A Review on State Complexity of Individual Operations

Yuan Gao, Nelma Moreira, Rogério Reis, Sheng Yu

Technical Report Series: DCC-2011-08
Version 1.1 September 2012

Departamento de Ciência de Computadores

Faculdade de Ciências da Universidade do Porto
Rua do Campo Alegre, 1021/1055,
4169-007 PORTO,
PORTUGAL

Tel: 220 402 900 Fax: 220 402 950
http://www.dcc.fc.up.pt/Pubs/

A Review on State Complexity of Individual Operations

Yuan Gao[1], Nelma Moreira[2], Rogério Reis[2], Sheng Yu[1]

[1] University of Western Ontario
London,Ontario, Canada
{ygao,syu}@csd.uwo.ca
[2] DCC-FC & CMUP – University of Porto
Rua do Campo Alegre 1021, 4169-007, Porto, Portugal
{nam,rvr}@dcc.fc.up.pt

Abstract. The state complexity of a regular language is the number of states of its minimal determinitisc finite automaton. The complexity of a language operation is the complexity of the resulting language seen as a function of the complexities of the operation arguments. In this report we review some of the results of state complexity of individual operations for regular and some subregular languages.

1 State Complexity and Nondeterministic State Complexity

The *state complexity* of a regular language L, $sc(L)$, is the number of states of its minimal DFA. The *nondeterministic state complexity* of a regular language L, $nsc(L)$, is the number of states of a minimal NFA that accepts L.

Since a DFA is in particular an NFA, for any regular language L one has $sc(L) \leq nsc(L)$. It is well known that any m-state NFA can be converted, via the *subset construction*, in an equivalent DFA with at most 2^m states [116] (we call this conversion *determination*). Thus, $sc(L) \leq 2^{nsc(L)}$. To show that

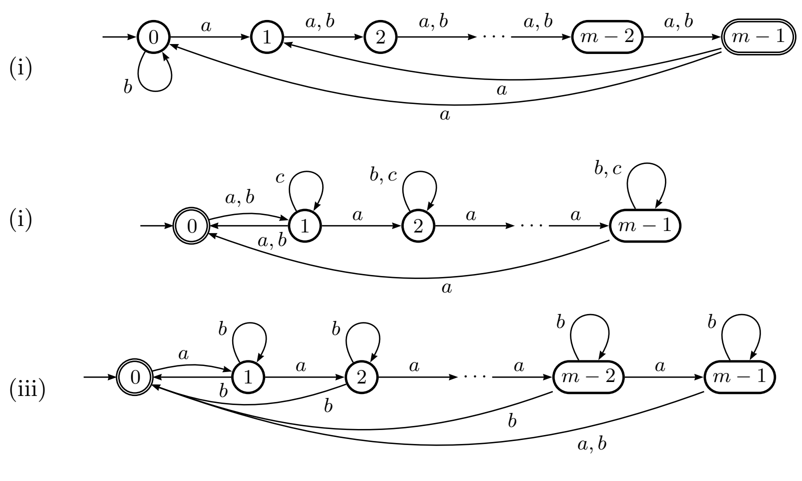

Fig. 1. Moore (i), Lupanov (ii), and Meyer & Fischer (iii) minimal m-state NFA's with equivalent minimal 2^m-state DFA's

this upper bound is tight one must exhibit a family of languages $(L_m)_{m\geq 1}$ such that $nsc(L_m) = m$ and $sc(L_m) = 2^m$, for every $m \geq 1$. In 1963, Lupanov [95] showed that this upper bound is tight using a family of ternary languages. In 1971, Moore [107] and Meyer and Fischer [105] presented different families of binary languages. All three families of NFA's are represented in Figure 1. However, for unary languages that upper bound is not achievable [96, 27, 28]. Chrobak [27, 28] proved that if L is a unary language with $nsc(L) = m$, then $sc(L) \leq O(F(m))$ where

$$F(m) = \max\{\text{lcm}(x_1, \dots, x_l) \mid x_1, \dots, x_l \geq 1 \text{ and } x_1 + \cdots + x_l = m\} \quad (1)$$

is the Landau's function and lcm denotes the least common multiple. It is known that $F(m) \in O(e^{\Theta\sqrt{m \ln m}})$, so $sc(L) = O(e^{\Theta\sqrt{m \ln m}})$. This asymptotic bound is tight, i.e., for every m there exists a unary language L_m such that $nsc(L) \leq m$ and $sc(L_m) = F(m-1)$. Other related bounds were studied by Meregethi and Pighizzini [104].

For a general finite language L, if $nsc(L) = m$ then $sc(L) = \Theta(k^{\frac{m}{1+\log k}})$ and this bound is tight [129]. In the case of finite binary languages, $\Theta(2^{\frac{m}{2}})$ is a tight bound. Mandl [98] had already proved that, for any finite binary language L, if $nsc(L) = m$ then $sc(L) \leq 2 \cdot 2^{m/2} - 1$ if m is even, and $sc(L) \leq 3 \cdot 2^{\lfloor m/2 \rfloor} - 1$ if m is odd, and these bounds are tight.

Finally, for finite unary languages, nondeterminism does not lead to significant improvements. If L is a finite unary language with $nsc(L) = m$, then $sc(L) \leq m + 1$ [98].

The possible gap between state complexity and nondeterministic state complexity for general regular languages lead to the notion of *magic number* introduced in 2000 by Iwama *et al.* [73, 74]. A number α, such that $\alpha \in [m, 2^m]$, is *magic* for m with respect to a given alphabet of size k, if there is no minimal m-state NFA whose equivalent minimal DFA has α states. This notion has been extensively researched in the last decade and has been extended to other gaps between two state complexity values [78, 50, 49, 77, 82, 63]. We summarize here some of the results. The general observation is that, apart from unary languages, magic numbers are hard to find. For binary languages, it was shown that if $\alpha = 2^m - 2^n$ or $\alpha = 2^m - 2^n - 1$, for $n \in [0, m/2 - 2]$ [73], and $\alpha = 2^m - n$ for $n \in [2, 2m - 2]$ and some coprimality condition holds for n [74], then α is not magic. Also, for a binary alphabet, all numbers $\alpha \in [m, m + 2^{\lfloor m/3 \rfloor}]$ have been shown to be non-magic [80], which improves previous results, $[m, m^2/2]$ [78] and $[m, 2^{m^{\sqrt{3}}}]$ [50]. For ternary or quaternary regular languages, and for languages over an alphabet of exponential growing size there are no magic numbers [78, 77, 82]. For the unary case, however, trivially all numbers between $e^{(1+o(1))\sqrt{m \ln m}}$ and 2^m are magic [96, 27, 49]. Moreover, it has been shown that there are much more magic than non-magic numbers in the range from m to $e^{(1+o(1))\sqrt{m \ln m}}$ [49]. In the case of finite languages, partial results were obtained by Holzer *et al.* [63]. All numbers $\alpha \in [m + 1, (\frac{m}{2})^2 + \frac{m}{2} + 1]$, if m even, and $\alpha \in [m + 1, (\frac{m-1}{2})^2 + m + 1]$, if m is odd, are non-magic. Moreover, all numbers of the form $3 \cdot 2^{\frac{m}{2}-1} + 2^i - 1$, if m is even, and $2^{\frac{m+1}{2}} + 2^i - 1$, if m is odd, for some integer $i \in [1, \lceil \frac{m-1}{2} \rceil]$ are non-magic.

1.1 State Complexity versus Quotient Complexity

Quotient complexity, introduced in 2009 by Brzozowski [10, 12], coincides, for regular languages, with the notion of state complexity but it is defined in terms of languages and their (left) quotients. The *left quotient* of a language L by a word w is defined as the language $w^{-1}L = \{x \in \Sigma^\star \mid wx \in L\}$. The *quotient complexity* of L is the number of distinct languages that are left quotients of L by some word (and is denoted by $\kappa(L)$). As it is well-known, for a regular language L, the number of left quotients is finite and is exactly the number of states of the minimal DFA accepting L. So, in the case of regular languages, state complexity and quotient complexity coincide. Considering that quotient complexity is given in terms of languages, and their left quotients, some language' algebraic properties can be used in order to obtain upper bounds for the complexity of operations over languages. Actually, the proof that the set of derivatives of a regular language is finite [9] was one of the earliest studies of state complexity. Quotient complexity can also be useful to show that an upper bound is tight. If a given operation can be expressed as a function of other operations (for example, $L_1 - L_2 = L_1 \cap \overline{L_2}$), then, witnesses for the complexity of those operations can be used to provide a witness for the complexity of the first operation.

2 State Complexity of Individual Operations

The *state complexity of an operation* (or *operational state complexity*) on regular languages is the worst-case state complexity of a language resulting from the operation, considered as a function of the state complexities of the operands.

Adapting a formulation from Holzer and Kutrib [67], given a binary operation $\circ$, the $\circ$-*language operation state complexity problem* can be stated as a decision problem:

- Given an m-state DFA A_1 and an n-state DFA A_2.
- How many states are sufficient and necessary, in the worst case, to accept the language $L(A_1) \circ L(A_2)$ by a DFA?

This formulation can be generalized for other operation arities, automata and languages.

Normally, an upper bound is obtained by providing an algorithm that, given the minimal complete DFA's for the operands, constructs a minimal complete DFA that accepts the resulting language. The number of states of this minimal DFA (as a function of the state complexities of the operands) is an upper bound for the state complexity of the referred operation. To show that an upper bound is tight, for each operand a family of languages (one language, for each possible value of the state complexity) must be given such that the resulting automata achieve that bound.

The same approach is used to obtain the nondeterministic state complexity of an operation on regular languages.

No proofs are here presented for the stated results, although several examples of families of languages, for which the operations achieve the given upper bound, are given.

There are very few results of the study of state complexity on the average case. However, whenever some results are known they will be mentioned together with the worst-case analysis.

In this section, the following notation is used. When considering unary operations, let L be regular language with $sc(L) = m$ ($nsc(L) = m$) and let $A = (Q, \Sigma, \delta, q, F)$ be the complete minimal DFA (a minimal NFA) such that $L = L(A)$. Furthermore, $|\Sigma| = k$ or $|\Sigma| = f(m)$ if a growing alphabet is taken into account, $|F| = f$, and $|F - \{q\}| = l$. In the same way, for binary operations let L_1 and L_2 be regular languages over the same alphabet with $sc(L) = m$ ($nsc(L) = m$) and $sc(L_2) = n$ ($nsc(L_2) = n$), and let $A_i = (Q_i, \Sigma, \delta_i, q_i, F_i)$ be complete minimal DFA's (minimal NFA's) such that $L_i = L(A_i)$, for $i \in [1, 2]$. Furthermore, $|\Sigma| = k$ or $|\Sigma| = f(m, n)$ if a growing alphabet is taken into account, $|F_i| = f_i$, and $|F_i - \{q_i\}| = l_i$, for $i \in [1, 2]$.

2.1 Basic Operations

In this section we review the main results related with state complexity (and nondeterministic state complexity) of some basic operations on regular languages: boolean operations (mainly union, intersection, and complement), catenation, star (and plus), and reversal. For some languages, left and right quotients are also considered. Because their particular characteristics, that were already pointed out in Section 1, for each operation the languages are divided into *regular* ($k \geq 2$ and infinite), *finite* ($k \geq 2$), *unary* (infinite) and *finite unary.* Whenever known, results on the range of complexities that can be reached for each operation are also presented.

There are some survey papers that partially review the results here presented and that were a reference to our presentation [138–140, 71, 141, 128, 67, 10, 66].

General Regular Languages Table 1 summarizes the results for general regular languages. The (fifth) third columns contains the smallest alphabet size of the witness languages for the (nondeterministic) state complexity given in the (fourth) second column, respectively. Columns with this kind of information will also appear in several tables to follow.

In 1994, Yu *et al.* [144] studied the state complexity of catenation, star, reversal, union, intersection, and left and right quotients. They also studied the state complexity of some operations for unary languages. Before, in 1970, Maslov [99] had presented some estimates for union, catenation, and star. Although Maslov considered possible incomplete DFA's, and the paper has some incorrections, the binary languages presented are tight witnesses for the upper bounds for that three operations [10].

Rabin and Scott [116] indicated the upper bound mn for the intersection (that also applies to union). Maslov and Yu *et al.* gave similar witnesses of tightness, both for union and intersection. The families of languages given by Yu *et al.* for intersection are $\{x \in \{a, b\} \mid \#_a(x) = 0 \pmod{m}\}$ and $\{x \in \{a, b\} \mid \#_b(x) = 0 \pmod{n}\}$. Their complements are witnesses for union. Hricko *et al.* [69]

Regular				
	sc	$\|\Sigma\|$	nsc	$\|\Sigma\|$
$L_1 \cup L_2$	mn	2	$m+n+1$	2
$L_1 \cap L_2$	mn	2	mn	2
$\overline{L}$	m	1	2^m	2
$(L_1 - L_2)$	mn	2		
$(L_1 \oplus L_2)$	mn	2		
L_1L_2	$m2^n - f_1 2^{n-1}$, if $m \geq 1,\ n > 1$	2	$m+n$	2
	m, if $m \geq 1$, $n = 1$	1		
$L^\star$	$2^{m-1} + 2^{m-l-1}$, if $m > 1$, $l > 0$	2	$m+1$	2
	m, if $m > 1$, $l = 0$	1		
	$m+1$, if $m = 1$	1		
L^+	$2^{m-1} + 2^{m-l-1} - 1$	2	m	2
L^R	2^m	2	$m+1$	2
$L_2 \setminus L_1$	$2^m - 1$	2		
L_1 / L_2	m	1		
$w^{-1}L$	m	1	$O(m+1)$	
Lw^{-1}	m	1	m	1

Table 1. State complexity and nondeterministic state complexity for basic operations on regular languages

showed that for any integers $m \geq 2$, $n \geq 2$, and $\alpha \in [1, mn]$ there exist binary languages L_1 and L_2 such that $sc(L_1) = m$, $sc(L_2) = n$, and $sc(L_1 \cup L_2) = \alpha$. The same holds for intersection.

Complementation for DFA's is trivial (one has only to exchange the final states) and thus, the state complexity of the complement is the same one of the original language, i.e., $sc(\overline{L}) = sc(L)$. For other boolean operations (set difference, symmetric difference, exclusive disjunction, etc.) the state complexity can be obtained by expressing them as a function of union, intersection and complement [10].

For catenation, Yu *et al.* gave the upper bounds $m2^n - f_1 2^{n-1}$, if $m \geq 1, n \geq 2$; and m, if $m \geq 1, n = 1$. They presented binary languages tight bound witnesses for $m \geq 1$, $n = 1$ and $m = 1$, $n \geq 2$, but ternary languages tight bound witnesses for $m > 1$, $n \geq 2$. But, for the following binary language families presented by Maslov the bound is tight: $\{w \in \{a,b\}^\star \mid \#_a(w) = (m-1) \pmod{m}\}$ and $L((a^\star b)^{n-2}(a+b)(b+a(a+b))^\star)$, for all $m, n \geq 2$ and $f_1 = 1$. Other families of binary languages for which the catenation achieves the upper bound were presented by Jirásková [79]. The same author [81] proved that, for all m, n and α such that either $n = 1$ and $\alpha \in [1, m]$, or $n \geq 2$ and $\alpha \in [1, m2^n - 2^{n-1}]$, there exist languages L_1 and L_2 with $sc(L_1) = m$ and $sc(L_2) = n$, defined over a growing alphabet, such that $sc(L_1L_2) = \alpha$. Moreover, Jirásek *et al.* [76] showed that the upper bound $m2^n - f_1 2^{n-1}$ on the catenation of two languages L_1 and L_2, with $sc(L_1) = m \geq 2$ and $sc(L_2) = n \geq 2$ respectively, are tight for any integer f_1 with $f_1 \in [1, m-1]$. The witness language families are binary and accepted by the DFA's presented in Figure 2.

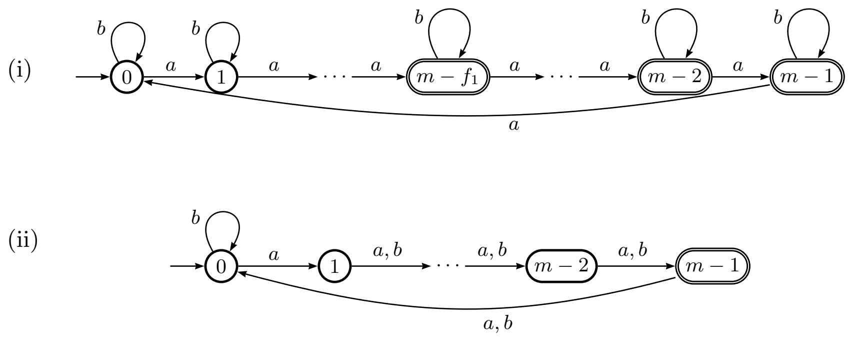

Fig. 2. Witness DFA's for all range of state complexities of the catenation

The state complexity for the star on a regular language L was studied by Yu *et al.*. A lower bound of 2^{m-1} was presented before by Ravikumar and Ibarra [120, 119]. If $sc(L) = 1$ then either $L = \Sigma^\star$, and $sc(L^\star) = 1$, $L = \emptyset$, and $sc(L^\star) = 2$. If $sc(L) = m > 1$, but $l = 0$, i.e., the minimal DFA accepting L has the initial state as the only final state, then $sc(L^\star) = m$, as $L = L^\star$. Finally, if $sc(L) = m > 1$, and $l > 0$, then $sc(L^\star) \leq 2^{m-1} + 2^{m-l-1}$. The upper bound $2^{m-1} + 2^{m-2}$ is achieved for the language $\{w \in \{a, b\} \mid \#_a(w) \text{ is odd}\}$, if $m = 2$; if $m > 2$, for the family of binary languages accepted by the DFA's presented in Figure 2:(ii). We note that although the upper bound given by Maslov is incorrect ($\frac{3}{4}2^m - 1$ instead of $\frac{3}{4}2^m$), the family of languages he presented are witnesses for the above bound (for $m > 2$). Those languages are accepted by the DFA's presented in Figure 3. Both DFA's are shown in Figure 3.

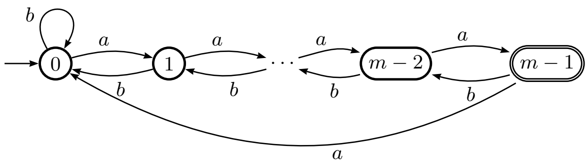

Fig. 3. Maslov's witness DFA's for the state complexity of the star

Jirásková [80] proved the following theorem.

Theorem 1 *For all integers m and α with either $m = 1$ and $\alpha \in [1, 2]$, or $m \geq 2$ and $\alpha \in [1, 2^{m-1} + 2^{m-2}]$, there exists a language L over an alphabet of size 2^m such that $sc(L) = m$ and $sc(L^\star) = \alpha$.*

The state complexity for the plus on a regular language L (i.e., $LL^\star$) coincide with the one for star in the first two cases, but for $m > 1$, $l > 0$ one state is saved (as a new initial state is not needed).

In 1966 Mirkin [106] pointed out that the reversal of the NFA's given by Lupanov as an example of a tight bound for determination (see Figure 1:(ii)), were deterministic. This yields that 2^m is a tight upper bound for the state complexity of reversal of a (at least ternary) language L such that $sc(L) = m$. Leiss [94] studied also this problem and proved the tightness of the bound for another family of ternary languages. Yu *et al.* presented also (independently) Lupanov example. Salomaa *et al.* [127] studied several classes of languages where the upper bound is achieved. Nevertheless, a family of binary languages therein presented as meeting the upper bound for $m \geq 5$ was later shown not to be so [83]. A family of binary languages for which the upper bound for reversal is tight was given by Jirásková and Sĕbej [88, 39] and their minimal DFA's are represented in Figure 4.

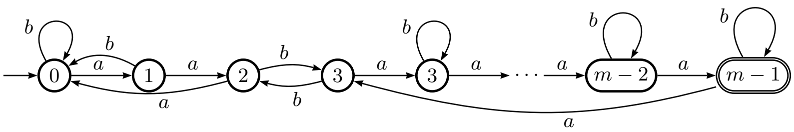

Fig. 4. Witness DFA's for the state complexity of the reversal

In the paper cited above [80], Jirásková proved that,

Theorem 2 *For all integers $m \geq 3$ and $\alpha \in [\log m, 2^m]$, there exists a language L over an alphabet of size 2^m such that has $sc(L) = m$ and $sc(L^R) = \alpha$. If $sc(L) = 2$ then $sc(L^R)$ can be 2, 3 or 4, and if $sc(L) = 1$ then $sc(L^R) = 1$.*

Yu *et al.* showed that the state complexity for the left quotient of a regular language L_1 by an arbitrary language L_2, $L_2 \setminus L_1$, is less or equal to $2^m - 1$, with $sc(L_1) = m$, and that this bound is tight for the family of binary languages given in Figure 2:(ii) and considering $L_2 = \Sigma^\star$. In 1971, Conway [29] had already stated that if L_2 is a regular language then $sc(L_2 \setminus L_1) \leq 2^m$. For the right quotient of a regular language L_1 by an arbitrary language L_2 one has $sc(L_1/L_2) \leq m$. The minimal DFA accepting L_1/L_2 coincides with the one for L_1, except that the set of final states is the set of states $q \in Q_1$ such that there exists a word of $w \in L_2$ such that $\delta_1(q, w) \in F_1$. The bound is tight for $L_2 = \{\varepsilon\}$. For the left and the right quotients of a regular language L by a word $w \in \Sigma^\star$ it is then easy to see that $sc(w^{-1}L) = sc(Lw^{-1}) \leq m$. As a family of languages for which the upper bound is tight consider $\{a^l \mid l = 0 \pmod{m}\}$ and $w \in \{a\}^\star$ [40].

The state complexity of basic operations on NFA's was first studied by Holzer and Kutrib [64], and also by Ellul [40]. We note that for state complexity purposes it is tantamount to consider NFA's with or without ε-transitions. NFA's are considered with only one initial state and trimmed, i.e., all states are accessible from the initial state and from all states a final state is reached.

For union, only a new initial state with ε transitions for each of the operands initial states is needed, thus $sc(L_1 \cup L_2) \leq m + n + 1$. To see that the upper bound is tight, consider the families $(a^m)^\star$ and $(b^n)^\star$ over a binary alphabet. For intersection, a product construction is needed.

The nondeterministic state complexity of the complementation is, trivially, at most 2^m. That this upper bound is tight even for binary languages was proved by Jirásková [79], using a *fooling-set lower-bound technique* [5, 52, 70]. Those languages are accepted by the NFA's presented in Figure 5 (for $m > 2$).

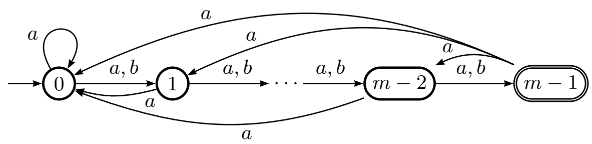

Fig. 5. Witness NFA's for the nondeterministic state complexity of complementation

See Holzer and Kutrib [67] for other witness languages. Using the same techniques, Jirásková and Szabari [76] proved the following theorem for languages over an alphabet of exponential growing size, and that was improved to a five-symbol alphabet by Jirásková [80]:

Theorem 3 *For all integers $m \geq 1$ and $\alpha \in [\log m, 2^m]$, there exists a language L such that $nsc(L) = m$ and $nsc(\overline{L}) = \alpha$.*

Mera and Pighizzini [103] proved a related *best case* result, i.e., for every $m \geq 2$ there exists a language L such that $nsc(L) = m$, $nsc(\overline{L}) \leq m + 1$ and $sc(L) = sc(\overline{L}) = 2^m$. However, as we will see below, this result does not hold if unary languages are considered.

The upper bound for the nondeterministic state complexity of catenation is $m+n$ and this bound can be reached considering the witness binary languages given for union. All the values $\alpha \in [1, m+n]$ can be obtained as nondeterministic state complexity of catenation of unary languages [81].

For the plus of a regular language L, we have $nsc(L^+) \leq nsc(L) = m$: an NFA accepting L^+ coincides with one accepting L except that each final state has also the transitions from the initial state. In the case of the star, one more state can be needed (if L does not accept the empty word), i.e., $sc(L^\star) \leq m+1$. Witness languages of the tightness of these bounds are $\{w \in \{a, b\}^\star \mid \#_a(w) = (m-1) \pmod{m}\}$. All range of values $\alpha \in [1, m+1]$ can be reached for the nondeterministic state complexity of the star of binary languages [80].

For the reversal, at most one more state will be needed, so $nsc(L^R) \leq m + 1$. Witness ternary languages were presented by Holzer and Kutrib, but the bound is tight even for the family of binary languages which minimal NFA's are presented in Figure 6 [79]. If $nsc(L) = m \geq 3$ the possible values for $nsc(L^R)$ are $m-1$, m or $m+1$ [80]. The first value is reached by the reversals of the above binary languages and the second considering the languages $\{w \in \{a, b\}^\star \mid \ |w| = 0 \pmod{m}\}$.

The nondeterministic state complexity of left and right quotients were studied by Ellul [40]. Given a minimal NFA $A_1 = (Q, \Sigma, \delta, q_0, F)$ accepting L, an NFA C accepting Lw^{-1}, for $w \in \Sigma$, coincides

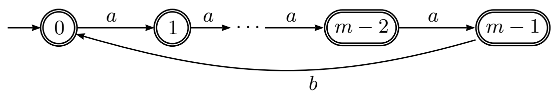

Fig. 6. Witness NFA's for the nondeterministic state complexity of reversal

with A_1 except that the set of final states is $\{q \in Q \mid \delta(q, w) \cap F \neq \emptyset\}$. Thus $nsc(Lw^{-1}) \leq nsc(L)$. The witness languages used for the state complexity of right quotient show that the bound is tight. An upper bound for $nsc(w^{-1}L)$ can be obtained by considering an NFA C with one new initial state q_0' and ε-transitions from q_0' to each state of A_1 reached when inputing w.

Unary Regular Languages Table 2 presents the main state complexity results of the basic operations on unary languages. Given the constraints on both DFA's and NFA's over a one symbol alphabet, and the results presented in Section 1, the state complexity for several operations on unary languages is much lower than what is predicted by the general results of state complexity.

average case Some results on the average case state complexity of operations on unary languages were presented by Nicaud [108, 109].

Unary Regular			
	sc	nsc	asc
$L_1 \cup L_2$	$\sim mn$	$m+n+1$, if $m \neq \dot{n}$	$\sim \frac{3\zeta(3)}{2\pi^2} mn$
$L_1 \cap L_2$	$\sim mn$	mn, if $(m,n) = 1$	$\sim \frac{3\zeta(3)}{2\pi^2} mn$
$\overline{L}$	m	$e^{\Theta(\sqrt{m \log m})}$	
$(L_1 - L_2)$	mn		
$(L_1 \oplus L_2)$	mn		
L_1L_2	$\sim mn$	$[m+n-1, m+n]$, if $m, n > 1$	$O(1)$, $n < P(m)$
$L^\star$	$(m-1)^2+1$, if $m > 1$, $l > 1$	$m+1$, if $m > 2$	$O(1)$
L^+	$(m-1)^2$	m, if $m > 2$	
L^R	m	m	
$w^{-1}L$	m	m	
Lw^{-1}	m	m	

Table 2. State complexity, nondeterministic state complexity and average state complexity of basic operations on unary languages. The upper bounds of state complexity for union, intersection and catenation are exact if $(m, n) = 1$. For the average state complexity of intersection and union, $\zeta(n)$ is the function ζ of Riemman. For the average state complexity of catenation, n must be bounded by a polynomial P in m.

A DFA that accepts a unary language is characterized by a noncyclic part (the tail) and a cyclic part (the loop). A characterization and the enumeration of minimal unary DFA's was given by Nicaud [108].

The state complexity of the reversal of a unary language L is trivially equal to the state complexity of L. For the boolean operations, the state complexity coincides asymptotically with the one for general regular languages. Yu [139] shown that the bound was tight for union (and thus, for intersection) if m and n are coprimes and the witness languages are $(a^m)^\star$ and $(a^n)^\star$. The state complexity of catenation and star was proved by Yu *et al.* [144] and the tightness for the first was also shown for m and n coprimes. The witnesses for the catenation are $(a^m)^\star a^{m-1}$ and $(a^n)^\star a^{n-1}$. For the star, if $m = 2$ a witness is $(aa)^\star$, and for each $m > 2$ a witness is $(a^m)^\star a^{m-1}$. The state complexity when m and n are not necessarily coprimes was studied by Pighizzini and Shallit [114, 113]. In this case, the tight bounds are given by the number of states in the tail and in the loop of the resulting automata. The state complexity for left and right quotient by a word on unary languages coincides with the general case.

Nicaud [108, 109] proved that the state complexity of union, intersection and catenation on two languages L_1 and L_2 is asymptotically equivalent to mn., where $m = sc(L_1)$ and $n = sc(L_2)$.

Let U_n be the set of unary (complete and initially connected) DFA's with n states. The *average state complexity* (asc) of a binary operation $\circ$ on regular languages is given by

$$\frac{\displaystyle\sum_{A_1 \times A_2 \in U_m \times U_n} sc(L(A_1) \circ L(A_2))}{|U_m \times U_n|}$$

This definition can be generalized to other operation arities, automata and languages. As shown in Table 2, the average state complexity of catenation and star on unary languages are bounded by a constant, and for intersection (and union) note that $\frac{3\zeta(3)}{2\pi^2} \approx 0.1826907423$.

The nondeterministic state complexity of basic operations on unary languages was studied by Holzer and Kutrib [65], and also by Ellul [40]. For union and intersection, the upper bound coincides with the general case. However, it was proved to be achievable for union if m is not a divisor or multiple of n. As in the deterministic case, the witnesses for intersection are $(a^m)^\star$ and $(a^n)^\star$, if m and n are coprimes. The nondeterministic state complexity of the complementation is $O(F(m))$ (where F is the Landau's function of equation (1)), which is directly related with the state complexity of determination. Holzer and Kutrib [65] proved that this upper bound is tight in order of magnitude:

Theorem 4 *For any integer $m > 1$ there exists a unary language L such that $nsc(L) = m$ and $nsc(\overline{L}) = \Omega(F(m))$.*

Moreover, Mera and Pighizzini [103] shown that for each $m \geq 1$ and unary language L, such that $nsc(L) = m$ and $sc(L) = sc(\overline{L}) = e^{O(\sqrt{m \log m})}$, then $nsc(\overline{L}) \geq m$.

The upper bound $m + n$ for the catenation of two unary languages is not know to be tight. The known lower bound is $m + n - 1$ achieved by the catenation of $\{a^l \mid l = m - 1 \pmod{m}\}$ and $\{a^l \mid l = (n-1) \pmod{n}\}$ [65]. The same languages can be used to show the tightness of the bound $m + 1$ for the star (and the plus) operation.

For the left and right quotients, notice that in the unary case $w^{-1}L = Lw^{-1}$, and the results for the general case apply.

Finite Languages Finite languages are an important subset of regular languages. They are accepted by complete DFA's that are acyclic apart from a loop on the *sink* (or dead) state, for all alphabetic symbols. Minimal DFA's have also special graph properties that lead to a linear time minimisation algorithm [121], and the length of the longest word accepted plays an important role.

Table 3 shows that the (nondeterministic) state complexity of operations on finite languages are, in general, lower than in the general case.

Câmpeanu *et al.* [22] presented the first formal study of state complexity of operations on finite languages. Yu [139] presented upper bounds of $O(mn)$ for the union and the intersection. The tight upper bounds were given by Han and Salomaa [59] using growing size alphabets. The upper bound for union and intersection cannot be reached with a fixed alphabet when m and n are arbitrarily large. Câmpeanu *et al.* gave tight upper bounds for catenation, star and reversal. For catenation the bound $(m - n + 3)2^{n-2} - 1$ is tight for binary languages if $m + 1 \geq n > 2$. The DFA's of the witness languages are presented in Figure 7.

For star, Câmpeanu *et al.* shown that the bound $2^{m-3} + 2^{m-4}$ is tight for ternary languages. The tight upper bound for the reversal of a finite binary language is $3 \cdot 2^{p-1} - 1$ if $m = 2p$, and $2^{p-1} - 1$ if $m = 2p - 1$.

Nondeterministic state complexity of basic operations on finite languages were studied by Holzer and Kutrib [64]. Minimal NFA's accepting finite languages without the empty word can be assumed to have only a final state (with no transitions); and if the empty word is in the language, the initial

Finite				
	sc	$\|\Sigma\|$	nsc	$\|\Sigma\|$
$L_1 \cup L_2$	$mn - (m+n)$	$f(m,n)$	$m+n-2$	2
$L_1 \cap L_2$	$mn - 3(m+n) + 12$	$f(m,n)$	mn	2
$\overline{L}$	m	1	$\Theta(k^{\frac{m}{1+\log k}})$	2
$(L_1 - L_2)$				
$(L_1 \oplus L_2)$				
L_1L_2	$O(mn^{f_1-1} + n^{f_1})$, if $l_1 > 1$ / $m+n-2$, if $l_1 = 1$	2 / 1	$m+n-1$	2
$L^\star$	$2^{m-3} + 2^{m-l-2}$, $l \geq 2$, $m \geq 4$ / $m-1$, if $f = 1$	3 / 1	$m-1$, $m > 1$	1
L^+	m	1	m, $m > 1$	1
L^R	$O(k^{\frac{m}{1+\log k}})$	2	m	2
$L_2 \setminus L_1$				
L_1/L_2				

Table 3. State complexity and nondeterministic state complexity of basic operations on finite languages

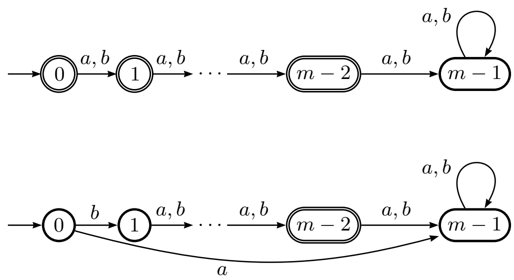

Fig. 7. Witness DFA's for the state complexity of catenation on finite languages

state is also final. Because there are no cycles, for the union of two finite languages three states can be avoided: no new initial state is need, and the initial states and the final states can be merged. The upper bound $m + n - 2$ is tight for the languages a^{m-1} and b^{n-1}, for $m, n \geq 2$. In the case of the intersection, the upper bound coincides with the general case, and it is tight for the binary languages $\{w \in \{a,b\}^\star \mid \#_a(w) = 0 \pmod{m}\}$ and $\{w \in \{a,b\}^\star \mid \#_b(w) = 0 \pmod{n}\}$. Considering the upper bound of determination for finite languages, the nondeterministic state complexity for complement is bounded by $O(k^{\frac{m}{1+\log k}})$. The tight bound $\Omega(k^{\frac{m}{2\log k}})$ is reached for alphabets $\Sigma = \{a_1, \ldots, a_k\}$ of size $k \geq 2$, and the languages $\Sigma^j a_1 \Sigma^i y$, where $i \geq 0$, $0 \leq j \leq i$, $y \in \Sigma \setminus \{a_1\}$, and $m > 2$.

For catenation of finite languages represented by NFA's, one state can be saved. Witness languages for the tightness of the bound $m + n - 1$ can be the ones used for union. Two states are also saved for the star, and for plus the nondeterministic state complexity coincides with the one for the general case. Witness languages are a^m and a^{m-1}, respectively.

NFA's for the reversal are exponentially more succinct then DFA's. In the case of finite languages, and like other operations, one state can be spared. Witness languages are $\{a,b\}^{m-1}$.

Finite Unary Languages Table 4 summarizes the state complexity and nondeterministic state complexity results of basic operations on finite unary languages [22, 139, 65]. State complexity of union, intersection and catenation on finite unary languages are linear, while they are quadratic for general unary languages. In this setting, nondeterminism is only relevant for the star (and plus), as unary regular languages are obtained. As already stated, for a finite unary language L, one has $sc(L) \leq nsc(L) + 1$, and $sc(L) - 2$ is the length of the longest word in the language. If a operation preserves finiteness, for state complexity only the longest words must be considered.

Finite Unary		
	sc	nsc
$L_1 \cup L_2$	$\max\{m,n\}$	$\max\{m,n\}$
$L_1 \cap L_2$	$\min\{m,n\}$	$\min\{m,n\}$
$\overline{L}$	m	$m+1$
$(L_1 - L_2)$	m	
$(L_1 \oplus L_2)$	$\max\{m,n\}$	
$L_1 L_2$	$m+n-2$	$m+n-1$
$L^\star$	2, if $m=3$ $m-1$, if $f=1$ $m^2-7m+13$, if $m>4$, $f\geq 3$	$m-1$
L^+	m	m
L^R	m	m

Table 4. State complexity and nondeterministic state complexity of basic operations on finite unary languages

2.2 Other Regularity Preserving Operations

Table 5 presents the results for the state complexity of some regularity preserving operations, that are detailed in the next paragraphs.

Proportional removals

Proportional removals preserving regularity were studied by Hartmamis [136] and were full characterized by Seiferas and McNaughton [132]. For any binary relation $r \subseteq \mathbb{N} \times \mathbb{N}$ and any language $L \subseteq \Sigma^\star$, let the language $P(r, L)$ be defined as

$$P(r, L) = \{x \in \Sigma^\star \mid \exists y \in \Sigma^\star \text{ such that } xy \in L \wedge r(|x|, |y|)\}.$$

A relation r is *regularity-preserving* if $P(r, L)$ is regular for every regular language L. Seiferas and McNaughton [132] gave sufficient and necessary conditions of regularity preservation in this context.

For the special case where r is the identity, the correspondent language is denoted by $\frac{1}{2}(L)$. Domaratzki [36] proved that for a regular language L, $sc(\frac{1}{2}(L)) = O(sc(L)F(sc(L)))$ (where F is the Landau's function of equation (1)) and this bound is tight for ternary languages. In the case of L be a unary language, one gets $sc(\frac{1}{2}(L)) = sc(L)$. Following Nicaud's work on average case complexity, mentioned above, Domaratzki showed that the average state complexity of the $\frac{1}{2}(\cdot)$ operation on a m-state unary automaton is asymptotically equivalent to $\frac{5}{8}m + c$, for some constant c.

Domaratzki also studied the state complexity of polynomial removals:

Theorem 5 *[36] Let $f \in \mathbb{Z}[x]$ be a strictly monotonic polynomial such that $f(\mathbb{N}) \subset \mathbb{N}$. Then the relation $r_f = \{(n, f(n)) \mid n \geq 0\}$ preserves regularity, and $sc(P(r_f, L)) \leq O(sc(L)F(sc(L)))$.*

In 1970, Maslov [99] had already studied the language $\frac{p}{q}(L)$, i.e., a language $P(r, L)$ such that r is defined by $\{(m, n) \mid mq = pn\}$ with $p, q \in \mathbb{N}$. An open problem is to obtain the state complexity of $P(r, L)$ where r belongs to the broader class of regularity preserving relations studied by Seiferas and McNaughton.

Power

Given a regular language L and $i \geq 2$, an upper bound of the state complexity of the language L^i is given by considering the state complexity of catenation. However, a tight upper bound is obtained if these operations is studied individually. Domaratzki and Okhotin [37] proved that $sc(L^i) = \Theta(m2^{(i-1)m})$, for $i \geq 2$. The bound is tight for a family of languages over a six-symbol alphabet. In the case $i = 3$, $sc(L^3) = \frac{6m-3}{8}4^m - (m-1)2^m - m$, for $m \geq 3$, and the tightness is witnessed by a family of languages over a four-symbol alphabet. The nondeterministic state complexity of L^i is proved to be mi. This bound is shown to be tight over a binary alphabet, for $m \geq 2$. The power of unary languages was studied by Rampersad [117]. If L is a unary language with $sc(L) = m \geq 2$, then $sc(L^i) = im + i + 1$.

Regular				
	sc	$\lvert\Sigma\rvert$	nsc	$\lvert\Sigma\rvert$
$\frac{1}{2}(L)$	$me^{\Theta(\sqrt{m \log m})}$	3		
	m	1		
L^i	$\Theta(m2^{(i-1)m})$	6	mi	2
	$im + i + 1$	1		
L^3	$\frac{6m-3}{8}4^m - (m-1)2^m - m$	4		
L^{CS}	$2^{m^2+m \log m - O(m)}$	4	1, if $m = 1$	
	$2^{\Theta(m^2)}$	2,3	$2m^2 + 1$, if $m \geq 2$	2
	m	1	m	1
$L_1 \sqcup\!\sqcup L_2$	$O(2^{mn} - 1)$	5	$O(mn)$	5
$L_1 \odot_\perp L_2$	$m2^{n-1} - 2^{n-2}$, if $m \geq 3$, $n \geq 4$	4	$m + n$	2
Unique Regular Operations				
$L_1 \mathring{\cup} L_2$	mn	2		
$L_1 \circ L_2$	$O(m3^n - f_1 3^{n-1})$		$\geq 2^{O(h)}$	
$L^{\circ 2}$	$m3^m - 3^{m-1}$	2		
L°	$O(3^{m-1} + (f+2)3^{m-f-1} - (2^{m-1} + 2^{m-f-1} - 2))$			

Table 5. State complexity and nondeterministic state complexity of some regularity preserving operations: proportional removals for the identity relation ($\frac{1}{2}(L)$); powerL^i where $i \geq 2$; cyclic shift L^{CS}; shuffle $L_1 \sqcup\!\sqcup L_2$; orthogonal catenation $L_1 \odot_\perp L_2$; unique operations: for unique star L°, $\varepsilon \notin L$; for the nondeterministic state complexity of $L_1 \circ L_2$, the combined state complexity of L_1 and L_2 is $O(h)$.

Cyclic Shift

The *cyclic shift* of a language L is defined as $L^{CS} = \{vu \mid uv \in L\}$. Maslov [99] gave an upper bound of $(m2^m - 2^{m-1})^m$ for the state complexity of cyclic shift and an asymptotic lower bound of $(m-3)^{m-3} \cdot 2^{(m-3)^2}$, by considering languages over a growing alphabet (if complete DFA's are considered). Jirásková and Okhotin [85] reviewed and improved Maslov results. Using a fixed four-symbol alphabet, they obtained a lower bound of $(m-1)! \cdot 2^{(m-1)(m-2)}$, which shows that $sc(L^{CS}) = 2^{m^2+m \log m - O(m)}$ for alphabets of size greater than 4. For binary and ternary languages, they proved that the state complexity is $2^{\Theta(n^2)}$. As this function grows faster than the number of DFA's for a given m, there must exist some *magic numbers* for the state complexity of the cyclic shift over languages of a fixed alphabet.

The nondeterministic state complexity of this operation was shown to be $2^{m^2} + 1$, for $m \geq 2$, and the upper bound is tight for binary languages. Although the hardness of this operation on the deterministic case, its nondeterministic state complexity is relatively low. For a unary language L, as $L^{CS} = L$, one gets $sc(L^{CS}) = nsc(L^{CS}) = sc(L)$.

Shuffle

The shuffle operation of two words $w_1, w_2 \in \Sigma^\star$ is defined by

$$w_1 \sqcup\!\sqcup w_2 = \{u_1 v_1 \dots u_m v_m \mid u_i, v_i \in \Sigma^\star,\ i \in [1, m],\ w_1 = u_1 \dots u_m \text{ and } w_2 = v_1 \dots v_m\}.$$

This operation is extended trivially to languages. If two regular languages are regular their shuffle is also a regular language. Câmpeanu *et al.* [24] obtained that the state complexity of the shuffle of two regular languages L_1 and L_2 is less or equal to $2^{mn} - 1$. They proved that this bound is tight for witness languages over a five symbols alphabet and if minimal incomplete DFA's are considered (see Figure 8). Thus, $sc(L_1 \sqcup\!\sqcup L_2)$ is at least $2^{(sc(L_1)-1)(sc(L_2)-1)}$.

Various restrictions and generalizations of the shuffle operation have been studied. Mateescu *et al.* [101] introduced the shuffle operation of two languages L_1 and L_2 on a set of trajectories $T \subseteq \{0, 1\}^\star$, $L_1 \sqcup\!\sqcup_T L_2$. When L_1, L_2, and T are regular languages $L_1 \sqcup\!\sqcup_T L_2$ is a regular language.

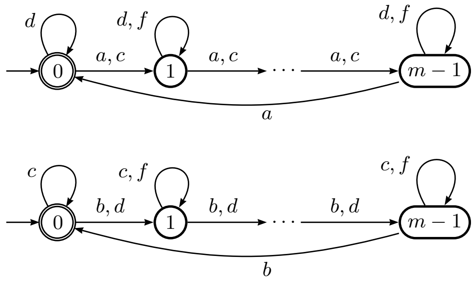

Fig. 8. Incomplete DFA's for the tight upper bound of state complexity of shuffle.

In particular, if $T = \{0,1\}^\star$, then $L_1 \sqcup\!\sqcup_T L_2 = L_1 \sqcup\!\sqcup L_2$; and if $T = \{0\}^\star\{1\}^\star$, then $L_1 \sqcup\!\sqcup_T L_2 = L_1L_2$. Domaratzki and Salomaa [38] studied the state complexity of the shuffle on regular trajectories. In general, $sc(L_1 \sqcup\!\sqcup_T L_2) \leq 2^{nsc(L_1)nsc(L_2)nsc(T)}$. If T belongs to special families of regular languages, tight bounds were also presented.

Orthogonal Catenation A language L is the *orthogonal catenation* of L_1 and L_2, and denoted by $L = L_1 \odot_\perp L_2$, if every word w of L can be obtained in just one way as a catenation of a word of L_1 and a word of L_2. If catenation uniqueness is not verified for every word of L, orthogonal catenation of L_1 and L_2 is undefined, otherwise L_1 and L_2 are *orthogonal*. Daley *et al.* [32, 33] studied the state complexity of orthogonal catenation and generalized orthogonality to other operations. Although it is a restricted operation, its state complexity is only half of the one for the general catenation, i.e., $m2^{n-1} - 2^{n-2}$ for $m \geq 3$ and $n \geq 4$. The tight bound was obtained for languages over a four-symbol alphabet. Concerning nondeterministic state complexity, one has $nsc(L_1 \odot_\perp L_2) = nsc(L_1) + nsc(L_2)$, which coincides with the one for (general) catenation. Witness languages presented for the catenation are orthogonal (see page 7), thus apply to orthogonal catenation.

Unique Regular Operations Similar to orthogonality is the concept of *unique operation* introduced by Rampersad *et al.* [118]. However, instead of demanding that every pair of words of the operand languages lead to a distinct word on the resulting language, the language resulting from a *unique operation* only contains the words that are uniquely obtained through the given operation. Rampersad *et al.* studied several properties of unique operations and of their *poly* counterpart (i.e. where each resulting word must be obtained in more than one way), such as closure, ambiguity, and membership and non-emptiness decision problems. Results on state complexity and nondeterministic state complexity where obtained for *unique union* ($L_1 \mathring{\cup} L_2$), *unique catenation* ($L_1 \circ L_2$), *unique square* ($L \circ L = L^{\circ 2}$), and *unique star* (L°). The state complexity of $L_1 \mathring{\cup} L_2$ is mn, and witness binary languages are $\{x \in \{a,b\} \mid \#_a(x) = (m-1) \pmod{m}\}$ and $\{x \in \{a,b\} \mid \#_b(x) = (n-1) \pmod{n}\}$, for $m, n \geq 3$ (that were also used by Maslov [99] for general union). For unique catenation, $sc(L_1 \circ L_2) \leq m3^n - f_1 3^{n-1}$ which is much higher than the one for general catenation. It is an open problem to know if this bound is tight, although several examples, for specific values of m and n, were presented. However, for the unique square $sc(L^{\circ 2}) = m3^m - 3^{m-1}$, and the bound is tight for binary languages and $m \geq 3$. For the nondeterministic state complexity of unique concatenation a exponential lower bound was provided. An upper bound for the state complexity of the unique star is $3^{m-1} + (f+2)3^{m-f-1} - (2^{m-1} + 2^{m-f-1} - 2)$. But, again, it is an open problem to know if this upper bound is tight.

2.3 Other Subregular Languages

Besides finite and unary languages, several other subregular languages are used in many applications and are now theoretically well studied. Prefix-free or suffix-free languages are examples of codes that are fundamental in coding theory [4, 89]. Prefix-closed, factor-closed, or subword-closed languages were studied by several authors [58, 137, 34, 51]. These languages belong to a boarder set of languages, the *convex languages*, for which a general framework have been recently addressed by Ang and Brzozowski [1] and Brzozowski *et al.* [20]. A detailed survey on complexity topics was presented by Brzozowski [11]. Partially based on that survey, here we summarize some of the results concerning the state complexity of preserving regularity operations over some of the convex subregular languages.

Star-free languages are other family of subregular languages well studied [102, 131]. We will briefly address recent results on the state complexity of basic regular operations on these languages.

Convex Subregular Languages We begin by some definitions and results on determination for these languages.

Let $\trianglelefteq$ be a partial order on $\Sigma^\star$, and let $\trianglerighteq$ be its converse. A language L is $\trianglelefteq$-convex if $u \trianglelefteq v$ and $v \trianglelefteq w$ with $u, w \in L$ implies $v \in L$. It is $\trianglelefteq$-free if $v \trianglelefteq w$ and $w \in L$ implies $v \notin L$. It is $\trianglelefteq$-closed if $v \trianglelefteq w$ and $w \in L$ implies $v \in L$. It is $\trianglerighteq$-closed if $v \trianglerighteq w$ and $w \in L$ implies $v \in L$. The closure and the converse closure operations are:

$$_{\trianglelefteq}L = \{v \mid v \trianglelefteq w \text{ for some } w \in L\},$$

$$L_{\trianglelefteq} = \{v \mid w \trianglelefteq v \text{ for some } \ w \in L\}.$$

The *freeness* operation, $L^{\trianglelefteq}$ can defined for a language L, by

$$L^{\trianglelefteq} \subseteq L \text{ and } \forall w \in L^{\trianglelefteq}, \forall v \in \Sigma^\star,\ v \triangleleft w \text{ implies } v \notin L^{\trianglelefteq}.$$

The following proposition is from [1], except for the last item.

Proposition 1 *Let $\trianglelefteq$ be an arbitrary relation on $\Sigma^\star$. Then*

1. *A language is $\trianglelefteq$-convex if and only if it is $\trianglerighteq$-convex.*
2. *A language is $\trianglelefteq$-free if and only if it is $\trianglerighteq$-free.*
3. *Every $\trianglelefteq$-closed language and every $\trianglerighteq$-closed language is $\trianglelefteq$-convex.*
4. *A language is $\trianglelefteq$-closed if and only if its complement is $\trianglerighteq$-closed.*
5. *A language L is $\trianglelefteq$-closed ($\trianglerighteq$-closed) if and only if $L =_{\trianglelefteq} L$ ($L = L_{\trianglelefteq}$).*
6. *A language L is $\trianglelefteq$-free if and only if $L = L^{\trianglelefteq}$.*

We consider $\trianglelefteq$ to be:

- $\leq$: if $u, v, w \in \Sigma^\star$ and $w = uv$, then u is *prefix* of w, and we write $u \leq w$.
- $\preceq$: if $u, v, w \in \Sigma^\star$ and $w = uv$, then v is *suffix* of w, and we write $v \preceq w$
- $\sqsubseteq$: if $u, v, w \in \Sigma^\star$ and $w = uxv$, then x is *factor* of w, and we write $x \sqsubseteq w$. Note that a prefix or suffix of w is also a factor of w. This relation is also called *infix*.
- $\Subset$: if $w = w_0a_1w_1 \cdots a_nw_n$, where $a_1, \ldots, a_n \in \Sigma$, and $w_0, \ldots, w_n \in \Sigma^\star$, then $v = a_1 \cdots a_n$ is a *subword* of w; and we write $v \Subset w$. Note that every factor of w is a subword of w.

If a language is both prefix- and suffix-convex it is *bifix-convex.* In the same way are defined *bifix-free* and *bifix-closed* languages.

Ideals are languages directly related with closed languages. A non-empty language $L \subseteq \Sigma^\star$ is a

- *right ideal* if $L = L\Sigma^\star$ (also called *ultimate definite* [111]); the complement is prefix converse-closed.
- *left ideal* if $L = \Sigma^\star L$ (also called *reverse ultimate definite* [111]); the complement is suffix converse-closed.
- *two-sided ideal* if $L = \Sigma^\star L \Sigma^\star$ (also called *central definite*); the complement is bifix converse-closed.
- *all-sided ideal* if $L = \Sigma^\star \sqcup\!\sqcup L$; the complement is subword converse-closed; also studied by Haines [58] and Thierrin [137].

Free					
$\leq$	$\|\Sigma\|$	$\preceq$	$\|\Sigma\|$	$\sqsubseteq$	$\|\Sigma\|$
$2^{m-1}+1$	3	$2^{m-1}+1$	3	$2^{m-2}+2$	3
$]m, 2^{m-1}+1]$		$]m, 2^{m-1}+1]$		$]m, 2^{m-2}+2]$	
Closed					
$\leq$	$\|\Sigma\|$	$\preceq$	$\|\Sigma\|$	$\sqsubseteq$	$\|\Sigma\|$
2^m	3	$2^{m-1}+1$	4	$2^{m-1}+1$	4
$]m, 2^m]$		$[m, 2^{m-1}+1]$		$]m, 2^{m-1}+1]$	
Ideal					
right	$\|\Sigma\|$	left	$\|\Sigma\|$	two-sided	$\|\Sigma\|$
2^{m-1}	2	$2^{m-1}+1$	3	$2^{m-2}+1$	3

Table 6. State complexity of determination of free, closed and ideal languages considering prefix, suffix and factor partial orders, respectively. For each free and closed of languages, the range of correspondent non-magic numbers appears on the second row.

Some of the languages defined above are also characterized in terms of properties of the finite automata that accept them. In particular: prefix-closed languages are accepted by NFA' s where all states are final; suffix-closed languages are accepted by NFA's where all states are initial; factor-closed languages are accepted by NFA's where all states are initial and final; prefix-free languages are accepted by non-exiting NFA's (i.e. there are no transitions from the final states); suffix-free languages are accepted by non-returning NFA's (i.e. there are no transitions to the initial state); and factor-free languages are accepted by non-returning and non-exiting NFA's.

The state complexity of the determination on some subregular languages (or for the kind of NFA's they are defined by) was recently studied by Bordihn *et al.* [8], Jui-Yi Kao *et al.* [90], and Jirásková *et al.* [83]. Table 6 presents some of the values for the languages considered above. The existence of magic numbers for some subregular languages was studied by Holzer *et al.* [63]. As can be seen in Table 6, m is the only magic number for all free languages and for both prefix- and factor-closed languages (except if $m = 1$, where m is non-magic). Suffix-closed languages have no magic numbers.

Prefix-free				
	sc	$\|\Sigma\|$	nsc	$\|\Sigma\|$
$L_1 \cup L_2$	$mn - 2$	2	$m + n$	2
$L_1 \cap L_2$	$mn - 2(m + n - 3)$	2	$mn - (m + n) + 2$	1
$\overline{L}$	m	1	2^{m-1}	3
$(L_1 - L_2)$	$mn - m - 2m + 4$	3	$(m - 1)2^{n-1} + 1$	4
$(L_1 \oplus L_2)$	$mn - 2$	2		
L_1L_2	$m + n - 2$	1	$m + n - 1$	1
$L^\star$	m $m - 2$	2 1	m	1
L^+				
L^R	$2^{m-2} + 1$	3	m	1
$L_2 \setminus L_1$				
L_1/L_2				
L^{CS}	$(2m - 3)^{m-2}$	6	$2m^2 - 4m + 3$	2

Table 7. State complexity and nondeterministic state complexity of some operations on prefix-free languages

Free languages Table 7 summarizes state complexity results of individual operations on prefix-free languages [62, 83, 16]. In the case of state complexity, the results are valid for boolean operations if $m, n \geq 3$; for catenation if $m, n \geq 2$; for star if $k = 1$, then $m \geq 3$, if $k = 2$ then $m \neq 3$, and else $m \geq 2$; and for star if $m \geq 4$ and the tight bound cannot be reached if $k = 2$ [83]. Note that the state complexity of the catenation and the star are much lower than on general regular languages.

Suffix-free				
	sc	$\|\Sigma\|$	nsc	$\|\Sigma\|$
$L_1 \cup L_2$	$mn - (m+n-2)$	2	$m+n-1$	2
$L_1 \cap L_2$	$mn - 2(m+n-3)$	2	$mn - (m+n-2)$	2
$\overline{L}$				
$L_1 - L_2$	$mn - (m+2n-4)$	4		
$L_1 \oplus L_2$	$mn - (m+n-2)$	5		
L_1L_2	$(m-1)2^{n-2}+1$	4		
$L^\star$	$2^{m-2}+1$	4		
L^R	$2^{m-2}+1$	3		

Table 8. State complexity and nondeterministic state complexity of some operations on suffix-free languages

Table 8 summarizes the state complexity of some regular operations on suffix-free languages. Han and Salomaa showed that all bounds, except for difference and symmetric difference, were tight [60, 61]. Jirásková and Olejár [87] provided binary witnesses for intersection and union. They also proved that for all integer α between 1 and the respective bound there are languages L_1 and L_2 such that $(n)sc(L_1 \circ L_2) = \alpha$, for $\circ \in \{\cap, \cup\}$ (and witnesses are all ternary, except for $L_1 \cap L_2$ for which they are over a four-symbol alphabet). The bounds for difference and symmetric difference are from Brzozowski and Jirásková [16].

Free				
		$\leq \cup \preceq$	$\sqsubseteq$	$\Subset$
	sc	$\|\Sigma\|$		
$L_1 \cup L_2$	$mn - m - n$	5	5	$\geq m+n-3$
$L_1 \cap L_2$	$mn - 3m - 3n + 12,\ m,n \geq 4$	3	3	$m+n-7$
$L_1 - L_2$	$mn - 2m - 3n + 9$	4	4	$\geq m+n-6$
$L_1 \oplus L_2$	$mn - m - n$	5	5	$m+n-3$
L_1L_2	$m+n-2,\ m,n > 1$	1	1	1
$L^\star$	$m-1,\ m > 2$	2	2	2
L^R	$2^{m-3}+2,\ m \geq 3$	2	2	$2^{m-3}-1$

Table 9. State complexity of basic operations on bifix-, factor-, and subword-free languages

If a language is subword-free then it is factor-free, and if it is factor-free then it is bifix-free. Table 9 summarizes the state complexity of some regular operations on bifix-, factor-, and subword-free languages [16]. The tight upper bounds for the state complexity of these operations on the three classes of languages coincide.

Closed Languages and Ideals Table 10 shows the state complexity results of some basic operations on prefix-, suffix-, factor-, and subword-closed languages [17]. A language is factor-closed if and only if it is subword-closed. So the state-complexity results of operations are the same for those classes. The state complexity of the closure on the respective partial orders is also considered. Subword and converse subword closures were first studied by Gruber and Holzer [56, 57] and Okhotin [110]. Brzozowski and Jirásková presented the tight upper bound. Given a regular language L with $sc(L) = m$, $nsc({}_{\Subset}L) = nsc(L_{\Subset}) = m$ and these upper bounds are tight for witness binary languages.

If L is a right (respectively, left, two-sided, all-sided) ideal, any language $G \subseteq \Sigma^\star$ such that $L = G\Sigma^\star$ (respectively, $L = \Sigma^\star G$, $L = \Sigma^\star G \Sigma^\star$, $L = \Sigma^\star \sqcup\!\sqcup G$) is a *generator* of L.

Prefix, suffix, and factor closures (respectively, ${}_{\leq}L$, ${}_{\preceq}L$, and ${}_{\sqsubseteq}L$) were studied by Kao *et al.* [90].

Brzozowski and Jirásková [15] studied state complexity on ideals. Table 11 presents the state complexity of basic operations on ideals. As stated before closed languages and ideals are related. In particular, the state complexity of basic operations on two-sided and all-sided ideals coincide.

Brzozowski [11] observed that for the four types of convex languages (prefix, suffix, factor and subword) the state complexity of the boolean operations is mn.

Closed							
	$\leq$	$\|\Sigma\|$	$\preceq$	$\|\Sigma\|$	$\sqsubseteq,\Subset$	$\|\Sigma\|_{\sqsubseteq}$	$\|\Sigma\|_{\Subset}$
$L_1 \cup L_2$	mn	4	mn	2	mn	2	2
$L_1 \cap L_2$	$mn-m-n+2$	2	mn	2	$mn-m-n+2$	2	2
$L_1 - L_2$	$mn-n+1$	2	mn	4	$mn-n+1$	2	2
$L_1 \oplus L_2$	mn	2	mn	2	mn	2	2
L_1L_2	$m2^{n-2}+2^{n-2}$	3	$mn-fn+f$	2	$m+n-1$	3	2
$L^\star$	$2^{m-2}+1$	3	m	2	2	2	2
L^R	2^{m-1}	2	$2^{m-1}+1$	3	$2^{m-2}+1$	3	$2m$
$_{\trianglelefteq}L$	m	1	2^{m-1}	2	2^m-1	2	
$_{\Subset}L$					$2^{m-2}+1$		$m-2$

Table 10. State complexity of some operations on prefix-, suffix-, factor-, and subword-closed languages. The last two columns correspond to factor and subword, respectively. The last but one row contains the state complexity of the closure of prefix, suffix, and factor respectively. The last row contains state complexity of the subword closure.

Ideal							
	right	$\|\Sigma\|$	left	$\|\Sigma\|$	-sided	$\|\Sigma\|_{\text{two}}$	$\|\Sigma\|_{\text{all}}$
$L_1 \cup L_2$	$mn-m-n+2$	2	mn	4	$mn-m-n+2$	2	2
$L_1 \cap L_2$	mn	2	mn	2	mn	2	2
$L_1 - L_2$	$mn-m+1$	2	mn	4	$mn-m+1$	2	2
$L_1 \oplus L_2$	mn	2	mn	2	mn	2	2
L_1L_2	$m+2^{n-2}$	1	$m+n-1$	1	$m+n-1$	1	3
$L^\star$	$m+1$	2	$m+1$	2	$m+1$	2	2
	If $\varepsilon \in L$, then $L = \Sigma^\star$ and $sc(L^\star) = 1$.						
L^R	2^{m-1}	2	$2^{m-1}+1$	3	$2^{m-2}+1$	3	$2m-4$

Table 11. State complexity of basic operations on ideals. The last two columns correspond to two-sided and all-sided ideals, respectively.

Unary convex languages In the case of unary languages, prefix, suffix, factor, and subword partial orders coincide. Table 12 summarizes the state complexity of basic operations on unary free, unary closed, unary ideals and unary convex languages.

Freeness Operations Here we analyse the state complexity of freeness operations for prefix, suffix, bifix and factor orders that were studied by Pribavkina and Rodaro [115]. Given a regular language L, the $\trianglelefteq$-free language $L^{\trianglelefteq}$ for $\trianglelefteq \in \{\leq, \preceq, \sqsubseteq\}$, is respectively [1]:

- prefix: $L^{\leq} = L - L\Sigma^+$
- suffix: $L^{\preceq} = L - \Sigma^+L$
- factor: $L^{\sqsubseteq} = L - (\Sigma^+L\Sigma^\star \cup \Sigma^\star L\Sigma^+)$

The bifix operation is defined by $L^b = L^{\leq} \cap L^{\preceq}$. If L is an ideal, prefix, suffix and factor operations were studied by Brzozowski and Jirásková [15]. In this case, the resulting languages are minimal generators for left, right and two sided ideals, respectively. Table 13 presents the state complexity of prefix, suffix, factor and bifix operations on regular languages (and correspondent ideals). The state complexity of this operations is much lower in the case of right and two-sided ideals than for general regular languages.

Star-free languages Star-free languages are the smallest class containing the finite languages and closed under boolean operations and catenation. This class of languages correspond exactly to the

[1] In [115] the superscripts for prefix, suffix and factor operations were respectively p, s and ι.

Unary				
	Free	Closed	Ideal	Convex
$L_1 \cup L_2$	$\max\{m,n\}$	$\max\{m,n\}$	$\min\{m,n\}$	$\max\{m,n\}$
$L_1 \cap L_2$	$m = n$	$\min\{m,n\}$	$\max\{m,n\}$	$\max\{m,n\}$
$L_1 - L_2$	m	m	n	$\max\{m,n\}$
$L_1 \oplus L_2$	$\max\{m,n\}$	$\max\{m,n\}$	$\max\{m,n\}$	$\max\{m,n\}$
$L_1 L_2$	$m+n-2$	$m+n-2$	$m+n-1$	$m+n-1$
$L^\star$	$m-2$	2	$m-1$	$n^2 - 7n + 13$
L^R	m	m	m	m

Table 12. State complexity of basic operations on unary convex languages

	Regular		Ideal	
	sc	$\lvert\Sigma\rvert$	sc	$\lvert\Sigma\rvert$
$L^{\leq}$	$m+1$	2	$m+1$	2
$L^{\preceq}$	$(m-1)2^{m-2}+2,\ m \geq 4$	4	$\frac{n(n-1)}{2}+2$	1
$L^{\sqsubseteq}$	$(m-2)2^{m-3}+3,\ m \geq 4$	3	$n+1$	1
L^{b}	$(m-2)2^{m-2}+3,\ m \geq 4$	4		

Table 13. State complexity of prefix, suffix, factor and bifix operations on regular languages and on ideals (right, left and two sided, respectively).

regular languages of star height 0. The minimal DFA's of star-free languages are *permutation-free* (i.e. no word performs a non-trivial permutation of a subset of its states). Bordhin *et al.* [8] showed that the state complexity of the determination of a star-free language L is $2^{nsc(L)}$. Figure 9 presents a family of ternary NFA's for which the bound is tight. Holzer *et al.* [63] showed that star-free languages have no magic numbers.

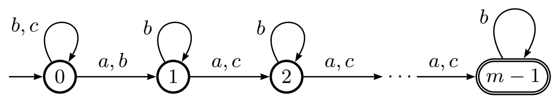

Fig. 9. Minimal m-state NFA's with equivalent minimal 2^m-state DFA for star-free languages

Brzozowski and Liu [19] studied the state complexity of the basic regular operations on star-free languages, and their results are summarized in Table 14. The bounds obtained for general regular languages are reached except in the catenation for $n = 2$, the reversal, and operations on unary languages.

2.4 Some more results

Here we briefly cite some more work on operational state complexity. Câmpeanu and Ho [21] and Brzozowski and Konstantinidis [18] considered uniform finite languages. Krieger *et al.* studied decimations of languages [92]. Câmpeanu and Konstantinidis [23] analysed a subword closure operation. Union-free languages were considered by Jirásková and Masopust [84]. Bassino *et al.* [3] provided upper bounds of the state complexity of basic operations on cofinite languages as a function of the size the complementary finite language (take as the summation of the lengths of all its words). The average state complexity on finite languages is addressed in two works. Gruber and Holzer [55] analysed the average state complexity of DFA's and NFA's based on a uniform distribution over finite languages whose longest word is of length at most n. Based on the size of finite languages as the summation of the lengths of all its words and a correspondent uniform distribution, Bassino *et al.* [2] establish that the average state complexities of the basic regular operations are asymptotically linear.

Star-free			
	sc	$\|\Sigma\|$	Unary
$L_1 \circ L_2$	mn	2	$\max\{m, n\}$
$L_1 L_2$	$(m-1)2^n + 2^{n-1}$, if $n \geq 3$ $[3m-2, 3m-1]$, if $n = 2$	4 3	$m + n - 1$
$L^\star$	2, if $m = 1$ $2^{m-1} + 2^{m-2}$, if $m \geq 2$	1 4	2, if $m = 1$ m, if $m \in [2, 5]$ $m^2 - 7m + 13$, if $m > 5$
L^R	$2^m - 1$	$m - 1$	m

Table 14. State complexity of basic regular operations on star-free regular and unary languages, where $\circ \in \{\cup, \cap, \setminus, \oplus\}$. For non-unary star-free languages and $n = 2$, $m \geq 2$. For non-unary star-free languages if $m \in [1, 2]$, the bound for reversal is tight for $|\Sigma| \geq m$.

References

1. Thomas Ang and Janusz A. Brzozowski. Languages convex with respect to binary relations, and their closure properties. *Acta Cybern.*, 19(2):445–464, 2009.
2. Frédérique Bassino, Laura Giambruno, and Cyril Nicaud. The average state complexity of rational operations on finite languages. *Int. J. Found. Comput. Sci.*, 21(4):495–516, 2010.
3. Frédérique Bassino, Laura Giambruno, and Cyril Nicaud. Complexity of operations on cofinite languages. In Alejandro López-Ortiz, editor, *LATIN 2010: Theoretical Informatics, 9th Latin American Symposium, Oaxaca, Mexico, April 19-23, 2010. Proceedings*, volume 6034 of *Lecture Notes in Computer Science*, pages 222–233. Springer, 2010.
4. Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*. Number 129 in Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2009.
5. Jean-Camille Birget. Intersection of regular languages and state complexity. *ACM SIGACT News*, 22(2), 1991.
6. Jean-Camille Birget. Partial orders on words, minimal elements of regular languages and state complexity. *Theor. Comput. Sci.*, 119(2):267–291, 1993.
7. Jean-Camille Birget. The state complexity of $\overline{\Sigma^\star \overline{L}}$ and its connection with temporal logic. *Inf. Process. Lett.*, 58(4):185–188, 1996.
8. Henning Bordihn, Markus Holzer, and Martin Kutrib. Determination of finite automata accepting subregular languages. *Theor. Comput. Sci.*, 410(35):3209–3222, 2009.
9. J. A. Brzozowski. Derivatives of regular expressions. *JACM*, 11(4):481–494, October 1964.
10. Janusz A. Brzozowski. Quotient complexity of regular languages. *CoRR*, abs/0907.4547, 2009.
11. Janusz A. Brzozowski. Complexity in convex languages. In Adrian Horia Dediu, Henning Fernau, and Carlos Martín-Vide, editors, *Language and Automata Theory and Applications, 4th International Conference, LATA 2010, Trier, Germany, May 24-28, 2010. Proceedings*, volume 6031 of *Lecture Notes in Computer Science*, pages 1–15. Springer, 2010.
12. Janusz A. Brzozowski. Quotient complexity of regular languages. *Journal of Automata, Languages and Combinatorics*, To appear.
13. Janusz A. Brzozowski, Elyot Grant, and Jeffrey Shallit. Closures in formal languages and Kuratowski's theorem. In Volker Diekert and Dirk Nowotka, editors, *Developments in Language Theory, 13th International Conference, DLT 2009, Stuttgart, Germany, June 30 - July 3, 2009. Proceedings*, volume 5583 of *Lecture Notes in Computer Science*, pages 125–144. Springer, 2009.
14. Janusz A. Brzozowski, Elyot Grant, and Jeffrey Shallit. Closures in formal languages: Concatenation, separation, and algorithms. *CoRR*, abs/0901.3763, 2009.
15. Janusz A. Brzozowski, Galina Jirásková, and Baiyu Li. Quotient complexity of ideal languages. In Alejandro López-Ortiz, editor, *LATIN 2010: Theoretical Informatics, 9th Latin American Symposium, Oaxaca, Mexico, April 19-23, 2010. Proceedings*, volume 6034 of *Lecture Notes in Computer Science*, pages 208–221. Springer, 2010.
16. Janusz A. Brzozowski, Galina Jirásková, and Joshua Smith. Quotient complexity of bifix-, factor-, and subword-free languages. *CoRR*, abs/1006.4843, 2010.
17. Janusz A. Brzozowski, Galina Jirásková, and Chenglong Zou. Quotient complexity of closed languages. In Farid M. Ablayev and Ernst W. Mayr, editors, *Computer Science - Theory and Applications, 5th International Computer Science Symposium in Russia, CSR 2010, Kazan, Russia, June 16-20, 2010. Proceedings*, volume 6072 of *Lecture Notes in Computer Science*, pages 84–95. Springer, 2010.

18. Janusz A. Brzozowski and Stavros Konstantinidis. State-complexity hierarchies of uniform languages of alphabet-size length. *Theor. Comput. Sci.*, 410(35):3223–3235, 2009.
19. Janusz A. Brzozowski and Bo Liu. Quotient complexity of star-free languages. *CoRR*, abs/1012.3962, 2010.
20. Janusz A. Brzozowski, Jeffrey Shallit, and Zhi Xu. Decision problems for convex languages. In Adrian Horia Dediu, Armand-Mihai Ionescu, and Carlos Martín-Vide, editors, *Language and Automata Theory and Applications, Third International Conference, LATA 2009, Tarragona, Spain, April 2-8, 2009. Proceedings*, volume 5457 of *Lecture Notes in Computer Science*, pages 247–258. Springer, 2009.
21. Cezar Câmpeanu and Wing Hong Ho. The maximum state complexity for finite languages. *Journal of Automata, Languages and Combinatorics*, 9(2/3):189–202, 2004.
22. Cezar Câmpeanu, Karel Culik II, Kai Salomaa, and Sheng Yu. State complexity of basic operations on finite languages. In Oliver Boldt and Helmut Jürgensen, editors, *Automata Implementation, 4th International Workshop on Implementing Automata, WIA'99, Potsdam, Germany, July 17-19, 1999, Revised Papers*, volume 2214 of *Lecture Notes in Computer Science*, pages 60–70. Springer, 2001.
23. Cezar Câmpeanu and Stavros Konstantinidis. State complexity of the subword closure operation with applications to DNA coding. *International Journal of Foundations of Computer Science*, 19(5):1099–1112, 2008.
24. Cezar Câmpeanu, Kai Salomaa, and Sheng Yu. Tight lower bound for the state complexity of shuffle of regular languages. *Journal of Automata, Languages and Combinatorics*, 7(3):303–310, 2002.
25. Cezar Câmpeanu, Nicolae Santean, and Sheng Yu. Minimal cover-automata for finite languages. *Theor. Comput. Sci.*, 267(1-2):3–16, 2001.
26. Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. Alternation. *J. ACM*, 28(1):114–133, 1981.
27. Marek Chrobak. Finite automata and unary languages. *Theor. Comput. Sci.*, 47(3):149–158, 1986.
28. Marek Chrobak. Errata to: "finite automata and unary languages". *Theor. Comput. Sci.*, 302(1-3):497–498, 2003.
29. John H. Conway. *Regular algebra and finite machines*. William Clowes& Sons, Great Britain, 1971.
30. Bo Cui, Yuan Gao, Lila Kari, and Sheng Yu. State complexity of catenation combined with star and reversal. *CoRR*, abs/1008.1648, 2010.
31. Bo Cui, Yuan Gao, Lila Kari, and Sheng Yu. State complexity of two combined operations: Reversal-catenation and star-catenation. *CoRR*, abs/1006.4646, 2010.
32. Mark Daley, Michael Domaratzki, and Kai Salomaa. State complexity of orthogonal catenation. *CoRR*, abs/0904.3366, 2009.
33. Mark Daley, Michael Domaratzki, and Kai Salomaa. Orthogonal concatenation: Language equations and state complexity. *Journal of Universal Computer Science*, 16(5):653–675, 2010.
34. A. de Luca and S. Varricchio. Some combinatorial properties of factorial languages. In R. Capocelli, editor, *Sequences*, pages 258–266. Springer, 1990.
35. A. de Luca and S. Varricchio. *Finiteness and Regularity in Semigroups and Formal Languages*. Springer, 1999.
36. Michael Domaratzki. State complexity of proportional removals. *Journal of Automata, Languages and Combinatorics*, 7(4):455–468, 2002.
37. Michael Domaratzki and Alexander Okhotin. State complexity of power. *Theor. Comput. Sci.*, 410(24-25):2377–2392, 2009.
38. Michael Domaratzki and Kai Salomaa. State complexity of shuffle on trajectories. *Journal of Automata, Languages and Combinatorics*, 9(2/3):217–232, 2004.
39. Juraj Sěbej. Reversal of regular languages and state complexity. Master's thesis, P. J. Šafárik, University of Košice, Slovakia, 2010.
40. Keith Ellul. Descriptional complexity measures of regular languages. Master's thesis, University of Waterloo, 2003.
41. Keith Ellul, Bryan Krawetz, Jeffrey Shallit, and Ming wei Wang. Regular expressions: New results and open problems. *Journal of Automata, Languages and Combinatorics*, 10(4):407–437, 2005.
42. Zoltán Ésik, Yuan Gao, Guangwu Liu, and Sheng Yu. Estimation of state complexity of combined operations. *Theor. Comput. Sci.*, 410(35):3272–3280, Oct 2030.
43. Yuan Gao, Zoltán Ésik, Guangwu Liu, and Sheng Yu. State complexity of multiple boolean and catenation operations. *Report*, UWO(715), 2008.
44. Yuan Gao, Kai Salomaa, and Sheng Yu. The state complexity of two combined operations: Star of catenation and star of reversal. *Fundamenta Informaticae*, 83:75–89, Jan 2008.
45. Yuan Gao, Kai Salomaa, and Sheng Yu. Transition complexity of incomplete dfas. *Fundam. Inform.*, 110(1-4):143–158, 2011.

46. Yuan Gao and Sheng Yu. State complexity approximation. *CoRR*, abs/0907.5124, 2009.
47. Yuan Gao and Sheng Yu. State complexity of union and intersection combined with star and reversal. *CoRR*, abs/1006.3755, 2010.
48. Viliam Geffert. (Non)determinism and the size of one-way finite automata. In Carlo Mereghetti, Beatrice Palano, Giovanni Pighizzini, and Detelf Wotschke, editors, *Proceedings of the 7th International Workshop on Descriptional Complexity of Formal Systems*, page 23–37. University of Milano, Italy, 2005.
49. Viliam Geffert. Magic numbers in the state hierarchy of finite automata. *Inf. Comput.*, 205(11):1652–1670, 2007.
50. Viliam Geffert. State hierarchy for one-way finite automata. *Journal of Automata, Languages and Combinatorics*, 12(1-2):139–145, 2007.
51. Arthur Gill and Lawrence T. Kou. Multiple-entry finite automata. *J. Comput. System Sci.*, 9:1–19, 1974.
52. Ian Glaister and Jeffrey Shallit. A lower bound technique for the size of nondeterministic finite automata. *Inf. Process. Lett.*, 59(2):75–77, 1996.
53. Jonathan Goldstine, Martin Kappes, Chandra M. R. Kintala, Hing Leung, Andreas Malcher, and Detlef Wotschke. Descriptional complexity of machines with limited resources. *J. UCS*, 8(2):193–234, 2002.
54. Hermann Gruber and Markus Holzer. Finding lower bounds for nondeterministic state complexity is hard. In Oscar H. Ibarra and Zhe Dang, editors, *Developments in Language Theory, 10th International Conference, DLT 2006, Santa Barbara, CA, USA, June 26-29, 2006, Proceedings*, volume 4036 of *Lecture Notes in Computer Science*, pages 363–374. Springer, 2006.
55. Hermann Gruber and Markus Holzer. On the average state and transition complexity of finite languages. *Theor. Comput. Sci.*, 387(2):155–166, 2007.
56. Hermann Gruber, Markus Holzer, and Martin Kutrib. The size of Higman-Haines sets. *Theor. Comput. Sci.*, 387(2):167–176, 2007.
57. Hermann Gruber, Markus Holzer, and Martin Kutrib. More on the size of Higman-Haines sets: Effective constructions. *Fundam. Inform.*, 91(1):105–121, 2009.
58. L. H. Haines. On free monoids partially ordered by embedding. *J. Combin. Theory*, 6:94–98, 1969.
59. Yo-Sub Han and Kai Salomaa. State complexity of union and intersection of finite languages. *Int. J. Found. Comput. Sci.*, 19(3):581–595, 2008.
60. Yo-Sub Han and Kai Salomaa. State complexity of basic operations on suffix-free regular languages. *Theor. Comput. Sci.*, 410(27-29):2537–2548, 2009.
61. Yo-Sub Han and Kai Salomaa. Nondeterministic state complexity for suffix-free regular languages. *CoRR*, abs/1008.1661, 2010.
62. Yo-Sub Han, Kai Salomaa, and Derick Wood. Nondeterministic state complexity of basic operations for prefix-free regular languages. *Fundam. Inform.*, 90(1-2):93–106, 2009.
63. Markus Holzer, Sebastian Jakobi, and Martin Kutrib. The magic number problem for subregular language families. *CoRR*, abs/1008.1653, 2010.
64. Markus Holzer and Martin Kutrib. State complexity of basic operations on nondeterministic finite automata. In Jean-Marc Champarnaud and Denis Maurel, editors, *Implementation and Application of Automata, 7th International Conference, CIAA 2002, Tours, France, July 3-5, 2002, Revised Papers*, volume 2608 of *Lecture Notes in Computer Science*, pages 148–157. Springer, 2003.
65. Markus Holzer and Martin Kutrib. Unary language operations and their nondeterministic state complexity. In Masami Ito and Masafumi Toyama, editors, *Developments in Language Theory, 6th International Conference, DLT 2002, Kyoto, Japan, September 18-21, 2002, Revised Papers*, volume 2450 of *Lecture Notes in Computer Science*, pages 162–172. Springer, 2003.
66. Markus Holzer and Martin Kutrib. Descriptional and computational complexity of finite automata. In Adrian Horia Dediu, Armand-Mihai Ionescu, and Carlos Martín-Vide, editors, *Language and Automata Theory and Applications, Third International Conference, LATA 2009, Tarragona, Spain, April 2-8, 2009. Proceedings*, volume 5457 of *Lecture Notes in Computer Science*, pages 23–42. Springer, 2009.
67. Markus Holzer and Martin Kutrib. Nondeterministic finite automata - recent results on the descriptional and computational complexity. *Int. J. Found. Comput. Sci.*, 20(4):563–580, 2009.
68. Markus Holzer, Kai Salomaa, and Sheng Yu. On the state complexity of k-entry deterministic finite automata. *Journal of Automata, Languages and Combinatorics*, 6(4):453–466, 2001.
69. M. Hricko, Galina Jirásková, and Alexander Szabari. Union and intersection of regular languages and descriptional complexity. In Carlo Mereghetti, Beatrice Palano, Giovanni Pighizzini, and Detlef Wotschke, editors, *Proceedings of the 7th International Workshop on Descriptional Complexity of Formal Systems*, pages 170–181, Milano, Italy, 2005. University of Milano.
70. Juraj Hromkovic. *Communication Complexity and Parallel Computing*. Springer, 1997.
71. Juraj Hromkovic. Descriptional complexity of finite automata: Concepts and open problems. *Journal of Automata, Languages and Combinatorics*, 7(4):519–531, 2002.

72. Masami Ito, Lila Kari, and Gabriel Thierrin. Shuffle and scattered deletion closure of languages. *Theor. Comput. Sci.*, 245(1):115–133, 2000.
73. Kazuo Iwama, Yahiko Kambayashi, and Kazuya Takaki. Tight bounds on the number of states of DFAs that are equivalent to n-state NFAs. *Theor. Comput. Sci.*, 237(1-2):485–494, 2000.
74. Kazuo Iwama, Akihiro Matsuura, and Mike Paterson. A family of NFA's which need 2^n -α deterministic states. In Mogens Nielsen and Branislav Rovan, editors, *Mathematical Foundations of Computer Science 2000, 25th International Symposium, MFCS 2000, Bratislava, Slovakia, August 28 - September 1, 2000, Proceedings*, volume 1893 of *Lecture Notes in Computer Science*, pages 436–445. Springer, 2000.
75. Tao Jiang and Bala Ravikumar. A note on the space complexity of some decision problems for finite automata. *Information Processing Letters*, 1:25–31, 1991.
76. Jozef Jirásek, Galina Jirásková, and Alexander Szabari. State complexity of concatenation and complementation of regular languages. *International Journal of Foundations of Computer Science*, 16(3):511–529, 2005.
77. Jozef Jirásek, Galina Jirásková, and Alexander Szabari. Deterministic blow-ups of minimal nondeterministic finite automata over a fixed alphabet. *Int. J. Found. Comput. Sci.*, 19(3):617–631, 2008.
78. Galina Jirásková. Note on minimal finite automata. In Jiri Sgall, Ales Pultr, and Petr Kolman, editors, *Mathematical Foundations of Computer Science 2001, 26th International Symposium, MFCS 2001 Marianske Lazne, Czech Republic, August 27-31, 2001, Proceedings*, volume 2136 of *Lecture Notes in Computer Science*, pages 421–431. Springer, 2001.
79. Galina Jirásková. State complexity of some operations on binary regular languages. *Theor. Comput. Sci.*, 330(2):287–298, 2005.
80. Galina Jirásková. On the state complexity of complements, stars, and reversals of regular languages. In Masami Ito and Masafumi Toyama, editors, *Developments in Language Theory, 12th International Conference, DLT 2008, Kyoto, Japan, September 16-19, 2008. Proceedings*, volume 5257 of *Lecture Notes in Computer Science*, pages 431–442. Springer, 2008.
81. Galina Jirásková. Concatenation of regular languages and descriptional complexity. In Anna E. Frid, Andrey Morozov, Andrey Rybalchenko, and Klaus W. Wagner, editors, *Computer Science - Theory and Applications, Fourth International Computer Science Symposium in Russia, CSR 2009, Novosibirsk, Russia, August 18-23, 2009. Proceedings*, volume 5675 of *Lecture Notes in Computer Science*, pages 203–214. Springer, 2009.
82. Galina Jirásková. Magic numbers and ternary alphabet. In Volker Diekert and Dirk Nowotka, editors, *Developments in Language Theory, 13th International Conference, DLT 2009, Stuttgart, Germany, June 30 - July 3, 2009. Proceedings*, volume 5583 of *Lecture Notes in Computer Science*, pages 300–311. Springer, 2009.
83. Galina Jirásková and Monika Krausová. Complexity in prefix-free regular languages. *Electron. Proc. Theor. Comput. Sci.*, 31:197–204, Aug 2010.
84. Galina Jirásková and Tomás Masopust. Complexity in union-free regular languages. In Yuan Gao, Hanlin Lu, Shinnosuke Seki, and Sheng Yu, editors, *Developments in Language Theory, 14th International Conference, DLT 2010, London, ON, Canada, August 17-20, 2010. Proceedings*, volume 6224 of *Lecture Notes in Computer Science*, pages 255–266. Springer, 2010.
85. Galina Jirásková and Alexander Okhotin. State complexity of cyclic shift. *Rairo-Theoretical Informatics nd Applications*, 42:335–360, Mar 2008.
86. Galina Jirásková and Alexander Okhotin. Nondeterministic state complexity of positional addition. *CoRR*, abs/0907.5072, 2009.
87. Galina Jirásková and Palov Olejár. State complexity of union and intersection of binary suffix-free languages. Workshop on Non-Classical Models of Automata and Applications (NCMA),August 31 - September 1, 2009, Wroclaw, Poland, 2009.
88. Galina Jirásková and Juraj Sebej. Note on reversal of binary regular languages. In Markus Holzer, Martin Kutrib, and Giovanni Pighizzini, editors, *Descriptional Complexity of Formal Systems - 13th International Workshop, DCFS 2011. Proceedings*, volume 6808 of *Lecture Notes in Computer Science*, pages 212–221, Gießen/Limburg, Germany, July 2011. Springer.
89. Helmut Jürgensen and Stavros Konstantinidis. Codes. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, volume 1, pages 511–607. Springer, 1997.
90. Jui-Yi Kao, Narad Rampersad, and Jeffrey Shallit. On NFAs where all states are final, initial, or both. *Theor. Comput. Sci.*, 410(47-49):5010–5021, 2009.
91. George Anton Kiraz. Compressed storage of sparse finite-state transducers. In Oliver Boldt and Helmut Jürgensen, editors, *Automata Implementation, 4th International Workshop on Implementing Automata, WIA'99, Potsdam, Germany, July 17-19, 1999, Revised Papers*, volume 2214 of *Lecture Notes in Computer Science*, pages 109–121. Springer, 2001.

92. Dalia Krieger, Avery Miller, Narad Rampersad, Bala Ravikumar, and Jeffrey Shallit. Decimations of languages and state complexity. *Theor. Comput. Sci.*, 410(24-25):2401–2409, 2009.
93. Ernst L. Leiss. Succint representation of regular languages by boolean automata. *Theor. Comput. Sci.*, 13:323–330, 1981.
94. Ernst L. Leiss. Succinct representation of regular languages by boolean automata ii. *Theor. Comput. Sci.*, 38:133–136, 1985.
95. O. B. Lupanov. A comparison of two types of finite sources. *Problemy Kibernetiki*, 9:321–326, 1963. German translation: Über den Vergleich zweier Typen endlicher Quellen. Probleme der Kybernetik 6 (1966), 328–335.
96. Yu I. Lyubich. Estimates for optimal determinization of nondeterministic autonomous automata. *Sibirskii Matematichskii Zhurnal*, 5:337–355, 1964. (in Russian).
97. Yu I. Lyubich. Estimates of the number of states that arise in the determinization of a nondeterministic autonomous automaton. *Dokllady Akademii Nauk SSSR*, pages 41–43, 1964. English translation in Soviet Mathematics Doklady, 5, 345–348 (1964).
98. Robert Mandl. Precise bounds associated with the subset construction on various classes of nondeterministic finite automata. In *Princeton Conference on Information and System Sciences*, page 263–267, 1973.
99. A. N. Maslov. Estimates of the number of states of finite automata. *Dokllady Akademii Nauk SSSR*, 194:1266–1268, 1970. English translation in Soviet Mathematics Doklady, 11 , 1373-1375 (1970).
100. A. N. Maslov. Cyclic shift operation for languages. *Probl. Inf. Transm*, 9:333–338, 1973.
101. Alexandru Mateescu, Grzegorz Rozenberg, and Arto Salomaa. Shuffle on trajectories: Syntactic constraints. *Theor. Comput. Sci.*, 197(1-2):1–56, 1998.
102. Robert McNaughton and Seymour Papert. Counter-free automata. In *Research Monographs*, volume 65. MIT Press, 1971.
103. Filippo Mera and Giovanni Pighizzini. Complementing unary nondeterministic automata. *Theor. Comput. Sci.*, 330(2):349–360, 2005.
104. Carlo Mereghetti and Giovanni Pighizzini. Optimal simulations between unary automata. *SIAM J. Comput.*, 30(6):1976–1992, 2000.
105. Albert R. Meyer and Michael J. Fischer. Economy of description by automata, grammars, and formal systems. In *Conference Record 1971 Twelfth Annual Symposium on Switching and Automata Theory, 13-15 October 1971, East Lansing, Michigan, USA*, pages 188–191, Los Alamitos, 1971. IEEE Society Press.
106. Boris G. Mirkin. On dual automata. *Kibernetika*, 2:7–10, 1966. English translation in Cybernetics 2, 6–9 (1966).
107. Frank R. Moore. On the bounds for state-set size in the proofs of equivalence between deterministic, nondeter- ministic, and two-way finite automata. *IEEE Transactions on computers*, 2:1211–1214, 1971.
108. Cyril Nicaud. Average state complexity of operations on unary automata. In Miroslaw Kutylowski, Leszek Pacholski, and Tomasz Wierzbicki, editors, *Mathematical Foundations of Computer Science 1999, 24th International Symposium, MFCS'99, Szklarska Poreba, Poland, September 6-10, 1999, Proceedings*, volume 1672 of *Lecture Notes in Computer Science*, pages 231–240. Springer, 1999.
109. Cyril Nicaud. *Étude du comportement en moyenne des automates finis et des languages rationels*. PhD thesis, Université de Paris 7, 2000.
110. Alexander Okhotin. On the state complexity of scattered substrings and superstrings. *Fundam. Inform.*, 99(3):325–338, 2010.
111. Azaria Paz and Bezalel Peleg. Ultimate-definite and symmetric-definite events and automata. *J. ACM*, 12(3):399–410, 1965.
112. Mika Perles, Michael O. Rabin, and E. Shamir. The theory of definite automata. *Electronic Computers, IEEE Transactions on*, EC-12(3):233 –243, june 1963.
113. Giovanni Pighizzini. Unary language concatenation and its state complexity. In Sheng Yu and Andrei Paun, editors, *Implementation and Application of Automata, 5th International Conference, CIAA 2000, London, Ontario, Canada, July 24-25, 2000, Revised Papers*, volume 2088 of *Lecture Notes in Computer Science*, pages 252–262. Springer, 2001.
114. Giovanni Pighizzini and Jeffrey Shallit. Unary language operations, state complexity and Jacobsthal's function. *International Journal of Foundations of Computer Science*, 13(1):145–159, 2002.
115. Elena V. Pribavkina and Emanuele Rodaro. State complexity of prefix, suffix, bifix and infix operators on regular languages. In Yuan Gao, Hanlin Lu, Shinnosuke Seki, and Sheng Yu, editors, *Developments in Language Theory*, volume 6224 of *Lecture Notes in Computer Science*, pages 376–386. Springer, 2010.
116. Michael O. Rabin and Dana Scott. Finite automata and their decision problems. *IBM J Research Development*, 3(2):114–129, 1959.

117. Narad Rampersad. The state complexity of L^2 and L^k. *Information Proceesing Lettes*, 98:231–234, 2006.
118. Narad Rampersad, Jeffrey Shallit, and Bala Ravikumar. State complexity of unique rational operations. *Theor. Comput. Sci.*, 410:2431–2441, 2009.
119. Bala Ravikumar. Some applications of a technique of sakoda and sipser. *SIGACT News*, 21(4):73–77, 1990.
120. Bala Ravikumar and Oscar H. Ibarra. Relating the type of ambiguity of finite automata to the succinctness of their representation. *SIAM J. Comput.*, 18(6):1263–1282, 1989.
121. Dominique Revuz. Minimisation of acyclic deterministic automata in linear time. *Theor. Comput. Sci.*, 92(1):181–189, 1992.
122. William J. Sakoda and Michael Sipser. Nondeterminism and the size of two way finite automata. In *Conference Record of the Tenth Annual ACM Symposium on Theory of Computing, 1-3 May 1978, San Diego, California, USA*, pages 275–286. ACM, 1978.
123. Arto Salomaa. On the reducibility of events represented in automata. *Annales Academiae Scientiarum Fennicae*, Series A(I. Mathematica 353), 1964.
124. Arto Salomaa. Theorems on the representation of events in Moore-automata. *Turun Yliopiston Julkaisuja Annales Universitatis Turkuensis*, Series A(69), 1964.
125. Arto Salomaa. *Theory of Automata*. Pergamon Press, Oxford, 1969.
126. Arto Salomaa, Kai Salomaa, and Sheng Yu. State complexity of combined operations. *Theor. Comput. Sci.*, 383(2-3):140–152, 2007.
127. Arto Salomaa, Derick Wood, and Sheng Yu. On the state complexity of reversals of regular languages. *Theor. Comput. Sci.*, 320(2-3):315–329, 2004.
128. Kai Salomaa. Descriptional complexity of nondeterministic finite automata. In Tero Harju, Juhani Karhumäki, and Arto Lepistö, editors, *Developments in Language Theory, 11th International Conference, DLT 2007, Turku, Finland, July 3-6, 2007, Proceedings*, volume 4588 of *Lecture Notes in Computer Science*, pages 31–35. Springer, 2007.
129. Kai Salomaa and Sheng Yu. NFA to DFA transformation for finite languages over arbitrary alphabets. *Journal of Automata, Languages and Combinatorics*, 2(3):177–186, 1997.
130. Kai Salomaa and Sheng Yu. On the state complexity of combined operations and their estimation. *International Journal of Foundations of Computer Science*, 18(4):683–698, Jul 2007.
131. M.P. Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965.
132. Joel I. Seiferas and Robert McNaughton. Regularity-preserving relations. *Theor. Comput. Sci.*, 2(2):147–154, 1976.
133. Jeffrey Shallit. State complexity and Jacobsthal's function. In Sheng Yu and Andrei Paun, editors, *Implementation and Application of Automata, 5th International Conference, CIAA 2000, London, Ontario, Canada, July 24-25, 2000, Revised Papers*, volume 2088 of *Lecture Notes in Computer Science*, pages 272–278. Springer, 2001.
134. Huei-Jan Shyr. *Free Monoids and Languages*. Hon Min Book Co., Taiwan, 2001.
135. Michael Sipser. Lower bounds on the size of sweeping automata. *J. Comput. Syst. Sci.*, 21(2):195–202, 1980.
136. Richard E. Stearns and Juris Hartmanis. Regularity preserving modifications of regular expressions. *Information and Control*, 6(1):55–69, 1963.
137. Gabriel Thierrin. Convex languages. In M. Nivat, editor, *ICALP*, pages 481–492. North-Holland, 1972.
138. Sheng Yu. Regular languages. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, volume 1, pages 41–110. Springer, 1997.
139. Sheng Yu. State complexity of regular languages. *Journal of Automata, Languages and Combinatorics*, 6(2):221–234, 2001.
140. Sheng Yu. State complexity of finite and infinite regular languages. *Bulletin of the EATCS*, 76:142–152, 2002.
141. Sheng Yu. State complexity: Recent results and open problems. *Fundam. Inform.*, 64(1-4):471–480, 2005.
142. Sheng Yu. On the state complexity of combined operations. In Oscar H. Ibarra and Hsu-Chun Yen, editors, *Implementation and Application of Automata, 11th International Conference, CIAA 2006, Taipei, Taiwan, August 21-23, 2006, Proceedings*, volume 4094 of *Lecture Notes in Computer Science*, pages 11–22. Springer, 2006.
143. Sheng Yu and Qingyu Zhuang. On the state complexity of intersection of regular languages. *ACM SIGACT News*, 22(7):52–54, 1991.
144. Sheng Yu, Qingyu Zhuang, and Kai Salomaa. The state complexities of some basic operations on regular languages. *Theor. Comput. Sci.*, 125(2):315–328, 1994.