\nalele ştiinţifice ale Universităţii „Al. I. Cuza" din Iaşi
'omul XXIII, S. I a, f. 2, 1977

REMARKS ON UNIFORM TAG SEQUENCES

BY

ADRIAN ŞOIL

1. In his paper [1] A. Cobham studies some properties of uniform ag sequences in connection with operations on such sequences and distri- ution of symbols.

By considering a new operation on the set of uniform tag sequences ver a finite alphabet, we reduce the study of distribution of blocks to the tudy of distribution of symbols in these sequences.

2. In this section we define a new operation for a class of tag sequen- es, the internal sequences of tag systems in the limited sense.

2.1. Let $\mathcal{T} = (B, b_0, w)$ be a tag system in the limited sense (TSL), i.e. B is a finite set, $b_0 \in B$ is the initial symbol of $\mathcal{T}$, $w : B \to B^*$ is a func- ion satisfying the constraints : b_0 begins $w(b_0)$, $|w(b)| \geqslant 1$, $\forall b \in B$, $b \neq b_0$, and $|w(b_0)| \geqslant 2$. Let $y = \operatorname{intseq}(\mathcal{T})$ be the internal sequence of $\mathcal{T}$, i.e. $y = \lim\limits_{n \to \infty} w^n(b_0)$. In the sequel the block $b_{i_1} b_{i_2} \dots b_{i_k} \in B^*$, with $b_{i_j} \in B$, $j = \overline{1, k}$, will be denoted by $b_1 b_2 \dots b_k$.

2.2. For every fixed $k \in N$, $k \geqslant 2$ we define the one-to-one function $h_k : B^k \to C_k$, by setting $b_1 b_2 \dots b_k \mapsto h_k(b_1 b_2 \dots b_k)$. Hence C_k will be the inite set of symbols $h_k(B^k)$. The function h_k has a natural extension to locks and sequences over B, which we also denote by h_k, defined by

$$b_1 b_2 \dots b_k b_{k+1} \dots b_{k+q} \mapsto h_k(b_1 b_2 \dots b_k)\, h_k(b_2 \dots b_k b_{k+1}) \dots h_k(b_{1+q} \dots b_{k+q})$$

$$b_1 b_2 \dots b_k b_{k+1} \dots \mapsto h_k(b_1 b_2 \dots b_k)\, h_k(b_2 \dots b_k b_{k+1}) \dots .$$

2.3. Proposition. *If* $y = \operatorname{intseq}(\mathcal{T})$, *then* $h_k(y)$ *is the internal sequence f a TSL* $\mathcal{T}_k$.

Proof. We define $\mathcal{T}_k$ by $\mathcal{T}_k = (C_k, \overline{b}_k, w_k)$ with $\overline{b}_k = h_k(\Pr_k(y))$, w_k : $C_k \to C_k^*$ by $h_k(b_1 b_2 \dots b_k) \to h_k(w(b_1) \Pr_{k-1}(w(b_2) \dots w(b_k)))$ (where $\Pr_i(p)$ de- notes the prefix of length i of p). The mapping w_k is well defined since $|w(b_1) \Pr_{k-1}(w(b_2) \dots w(b_k))| \geqslant k$. Let $y_k = \lim\limits_{n \to \infty} w_k^n(\overline{b}_k)$. To prove that $h_k(y) = y_k$ we need some lemmas.

2.4. Lemma. $|w_k(h_k(b_1 \dots b_k))| = |w(b_1)|$.

The lemma is immediate : for $p \in B^*$ with $|p| \geqslant k$, $|h_k(p)| = |p| - k + 1$, then

$$|w_k(h_k(b_1\dots b_k))| = |h_k(w(b_1)\,\mathrm{Pr}_{k-1}(w(b_2)\dots w(b_k)))| = |w(b_1)| +$$
$$+ |\mathrm{Pr}_{k-1}(w(b_2)\dots w(b_k))| - k + 1 = |w(b_1)|.$$

2.5. Lemma. *$w_k^n(\overline{b}^k) = h_k(w^n(b_0)\,\mathrm{Pr}_{k-1}(y - w^n(b_0)))$, where $y - w^n(b_0)$ is the sequence y after deleting his initial block $w^n(b_0)$, $n \in N^*$.*

The lemma will be proved by induction on n. For $n = 0$ the egality is true since $w_k^0 = 1_{C_k}$, $w^0 = 1_B$. Let's suppose the relation true for $n = q$ and prove it for $n = q + 1$.

$$w_k^{q+1}(\overline{b}_k) = w_k(w_k^q(\overline{b}_k)) = w_k(h_k(w^q(b_0)\,\mathrm{Pr}_{k-1}(y - w^q(b_0)))) =$$
$$= h_k(w^{q+1}(b_0)\,\mathrm{Pr}_{k-1}(y - w^{q+1}(b_0))).$$

since w_k and w "moves" every symbol $h_k(b_i b_{i+1}\dots b_{i+k-1})$, respectively b_i, on the next "floor" of w_k, respectively w, and "fills" it to $|w(b_i)| = |w_k h_k(b_i b_{i+1}\dots b_{i+k-1})|$; moreover the "concatenation" by exactly $k-1$ symbols of the arguments of two succesive h_k is kept. The lemma is proved.

Analogously it results:

2.6. Lemma. $w_k^n(h_k(b_1 b_2\dots b_k)) = h_k(w^n(b_1)\,\mathrm{Pr}_{k-1}(w^n(b_2)\dots w^n(b_k)))$, $\forall n \in N^*$.

Now, the proof of 2.3 is immediate. For every $n \geqslant 0$, we have

$$\mathrm{Pr}_{|w_k^n(\overline{b}_k)|}(y_k) = w_k^n(\overline{b}_k) = h_k(w^n(b_0)\,\mathrm{Pr}_{k-1}(y - w^n(b_0))) = \mathrm{Pr}_{|w^n(b_0)|}(h_k(y)).$$

But, using 2.4 and 2.5, it follows that $|w_k^n(\overline{b}_k)| = |w^n(b_0)|$, therefore $y_k = h_k(y)$.

2.7. For an internal sequence over C_k we define the function $\overline{h}_2 : C_k^2 \to C_{k+1}$ given by $h_k(b_1 b_2 b_3\dots b_k)\,h_k(b_2 b_3\dots b_k b_{k+1}) \mapsto h_{k+1}(b_1 b_2 b_3\dots b_k b_{k+1})$ and nondefined for "nonconcatened" elements of C_k^2. $\overline{h}_2$ has a natural extension to blocks and sequences. The function can always be applied to the sequence y_k, $k \geqslant 2$, since the adjacent elements are concatened by twos. Because the elements of y_k proceed from application of h_k to y, it follows that $\overline{h}_2(y_k) = h_{k+1}(y) = y_{k+1}$, therefore $\overline{h}_2$ keps the internal sequences.

3. In this section we introduce a function, analogous to h_k, for tag sequences.

3.1. Let $\mathcal{T} = (B, b_0, w, h, A)$ be a tag system in the general sense (TSG), i.e. (B, b_0, w) is a TSL, A a finite set of symbols and $h : B \to A$ a function. Let $x = \mathrm{seq}(\mathcal{T})$ be the external sequence of $\mathcal{T}$ (tag sequence), i.e. $x = h(\mathrm{intseq}(B, b_0, w))$. If $|w(b)| = m$, $\forall b$, $b \in B$, x will be called uniform tag sequence of modulus m (UTSm).

3.2. For every fixed $k \in N$, $k \geqslant 2$, we define the one-to-one mapping $g_k : A^k \to D_k$ by $a_1 a_2\dots a_k \mapsto g_k(a_1\dots a_k)$.

3.3. Proposition. *The set of tag sequences is closed under the mapping g_k.*

Proof. Given the TSG $\mathcal{T}$, we define the function $H_k : C_k \to D_k$ by $h_k(b_1 b_2\dots b_k) \mapsto g_k(h(b_1)\,h(b_2)\dots h(b_k))$. Then, by 2.3, $\mathcal{T}_k = (C_k, \overline{b}_k, w_k, H_k, D_k)$ satisfies $g_k(x) = x_k$, where $x_k = \mathrm{seq}(\mathcal{T}_k)$.

3.4. Corollary. *The set $U(m)$ of UTSm's is closed under the mapping g_k.*

Proof. This follows from 3.3 and 2.4. But 3.4 can be proved directely too, using the properties of UTS. Applying to $x = \mathrm{seq}(\mathcal{T})$, $x = a_0 a_1\dots a_{k-1}$ $a_{k-1}\dots$, the finite modification $x'(i) = x(k-1+i)$, $\forall i \geqslant 0$ we get the UTSm $x' = a_{k-1}a_k a_{k+1}\dots$. Applying to x' the finitary sequential transformation $(I, S, s_1, \delta, \lambda, C)$ with $I = A$, $S = A^{k-1}$, $C = D_k$, $s_1 = a_0 a_1\dots a_{k-1}$, $\delta : A^{k-1} \times A \to A^{k-1}$ by $\delta : (a_{i_1} a_{i_2}\dots a_{i_{k-1}}, a) \mapsto a_{i_2}\dots a_{i_{k-1}} a$ and $\lambda : A^{k-1} \times A \to D_k$ by $\lambda : (a_{i_1} a_{i_2}\dots a_{i_{k-1}}, a) \mapsto g_k(a_{i_1} a_{i_2}\dots a_{i_{k-1}} a)$ we remark that $x''(i) = \lambda(\delta(s_1, x'(0)x'(1)\dots x'(i-1)), x'(i))$, $\forall i \in N$, is an UTSm and $x'' = g_k(x)$.

3.5. *Example.* Let $\mathcal{T} = (\{0,1\}, 0, w)$ be a TSL, with $w(0) = 01$, $w(1) = 10$, which generates the Thue sequence

$$x = 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ \dots.$$

If we denote $C_3 = \{h_3(001), h_3(010), h_3(011), h_3(100), h_3(101), h_3(110)\}$, then the TSL $\mathcal{T}_3 = (C_3, h_3(011), w_3)$, with $w_3(h_3(001)) = h_3(0101) = h_3(010)\ h_3(101)$, $w_3(h_3(010)) = w_3(h_3(011)) = h_3(011)\ h_3(110)$, $w_3(h_3(100)) = w_3(h_3(101)) = h_3(100)\ h_3(001)$, $w_3(h_3(110)) = h_3(101)\ h_3(010)$, generates the sequence

$$x_2 = h_3(011)\,h_3(110)\,h_3(101)\,h_3(010)\,h_3(100)\,h_3(001)\dots.$$

4. The result from 3.4. suggests that in an uniform tag sequence the blocks and the symbols have the same distribution.

4.1. Definition. *Let's consider $M \in A^*$ a block with $|M| = k$, x an UTSm over A and $\pi(M, x, n)$ the number of occurrences of M among the first n symbols of x. The lower and upper k-asymptotic density of M in x are respectively the limits*

$$\underline{d}_k(M, x) = \liminf_{n\to\infty} \frac{\pi(M, x, n+k-1)}{n},$$

$$\overline{d}_k(M, x) = \limsup_{n\to\infty} \frac{\pi(M, x, n+k-1)}{n}.$$

If $\underline{d}_k(M, x) = \overline{d}_k(M, x)$, *their common value, denoted by $d_k(M, x)$, is called the k-asymptotic density of M in x.*

4.2. *Remark.* For $k = 1$ we get the asymptotic density $d(a, x)$ defined by A. Cobham in [1].

4.3. Lemma. $\underline{d}_k(M, x) = \underline{d}(g_k(M), x_k)$.

The lemma is immediate because

$$\frac{\pi(M, x, n+k-1)}{n} = \frac{\pi(g_k(M), x_k, n)}{n}$$

is true for all natural n, $n \geqslant 1$.

4.4. *Remark.* Analogously we get $\overline{d}_k(M, x) = \overline{d}(g_k(M), x_k)$ and $d_k(M, x) = d(g_k(M), x)$ (this, if and only if one of the two limits exists).

4.5. *Remark*. The k-logarithmic density of M in x can be introduced, e.g. the lower k-logarithmic density of M in x is

$$\underline{\delta}_k (M, x) = \lim_{n \to \infty} \inf \frac{1}{\log n} \sum_{i \in I_n} \frac{1}{i}$$

where $I_n = \{ i \in N^* \mid 0 \leqslant i \leqslant n,\ x(i)\, x(i+1) \dots x(i+k-1) = M \}$.

One can prove for the k-logarithmic density results analogous to 4.3 and 4.4. these involving together with the theorem 7 of [1], the existence of the k-logarithmic density $\delta_k (M, x)$ for every UTSm x over A and $M \in A^*$ with $|M| = k$.

4.6. *Remark*. A. Cobham's theorem 6 of [1] and 4.4 involve that if $d_k(M, x)$ exists, then it is rational.

Other theorems from the mentioned paper can be easily adapted to blocks of length k, by examining the block's "right context" of length $k-1$. For example the following:

4.7. Proposition. *Let x be an* UTS*m over A generated by $\mathcal{T}$ and $M \in A^*$, $|M| = k$. Then $d_k (M, x) = 0$ if and only if there exists an integer $v \geqslant 1$, such that for no integer $i, j \geqslant 0$ does M occur in the block* $(\mathrm{desc}_m (x, j, im^u + v))$ *$x((im^u + v + 1)m^j) \dots x((im^u + v + 1)m^j + k - 2)$, where u is the unique integer such that $m^{u-1} \leqslant v < m^u$ and* $\mathrm{desc}_m (x, j, i) = x(im^j)\, x(im^j + 1) \dots x(im^j + m^j - 1)$.

Proof. $d_k(M, x) = 0$ iff $d(g_k(M), x_k) = 0$, (see 4.4) that is by [1] Theorem 9 iff there exists an integer $v \geqslant 1$, such that for no integer i, $j \geqslant 0$ does $g_k(M)$ occur in $\mathrm{desc}_m (x_k, j, im^u + v)$, where u is the unique integer such that $m^{u-1} \leqslant v < m^u$, that is $g_k(M)$ does occur in $H_k(\mathrm{desc}_m (y_k, j, im^u + v))$, where $y_k = \mathrm{intseq}\, (\mathcal{T}_k)$. Using the proof of 2.3 and 2.6 we have $H_k(\mathrm{desc}_m (y_k, j, s)) = H_k w_k^j (y_k(s)) = H_k w_k^j (h_k (y(s)\, y(s+1) \dots$ $\dots y(s + k - 1))) = H_k h_k (w^j (y(s))\, \mathrm{Pr}_{k-1} (w^j y(s+1) \dots w^j y(s + k - 1))) =$ $= H_k h_k (\mathrm{desc}_m (y, j, s)(y((s + 1)m^j) \dots y((s + 1)m^j + k - 2))$, where $s = im^u + v$. Let Q be the argument of the last expression. It follows from the proof of **3.3** that $H_k h_k = g_k h$ and since g_k is one-to-one mapping $h = g_k^{-1} H_k h_k$. Then the assertion „does $g_k(M)$ occur in $H_k h_k (Q)$" is true iff the assertion „does M occur in $h(Q)$" is true as well.

REFERENCES

1. Cobham A. — *Uniform Tag Sequences*, Math. Systems Theory, 6 (1972), 164—192.

Received 27. VI. 1977

Computer Center
University of Iaşi
Romania

Analele ştiinţifice ale Universităţii „Al. I. Cuza" din Iaşi
Tomul XXIII, S. I a, f. 2, 1977

ON THE WEAK EQUIVALENCE PROBLEM OF EXTENDED SICK OL-SYSTEMS WITH SOME GENERATIVE DEVICES

BY

SORIN ISTRAIL

1. Introduction. In the following we investigate the weak equiva-
lence problem for some generative devices of formal languages. For the
classes of matrix and time-varying grammars with context-free rules we
give the answer to the problem: if for every member G of a given class of
generative devices $\mathcal{G}$, there exists an effectively constructed, equivalent
extended SICK OL-system $\mathcal{S}$.

For the two classes of considered grammars the answer is „Yes".

Let I be a finite alphabet of symbols, I^* the free semigroup genera-
ted by I, λ the unit of I^* and $I^+ = I^* \setminus \{\lambda\}$. The elements of I^* are called words and the subsets of I^* are called languages over I.

Definition 1. *A context-free grammar is a 4-tuple* $G = (I_N, I_T, x_0, F)$ *where:*

i) *I_N and I_T are disjoint finite sets of symbols;*

ii) *$x_0 \in I_N$, is the axiom of G;*

iii) *F is a finite set of pairs called the rules of G: $F \subset I_N \times (I_N \cup I_T)^*$.*

For $(x, u) \in F$ we write $x \to u$.

The basic concept of this paper — the SICK OL-system — is presented in the following definition.

Introduced in [2] as a generalization of OL-systems [1], [5], motivated by biological reasons, SICK OL-systems are proved to be a very useful tool for the simulation of the generative devices introduced in the formal language theory [3], [4], [6], [7].

Definition 2. *An extended SICK OL-system is a 3-tuple $\mathcal{S} = (S, \Pi, J)$ where $S = \langle V, P, w \rangle$, $\Pi = (\pi_a)_{a \in V}$, $J \subset V$ and the components of S satisfy the following conditions:*

a) *V is a finite alphabet of symbols,*

b) *P is a finite set of contect-free rules: $P \subset V \times V^*$, so that for every $a \in V$, there exists $u \in V^*$ with $(a, u) \in P$,*

c) *$w \in V^+$,*

d) *for every $a \in V$, $\pi_a = \langle V_a, E_a, F_a, n(a) \rangle$ where:*