all correspondence to the second author

A NOTE ON NUMERICALLY DISPERSED EOL LANGUAGES

by

A. Ehrenfeucht

and

G. Rozenberg †

#CU-CS-097-76 September, 1976

† Department of Mathematics
University of Antwerp, U.I.A.
Wilrijk, Belgium

working report

ABSTRACT

A result is provided which allows one to prove that certain laguages are not EOL languages.

INTRODUCTION

One of the obviously useful and still very much needed directions of research in L systems theory (see, e.g., [4] and [7]) is a search for results which would allow one to prove that certain languages are not in certain families of languages.

This note presents a work in this direction. It continues our work from [1] and [2]. In [1] we have considered EOL languages for which the length sets (after maybe erasing all letters from a certain subalphabet) were numerically dispersed (for each k from some moment on the difference between any two consecutive elements in the length set is larger than k). We have shown that in such languages letters must be distributed in the words in the special way. This result allows one to prove that certain languages are not EOL languages. However this result does not yield applications for, for example, EOL languages over a one letter alphabet. In this note we try to cover also these languages. To this aim we use results from [2]. We notice that if a language is numerically dispersed than it is determined in the sense of [2]. Then using the decomposition result from [2] we show that if K is an EOL language which is numerically dispersed than the "density" of its length set is either rather small (logarithmically bounded) or rather large (of an exponential type). In this way we get for example a short and rather elegant proof that $K=\{a^{2^{n}3^{m}}: n,m\geq 0\}$ is not an EOL language. (The problem whether K is an EOL language was posed by Salomaa in [9] and solved by J. Karhumaki in [5].)

PRELIMINARIES

We assume the reader to be acquainted with the basics of L systems theory (see, e.g., [4] and [7]) in particular with EOL and DOL systems.

There are two specific notions concerning languages that are central to this note, and so we will recall them now. (In what follows *Length* K denotes the length set of the language K.)

Definition 1. Let K be a language. It is called numerically dispersed if for every positive integer k there exists a positive integer n_k such that if u, v are elements of *Length* K with $u>v>n_k$ then $(u-v)>k$.

Definition 2. Let K be a language over an alphabet Σ and let Θ be nonempty subset of Σ. We say that K is Θ-determined if

$$(\forall k)_{N+}(\exists n_k)_{N+}(\forall x,y)_K$$

$$[\text{if } |x|,|y|>n_k, x=x_1ux_2, y=x_1vx_2 \text{ and } |u|,|v|<k \text{ then } h_\Theta(u)=h_\Theta(v)],$$

where h_Θ in a homomorphism erasing all elements from Σ/Θ and acting as identity on elements of Θ.

If K is Σ determined then we call it determined.

Numerically dispersed EOL languages were investigated in [1] and Θ-determined EOL languages were investigated in [2].

The following result which was proved in [2] will be very useful for us in the sequel.

Theorem 1. If K is a determined EOL language then there exists a finite set of PDOL languages $K_1,\ldots,K_f$ and a Λ-free homomorphism ψ such that $K=\bigcup_{i=0}^{f}\psi(K_i)$.

We will use also the following notation. For a language K and a positive integer q, $\mathcal{less}_q$ K={n:nε*Length* K and n<q}.

RESULTS

In this section we will prove our main result and provide its application.

We start with the following simple observation.

Lemma 1. Let K be a language over a one letter alphabet. If K is numerically dispersed then it is determined.

Proof.

Let $K=\{a^n: n\in Z\}$. Since K is numerically dispersed, for every positive integer q there exists a n_q such that for any two elements u_1,u_2 from Z if $u_2>u_1>n_q$ then $(u_2-u_1)>q$. But then
if $x,y\in K, x=a^{r_1}a^t a^{r_2}, y=a^{r_1}a^v a^{r_2}$ with $t<v<q$ and $r_1+t+r_2>n_q$,
then $(r_1+v+r_2)-(r_1+t+r_2)=v-t<q$,
implying that v=t. Consequently K is determined.

Lemma 2. If $\overline{K}$ is a PDOL language, ϕ is a Λ-free homomorphism and $K=\phi(\overline{K})$, then

either $(\exists a)_{R_{pos}}(\forall q)_{N^+}\ less_q(K)<a\log_2 q$,

or $(\exists a)_{R_{pos}}(\exists b)_{R_{pos}}(\forall q)_{N^+}\ less_q(K)\geq aq^b$.

Proof.

Let $G=<\Sigma,\delta,\omega>$ be a PDOL system such that $L(G)=\overline{K}$. Let $E(G)=\omega_0,\omega_1,\ldots$.

We will consider three (exclusive) possible cases.

(i) $\overline{K}$ is finite.

Then the first of the above two conditions trivially holds.

(ii) $\overline{K}$ is infinite but not polynomially bounded. Then by a result of Salomaa (see [8]), we know that there exists a letter x in Σ reachable from ω and that $\delta^m(x)=\alpha_1 x\alpha_2 x\alpha_3$ for some $m\geq 1$ and $\alpha_1,\alpha_2,\alpha_3$ in Σ^*.

Consequently there exists a positive real number d such that

$(\forall n)_N \quad |\phi(\omega_n)| \geq |\omega_n| > d^n$.

Now let q be a positive integer and let m be such that $|\phi(\omega_m)| < q$. Then $d^m < q$, hence $n \cdot \log_2 d < \log q$ and so $n < \frac{1}{\log_2 d} \cdot \log_2 q$. Consequently

$\#\{m: |\phi(\omega_m)| < q\} < \frac{1}{\log_2 d} \cdot \log_2 q$,

and so the first condition from the statement of the lemma holds.

(iii) $\overline{K}$ is infinite and polynomially bounded.

First of all by a result of Nielsen [6] we know that we can decompose G with a "step" c such that

$(\forall n)_N |\phi(\omega_{cn})| > |\phi(\omega_{c(n-1)})|$.

Now let c and d be such positive integers that the polynomial cn^d limits (from the above) the growth function of G.

Then, for a given q, if m is a positive integer such that $q > cm^d$, then $less_q(K) > \frac{m}{c}$.

Thus $less_q(K) > \frac{1}{c} \cdot Entier((\frac{q}{d})^{\frac{1}{a}})$, and so $less_q(K) > \frac{1}{2.c.b^{\frac{1}{a}}} q^{\frac{1}{a}}$, and consequently the second condition from the statement of the lemma holds.

From (i), (ii) and (iii) the lemma follows.

Here comes the main result of this section. Roughly speaking it says that for a numerically dispersed EOL language K the length set is of such a nature that either it is rather "thin" (bounded by a logarithm) or rather "dense" (of an exponential nature) but nothing "in-between".

Theorem 2. Let K be a numerically dispersed EOL language. Then either $(\exists a)_{R_{pos}} (\forall q)_{N+} \mathit{less}_q(K)<a \log q$,

or $(\exists a)_{R_{pos}} (\exists b)_{R_{pos}} (\forall q)_{N+} \mathit{less}_q(K)>a\, q^b$.

Proof.

Let $K \subseteq \Sigma^*$ and let ϕ be a coding that maps every element of Σ into one fixed letter, say a. That $\phi(K)$ is a numerically dispersed language over a one letter alphbet. Hence by Lemma 1, $\phi(K)$ is determined. But the class of EOL languages is closed w.r.t. codings (see, e.g., [4]) and so $\phi(K)$ is a determined EOL language. Thus by Theorem 1 there exists a Λ-free homomorphism ψ and a finite set of PDOL languages $K_1,\ldots,K_f$ such that $K=\bigcup_{i-1}^{f} K_i$.

Now the Theorem follows from Lemma 2.

Next we will show an application of Theorem 2, consisting of a rather elegant proof of a result (due to J. Karhumaki, see [5]) that $\{a^{2^n \cdot 3^m}: n,m \geq 0\}$ is not an EOL language.

Corollary. $K=\{a^{2^n \cdot 3^m}: n,m \geq 0\}$ is not an EOL language.

Proof.

(i) It was proved by Gelfond (see [3] p. 24) that for any fixed integer p there are only finitely many pairs <n,m> of nonnegative integers such that $2^n-3^m=p$. Hence K is numerically dispersed.

(ii) If q is a fixed positive integer and $2^n \cdot 3^m < q$ then both n and m are smaller than log q. Hence $\mathit{less}_q K<(\log_2 q)^2$.

(iii) Note that the number of n's such that $2^n < \sqrt{q}$ is not smaller than Entier($\log\sqrt{q}$) hence not smaller that $\frac{1}{2}\cdot\log_2 q - 1$. In the same way the number of m's such that $3^m < \sqrt{q}$ is not smaller than $\frac{1}{3}\cdot\log_3 q - 1$. Consequently (as $q = \sqrt{q}\cdot\sqrt{q}$), $less_q K \geq (\frac{1}{2}\log_2 q - 1)(\frac{1}{3}\log_2 q - 1)$.

(iv) From (ii) and (iii) it follows that

$$(\tfrac{1}{2}\log_2 q - 1)(\tfrac{1}{3}\log_3 q - 1) \leq less_q K < (\log_2 q)^2$$

and so from (i) and Theorem 2 it follows that K is not an EOL language.

ACKNOWLEDGEMENT

The authors greatfully acknowledge the financial support of NSF Grant DCR75-09972 and the financial support of the Belgian Foundation for Scientific Research (NFWO). We are indebted to L. Fosdick for providing us with an opportunity ot get together and work on this and other papers.

REFERENCES

[1] Ehrenfeucht, A. and Rozenberg, G., 1975. The number of occurrences of letters versus their distribution in some EOL languages, Information and Control, 26: 256-271.

[2] Ehrenfeucht, A. and Rozenberg, G., 1975. On Θ-determined EOL languages, In: A. Lindenmayer and G. Rozenberg (eds.), Automata, Languages, Development, North Holland, Amsterdam, 191-202.

[3] Gelfond, A., 1940. Sur la divisibilite´ de la difference des puissances de leux nombres entiers par une puissance d'un ideal premer, Math. Sbornik, 7: 7-24.

[4] Herman, G. T. and Rozenberg, G., 1975. Developmental Systems and Languages, North Holland Publishing Company, Amsterdam.

[5] Karhumaki, I., 1976. On length sets of informationless L systems, In: A. Lindenmayer and G. Rozenberg (eds.), Automata, Languages, Development, North Holland, Amsterdam, 227-242.

[6] Nielsen, M., 1976. On the decidability of sequence equivalence problems for DOL systems, Information and Control, 25: 166-193.

[7] Rozenberg, G. and Salomaa, A., 1974. The mathematical theory of L systems, In: J. Tou (ed.), Advances in Information Systems Science, Plenum Press, New York, 161-206.

[8] Salomaa, A., 1973. On exponential growth in Lindenmayer systems, Indogatioues Mathematicae, 35: 23-30.

[9] Salomaa, A., 1976. Iteration grammars and Lindenmayer AFL's, Lecture Notes in Computer Science, Springer Verlag, Heidelberg, 14: 523-533.