Acta Cybernetica, Vol. 10, Nr. 4, Szeged, 1992

Regularizing context-free languages by AFL operations: concatenation and Kleene closure

J. Dassow [*] A. Mateescu [†‡] G. Paun [§¶] A. Salomaa [‖]

Abstract

We consider the possibility to obtain a regular language by applying a given operation to a context-free language. Properties of the family of context-free languages which can be "regularized" by concatenation with a regular set or by Kleene closure are investigated here: size, hierarchies, characterizations, closure, decidability.

1 Introduction

The core of formal language theory is the study of the Chomsky hierarchy, especially of families of regular and of context-free languages. An important problem in this context is to understand the differences between "regularity" and "context-freeness". The question is approached, explicitly or implicitly, in many papers.

Here we follow [2], [3], [4], [7] and consider this problem in relation with operations with languages. Usually, the main topic dealt with when investigating operations with languages is the closure of various families (how much an operation can "complicate" a language). A dual natural question is "how much an operation can simplify languages in a given family". In particular, we are interested in transforming in this way context-free languages into regular languages.

Similar problems are investigated in [2], [4], whereas [3], [7] consider numerical measures of non-regularity of context-free languages and the influence of various operations on them.

Here we investigate the possibility of obtaining a regular language starting from a context-free language and using one of the six AFL operations: union, concatenation, intersection - all by regular sets -, Kleene closure, morphisms and inverse

---

[*] Otto von Guericke University of Magdeburg, Department of Computer Science, PSF 4120, O - 3010 Magdeburg, Germany

[†] Institute of Mathematics of the Romanian Academy of Sciences, Str. Academiei 14, 70109 Bucuresti, Romania

[‡] Research supported by the Academy of Finland, grant nr 11281

[§] Institute of Mathematics of the Romanian Academy of Sciences, Str. Academiei 14, 70109 Bucuresti, Romania

[¶] Research supported by Alexander von Humboldt Foundation

[‖] Academy of Finland and University of Turku, Department of Mathematics, SF - 20500 Turku 50, Finland

morphisms. We enter into details only for the right and left concatenation and for Kleene $*$, namely we study the properties of families of context-free languages which can lead to regular languages by left/right concatenation with regular sets of by Kleene $*$.

2 Notations

For an alphabet V, we denote by V^* the free monoid generated by V under the operation of concatenation; the null element of V^* is denoted by λ and $|x|$ denotes the length of $x \in V^*$. For $x \in V^*, a \in V$, we denote by $|x|_a$ the number of occurrences in x of the symbol a. We denote also by REG, LIN, CF the families of regular, linear and context-free languages.

For a language L we denote by $Pref(L), Suf(L), Sub(L)$ the sets of prefixes, suffixes, respectively subwords of strings in L.

The main problem of this paper is the following: given a language $L \in CF$ and an operation with languages, can we use this operation in such a way to obtain a regular language starting from L ?

In this form, the question is trivial for most AFL operations. For instance, for all context-free languages $L \subseteq V^*$, the languages

(i) $L \cup V^* = V^*$,
(ii) $h(L)$ for all $h : V^* \longrightarrow \{a\}^*$,
(iii) $L \cap R$ for all finite languages R,
(iv) $h^{-1}(L)$ for all $h : \{a\}^* \longrightarrow V^*$,

are regular. The question is not trivial for concatenation and Kleene closure:

(i) Concatenating (on the left side) the non-regular language

$$L_1 = \{a^n b^m \mid 1 \leq n \leq m\}$$

with

$$R = \{a^p \mid p \geq 1\},$$

we obtain a regular language, but no right or left concatenation of

$$L_2 = \{a^n b^n \mid n \geq 1\}$$

with a non-empty set will give a regular language (if $RL_2 \in REG$, for some R, then take $x \in R$ and intersect RL_2 with xa^*b^*; the obtained language is not regular, hence RL_2 is not regular, a contradiction).

(ii) For the above language L_2, the language L_2^* is not regular, but for

$$L_3 = L_2 \cup \{a, b\}$$

we have

$$L_3^* = \{a, b\}^*,$$

which is regular.

Thus, we are led to consider the families

$CL = \{L \in CF \mid \text{there is } R \in REG, R \neq \emptyset, \text{ such that } RL \in REG\}$,
$CR = \{L \in CF \mid \text{there is } R \in REG, R \neq \emptyset, \text{ such that } LR \in REG\}$,
$K = \{L \in CF \mid L^* \in REG\}$,
$K_n = \{L \in CF \mid \text{there is } 1 \leq m \leq n \text{ such that } \bigcup_{i=1}^{m} L^i \in REG\}$, for $n \geq 1$.

We shall investigate here only the families $CL, K, K_n, n \geq 1$; the results for CL are true also for CR, with obvious modifications.

3 The size of the families introduced above

The next relations follow from definitions.

Lemma 3.1 *(i)* $REG \subseteq CL \subseteq CF$,
(ii) $REG \subseteq K \subseteq CF$,
(iii) $REG = K_1 \subseteq K_2 \subseteq \ldots \subseteq CF$.

Lemma 3.2 $K_n \subseteq K$, *for all* $n \geq 1$.

Proof. Take $L \in K_n$. There is $m \leq n$ such that $\bigcup_{i=1}^{m} L^i \in REG$. Clearly, $L^* = (\bigcup_{i=1}^{m} L^i)^*$, hence also L^* is regular, that is $L \in K$. □

All these inclusions are proper.

Theorem 3.3 $REG \subset CL \subset CF$.

Proof. The language L_1 in the previous section is in CL but it is not regular, whereas the language L_2 in the previous section is not in $CL \cup CR$. □

Lemma 3.4 *(i) If an arbitrary language* $L \subseteq V^*$ *satisfies, for some* $k \geq 0$, *the relation* $V^k \subseteq L$, *then* $V^*L \in REG$. *In particular, if* $\lambda \in L$, *then* $V^*L \in REG$.

(ii) If an arbitrary language $L \in V^*$ *satisfies, for some* $k_1, k_2 \geq 0, k_1, k_2$ *relatively prime, the relation* $V^{k_1} \cup V^{k_2} \subseteq L$, *then* $L^* \in REG$.

Proof. (i) Under the previous conditions, we obtain

$$V^*L = V^*L_k,$$

for $L_k = \{x \in L \mid |x| \leq k\}$.

The inclusion $\subseteq$ is obvious. Conversely, take $x, y \in V^*L$, $x \in V^*, y \in L$. If $|y| \leq k$, then $y \in L_k, xy \in V^*L_k$. If $|y| > k$, then $y = y_1y_2, |y_2| = k$. As $xy_1 \in V^*$, we have again $xy = xy_1y_2 \in V^*L_k$.

The language L_k is finite, hence assertion (i) follows.

(ii) Note that, because k_1, k_2 are relatively prime, there exists $m_0, m_0 \in \mathbf{N}$, such that for any $n \geq m_0$ there are $i, j \in \mathbf{N}$ with $n = ik_1 + jk_2$. Thus L^* contains all words w such that $|w| \geq m_0$, hence $V^* - L^*$ is a finite set; consequently, L^* is regular. □

Corollary 3.5 *CL is incomparable with LIN.*

Proof. The above considered language L_2 proves the relation $LIN - CL \neq \emptyset$.

Conversely, take the Dyck language D over $\{a, b\}$. We have $D \in CF - LIN$. It contains the string λ, hence $D \in CL$ and $CL - LIN \neq \emptyset$ too. □

Corollary 3.6 *For every context-free language* $L, L \subseteq V^*$, *either* L *or* $V^* - L$ *is in* CL.

Proof. Obvious, as one of L and $V^* - L$ contains the null string.

Theorem 3.7 $REG \subset K \subset CF$.

Proof. For all $L \in CF, L \subseteq V^*$, the language $L' = L \cup V$ is in K, as $(L \cup V)^* = V^*$. For $L \in CF - REG$ we obtain $L' \notin REG$, hence $K - REG \neq \emptyset$.

Conversely, the language L_2 in the previous section is not in K (we have $L_2^* \cap a^+b^+ = L_2$), hence $L_2 \in CF - K$. □

Corollary 3.8 *K is incomparable with LIN.*

Proof. For $L \in CF - LIN, L' \notin LIN$, but $L_2 \in LIN - K$. □

Theorem 3.9 *The inclusions $K_n \subset K_{n+1}$ are proper for all $n \geq 1$.*

Proof. (1) $n = 1$.
The language

$$L_{a,b} = \{x \in \{a,b\}^* \mid |x|_a \neq |x|_b\}$$

is not regular (its complement, $\{x \in \{a,b\}^* \mid |x|_a = |x|_b\}$, is clearly non-regular), hence it is not in $K_1 = REG$.

However,

$$L_{a,b} \cup L_{a,b}L_{a,b} = \{a,b\}^+.$$

The inclusion $\subseteq$ is obvious. Conversely, if $x \in \{a,b\}^+$, $|x|_a \neq |x|_b$, then $x \in L_{a,b}$. If $|x|_a = |x|_b$, then either $x = ax'$, $|x'|_a < |x'|_b$ or $x = bx'$, $|x'|_a > |x'|_b$. In both cases $x' \in L_{a,b}$, and $a, b \in L_{a,b}$, therefore $x \in L_{a,b}L_{a,b}$.

On the other hand, $L_{a,b} \in CF$. Indeed, consider the context-free grammar

$$G = (\{S, A, B\}, \{a,b\}, S, P),$$

with P containing the following rules:

$$S \to AaA, S \to BbB,$$

$$A \to AA, A \to a, A \to \lambda, A \to aAb, A \to bAa,$$
$$B \to BB, B \to b, B \to \lambda, B \to aBb, B \to bBa.$$

Clearly, starting by $S \to AaA$ we generate strings x with $|x|_a > |x|_b$ and starting by $S \to BbB$ we obtain strings x with $|x|_a < |x|_b$ (from A one generates all the strings x with $|x|_a \geq |x|_b$ and from B one generates all the strings x with $|x|_a \leq |x|_b$).

(2) $n \geq 2$.
Consider the language

$$L_n = L_{a,b} \cup L_{a,b}\{c\}L_{a,b} \cup M_n,$$

for

$$M_n = \{x \in \{a,b,c\}^* \mid |x|_c \geq n\}.$$

Clearly, $L_n \in CF$, but

$$L_n \cap \{a,b\}^* = L_{a,b},$$

hence $L_n \notin REG$. In fact, for all $k, 1 \leq k \leq n$, we have

$$\begin{aligned}\bigcup_{i=1}^{k} L_n \quad \cap \quad & \{x \in \{a,b,c\}^* \mid |x|_c = k-1\} = \\ = \quad & \{x_1 c x_2 c \ldots c x_k x_{k+1} \mid x_i \in \{a,b\}^+, 1 \leq i \leq k+1, \\ & |x_j| \geq 2, 2 \leq j \leq k, \text{ and } x_1 \in L_{a,b}, \text{ or } x_{k+1} \in L_{a,b}\}.\end{aligned}$$

Denote this language by H. Indeed, $k - 1 < n$, hence $H \cap M_n^* = \emptyset$; it follows that

$$H \subseteq \bigcup (L_{a,b}\{c\}L_{a,b})^i L_{a,b}(L_{a,b}\{c\}L_{a,b})^j,$$

the union being taken for all $i, j \geq 0$ with $i + j = k - 1$.

The language H is not regular: $z_i = a^i caacaa \ldots aaca^i \in H$ for all $i \geq 1$, but every two strings z_i, z_j with $i \neq j$ are not congruent (the context (b^i, b^i) accepts only z_j).

However,

$$\begin{aligned}\bigcup_{i=1}^{n+1} L_n^i &= \{a,b\}^+ \cup M_n \cup \\ &\cup \ \{x_1 c x_2 c \ldots x_r c x_{r+1} \mid 1 \leq r \leq n-1, x_i \in \{a,b\}^+, \\ &\quad 1 \leq i \leq r+1, |x_j| \geq 2, 2 \leq j \leq n\},\end{aligned}$$

hence this language is regular.

The inclusion $\subseteq$ is obvious (note that $M_n^+ = M_n$). Conversely, $M_n \subseteq L_n$, $\{a,b\}^+ = L_{a,b} \cup L_{a,b}L_{a,b}$, and $x_1 c x_2 c \ldots x_r c x_{r+1} \in L_{a,b}(L_{a,b}\{c\}L_{a,b})^r L_{a,b}$ for all $1 \leq r \leq n-1$, $x_i \in \{a,b\}^+$, $1 \leq i \leq r+1$, $|x_j| \geq 2, 2 \leq j \leq r$. (The details are the same as in the first part of the proof.)

In conclusion, $L_n \in K_{n+1} - K_n$ and the proof is complete. □

Theorem 3.10 *$K_n \subset K$ for all $n \geq 1$.*

Proof. The language

$$L = \{a^n b^n \mid n \geq 1\} \cup \{a, b\}$$

is in K but $L \notin K_n$ for $n \geq 1$. Indeed, suppose that $\bigcup_{i=1}^{m} L^i$ is regular for some m. We have

$$\bigcup_{i=1}^{m} L^i \cap a^*b^* = \{x \in a^*b^* \mid -m \leq |x|_a - |x|_b \leq m\},$$

and this is not a regular language, a contradiction. □

The family CL is quite comprehensive and, in fact, the condition $R \in REG$ in its definition can be removed:

Theorem 3.11 *Assume that $L_1 \neq \emptyset$ and L_2 are arbitrary languages over the alphabet V such that $L_1 L_2 \in REG$. Then also $V^* L_2 \in REG$.*

Proof. Let $x \in L_1$ be a string such that the conditions

$$y \in L_1, \ |y| < |x|,$$

are satisfied for no string y. Since $L_1 L_2$ is regular, so is the left derivative

$$L_0 = d_x^l(L_1 L_2)$$

and, hence, also $V^* L_0$ is regular. Since x is shortest in L_1, we have also

$$L_0 = (d_x^l(L_1))L_2.$$

Hence,

$$V^* L_0 = (V^* d_x^l(L_1))L_2 \subseteq V^* L_2.$$

But $L_2 \subseteq L_0$ because $\lambda \in d_x^l(L_1)$. Consequently, $V^* L_2 \subseteq V^* L_0$, which implies that $V^* L_0 = V^* L_2$. Since $V^* L_0$ is regular, so is $V^* L_2$. □

Using right derivatives, it can be shown similarly that if $L_1L_2 \in REG$ and $L_2 \neq \emptyset$, then $L_1V^* \in REG$.

Remark 1. The proof is effective if one of the shortest strings in L_1 can be effectively found. This is the case when, for instance, L_1 is a context-free language.

Corollary 3.12 *$K \subset CL$, strict inclusion.*

Proof. Take $L \subseteq V^*, L \in K$. Therefore $L^* \in REG$. This implies $L^+ = L^* - \{\lambda\} \in REG$, too. Moreover, $L^+ = L^*L$.

According to the previous theorem, $L^*L \in REG$ implies $V^*L \in REG$, hence $L \in CL$ and we have obtained the inclusion $K \subseteq CL$.

This inclusion is proper. For instance, the language L_1 considered in Section 2 is in $CL - K$. Indeed, $L_1^* \cap a^*b^* = L_1$, which is not regular, hence L_1^* is not regular. □

Corollary 3.13 *A context-free language $L \subseteq V^*$ is in CL if and only if $V^*L \in REG$.*

This corollary is useful in showing that languages are not in CL, for instance, in the proof of Theorem 8.

Remark 2. The generality of this result (L_1, L_2 are arbitrary languages) can be compared with the known result (see [5], page 50) that the left quotient of a regular language by an arbitrary language is a regular language, as well as with Lemma 3.1 in [6], which states that also deleting from the strings of a regular language substrings which belong to an arbitrary language, we still obtain a regular language. The previous theorem is in some sense a dual to these results.

A sort of converse of Theorem 5 is natural to be looked for, namely given L_1L_2 regular, it is expected that for any $x \in L_1$, also $(L_1 - \{x\})L_2$ is regular. However, this is not true.

Theorem 3.14 *There are $L_1, L_2 \subseteq \{a, b\}^*, L_1$ linear, L_2 regular, and $x \in L_1$, such that L_1L_2 is regular, but $(L_1 - \{x\})L_2$ is not regular.*

Proof. Consider the language

$$L_1 = \{a^iba^j \mid 1 \leq i < j\} \cup \{a\}.$$

It is clearly linear and

$$\begin{aligned} L_1^* &= \{a^{i_1}ba^{i_2}b\dots a^{i_k}ba^{i_{k+1}} \mid k \geq 1, i_1 \geq 1, \\ &\qquad i_s \geq 3, 1 \leq s \leq k, i_{k+1} \geq 2\} \cup a^*. \end{aligned}$$

Consequently, $L_1^* \in REG$. We take $L_2 = L_1^*$. Obviously, $L_1L_2 = L_1^+$ is regular, too. However,

$$(L_1 - \{a\})L_2 \cap a^*ba^* = \{a^iba^j \mid 1 \leq i < j\},$$

which is not a regular language, hence $(L_1 - \{a\})L_2$ is not regular. □

The next theorem will give a characterization of languages in the family K. With this aim, the notion of *root* of a language in the sense of [1] is used (see also [8], pages 126 - 127).

Given a language $L \subseteq V^*$, we denote by $root(L)$ the smallest language $L_0 \subseteq L$ such that $L_0^* = L^*$; it is proved in [1] that such a language exists and it is unique.

Theorem 3.15 *A language $L \in CF$ is in K if and only if there is a regular language $L_0 \subseteq L$ such that $L \subseteq L_0^*$.*

Proof. The *if* part is obvious ($L_0 \subseteq L \subseteq L_0^*$, hence $L^* = L_0^* \in REG$).

Conversely, we have $root(L) = root(L^*)$. For all regular language, M, $root(M)$ is regular, too [1]. Therefore, for $L \in K$, $root(L^*) \in REG$. Thus, we can take $L_0 = root(L) = root(L^*)$, and all conditions in the theorem are satisfied. □

4 Closure and decidability properties

The families $CL, K, K_n, n \geq 2$, have rather poor closure properties.

Theorem 4.1 *The family CL is closed under morphisms and $Pref, Suf, Sub$, but it is not closed under union, concatenation, Kleene +, intersection by regular sets, inverse morphisms and mirror image.*

Proof.

Morphisms. If $L \in CL, L \subseteq V^*$ and $h : V^* \longrightarrow U^*$, then let $R \in REG$ be such that $RL \in REG$. As $h(RL) = h(R)h(L)$, we have $h(RL) \in REG$, hence $h(L) \in CL$.

Pref, Suf, Sub. As a consequence of Lemma 3 (i), if by an operation α, from a language L we obtain $\alpha(L)$ containing the empty string, then $\alpha(L) \in CL$. This is the case with $Pref, Sub, Suf$.

Union. Consider the languages

$$L_1 = \{a^n b^m \mid 0 \leq n \leq m\},$$

$$L_2 = \{c^n d^m \mid 0 \leq n \leq m\},$$

which are both in CL (take $R_1 = a^*, R_2 = c^*$). Since $\{a,b,c,d\}^*(L_1 \cup L_2)$ is not regular, we conclude by Corollary 2 of Theorem 5 that $L_1 \cup L_2 \notin CL$.

Concatenation. The languages

$$L_1 = \{b\},$$
$$L_2 = \{a^n b^m \mid 0 \leq n \leq m\},$$

are in CL, but L_1L_2 is not in CL, again by Corollary 2 of Theorem 5.

Kleene +. For the previous language L_2 we have $L_2^* \notin CL$ (indeed, $L_2^* \cap a^+b^+ = L_2$).

Intersection by regular sets. As we have seen, D, the Dyck language over $\{a, b\}$, is in CL, but

$$D \cap a^+b^+ = \{a^n b^n \mid n \geq 1\},$$

which is not in CL.

Inverse morphisms. Take the language

$$L = \{(baa)^n(ab)^m \mid 0 \leq n \leq m\}.$$

It belongs to CL. Consider also the morphism

$$h : \{a,b,c,d,e,f\}^* \longrightarrow \{a,b\}^*$$

defined by

$$h(a) = baa, h(b) = ab, h(c) = b, h(d) = aab, h(e) = aaa, h(f) = ba.$$

We obtain

$$\begin{aligned} h^{-1}(L) &= \{a^n b^m \mid 0 \leq n \leq m\} \cup \\ &\cup \{a^r c d^n e f^m c b^p \mid r, p \geq 0, 0 \leq r + n \leq m + p\} \cup \\ &\cup \{a^r f b d^n e f^m c b^p \mid r, p \geq 0, 0 \leq r + n + 1 \leq m + p\}. \end{aligned}$$

Again Corollary 2 of Theorem 5 shows that $h^{-1}(L) \notin CL$.

Mirror image. The language $\{a^n b^m \mid 0 \leq n \leq m\}$ is in CL, but its mirror image is not. □

Theorem 4.2 *The family K is closed under union, Kleene $*$ and morphisms, but it is not closed under concatenation, intersection by regular sets and inverse morphisms.*

Proof. The positive results follow from the next equalities:

$(L_1 \cup L_2)^* = (L_1^* \cup L_2^*)^*$ (union),

$(L^*)^* = L^*$ (Kleene closure),

$(h(L))^* = h(L^*)$ (morphisms).

Concatenation. Take the languages

$$L_1 = \{a^n b^n \mid n \geq 1\} \cup \{a, b\},$$
$$L_2 = \{c\},$$

both in K. However, $L_1 L_2 \notin K$, because

$$(L_1 L_2)^* \cap a^+ b^+ c = \{a^n b^n c \mid n \geq 1\},$$

a non-regular language.

Intersection by regular sets. For L_1 as above we have

$$L_1 \cap a^+ b^+ = \{a^n b^n \mid n \geq 1\},$$

which is not in K.

Inverse morphisms. Consider the language

$$L = \{a^{2n} b^{2n} \mid n \geq 1\} \cup \{a, b\},$$

which is in K, and the morphism $h : \{a, b\}^* \longrightarrow \{a, b\}^*$ defined by

$$h(a) = aa, h(b) = bb.$$

We have

$$h^{-1}(L) = \{a^n b^n \mid n \geq 1\},$$

which we have seen is not in K □.

Theorem 4.3 *The families $K_n, n \geq 2$, are closed under morphisms and Kleene $*$, but they are not closed under union, concatenation, intersection by regular sets and inverse morphisms.*

Proof.

Morphisms. Use the equality $h(\bigcup_{i=1}^{m} L^i) = \bigcup_{i=1}^{m} h(L^i), m \geq 1$.

Kleene $*$. Follows from the inclusion $K_n \subseteq K, n \geq 1$.

Union. Take

$$L_1 = \{a^s b a^t \mid s \neq t, s, t \geq 1\} \cup a^*,$$
$$L_2 = \{b^2\}.$$

We have

$$\begin{aligned} L_1 \cup L_1 L_1 \quad = \quad & \{a^s b a^t \mid s, t \geq 1\} \cup a^* \cup \\ & \cup \{a^s b a^t b a^r \mid s, r \geq 1, t \geq 2, \\ & \quad (s,t,r) \notin \{(1,2,1),(1,2,2),(2,2,1),(1,3,1),(2,3,2)\}\}, \end{aligned}$$

hence $L_1 \in K_2$; clearly, $L_2 \in K_1$. However, $L_1 \cup L_2 \notin K_n$, for all given n. Indeed, assume

$$L = \bigcup_{i=1}^{m} (L_1 \cup L_2)^i \in REG,$$

for some m. If $m = 2k, k \geq 1$, then we have

$$L \cap (a^* b a^* b^2)^k = \{a^s b a^t b^2 \mid s \neq t, s, t \geq 1\}^k,$$

which is not regular. If $m = 2k + 1, k \geq 1$, then

$$L \cap (a^* b a^* b^2)^k a^* b a^* = \{a^s b a^t b^2 \mid s \neq t, s, t \geq 1\}^k \{a^s b a^t \mid s \neq t, s, t \geq 1\},$$

which is non-regular, too.

Concatenation. For the above languages L_1, L_2, take $L_1 L_2$, then follow an argument similar as for union.

Intersection with regular sets. Take again L_1 and intersect it by $a^* b a^*$. We have

$$(\bigcup_{i=1}^{m} (L_1 \cap a^* b a^*)^i) \cap a^* b a^* = \{a^s b a^t \mid s \neq t, s, t \geq 1\},$$

which is not regular.

Inverse morphisms. Consider the language

$$L = \{(ab)^s b (ab)^t \mid s \neq t, s, t \geq 1\} \cup (ab)^*$$

and the morphism $h : \{a, b, c, d\}^* \longrightarrow \{a, b\}^*$, defined by

$$h(a) = a, h(b) = ba, h(c) = bba, h(d) = b.$$

As for L_1, we have $L \in K_2$. Clearly,

$$h^{-1}(L) = \{ab^{s-1} c b^{t-1} d \mid s \neq t, s, t \geq 1\} \cup \{ab^r d \mid r \geq 0\},$$

hence, for all $m \geq 1$,

$$(\bigcup_{i=1}^{m} h^{-1}(L)^i) \cap ab^* c b^* d = \{ab^{s-1} c b^{t-1} d \mid s \neq t, s, t \geq 1\},$$

which is not regular, hence $h^{-1}(L) \notin K_n$, for $n \geq 2$. □

Corollary 4.4 *No family $CL, CR, K, K_n, n \geq 2$, is an AFL or an anti-AFL.*

The following undecidability result is somewhat expected.

Corollary 4.5 *It is undecidable whether or not an arbitrarily given context-free language over an alphabet with at least two symbols is in CL (in K or in $K_n, n \geq 1$).*

Proof. Take $L \subseteq \{a,b\}^*$ arbitrary in CF and the morphism $h : \{a,b\}^* \longrightarrow \{a,b\}^*$, defined by

$$h(a) = bab, h(b) = baab.$$

Since $L = h^{-1}(h(L))$, the language $h(L)$ is regular iff L is regular.

We construct the language

$$L' = \{ba^3b\}h(L).$$

Then, $L' \in CL$ (and $L' \in K, L' \in K_n, n \geq 1$, respectively) iff L is regular (which is undecidable).

Indeed,

1. $\{a,b\}^*L' \in REG$ if and only if $L \in REG$.
   - (if) Obvious.
   - (only if) We have

$$L = h^{-1}(d^l_{ba^3b}(Suf(\{a,b\}^*L') \cap \{ba^3b\}\{a,b\}^*)).$$

2. $\bigcup_{i=1}^{n} L'^i \in REG$ if and only if $L \in REG$, for all $n = 2, 3, \ldots, \infty$.
   - (if) Obvious.
   - (only if) We have $L = h^{-1}(d^l_{ba^3b}(\bigcup_{i=1}^{n} L'^i \cap \{ba^3b\}\{a,b\}^*)), n \geq 2.$ □

References

[1] J. A. Brzozowski, Roots of star events, *Journal of the ACM*, 14 (1967), 466 - 477.

[2] W. Bucher, A. Ehrenfeucht, D. Haussler, On total regulators generated by derivation relations, *Theor. Computer Sci.*, 40 (1985), 131 - 148.

[3] J. Dassow, Gh. Paun, On the degree of non-regularity of context-free languages, *Intern. J. Computer Math.*, 36 (1990), 13 - 29.

[4] A. Ehrenfeucht, D. Haussler, G. Rozenberg, On regularity of context-free languages, *Theor. Computer Sci.*, 27 (1983), 311 - 332.

[5] M. A. Harrison, *Introduction to Formal Language Theory*, Addison Wesley, Reading, Mass., 1978.

[6] L. Kari, *On Insertion and Deletion in Formal Language Theory*, Ph.D. Thesis, Univ. of Turku, Dept. of Mathematics, 1991.

[7] E. Makinen, Two complexity measures for context-free languages, *Intern. J. Computer Math.*, 26 (1988), 29 - 34.

[8] A. Salomaa, *Theory of Automata*, Pergamon Press, New York, 1969.

[9] A. Salomaa, *Formal Languages*, Academic Press, New York, London, 1973.

Received December 3, 1992