

幼児早期出現語の理解-発話指標による

名詞学習の優位性の検証

南 泰浩¹・小林 哲生²

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

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

1. はじめに

幼児は1歳頃に初語を発話し始め、1歳後半に到達する頃に語彙を急速に獲得していく。この傾向は、どの幼児に対しても大きくは変わらない。しかし、発話する語の内容は、言語、文化、環境によって大きく異なる。幼児語彙学習の過程において、どのような語を学習していくのかを知ることは、言語を獲得するプロセスを解明する上で、科学的に重要な意味を持つだけでなく、工学的にどうやってコンピュータに言語を教えるかという課題への解法を提供する上でも価値がある。

これまで、Gentnerらは、幼児早期出現語には、動詞よりも名詞の方が多く現れることから、名詞学習が優位であると主張してきた[1]。Gentner らは、この理由を、図1に示すDivision of dominance continuumを用いて説明している[2]。

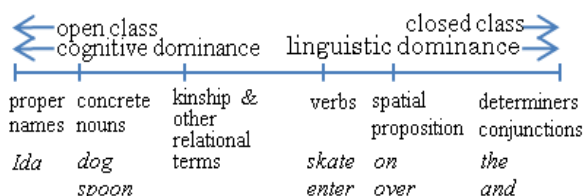


図1 Division of dominance continuum

図1に示すように、cognitive dominance からlinguistic dominanceへ変化する 抽象的な空間を定義し、語をその区間上に配置される連続体であると仮定した。そして、さらにnatural partitionsとrelational relativityという二つの仮説を導入し、cognitive dominanceの語の方がlinguistic dominance な語より早く学習できるという説明をしている。natural partitionsとは、知覚表象の密着力のある集まりはものとして概念化される傾向にあり、それらは名詞として言語化されるという仮説である。また、Relational relativityは、述語の方が、概念から言語へのマッピングが変化し易い、すなわち、関係性(述語)を言語化する方が、ものをあらす項目を言語化するより自由度大きいという仮説である。この

ことから、この空間上で、名詞が動詞より言語化し易く、名詞学習が優位であると考えている。

また、Maguire らは、図2のようなSICI continuumを仮定し、名詞学習の優位性を説明している[3]。ここでは、Shape, Individuation, Concreteness, Imageability を抽象的空間の軸と置き、それらの頭文字を取って、この空間に存在する語をSICI 連続体(SICI continuum)と呼んでいる。この図で、左に行くほど、Shape が簡単で、Individuation が簡単で、高い concreteness を持っており、高い Imageability を持っている。名詞と動詞は、その概念を考えると、それぞれ、左側、右側に配置される。そして、図のように名詞と動詞で重なりのある分布を仮定することで、名詞と動詞の学習の違いを説明している。

しかし、これらの抽象概念の軸と、語の学習の難易度を直接結びつける具体的な事実は得られていない。これは、語の難易度を示す直接的な手法が、理解及び発話する語の割合以外に提案されてこなかったからである。語の割合は語の学習開始時期が、母親の入力や、言語・文化によって大きく影響を受けるため、学習の優位性を測定するための指標として十分ではない。このため、名詞学習の優位性に対しては未だに議論が続いている[4,5]。

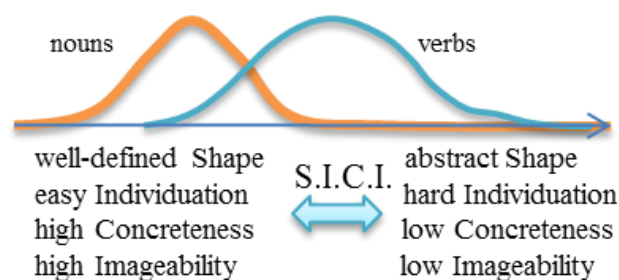


図2 SICI 連続体(SICI Continuum)

本稿では、幼児早期出現語彙に関する抽象的な空間の存在を検証するため、語の難易度を測定する直接的な指標を提案する。この指標を求めるために、横断的な日本語版 CDI による語彙チェックリスト[6,7]の結果から得られる語の理解と発話の習得率を

ロジスティック関数で近似する．そして，ロジスティック関数の 50% 習得率時の日齢を理解と発話に関して求め，その差分を指標とした．この指標のことをここでは，理解-発話指標(C2P index)と呼ぶ．

本稿では，2 節，3 節で用いたデータとその処理方法について述べ，4 節で理解-発話指標(C2P index)について詳説する．5 節では理解-発話指標を用いた名詞の優位性について議論していく．

2. 日本語 CDI データ

2006 年 4 月から約 6 年間に，実験調査のため NTT に来訪した，生後 10~32 ヶ月齢の幼児を持つ京都・大阪・奈良に在住の母親 1,699 名から，CDI [5,6]を用いて横断データを取得した．取得時には，プレールームにて調査者が記入上の注意事項を説明後，実験担当者が母親から質問を受け付け，それに回答し，母親が不明な点を解決した状態で母親が CDI の記入を行った．20 ヶ月以前は，2 種類ある CDI のうち「語と身振り」版(448 語)[6]を，21 ヶ月以降は「語と文法」版(711 語)[7]を使用した．語と身ぶり版では，各語に対し「わかる」と「わかる/言える」の回答欄が CDI にある．この欄はそれぞれ本論文での，語の理解及び発話にそれぞれ対応している．また，語と文法版では，各語に対し「言える」かどうかの判断をしてもらった．これは本論文での発話に対応する．なお，記入にかかる時間は，(月齢により変動するが)18 ヶ月以下であれば約 10-30 分，19 ヶ月以上であれば約 20-50 分であった．

3. ロジスティック関数による習得率のフィッティング

対象幼児の月齢に基づいて 1 ヶ月ごとにデータを分割し，各月齢における各語の習得率を計算した．ただし，十分な人数を確保できなかった月齢は，習得率のばらつきが大きくなるため，集計せずに欠損値とした．次に，各月齢(x)における習得率($f(x)$)をロジスティック関数でモデル化した．この式を以下に示す．

$$f(x) = \frac{ae^{cx+b}}{1+e^{cx+b}} \quad (1)$$

このとき関数のフィッティングには非線形最小自乗法を用いた．ただし， a が 1 より大きくなる場合は， a を 1 に固定し，再度非線形最小自乗法により b, c を計算した．この関数を利用し，語の習得率をフィッティングした例を図 2 に示す．この図では幼児の発話の習得率をモデル化した．

発話の習得率は，このようなモデル化を行うことができた．しかし，理解では，幼児の語彙習得速度が速いため，母親の語彙把握が難しく，早期にデータが取れなくなる．このため，発話の習得率が理解の習得率を超えてしまう語が数多く出現した．ここでは，これをさけるため，理解と発話の関係，すなわち，人は語を理解してから，それを発話につなげるという関係を利用する．ここで，理解のための，

ロジスティック関数を以下のように設定する．

$$f'(x) = \frac{a'e^{c'x+b'}}{1+e^{c'x+b'}} \quad (2)$$

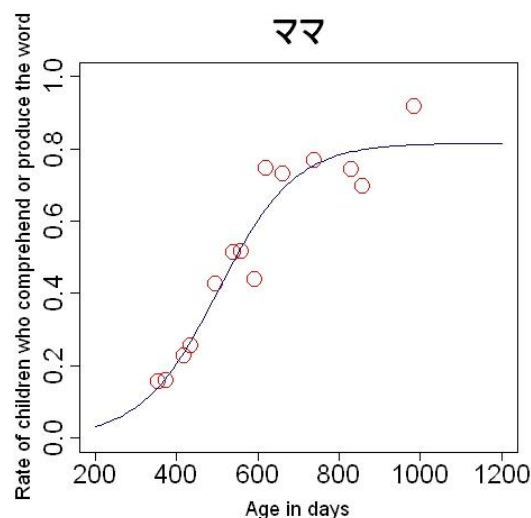


図 2 ロジスティック関数による”はい”の発話の習得率のフィッティング

このとき，上記の関係から $f(x) \leq f'(x)$ という拘束条件を設定する．この拘束条件下で a', b', c' を推定する．しかし， $f(x) \leq f'(x)$ の条件化で a', b', c' を求めるのは計算上困難であるため，ここでは $a' > a$ となるような拘束条件を置いて a', b', c' を計算する．これを行うため， a' が a より小さくなる場合には， a' を a と固定し非線形最小自乗法により b', c' を計算する．また， a' が 1 より大きくなる場合は， a' を 1 に固定し，再度，非線形最小自乗法により b', c' を計算した．

以上により，理解に関する習得率と発話に関する習得率をモデル化したロジスティック関数が求まる．

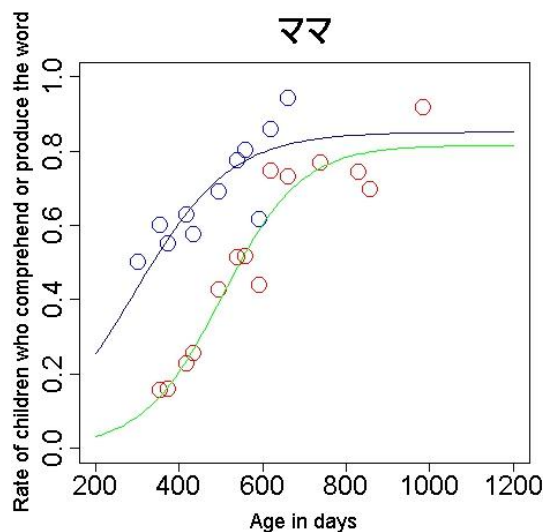


図 3 ロジスティック関数による”はい”の発話と理解の習得率のフィッティング

図3に理解と発話の習得率をロジスティック関数でフィッティングした例を示す・

4. 理解-発話指標の計算

Gentner が指摘しているように、語の理解や発話は、様々な環境からの入力に依存している[2]。筆者が考える語学習のプロセスを図4に示す。幼児の語学習プロセスは、親のインタラクションや言語の持っている特徴に影響される。その影響を図上部に記述した。語学習のプロセスは大まかに次のようになると考えている。まず、語の学習開始がインタラクションや言語に影響され決定される。そして、語を理解するまでも、様々な入力の影響は受け続ける。さらに、それらの入力の影響を受けながら、最終的に語が発話される。

Gentner らはこれらの入力の影響を考慮できれば、概念の固体化(単語化)の貢献だけが残るとしている[2]。しかし、実際にはこれを行うのは容易ではない。そこで、図3のように理解の日齢あるいは発話の日齢を調べ、ある一定の期間や日齢が上位の50単語の品詞の割合を調べ、名詞学習と動詞学習の優位差を決定していた[1]。しかし、この方法での問題は、様々な入力の影響を排除できないことである。特に、語の理解日齢や語の発話日齢の値は、語学習プロセスのスタートポイントに大きく依存するが、このポイントが明確には分からない。

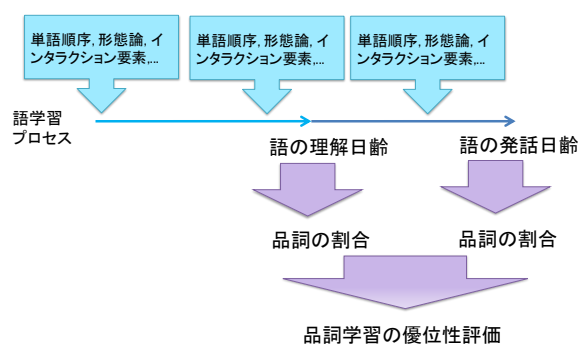


図4 これまでの名詞学習の優位性の評価手法

そこで、筆者らは図5のような指標を考える。これは、語の発話日齢から理解日齢を引いたものである。これにより語学習のスタートポイントの不確定さを相殺できる。また、影響を及ぼす入力も図4に比べ少なくなることがわかる。ただし、この手法では、この指標が理解から発話までしか考慮していないので、学習全体を把握していないという問題もある。しかし、語彙学習は、そもそも複雑な過程であり、それを部分的に分解し、その各々に関して分析

することは、語彙学習全体を見渡す上で非常に重要である。このため、部分的ではあっても学習の難易を示す指標を提案することは価値がある。

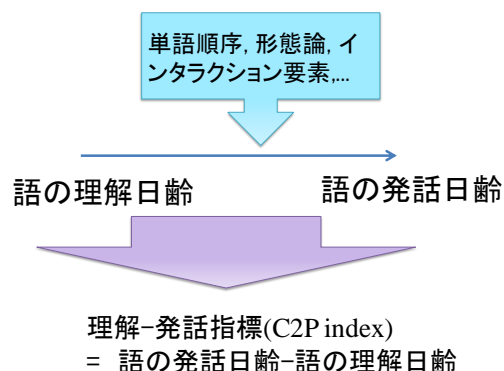


図5 提案する指標

この指標では、語の理解日齢と発話日齢を如何に計算するかが重要である。ここでは、幼児の50%が発話した日齢を語の発話日齢と定義した。発話日齢は、得られたロジスティック関数を用いて $f(x)=0.5$ を解くことによって求めた。これを解くために、本稿では、Rに実装されているブレント法を用いた。同様に語の理解日齢も幼児の50%が理解した日齢と定義した。これもロジスティック関数を使って $f'(x)=0.5$ を解くことにより求めた。最後に、得られた発話日齢と理解日齢の差を理解-発話指標(C2P index)とした。図6にこの例を示す。図中米印で示したものが語の理解日齢及び発話日齢である。

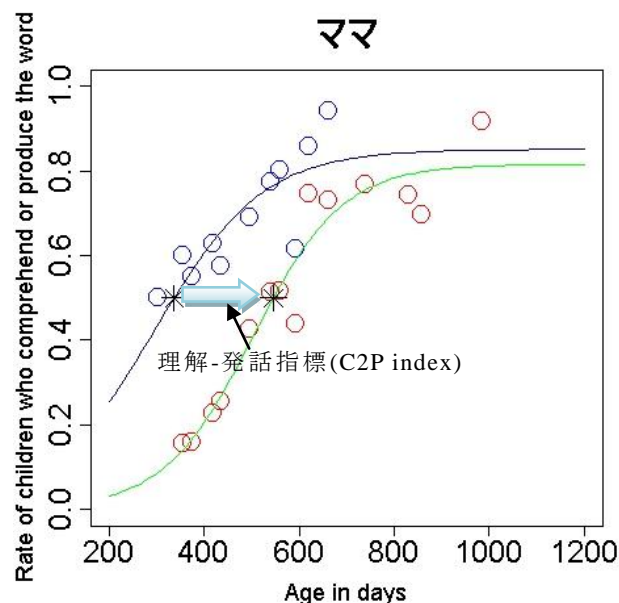


図6 理解-発話指標(C2P index)

5. 名詞優位性の検証

4. で示した理解-発話指標(C2P index)を CDI の全ての単語に対して求めた。ただし、ロジスティック関数が求められないものに対しては対象から除外した。語のカテゴリ分けは、Caselli らの基準[8]により決定した。そのうち、動詞と名詞に分けられる語のみを使った。図 7 に、動詞と名詞の理解-発話指標(C2P index)に対する分布を示す。この時の動詞の平均 C2P index が 249 であり、名詞の平均 C2P index が 189 であった。この差は 60 であり($8.8e-11 < 0.001$) 名詞学習の方が優位であることがわかる。この値は大まかに幼児が理解してから発話するまでの平均日齢の差(大まかと言ったのは求めているのは中央値であり、平均値ではないこと、仮定している分布の累積が 1 にならないことによる)を表している。

また、図 7 の形状が図 2 と近いこともわかる。図 2 の横軸は抽象的な空間であり、実際に幼児が語を学習する速度と直接的に結びついているかわからない。それに対して、図 7 では、理解から発話までかかる時間として明らかになっている。これらの結果から、著者らの手法を用いることにより、実際の難易の空間を用いても、図 7 で示したような連続体が存在することが確認できた。

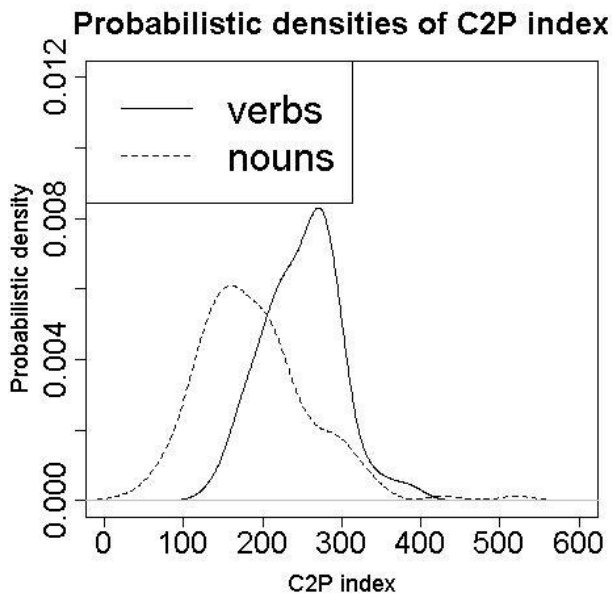


図 7 理解-発話指標(C2P index)に対する名詞と動詞の語分布

6. まとめ

本稿では、幼児早期出現語に関する難易を規定する抽象空間の存在を検証し、語学習の難易における名詞学習の優位性を検証した。この実際の難易度を表す指標として、ここでは、語学習のプロセスの部分的な指標である理解から発話までの日齢を示す認識-理解指標(C2P index)を提案した。この指標は、横

断的な日本語版 CDI による語彙チェックリストの結果から得られる語の理解と発話の習得率をロジスティック関数で近似し、そのロジスティック関数を用いて 50% 取得率の理解と発話の日齢を求め、それらの差分を計算したものである。この指標を品詞毎に分類した結果、名詞学習の優位性について確認した。また、この指標上の名詞と動詞の分布が連続的に重なっていることを示し、抽象空間上での単語連続体の存在の可能性も検証した。

参考文献

- [1] D. Gentner, "Cognitive and Linguistic Determinism: Object Reference and Relational Reference," Boston Child Language Meetings, 1988.
- [2] D. Gentner, L. Boroditsky, "Individuation, Relativity, and Early Word Learning," Language acquisition and conceptual development, vol. 3, pp. 215-256, 2001.
- [3] M. J. Maguire, K. Hirsh-Pasek, R. M. Golinkoff, "A Unified Theory Of word Learning: Putting Verb Acquisition in Context," Action Meets Word: How Children Learn Verbs, pp. 364-391, 2006.
- [4] T. Tardif, "Nouns Are Not Always Learned before Verbs: Evidence from Mandarin Speakers' Early Vocabularies," Developmental Psychology, vol. 32, no. 3, pp. 492, 1996.
- [5] M. Tomasello, ことばをつくる: 言語習得の認知言語学的アプローチ, 辻幸夫(訳), 野村益寛(訳), 出原健一(訳), 慶應義塾大学出版会, 2008.
- [6] 小椋たみ子, 綿巻徹, 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, 2004.
- [7] 綿巻徹, 小椋たみ子, 日本語マッカーサー乳幼児言語発達質問紙[語と文法の手引]. 京都国際社会福祉センター, 2004.
- [8] C. Caselli, P. Casadio, E. Bates, "A Comparison of the Transition from First Words to Grammar in English and Italian," Journal of Child Language, vol. 26, no. 1, pp. 69-111, 1999.