DECOMPOSITION THEOREMS FOR VARIOUS KINDS OF LANGUAGES PARALLEL IN NATURE

by

Sven Skyum

DAIMI PB-38
October 1974

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

Abstract

In this paper we give a method for decomposing subclasses of different families of languages, parallel in nature, into other families. These decomposition theorems can be used to produce languages not in a family by using examples of languages not belonging to some "smaller" family.

Keywords

Formal Languages, Parallelism, L systems, Definable sets, Level grammars, Decompositions.

INTRODUCTION

Within the last few years there has been a growing interest in various forms of parallelism in rewriting systems. This is mainly due to the large amount of work done in the area of L systems or developmental languages.

In this paper we will examine the ability of different systems to generate languages in which the words of the language are composed of words from languages belonging to other families.
These decomposition theorems can be used for examining the relation between various families of languages.

On this basis we can give examples of languages not belonging to a certain family by giving examples of languages not belonging to some smaller family.

Ehrenfeucht, Rozenberg and Skyum [3] employ this technique to show that the family of ET0L languages is properly included in the family of INDEX languages.

It is assumed that the reader is familiar with the basic notions concerning formal language theory. For unexplained notions we refer to Salomaa [12].

The following notations are used in this paper:

I denotes the set of nonnegative integers.

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

$|x|$ denotes the length of x.

$|x|_r$ denotes the number of occurrences in $x \in \Sigma^*$ of symbols belonging to some subalphabet $\Sigma_r \subseteq \Sigma$.

$\min(x)$ denotes the set of symbols occurring in x.

1. L SYSTEMS

For a general introduction to L systems we refer to [7, 11].

Definition 1

An E0L system is a 4-tuple $G = (V, P, w, \Sigma)$ where V (the alphabet) is a finite set of symbols, P (the productions) is a finite subset of $V \times V^*$, such that for each $A \in V$ there exists a $x \in V^*$ such that (A, x) is in P, w (the axiom) is a word in V^+, and Σ (the target alphabet) is a subset of V.

Definition 2

The E0L language $L(G)$ of an E0L system $G = (V, P, w, \Sigma)$ is

$$L(G) = \{x \in \Sigma^* \mid w \overset{*}{\underset{G}{\Rightarrow}} x\}$$

where $\overset{*}{\underset{G}{\Rightarrow}}$ is the transitive and reflexive closure of $\underset{G}{\Rightarrow}$ defined by $z \underset{G}{\Rightarrow} y$ iff $z = y = \lambda$ or there exist $a_1, a_2, \ldots, a_k \in V$ and $v_1, v_2, \ldots, v_k \in V^*$ such that $z = a_1 a_2 \ldots a_k$, $y = v_1 v_2 \ldots v_k$, and $(a_i, v_i) \in P$ for each $1 \leq i \leq k$.

Definition 3

An E0L system $G = (V, P, w, \Sigma)$ is deterministic (abbreviated ED0L) if for each $A \in V$ there exists exactly one $x \in V^*$ such that $(A, x) \in P$.

Definition 4

An ET0L system is a 4-tuple $G = (V, \mathscr{P}, w, \Sigma)$ where V, w, and Σ are as in the definition of an E0L system and $\mathscr{P}$ is a finite set (whose elements are called tables) such that for every $P \in \mathscr{P}$, (V, P, w, Σ) is an E0L system.

Definition 5

The ETOL language $L(G)$ of an ETOL system $G = (V, \mathcal{P}, w, \Sigma)$ is

$$L(G) = \{x \in \Sigma^* \mid w \overset{*}{\underset{G}{\Rightarrow}} x\}$$

where $\underset{G}{\Rightarrow}$ is defined by $z \underset{G}{\Rightarrow} y$ iff $z = y = \lambda$ or there exist $P \in \mathcal{P}$, $a_1, a_2, \ldots, a_k \in V$, and $v_1, v_2, \ldots, v_k \in V^*$ such that $z = a_1 a_2 \ldots a_k$, $y = v_1 v_2 \ldots v_k$, and $(a_i, v_i) \in P$ for each $1 \leq i \leq k$. (We will then write $z \underset{P}{\Rightarrow} y$.)

Definition 6

An ETOL system is deterministic iff each of the underlying EOL systems is deterministic.

Definition 7

An EFOL (ETFOL) system is defined as above, but here we allow a finite set Ω of axioms instead of a single axiom w. The language generated by such a system consists of the union of the languages generated by the system obtained by choosing in turn each element $w \in \Omega$ to be the axiom.

Definition 8

A OL (FOL, TOL, TFOL) system is an EOL (EFOL, ETOL, ETFOL) system $G = (V, P, w, \Sigma)$, $(G = (V, \mathcal{P}, w, \Sigma))$ where $\Sigma = V$.

For any class of systems, we will use the same notation for the family of languages generated by these systems.

Definition 9

The prefix H attached to the name of a language family indicates that we are considering homomorphic images of the languages in the family.

E.g. $L \subseteq \Sigma^*$ belongs to HDF0L iff there exists a DF0L system $G = (V, P, \Omega, V)$ and a homomorphism $h : V^* \to \Sigma^*$ such that $L = h(L(G))$.

Some of the relations between different L families are shown in Figure 1. If two nodes labelled X and Y are connected by an oriented edge then $X \subsetneq Y$ and if two nodes labelled X and Y are connected by a broken edge then X and Y are mutually incomparable.

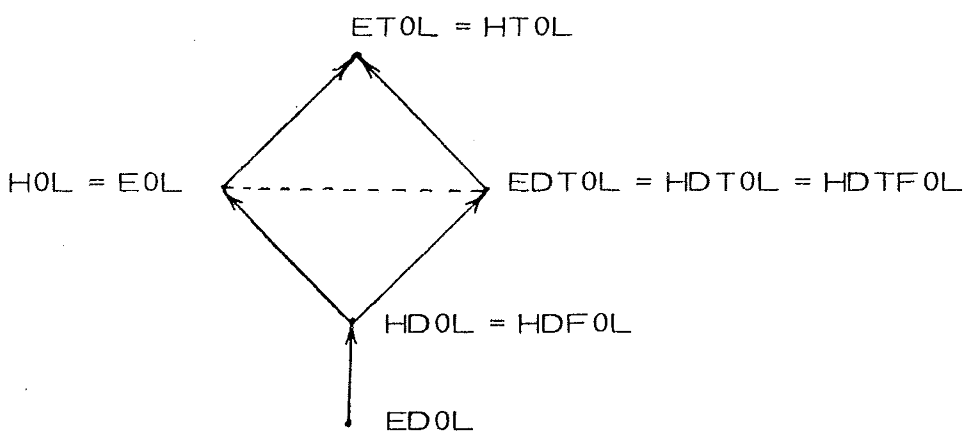

Figure 1.

The proofs of the relations can be found in [1, 9].

2. DEFINABLE AND EXTENDED DEFINABLE SETS SIMPLE AND EXTENDED RECURRENCE LANGUAGES

The following five definitions define notions introduced by Rose (1964).

Definition 10

A (n-ary) format is any triple $(\Sigma ; \xi ; F)$ where Σ (the alphabet) is a finite set of symbols, ξ is a n-tuple $(\xi_1, \ldots, \xi_n)$ of symbols (called variables) not in Σ, and F is a n-tuple $(F_1, \ldots, F_n)$ of finite subsets of $(\Sigma \cup \{\xi_1, \ldots, \xi_n\})^*$.

Definition 11

The generating function $g_{\Sigma ; \xi ; F}$ for a given (n-ary) format $(\Sigma ; \xi ; F)$ is defined thus:

For each n-tuple $W = (W_1, \ldots, W_n)$ of finite subsets of $(\Sigma \cup \{\xi_1, \ldots, \xi_n\})^*$,

$$g_{\Sigma ; \xi ; F}(W) = \Big(\bigcup_{\sigma \in R_{\Sigma ; \xi}(W)} \sigma(F_1), \ldots, \bigcup_{\sigma \in R_{\Sigma ; \xi}(W)} \sigma(F_n)\Big).$$

where $R_{\Sigma ; \xi}(W)$ is the set of all substitutions σ such that, for each $x \in \Sigma$, $\sigma(x) = \{x\}$ and $\sigma(\xi_i)$ is a subset of W_i with at most one element $(1 \le i \le n)$.

Definition 12

The approximating sequence $E(k) = (E_1(k), \ldots, E_n(k))$ $(k \in I)$ for a given (n-ary) format $(\Sigma ; \xi ; F)$ is defined thus:

$E_i(0) = \emptyset$ $(1 \le i \le n)$, and for all $k > 0$ $E(k) = g_{\Sigma ; \xi ; F}(E(k-1))$. The n-tuple $E = (\bigcup_{k \ge 0} E_1(k), \ldots, \bigcup_{k \ge 0} E_n(k))$ is said to be generated by $(\Sigma ; \xi ; F)$.

Definition 13

A language $L \subseteq \Sigma^*$ is said to be extended definable if it is the n'th coordinate of the n-tuple generated by some (n-ary) format.

We will denote the family of extended definable sets as ED.

Definition 14

The polynomial function $p_{\Sigma;\xi;F}$ for a given (n-ary) format $(\Sigma;\xi;F)$ is defined thus:

For each n-tuple $W = (W_1, \ldots, W_n)$ of finite subsets of $(\Sigma \cup \{\xi_1, \ldots, \xi_n\})^*$

$$p_{\Sigma;\xi;F}(W) = (S_\xi^W(F_1), \ldots, S_\xi^W(F_n))$$

where S_ξ^W is the substitution σ such that, for each $x \in \Sigma$, $\sigma(x) = \{x\}$ and $\sigma(\xi_i) = W_i$.

The following lemma belongs to Rose (1964).

Lemma 1

A language $L \subseteq \Sigma^*$ is definable (defined by Ginsburg and Rice (1962)) if and only if it is the n'th coordinate for the minimal fixpoint (mfp) of the polynomial function $p_{\Sigma;\xi;F}$ for some (n-ary) format $(\Sigma;\xi;F)$.

The mfp for $p_{\Sigma;\xi;F}$ is $D = (D_1, \ldots, D_n) = (\bigcup_{k\geq 0} D_1(k), \ldots, \bigcup_{k\geq 0} D_n(k))$ where $D_i(0) = \emptyset\ (1 \leq i \leq n)$ and for all $k \geq 1$

$$D(k) = p_{\Sigma;\xi;F}(D(k-1)).$$

We will denote the family of definable sets by D.

If we use the notion from definitions 10 and 14 we can give the following definition of the simple recurrence languages introduced by Herman (1973).

Definition 15

A recurrence system is a 4-tuple $R = (\Sigma;\xi;F;\alpha)$, where $(\Sigma;\xi;F)$ is a (n-ary) format and $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a n-tuple of finite subsets of Σ^*.

We define the simple recurrence language L(R) of R by

$$L(R) = \bigcup_{k\geq 0} D'_n(k)$$

where $D'(k) = (D'_1(k), \ldots, D'_n(k))$ is defined inductively by $D'(0) = (\alpha_1, \ldots, \alpha_n)$, and for $k \geq 1$ $D'(k) = p_{\Sigma;\xi;F}(D'(k-1))$.

The family of simple recurrence languages is denoted by SR.

Definition 16

Let $R = (\Sigma;\xi;F;\alpha)$ be a recurrence system. The extended recurrence language $L_E(R)$ of R is defined by

$$L_E(R) = \bigcup_{k\geq 0} E'_n(k)$$

where $E'(k) = (E'_1(k), \ldots, E'_n(k))$ is defined inductively by $E'(0) = (\alpha_1, \ldots, \alpha_n)$, and for $k \geq 1$ $E'(k) = g_{\Sigma;\xi;F}(E'(k-1))$.

The family of extended recurrence languages is denoted by ER.

Proposition 1

For every recurrence system $R = (\Sigma; \xi; F; \alpha)$ there exists a recurrence system $R' = (\Sigma; \xi'; F'; \alpha')$ such that $F' = (F'_1, F'_2, \ldots, F'_n)$ is a n-tuple of finite subsets of $\{\xi'_1, \xi'_2, \ldots, \xi'_n\}^*$ and for $1 \leq i \leq n$ α'_i is either empty or consists of a single element in Σ, $L(R) = L(R')$, and $L_E(R) = L_E(R')$. (The proof can be found in [6, 15].)

Definition 17

A level grammar is a 4-tuple $G = (V, P, S, \Sigma)$ where

V is the alphabet,

P (the productions) is a finite subset of $V \times V^*$,

$S \in V$ is the start symbol, and

$\Sigma \subseteq V$ is the terminal alphabet.

Definition 18

We say that $w(A,n)w'$ directly yields $w(A_1, n+1) \ldots (A_k, n+1)w'$ in G ($w(A,n)w' \underset{G}{\Rightarrow} w(A_1, n+1) \ldots (A_k, n+1)w'$) if $w, w' \in (V, I)^*$ and $(A, A_1 \ldots A_k) \in P$. $\overset{*}{\underset{G}{\Rightarrow}}$ is the transitive and reflexive closure of $\underset{G}{\Rightarrow}$. As usual we will write $\Rightarrow$ and $\overset{*}{\Rightarrow}$ if it is clear which grammar G is involved in.

Definition 19

The level language $L(G)$ is generated by a level grammar $G = (V, P, S, \Sigma)$ if

$$L(G) = h(\{w \in (V, I)^* \mid (S, 0) \overset{*}{\Rightarrow} w\}) \cap \Sigma^*$$

where $h:(V, I)^* \rightarrow V^*$ is a partial function only defined on strings, where all variables are associated with the same level number $n \in I$.

More specifically h is defined as follows:

(1) $h(\lambda) = \lambda$.

(2) For all $A_1, \ldots, A_k \in V$ and $n \in I$, $h((A_1, n)\ldots(A_k, n)) = A_1 \ldots A_k$.

(3) For all other strings in $(V, I)^+$, h is undefined.

We have that

$$L(G) = \bigcup_{n \geq 0} [h(\{w \in (V, n)^* \mid (S, 0) \overset{*}{\Rightarrow} w\}) \cap \Sigma^*] = \bigcup_{n \geq 0} L(G, n).$$

We say that $L(G, n)$ is the language of level n generated by G.

Example 1

Let $G = (\{S, a, b\}, \{(S, ab), (a, aa), (b, b), (b, bb)\}, S, \{a, b\})$.

Then

$$L(G, 0) = \emptyset$$

$$L(G, 1) = \{ab\}$$

$$L(G, 2) = \{aab, aabb\}$$

$$\vdots$$

$$L(G, n) = \{a^{2^{n-1}} b^i \mid 1 \leq i \leq 2^{n-1}\}$$

$$\vdots$$

$$L(G) = \bigcup_{n \geq 0} L(G, n) = \{a^{2^n} b^i \mid n \geq 0, 1 \leq i \leq 2^n\}$$

The family of level languages will be denoted by LL.

Definition 20

Let $G = (V, P, S, \Sigma)$ be a level grammar. We write

$w_1(A,n)w_2(A,n)\ldots w_{k-1}(A,n)w_k \Rightarrow_P w_1 ww_2 w \ldots w_{k-1} ww_k$ if $w_i \in ((V,I) \setminus (A,n))^*$ $(1 \leq i \leq k)$ and $(A,n) \Rightarrow w$. $\overset{*}{\Rightarrow}_P$ is again the transitive and reflexive closure of $\Rightarrow_P$.

We say that w derives w' in parallel if $w \overset{*}{\Rightarrow}_P w'$.

Definition 21

The parallel level language $L_P(G)$ generated by a level grammar $G = (V,P,S,\Sigma)$ is

$$L_P(G) = h(\{ w \in (V,I)^* \mid (S,0) \overset{*}{\Rightarrow}_P w\}) \cap \Sigma^*.$$

Again $L_P(G) = \bigcup_{n \geq 0} L_P(G,n)$ where

$$L_P(G,n) = h(\{ w \in (V,n)^* \mid (S,0) \overset{*}{\Rightarrow}_P w\}) \cap \Sigma^*.$$

Example 2

Let G be the level grammar from Example 1. Then

$$L_P(G,0) = \emptyset$$

$$L_P(G,1) = \{ab\}$$

$$L_P(G,2) = \{aab,\ aabb\}$$

$$L_P(G,3) = \{aaaab,\ aaaabb,\ aaaabbbb\}$$

$$\vdots$$

$$L_P(G,n) = \{a^{2^{n-1}} b^{2^{i-1}} \mid 1 \leq i \leq n\}$$

$$\vdots$$

$$L_P(G) = \bigcup_{n \geq 0} L_P(G,n) = \{a^{2^n} b^{2^i} \mid 0 \leq i \leq n\}$$

The family of parallel level languages is denoted by PLL.

Definition 22

A restricted level grammar $G = (V, P, S, \Sigma)$ is a level grammar with the restriction that for each $A \in \Sigma$, (A, A) is a production in P.

The corresponding families of restricted (parallel) level languages are denoted by RLL (RPLL).

Figure 2 shows the relations between the families defined in this section.

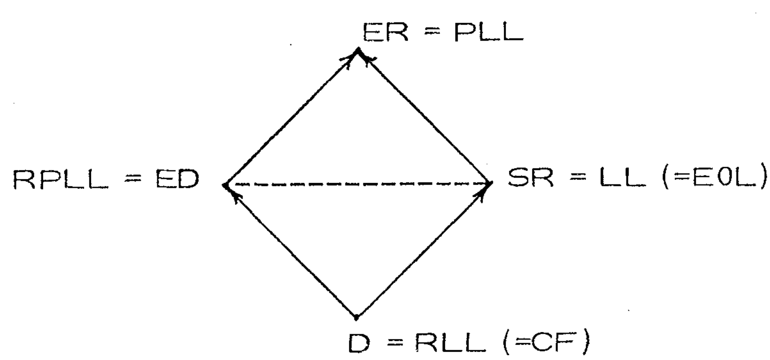

Figure 2.

The proofs of the relations can be found in [6, 10, 15].

3. RUSSIAN AND INDIAN PARALLELISM

For definitions and discussion see [8, 13, 14, 16].

Definition 23

A Russian parallel context-free grammar is a 5-tuple $G = (V, \Sigma, P_u, P_o, S)$ where the only difference to an ordinary context-free grammar is that the set of productions is divided into two sets of productions, P_u (the universal productions) and P_o (the ordinary productions).

Definition 24

The language generated by a Russian parallel context-free grammar is

$$L(G) = \{x \in \Sigma^* \mid S \overset{*}{\Rightarrow}_G x\}$$

where $\overset{*}{\Rightarrow}_G$ is the transitive and reflexive closure of $\Rightarrow_G$ defined by $z \Rightarrow_G y$ iff either

1) $z = z_1 A z_2$ and $y = z_1 v z_2$ for some $v, z_1, z_2 \in (V \cup \Sigma)^*$, and $A \in V$ such that (A, v) is in P_o or

2) $z = z_1 A z_2 A \ldots A z_k$ and $y = z_1 v z_2 v \ldots v z_k$ for some $v \in (V \cup \Sigma)^*$, $A \in V$, and $z_i \in ((V \cup \Sigma) \setminus \{A\})^*$, $1 \leq i \leq k$, such that $(A, v) \in P_u$.

We will denote the family of Russian parallel languages by RP.

Definition 25

An Indian parallel context-free grammar is a 4-tuple $G = (V, \Sigma, P, S)$ like a context-free grammar but P consists of universal productions only. That means that $(V, \Sigma, P, \emptyset, S)$ is a Russian parallel context-free grammar.

We will denote the family of Indian parallel languages by IP.

The relations between RP, IP, CF, and LIN (linear languages) are shown in figure 3.

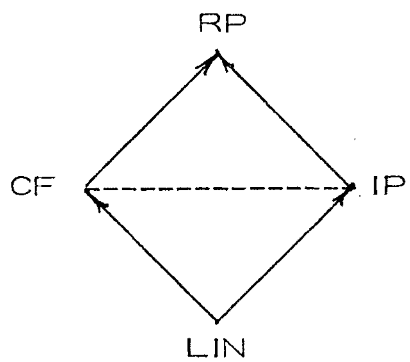

Figure 3.

The proofs of the nontrivial relations can be found in [12, 16].

In figure 4 we summarize the known inclusions between the families defined in sections 1 to 3.

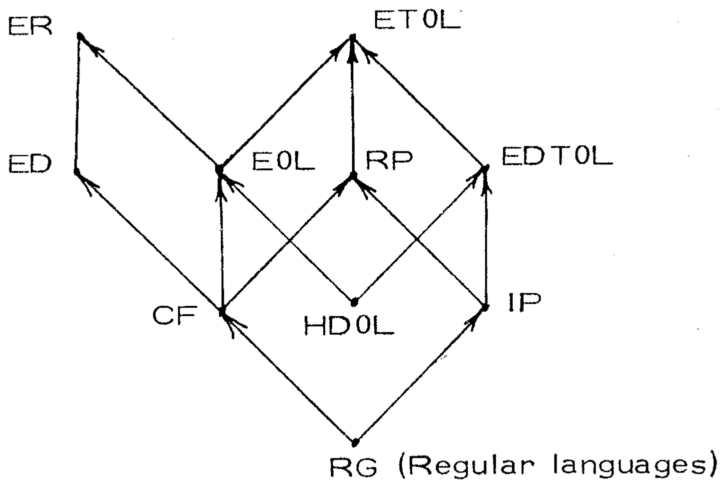

Figure 4.

4. RESULTS

In this section we will examine the possibility of decomposing some languages from a certain family into languages from some smaller family.

Theorem 1 Let Σ be some alphabet and $K \subseteq \Sigma^*$. Let $\# \notin \Sigma$ and $c_\#^n(K) = \{(w\#)^{n-1}w \mid w \in K\}$. The families HD0L, EDT0L, ED, ER, and IP are closed under the operators $c_\#^n$.

Proof

The proof is straightforward and is omitted.

Definition 26

Let $L = h(L(G))$, where $G = (V, \mathcal{P}, w, V)$ is a T0L system and $h : V^* \to \Sigma^*$ is a homomorphism.

$a \in V$ is called essentially nondeterministic with respect to h and G iff

1) There exist words $x_1, x_2, x_3 \in V^*$, such that $x_1 a x_2 a x_3 \in L(G)$.

2) There exists a sequence of tables $P_{i_1}, P_{i_2}, \ldots, P_{i_n}$ and words $w_1, w_2 \in V^*$ such that for $j = 1, 2$ $a \underset{P_{i_1}}{\Rightarrow} w_{j1} \underset{P_{i_2}}{\Rightarrow} w_{j2} \underset{P_{i_3}}{\Rightarrow} \ldots \underset{P_{i_n}}{\Rightarrow} w_{jn} = w_j$ and $h(w_1) \neq h(w_2)$.

Lemma 2

Let $L = h(L(G))$, where $G = (V, \mathcal{P}, w, V)$ is a T0L system and $h : V^* \to \Sigma^*$ a homomorphism. If there are no essentially nondeterministic symbols in V with respect to h and G then $L \in$ HDT0L.

Proof

Let $\overline{G} = (V, \overline{\mathcal{P}}, w, V)$ be the DT0L system where $\overline{\mathcal{P}}$ is defined by $P \in \overline{\mathcal{P}}$ iff $P \subseteq P'$ for some $P' \in \mathcal{P}$ and (V, P, w, V) is a D0L system.

Now it is obvious that $L(\overline{G}) \subseteq L(G)$.

Let $x = h(y)$ where $y \in L(G)$.

Let $w = w_0 \underset{P_{i_1}}{\Rightarrow} w_1 \underset{P_{i_2}}{\Rightarrow} \ldots \underset{P_{i_n}}{\Rightarrow} w_n = y$ be a derivation of y in G.

For $1 \leq j \leq n$ let P'_{i_j} be a table in $\overline{\mathcal{P}}$ such that $P'_{i_j} \subseteq P_{i_j}$ and if $a \in V$ only occurs once in w_{j-1} and is rewritten as z in the derivation then $(a, z) \in P'_{i_j}$. If a occurs several times in w_{j-1}, then just choose one production from P_{i_j} to be in P'_{i_j}. Note that because a is not essentially nondeterministic it does not matter which production one chooses.

Then $w = w_0 \underset{P'_{i_1}}{\Rightarrow} w'_1 \underset{P'_{i_2}}{\Rightarrow} \ldots \underset{P'_{i_n}}{\Rightarrow} w'_n = y'$ is a derivation in $\overline{G}$ and $h(y) = h(y')$.

Theorem 2

Let Σ be an alphabet and let $K \subseteq \Sigma^*$. Let $\# \notin \Sigma$.

(I) If $c^2_\#(K) \in ETOL$ then $K \in EDTOL$.

(II) If $c^3_\#(K) \in EOL$ then $K \in HDOL$.

(III) If $c^2_\#(K) \in RP$ then $K \in IP$.

(IV) If $c^2_\#(K) \in ED$ then $K \in ED$.

Proof

(I) Let $c^2_\#(K) = h(L(G)) \in HTOL = ETOL$ for some TOL system G and homomorphism h. Because of the form of $c^2_\#(K)$ it immediately follows that there is no essentially nondeterministic symbol in G. Therefore by Lemma 2 it follows that $c^2_\#(K) \in HDTOL = EDTOL$.

Let then $c^2_\#(K) = h(L(G))$, where $G = (V, \mathcal{P}, w, V)$ is a DTOL system and $h : V^* \to (\Sigma \cup \{\#\})^*$ is a homomorphism.

Define $V_\# \subseteq V$ to be the set satisfying $a \in V_\#$ iff there exists a x such that $a \overset{*}{\Rightarrow} x$ and $\# \in \min(h(x))$. Note that every symbol in $V_\#$ can occur at most once in every word in $L(G)$.

Define a DTOL system $H = (\overline{V}, \overline{\mathcal{P}}, \overline{w}, \overline{V})$ as follows.

$\overline{V} = V \times 2^{V_\#}$.

If $(a_0, a_1 a_2 \ldots a_{k_1} b_1 a_{k_1+1} \ldots a_{k_2} b_2 \ldots b_n a_{k_n+1} \ldots a_{k_{n+1}}) \in P$ $(n \geq 0)$, where $a_i \in V \setminus V_\#$, $1 \leq i \leq k_{n+1}$, $b_i \in V_\#$, $1 \leq i \leq n$, and $P \in \mathcal{P}$ then for all $N \subseteq V_\#$ $([a_0, N], [a_1, M][a_2, M] \ldots [a_{k_1}, M][b_1, M \cup \{b_1\}]$
$[a_{k_1+1}, M \cup \{b_1\}] \ldots [a_{k_2}, M \cup \{b_1\}][b_2, M \cup \{b_1, b_2\}] \ldots [b_n, M \cup$
$\{b_1, b_2, \ldots, b_n\}][a_{k_n+1}, M \cup \{b_1, b_2, \ldots, b_n\}] \ldots [a_{k_{n+1}}, M \cup \{b_1, b_2, \ldots, b_n\}])$
is in a corresponding table $\overline{P}$ in $\overline{\mathcal{P}}$. $M = \min(y) \cap V_\#$ for some y such that there exists a x, where $N \setminus \{a_0\} = \min(x)$ and $x \underset{P}{\Rightarrow} y$. Note that M is uniquely determined by N.

If $w = a_1 \ldots a_{k_1} b_1 a_{k_1+1} \ldots a_{k_2} b_2 a_{k_2+1} \ldots a_{k_n} b_n a_{k_n+1} \ldots a_{k_{n+1}}$ where $a_i \in V \setminus V_\#$, $1 \leq i \leq k_{n+1}$, and $b_i \in V_\#$, $1 \leq i \leq n$ then
$\overline{w} = [a_1, \emptyset] \ldots [a_{k_1}, \emptyset][b_1, \{b_1\}][a_{k_1+1}, \{b_1\}] \ldots [a_{k_2}, \{b_1\}][b_2, \{b_1, b_2\}]$
$[a_{k_2+1}, \{b_1, b_2\}] \ldots [a_{k_n}, \{b_1, \ldots, b_{n-1}\}][b_n, \{b_1, \ldots, b_n\}]$
$[a_{k_n+1}, \{b_1, \ldots, b_n\}] \ldots [a_{k_{n+1}}, \{b_1, \ldots, b_n\}]$.

If we now define a homomorphism $g : \overline{V}^* \to \Sigma^*$ by

$$g([a, N]) = \begin{cases} h(a) & \text{if } \# \notin \min(h(b)) \text{ for all } b \in N \\ v & \text{if } a \in N \text{ and } h(a) = v \# u \text{ for some } u \in \Sigma^* \\ \lambda & \text{otherwise} \end{cases}$$

then it follows that $K = g(L(H)) \in HDT0L = EDT0L$.

(II) Let $c_\#^3(K) = h(L(G))$ where $G = (V, P, w, V)$ is an 0L system and $h : V^* \to (\Sigma \cup \{\#\})^*$ is a lengthpreserving homomorphism. This is no restriction (see e.g. Ehrenfeucht and Rozenberg, to appear).
Let $n = |V|$, m an integer such that $|w| \leq m$ and if $(a, x) \in P$ then $|x| \leq m$.
Let $V_m \subseteq V$ be the set of mortal symbols in V, that means that $a \in V_m$ iff $a \Rightarrow^n x$ implies that $x = \lambda$.

Let $V_v = V \setminus V_m$ be the set of vital symbols.

Let $V_s \subseteq V$ be defined by $a \in V_s$ iff there exists a $l > 0$ such that if $x \in L(G)$, $|x|_v > l$ then $a \notin \min(x)$.

Let $V_b = V \setminus V_s$.

Observations

1) If $a \in V_b$ and $a \Rightarrow^* x$ then $\min(x) \subseteq V_b$.

2) There exists a $k > 0$ such that if $x \in L(G)$ and $|x|_v > k$ then $\min(x) \subseteq V_b$.

3) If $x \in L(G)$ and $|x| > m^n \cdot k$ then $\min(x) \subseteq V_b$.

To see the correctness of (3), let $w \Rightarrow^* w_1 \Rightarrow w_2 \Rightarrow \ldots \Rightarrow w_n = x$ be a derivation of $x \in L(G)$, where $|x| > m^n \cdot k$.

Assume that $a \in \min(x) \cap V_s$. Then because of (1) there exists a symbol $b \in \min(w_1) \cap V_s$ which implies that $|w_1|_v \leq k$ and $|x| \leq k \cdot m^n$ so our assumption cannot be true.

Let now $x \in L(G)$, $|x| > m^n \cdot k$, and $a \in \min(x)$. We will prove that if $a \Rightarrow^i w_1$ and $a \Rightarrow^i w_2$ for some $i > 0$ and $w_1, w_2 \in V^*$ then $h(w_1) = h(w_2)$. From that it will follow that $c^3_{\#}(K) \in HDF0L = HD0L$ because we can just choose one production for every symbol and choose $\{x \in L(G) \mid |x| \leq m^n \cdot k\}$ to be the set of axioms.

Now back to the statement.

Let $t = \max\{|w_1|, |w_2|\}$. Because of (3) above $a \in V_b$ and therefore there exists a word $y \in L(G)$ such that $a \in \min(y)$ and $|y|_v > 3t+2$.

Let $y = z_1 a z_2$ and $z_j \Rightarrow^i v_j$ for $j = 1, 2$. Then $|v_1 w_j v_2| > 3t+2$ for $j = 1, 2$ and since $|w_j| \leq t$ for $j = 1, 2$ $h(w_1)$ and $h(w_2)$ must be equal.

To prove that $K \in HD0L$ if $c^3_{\#}(K) \in HD0L$ we can use exactly the same technique as in the proof of $K \in HDT0L$ if $c^2_{\#}(K) \in HDT0L$.

(III) Let $c_{\#}^{2}(K) = L(G)$ where $G = (V, \Sigma \cup \{\#\}, P_u, P_o, S)$ is a Russian parallel context-free grammar. As for ordinary context-free grammars we can assume that all nonterminals are useful. That means that for all $A \in V$ there exist words $x, y \in (V \cup \Sigma \cup \{\#\})^*$ and $v \in (\Sigma \cup \{\#\})^*$ such that $S \Rightarrow^* xAy \Rightarrow^* v$.

Assume that $S \Rightarrow^* x_1 A x_2 A x_3 \Rightarrow^* v_1 A v_2 A \ldots A v_k \Rightarrow^* v_1 w_1 v_2 w_2 \ldots w_{k-1} v_k$ for some words $x_i \in (V \cup \Sigma \cup \{\#\})^*$, $1 \leq i \leq 3$, $v_j \in (\Sigma \cup \{\#\})^*$, $1 \leq j \leq k$, $w_i \in (\Sigma \cup \{\#\})^*$, $1 \leq i \leq k-1$, and some $A \in V$, such that A is not rewritten anywhere in the subderivation $x_1 A x_2 A x_3 \Rightarrow^* v_1 A v_2 A \ldots A v_k$.

Then $w_1 = w_2 = \ldots = w_{k-1}$ because $v_1 w_{i_1} v_2 w_{i_2} \ldots w_{i_{k-1}} v_k \in L(G)$ for all $1 \leq i_1, i_2, \ldots, i_{k-1} \leq k-1$.

It then follows that $c_{\#}(K) = L(G')$ where $G' = (V, \Sigma \cup \{\#\}, P_u \cup P_o, \emptyset, S)$, which means that $c_{\#}(K) \in IP$.

Define $V_{\#} \subseteq V \cup \{\#\}$ to be the set satisfying $A \in V_{\#}$ iff $A \Rightarrow^* x_1 \# x_2$ for some $x_1, x_2 \in \Sigma^*$.

Let $\overline{G} = (V, \Sigma, P', \emptyset, S)$ where P' is defined as follows.

If $(A, A_1 A_2 \ldots A_k B A_{k+1} A_{k+2} \ldots A_n) \in P_u \cup P_o$ where $A_i \in V \cup \Sigma$, $1 \leq i \leq n$, and $B \in V_{\#}$ then $(A, A_1 A_2 \ldots A_k B) \in P'$ if $B \neq \#$ and otherwise $(A, A_1 A_2 \ldots A_k) \in P'$.

(Note that in every sentential-form in G' there is exactly one occurrence of a letter in $V_{\#}$.)

If $(A, x) \in P_u \cup P_o$ and $V_{\#} \cap \min(x) = \emptyset$ then $(A, x) \in P'$.

No other productions are in P'.

It is easy to check that $K = L(\overline{G})$.

(IV) The proof of part (IV) is very similar to the last half of the proof of part (III) if we use the fact that $ED = RPLL$.

<u>Remark 1</u> Note that is is open whether $c_{\#}^{2}(K) \in EOL$ implies $K \in HDOL$ or not.

<u>Remark 2</u> If $c_{\#}^{2}(K) \in CF$ then $K \in RG$. This is in fact a special case of Theorem 2.3.2 in [5].

<u>Conjecture</u> If $c_{\#}^{2}(K) \in ER$ then $K \in ER$.

Theorem 2 is visualized in Figure 5. If two nodes labelled X and Y are connected by an oriented edge, then $c_{\#}^{3}(L) \in X$ implies that $L \in Y$.

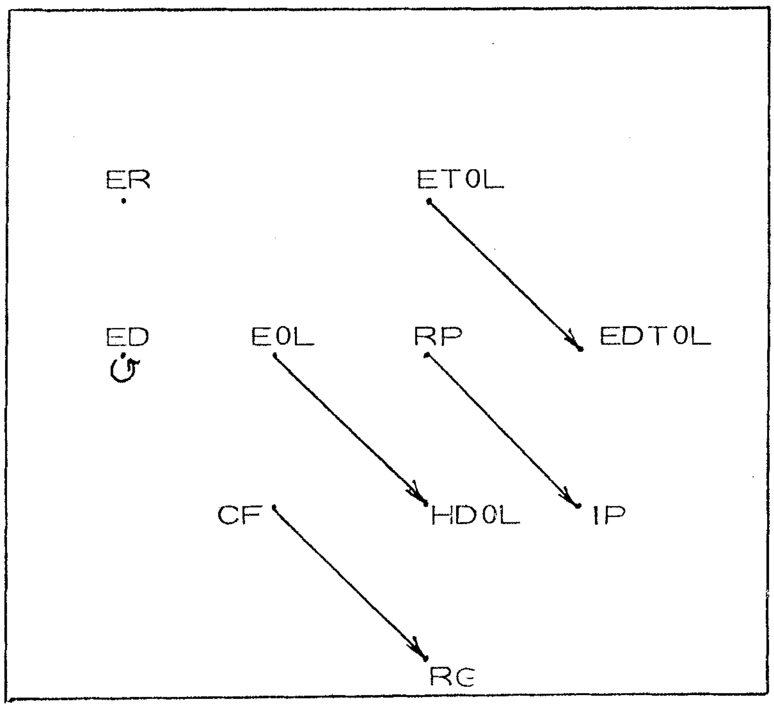

Figure 5.

We can give a more general theorem than Theorem 2, namely

<u>Theorem 3</u>

Let Σ be an alphabet and let $K_1, K_2 \subseteq \Sigma^*$. Let $\# \notin \Sigma$ and let $f : K_1 \to K_2$ be a bijective function from K_1 onto K_2. Let $K = \{w \# f(w) \mid w \in K_1\}$.

(I) If $K \in ETOL$ then $K, K_1, K_2 \in EDTOL$.

(II) If $K \in RP$ then $K, K_1, K_2 \in IP$.

(III) If $K \in ED$ then $K, K_1, K_2 \in ED$.

Proof

The proof is analogous to the proof of Theorem 2.

Remark

Note that it is not true, that $K \in EOL$ implies that $K_2 \in HDOL$.
Let $\Sigma = \{a,b\}$, $K_1 = K_2 = \Sigma^*$, and $f : K_1 \to K_2$ be defined by $w \in K_1$: $f(w) = mir(w)$. ($mir(w)$ denotes the mirror-image of w).
$K = \{w \# f(w) | w \in K_1\} \in EOL$ because it is generated by the following EOL system.
$(\{S,a,b,\#\}, \{S \to aSa,\ S \to bSb,\ S \to \#,\ a \to a,\ b \to b,\ \# \to \#\}, S, \{a,b,\#\})$.
But Σ^* is not a HDOL language.

Instead of having a special marker #, which divides the words into two parts we could have disjoint alphabets such that the words are concatenations of words in the alphabets.

Theorem 4

Let $\Sigma_1, \Sigma_2, \ldots, \Sigma_n$ be n alphabets, not necessarily disjoint, and let $f_i : \Sigma_1^* \to \Sigma_i^*$, $2 \leq i \leq n$, be homomorphisms. Let $K \subseteq \Sigma_1^*$ and $c_n'(K) = \{wf_2(w)f_3(w)\ldots f_n(w) \mid w \in K\}$.
EDTOL and HDOL are closed under the operator c_n'.

Proof Easy to check.

If f_i is bijective for $2 \leq i \leq n$ then we will denote the operator by c_n. Note that ED, ER, and IP are not closed under c_n'.

Lemma 3

Let Σ_1 and Σ_2 be two disjoint alphabets and let $K_1 \subseteq \Sigma_1^*$, $K_2 \subseteq \Sigma_2^*$. Let f be a bijective function from K_1 onto K_2. Let

$K = \{wf(w) \mid w \in K_1\}$.

If $K \in ER$ then $K \in SR$ (= E0L).

Proof

Let $K = L_E(R)$ where $R = (\Sigma; \xi_1, \xi_2, \ldots, \xi_n; F_1, F_2, \ldots, F_n; \alpha_1, \alpha_2, \ldots, \alpha_n)$ is a recurrence system satisfying the properties of proposition 1. Assume that there exist integers $1 \leq i, i_1, \ldots, i_m \leq n$, $l, q > 2$, words $w_k \in \{\xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_n\}^*$ for $1 \leq k \leq l$, and words $w_1^{(k)}, w_2^{(k)}, \in \{\xi_1, \ldots, \xi_n\}^*$ for $1 \leq k \leq m$ such that

(I) $w_1 \xi_i w_2 \xi_i w_3 \ldots \xi_i w_l \in F_{i_1}$

(II) $i_m = n$

(III) $w_1^{(k)} \xi_{i_k} w_2^{(k)} \in F_{i_{k+1}}$ for $1 \leq k \leq m$

(IV) If $\xi_r \in \min(w_1 w_2 \ldots w_l)$ then $E_r(q) \neq \emptyset$

(V) If $\xi_r \in \min(w_1^{(k)} w_2^{(k)})$ then $E_r(q+k) \neq \emptyset$ for $1 \leq k < m$

(VI) There exist words $v_1, v_2 \in E_i(q)$

Then there exist words $x_k \in \Sigma^*$, $1 \leq k \leq l$, $w, \overline{w} \in \Sigma_1^*$ such that $x_1 v_1 x_2 v_1 \ldots v_1 x_l = wf(w)$ and $x_1 v_2 x_2 v_2 \ldots v_2 x_l = \overline{w}f(\overline{w})$.
Since $w, \overline{w} \in \Sigma_1^*$ and $f(w), f(\overline{w}) \in \Sigma_2^*$ and f is a bijection, we must have that $v_1 = v_2$. Since the conclusion of our assumption is that $v_1 = v_2$, it does not matter whether we substitute in parallel or not. Therefore $K = L(R)$ and $K \in SR$ (= E0L).

Theorem 5

Let Σ_1, Σ_2 and Σ_3 be disjoint alphabets and let $K_i \subseteq \Sigma_i^*$, $1 \leq i \leq 3$. Let $f : K_1 \to K_2$ and $g : K_1 \to K_3$ be bijective functions. Let $K = \{wf(w)g(w) \mid w \in K_1\}$.

(I) If $K \in ETOL$ then $K_1, K_2, K_3, K \in EDTOL$.

(II) If $K \in EOL$ then $K_1, K_2, K_3, K \in HDOL$.

(III) If $K \in ER$ then $K_1, K_2, K_3, K \in HDOL$.

Proof

In [3] it is shown that if
$\{wf(w) \mid w \in K_1\} \in ETOL$ then $K_1, K_2, \{wf(w) \mid w \in K_1\} \in EDTOL$.
This statement is stronger than the one in this theorem.

(II) The proof is quite similar to the proof of Theorem 2 (II).

(III) Follows from Lemma 3 and part (II) of this theorem.

Theorem 6

Let Σ_1, Σ_2 be disjoint alphabets and let $K_1 \subseteq \Sigma_1^*$, $K_2 \subseteq \Sigma_2^*$. Let $f: K_1 \to K_2$ be a monotone bijective function. Monotone in the sense that $|x| \leq |y|$ implies that $|f(x)| \leq |f(y)|$. Let $K = \{wf(w) \mid w \in K_1\}$.

(I) If $K \in RP$ then $K \in LIN$ and $K_1, K_2 \in RG$.

(II) If $K \in ED$ then $K \in LIN$ and $K_1, K_2 \in RG$.

Proof

Let $\Sigma = \Sigma_1 \cup \Sigma_2$. Let $K = L(G)$, where $G = (V, \Sigma, P_u, P_o, S)$ is a Russian parallel context-free grammar. We will assume that there are no useless letters in V.

Assume $A \in V$ and $A \Rightarrow^* w_1$, $A \Rightarrow^* w_2$ for some $w_1, w_2 \in \Sigma^*$. We have then $S \Rightarrow^* x_1 A x_2 \Rightarrow^* v_1 A v_2 A \ldots A v_k$ for some $x_1, x_2 \in (V \cup \Sigma)^*$ and $v_i \in \Sigma^*$, $1 \leq i \leq k$ and therefore $v_1 w_1 v_2 w_1 \ldots w_1 v_k \in L(G)$ and $v_1 w_2 v_2 w_2 \ldots w_2 v_k \in L(G)$. Since f is a monotone bijection, we have that if $w_1 \neq w_2$ then A can occur at most once in a sentential-form and w_1, w_2 cannot both be words in Σ_1^* or Σ_2^*.

Now construct $H = (V, \Sigma, P, S)$ as follows. If $(A, x_1 B x_2) \in P_u \cup P_o$ for some $x_1, x_2 \in (V \cup \Sigma)^*$ and $B \in V$ such that B can generate more than one word, then $(A, v_1 B v_2) \in P$ where $x_1 \Rightarrow^* v_1 \in \Sigma_1^*$ and $x_2 \Rightarrow^* v_2 \in \Sigma_2^*$. Note that x_1, x_2, B, v_1, v_2 are unique.
If $(A, x) \in P_u \cup P_o$ and no symbol in x can generate more than one word then $(A, v) \in P$ where $x \Rightarrow^* v$. Note again that v is unique.
It is now clear that $L(G) = L(H)$ and that implies that $K \in LIN$ and $K_1, K_2 \in RG$.
(II) Similar to (I) when we observe that ED = RPLL.

Let Σ_1, Σ_2 and Σ_3 be three disjoint alphabets. Let $K_i \subseteq \Sigma_i^*$, $1 \leq i \leq 3$, and let $f_i : K_1 \to K_i$, $i = 2, 3$, be length preserving isomorphisms from K_1 onto K_i. Let $K = \{ w f_2(w) f_3(w) \mid w \in K_1 \}$.
Using Theorems 5 and 6 we can get the following Figure 6.

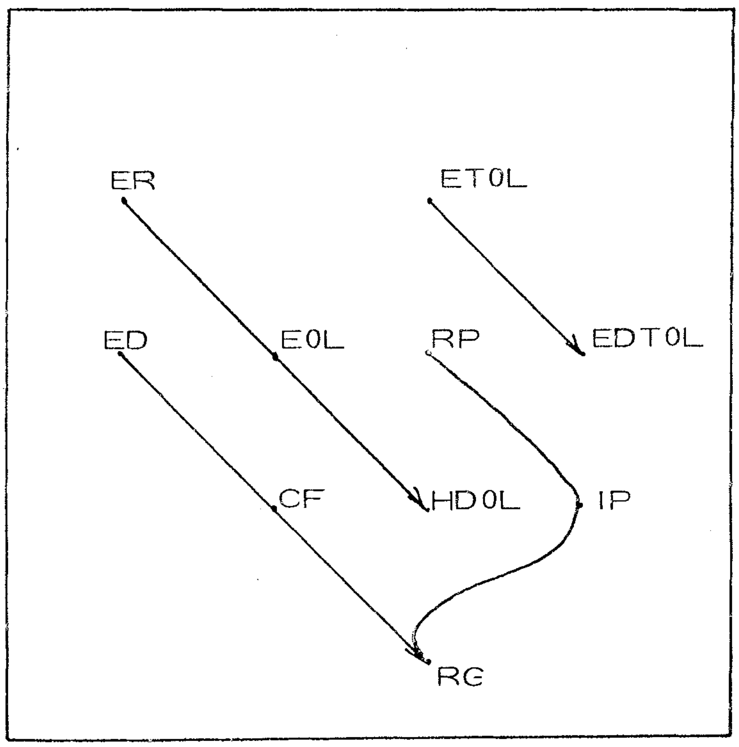

Figure 6.

If two nodes labelled X and Y are connected by an oriented edge then $K \in X$ implies that $K_1 \in Y$.

5. APPLICATIONS

In this section we will show how the theorems in Section 4 can be used to solve some of the open relations between families occurring in Figure 4.

It is known from the literature that mutual incomparability holds between families X and Y if there is no path from X to Y or from Y to X in Figure 4 with the following exceptions:

(ED, EDT0L), (ED, ET0L), (ER, EDT0L), (ER, ET0L), (ED, IP), (ED, RP), (ER, IP), and (ER, RP).

Mutual incomparability between

(ED, EDT0L), (ED, ET0L), (ER, EDT0L), and (ER, ET0L)

follows by the following theorem.

Theorem 7

ED $\not\subseteq$ ET0L and EDT0L $\not\subseteq$ ER.

Proof

In [2] it is proved that there exist context-free languages which are not EDT0L.

Let L be such a language. Then by using Theorems 1 and 2 we get that $c_{\#}^{2}(L)$ belongs to ED but not to ET0L. Hence ED $\not\subseteq$ ET0L.

Let $K_1 \subseteq \Sigma_1^*$ be a language in EDT0L\HD0L (e.g. $\{a,b\}^*$) and let $f_i : \Sigma_1^* \to \Sigma_i^*$, $i = 2,3$, be isomorphisms where Σ_1, Σ_2, and Σ_3 are disjoint alphabets. Then by Theorems 4 and 5 we get

$\{wf_1(w)f_2(w) \mid w \in K_1\} \in$ EDT0L\ER.

<u>Corollary</u>

$ED \not\subseteq RP$.

The remaining open problem is now whether or not IP is contained in ER.

<u>Acknowledgement</u>

I wish to thank Joost Engelfriet and Arto Salomaa for many useful discussions and comments.

REFERENCES

[1] Ehrenfeucht, A., Rozenberg, G.: The equality of E0L languages and codings of 0L languages. To appear in: International J. Computer Mathematics.

[2] Ehrenfeucht, A., Rozenberg, G.: On some context-free languages that are not deterministic ET0L languages, University of Colorado, Dept. of Computer Science, Technical Report #CU-CS-047-74, 1974.

[3] Ehrenfeucht, A., Rozenberg, G., Skyum, S.: A relationship between ET0L and EDT0L languages, University of Aarhus, Dept. of Computer Science, DAIMI-PB40, Oktober 1974.

[4] Ginsburg, S., Rice, H.G.: Two families of languages related to Algol, J. ACM 9, 350-371 (1962).

[5] Ginsburg, S.: The Mathematical Theory of Context-free Languages. McGraw-Hill (1966).

[6] Herman, G.T.: A Biologically Motivated Extension of ALGOL-like Languages, Information and Control 22, 487-502 (1973).

[7] Herman, G.T., Rozenberg, G.: Developmental systems and languages, North-Holland Publishing Co., to appear.

[8] Levitina, M.: O nekotoryh grammatikah s pravilami globalnoi podstanovki, Akad. Nauk SSSR Nauchno-Tekhn. Inform. Ser. 2, 32-36 (1973).

[9] Nielsen, M., Rozenberg, G., Salomaa, A., Skyum, S.: Nonterminals, homomorphisms and codings in different variations of 0L systems. I. Deterministic systems. II. Nondeterministic systems, to appear in Acta Informatica.

[10] Rose, G.F.: An extension of ALGOL-like languages, Comm. ACM 7, 52-61 (1964).

[11] Rozenberg, G., Salomaa, A. (ed.), L systems, Lecture Notes in Computer Science, Springer Verlag, 1974.

[12] Salomaa, A.: Formal Languages. Academic Press 1973.

[13] Salomaa, A.: Parallelism in rewriting systems, in Automata, Languages and Programming, Lecture Notes in Computer Science 14, 523-533, 1974.

[14] Siromoney, R., Krithivasan, K.: Parallel context-free languages, Information and Control 24, 155-161, 1974.

[15] Skyum, S.: On Extensions of ALGOL-like Languages, Information and Control 26, 1974.

[16] Skyum, S.: Parallel context-free languages, Information and Control 26, 1974.