

時系列データから要素の出現順序の 関係構造を抽出する方法

白井英俊*, 杉浦正利†

A Method to Induce the Underlying Constraints on the Occurrences
of Events from Time-Sequential Observations

SIRAI Hidetosi and SUGIURA Masatoshi

Abstract

We have developed a method that induces the hierarchy among the occurrences of the events which are longitudinally observed. It is intended as the research tool with which we find possible partial ordering on the development of morphology by children. We show two findings by using this method to analyze the developmental order of Japanese and English-speaking children.

1 はじめに

要素の出現の順序に何らかの(時間的な)制約があるものと想定される場合がある。ここでは、時系列的に観察された要素の出現から、その制約を推測することが問題となる。たとえば、 a, b, c, d という四要素において、 $a < b$ および $c < d$ (ここで $<$ は時間的な順序関係を表す) という制約があるが、 a と c 、 a と d 、 b と c 、 b と d の間には時間的な順序制約はないとする。この場合、 $a < b < c < d$ 、 $a < c < b < d$ 、 $a < c < d < b$ 、 $c < a < b < d$ 、 $c < a < d < b$ 、 $c < d < a < b$ という出現が観測される可能性がある¹⁾。本稿は、このように要素の出現になんらかの時間的な制約があるという条件のもとで、その時系列

的な観測から、要素間の時間的な順序制約(半順序構造)を抽出する方法を提案する。そして、子どもの言語獲得過程における形態素の習得の順序性をこの方法の有効な応用例として取り上げる。

2 研究の背景

我々は以下のような手法によって、日本語を母語とする子どもの言語発達の指標作成に関する研究を行ってきた(Sirai, 2001; Otomo, 2004)：

- 子どもの発話資料を縦断的に採取する。ただし、単なる養育者の直前の発話の繰り返しや、歌詞のような丸暗記された句などをマークして、以降の分析資料から除外する。また、子どもの発話を形態素解析して、MLU²⁾ や語の出現頻度などの統計資料を作

* 中京大学情報理工学部, 名古屋大学大学院国際開発研究科 2004 年度国内客員研究員

† 名古屋大学大学院国際開発研究科

成する。

- ・注目すべき文法現象（関係節や複文構造の出現など）や形態論的な規則の出現とみなされる現象・語の出現（否定、補助動詞、助動詞の使用など）、および意味、運用論的な意識の現われとみなされる語の出現（指示詞、終助詞の使用など）を予め設定し、それぞれの現象ごとに特徴的なパターンや語の出現を時系列的に数える。
- ・そのような文法的、形態論的、意味論的な規則や用法の習得の基準として想定された P レベルに基づき、P1 と P4 の基準からそれぞれの用法の習得時期をそれぞれの子どものデータに基づいて判定する。ここで、P レベルとは、異なる文脈における自発的な発話が行われたとみなされる時に 1 ずつ増えるものとする。従って、初出は P1、初出と異なる文脈における出現は P2 と計算され、Otomo (2004) では P4 を習得の基準として仮定している。
- ・形態素や文法現象が段階的に発達すると仮定し、その段階を特徴づける形態素や文法現象を特定し、日本語を母語とする幼児の言語発達指標として提案する。
- ・いろいろな年齢層の子どものデータを横断的に採取し、提案した言語発達指標と照らし合わせることにより、提案した言語発達指標の妥当性を検証する。また、これらの検証が十分行われた後、言語発達指標により対象となる子どもの言語発達段階を評価する。

この研究においては、いろいろな子どもを縦断的に観察して得られた発話データから、形態素や文法規則の習得に段階性があるかどうかを認定する方法が本質的な問題となった。習得過程には個人差があるものの、多くの子どもの発話データを集めれば、その習得過程には何らかの一般的な段階性が認められると

考えられる。その段階の認定をどのように捉えたらよいか、という問題である。これに対し Sugiura (2001) では SDOS (Sugiura's Developmental Order Scale) という方法を提案し、それぞれの対象者から得られた習得データをもとに、概略、以下のような行列を作成することで、要素間の順序性を捉えようとした。

調査対象とする現象数（要素数）を m とした時、 $m \times m$ の行列を作り、 (i, j) 成分 ($1 \leq i, j \leq m$, i 行 j 列の要素) には以下の値を与える：項目 i が項目 j よりも習得 (P レベル到達) が遅い場合に 2, 同時期なら 1, そうでなければ 0. このような行列が調査対象者の人数分だけできるので、それらを足し算した行列を作り、それぞれの行および列ごとの和を計算し、その値に基づいて項目の順番を並び替える：行は和の昇順、列は和の降順。(pp. 207-208)

例として、調査対象者 3 名の縦断的発話データから得られた P4 レベルの基準による終助詞³⁾ の習得過程に対する SDOS 行列を表 1 に示す (Shirai, 2001). 例えば、「ネ」行「ヨ」列の成分が 5 とあるが、これは「ネ」が「ヨ」よりも習得が遅かった対象者が 2 名、同時期に習得した対象者が 1 名であったことを意味する。この表から、(1) ヨ、ネ、ノ、(2) ワ、ナ (ア)、カ、カナ、(3) ノ-ネ、ヨ-ネ、ノ-カナという三つの終助詞のグループが段階的に習得されているらしいという大まかな傾向を読み取ることはできる。しかし、「モン」を第二グループにいれてよいかどうかは判断に迷う。また、ワとカやカナの間に出現の時間順序があることは見落とされる可能性がある。このように、要素（この場合は形態素）の出

表1 終助詞の習得データによるSDOS行列

	ヨ	ネ	ノ	ワ	モン	ナ(ア)	カ	カナ	ノ-ネ	ヨ-ネ	ノ-カナ	計
ヨ	3	1	2	0	0	0	0	0	0	0	0	6
ネ	5	3	2	0	0	0	0	0	0	0	0	10
ノ	4	4	3	0	0	0	0	0	0	0	0	11
ワ	6	6	6	3	3	2	0	0	0	0	0	26
モン	6	6	6	3	3	2	2	2	1	1	0	32
ナ(ア)	6	6	6	4	4	3	2	2	0	0	0	33
カ	6	6	6	6	4	4	3	2	0	0	0	37
カナ	6	6	6	6	4	4	4	3	0	0	0	39
ノ-ネ	6	6	6	6	5	6	6	6	3	2	1	53
ヨ-ネ	6	6	6	6	5	6	6	6	4	3	3	57
ノ-カナ	6	6	6	6	6	6	6	6	5	3	3	59
計	60	56	55	40	34	33	29	27	13	9	7	

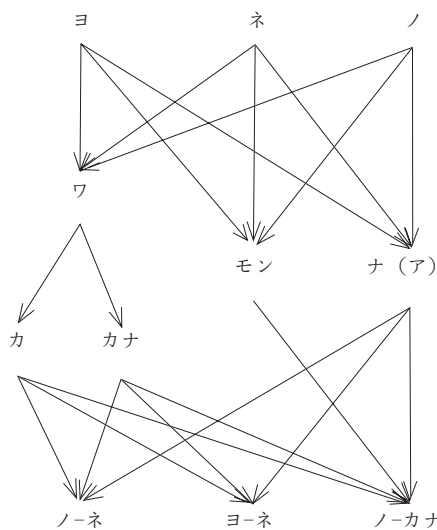


図1 調査対象者3名のデータから得られた終助詞の出現時期の半順序構造

現からどのような順序が想定できるかをこの行列からすべて導き出すことは難しい。

ここで、ある要素 (a) が別な要素 (b) よりも習得時期が早い(これを $a < b$ で表わす) 場合に a から b への有向枝で結ぶとする。この定義から $<$ は推移律が成り立つ。つまり、任意の要素 a, b, c に対し、 $a < b$ かつ $b < c$

が成り立っていれば $a < c$ が成り立つはずである。そしてそのような場合に、a から c への枝を除去することとする。表1からこのような手続きを経て得られた図が図1である。この図において「ヨ、ネ、ノ」の間には習得時期が早いとか遅いという時間的な関係は存在しないので、これらの間を結ぶ枝はない。つまり、要素間を結ぶ枝(の列)がないことにより、これらの終助詞の習得時期は重なることが示唆される。これは「ワ、モン、ナ(ア)」においても同様である。一方、「ヨ、ネ、ノ」それぞれを始点とし「ワ、モン、ナ(ア)」それぞれを終点とする方向枝が引かれている。これは「ワ、モン、ナ(ア)」という習得時期が重なる助詞が、「ヨ、ネ、ノ」という習得時期の重なる助詞の後に習得されることを示している。さらにこの図では「ワ」は「カ、カナ」に対して習得時期が早いことが示されているが、「モン」や「ナ(ア)」と「カ、カナ」の間にはそのような関係が成り立っていないことも示されている。言い替えれば、「ワ、カ、カナ」という終助詞のグループは「モン、ナ

(ア)」と習得時期が重なり、その中では「ワ」は「カ、カナ」よりも習得時期が先行する、ということが示されている。

これらの終助詞の習得は、一人一人の調査対象者では時間軸に沿って行われる全順序的な関係とみなせることもできるが、いろいろな調査対象者からのデータからこのような図を作成することで、ある終助詞のグループの間には習得順序がみられるが、別の終助詞の間には順序性がみられないという半順序関係を推測することができる。これは、表1から「読み取れる」よりも細かい情報を表わすものである。

このような考察を元に、我々は要素の出現の時系列データからまず SDOS 行列を求め、この SDOS 行列から要素間にどのような時間的な順序構造（半順序）が存在する可能性があるかを導出する方法を考案した。この具体的なアルゴリズムについては3節で述べる。また、5節では Brown (1973) の習得データを検証材料として、この方法の有効性を示す。

3 提案するアルゴリズム

本稿で提案する手法は、いくつかの要素の時系列データから、その要素の出現を生み出している要素の間の（半）順序構造の可能性を抽出するものである。ここで、本研究のそもその背景から、「要素」とは「幼児によって習得される形態素や文法規則」、「出現」とは「その習得時期」を指すものと考えて議論を進めるが、もちろんこれに限定されるものではない。

本アルゴリズムは二段階からなる：始めに SDOS 行列を求め、次に SDOS 行列から半順序構造を抽出する。

SDOS 行列は、2節で述べたように、以下の手続きにより求められる。

1. 調査対象の要素の個数を m とすると、 $m \times m$ の行列を作成し、 (i, j) 成分 ($1 \leq i, j \leq m$, i 行 j 列の要素) には以下の値を与える：項目 i が項目 j よりも習得 (Pレベル到達) が遅い場合に 2, 同時期なら 1, そうでなければ 0⁴⁾。
2. このような行列が観測データの個数 (我々の元々の研究の観点からいえば、調査対象者の人数分) だけできるので、それらの和を求める。すなわち、 i 番目の観測データから得られた行列を M_i とし、観測データ数を n とすると、 $\sum_{i=1}^n M_i$ を求める。
3. 得られた行列において、それぞれの行および列ごとの和を求め、その値に基づいて項目の順番を並び替える：行は和の昇順、列は和の降順。これにより表1のような行列が得られる。ただし複数の項目において和が一致する場合もある。その時は行の左から右に並ぶ項目の順序と列の上から下に並ぶ項目の順序が一致するように並べるものとする。

次に、このようにして得られた SDOS 行列から、以下の手続きにより半順序構造を抽出する。

1. SDOS 行列をグラフの隣接行列とみなして、要素を強連結成分に分解する。ただし、ある強連結成分を構成する要素と他の強連結成分を構成する要素の間の到達可能性が、それぞれの強連結成分に属する要素の間の順序構造となる⁵⁾。

表1に示す SDOS 行列を例にとれば、これにより、{ヨ、ネ、ノ} と、{ワ、モン、ナ (ア)、カ、カナ、ノ-ネ、ヨ-ネ、ノ-カナ} という二つの終助詞の集合に分けられ、前者のグループが上位 (早期に出現)、後者

表2 要素の出現制約に複数の可能性が考えられる例

	A	B	C	D	計
A	2	0	1	1	4
B	4	2	0	1	7
C	3	4	2	0	9
D	3	3	4	2	12
計	12	9	7	4	

のグループが下位（遅い出現）に位置することになる。

2. 同じ強連結成分を構成する要素の集合に対して以下を行い、同じ強連結成分に属す要素の間の半順序構造を求める。ここで、ある強連結成分に対応するSDOS行列の行（同じことだが列）が要素 a_1, \dots, a_n ($1 \leq i, j \leq n$) で表されるとする。

- (1) 要素 a_1, \dots, a_n に対し、有向枝 (a_i, \perp) を作る（枝の始点が a_i 、終点が \perp ）。

ここで、 \perp はどの要素 a_i に対しても遅く出現する仮想的な要素とする

- (2) それぞれの a_i に対し、以下を繰り返す： $1 \leq i < j \leq n$ なる i に対し、SDOS行列の (a_i, a_j) 成分が0であり、かつ a_i から有向枝によって到達可能な要素 a_k に対しても (a_i, a_k) 成分が0であれば、

- 有向枝 (a_i, a_j) を作る。
- (a_i, a_j) なる枝をもつすべての a_i に対し、枝 (a_i, a_j) があればそれを削除する

これは、時間的に順序関係がありそのような要素 a_i と a_j のペア ($a_i < a_j$ と仮定) に対し、他のデータとも時間的な矛盾がなければ (a_i, a_j) という枝をつくる、という作業である。

- (3) 自分を終点とする有向枝をもたない要素 a_i すべてに対し、枝 (\top, a_i) を

つくる。ここで、 \top とはどの要素 a_i にも先行して出現する仮想的な要素とする。

以上の手続きにより、与えられたデータから要素間の半順序構造が一つ得られる。前節で示した図1は、本稿で述べた手続きにより表1のSDOS行列から得られた構造である。もっともこのSDOS行列を強連結成分に分解したときに、{ヨ、ネ、ノ}は他の終助詞よりも上位の要素として分離されるため、図1は二つの構造を接合して書いてある。

ここで可能な半順序構造が複数個あった場合について述べておく。例えば、表2にあげるSDOS行列に対しては、 $\{A < B, C < D\}$ 、もしくは $\{B < C\}$ 、もしくは $\{A < B\}$ 、もしくは $\{C < D\}$ という関係が要素間にみられる、という分析が可能である。しかし、本手法では、SDOS行列において若い番号の要素に関する枝を優先した構造が選択される。すなわち、SDOS行列の要素の順番にしたがって $A < B$ という関係がまず得られ、次にこれと矛盾しない $C < D$ という関係が得られるため、最終的に、 $\{A < B, C < D\}$ という関係が抽出される。これは $A < B$ だけ、もしくは $C < D$ だけの関係を抽出するという分析を包含している。また、ここで $A < B$ という関係が得られた後に $C < D$ ではなく $B < C$ という関係の存在を仮定すると、推移律から $A < C$ とならなければならないはずだが、これは成り立たない (A 行 C 列の要素が0ではない) ため、棄却される。

4 本アルゴリズムの妥当性

本稿で提案したアルゴリズムにより、なぜ

半順序構造が得られるかについて、簡単に説明する。

まず、SDOS 行列の性質から、SDOS 行列の行（および列）の要素の並びは、その要素の出現の「平均的」な時間順になっている。もしも要素間に全順序が成り立っていたとすれば、つまり SDOS 行列の行を構成する要素を a_1, a_2, \dots, a_n とし、その出現に $a_1 < a_2 < \dots < a_n$ という関係があったとすれば、SDOS 行列は下三角行列になる⁶⁾。このことから、SDOS 行列の対角線よりも右上にある (a_i, a_j) ($i < j$) 成分に対して、その値が 0 であれば $a_i < a_j$ という関係が成り立つ ((a_j, a_i) という有向枝を作ってもよい)。また非 0 であれば a_i と a_j の「平均的な出現傾向」に対する反例が存在すること、即ち、この $a_i < a_j$ という関係を設定できない ((a_j, a_i) という有向枝を作ってはならない) ことが導かれる。このような考えに基づいて、時間的な順序関係が成り立つ要素間に枝を作っていけば、出現の順序を表した構造が得られることになる。なお、どの要素に対しても先行する仮想的な要素 \uparrow と後行する要素 \downarrow を設けることにより、任意の二つの要素に対する最大下界、最小上界の存在が保証される。

また、出現順序という時間的な順序関係は推移律が成り立つので、 $a_i < a_j$ かつ $a_j < a_k$ ならば、 $a_i < a_k$ が言える。このことから、 (a_i, a_j) かつ (a_j, a_k) という枝がある場合には (a_i, a_k) を削除することにより、全体の構造を分かりやすく提示することができる。

5 本アルゴリズムの適用例

本節では、英語を母語とする幼児から縦断的に得られた形態素の習得の順序データにこ

こで提案した手法を適用し、その有効性を検証したい。

Brown (1973) は Adam, Sarah, Eve という英語を母語とする 3 名の幼児の自発的な発話を縦断的に収集し、3 節で述べた習得基準によって、表 3 に示す習得過程が得られたとしている。なお、この表において、ローマ数字は Brown の習得段階を表すものである。

我々は、この表をもとに、14 個の形態素に対して表 5 (付録 B) にあげる習得年齢を想定した。これは実際の年齢を必ずしも反映していないが、形態素の習得の順序は保たれている⁷⁾。

この時系列データに対し、3 節で紹介したアルゴリズムを適用して得られた英語の形態素・文法規則の習得の順序構造を図示したものが図 2 である。この図から習得の順序性を議論するのは本稿の主旨ではないが、少なくとも Possessive と Articles, Past regular と Past irregular, Third person regular と Third person irregular という文法的に関連がある形態素の組み合わせにおいて、それぞれの習得に順序関係はなく、時期が重なって習得される⁸⁾ ことが示されていることは興味深い。

なお、de Villiers and de Villiers (1972) は、英語を母語とする幼児 21 名に対し Brown と同じ 14 個の形態素の使用を横断的に収集している。その実験で得られた幼児の平均の正解率を表 4 に示す。そして、彼らはその結果が Brown による縦断的なデータとスピアマンの順位相関係数の観点から相関が高い⁹⁾ と結論している。

我々の手法によって得られた図 2 の形態素間の習得の順序構造を、de Villiers らの結果と比べてみると次のようなことが分かる：枝の数でいえば、一致が 16 本あり、不一致は、

表3 Brownによる3人の幼児の14個の形態素の習得順序 (Brown, 1973: 275)

	Adam		Sarah		Eve
I (2;3)	—	I (2;3)	—	I (1;6)	—
II (2;6)	—	II (2;10)	—	II (1;9)	—
	Present progressive in on, plural		plural in, on		Present progressive, on
III (2;11)	—	III (3;1)	—	III (1;11)	—
	Uncontractible copula, past irregular		Present progressive, past irregular Possessive Uncontractible copula Articles		Present progressive, on in Plural, possessive Past regular
IV (3;2)	—	IV (3;8)	—	IV (2;2)	—
	Articles Third person irregular, possessive		Third person regular		
V (3;6)	—	V (4;0)	—	IV (2;3)	—
	Third person regular Past regular Uncontractible auxiliary Contractible copula Contractible auxiliary		Past regular Uncontractible auxiliary Contractible copula Third person irregular Contractible auxiliary		Uncontractible copula Past irregular Articles Third person regular Third person irregular Contractible copula Contractible auxiliary

注：年齢を“年；月”で表す

Possessive と Past regular の間, Articles と Third person irregular の間, Uncontractible auxiliary と Contractible copula の間の3本である。

Brown および de Villiers らは一本の時間軸上にすべての形態素の習得を並べようとし、彼らの結果の順位相関係数をもって相関が高いことを主張している。それに対し我々の分析は、形態素の習得を半順序関係と考え、結果として Brown および de Villiers らの結論と

相関が高く、形態素間の習得についての詳細な分析を与えているとみることができよう。

なお、de Villiers らによる平均の正解率を時系列データと読み替え、Brown の習得データと合わせて、階層構造を作り出すことも可能である (付録 C)。

6 議論と考察

本稿は、要素の出現に時間的な順序制約が

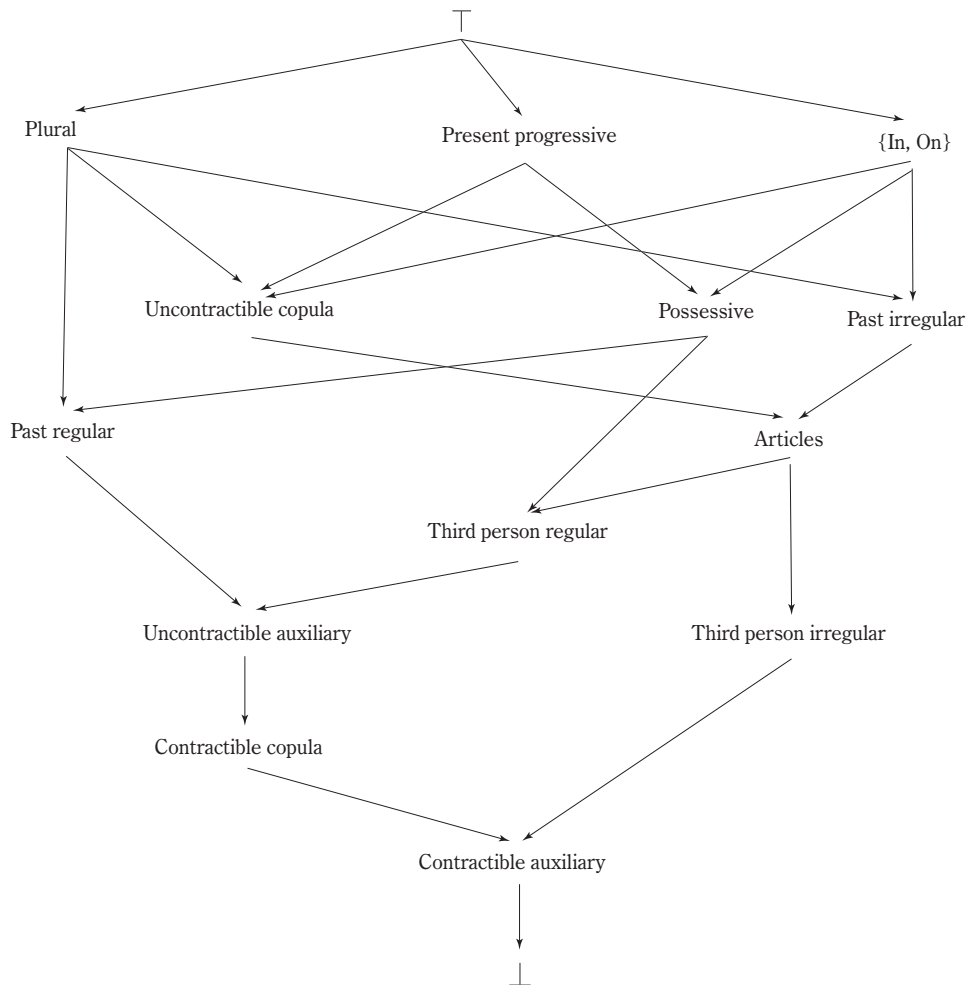


図2 Brown のデータから得られた 14 個の形態素の習得順序の構造

あると仮定できる時に、その出現の時系列的な観察データから、要素の出現の順序関係(半順序構造)を推定する手法を提案した。また、時系列データから順序関係を求めるウェブページも作成した(白井, 2005)。

本節では、これと関連する技法との比較を行い、次にこの手法を「形態素の習得順序に適用する」ことに対する問題について議論し、締めくくりとする。

6.1 関連技法との比較

システム工学における階層構造化手法の一つに ISM (Interpretive Structural Modeling) (Worfield, 1974; 木下, 1996) や DSM (Design Structure Matrix) (Steward, 1981) がある。これらの手法では、(a) 一般にブレインストーミングやインタビューにより取り組むべき問題の関連要素を抽出し、(b) この要素のペア比較を行って、要素 i が要素 j に影響を与えていれば 1, そうでなければ 0 として関係行列 (E) を作り、(c) これに単位行列 I を加えた

表4 de Villiersらによる21人の幼児による横断的な14個の形態素の習得状況. de Villiers & de Villiers(1972: 270-271)の表から作成

形態素	Brownのデータによる順位	順位	平均の正解率
Present progressive <i>-ing</i>	1	4	87.5
Postposition <i>in</i>	2.5	1	98.9
Postposition <i>on</i>	2.5	2	95.7
Plural <i>-s</i>	4	3	87.6
Uncontractible copula	5	10	51.2
Past irregular	6	5	70.0
Articles <i>a, the</i>	7	8	60.2
Third person regular <i>-s</i>	8	12	44.7
Possessive <i>-s</i>	9	11	50.0
Third person irregular	10	6	67.4
Past regular <i>-ed</i>	11	7	60.8
Uncontractible auxiliary	12	14	0
Contractible copula	13	9	55.7
Contractible auxiliary	14	13	35.0

行列 ($E + I$) を接続行列とみなしてグラフの到達可能性から、問題の要素の階層構造を求める。言葉の獲得問題にこれらの手法を素直に適用するとすれば、(a) 言葉の獲得の関連要素 (文法現象や形態素) を選びだし、(b) 要素 i が要素 j よりも先行して習得されていることを「要素 i の習得が要素 j の習得に影響を与えている」とみなして 1, 0 要素をもつ関係行列を作り、(c) 要素間の階層構造を求める、となり、一見極めて近い手法のように見える。

しかし、次のような違いが存在する：

- ISM や DSM では要素間の依存関係をもとに行列を作る。我々の手法は、依存関係が時系列データに現れたものと考え、時系列データから行列を作る。つまり、我々の手法は依存関係は与えられているものではなく、時系列データに隠れた依存関係を推測する手段を与えるものである。
- 構造分析方法が異なる。ISM や DSM にお

いて要素をグラフの節点 (ノード)、要素 i が要素 j に影響を与えることを点 i から点 j への弧 (枝) があることと読み替えると、ISM や DSM の階層構造の求め方はグラフの強連結成分に分解することと同等である。我々の方法では、強連結成分に分解した後に、その成分を構成する要素間の半順序構造を求める。

したがって、ISM や DSM の手法と本稿で提案する我々の手法は本質的に異なるものと考ええる¹⁰⁾。

6.2 「形態素の習得順序に適用する」ことに対する問題

何をもって「習得」したか、ということは、かなり難しい問題をはらんでいる。5節で述べた Brown (1973) の研究では、「形態素の習得」の判定基準とは「その形態素が義務的環境において90%以上の正解率で用いられるセッション (観察) が連続して3回」と

いうものであった。もっとも Brown は英語を母語とする幼児の言語獲得を対象としており、日本語を母語とする幼児の言語獲得の場合にはかなり事情が異なる。この研究のもととなった Sirai (2001), Otomo (2004) では要素の一定回数以上 (P4 レベル) の出現をもって要素の習得と仮定したが、確たる根拠をもつものではない。

大きく分けてここには次の問題があると考えられる：

1. P レベルは習得の判断基準として妥当か
2. 習得のレベルとして P4 以外の値を用いたときに本アルゴリズムの適用結果がどのように影響されるか
3. データを採取する対象者によって結果が変動する可能性があり、普遍的な習得順序をこの手法を用いて求めることはできないのではないか
4. この手法は要素の出現 (習得) 順序だけを問題としているが、要素の出現の時間間隔も考慮すべきではないか

1 に対しては、縦断的な観察から形態素の習得の P レベルを測定したデータが少ないため、我々はこれに答えることができない。しかし、我々のこれまでの研究観察から、よく用いられる形態素であれば、かなり習得が進んでいることが P4 レベルで表わされると考えている (Sirai, 2001; Otomo, 2004)。

2 は 1 と関連するが、使用頻度が少ない形態素 (Sirai (2001) の例では「のに」) の場合には P3 から P4 レベルに達するのに個人差が大きいなど、影響が大きいことが分かっている。より大きなデータに対して適用しながら P レベルをどう設定したら経験的な判断に合うものになるかということ調べるべきであ

り、これは今後の課題である。

3 に対しては、上記二つの問題とも関連し、より多くの対象者からデータを採取して、個人差によらない傾向を考える必要がある、というのがあるべき研究の方向であると考えている。ただデータ数が大きくなる場合に生じる問題は、SDOS 行列において「0」の部分が消えてしまう、つまりある要素間で明確に存在する出現に関する時間的な順序関係がみえなくなる、ということである。その場合には、ある閾値を設けてその閾値以下の値は「0」とみなす、という操作が必要になるかもしれない。

4 に対しては、要素間の距離 (時間差) を含めた構造を求める手法への拡張を考慮することができることを示す。距離の考え方として、要素の出現の時間差をそのまま採用する方法と、相対的な時間差を採用する方法とがある。前者は Sirai (2001) が用いた尺度であり、後者は Sugiura (2001) が用いた尺度である¹¹⁾。いずれにせよ、本稿で提案したアルゴリズムにより、要素間の階層構造が得られているので、要素間の時間差をその構造における要素間の距離とすればよい。それには、観測されたそれぞれの時系列において、要素間の時間間隔を求め、その平均値を求めれば、絶対的な時間差を尺度とした構造が得られることになる。相対的な時間を尺度とする場合も同様に計算される。このように、時間尺度を組み入れた階層構造も容易に得られる。

謝辞：Brown のデータを調べてくれた森田光宏氏に感謝する。

注

- 1) 付録 A に、このすべての可能性が現れた時、本稿で提案する手法により上記の制約が求められることを示す。もっとも、 $a < b = c < d$ のように、「同時」出現も含めると、もっと多くの可能性がある。
- 2) MLU (Mean Length of Utterance) とは平均発話長、つまり発話あたりの形態素の個数のことである。MLU と子どもの文法発達に相関があることを Brown (1973) が示して以来、MLU は子どものことばの発達を測る重要な尺度として用いられている。
- 3) Shirai (2001) では、個々の終助詞だけではなく「ヨ・ネ」のような終助詞の連鎖も調査対象としていた。
- 4) この手続きは Sugiura (2001) に基づいている。項目 j の出現が項目 i よりも時間的に先行することを項目 i の出現が項目 j の出現に依存しているとみなし、項目 i の出現が項目 j の出現より遅い場合に限り 1、それ以外は 0 を割り当てる（ただし、どの項目 i に対しても自分自身に依存するものとみなす）、という考え方もある。
- 5) 実際には強連結成分に分解するという手続きは不要で、二番目の手続きだけでも順序構造の抽出は可能である。しかし要素集合を大分類するには、この手続きが役立つ。
- 6) 下三角行列とは、行列の対角線よりも右上の要素がすべて 0 の行列のことである。
- 7) Brown 自身は詳細な年齢を与えていない。CHILDES (MacWhinney & Snow, 1985; MacWhinney, 2000) に Brown の 3 名の発話データが登録されているが、そのデータから「習得」されたとする実際の年齢を求めることはできなかった。
- 8) 図 2 の上下は時間軸を必ずしも表していないことに注意されたい。例えば Possessive の下に Articles があるが、Possessive から Articles への枝がないため、これらの習得時期には順序性がない、つまり同じような時期に対象者は前後して習得したことを示している。
- 9) 本稿では de Villiers の手法 2 (de Villiers & de Villiers, 1972: 271) の値を用いている。Brown のデータとのスピアマンの順位相関係数は + 0.78 である。
- 10) ISM や DSM の手法と本アルゴリズムは発想的に異なるものの、本アルゴリズムによってこ

れらの手法と同様に問題を構成する要素の間の階層構造を求めることもできる。

- 11) 習得には個人差があることから Sugiura (2001) は相対的な時間差を尺度として採用することを提案している。ここで相対的な時間差とは、観察対象の幼児それぞれにおいて観察された要素の最も早い出現と最も遅い出現との時間差を 1 としたときの、要素の出現の比で表される。

参考文献

- Brown, R. (1973). *A first language: The early stages*. Harvard University Press, Cambridge, Massachusetts.
- de Villiers, J. G. & de Villiers, P. A. (1972). A Cross-Sectional Study of the Acquisition of Grammatical Morphemes in Child Speech. *Journal of Psycholinguistic Research*, 2 (3), 267-278.
- 木下栄蔵 (1996). 『マネジメントサイエンス入門：経営・政策科学の戦略モデル』。近代科学社、東京。
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk, third edition*. Lawrence Erlbaum Associates.
- MacWhinney, B. & Snow, C. E. (1985). The Child Language Data Exchange System. *Journal of Child Language*, 12, 271-295.
- Otomo, K. (Ed.) (2004). *Comparative Research for a Developmental Index for First and Second Language of Japanese and English*. Tokyo Gakugei University. Report of the Grant-in-Aid for Scientific Research (B) (1) (2001-2003).
- Shirai, H. (2001). Acquisition of Sentence Final Particles and the Developmental Index for Japanese. In Shirai, H. (Ed.), *A Crosslinguistic Study for the Universal Developmental Index*, pp. 159-172. Chukyo University. Report of the Grant-in-Aid for Scientific Research (A) (2) (1999-2000).
- 白井英俊 (2005). 「時系列データから要素の習得における順序関係を表示するシステム」. manuscript.
- Sirai, H. (Ed.) (2001). *A Crosslinguistic Study for the Universal Developmental Index*. Chukyo University. Report of the Grant-in-Aid for Scientific Research (A) (2) (1999-2000).
- Steward, D. V. (1981). *The Design Structure*

System: A Method for Managing the Design of Complex Systems. *IEEE Transactions on Engineering Management*, 28, 71-74.

Sugiura, M. (2001). A Developmental Order Scaling. In Sirai, H. (Ed.), *A Crosslinguistic Study for the Universal Developmental Index*, pp. 207 -

218. Chukyo University. Report of the Grant-in-Aid for Scientific Research (A) (2) (1999-2000).

Warfield, J. N. (1974). Developing Interconnection Matrices in Structural Modeling. *IEEE Trans. on Systems, Man and Cybernetics*, 4 (1), 81-87.

A 付録 1

a, b, c, d という四要素に対し, $a < b < c < d, a < c < b < d, a < c < d < b, c < a < b < d, c < a < d < b, c < d < a < b$ という出現の観測が得られたとした時の本手法による分析結果を図 3 に示す.

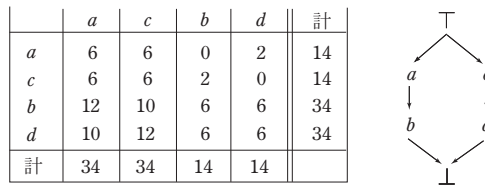


図 3 1 節の例題に対する本手法の適用

B 付録 2

Brown による 3 人の幼児の 14 個の形態素の習得の推定年齢を表 5 に示す.

表 5 Brown による 3 人の幼児の 14 個の形態素の習得の推定年齢

	Adam	Sarah	Eve
Present progressive	2 ; 6.0	3 ; 0.0	1 ; 10.0
In	2 ; 7.0	2 ; 11.0	1 ; 11.0
On	2 ; 8.0	2 ; 11.0	1 ; 10.0
Plural	2 ; 8.0	2 ; 10.0	2 ; 0.0
Uncontractible copula	2 ; 11.0	3 ; 1.0	2 ; 3.0
Past irregular	2 ; 11.0	3 ; 0.0	2 ; 4.0
Articles	3 ; 3.0	3 ; 2.0	2 ; 5.0
Third person irregular	3 ; 4.0	4 ; 3.0	2 ; 7.0
Possessive	3 ; 4.0	3 ; 0.15	2 ; 0.0
Third person regular	3 ; 6.0	3 ; 10.0	2 ; 6.0
Past regular	3 ; 7.0	4 ; 0.0	2 ; 1.0
Uncontractible auxiliary	3 ; 8.0	4 ; 1.0	2 ; 8.0
Contractible copula	3 ; 9.0	4 ; 2.0	2 ; 9.0
Contractible auxiliary	3 ; 10.0	4 ; 4.0	2 ; 10.0

注：年齢を“年；月”で表す

C 付録 3

de Villiers らによる平均の正解率を時系列データと読み替え、Brown の習得データと合わせて、作り出した階層構造を図4に示す。この構造においても、Brown の習得データでみられたように、Possessive と Articles, Past regular と Past irregular, Third person regular と Third person irregular という文法的に関連がある形態素の組み合わせにおいて、それぞれの習得に順序関係はなく、時期が重なって習得されることが示されている。

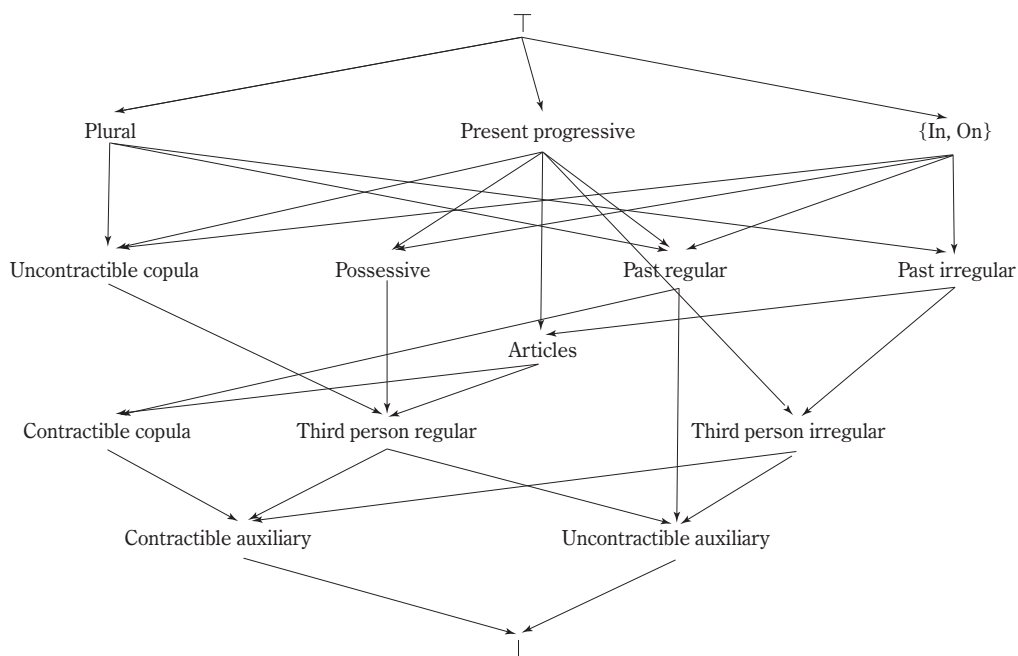


図4 Brown と de Villiers らのデータによる 14 個の形態素の習得順序の構造