ON ARITHMETIC SUBSTITUTIONS

OF EDTOL LANGUAGES

by

A. Ehrenfeucht

and

G. Rozenberg †

#CU-CS-096-76   September, 1976

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

working report

ABSTRACT

A family of languages is called arithmetic if each infinite language in it is such that its length set contains an arithmetic progression. It is proved that there exists an ETOL language which is not an arithmetic substitution of any EDTOL language. This result sheds some light on the question: How much more "complicated" are ETOL languages than EDTOL languages?

INTRODUCTION

One of the basic research topics in the theory of L systems is the role of the "deterministic restriction" (see, e.g., [4] or [5]). In other words one investigates the behavior of deterministic L systems versus the behavior of L systems without the deterministic restriction. This restriction is of grammatical nature: we call an L system deterministic if for each letter there exists precisely one production for each possible situation (table, context). The "ideal" result explaining the role of a deterministic restriction would be of the following kind: there exists a language operator ϕ such that every language in the L family X can be obtained as the result of applying ϕ to a language in the deterministic subfamily of X. Only partial results in this direction are known(see, e.g., [3]).

We feel that for some basic L families such a translation of a deterministic restriction on a grammar into a "natural" language operator is impossible. This paper is a step towards such a negative result for the family of ETOL languages. As a natural operator we have chosen a substitution into a family of languages L which has the property that the length set of each infinite language in L contains an infinite arithmetic progression. Then we demonstrate that there exists an ETOL language which *cannot* be obtained as a result of such a substitution from an EDTOL language. Thus the substitutions into regular languages, context free languages or even ETOL languages of finite index (see [6]) cannot make up for the loss of the language generating power resulting from imposing the deterministic restriction on ETOL systems.

PRELIMINARIES

We assume the reader to be familiar with the basics of the theory of ETOL systems (e.g. in the scope of [4]).

We use N, N^+, R and R^+ to denote the sets of nonnegative integers, positive integers, nonnegative reals and positive reals respectively. For a set Z, $\#Z$ denotes the cardinality of Z.

For a word x, $|x|$ denotes its length and *min* x denotes the set of letters occurring in x. For a language K, *Min* $K=\{a:(\exists x)_K[a \varepsilon min\ x]\}$ and *Length* $K=\{n \varepsilon N:(\exists x)_K[|x|=n]\}$.

If ϕ is a substitution (of type X) on Σ and K is a language over Σ such that for every element a of *Min* K, $\phi(a)$ is a language of type Y then we say that ϕ *is of type* Y *on* K. For the family of languages Y we use $\Phi_X(Y)$ to denote the family of all substitutions of type X of languages in Y.

We use $L(\text{ETOL})$ and $L(\text{EDTOL})$ to denote the families of ETOL languages and EDTOL languages respectively.

Now we will introduce several notions specific to this paper.

<u>Definition 1</u>. An infinite language K is called *arithmetic* if *Length* K contains an infinite arithmetic progression; otherwise K is called *antiarithmetic*. A family L of languages is called *arithmetic* if every infinite language in L is arithmetic; otherwise L is called *antiarithmetic*.

For example it is very well known (see, e.g., [7]) that both the family of context free languages and the family of regular languages are arithmetic.

<u>Definition 2</u>. Let ϕ be a substitution, $\phi:\Sigma \longrightarrow L$. If L is arithmetic than ϕ is called *arithmetic*.

The following basic property of arithmetic substitutions will be used in the sequel.

Lemma 1. Let L be an antiarithmetic language. If $L=\phi(\overline{L})$ where ϕ is an arithmetic substitution than ϕ is finite on $\overline{L}$.

Proof.

Let $L=\phi(\overline{L})$.

Let us assume, to the contrary, that $\overline{L}$ contains a word $w=w_1aw_2$ such that for the letter $a,\phi(a)$ is infinite. Since ϕ is arithmetic, $\phi(a)$ contains an infinite sequence of words $z_1,z_2,\ldots$ such that $|z_1|,|z_2|,\ldots$ form an arithmetic progression. But if $\overline{w}_1$ and $\overline{w}_2$ are fixed elements of $\phi(w_1)$ and $\phi(w_2)$ respectively then $\overline{w}_1z_1\overline{w}_2,\overline{w}_1z_2\overline{w}_2,\ldots$ is an infinite sequence of words in L such that $|\overline{w}_1z_1\overline{w}_2|,|\overline{w}_1z_2\overline{w}_2|,\ldots$ form an arithmetic progression. This however contradicts the fact that L is antiarithmetic.

Definition 3. A function f from R^+ into R^+ is called *slow* if $(\forall\alpha)_{R^+}(\exists n_\alpha)_{R^+}(\forall x)_{R^+}[\textit{if } x>n_\alpha \textit{ then } f(x)<x^\alpha]$.

Thus a constant function, $(\log x)^k$ and $(\log x)^{\log\log x}$ are examples of slow functions, whereas $\log x^{\log x}$, x^2, $\sqrt{x}$ are examples of functions that are not slow.

Definition 4. Let Σ be a finite alphabet and let f be a function from R^+ into R^+. A word w over Σ is called *f-random (over Σ)* if every two disjoint subwords of w which are longer than $f(|w|)$ are different.

The following result will be useful in the sequel. At the same time it nicely illustrates the notion of f-randomness.

<u>Lemma 2</u>. Let $f:R^+ \longrightarrow R^+$, where $f(x)=6 \cdot \log x$. Let $\Sigma=\{0,1\}$. Then $(\forall n)_{N^+}(\exists z)_{\Sigma^+}[|z|=2^n$ and z is f-random].

<u>Proof</u>.

Let $V=\{0,1,\$\}$. Let $n\varepsilon N^+$ and let $y_1,\ldots,y_{2^n}$ be an arbitrary, but fixed, ordering of all words over Σ of length n. Let $\alpha_n=y_1\$y_2\$\ldots\$y_{2^n}\$$. Clearly no two disjoint subwords of α_n that are of length at least 2n are identical. Let ψ be the homomorphism from V into Σ^* defined by $\psi(0)=0^3,\psi(1)=1^3$ and $\psi(\$)=101$. Let $\beta_n=\psi(\alpha_n)$. Clearly no two disjoint subwords of β_n that are of length at least 6n are identical.

Finally let z be the prefix of β_n of length 2^n. Obviously z do not contain two disjoint subwords of length at least 6n that are identical and so z is f-random.

As a direct corollary of the pumping theorem for EDTOL languages (see [1] or [2]) we get the following result.

<u>Theorem 1</u>. If f is a slow function and an EDTOL language K contains infinitely many f-random words then K is arithmetic.

RESULT

In this section we prove our main result.

<u>Theorem 2</u>. There exists an ETOL language K such that there do not exist an arithmetic substitution ϕ and an EDTOL language M with the property that $K=\phi(M)$.

<u>Proof</u>.

Let $G=<V,Q,S,\Sigma>$ be an ETOL system with

$V=\{S,F,0,1\}$,

$\Sigma=\{0,1\}$,

$Q=\{P_1,P_2,P_3\}$,

$P_1=\{S\longrightarrow S^2,F\longrightarrow F,0\longrightarrow F,1\longrightarrow F\}$,

$P_2=\{S\longrightarrow 0,S\longrightarrow 1,F\longrightarrow F,0\longrightarrow F,1\longrightarrow F\}$, and

$P_3=\{S\longrightarrow F,F\longrightarrow F,0\longrightarrow 0^3,1\longrightarrow 1^3\}$.

Let $K=L(G)$. We will prove now that for no EDTOL language M, K is an arithmetic substitution of M.

1) K is antiarithmetic.

<u>Proof</u>.

Clearly $Length$ $K=\{2^n\cdot 3^m:n,m\geq 0\}$. Let for $q\geq 1$, $less_q K=\{k\varepsilon Length\ K:k<q\}$. But if $2^n\cdot 3^m\leqslant q$ then $n<\log_2 q$ and $m<\log_3 q$. Consequently

$$\lim_{q\longrightarrow\infty}\frac{less_q K}{q}\leq\lim_{q\longrightarrow\infty}\frac{(\log_2 q)\cdot(\log_3 q)}{q}=0,$$

and so $Length$ K do not contain an arithmetic progression.

2) $(\forall m)_{N^+}(\exists n_m)_{N^+}(\forall a,b)_{N^+}(\forall w)_K(\forall\alpha,\beta)_{\Sigma^+}$

$[if\ a,b>n_m,|\alpha|<m,|\beta|<m,|w|=2^a\cdot 3^b\ and\ w=w_1\alpha w_2$

$then\ \overline{w}=w_1\beta w_2\varepsilon K\ implies\ \alpha=\beta]$.

<u>Proof</u>.

Let us take $n_m=m$.

(i)   First we prove that if $\bar{w}=w_1\beta w_2 \in K$ then $|\alpha|=|\beta|$.

If $\bar{w} \in K$ then there exist c and d such that $|w|=2^c \cdot 3^d$. Let us assume that $|\bar{w}|>|w|$ and let us consider $s=|\bar{w}|-|w|=2^c \cdot 3^d=2^a \cdot 3^b$. Clearly either $c \geq a$ or $d \geq b$ and so s is either divisible by 2^a or by 3^b. But $2^a>m$ and $3^b>m$ while obviously $|\bar{w}|-|w|<m$; a contradiction. Similarly if we assume that $|w|>|\bar{w}|$ we get a contradiction. Thus it must be that if $\bar{w} \in K$ then $|w|=|\bar{w}|$ and so $|\alpha|=|\beta|$.

(ii)   Now we will prove that if $|\alpha|=|\beta|$ and $\bar{w} \in K$ then $\alpha=\beta$.

Note that $w=w_1 \ldots w_{2^a}$ where each w_i is a word of length 3^b such that it either consists of 1's only or it consists of 0's only. Thus if we replace any subword α of w by a word β of the same length and obtain in this way a word in K then it must be $\alpha=\beta$.

3) K is not the result of an arithmetic substitution on an EDTOL language.

Proof.

Let us assume, to the contrary, that there exists an EDTOL language M and an arithmetic substitution ϕ such that $K=\phi(M)$.

By Lemma 1 we can assume that ϕ is a finite substitution. Let r be the maximal length of a word that ϕ can substitute for a single letter. Let $a>r, b>r$ and let us consider a word w in K such that $|w|=2^a \cdot 3^b$. Let z be a word in M such that $\phi(z)=w$. Then from 2) it immediately follows that for every x in $min(z), \phi(x)$ is a singleton.

Let us denote by $Singl_\phi M$ the set of all letters x from $MinM$ such that $\phi(x)$ is a singleton. From the above discussion we know that $Singl_\phi M \neq \emptyset$. Let $Z=M \cap (Singl_\phi M)^*$. Since the intersection of an EDTOL with a regular language is an EDTOL language and since a homomorphic

image of an EDTOL language is an EDTOL language, $\phi(Z)$ is an EDTOL

language. Also from the previous paragraph we know that

$K_r = \{w \epsilon K : |w| = 2^a \cdot 3^b, a > r \text{ and } b > r\} \subseteq \phi(Z)$.

Now let us consider the function f on positive integers such

that $f(x) = 6 \ 3^{r+1} \cdot \log_2 x$. Clearly f is a slow function. On the other

hand K_r contains infinitely many f-random words, which is seen as

follows. Let a>r and let us generate in G, using (a-1) times table P_1

and then the table P_2, a word x_a of length 2^a. Then, using (r+1) times

table P_3, let us substitute $0^{3^{r+1}}$ for each 0 in x and $1^{3^{r+1}}$ for each 1

in x obtaining in this way the word $y_{a,r+1}$. Now $y_{a,r+1}$ is f-random if

ond only if x_a is f-random. However by Lemma 2, for every a≥1 there

exists an f-random word over {0,1} of the length 2^a. Since every word

over {0,1} of length 2^a can be generated in G (in the way that x_a was

generated) K_r contains infinitely many f-random words.

Then however by Theorem 1 we get that $\phi(Z)$ (remember that

$K_r \subseteq \phi(Z)$) is arithmetic. Since $Z \subseteq M$, $\phi(Z) \subseteq \phi(M) \subseteq K$ and this implies

that K is arithmetic. This contradicts 1) and consequently 3) holds.

Corollary 1. Let X stand for either context free or regular or

ETOL of finite index. Then it is not true that $L(ETOL) = \Phi_X(EDTOL)$.

Proof.

It is well-known (see, e.g., [7]) that the families of regular

and context free languages are arithmetic. It is proved in [6] that

the family of ETOL languages of finite index is arithmetic.

ACKNOWLEDGEMENTS

REFERENCES

[1] Ehrenfeucht, A. and Rozenberg, G., 1975. On the (combinatorial) structure of L languages without interactions, Proceedings of the VIIth Annual ACM Symposium on Theory of Computing, Albuquerque, N. M., 137-144.

[2] Ehrenfeucht, A. and Rozenberg, G., 1975. A pumping theorem for deterministic ETOL languages, Revue Francaise d'Automatique, Informatique et Recherche Operationnele, R-2, 9: 13-23.

[3] 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.

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

[5] Rozenberg, G. and Salomaa, A., 1976. The mathematical theory of L systems, In: J. Tou (ed.), Advances in Information Systems Science, Plenum Press, New York, 161-206.

[6] Rozenberg, G. and Vermeir, D., 1975. On ETOL systems of finite index, Department of Mathematics, University of Antwerp, U.I.A., Technical Report, No. 75-13.

[7] Salomaa, A., 1973. Formal languages, Academic Press, London-New York.