

# 線形関数とプラトー割込による幼児語彙発達モデル化

南 泰浩      小林 哲生      杉山 弘晃

NTT コミュニケーション科学基礎研究所

{minami.yasuhiro, kobayashi.tessei, sugiyama.hiroaki}@lab.ntt.co.jp

## 1 はじめに

幼児の言語発達の研究は、人間の高度な情報処理過程を考察する上で科学的に重要な知見や示唆を提供し得るものでありかつ、音声認識、音声合成、言語処理などの言語に関わる工学的な研究へブレークスルーを提供し得るものでもある。しかし、現状は未解決の問題が多く、特に、音声認知や語彙学習、文法操作などの基本能力の中でも、語彙学習に関する科学的研究は、ほとんど進展が見られていない。幼児の個人性を捉える上で重要な特徴量のひとつは、語彙学習速度である。語彙学習速度は、1歳後半に起こるとされる語彙学習速度の急激な変化すなわち語彙爆発（またはボキャブラリー・スパート）の考察を抜きにして、求めることができない。

従来、発達心理学では、語彙爆発の現象を複数の言語で確認するため、語彙チェックリスト（親の回答に基づくアンケート調査）による大規模集団データ（横断データ）を用いた分析手法が取られてきた。この分析手法によると、平均の幼児語彙発達曲線が18~20ヶ月ころに変曲点を持つ曲線であることが分かっている[1, 2]。こうした事象から語彙爆発が一般的な現象であると考えられている。

語彙爆発事象の個人性に着目し、より細かく調べるためには、個人の長期に渡るデータ（縦断データ）を用いて幼児語彙発達曲線を解析する必要がある。しかし、横断データに比べ、縦断データ収集のための労力が莫大であるため、このような研究は数少ない。このため、これまで、語彙爆発の開始時期に関して、曖昧な定義[3]はあるものの、明確な定義はなされて来なかった。これに対して、我々は、日誌法による縦断データを用いて、幼児語彙発達曲線を折れ線により回帰し、その後半の直線を生成する最初のデータ点を語彙爆発開始時期と定義した[4]。さらに、その時期の前後で語彙学習速度を求め、幼児の特徴量とした。折れ線回帰では、二つの直線とデータとの二乗距離の和が最小になるように直線を決定した。Gangerらは、語彙学習

速度をロジスティック関数で近似する同様のモデルを提案している[5]が、この手法は語彙学習速度を平滑化する処理を利用するため、我々の手法より直接的でない。これまで、我々の提案したモデルが、ほとんどの場合、幼児語彙発達曲線に非常によく適合することを示してきた[4]。しかし、一部の幼児に対しては、折れ線が当てはまらなかったり、折れ線からの局所的なずれが見られたりした。我々は、この原因を調べるため、縦断データを個別に分析した。その結果、全ての幼児で、一単語も語彙学習が進まない期間（プラトー）が数多く見られることを発見した。我々は、このプラトーこそが語彙爆発の原因でもあると考え、線形の語彙発達関数に、プラトーが割り込むという新たなモデルを提案する。このモデルでは、見かけ上は語彙学習速度は変化するが、語彙学習速度を生成する機構は一定であり、内部機構の急峻な切り替わりを仮定しない。

急峻な語彙学習速度に関する内部機構の切り替わりを仮定しないモデルとして McMurray らのモデル[6]が最近提案されている。しかし、このモデルは、語彙の難易だけに依存するシンプルなモデルであるため、プラトーを含む複雑な発達曲線をうまくモデル化することはできない。本稿では、日誌法に基づく縦断データにより、提案モデルの検証を行う。

## 2 縦断データの収集

幼児語彙発達曲線を個人ごとに正確に求めるためには一人の幼児を長期間観察して結果を記録する縦断データが必要である。ここでは、データを記録する親への負担を軽減しつつ、かつ細かい時間ポイントでデータ取得が可能なデータ収集方法である日誌法を採用した。この方法は、幼児が単語を新たに学習（発話）した場合に、その日の日誌と共に、幼児が覚えた単語を記録するものである。この方法の有効性は科学的にも検証されている。

また、この方法の利点は、親にとっても比較的容易

に記録できる方式でありながら，記録年月日（幼児が新たな単語を覚えた年月日）と幼児の生年月日との差から，幼児が新たな単語を覚えた日齢を算出可能な点である．本稿では，日誌法を用いて，17人の幼児語彙発達曲線を求めた．

### 3 幼児語彙発達曲線の折れ線回帰

我々は，幼児語彙発達曲線を区分線形の考えを使って，二つの線形モデルを接合したモデルで表した [4]．二つの線形モデル（直線）で回帰することを，折れ線回帰と呼ぶ．ここで，回帰する2つの直線の前半の直線を  $y = ax + b$  とし，後半の直線を  $y = a'x + b'$  とする． $a$  及び  $b$  と  $a'$  及び  $b'$  はそれぞれ傾きと切片を表す．累積単語数の時系列を  $y_i$ ，獲得日齢を  $x_i$  とし ( $1 \leq i \leq I$ )，データを逐次2分割しながら，最小二乗法により最適な折れ線を求める．後半の折れ線を生成した最初のデータの日齢を語彙爆発開始時期とした．この操作を式 (1) に示す．ただし，この操作で直線の交点が妥当な場所に来ない候補は排除している．これにより，2つの直線  $y = ax + b$ ， $y = a'x + b'$  が得られる．

$$n' = \arg \min_n \left[ \min_{a,b} \sum_{i=1}^{n-1} (y_i - ax_i - b)^2 + \min_{a',b'} \sum_{i=n}^I (y_i - a'x_i - b')^2 \right] \quad (1)$$

図1，図2，図3に，幼児 C7-m, C8-m, C15-f の語彙発達曲線と折れ線回帰結果を示す．このIDのf, mは女性, 男性を表す．図中“ $\cdot$ ”が実際の語彙発達のプロットである．実線が折れ線回帰を表す．図を見ると，図3以外は折れ線が適合している．図3は，折れ線から逸脱する部分がかかなり多いことがわかる．また，図1，図2の折れ線も細かく見ると，幼児語彙発達曲線に適合しない部分が見受けられる．

### 4 線形関数とプラトー割込によるモデル化

我々は，図3のような不適合，及び，細かなずれがどこから来るのかを調べた．この結果から以下のことを発見した（図1，図2，図3参照）．

- 数日以上の上期間，一単語も語彙学習が進まない時期が頻繁に起きる．

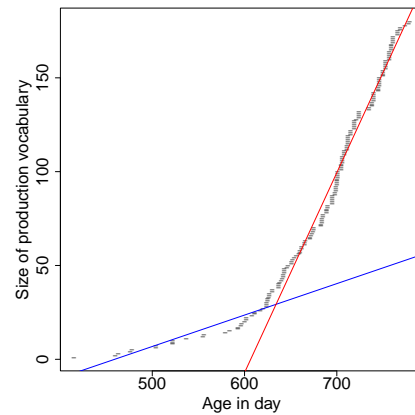


図1: 幼児語彙発達曲線とその折れ線回帰 (幼児 ID : C7-m)

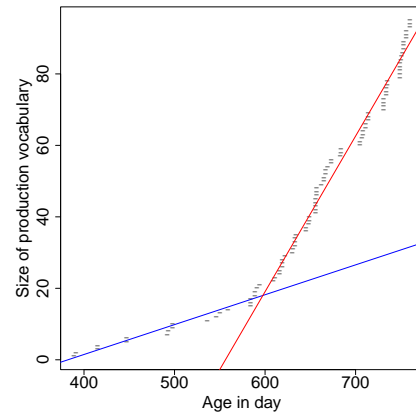


図2: 幼児語彙発達曲線とその折れ線回帰 (幼児 ID : C8-m)

- 語彙発達曲線の後半はほぼ直線となる．
- 語彙発達曲線は，傾きが一定の短い直線が層状に重なり合う．

この一単語も語彙学習が進まない期間を，英語の平原を意味し，心理学の分野では表層的な学習停滞を表す，プラトーと呼ぶことにした．以上のような観察から，我々は，語彙爆発がこのプラトーによって引き起こされ，その背景にある語彙発達の機構は極めてシンプルなものであると考えた．この考えに従い，幼児語彙発達のモデルとして，線形の幼児語彙発達曲線に，プラトーが割込むモデルを提案する．このモデルの妥当性を調べるため，プラトーを全て固定の期間に置き換え，語彙発達曲線が線形関数になるかどうか確認した．ここでは，6日間以上一単語も語彙学習が進まない期間

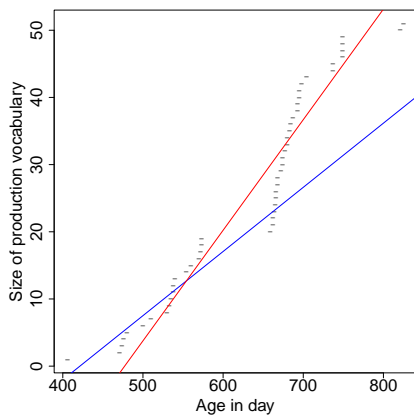


図 3: 幼児語彙発達曲線とその折れ線回帰 (幼児 ID : C15-f)

をプラトーとして定義した．これを全て固定の 1 日間に置き換えた．その後，得られた新たな幼児語彙発達曲線を最小二乗法により直線で回帰した．図 4 に幼児 C15-f のデータを示す．図中“-”が元々の語彙発達曲線を示し，“+”が検出されたプラトー開始位置を示す．プラトーを置き換えた語彙発達曲線を左側のプロットで示す．これを線形にモデル化したものを実線で示す．これを見ると図 3 では測定誤りとも疑われる可能性のある全く適合しなかった学習曲線が，直線に適合している様子が分かる．図 5 に残りの全ての幼児の語彙発達曲線とそのプラトー除去後の回帰直線を示す．ほぼ全ての幼児で，例外なく，プラトー除去後の語彙発達曲線が，直線に適合することがわかる．なお，ここでの検討では，パラメータ数に比べサンプル数が十分存在するので AIC などの手法によってモデルの善し悪しを比べることは不適切である．この結果得られた幼児毎の語彙学習速度（プラトー除去後の直線の傾き）を表 1 に示す．語彙学習速度は，0.44～1.35 語/日であり，プラトーを除去することにより，個人間のばらつきが少なくなることが分かる．

## 5 おわりに

本稿では，線形関数とプラトー割込によって幼児語彙発達のモデル化を行なった．17 人分の縦断データによって，モデルの妥当性を検証した．このモデルは，心理学で議論となっている語彙学習速度の変化を生み出す急峻な構造変化の存在を仮定しない．このモデルが示す表出のための語彙学習に必要な時間が線形であるということは，心理学のみならず，人間の学習機能

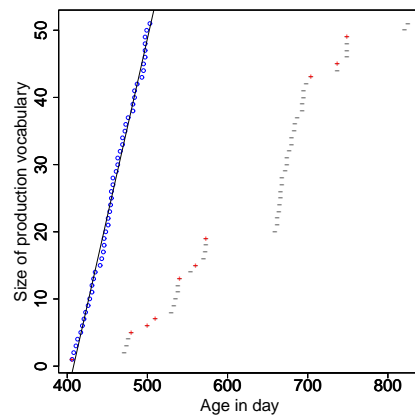


図 4: 線形関数とプラトー割込によるモデル化 (幼児 ID : C15-f)

の分析を参考にして来た音声認識・音声合成，言語処理の研究に大きな影響を与えられられる．今後の課題として，今回発見したプラトーが何によって引き起こされるのかを調べることが挙げられる．

## 参考文献

- [1] K. Nelson, "Structure and Strategy in Learning to Talk," Monographs of the Society for Research in Child Development, pp. 1-135, 1973.
- [2] L. Bloom, One Word at a Time: The Use of Single Word Utterances before Syntax: Mouton The Hague, 1973.
- [3] B. A. Goldfield, J. S. Reznick, "Early Lexical Acquisition: Rate, Content, and the Vocabulary Spurt," Journal of Child Language, vol. 17, no. 01, pp. 171-183, 1990.
- [4] 南 泰浩, 小林哲生, 杉山弘晃, "折れ線回帰による語彙爆発開始時期の推定," 電子情報通信学会技術研究報告 SP, 3 月, 2012.
- [5] J. Ganger, M. R. Brent, "Reexamining the Vocabulary Spurt," Developmental Psychology, vol. 40, no. 4, pp. 621-632, 2004.
- [6] B. McMurray, "Defusing the Childhood Vocabulary Explosion," Science, vol. 317, no. 5838, pp. 631-631, 2007.

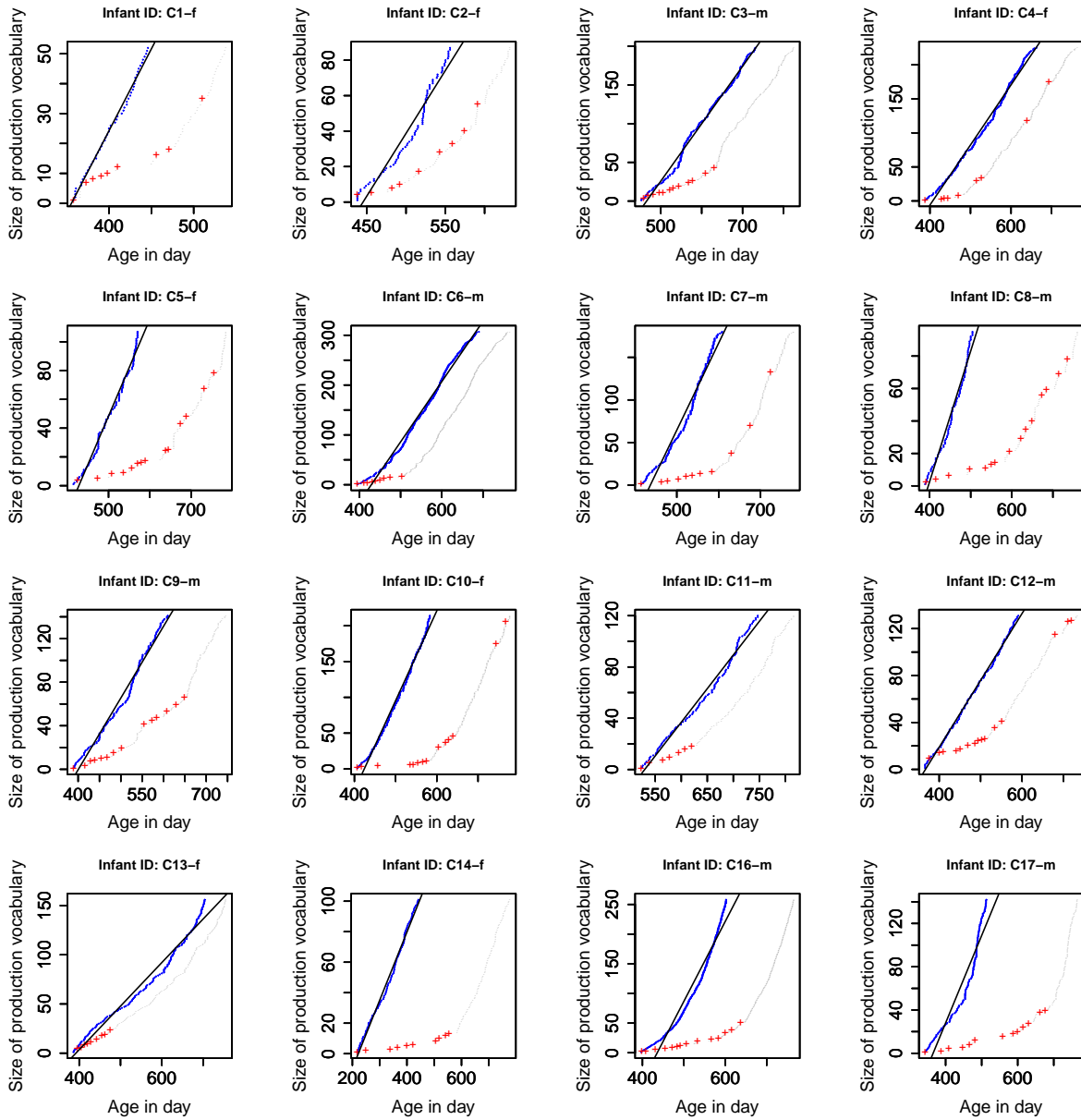


図 5: 線形関数とプラトー割込によるモデル化 (その他全て)

表 1: 幼児固有の語彙学習速度

幼児 ID	C1-f	C2-f	C3-m	C4-f	C5-f	C6-m	C7-m	C8-m	C9-m
速度 (語/日)	0.55	0.71	0.74	0.88	0.67	1.21	1.02	0.80	0.66
幼児 ID	C10-f	C11-m	C12-m	C13-f	C14-f	C15-f	C16-m	C17-m	-
速度 (語/日)	1.25	0.52	0.57	0.44	0.45	0.52	1.35	0.81	-