DEVELOPMENTAL SYSTEMS
WITH FRAGMENTATION

G. Rozenberg
K. Ruohonen
and
A. Salomaa

DAIMI PB-36
August 1974

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

Abstract.

The paper introduces a new class of L systems, where it is possible to continue derivations from certain specified subwords of the words obtained. Such L systems (called L systems with fragmentation or just JL systems) are of interest both from biological and formal language theory point of view. The paper deals with JL systems without interactions, discusses the basic properties of the language families obtained, as well as their position in the L hierarchy. Finally, two infinite hierarchies of language families are obtained by limited fragmentation, the notions being analogous to those of ultralinearity and finiteness of index for context-free languages.

CONTENTS

1. INTRODUCTION 2

2. DEFINITIONS 6

3. PROPERTIES OF J0L AND JT0L 10

4. THE POSITION OF J0L, JT0L AND SOME EXTENSIONS IN THE L HIERARCHY 15

5. HIERARCHIES DUE TO k-LIMITED FRAGMENTATION 21

REFERENCES 25

1. INTRODUCTION

The original aim of the theory of Lindenmayer systems (abbreviated L systems) was to provide mathematical models for the development of simple filamentous organisms. At first L systems were defined as linear arrays of finite automata, but were afterwards reformulated into the more suitable framework of grammar-like constructs. From then on, the theory of L systems has been developed essentially as a branch of formal language theory. In fact it constitutes today one of the most vigorously investigated areas of formal language theory.

One of the new aspects in formal language theory brought about by L systems is the study of the different ways of defining, "squeezing out", a language of a system. The traditional way of defining a language of a grammar is to intersect the set of all derivable words ("sentential forms") by V_T^*, where V_T is the specified terminal alphabet. Thus, this mechanism, usually referred to as E-mechanism in the theory of L systems, defines the language by excluding strings which are not of the right form. Another language defining mechanism widely studied in L systems theory is to apply a literal homomorphism (coding) to the set of all derivable words. This mechanism, usually referred to as C-mechanism, defines the language by transforming strings without excluding any of them. It has been shown in [1] and [2] that for the basic L families, the families of 0L and T0L languages, the generative capacity of these two mechanisms is the same, i.e., E0L = C0L and ET0L = CT0L. Further results concerning E, C and related mechanisms have been established in [5].

This paper introduces another mechanism, referred to as J-mechanism, which is similar to C in that it transforms strings without excluding any of them. However, otherwise, it is quite different from C. The basic idea is the following. The right sides of the productions may contain occurrences of a special symbol q. This symbol induces a cut in the string under scan, and the derivation may continue from any of the parts obtained. Thus, if we apply the productions $a \rightarrow aqa$, $b \rightarrow ba$, $c \rightarrow qb$ to the word abc, we obtain the words a, aba, and b.

The basic biological significance of the J-operator, which will be discussed in more detail at the end of this Introduction, is that it provides us with a new formalism for blocking communication, splitting the developing filament and cell death. The motivation of getting another way of squeezing out a language from a system was already discussed. Another motivation belonging to formal language theory is that this approach explores further the subword point of view, which has recently turned out to be very fruitful, cf [3] and the references given there.

In this paper, we shall discuss JL systems without interactions, i.e., rewriting happens in a context-free manner. JL systems with interactions will be discussed in a forthcoming paper by the middle author, K. Ruohonen, to whom also belongs the idea of considering fragmentation in the way described above.

For all unexplained notions concerning formal languages we refer to [7]. We also expect a basic knowledge of L systems (at least in the extent of [7] but also parts of [4] or [6] might be consulted) on part of the reader.

The contents of this paper will now be briefly outlined. In section 2, we discuss the equivalence of various definitions of fragmentation. Section 3 studies basic properties of the families J0L and JT0L and, Section

4, the position of these families, as well as some of their extensions, in the L hierarchy. In Section 5, two infinite hierarchies of language families are obtained by imposing an upper bound on the number of cuts. The two hierarchies correspond to inside and outside control in regulating the number of cuts. The situation is completely analogous to the study of ultralinearity and finiteness of index of context-free grammars and languages, although there are no nonterminals present.

We are grateful to A. Lindenmayer for the remainder of this Introduction, discussing the biological viewpoint in more detail.

Developmental systems with fragmentation can be viewed biologically as follows. Any kind of reproduction process of any cellular organism clearly must involve the separation of certain individual cells from the rest of the organism (the production of gametes or spores), or it must involve the breaking up of the organism into smaller fragments. In either case, cellular or subcellular mechanisms must exist which determine where separation occurs between adjacent cells. Fragmentation may also occur in an organism when certain cells or organs have to be discarded for physiological reasons rather then for reproductive purposes. The latter case obtains in most epithelial tissues, where a continous sloughing off takes place of the surface cells. Also, the abscision of leaves and of floral parts at predetermined intervals is of this kind of fragmenting process.

In general, fragmentation can be induced in two ways: either by cell death or by differentiation of cells. The first case involves the (pre-programmed) death of some cells (or cell layers) which are attached to other cells. Simply by the disintegration of the dead cells the organism fragments into several parts. This kind of mechanism is well known in filamentous algae and fungi , as well as being responsible for leaf abscision in higher plants. This case corresponds to production rules of the form a → q in JL-systems. The second case represents a mechanism by which certain cells develop a change in their wall structure at certain places, which results then in a mechanical weakening of their attachment to adjacent cells. This is what happens in the course of production of gametes or spores. This case corresponds to production rules of the forms a → bq and a → qb.

2. DEFINITIONS

Consider an alphabet Σ, let $q \in \Sigma$ and assume that $\Sigma_1 = \Sigma - \{q\}$ is not empty. A word x_1 over Σ_1 is a q-guarded subword of a word x over Σ iff either $x_1 = x$ or else there are words y_1 and y_2 such that one of the following equations is satisfied:

$$x = y_1 q x_1 q y_2, \qquad x = x_1 q y_2, \qquad x = y_1 q x_1 .$$

We now proceed to two different but equivalent definitions of J0L languages. The first is a recursive one. Consider a 0L system $G = (\Sigma, w, P)$, where Σ is the alphabet, w is the axiom and P is the set of productions. For a letter q (not necessarily belonging to Σ), define recursively the following languages:

$$L^{\circ}(G, q) = \{ x \mid x \text{ is a q-guarded subword of } w\},$$

$$L^{i+1}(G, q) = \{ x \mid \text{for some } Z_1, Z_2, Z_1 \in L^{i}(G,q), Z_1 \Rightarrow_G Z_2, \text{ and } x \text{ is a q-guarded subword of } Z_2\}, \text{for } i \geq 0.$$

Define now the operator J_q as follows:

$$J_q(G) = \bigcup_{i=0}^{\infty} L^{i}(G, q).$$

A language L is a J0L language iff there exist a 0L system G and a letter q such that $L = J_q(G)$. The family of all J0L languages is denoted simply by J0L. (Because there is no danger of confusion, we use an analogous notation throughout this paper and, thus, speak of language families E0L and ET0L).

Note that if q does not belong to Σ, then $J_q(G) = L(G)$. Thus, by definition, the family 0L is contained in the family J0L.

Our first theorem can be viewed as a representation lemma which gives an alternative definition of the family J0L.

Theorem 1. Every J0L language equals the set of all q-guarded subwords of the words in L(G), for some 0L system G and letter q such that $q \to q$ is the only production for q in G. Conversely, if G is a 0L system having at most the production $q \to q$ for the letter q (thus, q need not belong to the alphabet of G), then the set of all q-guarded subwords of the words in L(G) belongs to J0L.

Proof. Assume that $L = J_q(G)$, for some 0L system G. Then also $L = J_q(G_1)$, where G_1 is obtained from G by replacing all (if any) productions for q with the production $q \to q$. From this observation the theorem easily follows.

Thus, J0L is the family of languages obtained as collections of q-guarded subwords from 0L languages, with the additional assumption that the identity production $q \to q$ is the only production for q in the 0L system in question. From the biological point of view, this requirement for q is very natural: for example, it would be unnatural to "glue together" strings already separated. From the formal language theory point of view, if such gluing is allowed then the resulting language family will strictly contain J0L. This is seen by considering the 0L system G with axiom aq and productions

$$a \to a^2, \quad q \to bq, \quad b \to b.$$

The set of all q-guarded subwords of L(G) equals

$$\{a^{2^n} b^n \mid n \geq 0\} \cup \{\lambda\}$$

which is easily seen not to belong to JOL. (In this example, q works at the other end only. Similar examples for q working in the middle, more resembling to actual gluing, are easily constructed).

Theorem 1 and the subsequent discussion suggest the following general definition of JL languages, applicable to any variety of L systems. We say that a language is a $(q \to q)$ $\mathbb{X}$L language iff it is generated by an $\mathbb{X}$L system in which the only production for the terminal letter q is the identity $q \to q$. (Here the variable $\mathbb{X}$ refers to any variety of L systems, e.g., $\mathbb{X}$ = PDT0. If we are dealing with systems with tables, the definition is read: the identity is the only production for q in each table. If we are dealing with systems with interactions, the identity is the only production for q in each context. A language belongs to the family J$\mathbb{X}$L iff it equals the set of all q-guarded subwords of some $(q \to q)$ $\mathbb{X}$L language.

As mentioned before, we study in this paper only systems without interactions. Thus, our objects of study will be the families JOL and JTOL, and some of their variations,

<u>Remark 1.</u> From the formal language theory point of view, the notion of an $(q \to q)$ $\mathbb{X}$L system may seen somewhat unnatural but, as we have already indicated, the special role of q is quite essential from the biological point of view. This reflects also the general situation: from the biological point of view the language with a generating mechanism is certainly more important than the language itself.

<u>Remark 2.</u> The letter S or the letter F, being initial letters in the words, "split" and "fragment", would have been appropriate for the names of our language families: S$\mathbb{X}$L or F$\mathbb{X}$L. However, both S and F already

have fixed meanings in the theory of L systems. Therefore, much to the delight of the last two authors, we have chosen to use the letter J which is the initial letter of the corresponding Finnish word "jakautua".

3. PROPERTIES OF J0L and JT0L

We use the expression J0L system for any pair (G, q) where G is a 0L system and q is a letter such that G contains at most the production $q \to q$ for q. Then the language of a J0L system (G, q) equals the set of q-guarded subwords of L(G). The notion of a JT0L system is similarly understood.

By definition, J0L contains the family F0L (F standing for a finite set of axioms) and, hence, the family of finite languages. Our next theorem gives a sufficient condition for a J0L or JT0L language to be finite.

<u>Theorem 2.</u> Assume that L is generated by a J0L system or a JT0L system (G, q) where the right side of every production either equals the empty word or contains an occurrence of q. Then L is finite.

<u>Proof.</u> L cannot contain words longer than twice the length of the longest q-guarded subword appearing either in the axiom of G or on the right side of some production of G.

The following theorem strengthens Theorem 1 and gives a normal form for J0L systems.

<u>Theorem 3.</u> Every J0L language is generated by a J0L system (G, q) such that the right side of each production contains at most one occurrence of q.

<u>Proof.</u> An arbitrary J0L system (G', q) can be transformed to an equivalent one satisfying the condition of the theorem by repeated applica-

tions of the following trick: a production $a \to xqyqz$, where x and z do not contain occurrences of q, is replaced by the production $a \to xqz$ and, at the same time, the axiom is catenated from the left by the word yq. (Here it is assumed that a occurs in some word in the language L(G').).

We now investigate closure properties of J0L and JT0L. Both of the families turn out to be anti-AFL's, i.e., they are closed under none of the AFL-operations. The next theorem is a lemma needed in subsequent proofs.

Theorem 4. Neither one of the languages L_1 and L_2 defined by $L_1 = \{a^{2n+1} \mid n \geq 1\} \cup \{a^2\}$, $L_2 = L_3^+$ where $L_3 = \{a^{2^n} b a^{2^n} \mid n \geq 0\}$ belongs to the family JT0L.

Proof. Consider first L_1, assuming that it is generated by a JT0L system. If some table contains a production $a \to a^i$ then necessarily $i = 1$ (because, otherwise, either $\lambda \in L_1$ or else $a^{2i} \in L_1$ with $i > 1$). On the other hand, if all right sides of the productions contain an occurrence of q, then L_1 is finite by Theorem 2. Hence, some table contains the production $a \to a$. If it is the only production for a in all tables, L_1 is again finite. Therefore, some table t contains both of the productions

$a \to a$ and $a \to a^i x a^j$,

where $x \in \{q\} \cup \{q\}\{a, q\}^*\{q\}$ and $i, j \geq 0$.

Since $a \notin L_1$, an application of t to a^2 shows that either $i = 2$ and j is odd, or else i is odd and $j = 2$. Consider the alternative $i = 2$, $j = 2k-1$, the treatment of the other alternative being similar. A contradiction now arises by applying t to the word a^3 because we obtain

$$a^{2k-1} a a^2 = a^{2k+2} \in L_1 .$$

Consider then the language L_2, and assume it is generated by a JT0L system. We now make a sequence of observations as follows.

(i) If some table t contains a production

$a \to x$ where $x \in \{a, b, q\}^+ - \{a, b\}^+$

then $x \in L_3^+(\{q\} \cup \{q\}(L_3 \cup \{q\})^+\{q\})L_3^+$. This is seen by applying t to the word aba. Hence, if it contains also a production

$a \to y$ where $y \in \{a, b\}^*$

then necessarily $y \in L_3^*$. This is seen by applying t to the word a^2ba^2.

(ii) If some table t contains the production $a \to a$ then t must be the identity $[a \to a,\ b \to b,\ q \to q]$. This follows by (i) if we consider a word $a^{2^n}ba^{2^n}$, where n is sufficiently large.

(iii) Some table must contain a production $a \to a^i$ with $i > 1$ because, otherwise, the number of consecutive a's in the words of L_2 would be bounded which is not the case. Denote by T_1 the set of tables containing such a production. As in (ii) we see that the only production for b in tables belonging to T_1 is $b \to b$ and, furthermore, that no table in T_1 can contain two productions $a \to a^i$ and $a \to a^j$ with $i \neq j$.

(iv) Assume that some of the tables in T_1 contains the production $a \to a^i$, $i > 1$, and also the production

$a \to a^j x a^k$ where $x \in \{b\} \cup \{b\}\{a,b\}^*\{b\}$ and $j, k \geq 0$.

An application of this table to the word $a^{2^n}ba^{2^n}$ gives the word $y_n = a^{i(2^n-1)+j} x a^k b a^{i \cdot 2^n}$. But all words y_n do not belong to L_3^+. (In fact, a necessary condition for this would be that $k > i \cdot 2^n$, for all n, which is absurd). Hence, none of the tables in T_1 contains such a production $a \to a^j x a^k$. By (i), we may now conclude that the tables in T_1 are all of the form $[a \to a^i,\ b \to b,\ q \to q]$.

Consider now the word $a^{2^n}ba^{2^n+1}ba \in L_3^+$, where n is sufficiently large. Our JT0L system cannot generate this word. The only possibility to obtain it would involve an application of a table in T_1 at the last step of the derivation but there is no suitable direct ancestor in L_3^+. This contradiction proves Theorem 4.

Theorem 5. The family JT0L is an anti-AFL.

Proof. In order to apply the same argument also for some other results, we give examples where AFL-operations applied to 0L languages yield languages outside JT0L.

The language L_1 of Theorem 4 is the union of two 0L languages. Also $L_1 = h(L_4)$ where

$L_4 = \{a^{2^{n+1}} \mid n \geq 1\} \cup \{b\}$

and the homomorphism h is defined by: $h(a) = a$, $h(b) = a^2$.

Clearly, L_4 is 0L.

The language L_1 can also be expressed in the two forms

$$L_1 = L_5 \cap \{a\}^* = h_1^{-1}(L_5),$$

where

$$L_5 = \{a^{2^{n+1}} b \mid n \geq 1\} \cup \{c\} \cup L_1$$

and the homomorphism h_1 is defined by:

$$h_1(a) = a,\ h_1(b) = b^2,\ h_1(c) = c^2.$$

L_5 is generated by the 0L system with axiom c and productions

$$a \to a,\ b \to a^2 b,\ b \to \lambda,\ c \to a^3 b,\ c \to a^2.$$

Finally, we note that L_3 in Theorem 4 belongs to the family 0L. Hence, Theorem 5 follows.

Since by definition $0L \subseteq J0L \subseteq JT0L$, we have also established the following result.

Theorem 6. The family J0L is an anti-AFL.

Since each of the families 0L, F0L, T0L and FT0L contains 0L and is contained in JT0L, we have also given another proof for the following well-known result.

Theorem 7. Each of the families 0L, F0L, T0L and FT0L is an anti-AFL.

4. THE POSITION OF J0L, JT0L AND SOME EXTENSIONS IN THE L HIERARCHY

Comparisons will be made between J0L and JT0L on one hand and especially the videly studied L families 0L, T0L, E0L and ET0L, on the other.

We defined above the general family J$\boxtimes$L. Thus, the meaning of JF0L, JTF0L, JE0L, JET0L should be clear. We also apply the operators E and C to the J$\boxtimes$L families in the ordinary sense. Thus, a language belongs to CJT0L iff it is a coding of a JT0L language, and a language belongs to EJ0L iff it is of the form $L \cap V_T^*$, for some J0L language L and alphabet V_T.(From the definitional point of view, EJ0L is more natural than JE0L). The next theorem is an immediate consequence of the definitions.

Theorem 8. JF0L = J0L and JTF0L = JT0L.

We now prove a lemma useful for the investigation of the J families. For a language L and a letter q, denote by $\bar{J}_q$ (L) the collection of all q-guarded subwords of the words in L. (Thus, if q does not belong to the alphabet of L then $\bar{J}_q$ (L) = L). A language family K is closed under the operator $\bar{J}$ iff, for any $L \in K$ and any letter q, $\bar{J}_q(L) \in K$.

Theorem 9. If a language family K is closed under gsm mappings then it is closed under the operator $\bar{J}$.

Proof. Assume that $L \in K$ and q is a letter. Consider the gsm M defined by the following table:

	s_0	s_1	s_2	s_3
a	s_1/λ s_2/a	s_1/λ	s_2/a	s_3/λ
q	s_0/λ s_3/λ	s_0/λ	s_3/λ	s_3/λ

where a ranges over all letters of the alphabet of L different from q, and s_0 is the initial and s_1 the only non-final state. It is easy to see that $M(L) = \bar{J}_q(L)$, which proves Theorem 9.

Theorem 9 is applicable, for instance, to any language family which is a cone (a full trio).

Theorem 10. EOL = JEOL = EJOL = CJOL = JCOL and

ETOL = JETOL = EJTOL = CJTOL = JCTOL.

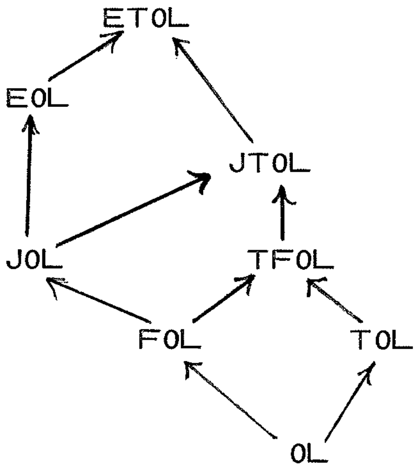
Proof. We prove the former equations, the proof for the latter being exactly the same. The inclusions

EOL $\subseteq$ JEOL, EOL $\subseteq$ EJOL, EOL $\subseteq$ CJOL

follow by the definitions and (the last one) by the equation EOL = COL. Each language in JEOL is obtained from a language in EOL by some operator $\bar{J}_q$. Since EOL is closed under gsm mappings (cf [4]), Theorem 9 implies that JEOL $\subseteq$ EOL. Hence, JOL $\subseteq$ EOL. The inclusions EJOL $\subseteq$ EOL and CJOL $\subseteq$ EOL now follow because EOL is closed under intersections with regular sets and under codings.

Theorem 10 remains valid with the same proof if the operator H (taking arbitrary homomorphisms) is considered instead of C.

Theorem 11. The following diagram holds true:



Here arrow denotes strict inclusion. Whenever the families are not connected by a directed path, they are incomparable.

Proof. It is well-known (cf [4] and [5]) that the lower part of the diagram involving OL, FOL, TOL, TFOL holds true and that EOL and TOL are incomparable and EOL $\subsetneq$ ETOL. The inclusions FOL $\subseteq$ JOL and TFOL $\subseteq$ JTOL follow by Theorem 8, and the inclusions JOL $\subseteq$ EOL and JTOL $\subseteq$ ETOL by Theorem 10. The language L_1 of Theorem 4 belongs to EOL - JTOL. From these facts the whole diagram follows, provided we still can present a language in the difference JOL - TFOL. Such a language is

$$L = \{aba^n \mid n \geq 1\} \cup \{a^n ba \mid n \geq 1\}.$$

Indeed, L is generated by the JOL system with the axiom aba and productions

$$a \to a, \quad b \to aba\ q\ aba.$$

(We note in passing that this system is also deterministic and, hence, L $\in$ JDOL. A further discussion of the family JDOL lies beyond the scope of this paper). On the other hand, $L \notin$ TF0L. (In a TF0L system for L, no production for a can contain b's on the right side, and each production for b contains exactly one b on the right side. By considering productions $a \to a^i$, $b \to a^j b a^k$ and words aba, a^2ba and aba^2, it is seen that $i = 1$, $j = k = 0$. Hence, every table would have to be the identity which is absurd).

Finally, we compare fragmentation with cell interactions. The most important result here is that the mechanism of fragmentation is even in its simplest form (J0L) capable of doing something which cannot be done with any interactions.

Theorem 12. There is a J0L language which is not an IL language. There is an IL language which is not in JT0L.

Proof. L_1 of Theorem 4 is an example for the second sentence. An example for the first sentence is provided by the language $L = K_1 \cup K_2 \cup K_3$, where

$$K_1 = \{a^n b a^{2n} c a^{2n} b a^n \mid n \ge 0\},$$

$$K_2 = \{a^{n+m} b a^{2n+m} \mid n \ge 1,\ m \ge 0\},$$

$$K_3 = mi\,(K_2).$$

Indeed, L is generated by the J0L system with the axiom bcb and productions

$$a \to a, \quad b \to aba, \quad c \to aca, \quad c \to aqa.$$

We prove by an indirect argument that L is not in IL. Assume the contrary: a (k, l) L system G with the production set δ generates L. We make a sequence of observations, finally arriving to a contradiction.

(i) Clearly

$$\delta\,(a^k,\ a,\ a^l) \in \{a\}^*$$

and

$$\delta(a^k, a, a^{l-1}b)\delta(a^k, a, a^{l-2}ba)\ldots\delta(ba^{k-1}, a, a^l)\in\{a,b\}^*.$$

(Here $\delta(a^k, a, a^l)$ denotes the right side of an arbitrary production for a in the environment (a^k, a^l). The same notation is used throughout this proof.) This implies that words in K_1 of sufficient length can be obtained directly only from words in K_1. Moreover, the productions used must be deterministic.

(ii) By (i),

$$\delta(a^k, a, a^l) = a^i$$

is the only production for (a^k, a, a^l). Furthermore, the language

$$K_4 = \{a^n ba^{2n} ca^{2n} ba^n \mid n \geq k+l\}$$

is generated (deterministically) by G from a finite number of starting words

$$a^{n_j} ba^{2n_j} ca^{2n_j} ba^{n_j} \qquad (j = 1, \ldots\ldots, N).$$

Assume first that $i > 1$. Then the starting word $a^{n_j} ba^{2n_j} ca^{2n_j} ba^{n_j}$ generates the language

$$K_{5j} = \{a^{h_n} ba^{2h_n} ca^{2h_n} ba^{h_n} \mid h_n = (n_j - x/(1-i))\, i^{n-n_j} + x/(1-i),\ n \geq n_j\},$$

where x is a constant not depending on j. But the union

$$\bigcup_{j=1}^{N} K_{5j}$$

is properly contained in K_4 because the number of words of the form

$$a^n ba^{2n} ca^{2n} ba^n, \qquad i^s \leq n < i^{s+1},$$

in the union is bounded by a constant not depending on s. Thus, we conclude that $i = 1$.

(iii) We consider now the language $K_8 = K_6 \cup K_7$, where

$$K_6 = \{a^{n+m} ba^{2n+m} \mid n \geq 1,\ m \geq 0,\ n+m \geq k+l\},$$

$$K_7 = mi\,(K_6).$$

We note first that, in each word of $K_2 \cup K_3$, the numbers n and m are uniquely determined by the exponents of a because the equations

$$n+m = n'+m'$$

$$2n+m = 2n'+m'$$

imply the equations $n = n'$ and $m = m'$. By (ii), every word in K_8 yields directly a word in K_8. We claim that also every word in K_6 (resp. K_7) yields directly only words in K_6 (resp. K_7). This follows because assuming that, for instance,

$$a^{n'+m'}ba^{2n'+m'} \Rightarrow a^{2n+m}ba^{n+m},$$

we obtain for sufficiently large p

$$a^{p-n'-n}ba^{p} \Rightarrow a^{p+n+m-2n'-m'}ba^{p+n+m-2n'-m'}$$

where the right side is not in K_8 although the left side

$$a^{p-n'-n}ba^{p} = a^{p-2(n'+n)+(n'+n)}ba^{p-2(n'+n)+2(n'+n)}$$

is in K_6.

Consider now an arbitrary relation

$$a^{n'+m'}ba^{2n'+m'} \Rightarrow a^{n+m}ba^{2n+m} .$$

If $n \neq n'$ then, for sufficiently large p, the word $a^{p+n-n'}ba^{p}$ (which is in K_6 if $n < n'$, and in K_7 if $n > n'$) yields directly a word not in K_8. Hence, $n = n'$. But this means that all words in K_6 are not generated by G.

5. HIERARCHIES DUE TO k-LIMITED FRAGMENTATION

The discussions in this section are restricted to JOL. A similar theory can be developed for JTOL.

Let k be nonnegative integer. We say that a language $L \in$ JOL is obtained by k-limited fragmentation under inside control, in symbols, $L \in IC(k)$ iff L is generated by a JOL system (G, q) such that no word in L(G) contains more than k occurrences of q. If $L \in$ JOL but $L \notin IC(k)$, for every k = 0, 1, 2, ..., we say that $L \in IC(\infty)$.

For instance, all OL languages belong to IC(0). Any FOL language generated by a system with two axioms belongs to IC(1). We say that a language $L \in$ JOL is obtained by k-limited fragmentation under outside control, in symbols, $L \in OC(k)$ iff L is generated by a JOL system (G, q) and, furthermore, every word in L is a q-guarded subword of such a word in L(G) which does not contain more than k occurrences of q. If $L \in$ JOL but $L \notin OC(k)$, for every k = 0, 1, 2, ..., we say that $L \in OC(\infty)$.

Note the analogy in context-free languages: $OC(\infty)$ corresponds to languages of infinite index, and $IC(\infty)$ to languages which are not ultralinear.

Theorem 13. For all $k \geq 0$,
$IC(k) \subsetneq IC(k+1)$, $OC(k) \subsetneq OC(k+1)$, $IC(k+1) \subsetneq OC(k+1)$. Furthermore, $IC(0) = OC(0)$ and there is a language in OC(1) belonging to $IC(\infty)$.

Proof. We prove first the second sentence. The equation $IC(0)=OC(0)$ is obvious by the definitions. Consider the JOL system with axiom abc and productions

$$a \to abc, \quad b \to bc, \quad b \to q, \quad c \to c.$$

The generated language is

$L = \{abc\,bc^2 bc^3 \ldots bc^i \mid i \geq 1\} \cup \{c^i \mid i \geq 1\} \cup \{c^i bc^{i+2} bc^{i+3} b \ldots bc^{i+j} \mid i \geq 1, j \geq 2\}$.

It is easy to see that every word in L not belonging to the first member L_1 of the union is obtained from a word in L_1 by making one cut at the end. Hence, $L \in OC(1)$. It is also easy to see that $L \notin IC(k)$, for all $k = 0, 1, 2, \ldots$. This follows because L is not generated by a J0L system where no production for b or c contains q on the right side and, on the other hand, no J0L system for L where q occurs on the right side of some production for b or c satisfies the requirements of k-limited fragmentation under inside control.

The inclusions in the first sentence, apart from being proper, follow by the definitions. It is now a consequence of the second sentence that the last inclusion is proper. Finally, the strictness of the first two inclusions follows because

$\{a_1^{2^n} \mid n \geq 1\} \cup \{a_2^{3^n} \mid n \geq 1\} \cup \ldots \cup \{a_{k+2}^{p_{k+2}^n} \mid n \geq 1\} \in IC(k+1) - OC(k)$,

where p_i is the ith prime.

Finally, we exhibit a language in the class $OC(\infty)$, i.e., a J0L language which can be viewed to have an infinite index.

Theorem 14. The language

$$L = \{b\} \cup \{ba^{2^n - 1} \mid n \geq 1\} (\{\lambda\} \cup \{b\})$$

is in the class $OC(\infty)$.

Proof. L is generated by the J0L system with the axiom bab and productions

$$a \to a^2, \quad b \to bqba.$$

We now claim that L does not belong to any of the classes OC(k), k= 0, 1, 2, ... To prove this, we consider an arbitrary J0L system G generating L. We again make a sequence of observations as follows.

(i) G does not contain any production $a \to x$, where b occurs in x but q does not occur in x. (Otherwise, we would obtain words with too many b's).

(ii) It is not possible that

$$a \to \lambda, \quad a \to a, \quad a \to x_1, \ldots, a \to x_n,$$

where each x_i contains an occurrence of q, are the only productions for a. (Otherwise, we could show in the same way as in the proof for Theorem 2 that G generates a finite language).

(iii) G contains a production $a \to a^i$, $i > 1$. (This is a direct consequence of (i) and (ii).)

(iv) There is no production for a such that b occurs on the right side. (By (i), such a production P would also contain an occurrence of q on the right side. We could now consider the word ba^{2^n-1}, for some large n, and apply the production $a \to a^i$ from (iii) to the first m a's and P to the (m+1)st a, for variable m. We would then obtain words with mi+j occurrences of a, for some constant j. But clearly some numbers mi + j are not of the form $2^n - 1$).

(v) Sufficiently long words of the form $ba^{2^n-1}b$ are generated directly only by words of the same form. (This is an immediate consequence of (iv)).

(vi) The production $b \to b$ is not in G. (Otherwise, we would get a contradiction by applying this production and the production $a \to a^i$ to words of the form $ba^{2^n-1}b$.)

(vii) The right side of every production for b must begin with bq. (Consider an arbitrary production P for b. The right side cannot begin with q because λ is not in the language. It cannot begin with a because there are no words beginning with a in L. Thus, by (vi), it must begin with bq, ba, or bb. The last two alternatives are ruled out because, otherwise, an application of P to the final b in the word $ba^{2^n-1}b$ and $a \to a^i$ to the occurrence of a immediately preceding it would give a word not in L.)

Our claim is now an immediate consequence of (v) and (vii).

REFERENCES.

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

[2] A. Ehrenfeucht and G. Rozenberg, Nonterminals versus homomorphisms in defining languages for some classes of rewriting systems, contained in Springer Lecture Notes for Computer Science, vol. 14.

[3] A. Ehrenfeucht, K.P. Lee and G. Rozenberg, Subword complexities of various classes of deterministic developmental languages without interactions, submitted for publication.

[4] G. Herman and G. Rozenberg, Developmental Systems and Languages, North-Holland Publ. Co. 1974.

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

[6] G. Rozenberg and A. Salomaa (ed.), L Systems, Springer Lecture Notes for Computer Science, Vol. 15, 1974.

[7] A. Salomaa, Formal Languages, Academic Press 1973.