

# 空間構造を導入した言語動力学によるクレオール化のシミュレーション ～言語入力量によるコミュニティ形成の差異について～

中村 誠<sup>†</sup>, 橋本 敬<sup>††</sup>, 東条 敏<sup>†</sup>

北陸先端科学技術大学院大学 { <sup>†</sup> 情報, <sup>††</sup> 知識 } 科学研究科

## 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 [1]. 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。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といわれている。

本研究の目的は、クレオールが発現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シミュレーションによって導出し、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。本稿においては、言語動力学によってモデル化されたこれまでの研究成果 [2] を踏まえ、空間構造を新たに導入し、そこに複数の学習エージェントを配置する。これにより、言語学習者の単位でクレオール化の過程を観察することが可能となり、個々の言語獲得環境とクレオール化の関係を調べることができる。また、空間構造を導入したモデル（空間ありモデル）と既存モデル（空間なしモデル）による振る舞いを比較する。

## 2 言語動力学と動的遷移行列

### 2.1 クレオール化を扱う言語動力学

Nowak et al. [3] の言語動力学 (Language Dynamics Equation) では、普遍文法を仮定することにより、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。ある言語集団において文法  $G_i$  を持つ言語話者の人口比率を  $x_i$  とすると ( $\sum_{i=1}^n x_i = 1$ )、 $x_i$  の時間変化は動力学系として表わされる。また、ここでは言語は抽象化され、以下の性質を持つ：

類似性 言語間の類似度を表わす。任意の2つの言

語間の類似度を示す行列  $S = \{s_{ij}\}$  で表わされる。要素  $s_{ij}$  は、 $G_i$  話者がランダムに文を発話したときの文法  $G_j$  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 言語間の遷移確率を表わす。 $G_i$  話者である親の子供が文法  $G_j$  を獲得する確率は行列  $Q = \{q_{ij}\}$  で表わされる。

我々はこれまでに、この言語動力学を修正し、言語の学習環境に関するパラメータを追加した動的遷移行列モデルを提案した [2]。Nowak et al. [3] のモデルが子供は親からしか言語を学習しなかったのに対して、本モデルでは集団内の親以外からも言語入力を受けようになっている。ここで、子供の言語獲得期において親以外の言語話者から言語入力を受け取る比率を導入し、これを *Exposure ratio*  $\alpha$  と呼ぶ。

動的遷移行列モデルは次の式で表わされる：

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

Nowak et al. [3] のモデルとの違いは次の通りである：(1) クレオールを話すことが人口増加に直接影響を及ぼすことがないことから、集団内において言語が通じる尺度を示す適応度関数を排除した。(2) 定数であった遷移確率  $Q = \{q_{ij}\}$  を集団内の人口構成比に依存し時間変化する関数  $\bar{Q}(t) = \{\bar{q}_{ij}(t)\}$  とした。この変化率を調整するパラメータが *Exposure ratio*  $\alpha$  である。

### 2.2 動的遷移行列

子供は、学習期間中に受け取った入力文を最も多く受取ることができる文法を  $\{G_1, \dots, G_n\}$  の中から選択する。本稿においては、学習アルゴリズムに Batch Learning Algorithm を用いている。親の文法、*Exposure ratio*  $\alpha$ 、言語間の類似度  $S = \{s_{ij}\}$ 、言語入力量  $w$  がわかれば、子供がどの文法を獲得するのかを確率分布で導出できる。なお、この確率行列の導出方法に関しては、Nakamura et al. [4] に詳述されている。

本稿においては、特に言語入力量  $w$  に注目する。これは、言語獲得期にある子供が受け取る言語入力のうち、特にその文法であることの証拠となり得るものの量を表している。すなわち、これは言語の獲得可能性を決定するパラメータとして考えられる。これをクレオールが発現する状況と照らし合わせてみると、言語入力量  $w$  が少ないときは、ピジンなどの単純な文の入力が多く、獲得する文法を特定する情報が少ない状況、また  $w$  が多いときは、クレオールが文法的に発達し、他の言語と同様に文法的に複雑な文を多く受け取る状況と考えることができる。

### 2.3 空間構造の導入

これまでの言語動力学では、全学習者が全員と同様に相互作用するモデルになっていた。ここで、空間構造を導入し、個々の学習者はローカルに相互作用するようにする。すなわち、その空間内で発話を行うエージェントが配置され、各エージェントの子供は、同じ場所に配置された親と近隣のエージェントに影響を受けながら学習を行う。

空間ありモデルの動的遷移行列  $\bar{Q}$  は、それぞれの学習者エージェントについて、近隣の言語分布から計算する。したがって、2節で定義した人口比率  $x_i$  を「位置  $l$  にいるエージェントの近隣にいる  $G_i$  話者の人口比率  $x_i^{(l)}$ 」とし、それぞれの学習者エージェントについて遷移確率を計算することで、動的遷移行列  $\bar{Q}^{(l)}(t)$  を導出する。

### 3 人口動力学上のクレオールの定義

本稿においては、言語学習者が獲得する文法  $\{G_1, \dots, G_m\}$  にクレオールが含まれていると考え、それを  $G_c$  とする。

ここで我々は、これまでの言語学的な定義 [1] とは大きく異なるが、人口動力学の視点からみたクレオールの定義をする。クレオールとは、新しい言語の発現である。すなわち、全体の言語分布を見たとき、初期に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、優勢クレオールとは、次のような言語  $L(G_c)$  であると定義される：

$$x_c(0) = 0, \quad x_c(t_s) > \theta_d. \quad (2)$$

ここで  $x_c(t_s)$  は人口動態が収束し、安定した  $t_s$  世代における  $G_c$  話者の人口比率、 $\theta_d$  は優勢 (dominant) 言語であるとみなすための人口比率の閾値を示し、本稿では  $\theta_d = 0.9$  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、最終的にはほとんどの人に使用される状態を表わしている。

### 4 実験とまとめ

クレオールが発現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発現する可能性がある環境である。3つの言語の初期の人口配分を  $(x_1, x_2, x_3) = (0.5, 0.5, 0)$  とすることにより、 $G_1$  と  $G_2$  を既存言語、 $G_3$  をクレオールとみなし、これが優勢言語となるための条件を考える。ここでは  $S$  行列を次のような対称行列とした：

$$S = \begin{pmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{pmatrix} \quad (3)$$

要素  $a$  は既存の2言語間の類似度であり、 $b, c$  はそれぞれ  $G_1$  とクレオール、 $G_2$  とクレオールの間の類似度を表わしている。全ての実験について、この類似行列の値を  $(a, b, c) = (0, 0.3, 0.4)$ 、Exposure ratio を  $\alpha = 0.7$  とした。以降、言語入力量  $w$  をパラメータとし、動力学の振る舞いを観察する。

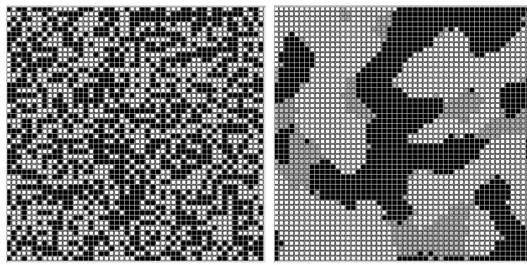
空間構造として、 $50 \times 50$  のグリッド上にエージェントを配置した。学習者エージェントは、そのマス目を囲む8つのエージェントの発話を聞いて、文法を獲得する。

#### 4.1 クレオール化の例

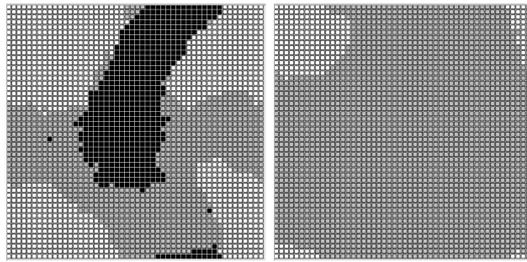
空間構造を持つダイナミクスの例を図1に示す。(a) 初期状態において、 $G_1$  と  $G_2$  話者だけがランダムに配置される。(b) 同一言語話者の局所的な集団 (以降、コミュニティと呼ぶ) が形成され始める。特に、複数の小さなクレオールコミュニティが早期に現れる。(c) 3言語の大集団が共存する準安定状態となる。(d) この試行では1552世代で  $G_3$  の人口比率が0.9を超え、クレオールが優勢言語となった。

4節冒頭で類似行列  $S$  を定義したとおり、クレオールは  $G_1, G_2$  のどちらの言語にも受け入れられる、最も効率のよい言語である。したがって、 $G_1$  と  $G_2$  話者に囲まれた学習者はクレオールを獲得しやすくなる。図1(b)をみると、クレオール話者は  $G_1$  話者グループと  $G_2$  話者グループの境界上にいくつか現れているのがわかる。

学習者の振る舞い、特に近隣のエージェントとのコミュニティの形成は、Exposure ratio  $\alpha$  や言



(a) 第 0 世代 (b) 第 50 世代



(c) 第 500 世代 (d) 第 1552 世代

図 1: 空間構造を持つダイナミクスの例 (white: $G_1$ , black: $G_2$ , gray: $G_3$ ;  $w = 10$ )

語入力量  $w$  によって変化する。Exposure ratio  $\alpha$  が大きいほど学習者は親よりも近隣のエージェントに影響を受けて言語を獲得するため、同じ言語を話すコミュニティを形成しやすくなる。また、語入力量  $w$  が多いほど遷移確率の分布が偏り、その学習者にとって最も効率のよい言語を選びやすくなる。つまり、 $\alpha$  が小さいと、親の言語をそのまま継承し、初期状態の分布から進行しにくいのに対し、 $\alpha$  が十分大きければコミュニティを形成しやすくなる。対照的に、語入力量  $w$  が少ないと、確率による選択のばらつきによって、世代間の言語の変化が大きい。結果として、大規模なコミュニティは形成されにくく、全ての言語が共存状態になる。

コミュニティの中にいる学習者は、近隣エージェントが親と同じ言語を発話するため、その言語を継承しやすくなる。したがって、一旦大きなコミュニティができると、その言語は消滅しにくい傾向にある。クレオールが優勢言語になるためには、早期の段階で大きなコミュニティを形成する必要がある。

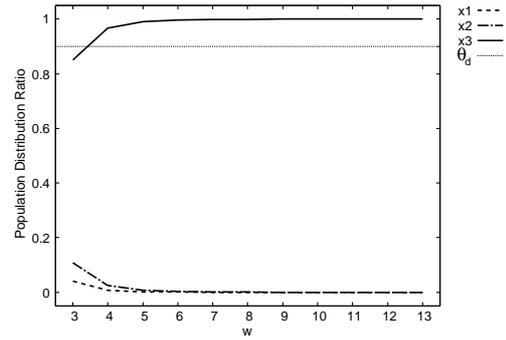


図 2: 空間なしモデルの言語入力量に対する収束世代の人口比率の変化

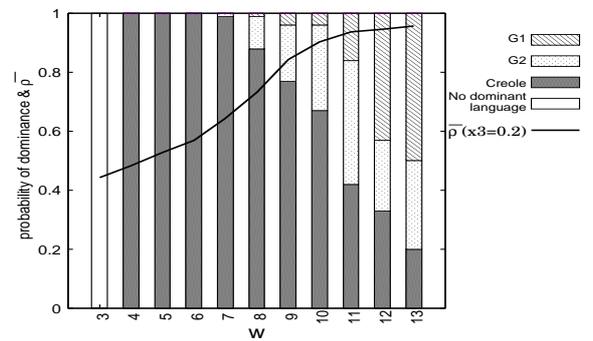


図 3: 空間ありモデルの言語入力量に対する優勢言語になる確率と  $\bar{p}$  の変化

#### 4.2 言語入力量とクレオール化

空間構造がないモデルは、(1) 式の微分方程式で表されるため、決定論的に解を求めることができる。言語入力量  $w$  に対する各言語話者の人口比率の変化を図 2 に示す。言語入力が増え、収束状態における優勢言語の人口比率は高くなる。言語入力量が  $w = 3$  のとき、 $G_3$  話者の人口比率は閾値  $\theta_d = 0.9$  を割り込んでいるため、クレオール化とみなされない。

次に、図 2 に対応する空間構造のモデルの結果を図 3 に示す。空間構造つきのモデルは、各言語学習者が動的遷移行列  $\bar{Q}$  を基に確率的に言語を獲得するため、その振る舞いは試行ごとに異なる。したがって、それぞれの  $w$  に対して 100 回ずつ試行を行い、各言語が優勢言語となった回数を表示している。なお、1 回の試行につき最大 1,000,000 世代まで追跡した。

図 2 と図 3 は、収束世代における人口比率と優勢言語になった割合という違いはあるものの、ク

レオール化の結果について注目すると、 $w$  が小さいうちは結果が類似していることがわかる。特に  $w = 3$  のとき、空間なしモデルではクレオール化とみなす閾値  $\theta_d$  を割り込んでいるのと同様に、空間ありモデルでも優勢言語が現れた試行がないことを示している。それに対し、 $w$  が大きくなるにつれて、2 つの結果に差が生じているのがわかる。空間なしモデルでは、クレオールの人口比率が  $x_3 \approx 1$  で収束しているのに対し、空間ありモデルでは、クレオールが優勢言語になった回数は徐々に減り、 $w \geq 12$  では、優勢言語になる割合は  $G_1$  のほうが高い。

4.1 節で述べたように、空間ありモデルの場合、早期のクレオールコミュニティの形成がクレオール化に求められる。学習期間中の言語入力が少ないと、近隣との同一言語コミュニティが形成されにくいいため、空間なしモデルとの差はあまり見られない。

ここで、早期のコミュニティ形成とクレオール化の関係を定量的に示すことを考える。空間構造における言語コミュニティ形成の定量的な評価をするために、境界密度  $\bar{\rho}$  を用いる [5]。これは、隣り合ったノードの状態が自分と異なっている割合の平均を表している。最大値  $\bar{\rho} = 1$  は、全てのエージェントが同じ言語を話している状態を表す。したがって、早期世代において  $\bar{\rho}$  を測定することで、早期のコミュニティ形成と最終的なクレオール化の関係を調べることができる。なお、コミュニティ形成の過程は試行ごとに異なるため、本稿においてはクレオール話者の人口比率  $x_3 = 0.2$  を初めて超えた世代を早期世代とした。図 3 の実線は、それぞれの試行における早期世代の  $\bar{\rho}$  の平均を表している。言語入力量  $w$  が増加するほど近隣と同じ言語を話す割合  $\bar{\rho}$  の値も増加しており、すなわち大規模な言語コミュニティが形成されていることを示している。

ピジンのように、その文法であることの証拠となるような言語入力が少ない状況では、学習者が獲得する言語が安定せず、コミュニティが形成されにくい。そのため、空間構造がないモデルとの差は生じにくい。それに対して言語入力量が多くなると、周りに同調しやすくなりコミュニティが形成されやすくなる。一旦大きな言語コミュニティができると消失しにくくなり、どの言語が生き残るのかは確率のゆらぎによって決まる。ここで、空間なしモデルとの差が生じると考えることができる。

## 5 まとめ

本稿では、これまでの言語動力学に空間構造を導入し、クレオール形成に関する両者の違いを比較した。実験においては、学習期間中の言語入力量、すなわち文法を獲得するために必要な証拠となる情報の多さによって 2 つのモデルの間に顕著な差が現れた。複数の言語が共存する準安定状態では、言語ごとのコミュニティが形成される状態と、コミュニティが形成されない完全な共存状態に大別され、それは学習時の言語入力量によって決定される。特に、言語入力が多い条件では、コミュニティが形成され、空間なしモデルとは異なる振り舞いをするのが観察された。空間構造を用いることの利点として、コミュニティの形成過程を可視化したことが挙げられる。その結果、クレオール化の条件としてコミュニティの形成に関する条件が存在し、それは言語入力量に依存することが確認された。今後の目標としては、エージェント間のネットワーク構造も考慮に入れ、実際のクレオールとの比較をしていきたい。

## 参考文献

- [1] Arends, J., Muysken, P. and Smith, N.(eds.): *Pidgins and Creoles*, John Benjamins Publishing Co., Amsterdam (1994).
- [2] 中村誠, 橋本敬, 東条敏: 言語動力学におけるクレオールの創発, *認知科学*, Vol. 11, No. 3, pp. 282–298 (2004).
- [3] Nowak, M. A., Komarova, N. L. and Niyogi, P.: Evolution of universal grammar, *Science*, Vol. 291, pp. 114–118 (2001).
- [4] Nakamura, M., Hashimoto, T. and Tojo, S.: Exposure Dependent Creolization in Language Dynamics Equation, *New Frontiers in Artificial Intelligence* (Sakurai, A., Hasida, K. and Nitta, K.(eds.)), Lecture Notes in Artificial Intelligence, Vol. 3609, Springer, pp. 295–304 (2006).
- [5] Miguel, M. S., Eguiluz, V. M., Toral, R. and Klemm, K.: Binary and Multivariate Stochastic Models of Consensus Formation, *Computing in Science and Engg.*, Vol. 7, No. 6, pp. 67–73 (2005).