Acta Informatica (2021) 58:281–299
https://doi.org/10.1007/s00236-020-00386-3

ORIGINAL ARTICLE

Operational complexity and right linear grammars

Jürgen Dassow[1]

Received: 28 January 2020 / Accepted: 7 July 2020

Abstract

For a regular language L, let $\mathrm{Var}(L)$ be the minimal number of nonterminals necessary to generate L by right linear grammars. Moreover, for natural numbers $k_1, k_2, \ldots, k_n$ and an n-ary regularity preserving operation f, let $g_f^{\mathrm{Var}}(k_1, k_2, \ldots, k_n)$ be the set of all numbers k such that there are regular languages $L_1, L_2, \ldots, L_n$ such that $\mathrm{Var}(L_i) = k_i$ for $1 \leq i \leq n$ and $\mathrm{Var}(f(L_1, L_2, \ldots, L_n)) = k$. We completely determine the sets g_f^{Var} for the operations reversal, Kleene-closures $+$ and $*$, and union; and we give partial results for product and intersection.

Mathematics Subject Classification 68Q05 · 68Q42 · 68Q45

1 Introduction and definitions

In the last 30 years, the following problem was studied very intensively: How behaves the complexity of languages under operations (on languages). More precisely, for a complexity measure K on regular languages and a regularity preserving n-ary function f on languages, one is interested in the set $g_f^K(k_1, k_2, \ldots, k_n)$ of all numbers k such that there are regular languages $L_1, L_2, \ldots, L_n$ such that $K(L_i) = k_i$ for $1 \leq i \leq n$ and $K(f(L_1, L_2, \ldots, L_n)) = k$ (in many cases only the maximal number $h_f^K(k_1, k_2, \ldots, k_n)$ in $g_f^K(k_1, k_2, \ldots, k_n)$ is considered).

Most of the research concerns the state complexity sc, where $\mathrm{sc}(L)$ is defined as the minimal number of states of a deterministic finite automata which accepts L. As an example we mention

$$g_{\cup}^{\mathrm{sc}}(m, n) = \{1, 2, \ldots, mn\} \quad \text{for } m \geq 2 \text{ and } n \geq 2$$

(see [12]). For further results concerning the behaviour of the state complexity under operations, we refer to the survey article [5].

Because the family of regular languages can be characterized by nondeterministic finite automata, by incomplete deterministic finite automata, regular expressions etc., it is natural to study the set g_f^K and the function h_f^K for measures K connected with those characterizations.

✉ Jürgen Dassow
dassow@iws.cs.uni-magdeburg.de

1 Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg, PSF 4120, 39016 Magdeburg, Germany

Table 1 Some results on ranges of nondeterministic state complexities for some operations

Union	$g^{\text{nsc}}_{\cup}(n,m) = \{1, 2, \ldots, m+n+1\}$ for $n \geq 2, m \geq 2$	[12]
Intersection	$g^{\text{nsc}}_{\cap}(n,m) = \{1, 2, \ldots, m \cdot n\}$ for $n \geq 1, m \geq 1$	[12]
Concatenation	$g^{\text{nsc}}_{\cdot}(n,m) = \{1, 2, \ldots, m+n\}$ for $n \geq 2, m \geq 2$	[13]
Reversal	$g^{\text{nsc}}_{R}(n) = \{n-1, n, n+1\}$ for $n \geq 2$	[14]
Kleene-closure	$g^{\text{nsc}}_{*}(n) = \{1, 2, \ldots, n+1\}$ for $n \geq 2$	[14]

The state complexity of a regular language coincides with the number of its left quotients by words. However, using the quotients instead of states gives some further interesting aspects. The behaviour of the quotient complexity under operations was intensively studied by Brzozowski and others. A summary is given in [1].

In 2003, Holzer and Kutrib started the investigation with respect to the nondeterministic state complexity nsc where $\text{nsc}(L)$ is defined as the minimal number of states of a nondeterministic finite automata which accepts L [10,11]. We mention here the facts given in Table 1.

A complete deterministic automaton with r states and s input symbols has $r \cdot s$ transitions. However, often it is not necessary to require completeness, i.e., it can be that, for some pairs of states and input symbols, no transition is defined. For such incomplete automata, one can study the transition complexity tc where $tc(L)$ is the minimal number of transitions in incomplete deterministic finite automata which accepts L. The behaviour of this measure was studied in [6,16]. For instance, in [6], it was shown that for all numbers $m \geq 1$ and $n \geq 1$,

$$m \cdot n + m + n - 1 \leq h^{\text{tc}}_{\cup}(m,n) \leq 2(m \cdot n + m + n).$$

The behaviour of the size of regular expressions was studied in [4,7].

We now introduce a further concept which characterizes regular languages. We start with some notation.

For a set M, by $\text{card}(M)$ we denote its cardinality. An alphabet is a non-empty finite set. The elements of an alphabet are called letters. A word over an alphabet V is a finite sequence of letters of V. The length of a word is the number of its letters (where we count the letters as often as they occur in the word); the length of a word w is denoted by $|w|$. A word w of length $n \geq 1$ is written as $w = a_1 a_2 \ldots a_n$, where $a_i \in V$ for $1 \leq i \leq n$. The empty word (of length 0) is designated by λ. The set of all non-empty words of V is denoted by V^+, and we set $V^* = V^+ \cup \{\lambda\}$. The reversal of a non-empty word $w = a_1 a_2 \ldots a_n$ is defined by $w^R = a_n a_{n-1} \ldots a_1$; moreover, we set $\lambda^R = \lambda$.

Any subset of V^* is called a language over V. For two languages L and L', we define the concatenation by $LL' = \{ww' \mid w \in L, w' \in L'\}$, the powers of L by $L^0 = \{\lambda\}$, $L^n = L^{n-1}L$ for $n \geq 1$, the positive Kleene-closure, the Kleene-closure and the reversal by

$$L^+ = \bigcup_{n \geq 1} L^n, \quad L^* = \bigcup_{n \geq 0} L^n \quad \text{and} \quad L^R = \{w^R \mid w \in L\}.$$

A right linear grammar is a quadruple $G = (N, T, P, S)$ where N and T are two disjoint (non-empty) alphabets, P is a finite subset of $N \times (T^*N \cup T^*)$, and $S \in N$. The elements of N and T are called nonterminals and terminals, respectively. The elements of P are called rules. For a rule (A, w), we mostly use the notation $A \to w$. S is called the axiom or the start symbol.

Table 2 Some results on ranges of Varcf-complexities for some operations

Union	$g_{\cup}^{\mathrm{Varcf}}(n,m) = \{1, 2, \ldots, m+n+1\}$ for $n \geq 1, m \geq 1$
concatenation	$\{1\} \cup \{\max\{n,m\}, \ldots, m+n+1\} \subseteq g_{\cdot}^{\mathrm{Varcf}}(n,m)$ for $n \geq 1, m \geq 1$
Reversal	$g_{R}^{\mathrm{Varcf}}(n) = \{n\}$ for $n \geq 1$
Kleene-closure*	$g_{*}^{\mathrm{Varcf}}(n) = \{1, 2, \ldots, n+1\}$ for $n \geq 1$

The result for reversal is not given in [3], but follows from the following fact: If L is generated by $G = (N, T, P, S)$, then L^R is generated by $G^R = (N, T, \{A \to w^R \mid A \to w \in P\}, S)$

For two words $x \in T^*N$ and $y \in T^*N \cup T^*$, we write $x \Longrightarrow y$ if and only if there is a rule $A \to w$ in P such that $x = vA$ and $y = vw$, and then we say that x derives y in one derivation step. Let $\Longrightarrow^*$ be the reflexive and transitive closure of $\Longrightarrow$. If $x \Longrightarrow^* y$, we say that y is derived from x.

The language $L(G)$ generated by a right linear grammar $G = (N, T, P, S)$ is defined by

$$L(G) = \{z \mid z \in T^*,\ S \Longrightarrow^* z\}.$$

It is known that a language L is regular if and only if it is generated by a right linear grammar.

For a right linear grammar $G = (N, T, P, S)$, we define the complexity measure Var and extend it for regular languages by setting

$$\mathrm{Var}(G) = \mathrm{card}(N)$$

and

$$\mathrm{Var}(L) = \min\{\mathrm{Var}(G) \mid G \text{ is a right linear grammar with } L(G) = L\}.$$

A right linear grammar $G = (N, T, P, S)$ is called to be in *normal form*, if all rules of P are of the form $A \to aB$ or $A \to \lambda$ where A and B are nonterminals and a is a terminal. Let Varnf be the complexity if one restricts to right linear grammars in normal form. Then it is known that $\mathrm{Varnf}(L) = \mathrm{nsc}(L)$. However, the complexities of concrete languages differ essentially. For instance, it is obvious that $\mathrm{nsc}(\{a^m\}^*) = m$ and $\mathrm{Var}(\{a^m\}^*) = 1$.

Obviously, context-free grammars (where the rules have the form $A \to w \in (N \cup T)^*$ with $A \in N$ and $w \in (N \cup T)^*$) are a generalization of right linear grammars, and one can define the Var-complexity for context-free grammars, which is denoted by Varcf. The behaviour of this measure under operations was studied in [3]. Some results are presented in Table 2 (there are no results for intersection and complement because the class of context-free languages is not closed under these two operations).

We mention that, again, the Var-complexities of languages can considerably differ. For instance, the set $U = \{aba\}^+\{ab^2a\}^+ \cdots \{ab^{2m}a\}^+$ has the complexities $\mathrm{Varcf}(U) = m$ (by Lemma 2.3 in [3]) and $\mathrm{Var}(U) = 2m$ (by Lemma 1 below).

In this paper we continue the research mentioned above by a (partial) determination of the sets g_f^{Var} with respect to the operations union, concatenation, Kleene-closure, intersection, and reversal.

By the above remarks, the study of the sets g_f^{Var} is an intermediate step between the investigation of g_f^{nsc} and the study of g_f^{Varcf}. Therefore our results sometimes look similar to those of [3,12–14]. In some cases, below, we can use modifications of the proofs in those papers. However, by the above mentioned differences in the complexities for certain languages, mostly, we cannot follow the ideas of that papers. For instance, to prove the first line of Table 1 the authors only use finite and unary languages. However, such languages cannot be used to prove an analogous statement for the measure Var because $\mathrm{Var}(L) \leq 2$ holds for any finite and/or unary language L.

2 The complexity of some special languages

In this section we determine the complexities of some languages, which are used later.

For two different letters a and b, a natural number p with $p \geq 1$, and pairwise different natural numbers $r_1, r_2, \ldots, r_p$ with $1 \leq r_i$ for $1 \leq i \leq p$, we set

$$P(r_1, r_2, \ldots, r_p) = \{ab^{r_1}a\}^+ \cdot \{ab^{r_2}a\}^+ \ldots \{ab^{r_p}a\}^+,$$
$$P'(r_1, r_2, \ldots, r_p) = \{ab^{r_1}a\}^* \cdot \{ab^{r_2}a\}^* \ldots \{ab^{r_p}a\}^*,$$
$$U(r_1, r_2, \ldots, r_p) = \{ab^{r_1}a, ab^{r_2}a, \ldots, ab^{r_p}a\}^+.$$

Lemma 1 *For two different letters a and b, natural numbers $n \geq 1$, $p_1, p_2, \ldots, p_n$, $p_i \geq 1$ for $1 \leq i \leq n$, and pairwise different natural numbers $r_{1,1}, r_{1,2}, \ldots, r_{1,p_1}, r_{2,1}, r_{2,2}, \ldots, r_{2,p_2}, \ldots, r_{n,p_n}$, $r_{i,j} \geq 1$ for $1 \leq i \leq n$, $1 \leq j \leq p_i$, let*

$$L = P(r_{1,1}, \ldots, r_{1,p_1}) \cup P(r_{2,1}, \ldots, r_{2,p_2}) \cup \ldots \cup P(r_{n,1}, \ldots, r_{n,p_n}).$$

Then we have

$$Var(L) = \begin{cases} p_1 + p_2 + \cdots + p_n + 1 & \text{for } n \geq 2 \\ p_1 & \text{for } n = 1 \end{cases}$$

and

$$\mathrm{Var}(P'(r_{1,1}, r_{1,2}, \ldots, r_{1,p_1})) = p_1.$$

Proof Let $G = (N, \{a, b\}, P, S)$ be a right linear grammar such that

$$L(G) = L \quad \text{and} \quad \mathrm{Var}(G) = \mathrm{card}(N) = \mathrm{Var}(L).$$

Further, let $t = \mathrm{card}(N)$ and $t' = \max\{|z| \mid A \to z \in P\}$.

For i and j, $1 \leq i \leq n$, $1 \leq j \leq p_i$, we consider the word

$$z_{i,j} = ab^{r_{i,1}}aab^{r_{i,2}}a \ldots ab^{r_{i,j-1}}a(ab^{r_{i,j}}a)^{(t+1)(5r_{i,j}+t')}ab^{r_{i,j+1}}aab^{r_{i,j+2}}a \ldots ab^{r_{i,p_i}}a$$

and its derivation. There are nonterminals $A_1, A_2, \ldots, A_{t+1}$ such that

$$S \Longrightarrow^* w_i A_1 \Longrightarrow^* w_i v_{i,1} A_2 \Longrightarrow^* w_i v_{i,1} v_{i,2} A_3 \Longrightarrow^* \ldots$$
$$\Longrightarrow^* w_i v_{i,1} v_{i,2} \ldots v_{i,t-1} A_t \Longrightarrow^* w_i v_{i,1} v_{i,2} \ldots v_{i,t-1} v_{i,t} A_{t+1} \Longrightarrow^* z_i$$

with $w_i = ab^{r_{i,1}}aab^{r_{i,2}}a \ldots ab^{r_{i,j-1}}aw'_i$, $|w'_i| \leq t'$, $2r_{i,j} + 3 \leq |v_{i,l}| \leq 2r_{i,j} + 3 + t'$ for $1 \leq l \leq t$. Because $|w'_i v_{i,1} v_{i,2} \ldots v_{i,t-1} v_{i,t}| \leq t' + (2r_{i,j} + 3 + t')t \leq (5r_{i,j} + t')(t+1)$, we get $w'_i v_{i,1} v_{i,2} \ldots v_{i,t-1} v_{i,t} = (ab^{r_{i,j}}a)^p w''$ with some natural number p and a proper prefix w'' of $ab^{r_{i,j}}a$. By the length condition for the words $v_{i,l}$, we obtain that $v_{i,l}$ contains at least one occurrence of $ab^{r_{i,j}}a$ as a subword.

Since N contains only t nonterminals, it follows that there are two numbers l_1 and l_2 such that $A_{l_1} = A_{l_2}$. We set $B_{i,j} = A_{l_1} = A_{l_2}$. Then the derivation

$$S \Longrightarrow^* y_{i,j} A_{l_1} \Longrightarrow^* y_{i,j} x_{i,j} A_{l_2} \Longrightarrow^* y_{i,j} x_{i,j} y'_{i,j} \in L(G)$$

can be written as

$$S \Longrightarrow^* y_{i,j}B_{i,j} \Longrightarrow^* y_{i,j}x_{i,j}B_{i,j} \Longrightarrow^* y_{i,j}x_{i,j}y'_{i,j} \in L(G)$$

and by construction $x_{i,j} = x'_{i,j}(ab^{r_{i,j}}a)^{p(i,j)}x''_{i,j}$ for some natural number $p(i,j) \geq 1$, a proper suffix $x'_{(i,j)}$ of $ab^{r_{i,j}}a$, and a proper prefix $x''_{(i,j)}$ of $ab^{r_{i,j}}a$.

If $B_{i,j} = B_{i,j'}$ for some $1 \leq j < j' \leq p_i$, then we have a derivation

$$\begin{aligned} S &\Longrightarrow^* y_{i,j'}B_{i,j'} \Longrightarrow^* y_{i,j'}x'_{i,j'}(ab^{r_{i,j'}}a)^{p(i,j')}x''_{i,j'}B_{i,j'} \\ &= \; y_{i,j'}x'_{i,j'}(ab^{r_{i,j'}}a)^{p(i,j')}x''_{i,j'}B_{i,j} \\ &\Longrightarrow^* y_{i,j'}x'_{i,j'}(ab^{r_{i,j'}}a)^{p(i,j')}x''_{i,j'}x'_{i,j}(ab^{r_{i,j}}a)^{p(i,j)}x''_{i,j}B_{i,j} \\ &\Longrightarrow^* y_{i,j'}x'_{i,j'}(ab^{r_{i,j'}}a)^{p(i,j')}x''_{i,j'}x'_{i,j}(ab^{r_{i,j}}a)^{p(i,j)}x''_{i,j}y'_{i,j} = z. \end{aligned}$$

By construction, $z \in L(G) = L$. But z contains an occurrence of $ab^{r_{i,j}}a$ after an occurrence of $ab^{r_{i,j'}}a$ with $j < j'$ which contradicts the structure of the words in L (more precisely, in $P(r_{i,1}, r_{i,2}, \ldots, p_{i,p_i})$).

If $B_{i,j} = B_{i',j'}$ for some $1 \leq i < i' \leq n$, $1 \leq j \leq p_i$, and $1 \leq j' \leq p_{i'}$, then we can analogously show that a word is generated which contains an occurrence of $ab^{r_{i,j}}a$ as well as an occurrence of $ab^{r_{i',j'}}a$ which contradicts the structure of words in L, again.

Thus N contains at least the $p_1 + p_2 + \cdots + p_n$ different letters $B_{i,j}$, $1 \leq i \leq n$, $1 \leq j \leq p_i$.

Moreover, if $n \geq 2$ and $B_{i,j}$ is the axiom for certain $1 \leq i \leq n$, $1 \leq j \leq p_i$, then we can as above produce a word which contains an occurrence of $ab^{r_{i,j}}a$ as well as an occurrence of $ab^{r_{i',j'}}a$, where $i' \neq i$. Again, we obtain a contradiction to the structure of L. Thus, N contains a start symbol different from all symbols $B_{i,j}$.

Hence we obtain $\text{Var}(L) \geq p_1 + p_2 + \cdots + p_n + 1$ for $n \geq 2$ and $\text{Var}(L) \geq p_1$ for $n = 1$.

On the other hand, the right linear grammars $G_1 = (N_1, \{a,b\}, P_1, C_1)$ and $G_{\geq 2} = (N_{\geq 2}, \{a,b\}, P_{\geq 2}, S)$ with

$$\begin{aligned} N_1 &= \{C_1, C_2, \ldots, C_{p_1}\}, \\ P_1 &= \{C_{p_1} \to ab^{r_{1,p_1}}aC_{p_1}, C_{p_1} \to ab^{r_{1,p_1}}a\} \\ &\quad \cup \bigcup_{i=1}^{p_1-1} \{C_i \to ab^{r_{1,i}}aC_i, C_i \to ab^{r_{1,i}}aC_{i+1}\}, \\ N_{\geq 2} &= \{S\} \cup \{C_{i,j} \mid 1 \leq i \leq n, 1 \leq j \leq p_i\}, \\ P_{\geq 2} &= \{S \to C_{i,1} \mid 1 \leq i \leq n\} \cup \bigcup_{i=1}^{n} \{C_{i,p_i} \to ab^{r_{i,p_i}}aC_{i,p_i}, C_{i,p_i} \to ab^{r_{i,p_i}}a\} \\ &\quad \cup \bigcup_{i=1}^{n} \bigcup_{j=1}^{p_i-1} \{C_{i,j} \to ab^{r_{i,j}}aC_{i,j}, C_{i,j} \to ab^{r_{i,j}}aC_{i,j+1}\} \end{aligned}$$

generate L for $n = 1$ and $n \geq 2$, respectively. Thus

$$\text{Var}(L) = \begin{cases} p_1 + p_2 + \cdots + p_n + 1 & \text{for } n \geq 2, \\ p_1 & \text{for } n = 1. \end{cases}$$

The proof that we need at least p_1 nonterminals for the generation of the set $P'(r_{1,1}, r_{1,2}, \ldots, r_{1,p_1})$ can be analogously given to the proof of this statement for $P(r_{1,1}, r_{1,2}, \ldots, r_{1,p_1})$. To show that that p_1 nonterminals are sufficient we add the rules

$C_i \to C_{i+1}$ for $1 \le i \le p_1 - 1$ and $C_{p_1} \to \lambda$ to the rules of G_1, which results in a grammar generating $P'(r_{1,1}, r_{1,2}, \ldots, r_{1,p_1})$ with p_1 nonterminals. □

Lemma 2 *For three pairwise different letters a, b and c, natural numbers $p \ge 1$ and $q \ge 1$, and pairwise different numbers $r_1, r_2, \ldots, r_p, s_1, s_2, \ldots, s_q$, $r_i \ge 1$ for $1 \le i \le p$ and $s_j \ge 1$ for $1 \le j \le q$, we have*

(i) $\mathrm{Var}(P(r_1, r_2, \ldots, r_p)\{c\}) = p,$
(ii) $\mathrm{Var}(\{c\}P(r_1, r_2, \ldots, r_p)) = p + 1,$
(iii) $\mathrm{Var}(\{c\}P(r_1, r_2, \ldots, r_p) \cup P(s_1, s_2, \ldots, s_q)) = p + q + 1.$

Proof (i) The proof that we need at least p nonterminals for the generation of $P(r_1, r_2, \ldots, r_p)\{c\}$ can be analogously given to the corresponding statement for the set $P(r_{1,1}, r_{1,2}, \ldots, r_{1,p_1})$ in the proof of Lemma 1. To show that p rules are sufficient we consider the right linear grammar $G_1' = (\{C_1, C_2, \ldots C_p\}, \{a, b, c\}, P_1', C_1)$ with

$$P_1' = \{C_p \to ab^{r_p}aC_p, C_p \to ab^{r_p}ac\} \cup \bigcup_{i=1}^{p-1}\{C_i \to ab^{r_i}aC_i, C_i \to ab^{r_i}aC_{i+1}\},$$

which generates $P(r_1, r_2, \ldots, r_p)\{c\}$ with p nonterminals.

(ii) As in the proof of Lemma 1, we can show the existence of p different nonterminals B_i, $1 \le i \le p$, with derivations

$$B_i \Longrightarrow^* u_i(ab^{r_i}a)^{p_i}v_iB_i \text{ where } p_i \text{ is a natural number,}$$
$$u_i \text{ and } v_i \text{ are a proper suffix and a proper prefix of } ab^{r_i}a, \text{ resp.}$$
$$B_i \Longrightarrow^* z_i \in \{a, b\}^*.$$

If one of these nonterminals is the axiom, say B_i, we have the derivation $B_i \Longrightarrow^* u_i(ab^{r_i}a)^{p_i}v_iB_i \Longrightarrow^* u_iab^{r_i}av_iz_i$ which produces a word which does not start with c. Thus, we need at least $p + 1$ nonterminals.
Moreover, we modify G_1' of part i) to G_1'' adding a new nonterminal C which is the axiom of G_1'', by replacing $C_p \to ab^{r_p}ac$ by $C_p \to ab^{r_p}a$ and adding the rule $C \to cC_1$. Then G_1'' generates $\{c\}P(r_1, r_2, \ldots, r_p)$ with $p + 1$ nonterminals.

(iii) As in the proof of Lemma 1, we can show that we need $p + q$ different nonterminals to generate the words $cab^{r_1}a \ldots ab^{r_{i-1}}a(ab^{r_i}a)^tab^{r_{i+1}}a \ldots ab^{r_p}a$, $1 \le i \le p$, and $ab^{s_1}a \ldots ab^{s_{j-1}}a(ab^{s_j}a)^tab^{s_{j+1}}a \ldots ab^{s_q}a$, $1 \le j \le q$, for sufficiently large t. Furthermore, as in the proof of Lemma 1, we can show that these nonterminals are pairwise different and different from the axiom. Therefore $\mathrm{Var}(\{c\}P(r_1, r_2, \ldots, r_p) \cup P(s_1, s_2, \ldots, s_q)) \ge p + q + 1$.
Because the grammar $G_2' = (\{S, C_1, C_2, \ldots, C_p, D_1, D_2, \ldots, D_q\}, \{a, b, c\}, P_2', S)$ with

$$P_2' = \{S \to cC_1, S \to D_1, C_p \to ab^{r_p}aC_p, C_p \to ab^{r_p}a, D_q \to ab^{s_q}aD_q, D_q \to ab^{s_q}a\}$$
$$\cup \bigcup_{i=1}^{p-1}\{C_i \to ab^{r_i}aC_i, C_i \to ab^{r_i}aC_{i+1}\} \cup \bigcup_{j=1}^{q-1}\{D_j \to ab^{s_j}aD_j, D_j \to ab^{s_j}aD_{j+1}\}$$

generates $\{c\}P(r_1, r_2, \ldots, r_p) \cup P(s_1, s_2, \ldots, s_q)$ with $p + q + 1$ nonterminals, the statement follows. □

Lemma 3 *For pairwise different letters a, b, and c, natural numbers $p \ge 1$, $q \ge 1$, and pairwise different numbers $r_1, r_2, \ldots, r_p, s_1, s_2, \ldots, s_q$, $r_i \ge 1$ for $1 \le i \le p$ and $s_j \ge 1$ for $1 \le j \le q$, we have*

(i) $\text{Var}(P(r_1, r_2, \ldots, r_p) \cup U(s_1, s_2, \ldots, s_q)) = p + 2$,
(ii) $\text{Var}(\{c\}P(r_1, r_2, \ldots, r_p) \cup U(s_1, s_2, \ldots, s_q) = p + 2$,
(iii) $\text{Var}(P'(r_1, r_2, \ldots, r_{p'})U(s_1, s_2, \ldots, s_q)P(r_{p'+1}, r_{p'+2}, \ldots, r_p)) = p+1$, *where* $p > p' \geq 1$ *holds in addition,*
(iv) $\text{Var}(U(s_1, s_2, \ldots, s_q)P(r_1, r_2, \ldots, r_p)) = p + 1$.

Proof (i) As in Lemma 1, we can show that the generation of

$$ab^{r_1}a \ldots ab^{r_{i-1}}a(ab^{r_i}a)^t ab^{r_{i+1}}a \ldots ab^{r_p}a,\ 1 \leq i \leq p \quad \text{and} \quad (ab^{s_1}a)^t$$

for sufficiently large t requires $p + 1$ different nonterminals which are also different from the axiom. Thus $\text{Var}((P(r_1, r_2, \ldots, r_p) \cup U(s_1, s_2, \ldots, s_q)) \geq p + 2$.
On the other hand, the grammar $H = (\{S, C, C_1, C_2, \ldots, C_p\}, \{a, b\}, P, S)$ with

$$P = \{S \to C, S \to C_1, C_p \to ab^{r_p}aC_p, C_p \to ab^{r_p}a\} \cup \bigcup_{j=1}^{q}\{C \to ab^{s_j}aC, C \to ab^{s_j}a\} \cup \bigcup_{i=1}^{p-1}\{C_i \to ab^{r_i}aC_i, C_i \to ab^{r_i}aC_{i+1}\}$$

generates $P(r_1, r_2, \ldots, r_p) \cup U(s_1, s_2, \ldots, s_q)$ with $p + 2$ nonterminals. (ii) can be analogously shown. (iii) As in Lemma 1, we can show that the generation of the words

$$(ab^{r_i}a)^t ab^{s_1}aab^{r_{p'+1}}a \ldots ab^{r_p}a \quad \text{for} \quad 1 \leq i \leq p',$$
$$(ab^{s_1}a)^t ab^{r_{p'+1}}a \ldots ab^{r_p}a,$$
$$ab^{s_1}aab^{r_{p'+1}}a \ldots ab^{r_{j-1}}a(ab^{r_j}a)^t ab^{r_{j+1}}a \ldots ab^{r_p}a \quad \text{for} \quad 1 \leq p' < j \leq p$$

for sufficiently large t requires $p + 1$ different nonterminals. Thus

$$\text{Var}(P'(r_1, r_2, \ldots, r_{p'})U(s_1, s_2, \ldots, s_q)P(r_{p'+1}, r_{p'+2}, \ldots, r_p)) \geq p + 1.$$

Since the grammar $H' = (\{C, C_1, C_2, \ldots C_p\}, \{a, b\}, P', C_1)$ with

$$P' = \{C_{p'} \to ab^{r_{p'}}aC_{p'}, C_{p'} \to C, C_p \to ab^{r_p}aC_p, C_p \to ab^{r_p}a\} \cup \bigcup_{i=1}^{p'-1}\{C_i \to ab^{r_i}aC_i, C_i \to C_{i+1}\} \cup \bigcup_{i=p'+1}^{p-1}\{C_i \to ab^{r_i}aC_i, C_i \to ab^{r_i}aC_{i+1}\} \cup \bigcup_{j=1}^{q}\{C \to ab^{s_j}aC, C \to ab^{s_j}aC_{p'+1}\}$$

generates $P'(r_1, r_2, \ldots, r_{p'})U(s_1, s_2, \ldots, s_q)P(r_{p'+1}, r_{p'+2}, \ldots, r_p)$ with $p + 1$ nonterminals, the statement follows. (iv) can be shown analogously. □

Lemma 4 (i) *For different letters a and b, we have* $\text{Var}(\{b\}\{a, b\}^*) = 2$.
(ii) *For different letters a and b, a natural* $p \geq 1$, *and pairwise different natural numbers* $r_1, r_2, \ldots, r_p, r_i \geq 1$ *for* $1 \leq i \leq p$, *we have* $\text{Var}(P(r_1, r_2, \ldots, r_p) \cup \{c\}\{a, b\}^*) = p+2$.

Proof (i) Let $G = (N, \{a, b\}, P, S)$ be a right linear grammar with $L(G) = \{b\}\{a, b\}^*$ and $\text{Var}(G) = \text{Var}(\{b\}\{a, b\}^*)$. Since $ba^t \in L(G)$ for all (sufficiently large) numbers $t \geq 1$, there is a nonterminal $A \in N$ such that $A \Longrightarrow^* a^r A$ for some $r \geq 1$. If $A = S$, then we have a derivation $S = A \Longrightarrow^* a^r A = a^r S \Longrightarrow^* a^r b$ (because $S \Longrightarrow^* b \in L(G) = \{b\}\{a, b\}^*$), which produces a word not in $\{b\}\{a, b\}^*$, which is a contradiction. Hence, $\text{Var}(\{b\}\{a, b\}^*) = \text{Var}(G) \geq 2$.

Because $H = (\{S, S'\}, \{a, b\}, \{S \to bS', S' \to aS', S' \to bS', S' \to \lambda\}, S)$ generates $\{b\}\{a, b\}^*$ with two nonterminals, we obtain $\mathrm{Var}(\{b\}\{a, b\}^*) = 2$.

(ii) can be shown analogously to the proofs of Lemma 3 and part i). □

3 Behaviour under operations

In this section we study the behaviour of the measure Var for regular languages (presented by right linear grammars) under some operations.

3.1 Reversal

We start with the operation reversal.

Theorem 1 *We have*

$$g_R^{\mathrm{Var}}(n) = \begin{cases} \{n-1, n, n+1\} & \text{for } n \geq 2, \\ \{1, 2\} & \text{for } n = 1. \end{cases}$$

Proof (i) We first prove that, for a language L with $\mathrm{Var}(L) = n$, $\mathrm{Var}(L^R) \leq n + 1$ holds. Since $\mathrm{Var}(L) = n$, L is generated by a right linear grammar $G = (N, T, P, S)$ with $\mathrm{card}(N) = n$. We construct the grammar $G^R = (N \cup \{S^R\}, T, P^R, S^R)$ where S^R is a new symbol, i.e., $S^R \notin N$, and P^R consists of all rules constructed as follows:

- if $A \to w \in P$, $w \in T^*$ and $A \in N$, then $S^R \to w^R A \in P^R$,
- if $A \to wB \in P$, $w \in T^*$, $A \in N$, and $B \in N$, then $B \to w^R A \in P^R$,
- if $S \to w \in P$ and $w \in T^*$, then $S^R \to w^R \in P^R$, and
- $S \to \lambda \in P^R$.

By construction, G^R is right linear and satisfies $L(G^R) = L^R$. Moreover, we have $\mathrm{Var}(G^R) = n + 1$. Therefore $\mathrm{Var}(L^R) \leq n + 1$.

(ii) We now prove that, for a language L with $\mathrm{Var}(L) = n$, $\mathrm{Var}(L^R) \geq n - 1$ holds. Assume that there is a language L such that $\mathrm{Var}(L) = n$ and $\mathrm{Var}(L^R) \leq n - 2$. Since $(L^R)^R = L$, we obtain from i) that $\mathrm{Var}(L) = \mathrm{Var}((L^R)^R) \leq \mathrm{Var}(L^R) + 1 \leq n - 1$ in contrast to our supposition.

(iii) We now present witnesses for the values n for $n \geq 1$, $n + 1$ for $n \geq 1$, and $n - 1$ for $n \geq 2$.
Let $n \geq 1$ and $L_n = P(1, 2, \ldots, n)$. Then we obtain $L_n^R = P(n, n-1, \ldots, 1)$. By Lemma 1, we get $\mathrm{Var}(L_n) = \mathrm{Var}(L_n^R) = n$.
Let $n \geq 1$ and $K_n = P(1, 2, \ldots, n)\{c\}$. Then $K_n^R = \{c\}P(n, n-1, \ldots, 1)$ and, by Lemma 2, $\mathrm{Var}(K_n) = n$ and $\mathrm{Var}(K_n^R) = n + 1$.
Let $n \geq 2$ and $M_n = \{c\}P(1, 2, \ldots, n-1)$. Then $M_n^R = P(n-1, n-2, \ldots, 1)\{c\}$. By Lemma 2, we have $\mathrm{Var}(M_n) = n$ and $\mathrm{Var}(M_n^R) = n - 1$. □

3.2 Kleene-closure

First, we consider the positive closure.

Theorem 2 *For $n \geq 1$, we have $g_+^{\mathrm{Var}}(n) = \{1, 2, \ldots, n\}$.*

Proof (i) We show that $\mathrm{Var}(L) = n$ implies $\mathrm{Var}(L^+) \leq n$.

Let L be a language with $\mathrm{Var}(L) = n$. Then there is a right linear grammar $G = (N, T, P, S)$ such that $\mathrm{card}(N) = n$ and $L(G) = L$. We construct the grammar $G' = (N, T, P', S)$ where

$$P' = P \cup \{A \to wS \mid A \to w \in P\}.$$

By construction, after finishing a derivation in G, we can start a new derivation in G'. Thus $L(G') = L^+$. Because $\mathrm{Var}(G') = \mathrm{card}(N) = n$, we obtain $\mathrm{Var}(L^+) \leq n$.

(ii) We now show that $n \in g_+^{\mathrm{Var}}(n)$ for $n \geq 1$. Let $L_n = P(1, 2, \ldots, n)$. By Lemma 1, $\mathrm{Var}(L_n) = n$.
Let $G = (N, \{a, b\}, P, S)$ be a right linear grammar such that $L(G) = L_n^+$. Since $L_n \subseteq L_n^+$, we can prove as in the proof of Lemma 1 that there are n different nonterminals $B_1, B_2, \ldots, B_n$ in N such that there are derivations

$$S \Longrightarrow^* y_i B_i \Longrightarrow^* y_i x_i B_i \Longrightarrow^* y_i x_i y_i'$$

where $x_i = x_i' a b^i a x_i''$ with $x_i', x_i'', y_i, y_i' \in \{a, b\}^*$. Thus we have $\mathrm{Var}(L_n^+) \geq n$. By i) we get $\mathrm{Var}(L_n^+) = n$. Hence $n \in g_+^{\mathrm{Var}}(n)$ for $n \geq 1$.

(iii) We now prove that all values k with $3 \leq k \leq n-1$ are in $g_+^{\mathrm{Var}}(n)$. Let

$$K_{n,k} = P(1) \cup P(2) \cup \ldots \cup P(n-k) \cup P(n-k+1, n-k+2, \ldots, n-1).$$

By Lemma 1, we have $\mathrm{Var}(K_{n,k}) = n$. Moreover, let

$$U = (\{aba, ab^2a, \ldots, ab^{n-k}a\} \cup P(n-k+1, n-k+2, \ldots, n-1))^+.$$

Because $ab^ia \in K_{n,k}$ for $1 \leq i \leq n-k$ and $P(n-k+1, n-k+2, \ldots, n-1) \subseteq K_{n,k}$, we get $U \subseteq K_{n,k}^+$. Furthermore, $(ab^ia)^j \in U$ for $1 \leq i \leq n-k$ and $j \geq 1$ and $P(n-k+1, n-k+2, \ldots, n-1) \subseteq U$ imply $K_{n,k}^+ \subseteq U$. Therefore we have $U = K_{n,k}^+$.
Because the right linear grammar $(\{S, A_1, A_2, \ldots, A_{k-1}\}, \{a, b\}, P, S)$ with

$$\begin{aligned} P = \{&S \to A_1, A_{k-1} \to ab^{n-1}aA_{k-1}, A_{k-1} \to ab^{n-1}a, A_{k-1} \to ab^{n-1}aS\} \\ &\cup \bigcup_{i=1}^{n-k} \{S \to ab^iaS, S \to ab^ia\} \\ &\cup \bigcup_{j=1}^{k-2} \{A_j \to ab^{n-k+j}aA_j, A_j \to ab^{n-k+j}aA_{j+1}\} \end{aligned}$$

generates U, we have $\mathrm{Var}(K_{n,k}^+) = \mathrm{Var}(U) \leq k$. On the other hand, let $G' = (N', \{a, b\}, P', S')$ be a right linear grammar such that $L(G') = K_{n,k}^+$. As in the proof of Lemma 1, we can show that, starting from the words $(aba)^t$ and $ab^{n-k+1}a \ldots ab^{n-k+j-1}a(ab^{n-k+j}a)^t ab^{n-k+j+1}a \ldots ab^{n-1}a$, $1 \leq j \leq k-1$, for sufficiently large t, there are k letters $B_1, B_{n-k+1}, \ldots, B_{n-1}$ with derivations

$$\begin{aligned} B_i &\Longrightarrow^* u_i(ab^ia)^{p_i}v_iB_i \text{ where } p_i \text{ is a natural number, } u_i \text{ and } v_i \text{ are} \\ &\qquad \text{a proper suffix and a proper prefix of } ab^ia, \text{ respectively,} \\ B_i &\Longrightarrow^* z_i \in \{a, b\}^*. \end{aligned}$$

As in Lemma 1, we can prove that all these nonterminals are different. Thus $\mathrm{Var}(K^+_{n,k}) \geq k$.
Combining these estimations we get $\mathrm{Var}(K^+_{n,k}) = k$. Thus $k \in g^{\mathrm{Var}}_+(n)$.

(iv) We show that $1 \in g^{\mathrm{Var}}_+(n)$ for $n \geq 3$. Let $M_n = P(1) \cup P(2) \cup \ldots \cup P(n-1)$. By Lemma 1, $\mathrm{Var}(M_n) = n$. Since $M^+_n = \{aba, ab^2a, \ldots, ab^{n-1}a\}^+$ is generated by the grammar

$$G = \left(\{S\}, \{a, b\}, \bigcup_{i=1}^{n-1} \{S \to ab^i aS, S \to ab^i a\}, S\right),$$

we have $\mathrm{Var}(M^+_n) = 1$. Therefore $1 \in g^{\mathrm{Var}}_+(n)$.

(v) We show that $2 \in g^{\mathrm{Var}}_+(n)$ for $n \geq 3$. Let $R_n = P(1) \cup P(2) \cup \ldots \cup P(n-2) \cup \{c\}P(n-1)$.
As in Lemma 2, we can prove that $\mathrm{Var}(R_n) = n$.
Moreover, we have $R^+_n = (\{aba, ab^2a, \ldots, ab^{n-2}a\} \cup \{c\}\{ab^{n-1}a\}^+)^+$. Again, as in the proof of Lemma 1, we can show the existence of a nonterminal B_{n-1} with derivations

$$\begin{aligned} B_{n-1} &\Longrightarrow^* u_{n-1}(ab^{n-1}a)^{p_{n-1}} v_{n-1} B_{n-1} \text{ where } p_{n-1} \text{ is a natural number,} \\ &\qquad u_{n-1} \text{and} v_{n-1} \text{ are a proper suffix and a proper prefix} \\ &\qquad \text{of } ab^{n-1}a, \text{ respectively,} \\ B_{n-1} &\Longrightarrow^* z_{n-1} \in \{a, b\}^*. \end{aligned}$$

If $\mathrm{Var}(R^+_n) = 1$, then B_{n-1} is the start symbol. Then there is a derivation

$$B_{n-1} \Longrightarrow^* u_{n-1}(ab^{r_{n-1}}a)^{p_{n-1}} v_{n-1} B_{n-1} \Longrightarrow^* u_{n-1}(ab^{r_{n-1}}a)^{p_{n-1}} v_{n-1} z_{n-1}$$

which derives a word not in R^+_n since it contains the subword $ab^{n-1}a$ but no c. Therefore $\mathrm{Var}(R^+_n) \geq 2$.
The right linear grammar $(\{S, S'\}, \{a, b, c\}, P, S)$ with

$$\begin{aligned} P = \{&S \to cS', S' \to ab^{n-1}aS', S' \to ab^{n-1}a, S' \to ab^{n-1}aS\} \\ &\cup \bigcup_{i=1}^{n-2} \{S \to ab^i aS, S \to ab^i a\} \end{aligned}$$

generates R^+_n, which gives $\mathrm{Var}(R^+_n) = 2$. This proves $2 \in g^{\mathrm{Var}}_+(n)$.

(vi) Finally, we prove $1 \in g^{\mathrm{Var}}_+(2)$.
Let $Q = \{a\} \cup \{a^3\}^*$. Assume that $\mathrm{Var}(Q) = 1$. Then there is a right linear grammar $(\{S\}, \{a\}, P, S)$ which generates Q. Obviously, P contains a rule $S \to a^k S$, $k \geq 1$, since otherwise a finite language is generated. Moreover, there is a derivation $S \Longrightarrow^* a$. Then we have the derivations

$$S \Longrightarrow a^k S \Longrightarrow^* a^k a = a^{k+1} \text{ and } S \Longrightarrow a^k S \Longrightarrow a^k a^k S \Longrightarrow^* a^k a^k a = a^{2k+1}.$$

Because $k + 1 > 1$, we obtain $k + 1 = 3p$ and $2k + 1 = 3q$ for certain positive integers p and q. We add these equations and get $3k + 2 = 3(p + q)$ or equivalently $2 = 3(p + q - k)$ which is impossible. This contradiction proves $\mathrm{Var}(Q) \geq 2$.
The right linear grammar $(\{S, S'\}, \{a\}, \{S \to a, S \to S', S' \to a^3 S', S' \to a^3\}, S)$ generates Q. Therefore $\mathrm{Var}(Q) = 2$.
Moreover, we have $Q^+ = \{a\}^+$ and $\{a\}^+$ is generated by the right linear grammar $(\{S\}, \{a\}, \{S \to aS, S \to a\}, S)$. Hence $\mathrm{Var}(Q^+) = 1$ and $1 \in g^{\mathrm{Var}}_+(2)$. □

Theorem 3 *For $n \geq 1$, we have $g_*^{\mathrm{Var}}(n) = \{1, 2, \ldots, n+1\}$.*

Proof (i) We first prove that $\mathrm{Var}(L) = n$ implies $\mathrm{Var}(L^*) \leq n+1$. We start with a grammar $G = (N, T, P, S)$ such that $L(G) = L$ and $\mathrm{Var}(G) = \mathrm{Var}(L) = n$. Then we construct the grammar $G' = (N, T, P', S)$ which generates L^+ as in part i) of the proof of Theorem 2. Now we modify G' to $G'' = (N \cup \{S'\}, T, P' \cup \{S' \to \lambda,\ S' \to S\}, S')$ where S' is a new symbol. Then $L(G'') = L^*$ and $\mathrm{Var}(G'') = n+1$. Thus $\mathrm{Var}(L^*) \leq n+1$.

(ii) Let $n \geq 1$. Let $L_n = P(1, 2, \ldots, n)\{c\}$. Then $\mathrm{Var}(L_n) = n$ by Lemma 2.
Let $G = (N, \{a, b, c\}, P, S)$ be a grammar such that $L(G) = L_n^*$ and $\mathrm{Var}(G) = \mathrm{Var}(L_n^*)$. Since $L_n \subseteq L_n^*$, as in Lemma 1, we can prove the existence of n different nonterminals $B_1, B_2, \ldots, B_n$ with a derivation

$$B_i \Longrightarrow^* u_i(ab^i a)^{p_i} v_i B_i \text{ where } p_i \text{ is a natural number,}$$
$$u_i \text{ and } v_i \text{ are a proper suffix and a proper prefix of } ab^i a, \text{ respectively,} \quad (1)$$

$1 \leq i \leq n$, in N. Moreover, since $\lambda \in L_n^*$, there is a derivation $S \Longrightarrow^* C \Longrightarrow \lambda$ where $C \to \lambda$ is applied in the last step. If $C = B_i$ for some i, $1 \leq i \leq n$, then we have a derivation

$$S \Longrightarrow^* C = B_i \Longrightarrow u_i(ab^i a)^{p_i} v_i B_i = u_i(ab^i a)^{p_i} v_i C \Longrightarrow u_i(ab^i a)^{p_i} v_i,$$

which produces a word in $L(G) = L_n^*$ which is not empty and contains no c. This contradicts the fact that a non-empty word of L_n^* contains a c. Thus we need an additional nonterminal, i.e., $\mathrm{Var}(L_n^*) \geq n+1$. By part i) of this proof, $\mathrm{Var}(L_n^*) = n+1$. Hence $n+1 \in g_*^{\mathrm{Var}}(n)$ for $n \geq 1$.

(iii) Let $n \geq 1$ and $L'_n = P'(1, 2, \ldots, n)$. Then we have $\mathrm{Var}(L'_n) = n$ by Lemma 1. Let $G = (N, \{a, b\}, P, S)$ be a grammar such that $L(G) = (L'_n)^*$ and $\mathrm{Var}(G) = \mathrm{Var}((L'_n)^*)$. Since $L'_n \subseteq (L'_n)^*$, as in Lemma 1, we can prove the existence of n different nonterminals $B_1, B_2, \ldots, B_n$ with a derivation (1). Hence $\mathrm{Var}((L'_n)^*) \geq n$. Because the grammar $H = (\{B_1, B_2, \ldots, B_n\}, \{a, b\}, P, B_1)$ with

$$P = \{B_n \to ab^n aB_n,\ B_n \to B_1,\ B_n \to \lambda\} \cup \bigcup_{i=1}^{n-1} \{B_i \to ab^i aB_i,\ B_i \to B_{i+1}\}$$

generates $(L'_n)^*$, we obtain $\mathrm{Var}((L'_n)^*) = n$.
Hence $n \in g_*^{\mathrm{Var}}(n)$ for $n \geq 1$.

(iv) $1, 2, \ldots, n-1 \in g_*^{\mathrm{Var}}(n)$ can be shown by the witnesses given in the proof of Theorem 2; we only have to add the rule $S \to \lambda$ for the axiom S because then the Kleene-closure of the witness is generated. □

3.3 Concatenation

For concatenation, we only present some partial results.

Lemma 5 *For any positive integers $n \geq 1$ and $m \geq 1$, languages L_n and K_m with $\mathrm{Var}(L_n) = n$ and $\mathrm{Var}(K_m) = m$, we have $\mathrm{Var}(L_n K_m) \leq n + m$.*

Proof Let $G_n = (N_n, T, P_n, S_n)$ and $H_m = (N_m, T, P_m, S_m)$ be two right linear grammars such that $L(G_n) = L_n$, $L(H_m) = K_m$, $\mathrm{card}(N_n) = \mathrm{Var}(L_n) = n$, and $\mathrm{card}(N_m) = \mathrm{Var}(K_m) = m$. We can assume that $N_n \cap N_m = \emptyset$. We consider the right linear grammar

$$G = (N_n \cup N_m, T, \{A \to wB \mid A \to wB \in P_n\} \cup \{A \to wS_m \mid A \to w \in P_n\} \cup P_m, S_n).$$

Since the finishing of a derivation in G_n by a rule $A \to w$ is replaced by an application of $A \to wS_m$, we have to continue with a derivation in H_m. Thus $L(G) = L_n K_m$. Because $\mathrm{Var}(G) = n + m$, we obtain the relation $\mathrm{Var}(L_n K_m) \leq n + m$. □

Theorem 4 *For $n \geq 1$, $m \geq 1$, we have*

$$\{\min\{m,n\}, \min\{m,n\}+1, \ldots, m+n\} \subseteq g_{\cdot}^{\mathrm{Var}}(m,n).$$

Proof (i) For $n \geq 1$ and $m \geq 1$, let

$$L_n = P(1,2,\ldots,n) \quad \text{and} \quad K_{m,n} = P(n+1, n+2, \ldots, n+m).$$

By Lemma 1, $\mathrm{Var}(L_n) = n$ and $\mathrm{Var}(K_m) = m$. Moreover,

$$L_n K_m = P(1,2,\ldots,n,n+1,\ldots,n+m).$$

Again, by Lemma 1, $\mathrm{Var}(L_n K_m) = n + m$. This proves $n + m \in g_{\cdot}^{\mathrm{Var}}(m,n)$.

(ii) Let $n \geq m \geq 2$ and $n - 1 \geq k' \geq 0$. We consider the languages

$$L_n = P'(1,2,\ldots,n) \text{ and } K_{m,k'} = U(k'+1, k'+2, \ldots, n)P(n+1, n+2, \ldots n+m-1).$$

Then we obtain

$$L_n K_{m,k'} = P'(1,2,\ldots,k')U(k'+1,k'+2,\ldots,n)P(n+1,n+2,\ldots n+m-1).$$

By Lemmas 1 and 3 , we get

$$\mathrm{Var}(L_n) = n, \ \mathrm{Var}(K_{m,k'}) = m, \quad \text{and} \quad \mathrm{Var}(L_n K_{m,k'}) = k' + m.$$

Thus $k \in g_{\cdot}^{\mathrm{Var}}(n,m)$ for $m \leq k \leq m + n - 1$.

If $n \geq m = 1$, then we consider the languages $L_n = P'(1,2,\ldots,n)$ and $K_{m,k'} = U(k'+1, k'+2, \ldots, n)$.

(iii) For $m \geq n \geq 2$ and $m - 1 \geq k' \geq 0$, we consider

$$L_{n,k'} = P(1,2,\ldots,n-1)U(n+1,n+2,\ldots,n+m-k'),$$
$$K_m = P'(n+1,n+2,\ldots n+m)$$

and obtain $k \in g_{\cdot}^{\mathrm{Var}}(n,m)$ for $n \leq k \leq m + n - 1$.

For $m \geq n = 1$, we consider the modification analogous to that in ii). □

With respect to numbers which are smaller than $\min\{n,m\}$, we only know that 1 can be obtained if $n \geq 2$ and $m \geq 2$.

Lemma 6 *For $n \geq 2$ and $m \geq 2$, we have $1 \in g_{\cdot}^{\mathrm{Var}}(n,m)$.*

Proof Let h be the homomorphism defined by $h(a) = b$ and $h(b) = a$.

If $n \geq 3$ and $m \geq 3$, we consider the languages

$$L_n = P(1,2,\ldots,n-2) \cup \{b\}\{a,b\}^* \text{ and } K_m = h(P(1,2,\ldots,m-2\} \cup \{b\}\{a,b\}^*).$$

Then we obtain $\mathrm{Var}(L_n) = n$ and $\mathrm{Var}(K_m) = m$ by Lemma 4 and symmetry. Moreover, $L_n K_m = \{a,b\}^+$, which gives $\mathrm{Var}(L_n K_m) = 1$.

If $n = 2$ or $m = 2$, we replace $P(1,2,\ldots,n-2)$ and $P(1,2,\ldots,m-2)$ by the empty set. □

With respect to state complexity, there are investigations where instead of all regular languages only members of some subregular families are considered. For instance, the restriction to prefix-free languages was considered in [8,15] (where a language L is prefix-free if and only if no proper prefix of a word in L belongs to L). We now show that such restrictions lead to the situation that small values cannot be obtained for the Var-complexity of $L_n K_m$.

Definition 1 Two languages L and K are well concatenated if and only if, for any words u and v, $u \in L$ and $uv \in LK$ implies $v \in K$.

As an example, the languages $L = \{x, x^2\}$ and $K = \{x^2\}$ are not well concatenated since $x \in L$ in $xx^3 = x^4 \in LK$, but x^3 is not in K. On the other hand, $L' = \{x^2, x^3\}$ and $K' = \{x^2, x^3\}$ are well concatenated.

Remark 1 If L is prefix-free and K is an arbitrary language, then L and K are well concatenated.

This can be seen as follows. Let us assume that L and K are not well concatenated. Then there are words u and v such that $u \in L$, $uv \in LK$, and $v \notin K$. Then there are words u' and v' such that $u'v' = uv$, $u' \in L$, and $v' \in K$. Obviously, $u \neq u'$. Therefore u is a proper prefix of u' or u' is a proper prefix of u. Since both words are in L, we obtain a contradiction to the prefix-freeness of L.

We now show that small numbers cannot occur if L and K are well concatenated.

Lemma 7 *Let L and K be well concatenated. Then* $\mathrm{Var}(LK) \geq \mathrm{Var}(K) - 1$.

Proof Let $G = (N, T, P, S)$ be a right linear grammar such that $L(G) = LK$ and $\mathrm{Var}(G) = \mathrm{Var}(LK)$. Let t be the maximal length of a terminal word w with $A \to wB \in P$ or $A \to w \in P$.

Let S' be a symbol not in $N \cup T$. For any derivation

$$D: \; S \Longrightarrow^* zX \Longrightarrow zuvY \Longrightarrow^* y \in L(G) \text{ or } D: \; S \Longrightarrow^* zX \Longrightarrow zuv \in L(G)$$

with $zu \in L$ (and $X, Y \in N$), we set $p_D = S' \to vY$ or $p_D = S' \to v$, respectively. Let P' be the set of all rules obtained in this way. Since we have only finitely many rules for any nonterminal X, P' is finite. Then we construct the the right linear grammar $G' = (N \cup \{S'\}, T, P \cup P', S')$ and prove that $L(G') = K$.

Let $w \in K$. We take a word $w' \in L$. Then $w'w \in LK$. Hence, in G, there are derivations

$$D: \; S \Longrightarrow^* zX \Longrightarrow zuvY \Longrightarrow^* zuvy \in L(G) \text{ or } D: \; S \Longrightarrow^* zX \Longrightarrow zuv \in L(G)$$

with $zu = w'$ and $vy = w$ or $v = w$, respectively. In the former case, by construction $S' \to vY \in P'$, and hence we have the derivation $S' \Longrightarrow vY \Longrightarrow^* vy = w$ in G' (because the derivation $Y \Longrightarrow^* y$ uses only rules of P), which proves $w \in L(G')$. Analogously, we conclude in the latter case. Thus $K \subseteq L(G')$.

Conversely, let $w \in L(G')$. Then there are derivations

$$D': \; S' \Longrightarrow vY \Longrightarrow^* vy' \text{ or } D': \; S' \Longrightarrow v$$

with $w = vy'$ or $w = v$. Since $S' \to vY$ or $S' \to v$ in P', there are derivations

$$D: \; S \Longrightarrow^* zX \Longrightarrow zuvY \Longrightarrow^* zuvy \in L(G) \text{ or } D: \; S \Longrightarrow^* zX \Longrightarrow zuv \in L(G)$$

in G with $zu \in L$. However, then we also have the derivations

$$D: \; S \Longrightarrow^* zX \Longrightarrow zuvY \Longrightarrow^* zuvy' \in L(G) \text{ or } D: \; S \Longrightarrow^* zX \Longrightarrow zuv \in L(G).$$

We obtain $zuvy' \in LK$ or $zuv \in LK$, respectively. Since $zu \in L$ and L and K are well concatenated, we get $vy' = w \in K$ or $v = w \in K$, respectively. Therefore $L(G') \subseteq K$.

By definition $\mathrm{Var}(K) \leq \mathrm{Var}(G') = \mathrm{card}(N) + 1 = \mathrm{Var}(LK) + 1$. □

We note that the witnesses given in the proof of Theorem 4 are not prefix-free. However, if we consider the prefix-free languages $L_n = P(1, 2, \ldots, n)\{c\}$ and $K_m = P(n+1, n+2, \ldots, n+m)\{d\}$, where c and d are additional letters, we also get $\mathrm{Var}(L_n) = n$, $\mathrm{Var}(K_m) = m$, and $\mathrm{Var}(L_n K_m) = n + m$, i.e., the upper bound $m + n$ is also obtained with prefix-free sets. We do not know whether Theorem 4 also holds for prefix-free languages (together with Lemma 7 and Remark 1 it would be a nearly optimal result).

3.4 Intersection

We start with a lemma which is an intermediate consequence of the commutativity of intersection.

Lemma 8 *For $n \geq 1$ and $m \geq 1$, we have $g_{\cap}^{\mathrm{Var}}(n, m) = g_{\cap}^{\mathrm{Var}}(m, n)$.* □

We now give a partial result concerning $g_{\cap}^{\mathrm{Var}}$. In contrast to concatenation, we only know that certain small numbers are in the intersection.

Lemma 9 (i) *For $n \geq 3$ and $m \geq 3$, we have $\{0, 1, 2, \ldots, n+m-3\} \subseteq g_{\cap}^{\mathrm{Var}}(n, m)$.*
(ii) *(i) For $n \geq 3$ and $m \in \{1, 2\}$, we have $\{0, 1, 2, \ldots, n\} \subseteq g_{\cap}^{\mathrm{Var}}(n, m)$.*

Proof By Lemma 8, it is sufficient to consider the situation that $n \geq m$.

We distinguish some cases and give witnesses L_n with $\mathrm{Var}(L_n) = n$ and K_m with $\mathrm{Var}(K_m) = m$ such that $\mathrm{Var}(L_n \cap K_m) = k$. Since the proofs that the languages have the required Var-complexities follow directly from Lemmas 1–4 or can be given by arguments analogous to those used in the proofs of Lemmas 1–4 in all cases, we only present the witnesses in Table 3 (in order to simplify the notation we omit the possible dependence of the languages by the parameters k or k'). □

We do not know an upper bound for the numbers in $g_{\cap}^{\mathrm{Var}}(n, m)$. This comes from the fact that all constructions of a right linear grammar for $L_n \cap K_m$ from right linear grammars G_n and H_m for L_n and K_m, respectively, do not only depend on n and m. For instance, we can transform the given grammars into right linear grammars $G'_n = (N_n, T, P_n, S_n)$ and $H'_m = (N_m, T, P_m, S_m)$ in normal form. Then the right linear grammar $(N_n \times N_m, T, P, S)$ with $S = (S_n, S_m)$ and

$$P = \{(A, B) \to a(C, D) \mid A \to aC \in P_n, B \to aD \in P_m\}$$
$$\cup\{(A, B) \to \lambda \mid A \to \lambda \in P_n, B \to \lambda \in P_m\}$$

generates $L_n \cap K_m$. However, the blowup of the number of nonterminals in the construction of the normal forms depends on the length of the right hand sides of the rules in the original grammars. Since this length can be arbitrarily large, we have no upper bound for $g_{\cap}^{\mathrm{Var}}(n, m)$ using this construction. The situation for other constructions—known to us—is similar.

3.5 Union

Finally we study the behaviour under union.

We start by proving that $g_{\cup}^{\mathrm{Var}}$ is a symmetric function.

Table 3 Table of the witnesses to prove Lemma 9

$n \geq m \geq 1$	$L_n = P(1, 2, \ldots, n)$
$k = 0$	$K_m = P(n+1, n+2, \ldots, n+m)$
	$L_n \cap K_m = \emptyset$
$n \geq m \geq 1$	$L_n = P'(1, 2, \ldots, n),$
$1 \leq k \leq m$	$K_m = P'(1, 2, \ldots k, n+1, \ldots, n+m-k)$
	$L_n \cap K_m = P'(1, 2, \ldots, k)$
$n \geq m \geq 3$	$L_n = U(1, 2, \ldots, m-2) \cup P(m-1, m, \ldots, m+k'-1)$
	$\cup P(m+k', m+k'+1, \ldots, n+m-4)$
$0 \leq k' \leq n-4$	$K_m = P(1, 2, \ldots, m-2) \cup U(m-1, \ldots, m+k'-1)$
$m \leq k = m+k' \leq m+n-4$	$L_n \cap K_m = P(1, \ldots, m-2) \cup P(m-1, \ldots, m+k'-1)$
$n \geq m \geq 3$	$L_n = U(1, 2, \ldots, m-2) \cup P(m-1, m, \ldots, m+n-4)$
$k = n+m-3$	$K_m = P(1, 2, \ldots, m-2) \cup U(m-1, \ldots, m+n-4)$
	$L_n \cap K_m = P(1, \ldots, m-2) \cup P(m-1, \ldots, m+n-4)$
$n \geq 3, m = 2$	$L_n = P(1, 2, \ldots, k) \cup h(P(k+1, k+2, \ldots, n-1))$
$1 \leq k \leq n-2$	$K_m = \{a\}\{a, b\}^*$
	$L_n \cap K_m = P(1, 2, \ldots, k)$
$n \geq 2, m = 2$	$L_n = \{c, \lambda\}P(1, 2, \ldots, n-1)$
$k = n-1$	$K_m = \{a\}\{a, b\}^*$
	$L_n \cap K_m = P(1, 2, \ldots, n-1)$
$n \geq 1, m = 2$	$L_n = P(1, 2, \ldots, n)$
$k = n$	$K_m = \{a\}\{a, b\}^*$
	$L_n \cap K_m = P(1, 2, \ldots, n)$
$n \geq 3, m = 1$	$L_n = P(1, 2, \ldots, k) \cup h(P(k+1, \ldots, n-1))$
$1 \leq k \leq n-2$	$K_m = \{a, b\}^*$
	$L_n \cap K_m = P(1, 2, \ldots, k)$
$n \geq 2, m = 1$	$L_n = \{c, \lambda\}P(1, 2, \ldots, n-1)$
$k = n-1$	$K_m = \{a, b\}^*$
	$L_n \cap K_m = P(1, 2, \ldots, n-1)$
$n \geq 1, m = 1$	$L_n = P(1, 2, \ldots, n)$
$k = n$	$K_m = \{a, b\}^*$
	$L_n \cap K_m = P(1, 2, \ldots, n)$

In each case, the intersection is given in the third/fourth line. The morphism h is given by $h(a) = c$ and $h(b) = d$

Lemma 10 *For $n \geq 1$ and $m \geq 1$, we have $g_{\cup}^{\mathrm{Var}}(n, m) = g_{\cup}^{\mathrm{Var}}(m, n)$.* □

The statement follows immediately from the commutativity of union.

Theorem 5 *For $n \geq 1$ and $m \geq 1$, we have $g_{\cup}^{\mathrm{Var}}(n, m) = \{1, 2, \ldots, m+n+1\}$*

Proof We first prove that $\mathrm{Var}(L_n) = n$ and $\mathrm{Var}(K_m) = m$ imply $\mathrm{Var}(L_n \cup K_m) \leq n+m+1$. Let $G_n = (N_n, T, P_n, S_n)$ and $H_m = (N_m, T, P_m, S_m)$ be right linear grammars with $L(G_n) = L_n$, $L(H_m) = K_m$, $\mathrm{Var}(G_n) = \mathrm{Var}(L_n) = n$, and $\mathrm{Var}(H_m) = \mathrm{Var}(K_m) = m$. We assume that $N_n \cap N_m = \emptyset$. Then, by the standard construction, the right linear grammar

$$G = (N_n \cup N_m \cup \{S\}, T, P_n \cup P_m \cup \{S \to S_n, S \to S_m\}, S)$$

generates $L_n \cup K_m$ and has $n+m+1$ nonterminals. Therefore, $\mathrm{Var}(L_n \cup K_m) \le n+m+1$.

For $n \ge 1$ and $m \ge 1$, let $L_n = P(1, 2, \ldots, n)$ and $K_m = P(n+1, n+2, \ldots, n+m)$. Then we have $\mathrm{Var}(L_n) = n$, $\mathrm{Var}(K_m) = m$, and $\mathrm{Var}(L_n \cup K_m) = n+m+1$ by Lemma 1.

For $n \ge 1, m \ge 1$, and $1 \le k \le n+m$, in Tables 4 and 5 , we now present witness languages L_n and K_m such that the relations $\mathrm{Var}(L_n) = n$, $\mathrm{Var}(K_m) = m$, and $\mathrm{Var}(L_n \cup K_m) = k$ hold. By Lemma 10, we can assume without loss of generality that $n \ge m$. Because in all cases the Var-complexity of the given languages follows directly by Lemmas 1–4 or can be given by arguments similar to those used in the proof of Lemmas 1–4, we only give the witnesses and their union. □

4 Conclusions

In this paper we have started the investigation of the sets $g_f^{\mathrm{Var}}(n, m)$ (and $g_f^{\mathrm{Var}}(n)$) for regularity preserving binary (and unary, respectively) operations f. Summarizing our results, we obtain Table 6.

We see that we have complete results for reversal, positive Kleene-closure, Kleene-closure and union. However, for concatenation and intersection, we have only given some partial results; an exact determination of $g_{\cdot}^{\mathrm{Var}}(n, m)$ and $g_{\cap}^{\mathrm{Var}}(n, m)$ remains to be done.

In comparison with the measures Varnf of right linear grammars in normal forms (which equals the nondeterministic state complexity nsc) and Varcf of context-free grammars given in the Tables 1 and 2 , we see that our results correspond more to those of Varnf (= nsc) (those for nsc are more complete) than to those of Varcf where the upper bound for concatenation and the range for reversal are different.

It is left open to determine the range for further operations as—for instance—complement, set-subtraction, and symmetric difference.

The behaviour of other complexity measures defined on special regular languages under operations is already investigated; for sc see [5]. Especially, the behaviour of finite and unary languages is studied. With respect to Var, the behaviour of finite and unary languages is not of interest since, for any finite language L, $\mathrm{Var}(L) = 1$ holds, and for any unary language K, $\mathrm{Var}(K) \le 2$ is valid. However, for other special languages as regular prefix-free languages or regular suffix-closed languages (any suffix of a word in L is in L, too) the behaviour can be studied.

Furthermore, for a right linear grammar $G = (N, T, P, S)$, we can also define the complexity measures

$$\mathrm{Prod}(G) = \mathrm{card}(P), \text{ and } \mathrm{Symb}(G) = \sum_{A \to w \in P} (|w| + 2),$$

which count the number of rules and the sum of symbols contained in rules, respectively. Then we can extend these measures for regular languages as in Section 1. These measures describe the complexity of a language in more precise manner. For context-free languages, the investigation of the behaviour of the corresponding complexity measures under operations was done in [2,9]. A study of the ranges for the measures Prod and Symb for right linear grammars is left as an open field.

Table 4 Table of the witnesses L_n with $\mathrm{Var}(L_n) = n$ and K_m with $\mathrm{Var}(K_m) = m$ for $n \geq 3$

$n \geq 3, m \geq 1$	$L_n = P(1) \cup P(2, 3, \ldots, n-1)$
$k = n + m$	$K_m = P(n+1, n+2, \ldots, n+m)$
	$P(1) \cup P(2, 3, \ldots, n-1) \cup P(n+1, n+2, \ldots, n+m)$
$n \geq m \geq 3$	$L_n = P(1) \cup P(2) \cup \cdots \cup P(n-1)$
$k = n + m - 1$	$K_m = P(n) \cup P(n+1) \cup \cdots \cup P(n+m-2)$
	$P(1) \cup P(2) \cup \cdots \cup P(n+m-2)$
$n \geq 3, m = 2$	$L_n = P(1) \cup P(2) \cup \cdots \cup P(n-1)$
$k = n + m - 1 = n + 1$	$K_m = \{c\}P(n)$
	$P(1) \cup P(2) \cup \cdots \cup P(n-1) \cup \{c\}P(n)$
$n \geq 3, m = 1$	$L_n = P(1) \cup P(2) \cup \cdots \cup P(n-1)$
$k = n + m - 1 = n$	$K_m = P(1)$
	$P(1) \cup P(2) \cup \cdots \cup P(n-1)$
$n \geq m \geq 3$	$L_n = P(1) \cup P(2) \cup \cdots \cup P(n-1)$
$n + 1 \leq k \leq m + n - 2$	$K_m = P(n+1) \cup \cdots \cup P(n+k') \cup P(1) \cup \cdots \cup P(m-k'-1)$
$k' = k - n$	$P(1) \cup \cdots \cup P(n-1) \cup P(n+1) \cup \cdots \cup P(n+k')$
$n \geq m \geq 3$	$L_n = P(1) \cup \cdots \cup P(n-2) \cup U(n+1, \cdots, n+m-2)$
$3 \leq k \leq n$	$K_m = U(k-2, \cdots, n-2) \cup P(n+1) \cup \cdots \cup P(n+m-2)$
	$P(1) \cup \cdots \cup P(k-3) \cup U(k-2, \ldots, n-2)$ $\cup U(n+1, \ldots, n+m-2)$
$n \geq 3, m \geq 3$	$L_n = \{c\}P(1, 2, \ldots, n-2) \cup \{cb\}\{a, b\}^*$
$k = 2$	$K_m = h(\{c\}P(1, 2, \ldots, m-2) \cup \{cb\}\{a, b\}^*)$
	$\{c\}\{a, b\}^+$
$n \geq 3, m \geq 3$	$L_n = P(1, 2, \ldots, n-2) \cup \{b\}\{a, b\}^*$
$k = 1$	$K_m = h(P(1, 2, \ldots, m-2) \cup \{b\}\{a, b\}^*)$
	$\{a, b\}^+$
$n \geq 3, m = 2$	$L_n = P(1) \cup \cdots \cup P(k-2) \cup \{c\}P(k-1) \cup \cdots \cup \{c\}P(n-1)$
$3 \leq k \leq n$	$K_m = \{c\}U(k-1, k, \ldots, n-1)$
	$P(1) \cup P(2) \cup \cdots \cup P(k-2) \cup \{c\}U(k-1, k, \ldots, n-1)$
$n \geq 3, m = 2$	$L_n = \{c\}P(1, 2, \ldots, n-2) \cup \{cb\}\{a, b\}^*$
$k = 2$	$K_m = h(\{cb\}\{a, b\}^*)$
	$\{c\}\{a, b\}^+$
$n \geq 3, m = 2$	$L_n = P(1, 2, \ldots, n-2) \cup \{b\}\{a, b\}^*$
$k = 1$	$K_m = h(\{b\}\{a, b\}^*)$
	$\{a, b\}^+$
$n \geq 3, m = 1$	$L_n = P(1) \cup \cdots \cup P(k-2) \cup P(k-1) \cup \cdots \cup P(n-1)$
$3 \leq k \leq n$	$K_m = U(k-1, k, \ldots, n-1)$
	$P(1) \cup P(2) \cup \cdots \cup P(k-2) \cup U(k-1, k, \ldots, n-1)$
$n \geq 3, m = 1$	$L_n = P(1) \cup \cdots \cup P(n-2) \cup \{c\}\{a, b\}^+$
$k = 2$	$K_m = \{a, b\}^+$

Table 4 continued

	$(\{c\} \cup \{\lambda\})\{a,b\}^+$
$n \geq 3, m = 1$	$L_n = P(1) \cup \cdots \cup P(k-2) \cup P(k-1) \cup \cdots \cup P(n-1)$
$k = 1$	$K_m = \{a,b\}^+$
	$\{a,b\}^+$

In each case, the union given in the third/fourth line, has complexity k. The morphism h (used two times) is given by $h(a) = b$, $h(b) = a$, and $h(c) = c$

Table 5 Table of the witnesses L_n with $\mathrm{Var}(L_n) = n$ and K_m with $\mathrm{Var}(K_m) = m$ for $n \leq 2$

$n = 2, m \geq 1$	$L_n = \{c\}P(1)$, $K_m = P(2, 3 \ldots, m+1)$
$k = n + m$	$\{c\}P(1) \cup P(2, 3 \ldots, m+1)$
$n = m = 1$	$L_n = \{a\}$, $K_m = \{a^{3i} \mid i \geq 1\}$
$k = n + m = 2$	$\{a\} \cup \{a^{3i} \mid i \geq 1\}$
$n = m = 2$	$L_n = K_m = \{c\}P(1)$
$k = 2$	$\{c\}P(1)$
$n = m = 2$	$L_n = \{b\}\{a,b\}^*$, $K_m = \{a\}\{a,b\}^*$
$k = 1$	$\{a,b\}^+$
$n = 2, m = 1$	$L_n = \{b\}\{a\}^+$, $K_m = \{a\}^+$
$k = 2$	$\{b,\lambda\}\{a\}^+$
$n = 2, m = 1$	$L_n = \{b\}\{a\}^+$, $K_m = \{a,b\}^+$
$k = 1$	$\{a,b\}^+$
$n = m = 1$	$L_n = K_m = \{a\}^+$
$k = 1$	$\{a\}^+$

In each case, the union given in the second line, has complexity k

Table 6 Summary of the results on $g_f^{\mathrm{Var}}(n,m)$ and $g_f^{\mathrm{Var}}(n)$

Union	$g_\cup^{\mathrm{Var}}(n,m) = \{1, 2, \ldots, n+m+1\}$ for $n \geq 1, m \geq 1$
Intersection	$\{0, 1, 2, \ldots, n+m-3\} \subseteq g_\cap^{\mathrm{Var}}(n,m)$ for $n \geq 3, m \geq 3$
Concatenation	$\{1\} \cup \{\min\{m,n\}, \min\{m,n\}+1, \ldots, m+n\} \subseteq g_\cdot^{\mathrm{Var}}(m,n)$ for $n \geq 1, m \geq 1$
Reversal	$g_R^{\mathrm{Var}}(n) = \{n-1, n, n+1\}$ for $n \geq 2$
Kleene-closure+	$g_+^{\mathrm{Var}}(n) = \{1, 2, \ldots, n\}$ for $n \geq 1$
Kleene-closure*	$g_*^{\mathrm{Var}}(n) = \{1, 2, \ldots, n+1\}$ for $n \geq 1$

Acknowledgements Open Access funding provided by Projekt DEAL. The author is very grateful to the referees for their comments which considerably improved the paper.

References

1. Brzozowski, J.: Quotient complexity of regular languages. J. Autom. Lang. Comb. **15**, 71–89 (2010)
2. Dassow, J., Harbich, R.: Descriptional complexity of union and star on context-free languages. J. Autom. Lang. Comb. **17**, 123–143 (2012)
3. Dassow, J., Stiebe, R.: Nonterminal complexity of some operations on context-free languages. Fundamenta Informaticae **83**, 35–49 (2008)
4. Ellul, K., Krawetz, B., Shallit, J., Wang, M.-W.: Regular expressions: new results and open problems. J. Autom. Lang. Comb. **10**, 407–437 (2005)
5. Gao, Y., Moreira, N., Reis, R., Yu, Sh: A survey on operational state complexity. J. Autom. Lang. Comb. **21**, 251–310 (2016)
6. Gao, Y., Salomaa, K., Yu, S.: Transition complexity of incomplete DFAs. Fundamenta Informaticae **110**, 143–158 (2011)
7. Gruber, H., Holzer, M.: Language operations with regular expressions of polynomial size. Theor. Comput. Sci. **418**, 3281–3289 (2009)
8. Han, Y.-S., Salomaa, K., Wood, D.: State complexity of prefix-free regular languages. In: Leung, H., Pighizzini, G. (eds.) Proceedings of 8th International Workshop of Descriptional Complexity of Formal Systems, pp. 165–176. University of Las Cruces, Las Cruces (2010)
9. Harbich, R.: Regel- und Symbolkomplexität kontextfreier Sprachen unter ausgewählten Operationen. Dissertation, Otto-von-Guericke-Universität Magdeburg (2019)
10. Holzer, M., Kutrib, M.: State complexity of basic operations on nondeterministic automata. In: Champernaud, J.-M., Maurel, D. (eds.) Implementation and Application of Automata, LNCS 2608, pp. 148–157 (2003)
11. Holzer, M., Kutrib, M.: Nondeterministic finite automata-recent results on the descriptional and computational complexity. Int. J. Found. Comput. Sci. **20**, 563–580 (2009)
12. Hricko, M., Jirásková, G., Szabari, A.: Union and intersection of regular languages and descriptional complexity. In: Mereghetti, C., Palano, B., Pighizzini, G., Wotschke, D. (eds.) Proceedings of 7th International Workshop of Descriptional Complexity of Formal Systems, pp. 170–181. University of Milano, Milan (2005)
13. Jirásková, G.: Concatenation of regular languages and descriptional complexity. In: Frid, A.E., Morozov, A., Rybalchenko, A., Wagner, K.W. (eds.) Fourth International Computer Science Symposium in Russia 2009, LNCS 5675, pp. 203–214 (2009)
14. Jirásková, G.: The ranges of state complexities for complement, star, and reversal of regular languages. Int. J. Found. Comput. Sci. **25**, 101–124 (2014)
15. Jirásková, G., Krausová, M.: Complexity in prefix-free regular languages. In: McQuillan, I., Pighizzini, G., Trost, B. (eds.) Proceedings 12th International Workshop of Descriptional Complexity of Formal Systems, pp. 236–244. University of Sasketchewan, Saskatoon (2010)
16. Maia, E., Moreira, N., Reis, R.: Incomplete operational transition complexity of regular languages. Inf. Comput. **244**, 1–22 (2015)