found that human infants acquire language effectively by utilizing learning biases such as shape bias and mutual exclusivity principle, etc.. In order that these biases benefit word learning, they need to be consistent with the language which the infants should learn. In this study, we have constructed a simple model of infant agent (IA) by implementing a vocabulary acquisition mechanism, which is derived from the assumption that an agent has consistent beliefs in communication, and have explored the following two problems: (1) How does language spoken by IA change with generation transitions? (2) Which biases does child-IA produce, depending on the language taught by parent-IA? Our simulation results show that the language evolves into the structure fitting for learning ability of IA, and IA develops various biases through learning the structured language and comes to acquire language more efficiently.

1. はじめに

近年の発達心理学の目覚ましい発展により、人間の幼児は、学習バイアスと呼ばれる様々な制約を利用して、効率良く言葉の学習を行っていると考えられるようになってきた^{6),12)}。

たとえば、ある対象を目の前にして新奇なラベルを聞く際、そのラベルが指し示すのは、対象全体なのか部分なのか、それとも対象が持つ何らかの属性なのかを決定する必要がある¹⁸⁾ が、幼児は新奇なラベルは事物全体を指し示すと考える傾向がある (事物全体バイアス)¹²⁾。

あるラベルが事物全体を指示するとしても、さらにそのラベルは他の対象にも適用可能なのか否か、つまり普通名詞なのか固有名詞なのかという問題が残る^{3),17)}。この場合、幼児は初めて聞くラベルを普通名詞と見なす傾向がある (事物カテゴリーバイアス)^{12),13)}。そしてその際、形状の類似した対象にラベルの適用範囲を

[†] 豊橋技術科学大学大学院工学研究科知識情報工学系
Department of Knowledge-based Information Engineering,
Graduate School of Engineering, Toyohashi University of Technology

^{††} 北陸先端科学技術大学院大学知識科学研究科
School of Knowledge Science, Japan Advanced Institute of Science and Technology

拡張する（形状類似バイアス^{5)~8),10),11)}。

ほかにも、幼児は各対象は1つのラベルのみを持つと考える傾向がある（相互排他性バイアス^{2),4),13)~15)}。たとえば既知の事物と新奇な事物を目の前にして新奇なラベルを聞いたとき、幼児はそのラベルを新奇な事物の名前であると見なす。このバイアスは、指差しなどが無い場合に指示対象の同定に役立つと考えられている¹⁵⁾。

幼児は、このようなバイアスを持つことで可能な仮説を限定し、効率良く言語を学習していくと考えられる。これらのバイアスは工学の分野からも注目されている。たとえば田口ら¹⁹⁾は、対話システムへの応用という観点から、学習の効率化を図るためにこれらのバイアスを組み込んだエージェントモデルを提案している。

以上に述べた語彙学習バイアスとは別に、ヒトの非論理的推論の一例として、対称性や刺激等価性の成立に関する研究が進められてきた²³⁾。その結果ヒトは、「AならばB」から「BならばA」を推論する傾向性を持つことが知られている。これは、たとえていうなら「いちごは赤くて丸い」と信じるとき「赤くて丸いものはいちごである」と信じるということである。しかし論理的にいえば、前者の言明が真であったとしても、後者の言明が真であるとは限らない。「赤くて丸いものはいちごである」という言明は、りんごやトマトのように、いちご以外に赤くて丸いものが存在すれば真ではなくなる。このように「AならばB」から「BならばA」を推論することは、論理的に間違っているにもかかわらず、年齢や能力などとは無関係に多くのヒトで容易に成立する。一方、ヒト以外の動物にとってはこのような対称性の成立は困難であることが知られており、ヒトとヒト以外の動物でなぜこのような違いを示すのかが注目されている^{9),16),26)}。

このようにヒトは、他の動物にはあまり見られない様々なバイアス（傾向性）を持つが、語彙学習バイアスの獲得と対称性の成立に関する研究は独立に行われており、現在のところ両者を縦断的に比較検討した研究はない²⁶⁾。本稿では、対称性の成立、すなわち「AならばB」から「BならばA」を推論する傾向性は、コミュニケーションの成立に重要な役割を果たすことを示したうえで、この傾向性が語彙学習バイアスの形成に与える影響について論じる。

語彙学習バイアスの形成を論じるにあたってまず着目したいのは、これらのバイアスと幼児が学習すべき教示言語との整合性である。事物カテゴリーバイアス

が言語学習に有効に作用するためには、教示言語に固有名詞よりも普通名詞が多く含まれていなければならないだろう。また形状類似バイアスは、教示言語が形状に関して構造化されていれば学習に有効であるが、他の属性、たとえば色に関して構造化された言語であれば、むしろ学習を阻害することになるだろう。つまり、バイアスが言語学習に有効に作用するためには、形成されるバイアスと教示言語が整合的でなければならない。

言語とバイアスの整合性に関して、言語がバイアスに適合したと考えるのか、それともバイアスが言語に適合したと考えるのか。どちらの問題ととらえるかによって、研究の手法は大きく異なってくる。両者の違いは、バイアスを生得的と見なすか否かという立場の違いと見ることもできる。バイアスの形成という問題に対して、これらの立場の違いに応じて、以下の3つの側面からのアプローチが可能であると思われる。

- (1) なぜ/どのようにしてバイアスを持つエージェントが進化してきたのか。
- (2) エージェントが何らかのバイアスを持つ場合、そのバイアスと整合的な言語がどのように進化してきたのか。
- (3) 教示言語に応じてバイアスがどのように形成されるのか。

(1)は生物学的な進化の問題であり、(2)は言語進化の問題である。また(3)は個体の発達過程に関する問題である。

Smith²⁴⁾は、同義語と同音異義語を回避するようなバイアス、つまりラベルと意味を1対1に対応付けようとするバイアスを持つエージェントをモデル化し、言語はエージェントが持つバイアスに適した形に進化すると述べている。上述のバイアスでいえば、同義語を回避するというバイアスは、相互排他性バイアスに相当する。彼はエージェントにこのバイアスをアプリオリに組み込んだうえで、同音異義語を回避するバイアスを持つ、同音異義語を好むバイアスを持つ、どちらでもない、という3種類のエージェントをモデル化し、同音異義語を回避するバイアスを持つエージェ

ここで、形状（色）に関して構造化された言語とは、同じ形（色）の対象は他の属性が違って同じラベルを持つが、他の属性が同じでも形（色）が異なれば異なるラベルを持つような言語である。

厳密にいえば、同義語を持たないというバイアスは、対比の原理¹⁾に相当するが、彼のモデルでは固有名詞と普通名詞のように階層関係にあるラベルを扱わないため、実質的に相互排他性バイアスと対比の原理を区別する必要はない。

ントが、進化的に安定であることも示した。

このように、Smith はバイアスを生得的なものと思わず、バイアスの形成を進化の問題としてとらえたうえで、問題(1)と(2)を扱っている。一方 Samuelson²⁰⁾ や日高ら²¹⁾ は、バイアスは学習によっても形成可能であるとの立場から、形状類似バイアスに関して(3)の問題を扱っている。

本稿では、問題(2)と(3)について論じる。両者の問題を同時に扱う場合、どのようなバイアスが生物学的進化によって生じた生得的なものであり、どのようなバイアスが発達過程における学習によって形成されるのか、という問題が生じる。

本研究では、語彙学習を行う単純な幼児エージェント (Infant Agent: IA) の数理モデルを構築し、このエージェントに「A ならば B」から「B ならば A」を推論するメカニズムを対称性バイアスとして組み込む。つまりこのバイアスを生得的と思わず。そのうえで、第1にこのエージェントにおいてどのような言語進化が見られるのか、第2にこのエージェントの発達過程においてどのような語彙学習バイアスが形成されるのか、という2つの問題について分析する。

2. モデル概要

本稿では、経験を通して言語を学習していく幼児エージェント (以下 IA と呼ぶ) をモデル化する。本章では、モデルの概要説明、および本稿で用いる用語の定義を行う。

2.1 言語

まず色属性の集合を $C = \{c_0, c_1, c_2, \dots\}$ 、形属性の集合を $F = \{f_0, f_1, f_2, \dots\}$ と表記する。モデルでは簡単のため、対象は色と形という2つの属性のみを持つこととし、色と形の対で表現するものとする。すなわち対象集合は、 $O = \{o \mid o \in C \times F\}$ と定義される。またすべての対象には1つのラベルが付与されるものとする。ここでラベル集合を $L = \{l_0, l_1, l_2, \dots\}$ と表記する。言語とは、対象とラベルの対応関係のことであるとし、 $Lang: O \rightarrow L$ と定義する。

2.2 経験

IA は、経験を通して言語を学習していくが、モデルでは学習の際に、ある対象を指差しながら、たとえ

ば「これはいちごです」と明示的に教えることはしない。IA は目の前にあるいくつかの対象を眺めていて、大人はそれらのうちのいずれかの対象について話している、という状況を想定する。つまり IA は対象集合の中からランダムに選ばれた N 個の対象を見ていて、たとえば「いちご」というラベルを耳にする。 N 個の対象の中に必ずいちごは存在するが、それらのうちどれがいちごなのかは教えられない。IA はこのような経験を繰り返すものとする。このような状況の中で、IA は「いちご」というラベルを耳にしたときに見ていたすべての対象が「いちご」と呼ばれたものと理解する。換言すれば、間違った事例も数多く経験することになる。

2.3 信念

IA は経験を通して

- 「ラベル l は対象 o である」という言明に対する確信度
- 「対象 o はラベル l である」という言明に対する確信度

という2つの確信度を形成していく。本稿ではこれらの確信度の形成過程を学習と呼ぶ。

前者の言明、たとえば「いちごは赤くて丸い」という言明は、赤や丸という属性が「いちご」と呼ばれる対象に共通の属性であるということをいい表している。一方後者の言明、たとえば「赤くて丸いものはいちごである」は赤や丸という属性がいちごに固有の属性であることをいい表している。この意味で本稿では、これらの言明に対する確信度を、各々共通度、固有度と呼ぶことにする。また共通度と固有度の組を信念と呼ぶ。

3. モデル

本章では、IA の学習過程をモデル化する。ただし本稿では、IA0, IA1, IA2 という学習能力の異なる3種類のモデルを構築する。

3.1 IA0

IA0 は、「様々な対象を眺めている状況の中であるラベルを聞く」という経験を繰り返す。IA0 はそれらの経験を通して、対象とラベルの各組に対して、 l_k というラベルを聞いたときに対象 o_i を見た頻度 $V(o_i|l_k)$ と o_i を見たときに l_k を聞いた頻度 $V(o_i|l_k)$ を記憶していく。

IA はこの記憶に基づいて、たとえば今までにいちごを見た経験が何度もあり、それらすべてが赤くて丸

言語に同音異義語が存在すれば、エージェント間の情報伝達が困難になる。たとえば1つのラベルしか存在せず、それがすべての対象を指示するような言語では、何の情報も伝えられない。また同義語は、それを学習し記憶するためにはコストがかかる一方で、言語にとっては冗長である。この観点からすれば、Smithの結果は妥当なものであると思われる。

ここで $N=1$ とすれば、明示的な教示と考えられる。

かったとすれば、「いちごは赤くて丸い」と確信するようになる。このような信念，すなわち「いちごは赤くて丸い」などの言明に対する確信度を定式化する方法として， $P((赤, 丸)| いちご)$ のような条件付き確率を利用することが考えられる。

しかし我々は，この方法には以下の2つの問題があると考える。第1は記憶の問題である。条件付き確率を用いるためには，今までにどのような属性を持ついちごを何回見たかということすべて記憶しておかなければならない。しかし我々は一般に，無限の記憶力があるわけではなく，遠い過去の経験ほど忘れやすいと考えられる。第2に，信念は経験の多寡に依存するという問題である。たとえば今までにいちごを一度しか見たことがなければ，そのいちごが赤くて丸かったとしても「いちごは赤くて丸い」との確信など持てないだろう。一方今までにいちごを見た経験が何度もあり，すべてのいちごが赤くて丸かったとすれば，かなりの確信を持って「いちごは赤くて丸い」と考えるだろう。つまり「いちごは赤くて丸い」という信念は，赤くて丸いいちごを見るという経験を繰り返すことによって徐々に形成されると考えられる。しかし確信度として条件付き確率を用いるならば，経験回数の違いが反映されない。

以上の2点を考慮し，モデルでは客観的な経験頻度をIAの内的な頻度に変換する。まず記憶の問題に関しては，IA0は遠い過去の記憶ほど忘れやすいと仮定し，記憶している頻度分布を以下の式を用いて更新する²⁵⁾。ただし V の初期値はすべて0とする。

$$V(o_i|l_k) \leftarrow V(o_i|l_k) + \alpha[r - V(o_i|l_k)] \quad (1)$$

$$r = \begin{cases} 1, & l_k \text{を聞き } o_i \text{を見たとき} \\ 0, & l_k \text{を聞き } o_i \text{を見なかったとき。} \end{cases}$$

$$V(l_k|o_i) \leftarrow V(l_k|o_i) + \alpha[r - V(l_k|o_i)] \quad (2)$$

$$r = \begin{cases} 1, & o_i \text{を見て } l_k \text{を聞いたとき} \\ 0, & o_i \text{を見て } l_k \text{を聞かなかったとき。} \end{cases}$$

この更新式は，強化学習の分野でモンテカルロ法と呼ばれるものである。ここで α ($0 \leq \alpha < 1$)は学習率と呼ばれるパラメータであり， α を大きくすればするほど最近の経験を優先する度合いが強まる。

次に，信念は経験の多寡に依存するという問題に関して述べる。本モデルでは，信念を形成された頻度を正規化することで計算する。ただし，信念は経験に依存して徐々に形成されると仮定し，ボルツマン分布

を用いて正規化する。ここで，共通度と固有度を各々 $B(o_i|l_k)$ ， $B(l_k|o_i)$ と表記することにし，

$$B(o_i|l_k) = \frac{\exp\{V(o_i|l_k)/\tau\}}{\sum_{o \in O} \exp\{V(o|l_k)/\tau\}} \quad (3)$$

$$B(l_k|o_i) = \frac{\exp\{V(l_k|o_i)/\tau\}}{\sum_{l \in L} \exp\{V(l|o_i)/\tau\}} \quad (4)$$

と定義する。ここで τ は温度と呼ばれる正定数であり，温度が高い場合にはすべての共通度（固有度）が同程度になるように設定される。一方温度が低い場合には， V の差がより増幅される。

このようにモデルでは，共通度と固有度を表すために条件付き確率を用いるのではなく，モンテカルロ法とボルツマン分布によって変換を施された B を用いる。

3.2 IA1

IA0において，頻度 V を記憶するためには， $2|O||L| = 2|C||F||L|$ のメモリ空間が必要とされる。IA1では，メモリ空間を節約するため，色と形の頻度を別々に記憶することにする。このため V について，以下のように色用 V_C と形用 V_F を用意する。このとき必要とされるメモリ空間は， $2(|C| + |F|)|L|$ になる。

$$V_C(c_i|l_k) \leftarrow V_C(c_i|l_k) + \alpha[r - V_C(c_i|l_k)] \quad (5)$$

$$r = \begin{cases} 1, & l_k \text{を聞き } c_i \text{を見たとき} \\ 0, & l_k \text{を聞き } c_i \text{を見なかったとき。} \end{cases}$$

$$V_C(l_k|c_i) \leftarrow V_C(l_k|c_i) + \alpha[r - V_C(l_k|c_i)] \quad (6)$$

$$r = \begin{cases} 1, & c_i \text{を見て } l_k \text{を聞いたとき} \\ 0, & c_i \text{を見て } l_k \text{を聞かなかったとき。} \end{cases}$$

$$V_F(f_j|l_k) \leftarrow V_F(f_j|l_k) + \alpha[r - V_F(f_j|l_k)] \quad (7)$$

$$r = \begin{cases} 1, & l_k \text{を聞き } f_j \text{を見たとき} \\ 0, & l_k \text{を聞き } f_j \text{を見なかったとき。} \end{cases}$$

$$V_F(l_k|f_j) \leftarrow V_F(l_k|f_j) + \alpha[r - V_F(l_k|f_j)] \quad (8)$$

$$r = \begin{cases} 1, & f_j \text{を見て } l_k \text{を聞いたとき} \\ 0, & f_j \text{を見て } l_k \text{を聞かなかったとき。} \end{cases}$$

なお V_C ， V_F の初期値に関しては，すべて0とする。

ここで $P((赤, 丸)| いちご)$ は，今までに「いちご」というベルを聞いた経験の中で赤くて丸い対象を見た割合を表す。

IA0では V の初期値を，またIA1では V_C ， V_F の初期値を各々0に設定している。これらの値はIAの経験頻度を反映するものであり，初期段階では何も経験していないという意味で0に設定した。ただし， V および V_C ， V_F の初期値を0以外の値に設定しても，以下に示す数値実験の結果は変わらない。これらの初期値をどんな値に設定したとしても，その値は経験に応じて修正され，経験を反映するものになる。つまり初期値の違いによる影響は，経験を積むことで速やかに消失する。

次に IA1 は、属性ごとに記憶された各頻度に対して、共通度と固有度を別々に計算する。

$$B_C(c_i|l_k) = \frac{\exp(V_C(c_i|l_k)/\tau)}{\sum_{c \in C} \exp(V_C(c|l_k)/\tau)} \quad (9)$$

$$B_C(l_k|c_i) = \frac{\exp(V_C(l_k|c_i)/\tau)}{\sum_{l \in L} \exp(V_C(l|c_i)/\tau)} \quad (10)$$

$$B_F(f_i|l_k) = \frac{\exp(V_F(f_i|l_k)/\tau)}{\sum_{f \in F} \exp(V_F(f|l_k)/\tau)} \quad (11)$$

$$B_F(l_k|f_i) = \frac{\exp(V_F(l_k|f_i)/\tau)}{\sum_{l \in L} \exp(V_F(l|f_i)/\tau)} \quad (12)$$

最後に、属性ごとに求めた値を掛け合わせることで、対象の共通度および固有度を計算する。

$$B((c_i, f_j)|l_k) = B_C(c_i|l_k)B_F(f_j|l_k) \quad (13)$$

$$B(l_k|(c_i, f_j)) = B_C(l_k|c_i)B_F(l_k|f_j) \quad (14)$$

3.3 「何？」と「どれ？」

IA2 の定義を行う前に、いくつかの質問に対する IA の回答法について述べる。はじめに、ある対象 o_i を見せられ「これ何？」と質問された場合の回答法を述べる。IA がこの質問に答えるためには、対象の入力に対し何らかのラベルを出力しなければならないが、その際に必要となるのは共通度ではない。たとえば赤くて丸い対象を示され「これは何？」と質問されて「これはいちごである」と答える際に必要とされるのは、「赤くて丸いものはいちごである」という固有度に基づく言明の正当性である。つまり「これは何？」という質問に対し IA は、提示された対象 o_i に対して最も固有度の高いラベルを選択するものとする。この質問に対する IA の解答を $What(o_i)$ と表記するとすれば、

$$What(o_i) = \arg \max_l B(l|o_i) \quad (15)$$

である。ただし、最も固有度の高いラベルが複数存在する場合は、それらの中から 1 つをランダムに選ぶ。

以上に示した質問以外に、あるラベル l_k を用いて「 l_k はどれ？」と聞かれることもあるだろう。また質問でなくても、「 l_k を取って」と依頼されたときには、どれが l_k なのかを判断する必要がある。「何？」の場合とは対照的に「どれ？」という質問の場合、回答方法としていくつかの候補が考えられる。

第 1 の候補として、式 (16) に示すような方法が考えられる。つまり、まず IA は各対象について、 $What$ を用いてそれが何かを判断する。次に、もしそれらの中で l_k と判断した対象あればそれを選ぶ。一方すべての対象が l_k ではないと判断した場合、「どれも l_k ではない (null)」と回答する。

$$Which1(l_k) = \begin{cases} o_i, & \exists o_i \in O, s.t. What(o_i) = l_k \\ null, & \forall o \in O, What(o) \neq l_k \end{cases} \quad (16)$$

ただし、 l_k と判断する対象が複数存在する場合、それらの中から 1 つをランダムに選ぶ。

$Which1$ を採用すると、「どれも l_k ではない」と判断する場合が起こりうる。この場合、 l_k は何も指示しない無意味なラベルになってしまう。またこの方法を採用すると、 l_k がどれかを判断するためだけに、すべての固有度を参照しなければならないため、計算量が多くなるという難点がある。

上記の問題を解消する方法として、 l_k に関する固有度のみを比較し、それらのうち最も値の高いものを選択するという方法が考えられる。

$$Which2(l_k) = \arg \max_o B(l_k|o) \quad (17)$$

ほかに、この質問に回答するには、ラベルを入力とし対象を出力しなければならないという点に着目するならば、共通度を使う、つまり l_k に関する共通度を比較し、それらのうち最も値の高いものを選択するという方法が考えられる。

$$Which3(l_k) = \arg \max_o B(o|l_k) \quad (18)$$

これら 2 つの方法を用いるとき、IA は任意のラベル入力に対して何らかの対象を回答として出力できる。また l_k に関する固有度あるいは共通度のみを比較すればよいと、 $Which1$ を用いる場合に比べて計算量が大幅に減少する。このように「どれ？」という質問に対する回答法として $Which2$ あるいは $Which3$ を採用すれば、 $Which1$ において生じる問題は解消される。

しかしこれらの方法を用いると、 $Which1$ を用いる場合には見られなかった新たな問題が発生する。具体的にいえば、ある対象に対して、たとえば「いちごはこれである」と一方で判断しながら、他方で「これはいちごではない」と判断する場合が起こりうる。このような信念を持つとコミュニケーションに不都合が生じる場合がある。

例を述べよう。A さんが B さんに「 l_k を取って」と依頼し、B に l_k を取ってもらった場面を想定する。ただし A と B はまったく同じ信念を持つと仮定する。ここで A が使用するラベル l_k とは、彼が取ってもらいたい対象 o_i に対して、彼が「 o_i は何？」と質問された

きに回答するラベルである．すなわち $l_k = \text{What}(o_i)$ である．

一方「 l_k を取って」と依頼された B が選択する対象 o_j とは「 l_k はどれ？」という質問をされた際に彼が返答する対象である．仮に彼らが判断に *Which3* を用いているとすれば、 $o_j = \text{Which3}(l_k)$ である．つまり彼らのコミュニケーションが成立する、すなわち $o_j = o_i$ であるためには、 $\text{Which3}(l_k) = o_i$ という関係が成立していなければならない．

ただし l_k が普通名詞の場合、A は対象 o_i そのものでなくても彼が l_k だと考える対象、すなわち $l_k = \text{What}(o_j) = \text{What}(o_i)$ を満たす対象 o_j を取ってもらえればよいだろう．このように条件を弱めるならば、A と B の間でコミュニケーションが成立するためには、 $\text{What}(\text{Which3}(l_k)) = l_k$ という関係が成立すればよい．逆にいえば、*What* と *Which* がこの関係を満たさないとき、A と B がまったく同じ信念を持っていたとしても、2 人の間でコミュニケーションが成立しない場合が生じる．この関係式を満足するということは、「 l_k はどれ？」と聞かれて o_j と回答するならば、「 o_j は何？」と聞かれた場合に l_k と回答しなければならないということの意味する．つまり「 l_k は o_j である」と判断するならば、「 o_j は l_k である」と判断しなければならない．しかし、*Which2* あるいは *Which3* と *What* の間でこの関係が満たされるとは限らない．

一方 *Which1* の場合は、判断に *What* を用いるため、 $\text{What}(\text{Which1}(l_k)) = l_k$ という関係は必ず成立する．したがって、コミュニケーションの場面において、以上に述べたような不都合が生じることはない．

以上に「どれ？」に対する 3 つの回答法を示したが、どの方法にも各々問題がある．我々人間が、3 つの回答法（あるいは他の方法もあるかもしれないが）のうちどの方法を採用しているのかを詮議することも重要ではあるが、どの方法を採用したとしても重大な問題が残る．

どの方法を使っても結果が一致するような信念を持つことができなだろうか？ もし IA がそのような信念を持つならば、たとえば *Which3* を使うことにより、*Which1* を採用した場合に生じる問題は解消される．また、*Which3* の出力結果が *Which1* と同じであるならば、*What* との間に齟齬をきたすこともない．次の IA2 では、どの方法を採用してもなるべく結果が一致するように信念を修正する傾向性を与える．

3.4 IA2

本節では IA2 の定義を行う．ただし本節では、IA1

からの変更点についてのみ述べる．3.3 節で述べたように、IA2 では「どれ？」に対する 3 つの回答法の出力結果がなるべく一致するように信念を修正する傾向性を与える．

まず *Which2* と *Which3* の結果が一致するためには、 β をある正定数として以下の関係が成立すればよい．

$$\forall o_i \in O, \forall l_k \in L, B(o_i|l_k) = \beta B(l_k|o_i) \quad (19)$$

次に、*Which2* あるいは *Which3* と *Which1* の結果に不都合が生じないためには、IA2 の信念は以下の条件を満たす必要がある．

$$\forall l_k \in L', \text{What}(\text{Which2}(l_k)) = l_k \quad (20)$$

$$\forall l_k \in L', \text{What}(\text{Which3}(l_k)) = l_k \quad (21)$$

ただし、 $L' = \{l | l \in L, o \in O, \text{What}(o) = l\}$ である．

モデルでは、これらの条件式をなるべく満足するように信念を修正する傾向性を IA2 に与える．ただし信念が条件式 (19) を満たすとき、*Which2* と *Which3* の出力結果は一致する．このため実際には、式 (20) か式 (21) どちらか一方と、式 (19) を満たすように信念を修正すればよい．以下では、条件式 (19) と (21) をなるべく満足させるように信念を修正する方法について述べる．

まず、信念が条件式 (19) を満足するということは、共通度 $B(o_i|l_k)$ と固有度 $B(l_k|o_i)$ が比例関係にある、つまり一方が高いなら他方も高いという関係にあることを意味する．次に条件式 (21) を満足するということは、IA2 は「 l_k はどれ？」と聞かれて o_i と答えるならば、「 o_i は何？」と聞かれた場合に l_k と答えなければならないということの意味する．これは、ある l_k に関して「 l_k は o_i である」という共通度 $B(o_i|l_k)$ が最も高いならば、「 o_i は l_k である」という固有度 $B(l_k|o_i)$ も最も高いということである．

モデルでは、以上のような信念を持つ傾向性を IA2 に与える．ただし IA2 は IA1 と同様に、色属性と形属性を独立に処理するため、属性ごとにこのような傾向性を与える．具体的にいえば、色、形各々の属性に関して共通度（固有度）が高いならば、固有度（共通度）も高いという関係を満足させるように信念を修正する．対象の共通度（固有度）は、色属性の共通度（固有度）と形属性の共通度（固有度）を掛け合わせたものである．このため色、形各々の属性について、固有度と共通度の間以上に以上のような関係が成立するならば、対象に関して必然的に共通度（固有度）が高いならば固有度（共通度）が高いという関係が成立する．本稿では、このような傾向性を、対称性バイアスと呼ぶことにする．

対称性バイアスを実装するための 1 つの方法として、モデルでは以下に示すように属性ごとに、共通度（固有度）を固有度（共通度）によって相互規定することにし、 $\forall c_i \in C, \forall f_j \in F, \forall l_k \in L$ に対して

$$B_C(l_k|c_i) = \frac{B_C(c_i|l_k)}{\sum_{l \in L} B_C(c_i|l)} \quad (22)$$

$$B_C(c_i|l_k) = \frac{B_C(l_k|c_i)}{\sum_{c \in C} B_C(l_k|c)} \quad (23)$$

$$B_F(l_k|f_j) = \frac{B_F(f_j|l_k)}{\sum_{l \in L} B_F(f_j|l)} \quad (24)$$

$$B_F(f_j|l_k) = \frac{B_F(l_k|f_j)}{\sum_{f \in F} B_F(l_k|f)} \quad (25)$$

という条件式を満足するように信念を修正する。各属性に関する IA2 の信念がこれらの条件式を満足するとき、対象の共通度 $B((c_i, f_j)|l_k)$ と固有度 $B(l_k|(c_i, f_j))$ の間には以下の関係が成立する。

$$B((c_i, f_j)|l_k) = \frac{|L|^2}{|O|} B(l_k|(c_i, f_j)) \quad (26)$$

つまり、対象に関して固有度と共通度は比例関係にあり、共通度（固有度）が高いならば固有度（共通度）が高いという関係、すなわち条件式 (19) が成立する。

IA2 は、IA1 と同様に算出された固有度と共通度を、条件式 (22), (23), (24), (25) を満足するように修正する。修正方法には様々なものが考えられるが、本稿では以下の方法を採用する。条件式を以下のような更新式に書き換える。

$$B_C(l_k|c_i) \leftarrow \frac{B_C(c_i|l_k)}{\sum_{l \in L} B_C(c_i|l)} \quad (27)$$

$$B_C(c_i|l_k) \leftarrow \frac{B_C(l_k|c_i)}{\sum_{c \in C} B_C(l_k|c)} \quad (28)$$

$$B_F(l_k|f_j) \leftarrow \frac{B_F(f_j|l_k)}{\sum_{l \in L} B_F(f_j|l)} \quad (29)$$

$$B_F(f_j|l_k) \leftarrow \frac{B_F(l_k|f_j)}{\sum_{f \in F} B_F(l_k|f)} \quad (30)$$

これらの更新式を見れば分かるように、共通度と固有度は互いに相互規定し合う関係にある。モデルでは、 $\forall c_i \in C, \forall f_j \in F, \forall l_k \in L$ に対してこれらの更新式を交互に繰り返し適用することで値を収束させる。なお対象の固有度（共通度）は、IA1 の場合と同様に

IA2 の信念がこれらの条件式を満足するとき、後述のように条件式 (19) は満たされる。しかし厳密に言えば条件式 (21) が必ず満たされるわけではない。モデルでは IA2 に条件式 (21) を満足するような傾向性を与えている。詳細は付録を参照されたい。

以下に示す数値実験では、繰り返し回数を 10 に設定した。この操作を 10 回繰り返せば、固有度および共通度の誤差は $10E-4$ 程度以下に収束する。

V_C, V_F を 0 に初期化する
 経験を積むたびに以下の処理を繰り返す：
 1. モンテカルロ法で V_C, V_F を更新
 2. ボルツマン分布で V_C, V_F を正規化 (B_C, B_F と表記)
 3. 対称性バイアスを適用し B_C, B_F を修正
 4. B_C と B_F を掛け合わせて B を算出

図 1 IA2 の学習手順

Fig. 1 Learning protocol for IA2.

式 (13), (14) を用いて計算する。

図 1 に IA2 の学習手順をまとめる。IA2 と IA1 の違いは、図中の処理 3 を行うか否かという点のみにある。また IA1 と IA0 の違いは、 V_C, V_F および B_C, B_F を用いて色と形の処理を別々に行うか、 V および B を用いて対象に対して処理を行うかという点にある。つまり IA0 では V を正規化することで直接 B を導くため、処理 4 を行う必要はない。

4. 数値実験

本章では、IA0, IA1, IA2 を用いて、言語の構造化およびバイアスの形成に関する数値実験を行う。パラメータの値は、各々 $\alpha = 0.01, \tau = 0.01$ に設定した。各属性の数とラベル数に関しては、色よりも形の属性数の方が多い ($|C| < |F|$) と仮定し、 $|C| = 16, |F| = 22, |O| = |C| \times |F| = 352, |L| = 22$ に設定した。

また IA が一度の経験において見る対象数は 5 に固定した。IA は、毎ステップ、対象集合の中からランダムに選ばれた 5 個の対象を見つつ何らかのラベルを聞くという経験を繰り返す。2.2 節でも述べたように、ここで IA が聞くラベルは、5 個の対象のうちのいずれかの名前であり、それは教示者である大人によって与えられる。IA はこの経験に基づいて信念（固有度と共通度）を形成していく。

4.1 言語の構造化

本節では主に言語の構造化に関する実験を行う。

4.1.1 実験 1

はじめに初期言語 $Lang1: O \rightarrow L$ をランダムに構成し、それを IA に対する教示言語とする。IA は、一定期間の学習を経て大人になると仮定し、大人になった IA が話す言語を次世代の IA に教示する言語とする。このような枠組みは、繰返し学習モデル (Iterated Learning Model: ILM) と呼ばれる²⁴⁾。ここで N 世代目の IA が教示される言語、換言すれば $N-1$ 世代目の IA が一定期間の学習後に話す言語を $LangN$ と表記することにする。

まず IA2 の実験結果を示す。この実験では IA が大

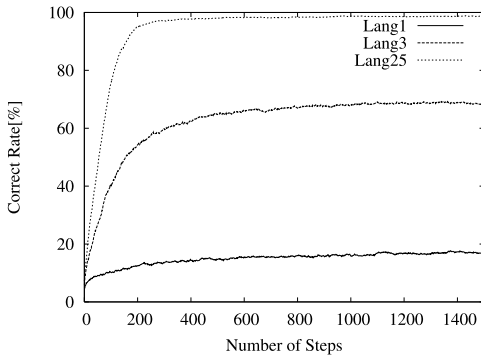


図 2 各世代の IA2 の正解率

Fig. 2 Correct rate of language learning by each generation of IA2.

人になるまでの期間を 1500 ターンに設定した．世代に関しては，25 世代まで実験を行った．図 2 に各世代の IA2 の正解率の時間発展を示す．ただし図では見やすさのため，1，3，25 世代目の IA，すなわち *Lang1*，*Lang3*，*Lang25* を教示された IA の結果のみを表示する．ここで正解率とは，ステップごとにすべての対象について各々「これは何？」と IA に質問し，その回答が教示言語と同じならば正解，異なるならば間違いとして正解の割合を表したものである．なおこの結果はランダムシードを変えて 20 回試行した結果を平均したものである．これは，特に断らない限り以下に示す実験結果においても同様である．

図から分かるように，世代を経るごとに最終的な正解率は高くなる．ランダムな初期言語 *Lang1* を与えられた 1 世代目の IA2 は，学習を繰り返しても 17% 程度の正解率しか達成できない．しかし世代を経るごとに正解率は上昇し，25 世代目での正解率はほぼ 100% である．

さて，1 世代目の IA2 は 17% 程度の正解率しか達成できない．正解率が 100% でないということは，換言すれば，IA2 は前の世代とは異なる言語を話しているということである．つまり世代を経るごとに言語は変化し，IA2 はそれにともなって効率良く高い正解率を達成するようになるといえる．言語がどのように変化しているかを見るために，図 3，図 4，図 5 にある試行における *Lang1*，*Lang3*，*Lang25* を示す．ここで X 軸， Y 軸は各々色属性，形属性を表す．また Z

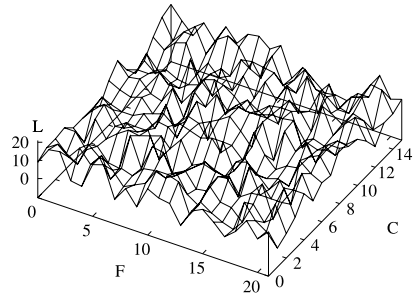


図 3 *Lang1*
Fig. 3 *Lang1*.

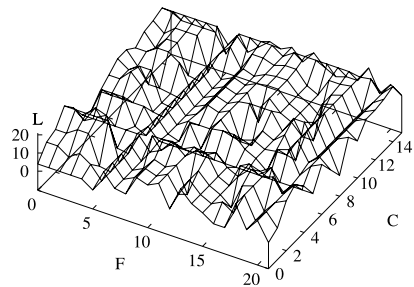


図 4 *Lang3*
Fig. 4 *Lang3*.

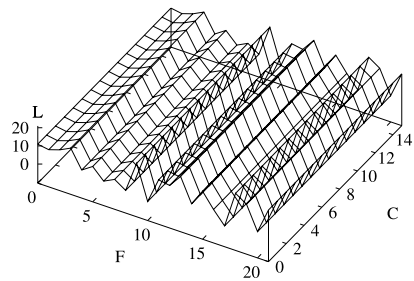


図 5 *Lang25*
Fig. 5 *Lang25*.

軸は，そのような属性を持つ対象に付されたラベルを表す．

これらの図から，世代を経るごとに言語が構造化されていくのが分かる．*Lang25* では，同じ形をした対象は色が異なっても同じラベルを付けられている．つまりこの言語は，形状に関して構造化されているといえる．ここで形状に関して構造化されるのは，色よりも形の属性数の方が多いという理由による．結果は省略するが，色の属性数の方を多くすれば，言語は色に関して構造化される．

ここで各世代の IA2 の話す言語が，色あるいは形に関してどれだけ構造化されているかを定量化するための尺度として，色とラベルの相互情報量 $I(C; L)$ およ

この実験では，学習期間を 1500 ステップに設定したが，どの世代においても 1500 ステップまでに正解率は飽和しており，これ以上学習期間を延長しても正解率はほとんど改善されない．また世代に関しても，25 世代目で数値実験を打ち切っているが，最終的な正解率は 12 世代目あたりでほぼ 100% に達し，これ以上世代交代を繰り返しても正解率はほとんど変わらない．

び形とラベルの相互情報量 $I(F; L)$ を導入する .

$$I(C; L) = H(C) + H(L) - H(C, L)$$

$$I(F; L) = H(F) + H(L) - H(F, L)$$

ただし $H(C)$, $H(F)$, $H(L)$ は各々, 色と形とラベルに関するエントロピーであり, 以下のように定義される .

$$H(C) = - \sum_{c \in C} P(c) \text{Log}_2 P(c)$$

$$H(F) = - \sum_{f \in F} P(f) \text{Log}_2 P(f)$$

$$H(L) = - \sum_{l \in L} P(l) \text{Log}_2 P(l)$$

ここで $P(c)$, $P(f)$, $P(l)$ は各々, すべての対象のうちで c という色属性を持つ対象の割合, すべての対象のうちで f という形属性を持つ対象の割合, すべての対象のうちでラベル l を付与された対象の割合である . また

$$H(C, L) = - \sum_{c \in C} \sum_{l \in L} P(c, l) \text{Log}_2 P(c, l)$$

$$H(F, L) = - \sum_{f \in F} \sum_{l \in L} P(f, l) \text{Log}_2 P(f, l)$$

であり, $P(c, l)$, $P(f, l)$ は各々, すべての対象のうちで c という色属性を持ちかつラベル l を付与された対象の割合, すべての対象のうちで f という形属性を持ち, かつラベル l を付与された対象の割合である . 言語が形に関して構造化されていけばいほど, $I(F; L)$ の値は大きくなる一方で, $I(C; L)$ の値は小さくなる . 図 6 に $I(C; L)$ と $I(F; L)$ を示す . この図から, 世代を経るごとに IA2 の話す言語が形に関して構造化されていくことが分かる .

IA1 においては世代を経るにつれて, すべての対象に対して同一のラベルを付与する言語, すなわちあるラベルがすべての対象の同音異義語になるような何の情報も持たない言語に変化する .

図 7 に IA1 と IA2 の各世代における言語が持つ情報量 $H(L)$ を示す . IA1 において $H(L)$ は, 6 世代目でほぼ 0 に収束する . これは, 世代を経ることで, すべての対象に対して同一のラベルを付与する言語に変化したことを示している . 一方 IA2 において $H(L)$ は, 世代を経ても変わらない . これは, 各ラベルが対象にほぼ均等に割り当てられていることを示している . このように IA2 は, 言語が持つ情報量を保持したまま, 言語を構造化させていく .

図 8 に 1 世代目の IA0 の正解率を示す . IA0 は IA2 とは異なり, ランダムな初期言語 $Lang1$ を完全に学

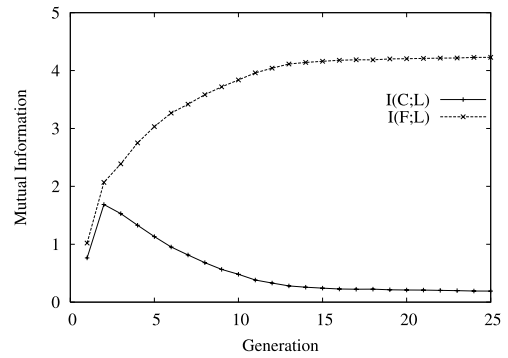


図 6 IA2 の各世代における言語の相互情報量

Fig. 6 Mutual Information of the language used by each generation of IA2.

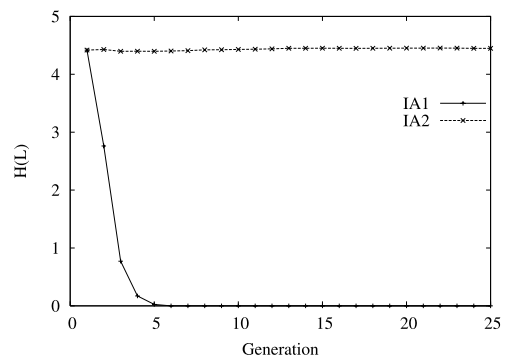


図 7 IA1 と IA2 の各世代における言語の情報量

Fig. 7 Entropy of the language used by each generation of IA1 (IA2).

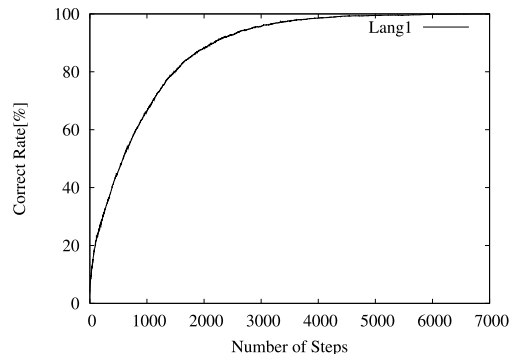


図 8 1 世代目の IA0 の正解率

Fig. 8 Correct rate of language learning by first generation of IA0.

習することができる . ただし正解率 100% を達成するためには, ほぼ 6000 ステップという非常に長期間の学習を要する . このように IA0 では完全な学習が可能であるため, 世代間で話す言語は変わらない . つまり世代を経ても $Lang1$ を使用し続けることになる . このため 2 世代以降の結果は省略した .

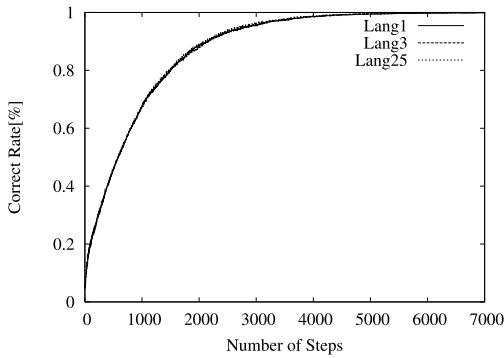


図 9 各世代の IA0 の正解率

Fig. 9 Correct rate of language learning by each generation of IA0.

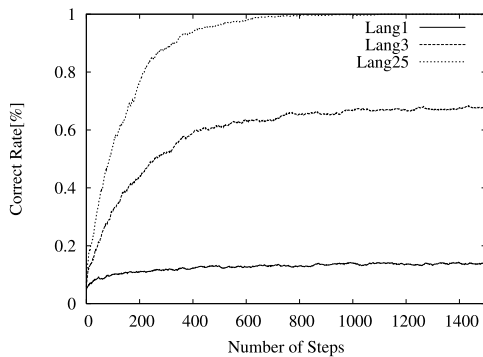


図 10 各世代の IA1 の正解率

Fig. 10 Correct rate of language learning by each generation of IA1.

4.1.2 実験 2

実験 1 で示したように，IA0 では世代を経ても言語はまったく変化しない．また IA1 では，世代を経ることで無意味な言語になってしまう．実験 2 では，IA0，IA1 と IA2 の学習能力を同一条件で比較するために，前項の図 3，図 4，図 5 で示した言語 *Lang1*，*Lang3*，*Lang25* を IA0 と IA1 にも強制的に与え，その学習能力を分析する．

図 9 に *Lang1*，*Lang3*，*Lang25* を与えた際の，IA0 の正解率の時間発展を示す．図から分かるように，どの言語を与えても，最終的な正解率および正解率が 100% に達するまでの学習期間は変わらない．

一方図 10 と図 11 は各々，IA1 と IA2 の正解率の時間発展を示す．IA1 と IA2 はともに，より構造化された言語を与えられるほど，正解率は上昇する．ただし，IA1 と IA2 を比較した場合，最終的な正解率はほとんど同じであるが，正解率が飽和するまでの時間は IA2 の方が短い．特に構造化された言語 *Lang25* に関してその傾向が顕著である．

以上に示した結果から，形に関して構造化された言

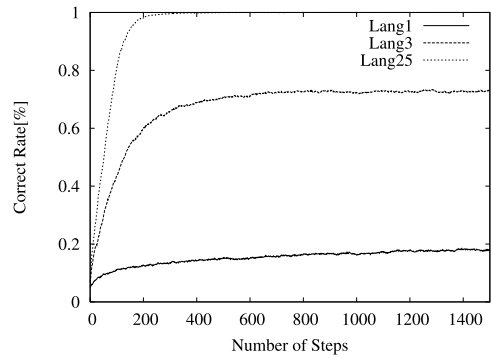


図 11 各世代の IA2 の正解率

Fig. 11 Correct rate of language learning by each generation of IA2.

語 *Lang25* を与えた場合，どの IA も完全な学習を達成できることが分かる．ただし，以下の 2 点に注意されたい．第 1 に学習を完了するまでに要する時間が大きく異なる．IA0，1，2 が 100% の正解率を達成するまでに要する時間は大まかに，各々 6000，900，300 ステップである．つまり IA2 の学習が圧倒的に速い．第 2 に，各 IA は *Lang25* を完全に学習する能力を持つが，この言語を作り上げることができるのは IA2 のみである．つまり IA0 や IA1 が，実際にこのような言語を持つことはない．

4.1.3 考察：言語の構造化

実験 1 で示したように，IA0 では世代を経ても言語はまったく進化しない．また IA1 では，言語はある 1 つのラベルがすべての対象を指示する同音異義語になってしまう．つまり IA2 のみに有意義な言語進化が見られる．本項ではこれらの理由について述べる．

本稿の実験設定において，言語の学習とは，対象とラベルの対応関係，つまり各対象に対してどのようなラベルが割り当てられるかということを学習することである．IA0 は， $B(l|o)$ によって，対象ごとにその対象を見たときに，各ラベルをどの程度の割合で聞いたのかを記憶していくことができる．また，教示言語 $Lang1: O \rightarrow L$ において，各対象に割り当てられるラベルは一意に定められており，各対象を見たときには，その対象に対応する特定のラベルを聞く頻度が高い．このため，十分多くの経験を積みめば，IA0 は各対象について，どのようなラベルが割り当てられるのかを判定できるようになる．このように IA0 は *Lang1* の完全な学習が達成できるため，話される言語は世代を経ても変わらず，言語は進化しない．

一方 IA1 や IA2 は，IA0 のように対象とラベルの対応関係を記憶する能力を持たない．これらの IA は，メモリ空間節約のため，対象とラベルの対応関係を，

表 1 教示言語の具体例

Table 1 An example of teaching language.

	c_0	c_1
f_0	l_0	l_1
f_1	l_1	l_0

表 2 $V(l|(c, f))$ の具体例: IA0Table 2 An example of $V(l|(c, f))$: IA0.

	(c_0, f_0)	(c_0, f_1)	(c_1, f_0)	(c_1, f_1)	合計
l_0	14	0	0	16	30
l_1	0	12	8	0	20
合計	14	12	8	16	

色とラベル, および形とラベルの対応関係に分解し, $B_C(l|c)$ と $B_F(l|f)$ によって記憶する. そして対象の固有度 $B(l|(c, f))$ を算出する際には, それらの値を掛け合わせるによって統合する. この分解と統合という操作によって, 対象とラベルの対応関係に関する多くの情報が失われる. このため *Lang1* のような無秩序な言語を完全に学習することはできない.

以上の状況を, 厳密さを犠牲にして簡単な具体例を用いて説明しよう. まず色属性と形属性は各々2種類しかないとし, それらを c_0, c_1 および f_0, f_1 と表記する. またラベルも2種類しかないとし, l_0, l_1 と表記する. 次に教示言語は, 表1のようであったとしよう.

さて, IA0 がこれまでの経験で各対象を見たときに各ラベルを聞いた頻度は表2のようであったとしよう. 以下では説明を簡単にするため, IA0 が内的に持つ頻度分布 $V(l|(c, f))$ は, 表に示した経験頻度そのものであるとする. このとき, たとえば対象 (c_0, f_0) に対して, それを見たときに l_0 を聞いた頻度と l_1 を聞いた頻度を比較すれば, l_0 を聞いた頻度の方が高い ($V(l_0|(c_0, f_0)) = 14 > V(l_1|(c_0, f_0)) = 0$). このため IA0 は, 「対象 (c_0, f_0) は何?」と質問されれば l_0 と回答する. 同様に, 各対象に対して V の高い方のラベルを選ぶことで, IA0 はすべての対象に対して正しいラベルを回答することができる.

一方 IA1 は表2のようなメモリ空間を持たず, 対象とラベルの対応に関する経験頻度を, 色とラベル, 形とラベルに分解して記憶する. このため IA0 とまったく同じ経験をしたとしても, IA1 が持つ頻度分布は表3, 表4のように記述されることになる. このとき,

実際に IA0 が回答の際に用いるのは, 頻度 V ではなく, 固有度 B である. ただし IA0 と IA1 において B は, V をボルツマン分布を用いて正規化したものであり, V を B に変換しても大小関係は保存される. このため, ここでは簡単のために, V を用いて議論を行う.

表 3 $V_C(l|c)$ の具体例: IA1Table 3 An example of $V_C(l|c)$: IA1.

	c_0	c_1	合計
l_0	14	16	30
l_1	12	8	20
合計	26	24	

表 4 $V_F(l|f)$ の具体例: IA1Table 4 An example of $V_F(l|f)$: IA1.

	f_0	f_1	合計
l_0	14	16	30
l_1	8	12	20
合計	22	28	

たとえば対象 (c_0, f_0) に関して, c_0 を見たときに聞いたラベルの頻度は l_0 の方が高く ($V_C(l_0|c_0) = 14 > V_C(l_1|c_0) = 12$), f_0 を見たときに聞いたラベルの頻度も l_0 の方が高い ($V_F(l_0|f_0) = 14 > V_F(l_1|f_0) = 8$). このため IA1 は, 「対象 (c_0, f_0) は何?」と質問されれば l_0 と回答する. また対象 (c_1, f_1) に関して, c_1 を見たときに聞いたラベルの頻度は l_0 の方が高く ($V_C(l_0|c_1) = 16 > V_C(l_1|c_1) = 8$), f_1 を見たときに聞いたラベルの頻度も l_0 の方が高い ($V_F(l_0|f_1) = 16 > V_F(l_1|f_1) = 12$). このため IA1 は, 「対象 (c_1, f_1) は何?」と質問されれば l_0 と回答する. 同様の判定を行えば, IA1 はすべての対象に対して l_0 と回答することになる.

上述の具体例では, l_0 を聞いた頻度は合計30回, l_1 を聞いた頻度は合計20回であった. このように経験頻度に偏りが見られる場合, IA1 は「何?」という質問に対して経験頻度の高いラベルを回答する傾向が強くなる. このような経験頻度の偏りは, 第1に IA は無限の経験を積めるわけではない, 第2に IA は無限の記憶力を持つわけではなくより最近の経験ほど優先して記憶する, という2つの理由により必ず生じる.

図12に, 実験1のある試行において, 1世代目の最終ステップ(1500ステップ目)で IA0 が各ラベルに関して持っていた内的頻度分布 $V_C(l), V_F(l)$ を示す. ただし $V_C(l) = \sum_{c \in C} V_C(l|c), V_F(l) = \sum_{f \in F} V_F(l|f)$ と定義する. 図から分かるように, IA1 はすべてのラベルを等頻度で聞くのではなく, ラベルごとの経験頻度に偏りが見られる. この試行においては, ラベル9, 18, 19, 20を聞いた経験頻度が高い. 図13に, 各対象に対して「これは何?」と質問された場合に, IA1 が回答したラベルの頻度を示す. 図から, 経験頻度の高いラベルを回答する割合が高いことが分かる. IA1 のこの回答が, 次世代の IA1 に対する教示言語 *Lang2* となる. このため, 2世代目の IA1

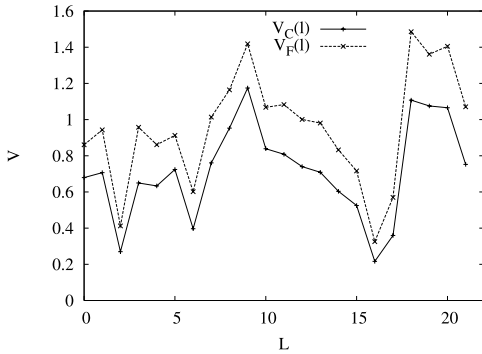


図 12 各ラベルに対する内的頻度分布 $V_C(l)$ と $V_F(l)$: IA1
Fig. 12 $V_C(l)$ and $V_F(l)$ for each label: IA1.

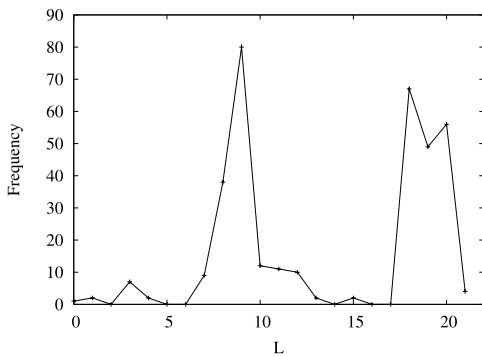


図 13 IA1 が回答したラベルの頻度
Fig. 13 Frequency of labels responded by IA1.

が各ラベルを聞く頻度の格差は、さらに広がる。このような過程を繰り返すことで、最終的に IA1 はすべての対象に対して同じラベルを割り当てようになる。このように、経験や記憶の有限性が、IA1 に同音異義語を好む傾向性を与える。

一方 IA2 は、頻度の記憶に関しては IA1 と同様であるが、これを固有度に変換する際に対称性バイアスを適用する。このため、頻度を固有度に変換した際に、IA0 や IA1 のように大小関係が保存されとは限らない。対称性バイアスを持つということは、たとえば色とラベルの対応関係に関して「 c は l である」と信じるならば、「 l は c である」と信じる傾向性を持つということである。ここで「 l は c である」という言明は、その対偶「 c でないものは l ではない」と論理的に同値である。つまり IA2 は「 c は l である」と信じるならば「 c でないものは l ではない」と信じる傾向性を持つ。上述の具体例でいえば、色属性とラベルは各々 2 種類しかない。このため、IA2 は「 c_i は l_k である」と信じるならば、「 $c_{i'}$ ($i' \neq i$) は $l_{k'}$ ($k' \neq k$) である」と信じることになる。この点は、形とラベルの対

表 5 $B_C(l|c)$ の具体例: IA2
Table 5 An example of $B_C(l|c)$: IA2.

	c_0	c_1
l_0	0.061	0.963
l_1	0.939	0.037

表 6 $B_F(l|f)$ の具体例: IA2
Table 6 An example of $B_F(l|f)$: IA2.

	f_0	f_1
l_0	0.731	0.269
l_1	0.269	0.731

応関係に關しても同様であり、「 f_j は l_k である」と信じるならば、「 $f_{j'}$ ($j' \neq j$) は $l_{k'}$ ($k' \neq k$) である」と信じることになる (ただし $i, i', j, j', k, k' \in \{0, 1\}$)。このため IA2 は、対象 (c_i, f_j) と $(c_{i'}, f_{j'})$ に対して、各々異なるラベル l_k と $l_{k'}$ を割り当てることになる。

実際に、表 3、表 4 で表される頻度に、対称性バイアスをかけて変換した固有度 $B_C(l|c)$ 、 $B_F(l|f)$ を、表 5、表 6 に示す。これらの表から、各色 (形) に対して最も固有度の高いラベルは、色 (形) ごとに異なっていることが分かる。

対象 (c_0, f_1) に関して、 c_0 に対する固有度は l_1 の方が高く ($B_C(l_0|c_0) = 0.061 < B_C(l_1|c_0) = 0.939$)、 f_1 に対する固有度も l_1 の方が高い ($B_F(l_0|f_1) = 0.269 < B_F(l_1|f_1) = 0.731$)。このため IA2 は「対象 (c_0, f_1) は何?」と質問されれば l_1 と回答する。一方対象 (c_1, f_0) に関しては、 c_1 に対する固有度は l_0 の方が高く ($B_C(l_0|c_1) = 0.963 > B_C(l_1|c_1) = 0.037$)、 f_0 に対する固有度も l_0 の方が高い ($B_F(l_0|f_0) = 0.731 > B_F(l_1|f_0) = 0.269$)。このため IA2 は、「対象 (c_1, f_0) は何?」と質問されれば l_0 と回答する。

このように対称性バイアスは、各色 (形) に割り当てられるラベルを色 (形) ごとに分散させようとする傾向性を生む。その結果、対象に割り当てられるラベルは対象ごとに分散する。すなわち、IA2 は同音異義語を回避しようとする傾向性を持つことになる。このため、IA2 では、IA1 のようにすべての対象に対して単一のラベルが割り当てられる言語に収束することはない。

最後に、言語が形状に關して構造化される理由について述べる。本稿で行った数値実験では、色よりも形の属性数の方が多いと仮定されている。換言すれば、本稿では、IA は色よりも形に対する識別能力の方が高い、と仮定している。ここで極端な例として、色は c_0, c_1 の 2 種類しかないが、形状に關しては無数の形 f_j ($j = 0 \sim \infty$) が存在する世界を考えよう。この世界では、各形状を持つ対象は各々 2 種類だけしか存

在しないのに対し、各色を持つ対象は各々無数に存在することになる。たとえば f_0 という形状を持つ対象は (c_0, f_0) と (c_1, f_0) の 2 種類である。一方 c_0 という色を持つ対象は、 $(c_0, f_0), (c_0, f_1), (c_0, f_2), \dots$ のように無数に存在する。

さて、教示言語 $Lang1$ において、各対象にはラベルがランダムに割り当てられている。このため IA2 は、 c_0 という色を持つ対象を見たときに、様々なラベルをランダムに聞くことになる。つまり、各色に対して特定のラベルの固有度が特に高くなるようなことはない。一方、 f_0 という形状を持つ対象を見たときには、 (c_0, f_0) と (c_1, f_0) に割り当てられた 2 種類のラベルを聞く頻度が非常に高い。このため、 (c_0, f_0) と (c_1, f_0) に割り当てられたラベルを各々 l_0, l_1 と表記することにすれば、 $B_F(l_0|f_0)$ と $B_F(l_1|f_0)$ がその他のラベルの固有度 $B_F(l_k|f_0)$ ($k \neq 0, 1$) に比べて突出して高くなる。このように形状に関しては、各形に対して特定のラベルの固有度が高くなる。このため、IA2 は色よりも形に強く依拠し、形に関する固有度に基づいて、対象が何であるかの判断を行うようになる。その結果として、形状が同じ対象には同じラベルを割り当てようとする傾向性が生まれる。ただしこの傾向性は、色よりも形の属性数の方が多いという仮定から導かれるものであり、IA2 だけではなく IA1 にも同様に生じる。

さて、形状に関して構造化された言語とは (1) 同じ形の対象は他の属性が違っても同じラベルを持つ、(2) 他の属性が同じでも形が異なれば異なるラベルを持つ、という 2 つの条件を満足する言語であるが、上述の傾向性は (1) を満足する言語を導く。実験 1 で示したように、IA1 はすべての対象に対して同一のラベルを割り当てる言語を生成する。この言語は (1) を満足するが (2) を満たさない。一方 IA2 は対称性バイアスを持つことで、各形に対して最も固有度の高いラベルを形ごとに分散させようとする傾向性を持つ。これが (2) を満足する言語を生成する要因となる。このため IA2 のみに (1) と (2) をともに満足する言語、すなわち形状に関して構造化された言語の生成が見られる。

4.2 学習バイアスの形成

本節では、学習バイアスの形成に焦点を当てて数値実験を行う。なお以下の実験で IA に与える言語は、前述の $Lang25$ とする。

4.2.1 実験 3：事物カテゴリーバイアス

この実験では、IA が見ることができる対象を一部に限定する。つまり IA はいくら経験を繰り返しても、

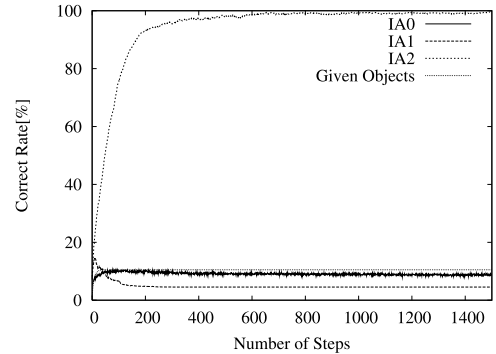


図 14 各 IA の汎化能力の差異

Fig. 14 Difference of generalization ability among IAs.

見たことのない未知の対象が数多く存在する。具体的には c_0 あるいは f_0 の属性を持つ対象のみを経験できるように設定した。このとき、すべての対象のうちで経験できる対象の割合は約 10% である。

図 14 に各ステップにおける各 IA の正解率、および IA がそのステップまでに見た対象の割合の時間経過を示す。見たことのある対象の割合 (図では Given Objects と表示) は、時間経過とともに上昇するが、実験の設定上すぐさま約 10% で飽和する。IA0 の正解率は、最終的に 10% を少し下回る程度で収束する。すなわち IA0 は、経験した対象についてはほぼ確実に学習できるが、未知の対象についてはまったく対処できない。IA1 の正解率は、最終的に 5% 程度になる。すなわち経験した対象についてもすべて学習できるわけではない。一方 IA2 は、未知の対象が数多くあるにもかかわらず、ほぼ 100% の正解率を達成した。このように IA2 のみがきわめて高い汎化能力を持つ。この結果は、IA2 が与えられたラベルを普通名詞と見なし、未学習の対象にまでラベルの適用範囲を拡張するという、つまり事物カテゴリーバイアスを持つことを示している。

4.2.2 実験 4：形状類似バイアス

この実験では、各ステップの学習後、IA が今までに見たことのないような新奇な色と形を持つ対象 O_{novel} を提示し、それに対して初めて聞く新奇なラベル l_{novel} を付与する。この教示を一度だけ行った後、 O_{novel} と色は同じだが形は異なるような対象群 O_C と、形は O_{novel} と同じだが色は異なるような対象群 O_F に含まれる各対象について、「これは何？」と質問する。

図 15, 図 16, 図 17 は各 IA が、 O_C と O_F 各々の対象群のうちで、 l_{novel} と判断した対象の割合を示す。IA0 は O_C と O_F のどちらにも含まれる対象に対しても、 l_{novel} と判断する割合は、10% 程度と非常に低い。

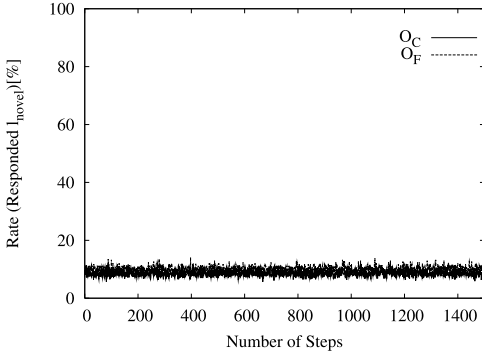


図 15 形状類似バイアス：IA0
Fig. 15 Shape bias: IA0.

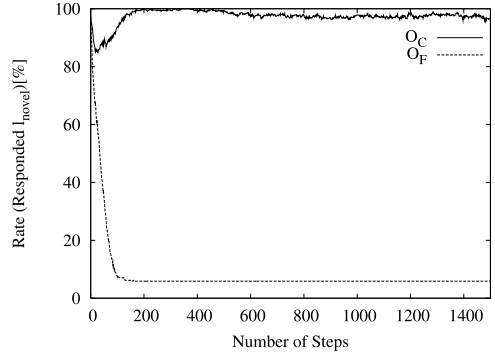


図 18 色類似バイアス：IA2
Fig. 18 Color bias: IA2.

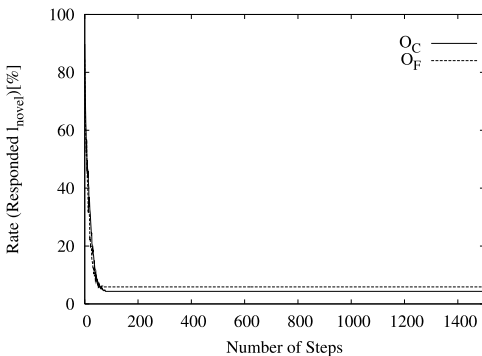


図 16 形状類似バイアス：IA1
Fig. 16 Shape bias: IA1.

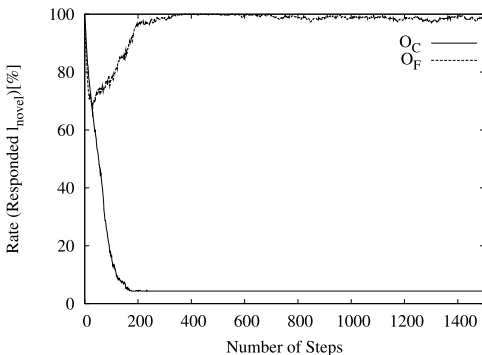


図 17 形状類似バイアス：IA2
Fig. 17 Shape bias: IA2.

ことはほとんどない。

IA2は、何も学習していない初期の段階においては、 O_C と O_F どちらにも含まれる対象も l_{novel} と判断する。つまり o_{novel} と形か色のどちらかが同じであれば、 l_{novel} と判断する。しかし学習が進行するにつれて、 l_{novel} と判断する対象の割合は減少していく。 O_C 内の対象に関しては、その割合は5%程度になるまで減少し続ける。一方 O_F 内の対象に関しては、その割合はいったん減少するが、30ステップあたりから上昇に転じ、400ステップあたりでほぼ100%に回復する。これは形が o_{novel} と同じであれば色が何であっても l_{novel} と判断するが、色が o_{novel} と同じであっても形が異なれば l_{novel} とは判断しないということを示している。つまり、IA2は学習が進行するにつれて、ある対象が l_{novel} であるか否かを、形状のみを基準に判断するようになる。このようにIA2では、学習の進行とともに形状類似バイアスが見出されるようになる。

また対比実験として、色に関して構造化された言語を作成し、これをIA2に学習させた場合、どのようなバイアスが形成されるかを分析した。図18から分かるように、このグラフは図17の O_C と O_F のグラフを入れ替えた形になっている。つまり、IA2は成長するにつれてある対象が l_{novel} であるか否かを、色のみを基準に判断するようになる。つまり色類似バイアスが形成される。

4.2.3 実験5：相互排他性バイアス

ここでは、各ステップの学習後、IAにすでに学習済みの既知の対象と新奇な対象 o_{novel} の2つを提示し、新奇なラベル l_{novel} を用いて「 l_{novel} はどちら？」と質問する。この質問をすべての既知の対象と o_{novel}

またその割合は学習の進行具合とは無関係である。

IA1は学習初期の段階では、 O_C と O_F どちらにも含まれる対象も60%以上の比較的高い割合で l_{novel} と判断する。しかし、学習の進行とともに、その割合は10%以下にまで減少する。このようにIA0およびIA1は、学習初期において違いが見られるものの、最終的には O_C と O_F に含まれる対象を l_{novel} と判断する

回答法には *Which1* を用いた。ただし、この場合選択対象は2つに限定されているので、評価はすべての対象ではなく2つの対象に対してのみ行う。

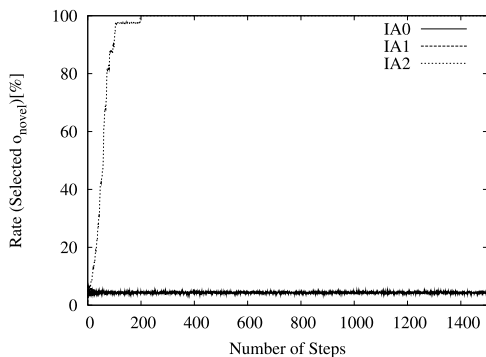


図 19 相互排他性バイアス

Fig. 19 Mutual exclusivity bias.

の間で行う。

図 19 は各 IA が新奇な対象 o_{novel} の方を選択した割合を示す。IA2 において、学習の進行とともに、新奇な対象を選択する割合が増加し、200 ステップあたりでその割合は 100% に達する。一方 IA0 および IA1 にそのような傾向は見られない。すなわち IA2 のみに相互排他性バイアスが見出された。

4.2.4 実験 6：同音異義語回避バイアス

この実験では、各ステップの学習後、以下の手続きで質問を行う。まず、を IA がその時点ですでに学習済みのラベル群 L_{known} に新奇なラベル l_{novel} を加えたテスト用のラベル群 $L_{test} = L_{known} \cup \{l_{novel}\}$ を用意する。次に、IA に新奇な対象 o_{novel} を提示し、「これ何？」と質問する。ただし、回答は L_{test} の中から強制的に選択させる。つまり IA は、 L_{test} の中から対象 o_{novel} に対して最も固有度の高いラベルを選択する。

図 20 に各 IA が、 l_{novel} を選択した割合を示す。ただしこの結果は、ランダムシードを変えて 500 回試行した結果を平均したものである。

IA0 と IA1 では、時間経過とともに l_{novel} を選択する割合は約 5% に収束する。つまり、 l_{novel} を選択する割合はチャンスレベルである。このように IA0 と IA1 は、同音異義語を好むわけでも、回避しようとするわけでもない。一方 IA2 は、学習が進行するにつれて、新奇な対象に新奇なラベルを割り当てるようになる。つまり IA2 は、同音異義語を回避するようなバイアスを持つようになる。

4.2.5 考察：学習バイアスの形成

以上の実験で、IA2 において対称性バイアスという 1 つの原理から、事物カテゴリー、形状類似、相互排他性、同音異義語回避など様々なバイアスが導出されることが示された。本項では、それらの理由について

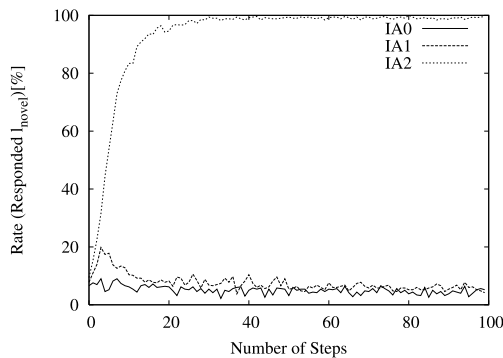


図 20 同音異義語回避バイアス

Fig. 20 Homonym avoidance bias.

表 7 実験 3 において IA に与えられる言語例

Table 7 An example of language given to IA in Experiment 3.

	c_0	c_1	c_2	...
f_0	l_0	l_0	l_0	l_0
f_1	l_1			
f_2	l_2			
...	...			

述べる。

まず事物カテゴリーバイアスに関して議論を行う。実験 3 において各 IA は、Lang25 のように形状に関して構造化された言語を与えられ、さらに c_0 あるいは f_0 という属性を持つ対象のみ経験することが許される。これは IA にとって見れば、表 7 に示したような言語を与えられている状況である。このような状況で、IA0 はいくら経験を積んでも、たとえば対象 (c_1, f_1) を見ることはない。したがって、IA0 がこの対象を見たときに各ラベルを聞いた頻度はすべて 0 である ($\forall l_k, V(l_k|(c_1, f_1)) = 0$)。つまり、IA0 にとって (c_1, f_1) は未知の対象であり続け、この対象にどのラベルを割り当てるのかは、ランダムに決定される。

一方 IA1 や IA2 にとっては、たとえ (c_1, f_1) を見たことがなくても c_1 と f_1 という個々の属性は (c_1, f_0) や (c_0, f_1) を見るという経験を通して既知であり、これらの経験を利用することで (c_1, f_1) に対して何らかの判定を行うことは可能である。表 7 から、経験可能な対象の多くに l_0 というラベルが割り当てられていることが分かる。つまり IA は、何かの対象を見たときに特定のラベル l_0 を聞く頻度が非常に高い。図 21、図 22 に、実験 3 のある試行の最終ステップにおいて、IA1 が持っていた内的頻度分布 $V_C(l|c)$ 、 $V_F(l|f)$ を示す。これらの図から分かるように、色に関しても

Lang25 においては $l_0 = 11$ である

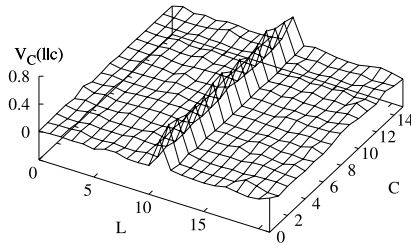


図 21 内的頻度分布 $V_C(l|c)$: IA1
Fig. 21 $V_C(l|c)$: IA1.

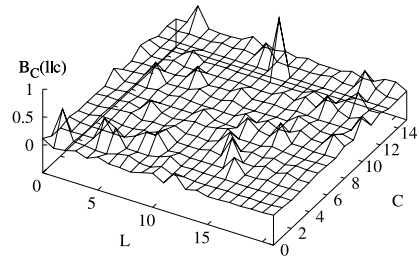


図 23 固有度 $B_C(l|c)$: IA2
Fig. 23 $B_C(l|c)$: IA2.

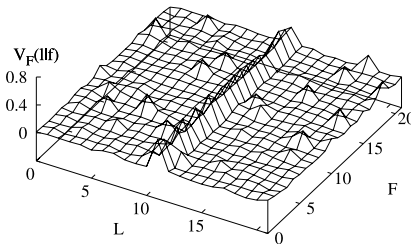


図 22 内的頻度分布 $V_F(l|f)$: IA1
Fig. 22 $V_F(l|f)$: IA1.

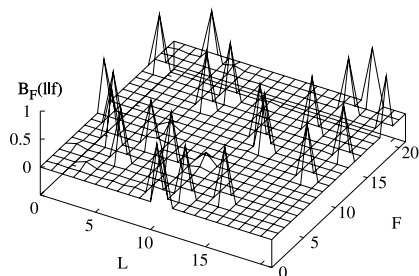


図 24 固有度 $B_F(l|f)$: IA2
Fig. 24 $B_F(l|f)$: IA2.

形に関しても、特定のラベル（ラベル 11）を聞く頻度が非常に高い。4.1.3 項でも述べたように、IA1 は経験頻度の高いラベルをすべての対象に対して割り当てようとする傾向性を持つ。このため、この試行において IA1 は、すべての対象に対してラベル 11 を割り当てることになる。

一方 IA2 は、内的に持つ頻度分布は IA1 と同様であるが、頻度から固有度を算出する際、単に正規化を行うのではなく、対称性バイアスによって対象に割り当てるラベルを対象ごとに分散させようとする傾向性を持つ。図 23、図 24 に色と形各々に関して IA2 が持つ固有度 $B_C(l|c)$ 、 $B_F(l|f)$ を示す。これらの図から、色に関しても形に関しても、各色（形）に対して最も固有度の高いラベルは色（形）ごとに分散する傾向にあることが見てとれる。ただしその傾向は、形に関する固有度 $B_F(l|f)$ においてより顕著である。また、各属性値に対する固有度の最大値も、 $B_C(l|c)$ よりも $B_F(l|f)$ の方が高い。これは、教示言語 *Lang25* が形状に関して構造化されているため、各形状に対応するラベルは一意に定められるが、各色には様々なラベルが対応付けられる、という理由による。このため、IA2 は形に強く依拠して、対象に割り当てられるラベルが何かを判定することになる。つまり IA2 は、たとえば対象 (c_1, f_1) を見たことがなくても、対象 (c_0, f_1) を見る経験によって f_1 については知っており、固有度 $B_F(l|f_1)$ に基づいて (c_1, f_1) が何かを判定するこ

とができる。その結果、事物カテゴリーバイアスが生じる。

次に IA2 において、形状類似バイアスが獲得される理由について述べる。実験 4 において、IA は新奇な対象 $o_{novel} = (c_{novel}, f_{novel})$ を提示され、それに新奇なラベル l_{novel} が割り当てられることを一度だけ教えられる。そのうえで、初めて見る対象、たとえば (c_{novel}, f_0) や (c_0, f_{novel}) に対してそれが何であるかについて判断を求められる。事物カテゴリーバイアスのところで述べたように、IA0 は初めて見る対象に対してはまったく対処できない。このため以下では、IA1 と IA2 について議論を行う。

IA1 と IA2 の差異は、対称性バイアス、すなわち共通度と固有度の相互作用を認めるか否かという点のみにある。IA1 においては、たとえば赤いものを見たときに「いちご」というラベルを聞く経験を何度も繰り返せば、「赤いものはいちごである」という確信度は高まる。また、「赤いものはいちごである」という確信度が高まれば、相対的に「赤いものはピーマンである」という確信度は低くなる。しかしその信念は、まったく無関係な対象に対する信念、たとえば「緑のものはピーマンである」という確信度に影響を与えることはない。

一方 IA2 では、共通度と固有度が互いに規定し合うために、「赤いものはピーマンである」という確信度が

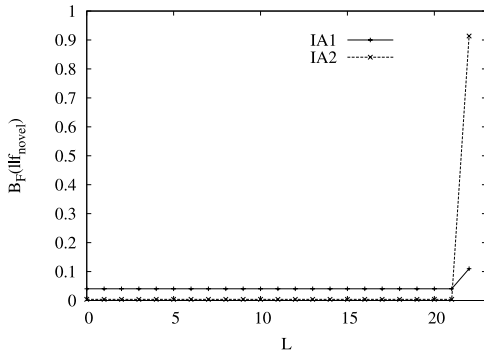


図 25 f_{novel} に対する各ラベルの固有度 $B_F(l|f_{novel})$
 Fig. 25 $B_F(l|f_{novel})$ for f_{novel} .

低下すれば、それに比例して「ピーマンは赤い」との確信度も低下する。その結果として相対的に「ピーマンは緑である」という確信度が高まる。さらに共通度が固有度に影響を与えるため、「緑のものはピーマンである」という確信度が高まる。このような信念の変化は他の信念にも波及し、信念全体が再編成されることになる。この点がバイアスの形成に深く関与する。

以上に述べたように、IA1 は対称性バイアスを持たないため、新奇な対象 (c_{novel}, f_{novel}) に新奇なラベル l_{novel} が付与され、それを教示されたとしても、 c_{novel} や f_{novel} に対する固有度は、それまでの学習成果とは無関係に決定される。形状に関していえば、 f_{novel} は教示によって一度だけ見ることができるが、そのときに聞くラベルは l_{novel} なので、 f_{novel} に関して IA1 が持つ内的頻度分布は、 $V_F(l_{novel}|f_{novel}) = \alpha = 0.01$ 、 $V_F(l_k|f_{novel}) = 0$ ($k \neq novel$) となる。固有度はこの頻度分布をボルツマン分布を用いて正規化したものであり、図 25 のようになる。ただし、ここでは $l_{novel} = 22$ に設定している。図から分かるように、 f_{novel} に対する l_{novel} の固有度は、その他のラベルと比較してわずかに高い程度である。 (c_{novel}, f_{novel}) に l_{novel} を付与する教示を何度も繰り返せば、 l_{novel} の固有度は高くなるが、一度だけの教示ではこの程度の格差しか生じない。結果は省略するが、この点は色に関して同様である。

一方 IA2 は対称性バイアスを持つため、新奇な対象に対する経験は、それまでに学んだ既知の対象に対する信念からの影響を強く受ける。形状に関して構造化された Lang25 のような言語を学習する場合、各色を見たときに聞くラベルはランダムである。このため各色（ラベル）に対して特定のラベル（色）の固有度（共通度）が突出して高くなるようなことはない。一方形に関しては、各形状を見るときにはその

形に対応する特定のラベルを聞く頻度が高い。したがって学習の進行とともに、ラベル（形状）ごとに特定の形状（ラベル）が非常に高い共通度（固有度）を持つようになる。このため、各形に対する l_{novel} の固有度 $B_F(l_{novel}|f_j)$ ($j \neq novel$) は相対的に低くなる。IA2 では対称性バイアスによって固有度と共通度が比例関係を満足するように修正されるので、固有度の低下にともなって l_{novel} に対する各形の共通度 $B_F(f_j|l_{novel})$ ($j \neq novel$) も低下することになる。その結果として相対的に l_{novel} に対する f_{novel} の共通度 $B_F(f_{novel}|l_{novel})$ は高くなる。さらに共通度が固有度に影響を与えるため、 f_{novel} に対する l_{novel} の固有度 $B_F(l_{novel}|f_{novel})$ が高くなる。つまり IA2 では、新奇な対象に対する教示がたとえ一度であっても、それまでの経験によって形成された他の信念からの影響を受けることで、固有度 $B_F(l_{novel}|f_{novel})$ は、他の固有度 $B_F(l_k|f_{novel})$ ($k \neq novel$) に比べて突出して高くなる（図 25）。このため IA2 は、 f_{novel} という形状を持つ対象 (c_i, f_{novel}) ($i \neq novel$) に対して l_{novel} を割り当てる傾向が強くなる。その結果 IA2 は、 O_F 内のすべての対象に対して l_{novel} と回答することになる。

一方色に関しては、固有度 $B_C(l_{novel}|c_{novel})$ が、形状の場合ほど突出することはない。したがって c_{novel} という色を持つ対象 (c_{novel}, f_j) ($j \neq novel$) に対して l_{novel} を割り当てる傾向は、それほど強くはない。対象 (c_{novel}, f_j) に対してはむしろ、 f_j という形状に対応する特定のラベルを割り当てようとする傾向の方が強い。このため O_C 内の対象に対しては、形状ごとに異なるラベルが割り当てられることになる。

このように、各形を見たときにはその形に対応する特定のラベルを聞く頻度が高いが、各色を見たときに聞くラベルはランダムであるという経験の非対称性と対称性バイアスの存在という 2 つの要因によって、形状類似バイアスが生み出される。換言すれば、IA2 にどのようなバイアスが形成されるかは教示言語に依存し、もしそれが色に関して構造化されていれば、IA2 は色類似バイアスを形成することになる。

最後に相互排他性バイアス、および同音異義語回避バイアスについて議論を行う。実験 5 と実験 6 では、IA に新奇な対象 (c_{novel}, f_{novel}) が提示されるが、この対象にどのようなラベルを割り当てるのかについては何も教示されない。このため IA0 がこの対象を見たときに各ラベルを聞いた頻度はすべて 0 になる ($\forall l_k, V(l_k|(c_{novel}, f_{novel})) = 0$)。また IA1 についても同様に、 c_{novel} や f_{novel} を見たときに各ラベルを

聞いた頻度はすべて 0 となる ($\forall l_k, V_C(l_k|c_{novel}) = 0, V_F(l_k|f_{novel}) = 0$). このため、この新奇な対象に割り当てられるラベルは、ランダムに決定されることになる。実験 5 では、新奇な対象とともに既知の対象が提示されるが、この対象にはすでに l_{novel} 以外のラベルが割り当てられている。このため、これらの IA は、「 l_{novel} はどっち？」という質問に対して、「どちらも l_{novel} ではない」と回答することになる。

さて、IA は経験を積むにつれて、各対象に対して正しいラベルを割り当ててようになる。逆にいえば、学習の進行にともなって、各ラベルは何らかの対象に割り当てられる。新奇な対象を提示されたとき、IA はこの対象に何らかのラベルを付与しようとするが、 l_{novel} 以外のラベル l_k ($k \neq novel$) はすでに何らかの対象に割り当てられている。このため、新奇な対象にも l_k を割り当てると、このラベルは複数の対象を指示する同音異義語になってしまう。4.1.3 項で述べたように、IA2 は対称性バイアスによって、各対象に割り当てられるラベルを対象ごとに分散させようとする傾向性を持つ。このため IA2 は、(c_{novel}, f_{novel}) について一度も教示されたことがなくても、いまだ指示対象を持たないラベル l_{novel} を割り当てようとするようになる。その結果、IA2 には相互排他性バイアスや同音異義語回避バイアスが見出されることになる。

5. まとめと議論

本稿では、IA0, IA1, IA2 という 3 種類の幼児エージェントのモデルを構築し、数値実験によって、言語を構造化する能力やバイアスの形成能力の比較を試みた。IA0 は、経験を通して対象とラベルの関係を記憶していき、その記憶（信念）に基づいて種々の判断を行うエージェントである。IA1 は、メモリ空間節約のため対象とラベルの関係を属性ごとに分解して記憶し、判断の際にはそれらを統合して使用する。IA2 はそれに加えて対称性バイアスを持つ。この IA は、論理的な正確さを犠牲にして、コミュニケーションにおける整合性を重視するエージェントであると見なすことができる。

記憶能力という観点からいえば、IA0 が最も優れている。IA0 は、その優れた記憶能力のために、対象とラベルの対応関係を対象ごとに個別に記憶することができる。このため、どのように無秩序な言語であっても構造化された言語と同様に学習可能である。ただし学習には、非常に長期間を要する。これは言語の進化という観点からいえば、世代を経ても言語は変わらないということの意味する。

一方 IA2 は、分解と統合という操作によって情報が損失するため、Lang1 のような無秩序な言語に関しては、それを忠実に再現することができない。逆にいえば、言語は世代を経るごとに変化せざるをえない。この変化が結果として言語の構造化をもたらす。IA1 においても世代を経るごとに言語は変化するが、数世代ですべての対象に同一のラベルを付与するという無意味な言語に収束する。すなわち、1 つのラベルがすべての対象を指示する同音異義語になる。

言語をコミュニケーションの道具であるとするならば、IA 間で情報伝達を行うことができ、また親世代から子世代へ速やかに言語を教授できる言語が良い言語である。この観点からすれば、IA2 にとって Lang1 は良い言語ではない。IA2 は言語の情報量を保持したまま、自らの学習能力に適した形に言語を改変する。その結果、Lang25 を獲得した IA2 は、親から子へ少ない事例で速やかにかつ正確に言語を伝えることを可能とする。

言語の学習という観点からいえば、このような言語の構造化は、IA2 が持つ記憶能力不足という制約を、事物カテゴリーバイアスという汎化能力に転化させる。IA0 はその優れた記憶力のため、教示された対象については確実に学習できるが、教示されない未知の対象についてはまったく対処できない。つまり IA0 は、ラベルを固有名詞と見なして学習を行うエージェントであるといえる。一方 IA1 においては、分解と統合という操作によって情報が失われるため、教示された対象さえ学習することを困難にする。IA2 においても Lang1 のような無秩序な言語を学習する際には、情報損失は、学習の失敗という否定的結果しか導かない。しかしそれは、Lang25 のような構造化された言語を学習する際には、汎化能力という肯定的結果をもたらす、学習を加速する要因となる。

IA2 は、ラベルを普通名詞と判断するが、その際ラベルの適用範囲を形状が同じ対象に拡張する。すなわち、IA2 は形状類似バイアスを持つ。ただし、形状類似バイアスを持つか否かは教示言語の構造に依存する。教示言語がたとえば色に関して構造化されていれば、色類似バイアスを形成する。つまり IA2 は、言語を学習していく過程の中で、言語の構造に応じたバイアスを柔軟に形成する。IA2 には、さらに相互排他性バイアスや同音異義語回避バイアスの形成も見られる。このように IA2 は、構造化された言語を学んでいく中で、各種の語彙学習バイアスを自律的に形成し、効率良く学習を行うようになる。本稿の結果は、幼児の言語学習過程において観察される様々な傾向性が、対称

性バイアスという1つの原理から導出できることを示唆する。

ここで、言語の構造化と語彙学習バイアスの形成との関係について述べたい。IA2をモデル化するにあたって、本稿で主に仮定したのは(1)対称性バイアスと、(2)色よりも形の方を細かく識別する能力の2つである。4.1.3項で述べたように(2)の仮定は、たとえば教示言語がランダムなものであっても、各形を見たときにはその形に対応する特定のラベルを聞く頻度が高いが、各色を見たときには様々なラベルをランダムに聞くという経験の非対称性を導く。また対称性バイアスは、IA2に各色(形)に割り当てられるラベルを色(形)ごとに分散させようとする傾向性を与える。この経験の非対称性と対称性バイアスの作用によって、形状類似バイアスが形成され、その結果としてわずかながらも言語の構造化が生じる。

4.2.5項で述べたように、このような言語の構造化は、それを教示言語とする次世代のIA2に、さらなる経験の非対称性をもたらす。この非対称性と対称性バイアスの作用によって、さらに強力な形状類似バイアスが形成される。このため、言語の構造化がさらに促進される。さらなる言語の構造化は、より強い経験の非対称性をもたらす。IA2は対称性バイアスによってより強力な形状類似バイアスを形成する。このように言語の構造化と形状類似バイアスの形成は、対称性バイアスの下で、正のフィードバックループをなす。

一方、相互排他性バイアス、および同音異義語回避バイアスは、4.2.5項で述べたように、対称性バイアスから直接導かれる。つまり、言語の構造化とこれらのバイアスの形成は無関係である。また事物カテゴリーバイアスは、形状類似バイアスと同じメカニズムによって導かれる。

最後に先行研究との比較を行いたい。1章で述べたように、日高ら²¹⁾は発達過程における形状類似バイアスの形成を扱っている。この点は本研究と共通しているが、彼らのモデルでは、バイアスの形成にニューラルネットワークの汎化能力を利用している。つまりバイアスの形成は対称性バイアスの有無とは無関係であり、本研究との関連性は低い。

Smith²⁴⁾は、進化の観点から、同音異義語回避バイアスの形成を扱っているが、このバイアスは語彙学習

バイアスというよりはむしろ、本研究で導入した対称性バイアスに近い。彼は、同音異義語を回避するバイアスを持つ、同音異義語を好むバイアスを持つ、どちらでもない、という3種類のエージェントをモデル化した。これらのエージェントを本モデルのIAに対応させるとすれば、IA2は同音異義語回避バイアスを持つエージェントに、IA0はどちらでもないエージェントに対応する。IA1は、実験6で示したように、初めて見る対象に対しては、あえて既知のラベルを割り当てようとするわけではない。この意味では、IA1はどちらでもないエージェントに対応する。しかし4.1.3項で述べたように、繰り返し経験を積んでいけば、経験頻度の高いラベルをすべての対象に割り当てようになる。この意味では、IA1は同音異義語を好むバイアスを持つともいえる。

Smithは、同音異義語回避バイアスを持つエージェントが、進化的に安定であることを示したが、このバイアスと形状類似バイアスや事物カテゴリーバイアスなどの語彙学習バイアスとの関連については言及していない。一方本研究では、対称性バイアスを持つIA2が、言語を学習に適した形に構造化させることを示しただけでなく、このバイアスが発達過程において語彙学習バイアスを形成する要因となることも示した。

以上に示した先行研究では、形状類似バイアスや同音異義語回避バイアスという単一のバイアスのみを扱っている。このため、複数の語彙学習バイアス間の関連性は定かではない。このように各種バイアスの形成を独立に扱った場合、個別に形成された複数のバイアス間での適用調整が問題となると思われる。

たとえば、目の前にすでに名前を知っている既知の対象が1つだけしかなく、それに新奇なラベルが付与されるという状況を考えよう。この状況において、そのラベルが目の前の既知の対象を指示すると解釈するならば、事物は1つの名前しか持たないという相互排他性バイアスを破ることになる。一方、相互排他性バイアスを遵守し、ラベルがその対象のある一部分や対象の一属性である材質などを指示すると考えるならば、事物全体バイアスに抵触する。つまりこの状況では、事物全体バイアスと相互排他性バイアスは両立できない。実際の幼児は、このような状況において、事物全体バイアスを緩和し、ラベルを部分や材質を表す語と見なす傾向があることが知られている^{13),14)}。

このように、複数のバイアスが両立不可能な場合、どのバイアスを優先するのか決定しなければならないが、各種バイアスを独立に扱った場合、その調整のために何らかの機構を新たに導入する必要があるだろう。

ただし(2)の能力は、IA2だけでなく他のIAでも仮定されている。

もしIA2が、形よりも色の方を細かく識別する能力を持つならば、色方向への言語の構造化が生じる。つまり、対称性バイアスさえ持てば、言語の構造化は生じる。

一方本モデルでは、複数の語彙学習バイアスが対称性バイアスという1つの原理のもとで関連しあっており、IA2は何の調整機構も持たずに、幼児と同様の適用調整を行うことができる²²⁾。

6. 今後の課題

本章では今後の課題について述べる。まずIA2では、対称性バイアスや色よりも形の方を細かく識別できるということが仮定されているが、これらの特性が生物進化の過程の中でどのように出現したのかという問題については、モデルを進化シミュレーションに拡張して分析する必要がある。

次に本稿のモデルでは、各対象に対して1つのラベルを付与するという条件を設定した。このため「りんご」と「果物」のように包含関係にある語を明示的に扱うことができない。このような語を扱うには、1つの対象に複数のラベルが付与されることを認める必要がある。またモデルでは、対象は色と形の2つの属性のみを持つと仮定した。今後は材質など他の属性も導入し、物体と物質の違いなども扱えるようにモデルを拡張したい。

謝辞 本研究の一部は、文部科学省21世紀COEプログラム「インテリジェントヒューマンセンシング」の補助による。

参 考 文 献

- Clark, E.: On the logic of contrast, *Journal of Child Language*, Vol.15, pp.317-335 (1988).
- Davidson, D. and Tell, D.: Monolingual and bilingual children's use of mutual exclusivity in the naming of whole objects, *Journal of Experimental Child Psychology*, Vol.92, pp.25-45 (2005).
- 針生悦子, 今井むつみ: 語意学習メカニズムにおける制約の役割とその生得性, 心の生得性—言語・概念獲得に生得的制約は必要か, 今井むつみ(編), pp.131-171, 共立出版, 東京(2000).
- Halberda, J.: The development of a word-learning strategy, *Cognition*, Vol.87, pp.B23-B34 (2003).
- Imai, M. and Gentner, D.: A crosslinguistic study of early word meaning: Universal ontology and linguistic influence, *Cognition*, Vol.62, pp.169-200 (1997).
- Imai, M., Gentner, D. and Uchida, N.: Children's theory of word meanings: The role of shape similarity in early acquisition, *Cognitive Development*, Vol.9, pp.45-75 (1994).
- Imai, M. and Haryu, E.: Learning proper nouns and common nouns without clues from syntax, *Child Development*, Vol.72, pp.787-802 (2001).
- Jones, S.S. and Smith, L.B.: How children name objects with shoes, *Cognitive Development*, Vol.13, pp.323-334 (1998).
- 藤田和生: 霊長類のコミュニケーション, コミュニケーションと思考, 乾 敏郎, 安西祐一郎(編), 岩波書店, 東京, pp.79-114 (2001).
- Landau, B., Smith, L.B. and Jones, S.S.: The importance of shape in early lexical learning, *Cognitive Development*, Vol.3, pp.299-321 (1988).
- Landau, B., Smith, L.B. and Jones, S.S.: Syntactic context and the shape bias in children's and adults' lexical learning, *Journal of Memory and Language*, Vol.31, pp.807-825 (1992).
- Markman, E.M.: *Categorization and naming in children: Problems of induction*, MIT Press, Cambridge (1989).
- Markman, E.M.: Constraints children place on word meanings, *Cognitive Science*, Vol.14, pp.121-157 (1990).
- Markman, E.M. and Wachtel, G.F.: Children's use of mutual exclusivity to constrain the meaning of words, *Cognitive Psychology*, Vol.20, pp.121-157 (1988).
- Markman, E.M., Wasow, J.L. and Hansen, M.B.: Use of the mutual exclusivity assumption by young word learners, *Cognitive Psychology*, Vol.47, pp.241-275 (2003).
- 田中正之, 松沢哲郎: シンボルの成立, 心の比較認知科学, 渡辺 茂(編), pp.226-268, ミネルヴァ書房, 京都(2000).
- 今井むつみ: 概念発達と言語発達における類似性の役割, 類似から見た心, 大西 仁, 鈴木宏昭(編), pp.148-178, 共立出版, 東京(2001).
- Quine, W.V.O.: *Word and Object*, MIT Press, Cambridge, MA (1960).
- 田口 亮, 木村優志, 篠原修二, 桂田浩一, 新田恒雄: Online-EM による語意学習機構と学習バイアスの適用, 信学技報, Vol.NLC2005-60 (/SP2005-93), pp.31-36 (2005).
- Samuelson, L.K.: Statistical Regularities in Vocabulary Guide Language Acquisition in Connectionist Models and 15-20-Month-Olds, *Developmental Psychology*, Vol.38, pp.1016-1037 (2002).
- 日高昇平, 齋木 潤: 幼児の新奇語カテゴリー化のモデル研究, 認知科学, Vol.12, pp.1-16 (2005).
- 篠原修二, 田口 亮, 橋本 敬, 桂田浩一, 新田恒雄: 語彙学習エージェントにおけるバイアスの自律調整について, 人工知能学会論文誌, Vol.22, No.2 (2007).

- 23) Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W. and Carrigan, P.: A search for symmetry in the conditional discriminations of rhesus monkeys, baboons and children, *Journal of the Experimental Analysis of Behavior*, Vol.37, pp.23-44 (1982).
- 24) Smith, K.: The evolution of vocabulary, *Journal of Theoretical Biology*, Vol.228, pp.127-142 (2004).
- 25) Sutton, R.S. and Barto, A.G.: *Reinforcement Learning — An Introduction*, MIT Press, Cambridge, MA (1998).
- 26) 山崎由美子, 岡ノ谷一夫, 入来篤史: 動物の論理を比較する: 意味の獲得へと導くメカニズム, 信学技報, Vol.NC2005-93 (2006-1), pp.19-24 (2006).

付 録

条件式 (22), (23) から $B_C(l_k|c_i)$ と $B_C(c_i|l_k)$ を消去すると

$$\sum_{l \in L} B_C(c_i|l) \sum_{c \in C} B_C(l_k|c) = 1$$

となる. $\forall c_i \in C, \forall l_k \in L$ に対してこの式が成立するためには, η, ξ を $\eta\xi = 1$ を満たす正定数として

$$\sum_{c \in C} B_C(l_k|c) = \eta$$

$$\sum_{l \in L} B_C(c_i|l) = \xi$$

という関係が満たされなければならない.

ここで

$$\forall l_k \in L, \sum_{c \in C} B_C(c|l_k) = 1$$

であるので

$$\begin{aligned} |L| &= \sum_{l \in L} \sum_{c \in C} B_C(c|l) \\ &= \sum_{c \in C} \sum_{l \in L} B_C(c|l) \\ &= \sum_{c \in C} \xi \\ &= \xi|C| \end{aligned}$$

となる. つまり $\xi = |L|/|C|$ である. また

$$\forall c_i \in C, \sum_{l \in L} B_C(l|c_i) = 1$$

であるので

$$|C| = \sum_{c \in C} \sum_{l \in L} B_C(l|c)$$

$$\begin{aligned} &= \sum_{l \in L} \sum_{c \in C} B_C(l|c) \\ &= \sum_{l \in L} \eta \\ &= \eta|L| \end{aligned}$$

となる. つまり $\eta = |C|/|L|$ である. すなわち, 固有度 $B_C(l_k|c_i)$ および共通度 $B_C(c_i|l_k)$ が, 条件式 (22), (23) を満たすならば, 両者の間には, $\forall c_i \in C, \forall l_k \in L$ に対して

$$B_C(c_i|l_k) = \frac{|L|}{|C|} B_C(l_k|c_i)$$

という関係が成立する.

形属性に関しても同様に, IA2 の信念が条件式 (24), (25) を満足するとき, 固有度 $B_F(l_k|f_j)$ と共通度 $B_F(f_j|l_k)$ の間には, $\forall f_j \in F, \forall l_k \in L$ に対して

$$B_F(f_j|l_k) = \frac{|L|}{|F|} B_F(l_k|f_j)$$

という関係が成立する. これらの両辺を各々掛け合わせると

$$\begin{aligned} B_C(c_i|l_k) B_F(f_j|l_k) &= \frac{|L|}{|C|} \frac{|L|}{|F|} B_C(l_k|c_i) B_F(l_k|f_j) \\ &= \frac{|L|^2}{|O|} B_C(l_k|c_i) B_F(l_k|f_j) \end{aligned}$$

となる. 対象 (c_i, f_j) の固有度 $B(l_k|(c_i, f_j))$ と共通度 $B((c_i, f_j)|l_k)$ は, 各々

$$B(l_k|(c_i, f_j)) = B_C(l_k|c_i) B_F(l_k|f_j)$$

$$B((c_i, f_j)|l_k) = B_C(c_i|l_k) B_F(f_j|l_k)$$

であるので,

$$B((c_i, f_j)|l_k) = \frac{|L|^2}{|O|} B(l_k|(c_i, f_j))$$

という関係が成立する.

(平成 18 年 2 月 16 日受付)

(平成 18 年 4 月 7 日再受付)

(平成 18 年 5 月 17 日採録)



篠原 修二

平成 11 年神戸大学大学院自然科学研究科博士後期課程修了. 博士 (理学). ベンチャー企業で, Web アプリケーションやデータベースシステムの開発に携わった後, 平成 17 年から豊橋技術科学大学で生体情報を用いた教育支援システムの研究開発に従事. 日本認知科学会, 進化経済学会各会員.



田口 亮 (学生会員)

平成 16 年豊橋技術科学大学大学院工学研究科修士課程修了。現在、同大学院博士後期課程在学中。人工知能学会会員。人間-エージェント対話に興味を持つ。



橋本 敬

平成 8 年東京大学大学院総合文化研究科博士後期課程修了。博士(学術)。理化学研究所脳科学総合研究センターでの基礎科学特別研究員を経て、平成 11 年より北陸先端科学技術大学院大学知識科学研究科助教授。言語のダイナミクスと進化、社会制度の形成と変化等の研究に従事。国際人工生命学会、日本認知科学会、日本進化学会、日本進化経済学会各会員。



桂田 浩一 (正会員)

平成 7 年大阪大学基礎工学部情報工学科卒業。平成 12 年同大学院基礎工学研究科博士後期課程修了。同年豊橋技術科学大学工学研究科助手。博士(工学)。マルチモーダル対話、知識処理に関する研究に従事。AAAI, 人工知能学会, 日本音響学会, 言語処理学会, ヒューマンインタフェース学会各会員。



新田 恒雄 (正会員)

昭和 44 年東北大学工学部電気工学科卒業(株)東芝総合研究所, 同社マルチメディア技術研究所を経て、平成 10 年豊橋技術科学大学大学院工学研究科教授。工学博士。音声認識・合成・文字認識, マルチモーダル対話システム, および概念獲得の研究に従事。IEEE, 電気情報通信学会, 人工知能学会, 日本音響学会各会員。