

絵本読み聞かせ場面の認識および発達順序体系を用いた発達段階推定

Estimation of the Developmental Stage based on the Recognition of the Scene of Picture Book Reading and Order of Developmental Stages

笠松 美歩[†] 宇津呂武仁[†] 齋藤 有^{††} 石川由美子^{†††}

[†] 筑波大学大学院 システム情報工学研究科 知能機能システム専攻
〒 305-8573 茨城県つくば市天王台 1-1-1

^{††} 聖徳大学 児童学部 〒 271-8555 千葉県松戸市岩瀬 550

^{†††} 宇都宮大学 共同教育学部 〒 321-8505 栃木県宇都宮市峰町 350

あらまし 発達心理学において、子どもは一定の順序で段階的に発達するとされている。本論文では、読み聞かせ中の子どもの反応を含めた発達順序体系を構築し、絵本読み聞かせという短時間の活動に対する観察から発達段階を推定する枠組みを提案する。さらに、絵本読み聞かせ場面の映像観察に要する時間を短縮するために、画像認識により要観察時点を推定する方式を導入し、その有用性の評価結果を示す。

キーワード 絵本, 認知発達の反応, 発達段階, 映像認識

1 はじめに

発達心理学において、子どもは一定の順序に従って段階的に発達し、その段階に応じた様々な反応を示すとされている。そのため、各子どもの発達がどの段階にあるかという情報は、発達心理学研究において非常に大きな意義を持っている。子どもは月齢に従って発達するが、その速度には個人差があるため、子どもの月齢のみで発達段階を決定することはできず、子どもの発達段階を知るためには、各個人に対して発達段階の診断を行う必要がある。この診断の方法としては、専門家が実際に子どもを 30~60 分程度観察する方法や、養育者や保育者が 100 以上の項目を含む質問紙に回答し、それを分析する方法等があるが、これらの実施には日常生活から離れた長時間の活動が必要であり、養育者・保育者および研究者にとって大きな負担となるため、容易に実施することができない。それに対して本論文は、こうした発達診断にかかる労力と時間を削減するために、日常生活の中で広く行われ、所要時間が短く、子どもの認知発達の反応を観察可能な場面である絵本の読み聞かせと、子どもの発達順序体系を用いて発達段階を推定する手法を提案する。

絵本の読み聞かせとは、養育者・保育者等が子どもに対して絵本を音読して聞かせる活動であり、家庭や保育現場における日常生活の中で広く行われ、その一回当たりの所要時間は一般的に数分程度である。この活動では、読み手から子どもへの絵本内容の伝達が行われるだけでなく、読み手と子どもとの間に内容に関する対話や質問等の相互作用が生じ、その中で子どもは指さしやページをめくる等の様々な認知発達の反応を示す。発達心理学においては、これらの絵本に関する子どもの発達と、

日常生活における子どもの発達との間には順序性があることが示されている [2]。そこで本論文ではまず、絵本に関する子どもの反応と、日常生活における子どもの反応の両方を含んだ子どもの発達順序体系を構築する。この体系の構築においては、絵本に関する反応と、それ以外の日常生活における反応について、各反応を子どもが示すかどうかを質問する質問紙調査を行うことで、各反応を子どもが示すようになる月齢を調査し、それらを月齢順に並べることにより、子どもが各反応を示すようになる順序を示した。次に本論文では、絵本読み聞かせ場面を収録した映像を分析し、そこに出現した反応を子どもが示すようになる月齢と、作成した発達順序体系とを対応付けることで、3 人の子どもに対して発達段階の推定を行った。この分析においては、映像中の指さしを自動検出することによる要観察時点の推定を行い、推定された場面の反応への観察のみを用いた発達段階の推定と、映像中に出現する全ての反応についての観察を用いた発達段階の推定を行った。その結果、いずれの手法においても、子どもが発達順序体系中の反応を示しているかどうかを高い精度で推定することができた。これにより、本論文の提案手法が子どもの発達段階の推定に対して有用であることを示すことができた。

2 子どもの発達順序体系

2.1 既存の発達順序体系

2.1.1 石川 [2] の体系

石川 [2] は、子どもの日常生活における発達と、絵本読み聞かせにおける発達との間には一定の順序性があることを明らかにした。石川 [2] では、0~74 か月の子どもを持つ 858 名の保護

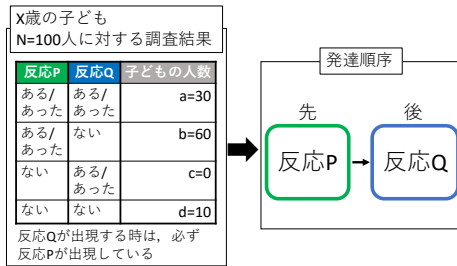


図1 Ordering Analysis [1] の概略

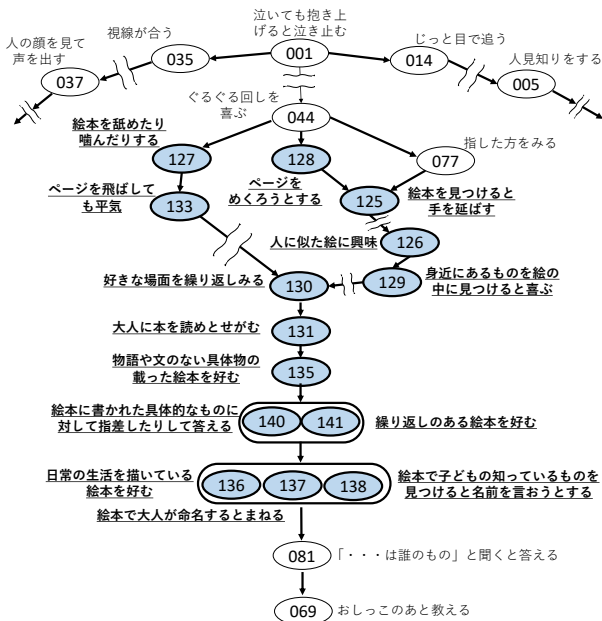


図2 石川 [2] における発達項目およびその順序性(抜粋)(下線・太字は絵本関連項目)

者に対して質問紙調査を実施し、0~74 か月の子どもが、98 の項目(日常生活関連 75 項目、絵本関連 23 項目)を通過する順序を調査している。この質問紙調査は、各質問項目を子どもが示すかどうかを「ある」、「あった」、「ない」の3段階で回答する形式をとっており、回答が「ある」または「あった」である場合に、子どもがその項目を通過していると定義している。そしてこの回答に対して、Ordering Analysis [1] の手法を適用し、各項目間の順序を判定している。この手法によって、項目 P と項目 Q の2項目の順序を判定する方法の概略を図1に示す。また、石川 [2] における子どもの発達の順序性を示すグラフの抜粋を、図2に示す。なお、この図中において、下線太字になっているものが絵本関連項目、それ以外が日常関連項目である。このグラフから、絵本に関する項目は、それ自身で独立したのではなく、日常生活における発達と関連を持ちながら通過されるものであることがわかる。なお、石川 [2] においては、子どもが各項目を通過する順序は示されているが、子どもが各項目を通過する月齢は示されていない。

2.1.2 KIDS 乳幼児発達スケール [5]

KIDS 乳幼児発達スケール [5] は、1989 年の調査に基づいて標準化された、子どもの発達段階を診断するための質問紙形式のテストである。このテストには、月齢および発達遅延傾向の

有無に応じ、タイプ A(0 歳 1 か月 ~ 0 歳 11 か月児用、117 項目。)、タイプ B(1 歳 0 か月 ~ 2 歳 11 か月児用、142 項目。)、タイプ C(3 歳 0 か月 ~ 6 歳 11 か月児用(就学児を除く)、133 項目。)、タイプ T(0 歳 1 か月 ~ 6 歳 11 か月児用(発達遅滞傾向児向き)、282 項目。)が用意されており、対象児に適したタイプの質問紙に対して、養育者・保護者等の日常的に対象児を観察している者が回答する。この質問紙においては、各質問項目の反応を子どもが示すことができるかどうかを、「○(できる)」、「×(できない)」の2段階で回答する。KIDS で使用される質問紙は、全国 38 都道府県の 6,090 名の幼児に対する調査によって標準化されており、0 歳 1 か月 ~ 6 歳 11 か月の幼児が、各月齢においてどのような項目を通過するかが示されている。KIDS の質問紙には、各タイプごとに、子どもができると回答された項目の数と、発達段階とを対応づけるための表が用意されており、テスト実施者は、質問紙の回答とこの表とを照らし合わせることで各子どもの発達段階を診断する。KIDS は、専門家による長時間の子どもの観察が不要であるという点で、比較的实施が容易な診断法であると言えるが、各タイプにおける質問項目数はいずれも 100 以上であり、乳幼児の養育者・保育者がこうした質問紙に回答するのは負担が大きい。

2.2 2020-2021 年度調査をふまえた発達順序体系

2.1.1 節、2.1.2 節で示したように、発達心理学分野においては、子どもの発達順序についての体系が存在する。しかし、これらの体系が作成されてから 2021 年現在までの間に、20 年以上が経過しているため、社会的な変化等の影響により、子どもの発達の順序体系にも変化が生じている可能性がある。また、これらの発達順序体系は子どもの発達段階を全て網羅しているとはいえず、さらにこれらの発達順序体系同士の関連も不明である。そこで本論文では、石川 [2] および KIDS の発達順序体系における項目と、独自に作成した項目を用いて、2020 年から 2021 年にかけて 0~41 か月の子どもを持つ母親 1,672 人に対して質問紙調査を実施し、新たな発達順序体系を構築した。この調査においては、石川 [2] で順序性が認められた全 98 項目と、KIDS 中の手指の動きの発達に関連する 64 項目に加えて、「指さし」と「ページをめくる」という2つの反応を細分化することで独自に作成した 32 項目を用いた。なお、この調査における質問紙は石川 [2] における質問紙の形式に準じ、各項目ごとに、「(その項目の反応を示すことが現在)ある」、「(その項目の反応を示したことがかつて)あった」、「(その項目の反応を示したことが)ない」の3段階で回答する形式とした。

調査で用いた全項目について、通過月齢を調査し、それらを通過月齢の昇順で並べた結果を表1、表2に示す。これらの表における項目 ID は、各項目の出典とそこでの番号を表しており、I から始まるものは石川 [2] で扱われた項目、K から始まるものは KIDS において扱われた項目であり、N から始まるものは本論文において新規に作成された項目である。また、通過月齢の判定においては KIDS における通過月齢の決定手法を参考にし、通過率が 60%以上となる月齢が 6 か月以上続くか、41 か月まで連続して通過率が 60%以上となる月齢の範囲の中で最小

表1 2020-2021 年度調査をふまえた発達順序体系 (抜粋) (1)

通過月齢	項目数	各通過月齢の項目 (日常関連項目は一部抜粋)		
		項目ID	領域	項目 (絵本関連は太字. 絵本の指し関連は太字下線.)
0	20	I35	Ⅲ やりとり行動	両親が話しかけたとき, 視線が合う
2	2	KA2-9	2. 操作	両手を触れ合わせたり, からめたりする.
3	3	KA1-9	1. 運動	うつぶせにした時, 両腕で胸や頭を持ち上げることができる.
4	2	N2	日常の指し	手を開いたままで, 興味をもった遠くにある物に向かって手を伸ばす.
5	1	N3	日常の指し	欲しい物や行きたいところが遠くにあつて手を伸ばしたり, 指さしたりした時, 欲求がすぐにならえられないと泣くなどして不満を示したり, 催促したりする.
6	6	I127	Ⅶ 絵本	絵本を持たせると, かどを舐めたり噛んだりする.
		I128	Ⅶ 絵本	自分でページをめくろうとする.
		I125	Ⅶ 絵本	手の届かない所に絵本を見つけると手をのびたり, 色彩のはっきりした絵に目を留めて手足をばたばたさせたり「アーアー」「ウーウ」などと声をだしたりする.
		I133	Ⅶ 絵本	絵本を読んでいるとき, ページをとばしても全然平気で見ている.
7	9	KA2-13	2. 操作	紙を引っぱってやぶる. (読んでいる新聞など)
8	3	KA2-18	2. 操作	落ちていた小さな物をひろう.
9	5	N4	ページをめくる	まだうまくページをめくれないが, 絵本の読み聞かせをしている時, 子どもが自分でページをめくろうとする.
		N5	ページをめくる	絵本の中身を見ているのではなく, ただページをめくるのを楽しんでいる様子でてたらめにページをめくる.
10	6	N7	日常の指し	はっきりと何かを指しているわけではないが, 手を指さしの形にしていることがある.
11	5	N10	日常の指し	手を指さしの形にして, 興味をもった遠くにある物に向かって手を伸ばす.
12	18	I126	Ⅶ 絵本	人の顔にいた形や表情のある絵に興味を示す.
		N11	絵本の指し	絵本中の絵を指さし, その時に「あ, あ」などの意味のない声を出す.
		N12	絵本の指し	明確に何を指しているかわからないが, 絵本に指さしをする.
		I131	Ⅶ 絵本	大人のところに本を持ってきて, しきりに読めとせがむ.
13	3	I54	IV 伝達行動	自分の要求するものや欲しいもののがはっきりしてきて, 目的のものを指さしたり, 催促したりする(「マンマ」と言って食事の催促をしたりするなど).
				繰り返しの絵と文で構成される絵本を好む (例えば, イナイイナイ … バアの本など).
14	1	KA2-26	2. 操作	積み木を2つ積み重ねる.
15	12	N13	ページをめくる	読み聞かせの際に, 子どもが自分のペースで読んでもらいたくてページをめくろうとする. (早く次の場面を読んでほしくて読み終わらないうちに次のページをめくろうとするなど.)
				N15
		I129	Ⅶ 絵本	身近にあるものを絵の中に見つけると喜ぶ.
16	8	N16	絵本の指し	子どもが絵本の絵の中の見慣れているものを指さした時, 大人がそれに反応すると喜ぶ.
		I130	Ⅶ 絵本	お気に入りの絵や場面がきまりその場面だけを繰り返し見る.
		I135	Ⅶ 絵本	物語や文がほとんどなく, 子供が毎日見ているような日常の具体的なもの (テレビ, ラジオ, 洗濯機, 掃除機, 果物, 食べ物, 三輪車, 自転車, ボールなど) が載っているような絵本を好む.
17	2	KB3-9	3. 理解言語	指定した本を1冊持ってくる.
18	5	I47	Ⅲ やりとり行動	大人の言った単語をそのまままねて繰り返す.

表2 2020-2021 年度調査をふまえた発達順序体系 (抜粋) (2)

通過月齢	項目数	各通過月齢の項目 (日常関連項目は一部抜粋)		
		項目ID	領域	項目 (絵本関連は太字. 絵本の指し関連は太字下線.)
19	4	I137	Ⅶ 絵本	絵本を見て大人が絵の中のものの名前を命名すると, それをまねて言おうとする.
		I140	Ⅶ 絵本	絵本に描かれている具体的なものにたいして, 大人が「ボールどこ」「ラジオどれ」等と聞くと, 指差したり, 答えたりできる.
		I138	Ⅶ 絵本	日常の具体的なものなどが描かれている絵本のなかに子供の知っているものを見つけると, 自分からそのものの名前を言おうとする.
20	1	I70	IV 伝達行動	「きれいね」「おいしいね」などという表現ができる.
21	1	I136	Ⅶ 絵本	子供が毎日, うちで体験しているような日常生活を描いている絵本 (お着替え, お風呂, ごはんを食べるなど) を好む.
22	8	N18	絵本の指し	絵本中の絵を指さし, 意味のある言葉を話す.
		N22	絵本の指し	絵本の絵の中の物を指さして, 指さした物の名前を言う.
		N23	絵本の指し	絵本に描かれている具体的なものにたいして, 大人が「ボールどこ」「ラジオどれ」等と聞くと, 絵を指さして答えることができる.
23	3	N25	絵本の指し	絵本の絵の中の物を指さして, 指さした物について説明する. (例: 泣いている顔を指さして「泣いている」と言う, 犬を指さして「ワンワンいた」と言うなど.)
24	9	I145	Ⅶ 絵本	日常的なもの身近なものが絵本の対象となっていなくても (冒険物語, 想像的な絵本) 興味をもって見ることができる.
		I134	ページをめくる	子どもが自分で絵本のページをちゃんとめくることができ, とばすと気づいてもとに戻ることができる.
25	3	N28	絵本の指し	絵本の中の絵を含む複数の物を同時に指さしたり, 順番に指さしたりする. (花を指さした後別の花を指さす, 人形の目, 手, 足などの一部を順に指さす, 「これが大きい」などと言いながら指さす, コップとポットなど関連のあるものを順に指さすなど.)
26	7	I139	Ⅶ 絵本	日常生活を描いている絵本を見て, 自分の生活に照らし合せ「これはパパが好き」「これはきれいだね」と言ったようなことを言う.
		N29	絵本の指し	絵本にでてくる物を指さして「それは, 何?」などと自分から質問することができる.
		I142	Ⅶ 絵本	単純な繰り返しのある文のある絵本でも (例えば, いいチョッキだねちょっと着せてよ … ちょっときついが似合うかなーの繰り返しなど), 大人が最初のページの部分を絵を見せながら読むと, 次に来る場面がわかり, 一緒に言ったり, 笑ったりできる.
		I147	Ⅶ 絵本	単純な繰り返しのある文のある絵本で, 大人が最初の場面を読んでから, 「次どうなるんだっけー」等の質問をすると, それに対して答えることができる.
28	2	KB1-16	1. 運動	鉄棒にぶらさがれる.
29	2	KB2-14	2. 操作	物をハンカチや新聞紙に包んで遊ぶ.
30	3	KB2-16	2. 操作	砂場で山を作る.
31	2	I143	Ⅶ 絵本	絵本を大人に読んでもらっているとき, 次の場面を先取りして説明したりすることがある.
		I146	Ⅶ 絵本	絵本の文中にでてくる知らないことばに対して「それは, 何?」などと質問したりできる.
33	1	I105	VI 遊び	鉛筆やクレヨンで丸を描くことができ, ときに丸の中に目や口らしいものをつける.
36	1	KC2-6	2. 操作	クレヨンで色を使い分けて絵を描く.
37	2	KB2-23	2. 操作	折り紙を半分に折ることができる.
38	1	I144	Ⅶ 絵本	かなり長い文章のある物語絵本でも, 内容を楽しむことができる.
39	6	I31	Ⅱ 興味・関心	時計を見て何時に興味を持つ.
40	5	I100	VI 遊び	まる, 三角などの形を描く.
42以上	19	I149	Ⅶ 絵本	絵本に描かれている文字を意識し, ひろい読みを自分から始める.

の月齢 X を項目 T の通過月齢とした。

なお、KIDS で扱われた項目について、本論文の調査および KIDS における通過月齢を比較したところ、本論文の調査における通過月齢が、KIDS における通過月齢よりも 3 か月以上早くなった項目が 13 項目¹、3 か月以上遅くなった項目が 10 項目²存在した。また、石川 [2] で扱われた項目については、絵本に対する興味に関する項目の順序が早くなる傾向があり、0~23 か月を対象とする絵本関連項目 14 項目中、9 項目で順序が早くなっていた (詳細については文献 [3] を参照)。

この体系を用いることで、子どもが各項目を通過する順序と月齢を知ることができるため、この体系中の子どもの反応 T が観察されたとき、その子どもは反応 T よりも通過月齢が低い反応は通過しており、 T よりも通過月齢が高い反応は未通過であると推定することが可能になる。

3 映像自動認識による要観察時点の推定



図 3 STAIR Actions データセット [7] 中の画像 (ファインチューニング用訓練データ. クリエイティブコモンズの YouTube 動画から抽出した静止画.) に対する指さし検出結果例

3.1 YOLO [6]

本論文では、物体検出モデルである YOLO [6] を用いて、絵本読み聞かせ場面における指さしの自動検出を行い、映像中の要観察時点の推定を行う。絵本読み聞かせの映像データの分析において、指さしは非常に重要な意義を持ち、観察が必須となる反応である。そこで、本論文においては、映像データから静止画を抽出し、それらの静止画の中に指さしが含まれるかどうかを自動的に検出することで、指さしが含まれる静止画が抽出された時刻を特定し、その時刻を映像中の要観察時点であると推

1: 「手に握らせた物を持ち上げようとするとしっかり握って離さない」、「授乳時に母親の服などを引っぱる」、「コップなどを両手で掴んで口に持っていく」、「自分で自分の口もとをふこうとする」、「自動販売機のボタンなどを押したがる」、「積み木を 3 つ積み重ねる」、「指定した本を 1 冊持ってくる」、「鉛筆で短いながらも直線を引く」、「転がって動いているボールを捕まえることができる」、「クレヨンで色を使い分けて絵を描く」、「人などを描く」、「脱いだ後、服をたためる」、「自動車、花など思ったものを絵にする (それらしく見えればよい)」が該当した。

2: 「あむけでミルクびんを自分で持って飲む」、「自動車などを手で走らせて遊ぶ」、「鉄棒にぶらさがれる」、「友達と手をつなげる」、「物をハンカチや新聞紙に包んで遊ぶ」、「服のスナップを自分ではずす」、「砂場で山を作る」、「折り紙を半分に分けることができる」、「ハサミを使って紙を切る」、「砂場で水を使って池を作る」が該当した。

定する。この自動検出においては、YOLOv5³ の COCO 2017 データセット [4] による事前学習済みモデルをファインチューニングしたものを使用した。このファインチューニングでは、訓練データとして STAIR Actions データセット [7] を用いた。STAIR Actions は、人が日常的な動作を行っている 5 秒間程度の短い動画を収集したデータセットであり、本論文では、この中の “pointing with finger” カテゴリの各動画から 1 秒ごとに 1 枚の静止画を抽出し、それらの静止画のうち、指さしが含まれていた静止画 471 枚を訓練データとした。ファインチューニングにおいては、各静止画に対して指さしが含まれている矩形範囲の座標の情報を付与するアノテーションを行い、各静止画を入力した時、その静止画中に存在する指さしを含む矩形範囲の座標が出力されるように訓練した。なお、訓練データ中の画像は、1 枚の画像単位で見ると、必ず指さしが含まれる画像 (正例) であるが、画像中には指さし以外の形の手や手以外の物体などが含まれており、画像中の物体単位で見ると、そうした指さし以外の物体が負例となってファインチューニングに用いられている。このファインチューニングによるモデルを用いて、訓練データ中の画像 (クリエイティブコモンズの YouTube 動画から抽出した静止画) に対して指さしの検出を行った結果の例を図 3 に示す。

3.2 画像データセット

本論文では、3 組の母子が絵本読み聞かせを行った場面を映像に収録し、そこから静止画を抽出することで、画像データセットを構築した。3 組の母子における子どもの月齢は、32 か月 (ID1)、26 か月 (ID2)、22 か月 (ID3) であり、絵本の読み手はいずれも母親で、読み聞かせは家庭において行われた。各被験者の映像データの秒数は、ID1 が 805 秒、ID2 が 276 秒、ID3 が 2463 秒であり、読み聞かせの回数は ID1 が 6 回、ID2 が 1 回、ID3 が 10 回であった。なお、ID3 の映像データには、読み聞かせ場面以外の映像 (玩具で遊ぶ場面や絵本を本棚から持ってくる場面など) が 1354 秒含まれていたため、人手で観察することによりこれらの場面を分析対象から除外し、読み聞かせ場面 1109 秒を抽出した。ID1、ID2 の映像データには、絵本読み聞かせ場面のみが収録されていたため、全ての場面の映像データを分析対象とした。本論文ではこれらの映像データから、1 秒間に 1 枚の静止画を抽出することで、画像データセットを作成した。これによって得られた静止画枚数は 2214 枚 (ID1: 817 枚、ID2: 278 枚、ID3: 1119 枚) であった。なお、静止画は 0 秒時点から 1 秒ごとに 1 枚抽出し、また抽出範囲の終端として映像ファイルの終了時を指定した場合、映像の終了時の静止画も抽出した。そのため、読み聞かせ 1 回ごとに 1 映像ファイルに分かれていた ID1、ID2 のデータにおいては、読み聞かせ場面の秒数に読み聞かせ回数の 2 倍を加えた静止画枚数が抽出されている。また、読み聞かせ場面が映像ファイルの途中で収録されていた ID3 のデータにおいては読み聞かせ場面の秒数に読み聞かせ回数を加えた静止画枚数が抽出されている。

3: <https://github.com/ultralytics/yolov5>

表3 指さし自動認識結果 (静止画枚数単位)

			人手による映像観察結果											
			ID1			ID2			ID3			3被験者全体		
			指さしあり	指さしなし	合計	指さしあり	指さしなし	合計	指さしあり	指さしなし	合計	指さしあり	指さしなし	合計
YOLOによる指さし検出結果	指さしあり	1箇所以上指さしを検出	23	-	23	75	-	75	60	-	60	158	-	158
		指さし以外を検出	5	121	126	8	72	80	9	108	117	22	301	323
	指さしなし				25	98	123	51	526	577	89	797	886	
	合計			41	295	335	108	170	278	120	634	754	269	1098
再現率 (指さしあり)[%]			68.3			76.9			57.5			66.9		
I_{od}/I_o			(23+5)/41			(75+8)/108			(60+9)/120			(158+22)/269		
適合率 (指さしあり)[%]			18.8			53.6			39.0			37.4		
I_{od}/I_d			(23+5)/(23+126)			(75+8)/(75+80)			(60+9)/(60+117)			(158+22)/(158+323)		
誤検出率 (指さしあり)[%]			81.2			46.5			61.0			62.6		
$1 - (I_{od}/I_d)$			1 - {(23+5)/(23+126)}			1 - {(75+8)/(75+80)}			1 - {(60+9)/(60+117)}			1 - {(158+22)/(158+323)}		
正答率 (指さしあり・なし)[%]			60.0			65.1			78.9			71.5		
$(I_{od} + I_{nond})/I_a$			{(23+5)+173}/335			{(75+8)+98}/278			{(60+9)+526}/754			{(158+22)+797}/1367		

表4 指さし場面自動認識結果 (秒数単位)

			人手による映像観察結果											
			ID1			ID2			ID3			3被験者全体		
			要観察場面	要観察場面以外	合計	要観察場面	要観察場面以外	合計	要観察場面	要観察場面以外	合計	要観察場面	要観察場面以外	合計
YOLOによる指さし検出結果	要観察場面		122	176	298	174	96	270	247	233	480	543	505	1048
	要観察場面以外		7	24	31	1	5	6	61	208	269	69	237	306
	合計		129	200	329	175	101	276	308	441	749	612	742	1354
再現率 (要観察場面)[%]			94.6			99.4			80.2			88.7		
S_{oe}/S_o			122/129			174/175			247/308			543/612		
適合率 (要観察場面)[%]			40.9			64.4			51.5			51.8		
S_{oe}/S_e			122/298			174/270			247/480			543/1048		
誤検出率 (要観察場面)[%]			59.1			35.6			48.5			48.2		
$1 - (S_{oe}/S_e)$			1 - (122/298)			1 - (175/270)			1 - (247/480)			1 - (544/1048)		
正答率 (要観察場面・それ以外の場面)[%]			44.4			64.9			60.7			57.6		
$(S_{oe} + S_{none})/S_a$			(122+24)/329			(174+5)/276			(247+208)/749			(543+237)/1354		

3.3 評価

3.2節で収集した映像データのうち、指さしが1回しか出現しなかったID1の読み聞かせ3回分と、指さしが2回以下しか出現しなかったID3の読み聞かせ4回分、ピントが合っていなかったID3の読み聞かせ1回分の静止画を除外し、読み聞かせ9回分(ID1: 3回, ID2: 1回, ID3: 5回), 1367枚(ID1: 335枚, ID2: 278枚, ID3: 754枚)の静止画を指さし検出対象データとして選定した。これらの静止画に対して、指さしの検出を行った結果を、表3に示す。この評価においては、指さしが含まれる静止画に対して、1つ以上指さしが検出されれば、実際には指さし以外のものが指さしとして検出されていても、その静止画を指さしが検出された静止画とした。これは、本論文における目的が、指さし場面として観察すべき時刻を推定することであり、他の物を指さしとして検出した結果であったとしても、指さしが出現した時刻を要観察時点として推定できていれば目的は果たされるためである。また、この評価においては、実際に指さしが観察される静止画枚数を I_o 、YOLOによって指さしが検出された静止画枚数を I_d 、実際に指さしが観察され、かつYOLOによって指さしが検出された静止画枚数を I_{od} 、実際に指さしが観察されず、YOLOによって指さしが検出されなかった静止画枚数を I_{nond} 、全静止画枚数を I_a とおき、次のように再現率 (指さしあり)、適合率 (指さしあり)、誤検出率 (指さし

あり)、正答率 (指さしあり・なし) を定義する。

$$\text{再現率 (指さしあり)} = \frac{I_{od}}{I_o}$$

$$\text{適合率 (指さしあり)} = \frac{I_{od}}{I_d}$$

$$\text{誤検出率 (指さしあり)} = 1 - \frac{I_{od}}{I_d}$$

$$\text{正答率 (指さしあり・なし)} = \frac{I_{od} + I_{nond}}{I_a}$$

この結果としては、ID1, ID2 においては再現率が70%以上となっており、3被験者全体でも再現率は67.2%となっていることから、指さしを漏れなく検出するという観点からは、一定以上の成果が出ていると言える。一方で、適合率は3被験者全体で40%未満、最も高いID2でも53.6%と低く、誤検出率が高い。この誤検出においては、ページをめくる際の手のような、指さし以外の形の指が指さしとして検出されている場合が多い。こうした誤検出が起きた理由としては、本論文においてファインチューニングに用いた訓練データ数が少なく、またそれらの訓練データは絵本読み聞かせに関するものではないことに加え、指さしが映っている画像のみを訓練データとして用いているため、誤検出が多い画像である、「絵本・指・手が含まれているが、指さしは含まれていない画像」が訓練データ中に存在しなかったために、指さしの手の形と絵本読み聞かせ場面で出現する他の手の形が十分に識別できていないということが考えられる。

こうした誤検出を減らす方策として、3.1節のファインチューニングにおいて、誤検出が多い画像である、「絵本・指・手が含まれているが、指さしは含まれていない画像」も併用したファインチューニングを行うことが挙げられる。この際の訓練データとしては、3.2節で収集した画像データ中の、指さしが含まれない画像を用いることが可能である。

また、これらの結果を踏まえて、指さしが検出された静止画の前後2秒間を要観察場面として推定した場合の検出率についても評価した。この評価においては、実際に指さしが観察された静止画の前後2秒間を要観察場面の正解(正解場面)とし、それらの正解と推定結果とを比較することによって評価を行った。正解場面の秒数を S_o 、推定された要観察場面の秒数を S_e 、要観察場面と推定された正解場面の秒数を S_{oe} 、正解場面ではなく、要観察場面と推定されなかった秒数を S_{none} 、全秒数を S_a とおき、次のように再現率(要観察場面)、適合率(要観察場面)、誤検出率(要観察場面)、正答率(要観察場面・それ以外の場面)を定義し、評価結果を表4に示す。

$$\text{再現率(要観察場面)} = \frac{S_{oe}}{S_o}$$

$$\text{適合率(要観察場面)} = \frac{S_{oe}}{S_e}$$

$$\text{誤検出率(要観察場面)} = 1 - \frac{S_{oe}}{S_e}$$

$$\text{正答率(要観察場面・それ以外の場面)} = \frac{S_{oe} + S_{none}}{S_a}$$

この結果としては、再現率はID1, ID2で90%以上、全体でも88.7%となっており、この推定によって観察から漏れる指さしは少ない。一方で、適合率は3被験者全体で50%程度と低く、指さしが含まれない場面500秒以上を要観察場面として推定している。しかし、この推定によって要観察場面とされた秒数の合計は実際の映像データの秒数の合計よりも約300秒少ない。対象データにおける1回あたりの絵本読み聞かせの所要時間の平均は150.8秒であることを鑑みると、この推定によって、読み聞かせ約2回分の時間を観察することなく、9回の絵本読み聞かせに出現する指さしのうち、90%近くを観察することができており、本論文における提案手法は、観察時間の短縮に一定以上の効果的があると言える。

3.4 被験者データを追加したファインチューニングと評価

本節では、訓練データとして3.2節で収集した画像データ中の静止画を追加し、3.1節と同様のファインチューニングを行い、そのモデルに対して3.3節と同様の評価を行った。ここでは、評価データから除外した8回分の読み聞かせ(ID1: 3回, ID3: 5回)の静止画から、正例として指さしが含まれる全ての静止画全18枚(ID1: 3枚, ID3: 15枚)と、負例として指さしが含まれない場面から無作為に抽出した静止画5枚(ID1: 2枚, ID3: 3枚)を訓練データに追加し、ファインチューニングを行った。このモデルに対して、3.3節と同様の評価を行った結果を、表5および表6に示す。3.3節における評価結果と比較すると、静止画枚数単位、秒数単位いずれの評価においても、全体として適合率がわずかに向上しているが、再現率は全体的に低下し

ている。ただし、秒数単位での評価においては、依然として全体の再現率は80%弱と高く、この時に要観察場面から除外された秒数は3.3節における評価結果よりも170秒以上増加しており、これは絵本読み聞かせ場面1回分の所要時間の平均を上回る秒数である。このことから、訓練データに被験者データを正例および負例データとして追加することは、適合率の向上および観察時間の短縮にある程度貢献できると考えられる。その一方で、本節では指さし検出精度の飛躍的な向上には至らなかった。この理由としては、訓練データ数が全体として少なく、特に絵本読み聞かせ場面における指さしのデータが少ないために、訓練データに被験者データを加えても、絵本場面における他の形の手と指さしとを十分に識別するには至らなかったということが考えられる。これに関する方策としては、さらに多くの被験者実験を行いデータを収集するなどして、訓練データ数を増やすことが挙げられる。

4 発達順序体系と絵本読み聞かせ映像観察による発達段階推定

4.1 映像手動認識後の場合

本節では、3.2節において画像データセットを収集した3人の子どもに対して、絵本読み聞かせの全場面に対する観察と、2.2節で作成した発達順序体系を用いて、次の手順で発達段階の推定を行う。

手順1. 絵本読み聞かせ場面の全時刻を人手で観察し、絵本関連項目(表1, 表2中で太字で示した項目)に該当する反応が出現するかどうかを判定する。

手順2. 手順1で出現した項目のうち、最も通過月齢が高い項目 T_1 の通過月齢を M_{T_1} とおく。

手順3. 通過月齢が M_{T_1} 以下の項目は全て通過しており、通過月齢が M_{T_1} よりも高い項目は全て未通過であると推定する。

3.2節の画像データセットの映像撮影依頼時においては、各子どもの母親に対して、発達順序体系の構築の際に用いた子どもの発達段階についての質問紙への回答を依頼した。本節では、それらの回答と本節における発達段階推定結果を比較することで、推定結果の評価を行った。なお、この質問紙調査においては、被験者の負担を考慮し、石川[2]における項目のうち、各被験者の月齢に適した項目と、本論文で独自に作成した「指さし」・「ページをめくる」に関する全項目のみ(ID1, ID2は合計70項目, ID3は合計102項目)を用いたため、評価においてもこれらの項目についてのみ実施した。ID1の T_1 は「I144: かなり長い文章のある物語絵本でも、内容を楽しむことができる」であり、 $M_{T_1} = 38$ であったため、ID1に対する質問紙中の70項目のうち、通過月齢が38か月以下である56項目を通過済と推定した。ID2については、 T_1 は「N29: 絵本にでてくる物を指さして「それは、何?」などと自分から質問することができる」であったため、 $M_{T_1} = 26$ で、ID2に対する質問紙中の70項目のうち、通過月齢が26か月以下である51項目を通過済と推定した。また、ID3の T_1 は、「I34: 子どもが自分で絵本のページをちゃんとめくることができ、とぼすと気づいてもとに

表5 被験者データを訓練データに追加しファインチューニングしたモデルにおける指さし自動認識結果 (静止画像枚数単位)

			人手による映像観察結果											
			ID1			ID2			ID3			3被験者全体		
			指さしあり	指さしなし	合計	指さしあり	指さしなし	合計	指さしあり	指さしなし	合計	指さしあり	指さしなし	合計
YOLOによる指さし検出結果	指さしあり	1箇所以上指さしを検出	19	-	19	74	-	74	46	-	46	139	-	139
		指さし以外を検出	9	102	111	9	63	72	3	71	74	21	236	257
	指さしなし		13	192	205	25	107	132	71	563	634	109	862	971
	合計		41	294	335	108	170	278	120	634	754	269	1098	1367
再現率 (指さしあり)[%]			68.3			76.9			40.8			59.5		
I_{od}/I_o			(19+9)/41			(74+9)/108			(46+3)/120			(139+21)/269		
適合率 (指さしあり)[%]			21.5			56.8			40.8			40.4		
I_{od}/I_d			(19+9)/(19+111)			(74+9)/(74+72)			(46+3)/(46+74)			(139+21)/(139+257)		
誤検出率 (指さしあり)[%]			81.2			46.5			61.0			62.6		
$1 - (I_{od}/I_d)$			1 - (19+9)/(19+111)			1 - (74+9)/(74+72)			1 - (46+3)/(46+74)			1 - (139+21)/(139+257)		
正答率 (指さしあり・なし)[%]			60.3			65.1			78.9			71.5		
$(I_{od} + I_{nond})/I_a$			((19+9)+192)/335			((74+9)+107)/278			((46+3)+563)/754			((139+21)+862)/1367		

表6 被験者データを訓練データに追加しファインチューニングしたモデルにおける指さし場面自動認識結果 (秒数単位)

			人手による映像観察結果											
			ID1			ID2			ID3			3被験者全体		
			要観察場面	要観察場面以外	合計	要観察場面	要観察場面以外	合計	要観察場面	要観察場面以外	合計	要観察場面	要観察場面以外	合計
YOLOによる指さし検出結果	要観察場面		114	136	250	174	83	257	181	185	366	469	404	873
	要観察場面以外		15	64	79	1	18	19	127	256	383	143	338	481
	合計		129	200	329	175	101	276	308	441	749	612	742	1354
再現率 (要観察場面)[%]			88.4			99.4			58.8			76.6		
S_{oe}/S_o			114/129			174/175			181/308			469/612		
適合率 (要観察場面)[%]			45.6			67.7			49.5			53.7		
S_{oe}/S_e			114/250			174/257			181/366			469/873		
誤検出率 (要観察場面)[%]			54.4			32.3			50.5			46.3		
$1 - (S_{oe}/S_e)$			1 - 114/250			1 - 174/257			1 - 181/366			1 - 469/873		
正答率 (要観察場面・それ以外の場面)[%]			54.1			69.6			58.3			59.6		
$(S_{oe} + S_{none})/S_a$			(114+64)/329			(174+18)/276			(181+256)/749			(469+338)/1354		

表7 映像手動認識後の通過項目推定結果

			質問紙回答結果								
			ID1(生後32か月)			ID2(生後26か月)			ID3(生後22か月)		
			通過	未通過	合計	通過	未通過	合計	通過	未通過	合計
映像からの推定結果	通過	観察	10	0	10	5	0	5	17	0	17
		推定	45	1	46	45	1	46	80	1	81
	未通過		0	14	14	12	7	19	1	3	4
	合計		55	15	70	62	8	70	98	4	102
	再現率 (通過項目)[%]		100.0			80.6			99.0		
	P_{qe}/P_q		(10+45)/55			(5+45)/62			(17+80)/98		
適合率 (通過項目)[%]		98.2			98.0			99.0			
P_{qe}/P_e		(10+45)/(10+46)			(5+45)/(5+46)			(17+80)/(17+81)			
正答率 (通過・未通過項目)[%]		98.6			81.4			98.0			
$(P_{qe} + U_{qe})/T_a$		((10+45)+14)/70			((5+45)+7)/70			((17+80)+3)/102			

戻ることができる」で、 $M_{T1} = 24$ だったため、ID3 に対する質問紙中の 102 項目のうち、通過月齢が 24 か月以下となる 98 項目を通過済と推定した。この評価においては、質問紙において通過と回答された項目数を P_q 、映像から通過と推定された項目数を P_e 、質問紙の回答においても映像からの推定においても通過となった項目数を P_{qe} 、質問紙の回答においても映像からの推定においても未通過となった項目数を U_{qe} 、全項目数を T_a

とおき、再現率 (通過項目)、適合率 (通過項目)、正答率 (通過・未通過項目) を次のように定義する。

$$(\text{再現率 (通過項目)}) = \frac{P_{qe}}{P_q}$$

$$(\text{適合率 (通過項目)}) = \frac{P_{qe}}{P_e}$$

$$(\text{正答率 (通過・未通過項目)}) = \frac{P_{qe} + U_{qe}}{T_a}$$

これらの評価結果を表7に示す。この評価において、質問紙回答においては通過となっている項目を、未通過であると推定してしまう誤りは、実際に被験者が通過している項目の通過月齢が M_{T1} よりも高い時、つまり被験者が実際に通過している最も通過月齢の高い項目を、映像では観察できなかった時に生じる。この誤りは、子どもが発達順序体系に従った発達をしている場合にも生じる誤りである。一方で、質問紙回答においては未通過となっている項目を通過していると推定してしまう誤りは、映像で観察された反応よりも通過月齢が低い項目を被験者が通過していない場合に生じる。この誤りは、被験者が発達順序体系に従った発達をしていない場合に起こる誤りであり、被験者の個性に起因するものであると考えられる。この評価結果において、いずれの被験者においても、適合率は 90% 以上であり、

再現率および正答率についても、ID1、ID3については90%以上、ID2でも80%以上と、十分高い精度で推定ができています。

この推定において、ID2の読み聞かせ回数は他の2人に比べて顕著に少なく、ID1が6回、ID3が10回に対してID2は1回であるため、ID2において実際に観察された項目の反応種類数は、ID1の半分、ID3の3分の1以下となっている。そのため、ID2の再現率・正答率は80%程度にとどまっており、他の2被験者と比較すると20%弱低い。しかし、それでも精度の低下はその程度にとどまっていることから、数分程度の読み聞かせ場面1回分に対する観察でも一定以上の精度で推定が可能であることが示された。

表8 映像自動認識後の通過項目推定結果

		質問紙回答結果								
		ID1(生後32か月)			ID2(生後26か月)			ID3(生後22か月)		
		通過	未通過	合計	通過	未通過	合計	通過	未通過	合計
映像からの推定結果	観察	5	0	5	3	0	3	4	0	4
	推定	46	0	46	47	1	48	87	1	88
	未通過	4	15	19	12	7	19	7	3	10
	合計	55	15	70	62	8	70	98	4	102
	再現率(通過項目)[%]	92.7			80.6			92.9		
	P_{qc}/P_g	(5+46)/55			(3+47)/62			(4+87)/98		
	適合率(通過項目)[%]	100.0			98.0			98.9		
	P_{qc}/P_e	(5+46)/(5+46)			(3+47)/(3+48)			(4+87)/(4+88)		
	正答率(通過・未通過項目)[%]	94.3			81.4			92.2		
	$(P_{qc}+U_{qc})/T_a$	((5+46)+15)/70			((3+47)+7)/70			((4+87)+3)/102		

4.2 映像自動認識後の場合

次に本節では、4.1節で発達段階の推定を行った3人の子どもに対して、3.3節において要観察場面として推定された場面に含まれる指さしのみに対する観察と、2.2節で作成した発達順序体系を用いて、次の手順で発達段階の推定を行う。

手順1. 要観察場面として推定された場面の映像を手手で観察する。

手順2. 各指さしが、絵本の指さし項目(表1、表2中で下線太字で示した項目)のどの項目に該当するかを判定する。

手順3. 手順2で子どもの反応として該当すると判定された項目のうち、最も通過月齢が高い項目 T_2 の通過月齢を M_{T_2} とおく。

手順4. 通過月齢が M_{T_2} 以下の項目は全て通過しており、通過月齢が M_{T_2} よりも高い項目は全て未通過であると推定する。

本節でも4.1節と同様に、各被験者の質問紙への回答と本節における発達段階推定結果を比較した。その結果を表8に示す。ID1の T_2 は「N29: 絵本にでてくる物を指さして「それは、何?」などと自分から質問することができる」であったため、 $M_{T_2} = 26$ となり、これは M_{T_1} よりも12か月低い。そのため、本節における推定結果では、通過済と推定された項目が4.1節の推定結果より5項目減少し、その結果再現率が7.3%、正答率が4.3%低下した。一方、ID2については、 T_2 は T_1 と同じ「N29: 絵本にでてくる物を指さして「それは、何?」などと自分から質問することができる」であった。そのため、 $M_{T_2} = M_{T_1}$

となり、本節における推定結果と、4.1節の推定結果は同一の結果となった。また、ID3の T_2 は、「N25: 絵本の絵の中の物を指さして、指さした物について説明する。(例: 泣いている顔を指さして「泣いている」と言う、犬を指さして「ワンワンいた」と言うなど。)」で、 $M_{T_2} = 23$ であった。これは M_{T_1} よりも1か月低く、このことにより、本節における推定結果では、通過済と推定された項目が4.1節の推定結果より6項目減少し、その結果再現率が6.1%、正答率が5.8%低下した。

本節における指さし場面のみに対する観察を用いた発達段階推定では、全場面の観察を用いた発達段階推定に比べて、再現率・正答率の低下がみられた。しかし、低下は数%にとどまっており、本節における推定結果においても、全ての被験者に対する再現率・正答率は依然として80%以上であることから、指さし場面のみを観察でも、十分な精度で発達段階の推定を行うことができていると考えられる。

5 おわりに

本論文では、絵本に関する子どもの反応を含んだ発達順序体系を構築し、その体系と絵本読み聞かせ場面を収録した映像に対する観察とを用いて、子どもの発達段階を推定する手法を提案した。映像に対する観察においては、指さし場面の自動検出を行うことによる要観察時点の推定を行い、これによって推定された時点の観察と、映像全体の観察それぞれを用いた発達段階の推定を行った。どちらの手法においても、高い精度で推定を行うことができ、提案手法が子どもの発達段階の推定に有用であることを示すことができた。

謝辞

本研究は科研費19K22858、19J20384の助成を受けたものである。

文献

- [1] P. W. Airasian and W. M. Bart. Ordering theory: A new and useful measurement model. *Educational Technology*, Vol. 13, No. 5, pp. 56–60, 1973.
- [2] 石川由美子, 前川久男. 絵本理解とその発達順序性: 発達援助としての絵本利用の基礎研究. *心身障害学*, Vol. 20, pp. 83–91, 1996.
- [3] 笠松美歩, 宇津呂武仁, 齋藤有, 石川由美子. 絵本に関する発達項目の発達順序性: 1996年・2020年の比較分析. 第35回人工知能学会全国大会論文集, 2021.
- [4] T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft COCO: Common objects in context. In *Proc. 13th ECCV*, pp. 740–755, 2014.
- [5] 三宅和夫, 大村政男, 山内茂, 高嶋正士, 橋本泰子. KIDS 乳幼児発達スケール. 発達科学研究教育センター, 1991.
- [6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Proc. 29th CVPR*, pp. 779–788, 2016.
- [7] Yuya Yoshikawa, Jiaqing Lin, and Akikazu Takeuchi. Stair actions: A video dataset of everyday home actions. *arXiv preprint arXiv:1804.04326*, 2018.