



日本語学習児の 初期語彙発達

基
般

小林 哲生 永田 昌明

NTT コミュニケーション科学基礎研究所

幼児の言語発達

ヒトの幼児は生後数年の間に言語をいとも簡単に身につけていく。最初はおっぱいを泣くことしかできないが、半年を過ぎると「バババ」「マママ」のような喃語を発し始め、1歳の誕生日までには母語に近い音声を発するようになる。またこの時期に、意味を伴う単語も発し始め、語彙の獲得が始まる。1歳後半になると、語彙を効率的に獲得できるようになり、複数の語を連結させた発話も開始する。2～3歳では、語順や項構造などの文法知識が格段に増え、言語による精緻なコミュニケーションができるようになる。こうした音声・語彙・文法などの各機能の発達過程は驚くほど複雑であり、これらの仕組みを解明するために、これまで発達心理学や心理言語学、認知科学などの分野で数多くの研究が行われてきた。

本稿では、言語発達の中でも1歳代の語彙発達に注目し、幼児がどのように語彙を学習していくかについて解説する。1歳頃の幼児には、語の理解 (comprehension) が発話 (production) に先行するという特徴的な性質がある。これは、ある語を学習してもその語を発話できるまでには時間的ラグが存在することを意味し、初期語彙発達を理解するにはこの両側面を研究することが重要となる。本稿では、私たちが行ってきた研究を題材にして、前半は語の発話データに関する研究を、後半は語の理解を探る実験研究を紹介する。

発話データ収集の問題点

1歳になると、幼児は特定の意味と音声結びついた初語 (first word) を発するようになり、語彙を少しずつ増やしていく。この時期は、「まんま」「パパ」などの一語からなる発話がほとんどで、何かを要求するための発話が多いのが特徴である。この時期で研究上問題となってきたのは、(1) 幼児がどんな語彙を初期に発話する傾向にあるのか、(2) どんな品詞 (名詞や動詞など) を覚えやすいのか、(3) 個人差や言語差はどの程度あるのかという問題である。これらを明らかにするには、幼児が実際にどんな語彙を発話するのかを調査すればよいのだが、データの性質によって研究結果の一般化が制約される場合があるので、注意が必要である。

従来、言語発達の分野で最も利用されてきたのは、語彙チェックリスト法 (幼児の発話できる語彙を親がチェックリストに記入する方法¹⁾) を用いて横断データ (cross-sectional data) を取得する方法である (図-1)。この横断データとは、月齢の異なる集団 (たとえば12, 18, 24カ月齢) から得られるデータのことである。短期間に多数のデータを取得できる点が長所である。しかし、同一幼児のデータが月齢間で揃っていないために個人差を検討できず、発達過程を再構成せざるを得ない点が問題であった。

こうした問題を回避する有効な方法は、縦断データ (longitudinal data) を取得することである。この縦断データとは、同一対象者を継続的に追跡して得られるデータのことである。従来は日誌法 (幼児が覚えた単語を随時日誌に記録する方法) や自然観察法

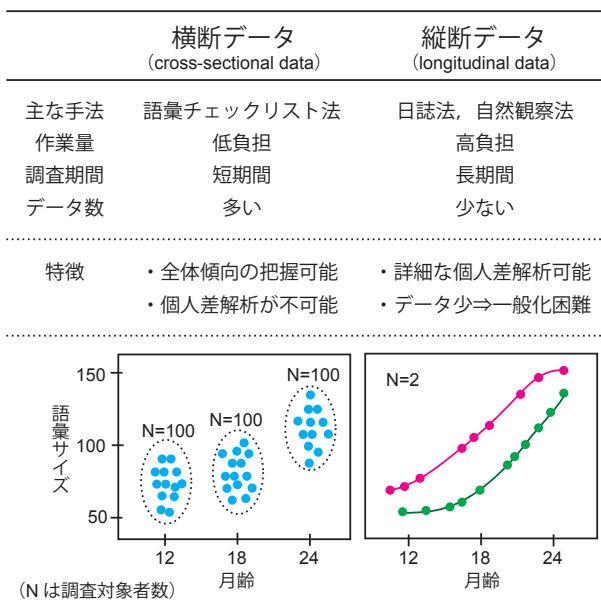


図-1 横断および縦断データを用いた研究の特徴

(幼児の発話場面を録画し解析する方法)によりデータ収集がなされてきた。この縦断データを用いた研究は、各個人の発達の変遷を明確にできる点で有効なアプローチであるが、データ取得に多大な時間と労力を要するため、これまでごくわずかな人数のデータ(たとえば、研究者自身の子どものデータ)しか取得できなかった。そのため、語彙発達に関する一般化や理論化が十分に進まず、多くの問題が未解決のまま放置されてきた。

なお、縦断データを用いた研究で最近話題になっているのは、わが子の生活を複数台のカメラで数年にわたって24時間記録し続けた、MITのDev Royによる研究である²⁾。この研究では、養育者からの言語入力と子どもの発話をすべて記録しており、それらの直接的な関係を解析できる点で期待が集まっている。ただし、たった1名のデータしかないため、そこから導かれる結果の一般化は難しい。

Webを利用した大規模縦断データ収集

従来研究から横断データによる全体的な傾向と、縦断データによる少数の幼児の発達過程は明らかになってきたが、初期語彙発達の詳細な検討や個人差の解析を行うには、多くの縦断データを集める必要



図-2 goo ベビー「赤ちゃん成長ダイアリー」の画面(本ツールの提供は2012年2月末で終了)

がある。こうした課題に対し私たちが行ったのは、Webを通じて縦断データを収集するという試みである。現在、日本のインターネット世帯利用率は9割を超え、研究に協力してくれる方が多数いれば、日本全国の家庭からWebを通じて大規模な縦断データを収集できると考え、Web上で日誌法を実施することにした(以下、「Web日誌法」)。

まず、育児中の親が多く集まる妊娠・育児情報サイト「gooベビー」(<http://baby.goo.ne.jp/>)の会員サービス上に、子どもがいつ、どんな単語を覚えたかを日齢単位で記録・管理できる日誌形式のWebツールを作成し、無料で公開した(図-2)。ユーザ会員には、研究目的や背景などを特に告知せず、わが子の成長記録として利用価値が高くなるような機能を付加することにより、データ投稿の促進を狙った。実際には、データの継続記録により語彙学習曲線や月齢別語彙獲得リストを自動表示するように

し、成長記録ツールとしての充実化を図った。また育児に忙しい親のことを考慮して、携帯電話からもデータを手軽に記録できるようにし、データ投稿の利便性を高める工夫も行った。

こうしたデータ収集の試みは2007年に開始し、日本全国の約800名のユーザから4年間で約20,000語のデータを投稿いただいた。この中には、初語開始時から2歳の誕生日前後までの語彙発達データが多数含まれており、大規模な縦断データを効率的に取得できた点で非常に有益な試みであった³⁾。

日本語の早期出現語の特徴

Web日誌法で得られた縦断データの重要な特徴は、ある語をいつ覚えたかというデータが多数の幼児から収集されることにある。この特徴により、各語が平均的に生後何カ月目に獲得されたか（心理学ではAoA [age of acquisition] と呼ぶ）を正確に特定できるので、日本全国の子どもが現在どんな語彙を初期に覚える傾向にあるかを容易に把握できる。そこで本データから各語のAoAを算出し、その値が小さい方から50番目までの語をリストアップした「早期出現語50語 (first 50 words)」を表-1に示した。最も早期に出現する傾向にあったのは、「まんま」「いないいないばあ」「はい」などの社会的な語（日課や挨拶）や、「ママ」「パパ」「ワンワン」などの人々や動物を表す語であった。また「アンパンマン」という語が早期に獲得されることも興味深く、日本の子どもが非常に幼い頃からアニメのキャラクターに注目していることが分かる。なお、本データはWebから取得したものであり、その信頼性が気になるが、このAoAの信頼性は予想以上に高く、横断データから算出した既存指標（50%獲得月齢）と高い一致率を示した。このことから、Web日誌法で取得したデータの信頼性がある程度裏付けられた⁴⁾。

また発達心理学の分野でよく使用される意味カテゴリ分類に従って、50語に占める各カテゴリの比率を算出すると、名詞が34%、社会的な語（日課

順位	発話語	AoA (月齢)	N (投稿数)
1	まんま	13.6	117
2	いないいないばあ	15.6	135
3	パパ	15.8	153
4	ママ	16.0	198
5	はい	16.1	117
6	ワンワン (犬)	16.3	186
7	ないない (片づける)	16.5	61
8	バイバイ	16.5	142
9	おっばい	16.8	82
10	よいしょ	16.8	52
11	ねんね (寝る)	17.2	103
12	くっく (靴)	17.7	59
13	ニャンニャン (猫)	17.8	107
14	アンパンマン	18.1	183
15	お父さん	18.2	72
16	お母さん	18.2	86
17	パン	18.3	65
18	たっち (立つ)	18.3	43
19	ある	18.4	80
20	どうぞ	18.4	120
21	バナナ	18.7	71
22	お茶	18.7	73
23	いや	18.8	119
24	葉っぱ	18.8	63
25	足	18.9	50
26	ぶーぶー (車)	19.0	78
27	だっこ	19.1	101
28	落ちる	19.2	53
29	痛い	19.2	123
30	じーじ/じいちゃん	19.3	100
31	おいしい	19.5	159
32	ばーば/ばあちゃん	19.6	112
33	耳	19.6	42
34	あっち	19.6	58
35	あつい	19.7	80
36	電車	19.7	78
37	ごちそうさま	19.8	56
38	目	19.9	49
39	ちょうだい	19.9	73
40	できる	20.0	53
41	チョウ	20.0	65
42	これ	20.0	68
43	ない	20.2	63
44	おいで	20.2	56
45	ニンジン	20.2	40
46	イチゴ	20.3	45
47	きれい	20.4	48
48	リンゴ	20.4	47
49	ジュース	20.4	55
50	おしっこ	20.5	67

表-1 Web日誌法から特定した早期出現語50語

・挨拶・会話語)が32%、述語(形容詞・動詞)が16%、人々に関する語が14%、機能語が4%となり、名詞と社会的な語が日本語の初期語彙に多く含まれることが分かった。

この早期出現語50語は、(データ取得法は異な



図-3 goo ベビー「子ども語辞書」のトップページ
URL は、<http://baby.goo.ne.jp/kodomogo/>



図-4 goo ベビー「子ども語辞書」の検索結果表示

るが)他の言語圏でも特定されており、それらと比較することにより言語間の違いを知ることができる。その際に使用される指標の1つが、早期出現語50語内の名詞と動詞の獲得数から算出するNV比(NV ratio = noun/noun+verb)である。先行研究では、英語学習児のNV比は0.92と非常に高く、初期の語彙のほとんどが名詞であったのに対し、韓国語学習児のNV比は0.64で、英語学習児よりも動詞を覚える傾向が強かった⁵⁾。同様に本データでもNV比を算出すると0.65になり、日本語の語彙カテゴリ構成が韓国語に非常に近く、英語と大きく異なることが分かった。

こうした言語間の差が生じる理由について、研究者の多くは、言語構造(項脱落により動詞が強調される言語か否か)や文化(幼児への発話が事物の名称を教えることに重きを置くか否か)の影響が強く反映していると考えており、母親の発話スタイルの特徴を言語間で比較する研究なども精力的に行われている⁶⁾。

データベースの公開

Web日誌法で取得したデータをもとに、私たちは各単語のAoAや発話の揺らぎなどをまとめたデータベースを作成した。これを使って、上述した初期語彙発達の基礎研究を進めるとともに、研究のアウトリーチとして一般公開も同時に行ってきた。その1つの試みが、上記のgooベビー上で公開中の「子ども語辞書」である(図-3)。

これは、0～3歳児の語彙発達過程を、意味・音声・月齢の点から検索・解析表示するWebツールのことで、主に育児中の親への情報提供を意図して作成した。たとえば、子ども語辞書上である単語(たとえば、わんわん)を入力し検索すると、その語がいつ(AoA:15.5カ月)、どんな意味(犬、動物、NHKのキャラクタ)で獲得される傾向にあるかが表示される(図-4)。また、わが子の成長にあわせて、ある月齢(たとえば18カ月)にどんな語彙(ブーブー[車]、ばーば[祖母])が獲得される傾向にあるのかを調べたり、関心の高い語(たとえば、パ

パヤママ) がいつ獲得されるかを調べたりするのも利用できる。私たちは、氾濫する情報の中でこうした科学的データに基づく情報提供こそが、子どもの発達に関する不安を解消し、楽しい育児につながるはずだと考えて、本データベースの公開を実施してきた。

なお、こども語辞書の「こども語」とは、私たちが考えた造語であり、幼児期に特有の幼児語／育児語(ないない [片づけ], おっちゃん [座る], ぶつぶー [車]) だけでなく、通常のことばや言い誤り(えべれーた [エレベータ], おかさな [お魚]) も含めた、こどもが発するすべての語を指している。また本データベースには、各語に対する発話の揺らぎ(電車 [でんちゃ, しゃ, えんしゃ]) も多数記録され、大人の言語から見れば「不自然な」言語表現も多い。幼児の言語に関するテキストデータを処理する必要性が生じた場合には、こうした揺らぎ特徴への対応が必須となるが、この問題に対しても本データベースは役立つと期待できる。

語の理解を探る実験アプローチ

発話データに基づく研究から、幼児がどんな語彙を初期に発話する傾向にあるかが明らかになってきた。しかし、語を学習しても発話するまでに時間的ラグが生じる(つまり、語の理解が発話に先行する)という幼児の特徴を考慮すると、発話データとは、語彙学習の結果(outcome)を示したものであり、幼児がどのような心的プロセスを経て語を学習しているかを知るには十分なデータとは言えない。この点を理解するには、語の理解の側面を詳細に検討していく必要がある。

語の理解を調べるには、幼児の行動反応を利用した実験アプローチを用いるのが有効である。心理学の分野では、言語報告を十分にできない幼い子どもの能力を探る際に、注視反応を指標とした実験法を主に用いてきた。これは、1970年代に乳幼児の知覚能力(視力や色覚、立体視等)の発達を調べる乳幼児心理物理学(infant psychophysics)の分野で

開発された方法で、現在は言語／認知処理能力を調べる際にも広く利用されている⁷⁾。代表的な方法には、馴化法(habituation paradigm)と選好注視法(preferential looking paradigm)がある。馴化法とは、幼児にある刺激を複数回提示して馴化(飽き)させた後に、新しい刺激を提示してそれに脱馴化(注視時間の上昇)が起こるかどうかを調べる方法である。脱馴化が起これば、2つの刺激の違いを幼児が区別していることが分かる。この方法はまた、ある刺激に馴化させるフェーズを必ず含むので、短時間に起こる学習プロセスを見る際にも有効な方法であり、語彙学習の分野でも頻繁に利用されている。一方、選好注視法とは2つの視覚刺激を同時に提示し、どちらを選好して注視するかを測定する方法である。言語能力を調べる場合には、ターゲット音声(「ワンワンはどっち?」)を提示し、2つの視覚刺激(犬 vs. 猫)のうちどちらを選好するかを測定すればよい。音声に該当する刺激(犬)をより長く注視するならば、その語を理解できていると考えられる。

こうした注視反応のほか、最近では、脳活動を直接測定する研究も増えている。特に、近赤外光を利用して頭皮の上から脳活動を非侵襲的に視覚化する「NIRS 脳計測装置」は、ある程度身動きのできる状態でデータ取得ができるため、長時間静止した状態を保つことが難しい幼児でも脳活動を容易に測定できる。こうした最新機器を用いて、心理学で示してきた行動反応を脳活動と結びつける試みが現在さかんに行われている⁷⁾。

語の指示対象をどのように特定するか

こうした実験アプローチを用いて私たちが現在取り組んでいるのは、幼児が新しい語を学習する際にその指示対象(referent)をどのように推定／特定するのかという問題である。新しい語を学習するには、(1)特定の音声刺激(語)を保持し、(2)語が指し示す対象(指示対象)を特定し、(3)語と指示対象を連合する(word-referent association)とい

った、3つの心的プロセスを経る必要があり、その中でも語の指示対象の特定は語彙学習の中心的な役割を果たしている。

たとえば、動物園で親が子どもに「シファカ！」と叫んだとしよう。もし子どもがその語を初めて聞いたとしたら、それが事物全体（動物名）を指しているのか、事物のパーツ（特徴的な耳）や材質（フサフサした毛）を指しているのか、それとも動作（特徴的な歩行形態）を指しているのかを特定するのは非常に難しい。語が指示する対象を特定する際に起こるこうした不確実性は、哲学者 Quine が「ギャバガイ問題」として提議した有名な問題であり⁸⁾、幼児がこの問題をどのように解決するかを巡って数多くの議論が沸き起こった。その中で多くの研究者が認める意見としては、幼児が知覚的情報（知覚的顕著性や時間的近接）や社会的情報（視線や文脈）、言語的情報（統語やプロソディ）といった複数の手がかりを総合的に利用して、語の指示対象を特定しているというものである。これは、私たち大人が普段利用する方略であり、理にかなった説明ではあるが、幼児が発達初期からそれらの手がかりをすべて利用しているとは思えず、その発達過程の解明が重要な研究課題となってきた。

そこで私たちが注目したのは、語の指示対象を特定する際に重要な手がかりとなる形態統語的情報 (morpho-syntactic cues) である。先ほどの例で「シファカがいるよ」と言われれば、私たち大人はその語に付随する形態統語的情報をたよりに、その語が事物を示す語（名詞）であることを容易に判断できる。また「シファカしてるよ」と言われれば、その語が動作を示す語だと推測できる。こうした手がかりを幼児がいつ利用できるようになるかを調べるために、私たちは1歳児を対象とした馴化法による実験を行った。馴化法を行うには、幼児の注意が散漫にならないように、**図-5**のような実験室に幼児をまず連れてくる必要がある。そこで、母親に子どもを抱っこしてもらって、モニタから2種類のムービー（ジャンプする動物とバウンスする車）を提示し、馴化させた。その際に、あるグループには名



図-5 馴化法を用いた実験の様子

詞文を用いた新奇語（セタ／モケがいるよ）を、もう一方のグループには動詞文を用いた新奇語（セタ／モケしてるよ）を聞かせた。この2つのムービーに馴化させた後、テストフェーズでは、新奇語を固定したまま、事物もしくは動作を入れ替えたムービーをそれぞれ提示して、彼らが新奇語を事物と動作のどちらに連合していたかを判定した。その結果、20カ月齢を対象にした実験では、名詞文を聞いた子どもが新奇語を事物に連合していたのに対し、動詞文を聞いた子どもは新奇語を動作に連合していることが分かった。これは、20カ月齢という幼い時期から形態統語的情報を利用して語の指示対象の特定を適切に行っていることを示している。一方で、それ以前の14カ月齢児では、名詞文と動詞文のいずれにおいても事物に新奇語を連合することが分かった。統制実験（新奇語を提示しない条件）で動作よりも事物への選好反応 (preference) があったことを考慮すると、14カ月齢の頃は、形態統語的情報を利用せずに、知覚的選好に左右される語彙学習を行っていることが推測され、語の指示対象の特定はまだ正確に行えない時期と言えるだろう⁹⁾。

なお、20カ月齢での成功は、従来想定されてきた時期よりも相当早い段階から形態統語的情報の抽象的知識を持っていることを示しており、既存の語彙学習理論に大きな影響を与える可能性がある。また、この20カ月齢という時期は、語彙学習速度が急激に上昇する語彙爆発 (word explosion, または vocabulary spurt) の開始時期¹⁰⁾と重なることから、形態統語的情報の利用可能性が実際の語彙学習速度

の上昇に寄与している可能性も考えられる。

このほかに私たちが取り組んでいるのは、(1) 幼児が述語項構造から動詞の意味を推測できるかどうか¹¹⁾、(2) ワンワンやブーブーなどの音韻反復の特徴を持つ幼児語が実際に学習しやすいかどうか¹²⁾、(3) 格助詞や語順などの知識がいつごろ獲得されるかという問題である。また、日本語だけでなく、英語や仏語などの言語構造が大きく異なる言語で同様の実験を実施することにより、語彙学習能力の共通性を探る試みも進行中である。こうした実験アプローチによる試みから、語彙学習能力を可能な限り明らかにし、初期語彙発達の全貌解明につなげたいと考えている。

最後に

本稿では、日本語学習児の初期語彙発達に関する私たちの研究を紹介しながら、幼児がどのように語彙を学習するのかについて解説した。紙面の都合上、触れられなかったテーマもあるが、その中でも今後重要になってくると思われるのは、語彙発達の個人差に関する問題である。世の中には、電車の名前ばかり覚える子どもや、社会的やりとりに関することばに長けている子ども、語彙の理解はできるがなかなかしゃべろうとしない子どもなど、いろいろなタイプの子どもが存在する。しかし、そのような言語発達の個人差がどのような要因で生じるかは詳細に分かっていない。その際に重要になってくるのは、やはり縦断データの蓄積である。語彙発達データに加えて、さまざまな発達指標や個人特性、生活資料などのデータを継続的に収集し、大規模なデータセットが揃えば、データマイニングを駆使することにより、個人差を生み出す要因が明らかになるだろう。その上で要因を統制した実験も併用すれば、初期語彙発達をより深く理解することができる。

参考文献

- 1) 代表的な語彙チェックリストに、マッカーサ乳幼児言語発達質問紙 (MacArthur Communicative Development Inventory) があり、日本語を含む複数の言語に翻訳され、研究現場で利用されている。日本語版は小椋・綿巻の監修で京都国際社会福祉センターから2004年に出版。
- 2) Roy, D.: The Birth of a Word, TED talks (2011), http://www.ted.com/talks/deb_roy_the_birth_of_a_word.html
- 3) 小林哲生, 永田昌明: ウェブを用いた幼児言語発達研究: 大規模縦断データの収集, 言語処理学会第15回年次大会発表論文集, pp.534-537 (2009).
- 4) 小林哲生, 永田昌明: ウェブ上で収集した幼児語彙発達データの信頼性検証, 言語処理学会第16回年次大会発表論文集, pp.403-406 (2010).
- 5) Kim, M., McGregor, K. K. and Thompson, C. K.: Early Lexical Development in English- and Korean-speaking Children: Language-general and Language-specific Patterns, *Journal of Child Language*, 27, pp.225-254 (2000).
- 6) Fernald, A. and Morikawa, H.: Common Themes and Cultural Variation in Japanese and American Mothers' speech to Infant, *Child Development*, 64, pp.637-656 (1993).
- 7) 山口真美, 金沢 創: 心理学研究法4: 発達, 誠信書房 (2011).
- 8) W・V・O・クワイン: ことばと対象, 勁草書房 (1984). (原題: Word and Object, MIT Press (1960))
- 9) Oshima-Takane, Y. and Kobayashi, T.: Early Word Learning in Young Japanese Children, *Studies in Language Sciences* (in press).
- 10) Minami, Y., Sugiyama, H. and Kobayashi, T.: Multiple Vocabulary Spurts in Japanese Children, Poster Presented at 12th International Congress for the Study of Child Language (IASCL), Montreal, Canada (2011).
- 11) Kobayashi, T. and Suzuki, T.: Japanese 28-month-olds' Inference about Verb Meaning from Syntactic Frames, Poster Presented at 12th International Congress for the Study of Child Language (IASCL), Montreal, Canada (2011).
- 12) Kobayashi, T. and Murase, T.: Learning Multiple Labels for a Single Object in Japanese Children, Poster Presented at the 36th Boston University Conference on Language Development (BUCLD36), Boston, USA (2011).

(2011年11月26日受付)

小林 哲生 ■ kobayashi.tessei@lab.ntt.co.jp

NTTコミュニケーション科学基礎研究所 協創情報研究部 言語知能研究グループ, 研究主任, 東京大学大学院総合文化研究科博士課程修了。博士(学術)。専門は発達心理学, 実験心理学。著書に「0-3さいはじめてのことば」(小学館)「モバイル社会の現状と行方」(NTT出版), 訳書に「数覚とは何か」(早川書房), 「動物のこころを探る」(新曜社)。

永田 昌明 (正会員) ■ nagata.masaaki@lab.ntt.co.jp

NTTコミュニケーション科学基礎研究所 協創情報研究部 言語知能研究グループ, 主幹研究員, グループリーダー, 京都大学大学院工学研究科修士課程修了。博士(工学)。専門は統計的機械翻訳, 日本語自然文検索など。著書に「質問応答システム」(コロナ社), 「言語と心理の統計」(岩波書店)など。