CLOSURE PROPERTIES OF LINEAR LANGUAGES UNDER OPERATIONS OF LINEAR DELETION

Tomáš MASOPUST, Doctoral Degree Programme (2)
Dept. of Information Systems, FIT, BUT
E-mail: masopust@fit.vutbr.cz

Supervised by: Prof. Alexander Meduna

ABSTRACT

In this paper, we give constructive prove that linear languages are closed under operations of regular deletion, and that they are not closed under operations of linear deletion. The operations are called random parallel, parallel, sequential, scattered sequential, and multiple scattered sequential deletion. In addition, we prove that the closure of linear languages under linear random parallel, parallel, or sequential deletion is the family of all recursively enumerable languages. In the conclusion, we formulate two open problems.

1 INTRODUCTION

The language operations that delete some parts of strings play an important role in modern informatics. So, it is no surprise that the formal language theory has payed a special attention to their study (see [2, 4, 5]).

In this paper, we constructively prove that linear languages are closed under operations of regular random parallel, parallel, sequential, scattered sequential, and multiple scattered sequential deletion, and that they are not closed under linear types of these operations. We also prove that the closure of linear languages under linear random parallel, parallel, or sequential deletion is the family of all recursively enumerable languages. In the conclusion of this paper, we formulate two open problems.

2 PRELIMINARIES

In this paper, we assume that the reader is familiar with the formal language theory (see [3]). Let RE, REC, CF, DCF, LIN, and REG denote the families of recursively enumerable, recursive, context-free, deterministic context-free, linear, and regular languages, respectively.

3 DEFINITIONS AND EXAMPLES

Let $L, K \subseteq \Sigma^*$ be two languages.

Definition 3.1. *Random Parallel Deletion* of L and K is denoted by $[\perp, L, K]$ and defined as the set $[\perp, L, K] = \{u_1u_2 \dots u_nu_{n+1} \in \Sigma^* : u_1x_1u_2 \dots u_nx_nu_{n+1} \in L,\, x_i \in K,\, 1 \le i \le n,\, n \ge 1\}$.

Example 3.1. Let $L = \{abababa, aababa, abaabaaba\}$ and $K = \{aba\}$.

- $[\perp, \{abababa\}, K] = \{baba, abba, abab, b\}$.
- $[\perp, \{aababa\}, K] = \{aba, aab\}$.
- $[\perp, \{abaabaaba\}, K] = \{abaaba, aba, \varepsilon\}$.
- $[\perp, L, K] = \{baba, abba, abab, b, aba, aab, abaaba, \varepsilon\}$.

Definition 3.2. *Parallel Deletion* of L and K is denoted by $[\perp_a, L, K]$ and defined as the set $[\perp_a, L, K] = \{u_1u_2 \dots u_nu_{n+1} \in \Sigma^* : u_1x_1u_2 \dots u_nx_nu_{n+1} \in L,\, x_j \in K,\, \{u_i\} \cap \Sigma^*(K \setminus \{\varepsilon\})\Sigma^* = \emptyset,\, 1 \le i \le n+1,\, 1 \le j \le n,\, n \ge 1\}$.

Example 3.2. Let $L = \{abababa, aababa, abaabaaba\}$ and $K = \{aba\}$.

- $[\perp_a, \{abababa\}, K] = \{b, abba\}$.
- $[\perp_a, \{aababa\}, K] = \{aba, aab\}$.
- $[\perp_a, \{abaabaaba\}, K] = \{\varepsilon\}$.
- $[\perp_a, L, K] = \{b, abba, aba, aab, \varepsilon\}$.

Definition 3.3. *Sequential Deletion* of L and K is denoted by $[\perp_1, L, K]$ and defined as the set $[\perp_1, L, K] = \{u_1u_2 \in \Sigma^* : u_1xu_2 \in L,\, x \in K\}$.

Example 3.3. Let $L = \{abababa, ab, aba\}$ and $K = \{aba\}$.

- $[\perp_1, \{abababa\}, K] = \{baba, abba, abab\}$.
- $[\perp_1, \{ab\}, K] = \emptyset$.
- $[\perp_1, \{aba\}, K] = \{\varepsilon\}$.
- $[\perp_1, L, K] = \{baba, abba, abab, \varepsilon\}$.

Definition 3.4. *Scattered Sequential Deletion* of L and K is denoted by $[\perp_{1s}, L, K]$ and defined as the set $[\perp_{1s}, L, K] = \{u_1u_2 \dots u_nu_{n+1} \in \Sigma^* : u_1x_1u_2 \dots u_nx_nu_{n+1} \in L,\, x_1x_2 \dots x_n \in K,\, n \ge 1\}$.

Example 3.4. Let $L = \{abacba\}$ and $K = \{ab, ca\}$.

- $[\perp_{1s}, \{abacba\}, \{ab\}] = \{acba, baca, abca\}$.
- $[\perp_{1s}, \{abacba\}, \{ca\}] = \{abab\}$.
- $[\perp_{1s}, \{abacba\}, K] = \{acba, baca, abca, abab\}$.

Definition 3.5. *Multiple Scattered Sequential Deletion* of L and K is denoted by $[\perp_s, L, K]$ and defined as the set $[\perp_s, L, K] = \{u_1 u_2 \dots u_n u_{n+1} \in \Sigma^* : u_1 x_1 u_2 \dots u_n x_n u_{n+1} \in L, x_1 x_2 \dots x_n \in K^+, n \geq 1\}$.

Example 3.5. Let $L = \{abacba\}$ and $K = \{ab, ca\}$.

- $[\perp_s, \{abacba\}, \{ab\}] = \{acba, baca, abca, ca\}$.
- $[\perp_s, \{abacba\}, \{ca\}] = \{abab\}$.
- $[\perp_s, \{abacba\}, \{ab, ca\}] = \{acba, baca, abca, ca, abab, ab\}$.

For any two families of languages $\mathcal{X}$ and $\mathcal{Y}$ denote by $\langle x, \mathcal{X}, \mathcal{Y} \rangle$ the set $\langle x, \mathcal{X}, \mathcal{Y} \rangle = \{[x, L, K] : L \in \mathcal{X}, K \in \mathcal{Y}\}$, where $x \in \{\perp, \perp_a, \perp_1, \perp_{1s}, \perp_s\}$.

4 RESULTS

Linear languages have been proved to be closed under operations of regular deletion (see [4]). The proofs given there are not constructive. Because of limited space, we describe only the constructions and omit the rigorous proofs.

Theorem 4.1. $\langle x, LIN, REG \rangle = LIN$, $x \in \{\perp, \perp_1, \perp_{1s}, \perp_s\}$.

Proof. Let $L \in LIN$, then $L = [x, L, \{\varepsilon\}] \in \langle x, LIN, REG \rangle$, $x \in \{\perp, \perp_1, \perp_{1s}, \perp_s\}$.

Let $L \in LIN$ and $K \in REG$. Without loss of generality, there is a proper linear grammar $G_L = (N_L, \Sigma_L, P_L, S_L)$ and a regular grammar $G_K = (N_K, \Sigma_K, P_K, S_K)$ such that S_K does not occur on the right-hand side of any rule, $L = \mathcal{L}(G_L)$, and $K = \mathcal{L}(G_K)$. ($S_K \to \varepsilon$ is the only possible ε-rule in G_K.) Construct linear grammar $G = (N, \Sigma_L, P, S)$, where $N = \{S\} \cup \{\langle x, B, y, U, V \rangle : x, y \in \Sigma_L^*, |x|, |y| \leq \max\{|u|, |v| : A \to uBv \in P_L\}, B \in N_L \cup \{\varepsilon\}, U, V \in N_K \cup \{\varepsilon\}\}$, and P contains rules of the following forms (depending on x):
$x = \perp$:

1) $S \to \langle \varepsilon, S_L, \varepsilon, S_K, \varepsilon \rangle$	
2) $\langle ax, A, yb, U, \varepsilon \rangle \to \langle ax, A, y, U, \varepsilon \rangle\, b$	
3) $\langle ax, A, yb, S_K, V \rangle \to a\, \langle x, A, yb, S_K, V \rangle$	
4) $\langle ax, A, yb, U, Y \rangle \to \langle x, A, yb, V, Y \rangle$	if $U \to aV \in P_K$, $V \in N_K \cup \{\varepsilon\}$
5) $\langle ax, A, yb, U, Y \rangle \to \langle ax, A, y, U, X \rangle$	if $X \to bY \in P_K$, $X \in N_K$
6) $\langle ax, A, yb, U, S_K \rangle \to \langle ax, A, y, U, S_K \rangle\, b$	
7) $\langle ax, A, yb, U, S_K \rangle \to \langle ax, A, yb, U, \varepsilon \rangle$	
8) $\langle ax, A, yb, \varepsilon, X \rangle \to a\, \langle x, A, yb, \varepsilon, X \rangle$	
9) $\langle ax, A, yb, \varepsilon, X \rangle \to \langle ax, A, yb, S_K, X \rangle$	
10) $\langle \varepsilon, A, \varepsilon, X, Y \rangle \to \langle x, B, y, X, Y \rangle$	if $A \to xBy \in P_L$
11) $\langle \varepsilon, \varepsilon, \varepsilon, X, X \rangle \to \varepsilon$	

$x = \perp_1$: We eliminate the rules allowing to delete more than one substring, i.e. rules of type 7 and 9.
$x = \perp_{1s}$: In each state, we can either generate a next symbol or delete it.

1) $S \to \langle \varepsilon, S_L, \varepsilon, S_K, \varepsilon \rangle$
2) $\langle ax, A, yb, U, Y \rangle \to a \langle x, A, yb, U, Y \rangle$
3) $\langle ax, A, yb, U, Y \rangle \to \langle ax, A, y, U, Y \rangle b$
4) $\langle ax, A, yb, U, Y \rangle \to \langle x, A, yb, V, Y \rangle$ if $U \to aV \in P_K$, $V \in N_K \cup \{\varepsilon\}$
5) $\langle ax, A, yb, U, Y \rangle \to \langle ax, A, y, U, X \rangle$ if $X \to bY \in P_K$, $X \in N_K$
6) $\langle \varepsilon, A, \varepsilon, X, Y \rangle \to \langle x, B, y, X, Y \rangle$ if $A \to xBy \in P_L$
7) $\langle \varepsilon, \varepsilon, \varepsilon, X, X \rangle \to \varepsilon$

$x = \perp_s$: As K^+ is regular, for K regular, the proof is the same as for $x = \perp_{1s}$. □

Theorem 4.2. $\langle \perp_a, LIN, REG \rangle = LIN$.

Proof. It is easy to see that $LIN \subseteq \langle \perp_a, LIN, REG \rangle$.

Let $L \in LIN$ and $K \in REG$. Without loss of generality, there is a proper linear grammar $G_L = (N_L, \Sigma_L, P_L, S_L)$ and a regular grammar $G_K = (N_K, \Sigma_K, P_K, S_K)$ such that S_K does not occur on the right-hand side of any rule, $L = \mathcal{L}(G_L)$, and $K = \mathcal{L}(G_K)$. ($S_K \to \varepsilon$ is the only possible ε-rule in G_K.) Construct linear grammar $G = (N, \Sigma_L, P, S)$, where $N = \{S\} \cup \{\langle x, B, y, U, V, M, N \rangle : x, y \in \Sigma_L^*, |x|, |y| \le \max\{|u|, |v| : A \to uBv \in P_L\}, B \in N_L \cup \{\varepsilon\}, U, V \in N_K \cup \{\varepsilon\}, M, N \subseteq N_K \cup \{\varepsilon\}\}$, and P contains rules of the following forms:

1) $S \to \langle \varepsilon, S_L, \varepsilon, S_K, \varepsilon, \{S_K\}, \{\varepsilon\} \rangle$
2) $\langle ax, A, yb, U, \varepsilon, M, N \rangle \to \langle ax, A, y, U, \varepsilon, M, N' \rangle b$ if $\varepsilon \notin M$, $S_K \notin N$, $b \in \Sigma_L$
3) $\langle ax, A, yb, S_K, V, M, N \rangle \to a \langle x, A, yb, S_K, V, M', N \rangle$ if $\varepsilon \notin M$, $S_K \notin N$, $a \in \Sigma_L$
4) $\langle ax, A, yb, U, Y, M, N \rangle \to \langle x, A, yb, V, Y, \{S_K\}, N \rangle$ if $\varepsilon \notin M$, $S_K \notin N$,
$U \to aV \in P_K$, $V \in N_K \cup \{\varepsilon\}$
5) $\langle ax, A, yb, U, Y, M, N \rangle \to \langle ax, A, y, U, X, M, \{\varepsilon\} \rangle$ if $\varepsilon \notin M$, $S_K \notin N$,
$X \to bY \in P_K$, $X \in N_K$
6) $\langle ax, A, yb, U, S_K, M, N \rangle \to \langle ax, A, y, U, S_K, M, N' \rangle b$ if $\varepsilon \notin M$, $S_K \notin N$, $b \in \Sigma_L$
7) $\langle ax, A, yb, U, S_K, M, N \rangle \to \langle ax, A, yb, U, \varepsilon, M, N \rangle$
8) $\langle ax, A, yb, \varepsilon, X, M, N \rangle \to a \langle x, A, yb, \varepsilon, X, M', N \rangle$ if $\varepsilon \notin M$, $S_K \notin N$, $a \in \Sigma_L$
9) $\langle ax, A, yb, \varepsilon, X, M, N \rangle \to \langle ax, A, yb, S_K, X, M, N \rangle$
10) $\langle \varepsilon, A, \varepsilon, X, Y, M, N \rangle \to \langle x, B, y, X, Y, M, N \rangle$ if $A \to xBy \in P_L$, $\varepsilon \notin M$, $S_K \notin N$
11) $\langle \varepsilon, \varepsilon, \varepsilon, X, X, M, N \rangle \to \varepsilon$ if $\varepsilon \notin M$, $S_K \notin N$

where $M' = \{S_K\} \cup \{D \in N_K \cup \{\varepsilon\} : A \to aD \in P_K, A \in M\}$ and $N' = \{\varepsilon\} \cup \{D \in N_K : D \to bC \in P_K, C \in N\}$. □

Now, we prove that the operations of linear deletion are very powerful—linear languages with the operation of linear deletion characterize recursively enumerable languages.

Theorem 4.3. $\langle x, LIN, LIN \rangle = RE$, $x \in \{\perp, \perp_a, \perp_1\}$.

Proof. It is not hard to construct a Turing machine accepting $\langle x, LIN, LIN \rangle$.

Now, suppose $L \in RE$, $L \subseteq \Sigma^*$, $\Sigma = \{a_1, \ldots, a_n\}$. *Extended Post correspondence problem* (EPCP), P, is a tuple $P = (\{(u_1, v_1), \ldots, (u_r, v_r)\}, (z_{a_1}, \ldots, z_{a_n}))$, where $u_i, v_i, z_a \in \{0,1\}^*$ for $i = 1, \ldots, r$, and $a \in \Sigma$. The language represented by P is the set $\mathcal{L}(P) = \{x_1 x_2 \ldots x_n \in \Sigma^* : \exists s_1, \ldots, s_l \in \{1, \ldots, r\}, l \ge 1, v_{s_1} \ldots v_{s_l} = u_{s_1} \ldots u_{s_l} z_{x_1} \ldots z_{x_n}\}$. For each recursively enumerable language, L, there is an EPCP, P, such that $\mathcal{L}(P) = L$ (see [1, Theorem 1]). Thus, $x_1 x_2 \ldots x_n \in L$ if and only if $x_1 x_2 \ldots x_n \in \mathcal{L}(P)$. Generate $x_1 x_2 \ldots x_n$ as follows:

$$
\begin{aligned}
S' \Rightarrow\ & \$S \Rightarrow \$z_{x_n}^R S x_n \Rightarrow \$z_{x_n}^R z_{x_{n-1}}^R S x_{n-1} x_n \\
\Rightarrow^*\ & \$z_{x_n}^R z_{x_{n-1}}^R \dots z_{x_1}^R S x_1 \dots x_{n-1} x_n \\
\Rightarrow^*\ & \$z_{x_n}^R z_{x_{n-1}}^R \dots z_{x_1}^R A \$ x_1 \dots x_{n-1} x_n \\
\Rightarrow^*\ & \$z_{x_n}^R z_{x_{n-1}}^R \dots z_{x_1}^R u_{s_l}^R A v_{s_l} \$ x_1 \dots x_{n-1} x_n \\
\Rightarrow^*\ & \$z_{x_n}^R z_{x_{n-1}}^R \dots z_{x_1}^R u_{s_l}^R \dots u_{s_1}^R \# v_{s_1} \dots v_{s_l} \$ x_1 \dots x_{n-1} x_n \\
=\ & \$(u_{s_1} \dots u_{s_l} z_{x_1} \dots z_{x_n})^R \# (v_{s_1} \dots v_{s_l}) \$ x_1 \dots x_n \\
=\ & \$w_1^R \# w_2 \$ x_1 x_2 \dots x_n
\end{aligned}
$$

and $x_1 x_2 \dots x_n \in L$ if and only if there are w_1, w_2 such that $w_1 = w_2$,

where $z_{x_i}, u_{s_j}, v_{s_j} \in \{0,1\}^*$, $\$, \# \notin \Sigma \cup \{0,1\}$.

In addition, there is a linear grammar, G', such that $\mathcal{L}(G') = \{\$w^R \# w\$: w \in \{0,1\}^*\}$. Thus, $L = [x, \mathcal{L}(G), \mathcal{L}(G')]$, $x \in \{\perp, \perp_a, \perp_1\}$. $\square$

Theorem 4.4. *REC* $\subset \langle x, LIN, LIN \rangle \subseteq$ *RE*, $x \in \{\perp_{1s}, \perp_s\}$.

Proof. Let L be recursively enumerable language, $L \subseteq \Sigma^*$, $\Sigma \cap \{0,1\} = \emptyset$. The proof follows from Theorem 4.3 because $L = [x, \mathcal{L}(G), \mathcal{L}(G')] \cap \Sigma^*$. If $[x, \mathcal{L}(G), \mathcal{L}(G')]$ is recursive, then so is L. Therefore, for $L \in RE - REC$ the language $[x, \mathcal{L}(G), \mathcal{L}(G')]$ is not recursive language. $\square$

5 OPEN PROBLEMS

Here we summarize two open problems:

1. Is it true that $\langle \perp_{1s}, LIN, LIN \rangle = RE$?
2. Is it true that $\langle \perp_s, LIN, LIN \rangle = RE$?

REFERENCES

[1] Geffert, V.: Context-free-like forms for the phrase-structure grammars. In Proc. Math. Found. Comput. Sci., volume 324 of Lecture Notes in Comput. Sci., pages 309–317, Springer-Verlag, 1988.

[2] Kari, L.: On insertion and deletion in formal languages. PhD thesis, University of Turku, Finland, Department of Mathematics, 1991.

[3] Meduna, A.: Automata and Languages: Theory and Applications. Springer, London, 2000.

[4] Meduna, A., Masopust, T.: Closure properties of trios under operations of regular deletion. Submitted.

[5] Meduna, A., Vítek, M.: New language operations in formal language theory. Schedae Informaticae, 13:123–150, 2004.