

幼児の語の意味の獲得における推論バイアスと語彙コーパスの役割

松村圭祐

京都学園大学大学院人間文化研究科

語の意味の般化において、形状バイアスや材質バイアスがあることが知られている。これらのバイアスには、英語話者の幼児における語彙の特性の関与が指摘されている。先行研究からも、この語彙が、バイアスの発現に重要な役割を果たしていることが示唆されている。それに加えて、獲得語彙数もバイアスの発現と関連があることが示唆されていた。そこで、本研究では、コネクショニストモデルによって獲得語彙数を段階的に増やすシミュレーションを行った。その結果、バイアスの発現と獲得語彙数とは、関連がないことがわかった。この結果を受けて、本研究ではさらに、語彙の特性について検証を行った。本研究では、幼児の語彙に含まれている、固体物であり形状の類似性によってまとめられる計数名詞と、非固体物であり材質の類似性によってまとめられる質量名詞の比率に注目した。これらの構成比率を操作し、ネットワークに提示することで、これらがバイアスの発現に及ぼす効果を検討した。シミュレーションの結果、すべての語彙においてバイアスの発現がみられたため、語彙の構成比率とバイアスの発現との関連性は明らかにならなかった。したがって、幼児に特有の語彙構成のみで、般化バイアスが生じるとはいえず、他の要因が決定的に関与しているといえる。

Background

子どもは、産出語彙数が50から100語を越えたあたりから、爆発的な勢いで語を獲得するようになることが知られている。この現象を、語彙爆発 (vocabulary spurt) という。このような子どもの急速な語の獲得は、大人が言葉を教えることによって成立しているのだろうか。確かに、大人が子どもに対して、言葉を教えることは日常的にみられることである。だがしかし、実際には、ある語が指し示す対象は、論理的には無数にあり、一義的に決定することができない (Quine, 1960)。たとえば、「ウサギ」という語は、ウサギ全体を指すのか、あるいは、長い耳や白い体毛、赤い目など、いろいろと考えられるのである。

だが、このような語の指示対象に対する決定不能の困難を抱えながらも、子どもは

新奇な語を一度聞いただけで、その語を概念に対応させて学習することができる。これは即時マッピング (fast mapping) とよばれる現象である。この即時マッピングを可能にするものとして、語の意味の推論に対する、制約やバイアスが提案されている。これらの代表的なものとして、Landau, Smith, & Jones (1988) の形状バイアス (shape bias) や、Markman (1989) の分類学的制約 (taxonomic constraint)、存在論的知識による制約 (Soja, Carey, & Spelke, 1991) などがある。

形状バイアスの発現には、発達初期の語彙のなかに存在する統計的規則性が、重要な役割を果たしているといわれている。Samuelson & Smith (1999) は、30ヵ月児に一般に知られている300の物体の名前を調べ、

それらの類似性の構造を検討した。その結果、以下の2つの特徴が明らかになった。1つは、対象児の語彙には、形状の類似性によってまとめられる固体物の名前が多数を占めることがわかったこと。2つ目に、材質の類似性によってまとめられる非固体物の名前は、ほとんどみられなかった。Colunga & Smith (2000) は、Samuelson & Smith (1999) の調査結果に基づいて、幼児の語の意味の般化を、コネクショニストモデルによるシミュレーションで検証した。その結果、幼児の語彙がもつ統計的規則性によって、般化バイアスが発現することが明らかになった。

Samuelson & Smith (1999) はまた、子どもはこれらの語の多くを知った後にだけ、形状の類似によって固体物を指す名前を般化し（形状バイアス）、材質の類似によって非固体物の新奇な名前を般化する（材質バイアス material bias）、ということも示した。この結果は、これらのバイアスは、彼らの語彙にある統計的規則性を学習した結果として発現する、ということを示唆している。

だが、これらの調査結果からは、なぜ獲得語彙が少ないと、般化バイアスが発現しないのか。この重要な点が、Colunga & Smithのシミュレーション結果では、明らかにされていない。

Simulation I

ここでは、形状バイアスと材質バイアスを取り上げ、コネクショニストモデルによるシミュレーションを行った。形状バイアスとは、未知の事物の名前を、それと同じ形状をもつ他の事物に対して適用するバイアスである。材質バイアスは、未知の物質の名前が、材質の類似性によって同じ材質でできている物質にも般化し得る、と考えるバイアスである。

このシミュレーションの目的は、獲得語彙数とバイアスの発現との関連性の検証を行うことであった。そこで、単一のネットワークに対して、語彙を段階的に学習させるシミュレーションを実行した。学習する語彙段階は、20, 50, 80, 100, 120, 150,

180, 200語の計8段階に分類された。ネットワークに提示される入力パターンは、Samuelson & Smith (1999) の調査結果を参照して作成した。ネットワークの学習課題は、この入力パターン（e. g., (実在物としての) リンゴ）に対して、特定の出力パターン（e. g., 「リンゴ」という名前）を学習することであった（入出力表現は、表1を参照）。

ネットワークのアーキテクチャは、入力層（28ユニット）、隠れ層（50ユニット）、出力層（10ユニット）の計3層からなるフィード・フォワード型のネットワークを構築した。出力誤差の修正には、Rumelhart & McClelland (1986) の誤差逆伝播

表1 学習課題の例.

ターゲット	入力パターン				教師信号
	形状次元	材質次元	固体性次元	統語次元	出力層
リンゴ	0, 1, ..., 0, 1	0, 1, 0, 3, ..., 0, 8, 0, 4	1, 0	1, 0	1, 0, ..., 0, 0
ミルク	0.7, 0.5, ..., 0.2, 0.6	1, 1, ..., 0, 0	0, 1	0, 1	0, 1, ..., 0, 0

法を使用した。学習に使用したパラメータは、学習率が0.1，モーメンタムは0.9とした。学習の精度を検証するために，学習セットから任意に入力パターンを抽出してテストセットを作成した。そして，これらの入力に対する出力パターンと教師信号との平均二乗誤差が0.5以下のときに，学習が完了したとみなし，つぎの学習段階に進めた。

バイアスの検証には，Colunga & Smith (2005) で使用された新奇語般化課題の概念版を使用した。この検証課題は，形状マッチテストと材質マッチテストの2種類からなっていた。

形状マッチテスト 標準刺激と選択刺激

(形状類似刺激と材質類似刺激)との類似度から選択率を算出した。選択刺激のうち，形状類似刺激は，標準刺激と形状次元，固体性次元，統語次元で一致し，材質次元のみ異なっていた。それに対して，材質類似刺激では，標準刺激と材質次元のみが一致し，その他の次元はすべて標準刺激と異なっていた。形状類似刺激の選択率がチャンスレベルを有意に上回っていたとき，形状バイアスが発現したといえる。

材質マッチテスト 形状マッチテストと同様に，標準刺激と選択刺激 (形状類似刺

激と材質類似刺激)との類似度から選択率を算出した。このテストでは，選択刺激のうち，材質類似刺激は，標準刺激と材質次元，固体性次元，統語次元で一致し，形状次元のみ異なっていた。それに対して，形状類似刺激では，標準刺激と形状次元のみが一致し，その他の次元はすべて標準刺激と異なっていた。材質類似刺激の選択率がチャンスレベルを有意に上回っていたとき，材質バイアスが発現したといえる。

学習の結果，すべてのネットワークの平均二乗誤差が0.5以下であったことを確認した。これをもって，単一のネットワークが，すべての学習段階において，学習課題を遂行したと評価した。

シミュレーションの結果，ネットワークは，すべての学習段階において，一貫した般化バイアスを示すことが明らかになった (形状バイアスはFig. 1, 材質バイアスはFig. 2)。しかしこの結果は，150語以上を獲得した後に，形状バイアスが現れる，というSamuelson & Smith (1999) による心理実験の結果とは，大きく異なっていた。この結果は，獲得語彙数とバイアスの発現とは，関連性がないことを示唆するものであると考えられる。

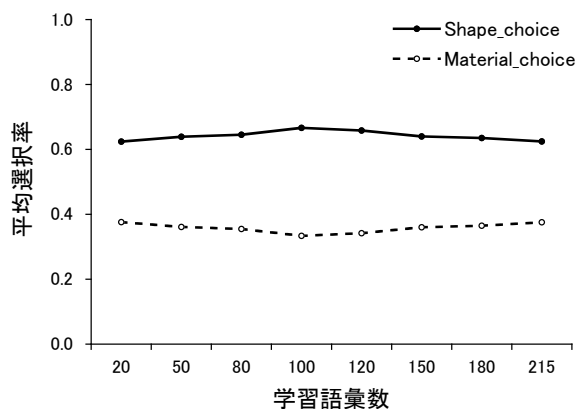


Fig. 1. 形状マッチテストにおける平均選択率

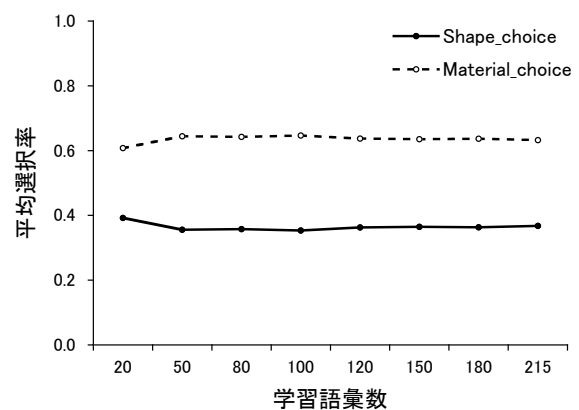


Fig. 2. 材質マッチテストにおける平均選択率

Simulation 2

Simulation 1の所見から, Samuelson & Smith (1999)の主張を再検証する必要があった。そこで, 語彙の構成比率を体系的に操作し, それがバイアスの発現に及ぼす効果を調べることを, このシミュレーションの目標とした。Samuelson & Smith (1999)の構成比率から作成した語彙(比率は0.5:0.1), 固体-形状類似-計数名詞と非固体物-材質類似-質量名詞の割合が0.5:0.5, 0.3:0.3, 0.1:0.1, 0.1:0.5の計5セットを作成し, ネットワークに提示した。

このシミュレーションでは, 分析を簡略にするために, 隠れ層のユニット数を10とした。Simulation 1では, 最初期の20語段階においても般化バイアスが発現していた。この結果は, 約150語以上を獲得した子どもにのみ, 強い形状バイアスが現れるというSamuelson & Smith (1999)の主張とは異なる。そこで, ここでは20語の学習段階のみを対象としたシミュレーションに限定した。学習課題, バイアスの検証方法ともにSimulation 1と同様とした。学習は, すべてのネットワークで5,000回行った。学習パラメータもSimulation 1と同様の値に設定した。

新奇語般化課題によって形状バイアスを検証した結果, すべての語彙において, 形状類似刺激の選択率が, 材質類似刺激のそれを上回っていた(Fig. 3)。すなわち, 異なる構成比率をもつ語彙で学習した結果においても, ネットワークによって形状バイアスが発現していた。この結果から, 語彙に含まれる固体物-形状類似-計数名詞と, 非固体物-材質類似-質量名詞のそれぞれの割合は, バイアスの発現には寄与しないことが示唆された。

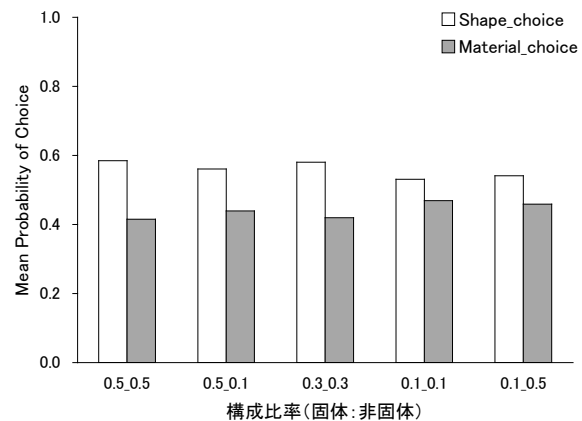


Fig. 3. 語彙コーパスの構成比率の差による形状類似刺激の選択率の差異

References

- Colunga, E. & Smith, L. B. (2000). Committing to an ontology: A connectionist account. *The Twenty Second Annual Meeting of the Cognitive Science Society*, **22**, pp. 89-94.
- Colunga, E. & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, **112**, 347-382.
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexicon learning. *Cognitive Development*, **3**, 299-321.
- Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.
- Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.
- Rumelhart, D. & McClelland, J. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition (Vol. 1)*. Cambridge, MA: MIT Press.
- Samuelson, L. K. & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, **73**, 1-33.
- Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meanings: Object terms and substance term. *Cognition*, **38**, 179-211.