

第一言語獲得研究における自然発話分析の役割について¹⁾

On the roles of naturalistic studies in child language research

野 村 潤

1. はじめに

第一言語獲得（L1獲得）の研究は、自然発話研究から始まったと言える。1980年代からは、コンピューターの普及を背景とした子どもの発話分析システムの開発により、発話記録の保存・共有・分析が飛躍的に容易になった。しかし同時に、理解能力や、自然発話では滅多に出現しない構文の産出能力を調べるため、幼児を対象とした実験研究も盛んに行われるようになった。この流れの中では、自然発話研究の限界が強く意識され、その役割が過小評価される状況も生じる可能性がある。

本稿は、自然発話研究の役割を改めて確認することを目的とし、以下の考察を行う。まず第2節では、自然発話研究の歴史、特にデジタルデータが普及し始めてからの流れを概観する。第3節では、自然発話データをいくつかの観点から分類する。第4節では、主に自然発話データのみを用いた研究例を紹介し、自然発話データがどのように研究成果を生み出してきたのかを考察する。第5節では、それまでの考察に基づいて自然発話研究の強みと限界を考える。第6節では、限界に対応しながら自然発話分析のみで意味のある結果を生み出すために注意すべき点をまとめる。

2. 自然発話記録を用いたL1研究の歴史

2.1 記録方法の変遷

L1研究は、自然発話の研究から始まったと言える（村田，1981）。初期の研究では、幼児の発話をその場で（あるいは記憶を頼りに）文字に記録していく、いわゆる日記データが用いられていた。一般に日記データでは、全ての発話を記録するのは不可能であり、主に記録者の注意を引いた発話を書き留められることとなる。また、調音や文法などの誤用が無意識に修正される可能性があることにも注意しなければならない。

その後、カセットテープやボイスレコーダーなどの録音機器の普及につれて、データの量と質が飛躍的に向上した。録音により、聴き直しが可能となる。このため、対象児はもちろん会話参加者全員の発話をほぼ全て書き留めることができ、誤用を認識することも容易になる。つまり、日記データにまつわる問題点をほぼ解消できる。むろん、子どもが動き回り、発話が重なり合い、様々な雑音が入る状況で一定の音質を保って録音し続けるにはある程度のノウハウが必要となるが、録音機器の性能の向上や、マイクを内蔵した服を子どもに着用させるなどの工夫により、明瞭な録音が可能となってきている。録音は、現在でも主要な記録方法の一つである。

録音よりもさらに多くの情報を記録できるのが録画である。録画に関しても、技術的ハードルは年々下がってきている。従来、録画に際して特に困難なのは、動き回る子どもを、確実に画面内に捉え続けることであった。しかし、近年では複数のビデオカメラを使用して部屋全体を映し、後に複数の映像を一つの画面に合成することが可能である。さらに、2020年以降の新型コロナウイルスのまん延により急速に普及したオンライン会議システムを用いれば、複数の映像と1つの音声を、ほぼリアルタイムで単一の動画ファイルとして保存することも技術的には可能である。このように、近年、録画の技術的ハードルは大変低くなっていると言える。ただし、複数のカメラとマイクを対象児の自

宅に設置するのは煩雑である。また、オンライン会議システムを使用する場合はインターネット環境も必要である。このため、複数のカメラを用いる場合は、研究機関の一室にあらかじめ機器を設置し、参加者に赴いてもらうほうが現実的であろう。

さらに、動画撮影の技術・運用面以外の大きな課題として、プライバシーが挙げられる。当然のことであるが、自宅での録画をためらう養育者は多い。また、自宅外で録画を行った場合でも、映像を学会で再生したり、他の研究者と共有したり、インターネット上に公開することに対する抵抗感は強いと思われる。プライバシーに関する懸念をクリアすることの難しさと、研究機関などで適切な録画用の部屋を用意することの難しさが相まって、現在でも記録形態として録画よりも録音が選択されることが多いのだと思われる。

2.2 CHILDESの登場

自然発話研究の進歩を後押ししてきたのは、録音・録画の技術だけではない。データのデジタル化に必要なコンピューターの普及を背景に、テキスト化した発話記録（トランスクリプト）の規格化や、分析・共有システムの開発も進められた。自然発話を用いたL1獲得研究の分野で現在、実質的にスタンダードとなっているデータ作成・分析・共有システムは、Child Language Data Exchange System（以下CHILDES、読み方は/'tʃaɪldz/または/'tʃaɪldəs/）である（MacWhinney, 2000）。CHILDESのプロジェクトは、トランスクリプトの統一フォーマットを定め、作成・分析を支援し、また発話記録を研究者間で共有することを目的として、1984年にBrian MacWhinneyらによって開始された。その後、同システムが子ども以外の発話の記録（学習者、バイリンガル、言語障害などの発話記録）にも使われるようになり、現在ではTalkBankと呼ばれるより大きなプロジェクトに発展している。

CHILDESの主なコンポーネントは2つである。1つは発話記録を作成および分析するためのソフトウェアCLAN（Computerized Language Analysis）である。CLANはWindows、Mac、Unixの各プラットフォームに対応してい

る。また、CHILDESウェブサイト上でも利用可能で²⁾、公開された発話記録の分析をブラウザで行うことができる。CHILDESのもう一つのコンポーネントは、CLANで分析するためのトランスクリプトのフォーマットのガイドライン、CHAT (Codes for the Human Analysis of Transcripts) である。以上のシステムおよび様々な研究者が公開した発話記録が、インターネット上で利用できるようになったことにより、自然発話研究を実施するための方法論的なハードルは大変低くなった。

CHILDESの開発は英語圏で始まったが、それを日本語に対応させるために様々な研究者が尽力した (Oshima-Takane & MacWhinney, 1998; 宮田・森川・村木, 2004など)。日本語対応に際しては主に2つのことが行われた。1つは、日本語をどのように単語に分かち書するかを定めた「分かち書きガイドライン」(最新版はWAKACHI2002 v.8.0、宮田・中, 1998; 宮田, 2018)の策定である。もう1つは、日本語用の形態素解析プログラムの開発である(最新版はJMOR 08、中・宮田, 1999; 宮田, 2018)。

以上のように、様々な研究者の貢献により、現在、CHILDESウェブサイトには英語や日本語を含む40以上の言語のデータ、CHATプログラム、様々な言語の形態素解析プログラム、各種マニュアルが公開されている。決められた文献の引用義務など、ルールを遵守することを条件に、それらは無償で利用することができる。データと分析ソフトをダウンロードしてローカルで利用することもできるが、上述のように、公開データの基本的な分析は、ブラウザ上でも可能である。

日本の研究者にとって1つの障壁となりうるのはCHILDESウェブサイト提供される情報が全て英語によるものであるという点かもしれない。しかしこれについても、総合的な日本語マニュアルとして機能する「CHILDES日本語版」³⁾のウェブサイト(宮田, 2012)の公開により、大部分が解決されている。

3. 発話記録の分類

「自然発話の記録」と一口に言っても、記録方法や参加者の属性により、その性質は様々である。以下では、発話記録をいくつかの側面から分類する。

3.1 縦断データと横断データ

縦断データとは、少数の対象児の発話を、長期にわたって記録したデータである。CHILDESウェブサイトで公開されている日本語の縦断データでは、1～数名を、半年から数年にわたって記録したものが多い。横断データは、多くの対象児を1回ないし数回記録したものである。CHILDESウェブサイトの日本語データでは、130名の記録であるOkayamaデータ（岡山, 1970-1974; Shirai, et al. 2001）が現在唯一の自然発話の横断データである。

縦断データは、発達過程の詳細な分析に向いているが、発達には個人差があることを常に意識し、過度な一般化が生じないように注意する必要がある。一方で横断データは、多くの子どもに見られる一般的な傾向を統計的に見出しやすいが、一人の子どものデータが限定的である。このため、ある現象について、一部の子どもは頻繁に発するが、他の子どもは全く発しない、と言ったことが起こりうる。発達過程を詳細に調査することも難しいことが多い。例えば産出語彙の発達を調べたい場合は一人の子どもの累積的なデータが必要になるため、横断データでは調べることは難しい。平均発話長の発達過程、子どもの発達に応じたインプットの変化などをつぶさに分析したい場合にも限界があるろう。

3.2 記録方法・形態

第2.1節で述べたように、発話データは、その場で発話をメモしてまとめた日記データと、録音・録画を文字に起こしたものに大別できる。日記データの場合は、前述のように、(1)対象児の発話を全て書き取れていない、(2)誤りが無

意識に修正されている、(3)発話速度の速い成人の発話が必要最小限しか書き取られていない、といった限界を認識する必要がある。録音または録画を書き起こした場合はそのような問題点はほぼ解消されるが、録音や調音が不明瞭な場合に、無意識に文脈から推測して書き起こされた可能性は残る。例えば、子どもがコワッタと発した場合に、文脈に応じてコワシタ、もしくはコワレタと記録されることがありうる。このような無意識の推測や修正は、例えば自動詞と他動詞の獲得を調査したい場合には大きな問題となるであろう。むろん、比較的多量のデータを量的に分析する場合は、そのような修正がそれほど重要にならないこともある。記録の不正確さをどの程度まで許容するかはケースバイケースで判断することとなるが、分析結果に影響を与えた可能性が否定できない場合にはその旨を明記しておく必要がある。

記録（収録）の長さや頻度にも注意する必要がある。CHILDESのサイトに公開されている、録画や録音に基づくデータは、おおよそ1週間から1ヶ月に1度、毎回40～60分程度記録されたものが多いが、時期によって頻度や1回あたりの収録時間にばらつきが見られる場合もある。

3.3 収録状況

自然発話データに分類されるものでも、その収録状況は様々である。収録場所については、現在公開されているCHILDESの日本語データは、ほとんどが対象児の自宅で収録されたものであるが、他の言語のデータでは研究機関で収録されたものも存在する。

対象児以外に、誰が会話に参加しているかも注意する必要がある。公開データでは、主たる養育者としての母親が対象児の相手をしていることが多いが、例えばIshiiデータ（Ishii, 1999）では、研究者である父親との会話がほとんどであり、主な養育者と思われる母親はほとんど参加していない。収録時のインプットが主たる養育者のものではないことに注意する必要がある。そのほか会話の参加者としては、研究者、兄弟姉妹、祖父母なども考えられる。

3.4 対象児の属性

定型発達児の場合、分析結果に影響を与える対象児の属性として、年齢、性別、方言が挙げられる。年齢については、CHILDESの公開日本語データの中でもよく使われる、3名の縦断データであるMiyataデータ（Miyata, 2004a; 2004b; 2004c）は、名古屋近郊に住む男児を1;5（1歳5か月）ごろから3;0ごろまで収録したデータとなっている。MiiProデータ（Miyata & Nishisawa, 2009; 2010; Nisisawa & Miyata, 2009; 2020）では、4名の子ども（男女2名ずつ）を、3;0前後から5;0前後まで収録している（1名のみ1;1から）。横断データであるOkayamaデータ（岡山, 1970-1974; Shirai, et al., 2001）においては、おおよそ2歳代から5歳代の子どもを1回ずつ記録している。

性別については、CHILDESの日本語公開データ、特に縦断データについては、ほとんどが男児となっているため、性差が結果に影響を及ぼす可能性がある場合には注意が必要である。

方言について、CHILDESの日本語公開データでは、Miyataデータが名古屋近郊、MiiProデータは関東、IshiiデータとOkayamaデータは関西地方の子どもを対象としている。方言の違いは、研究の目的によっては留意しておく必要がある。例えば木部（2019）は、成人による格助詞の省略率が方言によって異なることを指摘している。

3.5 音声・動画の有無

音声や動画の有無は、研究目的によっては、分析の精度に大きな影響を与える（第6.3節参照）。MiyataのTaiデータ、MiiProデータの大部分、Otaデータ（Ota, 2003）では音声、Ishiiデータではビデオが利用可能となっている。これらのCHILDESデータにおいては、音声・映像とテキストは同期されており、テキストデータと音声・映像を容易に行き来して、実際の発音や発話時の状況を確認することができる。

4. L1獲得研究における自然発話データの用い方

第2節で述べたように、自然発話データの作成・共有・分析の環境はここ数十年の間に飛躍的に向上した。それらの自然発話データの用い方は、大きく2つに分かれる。1つは、実験研究と組み合わせて用いる場合、もう1つは、自然発話データだけを用いる場合である。以下、4.1では実験研究と組み合わせて用いた研究例を、4.2以降では自然発話分析のみを用いた研究例を紹介していく。

4.1 実験研究と組み合わせて用いる

実験研究においては、自然発話データの分析は、実験材料の選択や制御に使われることが多い。例えば日本語児の動詞の屈折について調べた Tatsumi, Ambridge, & Pine (2018) では、過去・非過去についての誘導産出課題に使う動詞を選定するため、インプットにおける動詞の形態が過去・非過去のどちらに偏っているかを CHILDES で調べた。自然発話データが実験研究との組み合わせで用いられるもう1つのパターンは、インプットの調査、特に、子どもが持っている知識・能力が、インプットから獲得可能かを検討する場合である。例えば Suzuki & Yoshinaga (2013) は、遊離数量詞について、統語構造に依存した解釈を子どもが行うことを実験で示したが、そのような解釈を可能にする知識を獲得するための情報がインプットには不足していることを CHILDES データで示した。

4.2 任意の時点における知識・能力を推測する

L1獲得研究の一つの大きな目的は、任意の時点（年齢もしくは発達段階）において、子どもがどのような知識や能力を持っているのかを推測することである。それらの知識や能力を、実験をデザインして調べる場合も多いが、目的によっては自然発話データの分析を通して調べることも可能であるし、その方

が適切である場合もある。

例えば Sugisaki (2005) は、日本語児が発する VO 語順を持つ発話に着目した。動詞の後に文の構成素が生じる発話（後置発話）は非規範的で、疑問詞は後置要素になれないなどの制約に従う。もし、日本語児において語順に関するパラメータがすでに設定されていれば、子どもの発話は VO に関するさまざまな制約に従うはずであるとの仮定のもと、Miyata データのうち、Aki (Miyata, 2004a) および Ryo (Miyata, 2004b) のデータの 2 歳後半の部分から VO 発話を抽出して分析した。その結果、WH 要素が後置されているものは存在しなかった。Sugisaki は、2 歳後半の日本語児は語順に関するパラメータを設定済みであると結論づけた。

さらに Sugisaki (2011) では、子どもが格助詞と後置詞を区別しているかについて調査した。ハおよびモは、後置詞の後に続くことはできるが（祖母の家にも行った）、格助詞の後につづくとは非文となる（*祖母の家がも同じ町にある）。Sugisaki が Miyata データのうち、Aki (Miyata, 2004a) および Tai (Miyata, 2004c) の 2 歳代のデータを分析したところ、後置詞だけの発話 1464 例に対し後置詞 + ハ・モのパターンは 30 例存在したが、格助詞 1017 例に対し格助詞 + ハ・モのパターンは 1 例のみしか観察されなかった。このことから、対象児は格助詞と後置詞を明らかに区別していると推測された。

Tatsumi & Pine (2016) は、日本語児が初期に発する動詞の屈折語尾が、過去時制を一種のデフォルトとして位置付ける生成文法的な予測と、インプットの役割を重視する構成主義的な予測のどちらに近いパターンを示すかを調査した。Miyata データ 3 名分および Ishii データから、1;5 ~ 2;10 の期間の子どもの発話及びインプットを分析した結果、後者の予測を支持する結果が得られた。

以上の研究例では、2 つの理論（片方は一般的な論理や確率論でもよい）に基づいて、特定の期間における子どもの発話の性質に関する予測を立て、分析結果が一方のみを支持することを示している。このようなデザインは実験研究と共通であるが、1 歳 ~ 2 歳の、実験への参加が困難である子どもを対象とす

る場合には、自然発話データの分析が有効であるといえる。また、Tatsumi & Pine (2016) のように、インプットの影響を調査したい場合にも自然発話分析が不可欠である。ただし、自然発話分析においては、特定の構造の産出がいわば偶然に依存するため、その点を踏まえて研究を計画する必要がある（第6.2節参照）。

4.3 発達的变化を調べる

多くのL1獲得研究では、特定の時期に子どもがある知識や能力を持っているかどうかだけではなく、その知識や能力がいつ頃から、どのような過程を経て獲得されるのか、つまり発達的变化を調査し、その背景にあるメカニズムを探ることを目的とする。このような目的を持って自然発話データを分析する場合、特定の現象の出現頻度の絶対的・相対的变化や、出現順序、出現時期、出現の仕方などを調べることになる。

Matsuoka (1998) は、日本語児が、格助詞の獲得において、インプットとは関係なく同様のパターンを示すのかを調査した。3名の対象児（1歳台～3歳ごろのデータ）についてガ、ヲ、ニを調査した結果、獲得順序については、3名ともガ→オ、ニとなり、インプットの影響ではないことが推測された。また、格助詞の出現前に時制の形態素が出現しており、これは名詞句の格素性が機能的な主要部によりライセンスされるという仮説を支持するものであった。日本語では、2つの項が状態を表す述部とともに現れる時には、ガが主語と目的語の両方に使われうるが、対象児は全員、主格のガ1つのみを使用した。これは、複数のチェックがパラメータによるオプションであるという仮説に合致する。さらに、格助詞の誤用はガの過使用が圧倒的であったが、これは、日本語においては主格のガがデフォルトの格であると日本語児が仮定していることを示唆する。

Uno (2016) は、ハとガの談話語用論的な機能がどのように出現し発達するのかを、用法基盤アプローチに基づいて調査した。Miyataによる3名分のデータおよびNojiデータ（野地, 1973-77; Noji, Naka, & Miyata, 2004）を分析し

た。その結果、ハについては、最初はおもにWH疑問文の標識として用いられるが、次第に旧情報の命題を伝えるために使用される、という発達過程が全ての対象児について観察された。その一方でガは個人差が観察された。対象児のうち1名は、自分の体験を報告するために動作動詞とともに用いたのに対し、ほかの3人は、現在形の特定の状態述語とともに、指示物に対する主観的感情を表現するために用いた。Unoは、これらの結果の説明には、インプットの頻度が有効であると結論づけた。

Kido (2017) は、オシ-タオス、タベ-ハジメルなどの複合動詞の獲得について調査した。動詞が2つ連なる複合動詞には、語彙的複合動詞 (Lexical Compound Verbs, LCV) と統語的複合動詞 (Syntactic Compound Verbs, SCV) の2種類があると言われる。例えばオシ-タオスはLCVで、タベ-ハジメルはSCVである。両者は、2つ目の動詞をソウスルに置き換えることができるかどうかで判別可能である (*ソウ-シ-タオス vs. ソウ-シ-ハジメル)。日本語児によるこれら2種類の複合動詞の処理が異なるかどうかを調べることを目的に、KidoはMiyataのAkiデータとTaiデータ、およびNojiデータを分析した。その結果、(1)LCVの後にSCVが出現する、(2)SCV出現前には、それを構成する2つの動詞が単独で産出されるが、LCVに関してはそのようなことはない、などの結果が得られた。Kidoはこれらの結果をもとに、日本語児は2つの動詞を複合してSCVを産出しているのに対し、LCVについては単独の動詞として扱っていると結論した。

以上の研究例では、特定の構造や語彙の出現順序や発達過程が、パラメータの設定などの理論的背景やインプット、対象児内での知識の様態や処理過程といった要因による予測と合致するかどうかを検討することにより、おのおのの要因の妥当性について結論を導こうとしている。出現順序や出現時期を問題とする分析では、論理的には、統計的検定が可能な観察数を必ずしも必要としない。必要なのは、対象となる構造や語彙が最初に出現したのがいつかを特定することだけである。ただし、当該の知識・能力が獲得されているにもかかわらずたまたま発する機会がなかった可能性や、出現したからといって獲得されて

いるとは限らない点にも注意する必要がある（第5.2節参照）、この点を克服するための質的または量的な分析が必要になる場合も多い。

むろん、発達の変化を量的に検討する研究も多く存在する。例えば Nomura & Shirai (1997) は、伊藤 (1990) の、自動詞と他動詞の獲得に関するモデルを量的に検証するため、Noji データ (1:4-2:4) を分析した。伊藤のモデルでは、自動詞と他動詞は次の4段階を経て獲得されると主張する。

段階1. 他動の文脈で自動詞を使用

段階2. 使役を表す *-sase* の使用

段階3. *-sase* から *-sa* を削除

段階4. 正しい他動詞の獲得

このモデルの限界としては、量的なデータに基づいて提唱されたものではないことや、アク-アケルのように、自他が形態的にペアとなった動詞のみを対象としていることが挙げられた。Nomura & Shirai による分析の結果、初期においては確かに自動詞のみを用いる段階があり、その点では伊藤のモデルが支持されたが、他動の文脈で自動詞を用いる「過剰拡大」のパターンは必ずしもそうではなかった。具体的には、他動詞-自動詞のペアは大部分が正確に用いられていただけでなく、他動詞の正用が自動詞の過剰拡大の前に出現し、さらには自動の文脈で他動詞を用いる過剰拡大も観察された。これらの結果を、Nomura & Shirai は、インプットの頻度や、自動詞-他動詞ペアの形態的な類似などから説明した。

加山 & 大嶋 (2021) では、名詞句（および格助詞）の省略が頻繁な日本語において、動詞の自他の区別がどう獲得されるかを調査した。3組の母子の会話を1歳前後から3歳前後の5つの時点で録画し、母親と子どもが顕在的な項とともに発話した異なる動詞の累計数を算出した。その結果、(1)子どもの年齢が上がるほど、母親の顕在的な項が増えること、(2)顕在的な項を持つ4種類の文フレームで異なるタイプの自動詞・他動詞を多く発話した母親の子どもは、やはりそれぞれの文フレームで多くの自動詞・他動詞を産出することがわかった。つまり、多様な文フレームが含まれるインプットを与える親の方が項構造

の獲得を推進する、という考え方に合致する結果であった。

一般に、生成文法などの理論に基づいた研究は、実験研究においても自然発話研究においても、理論の予測を、1つまたはごく少数の差異にまで絞り込み、その差異が実際に観察されるかどうかを検証するというデザインを取ることが多い。一方で自然発話における発達的变化を量的に調査する研究では、予測を検証するだけでなく、分析結果の中に「ボトムアップ」的にパターンを見出し、その結果に影響すると思われる複数の要因を検討し、発達のメカニズムを推測する、という側面を持つことも多いようである。ただし、対象となる現象の分類方法や発達過程を先行研究に基づいて緻密に確立・推測し、また検証可能な仮説や明確な研究課題を設定することが必要なことは言うまでもない(第6.1節参照)。

4.4 言語差の有無を調べる

異なる言語を話す子どもの比較は、言語獲得の背景にある知識や認知、学習の仕組みについて、単一の言語を対象として分析するよりも豊かな証拠を提供してくれる可能性がある。例えばGuerrero et al. (2006)は、日本語児と英語児の複数の発達段階における母親との会話データを用い、指示語の形態(省略、代名詞、語彙的)と言語外の要因(文脈における名詞句が担う情報の新旧、指差し、視線など)の関係を調査した。2つの分析が行われた。分析1では日英語児6名ずつの、1;9と3;0の時の録画データを用いた。分析2においては、英語児については独自の発話記録、日本語児についてはMiyataによるAkiおよびTaiデータを用いた。分析の結果、英語児は早期に普遍的および言語に特有な原則に合致したパターンを示したが、日本語児は3;0になってもそうならなかった。また、インプットのパターンが一貫しているほど、子どもの指示のパターンも早期から一貫していた。

Suzuki & Nomura (2020)では、心の理論の発達における言語学的要因を調べるため、日本語児130名の横断データ(Okayamaデータ)、および英語児24名のデータ(Gleasonデータ、Gleason, 1980)を分析した。心の理論の発達は、

通常誤信念課題を使って調査されるが、日本語児の誤信念課題の通過は英語児より遅いことが知られている。また、心の理論の発達には、心的状態動詞や、定型の補文の獲得が関わっている (de Villiers, 2000) ことが提案されている。後者について例を挙げておくと、John thinks [that the U.S. President is a linguist] という文においては、[] でくくった定型の補文が、John の誤った信念の内容を表している。すなわち定型の補文は、誤信念を表現する枠組みを提供するという考え方である。Suzuki & Nomura の分析の結果、子どもとその母親の両方の発話において、心的状態動詞の頻度および定型の補文の生起率ともに、英語の方が日本語よりも高かった。つまり、心的状態動詞や定型の補文といった言語的要素の獲得が、心の理論の発達に関わっているという仮説に沿った知見が得られた。

言語間の比較を行う場合は、異なる言語でも同じパターンが観察されることを示して、普遍的なメカニズムの存在を主張する研究と、言語間の違いを積極的に見出して、その違いを生み出す要因を探る研究に分けられよう。特に後者の場合に注意すべきは、対象となる複数の言語に共通の判断基準や分類基準を先行研究に基づいて明確に確立しておかないと、観察された違いが本当に主張される要因によるものなのかがあやふやになってしまうという点である。例えば上述の Guerriero らの研究では、日英語の指示表現を省略・代名詞・語彙的名詞句の3種に分類しているが、コレ、this といった直示表現を、代名詞とみなすのか、語彙的表現とみなすのかという点に課題が残る (Nomura, 2015)。Suzuki & Nomura (2020) では、心的状態動詞の後に続く形の分類 (定形節、非定形節、名詞節など) について、日英語で共通の基準を設定することに多くの時間を費やした。

4.5 個人差を調べる

これまで紹介してきた研究例のいくつかは、個人差を問題にしていた。第4.3節で言及した Uno (2016) では、格助詞ガの使用に関する個人差をインプットの頻度に結びつけ、加山 & 大嶋 (2021) では養育者が多様な項構造をイン

プットとして与えると、子どもの項構造の獲得が促進される、と主張した。また、第4.4節で紹介したGuerriero et al. (2006) では、指示の形態について、インプットが一貫していると子どもの示すパターンも一貫している、とした。個人差を問題にする研究では、その要因をインプットに帰するケースが多く、言語獲得におけるインプットの役割を重視する理論的立場に立っているケースが多いようである。

5. 自然発話データ分析の強みと限界

前節まで、自然発話データのタイプや、自然発話データの分析に基づく研究例を紹介してきた。本節では、それらの考察から明らかになる、自然発話データ分析の強みと限界について論じる。

5.1 強み

実験研究に対する、自然発話研究特有の強みとして、2点挙げることができる。まず、実験研究が不可能な年齢の子どもも調査可能である。実際、前節で紹介した研究の多くが、実験の指示を理解できないであろう1歳代のデータを分析対象としている。また、インプットを調査できるという点も重要である。言語獲得におけるインプットの役割を重視する立場で研究を行う場合はもちろん、特定の構造の産出（または産出の欠如）の要因がインプットではないことを示したい場合にも有効なデータが得られる可能性がある。

そのほか、注意を払って行えば利点となる自然発話研究の性質としては、小さな子どもや人見知りする子どもであっても、一旦収録環境に慣れてしまえば、普段どおりの自然な状況下で能力を発揮できるという点が挙げられる。ただし、研究者が同席して行う収録は、リビングや子ども部屋における自由遊び場面であることが多く、食事や入浴、保育園の場面などを含む日常生活の中で、極めて限られた場面しか収録できない。会話の参加者も養育者と対象児、それに研究者のみであることが多い。収録場面が子どもの言語使用に与える影

響を考えると、収録内容が真の意味で「自然」であるのかという点については注意を払っておく必要がある。

5.2 限界

自然発話研究には、限界も多く存在する。第一に、理解能力を調査することはできない。養育者の発話に対して適切に返答しているように見えても、それが十分な理解に基づくものなのか、もしくは全く理解せずに返した発話がたまたま適切だったのかを判断することは極めて困難である。

2つ目の限界としては、稀な構文の産出能力を調査することができないことが挙げられる。例えば「誰がどこに行ったの？」のような二重WH疑問文の産出能力を調べたくても、統計的な信頼性を確保できるだけの数の発話を収録するのは大変困難であろう。また、「花子が太郎を押しした」といった、それほど珍しいと思えない文であっても、実はほとんど発せられることはない。語順や格助詞の獲得に関する実験研究では、このような、主語、目的語、格助詞がすべて明示され、かつ主語と目的語が有生性の観点から交換可能な文が使用されることがあるが、自然な状況でこのような発話が産出されることはめったにない。

この限界と関連するのが、ある発話が観察されなかった場合に、それが知識・能力の欠如によるのか、それともたまたま発せられなかったのかを決定できないという点である。さらに、対象となる発話が発せられたとしても、それが定型表現、つまり丸覚えしたひとかたまりの表現として発せられたのか、個々の語彙の活性化と文法的な処理を伴って発せられたものなのかは簡単に決定できない。

最後に、上に述べた限界の多くに関わる自然発話データの特徴として、通常は対象児の全発話の数%しか収録できないという点が挙げられる。例えば1日8時間、週に56時間、子どもが発話をおこなっているとすると、1週間に1時間の収録では、全発話の1.7%しか記録できないことになる。この弱点を克服するために、英語圏では週に5時間といった「濃いデータ (dense data)」

に基づいた研究も行われているが（例えばMaslen et al., 2004やRoy et al., 2009）、数は限られる。理由としては、研究者自身が養育者でない限りは参加者の同意を得ることが難しいほか、文字起こしの負担など方法論的な問題が考えられる。また、どれくらい濃くすればどのように分析の質が向上するのかが不明瞭なため、多くの研究者が、入手しやすい通常の密度（1週間から1ヶ月に1時間程度）のデータを前提に研究を計画するという背景もあるだろう。

6. 意味のある分析を行うために

近年、L1獲得研究においては、実験研究が盛んである。実験研究では、選好注視法、真偽値判断課題、誘導的産出法といった実験手法の発達により、理解能力や、偶然に任せては減多に発せられない構文の産出能力を調査できるようになってきているため、第5.2節で述べた自然発話分析の限界の多くを克服することができる。しかし、そのことによって、「自然発話データ分析は実験研究の補助にしかならない」といったように、その可能性や役割を過小評価すべきではない。本節では、特にCHILDESの公開データを使用して、自然発話分析のみに基づいて研究を行う場合に、意味のある結果を得るために注意すべき点をまとめる。

6.1 目的の明確化や理論に基づく精緻な予測の重要性

実験結果のデータとは異なり、自然発話データは「ノイズ」に満ちている。このため、CHILDESシステムで対象となる発話を検索する際に、精度の高い検索を行うには一定の習熟が必要になる。緩めの検索をかけておいて、手作業で対象発話を抽出していくことも多い。手作業で対象発話を抽出する際には、調音が不明瞭であったり、発話が重なり合う中で断片化したりして、当該の発話が対象発話かどうか、判断がつきづらくなることも多い。さらに、抽出した対象発話を何らかの基準で分類する場合、あらかじめ決めておいた分類基準が不十分で、分析を進めれば進めるほど基準が曖昧になったり一貫性が保てなく

なったりして、最終的には分析の目的さえ見失ってしまう、ということも起こりうる。

このような事態を防ぐためには、できるだけ多くの先行研究にあたり、明確な研究目的と具体的な研究課題を設定することが不可欠である。同様に、対象発話を分類する場合には、分類の基準を先行研究に基づいて可能な限り精緻に設定しておく必要があるが、上述のように、自然発話データはノイズに満ちている。このため、一部のデータについて予備分析を行い、予想外のケースをあらかじめ洗い出し、当初設定した分類基準を見直したのちに本分析に取り掛かるのが理想である。また、特定の理論に基づいて分析を行う場合には、理論の予測を可能な限り具体的に絞り込んでおく必要があろう（良い例として、第4.2節のSugisaki, 2005; 2011 参照）。

6.2 頻出する現象を扱う

第5.2節で述べた、「理解能力は調査できない」「稀な構文を調査できない」「産出されないことは、獲得されていないことを意味するとは限らない」「産出されることは獲得を意味するとは限らない」といった、自然発話研究の本質的な限界を認識しておくことは、意味のある結果を生み出すために不可欠である。これらの限界のいくつかについては、頻繁に出現する現象を研究対象とすることが最善の対応策と言える。量的分析のためには一定以上の観察数が必須である。また、事例が多ければ、構文の生産性などを推測することも容易になり、出現時期や出現順序の特定に貢献する。CHILDESの使い方がある程度習得しておけば、興味のある現象がどれくらい頻繁に出現するかを、それほど時間をかけずに推測することができる。対象発話が十分に産出されているという確証が得られなければ、実験研究を計画するべきであろう。

6.3 適切なデータが存在するかどうかを確認する

公開データを使って研究を行う場合、明確な目的と緻密な研究計画が存在しても、適切なデータが存在しなければ研究は成立しない。収録頻度や収録時期

などをあらかじめ見て、対象となる現象について十分な観察数が得られるかどうかを見極める必要がある。例えば年齢について言えば、現在CHILDESの公開日本語データには、5歳代よりも後のデータはほとんど存在しない。また、唯一の横断データであるOkayamaデータは、対象児130名と規模が大きいのが、学生が現場で行った書き取りがもとになっている（ただし同時に録音も行われ、それをもとに修正された）という点に注意する必要がある。音声・音韻的な現象や、語用論的もしくは非言語的要因の分析が必要な場合は、音声やビデオの有無が分析の精度を大きく左右するであろう。

7. まとめ

本稿では、L1獲得研究における自然発話分析の役割を考察した。第5節でまとめたように、自然発話データにはいくつかの大きな強みが存在するが、同時にいくつかの限界が指摘される。これらの限界を認識し、頻繁な現象について、明確な予測や分析方法に基づき、適切なデータを分析することにより、自然発話分析のみでも、本来の強みが活かされた、意味のある研究が可能となるであろう。

註

- 1) 本稿の内容の一部は、2021年11月21日にオンラインで開催された、言語系学会連合・日本言語学会による共催公開特別シンポジウム「データベースをつくる・つかう：課題と展望」における筆者の発表「CHILDES データのつくり方・つかい方と自然発話研究の役割について」に基づいている。
- 2) <https://childes.talkbank.org>
- 3) <http://www2.aasa.ac.jp/people/smiyata/CHILDESmanual/chapter01.html>

参考文献

- de Villiers, J. G. (2000). Language and theory of mind: What are the developmental relationships? In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (pp. 83-123). New York: Oxford University Press.

- Gleason, J. B. (1980). The acquisition of social speech and politeness formulae. In H. Giles, W. P. Robinson, & P. M. Smith (Eds.), *Language: Social psychological perspectives* (pp. 21-27). Oxford, UK: Pergamon.
- Guerriero, A. M. S., Oshima-Takane, Y., & Kuriyama, Y. (2006). The development of referential choice in English and Japanese: A discourse-pragmatic perspective. *Journal of Child Language*, 33(4), 823-857.
- Ishii, T. (1999). The JUN Corpus, unpublished.
- 伊藤克敏. (1990). 『こどものことば—習得と創造—』 東京：勤草書房.
- 加山裕子 & 大嶋百合子. (2021). 「日本語母語児の動詞項構造獲得におけるインプットの役割」. 『日本語科学会第22回年次国際大会ハンドブック』.
- 木部暢子 (2019). 「対格表示形式の地域差—無助詞形をめぐる—」『東京外国語大学 国際日本学研究報告』 5, 20-32.
- Kido, Y. (2017, July 1-2). On the acquisition of V-V compounds in child Japanese: An empirical study using CHILDES [Conference presentation]. The Japanese Society for Language Sciences 19th Annual International Conference (JLS2017), Kyoto Women's University, Kyoto, Japan.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum.
- Maslen, R. J. C., Theakston, A. L., Lieven, E. V. M., & Tomasello, M. (2004). A dense corpus study of past tense and plural overregularization in English. *Journal of Speech, Language, and Hearing Research*, 47(6), 1319-1333.
- Matsuoka, K. (1998). The acquisition of Japanese case particles and the theory of case checking. Doctoral dissertation, University of Connecticut.
- Miyata, S. (2004a). Japanese – Miyata – Aki Corpus. Pittsburgh, PA: TalkBank. ISBN 1-59642-055-3.
- Miyata, S. (2004b). Japanese – Miyata – Ryo Corpus. Pittsburgh, PA: TalkBank. ISBN 1-59642-056-1.
- Miyata, S. (2004c). Japanese – Miyata – Tai Corpus. Pittsburgh, PA: TalkBank. ISBN 1-59642-057-X.
- 宮田 Susanne. (2012). 「CHILDES日本語版」. <http://www2.aasa.ac.jp/people/smiyata/CHILDESmanual/chapter01.html>
- 宮田 Susanne. (2018). 「JCHAT入力フォーマット及びJMOR08形態素分析用分かち書きガイドライン」. <http://www2.aasa.ac.jp/people/smiyata/CHILDESmanual/datafiles/wakachi2002v8/00index.html>
- Miyata, S., & Nisisawa, H. Y. (2009). Japanese - MiiPro – Asato Corpus. Pittsburgh, PA: TalkBank. ISBN 1-59642-474-5.

- Miyata, S., & Nisisawa, H. Y. (2010). Japanese - MiiPro - Tomito Corpus. Pittsburgh, PA: TalkBank. ISBN 1-59642-472-9.
- 宮田 Susanne・森川尋美・村木恭子 (編). (2004). 『今日から使える発話データベース CHILDES 入門』, 東京: ひつじ書房.
- 宮田 Susanne・中則夫. (1998). 「分かち書きガイドライン WAKACHI98 v.1.1」. 『教育心理学会フォーラムレポート』, No. FR-98-003. 教育心理学会.
- 村田孝次. (1981). 『言語発達研究—その歴史と現代の動向—』, 東京: 培風館.
- 中則夫・宮田 Susanne (1999). 「CLANを用いた日本語形態素解析プログラム JMOR」. 『教育心理学フォーラムレポート』 No. FR-99-0001. 教育心理学会.
- Nisisawa, H. Y. & Miyata, S. (2009). Japanese - MiiPro - Nanami Corpus. Pittsburgh, PA: TalkBank. ISBN 1-59642-473-7.
- Nisisawa, H. Y. & Miyata, S. (2010). Japanese - MiiPro - ArikaM Corpus. Pittsburgh, PA: TalkBank. ISBN 1-59642-475-3.
- 野地潤家 (1973-1977). 『幼児の言語生活の実態 I-IV』 広島: 文化評論出版.
- Noji, J., Naka, N., & Miyata, S. (2004). Japanese - Noji Corpus. Pittsburgh, PA: TalkBank. ISBN 1-59642-058-8.
- Nomura, M., & Shirai, Y. (1997). Overextension of intransitive verbs in the acquisition of Japanese. In E. V. Clark (Ed.), *The proceedings of the twenty-eighth annual child language research forum* (pp. 233-242). Stanford, CA: Center Study Language & Information.
- Nomura, J. (2015). Factors affecting the assessment of Japanese-speaking children's referential behavior. *The 17th Annual International Conference of the Japanese Society for Language Sciences (JSLS 2015) Conference Handbook*, 116-119.
- 岡山禧子 (1970-1974). 『母と子の対話の集録: 家庭場面 (1:3歳児 2:2・4歳児)』 大阪: 安田生命社会事業団.
- Oshima-Takane, Y. MacWhinney, B., Sirai, H., Miyata, S., & Naka, N. (eds.) (1998). *CHILDES for Japanese* (2nd Ed.). Nagoya: Chukyo University.
- Ota, M. (2003). *The development of prosodic structure in early words: Continuity, divergence and change*. Amsterdam/Philadelphia: John Benjamins.
- Roy, B. C., Frank, M. C., & Roy, D. (2009). Exploring word learning in a high-density longitudinal corpus. *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society* (July 29-August 1, 2009, Vrije Universiteit, Amsterdam, Netherlands), 2106-2111.
- Shirai, Y., Miyata, S., Naka, N., & Sakazaki, Y. (2001). The acquisition of causative morphology in Japanese: A prototype account. In M. Nakayama (Ed.),

- Issues in East Asian language acquisition* (pp. 183-203). Tokyo: Kurocio.
- Sugisaki, K. (2005). Early acquisition of basic word order: New evidence from Japanese. *Proceedings of the Annual Boston University Conference on Language Development*, 29(2), 582-591.
- Sugisaki, K. (2011). The distinction between case markers and postpositions in early child Japanese: New evidence for children's grammatical conservatism. In N. Danis, K. Mesh, & H. Sung (Eds.). *Annual Boston University Conference on Language Development 35 Online Proceedings Supplement*. https://www.bu.edu/buclid/files/2011/05/35-Sugisaki_BUCLD35.pdf
- Suzuki, T. & Nomura, J. (2020). Mental state verbs used by mother-child dyads in Japanese and English: Implications for the development of Theory of Mind. *First Language*, 40(1), 84-106.
- Suzuki T, & Yoshinaga N. (2013). Children's knowledge of hierarchical phrase structure: Quantifier floating in Japanese. *Journal of Child Language*, 40(3), 628-655.
- Tatsumi, T., Ambridge, B., & Pine, J. M. (2018). Testing an input-based account of children's errors with inflectional morphology: An elicited production study of Japanese. *Journal of Child Language*, 45(5), 1144-1173.
- Tatsumi, T., & Pine, J. M. (2016). Comparing generativist and constructivist accounts of the use of the past tense form in early child Japanese. *Journal of Child Language*, 43(6), 1365-1384.
- Uno, M. (2016). A usage-based approach to early-discourse pragmatic functions of the Japanese subject markers *wa* and *ga*. *Journal of Child Language*, 43(1), 81-106.