

G-26

人間-エージェント間対話による画像・音声刺激からの
言語シンボル概念獲得に関する研究Knowledge Acquisition Method of Linguistic Symbol Through
Human-Computer Dialogues Using Image and Voice Stimuli

知久 慎太郎†

Shintaro CHIKU

荒井 秀一†

Shuichi ARAI

宮内 新†

Arata MIYAUCHI

1. あらまし

近年、計算機上に言語概念を構築する研究が多数行われている。これらの多くはシンボルの意味や関係を前提知識として与え、その上で概念の獲得を行うシンボリックな処理を行っている。

一般的に概念とは、経験される多くの事物に共通の内容を抽出して抽象化を行い、個々の事物にのみ属する偶然的な性質を捨て、捨象することである。一方言語シンボルとは、概念を指し示すラベルであり概念を象徴化したものであると言える。このことから言語シンボルは概念に先立って形成されるものではなく、後天的に獲得されるものであるため、言語概念を獲得する際に言語シンボルを用いる事はシンボルの変換を行っているにすぎず、言語概念を獲得しているとは言えない。

本研究では既に、シンボリック処理を行わずに外部からの刺激である画像・音声刺激と、その提示同時性により言語シンボル概念を獲得するモデルを提案している [1]。

また、獲得した概念は逐次的に学習していくのだが、一般的な教示による学習を用いた場合、人間と計算機の間インタラクションが無い場合、学習者の知識獲得度合を知るには認識テスト等によるしかなかった。即ち、獲得される概念が学習者独自の基準によるものであり、他者と共有できる概念であるかは確証が無い。このような偏った概念を共通化するには、外部とのインタラクションを通して自分の持つ概念の修正や拡張を行い、概念の擦り合わせを行う必要がある。そこで、本研究では獲得した概念を具象化し、人間-計算機間のインタラクションを行うことにより、学習者の知識獲得状況を把握しながら学習を行うことが可能な対話学習を用い、概念獲得モデルを実現している [2][3]。

上記のモデルにおいて、計算機が獲得している音声知識と画像知識との関係と、画像が表現できる構造・形状情報を用いることにより、名詞、形容詞にあたる概念が構築できていることを定量的に評価している。

2. 言語シンボル概念

本研究において獲得を目指している言語シンボル概念について、図1の概念図を元に説明する。

まず、外部から入力された画像・音声刺激を各々抽象化し、特徴量の分布で表現する。この画像・音声それぞれの特徴量分布を本研究では画像概念、音声概念と呼ぶこととする。

本研究では、同時に与えられた刺激には何らかの関係があるという提示同時性を利用することにより画像概念と音声概念との間にリンクを張り、そのリンクも含めた画像・音声概念をまとめて言語シンボル概念と定義する。

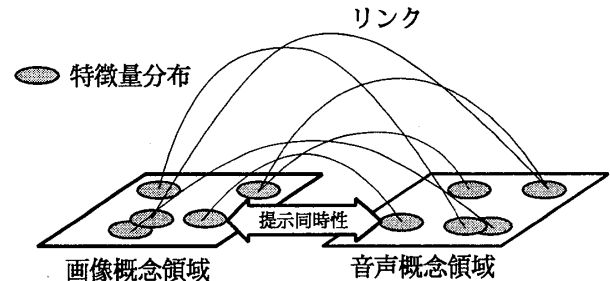


図1: 提示同時性による言語シンボル概念の表現

また、獲得した知識は入力刺激を認識して判断し逐次学習を行うので、再利用が可能でなければならない。さらに、完全に同一でなくても類似した刺激を同一な刺激であると思わせる表現形態でなければならないので、このような条件を満たす知識の表現形態として、音声認識の分野で広く用いられている HMM (Hidden Markov Model) を画像・音声それぞれに適応して概念を表現することとする。

3. 言語シンボル概念獲得のモデル化

ここで、言語的知識を用いない、画像・音声情報からの言語シンボル概念獲得のモデル化について述べる。

3.1 入力刺激

本研究における概念獲得モデルにおいては、シンボリックな情報は用いない。そこで、明示的な意味を持たない画像情報、音声情報のみを用いて概念を獲得する。画像情報は、物体の形状、構造特徴を扱うこととし、これらを十分に表現できる画像として、線画像を用いる。音声情報は、物体、形状、構造の概念を表現する単語音声を用いることとする。

3.2 概念獲得の方法

本研究では言語シンボルの意味の獲得を目的としているため、前提知識として言語的情報を利用することはできない。そこで、実際に言葉の意味をゼロから獲得している幼児が、先天的に持っている能力として、発話理解の能力、発話内容の学習能力、発話の能力は、前提知識として与えることとした。

3.3 概念獲得のモデル化

ここで、入力刺激の抽象化、認識・学習、概念の具象化のモデル化について図2を元に説明する。

(a) 抽象化

同時に提示された画像・音声刺激を各々抽象化し特徴量を算出する。

(b) 画像刺激及び音声刺激の認識

獲得されている各概念と入力刺激との比較を行う。

† 武蔵工業大学 Musashi Institute of Technology

ここで、構造的に矛盾がなく形状が似ている場合も適合するものとして認識結果の候補とする。そして、複数の候補の序列を定めるため、受容刺激の特徴量と知識との距離から確信度を求める。

- (c) 全体の認識
確信度によって序列が定まった候補群を、次に画像・音声間のリンク情報を用いてその組合せを見ながら総合的な判断を行う。
- (d) 学習
確信度を元に認識を行い、明確に判断されたもののみを学習することとした。この学習は、刺激そのものの学習と、画像・音声間のリンクの学習であり、既存のリンクが存在しない場合には新たなリンクを張り、既知のリンクに対しては強化学習を行う。
- (e) 発話
獲得した概念を具象化し、具体的な情報である、画像・音声刺激として表象する。

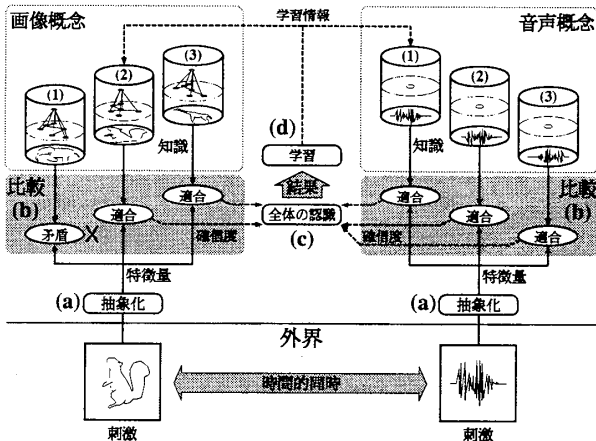


図 2: 概念獲得における情報の流れ

4. 部位・形状に関する概念の構築

獲得された言語シンボル概念を利用し、ある一つの音声概念で関連付けられた画像概念群の構造・形状情報から、形容詞・名詞にあたるカテゴリに分類できるかを実験にて評価する。

4.1 部位概念の構築

ゾウ, リス, キリン, ウサギの画像物体における形状を表現する音声刺激と部位を表現する音声刺激を教示し

表 1: 音声概念間の接続確率 (%)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	-	1.0	0	0	0	0	0	0	0	0	0	0	0
2	1.0	-	1.0	1.0	0	0	0	0	0.7	0.4	0.7	0.7	0
3	0	1.0	-	0	0	0	0	0	0	0	0	0	0
4	0	1.0	0	-	0	0	0	1.0	0	0	0	0	0
5	0	0	0	0	-	0	0	1.0	0	0	0	0	0
6	0	0	0	0	0	-	0	1.0	0	0	0	0	0
7	0	0	0	0	0	0	-	1.0	0	0	0	0	0
8	0	0	0	1.0	1.0	1.0	1.0	-	0.3	0.7	1.0	1.0	1.0
9	0	0.7	0	0	0	0	0	0.3	-	0	0	0	0
10	0	0.4	0	0	0	0	0	0.7	0	-	0	0	0
11	0	0.7	0	0	0	0	0	1.0	0	0	-	0	0
12	0	0.7	0	0	0	0	0	1.0	0	0	0	-	0
13	0	0	0	0	0	0	0	1.0	0	0	0	0	-

上段, 左辺の番号は以下の音声概念に対応(1. ミミ, 2. カオ, 3. ハナ, 4. クビ, 5. マエアシ, 6. ウシロアシ, 7. シッポ, 8. ドウタイ, 9. マルイ, 10. ナガイ, 11. ミジカイ, 12. フトイ, 13. ホソイ)

たものを用意し、そのデータを知識とした際の、獲得されている各音声概念の表現する部位概念の構築を行った。

実験結果として各音声概念の部位概念を表現する接続確率マトリックスを表 1 に示す。

これより、ドウタイには、クビ、マエアシ、ウシロアシ、シッポが、カオには、ミミ、ハナ、クビというように、人間が意図した部位に関する概念を獲得できていることが確認できた。しかしながら、ホソイ、フトイ、ミジカイといった人間が形状を示すものとして教示した音声もドウタイと高い接続確率を示してしまった。これは、この実験における知識状態では判断できなかったことを意味しているが、さらに教示を進めていくことで判断できると考えられる。

この結果から、知識内全ての音声概念群の中から部位を表現する音声概念を帰納的に選出が可能であることが確認できた。

4.2 形状概念の構築

キリン, リス, ウサギ, カンガルー, ソウ, それぞれ 4 姿勢の画像物体を部位に区切り、カオ, ハナ, クビ, マエアシ, ウシロアシ, シッポ, ミミ, ドウタイ, ミジカイ, ナガイ, フトイ, ホソイ, マルイと言う音声刺激と共に学習させた。そしてそのデータを知識としたとき、各音声概念が表現している形状特徴を示す概念を構築させる実験を行った。

これによって構築された各音声刺激が示す形状依存度を表 2 に示す。

表 2: 音声概念が表現する形状概念

音声概念	形状依存度		
ハナ	7.00		
クビ	9.26		
マエアシ	4.07		
ウシロアシ	6.38	ナガイ	6.44
シッポ	3.91	ミジカイ	6.51
ミミ	5.12	フトイ	21.74
ドウタイ	14.55	ホソイ	8.76
カオ	4.57	マルイ	24.61

この結果より、フトイ、マルイから成る形状を表現すべき音声概念の形状依存度は大きく、シッポ、マエアシ、カオなど、そうでないものは小さくなっていることが分かる。これより各々の音声概念に対してもっともらしい形状概念の構築が行えた。

参考文献

- [1] 荒井: “形状・構造に関する言語シンボルの概念を学習により獲得するための画像一言語知識表現法”, 信学論, D-II Vol. J80-D-II No. 9, (1997-09)
- [2] 伊与部, 荒井, 石川: “画像・音声情報による言語概念獲得のための人間-計算機間対話に関する研究”, 信学技報, AI99-63, (1999-11)
- [3] 黒田, 荒井, 石川: “画像・音声メディアを用いた人間-エージェント間対話による言語シンボル概念の獲得”, 電子情報通信学会, SAA2000 講演論文集, pp.297-304, (2000-11)