形式文法システムのネットワークと そのアルゴリズミックな進化

神戸大・自然科学研究科　　橋本　敬
e-mail: toshiwo@gradis.scitec.kobe-u.ac.jp

1　イントロダクション

本論では、計算能力の上昇するような進化のシミュレートを考える。計算能力は計算理論の分野においてオートマトンの計算能力という形で研究が進められている。オートマトンの計算能力の階層は、形式言語の語の受理能力を分類した Chomsky の階層によって特徴付けられる。[1]

ここでは、遺伝子上に形式文法がコードされており、各々の文法に則って語を発話・受理するルール・システム（RS）というものを考える。いくつかのルール・システムがネットワークをつくり、「会話」をおこなう。そしてルール・システムのもつ文法にはある割合で突然変異がおきる。文法の変異とセレクションの進化ダイナミクスにより、表現型とみられる語の受理能力がどのように進化するかをみる。

また、Crutchfield は言語の階層の再構成を行なっているが [2]、それはあるデータ列を記述できるオートマトンというかたちで行なっている。すなわち、記述の立場である。それに対して、ここでは構成的な立場からの言語の階層の再構成をこころみる。

2　モデル

2.1　ルールシステム（RS）の会話ネットワーク

ルール・システム（以下 RS と記述）は個々の文法を書き換えルールとして持っているシステムである。書き換えルールは $X \to A$ の形をしており、Xはひとつの非終端記号であり、A は 1 つ以上の終端記号または非終端記号の列である。終端記号とはそれ以上書き換えができない記号であり、非終端記号とはまだ書き換えることのできる記号である。記号列が終端記号のみによって構成されると、語の生成が終了する。

RS は非終端記号として X, Y, Sを、終端記号をして $0, 1$ を持っている。書き換えを始める最初の記号を初期記号と呼ぶが、ここでは Sである。また、左辺と同じ記号は右辺には現われないという制限を設ける。一方、RS の持つルールの数およびひとつのルールの右辺の長さは制限しない。個々の RS はそれぞれの持つ文法にしたがって語の発話・理解を行なう。

語を生成する過程では、ルールを左辺から右辺へあてはめる。生成は初期記号 Sから始め、記号列の頭からあてはまるルールを探し書き換えを行なっていく。あてはまるルールが２つ以上あるときはランダムにひとつのルールを選ぶ。１つのルールによる書き換えを 1 ステップとし、60 ステップ書き換えた時点で生成記号列に非終端記号が含まれていれば、発話失敗とみなす。また、生成記号列に非終端記号があるにもかかわらず、あてはまるルールがないときも発話失敗である。生成する語の長さは最大 6 とし、長さ 7 以上の語は 6 文字までで切った後の語を発話する。

生成過程とは逆に、受理過程ではルールを右辺から左辺にあてはめる。500 ステップ以内の書き換えで記号列が Sになれば理解できたとする。

10 人の RS が「会話」を行なうネットワークをつくる。すなわち、ある RS が発話した語を、発話した RS も含めた全 RS が受理を試みる。全員が一通り発話を行なうサイクルを 1 ステージとし、10 ステージで 1 タイムステップとする。

2.2 得点

RS は、語の発話、理解に応じて得点が得られる。さらに、自分の発した語が他の RS によって理解された場合にも得点がある。トータルの得点は

$$P_{tot} = R_{sp}P_{sp} + R_{rec}P_{rec} + R_{br}P_{br}$$

である。P_{sp}、P_{rec}、P_{br}はそれぞれ発話による得点、理解による得点、そして、聞き手（発話した RS 自身も含めて）に理解されたときの得点である。また、R_{sp}、R_{rec}、R_{br}は各得点の結合比である。

2.3 突然変異と選択

1timestep が終了した後、RS の持つルールに突然変異を起こす。しかし、これはその RS 自身のルールを変化させるのではなく、新たな RS として、もとの RS の持つルールの一部が変化したものをつくるのである。1timestep を RS のライフサイクルと考えれば、自分と全く同じ子孫と少し変化した子孫を残すと考えられる。

ここでは、３種類の変異を考える。それは、1) 付加（新たなルールをルールリストの最後に付加する）、2) 置換（新たなルールと古いルールを入れ換える）、3) 削除（ルールをひとつ取り除く）という３種である。ルールを付加する変異は、得点が平均以上の RS に対して変異率 m_{add}で起きるが、あとの２つの変異は全 RS に対してそれぞれ m_{rep}、m_{del}の変異率で起きる。

ルールリストの中から、ランダムに変異を起こすルールを選ぶ。ルールの変異のさせかたであるが、これも３種類ある。1) ルールの中でランダムに１文字を選び、その文字を別の文字に変える、2) その文字をとる、3) あるいはその文字の前か後に１文字入れるというように変異させる。

このようにしてできた新しい RS を得点の低かったものと入れ替え、全体で 10 人の RS がいるようにする。

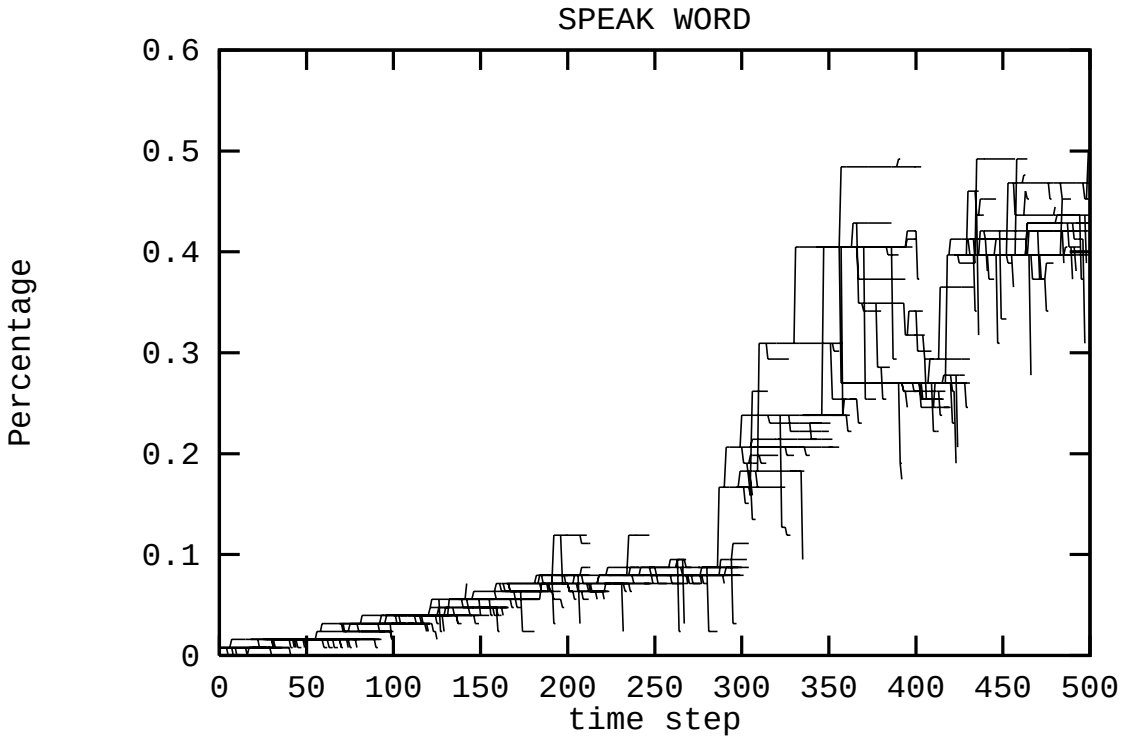

図 1: 系図（初期に発話できる RS を「始祖」とし、そこから別れて出ている線はその、変異体である。線が途中で切れているものはその時点で淘汰されたことを示す。）

3 結果

$R_{sp} = 3.0, R_{rec} = 1.0, R_{br} = -2.0$、$m_{add} = m_{rep} = m_{del} = 0.04$ の場合で示す。$R_{br} = -2.0$ なので、聞き手に理解されると点が下がる。よって、より多様な語を発話・理解する方向へと RS が進化することが予想される。

3.1 初期段階の発展

初期状態として、右辺が 1 記号のみのルールをランダムに 1 つ作る。Figure 1は横軸に timestep を、縦軸に全体の語の中で発話できる語の割合を書いたものである。これは、発話できる語の多様性の「系図」である。

より高い多様性を持った RS がところどころで現われている。RS の持つルールにより作られる語を生成木の形で表すことによりその発展の様子がよくわかるであろう。(Figure 2)

初めに存在する RS は生成木が sequencial な形をしている (a)。すなわち、$S \rightarrow 0$ のような単なる書き換えである。付加的突然変異により新たなルールを獲得し、別の語への書き換えが行なわれる。さらに、branch のしたでさらに branch がある（multi branch）構造が

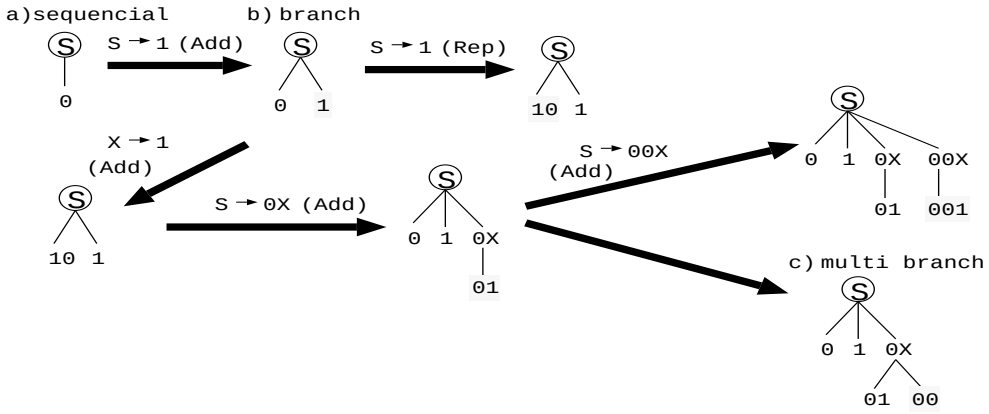

図 2: Rule System の初期の発展の様子 (付加的な変異により RS の発話できる語が増える)

現われる (c)。

3.2 アルゴリズム的進化（１）– モジュール型

初期の発展は前節のように、新たな RS が得る多様性はそれほど高いものではなくたかだか、数個の新たな語を発話・理解できるようになるだけである。しかし、Figure 1をみると $t = 192$ のところで、急に高い多様性を持ったものが現われている。このときの生成木の変化は Figure 3のようになっている。変異前のものを見ると多くの branch があり、そこに $X \rightarrow 00$ というルールが付け加わることで、たくさんの multi branch がつくられる。このとき $X \rightarrow 00$ というルールは、多数のルールからモジュールのように使われ多くの新たな語を生成する。このようにモジュールとなるルールを一つ得ることにより発話・理解できる語が飛躍的に増える。

3.3 アルゴリズム的進化（２）– ループ構造

$t = 310$ で、多様性が高い RS が登場している。モジュール型の進化までは、生成木が複雑になっていき葉が増えていく進化であったが、ここでの変化はこれまでのものとは違いルールの中でループを形成するものができるのである。(Figure 4)

生成木ではそのルールから作られる語をすべて表すことができない。

3.4 Ensemble の形成

モジュールをもって多様な語を発話・理解できる RS がでてもしばらくすると死んでしまう場合がある。たとえば、先ほど見た $t = 192$ でモジュール的に進化した RS は、その後、より発話の多様性の高い RS が現われるわけではないが淘汰されてしまう。また $t = 310$

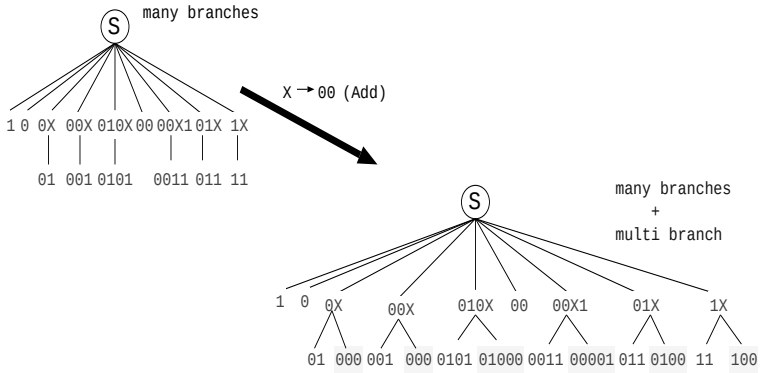

図 3: モジュール的進化の例（一つのルールがたくさんのルールから使われ、新たな語を生成する。）

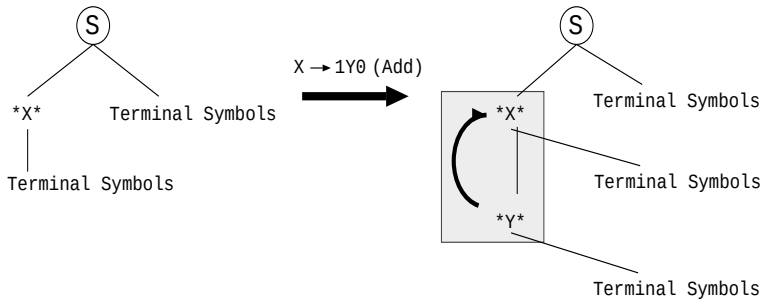

図 4: ループ構造の出現（$*X*$は右辺に Xが含まれていることを表している。これに、$X \rightarrow 1Y0$ というルールが付け加わると$*X*$と$*Y*$のあいだで無限に書き換えが可能である。このように、無限に書き換えが行なわれる場合、生成木ではそのルールから作られる語をすべて表すことができない。）

以前にもループを持ったRSは現われており、発話の多様性も高いものであったが、すぐに淘汰されている。

特に顕著な場合が、$t = 403$のあたりである。このように発話・理解の多様性が高いものが淘汰されるのは、Ensembleが形成されているからである。Ensembleとは、特定の語や、特定のルールにより生成される語を発話・理解しあう集団である。$t = 403$で存在するRSの持つルールを詳しく見ると、五つのEnsembleが形成されている。その中で小数派のRSは、発話・理解の多様性が高くても淘汰される。

4 結論

RSはモジュール的進化、ループ構造というかたちで、より多くの語を発話し、多くの語を理解する方向へ進化している。最終的にはすべての語を発話・理解することのできるオールマイティーへと向かう進化である。これは、得点の与え方とセレクションのかけかたにより予想されたものである。しかしそれぞれの発展の段階で、他のRSに理解されると得点が下がるにもかかわらず、同じルールを持つRSや、同じ語を話すRSがensembleを形成する。単に多くの語を発話できるだけではなく、集団に属しつつ新たな語を獲得していかなければいけない。すなわち、オールマイティーへ向かう進化対集団により保たれる文法という、多対多のネットワークであるがゆえの拮抗が生じる。

もし、ひとつの種のみによって値が決まるような評価関数をもちいて得点を付けたならばこのような拮抗は生じえないであろう。たとえば、この場合なら系図の縦軸に用いた発話あるいは理解の多様性で評価したなら、高い多様性を持ったものが淘汰されるようなことは起きない。ここでは、他のRSが話す語をどれだけ理解できるか、あるRSの話した語を他のRSが理解できるかということが大事である。これは、言い換えると外部に固定された環境があるわけではなく自分自身も含めてそこにいる種が環境を決定する。他にどのような種が存在するかにより環境が変化し、さらに自分が発した語により環境が変化すると考えられる。これまで、こういった世界は繰り返しジレンマゲームなどのゲーム理論によりモデル化されており、戦略はツリーを用いてコーディングされる[3]。しかし、そこでは戦略の論理的な進化を明確に捕えることができない。一方、ここではChomskyの階層によって進化の階梯を見ることができ、その結果としてツリー構造からループ構造への進化がみられた。

RSはそれぞれの特長を持った語を発するが、特定の語を理解できない。ensembleが一つの種を構成すると考えれば[4]、一つの種内に様々なアルゴリズムを用意し種内である程度のdiversityを保ちensemble内の個々のRSの持つ欠点を補いあい、他の種からの攻撃に対処したり[5]、種としての特定の同一性をもち様々な方向への進化の道を確保するのに役立っている。生物の進化では、ある方向への進化と多様性を広げるような進化が拮抗している。もっと大きなシステムサイズの場合、複数のensemble間で拮抗が生じ文法システムが分化していくようなさらに高次の階層が生じるであろう。また、文法の変異をもっと急激にすると（たとえば、交差や融合）分化したRSの集団の組み変わりなどが見られる

だろう。そうすれば、ある方向への進化と種の多様性が広がる進化が自然に構成されるであろう。

参考文献

[1] Hopcroft, J. E. , and Ullman, J. D. (1979) *Introduction to Automata Theory, Languages and Computation.*
(野崎昭弘、高橋正子、町田元、山崎秀記：オートマトン　言語理論　計算論 I,II , サイエンス社 (1986))

[2] Crutchfield, J. P. (1990) "Reconstructing Language Hierarchies"

[3] T. Ikegami, "From Genetic Evolution to Emergence of Game Strategy", Proceedings of Oji Conferende (1993)

[4] これはアイゲンの提出する「準種」の概念に相当する。M. Eigen (1993) SCIENTIFIC AMERICAN, July
日経サイエンス 1993 年９月

[5] 「準種」と混合戦略の議論に関しては池上高志　物性研究（この号）を参照