

## クレオール化をシミュレートする

中村 誠なかむら まこと

複数の言語が混在するコミュニティでクレオールが生まれる過程をシミュレートし、言語進化に関する普遍的な法則を導き出せないか。言語動力学に基づくモデルを使ったシミュレーションの原理と方法を説明し、その結果が何を意味するかを探る。

言語接触とクレオールに関する記述は本特集の他の先生方にお任せするとして、ここではクレオール化のシミュレーションについて考えてみよう。話を進める前に、話すべき項目をここで整理しておく。まず、クレオール化の過程をシミュレートすることのメリットとは何なのか、また、言語や言語変化というものを計算機上でどのように捉え、言語接触やクレオール化をどのようにモデル化するのか。これらを順に説明した上で、シミュレーションの結果を考察していきたい。

## 一 進化言語学とクレオール

クレオール化をシミュレートするという発想は、進化言語学の考え方から来ている。進化言語学とは、生物進化による人間の能力の向上とともに、言語がどのように出現し、発展したのかを問う学術分野である。言語の起源と進化は生物進化、学習、文化進化が相互作用する複雑な問題であるため、その研究領域は多くの学問の接点にあり、分野横断的である。そのなかでも、人工生命や人工知能といった分野の応用として、計算機を用いた手法、特に構成的アプローチが新し

い試みとして注目を浴びている。構成的アプローチを一言で表現すると、対象を「作って動かす」ことにより理解しようとする方法論である。例えば、人間の言語獲得のメカニズムは未解明の問題であるが、信憑性の高い仮説は存在する。これを基に簡単なモデル化を行うことで、その動作を検証しようという試みである。このとき、複数の言語学習者や言語獲得環境との相互作用により、そのメカニズムを分析するだけでは得られない、系全体としての振る舞いを観察することが可能となる。実験の結果を現実世界の事象と比較して、そのメカニズムの妥当性を示すことができる。計算機を用いた言語進化の研究に関しては、参考文献「1、2」に詳述されている。

クレオール化は観察可能な言語進化現象として進化言語学で注目されており、また、クレオール研究からみても、これをシミュレートすることのメリットは大きい。なぜなら、これまでの言語学的な側面から行われてきたヒシンやクレオールの調査においては、これらが出現する一般的な条件を実際のクレオールから求めることは現実的に不可能であったが、この構成的アプローチによって、それが可能となるからである。すなわち、クレオール化に関して何らかの仮説を立て、

言語と言語話者、複数言語が話されるコミュニティをモデル化し、求めたい条件をパラメータ化することで、クレオール化の条件を求めることができるのである。

ここでは「複数言語が混在するコミュニティでクレオールが発生する」という仮説を立てて、多言語コミュニティをモデル化してみよう。以降、モデル化するにあたって必要となる、構成要素について述べる。

## 二 言語と言語変化

言語の抽象化の方法は仮説に基づいて考えられる。ここでは、言語の話者数の変化という巨視的な観点から言語進化を数理的にモデル化する方法を考えてみよう。

まず、言語的制約として、普遍文法、特に原理とパラメータ理論を導入する。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であるとする考え方である。すなわち、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

を  $(G_1, \dots, G_n)$  としてあらかじめ定義する。すると言語の変化は、言語話者が所有するパラメータの変化として示され、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  $x_i$ 、すなわち  $M_i = \frac{G_i}{M}$  であるとするとき、 $t$  の時間変化は動力学系として表わされる。したがって、言語と言語変化を次のように定式化する。

**言語** 言語は他の言語間の相対的な類似性によって与えられる。すなわち、語彙や統語規則といった言語としての具体的な定義はなく、言語の集合は任意の二つの言語間の類似度を示す確率行列  $S = (s_{ij})$  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。

**言語変化** 言語の変化は、言語間における言語話者の遷移として定義される。したがって、言語の変化率は言語間の遷移確率によって表わされる。このとき、 $G_i$  話者である親の子供が文法  $G_j$  を獲得する確率は行列の  $(s_{ij} - \delta_{ij})/s_{ij}$  で表わされる。この行列を動的遷移行列

下、図1を参照しながら説明していく。

- 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、この図はある世代  $t$  における各言語話者の人口構成比を表す。
- ② 言語話者はそれぞれ子供を産む。このとき人口の増減はない。産まれた子供たちは周りの大人から言葉を聞き、文法を推定する。その結果、 $G_j$  話者の子供は  $m_{ij}(t)$  の確率で正常に  $G_j$  を獲得し、その全体に対する人口比率は  $q_{ij}(t)x_j(t)$  である。その他の子供は  $m_{ij}(t)$  の確率で  $G_i$  を身につける。
- ③ 次世代の  $G_i$  話者の人口比率  $x_i(t+1)$  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  $G_i$  話者の人口比は  $M_i^{-1} x_i(t+1)$  である。

通常、微分方程式で表される連続時間の人口力学モデルでは、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。ここで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に注意していただきたい。

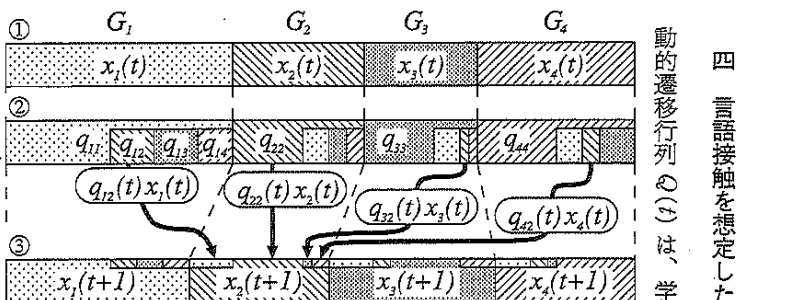


図1：人口変化の流れ

#### 四 言語接触を想定した学習アルゴリズム

と呼ぶ。遷移確率は学習アルゴリズムに対応し、学習アルゴリズムには  $S$  行列や言語話者の人口比率などがパラメータとして含まれる。言語話者の人口比率は時間変化するため、動的遷移行列も時間変化する関数である。

三 言語動力学

ノヴァツクラ「3」は言語獲得に関する理論を発展させることを目的として、数理生態学的な人口力学を応用したモデルを提案している。これを言語動力学 (Language Dynamic 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なった。そこでは大人を言語話者、子供を言語学習者とし、子供は親から言葉を聞いて文法を推定する。ノヴァツクラのモデルは、生物進化を考慮にいたった長いタイムスケールを対象にしており、また言語接触の度合いを考慮していないことから、多言語コミュニティに対応したモデルに修正する。言語動力学モデルは次の式のようになる。

$$\frac{dx_i(t)}{dt} = \sum_j q_{ij}(t)x_j(t) - x_i(t)$$

この式は次のような状況を描写していると解釈される。以

動的遷移行列の  $(s_{ij})$  は、学習アルゴリズムを数式に置き換えたものである。子供の言語獲得期において、さまざまな言語話者と接触する環境と、それに影響を受けて文法を獲得するような学習アルゴリズムを考えてみよう。なお、学習アルゴリズムから数式に展開する手順は割愛させていただきます。

まず、複数の言語が話されるコミュニティにおいて、子供が母語とそれ以外の言語と接触する比率を定義しよう。すなわち、他の言語と比べて親が話す言語に子供がどれ

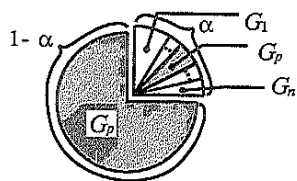


図2：接触比率  $\alpha$

と接触する比率はそれぞれの言語話者の人口比に比例すると仮定することにより、この  $\alpha$  から子供が接触するそれぞれの言語の比率を導き出すことができる。ある子供が接触する言語の分布を図2に示す。親が  $G_p$  話者であるこの子供は、影がかかった部分の割合で言語  $G_p$  と接することとなる。

次に接触比率  $\alpha$  を取り込んだ学習アルゴリズムを考える。本モデルは普通文法を仮定している。すなわち、人間が話すことができる文法の数は有限であり、言語学習とはその中から獲得するのに最もふさわしいと考えられるものをひとつ選択することに過ぎない。次にあげるアルゴリズムがこの学習のひとつであると考えられる。

1 子供の頭の中に、それぞれの文法ごとにスコア表があ

ると考える。

- 2 子供は大人が発話した文を受け取る。
- 3 全ての文法につき、その文が受理可能かどうか識別し、その文が受理できた文法のスコアに1ポイント加算する。
- 4 子供が言語獲得に十分な量と考えられる入力文を受け取るまで ( $w$  文) 2と3を繰り返す。
- 5 その子供は最も高いスコアを獲得した文法を採用する。

各言語に対する話者の人口比率と接触比率  $\alpha$  により、子供が接触する言語の比率が求まる。また、類似行列  $S$  は、それぞれの文法について入力文の受理能力を決定するものである。図3は  $G_2$  話者の子供が学習の結果  $G_2$  を獲得した例である。学習期間において、子供は、図中1から10と番号がふられた文を受け取る。入力文は接触比率  $\alpha$  に従い、二つのカテゴリに振り分けられる。ひとつは親以外の言語話者が発話した文、もう一つは親が発話したものである。前者において、それぞれの文は特定の文法を持った言語話者から発話されたものであるが、どの言語話者の発話であるかはそれぞれの言語話者の人口比率によって決定される。図3の例は、子供は

文1、4、5を  $G_1$  話者から、文3を  $G_3$  話者から聞いていることを表わしている。その他の文は全て  $G_2$  話者からである。言語学習者である子供は、入力文のうち、それぞれの文法で受理できる文の数をカウントし、最も多くの文を受理した文法を採用する。例えば、文1は  $G_1$  話者が発話したものであるが、 $G_2$  でも受理することができると、 $G_1$  と  $G_3$  のカウン

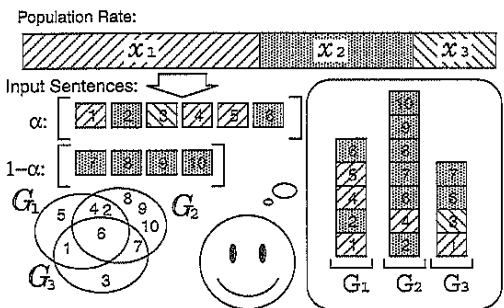


図3：学習アルゴリズム

タにそれぞれ1を加える。どの文がどの文法で受理することができるといことが図左下のベン図で示されており、 $S$  行列はこの関係を言語間の類似度として表わしたものである。

以上が多言語コミュニケーションのモデルである。パラメータとして言語間の類似行列  $S$ 、接触比率  $\alpha$ 、言語獲得期の言語入力量  $w$ 、および言語話者

の人口比の初期値  $(x_1(0), \dots, x_n(0))$  を与えることで、言語変化の様子を観察することができる。

### 五 言語動力学上のクレオール定義

言語動力学上でクレオールはどのように定義されるのだろうか。このモデルでは、原理とパラメータを仮定しており、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  $(G_1, \dots, G_n, \dots, G_n)$  となる。ここで、これまでの言語学的な定義とは大きく異なるが、言語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、あるコミュニケーションで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  $G_c$  であると定義される。

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

ここで  $\theta_c(t)$  は人口動態が収束し、安定した世代における  $G_c$  話者の人口比率、 $\theta_c$  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、ここでは  $\theta_c = 0.9$  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、最終的にはそのコミュニティでほとんどの人に使用される状態を表わしている。

### 六 実験と考察

クレオールが出現しうる最小システムとして、三つの言語 ( $G_1, G_2, G_3$ ) があり、そのうち二言語 ( $G_1, G_2$ ) の話者だけが初期に存在する状態からの時間変化をシミュレートした。すなわち、 $G_3$  がクレオールの候補である。このとき、言語間の類似性は  $3 \times 3$  行列で表される。既存の二言語間の類似度を  $a$  ( $\parallel S_{11} \parallel S_{21}$ ) とし、 $G_3$  とクレオール間、 $G_3$  とクレオール間の類似度をそれぞれ  $b$  ( $\parallel S_{13} \parallel S_{23}$ )、 $c$  ( $\parallel S_{31} \parallel S_{32}$ ) とした。また、対角要素はすべて1である。

言語動力学の振る舞いを図4に示す。これは、時間の変化に対する各言語話者の人口比率の変化を表している。各パラメータの値は、S行列の要素 ( $a, b, c$ ) ( $0.3, 0.2$ )、言語入力量  $\kappa=3$ 、接触比率  $\alpha=0.7$  とした。

接触比率  $\alpha$  を  $0.7$  と高い値に設定したことで、複数言語が混在するコミュニティを表している。その結果、初期状態 (7.0) では誰も話していなかった言語  $G_3$  が、安定状態にお

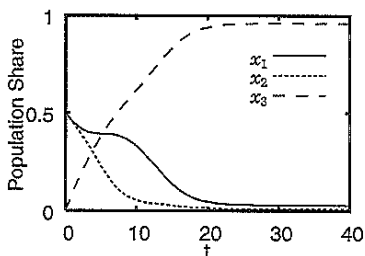


図4：クレオール化の例

触の度合いに応じてクレオールの発生を示したといえる。次はクレオールが出現する条件を求める。ここでは、言語間の類似度の各要素 ( $a, b, c$ ) をパラメータとして、クレオールが発生した領域を図5に示す。図5上の格子面の内側でクレオール化が観察された。下はその等高線図である。

既存言語の類似度が高いとコミュニケーションが取りやすいため、クレオールの必要性がない。したがって、既存の二言語間の類似度 ( $a$ ) が低いほど、クレオールが発生しやすい。また、既存の言語とクレオールとの類似度 ( $b, c$ ) は

各既存言語で同程度 ( $b, c$ ) であることが確認できる。一方の言語とクレオールが類似していると、その言語が優位になることが観察されている。

### 七 おわりに

ここに記述した内容は、北陸先端科学技術大学院大学情報科学研究科東条敏教授ならびに同大知識科学研究科橋本敬准教授との共同研究の成果である。我々は、このモデルを用いた実験によつて、ここで示したクレオール化における言語間の類似性に関する条件の推定に加え、言語獲得に必要な言語入力量について検証し、言語入力とクレオールの関係に

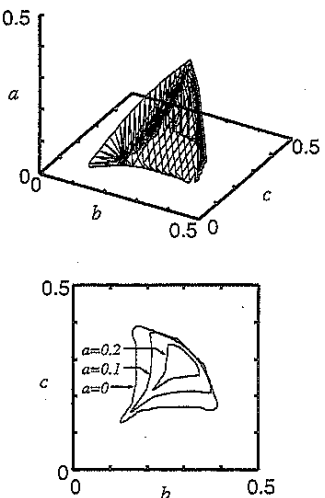


図5：クレオール化の類似性に関する条件

ついて定量的な分析を行っている [4, 5]。

言語動力学方程式は、解析的に解を求められる部分もあり、また、エージェントモデルに比べて長時間の振る舞いを簡単に調べられるという利点がある。しかし、各言語の特徴は類似度だけで表されるため、言語の具体的な特徴を言語動力学とどのように対応づけるかが今後の課題となる。

#### 【参考文献】

- [1] Briscoe, T. (ed.) (2002) *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, Cambridge University Press, Cambridge.
- [2] Lyon, C., Nehaniv, C. L. and Cangelosi, A. (eds.) (2007) *Emergence of Communication and Language*, Springer, London.
- [3] Nowak, M. A., Komarova, N. I. and Niyogi, P. (2001) Evolution of universal grammar, *Science*, Vol. 291, pp. 114-118.
- [4] 中本誠、橋本敬、東条敏 (1003) 「言語動力学におけるマンキーン問題」『認知科学』 Vol. 11, No. 3, pp. 282-298.
- [5] Nakamura, M., Hashimoto, T. and Tojo, S.: Simulation of Common Language Acquisition by Evolutionary Dynamics, *Proc. of 1st Intl Workshop on the Evolutionary Models of Collaboration*, pp. 21-26 (2007).

(北陸先端科学技術大学院大学／情報科学)