

## SIGVerse を用いた語彙学習シミュレーション

- 空間的概念の学習実験と語彙学習機能の印象評価実験 -

## Simulations of Vocabulary Learning by Robots using SIGVerse

- Learning Relative Location Concepts and Effects of Vocabulary Learning Mechanism on User Impression -

田口 亮\*<sup>1</sup>  
Ryo Taguchi東 拓実\*<sup>1</sup>  
Takumi Azuma沢登 京介\*<sup>1</sup>  
Kynosuke Sawanobori保黒 政大\*<sup>2</sup>  
Masahiro Hoguro梅崎 太造\*<sup>1</sup>  
Taizo Umezaki\*<sup>1</sup> 名古屋工業大学  
Nagoya Institute of Technology\*<sup>2</sup> 中部大学  
Chubu University

This paper describes two experiments of vocabulary learning by robots using SIGVerse that is a simulator developed by National Institute of Informatics. In the first experiment, we developed a method for learning relative location concepts and evaluated its effectiveness through a simulation. In the second experiment, we investigated effects of vocabulary learning mechanism on user impression by watching human-robot interaction on the simulator.

## 1. はじめに

近年、単語の意味をロボットの感覚・運動系の信号と対応付けて学習させる研究が行われ始めた[Araki 12]。ロボットによる語彙学習の研究は、人間の知能を構成論的に明らかにするという科学的な側面と、実環境で動作する音声対話ロボットを実現するという工学的な側面を合わせ持つ重要なテーマである。本稿では、国立情報学研究所で開発された社会的知能発生学シミュレータ SIGVerse[稲邑 12]を用いて実施した二つの語彙学習シミュレーション実験について報告する。一つ目の実験では、場所の名前や、相対的な位置関係を表す単語の学習機構を開発し、その有効性を検証した。二つ目の実験では、シミュレータ上で被験者とロボットの対話実験を行い、ロボットの語彙学習機能がコミュニケーションに与える影響について調査した。

## 2. 相対的な位置概念の学習実験

ロボットの語彙学習に関する先行研究の多くは、単語とセンサ情報の共起関係を、その単語の意味(本稿ではこれを概念と呼ぶ)として直接的に学習している。しかし、色や大きさ、位置などの概念は、他の物体や概念との比較により表される。基準となる物体や概念を参照点と呼ぶ。このような概念を学習するには、複数の候補から真の参照点を推定しながら学習する必要がある。我々は EM アルゴリズムを用いて参照点の推定と概念の学習を同時に行う手法を提案した[田口 13]。しかし、参照点の候補が多数ある場合に、参照点の推定を誤るという問題や、概念の解釈が特定の座標系に固定されており、学習できない概念が存在した。本研究では事前知識により参照点の推定誤りを軽減する手法、および複数の座標系候補から適切な座標系を推定する手法を提案する。さらに、被験者実験を通して位置概念の使用実態を調査し、提案手法の妥当性を評価する。

## 2.1 相対的な概念

例えば「大きい」という概念を、図 1 のようにカメラに映る画素数の分布として学習する。その際、アリと象に対して「大きい」と

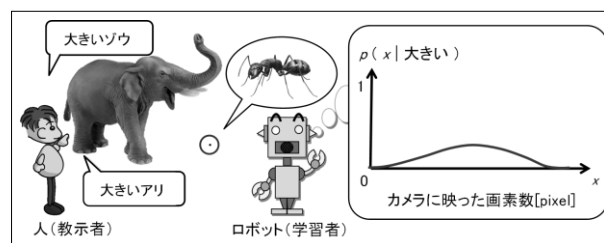
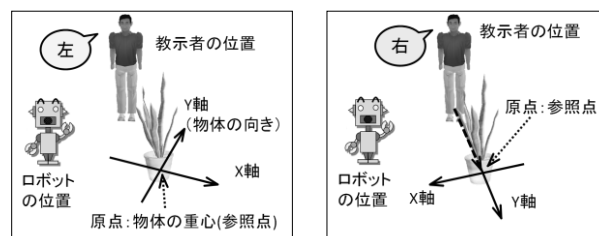


図1 参照点により言葉の意味が変わる例



(a) 参照点の向きがY軸の正 (b) 教師の向きがY軸の正

図 2 座標系により相対位置は変化する

教示されると、画素数が少ない場合(アリ)も、多い場合(象)も「大きい」と呼ばれるため、学習後の分布が広くなり「大きい」と画素数との関係が表現できない。実際には、人は象やアリの平均的な大きさ(画素数)と比べて「大きい」と表現しているのであり、画素数の絶対量に対して表現しているのではない。このように相対的な概念は、参照点により意味が変わるため、これを学習するためには参照点を推定する必要がある。

また、相対的な概念は座標系にも依存する。例えば、図 2 に示すように、話者やロボット、参照点となる物体の位置関係が同じでも、人が発話する際に想定した座標系により、ロボットの位置を表す単語が、「(植物の正面に対して)左」や「(教師者の位置から見て植物の)右」と変化する。従って、相対的な概念を適切に学習するには、参照点だけでなく、教示の際に使用された座標系も推定する必要がある。

## 2.2 相対的な位置概念の学習手法

ロボットが現在いる位置の名前をユーザが教示する。ここで、ロボットの位置を表す 2 次元座標を  $X^n = (x^n, y^n)$  とする。  $n$  は

学習サンプルのインデックスであり、 $1 \leq n \leq N$ である。また、座標系を $c$ とし、参照点 $k$ の重心を原点とした座標系におけるロボットの相対位置を $X_{c,k}^n$ と表す。参照点および座標系の候補は既知とするが、真の参照点と座標系は未知とする。

与えられた名前と相対位置の対応関係を平均 $\mu_c$ 、分散 $\sigma_c^2$ の正規分布 $N(\mu_c, \sigma_c^2)$ でモデル化する。学習サンプル $n$ において座標系を $c$ としたとき、参照点 $k$ が真の参照点である確率を $\pi_{c,k}^n$ とおく。また、サンプル $n$ における参照点候補の数を $M_n$ とする。各サンプルにおける参照点の確率の集合は $\pi_c^n = (\pi_{c,1}^n, \dots, \pi_{c,k}^n, \dots, \pi_{c,M_n}^n)$ となる。また、全参照点の確率の集合は $\pi_c = (\pi_c^1, \dots, \pi_c^n, \dots, \pi_c^N)$ となる。求めるモデルのパラメータは $\theta_c = (\mu_c, \sigma_c^2, \pi_c)$ となる。モデルの尤度 $L(\theta_c)$ を式(1)のように定義し、これを最大化するパラメータ $\hat{\theta}_c$ をEMアルゴリズムにより求める。

$$L(\theta_c) = \sum_{n=1}^N \ln \sum_{k=1}^{M_n} \pi_{c,k}^n \cdot N(X_{c,k}^n; \mu_c, \sigma_c^2) \quad (1)$$

ただし、本研究では簡単化のため、多次元正規分布の各次元を独立とみなし、式(2)のように近似する。

$$N(X_{c,k}^n; \mu_c, \sigma_c^2) \approx N(x_{c,k}^n; \mu_{c,x}, \sigma_{c,x}^2) \cdot N(y_{c,k}^n; \mu_{c,y}, \sigma_{c,y}^2) \quad (2)$$

EMアルゴリズムにより各座標系でモデルパラメータ $\hat{\theta}_c$ を求め、その後、モデル尤度が最大となる座標系を真の座標系 $\hat{c}$ とする。

$$\hat{c} = \arg \max_c L(\hat{\theta}_c) \quad (3)$$

EMアルゴリズムで用いる各パラメータの更新式は式(4)~(6)となる。それぞれ、式(4)はモデルの平均、式(5)はモデルの分散、式(6)は参照点の確率の更新を行う式である。

$$\mu_c = \frac{1}{N} \sum_{n=1}^N \frac{\sum_{k=1}^{M_n} \pi_{c,k}^n \cdot N(X_{c,k}^n; \mu_c, \sigma_c^2) \cdot X_{c,k}^n}{\sum_{k=1}^{M_n} \pi_{c,k}^n \cdot N(X_{c,k}^n; \mu_c, \sigma_c^2)} \quad (4)$$

$$\sigma_c^2 = \frac{1}{N} \sum_{n=1}^N \frac{\sum_{k=1}^{M_n} \pi_{c,k}^n \cdot N(X_{c,k}^n; \mu_c, \sigma_c^2) \cdot (X_{c,k}^n - \mu_c)^2}{\sum_{k=1}^{M_n} \pi_{c,k}^n \cdot N(X_{c,k}^n; \mu_c, \sigma_c^2)} \quad (5)$$

$$\pi_{c,k}^n = \frac{\pi_{c,k}^n \cdot N(X_{c,k}^n; \mu_c, \sigma_c^2)}{\sum_{k=1}^{M_n} \pi_{c,k}^n \cdot N(X_{c,k}^n; \mu_c, \sigma_c^2)} \quad (6)$$

参照点の確率 $\pi_{c,k}^n$ の初期値を設定する方法として、ランダムに値を設定する方法や、参照点の全候補を等確率に設定する方法が考えられる。しかし、実世界においては参照点の候補は無数にあり、それら全てを考慮することは現実的ではない。そこで、事前知識を用いて、参照点の候補を絞り込むことを考える。

人の注意の傾向性の1つとして、「近くにあるものほど注意が向きやすい」というヒューリスティクスをロボットに与える。具体的には式(7)を用いて、参照点の確率の初期値を設定する。

$$\pi_{c,k}^n = \frac{1/(d_k^n)^\alpha}{\sum_k 1/(d_k^n)^\alpha} \quad (7)$$

ここで、 $d_k^n$ はロボットから参照点候補 $k$ までの距離であり、 $\alpha$ は傾向性の強さを調整するためのパラメータである。

## 2.3 評価実験

### (1) 実験条件

学習の対象は相対的な位置を表す概念とし、「前、後、左、右、手前、向こう側」の6種類を教示する。ユーザは人型のアバターをコントローラで操作し、その後ろをロボットが追従する。ユーザは任意の位置で場所名をテキストで教示する(図3)。教示

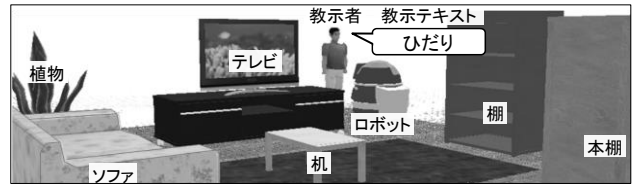


図3 教示の様子

表1 学習により得られた分布の2σ楕円

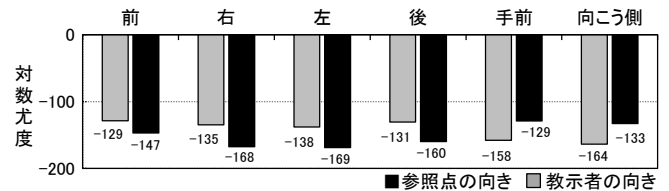
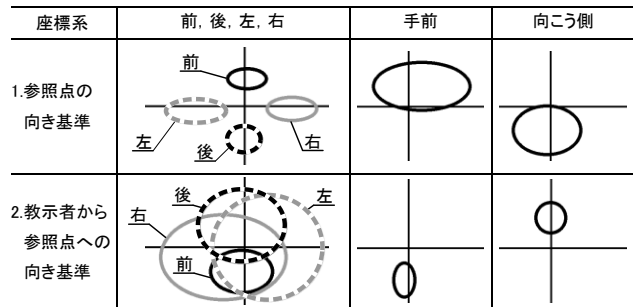


図4 学習により得られた各単語のモデル対数尤度

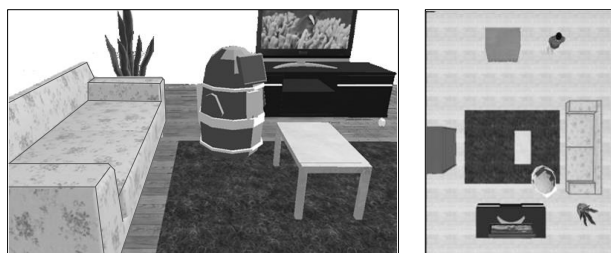
したテキストはロボットの位置座標、および人型のアバターの位置座標とともに保存される。

実験では「本棚」「テレビ」「ソファ」の3つのオブジェクトを、場所名を教示する際の参照点とする。参照点候補は「植物」「机」「棚」を加えた6つのオブジェクトとし、各オブジェクトの重心座標および向き(正面方向)は既知とした。また、座標系は図2に示す2種類とした。座標系1は、参照点の向きに基づいた座標系(図2(a))であり、参照点となる物体の重心を原点、参照点の正面方向をY軸の正、Y軸に直交する右方向をX軸の正とする右手系の座標系である。座標系2は、教示者の向きに基づいた座標系(図2(b))であり、参照点となる物体の重心を原点とし、教示者から参照点への向きをY軸の正、Y軸に直交する右方向をX軸の正とする右手系の座標系である。

実験では、参照点と相対位置を表す単語の組み合わせ(3×6=18通り)毎にロボットおよび人型アバターの位置を変えながら5箇所、合計90箇所を教示する。ただし、「前、後、左、右」は座標系1を用いて教示し、「手前、向こう側」は座標系2を用いて教示する。教示する座標系は単語単位で固定するが、どちらの座標系を用いたかは未知とする。

### (2) 実験結果と考察

表1に学習により得られた各単語の概念を表す相対位置分布を示す。図中の横棒はX軸(右が正)、縦棒はY軸(上が正)であり、原点は任意の参照点の重心となる。学習された正規分布を、平均を中心とする $2\sigma$ の楕円で表示した。図の上半分は座標系1で得られた各単語の学習結果、下半分は座標系2で得られた各単語の学習結果である。座標系1では「前、後、左、右」の分布が、座標系2では「手前、向こう側」の分布が、それぞれ単語の意味する適切な位置に収束している。



(a) 人型アバターの視界 (b) 俯瞰図  
図5 被験者実験に使用した画像の例

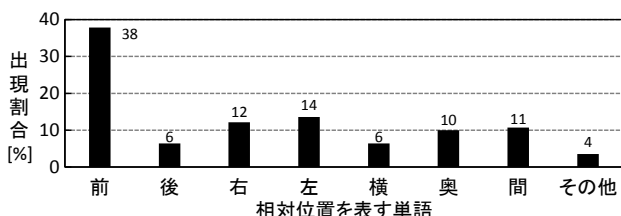


図6 被験者実験での相対位置を表す単語の出現割合

図 4 に学習により得られたモデルの対数尤度を示す。「前、後、左、右」は座標系 1 で学習したモデルの尤度が高く、「手前、向こう側」は座標系 2 で学習したモデルの尤度が高くなることからわかる。これよりモデルの尤度を用いて、複数の座標系候補から適切な座標系を推定可能と言えらる。

## 2.4 被験者実験による位置概念の使用実態調査

ユーザが位置の説明を行う際に用いる語句を、被験者実験により調べる。SIGVerse により作成したシミュレーション環境を用いて、図 5 のような画像を 20 種類作成した。画像はシミュレーション環境上の人型アバターの視界である。画像中には必ずロボットといずれかの物体が映る。20 枚の画像を 7 名の被験者に見せてロボットの位置を説明させ、使われた語句を記録した。記録した 140 発話の語句のうち、相対位置を表す概念を持つ単語の出現割合を、図 6 にまとめた。本研究で学習した 6 単語のうち、「前、後、右、左」は 70[%] の割合で使用されており、「手前、向こう側」は使用されていない。ただし、「前」という単語の一部は、本研究で学習した「手前」と同じ意味で使用されており、「奥」についても同様に、本研究での「向こう側」と同じ意味で使用されていた。「奥」については、「向こう側」の言い換えであるため、学習は可能である。「前」については、1 人の話者でも複数の座標系が混在して使用されているため、これに対応した学習が必要である。「横」については、本研究の手法では参照点が正しく推定できれば、左右に広い分布として学習は可能である。「間」については、参照点を複数持つ概念であるため、学習手法を改善する必要がある。

## 3. ヒューマン・ロボットコミュニケーションにおける語彙学習機構の有用性

ロボットの語彙学習に関する先行研究では、効率性や機能の側面から語彙学習の有用性が議論されてきた。しかし、語彙学習機構がロボットとの信頼性構築やコミュニケーションそのものを活性化させるという側面についての議論は少ない。人間とロボットの関係性に着目した研究で塩見らは、科学館において、ロボットが来訪者の名前を呼ぶ行動により、好印象を抱き易くなることを示した[塩見 06]。しかし、人間がロボットに名前を学習させることで好印象を感じるようになるかはまだ議論されていない。

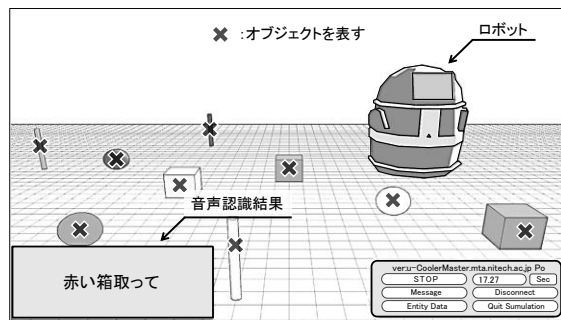


図7 第3章の実験に用いたシミュレーションの実行画面

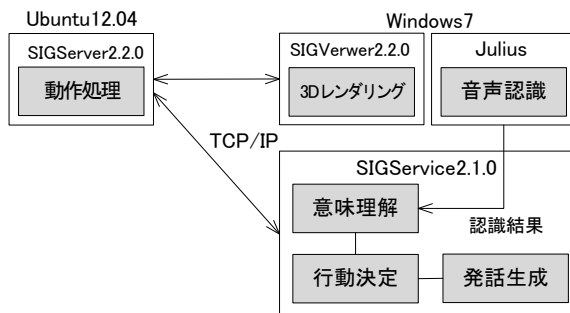


図8 ソフトウェアモジュール構成図

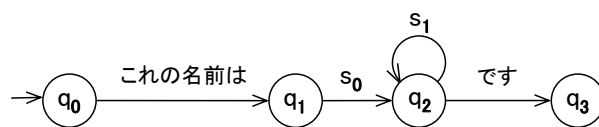


図9 単語学習用の文法

そこで本実験では、人がロボットに物体の名前を学習させることで、ロボットに対する印象がどのように変化するかを調査する。

### 3.1 実験システム

実験では、ユーザとロボットがシミュレータ上で音声インタラクションを行う。環境内にはロボットの他に 9 つのオブジェクトが配置されている。作成した実験環境の実行例を図 7 に、ソフトウェアモジュール構成図を図 8 に示す。なお、シミュレーションを行う際に用いた PC は、シミュレーションの物理演算を行う PC1 台、3Dレンダリングや、音声認識、ロボットの行動決定、発話処理等を行う PC2 台の計 2 台である。

ロボットの動作はユーザの発話によりすべて受動的に決定して、ユーザの発話に応じて①指示された方向へ向く、②指示された場所へ行く、③指示されたオブジェクトを取る、④指示されたオブジェクトを置く、⑤ 指示されたオブジェクトの名前をユーザの発話から学習する、の 5 種類の動作を行う。ロボットはオブジェクトが指示されると、その名前を音声合成により復唱する。オブジェクトが指示されない命令の場合は「分かりました」と発話する。さらに、ユーザの入力音声小さい、短いことなどが要因で音声認識エンジンがエラーを返した場合、「もう一度言って下さい」と発話を行い、ユーザの再発話を促す。

音声認識モジュールには Julius[李 09]を用い、記述文法と単語辞書により上記 5 種類の発話を認識出来るようにした。単語学習をするため、単語辞書に、音素の異なる清音、濁音、半濁音、拗音、長音、撥音から成る計 225 音を単語として登録した。学習用の定型文は「この名前は X です」とし、X には上記 255

単語の繰り返しを受理可能な文法を定義した。図9に単語学習用文法を示す。ただし、 $S_0$ は撥音を除いた224単語の集合、 $S_1$ は撥音を含む225単語の集合とする。これにより、任意の音素列が受理可能となり、得られた音素列を、オブジェクトを表す単語として新たに辞書に登録する。

### 3.2 実験条件

オブジェクトの名前を学習しないロボット(ロボットA)と学習するロボット(ロボットB)に対して、被験者がそれぞれ10分ずつ対話を行う。対話の後、それぞれのロボットに対するアンケートに回答する。内容はSD法による印象評価と、ロボットに対する印象をキーワードで挙げてもらうものとした。ロボットAとロボットBのどちらと先にコミュニケーションするかは、偏りが出ないようにランダムに決定した。被験者は男性11名、女性1名の計12名である。

### 3.3 実験結果と考察

ロボットAとロボットBにおいて、コミュニケーション内容がどのように変化していたのかを述べる。実験はオブジェクトが配置された環境であるため、オブジェクトを用いたコミュニケーションとそうでないものに分けられる。被験者全員の発話数の総和に対して、オブジェクトを用いたコミュニケーションは、ロボットAでは61[%]、ロボットBでは85[%]である。この結果より、ロボットが語彙学習機構を持つことで、コミュニケーションの傾向に影響を与えるという結果を示している。

SD法による印象評価の結果を図10に示す。また、キーワードによる自由記述において、キーワードの出現回数が、ロボットAとロボットBで2回以上の差が見られたものを図11に示す。

図10の印象プロフィールから、ロボットAよりロボットBの方が「人間的」、「飽きにくい」、「楽しい」、「親しみ易い」、「好き」という印象を受けやすいという結果が得られた。また、「不安-安心」の項目では、ロボットAよりロボットBの方がわずかに「不安」寄りの印象を受けるとい結果が得られた。

ロボットBの方が「飽きにくい」「楽しい」と評価された要因として、学習により単語や発話のパリエーションが増加したことや、教える行為自体の楽しさが考えられる。さらに、ロボットが聞き間違えた音素系列を発話することの楽しさも考えられる。これは、ただどく発話する幼児を連想させるため、ロボットBに対する印象記述で「幼い」が増加したと考えられる。また、聞き間違いはするが、命令は正しく遂行しようとする姿が「一生懸命」や「忠実」という印象記述を誘発したと考えられる。

ロボットBに対してわずかに「不安」寄りの印象を受ける要因として、ユーザにより教示された単語の音素列が、ロボットに正しく認識されないことが挙げられる。これは、ロボットBに対する印象の自由記述において「難聴」の出現数が増加し「従順」が減少したことから推察することが出来る。相手の行動や発話が全て予想出来たら「安心」ではあるが、それゆえに「機械的」であるとも言える。ロボットBの方がロボットAより「人間的」で「親しみ易い」と評価されたのは、語彙学習機能により生じる誤りが「人間的」であり、幼さや一生懸命さを演出したためであると考えられる。

### 4. まとめ

本稿ではSIGVerseを用いて行った二つの語彙学習シミュレーション実験について報告した。一つ目の実験では、参照点と座標系が未知となる教示から、相対的な概念を学習する手法を提案し、その有効性を調査した。また、被験者実験を通して、提案手法の適用範囲を調査するとともに、課題を明らかにした。二

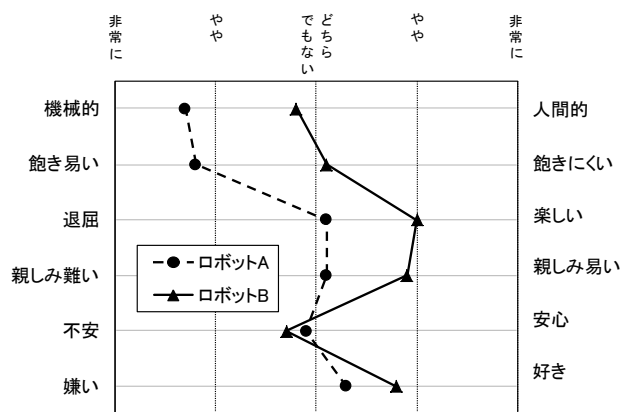


図10 ロボットに対する印象のプロフィール

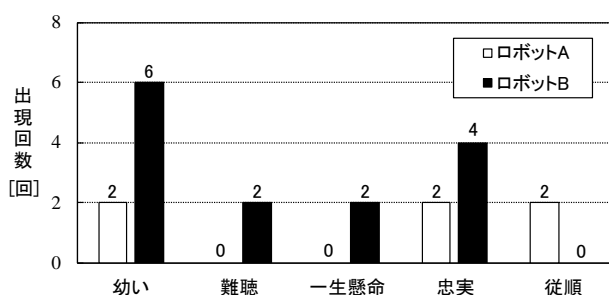


図11 出現数に2回以上差が見られたキーワード

つ目の実験では、シミュレータ上で被験者とロボットの対話実験を行い、ロボットの語彙学習機能がコミュニケーションに与える影響について調査した。実験により、ロボットが語彙学習機構を持つ方が、「人間的」で「親しみ易い」印象を受けるとい結果が得られた。本稿で紹介した両実験は異なるシステム上で実施したが、今後は両システムを統合し、より複雑な語彙学習機能をもつロボットとの長期間のインタラクションデータを収集したい。

### 参考文献

[Araki 12] Araki, T. et al.: "Online Object Categorization Using Multimodal Information Autonomously Acquired by a Mobile Robot", *Advanced Robotics*, 26(17), pp.1995-2020, 2012.

[稲島 12] 稲島 哲也: "社会的知能発生学シミュレータSIGVerseを用いた記号創発ロボティクスの展開", *人工知能学会誌*, Vol.27, No.6, pp.619-624, 2012.

[田口 13] 田口亮 他: "EMアルゴリズムを用いた参照点に依存した語意の学習", *情報処理学会 第75回全国大会講演論文集*, 4D-2, 2013.

[塩見 06] 塩見昌裕 他: RFIDタグを用いたコミュニケーションロボットによる科学館での展示案内, *日本ロボット学会誌*, Vol.24, No.4, pp.489-496, 2006.

[李 09] 李晃伸 他: "Juliusを用いた音声認識インタフェースの作成", *ヒューマンインタフェース学会誌*, Vol.11, No.1, pp.31-38, 2009.