

幼児の日常生活を収録したビデオデータの分析と名詞概念獲得システム SINCA の評価
 Analysis of Video Data of Daily Life and
 Evaluation of a System for Noun Concepts Acquisition (SINCA)

内田 ゆず[†] 荒木 健治[‡]

Yuzu Uchida Kenji Araki

1. まえがき

2001年に日本ロボット工業会(JARA)が発表した「21世紀におけるロボット社会創造のための技術戦略調査報告書」では、ロボットの市場規模は2010年に3兆円、2025年に8兆円にまで伸びると予測されている[1]。この市場の伸びは、生活分野や医療福祉分野などの非製造業分野におけるサービスロボットの普及によるところが大きい。しかし、ペットロボットやコミュニケーションロボット、掃除ロボットなどの一般家庭で動作するホームロボットが開発され、徐々に普及してきているものの、現状では大規模な市場の成長にはつながっていない。

ホームロボットの普及を妨げている原因の一つは、ユーザのロボットに対する過剰な期待であると考えられる。特に日本においては「ロボットはなんでもできる」というイメージが先行しているが、実際のロボットは人間とのコミュニケーションもままならない状況である。一方で、ユーザの期待に応えるためにロボットを高機能化していくと、ロボットを操作するユーザ側の負担が大きくなるというジレンマも抱えている。

言語処理の分野では、ロボットと人間のコミュニケーションを含め、多様なタスクに関する研究が行われている。それらの多くは、「人間なら簡単にできることが機械には難しい」という部分をいかに解決するかに焦点が当てられている。この問題は、人間と同等の言語理解能力を利用することにより解決することができる。しかし、現状ではシステムが人間と同じレベルで言語を理解することは不可能であるため、最終的には人間の手を加える必要がある。

人間の言語活動は、対話の中から言語を獲得する、記憶に基づいた対話を行う、質問に対して論理的な応答を行う、ユーモアや皮肉を理解する、環境や相手に適した言葉を選択する、感情を表現する、などの様々な能力から成り立っている。我々は、人間と同等の言語能力を持つロボットの実現を目指している。その研究の第一歩として、人間の言語獲得能力を手がかりとした言語獲得手法を考案した。そして、上記の手法を適用した名詞概念獲得システム SINCA (System for Noun Concepts Acquisition from utterances about Image) を開発し、研究を行ってきた[2]。SINCA は、ユーザが提示する画像と、画像に関する発話から名詞概念(画像に対するラベル)を獲得するシステムである。

ロボットを一般家庭で動作させるには、家庭内に存在する事物のラベルの知識が必要になる。しかし、あらゆる事物のラベルを予めデータベース化するには大変な労力がかかる。ユーザが直接ロボットに教示するとしても、事前に

教示方法を習得しなければいけないような方法ではユーザに多大な負担がかかってしまう。SINCA をロボットに搭載し、自然な発話でラベルを教示することが可能になれば、これらの問題は解決するものと考えられる。

SINCA の有効性については、様々な観点から評価が行われてきた[2],[3]。SINCA の重要なアルゴリズムの一つであるラベル獲得ルールについては次の三つの評価実験が行われた。

一つ目は、名詞概念を獲得するために必要な入力回数を、ラベル獲得ルールを利用する場合と利用しない場合と比較し、ラベル獲得ルールが名詞概念に与える効果の評価を行う実験である。その結果、ラベル獲得ルールを利用することで一つのラベルを獲得するために必要な平均入力回数が7.0回から5.6回に減少し、より効率的な名詞概念獲得が可能になることが明らかになった。

二つ目は、Google¹による検索ヒット件数を用いて各ルールの一般性を測定する実験である。その結果、Google の検索ヒット件数が10,000件以下のルールは、一般性が低く適用されにくいと判断でき、淘汰することが可能であることが明らかになった。

三つ目は、World Wide Web 上の日本語文書をコーパスとみなし、SINCA が生成するラベル獲得ルールの妥当性を調査する性能評価実験である。その結果、全体の26%のルールが80%以上の有効率で名詞の抽出が可能であることが明らかになった。

システム全体の評価としては、評定尺度法[4]を用いた印象評価実験を行った。実験の結果、評定尺度法で用いた全ての項目に対して中間点である4.0ポイントを上回るポジティブな評価を得られ、SINCA が幅広い層のユーザに好意的な印象を与えることが確認された。また、ELIZA[5]との比較によって、SINCA はキーワードやテンプレートを用いた簡単な対話システムよりも好印象を与えられることが明らかになった。

工学の分野では、これまでも言語獲得システムや言語理解システムの研究が行われてきた[6]~[8]。人間は言語獲得のエキスパートである。したがって、言語獲得期にある人間の幼児が実際に日常的に聞いている発話を参考にすることが有益であると考えられる。Roy ら[9]は統計的な手法によって、音声と画像の提示からロボットに語彙や簡単な文法の学習を行わせるシステムを開発し、実際の会話データを用いた実験を行っている。

我々は、人間の幼児が実際に日常的に聞いている対話データを分析し、SINCA に適用した実験を行った。大人と幼児の間で行われる実際の対話を、日本語を対象とした言語獲得システムに適用した研究はこれまでに存在しない。

[†] 青山学院大学理工学部 College of Science and Engineering, Aoyama Gakuin University

[‡] 北海道大学大学院情報科学研究科 Graduate School of Information Science and Technology, Hokkaido University

¹ <http://www.google.co.jp/>

本研究では、日本語を母語とする幼児が日常的に聞いている発話文をビデオデータから収集し、独自の対話データの作成を行った。また、そのデータを統計情報や特定の言語構造に依存しないアルゴリズムで構築された名詞概念獲得システムに適用している。これらの点に本研究の意義がある。

以下、2.で子どもの対話を対象とした既存コーパスについてまとめる。3.で SINCA の構成を述べ、4.では SINCA の処理過程を詳しく説明する。5.では、我々が作成した対話データの概要を述べ、6.では、対話データの分析結果について詳述する。7.では予備実験とその結果を示し、8.でビデオデータを用いた実験とその結果を示し、考察を述べる。最後に、9.で本論文をまとめる。

2. 子どもの対話コーパス

認知科学や心理学、言語学などの分野では、人が日常生活で行っている会話を収集し、研究資料として用いている。これらの分野では収集したデータを分析することが主な研究方針である。

子供の発話や子供と大人の会話を収集したコーパスには、以下のようなものがある。

- **CHILDES (Child Language Data Exchange System) [10]**
言語習得研究のための国際的な言語データ共有システムであり、英語や日本語をはじめ 26 か国語の発話データが収められている大規模コーパスである。
- **こども語辞書 [11]**
日本全国の母親が子供の言葉の成長をウェブサイト上で記録し、それらを集積・解析することで、1~3歳のこどもがどんな単語をいつごろ覚える傾向にあるかを検索・閲覧できるようにしたウェブツールである。対象言語は日本語である。
- **NTT 乳幼児音声データベース [12]**
3家庭 5名の幼児とその両親の自然発話を、幼児の誕生直後から最大 5年間にわたって断続的に録音した縦断的データベースである。対象言語は日本語である。
- **Vocabulary of First-Grade Children [13]**
5~8歳の 329名から収集した総単語数 286,108語、異なり単語数 6,412語の話し言葉のデータである。対象言語は英語である。
- **The Polytechnic of Wales Corpus [14]**
イギリスの South Wales 地域の 6~12歳の児童 120名より収集された約 65,000語の話し言葉コーパスである。対象言語は英語である。
- **The Bergen Corpus of London Teenager Language [15]**
London の 13~17歳の少年少女の自然な会話を録音した約 500,000語のコーパスである。対象言語は英語である。

このように、子供の話し言葉コーパスは徐々に整備されてきてはいるものの、特に日本語を対象としたコーパスは依然として少ない。

上述したコーパスは、単語単位のデータである点や、どのような場面での発話なのかがわからない点など、SINCAへの入力として使用するには問題が存在した。したがって、

本研究では、幼児のいる家庭にビデオカメラを設置し、大人と幼児の間で行われる日常的な会話を撮影するという方法で独自の話し言葉データを収集し、これを用いて評価実験を行った。日本語の話し言葉データの乏しさを考慮すると、本研究で収集したデータの分析によって幼児に向かつて行われる大人の発話の特徴などの新たな知見が得られ、認知科学や言語学への寄与も期待することができる。

3. システム概要

SINCA の概要を図 1 に示す。4.で個々の処理の詳細を述べる。

SINCA のインタフェースを図 2 に示す。なお、このインタフェースは Microsoft Visual Studio .NET 2005¹を用いて構築されている。ウィンドウ中の 2枚の画像のうち、左はユーザの画像撮影を補助するために表示される Web カメラからのプレビューである。右はキャプチャ済みの画像となっており、実際に入力画像として使用される。この場合、ユーザはシステム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。

右下の赤ちゃんに付随している吹き出しの内容は 4.5 で述べる出力である。赤ちゃんの画像の下には、「(今キャプチャされている画像は) 初めて見る画像である」「前に見たことがある」などの画像処理の結果や、「何か言っています」などのシステムの状態が表示される。

4. 処理過程

4.1 入力

SINCA への入力は常に一對の画像と文である。以下、4.1.1 において入力画像について、4.1.2 において入力文について詳しく述べる。

4.1.1 入力画像

入力画像は Web カメラからキャプチャされた画像(以降画像 P と呼ぶ)である。使用した Web カメラは USB-CAMCHAT2 (アイ・オー・データ機器。有効画素数: 30万画素)である。被写体に制限はなく、ユーザが自由に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。

4.1.2 入力文

入力文は画像 P に関する発話 1文(以降文 S と呼ぶ)である。発話内容、画像 P に含まれる物を指すラベルに制限はない。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理由は、漢字などの表意文字を使用して、入力された文字列自体に意味が含まれることを避けるためである。入力文の表記に揺れ(「消しゴム」、「けしごむ」など)が生じることを避ける目的もある。

形態素解析などの前処理を行わないのは、形態素解析結果の誤りの影響を避けるためと、未知語に対応できるようにするためである。

4.2 画像認識

この過程では、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。ただし、過去の全ての入力を対

¹ <http://www.microsoft.com/japan/msdn/vstudio/>.

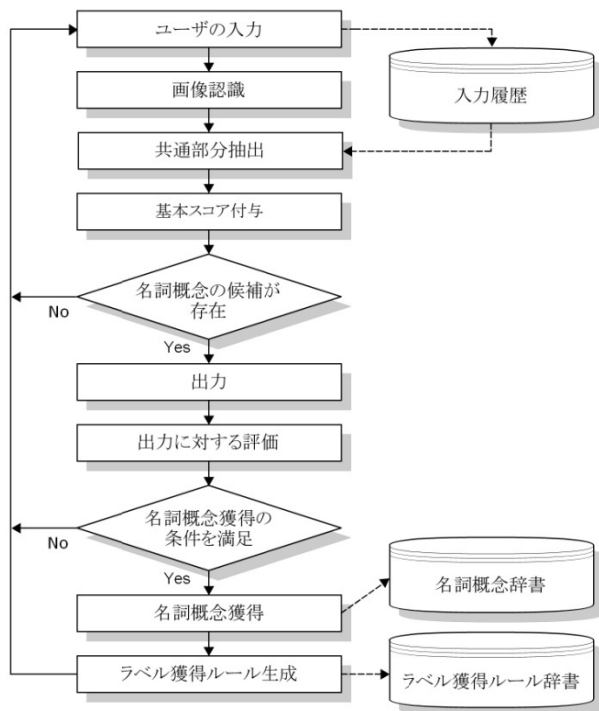


図1 処理の流れ



図2 インタフェース

象とすると入力回数が増えるにしたがって処理量が増大し、SINCA とユーザのインタラクションに支障が出る。したがって、画像認識の対象とする入力画像は、直近の 10 回分とした。

画像認識ツールには、エボリューション・ロボティクス社¹の"ERSP3.1 (Evolution Robotics Software Platform)"に含まれる"ERSP ビジョン"を用いた。

¹ <http://www.evolution.com/products/ersp/>

"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。

公表されている具体的な機能・性能を以下に示す。

- ・ 認識の対象となる物体の特性により、80~100%の認識率
- ・ 広範囲の視野角、レンズ歪み、画像ノイズ、照明条件でも動作
- ・ パターンの大部分 (最高 90%) が別の物体にさざぎられても動作
- ・ 複数個の物体を同時に認識することが可能
- ・ 計算量を大幅に増加させることなく、数百、さらには数千の視覚パターンを保存するデータベースを処理可能
- ・ 1,400MHz の PC を使用することで、ViPR は 208×160 ピクセルの画像を 1 秒間に約 14~18 フレーム処理可能

4.3 共通部分抽出

SINCA 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全て切り出す。この処理で切り出された文字列を共通部分と呼ぶ。以降の処理において、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。

一致した部分が一文字であっても共通部分として切り出されるためラベルの候補数が膨大になるが、短い共通部分は 4.4 で述べるスコア計算によって淘汰される。

4.4 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が与えられる。

基本スコア計算式は式 (1) のように定義する。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。

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

この式は、L の重みが異なる 4 種類のスコア計算式を用意し、SINCA に 1,000 回の入力を与えて名詞概念獲得の効率を調査する予備実験[16]によって決定したものである。

4.5 出力

4.4 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。

4.6 ユーザによる評価

SINCA の出力に対してユーザは次の三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。

獲得したラベル	: わんちゃん
入力文	: あっちに <u>わんちゃん</u> がいるよ
	↓
ラベル獲得ルール	: あっちに @ がいるよ

図3 ラベル獲得ルールの生成例

- ・ じょうず: ラベルとして適切である
- ・ おいしい: 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

本手法は語彙の知識を持たない初期状態を前提としている。SINCAはキーワードの意味を正確に理解しているのではなく、人間が相手の表情や声の調子で感じ取る様々な情報の代わりにこれらのキーワードを用いている。

出力された共通部分のスコアは、ユーザの評価が「じょうず」の場合は増加、「おいしい」の場合はわずかに減少、「ちがうよ」の場合は大幅に減少するように係数 β を乗じて再計算される。

本論文で述べる実験では、過去の実験結果を踏まえて、係数 β を「じょうず」の場合は1.5、「おいしい」の場合は0.8、「ちがうよ」の場合は0.2とした。

4.7 ラベル獲得

入力(4.1)からユーザによる評価(4.6)まで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。

4.8 ラベル獲得ルールの生成

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。

SINCAは文字列Sに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。具体的には、入力文に含まれるラベルの部分を変数化(以降、変数部を@と表す)することで、入力文を抽象化して生成する(図3参照)。以降、生成したラベル獲得ルールに合致する入力文が存在した場合、変数部@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。

5. 対話データの概要

幼児がいる家庭における日常生活をビデオカメラ(Everio GZ-MG255, Victor)を用いて撮影した。撮影対象となった家族は、日本語を母語としており、2歳7カ月の男児、12カ月の女兒(ビデオ撮影開始時の年齢)とその両親の4人で構成されている。



図4 ビデオ撮影風景の例

多くの場合、撮影場所は上記の家族の住宅の中であった。ただし、まれに自家用車内や祖父母宅、駅などの外出先で撮影された場合もある。撮影は母親か父親がビデオカメラを手で持った状態か、箆箭の上などのやや高い位置に固定した状態で行われた。図4に収録された場面の撮影を目的としたので、課題などは一切設けていない。

撮影は2007年7月21日から同年11月23日にかけて行われた。ただし、撮影協力者のプライバシーに配慮して、この期間中の「撮影を行っても支障がない状況」に限り、撮影するように依頼をした。その結果、約82時間分のビデオデータが得られた。

第一著者が得られた動画データを視聴し、その発話内容の書き起こし作業を行った。書き起こし対象は大人の発話(主に両親)であり、できる限り発話の通りにひらがな表記で記録を行った。音量が小さく内容が聞き取れない場合など、書き起こしが不可能な音声は"****"の記号で記録した。

実際の発話の一部を以下に示す。文中のDは長女の名前を表す。

- ・ もうあいすかわないよおとうさん(父親→長男)
- ・ ほらかまれるわありんこに(父親→長男)
- ・ ねむいのかなDちゃんは(父親→長女)
- ・ うえひこうきとんでるひこうき(母親→長男)
- ・ こおりとって(母親→長男)
- ・ りんごいまおうちにないもん(母親→長男)
- ・ なにたべてんのD(母親→長女)

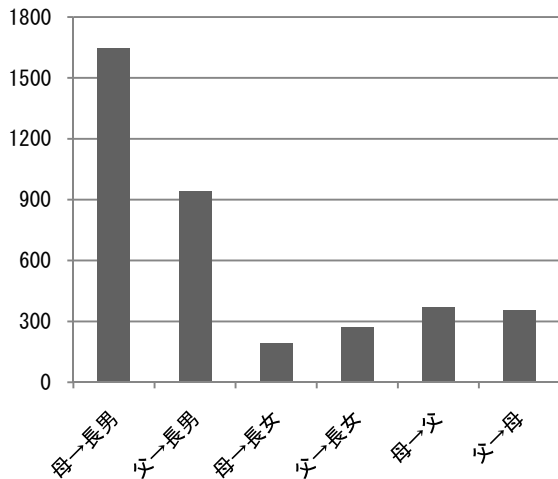


図5 発話回数

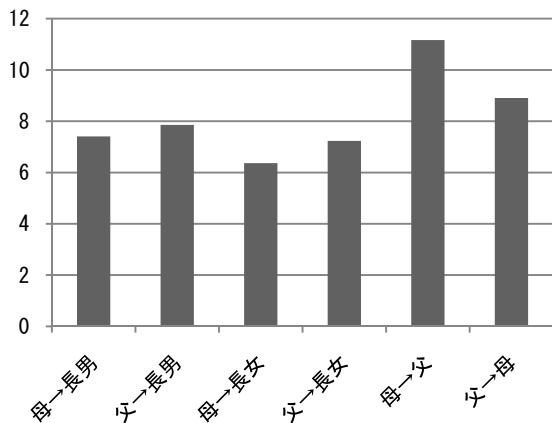


図6 発話の平均文字数

6. 対話データの分析

5.で述べた方法で書き起こしを行った対話データの一部を分析し、発話者と発話対象の違いが発話内容などにどのような影響を及ぼすのかの調査を行った。分析対象とした発話は3,954文である。

6.1 発話回数と発話の長さ

図5に発話者と発話対象ごとの発話回数を示す。母親から長男への発話が最も多く、全体の41.6%を占めている。それに対し、母親から長女への発話は全体の4.8%となっており、最も少ない。父親も長男への発話が多く、長女への発話は少ない傾向にある。長男は不完全ながらも1語～2語の発話を行うことができ、両親の発話に応答を返すことができる。自ら両親に向けて話しかけることもある。したがって、長男は両親と継続的な対話を行うことができる。一方で、長女はまだ話すことはできない。この違いが両親の発話回数に影響していると考えられる。

図6に発話者と発話対象ごとの発話の長さを示す。両親同士の発話は長く、子どもに向けた発話は短いという特徴

が見られる。長男への発話と長女への発話を比較すると、母親から長女への発話が平均して0.5～0.9文字ほど短くなっているが、顕著な差は見られなかった。

ビデオに収録された親子の対話の中で特徴的な現象として、幼児の発話に対して大人が行うオウム返しがあった。オウム返しは幼児の言語獲得に重要な役割を果たすことが報告されている[17]。今回分析を行ったデータの中には、91回のオウム返しが存在する。長女はまだ話すことができないこと、父親は母親に比べると長男の発話を正しく理解できないことから、必然的に母親が長男の発話に対して行うオウム返しとなっている。長男が正確ではない発音で発話を行い、母親はそれを正しく解釈し、正確な発音でオウム返しを行う様子が多く観察された。オウム返しが幼児の発音などを修正するためのフィードバックとして機能している点を参考にして、我々が開発した名詞概念獲得システムSINCAのフィードバック部分にもオウム返しを用いることを検討している。

ビデオデータの中には、呼びかけも多く含まれていた。両親の発話の中で、長男の名前が含まれているものは331文(全体の13.1%)、長女の名前が含まれているものは178文(全体の35.6%)存在した。大人同士の発話には呼びかけは非常に少ない(全体の0.8%)ため、子供に対する発話の特徴であると考えられる。

6.2 品詞の分布

発話者や発話対象によって発話に含まれる品詞に特定の傾向が現れるかを調査した。収集した発話文を話者、発話対象ごとに分類し、形態素解析を行った。分析対象となる発話文はオノマトペ¹を含んでいる、ひらがなで表記されている、などの特徴がある。形態素解析器は、連濁、反復形オノマトペを辞書に登録するのではなく、動的に認識を行うJUMAN²(ver.6.0)を使用した。

図7に両親から長男への発話、両親から長女への発話、両親の間で行われた発話にそれぞれ形態素解析を行った結果を示す。品詞の分類は、JUMANの仕様によるものである。この結果から、発話対象の違いには関係なく、名詞、動詞、助詞の割合が高いことが明らかになった。また、幼児を対象にした発話と大人を対象にした発話では、助詞の数に約6ポイントの差があることが明らかになった。幼児に対する発話は、1語発話が多く含まれることに加え、助詞によって格関係を明示する必要のない単純な内容であるためだと考えられる。

7. 予備実験

7.1 実験方法

SINCAの名詞概念獲得の性能のベースラインを決定するために、アンケートによって収集した入力文を用いた実験を行う。入力文はアンケートによって収集した文からランダムに選択する。入力文の収集方法については7.2で詳しく述べる。

¹ 擬声語、擬音語と擬態語の総称。

² <http://nlp.kuee.kyoto-u.ac.jp/nl-resource/juman.html>

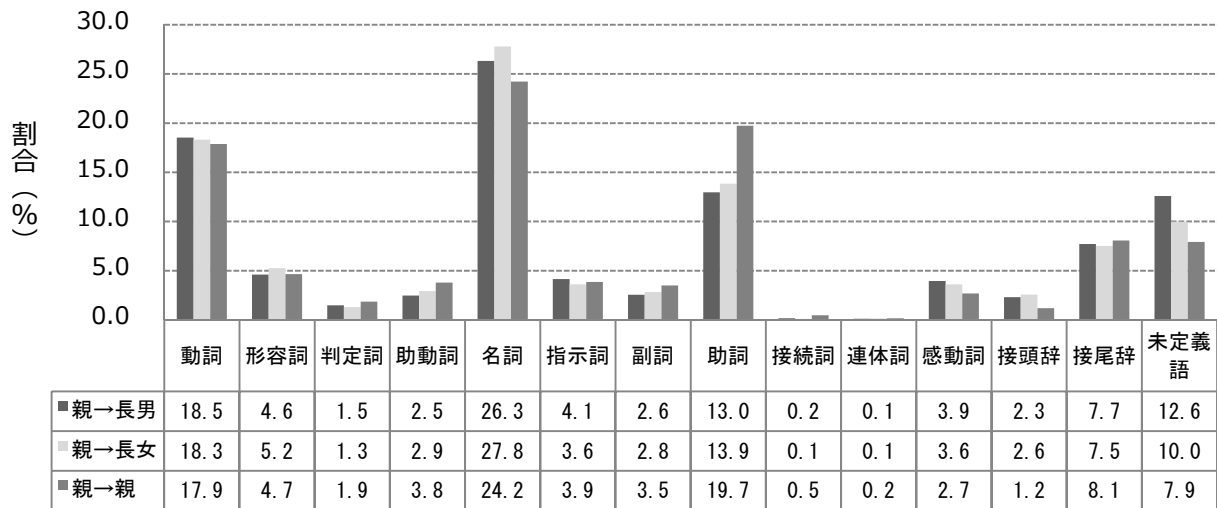


図7 発話者と発話対象ごとの発話に含まれる品詞の分布

SINCA は音声認識及び音素認識を用いて入力を行った場合にも名詞概念を獲得することが確認されているが[18], 本論文の実験では音声認識誤りが実験結果に影響を与えることを避けるために, 選択した文をキーボードから入力する. 同様に, 画像処理における誤認識が実験結果に与える影響を抑制するために, 準備した静止画像を入力ごとに Web カメラで撮影したものを入力画像とする.

7.2 アンケートによる入力文収集

10種類の事物(ウサギ, キリン, ネコ, クマ, りんご, かばん, 車, スプーン, ボール, ミルク)に関して「まだ話すことのできない赤ちゃんにその画像に含まれる物を見せながら話しかける」ことを想定した内容を尋ねるアンケートを実施した. 回答者は10代から60代までの31名(男性7名, 女性24名)であり, そのうち子育ての経験がある被験者は13名(男性1名, 女性12名)であった. 収集された全文数は324文であり, それらの平均モーラ長は10.98であった.

入力文の例を以下に示す.

- ・ かわいいねこがいるよ
- ・ おいしそうなりんごだね
- ・ あそこにいるのはくまさんだね
- ・ うさぎさんはにんじんを食べるんだよ
- ・ おもそうなかばんだね

7.3 実験結果

アンケートによって収集した入力文を用いた実験の結果, 10種類の画像に対して適切なラベルを対応付けることに成功した. 一つの名詞概念を獲得するまでに必要な入力回数は平均 6.2 回であり, 分散は 0.40 であった. 生成されたラベル獲得ルール数は, 52 個であった. ラベル獲得ルールが適用された回数は, 3 回であった.

8. SINCA を用いた実験

8.1 実験方法

6.で述べた対話データを SINCA に入力した場合, 名詞概念を獲得可能であるかを検証する実験を行う. ビデオ撮影

によって収集した発話文のうち, 親子の間で共同注意が成立している場面の発話を入力文の候補とする. さらにその中で出現頻度が高い 10 種類の名詞(あいす, あり, ばす, こおり, らーめん, おせんべい, とんねる, りんご, じゅーす)に関する 353 文からランダムに選択したものを入力文とする. 入力文はキーボードで入力し, 入力画像は上記の名詞に対応した静止画像を入力ごとに Web カメラで撮影したものをを用いた.

8.2 実験結果

実験の結果, ビデオから得られた対話データを入力に用いた場合も, SINCA は 10 種類の画像に対して適切なラベルを対応付けることに成功した. 一つの名詞概念を獲得するまでに必要な入力回数は, 平均 5.3 回であり, 分散は 0.21 であった. 生成されたラベル獲得ルール数は, 44 個であった. ラベル獲得ルールが適用された回数は, 1 回であった.

8.3 考察

ビデオ撮影によって収集した入力文を用いた実験から, SINCA は実際に幼児が聞いている発話データから名詞概念を獲得できることが明らかになった. また, 予備実験で示した, アンケートによって収集した入力文を用いた実験の結果と比較すると, ビデオデータを用いる方が効率的な学習が可能になることがわかる. 具体的には, 一つの名詞概念を獲得するまでに必要な入力回数は, ビデオデータによる入力文を用いた場合, アンケートによる入力文を用いた場合と比較して平均 0.9 回少なかった. このような実験結果が得られた要因として, 次の三つが考えられる.

一つ目は一語発話の数である. ビデオデータによって収集された入力文候補の平均モーラ長は 9.82, アンケートによって収集された文の平均モーラ長は 10.98 であった. このことから明らかなように, ビデオデータによる入力文はアンケートによる入力文より短いという傾向がある. アンケートによって収集された入力文の中には一語からなる文は全く含まれていなかったのに対し, ビデオデータから収集された入力文の候補のうち 66 文(全体の 18.6%)が一

表1 ラベルの直前・直後に出現する単語の異なり数と抜粋

ビデオデータ		アンケート	
直前	直後	直前	直後
全19種	全42種	全15種	全22種
も	そっち	は	は
の	たべ	が	だ
きみら	とった	かわいい	です
ほれ	きょう	この	か
ほら	いる	の	も
は	いっぱい	さん	いた
あと	だろ	さ	って
ちゃん	きた	そうな	から
そこ	おかたづけ	りっぱな	の
ね	ちょうだい	おでかけ	が
そんなに	のむ	わあ	を
それ	あけて	ながい	こっち
なんだ	すき	たかい	くび
いま	いった	...	はいて
...	しか		ひも
	のみ		...
	どうした		
	はき		
	うごいて		
	ない		
	そこ		
	...		

語発話であった。実際に実験で SINCA へ入力された文のうち、7文が一語発話であった。

二つ目は表現の多様性である。収集した文からランダムに100文ずつを抽出し、それらに含まれるラベルの直前と直後の1単語について分析を行った。ビデオデータによって収集した文では、ラベルの直前に出現する異なり単語数は19、直後では42であった。アンケートによって収集した文では、ラベルの直前に出現する異なり単語数は15、直後では22であった。ラベルの直前・直後に現れた単語の異なり数とその抜粋を表1に示す。アンケートの回答者はアンケート用紙に印刷された静止画に含まれる事物を説明しようとするため、画一的な表現を使用することが多い。しかし、ビデオデータによって収集された文は日常生活における様々な状況の中での発話であるため、表現が多岐にわたる。これは、ラベル獲得ルールの適用回数が少ないという結果にも結び付くと考えられる。

三つ目は助詞の欠落である。アンケートによって収集された文は書き言葉に近く、助詞が省略されることはまれである。一方、ビデオデータによって収集された文では「らーめんあちちだからふーふー」、「あいすたべたの」などのように助詞が含まれないことが多い。また、入力された文のうち全体の42.6%にあたる23文は本来必要であるはずの助詞が欠落していた。助詞はそれほど種類が豊富ではないため、SINCAが入力文から共通部分を抽出する際に、ラベルに助詞を伴った形で切り出されてしまうことがある。助詞が含まれない入力文を用いると、ラベル部分に隣接する文字の種類が大幅に増えるため、このような誤りを回避することが可能になり、正確なラベルの獲得が促進される。

このことから、SINCA への入力文には、日本語文法に即した正しい文よりも、大人が日常的に発話しているような文が適していると考えられる。

上記の三つの要因によって、ビデオデータから収集された入力文を用いた場合、共通部分抽出処理を行う際にラベルとして適切な文字列を切り出すことが容易になったと考えられる。

9. むすび

幼児の日常生活を収録したビデオデータを書き起こし、大人の発話の分析を行った。分析の結果、大人は短い発話で幼児に話しかけることが明らかになった。また、大人の発話に形態素解析を行った結果、幼児に対する発話には助詞が少ないことが明らかになった。

上記のデータを入力文として名詞概念獲得システム SINCA を用いた名詞概念獲得実験を行った。SINCA は形態素解析などを用いず、字面の一致のみを手がかりに名詞の候補を抽出する。したがって、名詞に隣接する文字が多様な(あるいは存在しない)文の集合が理想的な入力となる。これは大人が幼児に話しかける発話文の特徴に合致する。SINCA への入力文には、日本語文法に即した文よりも、大人が幼児に対して日常的に発話しているような文が適していることが示唆された。

一方で、日本語において助詞は動詞の格関係を示すなど、文の成立に重要な機能を果たしている。幼児が動詞を獲得する過程では助詞から得られる情報も必要だと考えられる。助詞と名詞獲得、動詞獲得との関係については今後の研究課題である。

参考文献

- [1] (社)日本機械工業連合会, (社)日本ロボット工業会, “平成12年度21世紀におけるロボット社会創造のための技術戦略調査報告書”, <http://www.jara.jp/other/dl/rt.pdf>(2001).
- [2] 内田ゆず, 荒木健治, “画像に対する発話を対象とした名詞概念獲得システム SINCA”, 知能と情報(日本知能情報ファジィ学会誌), Vol.20, No.5, pp.3-13(2008).
- [3] Yuzu Uchida and Kenji Araki, “A System for Acquisition of Noun Concepts from Utterances for Images Using the Label Acquisition Rules”, Springer-Verlag Lecture Notes in Artificial Intelligence (LNAI) 4830, pp.798-802(2007).
- [4] 末永俊郎, “社会心理学研究入門”, 東京大学出版会(1987).
- [5] Joseph Weizenbaum, “ELIZA - A computer program for the study of natural language communication between man and machine”, Communications of the Association for Computing Machinery, vol.9, no.1, pp.36-45(1966).
- [6] 須賀哲夫, 久野雅樹, “ヴァーチャルインファントー言語獲得の謎を解く”, 北大路書房(2000).
- [7] Terry Winograd, “Understanding Natural Language”, Academic Press Inc., Orland, FL, USA(1972).
- [8] Naoto Iwahashi, “Interactive learning of spoken words and their meanings through an audio-visual interface”, Trans. IEICE, vol.E91-D, no.2, pp.312-321(2008).
- [9] Deb Roy, Alex Pentland, “Learning Words from Sights and Sounds: A Computational Model”, Cognitive Science, 26(1), pp113-1468(2002).
- [10] Brian MacWhinney, “The CHILDES Project: Tools for Analyzing Talk”, Second Edition, Hillsdale, N.J.: Lawrence Erlbaum Associates(1995).
- [11] 小林哲生, 永田昌明, “ウェブを通じた初期語彙発達データ 収集の試みとその応用”, 日本赤ちゃん学会第8回学術集会, p.73(2008).
- [12] 天野成昭, 近藤公久, 加藤和美, “NTT 乳幼児音声データベースの構築”, 信学技報, vol.108, no.50, TL2008-6, pp.29-34(2008).

- [13] Alden J. Moe, Carol J. Hopkins, R. Timothy Rush, "Vocabulary of First-Grade Children", Charles C. Thomas Publisher, Springfield, IL(1982).
- [14] International Computer Archive of Modern English(ICAME), The ICAME Corpus Collection on CD-ROM, Version 2(1999).
- [15] Anna-Brita Stenstrom, Gisle Andersen, Ingrid K. Hasund, "Trends in Teenage Talk: Corpus Compilation", Analysis and Findings (Studies in Corpus Linguistics), John Benjamins Publishing Company, Amsterdam(2002).
- [16] 内田ゆず, 荒木健治, "幼児の普通名詞及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報告, vol.1, pp.21-27(2005).
- [17] 正高信男, "0歳児がことばを獲得するとき—行動学からのアプローチ", 中公新書(1993).
- [18] 内田ゆず, 荒木健治, "画像に対する発話からの名詞概念獲得システムにおける音素認識の導入について", 言語処理学会第13回年次大会発表論文集, pp.994-997(2007).