

# トランスポゾンから導かれる演算による チョムスキー言語の生成と言語理解機能 の数理モデルの試み

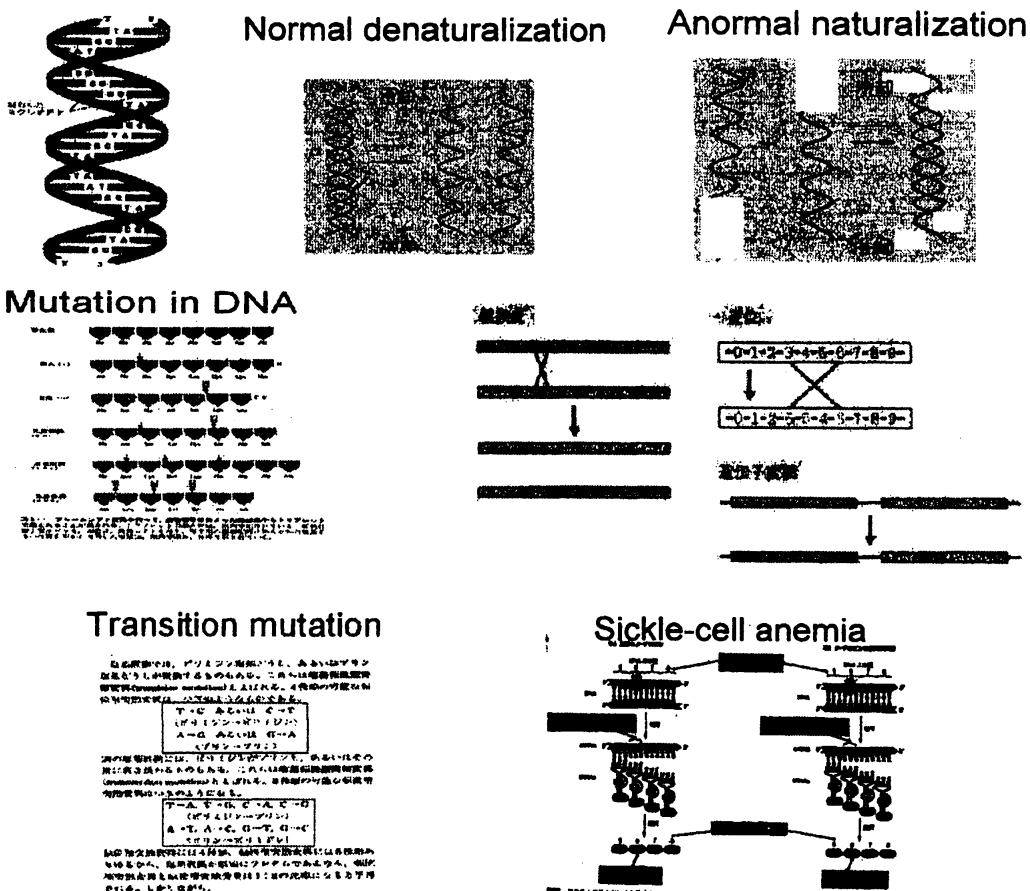
日本大学文理学部情報システム解析学科  
鈴木 理

## Abstract

チョムスキー言語に付随した DNA モデルを考え、DNA の持つ演算を用いてこの言語の生成を考える。その演算のひとつはトランスポゾンに基づいていることに注意する。これを用いてチョムスキー文の変形文法がどのように理解されるかを数理モデルを用いて考える。

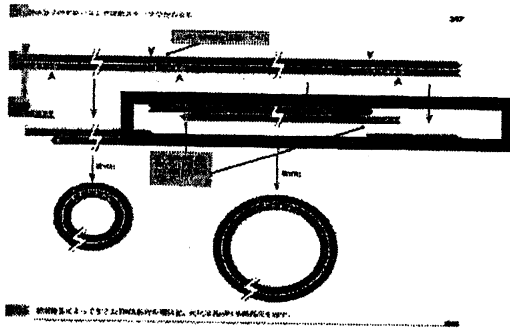
## 1. DNA モデル([2])

DNA は 4 種類の核酸 A, T, G, C を配列することにより得られる紐である(下図左上)。基本構造は(1)  $A \leftrightarrow T, G \leftrightarrow C$  となる相補性と(2) 突然変異である。突然変異は欠損、挿入、置換等からなっている(下図右上、下図左下)。Transition mutation についてもものべておく。これはあとで議論される(下図右下)。

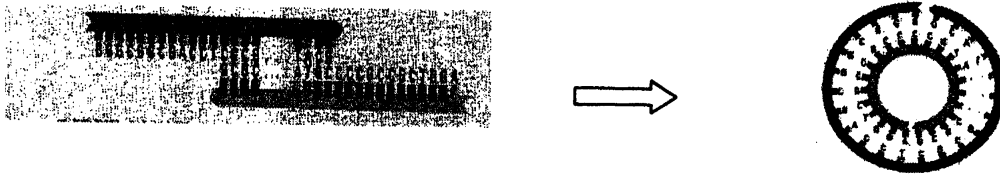


## 2. トランスポゾン ([2])

従来 DNA における進化は垂直方向(木構造)の進化を中心に考えられてきたが、ミトコンドリアあるいはラムダー・ファージのようにある遺伝子が別の遺伝子にはいりこむことにより突然変異をおこすものが確認されており、これを従来の変異と区別して水平変異と呼んでいる。類似の性質は DNA における突然変異のなかにも観測される。ここではこれもトランスポゾンということにする。



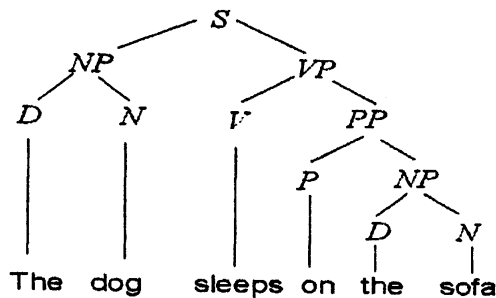
このように DNA の一部を取り去りこれを円環にして運搬し、またもとの線状 DNA に戻して別の DNA に進入することにより突然変異をおこす。スキームは次のようになる。



このほかにトランスポゾンは自分の分身を大量に生産し、DNA 内に分配しその組織を同じメカニズムで変異を起こし変質させる性質をもつ。

## 3. チョムスキー言語

英文で書かれた文章の木構造を考えることによりチョムスキー言語の基本構造が導入される ([7])。



この機構造に括弧列を考えることにより文脈自由文法が定義される。

`{{ The dog } { sleeps on { the sofa } } }`

`{ { } } { { } { { } } } }`

つぎにどのような括弧列が受理言語となるかを判定する。簡単な判定は TEX にかけることにより受理されるかどうかによりなされる。より正確には次のように述べることができる：現れる括弧列の個数を数えることによりなされる。各段階でそれまでの開き括弧の総和が閉じ括弧の総和を超えないことなく最後にその総和が等しくなるときにかぎり受理される。例をのべる：

Acceptable words

{{{{{{{{}}}}}}



1	2	2	3	4	4	4	5	5	5
0	0	1	1	1	2	3	3	4	5

Non-acceptable words

{{{{}}}}>{{}}



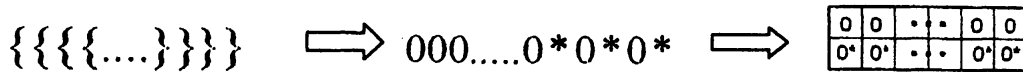
1	2	2	2	3	3	3	4	5	5
0	0	1	2	2	3	4	4	4	5

#### 4. チョムスキー言語の DNA モデル

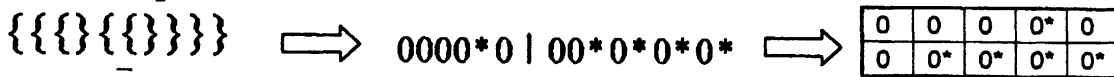
チョムスキー文のDNAモデルを構成する。現在、言語機能に対応するDNAの物質が何かは決定されるには至っておらず、我々の考察はたぶんに空想の域を出ないものであるが今後の発展を予想して述べておく([5])。まず設定を述べる：

(1)DNA の構成：DNAを構成する要素は $0, 0^*$ であり、その配列は次のようになされる：

Example 1 (標準形)



Example 2



(2)文章の分類：文章の幾つかのクラスを述べる

Primary sentences

0*	0*	...	0*	0*
----	----	-----	----	----

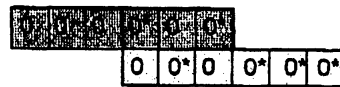
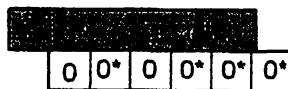
{{{{{{{{}}}}}}

Secondary sentence

0	0*	0*
---	----	----

{{}}

(3) Transposon: Transposon を次のように定める。上列と下列をずらしてその共通部分を貼り合わせる。共通部分でない部分の個数を次数という。下図ではそれぞれ degree 1 と degree 3 となる。



(4)相補性: 相補性は $0 \Leftrightarrow 0^*$ により定める(下図)。この対応においては相補性は一般に成り立たない。

Complementarities

0*	0*	0	0*	0	0*
----	----	---	----	---	----

$0 \leftrightarrow 0^*$

complementary

Sentence without complementarities

0	0*	0*	0*	0*
---	----	----	----	----

{{{{}}}}

この相補性は transposon の効果を考慮すると回復されることが示される。Transposon が存在するときその相補性を次ぎのように赤線の同一視により定める(下図)。

Transposon complementarities



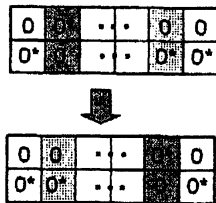
(1) Transposon of degree 1



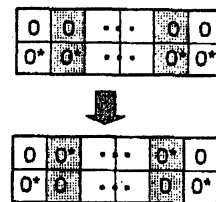
(2) Transposon degree 2,3

(5) 幾つかの演算：次に基本的な幾つかの演算をのべる：

(1) Substitution mutation



(2) Transition mutation



5. トランスポゾンによるチョムスキー文の生成

ここでは次のチョムスキー文章の生成に関する定理を証明する。

**Minimal Construction Theorem(Dynamical model)**

上記の  $0, 0^*$  の配列にチョムスキー文の相補性を導入して DNA 構造を与える。このとき

- (1) Transition mutation,
- (2) Substitution,
- (3) Transposon

によりすべてのチョムスキー文が DNA として実現できる。このときすべての演算は変形文法となる。すなわちチョムスキー文の全体は上記の演算で閉じている (§6)。また、この3つの演算の内からいずれかひとつでもはずすと構成されない文章が存在する。この意味において minimal といえる。

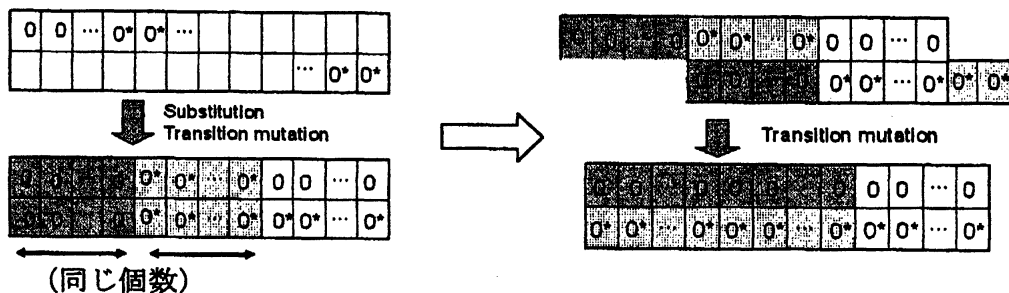
証明は難しくない。つぎの順序で標準形(Primary sentence)に帰される：

**Step 1:**最初に mutation と transition mutation を何回か行って下図左のように並べ替えることができる。

**Step 2:**つぎに transposon 演算を行って下図右のように並べ替えることができる。

**Step 3:**最後に transition mutation を何回か行って標準形にできる。

Reduction scheme



最後に minimality condition にふれる。

- (1) 長さが3となる文章はすべて (1), (3) により標準形にできる
- (2) 長さが4となる文章は次の文章以外は (1), (3) により標準形にできる。

0	0	0*	0*
0	0*	0*	0*

この文章に (2) を行うと標準形にできる。

### 6. チョムスキー言語理解の数理モデル

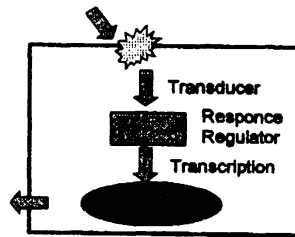
最後にシグナル伝達システムを基礎として、言語理解機能に関する数理モデルを構成する。一般に外界の刺激は受容体を経由して生命体に入り、キナーゼ等の蛋白質を刺激し核内にはいり DNA に変異を生じる。ここで刺激抑制がなされこの結果が反応として外界に出される([6])。この立場にたつて言語の理解がどのようになされるかを数理モデルを用いて考える。次のステップに従つて言語が理解されると考えてみる。

Step 1: チョムスキー文すなわち整合括弧列が読み取られる。これは単に 0 と 0\* の列が認識されることに他ならない。

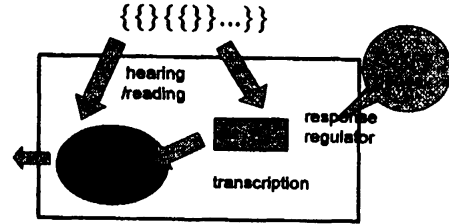
Step 2: つぎに Transposon, Transition mutation, mutation を次々を行うことにより、primary sentence にまでに reduce する。これにより文章は理解されることになる。

理解は Step 2 の演算の積がなんらかの信号に変換されて伝達することであると考える。

Signal transduction system

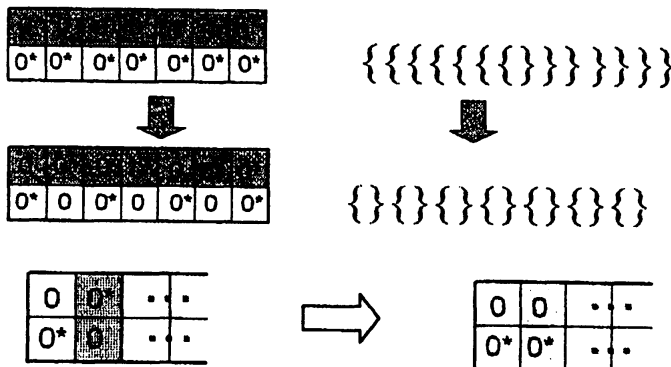


Signal system of understanding language)

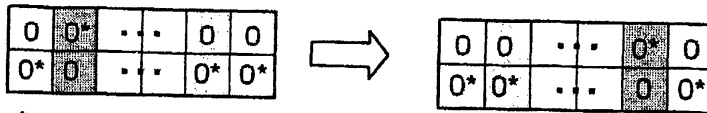


つぎに Step 2 に現れる演算の文法的な意味を考える。チョムスキー文に対する上記の演算は必ずしもチョムスキー文をチョムスキー文に写すとはかぎらない。そこでチョムスキー文を保存する演算を変形といいこれらの全体を変形文法という。以下どのような演算が変形になっているかを考える。演算が変形になるためにはチョムスキー文の成立条件がなりたちさえすればよい。無条件で変形になるものをいくつか述べる。

(1) **Transition deformation** : Transition mutation は文章を短い文章を長くしたりあるいは短い文章をまとめて長くする性質がある。次の演算は変形である。

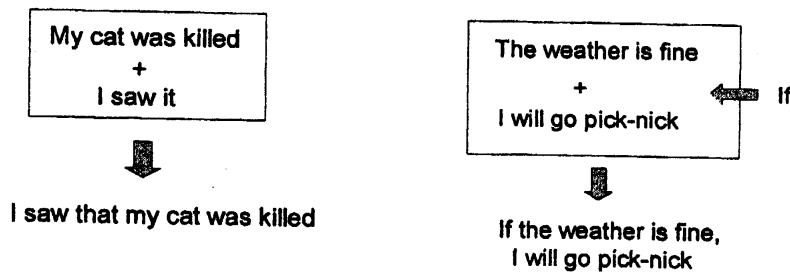


(2) **Substitution deformation** : Transition mutation は文章を短い文章を長くしたりあるいは短い文章をまとめて長くする性質がある。次の演算は変形である。



文の作成に *transposon* の作用がどのように貢献するかを考えてみる。基本的な文章を構成して次々と文章を挿入することにより、文章を構成することもある。また、幾つかの文章を想起してこれらを組み合わせて文章を構成することもある。これらの文章構成操作を行うのが *transposon* であると考えられる。次に例を取り上げこれについて述べる。

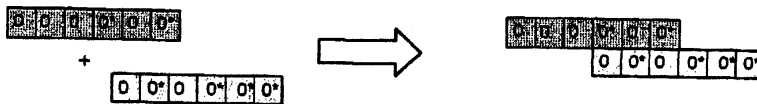
例文 I recognize the color red  
I saw that my cat was killed  
If the weather is fine, I will go pick-nick



上記 (1) の文章は意図された順序に書かれていると言ってよいであろう。これは文章としては *primary sequence* として書かれているとしてよい。

上記 (2) においてはまず驚きをもって「my cat was killed」があり、次に文章成立条件として「I saw that...」を用意するとも考えられる。

上記 (3) においては「The weather is fine」および「I go pick nick」があり、次に文章成立条件として「if」を用意したと考えられる。この最後のふたつについては次のようにも言える：



このように *transposon* に対応する演算はチョムスキー文のなかに多くみられる。今後これらについて専門家の助けをうけて考察したいと考えている。

## 7. Discussions

現在、ネアンデルタール人が言語を有しておらず、このことが現存するヒトと本質的に異なると考えられ、DNA レベルでこれを検証することがドイツマックスプランク研究所 Paabo 教授等により進められている ([5])。ヒトとチンパンジーを区別するものは *transposon* であろうと予想 (期待?) されている ([6])。我々の予想が審判されるのもそれほど遠い未来ではないであろう。今後の研究の進展が期待される。

## REFERENCE

- [1] 米田正明、広瀬貞樹(他 2 名): オートマトン・言語理論の基礎(2003)近代科学社
- [2] ハートル、ジョーンズ: エッセンシャル遺伝学(2003)培風館
- [3] G. パウン、G. ローゼンバーグ、A. サローマ: DNA コンピューティング(1999)シュプリングァー・フェアラーク東京
- [4] 斉藤成也(他 6 名): 遺伝子とゲノムの進化(シリーズ進化学 2)岩波書店
- [5] S. Paabo: Humann evolution, Millennium issue, Science(2000)14~116
- [6] 高橋秀夫: 分子遺伝学概論(1997)コロナ社
- [7] N. Chomsky: Sytactic Structures, Mouton, 1957