

Title	言語動力学におけるクレオール創発
Author(s)	中村, 誠; 橋本, 敬; 東条, 敏
Citation	認知科学, 11(3): 282-298
Issue Date	2004-09
Type	Journal Article
Text version	author
URL	<a href="http://hdl.handle.net/10119/7943">http://hdl.handle.net/10119/7943</a>
Rights	Copyright (C) 2004 日本認知科学会. 中村誠, 橋本敬, 東条敏, 認知科学, 11(3), 2004, 282-298.
Description	

## 言語動力学におけるクレオールの創発

中村 誠, 橋本 敬, 東条 敏

Creole is one of the main topics in the fields concerning the language change and evolution, such as sociolinguistics, the developmental psychology of language and so on. Our purpose in this paper is to develop an evolutionary theory of language to study the emergence of creole. We discuss how the emergence of creole is dealt with regard to population dynamics. We modify the language dynamics equations by Komarova et al., so as to include the generation parameter ' $t$ '. From the viewpoint of the population dynamics, we give the definition of creole as a language, which is predefined by the universal grammar together with pre-existing languages. We show experimental results, in which we could observe the emergence of creole. Furthermore, we analyze the condition of creolization in terms of similarity among languages.

Keywords: Creole (クレオール), Population Dynamics (人口動力学), Language Dynamics Equations (言語動力学)

### 1. はじめに

社会言語学や発達言語心理学などの分野において、ピジンやクレオールといわれる言語現象について盛んに研究が行われている (Arends, Muysken, & Smith, 1994; DeGraff, 1999). ピジンとは共通の言語を持っていないが、通商その他の目的で互いにコミュニケーションをとる必要がある人々の間に発達した伝達システムである。その語彙や文法は単純化されており、構造や言語使用の点ではるかに発達している土地の言語と並んで、補助言語として習得される。その後ピジンが発展し、ある共同体の母語となったものはクレオールと呼ばれ、元となったどの言語とも文法的に異なり、それ自身が文法的にも表現能力としても充実した土地の言語となる。

19世紀から20世紀初頭にかけて、ハワイのプランテーションで働く労働者のコミュニティで発生し

たハワイ・クレオールを具体例として挙げることができる。中国、フィリピン、日本、朝鮮、ポルトガル、プエルトリコなど世界中の国から集められた労働者は、共通の言語を持たない状況で英語話者であるオーナーと意志疎通を図る必要があった。その結果、英語を基盤とするハワイ・ピジンを話すようになった。その後、ピジン話者である労働者の間に子供が生まれると、その子供たちが言語獲得期において接する言語のほとんどがピジンであった。しかし、子供たちが身につけた第一言語は両親の母語でもハワイ・ピジンでもなく、独自の文法構造を持った新たな言語であった。この言語がハワイ・クレオールと呼ばれている (Bickerton, 1990)。

ピジンやクレオールは世界中で発見されており、それぞれが独自に発達した言語体系であるにも関わらず、非常に似通った特徴と文法構造を持っていることが大きな特徴である (亀井・河野・千野, 1996; 風間・長谷川, 1992)。これらを含む言語変化は人間の言語獲得に深く関わり、言語学習者の社会環境の急激な変化に対応するため、獲得する言語が変化した実例であると考えられる。ピジン話者のコミュニティからわずか一世代でクレオールが発生した例も報告されており、周りに獲得すべき言語がないときに、人間がもともと持っている生得的な言語が発

Emergence of Creole in Language Dynamics by Makoto Nakamura (School of Information Science, Japan Advanced Institute of Science and Technology), Takashi Hashimoto (School of Knowledge Science, Japan Advanced Institute of Science and Technology) and Satoshi Tojo (School of Information Science, Japan Advanced Institute of Science and Technology).

現すると Bickerton (1990) は主張している。このように、ピジンやクレオールを研究することは、人間の生得的な言語獲得のメカニズムの解明に直結するため、言語に関連するさまざまな分野から大きな関心を集めているのである。

本研究の目的は、クレオールが創発するための言語獲得能力と社会、言語環境についての条件を数理科学的に導出し、言語変化の過程を通時的な側面から一般化された形で定式化することである。これまで言語学的な側面からピジンやクレオールは定義されてきた。しかし、その元となった言語間の関係からもしくは各言語話者の人口構成比から、クレオールが創発する条件が存在することは想像に難くないが、それを実際のクレオールから厳密に求めることは現実的に不可能であった。近年ではこれらピジンやクレオールを含む言語の変化や言語進化に関する問題を解決する試みとして、言語の構造や人間の学習機構を直接解析するようなアプローチではなく、言語学習者を取り巻く環境を含めたシステムを構築することで理解しようとする構成論的手法が注目されている (Cangelosi & Parisi, 2001)。すなわち、言語学習に関するある特定の項目に関して仮説を立て計算機に実装する場合、それによって言語を学習する一語話者だけでなく、複数の個体と発話環境を含めたシステムを実装することにより、そこから発現するさまざまな事象を観察する。それを実際の言語現象と比較検証することによってその仮説が健全であることを主張する方法である。これにより、これまでのクレオールの実地調査からは得られなかった、クレオールが創発するための人口構成比、言語間の類似性、周りの環境に関する条件を計算機上のシミュレーションによって導き出すことが可能となる。

言語の変化を扱った研究は近年盛んに報告されており、その多くは自律的で能動的なエージェントが互いに文を発話し、その発話を認識することによって文法を学習するというものである (Hashimoto, 2001; Ono, Tojo, & Sato, 1996)。マルチエージェントを用いたモデルには、中村・東条 (2003) によってピジン化の過程をモデル化したものや、Briscoe (2002) によってクレオールの創発を観察したものも含まれる。しかし有限の数からなるエージェントによるシミュレーションからは、現実世界で発生する現象を一般化して結論づけることは困難である

と考えられる。それに対し、Komarova, Niyogi, and Nowak (2001) は普遍文法を用いた数理生態学的な言語動力学 (Language Dynamics Equations) を提案している。これは、個々のエージェントの能力に注目するのではなく、コミュニティ全体の言語話者人口の遷移を人口動力学 (Population Dynamics) として扱っており、上記に挙げたマルチエージェントモデルとは大きく異なるアプローチである。

本研究では、実際にクレオールが創発する環境に倣い、複数の言語が使用される特殊なコミュニティを仮定し、そこで子供が獲得する言語と各言語話者の人口変化の関係を調査する。これらの相互作用によって言語獲得に与える影響が言語変化の本質であり、これを促す環境を計算機上に実装することは、ピジン、クレオール研究に対して大きな貢献になると考えられる。その際、Komarova et al. (2001) の言語動力学は有用であると考えられるが、現実の状況と比較して非常に単純化されている。本稿における我々の目的は、この言語動力学をより現実なものに修正することにより、クレオールが創発する過程を示すことである。また、そのモデルからクレオールが創発するための言語間の類似性に関する条件を導き出すことである。

以降、第 2 節において言語動力学について説明し、その後我々の修正点について述べる。第 3 節では修正した動力学に基づいた具体的なモデルを示す。このモデルの健全性を示す実験およびその結果を第 4 節で述べ、次にクレオールの創発条件を求める実験とその結果を第 5 節に示す。最後に第 6 節で結論を述べる。

## 2. 文法獲得に関する人口動力学

本節ではまず最初に、文法の時間的な変化が人口動力学としてどのように表されるか述べる。次に Nowak and Komarova (2001), Komarova et al. (2001) によって提案された言語の動力学について説明し、その問題点を議論する。そして動力学モデルをより現実的なものにするために、これらのモデルの改善点を考える。

### 2.1 一般的な文法獲得モデル

子供の第一言語獲得に関するモデルを考える場合、言語学習者として子供を、その学習対象として

大人の言語をそれぞれ定義する。子供は学習期間において大人の言語に接することで第一言語の文法を獲得する。その後、ある世代における子供が次世代の大人になり、繰り返し学習をすることで系全体の言語の振る舞いや、大人が獲得した文法を観察するという手法が一般的である (Briscoe, 2002; Kirby, 2002)<sup>1)</sup>。

第一言語を獲得する際、その学習期間は限られており (Lenneberg, 1967)、学習者はそのコミュニティで話されている発話を聞き、そこからもっともらしい文法を身につけなければならない。このとき、有限の文から文法ルールを完全に特定することは不可能であることが数学的に示されているにも関わらず (Gold, 1967)、それでも同じコミュニティで育った子供は潜在的な文法のルールを正しく推論し、同じ言語を矛盾なく獲得するのである。文法獲得におけるこの困難性 (プラトン問題 (Chomsky, 1975)) を解決する考え方として、原理とパラメータ理論が提唱されている (Chomsky, 1981)。原理とパラメータ理論とは、普遍文法はすべての人間言語に共通な原則の体系、すなわち原理 (principle) とそれに付随するパラメータ (parameter) からなると仮定し、子供の文法獲得は普遍文法の原理に組み込まれたパラメータの値を言語経験により固定する過程と捉える考え方である (井上・原田・阿部, 1999)。

人口動力学モデルはこの原理とパラメータの考え方に基づいている (Komarova & Nowak, 2001; Niyogi & Berwick, 1995)。すなわち、その原理によって与えられた文法の探索空間は有限であると仮定し、言語話者が用いる文法は  $\{G_1, \dots, G_n\}$  としてあらかじめ定義される。すると言語の変化とは、言語話者が所有するパラメータの変化を示しており、その変化はそれぞれのパラメータ値に対応する言語間の人口遷移によって表現される。ここで、あるコミュニティにおいて文法  $G_i$  を持つ言語話者の人口比率を  $x_i$ 、すなわち  $\sum_{i=1}^n x_i = 1$  であるとすると、ある世代  $t$  における文法  $G_i$  の人口比率  $x_i$  の変化は動力学系として表される。このように人口動力学モデルは、各言語話者の人口比率の変化を追跡するものである。

以降、文法  $G_i$  から導出される言語を  $L(G_i)$  と

し、文法  $G_i$  を所有して言語  $L(G_i)$  を発話する大人を  $G_i$  話者と呼ぶことにする。

## 2.2 Komarova のモデル

Nowak と Komarova は文法獲得に関する人口動力学の数理的な理論を発展させることを目的として、進化ゲームの動力学 (Weibull, 1995) を応用したモデルを提案している (Nowak & Komarova, 2001; Komarova et al., 2001)。これを言語の動力学 (Language Dynamics Equations) と呼ぶ。彼らは子供の第一言語獲得が系全体で話される言語にどのように影響を与えるかについてモデル化を行った。このモデルでは大人を言語話者、子供を言語学習者として定義している。子供は親からのみ言葉を聞き、そこから文法を推定する。

Komarova らが提案した言語の動力学方程式は、言語が持つ性質を次のように与えている：

- 類似性：言語間の類似度を表す。任意の2つの言語間の類似性を示す確率行列  $S = \{s_{ij}\}$  で表される。この類似性は、言語の文法が似ている度合いを表すのではなく、ある言語の文が、他の言語の話者にどの程度理解されるかを表す確率である。すなわち、要素  $s_{ij}$  は、 $G_i$  話者がランダムに文を発話したとき、文法  $G_j$  を持つ聞き手に理解される確率として求められる。

- 遷移性：言語間の遷移度を表す。 $G_i$  話者である親の子供が文法  $G_j$  を獲得する確率は行列  $Q = \{q_{ij}\}$  で表される。

言語話者である大人は子供を産み、その数は各言語に与えられる適応度に比例する。ここで  $G_i$  の適応度は  $f_i(t) = \sum_{j=1}^n (s_{ij} + s_{ji})x_j(t)/2$  である。すなわち、より言葉を理解しやすかつ、理解された者が、より多くの子孫を残せると仮定している。これは  $G_i$  話者の発話が、コミュニティで理解される確率を表している。これらを用いて言語の動力学方程式は次のような微分方程式で与えられる：

$$\frac{dx_j(t)}{dt} = \sum_{i=1}^n q_{ij} f_i(t) x_i(t) - \phi(t) x_j(t) \quad (j = 1, \dots, n). \quad (1)$$

ここで  $\phi(t) = \sum_{i=1}^n f_i(t) x_i(t)$  であり、系の平均適応度を表している。これにより、 $-\phi(t)x_j(t)$  は、世

1) Kirby and Hurford (2001) は特にこの学習法を繰り返し学習モデル (Iterated Learning Model; ILM) と呼んでいる。

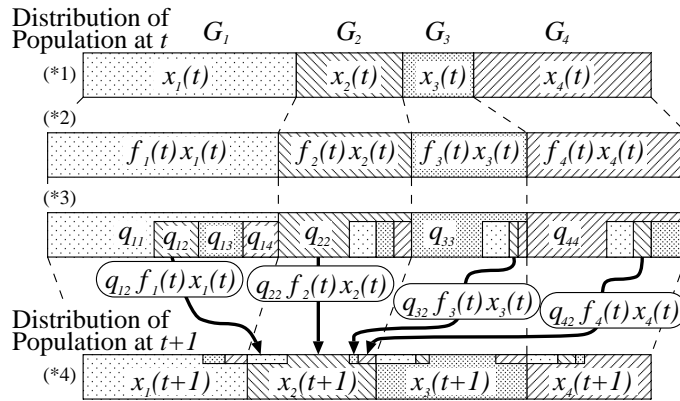


図1 人口変化の流れ  
 Fig. 1 Flow of population change

代<sup>2)</sup>を通じて総人口を一定に保つ働きをしている。

この式は次のような状況を描写していると解釈される(図1参照):

(\*1) あるコミュニティにおいて  $\{G_1, \dots, G_n\}$  の言語が話されている。各個体はその中のひとつを話すことができる。  $G_i$  話者人口の割合を  $x_i$  とすると、この図は各言語話者の人口構成比を表す。

(\*2) 言語話者は各言語の適応度に比例して子供を産む。すなわちコミュニケーション能力が次世代に子孫を残すための条件であることを表している。  $G_i$  話者の人口比率  $x_i$  に対して、産まれる子供の総数の比率は  $f_i x_i$  となる。この比率は大人の全人口に対するものであり、一般には  $\sum_i f_i x_i \neq 1$  である。

(\*3) 産まれた子供たちは親からのみことばを受け取り、文法を推定する。  $G_j$  話者の子供は  $q_{jj}$  の確率で正常に  $G_j$  を獲得し、その人口比率は  $q_{jj} f_j(t) x_j(t)$  である。その他の子供は  $q_{ji}$  の確率で  $G_i$  を身につける。このとき  $q_{jj}$  の値は学習アルゴリズムに依存するため、言語学習の精度(accuracy)と呼ばれる。

(\*4) 次世代の  $G_j$  話者の人口比率  $x_j(t+1)$  は親の言語を正しく継承したものと、他の言語話者からの流入の総和となる。すなわち  $t+1$  世代の  $G_j$  話者の人口比は  $\sum_{i=1}^n q_{ij} f_i(t) x_i(t)$  である。ここで  $\phi$  により総人口を一定に保つように調整する。

2) 通常、微分方程式で表される連続時間の人口動力学モデルでは、世代という考え方はなく、繁殖は同時に行われるものではないと仮定される。本稿においてはモデルを理解しやすいように  $t$  を世代と解釈し、モデルを説明していることを補足しておく。

上の解釈から、言語話者である親は多言語コミュニティの中で他の言語話者と会話をした結果、子供を産むのに対し、子供は親以外からことばを聞くことはない。しかし、このような状況において、子供が親の持つ文法を獲得するのに失敗し、他の文法を身につけるということは現実的に考えにくい。この文法獲得に失敗する可能性について、Niyogi (1996) が提案したモデルを用いて次節で論じることとする。

### 2.3 Niyogi のモデル

Niyogi (1996) は、言語学的な根拠に基づいた文法を基に、言語学習者がトリガー学習アルゴリズム (Trigger Learning Algorithm; TLA) (Gibson & Wexler, 1994)を用いて文法獲得を行うモデルを提案した。彼は言語学習者が文法を獲得するまでのパラメータセッティングの様子を、マルコフ過程図で示した。このモデルでも同様に、子供は親からしか文を受け取らない。またマルコフ過程図の各状態は、文法の獲得過程におけるパラメータ値を表しており、子供は学習を通じて各状態を遷移し、ある一定量の文を受け取った結果、その状態である文法を獲得したとする。このモデルから、2.2 節で我々が指摘した子供が親以外の文法を間違えて獲得する可能性、すなわち  $q_{ij} > 0 (i \neq j)$  の具体例を考えることができる。

A) 子供がマルコフ過程のある状態、すなわち目標文法とは異なるある文法を仮定している状態に陥ると、親が持つ目標文法から導出された正しい例文

のみを受け取っているにも関わらず、その状態から逃れられないという状態が存在する．この状態を吸収状態 (Absorbing State) と呼ぶ．

B) 子供は現在の状態では受理できない文を受け取った際、これを学習における刺激として学習によって状態を遷移させ、目標状態に近づいていく．しかし目標文法に到達する前に、子供に十分な刺激が与えられないまま学習期間が終了してしまうと、学習過程にある状態に対応した文法を誤って身につけてしまう．

$S$  行列は文法間の推移性に関わるため、 $Q$  行列は  $S$  行列に依存する (Komarova et al., 2001)．また言語獲得における精度もまた学習アルゴリズムに影響を受ける．そのため、Niyogi が用いた学習アルゴリズム (TLA) は  $q_{ij}$  ( $i \neq j$ ) の確率を不自然に高くする可能性がある．

### 3. 動的遷移行列モデル

本節では 2 節で述べた言語力学の問題を解消するために、新たなモデルを提案する．

#### 3.1 モデルの改良

ここで、2.3 節で示した Niyogi が述べている 2 つの記述について考えてみる．まず A) に関して、現実世界の子供は、親からいくら言葉を聞いても学習過程にある自分の文法を修正できなくなるという状態に陥ることは考えにくい．一般的に、子供は任意の言語について、その文を聞いて学習する限り、目標文法を誤りなく獲得できると考えられている．それゆえ、吸収状態は存在しないと考えるべきである．また B) については、たとえ子供が言語学習期間において十分な刺激を得られなかったとしても、その子供は他の言語を身につけるとことは考えにくく、Genie<sup>3)</sup> のように言葉を身につけることが出来なくなるだろう．このような理由から、現実世界において子供が親の文法の獲得に失敗し、他の文法を身につけるという確率は非常に小さいと考えられる．すなわち  $q_{ij} \approx 0$  ( $i \neq j$ ) と考えるべきだろう．

上記をふまえ、より現実世界に近いモデルを提案するため、我々は  $Q$  行列の定義を変えることから始める．Nakamura and Tojo (2002) が既に計算機

3) 誕生以来13年間父親によって小部屋に閉じ込められて育った子供で、第一言語を正常に獲得できなかった例としてしばしば取り上げられる (Pinker, 1994) ．

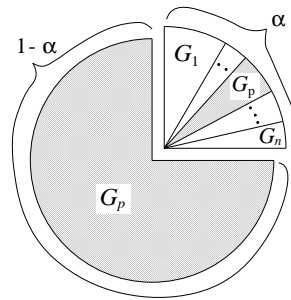


図2 接触確率  $\alpha$

Fig. 2 Exposure probability  $\alpha$

上の実験で示しているように、子供が習得する言語は、その言語学習期間において接触した言語とその接触頻度に大きく影響されることは明らかである．よって、言語学習者である子供は親からのみ発話を受け取り文法を獲得するという Komarova らのモデルを修正し、子供はコミュニティに属するさまざまな言語話者と接触し、そこから文法を学習すると考える．このとき、他言語話者との接触の結果、親の言語を正常に身につけられない可能性が考えられ、その確率を  $\bar{Q}$  行列として新たに定義することを提案する．ここでコミュニティの言語話者ごとの人口比率は世代によって変遷するため、 $\bar{Q}$  行列は時間に関するパラメータを持つようになる．それゆえ  $\bar{Q}(t) = \{\bar{q}_{ij}(t)\}$  となる．我々はこれを動的遷移行列 (Dynamic Transition Matrix) と呼ぶ．したがって(1) 式は次のように修正される (Nakamura, Hashimoto, & Tojo, 2003b) ．

$$\frac{dx_j(t)}{dt} = \sum_{i=1}^n \bar{q}_{ij}(t) f_i(t) x_i(t) - \phi(t) x_j(t) \quad (j = 1, \dots, n). \quad (2)$$

これを動的遷移行列モデル (Dynamic Transition Matrix Model) と呼ぶことにする．

#### 3.2 接触確率 $\alpha$ の導入

次に我々は、子供が親以外の言語話者と接触する確率を表すパラメータ  $\alpha$  を導入する．これを接触確率 (Exposure Probability) と呼ぶ．ここで子供が親の言葉を聞く確率は  $(1 - \alpha)$  である．このとき  $\alpha$  は親の言語以外の言語と接触する確率ではなく、親の言語も含めた多言語との接触確率である．例を図 2 に示す． $G_p$  はある子供の親の文法である．その子供は確率  $\alpha$  の割合で他の言語話者と集団の言語話者の比率に応じて接触する．すなわち、図中

の影がかかった部分の割合で子供は親の言語を聞くことになる。ここで  $\alpha = 0$  のとき、親からしか言語を学習することがないため、Komarova et al. が想定した状況と同じである。また逆に、 $\alpha = 1$  のとき、各言語話者の人口構成比に完全に比例した割合で言語と接触するため、どの言語話者の子供も獲得する言語の条件は等しくなる。

以上をまとめると、新たに定義した  $\bar{Q}(t)$  行列は、接触確率  $\alpha$  および各言語話者の人口構成比  $X(t) = (x_1(t), x_2(t), \dots, x_n(t))$  に依存する。

3.3 学習アルゴリズム

我々は Niyogi のモデルの問題を踏まえ、学習アルゴリズムに次のような制約を与えた：

- a) 言語学習者である子供は生まれた時点で特定の文法を持たない。これに対し、Niyogi のモデルでは初期値としてランダムにパラメータ値を与えるため、生まれてすぐになんらかの文法を持っていると仮定している。
- b) 子供は親からしかことばを聞かずに学習した場合、必ず親の文法を獲得する。これは Niyogi のモデルでは保証されず、子供の文法の獲得過程を示す状態遷移に依存する。また Komarova らのモデルでは、この状況における文法獲得の失敗確率を  $Q$  行列として定義している。
- c) 学習期間中は、目標文法の推定に十分な時間と例文が与えられる。

ここで上記制約を満たす単純な学習アルゴリズムを導入する。図 3 に示した学習の様子を以下に解説する：

- 1) 子供は言語話者によって発話された一文を聞く。この図では  $G_8$  話者から “S V O” という一文を受け取っている。
- 2) 子供は頭の中で文法の数だけカウンタを持っており、もしその文がある文法によって受理されるなら、その文法に対応したカウンタの値をひとつ上げる。これを全ての文法について行う。この図は “S V O” を受理可能な文法が  $G_2, G_4, G_5, G_6, G_8$  であることを表している。
- 3) 文法の推定に十分であると考えられる数の文を受け取り、その間、1) と 2) を繰り返す。この図では “S V O” 以降 “Adv S V O” “V S” ... “S V” の順に文を受け取っている。
- 4) 最も高い値を示したカウンタに対応した文法

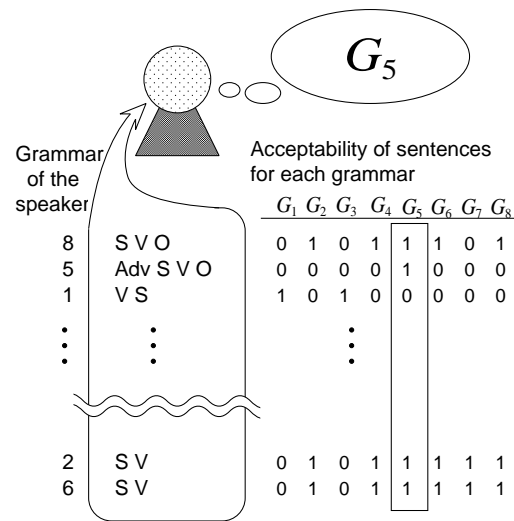


図3 単純な言語獲得アルゴリズムの導入  
Fig. 3 A Simple Learning Algorithm

を子供は採用する。この図は受け取った文を最も多く受理した文法が  $G_5$  であることを表している。

このアルゴリズムを定式化することを考える。学習対象が親の言語だけであった場合、上記の制約 b) から、子供が獲得する文法は、次のような  $G_{j^*}$  となる：

$$j^* = \operatorname{argmax}_j \eta s_{pj} = \operatorname{argmax}_j s_{pj} \quad (= p),$$

ここで制約 c) より、 $\eta$  は文法を推定するのに十分と考えられる入力文の数であり、また、 $p$  は親の文法のインデックスを意味する。

また子供は、コミュニティの各言語話者の人口に比例してそれぞれの言語を聞く機会がある。その場合、子供が獲得すると予想される文法は、次の式を満たす  $G_{j^*}$  となる：

$$j^* = \operatorname{argmax}_j \left\{ \eta \sum_{k=1}^n s_{kj} x_k(t) \right\} = \operatorname{argmax}_j \left\{ \sum_{k=1}^n s_{kj} x_k(t) \right\}.$$

ここで3.2 節で定義した、親以外の言語話者と接触する割合を表す接触確率  $\alpha$  を導入する。これにより、文法の選択は上記 2 式の線形結合となり、子供が推定する文法は次のような  $G_{j^*}$  となる：

$$j^* = \operatorname{argmax}_j \left\{ \alpha \sum_{k=1}^n s_{kj} x_k(t) + (1-\alpha) s_{pj} \right\}. \quad (3)$$

### 3.4 動的遷移行列 $\bar{Q}(t)$

動的遷移行列  $\bar{Q}(t) = \{\bar{q}_{ij}(t)\}$  の定義は、 $t$  世代における各言語の話者に影響を受けながら文法を学習した結果、 $G_i$  話者の子供が  $G_j$  に遷移する確率である。したがって(3) 式を確率関数に変換する必要がある。ここで(3) 式から、 $P^n(i, j) = \alpha \sum_{k=1}^n s_{kj} x_k(t) + (1-\alpha) s_{ij}$  とする。これは  $n$  種類ある言語のうち、 $G_i$  話者の子供が  $G_j$  によって受理することができる文を受け取る確率である。まず最初に2つの文法  $G_1$  と  $G_2$  しか存在しない場合を考えると、 $G_1$  を持っている言語話者の子供は、次のような条件を満たした場合  $G_1$  を獲得する：

$$P^2(1, 1) \geq P^2(1, 2)$$

両辺の値はそれぞれ独立して 0 から 1 までの範囲で値をとる。このとき子供の学習前の初期状態で、どちらの値もわからない場合の文法の採択確率を考える。ここで両辺の値が 0 から 1 までの範囲で一樣に分布すると仮定すると、 $G_1$  を採用する確率は左辺の値そのもの ( $0 \leq P^2(1, 1) \leq 1$ ) である。同様に  $n$  個の文法  $\{G_1, \dots, G_n\}$  のケースを考える。 $G_1$  を持っている言語話者の子供が  $G_1$  を獲得するためには、

$$P^n(1, 1) \geq P^n(1, i) \quad \text{for all } 2 \leq i \leq n$$

という条件を満たさなければならない。すなわち  $n-1$  個の文法と比較するため、 $G_1$  の採択確率は  $(P^n(1, 1))^{n-1}$  となる。同様に、 $G_i$  を持つ言語話者の子供が  $G_j$  を獲得する確率を、それぞれの文法が受理する確率から求めたものは次のようになる：

$$(P^n(i, j))^{n-1} = \left\{ \alpha \sum_{k=1}^n s_{kj} x_k(t) + (1-\alpha) s_{ij} \right\}^{n-1}. \quad (4)$$

これを  $j$  に関して正規化することによって  $\bar{q}_{ij}(t)$  を得る：

$$\bar{q}_{ij}(t) = \frac{(\alpha \sum_k s_{kj} x_k(t) + (1-\alpha) s_{ij})^{n-1}}{\sum_l (\alpha \sum_k s_{kl} x_k(t) + (1-\alpha) s_{il})^{n-1}}. \quad (5)$$

このとき  $\sum_{j=1}^n \bar{q}_{ij}(t) = 1$  である。

### 3.5 人口動力学上のクレオールの定義

我々は普遍文法の立場から、言語獲得のメカニズムとしてあらかじめ原理とパラメータを仮定している。そして、普遍文法にはクレオールも含まれるという立場を取り (Bickerton, 1981)、パラメータセッティングのレベルでのクレオールの出現を想定している。すなわち、クレオールの文法を  $G_c$  とすると、原理に

よって制限される文法集合は  $\{G_1, \dots, G_c, \dots, G_n\}$  となる。

また本稿では、 $S$  行列の値はビジン化の結果として与えられるものと仮定している。すなわち、2つの独立した言語間における類似性とは、表 1 のように語彙を共有し、埋め込み文のない単純な文を発話することによって類似性が現れると考えている。

ここで我々は、これまでの言語学的な定義 (亀井他, 1996; 風間・長谷川, 1992) とは大きく異なるが、人口動力学の視点から見たクレオールの定義をする。クレオールとは、新しい言語の創発現象である。すなわち、あるコミュニティである時点で存在しなかった言語が、後に存在するようになる現象と考えることができる。よって、次のように定義できる。

定義 1 (共存クレオール) 他の言語と共存するクレオールとは、次のような文法  $G_c$  である<sup>4)</sup>：

$$x_c(0) = 0, \quad x_c(t) > \theta_c.$$

定義 2 (優勢クレオール) 優勢言語となるクレオールとは、次のような文法  $G_c$  である：

$$x_c(0) = 0, \quad x_c(t) > \theta_d.$$

ここで  $x_c(t)$  は人口動態が収束し、安定した  $t$  世代における  $G_c$  話者の人口比率、 $\theta_c$  と  $\theta_d$  はそれぞれ共存 (coexistent) クレオールと優勢 (dominant) クレオールであるとみなすための人口比率の閾値を示す。本稿では  $\theta_c = 0.1$ ,  $\theta_d = 0.9$  としている。これらの定義は、初期状態では誰も話していなかった言語が、最終的にはある割合の話者を獲得することを表している。定義 1 は、少数ではあるが、一定数の個体が世代を通じて文法を維持することを意味し、定義 2 はそのコミュニティ内で使用される言語のほとんどがクレオールによって占有される状態を表している。

この節で論じたモデル、すなわち動的遷移行列  $\bar{Q}(t)$  の有効性を検証するために、実験を行い、人口構成比と接触確率の変化においてクレオールの創発を見る。また先行研究のモデルを修正する際に  $Q$  行列と並んで重要な役割を負っていた  $S$  行列 (類似性) についてもクレオール創発の条件を検証する。このため、本研究では、以下のように実験計画を立てる。

4) アフリカ西海岸のシエラレオネで話されているクリオ語は、約 50 万人の母語使用者に対し約 30 万人の非母語使用者がいる (亀井他, 1996)。この人口比が安定状態にあるとするならば、これは共存クレオールである。



実験 1 動的遷移行列モデルの検証

実験 2 優勢クレオールが創発する条件の検証

次節以降では、それぞれの実験について節を分けて実験の方法と結果について論じる。

#### 4. 実験 1 - 動的遷移行列モデルの検証

ここでは 3 節で提案した動的遷移行列を含む言語動力学の振る舞いを、クレオールの創発に注目して分析する。

##### 4.1 実験1の言語セットと $S$ 行列

実験1では Niyogi のモデル (Niyogi, 1996) と同様、Gibson and Wexler (1994) によって導出された8つの文法を採用した(付録 A参照)。各文は、文法項目、すなわち品詞の順序で指定されると考える。

一般的に、各文法とそこから導出される文の生成確率が求まれば  $S$  行列を求めることができる。ここでは、 $G_i$  を持つ言語話者がそれぞれの文を発話する確率が均一であると仮定し、 $S$  行列の要素  $s_{ij}$  は  $L(G_i)$  と  $L(G_j)$  にある共通な文の種類数を  $L(G_i)$  にある文の数で割った値として求めた。表 1 から、対角要素である  $s_{ii}$  は常に 1 であり、また  $s_{12}$  と  $s_{21}$  を求める場合、 $L(G_1)$ 、 $L(G_2)$  に含まれる文の総数はそれぞれ 12 と 18 なのに対し、共通な文は 6 個であるため、 $s_{12} = 6/12 = 1/2$ 、 $s_{21} = 6/18 = 1/3$  となる。

##### 4.2 実験1-1 - 人口動力学上でのクレオールの創発

本実験の目的は、接触確率  $\alpha$  をパラメータとすることによって、我々が提案した動的遷移行列モデルの振る舞いを観察することである。予備実験において最も顕著に特徴が現れたところとして、初期状態  $x_1(0) = 0.32$ 、 $x_3(0) = 0.32$ 、 $x_5(0) = 0.36$ 、その他は  $x_i(0) = 0$  の値を与えたときの実験を行った。この初期値を用いて接触確率  $\alpha$  を 0 から 1 の範囲で与え、モデルの振る舞いを観察する。動的遷移行列モデルの結果を図 4 に示す。

図 4(a) は  $\alpha = 0$  を与えたときの結果である。このとき言語学習者である子供は親からしか言語を学ばない。すなわち、(2) 式の動的遷移行列  $\bar{Q}(t) = \{\bar{q}_{ij}(t)\}$  は定数であるため、この動力学は (1) 式と同じ振る舞いをする。 $G_3$  話者の人口比率は世代を経るごとに増加し、最終的にはコミュニティ

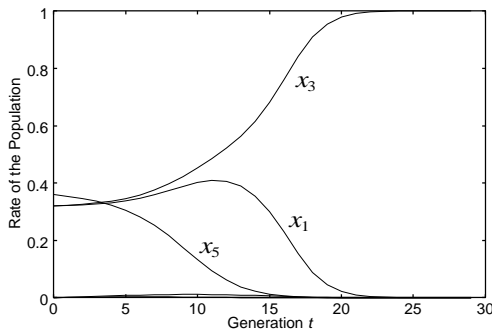
のほとんどが  $G_3$  を獲得する。これは文法の適応度に応じてそれぞれの文法を持つ人口が増減した結果であり、 $x_3$  が増加するのに対して、 $x_1$  と  $x_5$  は適応度の減少によって人口比率を減らしてしまうのである。

図 4(b) は  $\alpha = 0.627$  を与えたときの結果である。 $\alpha$  が増加するに従い、初期人口を与えられていない  $x_2$  が徐々に増加している。これは初期状態において誰も話していなかった  $G_2$  が、世代を経ることによってその話者を増加させていることを表している。この原因は、 $\alpha$  が増加することによって (2) 式中の特に  $q_{12}$  の値が増加し、 $G_1$  話者の子供が  $G_2$  を獲得するようになったためであると考えられる。我々はこの現象をクレオールの創発 (Arends et al., 1994; Bickerton, 1981; DeGraff, 1999) であると捉えている。これは、言語学習者である子供が、さまざまな言語話者と頻りに接触することによって、親の言語よりもそのコミュニティにおいて最もコミュニケーション能力の高い文法を選んだ結果である。しかし、この図 4(b) の場合、 $x_1$  と  $x_5$  のほとんどが  $x_2$  へ流出した後、 $x_2$  はそれ以上人口を増加させることができず、 $x_3$  に吸収され、最後には消失してしまう。

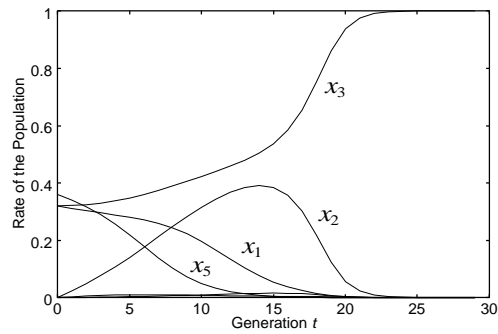
$\alpha = 0.627$  と  $\alpha = 0.628$  の間には大きな境目が存在することが図 4(c) からわかる。 $\alpha = 0.627$  では  $G_3$  が最終的な優勢言語であったのが、 $\alpha = 0.628$  において優勢言語は  $G_2$  に転じていることが観察された。すなわち定義 2 の優勢クレオールが創発したことを示している。これまでの  $\alpha$  の変化による系全体の振る舞いと比較して、 $\alpha$  が 0 から 0.6 付近までは緩やかな変化だったのに対し、この境目付近における振る舞いの変化は、それまでと比べて非常に急激なものであった。

その後さらに  $\alpha$  の値を増加していったところ、 $G_2$  は優勢クレオールであることを保ち続け、より安定していった(図 4(d)参照)。そのうえ、 $\alpha$  が増加すると、収束世代が短くなっていることが図 4(c) と図 4(d) を比較するとわかる。 $\alpha$  の定義から、 $\alpha = 1$  であるときに最もクレオールが創発しやすい条件であることが認められる。

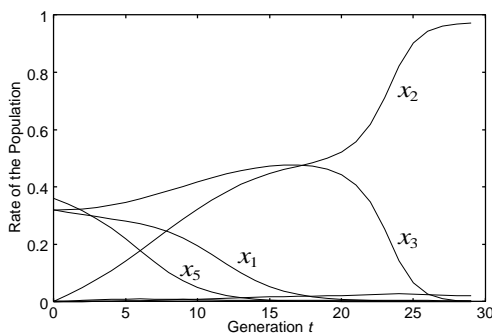
本実験において、特定の  $S$  行列と各言語の人口比率を初期値として与えたところ、クレオールの創発を観察した。またその創発は接触確率  $\alpha$  に依存することが確認された。



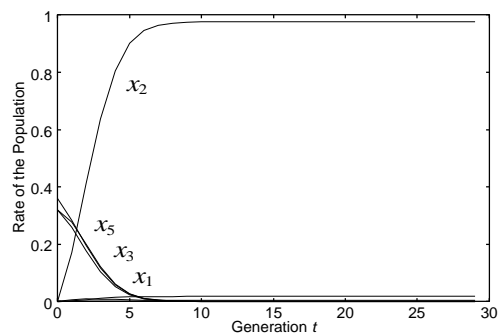
(a)  $\alpha = 0$  Not Creolized



(b)  $\alpha = 0.627$  Not Creolized



(c)  $\alpha = 0.628$  Creolized



(d)  $\alpha = 1$  Creolized

図4 動的遷移行列モデルの結果

Fig. 4 Result of the dynamic transition matrix model

4.3 実験1-2 - 初期条件に見る優勢言語の領域

4.2節で見た動的遷移行列モデルの振る舞いから、どの言語が優勢となるかは、初期人口比率と接触確率  $\alpha$  に依存することがわかる。このとき、初期人口を与えられていない言語が優勢言語となると、クレオールと認識されるのである。次の実験では上記2つの初期条件と、優勢言語となる言語の関係を示し、そこからクレオール化の条件を調査することを目的とする。

優勢言語となる言語間の境界を明白にするため、 $G_5$  の初期人口比率をパラメータとする。すなわち  $x_5(0)$  を 0 から 1 の範囲で与え、 $x_1(0)$  と  $x_3(0)$  には次のようにその残りを均等に配分する：

$$x_1(0) = x_3(0) = (1 - x_5(0))/2.$$

この  $\alpha - x_5(0)$  平面において、図 4(b)と図 4(c)のような、優勢言語が入れ替わる境界を調べた。その

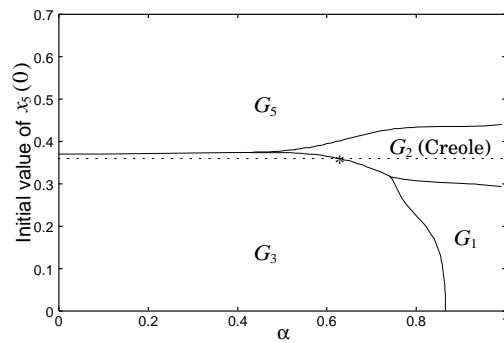


図5 優勢文法の領域の出力

Fig. 5 Regions of the dominant grammar

結果を図 5に示す。図中、実線は優勢言語が入れ替わる境目を表しており、それぞれの領域内に示されている文法  $G_i$  が優勢言語となった文法を表してい

る．その中にはクレオールとなった  $G_2$  も含まれている．破線は4.2節で行った一連の実験を表し，アスタリスク(\*)が描かれている点は図4(c)の初期値( $\alpha = 0.628, x_5(0) = 0.36$ )に対応する．

さらに図5を詳しく見ていくと， $G_2$  が図の右半分以降に現れていることから，クレオールが創発するための条件として， $\alpha$  の値がある程度大きくなければならないことがわかる．

図4(a)では， $G_5$  の初期人口比率が最も多いにも関わらず，初期人口が割り当てられている他の言語との類似度の低さから  $G_5$  の適応度は  $G_3$  のそれ以下となっている．そのため， $G_5$  は優勢言語になっていないが，それよりもさらに初期人口比が増えた  $x_5(0) = 0.371$  になると，類似性の低さを初期人口比率で補い，優勢言語となる．図4(a)，図4(b)に見るように， $x_5(0)$  の値が低い場合， $G_1, G_3$  の人口比が増す． $t = 0$  における  $G_1$  と  $G_3$  の適応度は同じであるが， $\alpha$  の値によって優勢言語は異なる．例えば  $\alpha = 0$  のとき，動的遷移行列は  $S$  行列だけで求まるため， $\bar{q}_{11}(t) \simeq 0.99219$  であるのに対し， $\bar{q}_{33}(t) \simeq 0.99998$  である．したがって  $G_1$  話者の子供は  $G_3$  話者の子供に比べ他の言語に遷移しやすい．また，他の言語話者の子供は  $G_1$  よりも  $G_3$  に遷移しやすいため， $G_3$  が優勢言語となる．一方， $G_1$  は  $G_3$  よりもクレオール  $G_2$  と類似度が高い．すなわち  $s_{13} = s_{31} = 2/12$  であるのに対し， $s_{12} = 6/12, s_{21} = 6/18$  である．そのため， $\alpha$  が大きいと図4(b)のように  $x_2$  が上昇し， $G_2$  に伴って  $G_1$  も適応度が増える．同時に  $G_1$  に関する動的遷移行列の値も影響を受け，特に  $\bar{q}_{21}(t)$  と  $\bar{q}_{12}(t)$  の値が増加する．このとき， $x_1(0)$  の値が小さいと， $\bar{q}_{21}(t) < \bar{q}_{12}(t)$  となり， $G_1$  話者の子供は  $G_2$  に遷移しやすくなり， $G_2$  が優勢クレオールとなる．反対に， $x_1(0)$  の値が大きいと， $\bar{q}_{21}(t) > \bar{q}_{12}(t)$  となり， $G_1$  が優勢言語となる．このようにして， $\alpha$  の値が小さいところでは  $G_3$ ，大きいところでは  $G_1$  が優勢言語となるのである．

初期の人口構成比に関してさまざまな組合わせで実験を行ったが， $\alpha = 0$  のときにクレオールが創発したケースは確認されていない． $\alpha = 0$  では我々のモデルは Komarova らのモデルと一致する．すなわち，動的遷移行列  $\bar{Q}(t)$  と接触確率  $\alpha$  を導入したことにより，クレオールの創発現象を実現することができた．

#### 4.4 実験1の考察

本節の実験においてクレオールが創発する現象を観察したが，初期の人口構成比や使用される言語によっては優勢クレオールは創発しない．例えば図5から，初期人口を  $(x_1, x_3, x_5) = (0.4, 0.4, 0.2)$  として与えると， $\alpha$  を増加させることによって優勢言語は  $G_3$  から  $G_1$  に変化するが，クレオールは優勢にならないことがわかる．

また，実験1-1の環境において， $G_1$  から  $G_8$  までさまざまな人口比率を初期値として与えたところ，クレオールとして創発した言語は  $G_2$  だけであった．クレオールとは，その定義から初期において誰も話していなかった言語である．したがって，ここで与えられた言語セットにおいては単純に  $G_2$  話者の人口が初期値として与えられる場合，すなわち  $x_2(0) > 0$  のとき，クレオールは創発しない．本実験で  $G_2$  だけがクレオールとなった原因は， $G_2$  と他の言語との類似度を表す  $S$  行列の要素， $s_{i2}$  または  $s_{2i}$  ( $1 \leq i \leq n$ ) が他の値と比べて比較的高いためであると考えられる．

したがって，実験1から得られた結論は，言語話者の人口遷移は，言語間の類似性，言語話者の人口構成比，および接触確率  $\alpha$  に大きく依存し，クレオールが出現するためには  $\alpha$  が十分大きく，他言語との類似度がある程度大きい必要がある，ということである．また，既存のモデルでは見られなかったクレオールの創発を観察することで，本稿で導入したモデルがより現実的であることを示した．

#### 5. 実験2 - 類似性に関する条件の検証

実験1で，クレオールが創発するためにはクレオールと他の既存の言語間の類似性について条件があることが示唆された．本節ではこの条件を一般的に導出することを目的とする (Nakamura, Hashimoto, & Tojo, 2003a) ．

##### 5.1 実験2の言語セットと $S$ 行列

ここでクレオールとして創発する言語とはどのようなものであるか考える．動的遷移行列モデルから動力学を導出するために必要な言語情報は  $S$  行列だけである． $S$  行列は本来文法セットから計算されるものであるが，本実験では逆にクレオール創発の条件を  $S$  行列に求めるため， $s_{ii} = 1$  ( $1 \leq i \leq n$ ) や， $s_{ij} = 0$  ならば  $s_{ji} = 0$  ( $i \neq j$ ) といった制約

のもとに  $S$  行列の各要素をパラメータとする．

### 5.2 実験2の初期条件とパラメータ

クレオールが創発するための最も単純なモデルとして3言語の場合を考える．すなわち2つの言語集団が接触した結果、第3の言語としてクレオールが創発する可能性がある環境である． $G_1$  と  $G_2$  を既存の言語とし、 $G_3$  をクレオールになりえる言語とする．人口構成比に対する影響を避けるため、既存の2言語の初期人口を等配分する．したがって初期に与える人口は  $(x_1(0), x_2(0), x_3(0)) = (0.5, 0.5, 0)$  である．また、接触確率  $\alpha$  の値はクレオールが最も創発しやすい値をとることとした．すなわち、実験1-2から明らかのように、 $\alpha = 1$  である．これらの値は以降の実験を通じて共通である．

ここで最も単純な場合として、 $S$  行列を次のような対称行列とした：

$$S = \begin{pmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{pmatrix}. \quad (6)$$

要素  $a (= s_{12} = s_{21})$  は既存の2言語間の類似度を表しており、 $b (= s_{13} = s_{31})$ 、 $c (= s_{23} = s_{32})$  はそれぞれの言語とクレオールとの類似度である．

### 5.3 実験2 - 優勢クレオールが創発する条件

3言語での優勢クレオールが創発した例を図6に示す．図6(a)はクレオール  $G_3$  が優勢である、すなわち  $x_3(t) > \theta_d = 0.9$  であることを示している．このときの(6)式の各要素の値は  $(a, b, c) = (0, 0.174, 0.174)$  である． $b$  と  $c$  が同じ値であるため、 $x_1$  と  $x_2$  の人口遷移は同じ振る舞いをする． $a$  の値が0であるということは、 $G_1$  と  $G_2$  に共通の文が全く存在せず、それぞれの言語話者の間の会話は全く成立しないことを意味する．図中、初期状態で  $G_3$  の話者が誰もいなかったが、時間が経つにつれ  $G_1$ 、 $G_2$  話者が移行することによって  $G_3$  話者の人口が増加し、最終的に  $\theta_d$  以上の比率を占めるようになり収束した様子を示している．このように3言語の場合でも実験1-1と同様、優勢クレオールが創発することがわかる．

$b$  と  $c$  の値を増加させたときの結果を図6(b)に示す． $S$  行列の各要素の値は  $(a, b, c) = (0, 0.176, 0.182)$  である．このときクレオール  $G_3$

はその話者人口を減少させ、優勢言語は既存の言語  $G_2$  に変わったことを表している．また、図6(c)はクレオール  $G_3$  が最も人口が多いが、収束点においてその人口比率が優勢クレオールであるとみなす閾値  $\theta_d$  よりもわずかに低い、すなわち  $x_3(t) < \theta_d = 0.9$  であることを表している．このときの  $S$  行列の値は  $(a, b, c) = (0, 0.188, 0.189)$  である．これは優勢言語が存在しない例である．

(6)式における  $S$  行列の各要素をパラメータとし、優勢クレオールが創発した値を  $a, b, c$  空間上に表したものが図7(a)である．図を見る限り平面に近く、クレオールが創発する範囲は非常に限られていることがわかる．ここから得られる  $S$  行列のおおよその条件は次のように表される：

$$a \lesssim 0.12 \quad (7)$$

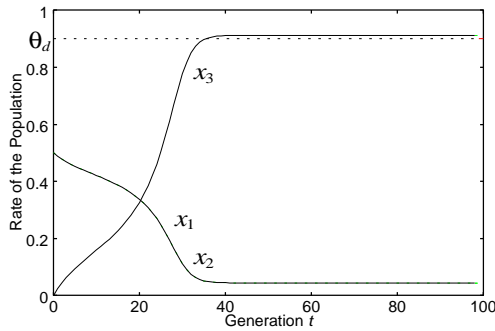
$$0.35a + 0.136 \lesssim b \simeq c \lesssim 0.2 \quad (8)$$

図中の破線で囲まれた部分を、 $b$ - $c$  空間上に拡大したものを図7(b)に示す．

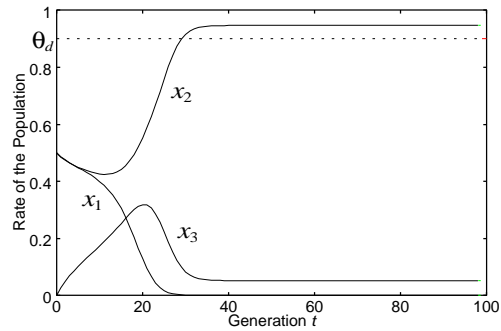
図6(a)と図6(b)において、それぞれの言語が優勢言語となるような  $S$  行列の値が存在することを示した．それぞれの要素の値が図7(b)中の  $\times$  で示された点 (a) から点 (d) に対応している．

ここで、点 (b) から点 (a) に向けて値を変えていくと、優勢言語が  $G_2$  から  $G_3$  に突然入れ替わる境目に遭遇する．その境界を示したのが図7(b)の実線である．実線で囲まれた領域の中の値が  $S$  行列の値として与えられると、優勢クレオールが創発する．最も外側にある実線が(6)式における  $a = 0$  のときの境界であり、以降、 $a$  の値が大きくなるごとにクレオールが創発する条件が限られて行くことがわかる． $a = 0$  であるということは、既存の2言語の話者間で全くコミュニケーションがとれない状況を表しており、このときが最も新言語が創発しやすく、共通言語となりやすい．また、 $a$  の値が大きくなるにつれ、既存の2言語間でコミュニケーションがとれるようになり、それぞれの言語間で人口の遷移が行われやすくなる．その結果、 $G_3$  への遷移が妨げられるため、 $a$  が大きくなるとクレオールが創発しにくくなるのである．図6(d)は図6(a)の初期値から  $a = 0.11$  に変更したときのモデルの振る舞いを表しており、図6(a)と比較して収束までの時間が遅くなっているのがわかる．

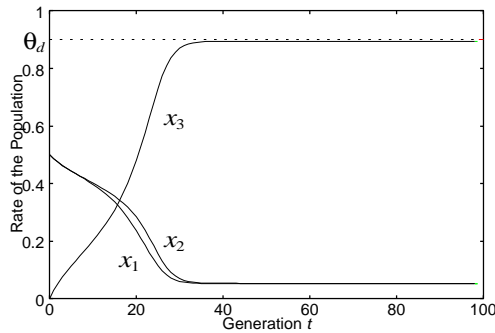
図7(b)中、点 (a) から点 (c) に向けて値を変化させたとき、 $x_3(t)$  は連続的に人口を減らし、図中



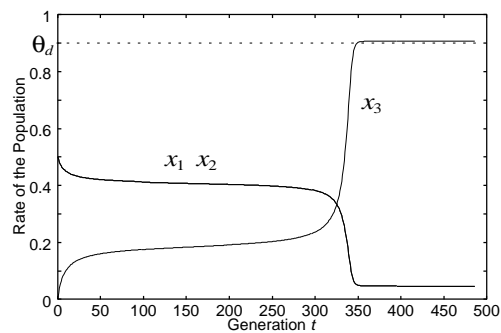
(a)  $(a, b, c) = (0, 0.174, 0.174)$ , Dominant, Creolized



(b)  $(a, b, c) = (0, 0.176, 0.182)$ , Dominant, Not-Creolized



(c)  $(a, b, c) = (0, 0.188, 0.189)$ , Not-Dominant, Creolized



(d)  $(a, b, c) = (0.11, 0.174, 0.174)$ , Dominant, Creolized

図6 3言語での優勢クレオールの創発

Fig. 6 The emergence of dominant creole in 3 languages

の実線部分の短辺を境にして  $G_3$  は優勢クレオールではなくなる。  $S$  行列の要素  $b, c$  の値が大きくなると、  $G_3$  の人口比率が高くなった場合においても、  $G_1, G_2$  への人口の流出が増大するため、それぞれの言語と共存し、クレオールが優勢言語にならないことが原因である。したがって、実線の長辺部分と短辺部分の境界で発生する変化の特徴は大きく異なる。図中の破線で示した境界は、  $G_1$  または  $G_2$  が優勢言語である領域から、優勢クレオールではないが  $G_3$  が最も人口比率の高い言語である領域に突然変化した境目を示している。

$b$  と  $c$  の値が大きいと、  $G_1$  と  $G_2$  からそれぞれ  $G_3$  へ多くの人口が遷移する。また逆に  $G_3$  の人口が増加すると  $G_1$  と  $G_2$  へ遷移するため、共存するようになる。しかし逆に  $b$  と  $c$  の値が小さいと、

人口比が安定するまでの世代数が増加する。  $b$  と  $c$  の値が小さいときの優勢クレオールが創発するケースを考えると、  $G_3$  の人口比率が  $G_1, G_2$  のそれを追い抜くまでに多くの時間を要し、さらに  $a = 0$  のときにおいて  $b \simeq c \lesssim 0.136$  を境に優勢クレオールが観察できなかった。さらに  $a$  の値に比例して収束世代が遅くなり、クレオールが創発しなくなっていく様子が図 6(d)と図 7(a)をみるとわかる。

5.4 実験2の考察

実験2において、クレオールが創発し、優勢言語となるための、言語間の類似性に関する条件を観察した。ここで自然言語を背景に、この条件がどのような意味を含んでいるのか考察する。あるコミュニティにおいて 2 つの言語が存在したとき、(7) 式

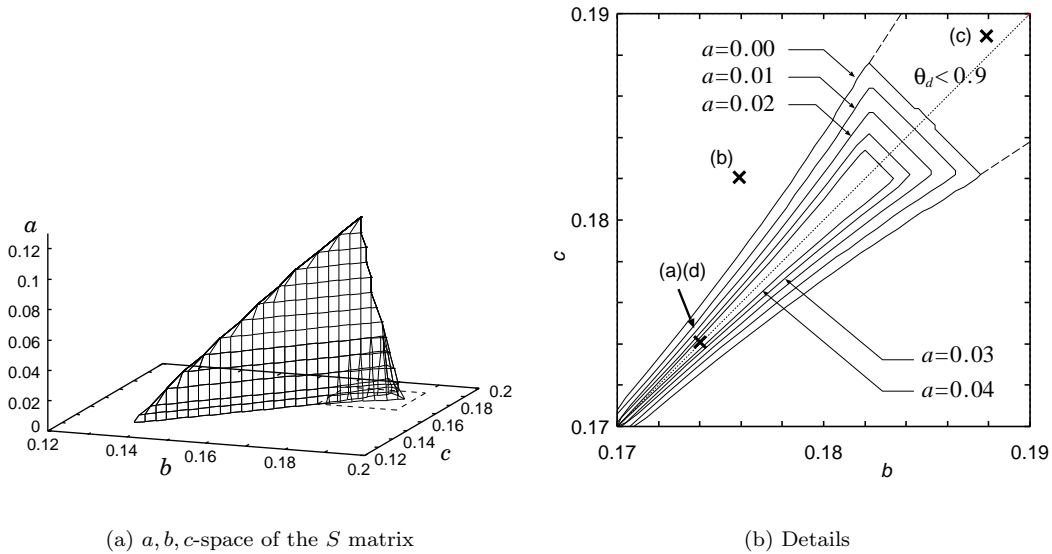


図7 優勢クレオールとなるための条件 ( $\theta_d = 0.9$ )  
 Fig. 7 Conditions for Dominant Creole ( $\theta_d = 0.9$ )

の条件はこれら 2 言語が互いに似ていないことを表している。もしこれらが十分に類似していれば、それぞれの言語話者はコミュニケーションをとることができるため、ピジンやクレオールといった新言語が誕生する必要がないだろう。また(8) 式は、クレオールは既存の言語と、既存の言語間の類似度よりも似ていなければならないが、あまり似すぎていてはならないことを表している。もし 2 言語のうちどちらかが新言語と必要以上に類似していると、クレオールは創発するが、その言語はクレオール話者とコミュニケーションがとれるため、新言語に移行せず、クレオールとその言語とで共存することになる。逆にクレオールと既存の言語とが必要以上に似ていないと、クレオールへの人口の移行に時間がかかり、クレオールは優勢言語になりにくい<sup>5)</sup>。さらに既存の 2 言語はクレオールとの類似度もほぼ同等でなければならないと制限している ( $b \simeq c$ )。この均衡が崩れると、クレオールは短期間において創発するものの、話者人口は安定することなく消滅し、クレオールに類似している言語が最終的に優勢

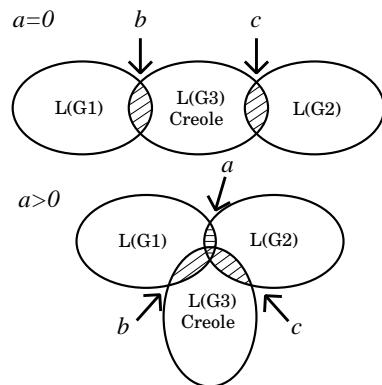


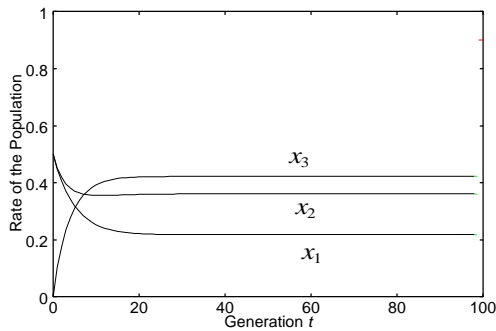
図8 言語空間上のクレオール創発の条件

言語となる。

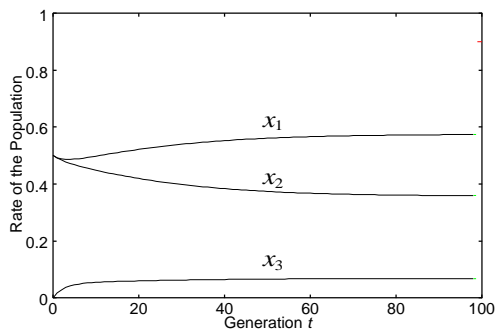
これらの条件から、クレオールと既存の 2 言語を言語空間上に配置したものを図 8 に示す。原理とパラメータ理論から考えると、これらの言語空間上の言語はあらかじめ原理によって与えられており、人間の第一言語獲得はパラメータを設定することによってこの空間上の言語を選択していることとなる。我々はクレオールの創発を観察した結果、この言語空間の言語の配置を導き出したと考えることができる。

パラメータ空間の全体を見ると、優勢言語とクレ

5) ここでいう言語の類似性とは 2 言語間の文法の類似性について言及しているわけではないため、クレオール文法とは既存の言語の文法が混ざりあった言語ではないという、実際のクレオールの観察結果に矛盾するものではない。



(a)  $(a, b, c) = (0.3, 0.4, 0.5)$ , Coexistent, Creolized



(b)  $(a, b, c) = (0.35, 0.25, 0.1)$ , Coexistent, Not-Creolized

図9 言語の共存

Fig. 9 Coexistent-language set

オールの創発に関して次の4つの領域に分類できることがわかった。

- i) 図6(a)のように、クレオールが優勢言語となる ( $x_3(t) \geq \theta_d = 0.9$ )。
- ii) 図6(b)のように、クレオールが創発しない ( $x_3(t) < \theta_c = 0.1$ ) が優勢言語が存在する。
- iii) 図9(a)のように、クレオールが創発するが、他の言語と共存する ( $\theta_c \leq x_3(t) < \theta_d$ )。
- iv) 図9(b)のように、クレオールが十分な話者を持たず ( $x_3(t) < \theta_c$ )、既存の言語が共存する。

実験において、上記 iii) および iv) の共存カテゴリーが現れるパラメータ領域はおおよそ  $a, b, c \geq 0.3$  であった。この値はそれぞれの言語は相対的に類似度が高く、またそれぞれの言語話者は他の言語話者と互いにコミュニケーションをとることができるこ

とを表している。この状況から考えると、これらの言語は互いに異なる独立した言語であるとみなすよりも、むしろ方言であると考えた方が良いように思われる。一般的に言語学の見方をすると、ある言語が他のある言語の方言であるのか、あるいは異なる言語であるのかということを明白に区別することはできない<sup>6)</sup>。一方、本稿で用いたような人口動力学的手法では、共存する言語間の類似性について述べることができる。ただし、本実験では接触確率  $\alpha = 1$  という特殊な状況における実験であるため、類似度が 0.3 以上という値自体が基準となるわけではない。

## 6. 結論

言語学分野におけるクレオールの主な調査対象は、これまでに発見されてきたクレオール間に見られる言語としての類似性と、そこから想像される普遍文法が存在する可能性についてである。ここで我々は普遍文法を仮定し、構成論的手法によってクレオールが創発するための条件を数理的に導き出すことを目的として実験を行った。言語の定義を各言語間の類似性から、クレオールの定義を言語話者の人口比率からそれぞれ与え、これまでの言語学的な視点と全く異なるアプローチで調査を行った。

我々は Komarova et al. (2001), Nowak and Komarova (2001) の言語動力学モデルをより現実的なモデルに改良するため、動的遷移行列モデルを提案した。このとき、現実的であるための条件として次の点を主張した。

- 子供の言語獲得は周りの言語話者が話す言語に影響を受ける。
- 親としか会話を行わなかったとき、子供は親の言語を身につける。

このモデルでは、言語間の遷移を特徴づけるパラメータが、人口比率に依存するので、遷移行列が動的に変化するものとなる。

このモデルの計算機実験による分析を通じて、以下のことが明らかになった。

- 1) 初期の人口構成比が最終的な優勢言語に影響を及ぼす。
- 2) 言語学習者が、どの言語から入力となる文を

6) このような境界線はしばしば政治的に設けられる。例えばボスニア・ヘルツェゴビナのセルビア語、クロアチア語、ボスニア語などがそれにあたる (Comrie, Matthews, & Polinsky, 1996)。

受け取るかが最終的な優勢言語に影響を及ぼす。

- 3) クレオールが存在するための、言語間の類似性の条件が存在する。

このうち、3)のクレオールが創発するための条件は次のようなものであった。

- 既存の言語は互いに類似しすぎていないこと。
- 既存のある言語から見て、他のどの既存の言語よりもクレオール言語がその言語と似ていること。
- 既存の言語とクレオールとの類似度は、各既存言語で同程度であること。

普遍文法を仮定した場合、人間が獲得することができる言語は、クレオールも含め、あらかじめ与えられていると考えられている。図 8はその言語間の関係をクレオールの創発条件を導いた結果から描くことができたものである。この結果は言語学の分野への大きな貢献であると考えられる。

本モデルにおいては、言語に関する全ての特性を、他の言語との類似度として表現したことが問題点として挙げられる。実験 1 では、特に原理とパラメータを仮定したことを明確に示すために、言語間の類似度をそれぞれの文法から導き出される文型(語順)の類似度として採用したが、本来ならば語彙や音韻などの情報も含めて導き出されるものである。本稿においてクレオールが創発するための類似度に関する条件を求めたが、ここから文法構造や音韻に関する具体的な条件を導くことはできない。数理モデルの上に文法的情報をいかに表現するかというのを今後の課題とする。

## 文 献

- Arends, J., Muysken, P., & Smith, N. (Eds.) (1994). *Pidgins and Creoles*. Amsterdam: John Benjamins Publishing Co.
- Bickerton, D. (1990). *Language and Species*. University of Chicago Press.
- Bickerton, D. (1981). *Roots of language*. Ann Arbor, MI: Karoma.
- Briscoe, E. J. (2002). Grammatical Acquisition and Linguistic Selection. In T. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, chap. 9. Cambridge University Press.
- Cangelosi, A. & Parisi, D. (2001). Computer Simulation: A New Scientific Approach to the Study of Language Evolution. In A. Cangelosi & D. Parisi (Eds.), *Simulating the Evolution of Language*, chap. 1. London: Springer.
- Chomsky, N. (1975). *Reflections on Language*. New York: Pantheon.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht, The Netherlands: Foris.
- Comrie, B., Matthews, S., & Polinsky, M. (Eds.) (1996). *The Atlas of Languages*. London: Quatro Publishing.
- DeGraff, M. (Ed.) (1999). *Language Creation and Language Change*. Cambridge, MA: The MIT Press.
- Gibson, E. & Wexler, K. (1994). Triggers. *Linguistic Inquiry*, **25** (3), 407–454.
- Gold, E.M. (1967). Language identification in the limit. *Information and Control*, **10**, 447–474.
- Hashimoto, T. (2001). The Constructive Approach to the Dynamic View of Language. In A. Cangelosi & D. Parisi (Eds.), *Simulating the Evolution of Language*, chap. 14. London: Springer.
- Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In T. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, chap. 6. Cambridge University Press.
- Kirby, S. & Hurford, J. R. (2001). The Emergence of Linguistic Structure: An Overview of the Iterated Learning Model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the Evolution of Language*, chap. 6. London: Springer.
- Komarova, N. L., Niyogi, P., & Nowak, M. A. (2001). The evolutionary dynamics of grammar acquisition. *Journal of Theoretical Biology*, **209** (1), 43–59.
- Komarova, N. L. & Nowak, M. A. (2001). Population Dynamics of Grammar Acquisition. In A. Cangelosi & D. Parisi (Eds.), *Simulating the Evolution of Language*, chap. 7. London: Springer.
- Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: John Wiley & Sons, Inc.
- Nakamura, M., Hashimoto, T., & Tojo, S. (2003a). Creole Viewed from Population Dynamics. *Proceedings of the Workshop/Course on Language Evolution and Computation in ESS-*



- LLI, 95–104.
- Nakamura, M., Hashimoto, T., & Tojo, S. (2003b). The Language Dynamics Equations of Population-Based Transition – a Scenario for Creolization. In H. R. Arabnia (Ed.), *Proceedings of the International Conference on Artificial Intelligence (IC-AI'03)*, 689–695. CSREA Press.
- Nakamura, M. & Tojo, S. (2002). The Emergence of Artificial Creole by the EM Algorithm. In S. Lange, K. Satoh, & C. H. Smith (Eds.), *Discovery Science*, Vol. 2534 of *Lecture Notes in Computer Science*, 374–381. Springer.
- Niyogi, P. & Berwick, R. C. (1995). The Logical Problem of Language Change. Report No. AIM-1516, AI Lab, MIT.
- Niyogi, P. (1996). *The Informational Complexity of Learning from Examples*. Ph. D. thesis, Massachusetts Institute of Technology.
- Nowak, M. A. & Komarova, N. L. (2001). Towards an evolutionary theory of language. *Trends in Cognitive Sciences*, 5 (7), 288–295.
- Ono, T., Tojo, S., & Sato, S. (1996). Common Language Acquisition by Multi-Agents. *International Computer Symposium (ICS'96), Proceedings on Artificial Intelligence*, 218–223.
- Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: William Morrow and Company.
- Weibull, J. (1995). *Evolutionary Game Theory*. Cambridge, MA: The MIT Press.
- 風間 喜代三・長谷川 欣也(編) (1992). 『言語学百科事典』. 東京: 大修館書店.
- 亀井 孝・河野 六郎・千野 栄一(編) (1996). 『言語学大辞典』. 東京: 三省堂.
- 中村 誠・東条 敏 (2003). マルチエージェント環境での人工ピジンの生成. 『認知科学』, 10 (2), 193–206.
- 井上 和子・原田 かつ子・阿部 泰明 (1999). 『生成言語学入門』. 東京: 大修館書店.

## 付 録

### A. 実験1において使用された言語

本モデルにおいては, Gibson and Wexler (1994) が提案し, Niyogi (1996) のモデルで用いられた言語(表 1)を使用した. これら8つの言語を導出する文法は, 3つのパラメータから求められる. 3つの

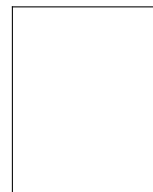
うち, 2つが X-バー理論に関するものであり, 句構造規則における Specifier-Head 関係および Head-Complement 関係に相当する. 以下のプロダクションルールは, これら両パラメータと一致する:

$$\begin{aligned} XP &\rightarrow \text{Spec } X' (p_1 = 0) \\ &\quad \text{or } X' \text{ Spec } (p_1 = 1), \\ X' &\rightarrow \text{Comp } X' (p_2 = 0) \\ &\quad \text{or } X' \text{ Comp } (p_2 = 1), \\ X' &\rightarrow X. \end{aligned}$$

3番目のパラメータは動詞の移動(Verb Movement)に関するものである. これはドイツ語やオランダ語等の平叙文にみられる, 動詞が文中の2番目の位置に移動するか否かを指定するパラメータである. この, 動詞が常に2番目に移動するルールは, 世界中の言語の中で現れるものと現れないものがあり, この多様性を V2 パラメータとして捉えている. 上記3つのパラメータによって得られた, 各文法ごとの埋め込みがない(Degree-0)言語( $L(G_1), \dots, L(G_8)$ )の一覧を表 1に示す.

(1994年5月16日受付)

(1994年5月16日採録)



中村 誠(正会員)

1972年生. 1995年九州工業大学情報工学部知能情報工学科卒業. 1997年北陸先端科学技術大学院大学情報科学研究科博士前期課程修了. 同年三洋電機(株)入社. 2004年北陸先端科学技術大学院大学情報科学研究科博士後期課程修了. 現在北陸先端科学技術大学院大学情報科学研究科助手. 博士(情報科学). 自然言語処理, 進化言語学などの研究に従事.

表1 言語群 (Gibson and Wexler (1994), Niyogi (1996)から引用)

Table 1 The Languages (cited from Gibson and Wexler (1994), Niyogi (1996))

Language	Spec	Comp	V2	Degree-0 unembedded sentences
$L(G_1)$	1	1	0	“V S” “V O S” “V O1 O2 S” “Aux V S” “Aux V O S” “Aux V O1 O2 S” “Adv V S” “Adv V O S” “Adv V O1 O2 S” “Adv Aux V S” “Adv Aux V O S” “Adv Aux V O1 O2 S”
$L(G_2)$	1	1	1	“S V” “S V O” “O V S” “S V O1 O2” “O1 V O2 S” “O2 V O1 S” “S Aux V” “S Aux V O” “O Aux V S” “S Aux V O1 O2” “O1 Aux V O2 S” “O2 Aux V O1 S” “Adv V S” “Adv V O S” “Adv V O1 O2 S” “Adv Aux V S” “Adv Aux V O S” “Adv Aux V O1 O2 S”
$L(G_3)$	1	0	0	“V S” “O V S” “O2 O1 V S” “V Aux S” “O V Aux S” “O2 O1 V Aux S” “Adv V S” “Adv O V S” “Adv O2 O1 V S” “Adv V Aux S” “Adv O V Aux S” “Adv O2 O1 V Aux S”
$L(G_4)$	1	0	1	“S V” “O V S” “S V O” “S V O2 O1” “O1 V O2 S” “O2 V O1 S” “S Aux V” “S Aux O V” “O Aux V S” “S Aux O2 O1 V” “O1 Aux O2 V S” “O2 Aux O1 V S” “Adv V S” “Adv V O S” “Adv V O2 O1 S” “Adv Aux V S” “Adv Aux O V S” “Adv Aux O2 O1 V S”
$L(G_5)$ (English, French)	0	1	0	“S V” “S V O” “S V O1 O2” “S Aux V” “S Aux V O” “S Aux V O1 O2” “Adv S V” “Adv S V O” “Adv S V O1 O2” “Adv S Aux V” “Adv S Aux V O” “Adv S Aux V O1 O2”
$L(G_6)$	0	1	1	“S V” “S V O” “O V S” “S V O1 O2” “O1 V S O2” “O2 V S O1” “S Aux V” “S Aux V O” “O Aux S V” “S Aux V O1 O2” “O1 Aux S V O2” “O2 Aux S V O1” “Adv V S” “Adv V S O” “Adv V S O1 O2” “Adv Aux S V” “Adv Aux S V O” “Adv Aux S V O1 O2”
$L(G_7)$ (Bengali, Hindi)	0	0	0	“S V” “S O V” “S O2 O1 V” “S V Aux” “S O2 O1 V Aux” “Adv S V” “Adv S O V” “S O V Aux” “Adv S O2 O1 V” “Adv S V Aux” “Adv S O V Aux” “Adv S O2 O1 V Aux”
$L(G_8)$ (German, Dutch)	0	0	1	“S V” “S V O” “O V S” “S V O2 O1” “O1 V S O2” “O2 V S O1” “S Aux V” “S Aux O V” “O Aux S V” “O1 Aux S O2 V” “O2 Aux S O1 V” “Adv V S” “Adv V S O” “Adv V S O2 O1” “Adv Aux S V” “Adv Aux S O V” “S Aux O2 O1 V” “Adv Aux S O2 O1 V”

## 橋本 敬(正会員)

1967年生。1990年神戸大学理学部物理学科卒業。1996年東京大学大学院総合文化研究科広域システム専攻修了。同年理化学研究所基礎科学特別研究員。1999年北陸先端科学技術大学院大学知識科学研究科助教授。1998年 Sony Computer Science Laboratory-Paris 客員研究員。2001-2002年イギリス, エディンバラ大学客員研究員。学術博士。複雑系, 人工生命, 進化言語学などの研究に従事。日本物理学会, 数理生物学会, 国際人工生命学会各会員。

## 東条 敏(正会員)

1981年東京大学工学部計数工学科卒業。1983年東京大学大学院工学系研究科修了。同年(株)三菱総合研究所入社。1986-1988年米国カーネギーメロン大学機械翻訳センター客員研究員。1995年北陸先端科学技術大学院大学情報科学研究科助教授。2000年同教授。1997-1998年ドイツ, シュトゥットガルト大学客員研究員。博士(工学)。自然言語の意味と形式化, マルチエージェントの研究に従事。情報処理学会, 日本ソフトウェア科学会, 言語処理学会, ACL, FoLLi 各会員。