

マルチモーダル幼児行動コーパスを用いた発話分析

桐山 伸也[†] 竹林 洋一[‡] 堀内 裕晃[†] 石川 翔吾^{*} 笠見 朋彦^{*} 北澤 茂良[†]

[†] 静岡大学情報学部

[‡] 静岡大学創造科学技術大学院

^{*} 静岡大学大学院情報学研究科

あらまし 人間の根源的なコモンセンス知識の抽出と良質な幼児学習環境の構築を目的としたマルチモーダル幼児行動コーパスを蓄積している。本稿では、本コーパスを用いた多角的な幼児発話分析について、発話行動の発達を可視化するシミュレーションモデル構築を目標とした音声言語獲得過程のマルチモーダル観察と、音声言語理解のための心的状況表現モデルの検討の2つを説明する。前者について、1歳児の指示表現の獲得過程をマルチモーダルに観察し、音韻・視線・身振りの発達変化を明らかにした。また観察結果を基に、幼児の保護者や幼児教室の講師向けに幼児の行動や意図を説明できる行動シミュレーションシステムの実現に向けた方策を検討した。後者について、2歳児の発話中の指示語観察に基づき、心的状況を記述するモデルを提案した。実例での検証により、提案モデルがダイナミックに変化する心的状況を表現できる能力を持つことを示した。

キーワード マルチモーダル幼児行動コーパス, 音声言語獲得,
心的状況記述モデル

Utterance Analyses using Multimodal Infant Behavior Corpus

Shinya Kiriya[†] Yoichi Takebayashi[‡] Hiroaki Horiuchi[†] Shogo Ishikawa^{*}
Tomohiko Kasami^{*} Shigeyoshi Kitazawa[†]

[†] Faculty of Informatics, Shizuoka University

[‡] Graduate School of Science and Technology, Shizuoka University

^{*} Graduate School of Informatics, Shizuoka University

Abstract This paper describes studies on infant utterance analyses using our constructing multimodal infant behavior corpus, which consists of audio and video data recording real-world infant classrooms, and includes annotation data from multiple viewpoints. Two studies have proceeded; one is multimodal observation on spoken language acquisition processes aiming to construct behavior models for developing a behavior simulation system. The other is to develop mental state description models for spoken language understanding. As for the former, multimodal observation revealed infant development in utterance, eye, and gesture at demonstrative expressions. The results also contributed to studying how to realize a behavior simulation system. For the latter, analysis results on demonstratives proposed a model for explaining mental state transition.

Key words multimodal infant behavior corpus, spoken language acquisition, mental state description model

1. はじめに

人間の根源的なコモンセンス知識の抽出と良質な幼児学習環境の構築を目的として幼児学習プロジェクトを進めている[1]。学内に幼児教室を設置して2005年6月から定期開催しており、複数カメラ・マイクによる行動収録環境と、親、先生、研究者による多層アノテーションによる行動観察環境を構築し、経年データに注釈を付与したマルチモーダル幼児行動コーパスの蓄積を進めている。幼児に専用マイクを装着させて収録した音声データをアノテートした幼児発話データは、本コーパスの主要部分である。

本稿では、構築したコーパスを用いた多角的な幼児発話分析について、**発話行動の発達を可視化するシミュレーションモデル構築を目標とした音声言語獲得過程のマルチモーダル観察と、音声言語理解のための心的状態表現モデルの検討**の2つを紹介する。

音声言語獲得過程の解明に関する研究は歴史が古く、本研究のように、幼児の観察に基づく研究は数多くなされてきた[2]。幼児の音声コーパス構築に基づく言語獲得研究の例としては、CHILDES プロジェクト[3]が世界的に有名である。各種の言語で音声コーパスと分析用のツール群をボランティアベースで開発し、研究共有リソースの整備を狙っている。日本語の音声コーパスとしては、NTT 日本人乳幼児音声データベース[4]がある。0歳から5歳までの長期に渡って月1時間以上のデータを収録している。しかしながらこれらのコーパスに基づく言語発達研究は、音声言語の単一の側面に着目するものであり、映像や各種センサ情報を包括的に活用したマルチモーダルな行動記録による観察を指向するものではない。

一方、MITのRoyらのグループでは、幼児の自然な行動を分析対象とし、生後～3歳までの幼児の行動を映像と音声で全て大容量ストレージに記録し、言語学習のモデル構築を狙ったプロジェクトを進めている[5]。しかしながら、心の内面に踏み込んだモデル構築を指向するものではない。また産総研の西田らは、乳幼児の事故防止のための行動シミュレーションシステムを開発[6]しており、実用性の高い応用ドメインでのサービス提供を狙っている。しかし、こちらも心的状態の考察は未着手であり、行動内部モデルの精緻化が課題である。

以下、2節で幼児発話の収録方法を説明する。3節で本コーパスを用いた幼児の音声言語獲得過程の分析について、4節で音声言語を手掛かりとした幼児の心的状況の記述モデルの検討につい

て述べる。

2. 幼児教室における発話収録

2005年6月より毎週1回幼児教室を定期開催している。2007年3月現在、1～3歳児を対象として年齢別に3クラスを設け、それぞれ親子3組ずつが参加している。21ヶ月に渡って通算146回の幼児教室を開催し、217時間分の幼児行動を映像・音声データとして収録した。

幼児教室は杉材のやぐらの内部に設置されている。やぐらの梁や柱に複数のカメラやマイクを設置して、幼児行動を収録している[7]。幼児発話の観察には雑音の入らない質の良い音声データが不可欠であるが、やぐらの梁に埋め込んだマイクでは、ノイズが大きく、音量が不安定で、発話者の特定が困難という問題があった。

これらの問題を解決するために、図1に示すような、ウェアラブル音声収録装置を開発した。左右の肩に近い部分にコンデンサマイクが仕込んであり、取り込んだ音声は、リュック内部のボイスレコーダに保存される。この装置により、ノイズは17dB程度軽減され、音量が格段に安定し、発話者の特定が容易にできることを確認した。

3. 音声言語獲得過程の観察

幼児の行動の意図やその発達変化を客観的に提示できるシステムが実現できれば、幼児の保護者や幼児教室の講師にとって有益である。この観点から、幼児行動のシミュレーションモデル構築を目標として、マルチモーダルな幼児行動観察を実践した。

具体的には、コーパス中の1歳児・2歳児の発話データを用いて、指示表現の獲得過程を観察



図1: ウェアラブル音声収録装置

した。ここでいう指示表現とは、発話や身振り・視線などにより、話し手が着目している特定の対象を聞き手に伝える表現全般を指す。

3.1 行動観察の方法

以下の手順に従って行動観察を実施した。

- (1) 指示語発話・手差し・指差しのいずれかが発現している発話行動をコーパスから抽出：1歳児の幼児1名について月齢14ヶ月～23ヶ月の10ヶ月間のデータを分析対象とし、1月当たり30分の授業時間の映像を観察して、240個の発話を抽出。
- (2) 自然言語によるアノテーションの付与：発話がなされた状況、発話内容、韻律的特徴、幼児の行動などの項目を発話別に記述。
- (3) 発話を特徴付ける記述項目の考察：自然言語記述のアノテーションを分析し、月齢・発話意図・発話内容・韻律・視線・身振りの6項目を指示表現の記述に必要な項目と決定。
- (4) 行動記述モデルの構築：(3)で決定した記述項目別に自然言語アノテーションを整理し、各々の語彙リストを決定して行動記述モデルを構築。
- (5) 行動記述モデルに従い、自然言語アノテーションを変換

行動記述の対象を、図2に示すような机上での授業の場面に限定した。これは、行動をシミュレートする場面に限定し、記述モデルの表現能力を明確にするためである。母親と先生を区別せず、両者を一括して「相手」と記述し、授業に用いられる教材や道具はすべて「対象物」と表現した。構築した行動記述モデルを表1、それによる行動記述例を表2にそれぞれ示す。

3.2 マルチモーダル行動観察結果

音韻獲得の過程[8]

最も頻出する /kore/ という指示詞の発話に着目し、音素単位でその単語を獲得するまでの変化を追跡した。当該発話が指示詞かどうかは

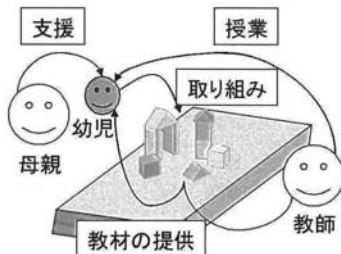


図2：行動記述の対象場面

表1：行動記述項目一覧

月齢	発話	韻律	目線	身振り	意図
14ヶ月	二語発話	平坦型	対象物	指差し	発見
15ヶ月	指示語	発見型	相手	指差し(交互)	欲求
16ヶ月	短母音	強調型	対象物→相手	指差し(詳細)	要望
17ヶ月	その他	呼びかけ型	対象物→対象物	指差し(叩く)	意見
18ヶ月		断定型	相手→対象物	取る	
19ヶ月		疑問型		身を乗り出す	
20ヶ月				差し込む	
21ヶ月				叩く	
22ヶ月				見せる	
23ヶ月					

表2：発話行動の記述例

月齢	発話	韻律	目線	身振り	意図
16ヶ月	もー	平坦型	遠くの対象物	指差し(交互)	意見
17ヶ月	おー	平坦型	対象物	指差し(詳細)	意見
18ヶ月	ぶーばっ	強調型	対象物→相手	手差し	欲求
18ヶ月	ぶー	平坦型	対象物→相手	指差し(詳細)	意見
18ヶ月	あ	断定型	対象物	指差し(詳細)	発見
18ヶ月	むた	強調型	対象物→相手	取る	欲求
23ヶ月	これ	断定型	対象物	指差し	要望
23ヶ月	ここ	断定型	相手→対象物	指差し(叩く)	意見
23ヶ月	これ	断定型	対象物→相手	指差し	発見
23ヶ月	これ	呼びかけ型	対象物→相手	指差し(叩く)	要望

手差し・指差しなど他のモダリティ情報から包括的に判断している。

音韻獲得の過程を図3に示す。月齢14ヶ月では、単純に母音を組み合わせた発話であり、子音は観測されなかった。1ヶ月程度経過すると不完全な子音を含む発話に変化する。月齢18ヶ月で、子音も安定して観測できるようになる。この頃には、独り言のような発話で /koe/ と /goe/ を連続して発声するなど両者の使用を決めかねている場面も見られた。音声波形にも /g/ のような有声破裂音のような子音が確認されている。その後、一時は /goe/ が頻繁に見られる

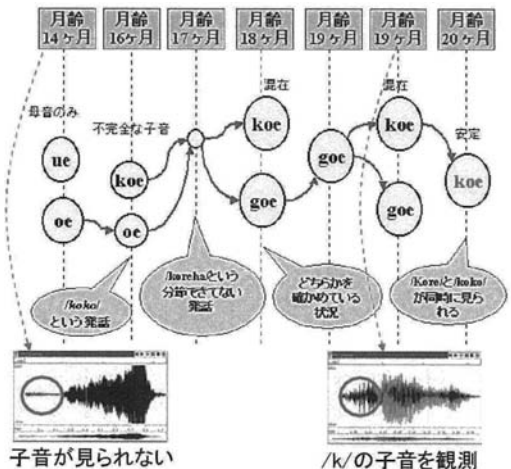


図3：指示詞「これ」の音韻変化

時期もあったが、月齢 20 ヶ月になると、/koe/ という正しい子音を安定して発話できるようになった。

視線・身振りの発達過程

視線の発達変化を図 4 に示す。月齢 15 ヶ月の時点では対象物のみを見ていたのが、成長するに従って相手を見るようになり、月齢 17 ヶ月では相手を見て相手の注意を引いてから対象物に視線を移すという動作も見られるようになる。

図 5 は身振りの発達変化を示す。月齢 14 ヶ月ですでに、欲しいときには手差しで発見したときなどは指差しを使うなどの使い分けを行っていることを確認した。さらに観察を進めていくと、欲しいときに手差しを使っていたのが、先生がその対象物をしまおうとしたときに指差しに変化するというように瞬間的に手差しと指差しを切り替えるというしぐさが見られた。

また、最初は対象全体を指差すことしかできなかったが、月齢 15 ヶ月になると対象の特定部分に指差すことができるようになった。

3.3 意図の発達に着目した観察

表 2 の行動記述データは、月齢と意図を入力として与えたとき、発動すべき行動の発話・韻律・視線・身振りを出力する行動シミュレーションのためのモデルとして活用できる。しかし、このモデルで表現できるのは、単発のアクションに過ぎず、欲しいものを入手する、相手の注意を引く、といったゴールを達成する過程での相手とのインタラクションを表現できない。

この観点から、単発でない、ゴールに向かう一連の行動を表現できるモデル構築へ向けた検討を実施した。幼児が「欲しいものを入手す

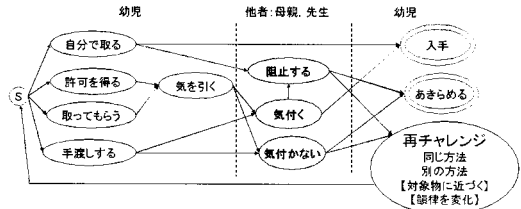


図 6: 意図の発達に着目した行動記述の検討

る」というゴールを持つ場面を相手とのインタラクションに着目して詳細に観察し、図 6 に示す状態遷移図にまとめた。行為の途中で意図が変わった場面や創意工夫する過程を記述できるため、成長による意図の発達を観察できる。

例えば、図 6 の「自分で取る」、「手渡しする」という行動は、初期には見られなかったものである。何かを欲しがるときに最初は相手に働き掛けるだけだったのが、成長によって自主的に物を取得する場面が見られるようになった。

4. 発話行動における心的状況の記述

発話から相手の心的態度を読み取って気の利いた応答ができるシステムの実現に向けた音声言語理解の枠組みを検討するため、指示語を伴う発話行動のマルチモーダル観察により、心的状況の記述モデルを検討した。ここでいう心的状況とは、指示語で言及される対象に対する発話者の思い入れの深さを意味する。すなわち、「これ」「あれ」「それ」という言語表象に込められた発話者の内面を、マルチモーダルな視点で客観的に記述する方法を検討した。

4.1 情報のなわ張り理論に基づく心的状況記述

言語学的な視点で構築された、神尾による「情報のなわ張り理論」[9]を行動観察に適用し、行動学・心理学的な視点を加えたマルチモーダルな観点からの行動分析を実施する。この理論では、話し手あるいは話し手が推測する聞き手にとって心的に近い情報が<内>、心的に遠い情報が<外>と定義される。話し手・聞き手と情報の心的な距離感を表現したものである。図 7 に例を示す。話し手にとって、「これ」という言葉は近い情報であり、「あれ」は遠い情報で

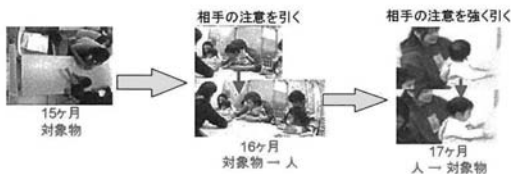


図 4: 視線の発達変化



図 5: 身振りの発達変化



図 7: 情報のなわ張り理論による指示語の表現



図 8: 心的状況記述モデル

ある。「それ」は、聞き手にとって近い情報について話し手が言及する場合に使用される。

4.2 心的状況記述モデルの検討

3 節と同様の手順により、心的状況記述モデルを検討した。具体的手順は以下ようになる。

- (1) 指示語発話の抽出：収録データを書き起こし、2 歳児 1 名 (2.0 歳～2.11 歳の 1 年間) の指示語を含む発話を 380 個抽出した。発話の内訳は、指示形態素「こ」を含む「こ」系の発話が 352 個、「あ」系が 28 個、「そ」系は 0 個であった。「そ」系は 3 歳児の発話に例が見られるが、数は少ない。今後も観察を継続することで、「そ」系の獲得変化の分析が可能である。
- (2) 自然言語によるアノテーション付与：抽出した指示語発話について、その発話がなされた状況 (話し手・聞き手・対象物・それらの位置関係など客観的に把握できる特徴) を自然言語で記述。
- (3) 想定される心的状況のカテゴリへの分類：自然言語アノテーションを手掛かりに、各々の発話の状況が、図 8 の 6 パターンのどれに属するかを考察し、分類。「話」は話し手、「聞」は聞き手、「★」は対象物をそれぞれ表す。実線が話し手のなわ張り、点線が聞き手のなわ張りであり、実線・点線の内側が<内> (= 心的に近い)、外側が<外> (= 心的に遠い) を意味する。実例を観察してみると、聞き手が対象物に気づいていないという場面が少なくなく、この場合を表現するために、<認識なし>を

表す「X」を新たに導入。

- (4) 心的状況記述モデルの構築：各カテゴリの自然言語アノテーションを吟味し、心的状況を表すカテゴリへの分類に必要な手掛かりを洗い出し、整頓して心的状況記述モデルを構築。

- (5) 心的状況記述モデルに従い、自然言語アノテーションを変換。

構築した心的状況記述モデルが図 8 である。しぐさ、対象物の属性、対象物との物理距離、取り組み、前の状況、後の行動という項目をセンシングすることにより、心的状況を推測できるモデルとなっている。

各カテゴリが具体的にどのような心的状態に対応するかの意味づけが重要である。380 発話を分類した状態で考察し、6 カテゴリが概ね「指向型」「欲求型」「共有型」「希薄型」「説明型」「配慮型」という特徴を持っていることを確認した。

4.3 心的状況の変化の分析

図 9 に心的状況記述の例を示す。場面 1 は、幼児が遠くの壁にかかっているおもちゃが気に

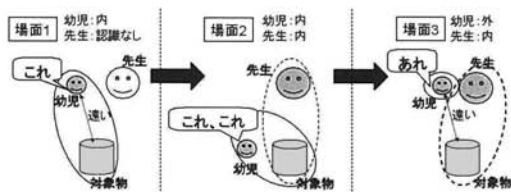


図 9: 指示詞を発話する際の心的状況の変化

なり、先生の注意を引くためにおもちゃを「これ」と指差したが、先生はそのおもちゃに気づいていないという状況である。場面2は、幼児の「これこれ」という強い注意喚起の発話を受けて、先生がおもちゃに気づいた状況を示す。場面3では、先生の「後でね」という抑制を受け、おもちゃを指示する発話が「あれ」と変化した状況を示す。場面1から場面2への変化は、心的状況と物理的状況のギャップによって生じる行動である。つまり、そのギャップを埋めるために近づくという行動が起こったと考察される。さらに所有という点から考察すると、幼児が、先生の心的状況が<認識なし>から<内>になったのを認識することで、より所有権を得るために対象物に近づき主張するという行動であると分析される。場面2から場面3への変化は、願望の抑圧という心的状況の変化を、心的状況の<内>から<外>への変化という形で表現できる。このように、構築した心的状況記述モデルは、心的状況のダイナミックな変化を表現する能力を持つことがわかる。

5. おわりに

マルチモーダル幼児行動コーパスの発話分析に基づき、指示表現の獲得過程の観察、心的状況の記述モデルの検討を行った。音韻獲得の様子や、視線や身振りの発達過程が、発話を手掛かりとしたマルチモーダルな行動観察によって観測できた。観測結果は、単発のアクションとしての行動に止まらず、ゴール達成に向けたインタラクション行動の分析にも有益であり、成長に伴って行動の意図が複雑化する様子も明らかになった。さらに、観測結果を幼児の行動モデルとしてシミュレーションシステムに活用できる見通しを得た。音声言語理解モデルの構築を狙い、発話中の指示語から心的状況を推定する方策を検討した。言語的な表象をトリガとした行動観察によって心的状況を表現する枠組みである。実例での検証により、他者の抑圧によって対象物に対する執着の強さが変化する様子が観測でき、他者とのインタラクションによって動的に変化する心的状況を表現できる能力を持つことがわかった。今後は、行動シミュレーションのプロトタイプ構築へ向けた検討を進める。

参考文献

- [1] 竹林 他: “幼児のコモンセンス知識抽出のためのマルチモーダル幼児行動収録観察環境の構築,” 音講論(春), 2-9-11, 2007.

- [2] Oller, D. K., “Metaphonology and infant vocalizations,” *Precursors of Early Speech*, pp.21-35, 1986.
- [3] Mac Whinney, B. (1995) *The CHILDES Project: Tools for Analyzing Talk*, second edition. Lawrence Erlbaum Associates.
- [4] 麦谷綾子, “乳幼児の音声言語獲得,” 電子情報通信学会技術研究報告. TL, 思考と言語 104(316),13-18, 2004-9.
- [5] Deb Roy, Rupal Patel, Philip DeCamp, Rony Kubat, Michael Fleischman, Brandon Roy, Nikolaos Mavridis, Stefanie Tellex, Alexia Salata, Jethran Guinness, Michael Levit, Peter Gormiak. (2006). *The Human Speechome Project. the Proceedings of the 28th Annual Cognitive Science Conference.*
- [6] K. Kitamura, Y. Nishida, N. Matsumoto, Y. Motomura, T. Yamanaka, H. Mizoguchi, “Development of Infant Behavior Simulator: Modeling Grasping Achievement Behavior Based on Developmental Behavior Model and Environmental Interest Induction Model,” *Journal of Robotics and Mechatronics*, Vol.17, No.6, 2005
- [7] 山本剛, 仲川淳, 坂根裕, 坂根信一, 桐山伸也, 竹林洋一: マルチモーダル情報を利用した幼児の行動観察の検討, 人工知能学会第20回全国大会, 1C2-1 (2006.6).
- [8] R. Tsuji, et al.: “Observations of the Spoken Language Acquisition Process based on a Multimodal Infant Behavior Corpus,” *Interspeech2006*, 2006.
- [9] 神尾昭雄: 情報のなわ張り理論, 大修館書店, 1990.