言語の創発と伝播

Self Re-organizing of Language triggered by Language Contact

前田 雄一郎† 佐々木 貴宏† 所 真理雄††
Yuichiro MAEDA Takahiro SASAKI Mario TOKORO

†慶応義塾大学大学院 理工学研究科 計算機科学専攻
††Sony IT 研究所 / Sony CSL

概 要

言語とは自己組織化の過程を経て創発し，絶えず変化する複雑なシステムである．こうした言語が持つ複雑性に関する研究は従来は遺伝的な時間スケールでの説明が多かったが言語の変動が歴史的な時間スケールで起こっていることは明らかである．一方，言語は外部に対して開かれた系であり，別の言語との接触を免れ得ない．このような言語の接触は言語の複雑性や動的な側面に大きな影響を及ぼす．本研究では，言語を歴史的な時間スケールで捉える“言語ゲーム”モデルを用いて言語の接触という現象について調べ，言語の動的な側面について考察する．

1 はじめに

言語とはエージェントの相互作用による自己組織化の過程を経て創発し，絶えず変化する複雑なシステムである．言語が持つ複雑性に関する研究は人工生命等の分野では遺伝的な時間スケールでの説明が多かったが言語の変動が歴史的な時間スケールで起こっていることは明らかである．一方，言語はその開放性ゆえに他の言語との接触を免れ得ないが，こうした言語の開放性は言語が持つ複雑性や多様性，流動性などといった動的な側面に影響を及ぼしている．

本研究では言語を構成する要素の内の一つである“語彙 (vocabulary) ”に着目して，最初に，エージェントが相互作用のモデルである言語ゲーム (Lanugage game) を通じて共通の語彙を発達させる過程を示す．本研究で用いる言語ゲームは Steels [2] [3] [4] のモデルを更に単純化したものである．このモデルで創発する言語は新たにエージェントやオブジェクトが付加されても適応し得るという点で開放性を有している [2]．次に，異なる言語を持つ異なるエージェント同士を接触させたときに言語がどのように適応するかを調べる．

2 言語の定義

エージェントの集合を $A = \{a_1, ..., a_m\}$，オブジェクトの集合を $O = \{o_1, ..., o_n\}$ とする．単語 w は，エージェント間で共有された有限のアルファベットの文字列で，本研究では“sa”，“ki”のように子音と母音の列として表す．辞書 $L \subset O \times W$ は，オブジェクト $o \in O$ と単語 $w \in W$ の間の関係を示す．この関係の各要素を 連想 と呼ぶ．ここで，W は単語の全ての集合を表す．各エージェント $a \in A$ はそれぞれ固有の辞書 L_a を一つ持つものとする．この辞書は最初は空で初期化する．

$u(\langle o, w\rangle, a)$ は，連想 $\langle o, w\rangle \in L_a$ が a によって用いられた回数である．$s(\langle o, w\rangle, a)$ は，以下で定義する言語ゲームで連想 $\langle o, w\rangle \in L_a$ を用いた際にゲームが成功した回数を示す．各連想には $m(\langle o, w\rangle, a) = s(\langle o, w\rangle, a)/u(\langle o, w\rangle, a)$ と定義するスコアが付いている．

3 言語ゲーム (Language game)

一回の言語ゲーム $g = \langle C, i, r, o\rangle$ は，コンテキスト (context) $C = \{o_j, ..., o_k\} \subseteq O$，イニシエー

タ (initiator) $i \in A$，リシピアント (recipient) $r \in A$，トピック $o \in C$ から成る．一連の言語ゲームの流れは以下のように構成されている．

1. あるコンテキスト C において，イニシエータ i は話題にするオブジェクトとなるトピック $o \in C$ を一つ選ぶ．
2. イニシエータはトピックを指す単語を辞書から探し，この単語を表現 e としてリシピアントに話す.
   $$speak(o, L_a) = \{e|e = \{w|\langle o, w\rangle \in L_a\}\}$$
3. リシピアントは表現 e を聞くことでこれが指すオブジェクトを辞書から探し，トピックになっているオブジェクトを H として予想する．
   $$hear(e, L_a) = \{H|H = \{o|\langle o, e\rangle \in L_a\}\}$$
4. リシピアントが予想したオブジェクトがコンテキストに含まれている場合，一連の言語ゲームは成功とみなす．

イニシエータが話す際には，表現 e は連想に伴うスコアがに応じた確率で選択することで一つに定める．ここではこの確率 p をボルツマン分布で以下のように与えた．

$$p = \exp((m(\langle o, w\rangle, a) - 1)/T)$$

なお，温度係数 T は 0.01 とした．

他の実験設定は以下のようになっている．

- イニシエータの辞書が不十分なとき： イニシエータの辞書にトピックを表現する単語 (連想) がない場合．このときは言語ゲームは失敗となり，イニシエータはある確率で新たに単語を創り出し (ここではこの確率を 0.10 とした)，これを新たな連想として辞書に付け加える．
- リシピアントの辞書が不十分なとき： リシピアントが聞いた表現 e に対応する単語が辞書にない場合．この場合，リシピアントは聞いた単語とコンテキスト中のオブジェクトを結びつけて，これを新たな連想として辞書に付け加える．ゲームは失敗となる．
- コミュニケーションの成立, 不成立： リシピアントが表現 e を聞き，その結果予想したオブジェクトの集合 H に応じて，更に以下の三つの場合に分けることができる．
  1. 成功： リシピアントが予想したオブジェクトがコンテキスト中に存在する場合，ゲームは成功．このときはイニシエータ，リシピアント双方の $s(\langle o, w\rangle, a)$ 及び $u(\langle o, w\rangle, a)$ を増加させる．
  2. 意味が異なる： リシピアントが予想したオブジェクトがコンテキスト中に一つも存在しない場合，ゲームは失敗．このときはイニシエータ，リシピアント双方が用いた連想の $u(\langle o, w\rangle, a)$ のみを増加させる．
  3. 曖昧： リシピアントが予想したオブジェクトが複数あり，その一部がコンテキスト中に存在する場合，コンテキストに存在するオブジェクトに関する連想の $s(\langle o, w\rangle, a)$ を増加させる．存在しないものについては $u(\langle o, w\rangle, a)$ のみ増加させる．この場合、ゲームは成功とみなす．

4 言語の創発

4.1 共通の語彙の出現

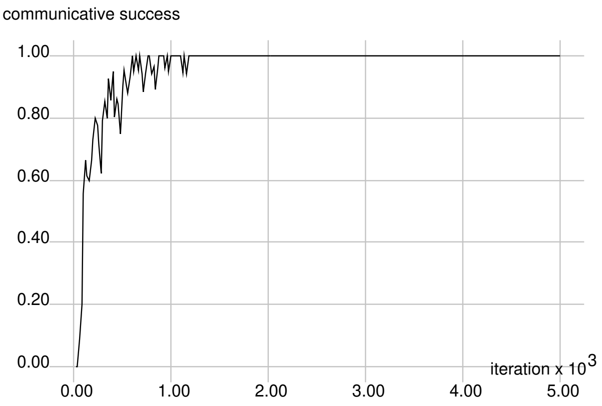

図 1　言語の創発の過程

図 1 は言語ゲームを繰り返すことでエージェント間に共通の言語が広まる様子を示している．この図において，横軸は言語ゲームの繰り返しの回数を，縦軸は全エージェントのコミュニケーションの成功率の平均を示している．ここではエージェントの数は 5, オブジェクトの数は 10 としている．

図 1 では，始めはエージェント自身がまだ単語をあまり創り出していないために辞書中の連想の数が不足していることと，あるオブジェクトについて話すときに用いる単語，すなわち会話に用いる連想がエージェントによって異なり，言語に一貫性がないためにコミュニケーションの成功率は低い．しかし，会話に用いる単語と意味の組合せ (連想) の

```
        --LANGUAGE_GAME_1--                    --LANGUAGE_GAME_2--
    obj: word: agent:                      obj: word: agent:
    o0   (bu)  [ a0 a1 a2 a3 a4 ]          o5   (di)  [ a5 a6 a8 a9 ]
    o1   (ha)  [ a0 a1 a2 a3 a4 ]               (la)  [ a7 ]
    o2   (ki)  [ a0 a1 a2 a3 a4 ]          o6   (lo)  [ a5 a6 a7 a8 a9 ]
    o3   (he)  [ a0 a1 a2 a3 a4 ]          o7   (di)  [ a5 a6 a7 a9 ]
    o4   (ce)  [ a0 a1 a2 a3 a4 ]               (na)  [ a8 ]
    o5   (ki)  [ a0 a1 a2 a3 a4 ]          o8   (bi)  [ a5 a6 a7 a8 a9 ]
    o6   (ni)  [ a0 a1 a2 a3 a4 ]          o9   (se)  [ a5 a6 a7 a8 a9 ]
    o7   (qe)  [ a0 ]                      o10  (po)  [ a5 a6 a7 a8 a9 ]
         (ni)  [ a1 a4 ]                   o11  (na)  [ a5 a6 a7 a8 a9 ]
         (co)  [ a2 a3 ]                   o12  (la)  [ a5 a6 a7 a8 a9 ]
    o8   (co)  [ a0 a1 a2 a3 ]             o13  (do)  [ a5 a6 a7 a8 a9 ]
         (ce)  [ a4 ]                      o14  (he)  [ a5 a6 a7 a8 a9 ]
    o9   (qe)  [ a0 a1 a2 a3 a4 ]
```

図 2　二つの言語ゲームで形成された典型的な言語の様子

幾つかがエージェント間で定まってくると，まだ定まっていない他の単語と意味の組 (連想) がとり得る組み合わせの数が減ることになる．こうした正のフィードバックの効果によって，一貫性のある語彙が急激に増加し，これに伴ってコミュニケーションの成功率も増加する．

5 言語の接触

第4節では，言語がエージェントの相互作用 (言語ゲーム) を通じて形成される過程を示した．既に第1節で述べたように，このようにして形成された言語は新たにエージェントやオブジェクトが加わっても適応することが [2] に示されている．

ここでは，別々の言語ゲームで言語を獲得した二つのエージェントの集団が接触したときに，異なる言語をもつエージェント同士のコミュニケーションを可能にするために，それぞれの言語がどのように変化するかを調べた．

5.1 非一貫性，曖昧性をもつ言語同士の接触

図2は一連の言語ゲームを通じてエージェント間に形成された言語の様子を示している．このモデルで形成される言語には，自然言語がそうであるように，同一のオブジェクトをエージェントによって異なる単語で呼ぶ非一貫性や，同一の単語が複数のオブジェクトを指すような曖昧性が存在する．

図2をみるとLANGUAGE_GAME_1 (以下LG1) で形成された言語ではオブジェクト o_7 ,o_8 に非一貫性があり (エージェントによって用いる単語が異なる)，単語 “ki” に曖昧性が存在している (o_2 と o_5 の二つを指す) ことがわかる．一方，LANUGUAGE_GAME_2 (以下LG2) では o_5 ,o_7 に非一貫性があり， 単語 “di” に曖昧性が存在している (o_5 ,o_7 の二つを指す) ．更に単語 “he” はLG1，LG2 両方の言語に存在しており，両者が接触することで曖昧性のある単語になる．ここでは o_5 ~ o_9 はLG1，LG2 双方の世界に共通して存在するオブジェクトであり，o_0 ~ o_4，o_{10} ~ o_{14} についてはそれぞれLG1，LG2 にのみ存在しており，相手の世界には存在しないオブジェクトになっている．

以上のような特徴を持った言語を用いる二つの集団中のエージェントのうち，LG1 の a_0 ,a_1 ,a_2 ,a_3 ,a_4 と LG2 の a_5 ,a_6 を接触させて言語ゲームを10000回実行した結果，図3に示した過程で言語は適応し，このときエージェント間に形成されていた言語は図4のようなものであった．

ここで接触した二つの集団は多数派，少数派の関

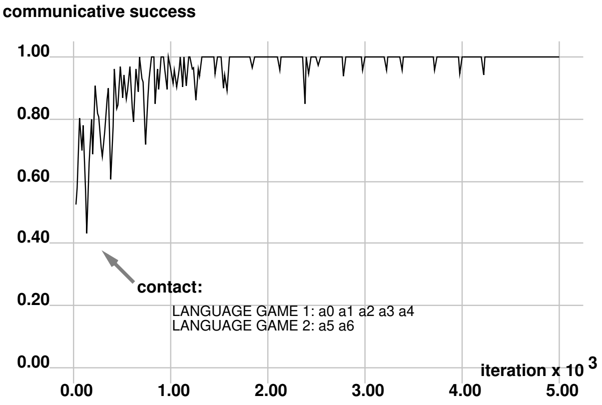

図 3　接触後の言語の適応

```
 obj: word: agent:
 o0   (bu)  [a0 a1 a2 a3 a4 a5 a6]
 o1   (ha)  [a0 a1 a2 a3 a4 a5 a6]
*o2   (lo)  [a0 a1 a2 a3 a4 a5 a6]
 o3   (he)  [a0 a1 a2 a3 a4 a5 a6]
 o4   (ce)  [a0 a1 a2 a3 a4 a5 a6]
*o5   (di)  [a0 a1 a2 a4 a5 a6]
      (ki)  [a3]
 o6   (ni)  [a0 a1 a2 a3 a4 a5 a6]
*o7   (ko)  [a0 a1 a2 a3 a4 a5 a6]
*o8   (bi)  [a0 a1 a2 a3 a4 a5 a6]
 o9   (qe)  [a0 a2 a3 a4 a5 a6]
      (se)  [a1]
 o10  (po)  [a0 a1 a2 a3 a4 a5 a6]
 o11  (na)  [a0 a1 a2 a3 a4 a5 a6]
 o12  (la)  [a0 a1 a2 a3 a4 a5 a6]
 o13  (do)  [a0 a1 a2 a3 a4 a5 a6]
*o14  (co)  [a0 a1 a2 a3 a4 a5]
      (he)  [a6]
```

図 4　接触後に形成された言語の様子

係にあり，なおかつ互いに相手にはないオブジェクトを持っているが，このような場合，言語は以下のように適応することが [5] に詳しく示されている．

- 多数派が用いていた言語が用いられるようになる．
- 自分の言語になかったオブジェクトについては直接，あるいは間接的に相手の言語から借用する．
- 自分の言語になかったオブジェクトについての単語を新たに創り出す．

接触する言語に非一貫性や曖昧性が存在する場合には，さらに以下のような傾向がある．

- 言語に非一貫性あるいは曖昧性が存在する場合，相手の言語から直接あるいは間接的に言語を借用することで，これが減少するように言語が適応する．
- 新たに単語を創り出すことで非一貫性，曖昧性を減少するように言語が適応する．

図 4 中の * 印は接触前に非一貫性あるいは曖昧性が存在していた言語を表しているが，これらはいずれも接触による言語の借用を通じて直接あるいは間接的に非一貫性，曖昧性を減少させており，より確実なコミュニケーションが可能になっている．

6　結論

本研究では単純な言語ゲームモデルを用いて言語の動的な側面，すなわち言語の接触によって引き起こされる現象に注目した．こうした言語の接触は自然界における言語でも日常的に生じており，その最も端的な例がピジン，クレオール言語で知られる諸現象 [1] であろう．本研究で用いたモデルは抽象的なモデルであり，自然言語におけるこうした現象を直接説明するものではないが，ここでの結果はこのような単純なモデルでも言語の持つ複雑性や多様性，流動性を説明し得るということを示唆するものである．

参考文献

[1] April M.S. McMahon. *Understanding language change.* Cambridge University Press, 1994.

[2] Luc Steels. Self-organising vocabularies. In *Proceedings of Artificial Life V*, 1996.

[3] Luc Steels. A self-organizing spacial vocabulary. *Artificial Life Journal*, 2(3), 1996.

[4] Luc Steels. The spontaneous self-organization of an adaptive language. In Muggleton S., editor, *Machine Intelligence*, Oxford, 1996. Oxford Univ. Press.

[5] Mario Tokoro Yuichiro Maeda, Takahiro Sasaki. Self re-organizing of language tirggered by 'language contact', 1997. Submitted to ECAL'97.