社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2006－NL－176（12）
2006／11／23

画像に対する発話からの名詞概念の獲得

内田 ゆず　　荒木 健治
北海道大学大学院 情報科学研究科

ロボットが人間の日常生活の中に浸透し，人間と関わりを持ちながら活動するためには，コミュニケーション能力の向上が重要な課題になる．オープンな環境で動作するロボットには，決められたタスクを遂行するための言語能力を組み込んでおくということができない．したがって，環境に合わせて動的に言語を獲得するメカニズムが必要である．本稿では，ユーザが画像を提示しながら行った発話から，その画像に適したラベルを獲得する手法について述べる．また，その手法を用いた名詞概念獲得システムを作成し，印象評価実験を行ったところ，幅広い層のユーザがこのようなシステムに好感を抱くという結果が得られ，本手法の有効性が確認された．

Acquisition of a Noun Concept from an Utterance for an Image

Yuzu Uchida　　Kenji Araki
Graduate School of Information Science and Technology, Hokkaido University

The communication ability of the robot entering our daily life needs to be enhanced. In case of a robot working in an open environment, the task-oriented language ability will not work properly. Therefore, it is necessary to adapt to the environment and acquire a language dynamically. In this paper, we propose a system which acquires a label of an image based on infant vocabulary acquisition process. By using the system, we have evaluated its impressions. The experiment results showed that the system achieved high evaluation results.

1. はじめに

近年, 様々な種類のコミュニケーションロボットが開発され, 徐々に我々の生活に浸透してきている．コミュニケーションロボットは”癒し”や”エンターテイメント性”といった新たな役割が与えられているため, これまでのロボットと異なり, 対話能力が大変重要である．しかし，我々と自然に意思の疎通を図ることができるロボットは今のところ存在しない．言い換えると，現在のロボットは，タスクを限定したり，予め使用できる言語表現決めておくなど, 何らかの負担をユーザに強いることで対話を成立させているのである．ロボットが人間の真のパートナーになるためには，対話能力の向上が必要不可欠である．

既存の対話システムは, 大人の言語処理能力をモデルとして構築されている．つまり，完璧に言語を使いこなせるモデルを一足飛びに作ろうとしている．しかし，人間の言語能力とは極めて複雑なものであるため, このようなアプローチでは汎用的な対話システムに到達することは非常に困難であると考えられる．そこで，我々は，人が言語能力を獲得する能力を模倣したシステムを作成することにより, 最終的に人間と同等の対話能力の実現が可能なのではないかと考えている[1].

人間の幼児の言語獲得過程にはいくつかの発達段階が見られる．生後数ヶ月はまだ言葉にならない音（喃語）を発するだけであるが，生後 1 歳から 1 歳半くらいになると, 意味のある一つの単語を発話するようになる．この時期は「一語期」と呼ばれる．また，この頃から幼児は一日 6～10 語という驚異的なスピードで語彙数を増大させていく（語彙爆発）．さらに生後２歳くらいまでに「二語期」といわれる段階に達し，次に「あっち，お兄ちゃんいる」のような「電報文」と呼ばれる段階がくる．３歳くらいまでには語彙数も数百語から千語以上になり, 発音も大人の言葉に近いものになる[2].

我々は, 言語獲得能力を工学的に実現する第一歩として, 幼児の「一語期」に数多く獲得される名詞語彙を獲得することを目指している.

これまでに, 言語獲得システムに関する研究が様々なアプローチで行われてきた. 岩橋は人間とロボットとの言語コミュニケーションによる相互理解のために, ロボットによる言語獲得に関する研究を行っている[3]. この手法は, 二つ以上の事物の関係の概念及び文法を学習するが, 名詞概念の学習は行っていない. 本稿では, 予め語彙および統語的知識を与えない状態で, 画像に対して行われたユーザの発話からその画像に適した名詞（ラベル）を獲得する手法を提案している[4][5][6]. 田中らは, 仮想世界の中に存在する複数のソフトウェアロボットと音声対話によってインタラクションすることができるシステムを開発している[7]. また, 須賀らは, 幼児の言語獲得過程をシミュレートするシステムとしてMlas (Multi-Language Acquisition System)を開発した[8]. これらの研究では, 仮想世界を構築し, その中での言語現象のみを対象としているが, 我々の研究は実データとして画像を用い, 自由発話を対象としている. 国外でも言語獲得に関する研究は盛んで, Rogers らは”The BABY Project”の中で, 与えられた例から言語規則などを学習するシステム Babbette を提案した[9]. また, 近年では Levinson らが実世界との相互作用から言語を獲得するロボットを用いた研究を進めている[10]. 本研究で用いる幼児の言語獲得能力に基づいたアルゴリズムや, 実データとしての画像や自由発話を入力としている点はこれらの研究との差異である.

本稿では, 作成したシステムの概要, 及び複数被験者によるシステムの印象評価実験, そして今後の展望について述べる.

2. 名詞獲得システム

2.1 幼児の言語獲得モデル

幼児は, ある言葉が指し示すものを的確に捉えることで, 効率的な言語獲得を成し遂げている. このように, 無数にある概念を有効に制限する能力が, 初期の言語獲得においては非常に重要である. この制限のひとつの方法が「制約の理論」[2][11]である.「制約の理論」のもとでは, いくつかの個別の「制約」が提案・検討されている. 本システムは, それらの諸制約のうち,「なじみのない物体につけられた未知の名前は物体の部分, 色, 素材, 特質などでなく物体全体の名前である.」という事物全体バイアス,「未知の物体に対してつけられた名前は, その物体を含むカテゴリに対する名前である.」という事物カテゴリバイアス,「幼児は獲得したラベルを形状が似ている他のものに拡張する」という形状類似バイアス,「異なる名詞が同じ対象を指示した場合, ふたつの語はまったく同一ではない.」という対比の原理, の 4 つの制約に基づいたモデルによって構築される.

2.2 タスク

幼児が言語を獲得するには, 次の 3 つの過程が必要であるとされている[12].

- 音声インプットを単語分割するため, 母語の音声的な特徴を分析する.
- 音声インプットから単語を切り出す.
- 切り出した単語に意味を付与する.

本研究が対象とするのはテキストから単語を切り出し, その単語に意味を付与するという過程の実現である. したがって, 上記の言語獲得過程のうち, 音声を分析する部分, 音声から単語を切り出す部分は研究の対象外としている.

本研究のタスクは, 画像を提示しながら行われた発話文からその画像に対応する名詞, つまりラベルを学習し, 獲得するというものである. これを, 前述の「制約の理論」をモデル化して得た言語獲得のための能力だけで実現する. つまり, 予めシステムに言語的知識は与えない. 与えられるのは, 名詞を獲得するためのいくつかのルールのみである.

2.3 処理過程

本システムは大まかに分けて, 入力・画像処理・共通部分抽出・基本スコア計算・出力・ユーザの評価・名詞獲得・ラベル獲得ルール生成の 8 つの処理部で構成される. 図 1 に処理の流れを示す. また, **2.3.1** から **2.3.8** で各処理部の詳細を述べる.

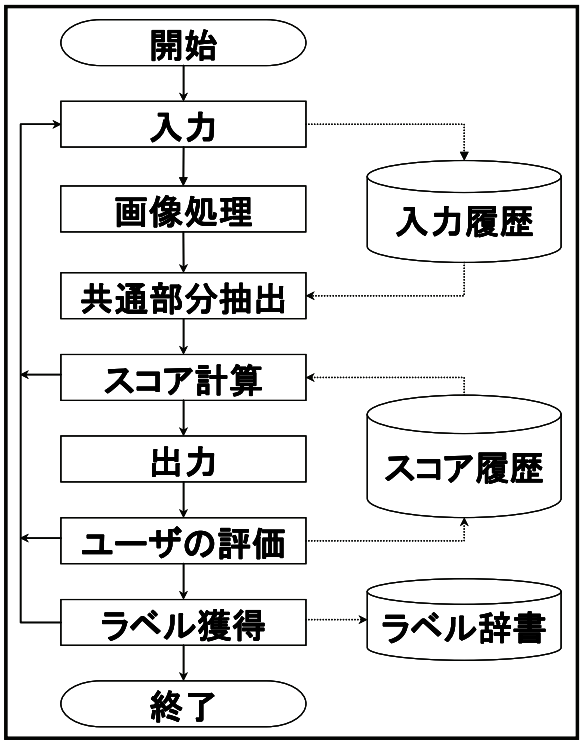

図 1 処理の流れ

2.3.1 入力

入力は画像と文の対である．入力画像は Web カメラ（USB-CAMCHAT2／アイ・オー・データ機器．有効画素数：30 万画素）からキャプチャされた画像（以降画像 P と呼ぶ），入力文は画像 P を見せながらユーザが幼児に話かける発話 1 文（以降文 S と呼ぶ）である．

入力画像は，ユーザが自由に被写体を選び撮影するものである．また，画像のサイズは 320×240 ピクセルで，なるべく被写体全体が画像内に収まるように撮影する．

入力文は全てひらがなで表記され，入力文に形態素解析などの前処理は一切施されない．ひらがなで表記するのは，入力された文字列自体に意味が含まれてしまうことを避けるためであり，形態素解析などを行わないのは，幼児が正確な品詞分割などの能力を持っていないと考えるためである．

2.3.2 画像処理

過去に同じ被写体が写った画像が入力されたかどうかを判断する．ここでは，エボリューション・ロボティクス社の”ERSP3.1”(Evolution Robotics Software Platform)[13]に含まれる”ERSP ビジョン”を用いた．ERSP 3.1 は，ロボット製品の作成を目的とした総合開発プラットフォームで，ERSP ビジョンは照明や物体の位置が管理されていない現実的な環境の中でも，ロボットや装置が２次元と３次元の物体を認識することができる画像認識ツールである．

2.3.3 共通部分抽出

システムは入力を得ると，過去に画像 P とともに入力された文と文 S を比較して，字面が一致する文字列を切り出す．この切り出された文字列を共通部分と呼ぶ．これ以降の処理で共通部分は，画像 P に対応するラベルの候補として扱われる．

2.3.4 基本スコア計算

抽出された共通部分には基本スコアが付与される．基本スコアとは，その共通部分のラベルとしての確からしさを表した値であり，出現頻度が高く，文字数が多く，他の画像と共に出現することのない共通部分ほど高いスコアを与えられる．基本スコア計算式は式(1)のようになる．

ここで，α は共通部分が他の画像とともに出現している場合スコアを減少させるように働く係数，F は共通部分が同一画像と共に出現した頻度，PN は画像の出現回数，L は共通部分の文字数である．

$$SCORE = \alpha \times \frac{F}{PN} \times \sqrt{L} \quad \cdots\cdots\cdots (1)$$

2.3.5 出力

2.3.4 で述べた方法で求めた基本スコアが閾値を超えた共通部分は，画像 P のラベルに適している可能性が高いと判断され，テキストで出力される．

2.3.6 ユーザの評価

システムの出力に対してユーザは次の 3 つのキーワードのうち，最も相応しいものを選び，入力する．

- 「じょうず」：ラベルとして適切である
- 「おしい」：ラベルとしては適切でないが意味はわかる
- 「ちがうよ」：意味がわからない

ユーザの反応によってその共通部分のスコアは再計算される. 幼児がこれらのキーワードを理解するとは考えられないが, 実際には, 大人の表情や声の調子で感じ取ることのできる情報は多い. 本手法ではそれらの代わりにキーワードを用いることとする. また, 今後は上に挙げたキーワード以外にも多様なキーワードを用意し, ユーザの自由な評価を許容できるようにする予定である.

2.3.7 名詞獲得

「入力」から「ユーザの評価」の処理を繰り返した結果, 再計算されたスコアが閾値を超え, さらに「じょうず」という評価を得たことがある共通部分は画像 P のラベルとして獲得される.

2.3.8 ラベル獲得ルールの生成[5]

ラベル獲得ルールとは, 再帰的な名詞獲得を行うためのルールである. 人間は過去に得た知識を活用し, より効率的に学習を進めていく. 本手法ではそのような再帰的な学習を次のようにして実現している.

獲得したラベル	:わんちゃん
過去の入力	:あっちにわんちゃんがいるよ
	↓
ラベル獲得ルール	:あっちに@1がいるよ

図 2 ラベル獲得ルールの例

システムが文字列 S をある事物に関する正しいラベルとして獲得すると, その事物に関する過去の入力文のうち, 文字列 S を含む文から, ラベル獲得ルールを生成する. ラベル獲得ルールとは, 図 2 のようにラベルの部分を変数とすることで, 入力文を抽象化したものである. 次に, 生成したラベル獲得ルールに合致する入力文があった場合, 変数部@1 に相当する部分を切り出し, スコアを上昇させる.

これは, 人間は様々な表現を聞いているうちに, どのような表現がラベルを示すものなのかを学習して, より効率的に学習を進めていると考え, その様子をモデル化したものである. 実際に, 語彙爆発期の幼児は, 一度大人が事物を指して言葉を発するのを聞いただけで, 正しくその言葉を使うことができる（即時マッピング）ことが知られている[2]. ラベル獲得ルールを生成することで, 本システムでもこれと似た現象を再現することができる.

2.4 ユーザインタフェース

本システムのユーザインタフェースを図 3 に示す. なお, Microsoft Visual Studio .NET 2005 を用いてこのインタフェースの構築を行った.

ウィンドウ中の 2 枚の画像のうち, 左は Web カメラからのプレビュー, 右はキャプチャ済みの画像となっている. この場合, ユーザは赤ちゃんに消しゴムを見せながら話しかけているという場面を表している. また, 右下の赤ちゃんに付随している吹き出しの内容は **2.3.5** で述べた出力である.

図 3 インタフェース

3. 印象評価実験

3.1 実験の構成

被験者に実際に本システムを使用してもらい, 実験を行った. 被験者に与えられた情報は, システムの操作方法, このシステムには提示した画像にラベルをつける能力があること, システムが何か出力した場合はそれに対して評価をすること, である. 入力文はキーボードを使って入力するように指示した. 一通りシステムに触れてもらった直後に, システムを使用してみて感じた印象を尋ねるアンケートの回答を依頼した. アンケートの詳細は 3.2 で述べる. また, コンピュータ習熟度や, ロボットに期待する機能などを質問紙によって回答してもらった.

実験の被験者は 20 代から 50 代までの男女 20 人（男：8 人・女：12 人）であった．

3.2 アンケートの詳細

システムの印象を評価するために評定尺度法を用いた．評価に用いた 20 の形容詞対を表 1 に示す．これらの形容詞対は，SD 法においてよく用いられる形容詞対から選出した[14]．以下では，各形容詞対についての 7 段階尺度（非常に・かなり・やや・どちらでもない・やや・かなり・非常に）の評定をポジティブな形容詞（表 1 の形容詞対のうち，右の語）が高くなるように 1 から 7 まで数値化して分析する．

3.3 実験結果

各形容詞対における評定の平均値･標準偏差を表 1 に示す．また，システムを使用して感じた印象についてのプロフィールを図 4 に示す．

表 1 評価に用いた形容詞対・平均値・標準

形容詞対		平均値	標準偏差
こわい	やさしい	4.95	0.91
わかりにくい	わかりやすい	4.89	1.20
退屈な	興味深い	5.32	1.16
感じの悪い	感じのよい	4.68	1.25
性的な	動的な	4.00	1.15
近づきにくい	近づきやすい	4.74	1.15
古い	新しい	5.47	1.02
陰気な	陽気な	4.95	0.71
親しみにくい	親しみやすい	5.32	1.16
消極的な	積極的な	4.42	1.02
つまらない	面白い	5.26	0.99
単純な	複雑な	4.42	1.22
嫌いな	好きな	5.05	1.08
わがままな	思いやりのある	4.26	0.56
空虚な	充実した	4.47	0.90
愚かな	賢い	5.11	0.94
にくらしい	かわいらしい	6.16	0.96
苦しい	楽しい	5.16	1.42
冷たい	暖かい	5.21	0.85
機械的な	人間的な	4.21	1.23

印象評価に男女差が存在するのかを調査するために，各形容詞対の評定の平均値について t 検定を用いて統計的有意性の検定を行った. その結果，有意水準 5%において，どの形容詞対についても統計的に有意な差は認められなかった．

同様にして，コンピュータの習熟度や，日常生活の場で活動するロボットに会話の機能を期待するかどうか, によって違いが存在するかを調査したところ, これらも有意な差は認められなかった．

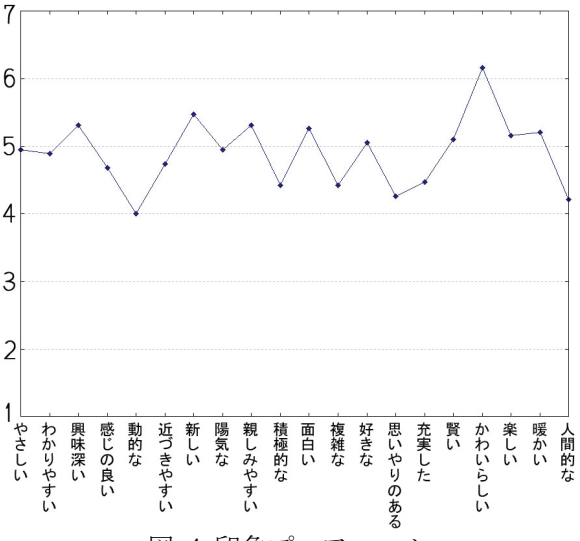

図 4 印象プロフィール

3.4 考察

まず，システムの印象について，評定の平均値から考察する．20 組の形容詞対全てについて，4 点以上を獲得している．ここから，言語を獲得するシステムはユーザに好印象を与えることがわかる．中でも，「興味深い」，「新しい」，「親しみやすい」，「面白い」，「好きな」，「賢い」，「かわいらしい」，「楽しい」，「暖かい」の項目では，平均値が 5.0 点以上と高得点であった．「かわいらしい」の結果が最も高得点であったが，これはシステムのインタフェースに大きく依存している可能性がある．しかし「新しい」，「面白い」，「賢い」などの項目の結果は, 純粋にこのシステムの機能を評価したものであると考えられる．また，それらの項目はコミュニケーションロボットに求められている”癒し”や”エンターテイメント性”に直結する．したがって，そのような項目が高得点であったことから, コミュニケーションロボットに対話の中から画像にラベルを付与する機能を持たせることでユーザ満足度の向上につながることが期待できる．

次に，t 検定による統計的有意性の検定結果について考察する．今回は性別，コンピュータ習熟度，ロボットと対話をしたいかどうか，の３つの観点について, 被験者の評定の平均値に有意な差が見られるかどうかを調査した. 全ての観点で有意な差は認められないという結果が得られたことから, 幅広い層のユーザにこのシステムが受け

入れられていると言える. 日常生活の中で人間と関わりを持って動作するコミュニケーションロボットにこのシステムの機能を搭載することを想定したとき，これは大きな利点である.

実験終了後, 自由記述により印象に残ったことについて解答を求めた結果, 最も多かったのはキーボード入力に対する抵抗感であった. 今回の実験では, 音声認識誤りなどの問題を排除するため, 入力文を音声認識ではなくキーボードでの入力を用いたが, 一般のユーザには不評であることがわかった. 我々はこれまでに，入力文に音声認識結果を用いてこのシステムを使用した場合, 音声認識誤りに対して頑健であることを実証している[6]. しかし，幼児は言語モデルに相当する知識を持たずに言語獲得を行っている. そこで，今後は，言語モデルを用いず，音素認識を用いて入力を行った場合でも言語獲得が可能であるかを検証する必要がある.

4. まとめ

本稿では, 我々が作成した対話の中から名詞を獲得するシステムの詳細と, そのシステムがユーザに与える印象についての印象評価実験の結果について報告した. 被験者は概ねこのシステムに好意的であることが明らかになった. また, 性別, コンピュータ習熟度, ロボットとの対話へのイメージによらず, この印象には大きな差がないことも確かめられた.

今後は, 音素認識による入力に対応したシステムを完成させ, 性能評価を行う予定である. また, 名詞獲得アルゴリズムの正当性を示すために, 他言語での実験も視野に入れている.

参考文献

[1] 荒木健治 : 自然言語処理ことはじめ―言葉を覚え会話のできるコンピュータ，森北出版，2004.

[2] 今井むつみ : ことばの学習のパラドックス，共立出版，1997.

[3] 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イムを目指して，人工知能学会誌，Vol.18，No.1，pp.49-58，2003.

[4] 内田ゆず，荒木健治 : 言語獲得システムにおける類似度に基づくラベル拡張手法の提案，平成 17 年 電気･情報関係学会北海道支部連合大会講演論文集，180，2005.

[5] 内田ゆず，荒木健治 : 幼児の普通名詞および固有名詞獲得モデルに基づく帰納的学習を用いた再帰的獲得手法の提案，言語獲得と理解研究会 (LAU)，Vol.1，No.1，pp.21-27，2005.

[6] 内田ゆず，荒木健治 : 画像とそれに対する発話を対象とした幼児の名詞獲得モデル，言語処理学会第 12 回論文集，pp.907-910，2006.

[7] 田中穂積，徳永健伸 : ロボットとの会話.人工知能からのアプローチ，情報処理，Vol.44，No.12，pp.1247-1252，2003.

[8] 須賀哲夫，久野雅樹 : ヴァーチャルインファント－言語獲得の謎を解く，北大路書房，2000.

[9] Yorick Wilks : Machine Conversations(Kluwer International Series in Engineering and Computer Science)，Kluwer Academic Pub，1997.

[10] S.E.Levinson，K.Squire，R.S.Lin，M.McClain : Automatic Language Acquisition by an Autonomous Robot，AAAI Spring Symposium on Developmental Robotics，March 21-23，2005.

[11] 小林郁夫，古川康一，今井むつみ，尾崎知伸 : 帰納論理プログラムによる幼児の名詞語彙獲得のモデル化，電子情報通信学会技術研究報告 言語理解とコミュニケーション研究会(NLC)，Vol.99，No.387，pp.29-36，1999.

[12] 今井むつみ，野島久雄 : 人が学ぶということ―認知学習論からの視点，北樹出版，2003.

[13] http://www.evolution.com/products/ersp/

[14] 末永俊郎 : 社会心理学研究入門, 東京大学出版会, 1987.

[15] Paul C. Quinn : Category representation in young infants，Current Directions in Psychological Science，Vol.11，No.2，pp.17-22，2002.

[16] Joost van de Weijer : How much does an infant hear in a day?, Proceedings of the GALA2001 Conference on Language Acquisition，pp.279-282

[17] 但馬香里 : 幼児における一語発話の獲得について―1 歳 10 ヶ月から 2 歳 0 ヶ月児の 3 人の幼児による観察報告，東京工芸大学工学部紀要，Vol.27，No.2，pp.59-64，2004.

[18] 須藤珠水，茂木健一郎 : 言語獲得期における語意学習とカテゴリー認知のメカニズム，電子情報通信学会 スマートインフォメディアシステム研究会 信学技報，SIS2004-4，pp.17-22，2004.