

自己シャフルされた記号列を入力とする
有限オートマトンについて

京産大・理 岩間 一雄

京大・工 上林 弥彦

1. まえがき

シャフル演算 \odot は、2つの記号列 x と y を、それぞれの列の記号出現順序を保存して"かみ合わせる"ことによって可能なすべての列をつくる演算であり、例えば、 $ab \odot cd = \{abcd, cdab, acbd, cabd, acdb, cadb\}$ である。この演算自体は Ginsburg⁽¹⁾によってかなり以前に導入されたが、最近それが非同期系の動作記述に利用できることが Shaw⁽²⁾によって指摘され、彼の導入したフロー表現は言語理論的にも興味深い研究対象になっている。⁽³⁾

我々は、自己シャフル入力有限オートマトンと呼ぶ

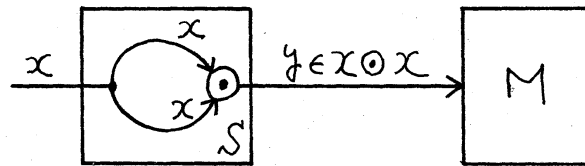


図1 自己シャフル入力有限オートマトン

れる新モデルを提案する。本モデルは、図1に示されるように、通常の有限オートマトン(受法器)Mの前段に、入力列 x に対して自己シャフル $x \odot x$ を実行して $y \in x \odot x$ を非決定

的に選択して M に送出する機構 S を有した系である。前置プロセッサ S を置くことによって、系の言語受理能力が通常の有限オートマトンと比べてどのように変化するかを議論することが本稿の目的である。本モデルの特徴がシャッフル機構にあることはいうまでもないが、注意しなければならないのは、シャッフルされる二つの列が常に等しいことである。より一般的には、シャッフルされる二つの列 x と y の間にあるかじめ知られた関係（例えば、 $x = f_1(z)$, $y = f_2(z)$ ）が常に成立するという条件のもとでのシャッフル演算を新たに提案してついでよ。

このような“条件付きシャッフル演算”の意味付けのために、これを導入するに至った動機を述べておく。著者は、複数の有限変換器を帰環なしで接続した系（図2）とその特殊な場合である縦続された系（図3）の性質について議論した⁽⁴⁾。そこでは、(i) 各有限変換器の間は容量に制限のないエラスタイクストア（一種の無限記憶）によって接続され、(ii) それらは互いに非同期的に動作し、(iii) 系へのある入力列に対し可能なすべての相対的動作順序が非決定的に選択されると仮定している。そして、縦続系では、各変換器間は無限記憶で接続されているにもかかわらずその能力は有限変換器のそれを越えないうが、一般の帰環なし接続系では有限変換器の能力を越

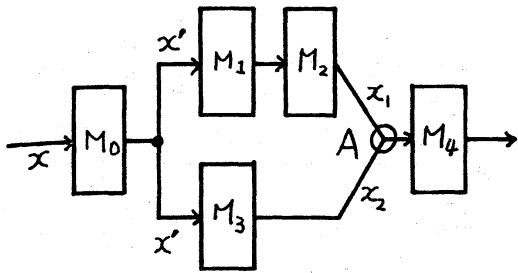


図2 帰環なし接続系

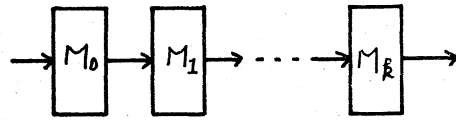


図3 縦続系

えることが示されている。図2の場合、入力列 x は2つの経路を通過して x_1 と x_2 に変換され、それが A 点で“かみ合わせ”されており、かつ、 x_1 と x_2 の間には明らかに一定の関係がある。図2と図3の2種類の形態の間に差が生じる要因がこの点に依存していることは容易に想像されよう。

2. 基本事項

Σ を有限アルファベットとする。シャフル (最近の慣例に従い \odot を用いる) は、 $\Sigma^* \times \Sigma^*$ から Σ^* の有限部分集合のクラス (2^{Σ^*}) の中への写像であり、 $x \odot y \equiv \{x_1 y_1 \dots x_k y_k \mid x = x_1 \dots x_k, y = y_1 \dots y_k, x_i, y_i \in \Sigma^*\}^0$ で定義される。自己シャフル は写像 $\pi: \Sigma^* \rightarrow 2^{\Sigma^*}$ であり、 $\pi(x) \equiv x \odot x$ である。言語 $L \subseteq \Sigma^*$ に対し、 $\pi(L) \equiv \bigcup_{x \in L} \pi(x)$ と拡張される。写像 $\bar{\pi}: \Sigma^* \rightarrow 2^{\Sigma^*}$ を $\bar{\pi}(x) \equiv \{y \mid x \in \pi(y)\}$ で定義し、逆自己シャフル と呼ぶ。 $\bar{\pi}(L)$ は π の場合と同様である。

例題 1 $\bar{\pi}(aabbbaabb) = \{abab, aabb\}$

本稿では、 L を正規言語としたとき、 $\bar{\pi}(L)$ がどのような

言語になるかという点に焦点をいぼって議論する。言語 \bar{L} を受理するモデルが前節で導入された 自己シャッフル入力有限オートマトン (Self-Shuffled Input Finite Automaton, SSIFA) である。SSIFA K の部分系 S は、入力列 x に対して $\bar{L}(x)$ の語を非決定的に選択して出力し、もしその語が有限オートマトン M によって受理されるもの (つまり $\bar{L}(x)$ に M の通常の意味での受理語が入っているもの) またその時に限って、 x は K によって受理される。 K の受理する語の全体 (K の受理言語) を $W(K)$ で表わし、 K の部分系 M の通常の意味での受理言語を $T(K)$ で表わす。 \bar{L} を受理させるためには、 $T(K) = \bar{L}$ である SSIFA K を構成すればよい。

SSIFA の関連モデルとして、 非決定性一方向スヘッド有限オートマトン⁽⁵⁾ (2-FA, 特に決定性のものを 2-DA で表わす) を思いつくが、両者の間には以下に述べるような違いがある。2-FA は図 4 に示されるように 2 つのヘッド h_1 と h_2 と有限制御部 FC を有し、 FC の状態集合は S_1 と S_2 に分割されている (テープ右端記号は使用しないモデルを考える)。もし現在の状態 λ が S_i ($i=1$ or 2) に属するもの必ずヘッド h_i の下の記号を読み (即ち FC はそれぞれのヘッドを選択するか指定できる)、 そのヘッドを 1 コマ進める。ただし、次状態は読まれた記号と λ から次状態関数で指定されるいくつかの候補の中

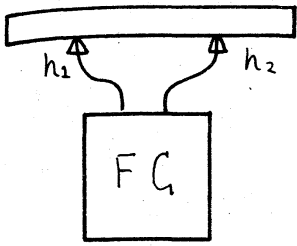


図4 2-FA

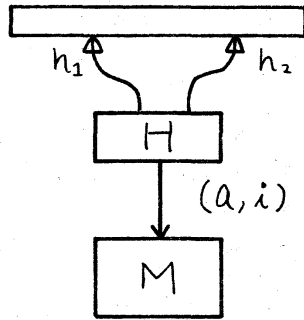


図5 2-FAと同等モデル

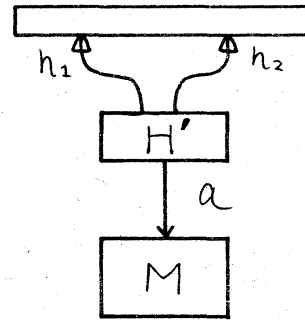


図6 SSIFAと同等モデル

から非決定的に選択できる。ところが見方を変えると、図5の様相モデルも全く同等の能力を有する(証明は省略する)。ここで、 M は通常の(決定性)有限オートマトンであり、 H は2つのヘッド h_1 か h_2 のいずれかを M とは独立に非決定的に選択してその下の記号を読み、その記号と選択された方のヘッドの番号 i ($=1$ or 2)の組を M へ出力する機構である(この考え方は著者が以前に導入した2入出力オートマトン⁽⁶⁾の考え方と類似している)。一方、SSIFAは図6の様相モデルと同等である。即ち、 H' はヘッド h_1 か h_2 のいずれかを非決定的に選択してその下の記号を M に送るところまでは図5の H と同じであるが、いずれのヘッドが選択されたかの情報が M に伝達されない。よって、2-FAが少なくともSSIFAと同等の能力を有していることは直観的に明らかである。なお、2-DAを図5のモデルとしてみた場合、 M を不完全指定オートマトンにすればよい。つまり、 M の状態集合を S_1 と S_2 に分割し、 $s_1 \in S_1$ に対しては入力 $(a, 2)$ ($\forall a \in \Sigma$)による次状態を、

又 $s_2 \in S_2$ に対しては入力 $(a, 1)$ ($\forall a \in \Sigma$) に対する次状態を未定義にする。

3. SSIFA の言語受理能力

以下, SSIFA が受理できる言語のクラスを \mathcal{L}_{SSIFA} で表わし, その範囲を考察する。2-FA, 2-DA が受理できる言語のクラスをそれぞれ \mathcal{L}_{2-FA} , \mathcal{L}_{2-DA} で, また正規, 文脈自由, 文脈依存言語のクラスをそれぞれ \mathcal{L}_{RL} , \mathcal{L}_{CFL} , \mathcal{L}_{CSL} で表わす。

補題 1 $\mathcal{L}_{SSIFA} \subseteq \mathcal{L}_{2-FA} \subsetneq \mathcal{L}_{CSL}$ 。

(証明) $\mathcal{L}_{SSIFA} \subseteq \mathcal{L}_{2-FA}$ は前節で述べた。 $\mathcal{L}_{2-FA} \subsetneq \mathcal{L}_{CSL}$ は文献 5 に述べられている。□

以下, SSIFA が受理できる文脈自由言語 (補題 2), 文脈依存言語 (補題 3) が存在するが, 他方 SSIFA が受理できない正規言語 (補題 4) も存在すること示す。また, 2-DA では受理できないが, SSIFA では受理できる言語 (補題 5) が存在することも示す。

補題 2 $L_1 \triangleq \{ab^mcd^ne \mid n \geq 0\} \in \mathcal{L}_{SSIFA}$ 。

(証明) $T(K_1) = \{ab^i ca (bd)^j ced^k e \mid i, j, k \geq 0\}$ とする SSIFA K_1 が存在することは明らかであろう。以下 $W(K_1) = L_1$ になることを示す。なお, $L_1 \subseteq W(K_1)$ はほぼ明らかであるから, $L_1 \supseteq W(K_1)$ のみ証明する。

先ず, $x \in W(K_1)$ なる $x = ab^l cd^m e$ の形をしていなければならないことを示す。 $T(K_1)$ に入っているある語 $y = ab^i ca (bd)^j c ed^k e$ が, 列 x_1 と x_2 の適当な組み合わせの結果得られたとする (x_1 と x_2 は等しい列 x であるが呼び出しのために添字を付す)。すると,

(1) x_i は必ず $ab^i c$ の形で始まらねばならず, $d^k e$ の形で終わらねばならない。例えば, y の最初に現れる $ab^i c$ が組み合わせの結果生じたものでないことは記号 a が 1 回しか現れていないことより容易に判る。よって, x_i は $ab^i c d^k e$ ($d \in \{a, b, c, d, e\}^*$) の形である。

(2) d に記号 a, c, e が現れることはありえない。なぜなら, y に記号 a, c, e はちょうど 2 回ずつ現れているからである。

(3) d に記号 b が現れることはありえない。 y に 2 回ずつ現れている a, c, e に添字を付して, $y = a_1 b^i c_1 a_2 (bd)^j c_2 e_1 d^k e_2$ と表わしてみよう。すると, (1) より $a_1 b^i c_1$ は一般性を失うことなく x_1 のプレフィクスであると仮定できるから, a_2 と c_2 は x_2 から得られた記号でなければならぬ。ところが y の c_2 より後には記号 b は現れていない。よってもし d に記号 b が現れるならば明らかに矛盾が生じる。

従って, d が空系列でないとしたら現れる記号は d のみである。よって " $x \in W(K_1)$ なる $x = ab^l cd^m e$ の形である" こと

が証明された。 $l=m$ でなければならず、このことは明らかであり、よって " $x \in W(K_1)$ ならば $x \in L_1$ である" ことが証明された。 \square

言語 L_1 において、区切り記号 a, c, e が重要な役割をはたしていることに注意すべきである。言語 $\{b^n d^m \mid n \geq 0\}$ は SSIFA では受理できないことが示せる (詳細略)。

補題3 $L_2 \equiv \{ab^m cd^n ef^m g \mid m \geq 0\} \in \mathcal{L}_{SSIFA}$ 。

(証明) 補題2と同様の方針で示せる。 \square

補題4 $L_3 \equiv \{0^a 10^b 10^c \mid a, c \geq 0, b \geq 2\} \notin \mathcal{L}_{SSIFA}$ 。

(証明) L_3 を受理する SSIFA K_3 ($W(K_3) = L_3$) が存在すると仮定して矛盾を導く。一般性を失うことなく、任意の $x \in T(K_3)$ は $x = 0^i 10^j 10^k 10^l 10^m$ (ただし、 $i+j+k+l+m$ は偶数) の形をしていると仮定できる (もしそうでなければ、直積機械の考え方を利用してそのような SSIFA を新たに作る)。そこで、

$$0^i 10^j 10^k 10^l 10^m \in 0^a 10^b 10^c \circ 0^a 10^b 10^c$$

とすると、明らかに $i \geq a$ かつ $m \geq c$ である。よって、任意に大きい整数 N_1 と N_2 に対し、 $i \geq N_1$ かつ $m \geq N_2$ である語 $0^i 10^j 10^k 10^l 10^m$ が $T(K_3)$ に入っているなければならない。この事実と $T(K_3)$ が正規集合であることに着目すると、条件

$$i \geq j+k \quad \text{かつ} \quad m \geq l+k \quad (*)$$

を満たす語 $0^i 10^j 10^k 10^l 10^m$ が $T(K_3)$ に必ず入ってのなければならぬ。以下この条件を満たす語について考える。

(1) $i+j$ が奇数でかつ $k \geq 1$ の場合。この場合 $k+l+m$ もまた奇数である。上記条件(*)より, $i-j \geq 1, m-l-k \geq 0$ 等が成り立つので, よって

$$\begin{aligned} 0^i 10^j 10^k 10^l 10^m &= 0^{\frac{i+j-1}{2} + (\frac{i-j}{2})} 10^{(j)} 10^{k+1+(1)} 10^l 10^{\frac{m-l-k+1}{2} + (\frac{m+l+k-1}{2})} \\ &\in 0^{\frac{i+j-1}{2}} 110^{k+l+\frac{m-l-k+1}{2}} \odot 0^{\frac{i-j-1}{2} + j+1} 110^{\frac{m+l+k-1}{2}} \\ &= 0^{\frac{i+j-1}{2}} 110^{\frac{m+l+k-1}{2}} \odot 0^{\frac{i+j-1}{2}} 110^{\frac{m+l+k-1}{2}} \end{aligned}$$

とかけると, このことは, $0^a 110^c$ の形の語が $W(K_3)$ に入ってしまうことになり, これは $W(K_3) = L_3$ に矛盾する。

(2) $i+j$ が奇数でかつ $k=0$ の場合。 $l+m$ も奇数になる。 $j=0$ 又は $l=0$ の場合には, $0^a 110^c$ の形の語が $W(K_3)$ に入ることが容易に示せるので, 以下, $j \neq 0$ かつ $l \neq 0$ の場合を考える。条件(*)より,

$$\begin{aligned} 0^i 10^j 110^l 10^m &= 0^{\frac{i+j-1}{2} + (\frac{i-j}{2})} 10^{1+(j-1)} 110^{l-1+(1)} 10^{\frac{m-l+1}{2} + (\frac{m+l-1}{2})} \\ &\in 0^{\frac{i+j-1}{2}} 1010^{l-1+\frac{m-l+1}{2}} \odot 0^{\frac{i-j-1}{2} + j-1} 1010^{\frac{m+l-1}{2}} \\ &= 0^{\frac{i+j-1}{2}} 1010^{\frac{m+l-1}{2}} \odot 0^{\frac{i+j-1}{2}} 1010^{\frac{m+l-1}{2}} \end{aligned}$$

とかけると, $0^a 1010^c$ の形の語が $W(K_3)$ に入ってしまうことになり矛盾する。

(3) $i+j$ が偶数の場合。

$$0^i 10^j 10^k 10^l 10^m \in 0^{\frac{i+j}{2}} 110^{\frac{m+l+k}{2}} \odot 0^{\frac{i+j}{2}} 110^{\frac{m+l+k}{2}}$$

とかけ, (1)と同様矛盾が示せる。□

補題5 $L_4 \triangleq \{ab^i cd^j e f^k g \mid i+j+k \geq 1, i=j \text{ or } j=k \text{ or } k=i\} \in \mathcal{L}_{SSIFA}$ 。

(証明) $T(K_4) = \{ab^i ca (bd)^j ced^j e f^k g f^{k'} g \mid i+i+j+k+k' \geq 2, i, i', j, k, k' \geq 0\} \cup \{ab^i cd^j ab^i ce (df)^j g e f^k g \mid i+i+j'+j+k \geq 2, i, i', j', j, k \geq 0\} \cup \{ab^i cd^j e a (fb)^k c g d^j e f^k g \mid i+j+j'+k+k' \geq 2, i, j, j', k, k' \geq 0\}$ となる SSIFA K_4 を構成すればよい (構成できることは明らかであろう)。□

定理1 (1) $\mathcal{L}_{SSIFA} \subsetneq \mathcal{L}_{2-FA} \subsetneq \mathcal{L}_{CSL}$ 。

(2) $\mathcal{L}_{SSIFA} \not\subset \mathcal{L}_{CFL}$, $\mathcal{L}_{SSIFA} \not\subset \mathcal{L}_{RL}$, $\mathcal{L}_{SSIFA} \not\subset \mathcal{L}_{2-DA}$ 。

(証明) $L_2 \notin \mathcal{L}_{CFL}$ であることはよく知られている。また明らかに $L_3 \in \mathcal{L}_{RL}$ である。 $L_4 \notin \mathcal{L}_{2-DA}$ であることは文献(7)と同様にして示せる。2-DA や 2-FA が任意の正規言語を受理できることは自明である。よって, 補題1, 3, 4, 5より定理の結果を得る。□

以上の議論より, SSIFA における自己シャッフル機構 S は, 有限オートマトンの能力を上げるように働く場合もあるし又下げるように働く場合もあることが判った。そこで, 次に興味の対象に於けるのが, 任意の有限言語が SSIFA で受理できるかどうかということである。要するに, 記号列 x の符号語として $\bar{x}(x)$ の中から一つの列 y を選ぶという符号化方式において, $\bar{x}(y) = \{x\}$ (一意復号可能) である y を常に構成できる

かどうかという点である。本問題は現在のところ未解決であるが、有限言語 L が SSIFA で受理できるための一つの十分条件が得られているのでそれを次に記す。なお、この定理より、以前に述べた言語 L_3 の任意の有限部分集合が SSIFA で受理できることが判る。従って補題 3 の結果は L_3 が無限言語であることに真に依存していることを示しており、この面からも興味深い。

定理 2 $\Sigma = \{0, 1\}$ 上の有限言語 L が、条件 " $x \in L$ ならば x に現れる記号 1 (又は 0) の数は高々 3 個である " を満たすならば、 $L \in \mathcal{L}_{SSIFA}$ である。

(証明) 条件を満たす一語の言語が SSIFA で受理できることを示せば十分である。以下では 1 の個数が 3 個の場合のみを示し、1 の個数が 2 個以下の場合は省略する。また以下の文章中の 0 と 1 を入れ換えることによって 0 の個数が 3 個の場合が証明される。

与えられた一語言語 $L = \{x\} = \{0^i 1 0^j 1 0^k 1 0^l\}$ に対し、 $T(K_5) = \{y\} = \{0^{i'} 1 0^{j'} 1 0^{k'} 1 0^{l'}\}$ とする SSIFA K_5 を構成する。この K_5 が求める SSIFA であることを示すが、 $W(K_5) \supseteq L$ は明らかであるから $W(K_5) \subseteq L$ のみを以下で証明する。即ち、 $y = x_1 \circ x_2$ ($x_1 = x_2 = 0^{i'} 1 0^{j'} 1 0^{k'} 1 0^{l'}$) ならば $x_1 = x$ ($i=i', j=j', k=k', l=l'$) と示す。 x_1 と x_2 のかみ合わせ

で y をつくる時の可能なすべての場合について、いずれの場合も $x_1 = x_2$ となることを示せばよいのであるが、ここでは、代表的な2つの場合についてだけ記述する。他の場合も同様に示せる。なお、 y の左から m 番目の記号 1 を 1_m ($m=1, \dots, 6$) で示す。

(場合1) y の 1_1 と 1_3 と 1_4 を x_1 の 1 から選ぶ場合。即ち、

$$y = 0^i \mathbf{1} 0^{i+j} 1 0^j \mathbf{11} 0^k 1 0^{k+l} 1 0^l$$

のとき (x_2 から選ばれた 1 を太字で表わしている) である。

y の 1_3 と 1_4 は連続しているから、

$$k' = 0 \tag{1}$$

でなければこのような選ぶ方はできない。すなわち y の 1_5 と 1_6 の間の 0 はすべて x_1 の最後の l 個の 0 の中から選ばれねばならない。よって、

$$l' \geq k + l \tag{2}$$

y の最後の l 個の 0 は x_2 の l' 個の 0 をすべて含む。よって、

$$l \geq l' \tag{3}$$

(1), (2), (3) より、

$$k = k' (= 0), \quad l = l' \tag{4}$$

が導かれた。そこで、

$$y = 0^i \mathbf{1} 0^{i+j} 1 0^j \mathbf{11} 1 0^l 1 0^l$$

とかける。 y の 1_1 と 1_3 の間にある $i+2j$ 個の 0 は, x_1 の j' 個の 0, x_2 の j' 個の 0 をすべてと, x_2 の先頭の i' 個の 0 の一部より成っている。よって,

$$i'+2j' \geq i+2j \quad (5)$$

さらに, y の 1_2 と 1_5 の間の j 個の 0 は x_2 の j' 個の 0 をすべて含む。よって,

$$j \geq j' \quad (6)$$

y の先頭の i 個の 0 は x_1 の先頭の i' 個の 0 をすべて含む。よって,

$$i \geq i' \quad (7)$$

(5), (6), (7) より, $i=i'$ かつ $j=j'$ が導かれ, (4) と合わせて $x_1 = x$ が得られる。

(場合 2) y の 1_1 と 1_2 と 1_3 を x_1 から選ぶ場合。即ち,

$$y = 0^i 1 0^{i+j} 1 0^j 1 1 0^k 1 0^{k+l} 1 0^l$$

の場合。以下の各式が成立することはほぼ明らかであろう。

$$i \geq i' \quad (8)$$

$$i+j \geq j' \quad (9)$$

$$j \geq k' \quad (10)$$

$$k \geq j' \quad (11)$$

$$k+l \geq k' \quad (12)$$

$$l \geq l' \quad (13)$$

ところで, I_1 と I_2 の間の $i+j$ 個の 0 は, x_2 の j' 個の 0 すべてと, x_2 の先頭の i 個の 0 の一部より成っている。よって,

$$i'+j' \geq i+j \quad (14)$$

(8) と (14) より,

$$j' \geq j \quad (15)$$

同様の考察を y の右半分に適用して,

$$k'+l' \geq k+l \quad (16)$$

$$k' \geq k \quad (17)$$

を得る。(15), (16), (17), (11) より $j' \geq j \geq k' \geq k \geq j'$ を得る。

よって,

$$j = j' = k = k' \quad (18)$$

(8), (14), (18) より

$$i = i' \quad (19)$$

(13), (16), (18) より

$$l = l' \quad (20)$$

(18), (19), (20) より $x_1 = x$ がいえた。□

4. あとがき

本稿のモデルを非同期系等の解析に適用する場合に問題になるのが, 前置シャトル機構 S から有限オートマトン M へ送られる列 y が $x \otimes x$ の中から "非決定的に選択される" と

うな事であろう。言語 L が SSIFA K で受理されるとする。すると、 L の補集合 \bar{L} の語 x は、どの $y \in \bar{L}(x)$ も K の部分系 M によって受理されるという性質、つまりかえると、 x は、すべての $y \in \bar{L}(x)$ が $M(T(K))$ の補集合を受理する有限オートマトン) によって受理されるという性質を有している。この事に着目するから、非決定的に選択されるという数学的概念が実際的应用に当たって必ずしも障害になる場合も多くと考えられる。例えば、SSIFA K の部分系 M の最終状態をある系のデッドロック状態に対応させて考えるから、 $\overline{W(K)}$ はその系にとって安全な入力集合、 $W(K)$ は安全とはいえない(デッドロックに陥る可能性のある)入力集合を表わしている。

又 SSIFA, 自己シャッフル自身, 条件付きシャッフルのより詳しい性質, スヘッド機械の条件付きシャッフルを利用した特性化, 決定問題 " $\bar{L}(x) = \phi?$ " の複雑さ等, いくつかの問題が残されているが, これらについては稿を改めて述べる予定である。

謝辞 貴重なる御意見をいただいた京都大学矢島脩三教授に感謝いたします。本研究は一部文部省科学研究費による。

文献 (1) S. Ginsburg and E.H. Spanier, "Mapping of Languages

by Two-Tape Devices," J.ACM, 12, 3, pp423-434, 1965.

(2) A.C. Shaw, "Software Descriptions with Flow Expressions," IEEE Trans. Software Eng., SE-4, 3, p242, 1978.

(3) 荒木, 都倉, "フロー表現の等価問題の決定不能性" 信学論, J62-D, 6, pp427-431, 1979.

(4) 岩間, 上林, 矢島, "部分自律有限オートマトンのル-フ-フリー-接続" 情報処理才20回大会, 3E-8, 1979.

(5) A.L. Rosenberg, "On Multi-Head Finite Automata," IBM J. Res. and Develop., 10, pp388-394, 1966.

(6) K.Iwama, Y.Kambayashi and S.Yajima, "Computer Communication Interfaces Based on Two-Input/Output-Pair Automata and Their Implementation in the LABOLINK Network," Proc. 4th ICC, Sept. 1978.

(7) O.H. Ibarra and C.E. Kim, "On 3-Head Versus 2-Head Finite Automata," Acta Informatica, 4, pp193-200, 1975.