EOL AND ETOL SYSTEMS WITH CONTROL DEVICES

by

Mogens Nielsen

DAIMI PB-37

September 1974

Institute of Mathematics University of Aarhus
DEPARTMENT OF COMPUTER SCIENCE
Ny Munkegade - 8000 Aarhus C - Denmark
Phone 06-12 83 55

Contents

Introduction 1

Definitions 2

General Results 7

RC-PartETOL is properly included in RC-PartETOL^{ac} 15

PartETOL systems with matrix and vector control ... 24

References 30

Summary

This paper is concerned with extended 0L systems and the effects (with respect to the generative capacity) on these systems caused by

1) regular control on the use of subsets of productions,
2) appearance checking,
3) minimal table interpretation,

and their combinations. Among other things it is proven that the effect of 3) is strictly stronger than the effect of both 1) and 2), and equal to the effect of the combination of 1) and 2). This implies among other things that the effect of appearance checking on the systems with regular control is significant.

Finally the notions of matrix and vector control are introduced, and the effects of these mechanisms are investigated. This leads to results very much different from corresponding well-known results for context-free grammars.

Introduction

In the theory of formal languages one has studied intensively the systems of ordinary Chomsky grammars with various kinds of regulated rewriting (see, e.g., Salomaa (1973)), that is, systems in which at any step in a derivation of the grammar, the choice of the production to be applied is somehow restricted – the restriction being determined by some kind of a control mechanism. The idea of regulated rewriting may also be introduced in the so-called Lindenmayer systems (L systems), which differ from Chomsky grammars essentially in the way of deriving words: in one step of a derivation of an L system, productions are applied in parallel to all occurrences of symbols in the string considered, i.e., a whole subset of productions is applied. But then an L system with regulated rewriting is naturally defined as a system in which at any step of a derivation the choice of this subset is somehow restricted.

In the theory of grammars with control devices one has studied in particular the power of various control mechanisms added to context-free grammars (that is, power with respect to the generative capacity). In this paper we shall study the power of some of the corresponding control mechanisms added to the class of L-systems corresponding to context-free grammars, the so-called extended informationless L-systems (E0L systems). In standard L-notation a subset of the set of productions is called a partial table, so, in other words, we shall study E0L systems with a specified set of partial tables (denoted Part ET0L systems) and with control on the use of these tables. It is well known that the power of Part ET0L systems is strictly greater than the power of E0L systems.

One of the main open questions in the theory of Chomsky grammars with regulated rewriting, is whether or not the appearance checking facility affects the generative capacity of context-free grammars with regular control. In Rozenberg (1971) partial table 0L systems, with a special interpretation of the use of a table – in this paper called the minimal table interpretation – were considered, and it was proved that the appearance checking facility does not affect the generative capacity of those systems with regular control.

In this paper we shall examine the powers of 1) regular control, 2) appearance checking, 3) minimal table interpretation, and their combinations, when added as control mechanisms to Part ET0L systems. Among other things, it is proven that the effect of 3) is strictly stronger than the effect of both 1) and 2), and equal to the effect of the combination of 1) and 2). This implies then directly that the effect of appearance checking on (extended) partial table 0L systems with regular control is significant; in other words, the results referred to from Rozenberg (1971) do not carry over to the case without the minimal table interpretation.

It should be noted that the results in Rozenberg (1971) are stated for programmed partial table 0L systems. It is, however, easy to prove that the generative capacity of programmed (extended) partial table 0L systems (with appearance checking) is equal to the generative capacity of (extended) partial table 0L systems with regular control (with appearance checking). The methods used in this proof are the same as in the case of context-free grammars.

In the proof of the main result of this paper, a characterization result for ET0L languages is proved somewhat related to the result proved in Ehrenfeucht and Rozenberg (1973) – ideas and notation have been taken from this paper.

In the last section, the notions of matrix and vector control are introduced in extended partial table 0L systems, and some results on the effects of these mechanisms are proved. These results turn out to be very much different from the corresponding well-known results for context-free grammars. The differences illustrate in some sense the different nature of control in context-free grammars and extended partial table 0L systems.

Notation

The following notation is used in this paper:

Let Σ be a finite alphabet, and $x \in \Sigma^*$, then

$|\Sigma|$ denotes the cardinality of Σ;

$|x|$ denotes the length of x;

$\min(x) = \{\sigma \in \Sigma \mid \sigma \text{ occurs in } x\}$;

Let t be a finite subset of $\Sigma \times \Sigma^*$, then

$reg(t) = \{\sigma \in \Sigma \mid \exists\ x \in \Sigma^* : (\sigma, x) \in t\}$;

$\emptyset$ denotes the empty set;

λ denotes the empty string;

N denotes the set of natural numbers;

Definitions

Definition 1

A partial ETOL system (Part ETOL) is an ordered sixtuple:
$S = \langle \Sigma, P, L, L^{ac}, \omega, \Delta \rangle$, where

Σ is a finite alphabet;

P is a finite, nonempty collection of finite subsets of $\Sigma \times \Sigma^*$ – the elements of P are called tables – the elements of the tables are called productions;

L is a set of labels of P (that is a finite alphabet with an associated one-to-one correspondance between the elements of P and L);

L^{ac} is a subset of L;

ω is an element of Σ^+ – the axiom of the system;

Δ is a subset of Σ – the set of terminals;

Definition 2

Let $S = \langle \Sigma, P, L, L^{ac}, \omega, \Delta \rangle$ be a Part ETOL system, and let $x, y \in \Sigma^*$, $l \in L$. x is said to derive y directly applying the table labelled l in S, $x \underset{S}{\Rightarrow} y\ (l)$, iff

1) $x = x_1 x_2 \ldots x_k$, for some k where $x_i \in \Sigma$ for $i = 1, 2, \ldots, k$;

2) $y = y_1 y_2 \ldots y_k$, where $y_i \in \Sigma^*$ for $i = 1, 2, \ldots, k$;

3) $(x_i, y_i) \in t$ for $i = 1, 2, \ldots, k$, where t is the table labelled l;

x is said to derive y directly under the minimal table interpretation applying the table labelled l, $x \underset{S\,mt}{\Rightarrow} y\ (l)$, iff

1), 2) and 3) above are satisfied and

4) $reg(t) = min(x)$;

Definition 3

Let $S = \langle \Sigma, P, L, L^{ac}, \omega, \Delta \rangle$ be a Part ETOL system and let $x \in \Sigma^+$, $y \in \Sigma^*$, and $z \in L^*$. Then x is said to derive y with control word z (under the minimal table interpretation), $x \overset{*}{\underset{S}{\Rightarrow}} y\ (z)$ $(x \overset{*}{\underset{S\,mt}{\Rightarrow}} y\ (z))$, iff

1) $z = l_1 l_2 \cdot \ldots \cdot l_d$, where $l_i \in L$ for $i = 1, 2, \ldots, d$;

2) there exist words $x_0, x_1, \ldots, x_d$ from Σ^* such that $x_0 = x$, $x_d = y$, and $x_{i-1} \underset{S}{\Rightarrow} x_i(l_i) (x_{i-1} \underset{S\,mt}{\Rightarrow} x_i(l_i))$ for every $i = 1, 2, \ldots, d$;

x is said to derive y with control word z (under the minimal table interpretation) with appearance checking, $x \overset{*\,ac}{\underset{S}{\Rightarrow}} y\ (z)$ $(x \overset{*\,ac}{\underset{S\,mt}{\Rightarrow}} y\ (z))$ iff

1) $z = l_1 l_2 \cdot \ldots \cdot l_d$, where $l_i \in L$ for $i = 1, 2, \ldots, d$;

2) there exist words $x_0, x_1, \ldots, x_d$ from Σ^* such that $x_0 = x$, $x_d = y$, and for every $i = 1, 2, \ldots, d$ (t_i denotes the table associated with the label l_i):

if $min(x_{i-1}) \subseteq reg\ (t_i)$ $(min(x_{i-1}) = reg\ (t_i))$ then $x_{i-1} \underset{S}{\Rightarrow} x_i(l_i)$

$(x_{i-1} \underset{S\,mt}{\Rightarrow} x_i(l_i))$, otherwise $l_i \in L^{ac}$ and $x_{i-1} = x_i$;

Now for any $x, y \in \Sigma^*$ and $z = l_1 l_2 \ldots l_d \in L^*$ for which $x \underset{S^{\alpha}}{\Rightarrow}^{\beta} y\ (z)$, where α may be the index mt or not, and correspondingly β may be the index ac or not, you may define a derivation D of y from x with control word z as a complete tree structure containing information about

1) strings x_i satisfying the proper definition above, $i = 1, 2, \ldots, d$;
2) which productions from the table labelled l_i are used on the specific occurences of symbols in x_{i-1};

To save a lot of cumbersome notation, a derivation D will just be written on the form

$$D : \quad x = x_0 \underset{l_1}{\Rightarrow} x_1 \Rightarrow \ \ldots \ \underset{l_d}{\Rightarrow} x_d = y.$$

The length of D is defined as the length of D's control word. Furthermore, for α and β as defined above, you may define the corresponding languages generated by S as

$$L_{\alpha}^{\beta}(S) = \{x \in \Delta^* \mid \exists\, z \in L^* : \omega \underset{S\alpha}{\overset{*\beta}{\Rightarrow}} x\ (z)\ \}$$

Definition 4

A Part ETOL system with regular control, RC-Part ETOL, is a seventuple $S = <\Sigma, P, L, L^{ac}, \omega, \Delta, R>$, where $S' = <\Sigma, P, L, L^{ac}, \omega, \Delta>$ is a Part ETOL system, and R is a regular language over L. Let α and β as in Definition 3 be variables that index the relations defined in Definition 3 properly, then the corresponding languages generated by S are defined by

$$L_{\alpha}^{\beta}(S) = \{x \in \Delta^* \mid \exists z \in R : \omega \overset{*}{\underset{S',\alpha}{\Rightarrow}}{}^{\beta} x(z)\}$$

Definition 5

A Part ETOL system $S = \langle \Sigma, P, L, L^{ac}, \omega, \Delta \rangle$ is called complete or usually just an ETOL system, ETOL, iff for every $t \in P$ and every $\sigma \in \Sigma$ there exists an $x \in \Sigma^*$ such that $(\sigma, x) \in t$. S is called deterministic, Part EDTOL, iff for every $t \in P$ and every $\sigma \in \Sigma$ there exists at most one $x \in \Sigma^*$ such that $(\sigma, x) \in t$. S is called propagating, Part EPTOL, iff for every $t \in P$ and every $\sigma \in \Sigma$, $(\sigma, \lambda) \notin t$. S is called total or usually just a Part TOL system, Part TOL, iff $\Sigma = \Delta$. One may also combine these notions and speak, for instance, of an EPDTOL system.

Notation

In the following the abbreviations introduced for the various kinds of systems and the languages generated by these systems, are also used for the corresponding classes of languages generated. E.g., Part ETOL_{mt}^{ac} denotes the class of languages generated by Part ETOL systems under the minimal table interpretation and with appearance checking. The next two sections will be an investigation of the relations between the following classes of languages

ETOL

Part ETOL	Part ETOL_{mt}	Part ETOL^{ac}	Part ETOL_{mt}^{ac}
RC-Part ETOL	RC-Part ETOL_{mt}	RC-Part ETOL^{ac}	RC-Part ETOL_{mt}^{ac}

General Results

Theorem 1

$$\text{Part ETOL} = \text{Part ETOL}^{ac}$$

$$\text{Part ETOL}_{mt} = \text{Part ETOL}_{mt}^{ac}$$

Proof

The theorem follows directly from the observation that for any Part ETOL system, S, $L(S) = L^{ac}(S)$ $(L_{mt}(S) = L_{mt}^{ac}(S))$.

The following theorem is trivially seen to be true, and is therefore stated without any proof.

Theorem 2 ETOL = Part ETOL.

Theorem 3   Part ETOL = RC-Part ETOL.

Proof

Obviously it is sufficient to prove that for any RC-Part ETOL system there exists a Part ETOL system generating the same language. Let $S = <\Sigma, P, L, L^{ac}, \omega, \Delta, R>$ now be an RC-Part ETOL system, and let $M = <Q, L, \delta, F, q_0>$ be an ordinary finite-state acceptor for R, where Q is the set of states, L the input alphabet, δ the transition function, $\delta : Q \times L \rightarrow Q$, F the set of accepting states, $F \subseteq Q$, and q_0 the initial state, $q_0 \in Q$. Now a Part ETOL system, S', will be constructed such that $L(S) = L(S')$. The alphabet of S' is $\Sigma' = \Sigma \times Q \cup \Delta$, and the terminal alphabet Δ. The basic idea in the construction of S' is that the second component of the nonterminals in the alphabet of S' will keep track of the used control word with respect to membership of R. If $\omega = \sigma_1\sigma_2 \cdot \ldots \cdot \sigma_p$, where $\sigma_i \in \Sigma$ for $i = 1, 2, \ldots, p$, then the axiom of S' is equal to the string $\omega' = (\sigma_1, q_0) \cdot \ldots \cdot (\sigma_p, q_0)$. The tables of S', P', are constructed as follows:

1) for every table $t \in P$ with associated label $l \in L$, containing productions

$$(\sigma_i, \sigma_{i_1} \sigma_{i_2} \cdot \ldots \cdot \sigma_{i_{n_i}}) \text{ for } i = 1, 2, \ldots, n$$

where the σ_i's and the σ_{i_j}'s belong to Σ, and for every $q \in Q$, P' contains a table with productions

$$((\sigma_i, q), (\sigma_{i_1}, \delta(q, l)) \cdot \ldots \cdot (\sigma_{i_{n_i}}, \delta(q, l))) \text{ for } i = 1, 2, \ldots, n$$

2) for every $q \in F$, P' contains a table with productions

$$((\sigma, q), \sigma) \text{ for every } \sigma \in \Delta.$$

So $|P'| = |P| \cdot |Q| + |F|$.

Now let $S' = <\Sigma', P', L', \emptyset, \omega', \Delta>$, where L' is a set of labels of the constructed P'. Clearly S' simulates both the derivations of S and the behaviour of M (reading the control words of these derivations). A terminal word is generated (using tables from 2) above) in S' iff the word generated in the simulated derivation from S is terminal (from Δ^*) and the control word used is accepted by M. This proves the theorem.

Theorem 4 Part ETOL_{mt} = RC-Part ETOL_{mt}.

Proof

The proof is identical to the proof of Theorem 3, with the exception that the set of tables of the constructed system, S', contains some more "terminal" tables.

Theorem 5 RC-Part $\mathrm{ETOL}^{ac} \subseteq$ RC-Part ETOL_{mt}.

Proof

Let $S = <\Sigma, P, L, L^{ac}, \omega, \Delta, R>$ be an RC-Part ETOL system. Another RC-Part ETOL system $S' = <\Sigma, P', L', \emptyset, \omega, \Delta, R'>$ will be constructed such that $L^{ac}(S) = L_{mt}(S')$. Let K be a set of labels for the set of nonempty subsets of Σ, then $L' = L \times K$. For every table $t \in P$ with associated label l and every nonempty subset of Σ, $\Sigma' \subseteq \Sigma$, with associated label $k \in K$, construct the following table $t_{\Sigma'}$, and associate with it the label $(l, k) \in L'$:

1) if $\Sigma' \subseteq reg(t)$ then $t_{\Sigma'}$ contains all productions from t of the form (σ, x) where $\sigma \in \Sigma'$,
2) otherwise $t_{\Sigma'}$ contains identity productions, (σ, σ), for all $\sigma \in \Sigma'$.

Let now P' be the set of all the constructed tables $t_{\Sigma'}$, and let R' be the regular language over L' obtained from R by the following finite substitution $\varphi : L \to L'$,

$$\forall\ l \in L^{ac} : \varphi(l) = \{(l, k) \mid k \in K\}$$

$$\forall\ l \in L\backslash L^{ac} : \varphi(l) = \{(l, k) \mid reg(t) \supseteq \Sigma', \text{ where } l \text{ and } k \text{ are the labels associated with } t \text{ and } \Sigma' \text{ resp.}\}$$

Then clearly $L^{ac}(S) = L_{mt}(S)$, and the theorem is proved.

Theorem 6 RC-Part $ETOL^{ac}_{mt}$ = RC-Part $ETOL_{mt}$.

Proof

RC-Part $ETOL_{mt}$ is included in RC-Part $ETOL^{ac}_{mt}$ by definition.

Let S then be any RC-Part ETOL system. Another RC-Part ETOL system S' is now constructed exactly like in the proof of Theorem 5 with the following exceptions:

a) The construction of the tables $t_{\Sigma'}$ is now:

1) if $\Sigma' = reg(t)$ then $t = t_{\Sigma'}$,

2) otherwise $t_{\Sigma'}$ contains identity productions for all symbols from Σ';

b) the finite substitution $\varphi : L \to L'$ is now defined by:

$\forall\ l \in L^{ac}$ $\quad : \varphi(l) = \{(l,k) \mid k \in K\}$

$\forall\ l \in L \setminus L^{ac}$, where l is associated with the table t: $\varphi(l) = \{(l,k) \mid k$ is the label associated with $reg(t)\}$

Clearly the constructed system S' satisfies $L^{ac}_{mt}(S) = L_{mt}(S')$, and the theorem is proved.

<u>Theorem 7</u> $\quad$ RC-Part $ETOL_{mt} \subseteq$ RC-Part $ETOL^{ac}$.

<u>Proof</u>

Let $S = \langle \Sigma, P, L, L^{ac}, \omega, \Delta, R \rangle$ be any RC-Part ETOL system with $\Sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$. Another RC-Part ETOL system will now be constructed such that $L_{mt}(S) = L^{ac}(S')$.
Let $\overline{\Sigma} = \{\overline{\sigma}_1, \ldots, \overline{\sigma}_n\}$ be a set of barred versions of the symbols from Σ, $\overline{\Sigma}$ disjoint from Σ, and let ¢ be a symbol not in $\Sigma \cup \overline{\Sigma}$. The alphabet of S' will then be $\Sigma' = \Sigma \cup \overline{\Sigma} \cup \{¢\}$. The set of labels for the tables of S' will be $L' = L \times \{1, 2, \ldots, n+2\}$. The tables of S', P',

are constructed and labeled as follows:

for every $t \in P$ with associated label $l \in L$, let

1) for $i = 1, 2, \ldots, n$, t_i be the table containing productions of the form $(\sigma, \bar{\sigma})$ for every $\sigma \in \mathrm{reg}\,(t) \setminus \{\sigma_i\}$ and associate with t_i the label $(l, i) \in L'$;
2) t_{n+1} be the table containing productions $(\sigma, \bar{\sigma})$ and $(\bar{\sigma}, ¢)$ for every $\sigma \in \mathrm{reg}\,(t)$, and associate with t_{n+1} the label $(l, n+1) \in L'$;
3) t_{n+2} be the table containing productions $(\sigma, ¢)$ for every $\sigma \in \Sigma$, and $(\bar{\sigma}, x)$ for every $(\sigma, x) \in t$, and associate with it the label $(l, n+2) \in L'$;

P' consists of all these $|P| \cdot (n+2)$ tables.

Let φ be the homomorphism from L into L' defined by:

if $l \in L$ is associated with $t \in P$, and $\mathrm{reg}\,(t) = \{\sigma_{i_1}, \sigma_{i_2}, \ldots, \sigma_{i_p}\}$ then $\varphi(l) = (l, i_1)(l, i_2) \cdot \ldots \cdot (l, i_p)(l, n+1)(l, n+2)$.

Finally define $S' = <\Sigma', P', L', L', \omega, \Delta, \varphi(R)>$. Then S' is a well-defined RC-Part ETOL system for which $L_{mt}(S) = L^{ac}(S')$, and the theorem is proved.

Now since by definition RC-Part ETOL is included in RC-Part ETOL^{ac}, the results from this section can be summarized in the following diagram:

$$
\begin{array}{l}
\text{RC-Part ETOL}^{ac}_{mt} = \\
\text{Part ETOL}_{mt} = \text{Part ETOL}^{ac}_{mt} = \\
\text{RC-Part ETOL}^{ac} = \text{RC-Part ETOL}_{mt} \\
\quad \cup\!\! \mid \\
\text{RC-Part ETOL} = \text{ETOL} = \\
\text{Part ETOL} = \text{Part ETOL}^{ac}
\end{array}
$$

Diagram 1.

Some interesting corollaries of the results of this section will now be discussed.

In Salomaa (1968) the notion of full checking is introduced (for time varying grammars). A Chomsky grammar with regulated rewriting is said to work under the full checking interpretation iff all productions are applied with the appearance checking interpretation. Correspondingly, in the notation of this paper, a $\mathrm{PartETOL}^{ac}$ system, $S = <\Sigma, P, L, L^{ac}, \omega, \Delta>$, is said to work under the full checking interpretation iff $L = L^{ac}$.

It follows from results in Salomaa (1968) that the generative power of context-free grammars with regular control and with full checking is strictly smaller than the one of context-free grammars with regular control and with ordinary appearance checking. Note, however, that it follows directly from the proofs of Theorems 5 and 7, that any $\mathrm{RC\text{-}PartETOL}^{ac}$ system can be simulated by an RC-PartETOL system with full checking, i.e., the following corollary follows.

<u>Corollary 1</u>

The generative capacity of $\mathrm{RC\text{-}Part\ ETOL}^{ac}$ is equal to the generative capacity of RC-Part ETOL systems working under the full checking interpretation.

Let us also consider Diagram 1 for deterministic systems.

<u>Theorem 8</u> $\quad$ $\mathrm{Part\ EDTOL} \supseteq \mathrm{Part\ EDTOL}_{mt}$.

Proof

Let $S = <\Sigma, P, L, \emptyset, \omega, \Delta>$ be a Part EDTOL system. The following observation is easily checked to be true: For any string $x \in \Sigma^*$ and any table $t \in P$ with associated label $l \in L$, the min-value of the string y satisfying $x \underset{l}{\Rightarrow} y$, is uniquely determined by t and the min-value of x (note that this observation is only true in the deterministic case). But now the mt-interpretation is equivalent to squeezing out from L^* those control words, z, for which the corresponding sequence of reg-values is equal to the sequence of min-values associated with the (unique) derivation from ω with control word z. But from the observation above it follows immediately that the set of all control words with this property is a regular language over L, i.e., $L_{mt}(S) = L(S,R)$ for some regular language $R \subseteq L^*$. But now the construction of Theorem 3 will give you a Part EDTOL system S', such that L(S,R) = L(S'), and this proves the theorem (all technical details are left to the reader).

The reader may verify that indeed all theorems of this sections are also true for deterministic systems as well. But from this and Theorem 8 you get the following corollary which is interesting in that it will be shown in the next section that the corresponding result is not true for nondeterministic systems.

Corollary 2

Let α (resp. β) be the index mt (resp. ac) or missing, then

$$\text{EDTOL} = \text{Part EDTOL}_{\alpha}^{\beta} = \text{RC-Part EDTOL}_{\alpha}^{\beta} .$$

RC-Part ETOL is properly included in RC-Part ETOL^{ac}

In this section it will be proved that the inclusion of Diagram 1 is proper.

Definition 6

Let L be a language over some alphabet Δ. $\delta \in \Delta$ is said to be distributed iff for every integer n, there exists a word $x \in L$, such that x contains more than two occurrences of δ and the distance between any two occurrences of δ is greater than n. δ is said to be periodically occurring in L iff for every word $x \in L$ there exist integers n and m greater than or equal to two, such that $x \in (\delta(\Delta - \{\delta\})^n)^m$. δ is said to be tied in L iff for every integer k, there exists an integer n_k, such that for any word $x \in L$, if x contains less than k occurrences of δ, then the distance between any two consecutive occurrences is less than n_k.

Definition 7

Let $S = <\Sigma, P, L, L^{ac}, \omega, \Delta>$ be a Part ETOL system. The PartEDTOL system associated with S, denoted Assoc (S), is defined by

$$\text{Assoc}\ (S) = <\Sigma, P', L', L'^{ac}, \omega, \Delta>,$$

where L' is a set of labels for P'; for any table $t \in P$ and any deterministic subset of t, P' contains one (deterministic) table, t', identical to this subset; the label of t' is in L'^{ac} iff the label of t is in L^{ac}.

Lemma 1

Let S be a Part ETOL system generating a language $L(S)$ over some alphabet Δ. If $\delta \in \Delta$ is periodically occurring (tied) in $L(S)$, then δ is also periodically occurring (tied) in $L(\text{Assoc}(S))$.

Proof

It follows from the definition of Assoc (S) that $L(\text{Assoc}(S)) \subseteq L(S)$, and from this observation the lemma follows directly.

Lemma 2

Let S be a Part ETOL system, $S = <\Sigma, P, L, L^{ac}, \omega, \Delta>$, generating a language $L(S)$ over Δ. If $\delta \in \Delta$ is distributed and periodically occurring in $L(S)$, then δ is distributed in $L(\text{Assoc}(S))$.

Proof

Let n be any integer and let

$$D : \omega = x_0 \underset{l_1}{\Rightarrow} x_1 \underset{l_2}{\Rightarrow} \dots \underset{l_d}{\Rightarrow} x_d = x$$

be a derivation in S of a word $x \in (\delta(\Delta - \{\delta\})^k)^m$, where $k \geq n$ and $m \geq 2$. For $i = 0, 1, \dots, d$, x_i can be written on the form $x_i = y_i \sigma_i z_i$ where $y_i, z_i \in \Sigma^*$, and $\sigma_i \in \Sigma$ is the unique symbol from x_i that generates in D the rightmost occurrence of δ in x. Now, for any $i = 1, 2, \dots, d$, define s_i as any (but fixed) deterministic subset of t_i, the table labelled l_i, satisfying

1) reg (s_i) = reg (t_i);

2) the production used in D on the specific occurrence of σ_i mentioned above (generating the rightmost occurrence of δ in x) belongs to s_i;

Now consider first the derivation D' in S, obtained from D with the modification that only productions from s_i are used on symbols occurring in z_{i-1}. Then clearly the derivation of y_d is left unchanged and since δ is periodically occurring in L(S) this implies that the word derived in D', x', belongs to $y_d(\delta(\Delta - \{\delta\})^k)^+$, $x' = y_d z'_d$. Consider next the derivation D'' in S, obtained from D' with the modification that only productions from s_i are used in the i'th step of D''. Then clearly the derivation of z'_d is left unchanged, and this implies, since δ is periodically occurring in L(S), that the word derived in D'' belongs to $(\delta(\Delta - \{\delta\})^k)^{m'}$ for some $m' \geq 2$. But D'' is a derivation in Assoc (S) and this proves the lemma.

<u>Definition 8</u>

Let $S = \langle \Sigma, P, L, L^{ac}, \omega, \Delta \rangle$ be a Part EPTOL system, and let

$$D : x_0 \underset{l_1}{\Rightarrow} x_1 \underset{l_2}{\Rightarrow} \ldots \underset{l_d}{\Rightarrow} x_d$$

be a derivation in S. Then a step in D, $x_{i-1} \underset{l_i}{\Rightarrow} x_i$, is called <u>essential</u> iff $|x_{i-1}| < |x_i|$. The <u>essential length</u> of D is defined as the number of essential steps in D.

Lemma 3

Let $S = \langle \Sigma, P, L, L^{ac}, \omega, \Delta \rangle$ be a Part EPDTOL system. Then any string $y \in \Sigma^+$ derived from some string $x \in \Sigma^+$ in a derivation of essential length less than k, is also derived from x, in a derivation in S of length less than $k \cdot |\Sigma|! \cdot 2^{|\Sigma|}$.

Proof

Let D be a derivation as in Definition 8 of essential length less than k, $x_0 = x$, $x_d = y$, and $d > k \cdot |\Sigma|! \cdot 2^{|\Sigma|}$. Then there exists a sequence of consecutive steps in D, $x_i \underset{l_{i+1}}{\Rightarrow} x_{i+1} \underset{l_{i+2}}{\Rightarrow} \ldots \underset{l_j}{\Rightarrow} x_j$ including no essential steps, such that $j-i > |\Sigma|! \cdot 2^{|\Sigma|}$. But then there exists a sequence of indices, $i_1, i_2, \ldots, i_m$, satisfying

1) $i \leq i_1 < i_2 < \ldots < i_m \leq j$;
2) $m > |\Sigma|!$;
3) $\min(x_{i_1}) = \min(x_{i_p})$ for $p = 2, 3, \ldots, m$;

Now since S is deterministic and propagating the tables applied in D between any of these x_{i_p}'s, $p = 1, 2, \ldots, m$, determine uniquely a permutation in $\min(x_{i_1})$. But since the group of permutations of this set contains no more than $|\Sigma|!$ elements it follows that there exist i_q and i_r such that $i_q \neq i_r$ and $x_{i_q} = x_{i_r}$. But now obviously the steps in D between x_{i_q} and x_{i_r} may be omitted, i.e., there exists a derivation D' of y from x of length equal to the length of D subtracted by $(i_r - i_q)$. This process may be repeated until a derivation of y from x of length less than $k \cdot |\Sigma|! \cdot 2^{|\Sigma|}$ is obtained, and this proves the lemma.

Lemma 4

Let S be a Part ETOL system generating a language $L(S)$ over some alphabet Δ. If $\delta \in \Delta$ is distributed and periodically occurring in $L(S)$, then δ is not tied in $L(S)$.

Proof

It follows from Nielsen, Rozenberg, Salomaa and Skyum (1974) and Theorem 2, that there exists an EPTOL system T such that $L(T) = L(S)\setminus\{\lambda\}$. Now the assumptions of the lemma imply that δ is distributed and periodically occurring in $L(\text{Assoc}(T))$ (Lemma 1 and Lemma 2). It will now be proved that δ cannot be tied in $L(\text{Assoc}(T))$ which then proves the lemma by Lemma 1.

Let $\text{Assoc}(T) = \langle\Sigma, P, L, L^{ac}, \omega, \Delta\rangle$.

Define $g = |\Sigma|! \cdot 2^{2\cdot|\Sigma|}$, and $n = \max\{|x| \mid x$ is derived in Assoc (T) in less than g essential steps from ω or from some $\sigma \in \Sigma\}$. Note that n is welldefined from Lemma 3.

Let x now be a word generated by Assoc (T), $x \in (\delta(\Delta - \{\delta\})^k)^m$, where $k > n$ and $m \geq 2$, and let D be a derivation of x in Assoc (T),

$$D : \omega = x_0 \underset{l_1}{\Rightarrow} x_1 \underset{l_2}{\Rightarrow} \cdots \underset{l_d}{\Rightarrow} x_d = x.$$

Define i_0 as the largest integer for which there is an occurrence of a letter $\sigma \in \Sigma$ in x_{i_0} generating two occurrences of δ in x in D, and $i_0 = 0$ if no such x_{i_0} exists. It follows from the definition of n that the derivation

$$D' : x_{i_0} \underset{l_{i_0+1}}{\Rightarrow} x_{i_0+1} \Rightarrow \cdots \underset{l_d}{\Rightarrow} x_d = x$$

contains at least g essential steps. This implies that there exists a sequence $i_1, i_2, \ldots, i_f$ such that

1) $i_0+2 \leq i_1 < i_2 < \ldots < i_f \leq d$;
2) $f > |\Sigma|! \cdot 2^{|\Sigma|}$;
3) $x_{i_p-1} \underset{l_{i_p}}{\Rightarrow} x_{i_p}$ is an essential step in D for $p = 1, 2, \ldots, f$;
4) $\min(x_{i_1}) = \min(x_{i_p})$ for $p = 2, 3, \ldots, f$;

Now define for every i_p in this sequence $\min_\delta(x_{i_p})$ as the set of symbols in x_{i_p} that contributes in D to an occurrence of a δ in x. Then clearly there exists a subsequence, $j_1, j_2, \ldots, j_h$, of the above sequence, such that

1) $i_0+2 \leq j_1 < j_2 < \ldots < j_h \leq d$;
2) $h > |\Sigma|!$;
3) $x_{j_p-1} \underset{l_{j_p}}{\Rightarrow} x_{j_p}$ is an essential step in D for $p = 1, 2, \ldots, h$;
4) $\min(x_{j_1}) = \min(x_{j_p})$ for $p = 2, 3, \ldots, h$;
5) $\min_\delta(x_{j_1}) = \min_\delta(x_{j_p})$ for $p = 2, 3, \ldots, h$;

Now the deterministic tables used in D between these x_{j_p}'s, determine uniquely a permutation in $\min_\delta(x_{j_1})$. But since the group of permutations in this set contains no more than $|\Sigma|!$ elements, it follows that there exist j_q and j_r such that $j_q < j_r$ and the permutation in $\min_\delta(x_{j_1})$ defined by the tables used in D between x_{j_q} and x_{j_r} is the identity.

Consider now for any integer i the derivation in Assoc (T) from ω with control word

$$z_i = l_1 l_2 \cdot \ldots \cdot l_{j_q} (l_{j_q+1} \cdot \ldots \cdot l_{j_r})^i l_{j_r+1} \cdot \ldots \cdot l_d$$

From the construction above it follows that

1) For every integer i, the word derived in Assoc (T) from ω with control word z_i contains m occurrences of δ;
2) for every integer n there exists an integer i such that the word derived in Assoc (T) from ω with control word z_i is of length longer than n;

Now 1) and 2) above and the fact that δ is periodically occurring in L (Assoc (T)) shows that δ cannot be tied in L (Assoc (T)), and this proves the lemma.

Finally we are able to prove the main results of this section.

<u>Theorem 9</u> The language $L = \{(ab^m)^n \mid 2 \leq m \leq n\}$ does not belong to Part ETOL.

<u>Proof</u>

Follows directly from Lemma 4, since the symbol a is distributed, periodically occurring and tied in L.

<u>Theorem 10</u> The language L of Theorem 9 does belong to RC-Part $ETOL_{mt}$.

<u>Proof</u>

Consider the RC-Part ETOL system $S = <\Sigma, P, L, L^{ac}, \omega, \Delta, R>$, where

$\Sigma = \{S, S_1, X, X_1, X_2, X_3, a, b\}$;

P consists of the following tables, labelled with the set $L = \{1, 2, 3, 4, 5\}$:

$$1: \begin{Bmatrix} (S_1 & , & S_1 S) \\ (S & , & S) \\ (X_1 & , & X_1 X) \\ (X & , & X) \end{Bmatrix} \qquad 2: \begin{Bmatrix} (S_1 & , & Sb) \\ (S & , & Sb) \\ (X_1 & , & X) \\ (X & , & X) \end{Bmatrix}$$

$$3: \begin{Bmatrix} (S & , & Sb) \\ (b & , & b) \\ (X & , & X_2) \\ (X & , & X_3) \end{Bmatrix} \qquad 4: \begin{Bmatrix} (S & , & S) \\ (b & , & b) \\ (X_2 & , & X) \\ (X_3 & , & \lambda) \end{Bmatrix} \qquad 5: \begin{Bmatrix} (S & , & a) \\ (b & , & b) \\ (X & , & \lambda) \end{Bmatrix}$$

$L^{ac} = \emptyset$;

$\omega = S_1 S X_1 X$;

$\Delta = \{a, b\}$;

$R = 1^* \; 2 \; (3 \; 4)^* \; 5$;

It is now very easy to prove that $L = L_{mt}(S)$.

Theorem 3, Theorem 7 and the last two theorems imply directly the following theorem, which states that the inclusion in the final diagram of the previous section is proper.

Theorem 11 RC-Part ETOL $\subsetneq$ RC-Part ETOL^{ac}.

Theorem 11 has the following corollary, which is of some interest in relation to the results in Rozenberg (1971).

Corollary 3 RC-Part TOL $\subsetneq$ RC-Part TOL^{ac}.

Part ETOL systems with matrix and vector control

As mentioned in the introduction, it is still an open question, whether or not the appearance checking facility affects the generative capacity of context-free grammars with regular control. One might think that a positive solution to this question could be given along the same lines as the proof of Theorem 11. However, the reader who is familiar with the theory of context-free grammars with regulated rewriting, will have noticed that not only is the parallelism of Part ETOL systems used many times in the proofs of the previous sections, but the whole nature of (regular) control in Part ETOL systems is very much different from the one of (regular) control in context-free grammars. This difference will be illustrated in this section, in which matrix and vector control will be defined for Part ETOL systems. Results on the effect of these control mechanisms will be proved, and some of these results turn out to be very much different from corresponding well-known results for context-free grammars.

Definition 9

Let $x = \sigma_1 \sigma_2 \cdot \ldots \cdot \sigma_n$ be a string over some alphabet Σ, $\sigma_i \in \Sigma$, $1 \leq i \leq n$. $p(x)$ is defined as the finite set of all permutations of x, i.e.,

$$p(x) = \{\sigma_{\pi(1)} \cdot \ldots \cdot \sigma_{\pi(n)} \mid \pi \text{ is a permutation of } \{1, 2, \ldots, n\}\}.$$

Let X be a set of strings, then

$$p(X) = \bigcup_{x \in X} p(x)$$

Definition 10

Let x and y be two strings over some alphabet Σ. Then merge(x, y) is defined as the following finite set of strings over Σ,

$$\text{merge}(x, y) = \{ z \in \Sigma^* \mid \text{there exists a natural number } n, \text{ and strings } x_i, y_i \in \Sigma^*,\ 1 \leq i \leq n, \text{ such that } z = x_1 y_1 x_2 y_2 \cdot \ldots \cdot x_n y_n,\ x = x_1 x_2 \cdot \ldots \cdot x_n,\ y = y_1 y_2 \cdot \ldots \cdot y_n \}$$

Notice that from the definition it follows immediately that for any two strings x and y

$$\text{merge}(x, y) = \text{merge}(y, x)$$

Definition 11

Let X be a finite set of strings over some alphabet Σ, $X = \{x_1, x_2, \ldots, x_n\}$. Then the set of non-recursive mergings of X, NRM(X), is defined as follows,

$$NRM_1(X) = \{ x \in \Sigma^* \mid \exists\, k \in N : x = x_1^k \};$$

$$NRM_i(X) = \{ x \in \Sigma^* \mid \exists\, k \in N,\ \exists\, y \in NRM_{i-1}(X) : x = \text{merge}(x_i^k, y) \},\quad 2 \leq i \leq n;$$

$$NRM(X) = NRM_n(X);$$

The set of recursive mergings of X, RM(X), is defined as follows,

$$RM_1(X) = X;$$

$$RM_i(X) = \{ x \in \Sigma^* \mid \exists\, x' \in X,\ \exists\, x'' \in RM_{i-1}(X) : x = \text{merge}(x', x'') \},\quad 2 \leq i;$$

$$RM(X) = \bigcup_{i \geq 1} RM_i(X);$$

Notice that in the definition of NRM(X) a specific ordering of the elements of X is used. It follows, however, from the remark following Definition 10, that the definition of NRM(X) does not depend on which ordering is chosen.

Let S' be a Part ETOL system, $S' = \langle \Sigma, P, L, \emptyset, \omega, \Delta \rangle$, and let X be a finite set of strings over L, the elements of which are called <u>matrices.</u> X may be interpreted in various ways as the basis of control on the derivations in S', as shown in Table 1. Note that the concepts of Table 1 were introduced for grammars in Cremers and Mayer (1973).

Interpretation	Notation	Control set
Matrix	M-Part ETOL	$M(X) = X^*$
Unordered Matrix	UM-Part ETOL	$UM(X) = (p(X))^*$
Vector	V-Part ETOL	$V(X) = NRM(X)$
Unordered Vector	UV-Part ETOL	$UV(X) = NRM(p(X))$
Generalized Vector	V^∞-Part ETOL	$V^\infty(X) = RM(X)$
Generalized Unordered Vector	UV^∞-Part ETOL	$UV^\infty(X) = RM(p(X))$

<u>Table 1.</u>

The language generated by the I-Part ETOL system, $S = (S', X)$, where I is one of the notations introduced in Table 1, is defined by

$$L_I(S) = \{x \in \Delta^* \mid z \in I(X) : \omega \overset{*}{\underset{S'}{\Rightarrow}} x(z)\}.$$

As in the previous sections, the notations introduced in Table 1 will also denote the corresponding classes of languages.

In Cremers and Mayer (1973) and (1974) the relations of Diagram 2 were proven for context-free grammars, CF, with control mechanisms corresponding to the ones introduced above (see Cremers and Mayer (1973) for the exact definitions). Some relations between Part ETOL systems with the same control mechanisms will be proven in the next theorems.

$$
\begin{array}{c}
V - CF = V^{\infty} - CF \\
M - CF = UM - CF \\
\rotatebox{90}{$\subsetneq$} \\
UV - CF = UV^{\infty} - CF \\
\rotatebox{90}{$\subsetneq$} \\
CF
\end{array}
$$

Diagram 2

Theorem 12 Let I be any of the control notations introduced in Table 1, then Part ETOL $\subseteq$ I - Part ETOL.

Proof

Let S be a Part ETOL system with L as its set of labels. Define X = L as a set of matrices of S, i.e., X contains only matrices of length 1, one for each element of L. Then obviously $L(S) = L_I(S, X)$.

Theorem 13 Part ETOL = M-Part ETOL = UM-Part ETOL

= V-Part ETOL = UV-Part ETOL.

Proof

Follows from Theorems 3 and 12 and the observation, that the control sets corresponding to the four control interpretations are all regular.

<u>Theorem 14</u> Part ETOL $\subsetneq$ UV^{∞} -- Part ETOL.

<u>Proof</u>

The inclusion follows from Theorem 12, and it is seen to be proper from the following example.

Let $S = \langle \Sigma, P, L, \emptyset, \omega, \Delta \rangle$ be a Part ETOL system with

$\Sigma = \{A, B, C, a, b\}$;

P as the following set of tables, labelled with the set

$L = \{1, 2, 3\}$,

$1 : \{(A,\ AB),\ (B,\ B)\}$

$2 : \{(A,\ Cb),\ (B,\ Cb),\ (C,\ Cb)\}$

$3 : \{(C,\ ab)\}$;

$\omega = A$;

$\Delta = \{a, b\}$;

Define X as the following set of matrices associated with S, $X = \{12,\ 3\}$, it is then easy to see that the language generated by (S, X) with UV^{∞}-control is equal to

$$L = \{(ab^n)^n \mid n \geq 2\}.$$

(Note that L is also the language generated by (S, X) with V^{∞}-control, since the only control words from both $UV^{\infty}(X)$ and $V^{\infty}(X)$ associated with derivations of terminal words in S are the words in the set $\{1^n 2^n 3 \mid n \geq 1\}$.)

But now obviously the letter a is periodically occurring, distributed and tied in L, and hence it follows from Lemma 4 that L does not belong to Part ETOL (this was also proven in Ehrenfeucht and Rozenberg (1973)), and this completes the proof of Theorem 14.

<u>Theorem 15</u> UV^{∞}-Part ETOL $\subseteq V^{\infty}$ -Part ETOL.

<u>Proof</u>

For any Part ETOL system S, and any set X of associated matrices, $L_{UV^{\infty}}(S, X) = L_{V^{\infty}}(S, p(X))$.

It is not known whether the inclusion of Theorem 15 is proper or not.

The results of this section are summarized in Diagram 3.

V^{∞}-Part ETOL

$\rotatebox{90}{$\subseteq$}$

UV^{∞}-Part ETOL

$\rotatebox{90}{$\subsetneq$}$

V-Part ETOL = UV-Part ETOL =

M-Part ETOL = UM-Part ETOL =

Part ETOL

<u>Diagram 3</u>

Notice the interesting differences between the diagrams of Diagrams 2 and 3, especially with respect to the effects of the concepts "un-ordering" and "generalization" on vector control.

References

A.B. Cremers and O. Mayer (1973), On matrix languages, Information and Control 23, 86-96.

A.B. Cremers and O. Mayer (1974), On vector languages, Journal of Computer and System Sciences 8, 158-166.

A. Ehrenfeucht and G. Rozenberg (1973), A characterization theorem for a subclass of ETOL languages, submitted for publication.

M. Nielsen, G. Rozenberg, A. Salomaa and S. Skyum (1974), Nonterminals, homomorphisms and codings in different variations of OL systems, II. Nondeterministic systems, Acta Informatica, 3, 357 - 364.

G. Rozenberg (1971), On OL systems with restricted use of productions, to appear in Journal of Computer and System Sciences.

A. Salomaa (1973), Formal Languages, Academic Press.

A. Salomaa (1968), On grammars with restricted use of productions, Ann. Acad. Sci. Fennicae, Ser. AI 454.