


# Syntactic Complexity of Bifix-Free Languages

Marek Szykuła<sup>1</sup> and John Wittnebel<sup>2</sup>

<sup>1</sup> Institute of Computer Science, University of Wrocław,  
Joliot-Curie 15, 50-383 Wrocław, Poland  
`msz@cs.uni.wroc.pl`

<sup>2</sup> David R. Cheriton School of Computer Science,  
University of Waterloo, Waterloo, ON N2L 3G1, Canada  
`jkwittnebel@hotmail.com`

**Abstract.** We study the properties of syntactic monoids of bifix-free regular languages. In particular, we solve an open problem concerning syntactic complexity: We prove that the cardinality of the syntactic semigroup of a bifix-free language with state complexity  $n$  is at most  $(n-1)^{n-3} + (n-2)^{n-3} + (n-3)2^{n-3}$  for  $n \geq 6$ . The main proof uses a large construction with the method of injective function. Since this bound is known to be reachable, and the values for  $n \leq 5$  are known, this completely settles the problem. We also prove that  $(n-2)^{n-3} + (n-3)2^{n-3} - 1$  is the minimal size of the alphabet required to meet the bound for  $n \geq 6$ . Finally, we show that the largest transition semigroups of minimal DFAs which recognize bifix-free languages are unique up to renaming the states.

## 1 Introduction

The *syntactic complexity* [11]  $\sigma(L)$  of a regular language  $L$  is defined as the size of its syntactic semigroup [17]. It is known that this semigroup is isomorphic to the transition semigroup of the quotient automaton  $\mathcal{D}$  and of a minimal deterministic finite automaton accepting the language. The number  $n$  of states of  $\mathcal{D}$  is the *state complexity* of the language [19], and it is the same as the *quotient complexity* [2] (number of left quotients) of the language. The *syntactic complexity of a class* of regular languages is the maximal syntactic complexity of languages in that class expressed as a function of the quotient complexity  $n$ .

Syntactic complexity is related to the Myhill equivalence relation [16], and it counts the number of classes of non-empty words in a regular language which act distinctly. It provides a natural bound on the time and space complexity of algorithms working on the transition semigroup. For example, a simple algorithm checking whether a language is *star-free* just enumerates all transformations and verifies whether none of them contains a non-trivial cycle [15].

Syntactic complexity does not refine state complexity, but used as an additional measure it can distinguish particular subclasses of regular languages from the class of all regular languages, whereas state complexity alone cannot.

---

M. Szykuła—Supported in part by the National Science Centre, Poland under project number 2014/15/B/ST6/00615.

For example, the state complexity of basic operations in the class of star-free languages is the same as in the class of all regular languages (except the reversal, where the tight upper bound is  $2^{n-1} - 1$  see [8]).

Finally, the largest transition semigroups play an important role in the study of *most complex* languages [3] in a given subclass. These are languages that meet all the upper bounds on the state complexities of Boolean operations, product, star, and reversal, and also have maximal syntactic semigroups and most complex atoms [10]. In particular, the results from this paper enabled the study of most complex bifix-free languages [12].

A language is *prefix-free* if no word in the language is a proper prefix of another word in the language. Similarly, a language is *suffix-free* if there is no word that is a proper suffix of another word in the language. A language is *bifix-free* if it is both prefix-free and suffix-free. Prefix-, suffix-, and bifix-free languages are important classes of codes, which have numerous applications in such fields as cryptography and data compression. Codes have been studied extensively; see [1] for example.

Syntactic complexity has been studied for a number of subclasses of regular languages (e.g., [4–6, 8, 13, 14]). For bifix-free languages, the lower bound  $(n - 1)^{n-3} + (n - 2)^{n-3} + (n - 3)2^{n-3}$  for the syntactic complexity for  $n \geq 6$  was established in [6]. The values for  $n \leq 5$  were also determined.

The problem of establishing tight upper bound on syntactic complexity can be quite challenging, depending on the particular subclass. For example, it is easy for prefix-free languages and right ideals, while much more difficult for suffix-free languages and left ideals. The case of bifix-free languages studied in this paper requires an even more involved proof, as the structure of maximal transition semigroup is more complicated.

Our main contributions are as follows:

1. We prove that  $(n - 1)^{n-3} + (n - 2)^{n-3} + (n - 3)2^{n-3}$  is also an upper bound for syntactic complexity for  $n \geq 8$ . To do this, we apply the general method of injective function (cf. [7, 9]). The construction here is much more involved than in the previous cases, and uses a number of tricks for ensuring injectivity.
2. We prove that the transition semigroup meeting this bound is unique for every  $n \geq 8$ .
3. We refine the witness DFA meeting the bound by reducing the size of the alphabet to  $(n - 2)^{n-3} + (n - 3)2^{n-3} - 1$ , and we show that it cannot be any smaller.
4. Using a dedicated algorithm, we verify by computation that two semigroups  $\mathbf{W}_{\text{bf}}^{\leq 5}$  and  $\mathbf{W}_{\text{bf}}^{\geq 6}$  (defined below) are the unique largest transition semigroups of a minimal DFA of a bifix-free language, respectively for  $n = 5$  and  $n = 6, 7$  (whereas they coincide for  $n = 3, 4$ ).

In summary, for every  $n$  we have determined the syntactic complexity, the unique largest semigroups, and the minimal sizes of the alphabets required; this completely solves the problem for bifix-free languages.

The full version of this paper is available at [18].

## 2 Preliminaries

Let  $\Sigma$  be a non-empty finite alphabet, and let  $L \subseteq \Sigma^*$  be a language. If  $w \in \Sigma^*$  is a word,  $L.w$  denotes the *left quotient* or simply quotient of  $L$  by  $w$ , which is defined by  $L.w = \{u \mid wu \in L\}$ . The number of quotients of  $L$  is its *quotient complexity* [2]  $\kappa(L)$ . From the Myhill-Nerode Theorem, a language is regular if and only if the set of all quotients of the language is finite. We denote the set of quotients of regular  $L$  by  $K = \{K_0, \dots, K_{n-1}\}$ , where  $K_0 = L = L.\varepsilon$  by convention.

A *deterministic finite automaton (DFA)* is a tuple  $\mathcal{D} = (Q, \Sigma, \delta, q_0, F)$ , where  $Q$  is a finite non-empty set of *states*,  $\Sigma$  is a finite non-empty *alphabet*,  $\delta: Q \times \Sigma \rightarrow Q$  is the *transition function*,  $q_0 \in Q$  is the *initial state*, and  $F \subseteq Q$  is the set of *final states*. We extend  $\delta$  to a function  $\delta: Q \times \Sigma^* \rightarrow Q$  as usual.

The *quotient DFA* of a regular language  $L$  with  $n$  quotients is defined by  $\mathcal{D} = (K, \Sigma, \delta_{\mathcal{D}}, K_0, F_{\mathcal{D}})$ , where  $\delta_{\mathcal{D}}(K_i, w) = K_j$  if and only if  $K_i.w = K_j$ , and  $F_{\mathcal{D}} = \{K_i \mid \varepsilon \in K_i\}$ . Without loss of generality, we assume that  $Q = \{0, \dots, n-1\}$ . Then  $\mathcal{D} = (Q, \Sigma, \delta, 0, F)$ , where  $\delta(i, w) = j$  if  $\delta_{\mathcal{D}}(K_i, w) = K_j$ , and  $F$  is the set of subscripts of quotients in  $F_{\mathcal{D}}$ . A state  $q \in Q$  is *empty* if its quotient  $K_q$  is empty. The quotient DFA of  $L$  is isomorphic to each complete minimal DFA of  $L$ . The number of states in the quotient DFA of  $L$  (the quotient complexity of  $L$ ) is therefore equal to the state complexity of  $L$ .

In any DFA  $\mathcal{D}$ , each letter  $a \in \Sigma$  induces a transformation on the set  $Q$  of  $n$  states. We let  $\mathcal{T}_n$  denote the set of all  $n^n$  transformations of  $Q$ ; then  $\mathcal{T}_n$  is a monoid under composition. The *image* of  $q \in Q$  under transformation  $t$  is denoted by  $qt$ , and the *image* of a subset  $S \subseteq Q$  is  $St = \{qt \mid q \in S\}$ . If  $s, t \in \mathcal{T}_n$  are transformations, their composition is denoted by  $st$  and defined by  $q(st) = (qs)t$ . The identity transformation is denoted by  $\mathbf{1}$ , and we have  $q\mathbf{1} = q$  for all  $q \in Q$ . By  $(S \rightarrow q)$ , where  $S \subseteq Q$  and  $q \in Q$ , we denote a *semiconstant* transformation that maps all the states from  $S$  to  $q$  and behaves as the identity function for the states in  $Q \setminus S$ . A *constant* transformation is the semiconstant transformation  $(Q \rightarrow q)$ , where  $q \in Q$ . A *unitary* transformation is  $(\{p\} \rightarrow q)$ , for some distinct  $p, q \in Q$ ; this is denoted by  $(p \rightarrow q)$  for simplicity.

The *transition semigroup* of  $\mathcal{D}$  is the semigroup of all transformations generated by the transformations induced by  $\Sigma$ . Since the transition semigroup of a minimal DFA of a language  $L$  is isomorphic to the syntactic semigroup of  $L$  [17], the syntactic complexity of  $L$  is equal to the cardinality of the transition semigroup of  $\mathcal{D}$ .

The *underlying digraph* of a transformation  $t \in \mathcal{T}_n$  is the digraph  $(Q, E)$ , where  $E = \{(q, qt) \mid q \in Q\}$ . We identify a transformation with its underlying digraph and use usual graph terminology for transformations: The *in-degree* of a state  $q \in Q$  is the cardinality  $|\{p \in Q \mid pt = q\}|$ . A *cycle* in  $t$  is a cycle in its underlying digraph of length at least 2. A *fixed point* in  $t$  is a self-loop in its underlying digraph. The *orbit* of a state  $q \in Q$  in  $t$  is a connected component containing  $q$  in its underlying digraph, that is, the set  $\{p \in Q \mid pt^i = qt^j \text{ for some } i, j \geq 0\}$ . Note that every orbit contains either exactly one cycle or one fixed point. The *distance* in  $t$  from a state  $p \in Q$  to a state  $q \in Q$  is the length of the path in

the underlying digraph of  $t$  from  $p$  to  $q$ , that is,  $\min\{i \in \mathbb{N} \mid pt^i = q\}$ , and is undefined if no such path exists. If a state  $q$  does not lie in a cycle, then the *tree* of  $q$  is the underlying digraph of  $t$  restricted to the states  $p$  such that there is a path from  $p$  to  $q$ .

### 2.1 Bifix-Free Languages and Semigroups

Let  $\mathcal{D}_n = (Q, \Sigma, \delta, 0, F)$ , where  $Q = \{0, \dots, n - 1\}$ , be a minimal DFA accepting a bifix-free language  $L$ , and let  $T(\mathcal{D}_n)$  be its transition semigroup. We also define  $Q_M = \{1, \dots, n - 3\}$  (the set of the “middle” non-special states).

The following properties of bifix-free languages, slightly adapted to our terminology, are well known [6]:

**Lemma 1.** *A minimal DFA  $\mathcal{D}_n = (Q, \Sigma, \delta, 0, F)$  of a bifix-free languages  $L$  satisfies the following properties:*

1. *There is an empty state, which is  $n - 1$  by convention.*
2. *There exists exactly one final quotient, which is  $\{\varepsilon\}$ , and whose state is  $n - 2$  by convention, so  $F = \{n - 2\}$ .*
3. *For  $u, v \in \Sigma^+$ , if  $L.v \neq \emptyset$ , then  $L.v \neq L.uv$ .*
4. *In the underlying digraph of every transformation of  $T(\mathcal{D}_n)$ , there is a path starting at 0 and ending at  $n - 1$ .*

The items (1) and (2) are sufficient and necessary for prefix-free languages, while (3) and (4) follow from the properties of suffix-free languages. Following [9], we say that an (unordered) pair  $\{p, q\}$  of distinct states in  $Q_M$  is *colliding* (or  $p$  *collides* with  $q$ ) in  $T(\mathcal{D}_n)$  if there is a transformation  $t \in T(\mathcal{D}_n)$  such that  $0t = p$  and  $rt = q$  for some  $r \in Q_M$ . A pair of states is *focused by* a transformation  $u \in T(n)$  if  $u$  maps both states of the pair to a single state  $r \in Q_M \cup \{n - 2\}$ . We then say that  $\{p, q\}$  is *focused to the state*  $r$ . By Lemma 1(3), it follows that if  $\{p, q\}$  is colliding in  $T(\mathcal{D}_n)$ , then there is no transformation  $u \in T(\mathcal{D}_n)$  that focuses  $\{p, q\}$ . Hence, in the case of bifix-free languages, colliding states can be mapped to a single state only if the state is  $n - 1$ . In contrast with suffix-free languages, we do not consider the pairs from  $Q_M \times \{n - 2\}$  being colliding, as they cannot be focused.

For  $n \geq 2$  we define the set of transformations

$$\mathbf{B}_{\text{bf}}(n) = \{t \in \mathcal{T}_n \mid 0 \notin Qt, (n - 1)t = n - 1, (n - 2)t = n - 1, \text{ and for all } j \geq 1, \\ 0t^j = n - 1 \text{ or } 0t^j \neq qt^j \ \forall q, 0 < q < n - 1\}.$$

In [6] it was shown that the transition semigroup  $T(\mathcal{D}_n)$  of a minimal DFA of a bifix-free language must be contained in  $\mathbf{B}_{\text{bf}}(n)$ . It contains all transformations  $t$  which fix  $n - 1$ , map  $n - 2$  to  $n - 1$ , and do not focus any pair which is colliding from  $t$ .

Since  $\mathbf{B}_{\text{bf}}(n)$  is not a semigroup, no transition semigroup of a minimal DFA of a bifix-free language can contain all transformations from  $\mathbf{B}_{\text{bf}}(n)$ . Therefore, its cardinality is not a tight upper bound on the syntactic complexity of bifix-free languages. A lower bound on the syntactic complexity was established in [6].

We study the following two semigroups that play an important role for bifix-free languages.

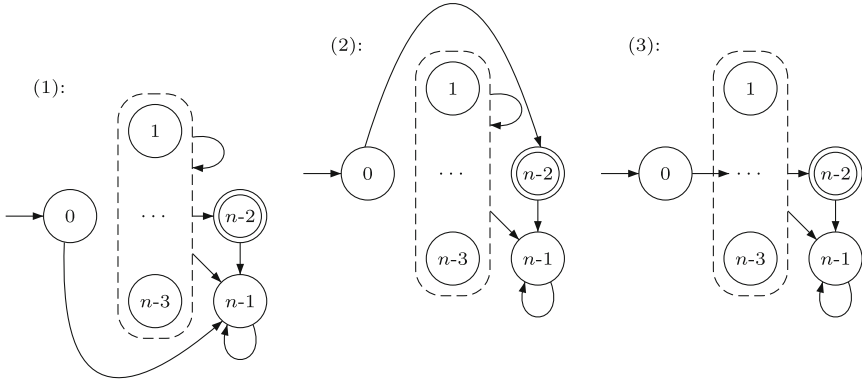
**Semigroup  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$ .** For  $n \geq 3$  we define the semigroup:

$$\mathbf{W}_{\text{bf}}^{\geq 6}(n) = \{t \in \mathbf{B}_{\text{bf}}(n) \mid 0t \in \{n-2, n-1\}, \text{ or } 0t \in Q_M \text{ and } qt \in \{n-2, n-1\} \text{ for all } q \in Q_M\}.$$

The following remark summarizes the transformations of  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  (illustrated in Fig. 1):

*Remark 2.*  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  contains all transformations that:

1. map  $\{0, n-2, n-1\}$  to  $n-1$ , and  $Q_M$  into  $Q \setminus \{0\}$ , or
2. map  $0$  to  $n-2$ ,  $\{n-2, n-1\}$  to  $n-1$ , and  $Q_M$  into  $Q \setminus \{0, n-2\}$ , or
3. map  $0$  to a state  $q \in Q_M$ , and  $Q_M$  into  $\{n-2, n-1\}$ . ■



**Fig. 1.** The three types of transformations in  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  from Remark 2.

The cardinality of  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  is  $(n-1)^{n-3} + (n-2)^{n-3} + (n-3)2^{n-3}$ .

**Proposition 3.**  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  is the unique maximal transition semigroup of a minimal DFA  $\mathcal{D}_n$  of a bifix-free language in which there are no colliding pairs of states.

In [6] it was shown that for  $n \geq 5$ , there exists a witness DFA of a bifix-free language whose transition semigroup is  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  over an alphabet of size  $(n-2)^{n-3} + (n-3)2^{n-3} + 2$  (and 18 if  $n = 5$ ). Now we slightly refine the witness from [6, Proposition 31] by reducing the size of the alphabet to  $(n-2)^{n-3} + (n-3)2^{n-3} - 1$ , and then we show that it cannot be any smaller.

**Definition 4 (Bifix-free witness).** For  $n \geq 4$ , let  $\mathcal{W}(n) = (Q, \Sigma, \delta, 0, \{n-2\})$ , where  $Q = \{0, \dots, n-1\}$  and  $\Sigma$  contains the following letters:

1.  $b_i$ , for  $1 \leq i \leq n - 3$ , inducing the transformations  $(0 \rightarrow n - 1)(i \rightarrow n - 2)(n - 2 \rightarrow n - 1)$ ,
2.  $c_i$ , for every transformation of type (2) from Remark 2 that is different from  $(0 \rightarrow n - 2)(Q_M \rightarrow n - 1)(n - 2 \rightarrow n - 1)$ ,
3.  $d_i$ , for every transformation of type (3) from Remark 2 that is different from  $(0 \rightarrow q)(Q_M \rightarrow n - 1)(n - 2 \rightarrow n - 1)$  for some state  $q \in Q_M$ .

Altogether, we have  $|\Sigma| = (n - 3) + ((n - 2)^{n-3} - 1) + (n - 3)(2^{n-3} - 1) = (n - 2)^{n-3} + (n - 3)2^{n-3} - 1$ . For  $n = 4$  three letters suffice, since the transformation of  $b_1$  is induced by  $c_i d_i$ , where  $c_i: (0 \rightarrow 2)(2 \rightarrow 3)$  and  $d_i: (0 \rightarrow 1)(1 \rightarrow 2)(2 \rightarrow 3)$ .

**Proposition 5.** *The transition semigroup of  $\mathcal{W}(n)$  is  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$ .*

**Proposition 6.** *For  $n \geq 5$ , at least  $(n - 2)^{n-3} + (n - 3)2^{n-3} - 1$  generators are necessary to generate  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$ .*

**Semigroup  $\mathbf{W}_{\text{bf}}^{\leq 5}(n)$ .** For  $n \geq 3$  we define the semigroup

$$\mathbf{W}_{\text{bf}}^{\leq 5}(n) = \{t \in \mathbf{B}_{\text{bf}}(n) \mid \text{for all } p, q \in Q_M \text{ where } p \neq q, pt = qt = n - 1 \text{ or } pt \neq qt\}.$$

**Proposition 7.**  *$\mathbf{W}_{\text{bf}}^{\leq 5}(n)$  is the unique maximal transition semigroup of a minimal DFA  $\mathcal{D}_n$  of a bifix-free language in which all pairs of states from  $Q_M$  are colliding.*

In [6] it was shown that for  $n \geq 2$  there exists a DFA for a bifix-free language whose transition semigroup is  $\mathbf{W}_{\text{bf}}^{\leq 5}(n)$  over an alphabet of size  $(n - 2)!$ . We prove that this is an alphabet of minimal size that generates this transition semigroup.

**Proposition 8.** *To generate  $\mathbf{W}_{\text{bf}}^{\leq 5}(n)$  at least  $(n - 2)!$  generators must be used.*

### 3 Upper Bound on Syntactic Complexity

Our main result shows that the lower bound  $(n - 1)^{n-3} + (n - 2)^{n-3} + (n - 3)2^{n-3}$  on the syntactic complexity of bifix-free languages is also an upper bound for  $n \geq 8$ .

We consider a minimal DFA  $\mathcal{D}_n = (Q, \Sigma, \delta, 0, \{n - 2\})$ , where  $Q = \{0, \dots, n - 1\}$  and whose empty state is  $n - 1$ , of an arbitrary bifix-free language. Let  $T(\mathcal{D}_n)$  be the transition semigroup of  $\mathcal{D}_n$ . We will show that  $T(\mathcal{D}_n)$  is not larger than  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$ .

Note that the semigroups  $T(\mathcal{D}_n)$  and  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  share the set  $Q$ , and in both of them  $0, n - 2$ , and  $n - 1$  play the role of the initial, final, and empty state, respectively. When we say that a pair of states from  $Q$  is *colliding* we always mean that it is colliding in  $T(\mathcal{D}_n)$ .

First, we state the following lemma, which generalizes some arguments that we use frequently in the proof of the main theorem.

**Lemma 9.** Let  $t, \hat{t} \in T(\mathcal{D}_n)$  and  $s \in \mathbf{W}_{\text{bf}}^{\geq 6}(n)$  be transformations. Suppose that:

1. All states from  $Q_M$  whose mapping is different in  $t$  and  $s$  belong to  $C$ , where  $C$  is either an orbit in  $s$  or is the tree of a state in  $s$ .
2. All states from  $Q_M$  whose mapping is different in  $\hat{t}$  and  $s$  belong to  $\hat{C}$ , where  $\hat{C}$  is either an orbit in  $s$  or is the tree of a state in  $s$ .
3. The transformation  $s^i t^j$ , for some  $i, j \geq 0$ , focuses a colliding pair whose states are in  $C$ .

Then either  $C \subseteq \hat{C}$  or  $\hat{C} \subseteq C$ . In particular, if  $C$  and  $\hat{C}$  are both orbits or both trees rooted in a state mapped by  $s$  to  $n - 1$ , then  $C = \hat{C}$ .

The following is our main theorem:

**Theorem 10.** For  $n \geq 8$ , the syntactic complexity of the class of bifix-free languages with  $n$  quotients is  $(n - 1)^{n-3} + (n - 2)^{n-3} + (n - 3)2^{n-3}$ .

*Proof (Idea).* We construct an injective mapping  $\varphi: T(\mathcal{D}_n) \rightarrow \mathbf{W}_{\text{bf}}^{\geq 6}(n)$ . Since  $\varphi$  will be injective, this will prove that  $|T(\mathcal{D}_n)| \leq |\mathbf{W}_{\text{bf}}^{\geq 6}(n)| = (n - 1)^{n-3} + (n - 2)^{n-3} + (n - 3)2^{n-3}$ .

The mapping  $\varphi$  is defined by 23 (sub)cases covering all possibilities for a transformation  $t \in T(\mathcal{D}_n)$ . Let  $t$  denote a transformation of  $T(\mathcal{D}_n)$ , and  $s$  denote the assigned transformation  $\varphi(t)$ .

The whole proof is split into three Supercases, depending on  $t$ . Supercase 2 and Supercase 3 are split into a number of cases, and the cases are split into subcases. To show injectivity, in every (sub)case we prove *external injectivity*, which is that there is no other transformation  $\hat{t}$  that fits to one of the previous (sub)cases and results in the same  $s$ , and we prove *internal injectivity*, which is that no other transformation  $\hat{t}$  that fits to the same (sub)case results in the same  $s$ . We use there various kinds of arguments of analysis orbits, cycles, longest paths, and focused states. Often, we use Lemma 9 to argue that if another  $\hat{t}$  yields the same  $s$  (so  $\varphi$  is not injective) and  $s$  is obtained by a local modification of  $t$  or  $\hat{t}$ , then the difference between  $t$  and  $\hat{t}$  is also only local – restricted to the same orbit or tree. All states and variables related to  $\hat{t}$  are always marked by a hat.

**Supercase 1:**  $t \in \mathbf{W}_{\text{bf}}^{\geq 6}(n)$ .

We take  $s = t$ . The internal and external injectivity are obvious. ◁  
 For all the remaining cases let  $p = 0t$ . Note that all  $t$  with  $p \in \{n - 2, n - 1\}$  fit in Supercase 1. Let  $k \geq 0$  be a maximal integer such that  $pt^k \notin \{n - 2, n - 1\}$ . Then  $pt^{k+1}$  is either  $n - 1$  or  $n - 2$ , and we have two supercases covering these situations.

**Supercase 2:**  $t \notin \mathbf{W}_{\text{bf}}^{\geq 6}(n)$  and  $pt^{k+1} = n - 1$ .

Here we have the chain

$$0 \xrightarrow{t} p \xrightarrow{t} pt \xrightarrow{t} \dots \xrightarrow{t} pt^k \xrightarrow{t} n - 1.$$

Within this supercase, we always assign transformations  $s$  focusing a colliding pair, and this will make them different from the transformations of Supercase 1.

Also, we use only transformations  $s$  of type 1 from Remark 2, that is, we will always have  $0s = n - 1$ .

As an example, we show the full proof of the first case:

**Case 2.1:**  $t$  has a cycle.

Let  $r$  be the minimal state among the states that appear in cycles of  $t$ , that is,

$$r = \min\{q \in Q \mid q \text{ is in a cycle of } t\}.$$

Let  $s$  be the transformation illustrated in Fig. 2 and defined by:

$$\begin{aligned} 0s &= n - 1, ps = r, \\ (pt^i)s &= pt^{i-1} \text{ for } 1 \leq i \leq k, \\ qs &= qt \text{ for the other states } q \in Q. \end{aligned}$$

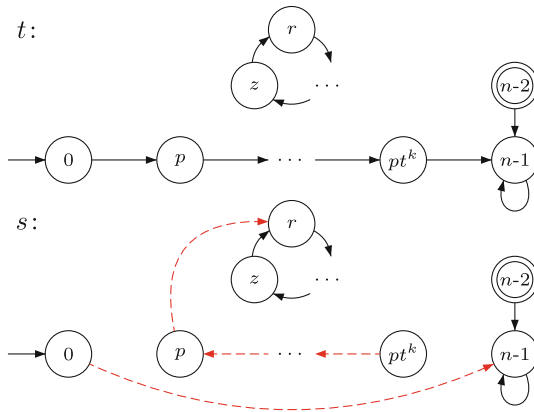


Fig. 2. Case 2.1.

Let  $z$  be the state from the cycle of  $t$  such that  $zt = r$ . We observe the following properties:

- (a) Pair  $\{p, z\}$  is a colliding pair focused by  $s$  to state  $r$  in the cycle, which is the smallest state of all states in cycles. This is the only colliding pair which is focused to a state in a cycle.  
*Proof:* Note that  $p$  collides with any state in a cycle of  $t$ , in particular, with  $z$ . The property follows because  $s$  differs from  $t$  only in the mapping of states  $pt^i$  ( $0 \leq i \leq k$ ) and  $0$ , and the only state mapped to a cycle is  $p$ .  $\triangleleft$
- (b) All states from  $Q_M$  whose mapping is different in  $t$  and  $s$  belong to the same orbit in  $s$  of a cycle. Hence, all colliding pairs that are focused by  $s$  consist only of states from this orbit.
- (c)  $s$  has a cycle.



(d) For each  $i$  with  $1 \leq i < k$ , there is precisely one state  $q$  colliding with  $pt^{i-1}$  and mapped by  $s$  to  $pt^i$ , and that state is  $q = pt^{i+1}$ .

*Proof:* Clearly  $q = pt^{i+1}$  satisfies this condition. Suppose that  $q \neq pt^{i+1}$ . Since  $pt^{i+1}$  is the only state mapped to  $pt^i$  by  $s$  and not by  $t$ , it follows that  $qt = qs = pt^i$ . So  $q$  and  $pt^{i-1}$  are focused to  $pt^i$  by  $t$ ; since they collide, this is a contradiction.  $\triangleleft$

*External injectivity:* By (a),  $\{p, z\}$  is a colliding pair focused by  $s$ , therefore  $t$  and  $s$  cannot be both present in  $T_n$  and so  $s$  was not used in Supercase 1.  $\triangleleft$

*Internal injectivity:* Let  $\hat{t}$  be any transformation that fits in this case and results in the same  $s$ ; we will show that  $\hat{t} = t$ . From (a), there is the unique colliding pair  $\{p, z\}$  focused to a state in a cycle, hence  $\{\hat{p}, \hat{z}\} = \{p, z\}$ . Moreover,  $p$  and  $\hat{p}$  are not in this cycle, so  $\hat{p} = p$  and  $\hat{z} = z$ , which means that  $0\hat{t} = 0t = p$ . Since there is no state  $q \neq 0$  such that  $qt = p$ , the only state mapped to  $p$  by  $s$  is  $pt$ , hence  $p\hat{t} = pt$ . From (d) for  $i = 1, \dots, k - 1$ , state  $pt^{i+1}$  is uniquely determined, hence  $p\hat{t}^{i+1} = pt^{i+1}$ . Finally, for  $i = k$  there is no state colliding with  $pt^{k-1}$  and mapped to  $pt^k$ , hence  $p\hat{t}^{k+1} = pt^{k+1} = n - 1$ . Since the other transitions in  $s$  are defined exactly as in  $t$  and  $\hat{t}$ , we have  $\hat{t} = t$ .  $\triangleleft$

Then we have four other cases, which together cover all possibilities for  $t$ .

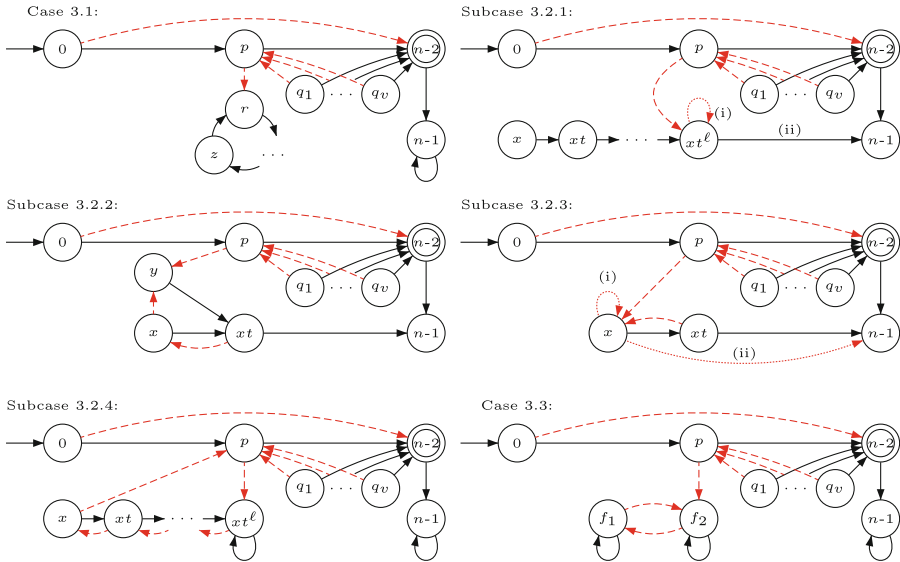


Fig. 3. Map of the (sub)cases of Supercase 3 in the proof of Theorem 10 (part 1).

**Supercase 3:**  $t \notin \mathbf{W}_{\text{bf}}^{\geq 6}(n)$  and  $pt^{k+1} = n - 2$ .

Here we have the chain

$$0 \xrightarrow{t} p \xrightarrow{t} pt \xrightarrow{t} \dots \xrightarrow{t} pt^k \xrightarrow{t} n - 2 \xrightarrow{t} n - 1.$$

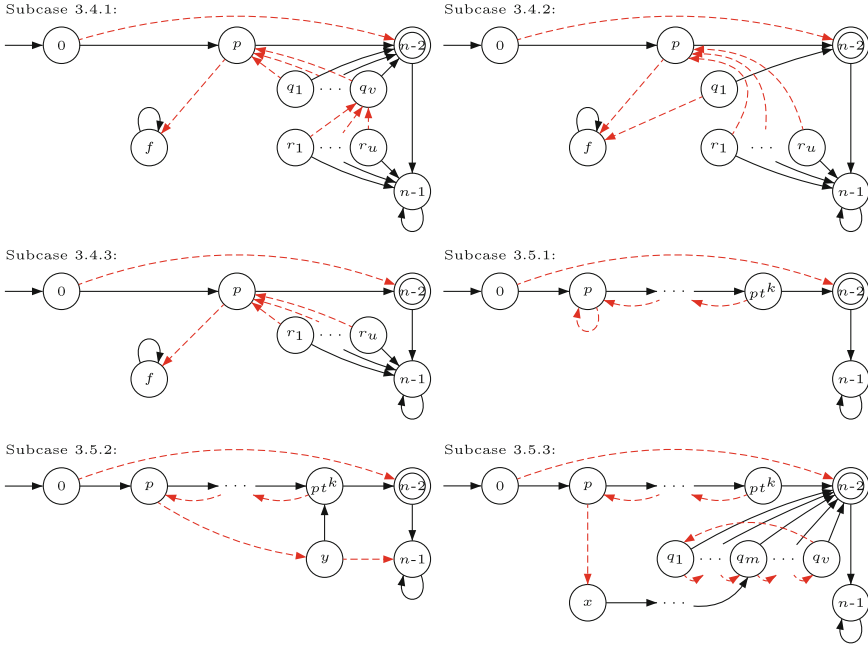


Fig. 4. Map of the (sub)cases of Supercase 3 in the proof of Theorem 10 (part 2).

We always assign transformations  $s$  such that  $s$  together with  $t$  generate a transformation that focuses a colliding pair, which distinguishes such transformations  $s$  from those of Supercase 1. Moreover, we always assign transformations of type 2 from Remark 2, that is, we always have  $0s = n - 2$ . This distinguishes  $s$  from all the transformations used in Supercase 2.

To show briefly how the construction looks like, in Figs. 3 and 4 we present a map of the (sub)cases for Supercase 3. The black solid edges are the edges of  $t$ , and the dashed edges (also red in a color printout) are the edges of the corresponding  $s$ . ◁

### 4 Uniqueness of Maximal Semigroups

Here we show that  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  for  $n \geq 6$  and  $\mathbf{W}_{\text{bf}}^{\leq 5}(n)$  for  $n \in \{3, 4, 5\}$  (whereas  $\mathbf{W}_{\text{bf}}^{\geq 6}(n) = \mathbf{W}_{\text{bf}}^{\leq 5}(n)$  for  $n \in \{3, 4\}$ ) have not only the maximal sizes, but are also the unique largest semigroups up to renaming the states in a minimal DFA  $\mathcal{D}_n = (Q, \Sigma, \delta, 0, \{n - 2\})$  of a bifix-free language.

**Theorem 11.** *If  $n \geq 8$ , and the transition semigroup  $T(\mathcal{D}_n)$  of a minimal DFA  $\mathcal{D}_n$  of a bifix-free language has at least one colliding pair, then*

$$|T(\mathcal{D}_n)| < |\mathbf{W}_{\text{bf}}^{\geq 6}(n)| = (n - 1)^{n-3} + (n - 2)^{n-3} + (n - 3)2^{n-3}.$$

*Proof (Idea).* This is done by finding one more  $s$  (under the assumption that there exists a colliding pair) that was not assigned by  $\varphi$  in the proof of Theorem 10. Thus, since  $\varphi$  is injective and  $\varphi(T(\mathcal{D}_n)) \subseteq \mathbf{W}_{\text{bf}}^{\geq 6}(n)$ ,  $s \in \mathbf{W}_{\text{bf}}^{\geq 6}(n)$  but  $s \notin \varphi(T(\mathcal{D}_n))$ , it follows that  $\varphi(T(\mathcal{D}_n)) \subsetneq \mathbf{W}_{\text{bf}}^{\geq 6}(n)$ , so  $|T(\mathcal{D}_n)| < |\mathbf{W}_{\text{bf}}^{\geq 6}(n)|$ .  $\square$

**Corollary 12.** *For  $n \geq 8$ , the transition semigroup  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  is the unique largest transition semigroup of a minimal DFA of a bifix-free language.*

*Proof.* From Theorem 11, a transition semigroup that has a colliding pair cannot be largest. From Proposition 3,  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  is the unique maximal transition semigroup that does not have colliding pairs of states.  $\square$

The following theorem solves the remaining cases of small semigroups:

**Theorem 13.** *For  $n \in \{6, 7\}$ , the largest transition semigroup of minimal DFAs of bifix-free languages is  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  and it is unique. For  $n = 5$ , the largest transition semigroup of minimal DFAs of bifix-free languages is  $\mathbf{W}_{\text{bf}}^{\leq 5}(n)$  and it is unique. For  $n \in \{3, 4\}$ ,  $\mathbf{W}_{\text{bf}}^{\geq 6}(n) = \mathbf{W}_{\text{bf}}^{\leq 5}(n)$  is the unique largest transition semigroup of minimal DFAs of bifix-free languages.*

*Proof (Idea).* We have verified this with the help of computation, basing on the idea of conflicting pairs of transformations from [6, Theorem 20]. We have developed an algorithm which verified for a given  $n \leq 7$  that no transformation from  $\mathbf{B}_{\text{bf}}(n)$  can belong to a transition semigroup of a minimal DFA  $\mathcal{D}$  of a bifix-free language of size at least  $\max\{\mathbf{W}_{\text{bf}}^{\leq 5}(n), \mathbf{W}_{\text{bf}}^{\geq 6}(n)\}$  that is different from  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  and  $\mathbf{W}_{\text{bf}}^{\leq 5}(n)$ .  $\square$

Since the largest transition semigroups are unique, from Propositions 6 and 8 we infer the sizes of the alphabets required in order to meet the bound for the syntactic complexity.

**Corollary 14.** *To meet the bound for the syntactic complexity of bifix-free languages,  $(n - 2)^{n-3} + (n - 3)2^{n-3} - 1$  letters are required and sufficient for  $n \geq 6$ , and  $(n - 2)!$  letters are required and sufficient for  $n \in \{3, 4, 5\}$ .*

## 5 Conclusions

We have solved the problem of syntactic complexity of bifix-free languages and identified the largest semigroups for every number of states  $n$ . In the main theorem, we used the method of injective function (cf. [7, 9]) with new techniques and tricks for ensuring injectivity (in particular, Lemma 9 and the constructions in Supercase 3). This stands as a universal method for solving similar problems concerning maximality of semigroups. Our proof required an extensive analysis of 23 (sub)cases and much more complicated injectivity arguments than those for suffix-free (12 cases), left ideals (5 subcases) and two-sided ideals (8 subcases). The difficulty of applying the method grows quickly when characterization of the class of languages gets more involved.

It may be surprising that we need a witness with  $(n-2)^{n-3} + (n-3)2^{n-3} - 1$  (for  $n \geq 6$ ) letters to meet the bound for syntactic complexity of bifix-free languages, whereas in the case of prefix- and suffix-free languages only  $n+1$  and five letters suffice, respectively (see [6, 9]).

Finally, our results enabled establishing existence of most complex bifix-free languages ([12]).

## References

1. Berstel, J., Perrin, D., Reutenauer, C.: Codes and Automata. Cambridge University Press, Cambridge (2009)
2. Brzozowski, J.A.: Quotient complexity of regular languages. *J. Autom. Lang. Comb.* **15**(1/2), 71–89 (2010)
3. Brzozowski, J.A.: In search of the most complex regular languages. *Int. J. Found. Comput. Sci.* **24**(6), 691–708 (2013)
4. Brzozowski, J.A., Li, B.: Syntactic complexity of  $R$ - and  $J$ -trivial languages. *Int. J. Found. Comput. Sci.* **16**(3), 547–563 (2005)
5. Brzozowski, J.A., Li, B., Liu, D.: Syntactic complexities of six classes of star-free languages. *J. Autom. Lang. Comb.* **17**, 83–105 (2012)
6. Brzozowski, J.A., Li, B., Ye, Y.: Syntactic complexity of prefix-suffix-, bifix-, and factor-free regular languages. *Theoret. Comput. Sci.* **449**, 37–53 (2012)
7. Brzozowski, J., Szykuła, M.: Upper bounds on syntactic complexity of left and two-sided ideals. In: Shur, A.M., Volkov, M.V. (eds.) DLT 2014. LNCS, vol. 8633, pp. 13–24. Springer, Cham (2014). doi:[10.1007/978-3-319-09698-8\\_2](https://doi.org/10.1007/978-3-319-09698-8_2)
8. Brzozowski, J.A., Szykuła, M.: Large aperiodic semigroups. *Int. J. Found. Comput. Sci.* **26**(07), 913–931 (2015)
9. Brzozowski, J., Szykuła, M.: Upper bound on syntactic complexity of suffix-free languages. In: Shallit, J., Okhotin, A. (eds.) DCFS 2015. LNCS, vol. 9118, pp. 33–45. Springer, Cham (2015). doi:[10.1007/978-3-319-19225-3\\_3](https://doi.org/10.1007/978-3-319-19225-3_3)
10. Brzozowski, J.A., Tamm, H.: Theory of automata. *Theoret. Comput. Sci.* **539**, 13–27 (2014)
11. Brzozowski, J., Ye, Y.: Syntactic complexity of ideal and closed languages. In: Mauri, G., Leporati, A. (eds.) DLT 2011. LNCS, vol. 6795, pp. 117–128. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-22321-1\\_11](https://doi.org/10.1007/978-3-642-22321-1_11)
12. Ferens, R., Szykuła, M.: Complexity of bifix-free languages. In: Carayol, A., Nicaud, C. (eds.) CIAA 2017. LNCS, vol. 10329, pp. 76–88. Springer, Cham (2017)
13. Holzer, M., König, B.: On deterministic finite automata and syntactic monoid size. *Theoret. Comput. Sci.* **327**, 319–347 (2004)
14. Iván, S., Nagy-György, J.: On nonpermutational transformation semigroups with an application to syntactic complexity (2014). <http://arxiv.org/abs/1402.7289>
15. McNaughton, R., Papert, S.A.: Counter-free automata (M.I.T. Research Monograph No. 65). The MIT Press (1971)
16. Myhill, J.: Finite automata and representation of events. Wright Air Development Center Technical report, pp. 57–624 (1957)
17. Pin, J.E.: Syntactic semigroups. In: Rozenberg, G., Salomaa, A. (eds.) Handbook of Formal Languages, Volume 1 Word, Language, Grammar, pp. 679–746. Springer, Heidelberg (1997)
18. Szykuła, M., Wittnebel, J.: Syntactic complexity of bifix-free languages (2017). <http://arxiv.org/abs/1604.06936>
19. Yu, S.: State complexity of regular languages. *J. Autom. Lang. Comb.* **6**, 221–234 (2001)