ピジン発生過程のマルチエージェントシミュレーション

中島 千智 †　内海 彰 †　中村 誠 ††
† 電気通信大学 電気通信学研究科　†† 北陸先端科学技術大学院大学 情報科学研究科

1　はじめに

社会言語学や発達心理学などの分野では，ピジンといわれる言語現象について色々な研究がなされている．ピジンとは共通の言語を持たない者同士がコミュニケーションをとる必要がある場合に発生する言語である．ピジン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，(1) 格助詞と活用の欠如，(2) 語彙の共有，(3) 語順が固定という共通した特徴がある [1]．よってピジンの発生過程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を知る上で重要である.

先行研究 [2] では，マルチエージェントの枠組みで 2 言語間での接触においてピジンが発生する過程のモデルを提案し，シミュレーションを行っている．しかし，実際のピジン発生時の環境（3 言語以上の話者が接触し，各言語話者の人数比と会話接触比が異なる）を正確にモデル化していないという問題がある．

そこで本研究では，これらの環境設定を反映させたピジン発生の計算モデルを提案し，その妥当性についてシミュレーション実験を通じて検証する．

2　ピジン発生過程の計算モデル

2.1　エージェント同士の会話

エージェントの会話を説明する前に文法と中間表現について説明する．本研究で用いる文法とは文を生成する際に用いる書き換え規則の集合であり，語彙規則も含まれる．また中間表現とは異言語間で会話が通じたかどうかを判断するために設けた共通の意味を表す表現で，格構造をもとに構成される．

エージェント同士の会話は，まず話し手がランダムに中間表現を発生させ，その中間表現と持っている文法を用いて文を生成し，聞き手にその文を発話する．聞き手は持っている文法を用いて受け取った文を構文解析して中間表現を導出し，話し手に返す．図 1 は，以上の過程を模式的に示しており，話し手が中間表現 $[^{*}ran(agt^{*}:I)]$ から生成した文を発話し，聞き手から中間表現を受け取る様子を示している．

先行研究 [2] では，各エージェントは他の全てのエージェントと同じ回数だけ会話を行うことを仮定しているため，各言語エージェントの人数比と会話の回数比は同じであった．しかし前述したように実際には人数比と会話頻度比は一致しない．そこで本研究では，L_i 言語エージェント群が L_j 言語エージェントと接触する確率 (接触率)ci_{L_i,L_j} という新たなパラメータを導入して，エージェントの人数比とは独立にエージェントの会話する人数比を変化させる．

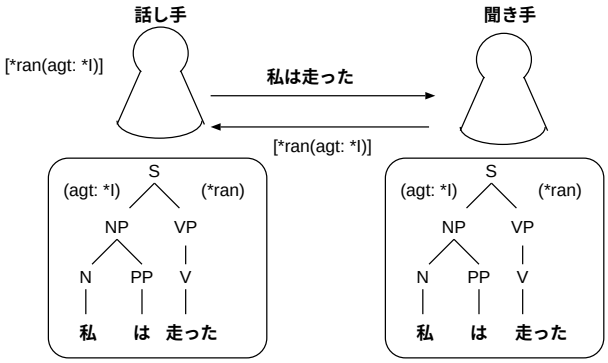

図 1: エージェントの発話例

2.2　GA によるエージェントの学習

このモデルでは，遺伝的アルゴリズム (GA) による学習を行い，会話で得られた得点を適応度と見ることによって，エージェントが持っているいくつかの文法候補から次の世代に残す文法を選ぶ操作を行う．

2.2.1　エージェントの文法

エージェント m は世代 L において文法 $G_{L,m}$ と複数の染色体を持っている．文法 $G_{L,m}$ は 1 つの染色体を用いて新しい文法 $G'_{L,m}$ を作り出すことができる．染色体は 3 つの遺伝子から成り，各遺伝子は文法規則の変換ルールに対応している．具体的には「何もしない」,「語彙を変換する」,「基本語順を変える」,「格助詞を消す」,「格助詞を加える」という変換ルールを用いる．

2.2.2　GA を用いた学習

本モデルでは前世代から受け継いだ 10 個の染色体に対して，交差や突然変異などの操作を施して 20 個の染色体を新たに生成し，それらのうち会話で得られた適応度の高い 10 個の染色体を次世代に残すという処理を行う．なお，以上の処理では元の文法 $G_{L,m}$ は変化しない．そこで一定の世代ごとに最も適応度が高かった染色体で変換した文法を次世代の文法 $G_{L,m}$ とする．

2.3　適応度計算

会話中に以下の手順で各染色体に対して，適応度を計算する．まずある話し手と聞き手が 1 回会話した場合に互いに得られる得点 $score$ は次式で定義される．

$$score = c_w + c_g + c_p \tag{1}$$

ここで c_w, c_g, c_p はそれぞれ，聞き手が話し手の発した文中の語彙のうちで理解できた数，聞き手と話し手の用いた文法中の共通する文法規則数，聞き手と話し手の中間表現が一致したときの得点である．次に同言語エージェントと会話したときの重要度 p_c と，異言語エージェントと会話したときの重要度 p_f を導入する．これらは実社会での言語間の力関係などを表すパラメータである．このパラメータを用いると，L_i 言語エージェントが L_j 言語を話すエージェントと会話したときに L_i 言語エージェントが得る得点 $score(L_i, L_j)$ は次式で与えられる．

$$score(L_i, L_j) = \begin{cases} score * p_{cL_j} & (L_i = L_j) \\ score * p_{fL_j} & (L_i \neq L_j) \end{cases} \tag{2}$$

3　実験 1：接触率を変化させた場合

3.1　シミュレーション対象とパラメータ設定

ハワイのプランテーションにおいて英語を母語としない労働者の間で用いられていた PPE（Plantation Pidgin English）は一般のピジンとは異なり，母語の特徴を残した英語基盤のピジンで，母語に応じて複数の種類が存在していた [3]．PPE が発生した背景として，一般的なピジンが発生する場合に比べて母語以外を使う必要がなく，異言語話者との接触する回数も少なかったことがあげられる．

そこで実験 1 では，日本語文法を習得している日本語エージェント群と英文法を習得している英語エージェント群を用いて接触率 $ci_{ee}(= ci_{jj})$ を変化させることにより，ピジンが 1 つに収束するのか，又は PPE のような異なるピジンに収束するかを観察した．ci_{ee} を高くし

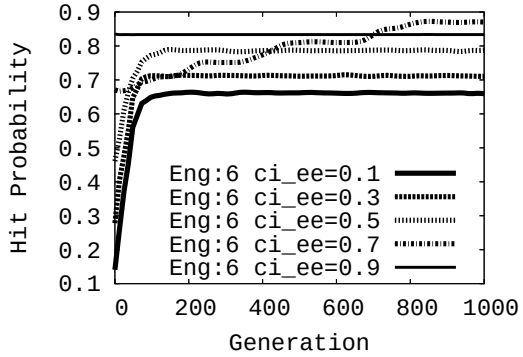

図 2: 実験 1 の世代ごとのヒット率の結果

表 1: 実験 1 の発話例

	英語エージェント	日本語エージェント
0 世代	I ran I saw him I went to Tokyo	私は走った 私は彼を見た 私は東京へ行った
$ci_{ee} = 0.1$ $ci_{ej} = 0.9$ 999 世代	私 ran 私 彼 見た 私 went Tokyo	
$ci_{ee} = 0.9$ $ci_{ej} = 0.1$ 999 世代	I ran I saw He I went to Tokyo	私 走った 私 彼 見た 私 東京 行った

て同言語との会話回数が増えることが，PPE が発生した状況と対応している．英語，日本語エージェント数をそれぞれ 6，6 に固定し，同言語エージェントの接触率 ci_{ee} を 0.1，0.3, 0.5, 0.7, 0.9 と変化させて行った．実験の実行世代は 1000 世代，文法の更新世代は 20 世代，重要度 p_c,p_f は全て 1 とした．なお，日本語と英語の文法間には格の指定に助詞をマーカとして用いるか，名詞句の表層位置と代名詞を活用すること，基本語順が英語は SVO，日本語が SOV という違いがある．

3.2 結果と考察

実験 1 の結果を図 2 と表 1 に示す．実験の評価指標としてヒット率を用いる．ヒット率はエージェントが正しく理解した会話の割合である．図 2 のグラフは 5 回のシミュレーションのヒット率の平均，縦軸がヒット率，横軸が GA の実行世代である．図 2 を見ると $ci_{ee} = 0.9$ 以外は世代を重ねるごとヒット率が上昇し，一定のところで収束している．これはエージェントが共通の文法を獲得したことを示している．表 1 は ci_{ee} が 0.1 と 0.9 の場合のエージェントの発話例である．$ci_{ee} = 0.1$ のとき英語，日本語エージェントは同一の文法を獲得しているが，$ci_{ee} = 0.9$ のときには英語，日本語エージェントで別々の文法を獲得している．これは ci_{ee} の値が高くなると異言語エージェントとの接触が少なくなるので，同言語エージェント内で独自のピジンを獲得したと考えられる．また語順は SVO，SOV のいずれかとなり，ci_{ee} が高いほど母語の影響を強く受ける結果となった．また発話に 1 節で述べたようなピジンの特徴が見られた．

先行研究では（1）人数比の高い集団に少ない集団が吸収され，1 つの共通文法を獲得するか，(2) 人数比が等しいために，共通文法が獲得できないという結果しか得られなかった．しかし本研究では接触率を導入することによって新たに（3）集団ごとで共通文法を獲得するという実際に観察される結果が得られるようになった．

4 実験 2：3 言語間のピジン発生

4.1 シミュレーション対象とパラメータ設定

ハワイで発生したハワイピジン [3] は英語話者の人口が他言語話者に比べて非常に少数であり，さらに他言語話者との接触が非常に限られた環境にも関わらず，英語の語彙を用いたピジンとなった例である．ハワイピジンは (a) 雇用者として権力を持っていた英語話者，(b) 現地語を話すハワイ語話者，(c) 人口は多い労働者の言語話者という 3 つの話者集団で発生したピジンである．

そこで実験 2 では，実験 1 で用いた日本語エージェント群と英語エージェント群，そして別の仮想的な文法を習得している仮想言語エージェント群をそれぞれ (a) を英語エージェント，(b) を仮想言語エージェント，(c) を日本語エージェントとしてハワイピジンの発生過程をシミュレートする．人数比をエージェント数，また話者間の力関係を重要度 p_f で表すことによってハワイピジンとの対応付けを行う．英語，仮想言語，日本語エージェント数をそれぞれ 3, 8, 10 と固定し，異言語と会話したときの重要度 p_f を 20, 4, 2 とした場合と，重み付けをしない場合について行う．実験の実行世代，文法の更新世代，同言語話者間の重要度 p_c は実験 1 と同じとした．なお実験 2 では接触率を考えず，エージェントは等しく会話するとした．仮想文法は格の指定に格助詞を用い，語順 VOS となる文法とする．

表 2: 実験 2 のエージェントの発話例

	英語	仮想言語	日本語
0 世代	I ran I saw him I went to Tokyo	laufe as ich sehe es er as ich fahre em Tokio as ich	私は走った 私は彼を見た 私は東京へ行った
999 世代 重み付け無	laufe は ich，　見た を er は ich， 行った へ 東京 は ich		
999 世代 重み付け有	私 laufe，　私 saw er，　私 went Tokyo		

表 3: 実験 2 の獲得された文法の影響

	重要度による重み付け無		重要度による重み付け有	
	語彙規則	文法規則	語彙規則	文法規則
英語	0 %	12 %	40 %	62 %
仮想言語	51 %	88 %	40 %	13 %
日本語	49 %	0 %	20 %	0 %
その他	0 %	0 %	0 %	25 %

4.2 結果と考察

実験 2 の結果を表 2，3 に示す．表 2 はエージェントの発話例である．表 2 より最終的に獲得された文法による発話は重要度による重み付けを用いた場合，重み付けしない場合に比べて，エージェント数が少ない英語エージェント群の元々の文法の影響を受けていることが分かる．表 3 はエージェントが獲得した文法規則が元々どの言語であったかの割合を示している．重み付けをしない場合にはエージェント数の多い言語の影響を受けているが，重み付けを用いることにより重要度が高くエージェント数の少ない言語の影響も受ける結果となっている．これは先行研究では見られなかった結果であり，観察された現象を再現しているといえる．

5 おわりに

本研究では，接触率を考慮するような 3 言語間のピジンの発生過程モデルを提案し，実際に観察された現象を計算機上で再現した．接触率を変化させた場合に先行研究では再現できなかった 2 つの独立するピジンが得られ，また 3 言語でのシミュレーションでは重要度を用いた重み付けにより獲得される文法が，人口比だけではなく実際の力関係にも依存するような結果が得られた．

今後の課題として，より実現象が再現できるようなモデルや学習方法を提案し，検討することがあげられる．

参考文献

[1] ロレト・トッド (著), 田中 幸子 (訳) :ピジン・クレオール入門, 大修館書店 (1986).

[2] 中村 誠，東条 敏: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, Vol. 10, No.2, pp.193-206 (2003).

[3] 亀井 孝:言語学大事典, 三省堂 (1996).