

ウェブを用いた幼児言語発達研究:大規模縦断データ収集の試み

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 協創情報研究部

〒619-0237 京都府相楽郡精華町光台 2-4

発表要旨 本研究では、幼児の初期語彙発達に関する縦断データを大規模に収集するために、わが子が覚えた単語を記録できる日誌形式のツールをウェブ上で公開し、語彙発達データの投稿を呼びかけた。その結果、20 ヶ月間に約 800 名の親から約 12,000 語の単語データをいただき、それを分析することにより、初期語彙発達の特徴を一部明らかにできた。また本データから言語発達の辞書(こども語辞書)を作成し、育児情報ツールとしてウェブ上で公開した。

1. 背景と目的

こどもがどのようにことばを獲得していくのかという問題は、人間科学の分野における重要な問題のひとつであり、この問題を明らかにするには、初期の語彙発達過程に注目し、その特徴を特定することが重要である。

従来は、こどもがその時点で発話できる単語を、親がチェックリストに記入する形で報告する「語彙チェックリスト法」を用いて、研究が進められてきた(Caselli et al., 1995; Fenson et al., 1994; 小椋, 2007; 小椋・綿巻, 2004)。この方法は、複数の月齢群(例えば、12、18、24 ヶ月齢)の横断データ(cross-sectional data)を短時間で多数取得することが可能なため、比較的簡便な方法と言える。しかし、各月齢データを異なる被験者のデータで比較することになるため、個人差などをいっさい考慮できず、発達過程を再構成せざるを得ないという問題点が指摘されている。

より有効でかつ直接的な方法は、こどもが覚えた単語を随時日誌に記録する「日誌法」や、こどもの発話場面を撮影して書き起こす「自然観察法」などにより、特定の個人を追跡調査し、縦断データ(longitudinal data)を取得する方法である。しかし、こうした縦断データの取得には、多大な時間と労力を要するため、これまでは、ごく僅かな人数のデータ(例えば、研究者のお子さんのデータ)しか取得できなかった(天野・近藤・加藤, 2008; MacWhinney, 2000; Tomasello, 1992)。そのため、語彙発達に関する一般化・理論化を十分

検討できないまま、多くの問題が残されたままになってきた。

こうした研究上の問題を解決する方法のひとつとして考えられるのが、ウェブを利用して大規模に縦断データを収集することである。現在、日本のウェブ普及率は高くなり、日本全国の家庭からウェブを通じてデータを収集する環境は整ってきているように思える。こうした方法が有効に機能するようであれば、多大な労力をかけずに、大規模な縦断データを取得できる可能性が高いと考えられる。

そこで本研究では、ウェブで日誌法を実施することを考え(以下、「ウェブ日誌法」; 小林・永田, 2008a)、こどもがどんな単語をいつ覚えたかを日齢単位で記録・管理できる日誌形式のウェブツールを作成・公開し、育児中の親が記録したデータを用いて、初期語彙発達に関する縦断データを大規模に取得する試みを行った。

2. 方法

2.1. 実施期間

2007 年 4 月から 2008 年 12 月までの約 20 ヶ月間、ウェブを利用したデータ収集を実施した。

2.2. 実施方法

NTT レゾナント社のポータルサイト goo にある育児支援サイト「goo ベビー」(<http://baby.goo.ne.jp/>)の会員ユーザを対象に、初期語彙発達のデータ収集実験を行った。goo ベビーの会員サービスのひとつである

「赤ちゃん成長ダイアリー」上に、こどもがいつ、どんな単語を覚えたかを日齢単位で管理できる日誌形式のツールを作成し、会員ユーザに無料で提供した。本ツールの開始は、goo ベビー会員用のメールマガジンと、サイト上でアナウンスした。

本ツールは、記録されたデータから「成長グラフ」と「月齢別リスト」が自動表示される仕様になっており、ユーザがいつでもことばの成長記録を確認できるように工夫した。また日記や写真、身長・体重なども同時に記録できる仕様になっており、成長記録ツールとして利用価値の高いものにした(図1)。

具体的には、こどもが新しい単語を発話するようになったら、赤ちゃん成長ダイアリー上の日付をクリックし、その日付のページに単語の発音(例、わんわん)と、意味(例、イヌ)を該当欄にそれぞれ記入する(図1)。登録済みの単語は重複投稿できないようにし、いつどの単語を覚えたかを記録するツールとして設計した。

Figure 1 displays two screenshots of the 'goo Baby' 'Baby Growth Diary' tool. The left screenshot shows the data entry interface, which includes a photo upload section, a date management calendar, a diary section, a body height/weight section, a section for recording words learned (with fields for word and meaning), and a monthly list of words. The right screenshot shows the growth graph and the monthly list of words.

図1. goo ベビー「赤ちゃん成長ダイアリー」のデータ入力画面(左)と成長閲覧画面(右)。入力の際は、はじめて覚えた単語を、「発話」と「意味」の点から記入。

2.3. 取得データ

本研究で取得したデータは、(1)発話、(2)意味、(3)覚えた日付(以上、ダイアリーより)、(4)生年月日、(5)性別、(6)居住地(都道府県)、(7)出生順位(以上、会員登録のデータより)であった。また、単語を覚えた日付と生年月日の差を算出し、各単語の獲得月齢(age of acquisition, 以下 AOA)とした。

3. 結果

3.1. データ概要

2007 年 4 月より 20 ヶ月間、実験を行った結果、約 800 名のユーザから約 12,000 語の単語データを投稿いただいた。単語の種類は約 2,000 語であった。ただし、ユーザ全員が継続的にデータを投稿するわけではないので、縦断データとして利用できるものは、全体の約1割の約 80 名程度のデータであった。

3.2. ユーザのプロフィール

投稿いただいたユーザの 95%が母親で、平均年齢は 33.7 歳であった。対象となった被験者は、52%が男児、48%が女児で、やや男児が多い傾向にあった。出生順位は、第1子が 77%、第2子が 20%、第3子が 3%となった。また、地域別に見ると、関東が 49%、関西が 19%、中部が 12%で、都市部からの投稿が多い結果となった(図2)。

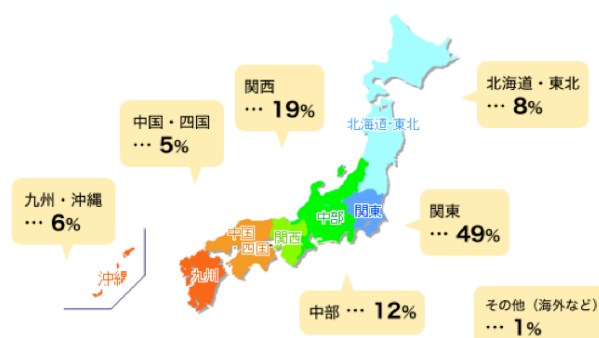


図2. データ投稿者の地域別分布。関東が約5割。

3.3. データ分析

日本語を母語とする幼児の語彙発達の特徴を特定するために、投稿数の多かった上位 50 語の平均獲得月齢(AOA)を算出し、それらを昇順に並べたものを「最初の 50 語(first 50 words)」とした(表1参照; 小林・永田, 2008b)。平均獲得月齢の上位 5 語は、まんま(食べ物)、おっぱい、ばー(いないいないばあ)、ま(ママ)、はい(返事)であった。

次に、この 50 語を、言語発達分野でよく使用されるジャンル(Fenson et al., 1994)で分類したところ、最も多かったのは、社会的な語(挨拶・日課・会話語)で 32%、普通名詞が 26%であった。また、様子・性質の語が 16%、人々が 12%、動作語が 8%、代名詞・疑問詞が 6%であった。これらの結果から、日本語を母語とする幼児の場合、社会的な語が初期語彙に最も多く含まれていることがわかった。

表1. ウェブ日誌法を用いて特定した、日本語を母語とする
幼児の「最初の50語」(小林・永田, 2008b)

順位	単語	平均獲得月齢	N	平均獲得日齢	SD
1	まんま(ごはん)	15.4	73	469.6	151.9
2	おっぱい	15.5	33	470.2	87.8
3	いないいないばあ	15.5	57	471.9	93.0
4	ママ	15.6	80	474.5	121.8
5	はい(返事)	15.8	34	479.6	110.0
6	ワンワン(犬)	15.9	78	483.8	111.9
7	ねんね(寝る)	16.0	32	485.3	88.1
8	パパ	16.0	60	486.1	118.6
9	バイバイ	16.3	56	496.8	123.5
10	よいしょ	16.5	21	501.8	104.8
11	どうぞ	17.5	43	533.2	138.3
12	お母さん	17.6	38	534.6	137.2
13	お父さん	17.7	30	538.5	153.0
14	ニャンニャン(猫)	17.7	42	538.6	121.2
15	くっく(靴)	17.7	25	539.0	82.1
16	ある/あった	18.0	21	546.8	170.6
17	痛い	18.0	42	547.2	100.0
18	ないない(片づける)	18.1	31	551.3	186.2
19	バナナ	18.2	21	554.9	111.9
20	ブーブー(車)	18.5	38	563.7	120.6
21	アンパンマン	18.8	75	571.8	129.4
22	いや/いやだ	18.9	44	573.5	174.4
23	おいしい	18.9	60	574.1	114.5
24	じいじ/おじいちゃん	18.9	42	574.6	105.4
25	お茶	18.9	31	576.1	151.0
26	ごちそうさま	19.3	26	586.8	197.1
27	これ	19.3	29	586.9	133.7
28	だっこ	19.4	38	590.0	128.4
29	あっち	19.4	21	591.0	78.4
30	ありがとう	19.4	21	591.6	102.6
31	ちょうだい	19.6	26	596.3	143.2
32	きれい	19.7	22	598.0	94.8
33	おしっこ	19.7	25	599.2	94.5
34	あつい	19.8	31	601.5	103.8
35	できる/できた	19.8	20	601.8	122.2
36	チョウ	19.9	20	604.1	74.5
37	ばあば/おばあちゃん	20.0	47	608.7	134.4
38	牛乳/ミルク	20.0	21	609.0	82.6
39	おいで	20.2	24	615.9	130.7
40	なに/なんで	20.3	22	616.7	130.9
41	電車	20.3	23	617.6	92.3
42	かわいい	20.4	21	620.3	114.2
43	ゾウ	20.8	24	633.6	110.6
44	ジュース	21.0	22	638.1	100.9
45	こわい	21.5	22	654.9	86.9
46	行く/行った	21.8	30	663.6	165.6
47	キリン	22.0	21	668.9	98.5
48	赤	22.0	20	670.4	122.2
49	お帰りのさい	22.5	21	683.7	106.0
50	大きい	23.1	21	703.0	100.2

こうした「最初の 50 語」は、他の言語圏でも特定されており、それらと比較することで各言語の初期語彙発達の特徴をある程度知ることができる。その際に使用される指標のひとつが、「NV 比(noun-verb ratio)」である。これは、最初の 50 語における名詞と動詞の比であり、その言語で名詞(あるいは動詞)がどの程度優勢であるかを検討できる。先行研究によると、英語圏

(米国)では NV 比が 12:1、韓国語では 2:1 であった(Kim et al., 2000)。そこで本データでも同様に NV 比を算出すると、約 3:1 であった。これらの結果から、日本語の NV 比は、韓国語に非常に近く、英語圏の NV 比と大きく異なることがわかった。

4. データの公開(こども語辞書)

本研究では、ウェブで取得したデータを幼児語彙発達研究の基礎資料として利用するだけでなく、本研究のアウトリーチとしてデータの公開も行ってきた。

そのひとつの試みが、「こども語辞書」である(図 3)。これは、0-3 歳児の語彙発達過程を、意味・音声・月齢の点から解析表示するウェブツールのことであり、2008 年 1 月より、NTT の新技術の実験サイトである「goo ラボ」(<http://labs.goo.ne.jp/>)で公開している。

こども語辞書では、ある単語(例: わんわん)を検索すると、それがどんな意味(例: 犬、動物、キャラクター)でいつ頃(平均月齢: 15.5 ヶ月齢, 習得分布グラフ)獲得されるかを解析表示することができる(図 4)。

使用例としては、幼児語(ぶーぶー、じいじ、おっちゃん)などの、この時期に特有のことばを調べたり、ある月齢(例えば、18 ヶ月)にどんな単語を覚える傾向にあるかを調べたりする場合に有用である。本研究では、こうした情報ツールが、楽しい育児につながるかどうかを検証中である。



図 3. goo ラボ「こども語辞書」のトップページ



図 4. goo ラボ「こども語辞書」の検索結果(わんわん)

なお、こども語辞書の公開のもうひとつの理由は、「ユーザの手で作る辞書」という仕組みを導入することにより、データ投稿を促していきたいという意図があった。現在、ウィキペディア百科事典に代表されるような CGM (consumer generated media) は、群衆の叡智 (wisdom of crowds) と呼ばれ、ユーザによる知の創造活動として大きな成功を収めている (e.g., 佐藤, 2008; Surowiecki, 2005)。この意味では、こども語辞書は、幼児の言語発達というカテゴリーにおける CGM であり、今後、新しく親になる方々のために、有用な育児情報を提供できると考えている。

5. 今後の展開

本研究では、ウェブを通じて初期語彙発達に関するデータを多数収集することができたが、ユーザの多くが継続的にデータを記録するわけではないので、縦断データを大規模に収集できたというレベルにはまだない。従って、今後も継続的にデータを収集していく必要がある。

今後、大規模な縦断データを収集できたならば、初期語彙発達の解明や個人差の特定、発達障がいのスクリーニング技術などにつなげていくことができる。

6. まとめ

本研究では、幼児の初期語彙発達に関する縦断データを大規模に収集するために、わが子が覚えた単語を日齢単位で記録できる日誌形式のツールをウェブ上で公開し、語彙発達データの投稿を呼びかけた結果、約 20 ヶ月間に約 12,000 語の単語データをいただき、それを分析することで、初期語彙発達の特徴を一部明らかにできた。また本データから言語発達の辞書を作成し、育児情報ツールとしてウェブ上で公開した。今後の展開として、データ収集を継続し、大規模な縦断データに基づくデータベースを構築し、言語発達研究の貴重な資料としたい。

文 献

- ・ 天野成昭・近藤公久・加藤和美 (2008). NTT 乳幼児音声データベース (INFANT). 国立情報学研究所音声資源コンソーシアム.
- ・ Caselli, M.C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sander, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, 10, 159-199.
- ・ 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*, Vol. 59, No. 5, Serial # 242.
- ・ Kim, M., McGregor, K.K., & Thompson, C.K. (2000). Early lexical development in English- and Korean-speaking children: language-general and language-specific patterns. *Journal of Child Language*, 27, 225-254.
- ・ 小林哲生・永田昌明 (2008a). ウェブを通じた初期語彙発達データ収集の試みとその応用. 第 8 回日本赤ちゃん学会発表要旨.
- ・ 小林哲生・永田昌明 (2008b). 日本語を母語とする幼児の初期語彙発達: ウェブ日誌法による早期出現語彙の特定. 第 72 回日本心理学会 大会論文集, pp.1122.
- ・ MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. Volume 1: Transcription format and programs. Volume 2: The Database.* Mahwah, NJ: Lawrence Erlbaum Associates.
- ・ 小椋たみ子 (2007). 日本の子どもの初期の語彙発達. *言語研究*, 132, p.29-53.
- ・ 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[語と身ぶりの手引]. 京都国際社会福祉センター.
- ・ 佐藤良明 (2008). 投稿を意識しない CGM. *月刊ビジネスコミュニケーション*.
- ・ Surowiecki, J. (2005). *The wisdom of crowds.* Anchor Books.
- ・ Tomasello, M. (1992). *First verbs: A case study of early grammatical development.* Cambridge Univ. Press.

(Tessei Kobayashi & Masaaki Nagata)