

# 縦断および横断データを用いた幼児早期出現語の獲得月齢の特定

小林 哲生<sup>1</sup>・南 泰浩<sup>2</sup>・永田 昌明<sup>1</sup>

NTT コミュニケーション科学基礎研究所<sup>1</sup> 協創情報研究部/<sup>2</sup>メディア情報研究部

{kobayashi.tessei, minami.yasuhiro, nagata.masaaki}@lab.ntt.co.jp

幼児は1歳になると、「まんま」や「パパ」といった初語を発し始め、1歳後半になると語彙を急速に学習していく。この1歳の時期に幼児がいつ、どんな語を獲得していく傾向にあるかを正確に突き止めることができれば、幼児の初期語彙発達の特徴を明確にできる。

この問題に対し直接的で且つ理想的な方法は、多数の縦断データ(longitudinal data)を取得し、そのデータから各語の獲得月齢(age of acquisition)を特定することである。縦断データとは、同一対象者のデータを継続的に取得したデータのこと、子どもが覚えた語を随時日誌に記録する「日誌法」や母子対話を自然場面で記録する「自然観察法」などにより取得できる。しかし、縦断データの取得には、多大な時間と労力を要するため、これまで、ごく僅かな人数のデータしか取得できなかった。そのため、各語の獲得月齢を算出するにはサンプル数が少なすぎるという問題があった。

縦断データを用いた研究のほかには、語彙チェックリスト法(子が発話できる語を、親がチェックリストに記入する方法)を用いて、異なる月齢サンプルから横断データ(cross-sectional data)を多数取得する方法がある。この場合、特定の子どもに対しデータ取得は1回限りでよい。ため、調査を依頼する親の負担も少なく、短期間に多数のデータを取得できる点で簡便な方法と言える。また、マッカーサー乳幼児言語発達質問紙(MacArthur-Bates Communicative Developmental Inventories: 略して CDI<sup>[1]-[3]</sup>)と呼ばれる標準化された語彙チェックリストも世界各国で出版されており、調査は比較的容易に実施できる。但し、この横断データには各語をいつ獲得したかという情報がないために、たとえ多数のデータが存在しても、各語の獲得月齢を直接的に算出することはできない。

こうした問題に対し従来用いられてきたのは、各語に対する「50%到達月齢(arrival age at 50%)」を算出する方法である<sup>[4]</sup>。50%到達月齢とは、1ヶ月ごとに分割した横断データから各語に対する獲得割合を算出し、その割合が最初に50%に到達した月齢のことを指す。例えば、「パパ」という語を、14ヶ月では42%、15ヶ月では51%、16ヶ月では64%の子どもが獲得していたとすると、パパの50%到達月齢は15ヶ月となる。この値をこれまでは「間接的」な獲得月齢とみなし、初期語彙発達の研究に役立ててきたが<sup>[5]</sup>、

問題はそれがどういった特徴を持つ指標かを十分に検証せずに使用している点にある。この点が明らかになれば、世界各国で使用頻度の高い語彙チェックリストに基づく横断データを有効に活用して、語彙発達に関するより精緻な研究が実現できるはずである。

本研究では、縦断データと横断データを大規模に取得し、前者から平均獲得月齢を、後者から50%到達月齢をそれぞれ算出し、早期出現語50語における両指標を比較する試みを行った。データ取得が難しい縦断データは、以前私たちが提案した「ウェブ日誌法」<sup>[6]</sup>で取得したものを使用した。横断データは、CDIを用いて自ら取得したデータを使用した。こうした試みにより、両データの関係性をより明確にし、日本語の早期出現語における獲得月齢のより正確な特定を行うとともに、語彙発達研究におけるデータ取得及び解析手法の問題点を検討する。

## 方法

### データ収集

縦断データ: 妊娠・育児支援サイト「goo ベビー」(<http://baby.goo.ne.jp/>)の会員を対象にして、2007年4月より約4年間、データ収集を行った。会員サービス「赤ちゃん成長ダイアリー」上に、子どもがいつ、どんな語を覚えたかを日齢単位で記録できる日誌形式のツールを作成し、ユーザに提供した。実際には、子どもが新しい語を発話できるようになったら、該当日時に語の発音(例:わんわん)と意味(例:犬)を記入する。これにより、語をいつ獲得したか(例:401日目)というデータが得られた。

本調査の開始は、会員用メーリングリストで告知し、研究目的や背景は特に告知せず、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(語彙成長グラフなど)を付加することにより、データ投稿の促進を狙った。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。

その結果、約4年間で約800名の会員ユーザから約20,000語の投稿を得た。そのうち、意味を特定できないデータと、対象月齢(9ヶ月齢以上36ヶ月齢未満)以外のデータを取り除き、合計18,611語(783名分)を有効データとして分析を行った。



## データの比較

縦断データから得られた平均獲得月齢と、横断データから得られた 50%到達月齢の関係を見るために、ピアソンの積率相関係数を算出した結果、両指標が正の相関を示すことがわかった( $r = .585$ ,  $p < .001$ )。これは、両指標がある程度一致することを示している。また、両指標の関係をプロットすると、横断データは縦断データよりも全体的に大きい値を示した(図 2)。このことから、横断データの 50%到達月齢は、縦断データの平均獲得月齢よりも一貫して遅い獲得時期を示す傾向にあることがわかった。

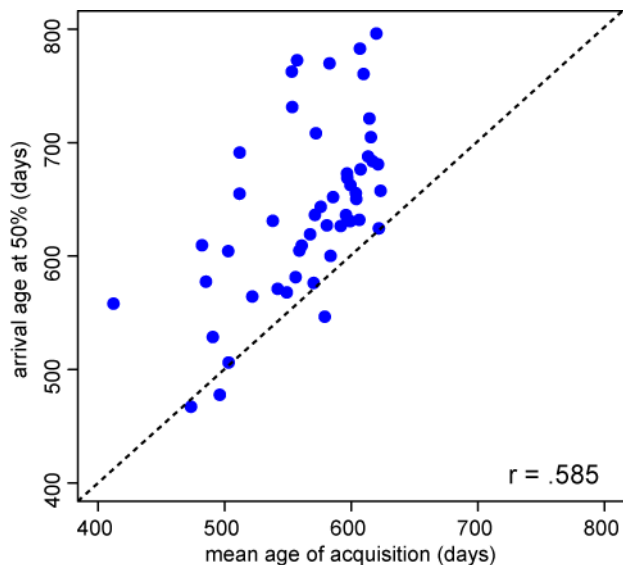


図 2. 平均獲得月齢と 50%到達月齢の散布図

## 考察

縦断データと横断データという特徴の異なるデータから早期出現語の獲得月齢(平均獲得月齢と 50%到達月齢)を算出した結果、両指標は正の相関を示すものの、完全には一致せず、横断データが一貫して高い値を示す結果となった。これは、先行研究で指摘されたことのない傾向であり、縦断および横断データを用いた研究を実施する場合には注意が必要であることを示唆している。

両指標が十分に一致しないひとつの理由は、横断データにおいてロジスティック関数での近似が十分でない語が一部存在するためだと考えられる。例えば、「まんま」や「くっく」などの語は、24 ヶ月歳以降の獲得割合がそれ以前よりも大きく減少する値があるため、近似精度を著しく低める結果となり、50%到達月齢がかなり遅く見積もられてしまう(図 3)。こうした獲得割合の減少は、通常の語彙発達では考えにくい傾向であるが、語によっては、母親がチェックリスト回答時に見落としや迷いが生じてエラーが起こる可能性があり得る。たしかに、「まんま」という最も早期に出現する語は、2 歳になると「ごはん」などの正式な語に替わっている可能性が高く、母親は「現時点で使用しない」と判断して、チェックから外す可能性もあり得る。ま

た子どもが「まんま」という語を早期にしか使用しなかったために、チェックリスト回答時に記憶から呼び起こせなかった可能性もあり得る。

現時点では、こうした可能性を考慮した定量的分析を実施できていないが、今回分析した各語の「50%到達月齢」だけでなく、そのほかの特徴量(ロジスティック曲線の傾斜など)を考慮することにより、早期出現語を複数のクラスに分類することを考えている。こうした分類ができれば、クラスごとの補正を行うことにより、横断データにおける 50%到達月齢のエラーが少なくなると同時に、語のタイプ(社会語や幼児語、述語など)の解析にも利用できる可能性が高い。

一方、縦断データにおいては、日誌法の性質上、ある特定の語に対して発話した子どものデータしかカウントされない。そのため、発話しない子どもも含める CDI データと比べると、日誌法では、対象となるデータ総数が少なくなる傾向にあり、こうした影響が平均獲得月齢を多少早期にシフトさせている可能性も否定できない。今後は、この点を考慮に入れた解析も考えていく必要がある。

このように、早期出現語の獲得月齢を特定するという試みは、大規模なデータを取得した現在でも問題が残る。しかし、本研究で得られた早期出現語は、ここ数年の間に得られた大規模な縦断データから特定されたものであり、現時点ではどの先行研究よりも、現在の幼児の語彙発達を反映していると考えている。私たちは、今後もデータ解析の精緻化を進めるとともに、こうしたノウハウから幼児の語彙数を正確に測定するテストの考案などにつなげていきたいと考えている。

## 参考文献

- [1] Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5), Serial # 242.
- [2] 小椋たみ子・綿巻徹(2004). 日本語マッカーサー乳幼児言語発達質問紙[語と身ぶり版]. 京都国際社会福祉センター.
- [3] 綿巻徹・小椋たみ子(2004). 日本語マッカーサー乳幼児言語発達質問紙[語と文法版]. 京都国際社会福祉センター.
- [4] 小椋たみ子・綿巻徹(2008). 日本の子どもの語彙発達の規準研究: 日本語マッカーサー乳幼児言語発達質問紙から. 発達・療育研究, 24, pp.1-42. 京都国際社会福祉センター.
- [5] Ma, W., Golinkoff, R.M., Hirsh-Pasek, K., McDonough, C., & Tardiff, T. (2009). Imageability predicts verb learning in Chinese children. *Journal of Child Language*, 36, 2, 405-425.
- [6] 小林哲生・永田昌明(2009). 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.

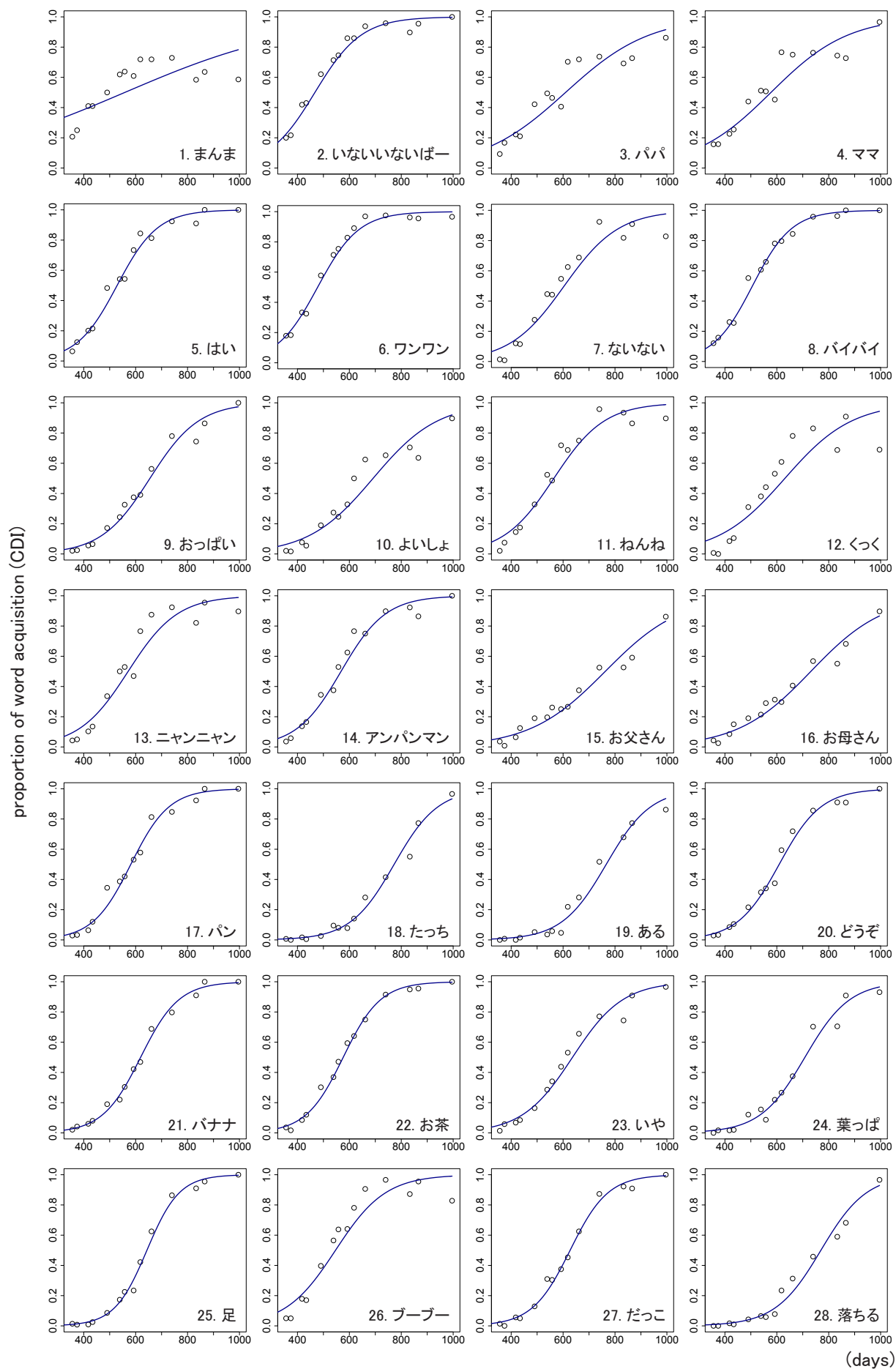


図 3. 横断データ (CDI) における各語の獲得割合をロジスティック関数で近似した図