Theoretical Computer Science 449 (2012) 37–53

Contents lists available at SciVerse ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Syntactic complexity of prefix-, suffix-, bifix-, and factor-free regular languages [☆]

Janusz Brzozowski [a], Baiyu Li [a,*], Yuli Ye [b]

[a] David R. Cheriton School of Computer Science, University of Waterloo. Waterloo, ON, Canada N2L 3G1

[b] Department of Computer Science, University of Toronto, Toronto, ON, Canada M5S 3G4

ARTICLE INFO

Keywords:
Bifix-free
Factor-free
Finite automaton
Monoid
Prefix-free
Regular language
Reversal
Semigroup
Suffix-free
Syntactic complexity

ABSTRACT

The syntactic complexity of a regular language is the cardinality of its syntactic semigroup. The syntactic complexity of a subclass of the class of regular languages is the maximal syntactic complexity of languages in that class, taken as a function of the state complexity n of these languages. We study the syntactic complexity of prefix-, suffix-, bifix-, and factor-free regular languages. We prove that n^{n-2} is a tight upper bound for prefix-free regular languages. We present properties of the syntactic semigroups of suffix-, bifix-, and factor-free regular languages, conjecture tight upper bounds on their size to be $(n-1)^{n-2}+(n-2)$, $(n-1)^{n-3}+(n-2)^{n-3}+(n-3)2^{n-3}$, and $(n-1)^{n-3}+(n-3)2^{n-3}+1$, respectively, and exhibit languages with these syntactic complexities.

1. Introduction

A language is *prefix-free* (respectively, *suffix-free*, *factor-free*) if it does not contain any pair of words such that one is a proper prefix (respectively, suffix, factor) of the other. It is *bifix-free* if it is both prefix- and suffix-free. We refer to prefix-, suffix-, bifix-, and factor-free languages as *free* languages. Nontrivial prefix-, suffix-, bifix-, and factor-free languages are also known as prefix, suffix, bifix, and infix codes [1,23], respectively, and have many applications in areas such as cryptography, data compression and information processing.

The *state complexity* of a regular language is the number of states in the minimal deterministic finite automaton (DFA) recognizing that language. An equivalent notion is that of *quotient complexity*, which is the number of left quotients of the language. State complexity of regular operations has been studied quite extensively; for surveys of this topic and lists of references we refer the reader to [3,25]. With regard to the state complexity of free regular languages, Han, Salomaa and Wood [11] examined prefix-free regular languages, and Han and Salomaa [10] studied suffix-free regular languages. Bifix- and factor-free regular languages were studied by Brzozowski et al. [4].

The notion of quotient complexity can be derived from the Nerode right congruence [18], while the Myhill congruence [17] leads to the syntactic semigroup of a language and to its *syntactic complexity*, which is the cardinality of the syntactic semigroup. It was pointed out in [6] that syntactic complexity can be very different for regular languages with the same quotient complexity. Thus, for a fixed n, languages with quotient complexity n may possibly be distinguished by their syntactic complexities.

☆ This work was supported by the Natural Sciences and Engineering Research Council of Canada under grant No. OGP0000871, a Postgraduate Scholarship, and by a Graduate Award from the Department of Computer Science, University of Toronto.

* Corresponding author. Tel.: +1 5197815017.
E-mail addresses: brzozo@uwaterloo.ca (J. Brzozowski), b5li@uwaterloo.ca (B. Li), y3ye@cs.toronto.edu (Y. Ye).
URLs: http://maveric.uwaterloo.ca/~brzozo/ (J. Brzozowski), http://cs.uwaterloo.ca/~b5li/ (B. Li), http://www.cs.utoronto.ca/~y3ye/ (Y. Ye).

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2012.04.011

In contrast to state complexity, syntactic complexity has not received much attention. In 1970, Maslov [15] stated without proof that n^n was a tight upper bound on the number of transformations performed by a DFA of n states; this number is the same as the syntactic complexity of the language of the DFA [16]. In 2003–2004, Holzer and König [12], and Krawetz et al. [14] studied the syntactic complexity of unary and binary languages. In 2010, Brzozowski and Ye [6] examined the syntactic complexity of ideal and closed regular languages, and in 2011, Brzozowski and Li [5] studied the syntactic complexity of star-free languages. Here, we deal with the syntactic complexity of prefix-, suffix-, bifix-, and factor-free regular languages, and their complements.

Basic definitions and facts are stated in Sections 2 and 3. In Section 4, we obtain a tight upper bound on the syntactic complexity of prefix-free regular languages. In Section 5–7, we study the syntactic complexity of suffix-, bifix-, and factor-free regular languages, respectively. We state conjectures about tight upper bounds for these classes, and exhibit languages in these classes that have large syntactic complexities. In Section 8, we show that the upper bounds on the quotient complexity of reversal of prefix-, suffix-, bifix-, and factor-free regular languages can be met by our languages with largest syntactic complexities. Section 9 concludes the paper.

2. Transformations

A *transformation* of a set Q is a mapping of Q into itself. In this paper, we consider only transformations of finite sets, and we assume without loss of generality that $Q = \{1, 2, \dots, n\}$. Let t be a transformation of Q. If $i \in Q$, then it is the *image* of i under t. If X is a subset of Q, then $Xt = \{it \mid i \in X\}$, and the *restriction* of t to X, denoted by $t|_X$, is a mapping from X to Xt such that $it|_X = it$ for all $i \in X$. The *composition* of two transformations t_1 and t_2 of Q is a transformation $t_1 \circ t_2$ such that $i(t_1 \circ t_2) = (it_1)t_2$ for all $i \in Q$. We usually drop the composition operator "$\circ$" and write t_1t_2 for short. An arbitrary transformation can be written in the form

$$t = \begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ i_1 & i_2 & \cdots & i_{n-1} & i_n \end{pmatrix},$$

where $i_k = kt$, $1 \leqslant k \leqslant n$, and $i_k \in Q$. The *domain* $\mathrm{dom}(t)$ of t is Q. The *range* $\mathrm{rng}(t)$ of Q under t is the set $\mathrm{rng}(t) = Qt$. We also use the notation $t = [i_1, i_2, \dots, i_n]$ for the transformation t above.

A *permutation* of Q is a mapping of Q *onto* itself. In other words, a permutation π of Q is a transformation where $\mathrm{rng}(\pi) = Q$. The *identity* transformation maps each element to itself, that is, $it = i$ for $i = 1, \dots, n$. A transformation t is a *cycle* of length k if there exist pairwise different elements $i_1, \dots, i_k$ such that $i_1t = i_2, i_2t = i_3, \dots, i_{k-1}t = i_k$, and $i_kt = i_1$. A cycle is denoted by $(i_1, i_2, \dots, i_k)$. For $i < j$, a *transposition* is the cycle (i, j). A *singular* transformation, denoted by $\binom{i}{j}$, has $it = j$ and $ht = h$ for all $h \neq i$. A *constant* transformation, denoted by $\binom{Q}{j}$, has $it = j$ for all i.

The set of all permutations of a set Q of n elements is a group, denoted by $\mathfrak{S}_Q$ and called the *symmetric group* of degree n. Piccard [19] showed in 1935 that two generators are sufficient to generate the symmetric group of degree n.

Theorem 1 (*Permutations*). *The symmetric group $\mathfrak{S}_Q$ of size $n!$ can be generated by any cyclic permutation of n elements together with any transposition. In particular, $\mathfrak{S}_Q$ can be generated by $c = (1, 2, \dots, n)$ and $t = (1, 2)$.*

The set of all transformations of a set Q, denoted by $\mathcal{T}_Q$, is a finite semigroup, in fact, a monoid. We refer the reader to the book of Ganyushkin and Mazorchuk [8] for a detailed discussion of finite transformation semigroups. In 1935, Piccard [19] proved that three transformations of Q are sufficient to generate the monoid $\mathcal{T}_Q$. In the same year, Eilenberg showed that fewer than three generators are not possible, as reported by Sierpiński [24]. Dénes [7] (apparently unaware of the earlier work) studied more general generators in 1968; we use his formulation.

Theorem 2 (*Transformations*). *The complete transformation monoid $\mathcal{T}_Q$ of size n^n can be generated by any cyclic permutation of n elements together with any transposition and any singular transformation. In particular, $\mathcal{T}_Q$ can be generated by $c = (1, 2, \dots, n)$, $t = (1, 2)$ and $r = \binom{n}{1}$.*

3. Quotient complexity and syntactic complexity

If Σ is a non-empty finite alphabet, then Σ^* is the free monoid generated by Σ, and Σ^+ is the free semigroup generated by Σ. A *word* is any element of Σ^*, and the empty word is ε. The length of a word $w \in \Sigma^*$ is $|w|$. A *language* over Σ is any subset of Σ^*. If $w = uxv$ for some $u, x, v \in \Sigma^*$, then u is a *prefix* of w, v is a *suffix* of w, and x is a *factor* of w. Both u and v are also factors of w. A *proper* prefix (suffix, factor) of w is a prefix (suffix, factor) of w other than w.

The *left quotient*, or simply *quotient*, of a language L by a word w is the language $L_w = \{x \in \Sigma^* \mid wx \in L\}$. For any $L \subseteq \Sigma^*$, the *Nerode right congruence* [18] $\sim_L$ of L is defined as follows: For all $x, y \in \Sigma^*$,

$$x \sim_L y \quad \text{if and only if} \quad xv \in L \Leftrightarrow yv \in L, \text{ for all } v \in \Sigma^*.$$

Clearly, $L_x = L_y$ if and only if $x \sim_L y$. Thus each equivalence class of this right congruence corresponds to a distinct quotient of L.

The *Myhill congruence* [17] $\approx_L$ of L is defined as follows: For all $x, y \in \Sigma^*$,

$$x \approx_L y \quad \text{if and only if} \quad uxv \in L \Leftrightarrow uyv \in L \text{ for all } u, v \in \Sigma^*.$$

This congruence is also known as the *syntactic congruence* of L. The quotient set $\Sigma^+/\approx_L$ of equivalence classes of the relation $\approx_L$ is a semigroup called the *syntactic semigroup* of L, and $\Sigma^*/\approx_L$ is the *syntactic monoid* of L. The *syntactic complexity* $\sigma(L)$ of L is the cardinality of its syntactic semigroup. The *monoid complexity* $\mu(L)$ of L is the cardinality of its syntactic monoid. If the equivalence class containing ε is a singleton in the syntactic monoid, then $\sigma(L) = \mu(L) - 1$; otherwise, $\sigma(L) = \mu(L)$.

A *deterministic finite automaton* (DFA) is a quintuple $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$, where Q is a finite, non-empty set of *states*, Σ is a finite non-empty *alphabet*, $\delta : Q \times \Sigma \to Q$ is the *transition function*, $q_1 \in Q$ is the *initial state*, and $F \subseteq Q$ is the set of *final states*. We extend δ to $Q \times \Sigma^*$ in the usual way. The DFA $\mathcal{A}$ accepts a word $w \in \Sigma^*$ if $\delta(q_1, w) \in F$. The set of all words *accepted* by $\mathcal{A}$ is $L(\mathcal{A})$. By the *language of a state* q of $\mathcal{A}$ we mean the language accepted by the DFA $(Q, \Sigma, \delta, q, F)$. A state is *empty* if its language is empty.

Let L be a regular language. The *quotient DFA* of L is $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$, where $Q = \{L_w \mid w \in \Sigma^*\}$, $\delta(L_w, a) = L_{wa}$, $q_1 = L_\varepsilon = L$, $F = \{L_w \mid \varepsilon \in L_w\}$. A quotient L_w is *final* if $\varepsilon \in L_w$. The number $\kappa(L)$ of distinct quotients of L is the *quotient complexity* of L. The quotient DFA of L is the minimal DFA accepting L, and so quotient complexity is the same as state complexity. The quotient viewpoint is often useful for deriving upper bounds, while the state approach may be more convenient for proving lower bounds.

In terms of automata, each equivalence class $[w]_{\sim_L}$ of $\sim_L$ is the set of all words w that take the automaton to the same state from the initial state, and each equivalence class $[w]_{\approx_L}$ of $\approx_L$ is the set of all words that perform the same transformation on the set of states [16]. In terms of quotients, $[w]_{\sim_L}$ is the set of words w that can be followed by the same quotient L_w.

Let $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$ be a DFA. For each word $w \in \Sigma^+$, the transition function for w defines a transformation t_w of Q by the word w: for all $i \in Q$, $it_w \overset{\text{def}}{=} \delta(i, w)$. The set $T_{\mathcal{A}}$ of all such transformations by non-empty words forms a subsemigroup of $\mathcal{T}_Q$, called the *transition semigroup* of $\mathcal{A}$ [20]. Conversely, we can use a set $\{t_a \mid a \in \Sigma\}$ of transformations to define δ, and so the DFA $\mathcal{A}$. When the context is clear we simply write $a = t$, where t is a transformation of Q, to mean that the transformation performed by $a \in \Sigma$ is t.

If $\mathcal{A}$ is the quotient DFA of L, then $T_{\mathcal{A}}$ is isomorphic to the syntactic semigroup T_L of L [16], and we represent elements of T_L by transformations in $T_{\mathcal{A}}$.

We attempt to obtain tight upper bounds on the syntactic complexity $\sigma(L) = |T_L|$ of L as a function of the quotient complexity $\kappa(L)$ of L. First we consider the syntactic complexity of regular languages over a unary alphabet, where the concepts prefix-, suffix-, bifix-, and factor-free, coincide. So we may consider only unary prefix-free regular languages L with quotient complexity $\kappa(L) = n$. When $n = 1$, the only prefix-free language is $L = \emptyset$ with $\sigma(L) = 1$. For $n \geqslant 2$, a prefix-free language L must be a singleton, $L = \{a^{n-2}\}$. The syntactic semigroup T_L of L consists of $n-1$ transformations t_w by words $w = a^i$, where $1 \leqslant i \leqslant n-1$. Thus we have

Proposition 3 (*Unary Free Regular Languages*)**.** *If L is a unary free regular language with $\kappa(L) = n \geqslant 2$, then $\sigma(L) = n - 1$.*

The tight upper bound for regular unary languages [12] is n.

We assume that $|\Sigma| \geqslant 2$ in the following sections. Since the syntactic semigroup of a language is the same as that of its complement, we deal only with prefix-, suffix-, bifix-, and factor-free languages. All the syntactic complexity results, however, apply also to the complements of these languages.

4. Prefix-free regular languages

To simplify notation we write ε for the language $\{\varepsilon\}$. Recall that a regular language L is prefix-free if and only it has exactly one final quotient, and that quotient is ε [11].

Theorem 4 (*Prefix-Free Regular Languages*)**.** *If L is regular and prefix-free with $\kappa(L) = n \geqslant 2$, then $\sigma(L) \leqslant n^{n-2}$. Moreover, this bound is tight for $n = 2$ if $|\Sigma| \geqslant 1$, for $n = 3$ if $|\Sigma| \geqslant 2$, for $n = 4$ if $|\Sigma| \geqslant 4$, and for $n \geqslant 5$ if $|\Sigma| \geqslant n + 1$.*

Proof. If L is prefix-free, the only final quotient of L is ε. Thus L also has the empty quotient, since $\varepsilon_a = \emptyset$ for $a \in \Sigma$. Let $\mathcal{A} = (Q, \Sigma, \delta, 1, \{n-1\})$ be the quotient DFA of L, where, without loss of generality, $n - 1 \in Q$ is the only final state, and $n \in Q$ is the empty state. For any transformation $t \in T_L$, $(n-1)t = nt = n$. Thus we have $\sigma(L) \leqslant n^{n-2}$.

The only prefix-free regular language for $n = 1$ is $L = \emptyset$ with $\sigma(L) = 1$; here the bound n^{n-2} does not apply. For $n = 2$ and $\Sigma = \{a\}$, the language $L = \varepsilon$ meets the bound. For $n = 3$ and $\Sigma = \{a, b\}$, $L = b^*a$ meets the bound. For $n \geqslant 4$, let

$$\mathcal{A}_n = (\{1, 2, \ldots, n\}, \{a, b, c, d_1, d_2, \ldots, d_{n-2}\}, \delta, 1, \{n-1\}),$$

where $a = \binom{n-1}{n}(1, 2, \ldots, n-2)$, $b = \binom{n-1}{n}(1, 2)$, $c = \binom{n-1}{n}\binom{n-2}{1}$, and $d_i = \binom{n-1}{n}\binom{i}{n-1}$ for $i = 1, 2, \ldots, n-2$. DFA $\mathcal{A}_6$ is shown in Fig. 1, where $\Gamma = \{d_1, d_2, \ldots, d_{n-2}\}$. For $n = 4$, input a coincides with b; hence only 4 inputs are needed.

Any transformation $t \in T_L$ has the form

$$t = \begin{pmatrix} 1 & 2 & \cdots & n-2 & n-1 & n \\ i_1 & i_2 & \cdots & i_{n-2} & n & n \end{pmatrix},$$

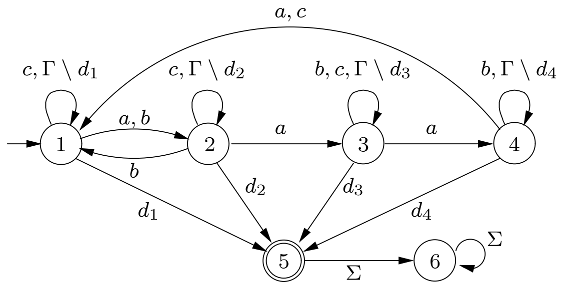

Fig. 1. Quotient DFA $\mathcal{A}_6$ of prefix-free regular language with 1,296 transformations.

where $i_k \in \{1, 2, \ldots, n\}$ for $1 \leqslant k \leqslant n-2$. There are three cases:

1. If $i_k \leqslant n-2$ for all k, $1 \leqslant k \leqslant n-2$, then by Theorem 2, $\mathcal{A}_n$ can do t.

2. If $i_k \leqslant n-1$ for all k, $1 \leqslant k \leqslant n-2$, and there exists some h such that $i_h = n-1$, then there exists some j, $1 \leqslant j \leqslant n-2$ such that $i_k \neq j$ for all k, $1 \leqslant k \leqslant n-2$. For all $1 \leqslant k \leqslant n-2$, define i'_k as follows: $i'_k = j$ if $i_k = n-1$, and $i'_k = i_k$ if $i_k \neq n-1$. Let
$$s = \begin{pmatrix} 1 & 2 & \cdots & n-2 & n-1 & n \\ i'_1 & i'_2 & \cdots & i'_{n-2} & n & n \end{pmatrix}.$$
By Case 1 above, $\mathcal{A}_n$ can do s. Since $t = sd_j$, $\mathcal{A}_n$ can do t as well.

3. Otherwise, there exists some h such that $i_h = n$. Then there exists some j, $1 \leqslant j \leqslant n-2$, such that $i_k \neq j$ for all k, $1 \leqslant k \leqslant n-2$. For all $1 \leqslant k \leqslant n-2$, define i'_k as follows: $i'_k = n-1$ if $i_k = n$, $i'_k = j$ if $i_k = n-1$, and $i'_k = i_k$ otherwise. Let s be as above but with new i'_k. By Case 2 above, $\mathcal{A}_n$ can do s. Since $t = sd_j$, $\mathcal{A}_n$ can do t as well.

Therefore, the syntactic complexity of $\mathcal{A}_n$ meets the desired bound. □

We conjecture that the alphabet sizes cannot be reduced. As shown in Table 2, we have verified this conjecture for $n \leqslant 5$ by enumerating all prefix-free regular languages with $n \leqslant 5$ using *GAP* [9].

5. Suffix-free regular languages

For any regular language L, a quotient L_w is *uniquely reachable* [3] if $L_w = L_x$ implies that $w = x$. It is known from [10] that, if L is a suffix-free regular language, then $L = L_\varepsilon$ is uniquely reachable by ε, and L has the empty quotient. Without loss of generality, we assume that 1 is the initial state, and n is the empty state in the quotient DFA of L. We will show that the cardinality of $\mathbf{B}_{\rm sf}(n)$, defined below, is an upper bound (**B** for "bound") on the syntactic complexity of suffix-free regular languages with quotient complexity n. For $n \geqslant 2$, let

$$\mathbf{B}_{\rm sf}(n) = \{t \in \mathcal{T}_Q \mid 1 \notin \mathrm{rng}(t),\ nt = n,\ \text{and for all } j \geqslant 1,\ 1t^j = n \text{ or } 1t^j \neq it^j\ \ \forall i,\ 1 < i < n\}.$$

Proposition 5. *If L is a regular language with quotient DFA $\mathcal{A}_n = (Q, \Sigma, \delta, 1, F)$ and syntactic semigroup T_L, then the following hold:*

1. *If L is suffix-free, then T_L is a subset of $\mathbf{B}_{\rm sf}(n)$.*
2. *If L has the empty quotient, only one final quotient, and $T_L \subseteq \mathbf{B}_{\rm sf}(n)$, then L is suffix-free.*

Proof. 1. Let L be suffix-free, and let $\mathcal{A}_n$ be its quotient DFA. Consider an arbitrary $t \in T_L$. Since the quotient L is uniquely reachable, $it \neq 1$ for all $i \in Q$. Since the quotient corresponding to state n is empty, $nt = n$. Since L is suffix-free, for any two quotients L_w and L_{uw}, where $u, w \in \Sigma^+$ and $L_w \neq \emptyset$, we must have $L_w \cap L_{uw} = \emptyset$, and so $L_w \neq L_{uw}$. This means that, for any $j \geqslant 1$, if $1t^j \neq n$, then $1t^j \neq it^j$ for all i, $1 < i < n$. So $t \in \mathbf{B}_{\rm sf}(n)$, and $T_L \subseteq \mathbf{B}_{\rm sf}(n)$.

2. Assume that $T_L \subseteq \mathbf{B}_{\rm sf}(n)$, and let f be the only final state. If L is not suffix-free, then there exist non-empty words u and v such that $v, uv \in L$. Let t_u and t_v be the transformations by u and v, and let $i = 1t_u$; then $i \neq 1$. Assume without loss the generality that n is the empty state. Then $f \neq n$, and we have $1t_v = f = 1t_{uv} = 1t_ut_v = it_v$, which contradicts the fact that $t_v \in \mathbf{B}_{\rm sf}(n)$. Therefore L is suffix-free. □

Let $\mathrm{b}_{\rm sf}(n) = |\mathbf{B}_{\rm sf}(n)|$. We now prove that $\mathrm{b}_{\rm sf}(n)$ is an upper bound on the syntactic complexity of suffix-free regular languages.

With each transformation t of Q, we associate a directed graph G_t, where Q is the set of nodes, and $(i, j) \in Q \times Q$ is a directed edge from i to j if $it = j$. We call such a graph G_t the *transition graph* of t. For each node i, there is exactly one edge leaving i in G_t. Consider the infinite sequence $i, it, it^2, \ldots$ for any $i \in Q$. Since Q is finite, there exists least $j \geqslant 0$ such that $it^{j+1} = it^{j'}$ for some $j' \leqslant j$. Then the finite sequence $\mathfrak{s}_t(i) = i, it, \ldots, it^j$ contains all the distinct elements of the above infinite sequence, and it induces a directed path $P_t(i)$ from i to it^j in G_t. In particular, if $n \in \mathfrak{s}_t(1)$, and $nt = n$, then we call $\mathfrak{s}_t(1)$ the *principal sequence* of t, and $P_t(1)$, the *principal path* of G_t.

Table 1
The number $S_m(h)$ of labeled rooted trees with m nodes and height at most h.

h/m	1	2	3	4	5	6	7
0	1	0	0	0	0	0	0
1	1	2	3	4	5	6	7
2	1	2	9	40	205	1176	7399
3	1	2	9	64	505	4536	46249
4	1	2	9	64	625	7056	89929
5	1	2	9	64	625	7776	112609
6	1	2	9	64	625	7776	117649

Proposition 6. *There exists a principal sequence for every transformation $t \in \mathbf{B}_{\mathrm{sf}}(n)$.*

Proof. Suppose $t \in \mathbf{B}_{\mathrm{sf}}(n)$ and $\mathfrak{s}_t(1) = 1, 1t, \ldots, 1t^j$. If t does not have a principal sequence, then $n \notin \mathfrak{s}_t(1)$, and $1t^{j+1} = 1t^{j'} \neq n$ for some $j' \leqslant j$. Let $i = 1t^{j+1-j'}$; then $i \neq 1$ and $1t^{j'} = it^{j'}$, violating the last property of $\mathbf{B}_{\mathrm{sf}}(n)$. Therefore there is a principal sequence for every $t \in \mathbf{B}_{\mathrm{sf}}(n)$. $\square$

Fix a transformation $t \in \mathbf{B}_{\mathrm{sf}}(n)$. Let $i \in Q$ be such that $i \notin \mathfrak{s}_t(1)$. If the sequence $\mathfrak{s}_t(i)$ does not contain any element of the principal sequence $\mathfrak{s}_t(1)$ other than n, then we say that $\mathfrak{s}_t(i)$ has *no principal connection*. Otherwise, there exists least $j \geqslant 1$ such that $1t^j \neq n$ and $1t^j = it^{j'} \in \mathfrak{s}_t(i)$ for some $j' \geqslant 1$, and we say that $\mathfrak{s}_t(i)$ has a *principal connection* at $1t^j$. If $j' < j$, the principal connection is *short*; otherwise, it is *long*.

Lemma 7. *For all $t \in \mathbf{B}_{\mathrm{sf}}(n)$ and $i \notin \mathfrak{s}_t(1)$, the sequence $\mathfrak{s}_t(i)$ has no long principal connection.*

Proof. Let t be any transformation in $\mathbf{B}_{\mathrm{sf}}(n)$. Suppose for some $i \notin \mathfrak{s}_t(1)$, the sequence $\mathfrak{s}_t(i)$ has a long principal connection at $1t^j = it^{j'} \neq n$, where $j \leqslant j'$. Hence $it^{j'-j} \neq n$, and $1t^j = (it^{j'-j})t^j$, which is a contradiction. Therefore, for all $i \notin \mathfrak{s}_t(1)$, $\mathfrak{s}_t(i)$ has no long principal connection. $\square$

To calculate the cardinality of $\mathbf{B}_{\mathrm{sf}}(n)$, we need the following observation.

Lemma 8. *For all $t \in \mathbf{B}_{\mathrm{sf}}(n)$ and $i \notin \mathfrak{s}_t(1)$, if $\mathfrak{s}_t(i)$ has a principal connection, then there is no cycle incident to the path $P_t(i)$ in the transition graph G_t.*

Proof. This observation can be derived from Theorem 1.2.9 of [8]. However, our proof is shorter. Pick any $i \notin \mathfrak{s}_t(1)$ such that $\mathfrak{s}_t(i)$ has a principal connection at $1t^j = it^{j'}$ for some i, j and j'. Then the sequence $\mathfrak{s}_t(i)$ contains n, and the path $P_t(i)$ does not contain any cycle. Suppose C is a cycle which includes node $x = it^k \in P_t(i)$. Since there is only one outgoing edge for each node in G_t, the cycle C must be oriented and must contain a node $x' \notin P_t(i)$ such that (x', x) is an edge in C. Then the next node in the cycle must be it^{k+1} since there is only one outgoing edge from x. But then x' can never be reached from $P_t(i)$, and so no such cycle can exist. $\square$

By Lemma 8, for any $1t^j \in \mathfrak{s}_t(1)$, where $j \geqslant 1$, the union of directed paths from various nodes i to $1t^j$, if $i \notin \mathfrak{s}_t(1)$ and $\mathfrak{s}_t(i)$ has a principal connection at $1t^j$, forms a labeled tree $T_t(j)$ rooted at $1t^j$. Suppose there are $r_j + 1$ nodes in $T_t(j)$ for each j, and suppose there are r elements of Q that are not in the principal sequence $\mathfrak{s}_t(1)$ nor in any tree $T_t(j)$, for some $r_j, r \geqslant 0$. Note that, $1t^j$ is the only node in $T_t(j)$ that is also in the principal sequence $\mathfrak{s}_t(1)$. Each tree $T_t(j)$ has height at most $j - 1$; otherwise, some $i \in T_t(j)$ has a long principal connection. In particular, tree $T_t(1)$ has height 1; so it is trivial with only one node $1t$. Then $r_1 = 0$, and we need to consider trees $T_t(j)$ only for $j \geqslant 2$. Let $S_m(h)$ be the number of labeled rooted trees with m nodes and height at most h. This number can be found in the paper of Riordan [21]; the calculation is somewhat complex, and we refer the reader to [21] for details. For convenience, we include the values of $S_m(h)$ for small values of m and h in Table 1, where the row number is h and the column number is m.

Since each of the m nodes can be the root, there are $S'_m(h) = \frac{S_m(h)}{m}$ labeled trees rooted at a fixed node and having m nodes and height at most h. The following is an example of trees $T_t(j)$ in transformations $t \in \mathbf{B}_{\mathrm{sf}}(n)$.

Example 9. *Let $n = 15$. Consider any transformation $t \in \mathbf{B}_{\mathrm{sf}}(15)$ with principal sequence $\mathfrak{s}_t(1) = 1, 2, 3, 4, 5, 15$. There are 9 elements of Q that are not in $\mathfrak{s}_t(1)$, and some of them are in the trees $T_t(j)$ for $2 \leqslant j \leqslant 4$. Consider the cases where $r_2 = 2, r_3 = 3$, $r_4 = 1$, and $r = 3$. Fig. 2 shows one such transformation t.*

For $j = 2$, the tree $T_t(2)$ has height at most 1, and there are $S'_{r_2+1}(1) = \frac{S_{r_2+1}(1)}{r_2+1} = \frac{3}{3} = 1$ possible $T_t(2)$. For $j = 3$, there are $S'_{r_3+1}(2) = \frac{S_{r_3+1}(2)}{r_3+1} = 10$ possible $T_t(3)$, which are of one of the three types shown in Fig. 3. Among the 10 possible $T_t(3)$, one is of type (a), three are of type (b), and six are of type (c). For $j = 4$, there are $S'_{r_4+1}(3) = \frac{S_{r_4+1}(3)}{r_4+1} = 1$ possible $T_t(4)$.

Let C^n_k be the binomial coefficient, and let $C^n_{k_1,\ldots,k_m}$ be the multinomial coefficient. Then we have the following lemma.

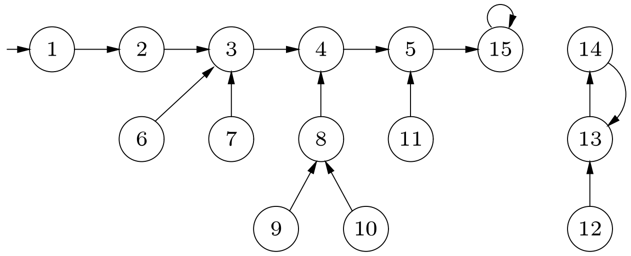

Fig. 2. Transition graph of some $t \in \mathbf{B}_{\mathrm{sf}}(15)$ with principal sequence 1, 2, 3, 4, 5, 15.

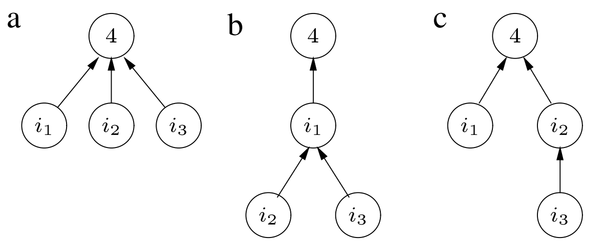

Fig. 3. Three types of trees of the form $T_t(3)$, where $\{i_1, i_2, i_3\} = \{8, 9, 10\}$.

Lemma 10. *For $n \geqslant 2$, we have*

$$\mathrm{b}_{\mathrm{sf}}(n) = \sum_{k=0}^{n-2} C_k^{n-2} k! \sum_{\substack{r_2+\cdots+r_k+r\\=n-k-2}} C_{r_2,\ldots,r_k,r}^{n-k-2} (r+1)^r \prod_{j=2}^{k} S'_{r_j+1}(j-1). \tag{1}$$

Proof. Let t be any transformation in $\mathbf{B}_{\mathrm{sf}}(n)$. Suppose $\mathfrak{s}_t(1) = 1, 1t, \ldots, 1t^k, n$ for some k, $0 \leqslant k \leqslant n-2$. There are $C_k^{n-2}k!$ different principal sequences $\mathfrak{s}_t(1)$. Now, fix $\mathfrak{s}_t(1)$. Suppose $n-k-2 = r_2 + \cdots + r_k + r$, where, for $2 \leqslant j \leqslant k$, tree $T_t(j)$ contains $r_j + 1$ nodes, for some $r_j \geqslant 0$. There are $C_{r_2,\ldots,r_k,r}^{n-k-2}$ different tuples $(r_2, \ldots, r_k, r)$. Each tree $T_t(j)$ has height at most $j-1$, and it is rooted at $1t^j$. There are $S'_{r_j+1}(j-1) = \frac{S_{r_j+1}(j-1)}{r_j+1}$ different trees $T_t(j)$. Let E be the set of the remaining r elements x of Q that are not in any tree $T_t(j)$ nor in the principal sequence $\mathfrak{s}_t(1)$. The image xt can only be chosen from $E \cup \{n\}$. There are $(r+1)^r$ different mappings of E. Altogether we have the desired formula. □

From Proposition 5 and Lemma 10 we have

Proposition 11. *For $n \geqslant 2$, if L is a suffix-free language with quotient complexity n, then its syntactic complexity $\sigma(L)$ satisfies that $\sigma(L) \leqslant \mathrm{b}_{\mathrm{sf}}(n)$, where $\mathrm{b}_{\mathrm{sf}}(n)$ is the cardinality of $\mathbf{B}_{\mathrm{sf}}(n)$, and it is given by Equation Eq. (1).*

Note that $\mathbf{B}_{\mathrm{sf}}(n)$ is not a semigroup for $n \geqslant 4$ because $s_1 = [2, 3, n, \ldots, n, n]$, $s_2 = [n, 3, 3, \ldots, 3, n] \in \mathbf{B}_{\mathrm{sf}}(n)$, but $s_1 s_2 = [3, 3, n, \ldots, n, n] \notin \mathbf{B}_{\mathrm{sf}}(n)$. Hence, although $\mathrm{b}_{\mathrm{sf}}(n)$ is an upper bound on the syntactic complexity of suffix-free regular languages, that bound is not tight. Our objective is to find the largest subset of $\mathbf{B}_{\mathrm{sf}}(n)$ that is a semigroup. For $n \geqslant 2$, let

$$\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n) = \{t \in \mathbf{B}_{\mathrm{sf}}(n) \mid \text{for all } i, j \in Q \text{ where } i \neq j, \text{ we have } it = jt = n \text{ or } it \neq jt\},$$

where **W** stands for "witness", and the superscript $\leqslant 5$ will be explained in Theorem 16.

Proposition 12. *For $n \geqslant 2$, $\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$ is a semigroup contained in $\mathbf{B}_{\mathrm{sf}}(n)$, and its cardinality is*

$$\mathrm{w}_{\mathrm{sf}}^{\leqslant 5}(n) = |\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)| = \sum_{k=1}^{n-1} C_k^{n-1}(n-1-k)!\,C_{n-1-k}^{n-2}.$$

Proof. We know that any t is in $\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$ if and only if the following hold:

1. $it \neq 1$ for all $i \in Q$, and $nt = n$;
2. for all $i, j \in Q$, such that $i \neq j$, either $it = jt = n$ or $it \neq jt$.

Clearly $\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n) \subseteq \mathbf{B}_{\mathrm{sf}}(n)$. For any transformation $t_1, t_2 \in \mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$, consider the composition $t_1 t_2$. Since $1 \notin \mathrm{rng}(t_2)$, we have $1 \notin \mathrm{rng}(t_1 t_2)$. We also have $nt_1t_2 = nt_2 = n$. Pick any $i, j \in Q$ such that $i \neq j$. Suppose $it_1t_2 \neq n$ or $jt_1t_2 \neq n$. If $it_1t_2 = jt_1t_2$, then $it_1 = jt_1$ and thus $i = j$, a contradiction. Hence $t_1t_2 \in \mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$, and $\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$ is a semigroup contained in $\mathbf{B}_{\mathrm{sf}}(n)$.

Let $t \in \mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$ be any transformation. Note that $nt = n$ is fixed. Let $Q' = Q \setminus \{n\}$, and $Q'' = Q \setminus \{1, n\}$. Suppose k elements in Q' are mapped to n by t, where $0 \leqslant k \leqslant n-1$; then there are C_k^{n-1} choices of these elements. For the set D of the remaining $n-1-k$ elements, which must be mapped by t to pairwise distinct elements of Q'', there are $C_{n-1-k}^{n-2}(n-1-k)!$ choices for the mapping $t|_D$. When $k = 0$, there is no such t since $|Dt| = n-1 > n-2 = |Q''|$. Altogether, the cardinality of $\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$ is $|\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)| = \sum_{k=1}^{n-1} C_k^{n-1}(n-1-k)!C_{n-1-k}^{n-2}$. $\square$

Remark 13. A *partial injective transformation* of a set Q is a partial injective mapping of Q into itself. The set of all such transformations of Q is a semigroup, usually called the *symmetric inverse semigroup* [8] and denoted by $\mathcal{IS}_Q$. Let $Q' = Q \setminus \{n\}$. The number $\mathrm{w}_{\mathrm{sf}}^{\leqslant 5}(n)$ coincides with the number of *nilpotents* in $\mathcal{IS}_{Q'}$, which are the transformations $t \in \mathcal{IS}_{Q'}$ such that $\mathrm{dom}(t^k) = \emptyset$ for some $k \geqslant 1$. Riordan [22] reported that $\mathrm{w}_{\mathrm{sf}}^{\leqslant 5}(n)$ has the asymptotic approximation

$$\mathrm{w}_{\mathrm{sf}}^{\leqslant 5}(n) \sim \frac{1}{\sqrt{2e}}(n-1)^{n-\frac{4}{5}}e^{-(n-1)+2\sqrt{n-1}}.$$

We now construct a generating set $\mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(n)$ (**G** for "generators") of size n for $\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$, which will show that there exist DFA's accepting suffix-free regular languages with quotient complexity n and syntactic complexity $\mathrm{w}_{\mathrm{sf}}^{\leqslant 5}(n)$.

Proposition 14. *When $n \geqslant 2$, the semigroup $\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$ is generated by the following set $\mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(n)$ of transformations of Q:*

$$\mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(2) = \{a_1\}, \quad \textit{where } a_1 = [2, 2];$$
$$\mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(3) = \{a_1, a_2\}, \quad \textit{where } a_1 = [3, 2, 3] \textit{ and } a_2 = [2, 3, 3];$$

and for $n \geqslant 4$, $\mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(n) = \{a_0, \ldots, a_{n-1}\}$, where

- $a_0 = \binom{1}{n}(2, 3)$,
- $a_1 = \binom{1}{n}(2, 3, \ldots, n-1)$,
- *for $2 \leqslant i \leqslant n-1$, $ja_i = j+1$ for $j = 1, \ldots, i-1$, $ia_i = n$, and $ja_i = j$ for $j = i+1, \ldots, n$.*

For $n = 4$, a_0 and a_1 coincide, and three transformations suffice.

Proof. We have $\mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(n) \subseteq \mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$, and so $\langle \mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(n) \rangle$, the semigroup generated by $\mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(n)$, is a subset of $\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$. We now show that $\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n) \subseteq \langle \mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(n) \rangle$.

It is easy to verify the cases for $n = 2, 3$. Assume $n \geqslant 4$. Pick any t in $\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$. Note that $nt = n$ is fixed. Let $Q' = Q \setminus \{n\}$, $E_t = \{j \in Q' \mid jt = n\}$, $D_t = Q' \setminus E_t$, and $Q'' = Q \setminus \{1, n\}$. Then $D_t t \subseteq Q''$, and $|E_t| \geqslant 1$, since $|Q''| < |Q'|$. We prove by induction on $|E_t|$ that $t \in \langle \mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(n) \rangle$.

First, note that $\langle a_0, a_1 \rangle$, the semigroup generated by $\{a_0, a_1\}$, is isomorphic to the symmetric group $\mathfrak{S}_{Q''}$ by Theorem 1. Consider $E_t = \{i\}$ for some $i \in Q'$. Then $ia_i = it = n$. Moreover, since $D_t a_i, D_t t \subseteq Q''$, there exists $\pi \in \langle a_0, a_1 \rangle$ such that $(ja_i)\pi = jt$ for all $j \in D_t$. Then $t = a_i \pi \in \langle \mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(n) \rangle$.

Assume that any transformation $t \in \mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$ with $|E_t| < k$ can be generated by $\mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(n)$, where $1 < k < n-1$. Consider $t \in \mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$ with $|E_t| = k$. Suppose $E_t = \{e_1, \ldots, e_{k-1}, e_k\}$. Let $s \in \mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$ be such that $E_s = \{e_1, \ldots, e_{k-1}\}$. By assumption, s can be generated by $\mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(n)$. Let $i = e_k s$; then $i \in Q''$, and $e_j(sa_i) = n$ for all $1 \leqslant j \leqslant k$. Moreover, we have $D_t(sa_i) \subseteq Q''$. Thus, there exists $\pi \in \langle a_0, a_1 \rangle$ such that, for all $d \in D_t$, $d(sa_i\pi) = dt$. Altogether, for all $e_j \in E_t$, we have $e_j(sa_i\pi) = e_j t = n$, for all $d \in D_t$, $d(sa_i\pi) = dt$, and $n(sa_i\pi) = nt = n$. Thus $t = sa_i\pi$, and $t \in \langle \mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(n) \rangle$.

Therefore $\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n) = \langle \mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(n) \rangle$. $\square$

Theorem 15. *For $n \geqslant 2$, let $\mathcal{A}_n = (Q, \Sigma, \delta, 1, \{n-1\})$ be a DFA with alphabet Σ, where each $a \in \Sigma$ defines a distinct transformation in $\mathbf{G}_{\mathrm{sf}}^{\leqslant 5}(n)$ as in Proposition 14. Then $L = L(\mathcal{A}_n)$ has quotient complexity $\kappa(L) = n$, and syntactic complexity $\sigma(L) = \mathrm{w}_{\mathrm{sf}}^{\leqslant 5}(n)$. Moreover, L is suffix-free.*

Proof. The cases for $n = 2, 3$ are easy to verify. Assume $n \geqslant 4$. First we show that all the states of $\mathcal{A}_n$ are reachable: 1 is the initial state, state n is reached by a_1, and for $2 \le i \le n-1$, state i is reached by a_i^{i-1}. For $1 \le i \le n-1$, the word a_{n-1}^{n-1-i} is accepted only by state i. Also n is the empty state. Thus all the states of $\mathcal{A}_n$ are distinct, and $\kappa(L) = n$.

By Proposition 14, the syntactic semigroup of L is $\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$; hence $\sigma(L) = |\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)| = \mathrm{w}_{\mathrm{sf}}^{\leqslant 5}(n)$. By Proposition 5, L is suffix-free. $\square$

As shown in Table 2, the size of Σ cannot be decreased for $n \leqslant 5$.

Theorem 16. *For $2 \leqslant n \leqslant 5$, if a suffix-free regular language L has quotient complexity $\kappa(L) = n$, then its syntactic complexity satisfies that $\sigma(L) \leqslant \mathrm{w}_{\mathrm{sf}}^{\leqslant 5}(n)$, and this is a tight upper bound.*

Proof. By Proposition 5, the syntactic semigroup of a suffix-free regular language L is contained in $\mathbf{B}_{\rm sf}(n)$. For $n \in \{2,3\}$, $\mathrm{w}_{\rm sf}^{\leqslant 5}(n) = \mathrm{b}_{\rm sf}(n)$. So $\mathrm{w}_{\rm sf}^{\leqslant 5}(n)$ is an upper bound, and it is met by the language $L=\varepsilon$ for $n=2$ and by $L=ab^*$ for $n=3$. For $n=4$, we have $|\mathbf{B}_{\rm sf}(4)|=15$ and $|\mathbf{W}_{\rm sf}^{\leqslant 5}(4)|=13$. Two transformations, $s_1=[4,2,2,4]$ and $s_2=[4,3,3,4]$, in $\mathbf{B}_{\rm sf}(4)$ are such that s_1 conflicts with $t_1=[3,2,4,4]\in \mathbf{W}_{\rm sf}^{\leqslant 5}(4)$ (because $t_1s_1=[2,2,4,4]\notin \mathbf{B}_{\rm sf}(4)$), and s_2 conflicts with $t_2=[2,3,4,4]$ (because $t_2s_2=[3,3,4,4]\notin \mathbf{B}_{\rm sf}(4)$). Thus $\sigma(L)\leqslant 13$. Let L be the language accepted by the DFA $\mathcal{A}_4$ in Theorem 15; then $\kappa(L)=4$ and $\sigma(L)=13$. So the bound is tight.

For $n=5$, we have $|\mathbf{B}_{\rm sf}(5)|=115$ and $|\mathbf{W}_{\rm sf}^{\leqslant 5}(5)|=73$. Suppose $\mathbf{B}_{\rm sf}(5)\setminus \mathbf{W}_{\rm sf}^{\leqslant 5}(5)=\{s_1,\dots,s_{42}\}$. For each s_i, we enumerated transformations in $\mathbf{W}_{\rm sf}^{\leqslant 5}(5)$ using *GAP* and found a unique $t_i\in \mathbf{W}_{\rm sf}^{\leqslant 5}(5)$ such that the semigroup $\langle t_i,s_i\rangle$ is not contained in $\mathbf{B}_{\rm sf}(5)$. Thus at most one transformation in each pair $\{t_i,s_i\}$ can appear in the syntactic semigroup of L. So we reduce the upper bound to 73. By Theorem 15, this bound is tight. □

When $n\geqslant 6$, the semigroup $\mathbf{W}_{\rm sf}^{\leqslant 5}(n)$ is no longer the largest semigroup contained in $\mathbf{B}_{\rm sf}(n)$; hence the upper bound in Theorem 16 does not apply. In the following, we define and study another semigroup $\mathbf{W}_{\rm sf}^{\geqslant 6}(n)$, which is larger than $\mathbf{W}_{\rm sf}^{\leqslant 5}(n)$ and is also contained in $\mathbf{B}_{\rm sf}(n)$. For $n\geqslant 2$, let

$$\mathbf{W}_{\rm sf}^{\geqslant 6}(n)=\{t\in \mathbf{B}_{\rm sf}(n)\mid 1t=n \text{ or } it=n\ \forall\, i,\ 2\leqslant i\leqslant n-1\}.$$

Proposition 17. *For $n\geqslant 2$, the set $\mathbf{W}_{\rm sf}^{\geqslant 6}(n)$ is a semigroup contained in $\mathbf{B}_{\rm sf}(n)$, and its cardinality is*

$$\mathrm{w}_{\rm sf}^{\geqslant 6}(n)=|\mathbf{W}_{\rm sf}^{\geqslant 6}(n)|=(n-1)^{n-2}+(n-2).$$

Proof. Pick any t_1,t_2 in $\mathbf{W}_{\rm sf}^{\geqslant 6}(n)$. If $1t_1=n$, then $1(t_1t_2)=n$ and $t_1t_2\in \mathbf{W}_{\rm sf}^{\geqslant 6}(n)$. If $1t_1\neq n$, then, for all $i\in\{2,\dots,n-1\}$, $it_1=n$ and $i(t_1t_2)=n$; so $t_1t_2\in \mathbf{W}_{\rm sf}^{\geqslant 6}(n)$ as well. Hence $\mathbf{W}_{\rm sf}^{\geqslant 6}(n)$ is a semigroup contained in $\mathbf{B}_{\rm sf}(n)$.

For any $t\in \mathbf{W}_{\rm sf}^{\geqslant 6}(n)$, $nt=n$ is fixed. There are two possible cases:

1. $1t=n$: for each $i\in\{2,\dots,n-1\}$, it can be chosen from $\{2,\dots,n\}$. Then there are $(n-1)^{n-2}$ different t's in this case.
2. $1t\neq n$: now $1t$ can be chosen from $\{2,\dots,n-1\}$. For each $i\in\{2,\dots,n-1\}$, $it=n$ is fixed. Then, for any $t'\in \mathbf{W}_{\rm sf}^{\geqslant 6}(n)$ such that $1t'\neq n$, t differs from t' if and only if $1t\neq 1t'$. So there are $n-2$ different t's in this case.

Therefore $\mathrm{w}_{\rm sf}^{\geqslant 6}(n)=(n-1)^{n-2}+(n-2)$. □

When $n\geqslant 6$, one verifies that $\mathrm{w}_{\rm sf}^{\geqslant 6}(n)>\mathrm{w}_{\rm sf}^{\leqslant 5}(n)$. Hence $\mathbf{W}_{\rm sf}^{\geqslant 6}(n)$ is a larger semigroup than $\mathbf{W}_{\rm sf}^{\leqslant 5}(n)$. Table 2 contains values of $\mathrm{w}_{\rm sf}^{\leqslant 5}(n)$ and $\mathrm{w}_{\rm sf}^{\geqslant 6}(n)$ for small n's. For $n\in\{2,3\}$, we have $\mathbf{W}_{\rm sf}^{\geqslant 6}(n)=\mathbf{W}_{\rm sf}^{\leqslant 5}(n)$. From now on, we are only interested in larger values of n.

Proposition 18. *For $n\geqslant 4$, the semigroup $\mathbf{W}_{\rm sf}^{\geqslant 6}(n)$ is generated by the set $\mathbf{G}_{\rm sf}^{\geqslant 6}(n)=\{a_1,a_2,a_3,b_1,\dots,b_{n-2},c\}$ of transformations of Q, where*

- $a_1=\binom{1}{n}(2,\dots,n-1)$, $a_2=\binom{1}{n}(2,3)$, $a_3=\binom{1}{n}\binom{n-1}{2}$;
- *For* $1\leqslant i\leqslant n-2$, $b_i=\binom{1}{n}\binom{i+1}{n}$;
- $c=\binom{Q\setminus\{1\}}{n}\binom{1}{2}=[2,n,\dots,n]$.

For $n=4$, a_1 and a_2 coincide, and five transformations suffice.

Proof. Clearly $\mathbf{G}_{\rm sf}^{\geqslant 6}(n)\subseteq \mathbf{W}_{\rm sf}^{\geqslant 6}(n)$, and $\langle \mathbf{G}_{\rm sf}^{\geqslant 6}(n)\rangle\subseteq \mathbf{W}_{\rm sf}^{\geqslant 6}(n)$. We show in the following that $\mathbf{W}_{\rm sf}^{\geqslant 6}(n)\subseteq \langle \mathbf{G}_{\rm sf}^{\geqslant 6}(n)\rangle$.

Let $Q'=\{2,\dots,n-1\}$. By Theorem 2, a_1, a_2 and a_3 together generate the semigroup

$$\mathbf{Y}=\{t\in \mathbf{W}_{\rm sf}^{\geqslant 6}(n)\mid \text{ for all } i\in Q',\ it\in Q'\},$$

which is isomorphic to $\mathcal{T}_{Q'}$ and is contained in $\mathbf{W}_{\rm sf}^{\geqslant 6}(n)$. Next, consider any $t\in \mathbf{W}_{\rm sf}^{\geqslant 6}(n)\setminus \mathbf{Y}$. We have two cases:

1. $1t=n$: Let $E_t=\{i\in Q'\mid it=n\}$. Since $t\notin \mathbf{Y}$, $E_t\neq\emptyset$. Suppose $E_t=\{i_1,\dots,i_k\}$, for some $1\leqslant k\leqslant n-2$. Then there exists $t'\in \mathbf{Y}$ such that, for all $i\notin E_t$, $it'=it$. Let $s=b_{i_1-1}\cdots b_{i_k-1}$. Note that $E_ts=\{n\}$, and, for all $i\notin E_t$, $i(t's)=(it')s=it$. So $t=t's\in\langle \mathbf{G}_{\rm sf}^{\geqslant 6}(n)\rangle$.
2. $1t\neq n$: If $1t=2$, then $t=c$. Otherwise, $1t\in\{3,\dots,n-1\}\subseteq Q'$, and we know from the above case that there exists $t'\in \mathbf{G}_{\rm sf}^{\geqslant 6}(n)$ such that $2t'=1t$. Then $1(ct')=1t$, and $i(ct')=(ic)t'=n=it$, for all $i\in Q'$. Hence $t=ct'\in\langle \mathbf{G}_{\rm sf}^{\geqslant 6}(n)\rangle$.

Therefore $\langle a_1,a_2,a_3,b_1,\dots,b_{n-2},c\rangle=\mathbf{W}_{\rm sf}^{\geqslant 6}(n)$. □

Theorem 19. *For $n\geqslant 4$, let $\mathcal{A}'_n=(Q,\Sigma,\delta,1,\{2\})$ be a DFA with alphabet $\Sigma=\{a_1,a_3,b_1,b_2,c\}$ if $n=4$ or $\Sigma=\{a_1,a_2,a_3,b_1,\dots,b_{n-2},c\}$ if $n\geqslant 5$, where each letter defines a transformation as in Proposition 18. Then $L'=L(\mathcal{A}'_n)$ is suffix-free with quotient complexity $\kappa(L')=n$ and syntactic complexity $\sigma(L')=\mathrm{w}_{\rm sf}^{\geqslant 6}(n)$.*

Proof. First we show that $\kappa(L') = n$. From the initial state, we can reach state 2 by c and state n by a_1. From state 2 we can reach state i, $3 \leqslant i \leqslant n-1$, by a_1^{i-1}. So all the states in Q are reachable. The word c is accepted only by state 1. For $2 \leqslant i \leqslant n-1$, the word a_1^{n-i} is accepted only by state i. State n is the empty state, which rejects all words. Thus all the states in Q are distinct.

By Proposition 18, the syntactic semigroup of L' is $\mathbf{W}_{\text{sf}}^{\geqslant 6}(n)$, and $\sigma(L') = \mathrm{w}_{\text{sf}}^{\geqslant 6}(n)$. Also L' is suffix-free by Proposition 5. □

Theorem 20. *If L is a suffix-free regular language with $\kappa(L) = 6$, then $\sigma(L) \leqslant \mathrm{w}_{\text{sf}}^{\geqslant 6}(6) = 629$ and this is a tight bound.*

Proof. Note that $|\mathbf{B}_{\text{sf}}(6)| = 1169$ and $|\mathbf{W}_{\text{sf}}^{\geqslant 6}(6)| = 629$. Suppose $\{s_1, \dots, s_{540}\} = \mathbf{B}_{\text{sf}}(6) \setminus \mathbf{W}_{\text{sf}}^{\geqslant 6}(6)$. For each i, we enumerated transformations in $\mathbf{W}_{\text{sf}}^{\geqslant 6}(6)$ using *GAP* and found a unique $t_i \in \mathbf{W}_{\text{sf}}^{\geqslant 6}(6)$ such that $\langle t_i, s_i \rangle \not\subseteq \mathbf{B}_{\text{sf}}(6)$. As in the proof of Theorem 16, for each i, at most one transformation in $\{t_i, s_i\}$ can appear in the syntactic semigroup of L. Then we can reduce the upper bound to 629. This bound is met by the language L' in Theorem 19; so it is tight. □

We know that the upper bound on the syntactic complexity of suffix-free regular languages is achieved by the largest semigroup contained in $\mathbf{B}_{\text{sf}}(n)$. We conjecture that $\mathbf{W}_{\text{sf}}^{\geqslant 6}(n)$ is such a semigroup also for $n \geqslant 7$.

Conjecture 21 (*Suffix-Free Regular Languages*)**.** *If L is a suffix-free regular language with $\kappa(L) = n \geqslant 7$, then $\sigma(L) \leqslant \mathrm{w}_{\text{sf}}^{\geqslant 6}(n)$.*

6. Bifix-free regular languages

Let L be a regular bifix-free language with $\kappa(L) = n$. From Sections 4 and 5 we have:

1. L has ε as a quotient, and this is the only final quotient;
2. L has $\emptyset$ as a quotient;
3. L as a quotient is uniquely reachable.

Let $\mathcal{A}$ be the quotient DFA of L, with Q as the set of states. We assume that 1 is the initial state, $n-1$ corresponds to the quotient ε, and n is the empty state. For $n \geqslant 2$, consider the set

$$\mathbf{B}_{\text{bf}}(n) = \{t \in \mathbf{B}_{\text{sf}}(n) \mid (n-1)t = n\}.$$

The following is an observation similar to Proposition 5.

Proposition 22. *If L is a regular language with quotient complexity n and syntactic semigroup T_L, then the following hold:*

1. *If L is bifix-free, then T_L is a subset of $\mathbf{B}_{\text{bf}}(n)$.*
2. *If ε is the only final quotient of L, and $T_L \subseteq \mathbf{B}_{\text{bf}}(n)$, then L is bifix-free.*

Proof. 1. Since L is suffix-free, $T_L \subseteq \mathbf{B}_{\text{sf}}(n)$. Since L is also prefix-free, it has ε and $\emptyset$ as quotients. By assumption, $n-1 \in Q$ corresponds to the quotient ε. Thus for any $t \in T_L$, $(n-1)t = n$, and so $T_L \subseteq \mathbf{B}_{\text{bf}}(n)$.

2. Since ε is the only final quotient of L, L is prefix-free, and L has the empty quotient. Since $T_L \subseteq \mathbf{B}_{\text{bf}}(n) \subseteq \mathbf{B}_{\text{sf}}(n)$, L is suffix-free by Proposition 5. Therefore L is bifix-free. □

Lemma 23. *We have $|\mathbf{B}_{\text{bf}}(2)| = 1$, and for $n \geqslant 3$, $|\mathbf{B}_{\text{bf}}(n)| = M_n + N_n$, where*

$$M_n = \sum_{k=1}^{n-2} C_{k-1}^{n-3}(k-1)! \sum_{\substack{r_2+\cdots+r_k+r \\ =n-k-2}} C_{r_2,\dots,r_k,r}^{n-k-2}(r+1)^r \prod_{j=2}^{k} S'_{r_j+1}(j-1), \tag{2}$$

$$N_n = \sum_{k=0}^{n-3} C_k^{n-3} k! \sum_{\substack{r_2+\cdots+r_k+r \\ =n-k-3}} C_{r_2,\dots,r_k,r}^{n-k-3}(r+2)^r \prod_{j=2}^{k} S'_{r_j+1}(j-1). \tag{3}$$

Proof. It is easy to see that $\mathbf{B}_{\text{sf}}(2) = \{[2,2]\}$. Assume $n \geqslant 3$. Let t be any transformation in $\mathbf{B}_{\text{bf}}(n)$. Suppose $\mathfrak{s}_t(1) = 1, 1t, \dots, 1t^k, n$, where $0 \leqslant k \leqslant n-2$. For $2 \leqslant j \leqslant k$, suppose tree $T_t(j)$ contains $r_j + 1$ nodes, for some $r_j \geqslant 0$; then there are $S'_{r_j+1}(j-1)$ different trees $T_t(j)$. Let E be the set of elements of Q that are not in any tree $T_t(j)$ nor in the principal sequence $\mathfrak{s}_t(1)$. Then there are two cases:

1. $n-1 \in \mathfrak{s}_t(1)$: Since $(n-1)t = n$, we must have $1t^k = n-1$, and $k \geqslant 1$. So there are $C_{k-1}^{n-3}(k-1)!$ different $\mathfrak{s}_t(1)$. Let $r = |E| = (n-k-2) - (r_2 + \cdots + r_k)$. Then there are $C_{r_2,\dots,r_k,r}^{n-k-2}$ tuples $(r_2, \dots, r_k, r)$. For any $x \in E$, its image xt can be chosen from $E \cup \{n\}$. Then the number of transformations t in this case is M_n.
2. $n-1 \notin \mathfrak{s}_t(1)$: Then $k \leqslant n-3$, and there are $C_k^{n-3}k!$ different $\mathfrak{s}_t(1)$. Note that $n-1 \in E$, and $(n-1)t = n$ is fixed. Let $r = |E \setminus \{n-1\}| = (n-k-3) - (r_2 + \cdots + r_k)$. Then there are $C_{r_2,\dots,r_k,r}^{n-k-3}$ tuples $(r_2, \dots, r_k, r)$. For any $x \in E \setminus \{n-1\}$, xt can be chosen from $E \cup \{n\}$. Thus the number of transformations t in this case is N_n.

Altogether we have the desired formula. □

Let $\mathrm{b}_{\mathrm{bf}}(n) = |\mathbf{B}_{\mathrm{bf}}(n)|$. From Proposition 22 and Lemma 23 we have

Proposition 24. *For $n \geqslant 2$, if L is a bifix-free regular language with quotient complexity n, then its syntactic complexity $\sigma(L)$ satisfies that $\sigma(L) \leqslant \mathrm{b}_{\mathrm{bf}}(n)$, where $\mathrm{b}_{\mathrm{bf}}(n)$ is the cardinality of $\mathbf{B}_{\mathrm{bf}}(n)$ as in Lemma 23.*

When $2 \leqslant n \leqslant 4$, the set $\mathbf{B}_{\mathrm{bf}}(n)$ is a semigroup. But for $n \geqslant 5$, it is not a semigroup because $s_1 = [2, 3, n, \dots, n, n]$, $s_2 = [n, 3, 3, n, \dots, n, n] \in \mathbf{B}_{\mathrm{bf}}(n)$ while $s_1 s_2 = [3, 3, n, \dots, n, n] \notin \mathbf{B}_{\mathrm{bf}}(n)$. Hence $\mathrm{b}_{\mathrm{bf}}(n)$ is not a tight upper bound on the syntactic complexity of bifix-free regular languages in general. We look for a large semigroup contained in $\mathbf{B}_{\mathrm{bf}}(n)$ that can be the syntactic semigroup of a bifix-free regular language. For $n \geqslant 2$, let

$$\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n) = \{t \in \mathbf{B}_{\mathrm{bf}}(n) \mid \text{for all } i, j \in Q \text{ where } i \neq j, \text{ we have } it = jt = n \text{ or } it \neq jt\}.$$

(The reason for using the superscript $\leqslant 5$ will be made clear in Theorem 29.)

Proposition 25. *For $n \geqslant 2$, $\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$ is a semigroup contained in $\mathbf{B}_{\mathrm{bf}}(n)$ with cardinality*

$$\mathrm{w}_{\mathrm{bf}}^{\leqslant 5}(n) = |\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)| = \sum_{k=0}^{n-2} \left(C_k^{n-2}\right)^2 (n-2-k)!.$$

Proof. First, note that $\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n) = \mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n) \cap \mathbf{B}_{\mathrm{bf}}(n)$, and that $\mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$ is a semigroup contained in $\mathbf{B}_{\mathrm{sf}}(n)$ by Proposition 12. For any $t_1, t_2 \in \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$, we have $t_1 t_2 \in \mathbf{W}_{\mathrm{sf}}^{\leqslant 5}(n)$, and $(n-1)t_1 t_2 = nt_2 = n$; so $t_1 t_2 \in \mathbf{B}_{\mathrm{bf}}(n)$. Then $t_1 t_2 \in \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$, and $\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$ is a semigroup contained in $\mathbf{B}_{\mathrm{bf}}(n)$.

Pick any $t \in \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$. Note that $(n-1)t = n$ and $nt = n$ are fixed, and $1 \notin \mathrm{rng}(t)$. Let $Q' = Q \setminus \{n-1, n\}$, $E = \{i \in Q' \mid it = n\}$, and $D = Q' \setminus E$. Suppose $|E| = k$, where $0 \leq k \leq n-2$; then there are C_k^{n-2} choices of E. Elements of D are mapped to pairwise different elements of $Q \setminus \{1, n\}$; then there are $C_{n-2-k}^{n-2}(n-2-k)!$ different mappings $t|_D$. Altogether, we have $|\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)| = \sum_{k=0}^{n-2} \left(C_k^{n-2}\right)^2 (n-2-k)!$. □

Remark 26. Assume $n \geqslant 3$, and let $Q' = Q \setminus \{n-1, n\}$. Then the semigroup $\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$ is isomorphic to the symmetric inverse semigroup $\mathcal{IS}_{Q'}$; so $\mathrm{w}_{\mathrm{bf}}^{\leqslant 5}(n) = |\mathcal{IS}_{Q'}|$. Janson and Mazorchuk [13] showed that, for large n, the number $\mathrm{w}_{\mathrm{bf}}^{\leqslant 5}(n)$ is asymptotically

$$\mathrm{w}_{\mathrm{bf}}^{\leqslant 5}(n) \sim \frac{1}{\sqrt{2e}} e^{2\sqrt{n-2}-n+2} (n-2)^{n-\frac{7}{4}}.$$

Proposition 27. *For $n \geqslant 2$, let $Q' = Q \setminus \{n-1, n\}$ and $Q'' = Q \setminus \{1, n\}$. Then the semigroup $\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$ is generated by*

$$\mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n) = \{t \in \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n) \mid Q't = Q'' \text{ and } it \neq jt \text{ for all } i, j \in Q'\}.$$

Proof. The case for $n = 2$ is trivial since $\mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(2) = \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(2)$. Assume $n \geqslant 3$. We want to show that $\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n) = \langle \mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n) \rangle$. Since $\mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n) \subseteq \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$, we have $\langle \mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n) \rangle \subseteq \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$. Let $t \in \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$. By definition, $(n-1)t = nt = n$. Let $E_t = \{i \in Q' \mid it = n\}$. If $E_t = \emptyset$, then $t \in \mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n)$; otherwise, there exists $x \in Q''$ such that $x \notin \mathrm{rng}(t)$. We prove by induction on $|E_t|$ that $t \in \langle \mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n) \rangle$.

First note that, for all $t \in \mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n)$, $t|_{Q'}$ is an injective mapping from Q' to Q''. Consider $E_t = \{i\}$ for some $i \in Q'$. Since $|E_t| = 1$, $\mathrm{rng}(t) \cup \{x\} = Q''$. Let $t_1, t_2 \in \mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n)$ be defined by

1. $jt_1 = j+1$ for $j = 1, \dots, i-1$, $it_1 = n-1$, $jt_1 = j$ for $j = i+1, \dots, n-2$,
2. $1t_2 = x$, $jt_2 = (j-1)t$ for $j = 2, \dots, i$, $jt_2 = jt$ for $j = i+1, \dots, n-2$.

Then $t_1 t_2 = t$, and $t \in \langle \mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n) \rangle$.

Assume that any transformation $t \in \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$ with $|E_t| < k$ can be generated by $\mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n)$, where $1 < k < n-2$. Consider $t \in \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$ with $|E_t| = k$. Suppose $E_t = \{e_1, \dots, e_{k-1}, e_k\}$, and let $D_t = Q' \setminus E_t = \{d_1, \dots, d_l\}$, where $l = n-2-k$. By assumption, all $s \in \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$ with $|E_s| = k-1$ can be generated by $\mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n)$. Let s be such that $E_s = \{1, \dots, k-1\}$; then $1s = \cdots = (k-1)s = n$. In addition, let $ks = x$, and let $(k+j)s = d_j t$ for $j = 1, \dots, l$. Let $t' \in \mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n)$ be such that $e_j t' = j$ for $j = 1, \dots, k-1$, $kt' = n-1$, and $d_j t' = k+j$ for $j = 1, \dots, l$. Then $t's = t$, and $t \in \langle \mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n) \rangle$. Therefore, $\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n) = \langle \mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n) \rangle$. □

Theorem 28. *For $n \geqslant 2$, let $\mathcal{A}_n = (Q, \Sigma, \delta, 1, \{n-1\})$ be a DFA with alphabet Σ of size $(n-2)!$, where each $a \in \Sigma$ defines a distinct transformation $t_a \in \mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n)$. Then $L = L(\mathcal{A}_n)$ has quotient complexity $\kappa(L) = n$, and syntactic complexity $\sigma(L) = \mathrm{w}_{\mathrm{bf}}^{\leqslant 5}(n)$. Moreover, L is bifix-free.*

Proof. The case for $n = 2$ is easy to verify. Assume $n \geqslant 3$. We first show that all the states of $\mathcal{A}_n$ are reachable. Note that there exists $a \in \Sigma$ such that $t_a = [2, \ldots, n-1, n, n] \in \mathbf{G}_{\mathrm{bf}}^{\leqslant 5}(n)$. State $1 \in Q$ is the initial state, and a^{i-1} reaches state $i \in Q$ for $i = 2, \ldots, n$. Furthermore, for $1 \le i \le n-1$, state i accepts a^{n-1-i}, while for $j \neq i$, state j rejects it. Also, n is the empty state. Thus all the states of $\mathcal{A}_n$ are distinct, and $\kappa(L) = n$.

By Proposition 27, the syntactic semigroup of L is $\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$. So the syntactic complexity of L is $\sigma(L) = \mathrm{w}_{\mathrm{bf}}^{\leqslant 5}(n)$. By Proposition 22, L is bifix-free. □

Theorem 29. *For $2 \leqslant n \leqslant 5$, if a bifix-free regular language L has quotient complexity $\kappa(L) = n$, then $\sigma(L) \leqslant \mathrm{w}_{\mathrm{bf}}^{\leqslant 5}(n)$, and this bound is tight.*

Proof. We know by Proposition 22 that the upper bound on the syntactic complexity of bifix-free regular languages is reached by the largest semigroup contained in $\mathbf{B}_{\mathrm{bf}}(n)$. Since $\mathrm{w}_{\mathrm{bf}}^{\leqslant 5}(n) = \mathrm{b}_{\mathrm{bf}}(n)$ for $n = 2, 3$, and 4, $\mathrm{w}_{\mathrm{bf}}^{\leqslant 5}(n)$ is an upper bound, and it is tight by Theorem 28.

For $n = 5$, we have $\mathrm{b}_{\mathrm{bf}}(5) = |\mathbf{B}_{\mathrm{bf}}(5)| = 41$, and $\mathrm{w}_{\mathrm{bf}}^{\leqslant 5}(5) = |\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(5)| = 34$. Let $\mathbf{B}_{\mathrm{bf}}(5) \setminus \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(5) = \{\tau_1, \ldots, \tau_7\}$. We found for each τ_i a unique $t_i \in \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(5)$ such that the semigroup $\langle \tau_i, t_i \rangle$ is not a subset of $\mathbf{B}_{\mathrm{bf}}(5)$:

$$\begin{aligned}
\tau_1 &= [2,4,4,5,5], & t_1 &= [3,4,2,5,5];\\
\tau_2 &= [3,4,4,5,5], & t_2 &= [3,5,2,5,5];\\
\tau_3 &= [4,2,2,5,5], & t_3 &= [2,4,3,5,5];\\
\tau_4 &= [4,3,3,5,5], & t_4 &= [2,5,3,5,5];\\
\tau_5 &= [5,2,2,5,5], & t_5 &= [3,2,4,5,5];\\
\tau_6 &= [5,3,3,5,5], & t_6 &= [2,3,4,5,5];\\
\tau_7 &= [5,4,4,5,5], & t_7 &= [3,2,5,5,5].
\end{aligned}$$

Since $\langle \tau_i, t_i \rangle \subseteq T_L$, if both τ_i and t_i are in T_L, then $T_L \nsubseteq \mathbf{B}_{\mathrm{bf}}(5)$, and L is not bifix-free by Proposition 22. Thus, for $1 \leqslant i \leqslant 7$, at most one of τ_i and t_i can appear in T_L, and $|T_L| \leqslant 34$. Since $|\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(5)| = 34$ and $\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(5)$ is a semigroup, we have $\sigma(L) \leqslant 34 = \mathrm{w}_{\mathrm{bf}}^{\leqslant 5}(5)$ as the upper bound for $n = 5$. This bound is reached by the DFA $\mathcal{A}_5$ in Theorem 28. □

When $n \geqslant 6$, the semigroup $\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$ is no longer the largest semigroup contained in $\mathbf{B}_{\mathrm{bf}}(n)$, and the upper bound in Theorem 29 does not apply. We find another large semigroup $\mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n)$ suitable for bifix-free regular languages. For $n \geqslant 3$, let

$$\begin{aligned}
\mathbf{U}_n^1 &= \{t \in \mathbf{B}_{\mathrm{bf}}(n) \mid 1t = n\},\\
\mathbf{U}_n^2 &= \{t \in \mathbf{B}_{\mathrm{bf}}(n) \mid 1t = n-1\},\\
\mathbf{U}_n^3 &= \{t \in \mathbf{B}_{\mathrm{bf}}(n) \mid 1t \notin \{n, n-1\}, \text{ and } it \in \{n-1, n\} \text{ for all } i \neq 1\},
\end{aligned}$$

and let $\mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n) = \mathbf{U}_n^1 \cup \mathbf{U}_n^2 \cup \mathbf{U}_n^3$. Note that $\mathbf{U}_3^3 = \emptyset$.

Proposition 30. *For $n \geqslant 3$, $\mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n)$ is a semigroup contained in $\mathbf{B}_{\mathrm{bf}}(n)$ with cardinality*

$$\mathrm{w}_{\mathrm{bf}}^{\geqslant 6}(n) = |\mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n)| = (n-1)^{n-3} + (n-2)^{n-3} + (n-3)2^{n-3}.$$

Proof. First we show that $\mathbf{U}_n^1$ is a semigroup. For any $t_1, t_1' \in \mathbf{U}_n^1$, since $1(t_1 t_1') = (1t_1)t_1' = nt_1' = n$, we have $t_1 t_1' \in \mathbf{U}_n^1$. Next, let $t_2 \in \mathbf{U}_n^2$ and $t \in \mathbf{U}_n^1 \cup \mathbf{U}_n^2$. If $t \in \mathbf{U}_n^1$, then $1(t_2 t) = (n-1)t = n$ and $1(tt_2) = nt_2 = n$; so $t_2 t, tt_2 \in \mathbf{U}_n^1$. If $t \in \mathbf{U}_n^2$, then $1(t_2 t) = (n-1)t = n$ and $1(tt_2) = (n-1)t_2 = n$; so $t_2 t, tt_2 \in \mathbf{U}_n^1$ as well. Thus $\mathbf{U}_n^1 \cup \mathbf{U}_n^2$ is also a semigroup. For any $t_3 \in \mathbf{U}_n^3$ and $t' \in \mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n)$, since $it_3 \in \{n-1, n\}$ for all $i \neq 1$, and $(n-1)t' = nt' = n$, we have $i(t_3 t') = n$, and $t_3 t' \in \mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n)$. Also $1(t' t_3) = (1t')t_3 \in \{n-1, n\}$, so $t' t_3 \in \mathbf{U}_n^1 \cup \mathbf{U}_n^2$. Hence $\mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n)$ is a semigroup contained in $\mathbf{B}_{\mathrm{bf}}(n)$.

Note that $\mathbf{U}_n^1$, $\mathbf{U}_n^2$, and $\mathbf{U}_n^3$ are pairwise disjoint. For any $t \in \mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n)$, there are three cases:

1. $t \in \mathbf{U}_n^1$: for any $i \notin \{1, n-1, n\}$, it can be chosen from $Q \setminus \{1\}$. Then $|\mathbf{U}_n^1| = (n-1)^{n-3}$;
2. $t \in \mathbf{U}_n^2$: for any $i \notin \{1, n-1, n\}$, it can be chosen from $Q \setminus \{1, n-1\}$. Then $|\mathbf{U}_n^2| = (n-2)^{n-3}$;
3. $t \in \mathbf{U}_n^3$: now, $1t$ can be chosen from $Q \setminus \{1, n-1, n\}$. For any $i \notin \{1, n-1, n\}$, it has two choices: $it = n-1$ or n. Then $|\mathbf{U}_n^3| = (n-3)2^{n-3}$.

Therefore we have $|\mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n)| = (n-1)^{n-3} + (n-2)^{n-3} + (n-3)2^{n-3}$. □

Table 3 contains values of $\mathrm{w}_{\mathrm{bf}}^{\leqslant 5}(n)$ and $\mathrm{w}_{\mathrm{bf}}^{\geqslant 6}(n)$ for small n's. When $n \in \{3, 4\}$, we have $\mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n) = \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$, and these cases were already discussed. So we are only interested in larger values of n. When $n \geqslant 6$, one verifies that $\mathrm{w}_{\mathrm{bf}}^{\geqslant 6}(n) > \mathrm{w}_{\mathrm{bf}}^{\leqslant 5}(n)$; hence $\mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n)$ is larger than $\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(n)$.

Proposition 31. *For $n \geqslant 5$, the semigroup $\mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n)$ is generated by $\mathbf{G}_{\mathrm{bf}}^{\geqslant 6}(n) = \{a_1, a_2, a_3, b_1, \ldots, b_{n-3}, c_1, \ldots, c_m, d_1, \ldots, d_l\}$, where $m = (n-2)^{n-3} - 1$ and $l = (n-3)(2^{n-3} - 1)$, and*

- $a_1 = \binom{1}{n}\binom{n-1}{n}(2, \ldots, n-2)$, $a_2 = \binom{1}{n}\binom{n-1}{n}(2, 3)$, $a_3 = \binom{1}{n}\binom{n-1}{n}\binom{n-2}{2}$;

- *for* $1 \leqslant i \leqslant n-3$, $b_i = \binom{1}{n}\binom{n-1}{n}\binom{i+1}{n-1}$*;*
- *each* c_i *defines a distinct transformation in* $\mathbf{U}_n^2$ *other than* $[n-1, n, \ldots, n, n]$*;*
- *each* d_i *defines a distinct transformation in* $\mathbf{U}_n^3$ *other than* $[j, n, \ldots, n, n]$ *for all* $j \in \{2, \ldots, n-2\}$*.*

For $n = 5$*,* a_1 *and* a_2 *coincide, and 18 transformations suffice.*

Proof. Since $\mathbf{G}_{\mathrm{bf}}^{\geqslant 6}(n) \subseteq \mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n)$, we have $\langle \mathbf{G}_{\mathrm{bf}}^{\geqslant 6}(n)\rangle \subseteq \mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n)$. It remains to be shown that $\mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n) \subseteq \langle \mathbf{G}_{\mathrm{bf}}^{\geqslant 6}(n)\rangle$. Let $Q' = Q \setminus \{1, n-1, n\}$.

1. First consider $\mathbf{U}_n^1$. By Theorem 2, a_1, a_2 and a_3 together generate the semigroup
$$\mathbf{Y}' = \{t \in \mathbf{U}_n^1 \mid \text{ for all } i \in Q', it \in Q'\},$$
which is contained in $\mathbf{U}_n^1$. For any $t \in \mathbf{U}_n^1 \setminus \mathbf{Y}'$, let $E_t = \{i \in Q \mid it = n-1\}$; then $E_t \neq \emptyset$. Suppose $E_t = \{i_1, \ldots, i_k\}$, where $1 \leqslant k \leqslant n-3$. Then there exists $t' \in \mathbf{Y}'$ such that, for all $i \notin E_t$, $it' = it$. Let $s = b_{i_1-1} \cdots b_{i_k-1}$. Note that $E_t s = \{n-1\}$, and, for all $i \notin E_t$, $i(t's) = (it')s = it$. So $t's = t$, and $\langle a_1, a_2, a_3, b_1, \ldots, b_{n-3}\rangle = \mathbf{U}_n^1$.
2. Next, the transformations that are in $\mathbf{U}_n^2 \cup \mathbf{U}_n^3$ but not in $\mathbf{G}_{\mathrm{bf}}^{\geqslant 6}(n)$ are $t_i = [i, n, \ldots, n, n]$, where $2 \leqslant i \leqslant n-1$. Note that $d = \binom{1}{2}\binom{n-1}{n}\binom{Q'}{n-1} \in \mathbf{G}_{\mathrm{bf}}^{\geqslant 6}(n)$, and, for each $i \in \{2, \ldots, n-1\}$, $s_i = \binom{1}{n}\binom{n-1}{n}\binom{2}{i} \in \mathbf{U}_n^1$. Then $t_i = ds_i \in \langle \mathbf{G}_{\mathrm{bf}}^{\geqslant 6}(n)\rangle$, and $\mathbf{U}_n^2 \cup \mathbf{U}_n^3 \subseteq \langle \mathbf{G}_{\mathrm{bf}}^{\geqslant 6}(n)\rangle$.

Therefore $\mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n) = \langle \mathbf{G}_{\mathrm{bf}}^{\geqslant 6}(n)\rangle$. $\square$

Theorem 32. *For* $n \geqslant 5$*, let* $\mathcal{A}_n' = (Q, \Sigma, \delta, 1, \{n-1\})$ *be a DFA with alphabet* Σ *of size* 18 *if* $n = 5$ *or* $(n-2)^{n-3} + (n-3)2^{n-3} + 2$ *if* $n \geqslant 6$*, where each letter defines a transformation as in Proposition 31. Then* $L' = L(\mathcal{A}_n')$ *has quotient complexity* $\kappa(L') = n$*, and syntactic complexity* $\sigma(L') = \mathrm{w}_{\mathrm{bf}}^{\geqslant 6}(n)$*. Moreover,* L' *is bifix-free.*

Proof. First, for all $i \in Q \setminus \{1\}$, there exists $a \in \Sigma$ such that $t_a = [i, n, \ldots, n, n] \in \mathbf{G}_{\mathrm{bf}}^{\geqslant 6}(n)$, and state i is reachable by a. So all the states in Q are reachable. Next, there exist $b, c \in \Sigma$ such that $t_b = [n-1, n, \ldots, n, n] \in \mathbf{G}_{\mathrm{bf}}^{\geqslant 6}(n)$ and $t_c = [n, 3, 4, \ldots, n-1, n, n] \in \mathbf{G}_{\mathrm{bf}}^{\geqslant 6}(n)$. The initial state accepts b, while all other states reject it. For $2 \leqslant i \leqslant n-2$, state i accepts c^{n-i-1}, while all other states reject it. Also, state $n-1$ is the only final state, and state n is the empty state. Then all the states in Q are distinct, and $\kappa(L') = n$.

By Proposition 31, the syntactic semigroup of L' is $\mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(n)$; so $\sigma(L') = \mathrm{w}_{\mathrm{bf}}^{\geqslant 6}(n)$. By Proposition 22, L' is bifix-free. $\square$

Theorem 33. *If* L *is a bifix-free regular language with* $\kappa(L) = 6$*, then* $\sigma(L) \leqslant \mathrm{w}_{\mathrm{bf}}^{\geqslant 6}(6) = 213$ *and this is a tight bound.*

Proof. Since $|\mathbf{B}_{\mathrm{bf}}(6)| = 339$ and $|\mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(6)| = 213$, there are 126 transformations $\tau_1, \ldots, \tau_{126}$ in $\mathbf{B}_{\mathrm{bf}}(6) \setminus \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(6)$. For each τ_i, we enumerated transformations in $\mathbf{W}_{\mathrm{bf}}^{\geqslant 6}(6)$ using *GAP* and found a unique $t_i \in \mathbf{W}_{\mathrm{bf}}^{\leqslant 5}(6)$ such that $\langle t_i, \tau_i\rangle \not\subseteq \mathbf{B}_{\mathrm{bf}}(6)$. Thus, for each i, at most one of t_i and τ_i can appear in the syntactic semigroup T_L of L. So we lower the bound to $\sigma(L) \leqslant 213$. This bound is reached by the DFA $\mathcal{A}_6'$ in Theorem 32; so it is a tight upper bound for $n = 6$. $\square$

Conjecture 34 (*Bifix-Free Regular Languages*)**.** *If* L *is a bifix-free regular language with* $\kappa(L) = n \geqslant 7$*, then* $\sigma(L) \leqslant \mathrm{w}_{\mathrm{bf}}^{\geqslant 6}(n)$*.*

7. Factor-free regular languages

Let L be a factor-free regular language with $\kappa(L) = n$. Since factor-free regular languages are also bifix-free, L as a quotient is uniquely reachable, ε is the only final quotient of L, and L also has the empty quotient. As in Section 6, we assume that Q is the set of states of quotient DFA of L, in which 1 is the initial state, and states $n-1$ and n correspond to the quotients ε and $\emptyset$, respectively. For $n \geqslant 2$, let

$$\mathbf{B}_{\mathrm{ff}}(n) = \{t \in \mathbf{B}_{\mathrm{bf}}(n) \mid \text{ for all } j \geqslant 1, 1t^j = n-1 \Rightarrow it^j = n \ \forall\, i, 1 < i < n-1\}.$$

We first have the following observation.

Proposition 35. *If* L *is a regular language with quotient complexity* n *and syntactic semigroup* T_L*, then the following hold:*

1. *If* L *is factor-free, then* T_L *is a subset of* $\mathbf{B}_{\mathrm{ff}}(n)$*.*
2. *If* ε *is the only final quotient of* L*, and* $T_L \subseteq \mathbf{B}_{\mathrm{ff}}(n)$*, then* L *is factor-free.*

Proof. 1. Assume L is factor-free. Then L is bifix-free, and $T_L \subseteq \mathbf{B}_{\mathrm{bf}}(n)$ by Proposition 22. For any transformation $t_w \in T_L$ performed by some non-empty word w, if $1t_w^j = n-1$ for some $j \geqslant 1$, then $w^j \in L$. If we also have $it_w^j \neq n$ for some $i \in Q \setminus \{1\}$, then $i \notin \{n-1, n\}$ as $(n-1)t = nt = n$ for all $t \in \mathbf{B}_{\mathrm{ff}}(n)$. Thus there exist non-empty words u and v such that state i is reachable by u, and state it_w^j accepts v. So $uw^jv \in L$, which is a contradiction. Hence $T_L \subseteq \mathbf{B}_{\mathrm{ff}}(n)$.

2. Since ε is the only final state and $\mathbf{B}_{\mathrm{ff}}(n) \subseteq \mathbf{B}_{\mathrm{bf}}(n)$, L is bifix-free by Proposition 22. If L is not factor-free, then there exist non-empty words u, v and w such that $w, uwv \in L$. Thus $1t_w = n-1$, and $1t_{uwv} = 1(t_ut_wt_v) = n-1$. Since L is bifix-free, $1t_u \neq 1$ and $nt_v = n$; thus $(1t_u)t_w \neq n$, which contradicts the assumption that $t_w \in T_L \subseteq \mathbf{B}_{\mathrm{ff}}(n)$. Therefore L is factor-free. $\square$

The properties of suffix- and bifix-free regular languages still apply to factor-free regular languages. Moreover, we have

Lemma 36. *For all $t \in \mathbf{B}_{\rm ff}(n)$ and $i \notin \mathfrak{s}_t(1)$, if $n-1 \in \mathfrak{s}_t(1)$, then $n \in \mathfrak{s}_t(i)$.*

Proof. Suppose $n-1 = 1t^k \in \mathfrak{s}_t(1)$ for some $k \geqslant 1$. If $n \notin \mathfrak{s}_t(i)$, then for all $j \geqslant 1$, $it^j \neq n$. In particular, $it^k \neq n$, which contradicts the definition of $\mathbf{B}_{\rm ff}(n)$. Therefore $n \in \mathfrak{s}_t(i)$. □

Lemma 37. *We have $|\mathbf{B}_{\rm ff}(2)| = 1$, and for $n \geqslant 3$, $|\mathbf{B}_{\rm ff}(n)| = N_n + O_n$, where*

$$O_n = 1 + \sum_{k=2}^{n-2} C_{k-1}^{n-3}(k-1)! \sum_{\substack{r_2+\cdots+r_k+r\\ =n-k-2}} C_{r_2,\ldots,r_k,r}^{n-k-2} S'_{r+1}(k) \prod_{j=2}^{k} S'_{r_j+1}(j-1),$$

and N_n as given in Equation Eq. (3).

Proof. First we have $\mathbf{B}_{\rm ff}(2) = \{[2,2]\}$ and $|\mathbf{B}_{\rm ff}(2)| = 1$. Assume $n \geqslant 3$. Let $t \in \mathbf{B}_{\rm ff}(n)$ be any transformation. Suppose $\mathfrak{s}_t(1) = 1, 1t, \ldots, 1t^k, n$, where $0 \leqslant k \leqslant n-2$. Then there are two cases:

1. $n-1 \in \mathfrak{s}_t(1)$. Since $(n-1)t = n$, we have $n-1 = 1t^k$, and $k \geqslant 1$. If $k = 1$, then $1t = n-1$, and $it = n$ for all $i \neq 1$; such a t is unique. Consider $k \geqslant 2$. There are $C_{k-1}^{n-3}(k-1)!$ different $\mathfrak{s}_t(1)$. For $2 \leqslant j \leqslant k$, suppose there are $r_j + 1$ nodes in tree $T_t(j)$; then there are $S'_{r_j+1}(j-1)$ such trees. Let E be the set of elements x that are not in any tree $T_t(j)$ nor in $\mathfrak{s}_t(1)$, and let $r = |E| = (n-k-2) - (r_2 + \cdots + r_k)$. By Lemma 36, $n \in \mathfrak{s}_t(x)$ for all $x \in E$. Then the union of paths $P_t(x)$ for all $x \in E$ form a labeled tree rooted at n with height at most k, and there are $S'_{r+1}(k)$ such trees. Thus the number of transformations in this case is O_n.
2. $n-1 \notin \mathfrak{s}_t(1)$. Now, for all $j \geqslant 1$, $1t^j \neq n-1$. Then $t \in \mathbf{B}_{\rm bf}(n)$. As in the proof of Lemma 23, the number of transformations in this case is N_n.

   Altogether we have the desired formula. □

Let $\mathrm{b}_{\rm ff}(n) = |\mathbf{B}_{\rm ff}(n)|$. From Proposition 35 and Lemma 37 we have

Proposition 38. *For $n \geqslant 2$, if L is a factor-free regular language with quotient complexity n, then its syntactic complexity $\sigma(L)$ satisfies that $\sigma(L) \leqslant \mathrm{b}_{\rm ff}(n)$, where $\mathrm{b}_{\rm ff}(n)$ is the cardinality of $\mathbf{B}_{\rm ff}(n)$ as in Lemma 37.*

The tight upper bound on the syntactic complexity of factor-free regular languages is reached by the largest semigroup contained in $\mathbf{B}_{\rm ff}(n)$. When $2 \leqslant n \leqslant 4$, $\mathbf{B}_{\rm ff}(n)$ is a semigroup. The languages $L_2 = \varepsilon$, $L_3 = a$ over alphabet $\{a, b\}$, and $L_4 = ab^*a$ have syntactic complexities $1 = \mathrm{b}_{\rm ff}(2)$, $2 = \mathrm{b}_{\rm ff}(3)$, and $6 = \mathrm{b}_{\rm ff}(4)$, respectively. So $\mathrm{b}_{\rm ff}(n)$ is a tight upper bound for $n \in \{2,3,4\}$. However, the set $\mathbf{B}_{\rm ff}(n)$ is not a semigroup for $n \geqslant 5$, because $s_1 = [2, 3, \ldots, n-1, n, n]$, $s_2 = \binom{n-1}{n}\binom{2}{n-1}\binom{1}{n} = [n, n-1, 3, \ldots, n-2, n, n] \in \mathbf{B}_{\rm ff}(n)$ but $s_1s_2 = [n-1, 3, \ldots, n-2, n, n, n] \notin \mathbf{B}_{\rm ff}(n)$.

Next, we find a large semigroup that can be the syntactic semigroup of a factor-free regular language. For $n \geqslant 3$, let $t_0 = \binom{Q\setminus\{1\}}{n}\binom{1}{n-1} = [n-1, n, \ldots, n]$, and let $\mathbf{W}_{\rm ff}(n) = \mathbf{U}_n^1 \cup \{t_0\} \cup \mathbf{U}_n^3$.

Proposition 39. *For $n \geqslant 3$, $\mathbf{W}_{\rm ff}(n)$ is a semigroup contained in $\mathbf{B}_{\rm ff}(n)$ with cardinality*

$$\mathrm{w}_{\rm ff}(n) = |\mathbf{W}_{\rm ff}(n)| = (n-1)^{n-3} + (n-3)2^{n-3} + 1.$$

Proof. As we have shown in the proof of Proposition 30, $\mathbf{U}_n^1$ is a semigroup. For any $t \in \mathbf{U}_n^1 \cup \{t_0\}$, since $t_0 \in \mathbf{U}_n^2$, we have $tt_0, t_0t \in \mathbf{U}_n^1$; so $\mathbf{U}_n^1 \cup \{t_0\}$ is also a semigroup. We also know that, for any $t_3 \in \mathbf{U}_n^3$ and $t' \in \mathbf{W}_{\rm ff}(n)$, since $\mathbf{W}_{\rm ff}(n) \subseteq \mathbf{W}_{\rm bf}^{\geqslant 6}(n)$, $i(t_3t') = n$ for all $i \neq 1$; so $t_3t' \in \mathbf{W}_{\rm ff}(n)$. If $t' \in \mathbf{U}_n^1 \cup \{t_0\}$, then $1t't_3 = n$ and $t't_3 \in \mathbf{U}_n^1$; otherwise, $t' \in \mathbf{U}_n^3$, and $t't_3 = t_2$ or $\binom{Q}{n} \in \mathbf{U}_n^1$. Hence $\mathbf{W}_{\rm ff}(n)$ is a semigroup.

For any $t \in \mathbf{U}_n^1$, since $1t = n$, we have $t \in \mathbf{B}_{\rm ff}(n)$. For any $t \in \mathbf{U}_n^3$, $1t \neq n-1$, and $it^2 = n$ for all $i \in \{2, \ldots, n\}$; then $t \in \mathbf{B}_{\rm ff}(n)$ as well. Clearly $t_0 \in \mathbf{B}_{\rm ff}(n)$. Hence $\mathbf{W}_{\rm ff}(n)$ is contained in $\mathbf{B}_{\rm ff}(n)$.

We know that $|\mathbf{U}_n^1| = (n-1)^{n-3}$ and $|\mathbf{U}_n^3| = (n-3)2^{n-3}$. Therefore $|\mathbf{W}_{\rm ff}(n)| = (n-1)^{n-3} + (n-3)2^{n-3} + 1$. □

For $n \in \{3, 4\}$, we have $\mathbf{W}_{\rm ff}(n) = \mathbf{B}_{\rm ff}(n)$. So we are interested in larger values of n in the rest of this section.

Proposition 40. *For $n \geqslant 5$, the semigroup $\mathbf{W}_{\rm ff}(n)$ is generated by $\mathbf{G}_{\rm ff}(n) = \{a_1, a_2, a_3, b_1, \ldots, b_{n-3}, c_1, \ldots, c_m\}$, where $m = (n-3)(2^{n-3}-1)$, and*

- $a_1 = \binom{1}{n}\binom{n-1}{n}(2, \ldots, n-2)$, $a_2 = \binom{1}{n}\binom{n-1}{n}(2,3)$, $a_3 = \binom{1}{n}\binom{n-1}{n}\binom{n-2}{2}$;
- *for $1 \leqslant i \leqslant n-3$, $b_i = \binom{1}{n}\binom{n-1}{n}\binom{i+1}{n-1}$;*
- *each c_i defines a distinct transformation in $\mathbf{U}_n^3$ other than $[j, n, \ldots, n, n]$ for all $j \in \{2, \ldots, n-2\}$.*

For $n = 5$, a_1 and a_2 coincide, and 10 transformations suffice.

Proof. We know from the proof of Proposition 31 that $\mathbf{U}_n^1$ is generated by $\{a_1, a_2, a_3, b_1, \dots, b_{n-3}\}$. Also, the transformations that are in $\{t_0\} \cup \mathbf{U}_n^3$ but not in $\mathbf{G}_{\rm ff}(n)$ are $t_j = [j, n, \dots, n, n]$, where $j \in \{2, \dots, n-1\}$. Let $Q' = Q \setminus \{1, n-1, n\}$. Each t_j is a composition of $d = \binom{n-1}{n}\binom{Q'}{n-1}\binom{1}{2} \in \mathbf{G}_{\rm bf}^{\geqslant 6}(n)$ and $s_j = \binom{1}{n}\binom{n-1}{n}\binom{2}{j} \in \mathbf{U}_n^1$. Therefore $\langle \mathbf{G}_{\rm ff}(n)\rangle = \mathbf{W}_{\rm ff}(n)$. □

Theorem 41. *For $n \geqslant 5$, let $\mathcal{A}_n = (Q, \Sigma, \delta, 1, \{n-1\})$ be a DFA with alphabet Σ of size 10 if $n = 5$ or $(n-3)2^{n-3} + 3$ if $n \geqslant 6$, where each letter defines a transformation as in Proposition 40. Then $L = L(\mathcal{A}_n)$ has quotient complexity $\kappa(L) = n$, and syntactic complexity $\sigma(L) = {\rm w}_{\rm ff}(n)$. Moreover, L is factor-free.*

Proof. Since $\mathbf{G}_{\rm ff}(n) \subseteq \mathbf{G}_{\rm bf}^{\geqslant 6}(n)$, the DFA $\mathcal{A}_n$ can be obtained from the DFA $\mathcal{A}'_n$ of Theorem 32 by restricting the alphabet. The words used to show that all the states of $\mathcal{A}'_n$ are reachable and distinct still exist in $\mathcal{A}_n$. Then we have $\kappa(L) = n$. By Proposition 40, the syntactic semigroup of L is $\mathbf{W}_{\rm ff}(n)$; so $\sigma(L) = {\rm w}_{\rm ff}(n)$. By Proposition 35, L is factor-free. □

Theorem 42. *For $n \in \{5, 6\}$, if L is a factor-free regular language with $\kappa(L) = n$, then $\sigma(L) \leqslant {\rm w}_{\rm ff}(n)$ and this is a tight upper bound.*

Proof. For $n = 5$, $|\mathbf{B}_{\rm ff}(5)| = 31$, and $|\mathbf{W}_{\rm ff}(5)| = 25$. There are 6 transformations $\tau_1, \dots, \tau_6$ in $\mathbf{B}_{\rm ff}(5) \setminus \mathbf{W}_{\rm ff}(5)$. For each τ_i, $1 \leqslant i \leqslant 6$, we found a unique $t_i \in \mathbf{W}_{\rm ff}(5)$ such that $\langle t_i, \tau_i\rangle \not\subseteq \mathbf{B}_{\rm ff}(5)$:

$$\begin{aligned}
&\tau_1 = [2, 3, 4, 5, 5], && t_1 = [5, 2, 2, 5, 5],\\
&\tau_2 = [2, 3, 5, 5, 5], && t_2 = [5, 4, 2, 5, 5],\\
&\tau_3 = [2, 5, 3, 5, 5], && t_3 = [5, 3, 3, 5, 5],\\
&\tau_4 = [3, 2, 5, 5, 5], && t_4 = [5, 2, 4, 5, 5],\\
&\tau_5 = [3, 4, 2, 5, 5], && t_5 = [5, 3, 2, 5, 5],\\
&\tau_6 = [3, 5, 2, 5, 5], && t_6 = [5, 3, 4, 5, 5].
\end{aligned}$$

For each $1 \leqslant i \leqslant 6$, at most one of t_i and τ_i can appear in the syntactic semigroup T_L of a factor-free regular language L. Then $\sigma(L) = |T_L| \leqslant 25$. By Theorem 41, this upper bound is tight for $n = 5$.

For $n = 6$, $|\mathbf{B}_{\rm ff}(6)| = 246$, and $|\mathbf{W}_{\rm ff}(6)| = 150$. There are 96 transformations $\tau_1, \dots, \tau_{96}$ in $\mathbf{B}_{\rm ff}(6) \setminus \mathbf{W}_{\rm ff}(6)$. For each τ_i, $1 \leqslant i \leqslant 96$, we enumerated the transformations in $\mathbf{W}_{\rm ff}(6)$ using *GAP* and found a unique $t_i \in \mathbf{W}_{\rm ff}(6)$ such that $\langle t_i, \tau_i\rangle \not\subseteq \mathbf{B}_{\rm ff}(6)$. Thus 150 is a tight upper bound for $n = 6$. □

Conjecture 43 (*Factor-Free Regular Languages*). *If L is a factor-free regular language with $\kappa(L) = n$, where $n \geqslant 7$, then $\sigma(L) \leqslant {\rm w}_{\rm ff}(n)$.*

8. Quotient complexity of the reversal of free languages

It has been shown in [6] that for certain regular languages with maximal syntactic complexity, the reverse languages have maximal quotient complexity. This is also true for some free languages, as we now show.

In this section, we consider *non-deterministic finite automata* (NFA). An NFA $\mathcal{N}$ is a quintuple $\mathcal{N} = (Q, \Sigma, \delta, I, F)$, where Q, Σ, and F are as in a DFA, $\delta : Q \times \Sigma \to 2^Q$ is the non-deterministic transition function, and I is the set of initial states. For any word $w \in \Sigma^*$, the *reverse* of w is defined inductively as follows: $w^R = \varepsilon$ if $w = \varepsilon$, and $w^R = u^R a$ if $w = au$ for some $a \in \Sigma$ and $u \in \Sigma^*$. The *reverse* of any language L is the language $L^R = \{w^R \mid w \in L\}$. For any finite automaton (DFA or NFA) $\mathcal{M}$, we denote by $\mathcal{M}^R$ the NFA obtained by reversing all the transitions of $\mathcal{M}$ and exchanging the roles of initial and final states, and by $\mathcal{M}^D$, the DFA obtained by applying the subset construction to $\mathcal{M}$. Then $L(\mathcal{M}^R) = (L(\mathcal{M}))^R$, and $L(\mathcal{M}^D) = L(\mathcal{M})$. To simplify our proofs, we use an observation from [2] that, for any NFA $\mathcal{N}$ that does not have any empty states, if the automaton $\mathcal{N}^R$ is deterministic, then the DFA $\mathcal{N}^D$ is minimal if only reachable subsets are included in the subset construction.

Theorem 44. *For $n \geqslant 4$, the reverse of the prefix-free regular language accepted by the DFA $\mathcal{A}_n$ of Theorem 4 restricted to $\{a, c, d_{n-2}\}$ has $2^{n-2} + 1$ quotients, which is the maximum possible for a prefix-free regular language.*

Proof. Let $\mathcal{B}_n$ be the DFA $\mathcal{A}_n$ restricted to $\{a, c, d_{n-2}\}$. Since $L(\mathcal{A}_n)$ is prefix-free, so is $L_n = L(\mathcal{B}_n)$. We show that $\kappa(L_n^R) = 2^{n-2} + 1$.

Let $\mathcal{N}_n$ be the NFA obtained by removing unreachable states from the NFA $\mathcal{B}_n^R$. (See Fig. 4 for $\mathcal{N}_6$.) We first prove that the following $2^{n-2} + 1$ sets of states of $\mathcal{N}_n$ are reachable: $\{\{n-1\}\} \cup \{S \mid S \subseteq \{1, \dots, n-2\}\}$.

The singleton set $\{n-1\}$ of initial states of $\mathcal{N}_n$ is reached by ε. From $\{n-1\}$ we reach the empty set by a. The set $\{n-2\}$ is reached by d_{n-2} from $\{n-1\}$, and from here, $\{1\}$ is reached by a^{n-3}. From any set $\{1, 2, \dots, i\}$, where $1 \leqslant i < n-2$, we reach $\{1, 2, \dots, i, i+1\}$ by ca^{n-3}. Thus we reach $\{1, 2, \dots, n-2\}$ from $\{1\}$ by $(ca^{n-3})^{n-3}$. Now assume that any set S of cardinality $l \leqslant n-2$ can be reached; then we can get a set of cardinality $l-1$ by deleting an element j from S by applying $a^j d_{n-2} a^{n-2-j}$. Hence all the subsets of $\{1, 2, \dots, n-2\}$ can be reached.

The automaton $\mathcal{N}_n^R$ is a subset of $\mathcal{A}_n$, and it is deterministic. Then $\mathcal{N}_n^D$ is minimal. Hence $\kappa(L_n^R) = 2^{n-2} + 1$, which is the maximal quotient complexity of reversal of prefix-free languages as shown in [11]. □

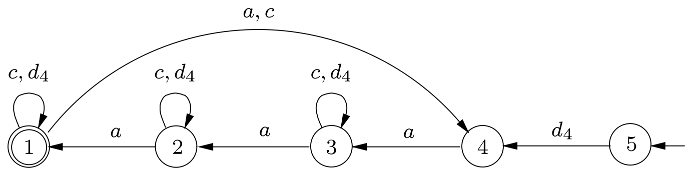

Fig. 4. NFA $\mathcal{N}_6$ of L_6^R with quotient complexity $\kappa(L_6^R) = 17$; empty state omitted.

It is interesting that, for suffix-, bifix-, and factor-free regular languages, although we do not have tight upper bounds on their syntactic complexities, some languages in these classes with large syntactic complexities have their reverse languages reaching the upper bounds on the quotient complexities for the reversal operation.

Theorem 45. *For $n \geqslant 4$, the reverse of the suffix-free regular language accepted by the DFA $\mathcal{A}'_n$ of Theorem 19 restricted to $\{a_1, a_2, a_3, c\}$ has $2^{n-2} + 1$ quotients, which is the maximum possible for a suffix-free regular language.*

Proof. Let $\mathcal{C}_n$ be the DFA $\mathcal{A}'_n$ restricted to the alphabet $\{a_1, a_2, a_3, c\}$. Since $L(\mathcal{A}'_n)$ is suffix-free, so is $L'_n = L(\mathcal{C}_n)$. Let $\mathcal{N}'_n$ be the NFA obtained from $\mathcal{C}_n^R$ by removing unreachable states. Fig. 5 shows the NFA $\mathcal{N}'_6$.

Applying the subset construction to $\mathcal{N}'_n$, we get a DFA $\mathcal{N}'^D_n$. Its initial state is a singleton set $\{2\}$. From the initial state, we can reach state $\{2, 3, \ldots, i\}$ by $(a_3a_1^{n-3})^{i-2}$, where $3 \leqslant i \leqslant n-1$. Then the state $\{2, 3, \ldots, n-1\}$ is reached from $\{2\}$ by $(a_3a_1^{n-3})^{n-3}$. Assume that any set S of cardinality l can be reached, where $2 \leqslant l \leqslant n-2$. If $j \in S$, then we can reach $S' = S \setminus \{j\}$ from S by $a_1^{j-1}a_3a_1^{n-j-1}$. So all the non-empty subsets of $\{2, 3, \ldots, n-1\}$ can be reached. We can also reach the singleton set $\{1\}$ from $\{2\}$ by c, and, from there, the empty state by c again. Hence $\mathcal{N}'^D_n$ has $2^{n-2} + 1$ reachable states.

Since the automaton $\mathcal{N}'^R_n$, the reverse of $\mathcal{N}'_n$, is a subset of $\mathcal{C}_n$, it is deterministic; hence $\mathcal{N}'^D_n$ is minimal. Then the quotient complexity of L'^R_n is $2^{n-2} + 1$, which meets the upper bound for reversal of suffix-free regular languages [10]. $\square$

Theorem 46. *For $n \geqslant 5$, the reverse of the factor-free regular language accepted by the DFA $\mathcal{A}_n$ of Theorem 41 restricted to the alphabet $\{a_1, a_2, a_3, c\}$, where $c = [2, n-1, n, \ldots, n, n] \in \mathbf{G}_{\mathrm{ff}}(n)$, has $2^{n-3} + 2$ quotients, which is the maximum possible for a bifix- or factor-free regular language.*

Proof. Let $\mathcal{D}_n$ be the DFA $\mathcal{A}_n$ restricted to the alphabet $\{a_1, a_2, a_3, c\}$; then $L''_n = L(\mathcal{D}_n)$ is factor-free. Let $\mathcal{N}''_n$ be the NFA obtained from $\mathcal{D}_n^R$ by removing unreachable states. An example of $\mathcal{N}''_n$ is shown in Fig. 6.

Note that $\mathcal{N}''_n$ can be obtained from the NFA $\mathcal{N}'_{n-1}$ in Theorem 45 by adding a new state $n-1$, which is the only initial state in $\mathcal{N}''_n$, and the transition from $\{n-1\}$ to $\{2\}$ under input c. We know that all non-empty subsets of $\{2, 3, \ldots, n-2\}$ are reachable from $\{2\}$. The final state $\{1\}$ is also reachable from $\{2\}$. From the initial state $n-1$, we reach the empty state under input a_1. Then $\mathcal{N}''^D_n$ has $2^{n-3} + 2$ reachable states.

Since $\mathcal{N}''^R_n$ is a subset of $\mathcal{D}_n$ and it is deterministic, the DFA $\mathcal{N}''^D_n$ is minimal. Therefore $\kappa(L''^R_n) = 2^{n-3} + 2$, and it reaches the upper bound for reversal of both bifix- and factor-free regular languages with quotient complexity n [4]. $\square$

9. Conclusions

Our results are summarized in Tables 2 and 3. Each cell of Table 2 shows the syntactic complexity bounds of prefix- and suffix-free regular languages, in that order, with a particular alphabet size. Table 3 is structured similarly for bifix- and factor-free regular languages. The figures in bold type are tight bounds verified by *GAP*. To compute the bounds for suffix-, bifix-, and factor-free languages, we enumerated semigroups generated by elements of $\mathbf{B}_{\mathrm{sf}}(n)$, $\mathbf{B}_{\mathrm{bf}}(n)$, and $\mathbf{B}_{\mathrm{ff}}(n)$ that are contained in $\mathbf{B}_{\mathrm{sf}}(n)$, $\mathbf{B}_{\mathrm{bf}}(n)$, and $\mathbf{B}_{\mathrm{ff}}(n)$, respectively, and recorded the largest ones. By Propositions 5, 22 and 35, we obtained the desired bounds from the enumeration. The asterisk $*$ indicates that the bound is already tight for a smaller alphabet. In Table 2, the last four rows include the tight upper bound n^{n-2} for prefix-free languages, $\mathrm{w}_{\mathrm{sf}}^{\leqslant 5}(n)$, which is a tight upper bound for $2 \leqslant n \leqslant 5$ for suffix-free languages, conjectured upper bound $\mathrm{w}_{\mathrm{sf}}^{\geqslant 6}(n)$ for suffix-free languages, and a weaker upper bound $\mathrm{b}_{\mathrm{sf}}(n)$ for suffix-free languages. In Table 3, the last four rows include $\mathrm{w}_{\mathrm{bf}}^{\leqslant 5}(n)$, which is a tight upper bound for bifix-free languages for $2 \leqslant n \leqslant 5$, conjectured upper bounds $\mathrm{w}_{\mathrm{bf}}^{\geqslant 6}(n)$ for bifix-free languages and $\mathrm{w}_{\mathrm{ff}}(n)$ for factor-free languages, and weaker upper bounds $\mathrm{b}_{\mathrm{bf}}(n)$ for bifix-free languages and $\mathrm{b}_{\mathrm{ff}}(n)$ for factor-free languages.

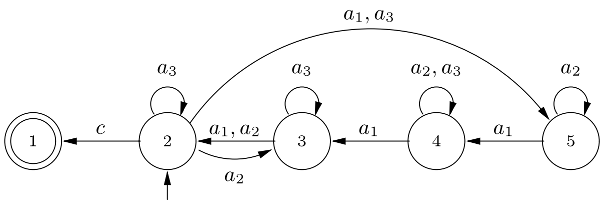

Fig. 5. NFA $\mathcal{N}'_6$ of L'^R_6 with quotient complexity $\kappa(L'^R_6) = 17$; empty state omitted.

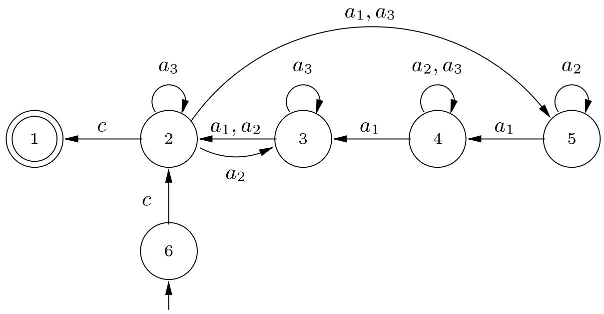

Fig. 6. NFA $\mathcal{N}_7''$ of $L_7''^R$ with quotient complexity $\kappa(L_7''^R) = 18$; empty state omitted.

Table 2
Syntactic complexities of prefix- and suffix-free regular languages.

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
$\|\Sigma\| = 1$	**1**	**2**	**3**	**4**	**5**
$\|\Sigma\| = 2$	$*$	**3/3**	**11/11**	**49/49**	?
$\|\Sigma\| = 3$	$*$	$*$	**14/13**	**95/61**	?
$\|\Sigma\| = 4$	$*$	$*$	**16**/$*$	**110/67**	?
$\|\Sigma\| = 5$	$*$	$*$	$*$	**119/73**	?
$\|\Sigma\| = 6$	$*$	$*$	$*$	**125**/ $*$	? /501
$\|\Sigma\| = 7$	$*$	$*$	$*$	$*$	**1296**/ ?
$\|\Sigma\| = 8$	$*$	$*$	$*$	$*$	$*$ /**629**
...					
n^{n-2}	**1**	**3**	**16**	**125**	**1296**
$w_{sf}^{\leqslant 5}(n)$	**1**	**3**	**13**	**73**	501
$w_{sf}^{\geqslant 6}(n)$	**1**	**3**	11	67	**629**
$b_{sf}(n)$	**1**	**3**	15	115	1169

Table 3
Syntactic complexities of bifix- and factor-free regular languages.

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
$\|\Sigma\| = 1$	**1**	**2**	**3**	**4**	**5**
$\|\Sigma\| = 2$	$*$	$*$	**7/6**	**20/12**	?
$\|\Sigma\| = 3$	$*$	$*$	$*$	**31/16**	?
$\|\Sigma\| = 4$	$*$	$*$	$*$	**32/19**	?
$\|\Sigma\| = 5$	$*$	$*$	$*$	**33/20**	?
$\|\Sigma\| = 6$	$*$	$*$	$*$	**34**/ ?	?
...					
$w_{bf}^{\leqslant 5}(n)$	**1**	**2**	**7**	**34**	209
$w_{bf}^{\geqslant 6}(n)$	**1**	**2**	**7**	33	**213**
$w_{ff}(n)$	**1**	**2**	**6**	**25**	**150**
$b_{bf}(n)/b_{ff}(n)$	**1/1**	**2/2**	**7/6**	41/31	339/246

References

[1] J. Berstel, D. Perrin, C. Reutenauer, Codes and Automata, Cambridge University Press, 2009.
[2] J. Brzozowski, Canonical regular expressions and minimal state graphs for definite events, in: Mathematical Theory of Automata, in: Polytechnic Press, Polytechnic Institute of Brooklyn, N.Y., Vol. 12, 1962, pp. 529–561.
[3] J. Brzozowski, Quotient complexity of regular languages, J. Autom. Lang. Comb 15 (2010) 71–89.
[4] J. Brzozowski, G. Jirásková, B. Li, J. Smith, Quotient complexity of bifix-, factor-, and subword-free regular languages, in: P. Dömösi, S. Iván (Eds.), Automata and Formal Languages 13th International Conference AFL 2011, in: College of Nyíregyháza, Debrecen, Hungary, 2011, pp. 123–137.
[5] J. Brzozowski, B. Li, Syntactic complexity of star-free languages, 2011. http://arxiv.org/abs/1109.3381.
[6] J. Brzozowski, Y. Ye, Syntactic complexity of ideal and closed languages, in: G. Mauri, A. Leporati (Eds.), 15th International Conference on Developments in Language Theory, DLT 2011, in: LNCS, vol. 6795, Springer, Berlin / Heidelberg, 2011, pp. 117–128.
[7] J. Dénes, On transformations, transformation semigroups and graphs, in: P. Erdös, G. Katona (Eds.), Theory of Graphs. Proceedings of the Colloquium on Graph Theory held at Tihany, 1966, Akadémiai Kiado, 1968, pp. 65–75.
[8] O. Ganyushkin, V. Mazorchuk, Classical Finite Transformation Semigroups: An Introduction, Springer, 2009.
[9] GAP-Group, GAP - Groups, Algorithms, Programming - a System for Computational Discrete Algebra, 2010. http://www.gap-system.org/.
[10] Y.S. Han, K. Salomaa, State complexity of basic operations on suffix-free regular languages, Theoret. Comput. Sci. 410 (2009) 2537–2548.
[11] Y.S. Han, K. Salomaa, D. Wood, Operational state complexity of prefix-free regular languages, in: Z. Ésik, Z. Fülöp (Eds.), Automata, Formal Languages, and Related Topics, Inst. of Informatics, University of Szeged, Hungary, 2009, pp. 99–115.
[12] M. Holzer, B. König, On deterministic finite automata and syntactic monoid size, Theoret. Comput. Sci. 327 (2004) 319–347.
[13] S. Janson, V. Mazorchuk, Some remarks on the combinatorics of $\mathcal{IS}_n$, Semigroup Forum 70 (2005) 391–405.

[14] B. Krawetz, J. Lawrence, J. Shallit, State complexity and the monoid of transformations of a finite set, 2003. http://arxiv.org/abs/math/0306416v1.
[15] A.N. Maslov, Estimates of the number of states of finite automata, Dokl. Akad. Nauk SSSR 194 (1970) 1266–1268. (Russian).English translation: Soviet Math. Dokl. 11 (1970), pp. 1373–1375.
[16] R. McNaughton, S.A. Papert, Counter-Free Automata, in: M.I.T. Research Monographs, vol. 65, The MIT Press, 1971.
[17] J. Myhill, Finite automata and the representation of events, Wright Air Dev. Center Tech. Rep. (1957) 57–624.
[18] A. Nerode, Linear automaton transformations, Proc. Amer. Math. Soc. 9 (1958) 541–544.
[19] S. Piccard, Sur les fonctions définies dans les ensembles finis quelconques, Fund. Math. 24 (1935) 298–301.
[20] J.E. Pin, Syntactic semigroups, in: Handbook of Formal Languages, in: Word, Language, Grammar, vol. 1, Springer-Verlag, New York, Inc., New York, NY, USA, 1997, pp. 679–746.
[21] J. Riordan, The enumeration of trees by height and diameter, IBM J. Res. Dev. 4 (1960) 473–478.
[22] J. Riordan, Forests of labeled trees, J. Combin. Theory 5 (1968) 90–103.
[23] H.J. Shyr, Free Monoids and Languages, third edition, Hon Min Book Co., Taiwan, 2001.
[24] W. Sierpiński, Sur les suites infinies de fonctions définies dans les ensembles quelconques, Fund. Math. 24 (1935) 209–212.
[25] S. Yu, State complexity of regular languages, J. Autom. Lang. Comb. 6 (2001) 221–234.