COMMUTATIVE CLOSURE OF CONTEXT-FREE LANGUAGES*

L. P. Lisovik

UDC 62-5:681.3:007

Let $L$ be a language in an alphabet $\Sigma$, where $\Sigma = \{a_1, \ldots, a_n\}$. By $\pi(L)$ we denote the commutative closure of language $L$, being the set of all words obtained by all possible permutations of letters in the words of language $L$.

In the present paper we have indicated a necessary and sufficient condition for language $\pi(L)$ to be contextless for any contextless (or, what is the same, context-free) language $L$ closed relative to the iteration operation. The condition mentioned is algorithmically verifiable. A short account of the results of the present paper appeared in [2]. The generalized solution of one special problem posed in [1] (p. 238), given in [3], is obtained as a corollary.

1. Definitions

The iteration operation ($*$) associates with each language $L$ the language $L^*$ made up from all possible products of words of language $L$. Here $e \in L^*$, where $e$ is the empty word. For any word $w$ we assume that $w^* = \{e, w, w^2, \ldots \}$.

$$\Sigma = \{a_1, \ldots, a_n\}, \quad N = \{0, 1, \ldots\}, \quad N^+ = N \setminus \{0\}.$$  

For each ordered collection of words $w = (w_1, \ldots, w_n)$ we assume that $f_w$ is a mapping of set $N^n$ into the set $\bigcup_n w_n^*$ such that if $c = (c_1, \ldots, c_n) \in N^n$, then $f_w(c) = w_1^{c_1} \cdot \cdots \cdot w_n^{c_n}$.

We assume that $f_0 = f_w$, where $w = (a_1, \ldots, a_n)$; $0 \in N^k$ is the null vector.

For a system $P = \{p_1, \ldots, p_m\} \subseteq N^k$ and a vector $c \in N^k$ we assume that $L(c, P) = \{x \mid x = c + \sum_{i=1}^m k_ip_i, k_i \in N\}$. Set $L(c, P)$ is called a linear set with preperiod $c$ and system $P$ of periods.

The finite union of linear sets is called a semilinear set.

*The general concepts used in the article are defined in accord with [1].


---

**LITERATURE CITED**

For finite sets $C$, $P \subseteq \mathbb{N}^k$ we assume that $L(C; P) = \{x \mid x = y + z, y \in C, z \in L(\theta, P)\}$.

A vector $p \in \mathbb{N}^k$ is said to be stratified if it has no more than two nonzero coordinates.

A system of vectors $Q \subseteq \mathbb{N}^k$ is said to be stratified if each vector $q \in Q$ is stratified and no vectors $q = (q_1, \ldots, q_k) \in Q$, $p = (p_1, \ldots, p_k) \in Q$ and numbers $1 \leq f < s < r < t \leq k$ exist such that $q_f > 0$, $q_r > 0$, $p_s > 0$, $p_t > 0$.

We assume that system $P \subseteq \mathbb{N}^k$ is stratifiable with coefficient $l \in \mathbb{N}^+$ if for each $w \in L(\theta, P)$ there exists a stratified subsystem $Q$ of system $P$ such that $lw \in L(\theta, Q)$.

A vector $p \in \mathbb{N}^k$ is said to be stratified if it has no more than two nonzero coordinates.

A system of vectors $Q \subseteq \mathbb{N}^k$ is said to be stratified if each vector $q \in Q$ is stratified and no vectors $q = (q_1, \ldots, q_k) \in Q$, $p = (p_1, \ldots, p_k) \in Q$ and numbers $1 \leq f < s < r < t \leq k$ exist such that $q_f > 0$, $q_r > 0$, $p_s > 0$, $p_t > 0$.

We assume that system $P \subseteq \mathbb{N}^k$ is stratifiable with coefficient $l \in \mathbb{N}^+$ if for each $w \in L(\theta, P)$ there exists a stratified subsystem $Q$ of system $P$ such that $lw \in L(\theta, Q)$.

A finite sequence of positive integers $\tau = \alpha_1, \ldots, \alpha_k$, where $1 \leq \alpha_1 \leq n$, $1 \leq i \leq k$, is called a sequence.

For any $P \subseteq \mathbb{N}^n$ and sequence $\tau = \alpha_1, \ldots, \alpha_k$, by $P_\tau$ we denote the set of all vectors $q = (q_1, \ldots, q_k)$ for which a vector $p = (p_1, \ldots, p_n) \in P$ exists such that $q$ and $p$ have an equal number of nonzero coordinates and the following conditions are fulfilled: if $q_i > 0$, then $q_i = p_{\alpha_i}$; if $q_i > 0$ and $q_j > 0$, then $\min(\alpha_i, \alpha_j)$.

We assume that $\varphi$ is a mapping of set $\Sigma^*$ into set $\mathbb{N}^n$ such that for any word $w \in \Sigma^*$ we have $\varphi(w) = (k_1, \ldots, k_n)$, where $k_i = |w|_i$.

**2. Necessary and Sufficient Condition for a Language to be Contextless**

**THEOREM 1.** Let $P \subseteq \mathbb{N}^n$ be a finite system of stratified vectors and let $L = f_0(L(\theta, P))$. Then the set $\pi(L)$ is a contextless language if and only if system $P_\tau$ is stratifiable for any sequence $\tau = \alpha_1, \alpha_2, \alpha_3, \alpha_4$.

**Proof.** Necessity. A set $C \subseteq \mathbb{N}^n$ exists such that $f_w(L(C; P_\tau)) = \pi(L)$ is a contextless language if and only if system $P_\tau$ is stratifiable for any sequence $\tau = \alpha_1, \alpha_2, \alpha_3, \alpha_4$.

Sufficiency. Let $P = \{p_1, \ldots, p_m\}$ and $p_i = (p_{i1}, \ldots, p_{in})$, $1 \leq i \leq m$. Let $d_j = \max\{p_{ij}, 1 \leq i \leq m\}$ and $d = \max\{d_j, 1 \leq j \leq n\}$.

We take a number $l_0 \in \mathbb{N}$ such that each system $P_\tau$ in the condition of the theorem is stratifiable with coefficient $l_0$.

Let us consider an automaton with a stack memory $M = (K, \Sigma, \Gamma, \delta, Z_0, q_0, q^*)$, where:

- $K \subseteq \mathbb{N}^n$ is the state set, $K = \{(a_1, \ldots, a_n) \mid 0 \leq a_i \leq d_i, 1 \leq i \leq n\}$;
- $\Sigma = \{a_1, \ldots, a_n\}$ is the set of input symbols;
- $\Gamma = \Sigma \cup \{Z_0\}$ is the set of symbols of the stack memory;
- $\delta$ is the mapping of set $K \times (\Sigma \cup \{\varepsilon\}) \times \Delta$ into the set of all finite subsets of set $K \times \Gamma^*$, where $\Delta = \{w \mid w \in \Gamma^*, |w|_i \leq nm^3l_0(d + 1)^{t_i} \}$ (defined below);
- $Z_0 \in \Gamma$ is a marker in the stack memory;
- $q_0 \in K$ is the initial state, $q_0 = (0, \ldots, 0)$;
- $q^* = q_0$ is the final state.

The mapping $\delta$ is defined by the following schemes of transition rules of types A, B, and C:

- **A:** $(g, \gamma) \in \delta((q, a_i, \gamma), q = (a_1, \ldots, a_n), a_i < d_i, 1 \leq i \leq n)$, and $g = q + \varphi(a_i)$;
  $(q, \gamma) \in \delta((q, a_i, \gamma), q = (a_1, \ldots, a_n), a_i = d_i, 1 \leq i \leq n)$, and $\gamma_i = \gamma \cdot a_i$.

Transitions of type A signify the entry of letters into the finite or stack memory.

- **B:** $(g, \gamma) \in \delta((q, a_i, \gamma), g = q - p_i$ for some $i \in 1, m$.

Transitions of type B signify the release of the finite memory with respect to some period $p_i \in P$.

- **C:** $(g, \gamma) \in \delta((q, \varepsilon, \gamma), q = (a_1, \ldots, a_n), a_i < d_i, 1 \leq i \leq n)$, $\gamma = Z_0 \cdot \varphi(a_i)$, and $\gamma_i = Z_0 \cdot \varphi(a_i)$.

Transitions of type C signify the rewriting of letters from the upper block of the stack memory of length $nm^3l_0(d + 1)^{t_i}$ into the finite memory of the automaton.