3E3-3

言語獲得に適した意味表現について

A Semantic Representation Suited for Language Acquisition

麻生 英樹[*1] 伊東 幸宏[*2] 高木 朗[*1,3]
Hideki ASOH Yukihiro ITOU Akira TAKAGI

[*1] 産業技術総合研究所 AIST
[*2] 静岡大学情報学部 Faculty of Informatics, Shizuoka University
[*3] 言語処理研究所 Language Processing Laboratory

A semantic representation which is suited for language acquisition is proposed. The representation of the result of visual cognition is assumed to have rich combinatorial structure and to be the semantic representation of the language. A method for learning both lexicon and grammar from recognized visual and auditory inputs is described.

1. はじめに

言語獲得は，人間が行う学習の中でも最も驚くべきものの一つであり，そのメカニズムは未だ解明されてはいない．言語獲得のための入力情報は，音声言語情報に加えて，様々な感覚情報であり，特に視覚情報は重要な役割を果たすと考えられる．言語獲得研究のゴール，すなわち人間の言語能力を定義することは難しいが，ここでは，与えられた視覚情報に対して適切な音声言語情報を生成できるようになること，あるいは，与えられた音声言語情報が，視覚情報と適合しているか否かを判断できるようになること，と考える．このように考えた場合，言語獲得の問題は，以下のような問題に分解できる．

- 視覚情報および音声言語情報を分節化し，認識するための適切なカテゴリの獲得
- 獲得したカテゴリを使って視覚情報および音声言語情報を認識する能力の獲得
- 視覚認識結果と音声言語情報認識結果を関係づけ，相互に変換する能力の獲得

人間の言語獲得では，これらの学習が並行的に進行すると考えられるが，本発表では，このうちの 3 つ目の問題を取り上げる．視覚情報や音声言語情報を認識する能力は既に獲得されているとして，それらの認識結果を入力とし，その間の関係に関する知識である単語の意味と文法を学習する方法と，それに適した意味表現の構造について述べる．

2. 問題設定

最も簡単な言語獲得課題として，視覚（静止画）入力と，それに対応する音声言語入力から，相互の関係を学習するという課題について考える．具体的な状況としては，たとえば[Iwahashi 04]や[田口 07]で扱われているように，教示者がカメラの前に単独の物体（ボール，人形等）を提示し，その物体に関する発話を行うことを想定する．ただし，[Iwahashi 04]や[田口 07]では，発話されるのは「ボール」，「赤い」のような物体の名前や属性を表す単語であるが，以下では，「赤いボール」などの名詞句や，「ボールの色は赤い」などの文も発話されるものとする．

既に述べたように，視覚情報および音声言語情報の認識は終了し，それぞれの認識結果表現が得られているとする．従来の言語獲得システムの多くでは，視覚情報から色や形などの視覚特徴を数値を並べた特徴ベクトルの形で抽出し，その特徴ベクトルと，表層の言語表現との間の対応関係を直接的に学習しようとしているものが多い．すなわち，言語情報に内在する構造に関する情報は視覚情報認識結果の中にはほとんど存在せず，もっぱら，言語情報のみに基づいて学習されることを前提としている．

これに対して，我々は，語彙，係り受け関係，統語カテゴリなどの言語情報が持つ構造は，人間が世界を認知する際の構造を反映したものであると考える．すなわち，視覚などの感覚認識結果自体が，数値ベクトルではなく，より豊かな構造を持ち，表層の言語表現の構造はその構造を起源として，それを1次元の語の並びで表現するために発生したものである，と考える．

このように考えることは，大きく分けて二つの利点を持つ．まず，感覚認識結果の表現が，係り受け関係や，「名詞」や「動詞」などの表層表現における統語的カテゴリの種・起源となるような構造を持つことで，語彙や文法，特に文法の獲得が容易になることが期待される．次に，認識結果表現の構造が，ひとつの認識結果を，比較的簡単な変形等によって多様な表層表現と対応づけることを可能にするようなものであれば，システムが多様な言語表現を受理しつつ言語獲得を進めることが可能になると考えられる．このとき，言語学習の容易さは，認識結果表現の構造に大きく依存する．以下では，[高木 82, 87]において提案した意味表現構造が，語彙と文法の獲得に適していることを示す．

3. 認識結果表現＝意味表現の構造

Jackendoff は，視覚情報等の感覚情報と音声言語情報が結びつく場として，概念構造(Conceptual Structure)を提案し，その構造について検討するとともに，それが言語の意味構造に他ならないと主張している[Jackendoff 83, 02]．本研究でも同様の考えに従って，視覚情報等の感覚情報の認識結果表現が，すなわち，言語の意味表現であると考える．

自然言語の表層表現は自由度が高く，ほぼ同じ意味内容を多様な形で表現できる．従って，意味表現は，多様な表層表現の構造にあまり依らずに，類似の意味内容を持つ表層表現が類似の意味表現を持つようなものであることが望ましい．この点において，述語論理式による意味表現などは不適切である．また，文法の学習を容易にするためには，意味表現の構造に，表層表現における依存関係が，詳細かつ正確に反映されていることが望ましい．表層表現の同義変形を根拠として[高木 82, 87]で提案した意味表現はそうした条件を満たしている．以下では，上記の課題で必要とされる形容詞および名詞の意味表現を中心に，その概要を説明する．

連絡先：麻生英樹，産業技術総合研究所情報技術研究部門，
〒305-8568 茨城県つくば市梅園1－1－1中央第2，
h.asoh@aist.go.jp

3.1 基本的な考え方

「赤いボール」という名詞句は，たとえば「赤い色を持つボール」という句とほぼ同じ意味を持つ．ここで，二番目の表現における「した」はどこから生じたと考えればよいだろうか？「赤い」が，「赤い色を持つ」と言い換えられることから，「持つ」は「赤い」の意味に含まれている，と考えるのは自然である．しかしその一方で，「ものは色を持つ」というのはボールを含む「もの」一般に関する基本的な知識であるため，「持つ」は「ボール」の意味の中に含まれているとも考えられる．

[高木 82]では，日本語，英語，中国語等におけるこのような同義文に関する考察に基づき，「赤いボール」のように「赤い」が「ボール」を修飾できるのは，両者の意味表現に共通する情報が含まれており，それを糊しろとして，両者の意味表現を重ね合わせて接続することが可能であるためである，という考えを提案した．具体的には，「赤い」と「ボール」の意味表現をそれぞれ図 1 のようにすることを考案した．

図 1 では，「赤い」の意味が「赤に等しい色相を内包する色を内包するところの」であり，「ボール」の意味が「X に等しい色相を内包する色を内包するところのものであり，球に等しい形を内包するところのものであり，Y がそれを投げるところのものであり，・・・」である，ということが，それぞれ表現されている．この例からわかるとおり，形容詞も名詞も非常に豊かな内部構造を持ち，両者は，多くの部分を共有している．「赤い」が「ボール」を修飾している「赤いボール」という表層表現に対応する意味表現は，二つの語彙の意味表現を，共有部分を糊しろとして接続した図 2 のような構造となる．

3.2 意味表現に使われる記号と構成規則

上記の意味表現では，以下のような記号が使われている．

- 実体や属性を表す名詞的な概念構成素を○で表す
- 現象を表す動詞的な概念構成素を◎で表す
- 英語の関係代名詞相当の構成素を●で表す
- ある●が○を指示することを＝で示す
- 格概念を→（矢印）で表す

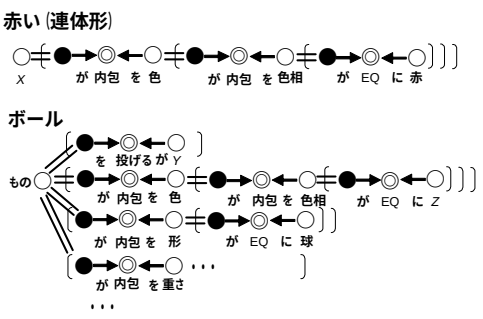

図 1 「赤い（連体形）」と「ボール」の意味表現

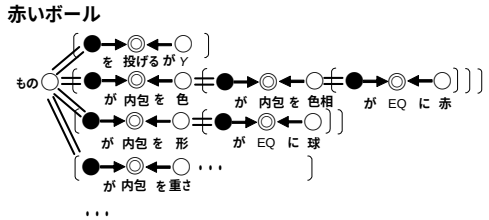

図 2 「赤いボール」の意味表現

- 不定な概念構成素名を示すのに変数 X, Y などを用いる．

これらの記号は，下記のような制約を満たすように接続される．

- ○，●，変数は単一の→の始点に接続することができる
- ◎は複数の→の終点に接続することができる．（ただし，→の種類によっては一つしか接続できない．）
- ○は複数の●と＝で接続することができる

図 1, 2 に示すように，意味構造は一般に，○や◎が→や＝で接続された木構造となる．木構造のルートをヘッドと呼ぶ．「ボール」のような実体を表す語の意味表現のヘッドは○であり，「持つ」のような現象を表す語の意味表現のヘッドは◎である．節に対応する意味構造のまとまりを（ ）によって表す．上述したように，表層表現において，語（あるいは文節）A が B を修飾しているときには，A の意味表現と B の意味表現は共有する構造を持ち，その部分を糊しろとして A の意味表現を B の意味表現のヘッドに接続することで全体の意味表現が形成される．

概念構成素として具体的にどのようなものを用意するかは，概念体系の設計による．以下，本稿の議論の範囲では次のような概念構成素を使用する．

- ○：もの，色，形，色相，など
- ◎：内包，EQ（値保持），など
- →：主格（が格），目的格（を格），対象格（に格），など

ただし，上記の意味表現の構造は，具体的な構成素として何を用いるかとは独立に，意味表現の構造および接続関係を規定していることに注意しておきたい．

[高木 82, 87]では，この意味表現を用いることで，「赤い色を持つボール」「色が赤いボール」などの多様な表層表現が，ほぼ同じ構造の意味表現を持つようになることを示している．

3.3 視覚特徴抽出結果との関係

視覚情報を処理して対象物体の特徴抽出をした結果は，それぞれの特徴に対応する数値を集めたベクトルとして表現されることが多い．たとえば[田口 07]では，色は RGB 表色系から HSV 表色系へ変換され，最終的に，明度と色相の 2 次元の数値で表されている．また，形は，高次局所自己相関特徴を用いて 25 次元の特徴ベクトルで表されている．

こうした特徴ベクトルによる認識結果の表現と，図 2 のような意味表現との間は遠く離れているように見えるが，図 2 の意味表現において，「○赤」，あるいは，「○球」，という記号が置かれているところに，それぞれ色相や形を現す数値ベクトルを代入することで，視覚特徴抽出結果から図 2 のような構造を持つ表現を容易に作ることができる（ただし，各○，◎，→につけられている「もの」や「内包」といったラベルは，ノードの種別を表すだけのものであり，この段階では特定の意味を持つわけではない）．

このことは，視覚情報を認識した結果に，「もの」が「色」という属性を内包し，「色」が「色相」という属性を内包している，というような，人間の世界認知の基本的な構造が埋め込まれている，と考えることに対応しており，図 2 の表現は，そのことを明示的に表したものになっている．これに対して，数値ベクトル表現は，本来図 2 のような構造を持つ認識結果表現を，極めて簡略化し，数値部分のみを取り出して並べたもの，と考えられる．

4. 語彙と表層の文法の獲得

上記の認識結果表現（＝意味表現）構造を用いて表現される視覚認識結果と，音声言語認識結果（文字列）とを付き合わせて，語彙と表層の文法を獲得してゆくプロセスについて述べる．前提として，人間は，認識結果表現の構造に関する豊富な知識を，言語の獲得に先立ち，生得的に持っていると仮定する．すなわち，認識結果表現が，要素からどのように構成されるか，ど

のように要素に分解され得るか，分解された要素間の依存関係はどのようなものになるか，等の知識が言語獲得に利用可能であると仮定する．また，Quine が「ガヴァガーイ問題」として指摘したように，ある情景に含まれる膨大な情報の中のどれに注目するかには，莫大な可能性があり，すべての可能性を考慮していては，現実的な数の例から語の意味を学習することは不可能である[Quine 60]．そこで，人間はいくつかの「認識の構え」を持ち，注意を向ける情報をバイアスを加えて選んでいると考えられている[Markman 89, 今井 08]．こうした認識の構えに関する知識も言語獲得に利用可能であると仮定する．さらに，同じ情報、現象に注意を向けた場合でも，それにかかわる複数の要素のうちのどれを中心として認識するかには複数の可能性がある．それによって認識結果表現の構造も変化すると仮定する．

言語表現は単語から構成されている．そこで，言語表現と視覚認識結果との関係を学習するためには，

- 言語認識結果および視覚認識結果の分解と，分解された部分同士の対応づけの学習（語彙学習に相当）
- 語順と係り受け関係の学習（表層の文法の学習に相当）
- 表層の統語的カテゴリの学習

を並行して進める必要がある．以下，それぞれについて基本的な考え方を述べる．

4.1 入力文と視覚認識結果の分解と対応づけの学習

「赤いボール」という入力文と視覚認識結果のペアが与えられたとする．「赤いボール」のうち，たとえば「赤い」が既知語で，それと対応する視覚認識表現とともに語彙辞書に登録されている場合には，その情報を使って，入力文を「赤い」と「ボール」に分解するとともに，認識結果表現をも分解することができる．その結果，未知語であった「ボール」とそれに対応する認識結果表現の部分が得られるので，それを語彙辞書に登録すればよい．一方，「赤い」も「ボール」も未知語である場合には，とりあえず，「赤いボール」全体を一つの単語と仮定し，認識結果表現とあわせて語彙辞書に登録するしかない．しかし，こうした語を次第に単語に分解してゆかないと，システムは柔軟な言語能力を獲得できない．

「青いボール」のように，登録してある単語と共通部分を持つ入力が与えられたときが，分解を行なうチャンスとなる．対応する視覚認識結果を比較して，異なっている部分を変数化し，共通の単語「ボール」の意味とする．また，「青い」「赤い」などは変数化する要素を持つ枝の情報を記述する単語であると考えられるため，その枝を根本から分離してそれらの語の意味とし，単語と意味のペアを語彙辞書に登録する．変数を含む枝が複数ある場合には，可能な候補を語彙候補として保持する．

このような単語分解，意味表現分解，それぞれの分解結果の対応づけ，をくり返すことで，語彙項目を含む語彙辞書を獲得することが可能になると考えられる．ただし，実際には，未知語に対応する視覚認識結果表現の部分構造を同定することは，特に，手がかりの少ない言語獲得の初期においては，高度な推論能力や試行錯誤を必要とする難しい問題であり，複数の入力事例を通じて，安定的な分解と対応づけが獲得されていくと考えられる．

4.2 語順と係り受け関係の学習

「赤いボール」が「赤い」と「ボール」に分解できることがわかれば，「赤い」→「ボール」という語順を学習することができる．また，それぞれの意味構造の形体から，「赤い」の意味が「ボール」の意味を修飾している，ということもわかる．このようにして，語順と係り受け関係に関する知識を得ることができる．

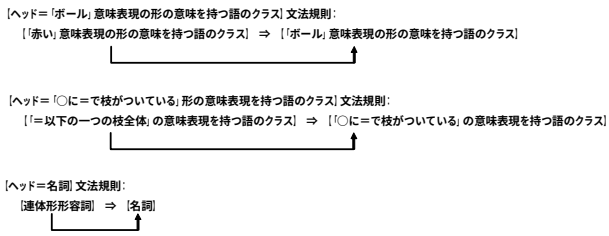

図 3 文法規則の獲得の例

しかしながら，表層の単語を用いて知識を記述していては，得られる規則に汎用性が無いため，莫大な数の規則を学習することが必要になり，学習に非常に多くの時間と事例を必要とする．そこで，同じような機能と意味構造を持つ単語を集めた統語カテゴリを学習し，カテゴリを使って知識を表現することが必要になる．

4.3 統語的カテゴリの生成

語彙辞書の項目が蓄積されるにつれて，同じような意味構造を持つ語が存在することが見えてくる．たとえば，「ボール」と「箱」は，いずれも○をヘッドとして持ち，その下に＝で始まる関係節構造が複数接続する構造を持つ．また，語順と係り受け関係の知識が蓄積されるにつれて，同型の構造をもつ語同士は，語順や係り受け関係においても，類似の機能を持つことも見えてくる．このような，意味構造パタン，語順における性質，語の活用の性質，などを総合的に統合して抽象化を行い，語をカテゴライズしてゆく．この結果として，最終的に，名詞や形容詞連体形などに相当する統語的カテゴリを学習することができる．

こうした統語的カテゴリの生成と並行して，語順や係り受けに関する規則もカテゴリを使って記述するように抽象化してゆけば，ある語が特定の統語カテゴリに属していることがわかった瞬間に，そのカテゴリが関与する多くの知識が適用可能になるため，学習が効率化されると期待できる．文法規則の獲得過程の様子を図 3 に示した．図の⇒は語順を表し，→は係り受け関係を表している．

5. 関連研究

近年，人間の言語能力の解明を目指して，あるいは，自然言語による情報機器やロボットとのインタフェースの実現を目指して，ロボットや知的エージェントに対話的に言語を教示する研究が盛んになっている[赤穂 97, Roy 02a, Roy 02b, 岩橋 03, Iwahashi 04, Yu 04, 新田 05, 小島 05, 田口 07]．しかし，それらの多くは，視覚情報や音声言語情報の分節化とカテゴリ化の問題を主たる対象とし，コップや人形のような独立した物体を提示しながら「ボール」，「赤」，のようにその名前や属性を発話し，ものや属性の概念を学習させる課題を扱うものである．

[岩橋 03]では，複数の物体が関与するような物体の動きを提示しながら，それに対応する文を発話し，現象の概念と文法を同時に学習させる課題も取り上げているが，そこで使われている言語表現は「緑 大きい カーミット 青い 箱 乗せる」のような人工言語に近い定型的な文である．一方，[Roy 02a]や[Yu 04]では，実際の人間と幼児の間の対話文や，人間同士の対話文が音声言語入力として使われているが，いずれも，単語の意味の学習を課題としており，文法の学習は行っていない．[Roy 02b]では，物体とその位置関係の記述を対象として，語彙だけでなく，統語的カテゴリや文法も学習させているが，そこで獲得されているものは統計的な単語バイグラムである．[Jackendoff 83, 02]では，[Marr 82]なども考慮しつつ，感覚情報の認識結果

と言語情報の関係について考察し, それらを付き合わせる場として, 概念構造が提案されているが, それを用いた言語獲得については論じられていない.

視覚情報や音声言語情報の認識が終わっていることを前提として, 語彙と文法の学習をモデル化している研究としては, 古くは, 言語表現とその意味を表す意味ネットワークを入力として ATN の文法を学習するシステム LAS[Anderson 77], Schank による概念依存構造を意味表現として用いた言語獲得システム CHILD[Selfridge 83], 入力文と意味役割り記述から決定的パーザの構文解析規則を逐次的に学習するモデル[Berwick 85], などがよく知られている.

最近の研究としては, 視覚認識結果の記述と文とから, 語の意味に相当するフィルタおよび統語規則, 統語的カテゴリを獲得するシステム Rhea[錦見 92], 言語とその意味表現とから, 語彙と文法を獲得するダイナミカルシステム CAMiLLe[Culicover 03]などがあげられる. しかし, これらの研究においても, 視覚認識結果の表現や意味表現は比較的単純なものであり, 我々のように言語構造の源となるような十分に豊かで柔軟な内部構造を持つわけではない. その結果として, 獲得されている文法知識も比較的単純なものに限られている.

6. おわりに

言語獲得に適した認識結果表現＝意味表現の構造を提案し, それを用いた文法と語彙の獲得手続きを示した. 具体例として, 単独物体の名前や属性に関する言語表現の語彙や文法が獲得できることを示した.

ここで述べた言語獲得モデルの最大の特徴は, 視覚認識結果の表現が豊かな構造を持つことであり, 表層言語表現の文法構造は, そうした認識結果表現の文法を源とし, それを写し取ることによって形成されている, と考える点である. このことが, 比較的少数の言語情報から安定的に, 語彙や文法を獲得することを可能にしていると考えられる.

今後の課題としては, まず, 今回提案した言語獲得プロセスを実際のデータに適用して, その性能を評価することがあげられる. また, 今回検討したのは, 単独の物体の静止画を入力として, その物体の属性について言及する場合だけだが, [高木 87]では, より複雑な, 複数の物体が位置関係などを変化させてゆく場合などについても, 同様の仕組みでの言語獲得が可能であることを示している. こうした場合についての性能評価も今後の課題である. さらに, 豊かな構造を持つ認識結果表現が, 進化の過程でどのように発生したのか, も興味深い問題である.

参考文献

[赤穂 97] 赤穂, 長谷川, 吉村, 麻生, 速水: 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, No.9, pp.1546-1553, 1997.

[Anderson 77] Anderson, J. R.: Induction of augmented transition networks, Cognitive Science, Vol.1, pp.125-157, 1977.

[Berwick 85] Berwick, R. C.: The Acquisition of Syntactic Knowledge, MIT Press, 1985.

[Culicover 03] Coulicover, P. W., Nowak, A.: Computational simulation of language acquisition: CAMiLLe, in "Dynamical Grammar", Oxford University Press, 2003.

[今井 08] 今井, 針生: レキシコンの構築:子どもはどのように語と概念を学んでいくのか, 岩波書店, 2008.

[岩橋 03] 岩橋直人: ロボットによる言語獲得: 言語処理の新しいパラダムを目指して, 人工知能学会誌, Vol.18, No.1, pp.49-58, 2003.

[Iwahashi 04] Iwahashi N.: Active and unsupervised learning for spoken word acquisition through multimodal interface, Proceedings of 13th IEEE Workshop Robot and Human Interactive Communication, pp.437-442, 2004.

[Jackendoff 83] Jackendoff, R.: Semantics and Cognition, MIT Press, 1983.

[Jackendoff 02] Jackendoff, R.: Foundations of Language, Oxford Univ. Press, 2002（郡司訳: 言語の基盤, 岩波書店, 2005）.

[小島 05] 小島, 長谷川: 自己増殖型ニューラルネットを用いたヒューマノイドロボット上の発達的シンボルグラウンディング, 人工知能学会第 19 回全国大会, 1B3-03, 2005.

[Markman 87] Markman, E. M.: Categorization and Naming in Children, MIT Press, 1987.

[Marr 82] Marr, D.: Vision, W.H. Freeman and Co., 1982 (乾, 安藤訳: ビジョン－視覚の計算理論と脳内表現, 産業図書, 1987).

[錦見 92] 錦見, 中島, 松原: 一般学習機構を用いた言語獲得の計算機モデル, 認知科学の発展, Vol.5, 1992.

[新田 05] 新田, 小玉, 田口, 木村, 桂田: 生得的学習バイアスを適用した Infant Agent による概念獲得, 情報処理学会研究報告, 2005-SLP-69, pp 69-74, 2005.

[Quine 60] Quine, W. V.: Word and Object, MIT Press, 1960.

[Roy 02a] Roy, D. K., Pentland, A..: Learning words from sights and sounds: A computational model, Cognitive Science, Vol.26, No.2, pp.113-146, 2002.

[Roy 02b] Roy, D. K.: Learning visually-grounded words and syntax for a scene description task, Computer Speech and Language, Vol.16, No.3, 2002.

[Selfridge 86] Selfridge, M.: A computer model of child language learning, Artificial Intelligence, Vol.29, pp.171-216, 1986.

[田口 07] 田口, 木村, 小玉, 篠原, 入部, 桂田, 新田: 幼児の学習バイアスを利用したエージェントによる語意学習の効率化, 人工知能学会論文誌, Vol.22, No.4, pp.444-453, 2007.

[高木 82] 高木, 小原: 属性形容詞の意味構造－意味表現方法の一つの試み－, 電子情報通信学会論文誌 D, Vol.J65-D, No.11, pp.1427-1434, 1982.

[高木 85] 高木, 芦沢, 太田, 蔦田, 伊東, 小原: 簡単な日本語の意味と文法を学習するシステム, 情報処理学会研究報告, ICS Vol.1984, No.74, pp.17-24, 1985.

[高木 87] 高木, 伊東: 自然言語の処理, 丸善, 1987.

[Yu 04] Yu, C. and Ballard, D.: On the integration of grounding language and learning objects, 19th National Conference on Artificial Intelligence, 2004.