ON EXTENSIONS OF ALGOL-LIKE LANGUAGES

by

Sven Skyum

DAIMI PB-25
March 1974

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

On Extensions of ALGOL-like Languages

by

Sven Skyum
Department of Computer Science
University of Aarhus
DENMARK

Mailing address:

Sven Skyum
Dept. of Computer Science
University of Aarhus
Ny Munkegade
8000 Aarhus C
Denmark

26 pages.

1 table.

Proposed running head: Extensions of ALGOL-like Languages.

$\mathbb{P}(A)$ denotes the set of subsets of A, λ denotes the empty word, and φ denotes the empty set.

CONTENTS:

Abstract

In this paper we introduce the notion of a level grammar and a level language. We define an extension to the extended definable sets and we characterize ALGOL-like languages, extended definable sets, and extensions of those as languages generated by level grammars with different kinds of restrictions on the productions and the use of productions. Finally we investigate some relations between the families defined.

On Extensions of ALGOL-like Languages

by

Sven Skyum

1. INTRODUCTION

Ginsburg and Rice (1962) introduced the notion of definable sets (ALGOL-like languages), by using formal systems of equations used to define constituent parts of ALGOL 60 (Naur (1960)). Ginsburg and Rice (1962) showed that the family of definable sets coincides with the family of context-free languages and they also pointed out that significant classes of programs in ALGOL were not representable as definable sets.

Rose (1964) then introduced the notion of extended definable sets, obtained by a greater class of formal systems generating languages. No characterization of extended definable sets by systems similar to context-free grammars has been given up to now.

Also Herman (1973) has defined an extension of the definable sets, the socalled simple recurrence languages, and shown that this family coincides with the family of EOL languages (see e.g. Herman and Rozenberg (1974)).

In this paper we introduce the notion of level grammars which is defined as context-free grammars, except that production rules are specified for all symbols (like for EOL-systems) and that the use of productions is restricted by the association of a so-called level number to each symbol in the sentential forms.

In this framework we are able to characterize the definable, extended definable, simple recurrence, and extended recurrence languages by simple restrictions on the level grammars and the use of productions. The latter family, extended recurrence languages, is an extension of extended definable sets in the same way as simple recurrence languages are an extension of definable sets.

Finally we investigate the relations between the families mentioned above.

2. PRELIMINARY DEFINITIONS AND RESULTS

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

Definition 1

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 2

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 $(1 \leq i \leq n)$.

Definition 3

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

Definition 4

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 $\mathfrak{F}_{ED}$.

Definition 5

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 theorems belong to Rose (1964) and Ginsburg and Rice (1962).

Theorem 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 $D_i(0)=\varphi$ $(1 \leq i \leq n)$ and for all $k \geq 1$

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

We will denote the family of definable sets by $\mathfrak{F}_D$.

Theorem 2

The family $\mathfrak{F}_D$ equals the family of context-free languages $(\mathfrak{F}_{CF})$.

If we use the notion from definition 1 and 5 we can give the following definition of the simple recurrence languages introduced by Herman (1973).

Definition 6

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 $\mathfrak{F}_{SR}$.

As simple recurrence languages are an extension to the definable languages, we will define an extension to the extended definable languages as follows.

Definition 7

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 $k \geq 1$ $E'(k) = g_{\Sigma;\ \xi;\ F}(E'(k-1))$.

We will denote the family of extended recurrence languages by $\mathfrak{F}_{ER}$.

Proposition 1

For every recurrence system $R = (\Sigma;\ \xi;\ F;\ \alpha)$ there exists a recurrence system $\overline{R} = (\Sigma;\ \overline{\xi};\ \overline{F};\ \overline{\alpha})$ such that

i) For all $1 \leq i \leq \overline{n}$, $\overline{F}_i$ does not contain λ,
ii) for all $1 \leq i \leq \overline{n}$, $\overline{\alpha}_i$ does not contain λ, and
iii) $L_E(\overline{R}) = L_E(R) \setminus \{\lambda\}$.

Proof

Let $R = (\Sigma;\ \xi_1, \ldots, \xi_n;\ F_1, \ldots, F_n;\ \alpha_1, \ldots, \alpha_n)$ be an arbitrary recurrence system and let $E'(k) = (E'_1(k), \ldots, E'_n(k))$, $k \in I$, be defined as in definition 7.

For $k \geq 0$, let

$$\pi(k) = \{ i \mid \lambda \in E_i^!(k) \}.$$

If $\pi(k_1) = \pi(k_2)$ then $\pi(k_1+1) = \pi(k_2+1)$. This follows because, if $i \in \pi(k_1+1)$ then there is a word x in F_i such that $x = \xi_{i_1} \xi_{i_2} \ldots \xi_{i_m}$, $m \geq 0$, where $\lambda \in E_{i_j}^!(k_1)$ for $1 \leq j \leq m$. Now if $\pi(k_1) = \pi(k_2)$ then $\lambda \in E_{i_j}^!(k_2)$ for $1 \leq j \leq m$ which means that $i \in \pi(k_2+1)$.
Since $\pi(k) \subseteq \{1, \ldots, n\}$, there must exist a k_1 and a k_2 such that $\pi(k_1) = \pi(k_2)$ and $k_1 \neq k_2$.
Let p and q be integers such that $0 \leq p < q$ and $\pi(p) = \pi(q)$. Let $d = q-p$.
Now let $\overline{R} = (\Sigma;\ \overline{\xi}_1, \ldots, \overline{\xi}_{nd+1};\ \overline{F}_1, \ldots, \overline{F}_{nd+1};\ \overline{\alpha}_1, \ldots, \overline{\alpha}_{nd+1})$ be the recurrence system defined as follows.

(i) For $0 \leq i \leq n-1$, $\overline{F}_{id+1} = \bigcup\limits_{\sigma \in M_d} \sigma(F_{i+1}) \setminus \{\lambda\}$.

(ii) For $2 \leq j \leq d$, $0 \leq i \leq n-1$, $\overline{F}_{id+j} = \bigcup\limits_{\sigma \in M_{j-1}} \sigma(F_{i+1}) \setminus \{\lambda\}$.

(iii) $\overline{F}_{nd+1} = \{ \overline{\xi}_{(n-1)d+1}, \ldots, \overline{\xi}_{nd} \}$.

where M_r, $1 \leq r \leq d$, is the set of all substitutions σ such that, for each $x \in \Sigma$, $\sigma(x) = \{x\}$ and $\sigma(\xi_i)$ is either $\overline{\xi}_{(i-1)d+r}$ or λ (λ only if $i \in \pi(p+r-1)$).

(iv) For $0 \leq i \leq n-1$, $\overline{\alpha}_{id+1} = E_{i+1}^!(p) \setminus \{\lambda\}$.

(v) For $0 \leq i \leq n-1$, $2 \leq j \leq d$, $\overline{\alpha}_{id+j} = \varphi$.

(vi) $\overline{\alpha}_{nd+1} = \bigcup\limits_{i=1}^{n-1} E_n^!(i) \setminus \{\lambda\}$.

k	1	2	3	.....d	...	(n−1)d+1	(n−1)d+2	nd	nd+1
0	$E'_1(p)\setminus\{\lambda\}$	φ	φ	φ		$E'_n(p)\setminus\{\lambda\}$	φ	φ	$\bigcup_{i=1}^{p-1} E'_n(i)\setminus\{\lambda\}$
1	$F_1 \cap \Sigma^+$	$E'_1(p+1)\setminus\{\lambda\}$	$F_1 \cap \Sigma^+$	$F_1 \cap \Sigma^+$		$F_n \cap \Sigma^+$	$E'_n(p+1)\setminus\{\lambda\}$	$F_n \cap \Sigma^+$	$E'_n(p)\setminus\{\lambda\}$
2	$\overline{E}'_1(2)$	$\overline{E}'_2(2)$	$E'_1(p+2)\setminus\{\lambda\}$	$\overline{E}'_d(2)$		$\overline{E}'_{(n-1)d+1}(2)$	$\overline{E}'_{(n-1)d+2}(2)$	$\overline{E}'_{nd}(2)$	$E'_n(p+1)\setminus\{\lambda\} \cup F_n \cap \Sigma$
. . .			$\ddots$				$\ddots$		$\vdots$
d−1				$E'_1(q-1)\setminus\{\lambda\}$				$E'_n(q-1)\setminus\{\lambda\}$	$\vdots$
d	$E'_1(q)\setminus\{\lambda\}$					$E'_n(q)\setminus\{\lambda\}$			$\vdots$
d+1		$E'_1(q+1)\setminus\{\lambda\}$					$E'_n(q+1)\setminus\{\lambda\}$		$\vdots$
. . .		$\ddots$					$\ddots$		$\vdots$

Table 1.

The following equations hold true for $0 \leq i \leq n-1$, $1 \leq j \leq d$, and $0 \leq r$. (See table 1.)

(1) $\overline{E}'_{id+j}(rd+j-1) = E'_{i+1}(p+rd+j-1) \setminus \{\lambda\}$.

We show (1) by induction on $k = rd+j \geq 1$, where $1 \leq j \leq d$.
Assume that $j = 1$ and $r = 0$.
For $0 \leq i \leq n-1$ we then have $\overline{E}'_{id+1}(0) = \overline{\alpha}_{id+1} = \overline{E}_{i+1}(p) \setminus \{\lambda\}$.

Assume that (1) holds true for $k = r_1 d+j_1$. Then there are two possibilities:

a) $j_1 = d$, and therefore $k+1 = (r_1+1)d+1$.
Then for $0 \leq i \leq n-1$
$w \in \overline{E}'_{id+1}((r_1+1)d)$
iff
there exist a word $w_1 \overline{\xi}_{i_1 d} w_2 \overline{\xi}_{i_2 d} \ldots \overline{\xi}_{i_k d} w_{k+1}$ in $\overline{F}_{id+1}$ where $w_m \in \Sigma^*$ for $1 \leq m \leq k+1$, and words $\overline{w}_m \in \overline{E}'_{i_m d}(r_1 d+d-1)$ for $1 \leq m \leq k$, such that $\overline{w}_{m_1} = \overline{w}_{m_2}$ if $i_{m_1} = i_{m_2}$ and
$w = w_1 \overline{w}_1 w_2 \overline{w}_2 \ldots \overline{w}_k w_{k+1}$
iff
there exists a word $w'_1 \xi_{i_1} w'_2 \xi_{i_2} \ldots \xi_{i_k} w'_{k+1}$ in $F_{i+1} \setminus \{\lambda\}$ where $w'_m \in (\Sigma \cup \{\xi_1, \ldots, \xi_n\} \setminus \{\xi_{i_1}, \ldots, \xi_{i_k}\})^*$ for $1 \leq m \leq k+1$, and $\lambda \in E'_t(p+r_1 d+d-1)$ if ξ_t occurs in some of the words w'_m, and there exist words $\overline{w}_{m_1} = \overline{w}_{m_2}$ if $i_{m_1} = i_{m_2}$ and $w = w_1 \overline{w}_1 w_2 \overline{w}_2 \ldots \overline{w}_{k+1}$ where w_m is the word we obtain by substituting λ for all occurrences of ξ_t, $1 \leq t \leq n$, in w'_m.
iff
$w \in E'_{i+1}(p+(r_1+1)d) \setminus \{\lambda\}$.

b) $j_1 < d$, and therefore $k+1 = r_1 d+j_1+1$.
Then for $0 \leq i \leq n-1$ we can show that
$\overline{E}'_{id+j_1+1}(r_1 d+j_1) = E'_{i+1}(p+r_1 d+j_1) \setminus \{\lambda\}$ in a similar way.

Because of equation (1), $\overline{E}'_{nd+1}(r) = \bigcup_{i=(n-1)d+1}^{nd} \overline{E}'_i(r-1)$ for $r \geq 1$, and $\overline{E}'_{nd+1}(0) = \bigcup\limits_{i=0}^{n-1} E'_n(i)$ we have that $L_E(R) \setminus \{\lambda\} \subseteq L_E(\overline{R})$.

To show the other inclusion it suffices to show that for $0 \leq i \leq n-1$, $1 \leq j \leq d$, $0 \leq r$, $0 < s < d$, and $0 \leq t = rd+j-1-s$ the following inclusion holds true.

(2) $\overline{E}'_{id+j}(t) \subseteq \overline{E}'_{id+j}(rd+j-1)$.

For $i \in I$ let $W(i)$ denote the n-tuple $(\overline{E}'_j(rd+j-1), \ldots, \overline{E}'_{n-1\,d+j}(rd+j-1))$ where $i = rd+j-1$, and $1 \leq j \leq d$.

Let $r \geq 0$, $1 \leq j \leq d$, $0 < s < d$, and $0 \leq t = rd+j-1-s$.
Then

$$(\overline{E}'_{s+1}(0), \ldots, \overline{E}'_{(n-1)d+s+1}(0)) = (\varphi, \ldots, \varphi) \subseteq W(s)$$

and therefore because of the construction of $\overline{R}$

$$(\overline{E}'_{s+2}(1), \ldots, \overline{E}'_{(n-1)d+s+2}(1)) \subseteq W(s+1)$$

$$\vdots$$

$$(\overline{E}'_d(d-s-1), \ldots, \overline{E}'_{nd}(d-s-1)) \subseteq W(d-1)$$

$$\vdots$$

$$(\overline{E}'_d(rd-s-1), \ldots, \overline{E}'_{nd}(rd-s-1)) \subseteq W(rd-1)$$

$$\vdots$$

$$(\overline{E}'_j(rd+j-s-1), \ldots, \overline{E}'_{(n-1)d+j}(rd+j-s-1)) \subseteq W(rd+j-1).$$

This inclusion is equivalent to inclusion (2).

Corollary

For a given recurrence system R it is decidable whether $L_E(R)$ contains λ or not.

Definition 8

An EOL-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 $\mathbb{P}((V, V^*))$, such that for every $A \in V$, there exists a $x \in V^*$ such that (A, x) is in P, w (the axiom) is a word in V^+ (it is no restriction to assume that $w \in V$), and Σ (the terminal alphabet) is a subset of V.

Definition 9

The EOL-language L(G) of an EOL-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 there exist $u_1, \ldots, u_k$ in V and $v_1, \ldots, v_k$ in V^* such that $z = u_1 \ldots u_k$, $y = v_1 \ldots v_k$, and (u_i, v_i) is in P for each $1 \leq i \leq k$.

The following theorem belongs to Herman (1973).

Theorem 3

The family $\mathfrak{F}_{SR}$ equals the family of EOL-languages ($\mathfrak{F}_{EOL}$).

3. LEVEL GRAMMARS AND LEVEL LANGUAGES

Definition 10

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 $\mathbb{P}((V, V^*))$,
$S \in V$ is the start symbol, and
$\Sigma \subseteq V$ is the terminal alphabet.

Definition 11

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.

Definition 12

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 $h : (V, I)^* \to V^*$ 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

$$
\begin{aligned}
L(G,0) &= \varphi \\
L(G,1) &= \{ab\} \\
L(G,2) &= \{aab, aabb\} \\
&\vdots \\
L(G,n) &= \{a^{2^{n-1}} b^i \mid 1 \le i \le 2^{n-1}\} \\
&\vdots \\
L(G) &= \bigcup_{n \ge 0} L(G,n) = \{a^{2^n} b^i \mid n \ge 0,\ 1 \le i \le 2^n\}
\end{aligned}
$$

The family of level languages will be denoted by $\mathfrak{F}_{LL}$.

It is easy to check that the following theorem is true.

Theorem 4

$$\mathfrak{F}_{LL} = \mathfrak{F}_{EOL} \ (= \mathfrak{F}_{SR}).$$

Definition 13

Let $G = (V, P, S, \Sigma)$ be a level grammar.
We write $w_1(A,n)w_2(A,n)\ldots w_{k-1}(A,n)w_k \Rightarrow_P w_1 w w_2 w \ldots w_{k-1} w w_k$
if $w_i \in ((V, I) \setminus (A,n))^*$ $(1 \le i \le k)$ 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 14

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 $L_P(G) = L_P(G,n)$ 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

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

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

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

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

$\vdots$

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

$\vdots$

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

The family of parallel level languages is denoted by $\mathfrak{F}_{PLL}$.

We will now give a normal form for a level grammar.

Definition 15

A level grammar $G = (V,\ P,\ S,\ \Sigma)$ is said to be in normal form if there exists a special symbol $\#$ in $V \backslash \Sigma$ such that

1) P is a finite subset of $\mathbb{P}((N, N^* \cup \Sigma\ )) \cup \mathbb{P}(\Sigma \cup \{\#\},\ \#)$ and
2) $S \in N$, where $N = V \backslash \Sigma \cup \{\#\}$.

Proposition 2

For every level grammar G there exists a level grammar G' in normal form such that $L(G) = L(G')$ and $L_P(G) = L_P(G')$.

Proof

Let $G = (V,\ P,\ S,\ \Sigma)$ be an arbitrary level grammar.

Let $\Sigma' = \{A' \mid A \in \Sigma\}$ be disjoint from V.

If $y = x_1 A_1 x_2 \ldots x_{k-1} A_{k-1} x_k$, where $A_i \in \Sigma$ $(1 \leq i \leq k-1)$ and $x_i \in (V \backslash \Sigma)^*$ $(1 \leq i \leq k)$ then we will denote $x_1 A_1' x_2 \ldots x_{k-1} A_{k-1}' x_k$ by y'.

Now define $G' = (V', P', S', \Sigma)$, which will be in normal form, as follows.

$$V' = V \cup \Sigma' \cup \{\#\} \quad (\# \notin V \cup \Sigma').$$

(I) For all $(A, x) \in P, (A', x') \in P'$;

(II) For all $A \in \Sigma$, $(A', A) \in P'$ and $(A, \#) \in P'$;

(III) $(\#, \#) \in P'$.

Only those ordered pairs are in P' which are defined to be there in virtue of (I) to (III).

We will prove that $L_P(G) = L_P(G')$. The equation $L(G) = L(G')$ is proved in the same way.

Let $x \in L_P(G)$, then there exists a sequence $w_0, w_1, \ldots, w_k$ of words in $(V, I)^*$ such that $w_0 = (S, 0)$, $w_i \underset{G}{\Rightarrow} P \; w_{i+1}$ for $0 \leq i \leq k-1$, and $h(w_k) = x$. Because of (I) and (II) above we then have that there exists a sequence $v_0, v_1, \ldots, v_{k+j}$ of words in $(V', I)^*$ such that $v_0 = (S', 0)$, $v_i \underset{G'}{\Rightarrow} P \; v_{i+1}$ for $0 \leq i \leq k+j-1$, $h(v_k) = h(w_k)' = x'$ and $h(v_{k+j}) = x$. This proves that $L_P(G) \subseteq L_P(G')$.

Now let $x \in L_P(G')$, then there exists a sequence $v_0, v_1, \ldots, v_k$ of words in (V', I) such that $v_0 = (S', 0)$, $v_i \underset{G'}{\Rightarrow} P \; v_{i+1}$ for $0 \leq i \leq k-1$ and $h(v_k) = x$. Assume that $v_k \in (\Sigma, n)^*$. Then there exists a sequence $v_0', \ldots, v_p'$ such that $v_0' = (S', 0)$, $v_i' \underset{G'}{\Rightarrow} P \; v_{i+1}$ for $0 \leq i \leq p-1$, $v_p' \in (\Sigma', n-1)^*$ and $h(v_p') = x'$. Because of (I) and (II) above we then have that there exists a sequence $w_0, \ldots, w_p$ such that $w_0 = (S, 0)$, $w_i \underset{G}{\Rightarrow} P \; w_{i+1}$ for $0 \leq i \leq p-1$, and $h(w_p) = x$.

Theorem 5

$$\mathfrak{F}_{PLL} = \mathfrak{F}_{ER}.$$

Proof

Let $R = (\Sigma; \xi_1, \ldots, \xi_n; F_1, \ldots, F_n; \alpha_1, \ldots, \alpha_n)$ be an arbitrary recurrence system, and $\mathbf{L}_E(R) = \bigcup_{k \geq 0} E_n'(k)$ the corresponding extended recurrence language.

Let $\mathbf{\Sigma}' = \{A' \mid A \in \Sigma\}$ be disjoint from $\Sigma \cup \{\xi_1, \ldots, \xi_n\}$.

Because of proposition 1 we can, without loss of generality, assume that for all $1 \leq i \leq n$ λ is neither in F_i nor in α_i.

We now define $G = (V, P, S, \Sigma)$ as follows.

$V = \Sigma \cup \Sigma' \cup \{\xi_1, \xi_2, \ldots, \xi_n\} \cup \{\#\}$.

(I) For all $A \in \Sigma \cup \{\#\}$, $(A, \#)$ is in P;
(II) For all $A \in \Sigma$, (A', A') and (A', A) are in P;
(III) For all $1 \leq i \leq n$ and $y \in F_i$, (ξ_i, y') is in P;
(IV) For all $1 \leq i \leq n$ and $y \in \alpha_i$, (ξ_i, y) is in P;
These are the only productions in P.
$S = \xi_n$.

For $1 \leq i \leq n$, let $G_i = (V, P, \xi_i, \Sigma)$.

We prove by induction on j that, for $1 \leq i \leq n$

(1) $L_P(G_i, j) = E_i^!(j-1)$.

Assume that $j = 1$. Then

$w \in L_P(G_i, 1)$ iff $w \in \Sigma^+ \wedge (\xi_i, w) \in P$ iff $w \in \alpha_i$ iff $w \in E_i^!(0)$.

Assume that (1) holds true for all $j \leq m$ and $1 \leq i \leq n$.
Then

$w \in L_P(G_i, m+1)$

iff

there exist a production $(\xi_i, A_1^! A_2^! \ldots A_k^!) \in P$ where $A_i^! \in \Sigma' \cup \{\xi_1, \xi_2, \ldots, \xi_n\}$ and words w_q, $1 \leq q \leq k$, such that $w_q = A_q$ if $A_q^! \in \Sigma'$, $w_q \in L_P(G_{i_q}, m)$ if $A'_q = \xi_{i_q}$, $w_{q_1} = w_{q_2}$ if $A_{q_1}^! = A_{q_2}^!$, and $w = w_1 w_2 \ldots w_k$

iff

there exist a word $A_1 A_2 \ldots A_k \in F_i$ and words w_q, $1 \leq q \leq k$, such that $w_q = A_q$ if $A_q \in \Sigma$, $w_q \in E_{i_q}^!(m-1)$ if $A_q = \xi_{i_q}$, $w_{q_1} = w_{q_2}$ if $A_{q_1} = A_{q_2}$, and $w = w_1 w_2 \ldots w_k$

iff

$w \in E_i(m)$.

This proves that $\mathfrak{F}_{ER} \subseteq \mathfrak{F}_{PLL}$.

To prove the other inclusion consider an arbitrary level grammar $G = (V, P, S, \Sigma)$.
We know that there exists a level grammar G' in normal form such that $L_P(G) = L_P(G')$. We can assume that

$$G' = (\{\xi_1, \xi_2, \ldots, \xi_n\} \cup \Sigma, P', \xi_n, \Sigma).$$

Now define a recurrence system

$$R = (\Sigma; \xi_1, \xi_2, \ldots, \xi_n; F_1, \ldots, F_n; \alpha_1, \ldots, \alpha_n)$$

where $F_i = \{w \in (V \backslash \Sigma)^* \mid (\xi_i, w) \in P'\}$ and $\alpha_i = \{w \in \Sigma \cup \varphi \mid (\xi_i, w) \in P'\}$.

For $1 \leq i \leq n$, let $G_i' = (\{\xi_1, \xi_2, \ldots, \xi_n\} \cup \Sigma, P', \xi_i, \Sigma)$.

We prove by induction on j that, for $1 \leq i \leq n$

(2) $L_P(G_i', j) = E_i'(j-1)$.

Assume that $j = 1$. Then

$w \in L_P(G_i', 1)$ iff (ξ_i, w) is in P' and $w \in \Sigma^*$ iff $w \in \alpha_i$ iff $w \in E_i'(0)$.

Assume that (2) holds true for all $j \leq m$ and $1 \leq i \leq n$.
Then

$w \in L_P(G_i', m+1)$

iff

there exist a production $(\xi_i, \xi_{i_1} \xi_{i_2} \ldots \xi_{i_k})$ in P' and words w_q $1 \leq q \leq k$ such that $w_q \in L_P(G_{i_q}', m)$, $w_{q_1} = w_{q_2}$ if $\xi_{i_{q_1}} = \xi_{i_{q_2}}$, and $w = w_1 w_2 \ldots w_n$

iff

there exist a word $\xi_{i_1} \xi_{i_2} \ldots \xi_{i_k}$ in F_i and words w_q, $1 \leq q \leq k$, such that $w_q \in E_{i_q}'(m-1)$, $w_{q_1} = w_{q_2}$ if $\xi_{i_{q_1}} = \xi_{i_{q_2}}$, and $w = w_1 w_2 \ldots w_n$

iff

$w \in E_i(m)$.

This completes the proof of theorem 5.

As a consequence of the constructions in the proof of theorem 5 we have the following corollary.

<u>Corollary 1</u>

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

<u>Remark</u>

It can be shown that corollary 1 remains true if we write $L(R) = L(R')$ instead of $L_E(R) = L_E(R')$. This is indeed already pointed out by Herman (1973).

<u>Definition 16</u>

A restricted level grammar $G = (V, P, S, \Sigma)$ is a level grammar as defined in definition 10 with the restriction that for all $A \in \Sigma$, $(A, A) \in P$.

<u>Proposition 3</u>

If $G = (V, P, S, \Sigma)$ is a restricted level grammar then there exists a grammar $G' = (V', P', S', \Sigma)$ such that for all $A \in \Sigma$, (A, A) is the only production in P with A as the first component, $L(G) = L(G')$ and $L_P(G) = L_P(G')$.

<u>Proof</u>

Let $G = (V, P, S, \Sigma)$ be an arbitrary restricted level grammar. Define $G' = (V', P', S', \Sigma)$ as follows.

$V' = \{A' \mid A \in V\} \cup \Sigma$,

(I) For all $(A, x) \in P$, $(A', x') \in P'$,

(II) for all $A \in \Sigma$, (A', A) and (A, A) is in P'.

These are the only productions in P'.

We will prove that $L(G) = L(G')$. The equation $L_P(G) = L_P(G')$ is proved in the same way.

It is clear from the definition that $L(G) \subseteq L(G')$. Now let $x \in L(G')$, then there exists a sequence $w_0, w_1, \ldots, w_k$ of words in $(V', I)^*$ such that $w_0 = (S', 0)$, $w_i \underset{G'}{\Rightarrow} w_{i+1}$ for $0 \leq i \leq k-1$, and $h(w_k) = x$. If we remove the markers in all the words w_i then we get a sequence $v_0, \ldots, v_k$ of words in $(V, I)^*$ such that $v_0 = (S, 0)$, $v_i \underset{G}{\Rightarrow} v_{i+1}$ for $0 \leq i \leq k-1$, (this is trivially true if the production used is defined in (I). If a production defined in (II) is used it remains true because we know that $(A, A) \in P$ for all $A \in \Sigma$.), and $h(v_k) = x$.

This completes the proof.

The family of restricted (parallel) level languages is denoted by $\mathfrak{F}_{RLL}$ $(\mathfrak{F}_{RPLL})$.

Theorem 6

$$\mathfrak{F}_{RLL} = \mathfrak{F}_{CF} \; (= \mathfrak{F}_{D}).$$

Proof

Easy to check.

Theorem 7

$$\mathfrak{F}_{RPLL} = \mathfrak{F}_{ED}.$$

Proof

Let $(\Sigma;\ \xi_1, \ldots, \xi_n;\ F_1, \ldots, F_n)$ be an arbitrary (n-ary) format, and $E(k) = (E_1(k), \ldots, E_n(k))$ the approximating sequence.

We define $G = (V, P, S, \Sigma)$ as follows.

$$V = \Sigma \cup \{\xi_1, \xi_2, \ldots, \xi_n\}.$$

(I) For all $A \in \Sigma$, (A, A) is in P;

(II) for all $1 \leq i \leq n$, and $y \in F_i$, (ξ_i, y) is in P.

These are the only productions in P.

$S = \xi_n$.

For $1 \leq i \leq n$, let $G_i = (V, P, \xi_i, \Sigma)$.

We prove by induction on j that, for $1 \leq i \leq n$

(3) $L_P(G_i, j) = E_i(j)$.

Assume that $j = 0$. Then

$$L_P(G_i, 0) = \varphi = E_i(0).$$

Assume that (3) holds true for $j \leq m$ and $1 \leq i \leq n$.
Then

$w \in L_P(G_i, m+1)$

iff

there exist a word $A_1 A_2 \dots A_k$ in V^* such that $(\xi_i, A_1 A_2 \dots A_k) \in P$, and words w_q, $1 \leq q \leq k$ such that $w_q = A_q$ if $A_q \in \Sigma$, $w_q \in L_P(G_{i_q}, m)$ if $A_q = \xi_{i_q}$, $w_{q_1} = w_{q_2}$ if $A_{q_1} = A_{q_2}$, and $w = w_1 w_2 \dots w_k$

iff

there exist a word $A_1 A_2 \dots A_k$ in F_i and words w_q, $1 \leq q \leq k$ such that $w_q = A_q$ if $A_q \in \Sigma$, $w_q \in E_{i_q}(m)$ if $A_q = \xi_{i_q}$, $w_{q_1} = w_{q_2}$ if $A_{q_1} = A_{q_2}$, and $w = w_1 w_2 \dots w_k$

iff

$w \in E_i(m+1)$.

This proves that $\mathfrak{F}_{ED} \subseteq \mathfrak{F}_{RPLL}$.

Now let $G = (V, P, S, \Sigma)$ be an arbitraty restricted level grammar. We can assume that $V = \{\xi_1, \dots, \xi_n\} \cup \Sigma$ and $S = \xi_n$.

If $S \in \Sigma$, then $L_P(G) = \{S\} \in \mathfrak{F}_{ED}$ because (S, S) is the only production for S.

Define an n-ary format $(\Sigma; \xi_1, \dots, \xi_n; F_1, \dots, F_n)$ where $F_i = \{w \mid (\xi_i, w) \text{ is in } P\}$.

Now we can use the same proof as above to show that $L_P(G_i, j) = E_i(j)$ for all $j \geq 0$, and $1 \leq i \leq n$.

This completes the proof of $\mathfrak{F}_{ED} = \mathfrak{F}_{RPLL}$.

4. SOME RELATIONS BETWEEN THE FAMILIES $\mathfrak{F}_{LL}$, $\mathfrak{F}_{PLL}$, $\mathfrak{F}_{RLL}$, and $\mathfrak{F}_{RPLL}$.

Theorem 8

$\mathfrak{F}_{RLL} \subsetneq \mathfrak{F}_{LL}$.

Proof

From Theorems 6 and 4 we know that $\mathfrak{F}_{RLL} = \mathfrak{F}_{CF}$ and $\mathfrak{F}_{LL} = \mathfrak{F}_{EOL}$. The theorem follows now from the fact that $\mathfrak{F}_{CF} \subsetneq \mathfrak{F}_{EOL}$ (see e.g. Herman and Rozenberg (1974)).

Theorem 9

$\mathfrak{F}_{RLL} \subsetneq \mathfrak{F}_{RPLL}$.

Proof

We know that $\mathfrak{F}_{RLL} = \mathfrak{F}_{D}$ and $\mathfrak{F}_{RPLL} = \mathfrak{F}_{ED}$ from theorems 6 and 7 respectively. From Rose (1964) we know that $\mathfrak{F}_{D} \subsetneq \mathfrak{F}_{ED}$ which completes the proof.

Theorem 10

$\mathfrak{F}_{LL} \subsetneq \mathfrak{F}_{PLL}$.

Proof

Let $G = (V, P, S, \Sigma)$ be an arbitrary level grammar in normal form.

Let $t = \max\{ |w| \,|\, \exists A \in V : (A, w) \in P\}$.

Define $G' = (V', P', S', \Sigma)$ as follows.

$$V' = \bigcup_{A \in V \setminus \Sigma} \{A^{(1)}, A^{(2)}, \ldots, A^{(t)}\} \cup \Sigma$$

(I) If $(A, A_1 A_2 \ldots A_k)$ is a production in P where $0 \leq k \leq t$ and $A_i \in V \setminus \Sigma$ for $1 \leq i \leq k$ then, for all $1 \leq i \leq t$, $(A^{(i)}, A_1^{(1)}, A_2^{(2)} \ldots A_k^{(k)})$ is a production in P'.

(II) If (A, B) is in P and $A \in V \setminus \Sigma$, $B \in \Sigma$ then, for all $1 \leq i \leq t$, $(A^{(i)}, B)$ is in P'.

(III) If (A, B) is in P and $A \in \Sigma$ then $(A, B^{(1)})$ is in P'.

These are the only productions in P'.

$S' = S^{(1)}$.

We can now show that

(4) $L(G) = L_P(G') \quad (= L(G'))$.

For each $A \in V$ define

$$G_A = (V, P, A, \Sigma),$$

and for each $A \in V'$ define

$$G'_A = (V', P', A, \Sigma).$$

Now we prove by induction on j that, for all $1 \leq i \leq t$ and all $A \in V$

(5) $L(G_A, j) = L_P(G'_{A^{(i)}}, j) \quad (= L(G'_{A^{(i)}}, j))$

Let $j = 1$. Then

$w \in L(G_A, 1)$

iff

$w \in \Sigma^* \wedge (A, w) \in P$

iff

$w \in \Sigma^* \wedge (A^{(i)}, w) \in P'$ for all $1 \leq i \leq t$

iff

$w \in L_P(G'_{A^{(i)}}, 1) \quad (\wedge\ w \in L(G'_{A^{(i)}}, 1))$ for all $1 \leq i \leq t$.

Now assume that (5) holds true for $j = m \geq 1$ and $1 \leq i \leq t$.
Then

$w \in L(G_A, m+1)$

iff

there exist a word $A_1 A_2 \ldots A_k \in (V \setminus \Sigma)^*$ (G is in normal form) such that $(A, A_1 A_2 \ldots A_k) \in P$, and words $w_q \in L(G_{A_q}, m)$, $1 \leq q \leq k$ such that $w = w_1 w_2 \ldots w_k$

iff

there exist a word $A_1 A_2 \ldots A_k \in (V \setminus \Sigma)^*$ such that for all $1 \leq i \leq t$, $(A^{(i)}, A_1^{(1)} A_2^{(2)} \ldots A_k^{(k)}) \in P'$ and words $w_q \in L(G'_{A_q^{(q)}}, m)$, $1 \leq q \leq k$, such that $w = w_1 w_2 \ldots w_k$

(iff

$w \in L(G'_{A^{(i)}}, m+1)$ for all $1 \leq i \leq t$)

iff

there exist a word $A_1 A_2 \dots A_k \in (V \setminus \Sigma)^*$ such that for all $1 \le i \le t$ $(A^{(i)}, A_1^{(1)} A_2^{(2)} \dots A_k^{(k)}) \in P'$ and words $w_q \in L(G'_{A_q^{(q)}}, m)$ $1 \le q \le k$ such that $w_{q_1} = w_{q_2}$ if $A_{q_1}^{(q_1)} = A_{q_2}^{(q_2)}$ (this can never happen!) and $w = w_1 w_2 \dots w_k$

iff

$w \in L_P(G'_{A^{(i)}}, m+1)$ for all $1 \le i \le t$.

(4) is now established by

$$L(G) = \bigcup_n L(G, n) = \bigcup_n L(G_S, n) = \bigcup_n L_P(G'_{S^{(1)}}, n) = \bigcup_n L_P(G', n) = L_P(G')$$

$(= L(G'))$

The inclusion is proper because the language $L_1 = \{(ab^m)^{2^n} \mid m, n \ge 0\}$ is not in the family $\mathfrak{F}_{LL}$ (Herman (to appear)) but is parallel generated by the following (restricted) level grammar

$$G = (\{S, B, a, b\}, \{(S, SS), (S, a), (S, aB), (B, bB), (B, b), (a, a), (b, b)\}, S, \{a, b\}).$$

<u>Theorem 11</u>

$$\mathfrak{F}_{RPLL} \subsetneq \mathfrak{F}_{PLL}.$$

<u>Proof</u>

The inclusion is true by definition. That it is proper follows from the language

$$L_2 = \{a^n b^n a^n \mid n \ge 1\}$$

which does not belong to $\mathfrak{F}_{PRLL}$ $(= \mathfrak{F}_{ED})$ (see e.g. Rose (1964)) but is parallel generated by the following level grammar in normal form

$$G = (\{A, A', B, B', a, b, \#, S\}, \{(A, AA'), (A, a), (A', A'), (A', a), (B, BB'), (B, b), (B', B'), (B', b), (a, \#), (b, \#), (\#, \#), (S, ABA)\}, S, \{a, b\}).$$

<u>Theorem 12</u>

The families $\mathfrak{F}_{LL}$ and $\mathfrak{F}_{RPLL}$ are incomparable.

<u>Proof</u>

The language L_1 from above belongs to $\mathfrak{F}_{RPLL} \backslash \mathfrak{F}_{LL}$ and L_2 belongs to $\mathfrak{F}_{LL} \backslash \mathfrak{F}_{RPLL}$.

Theorems 8-12 can be summarized in the following diagram:

$$\begin{array}{ccccc} & & \mathfrak{F}_{PLL} & & \\ & \subsetneq & & \supsetneq & \\ \mathfrak{F}_{LL} & & & & \mathfrak{F}_{RPLL} \\ & \supsetneq & & \subsetneq & \\ & & \mathfrak{F}_{RLL} & & \end{array}$$

References

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

Herman, G.T. (to appear), Closure Properties of Some Families of Languages Associated with Biological Systems, Information and Control.

Herman, G.T. (1973), A Biologically Motivated Extension of ALGOL-like languages, Information and Control 22, 487-502.

Herman, G.T. and Rozenberg, G. (1974), Developmental systems and languages, North Holland Publishing Co., to appear.

Naur, P. (Ed.) et al. (1960), Report on the algorithmic language ALGOL 60, Comm. ACM 3, 299-314.

Rose, G.F. (1964), An extension of ALGOL-like languages, Comm. ACM 7, 52-71.