〔東京家政大学研究紀要　第46集 (2), 2006, pp. 79～86〕

パターン形成と言語理論

橋本　直樹

(平成17年10月6日受理)

Patterns and Formal Language

HASHIMOTO, Naoki

(Received on October 6, 2005)

キーワード：パターン，漸化式，形式言語

Key words：Pattern, Recursion, Formal Language

1．はじめに

自然界に存在する種々のパターンを科学的に説明する多くの試みがある．その中で，化学反応により動物の表皮の模様をダイナミクスとして記述し再現することに成功している研究がある．それは，A.Turing[1]に始まる研究で，形態形成を化学反応系の連立微分方程式として解く方法である．これは，ある原理をもとにして演繹的に自然現象を理解しようとする考えに基づくアプローチの一つである．これらの方法によるパターン形成の原点はそのTuringモデルであるが，それを様々な系に対して適応し，またより簡単に解くために数学からコンピュータ工学或いは生物学の分野までの広範囲の研究対象となっている．その中でS.Wolfram[2]により発展させられたCA（セルオートマトン）に基づく研究も重要である．それは，2進数でパターンを表現できるからで，自己形態生成の簡単で強力な研究手段となっている．

本論文では，Turingモデルの平衡点に対する空間方向の広がりに対する解，及び初期条件・境界条件について議論する．これらの条件は漸化式で与えられるので，2節でその一般解を求める．そして，漸化式を新たな出発点として，Turingモデルの枠内でどのような定常的なパターンが構成できるのかを調べる．実際の簡単なパターンには，周期性を持つものが多いので，周期解が得られる漸化式を特殊例として構成しその性質を調べる．本来Turingモデルは，パターンの時間的な変化を記述するものであるが，我々の議論は，静的な空間方向の解の条件を与える．

3節では，言語理論と形態形成を記述する漸化式との関係を議論する．同種の研究として，植物の成長構造を記述するL-System[3]という方法がある．また，RNAをA，T，G，Cの文字列の文法を持つ言語として扱う方法[4]や，岡村[5]らによる鳥の鳴き声を文法構造を持つ言語として捉えるという方法がある．我々は，自然界を含めたパターンを言語構造，特にその文法構造として分類できるのかという動機から上記の平衡点でのTuringモデルから得られる空間方向のセルの広がりパターンを言語として捉えて，その文法構造を形式言語理論の立場から議論する．

2．パターンと漸化式

Turingモデルは，生物の模様などのパターンが，その細胞（セル）内の2種類の化学物質，すなわち活性化因子と阻害因子の2つの物質が相互に反応しながら拡散していくという考え方に基づいている．空間1次元のモデル方程式を，平衡状態の周りで展開した形で書くと，

$$\begin{aligned} \frac{dx_r}{dt} &= Ax_r + By_r + \mu(x_{r+1} - 2x_r + x_{r-1}) \\ \frac{dy_r}{dt} &= Cx_r + Dy_r + \gamma(y_{r+1} - 2y_r + y_{r-1}) \end{aligned} \tag{1}$$

ただし，x_r, y_rは，r番目（$1 \leq r \leq N$）のセルの活性化因子と阻害因子の濃度である．Nは，セルの空間的な広がりの数で，周期Nの境界条件を置く．μは活性化因子x_rの近傍のセルへの拡散定数で，γは阻害因子y_rの近傍

英語英文学科　第3英語コミュニケーション学研究室

のセルへの拡散定数である．Ax_r+By_r，Cx_r+Dy_rは，活性化因子，阻害因子それぞれの増加率である．A. Turing[1] は，阻害因子の時間変化が活性化因子より速いとき，定常的な空間パターンが出現すると結論している．我々は，平衡状態の極限，もしくは漸近的な平衡状態が存在するときの解を求めたい．しかし，(1) 式の一般解で，$t\to+\infty$として条件を満たす一般的な状態が得られると期待するが実際的に求めるのは大変難しい．(1) 式の一般解を複素フーリエ級数の形でTuringが求めているが，それを初期条件の設定が容易になるように実フーリエ級数の形で書くと次のようになる．

$$x_r=\sum_{s=1}^{N}(E_s e^{p_s t}+F_s e^{p_s' t})\cos(\frac{2\pi r}{N}s)$$
$$y_r=\sum_{s=1}^{N}(G_s e^{p_s t}+H_s e^{p_s' t})\cos(\frac{2\pi r}{N}s) \tag{2}$$

ここで，E_s,F_s,G_s,H_s は，初期条件すなわち $t=0$ の状態から定められる任意定数である．

また，この展開係数の中のp_s,p_s'は，次の各 s ごとの p についての特性方程式の根であり，Turingが求めた複素フーリエ級数の場合の条件と同じになる[1]．

$$(p-A+4\mu\sin^2(\frac{\pi s}{N}))(p-D+4\gamma\sin^2(\frac{\pi s}{N}))=BC \tag{3}$$

$t=0$ の初期条件から，これらの定数 E_s,F_s,G_s,H_sを定めるには，2 N個の連立方程式を解いて決定せねばならず，特別な状況以外正確に解くのは難しいが，$t\to+\infty$ の挙動は少しわかる．(3) 式を解いて，その解が実数ならば，$t\to+\infty$のとき，解 (2) は，減衰または増大する．また，解 (2) が複素数解の場合は，$t\to+\infty$のとき，その解は振動すると考えられる．

我々は，平衡状態のセルの配位を求めるため (1) 式に条件 $\frac{dx_r}{dt}=\frac{dy_r}{dt}=0$ を置いて出発する．すなわち，はじめから時間に依存しない解をTuringモデルの枠内で求める．これは，モデルが満たすべき初期配位を求めることと同じである．ただし，N個のセルに対する周期境界条件は後で議論する．

$$Ax_r+By_r+\mu(x_{r+1}-2x_r+x_{r-1})=0$$
$$Cx_r+Dy_r+\gamma(y_{r+1}-2y_r+y_{r-1})=0 \tag{4}$$
$$(r=1,\dots,N)$$

この式の (x_r,y_r)，$(r=1,\dots,N)$ に対する一般解を求める．しかし，我々が欲する量は，活性化因子に対応する量x_rで，y_rはその際の補助的な量である．これは，CAなどの他の方法による計算でも同様である．このため，(4) 式で，変数y_rを消去した漸化式を作る．式 (4) の定数を書き換えて次のように置く．

$$x_{r+1}=ax_r+by_r+ex_{r-1}$$
$$y_{r+1}=cx_r+dy_r+fy_{r-1} \tag{5}$$

ただし，$a=(2-\frac{A}{\mu})$，$b=-\frac{B}{\mu}$，$c=-\frac{C}{\gamma}$，$d=(2-\frac{D}{\gamma})$，$e=-1$，$f=-1$である．

式 (5) から，y_rを代数的に消去すると活性化因子の変数だけからなる次式が得られる．

$$x_{r+2}-(a+d)x_{r+1}-(e+f+bc-ad)x_r-(af+ed)x_{r-1}-ef\,x_{r-2}=0 \tag{6}$$

これは，定数係数の 4 階線形同次差分方程式である．このとき，n 階線形同次差分方程式の一般論[6] から，(6) 式の一般解を求めることができる．すなわち，(6) 式の特性方程式$\Phi(t)$を，tを不定元として

$$\Phi(t)=t^4-(a+d)t^3-(e+f+bc-ad)t^2-(af+ed)t-1=0 \tag{7}$$

とすると，代数学からこの 4 次代数方程式の解の公式が存在し，その解法もよく知られている．(7) 式が既約な場合は，Ferroriの解法やCardanoの公式，またはLagrangeの解法など[7] によりその一般解が求められ，また，既約でない場合は 3 次方程式以下の解の公式による．今，方程式 (7) の異なる根を$\alpha_1,\dots,\alpha_\lambda$，その重複度をそれぞれ $m_1,\dots,m_\lambda$ とすれば，(6) の一般解は，$r^k\alpha_i^r$ $(k=0,1,\dots,m_i-1;i=1,\dots,\lambda)$ を基底とする複素数体上の 1 次結合として表せる[8]．ただし，$0\le\lambda\le4$ で，かつ $\sum_{i=1}^{\lambda}m_i=4$ である．すなわち，x_rは，4 つの基本解の 1 次結合で表される．

$$x_r=\sum_{i=1}^{\lambda}\sum_{k=0}^{m_i-1}g_{ik}r^k\alpha_i^r \tag{8}$$

ここで，g_{ik}は，高々 4 つの項が非零の定数で，漸化式の初期条件により決定される．重複度 m_i がすべて 1 の場合は，$\lambda=4$ で $g_{ik}=g_i$ $(i=1,\dots,4)$ となる．

例 1　$x_{r+2}+2x_{r+1}-6x_r+2x_{r-1}+x_{r-2}=0$

この漸化式の特性方程式の根は，$x=1$（重複度 2），$x=-2\pm\sqrt{3}$であるので，一般解x_rは，

$$x_r = a_1 + a_2 r + a_3(-1+\sqrt{3})^r + a_4(-1-\sqrt{3})^r$$

と書ける．ここで，a_1, a_2, a_3, a_4 は，任意定数である．

例１の漸化式は，活性化因子の単純増大（減少）列となり有界でない．セルの個数から要請される周期境界条件は満たさないので，人為的にそのセルの周期を設定することになる．このような解は，パターン形成としては余り適切でない．

次に，(6) 式の周期解について考える．(6) 式の一般解を求めそしてその初期値を与えたすべての解の中から周期解を把握するのはかなり大変である．しかしより簡単に周期性の存在がわかる方法が知られている[6]．その方法で調べるため，(6) 式の $\{x_r\}$ に対して，$X_r = {}^t(x_{r-2}, x_{r-1}, x_r, x_{r+1}, x_{r+2})$ とおく．このとき，(6) 式は，

$$X_{r+1} = MX_r \tag{9}$$

とかける．ただし，M は，４×４の行列で次の値を持つ．

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -ef & -(af+ed) & -(ad-e-f-bc) & -(a+d) \end{pmatrix} \tag{10}$$

このとき，次の結果が知られている．

定理２［文献６］ 漸化式 (8) が，周期Ωの周期解を持つための必要十分条件は

$$\det(M^{\Omega} - E) = 0 \tag{11}$$

となることである．ここでＥは，単位行列である．すなわち，M^{Ω} が固有値１を持つことである．

この定理から，(6) 式の係数，すなわちその係数行列 (9) が与えられれば，周期解の存在を調べることが出来る．我々は，漸化式で得られる値すなわち値域が整数になるものを求めたい．これは次節での言語的な扱いをしたいからである．この点を考慮して，係数を変えて調べた結果，いくつかの周期解を見つけることができた．その１つの簡単な例が次の漸化式である．

例３ 漸化式 $x_{r+2} = -x_{r+1} - x_r - x_{r-1} - x_{r-2}$ の行列表現による係数行列

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -1 & -1 & -1 \end{pmatrix}$$

は，$M^5 = E$ を満たすので，周期 $\Omega = 5$ である．すなわち，$x_{r+5} = x_r$ となる．

初期値を1111とするとその数列のパターンは，

1111-41111-41111-41111-41111-4..........

となる．

漸化式により作成される整数（パターン）は，その初期値に依存する．我々は，なるべく簡単なパターンを得るため整数Ｚ上で初期値を与え，できるだけ漸化式の値域の集合が小さくなるものを探した．ただし，例１のような列は，増大列なので値域の集合は無限集合になる．漸化式で得られる数値は，活性化因子の値である．この列の値域が，いかにパターンとして表されるかというダイナミクスは化学反応として記述されるべきで，我々の方法では不可能である．生体反応も記述することが必要だからである．また，化学反応自体をTuringモデルの枠内で扱っている文献も多く存在する．我々は，Turingモデルの結果を抽象化し，単純に数値としてのパターンないしはそれを別の記号や文字に置き換えたもので得られるパターンを考える．この抽象化により，必ずしも自然現象だけでなく人工的にあるいは計算的に作られるパターンも研究対象としていることになる．Turingモデル (1) は，セルに対して周期境界条件を置くのが一般的である．しかし，例３のように，自発的に解が周期性を持つものが存在することがわかる．ただし，係数がＺ上の漸化式で，長周期の解を求めることは難しい．係数および値域がＺ上にあることのメリットは，数値計算が "exact" にできるからである．しかし，特性方程式から得られる一般解は，通常は，根号で表現されるものが多く，特にいろいろな周期解を得るには，係数体が有理数体ではなく実数体上で考える必要がある．この場合，コンピュータで計算すると精度は高くできるがあくまでそれは近似計算である．近似計算のデメリットは，計算回数が多くなる長周期の解については，丸めの誤差が積もって不正確な計算になる場合が生じることである．すなわち数値計算のみで，周期性の議論をすることが危う

くなる．このため，我々は，以下で (6) 式のある種の漸化式が長周期の解を持つことを数値計算ではなく，数学的に証明する．

補題 4　Kを自然数とする（$K \geq 5$）．このとき，漸化式

$$x_{r+2} - 2\cos(\frac{2\pi}{K}) \cdot x_{r+1} + 2x_r - 2\cos(\frac{2\pi}{K}) \cdot x_{r-1} + x_{r-2} = 0 \quad (12)$$

は，次の周期Ωを持つ．ただし，pを奇数，qを自然数とする．

$K = 4 \cdot q$のとき，周期$\Omega = K$
$K = 2 \cdot p$のとき，周期$\Omega = 2K$
$K = p$のとき，周期$\Omega = 4K$

証明

(12) 式を行列の形で書く．$X_r = {}^t(x_{r-2}, x_{r-1}, x_r, x_{r+1}, x_{r+2})$として，かつ

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & (e^{\frac{2\pi}{K}i} + e^{-\frac{2\pi}{K}i}) & 2 & (e^{\frac{2\pi}{K}i} + e^{-\frac{2\pi}{K}i}) \end{pmatrix} \quad (13)$$

とおくと (12) 式は，

$$X_{r+1} = AX_r$$

となる．この形の行列に対しては，その固有多項式$\Phi(t)$と，最小多項式 $m(t)$（t は不定元）が一致する[8]．複素数体上で

$$\Phi(t) = (t-i)(t+i)(t - e^{\frac{2\pi}{K}i})(t - e^{-\frac{2\pi}{K}i}) \quad (14)$$

となる．Jordan標準形の一般論より，複素数体上の正則行列Pが存在し，そして固有多項式の固有値から具体的にPが構成できて

$$J = P^{-1}AP \quad (15)$$

となる．$K \geq 5$であるので，Jordan標準形Jは，

$$J = J(i,1) \oplus J(-i,1) \oplus J(e^{\frac{2\pi}{K}i},1) \oplus J(e^{-\frac{2\pi}{K}i},1)$$

となる．このJordan標準形を行列の形に書けば

$$J = \begin{pmatrix} i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & e^{-\frac{2\pi}{K}i} & 0 \\ 0 & 0 & 0 & e^{\frac{2\pi}{K}i} \end{pmatrix} \quad (16)$$

である．我々が計算したい量は，A^m（mは自然数）であるので，Pの正則性を用いて，

$$A^m = PP^{-1}APP^{-1}AP \ldots\ldots\ldots P^{-1}APP^{-1} = PJ^mP^{-1}$$

となるので次にJ^mを計算する．このJは，対角行列なので

$$J^m = \begin{pmatrix} i^m & 0 & 0 & 0 \\ 0 & (-i)^m & 0 & 0 \\ 0 & 0 & e^{-\frac{2\pi m}{K}i} & 0 \\ 0 & 0 & 0 & e^{\frac{2\pi m}{K}i} \end{pmatrix}$$

となる．ここで，Kの各値に対してこの対角要素を評価する．$K = 4q$（qは自然数）に対しては，$m = K$のとき，J^mの対角要素は，$diag(J^m) = (1,1,1,1)$となる．同様に，$K = 2p$（pは奇数）に対しては$m = 2K$，$K = p$に対しては，$m = 4K$とすると，それぞれの場合に対しJ^mの対角要素は，(1,1,1,1) となる．すなわち，これらのmの値のとき，$J^m = E$（Eは単位行列）となる．従って，

$$A^m = PJ^mP^{-1} = PP^{-1} = E$$

となるので，定理 2 からそのmは周期となる．

（証明終わり）

例 5　$K = 10$のとき，漸化式 (12) は，

$$x_{r+2} - \frac{(\sqrt{5}+1)}{2}x_{r+1} + 2x_r - \frac{(\sqrt{5}+1)}{2}x_{r-1} + x_{r-2} = 0 \quad (17)$$

となり，周期Ωは，$\Omega = 2K = 20$である．
この漸化式に，初期値$x_1 = 1, x_2 = -1, x_3 = 1, x_4 = 1$を設定して数値を求め，その値が正のとき文字 '1'，負のとき文字 '0' としてパターンを作ると

1011000011011000100111011000011011000100 11 (18)

のようになる．数値を射影して別の記号，文字に置き換えることによりパターンを得ることができる．これは次節の言語理論と関係する．

以上の結果により，Turingモデルのセルの周期Nに漸化式の周期を選ぶことができる．また漸化式に周期解がない場合でも，有限範囲に限定した数列と考えれば，人為的ではあるがパターンを得ることができる．しかし，得られた解をTuringモデルの真の静的解として理解するためにはその解の安定性を示す必要があるがここではその議論は略す．

3．パターン形成と形式言語

本節では，Turingモデル等の形態形成を記述する方法の結果で，特に規則性を持つパターン形成に注目し，それを言語として捉えることができるのかを議論する．言語理論は，広範囲にわたっており，また，パターンも自然現象によるもの，人工的なものなどさまざまである．ここでは，一般論ではなく，ある種の特別なパターンが，ある特定の言語文法に属しているのかを調べ，より一般的な理論構成の足がかりとするのが目的である．

以下，言語とは，形式言語（Formal Language）理論の意味で定義されるものとする．この言語には，チョムスキー階層と呼ばれる4つの言語分類とそれを規定する形式文法（Formal Grammar）が知られている．本稿で関係するのは，その中の3つで，文脈依存文法に基づく1型言語（文脈依存言語），ならびに文脈自由文法に基づく2型言語（文脈自由言語），それと正規文法に基づく3型言語（正規言語）である．これらの言語には包含関係があり，いずれも互いに前者のほうが大きい集合で，それぞれは，その真部分集合であることが知られている．

ここでは，漸化式やCAなどの計算により与えられた数値パターンを，文字としてとらえ，それを形式言語として形式文法構造を調べる．パターンを文字として認識する方法には，不確定性があり，確定させるためにはある種の条件が必要になる．実数計算による結果は，多くの場合，視覚的にわかりやすくするために色とか濃淡により図形表示またはグラフ表示されるが，数学的にはその結果の把握が難しい．我々は，結果を離散的な対象にするため最も簡単な射影変換を行いそれを言語として捕らえる．

文字 $\{0,1,\ldots,b-1\}$ 上で，フィボナッチ数 F_n を，b 進数にエンコードしたものからなる言語を F_b とする．特に F_1 は，$\{0^{F_n} \mid n \geq 1\} = \{0,00,000,00000,\ldots\}$ により定義される．Moll & Venkasen[9] は，F_b（$b>1$）が文脈自由言語ではないことを証明した．$b=1$ の場合は，同様のことが簡単に判る．また，Mootha[10] は，F_1 が文脈依存言語であることを示した．フィボナッチ数列は比較的簡単な数列であるが，その言語としての文法構造を決定するのは簡単でない．

2節で得た漸化式は，いかなる言語構造を持つパターンを導出できるのかという問題意識の下で，形式言語への変換方法及びその言語の文法構造を調べた．本稿ではその中の一部を報告する．はじめに，文脈自由言語に対して次が成立する．

定理6［文献11］ 文脈自由言語は，集合和，連接，閉包，準同型写像のそれぞれの演算で閉じている．

ここで，集合和は2つの言語LとMがあったとき，LとMの和集合 $L \cup M$ のことで，連接とは，Lに属する列に対し，Mに属する列を連接することによりできる列全体をいう．また，言語Lの閉包とは，Lの中の有限個の列を取り出し，それらを連接してできる列全体をいう．準同型写像は，文字列の中の各文字を特定の文字で置き換える演算である．定理6と同様の主張は，正規言語に対しても成立する[11]．もし与えられた言語が，文脈自由言語（または正規言語）ならば，定理の準同型写像により文字を置き換えても言語構造は変わらないので，扱いやすい文字に置き換えればよい．しかし，文脈依存言語は，準同型写像に対して閉じるために，その写像にある条件（ε-free）を要する[12]．これらのことにより，前節の漸化式で得られた数値結果を文字に変換できる．特に整数列の場合には1対1対応で可能である．実数の場合には射影を行う．例5で示したように，最も簡単な射影は数値が正の場合と負の場合に分ける方法である．この操作は，セルオートマトンによる研究でも普通に行われている方法なので我々も用いる．

例3で，準同型写像により「1」→ 'a'，「-4」→ 'b' に置き換えるとその文字パターンは次のようになる．

aaaabaaaabaaaabaaaabaaaab...... (19)

これは，周期 $\Omega=5$ の単純なパターンである．しかし，初期値を変えた場合，出現パターンは異なる．また，例5による周期20のパターンは，(18) の「1」を 'a' に，「0」

を 'b' に準同型写像により置き換えると

abaabbbaabaabbbabbaa (20)

が繰り返されるパターンとなる.

次に周期パターンではない例を示す. 経験的に漸化式の計算は, 周期性のある漸化式の係数に摂動を与えると, その連続性から周期性はないが急な増減をしない例が作成できる. この方法により例 3 から派生した次の例を計算した.

例 7　漸化式 $x_{r+2}+\frac{11}{10}x_{r+1}+x_r+x_{r-1}+x_{r-2}=0$

初期値は, $x_1=x_2=x_3=x_4=1$ として計算し, その値が正のとき 'a', 負のときを 'b' とすると, 次の非周期の無限列が得られる.

aaaabaaaabaaaabaaaabaaaabaaaabaababaababaababaababaababaababaababbabab (21)

以上のように漸化式からは, いろいろな文字列パターンを得ることができる. これらを言語と見なしたとき, それはどのような文法構造を持っているのかということを調べる.

ある言語が, 文脈自由言語なのか, あるいは正規言語なのかを判定できる定理がある. 反復補題 (Pumping Lemma) と呼ばれるもので, 文脈自由言語, 正規言語のそれぞれに互いに類似した定理がある. 文脈依存言語に対応する定理を一般的には考えない. 文脈依存言語でないことを証明することはほとんどないからである. 次はその文脈自由言語の反復補題である.

定理 8 [文献11]　L を文脈自由言語とする. このとき, ある定数nが存在して, zを長さn以上のLの列とすれば, $z=uvwxy$と書けて, 次の条件を満たす.

(i) $|vwx|\le n$ ($|a|$は, aの文字列の長さを表す)

(ii) $vx\ne\varepsilon$ (ε は空列)

(iii) 任意の $i\ge 0$ に対し $uv^iwx^iy\in L$

この定理は, 言語Lが文脈自由言語でないことを示すのに威力を発する. また, ある言語Lが, 特定の文脈自由文法を持つことを示すには, 通常はその文法を構成する. 一方, ある文字列または言語が特定の文脈自由言語に所属するかどうかは決定可能で, CYK アルゴリズム[11, 15]により簡単に調べることができる. CYK アルゴリズムは, 言語Lの文脈自由文法G = (V, T, P, S) から出発し, n文字の入力文字列wに対し, $O(n^3)$ 時間で $w\in L(G)$ か否かを決定できる. ここでVは変数の集合, Tは終端記号の集合, Pは生成規則の集合, そしてSは出発記号を表す. このアルゴリズムでは, 入力列 $w=a_1a_2....a_n$ に対し三角形の表を作成する. 表の要素 X_{ji}は, $A\overset{*}{\Rightarrow}a_ia_{i+1}......a_j$ となる文法を定義する変数の集合である. $\overset{*}{\Rightarrow}$は導出を表す. 表は, $X_{ii}(i=1,....,n)$ から順に作成する. X_{ii}は, $A\rightarrow a_i$となる規則中の変数の集合である. 以下順次$j>i$の条件の下でX_{ji}の変数の集合を求めていく. CYKアルゴリズムの原理は, その言語または文字列を生成するすべての並べ方を調べるものである. これらの方法論により, 上で求めた文字パターンを調べる. 周期性がある文字列については, 次のことが知られている.

補題 9　次が成立する.

(i) 言語$COPY=\{ww\,|\,w\in\{a,b\}^*\}$は文脈自由言語ではない.

(ii) 言語$COPY$に属さない文字a, bの列は, 文脈自由言語である.

(iii) 言語$COPY$は, 文脈依存言語である.

証明

(i) 定理 8 を使って証明できる. あるnに対して, $COPY$に属する言語$z=a^nb^na^nb^n$を選ぶと定理 8 の条件を満たさないことを容易に示せる[11].

(ii) 実際に文脈自由文法Gを構成する方法で証明できる[13].

S → A | B|AB | BA

A → a | bAb | bAa | aAb | aAa

B → b | bBb | bBa | aBb | aBa

ここで→は生成記号, Sは出発記号, A, B, Cは変数, a, bは終端記号である.

(iii) 実際に文脈依存文法を構成する方法で示す. 例えば文献14にその構成例がある.

(証明終わり)

補題にある言語$COPY$はよく知られた言語で, すべての自然言語は文脈自由言語であるかどうかの議論に使われる. 我々の場合は, 周期性のある漸化式から得られる場合に適応できる. 漸化式が周期解のみを生成する場合はこの補題により, その言語パターンは, 文脈自由言語

ではなく文脈依存言語に属する．したがって，漸化式(12)で，Kが5以上の自然数の場合には，すべて文脈依存言語を生成する．

一般の数列は，必ずしも周期解だけではなく，単調に増大・減少をする解も生成する．このとき，*COPY*のような性質のパターンが含まれていても文脈依存言語であるということはわからない．そのような多くの場合，その言語は文脈自由言語であると予想できる．このため，なるべく多くの文字パターンを文脈自由言語として許容するように次の文脈自由文法Gを定義した．CYKアルゴリズムは，チョムスキー標準形（CNF）[11]が必要なので，CNFに変換したもので書く．

例10　CNF文法Gを次で定義する．

$S \rightarrow AB \mid AA$

$A \rightarrow AB \mid AC$

$A \rightarrow a$

$B \rightarrow b$

$C \rightarrow a$

ここで→は生成記号，Sは出発記号，A，B，Cは変数，a，bは終端記号である．

この文法Gに対する言語をL（G）と書く．与えられた文字列がこのL（G）に所属するか否かは上述のCYKアルゴリズムで調べる．このアルゴリズムによる判定時間はn^3に比例するから(21)のような文字列を調べるためには，手作業による実行はほとんど不可能である．そのため，プログラミング言語によりCYKアルゴリズムによる所属言語判定プログラムを作成した．現時点では，文法ごとに異なるプログラムを作成する必要があるが，(21)の判定に，1秒もかからない．その結果，(21)の文字列はL（G）に所属するという結論を得た．このL（G）には，(19)のような周期のある文字列も所属する．しかし補題9（i）は，周期性のあるものだけの文脈自由文法は構成できないということを主張していることに注意する．他に，Turingモデルに基づいたCAにより得られるパターンについても例10の文法への所属問題を調べた．CAからのパターンはかなりランダムに近い．その場合調べたすべてがそのL（G）に所属した．CAについての詳細は別途報告する予定である．

漸化式から得られるパターンの中で，文脈自由言語のものと文脈依存言語のものが多いことがわかる．特にすべての列が周期を有する漸化式から得られるパターンは文脈依存言語となる．自然言語と比べるとチョムスキー階層のより上位に分類されるものが多い．すなわちそのパターンは，自然言語より複雑な形式言語構造を持っていると考えられる．

参考文献

1．Turing, A.M., (1952) The chemical basis of morphogenesis, Phil. Trans. Roy.Soc. London B237, 37-72
2．Wolfram, S., (1983) Statistical mechanics of cellular automata, Rev.Mod.Phys. 55, 601-644
3．Prusinkiewiz, P., & Lindenmayer, A., (1996) The Algorithm Beauty of Plants, Springer-Verlag, New York
4．Dowell,R.D. & Eddy, S.R., (2004) Evaluation of Several lightweight stochastic context-free grammars for RNA secondary structure prediction, BMC Bioinformatics, 4, 71
5．岡ノ谷一夫，(2000)「言語の発生（心の比較認知科学)」，ミネルヴァ書房
6．杉山昌平，(1971)「差分・微分方程式」，共立出版株式会社
7．高木貞治，「代数学講義」，共立出版株式会社
8．杉浦光夫，「Jordan標準形と単因子論Ⅱ（基礎数学講座)」，岩波書店
9．Moll, R.J. & Venkatesav, S.M., (1991) Fibonacci numbers are not context-free, Fibonacci Quarterly, 29 (1), 59-61
10．Mootha, V.K., (1993) Unary Fibonacci numbers are context-sensitive, Fibonacci Quarterly, 31 (1), 41-43
11．Hopcroft, J.E., Motwani, R., & Ullman, J.D., (1979) Introduction to Automata Theory, Languages, and Computation, Addison-Wesley
12．Rus, T., http://www.cs.uiowa.edu/~rus/Courses/Theory/cfl3.pdf
13．Nievergelt, J., http://www.jn.inf.ethez.ch/script/chapter5.pdf
14．Gallier, J., (2005), Phrase-structure grammars, context-sensitive grammars, CIS511

15. Kasami, T., (1965). An efficient recognition and syntax-analysis algorithm for context-free languages. Scientific report AFCRL-65-758, Air Force Cambridge Research Lab, Bedford, MA., Younger, D.H., (1967), Recognition and parsing of context-free languages in time, Information and Control 10 (2) : 189-208., Cocke,J., & Schwartz, J.T., (1970), Programming languages and their compilers: Preliminary notes. Technical report, Courant Institute of Mathematical Sciences, New York University

Abstract

In this paper, we show a pattern generation of the cell through the recursion which is defined by the equilibrium point of the Turing model. As a specific example, we construct the recursion with long periodicity. The obtained some patterns relate to the formal languages produced by the context free grammars and the context sensitive grammars.