Upper Bound on Syntactic Complexity of Suffix-Free Languages*

Janusz Brzozowski[1] and Marek Szykuła[2]

[1] David R. Cheriton School of Computer Science, University of Waterloo,
Waterloo, ON, Canada N2L 3G1
{brzozo@uwaterloo.ca}
[2] Institute of Computer Science, University of Wrocław,
Joliot-Curie 15, PL-50-383 Wrocław, Poland
{msz@cs.uni.wroc.pl}

Abstract. We solve an open problem concerning syntactic complexity: We prove that the cardinality of the syntactic semigroup of a suffix-free language with n left quotients (that is, with state complexity n) is at most $(n-1)^{n-2}+n-2$ for $n \geqslant 7$. Since this bound is known to be reachable, this settles the problem. We also reduce the alphabet of the witness languages reaching this bound to five letters instead of $n+2$, and show that it cannot be any smaller. Finally, we prove that the transition semigroup of a minimal deterministic automaton accepting such a witness language is unique for each $n \geqslant 7$.

Keywords: regular language, suffix-free, syntactic complexity, transition semigroup, upper bound

1 Preliminaries

1.1 Introduction

The *syntactic complexity* [7] $\sigma(L)$ of a regular language L is the size of its syntactic semigroup [10]. This semigroup is isomorphic to the transition semigroup of the quotient automaton $\mathcal{D}$ (a minimal deterministic finite automaton) accepting the language. The number n of states of $\mathcal{D}$ is the *state complexity* of the language [12], and it is the same as the *quotient complexity* [3] (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.

If $w = uxv$ for some $u, v, x \in \Sigma^*$, then u is a *prefix* of w, v is a *suffix* of w and x is a *factor* of w. A suffix of w is also a factor of w. A language L is *prefix-free* (respectively, *suffix-free, factor-free*) if $w, u \in L$ and u is a prefix (respectively, *suffix, factor*) of w, implies that $u = w$. A language is *bifix-free* if it

* This work was supported by the Natural Sciences and Engineering Research Council of Canada grant No. OGP000087, and by Polish NCN grant DEC-2013/09/N/ST6/01194.

is both prefix- and suffix-free. These languages play an important role in coding theory, have applications in such areas as cryptography, data compression, and information transmission, and have been studied extensively; see [2] for example. In particular, suffix-free languages (with the exception of $\{\varepsilon\}$, where ε is the empty word) are suffix codes. Moreover, suffix-free languages are special cases of suffix-convex languages, where a language is *suffix-convex* if it satisfies the condition that, if a word w and its suffix u are in the language, then so is every suffix of w that has u as a suffix [1, 11]. We are interested only in regular suffix-free languages.

The syntactic complexity of prefix-free languages was proved to be n^{n-2} in [4]. The syntactic complexities of suffix-, bifix-, and factor-free languages were also studied in [4], and the following lower bounds were established $(n-1)^{n-2}+n-2$, $(n-1)^{n-3}+(n-2)^{n-3}+(n-3)2^{n-3}$, and $(n-1)^{n-3}+(n-3)2^{n-3}+1$, respectively. It was conjectured that these bounds are also upper bounds; we prove the conjecture for suffix-free languages in this paper.

A full version of the paper is available in [5].

1.2 Languages, Automata and Transformations

Let Σ be a finite, non-empty alphabet and let $L \subseteq \Sigma^*$ be a language. The *left quotient* or simply *quotient* of a language L by a word $w \in \Sigma^*$ is denoted by $L.w$ and defined by $L.w = \{x \mid wx \in L\}$. A language is regular if and only if it has a finite number of quotients. We denote the set of quotients by $K = \{K_0, \ldots, K_{n-1}\}$, where $K_0 = L = L.\varepsilon$ by convention. Each quotient K_q can be represented also as $L.w_q$, where $w_q \in \Sigma^*$ is such that $L.w_q = K_q$.

A *deterministic finite automaton (DFA)* is a quintuple $\mathcal{D} = (Q, \Sigma, \delta, q_0, F)$, where Q is a finite non-empty set of *states*, Σ is a finite non-empty *alphabet*, $\delta\colon Q \times \Sigma \to Q$ is the *transition function*, $q_0 \in Q$ is the *initial* state, and $F \subseteq Q$ is the set of *final* states. We extend δ to a function $\delta\colon Q \times \Sigma^* \to Q$ 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 $\delta_{\mathcal{D}}(K_q, w) = K_p$ if and only if $K_q.w = K_p$, and $F_{\mathcal{D}} = \{K_q \mid \varepsilon \in K_q\}$. To simplify the notation, without loss of generality we use the set $Q = \{0, \ldots, n-1\}$ of subscripts of quotients as the set of states of $\mathcal{D}$; then $\mathcal{D}$ is denoted by $\mathcal{D} = (Q, \Sigma, \delta, 0, F)$, where $\delta(q, w) = p$ if $\delta_{\mathcal{D}}(K_q, w) = K_p$, and F is the set of subscripts of quotients in $F_{\mathcal{D}}$. The quotient corresponding to $q \in Q$ (known also as the *right language* of q) is $K_q = \{w \mid \delta_{\mathcal{D}}(K_q, w) \in F_{\mathcal{D}}\}$. The quotient $K_0 = L$ is the *initial* quotient. A quotient is *final* if it contains ε. A state 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, each letter $a \in \Sigma$ induces a transformation of the set Q of n states. Let $\mathcal{T}_Q$ be the set of all n^n transformations of Q; then $\mathcal{T}_Q$ is a monoid under composition. The *image* of $q \in Q$ under transformation t is denoted by qt. If s, t are transformations of Q, their composition is denoted $s \circ t$ and defined by $q(s \circ t) = (qs)t$; the $\circ$ is usually omitted. The *in-degree* of a state q in a transformation t is the cardinality of the set $\{p \mid pt = q\}$.

The *identity* transformation $\mathbf{1}$ maps each element to itself. For $k \geqslant 2$, a transformation (permutation) t of a set $P = \{q_0, q_1, \ldots, q_{k-1}\} \subseteq Q$ is a *k-cycle* if $q_0 t = q_1, q_1 t = q_2, \ldots, q_{k-2} t = q_{k-1}, q_{k-1} t = q_0$. A k-cycle is denoted by $(q_0, q_1, \ldots, q_{k-1})$. If a transformation t of Q is a k-cycle of some $P \subseteq Q$, then t *has a k-cycle*. A transformation *has a cycle* if it has a k-cycle for some $k \geqslant 2$. A 2-cycle (q_0, q_1) is called a *transposition*. A transformation is *unitary* if it changes only one state p to a state $q \neq p$; it is denoted by $(p \to q)$. A transformation mapping a subset P of Q to a single state and acting as the identity on $Q \setminus P$ is denoted by $(P \to q)$.

The binary relation ω_t on $Q \times Q$ is defined as follows: For any $p, q \in Q$, $p \,\omega_t\, q$ if and only if $pt^k = qt^\ell$ for some $k, \ell \geqslant 0$. This is an equivalence relation, and each equivalence class is called an *orbit* [8] of t. For any $q \in Q$, the orbit of t containing q is denoted by $\omega_t(q)$. An orbit contains either exactly once cycle and no fixed points or exactly one fixed point and no cycles. The set of all orbits of t is a partition of Q.

If $w \in \Sigma^*$ induces a transformation t, we denote this by $w\colon t$.

The *transition semigroup* of a DFA $\mathcal{D} = (Q, \Sigma, \delta, 0, F)$ is the semigroup of transformations of Q generated by the transformations induced by the letters of Σ. Since the transition semigroup of a minimal DFA of a language L is isomorphic to the syntactic semigroup of L [10], syntactic complexity is equal to the cardinality of the transition semigroup.

1.3 Suffix-Free Languages

For any transformation t, consider the sequence $(0, 0t, 0t^2, \ldots)$; we call it the *0-path* of t. Since Q is finite, there exist i, j such that $0, 0t, \ldots, 0t^i, 0t^{i+1}, \ldots, 0t^{j-1}$ are distinct but $0t^j = 0t^i$. The integer $j - i$ is the *period* of t and if $j - i = 1$, t is *initially aperiodic*. Let $Q = \{0, \ldots, n-1\}$, let $\mathcal{D}_n = (Q, \Sigma, \delta, 0, F)$ be a minimal DFA accepting a language L, and let T_n be its transition semigroup. The following is known [4, 9]:

Lemma 1. *If L is a suffix-free language, then*

1. *There exists $w \in \Sigma^*$ such that $L.w = \emptyset$; hence $\mathcal{D}_n$ has an empty state, which is state $n-1$ by convention.*
2. *For $w, x \in \Sigma^+$, if $L.w \neq \emptyset$, then $L.w \neq L.xw$.*
3. *If $L.w \neq \emptyset$, then $L.w = L$ implies $w = \varepsilon$.*
4. *For any $t \in T_n$, the 0-path of t in $\mathcal{D}_n$ is aperiodic and ends in $n-1$.*

An (unordered) pair $\{p, q\}$ of distinct states in $Q \setminus \{0, n-1\}$ is *colliding* (or p *collides* with q) in T_n if there is a transformation $t \in T_n$ such that $0t = p$ and $rt = q$ for some $r \in Q \setminus \{0, n-1\}$. A pair of states is *focused* by a transformation u of Q if u maps both states of the pair to a single state $r \notin \{0, n-1\}$. We then say that $\{p, q\}$ is *focused to state r*. If L is a suffix-free language, then from Lemma 1 (2) it follows that if $\{p, q\}$ is colliding in T_n, there is no transformation $t' \in T_n$ that focuses $\{p, q\}$. So colliding states can be mapped to a single state by a transformation in T_n only if that state is the empty state $n-1$.

Remark 1. If $n = 1$, the only suffix-free language is the empty language $\emptyset$ and $\sigma(\emptyset) = 1$. If $n \geqslant 2$ and $\Sigma = \{a\}$, the language $L = a^{n-2}$ is the only suffix-free language of quotient complexity n, and its syntactic complexity is $\sigma(L) = n-1$.

Assume now that $|\Sigma| \geqslant 2$. If $n = 2$, the language $L = \varepsilon$ is the only suffix-free language, and $\sigma(L) = 1$. If $n = 3$, the tight upper bound on syntactic complexity of suffix-free languages is 3, and $L = ab^*$ over $\Sigma = \{a, b\}$ meets this bound [4].

If $n = 4$ and $n = 5$, the tight upper bounds are 13, and 73 [4]. In [4] it was shown that there is a suffix-free witness DFA with n states and an alphabet of size $n + 2$ that meets the bound $(n-1)^{n-2} + n - 2$ for $n \geqslant 4$. For $n = 4$ and $n = 5$, these bounds are 11 and 67, and so are smaller than the bounds above. For $n \geqslant 6$, $(n-1)^{n-2} + n - 2$ is the largest known lower bound. ■

2 Lower Bound for Suffix-Free Languages

The lower bound of $(n-1)^{n-2}+n-2$ on the complexity of suffix-free languages was established in [4] using a witness DFA with an alphabet with $n+2$ letters. Our first contribution is to simplify the witness of [4] by using an alphabet with only five letters, as stated in Definition 1. The transitions induced by inputs a, b, c, and e are the same as in [4].

Definition 1 (Witness). *For $n \geqslant 4$ define the DFA $\mathcal{W}_n = (Q, \Sigma_{\mathcal{W}}, \delta_{\mathcal{W}}, 0, \{1\})$, where $Q = \{0, \ldots, n-1\}$, $\Sigma_{\mathcal{W}} = \{a, b, c, d, e\}$, and $\delta_{\mathcal{W}}$ is defined by the transformations a: $(0 \to n-1)(1, \ldots, n-2)$, b: $(0 \to n-1)(1, 2)$, c: $(0 \to n-1)(n-2 \to 1)$, d: $(\{0, 1\} \to n-1)$, and e: $(Q \setminus \{0\} \to n-1)(0 \to 1)$. For $n = 4$, a and b coincide, and we can use $\Sigma_{\mathcal{W}} = \{b, c, d, e\}$. Let S_n be the transition semigroup of $\mathcal{W}_n$.*

The structure of $\mathcal{W}_n$ is illustrated in Figure 1 for $n = 5$. We claim that no pair of states from Q is colliding in S_n. If $0t = p \notin \{0, n-1\}$, then t is not the identity but must be induced by a word of the form ew for some $w \in \Sigma^*$. Such a word maps every $r \notin \{0, n-1\}$ to $n-1$; so $q = rt = n-1$, and p and q do not collide.

Proposition 1. *For $n \geqslant 4$ the DFA of Definition 1 is minimal, suffix-free, and its transition semigroup S_n has cardinality $(n-1)^{n-2}+n-2$. In particular, S_n contains (a) all $(n-1)^{n-2}$ transformations that send 0 and $n-1$ to $n-1$ and*

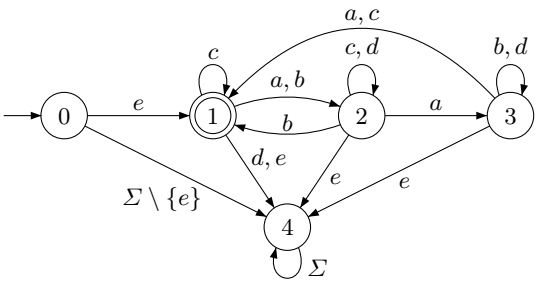

Fig. 1. Witness DFA $\mathcal{W}_5$.

map $Q \setminus \{0, n-1\}$ to $Q \setminus \{0\}$, and (b) all $n-2$ transformations that send 0 to a state in $Q \setminus \{0, n-1\}$ and map all the other states to $n-1$.

3 Upper Bound for Suffix-Free Languages

Our second result shows that the lower bound $(n-1)^{n-2}+n-2$ on the syntactic complexity of suffix-free languages is also an upper bound. Our approach is as follows: We consider a minimal DFA $\mathcal{D}_n = (Q, \Sigma, \delta, 0, F)$, where $Q = \{0, \ldots, n-1\}$, of an arbitrary suffix-free language with n quotients and let T_n be the transition semigroup of $\mathcal{D}_n$. We also deal with the witness DFA $\mathcal{W}_n = (Q, \Sigma_{\mathcal{W}}, \delta_{\mathcal{W}}, 0, \{1\})$ of Definition 1 that has the same state set as $\mathcal{D}_n$ and whose transition semigroup is S_n. We shall show that there is an injective mapping $\varphi\colon T_n \to S_n$, and this will prove that $|T_n| \leqslant |S_n|$.

The image of our mapping φ of a transition t in T_n depends on the properties of t. We separate these properties into 12 mutually disjoint cases that cover all the possibilities. The cases are structured as follows: We begin with an arbitrary transformation $t \in T_n$. Case 1 consists of transformations t that are also in S_n, and the remainder, R_1, of the cases has $t \notin S_n$. Having reached Case i, we define Case $(i+1)$ as all the transformations that do not fit in Cases 1 to i and satisfy a property P_{i+1}. The remainder R_{i+1} consists of all the transformations that do not fit in Cases 1 to i, and do not satisfy P_{i+1}. Because of this structure it is evident that the cases are mutually disjoint. In view of Case 12, they exhaust all the possibilities. The proof for each case is similar: we prove that $s = \varphi(t)$ differs from all the images s defined in previous cases and also from all the other images defined in the present case.

A note about terminology may be helpful to the reader. The semigroups T_n and S_n share the set Q. When we say that a pair of states from Q is *colliding* we mean that it is colliding in T_n; there is no room for confusion because no pair of states is colliding in S_n. Since we are dealing with suffix-free languages, a transformation that focuses a colliding pair cannot belong to T_n.

In Cases 2–11 of the proof p always denotes $0t$.

Theorem 1 (Tight Bound). *For $n \geqslant 6$ the syntactic complexity of the class of suffix-free languages with n quotients is $(n-1)^{n-2} + n - 2$.*

Proof. The case $n = 6$ has been proved in [4]; hence assume that $n \geqslant 7$. In [4] and in Proposition 1 it was shown that $(n-1)^{n-2} + n - 2$ is a lower bound for $n \geqslant 7$; hence it remains to prove that it is also an upper bound, and we do this here. We have the following cases:

Case 1: $t \in S_n$. Let $\varphi(t) = t$; obviously φ is injective.

Case 2: $t \notin S_n$, and t has a cycle. By Lemma 1 (4) we have the chain $0 \xrightarrow{t} p \xrightarrow{t} pt \xrightarrow{t} \cdots \xrightarrow{t} pt^k \xrightarrow{t} n-1$, where $k \geqslant 0$. Observe that pairs $\{pt^i, pt^j\}$ for $0 \leqslant i < j \leqslant k$ are colliding, since transformation t^{i+1} maps 0 to pt^i and pt^{j-i-1} to pt^j. Also, p collides with any state from a cycle of t and any fixed point of t other than $n-1$.

Let r be minimal 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

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

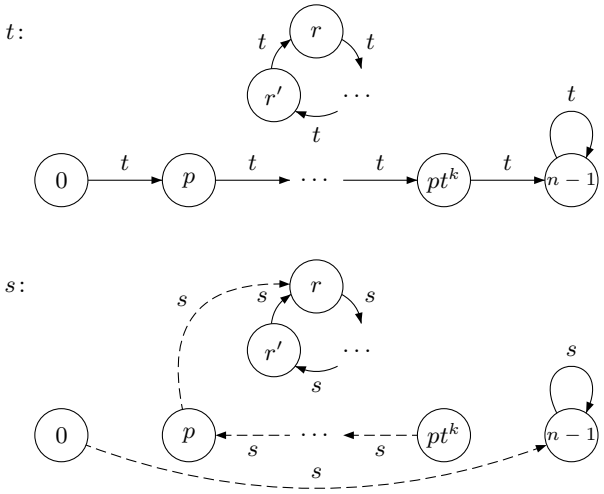

Fig. 2. Case 2 in the proof of Theorem 1.

By Proposition 1, $\varphi(t) = s$ is in S_n, since it maps 0 to $n-1$, fixes $n-1$, and does not map any states to 0. Note that the sets of cyclic states in both t and s are the same. Let r' be the state from the cycle of t such that $r't = r$; then transformation s has the following properties:

(a) Since p collides with any state in a cycle of t, $\{p, r'\}$ is a colliding pair focused by s to state r in the cycle. Moreover, if q' is a state in a cycle of s, and $\{q, q'\}$ is colliding and focused by s to a state in a cycle, then that state must be r (the minimal state in the cycles of s), q must be p, and q' must be r'. This follows from the definition of s. Since s differs from t only in the mapping of states pt^i and 0, any colliding pair focused by s contains pt^i for some i, $0 \leqslant i \leqslant k$. Only p is mapped to r, which is in a cycle of t, and r' is the only state in that cycle that is mapped to r.

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

(c) Every focused colliding pair consists of states from the orbit of p. This follows from the fact that all the states except 0 that are mapped by s differently than by t belong to the orbit of p.

(d) s has a cycle.

From (a), $s \notin T_n$ and so s is different from the transformations of Case 1.

Given a transformation s from this case we will construct a unique t that results in s when the definition of s given above is applied. This will show that our mapping φ has an inverse, and so is injective. From (a) there is the unique colliding pair focused to a state in a cycle. Moreover, one of the states in the pair, say p, is not in this cycle and another one, say r', is in this cycle. It follows that $0t = p$. Since there is no state $q \neq 0$ such that $qt = p$, the only state mapped to p by s is pt. From (b) for $i = 1, \ldots, k-1$ state pt^{i+1} is uniquely determined. Finally, for $i = k$ there is no state colliding with pt^{k-1} and mapped to pt^k; so $pt^{k+1} = n-1$. Since the other transitions in s are defined exactly as in t, this procedure defines the inverse function φ^{-1} for the transformations of this case.

Case 3: $t \notin S_n$, t has no cycles, but $pt \neq n-1$. Let s be the transformation defined by

$$0s = n-1, \quad ps = p, \quad (pt^i)s = pt^{i-1} \text{ for } 1 \leqslant i \leqslant k,$$
$$qs = qt \text{ for the other states } q \in Q.$$

Observe that s has the following properties:

(a) $\{p, pt\}$ is the only colliding pair focused by s to a fixed point. Moreover the fixed point is contained in the pair, and has in-degree 2. This follows from the definition of s, since any colliding pair focused by s contains pt^i, for some i with $0 \leqslant i \leqslant k$, and only pt is mapped to p, which is a fixed point. Also, no state except 0 is mapped to p by t since this would violate suffix-freeness; so only p and pt are mapped by s to p, and p has in-degree 2.

(b) For each i with $1 \leqslant i < k$, there is precisely one state q colliding with pt^{i-1} and mapped to pt^i, and that state is $q = pt^{i+1}$. This follows exactly like Property (b) from Case 2.

(c) Every colliding pair focused by s consists of states from the orbit of p. This follows exactly like Property (c) from Case 2.

(d) s does not have a cycle, but has a fixed point $f \neq n-1$ with in-degree $\geqslant 2$, which is p.

From (a), $s \notin T_n$ and so s is different from the transformations of Case 1. Here s does not have a cycle in contrast with the transformations of Case 2.

As before, s uniquely defines the transformation t from which it is obtained: From (a) there is the unique colliding pair $\{p, pt\}$ focused to the fixed point p. Thus $0t = p$. Then, as in Case 2, for $i = 1, \ldots, k-1$ state pt^{i+1} is uniquely defined, and $pt^k = n-1$. Since the other transitions in s are defined exactly as in t, this procedure yields the inverse function φ^{-1} for this case.

Case 4: t does not fit in any of the previous cases, but there is a fixed point $r \in Q \setminus \{0, n-1\}$ with in-degree $\geqslant 2$. Let s be the transformation defined by

$$0s = n-1, \quad ps = r,$$
$$qs = qt \text{ for the other states } q \in Q.$$

Observe that s has the following properties:

(a) $\{p, r\}$ is the only colliding pair focused by s to a fixed point, where the fixed point is contained in the pair. Also, the fixed point has in-degree at least 3. Since s differs from t only by the mapping of states 0 and p, it follows that all focused colliding pairs contain p. Since p is mapped to r, the second state in the pair must be the fixed point r. Since r has in-degree at least 2 in t, and s additionally maps p to r, r has in-degree at least 3.

(b) s does not have a cycle, but has a fixed point other than $n-1$ with in-degree $\geqslant 3$, which is r.

From (a) we have $s \notin T_n$, and so s is different from the transformations of Case 1. Here s does not have a cycle in contrast with the transformations of Case 2. Also from (a) we know that the fixed point in the distinguished colliding pair has in-degree $\geqslant 3$, whereas in Case 3 it has in-degree 2. From (a) we see that the colliding pair $\{p, r\}$ in which r is a fixed point and p is not is uniquely defined. Hence $0t = p$ and $pt = n-1$, and t is again uniquely defined from s.

Case 5: t does not fit in any of the previous cases, but there is a state r with in-degree $\geqslant 1$ that is not a fixed point and satisfies $rt \neq n-1$.

Since there are no fixed points in s with in-degree $\geqslant 2$ other than $n-1$, and there are no cycles, it follows that r belongs to the orbit of $n-1$. Hence we can choose r such that $rt \neq n-1$ and $rt^2 = n-1$.

Let s be the transformation defined by

$$0s = n-1, \quad ps = rt,$$
$$qs = qt \text{ for the other states } q \in Q.$$

Observe that s has the following properties:

(a) All focused colliding pairs contain p, and the second state from such a pair has in-degree $\geqslant 1$.

This follows since s differs from t only in the mapping of 0 and p.

(b) The smallest i with $ps^i = n-1$ is 2.

(c) s has neither a cycle nor a fixed point with in-degree $\geqslant 2$ other than $n-1$.

Note that p and r collide. Since $\{p, r\}$ is focused to rt, we have $s \notin T_n$ and so s is different from the transformations of Case 1. Here s does not have a cycle in contrast with the transformations of Case 2. Also s does not have a fixed point other than $n-1$, and so is different from the transformations of Cases 3 and 4.

From (a) all focused colliding pairs contain p. If there are two or more such pairs, p is the only state in their intersection. If there is only one such pair, then it must be $\{p, r\}$, and p is uniquely determined, since it has in-degree 0 and r has in-degree $\geqslant 1$. Hence $0t = p$ and $pt = n-1$, and again t is uniquely defined from s.

Case 6: t does not fit in any of the previous cases, but there is a state $r \in Q \setminus \{0, n-1\}$ with in-degree $\geqslant 2$. Clearly $r \neq p$, since the in-degree of p is 1.

Also $rt = n-1$, as otherwise t would fit in Case 5. Let $R = \{r' \in Q \mid r't = r\}$; then $|R| \geqslant 2$. We consider the following two sub-cases. If $p < r$, let q_1 be the smallest state in R and let q_2 be the second smallest state; so $q_1 < q_2$. If $p > r$, let q_1 be the second smallest state in R, and let q_2 be the smallest state; so $q_2 < q_1$. Let s be the transformation defined by

$$0s = n-1, \quad ps = q_1, \quad rs = q_1, \quad q_1 s = q_2, \quad q_2 s = n-1,$$
$$qs = qt \text{ for the other states } q \in Q.$$

Observe that s has the following properties:

(a) There is only one focused colliding pair, namely $\{p, r\}$, mapped to q_1. Clearly p and r collide. Note that no state can be mapped by t to q_1 or q_2, since this would satisfy Case 5. Because q_1 is the only state mapped by s to q_2, it does not belong to a focused colliding pair. Also 0 and q_2 are mapped to $n-1$. Since the other states are mapped exactly as in t, it follows that s does not focus any other colliding pairs.

(b) The smallest i with $ps^i = n-1$ is 3.

(c) s has neither a cycle nor a fixed point $\neq n-1$ with in-degree $\geqslant 2$. This follows since t does not have a cycle, and the states $0, p, r, q_1, q_2$ that are mapped differently by s are in the orbit of $n-1$.

Since s focuses the colliding pair $\{p, r\}$, s is different from the transformations of Case 1. Also s has neither a cycle nor a fixed point $\neq n-1$ and so is different from the transformations of Cases 2, 3 and 4. In Case 5, transformation s^2 maps a colliding pair to $n-1$, and here s^2 maps the unique colliding pair to $q_2 \neq n-1$. Thus, s is different from the transformations of Case 5.

From (a) we have the unique colliding pair $\{p, r\}$ focused to q_1. Then $q_1 < q_1 s = q_2$ means that $p < r$, and so p is distinguished from r. Similarly, $q_1 > q_2$ means that $p > r$. Thus $0t = p$, $pt = n-1$, $q_1 t = r$, $q_2 t = r$, and $rt = n-1$, and t is again uniquely defined from s.

Case 7: t does not fit in any of the previous cases, but there are two states $q_1, q_2 \in Q \setminus \{0, n-1\}$ that are not fixed points and satisfy $q_1 t \neq n-1$ and $q_2 t \neq n-1$. Since this is not Case 5 we may assume that $q_1 t^2 = n-1$ and $q_2 t^2 = n-1$. Let $r_1 = q_1 t$ and $r_2 = q_2 t$; clearly $p \neq r_1$ and $p \neq r_2$. The in-degree of both q_1 and q_2 is 0; otherwise t would fit in Case 5. We consider the following two sub-cases. If $p < r_1$ then **(i)** let s be the transformation defined by

$$0s = n-1, \quad ps = q_1, \quad r_1 s = q_1, \quad q_1 s = n-1,$$
$$qs = qt \text{ for the other states } q \in Q.$$

If $p > r_1$ then **(ii)** let s be the transformation defined by

$$0s = n-1, \quad ps = q_1, \quad r_1 s = q_1, \quad q_1 s = q_2,$$
$$qs = qt \text{ for the other states } q \in Q.$$

Case 8: t does not fit in any of the previous cases, but it has two fixed points r_1 and r_2 in $Q \setminus \{0, n-1\}$ with in-degree 1; assume that $r_1 < r_2$.
Let s be the transformation defined by

$$0s = n-1, \quad ps = r_2, \quad r_1 s = r_2, \quad r_2 s = r_1,$$
$$qs = qt \text{ for the other states } q \in Q.$$

Case 9: t does not fit in any of the previous cases, but there is a state $q \in Q \setminus \{0, n-1\}$ that is not a fixed point and satisfies $qt \neq n-1$, $p < qt$, and there is a fixed point $f \neq n-1$.

Let $r = qt$; then $rt = n-1$ because otherwise this would fit in Case 5. Here q is the only state from $Q \setminus \{0\}$ that is not a fixed point and is not mapped to $n-1$, as otherwise t would fit in Case 7. Similarly, f is the only fixed point $\neq n-1$, as otherwise t would fit in either Case 4 or Case 8.

Let s be the transformation defined by

$$0s = n-1, \quad ps = r, \quad rs = q, \quad qs = p, \quad fs = r,$$
$$qs = qt \text{ for the other states } q \in Q.$$

Case 10: t does not fit in any of the previous cases, but there is a state $q \in Q \setminus \{0, n-1\}$ that is not a fixed point and satisfies $qt \neq n-1$, and a fixed point $f \in Q \setminus \{0, n-1\}$. Let $r = qt$; then $rt = n-1$ since this is not Case 5. Now, in contrast to Case 9, we have $p > r$. Let s be the transformation defined by

$$0s = n-1, \quad ps = q, \quad rs = q, \quad qs = n-1,$$
$$qs = qt \text{ for the other states } q \in Q.$$

Case 11: t does not fit in any of the previous cases, but there is a state $q \in Q \setminus \{0, n-1\}$ that is not a fixed point and satisfies $qt \neq n-1$.
As shown in Case 9, q is the only state from $Q \setminus \{0\}$ that is not mapped to $n-1$, and also t has no fixed points other than $n-1$, as otherwise it would fit in one of the previous cases. Hence, all states from $Q \setminus \{0, q\}$ are mapped to $n-1$. Let $r = qt$. Here we use the assumption that $n \geqslant 7$. So in $Q \setminus \{0, p, q, r, n-1\}$ we have at least 2 states, say r_1 and r_2, that are mapped to $n-1$.
Sub-case (i): $p < r$. Let s be the transformation defined by

$$0s = n-1, \quad ps = q, \quad rs = q, \quad qs = n-1,$$
$$qs = qt \text{ for the other states } q \in Q.$$

Sub-case (ii): $p > r$. Let s be the transformation defined by

$$0s = n-1, \quad ps = q, \quad rs = q, \quad qs = n-1, \quad r_1 s = r_2, \quad r_2 s = r_1,$$
$$qs = qt \text{ for the other states } q \in Q.$$

Case 12: t does not fit in any of the previous cases.
Here t must contain exactly one fixed point $f \in Q \setminus \{n-1\}$, and every state from $Q \setminus \{0, f\}$ is mapped to $n-1$. If all states from $Q \setminus \{0\}$ would be mapped to $n-1$, then by Proposition 1, t would be in S_n and so would fit in Case 1.

Because $n \geqslant 7$, in $Q \setminus \{0, p, f, n-1\}$ we have at least 2 states, say r_1 and r_2, that are mapped to $n-1$. Let s be the transformation defined by

$$0s = n-1, \quad ps = f, \quad r_1 s = r_2, \quad r_2 s = r_1,$$
$$qs = qt \text{ for the other states } q \in Q.$$

□

4 Uniqueness of Maximal Witness

Our third contribution is a proof that the transition semigroup of a DFA $\mathcal{D}_n = (Q, \Sigma, \delta, 0, F)$ of a suffix-free language with syntactic complexity $(n-1)^{n-2}+n-2$ is unique.

Lemma 2. *If $n \geqslant 4$ and $\mathcal{D}_n$ has no colliding pairs, then $|T_n| \leqslant (n-1)^{n-2}+n-2$ and T_n is a subsemigroup of S_n.*

Lemma 3. *If $n \geqslant 7$ and $\mathcal{D}_n$ has at least one colliding pair, then $|T_n| < (n-1)^{n-2}+n-2$.*

Proof. Let φ be the injective function from the proof of Theorem 1 and assume that there is a colliding pair $\{p, r\}$. Let r_1, r_2 and r_3 be three distinct states from $Q \setminus \{0, p, r, n-1\}$; there are at least 3 such states since $n \geqslant 7$. Let s be the following transformation:

$$0s = n-1, \quad ps = r, \quad rs = r, \quad r_1 s = r_2, \quad r_2 s = r_3, \quad r_3 s = r_1,$$
$$qs = qt \text{ for the other states } q \in Q.$$

We can show that s is not defined in any case in the proof of Theorem 1. Note that s focuses the colliding pair $\{p, r\}$, and so it cannot be present in T_n; hence it is not defined in Case 1. We can follow the proof of injectivity of the transformations in Case 12 of Theorem 1, and show that s is different from all the transformations of Cases 2–11. For a distinction from the transformations of Case 12, observe that they each have a 2-cycle, and here s has a 3-cycle.

Thus $s \notin \varphi(T_n)$, but $s \in S_n$, and so $\varphi(T_n) \subsetneq S_n$. Since φ is injective, it follows that $|T_n| < |S_n| = (n-1)^{n-2}+n-2$. □

Corollary 1. *For $n \geqslant 7$, the maximal transition semigroups of DFAs of suffix-free languages are unique.*

Finally, we show that Σ cannot have fewer than five letters.

Theorem 2. *If $n \geqslant 7$, $\mathcal{D}_n = (Q, \Sigma, \delta, 0, F)$ is a minimal DFA of a suffix-free language, and $|\Sigma| < 5$, then $|T_n| < (n-1)^{n-2}+n-2$.*

Proof. DFA $\mathcal{D}_n$ has the initial state 0, and an empty state, say $n-1$. Let M be the set of the remaining $n-2$ "middle" states. From Lemma 1 no transformation can map any state in Q to 0, and every transformation fixes $n-1$.

Suppose the upper bound $(n-1)^{n-2}+n-2$ is reached by T_n. From Proposition 1 and Corollary 1 all transformations of M must be possible, and it is well known that three generators are necessary to achieve this. Let the letters a, b, and c correspond to these three generators, t_a, t_b and t_c. If $0t_a \neq n-1$, then t_a must be a transformation of type (b) from Proposition 1, and so $qt_a = n-1$ for any $q \in M$. So t_a cannot be a generator of a transformation of M. Hence we must have $0t_a = n-1$, and also $0t_b = 0t_c = n-1$.

So far, the states in M are not reachable from 0; hence there must be a letter, say e, such that $0t_e = p$ is in M. This must be a transformation of type (b) from Proposition 1, and all the states of M must be mapped to $n-1$ by t_e.

Finally, to reach the upper bound we must be able to map any proper subset of M to $n-1$. The letter e will not do, since it maps *all* states of M to $n-1$. Hence we require a fifth letter, say d. □

5 Conclusions

We have shown that the upper bound on the syntactic complexity of suffix-free languages is $(n-1)^{n-2}+n-2$. Since it was known that this is also a lower bound, our result settles the problem. Moreover, we have proved that an alphabet of at least five letters is necessary to reach the upper bound, and that the maximal transition semigroups are unique.

In our proof we exhibited an injective function from the transition semigroup of a minimal DFA of an arbitrary suffix-free semigroup to the transition semigroup of the witness DFA attaining the upper bound for suffix-free languages. This approach is generally applicable for other subclasses of regular languages. For example, in [6] we have used this method to establish the upper bound for left and two-sided ideals.

References

1. Ang, T., Brzozowski, J.: Languages convex with respect to binary relations, and their closure properties. Acta Cybernet. 19(2), 445–464 (2009)
2. Berstel, J., Perrin, D., Reutenauer, C.: Codes and Automata. Cambridge University Press (2009)
3. Brzozowski, J.: Quotient complexity of regular languages. J. Autom. Lang. Comb. 15(1/2), 71–89 (2010)
4. Brzozowski, J., Li, B., Ye, Y.: Syntactic complexity of prefix-, suffix-, bifix-, and factor-free regular langauges. Theoret. Comput. Sci. 449, 37–53 (2012)
5. Brzozowski, J., Szykuła, M.: Upper bound for syntactic complexity of suffix-free languages (2014), `http://arxiv.org/abs/1412.2281`
6. 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 (2014)
7. 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 (2011)
8. Ganyushkin, O., Mazorchuk, V.: Classical Finite Transformation Semigroups: An Introduction. Springer (2009)
9. Han, Y.S., Salomaa, K.: State complexity of basic operations on suffix-free regular languages. Theoret. Comput. Sci. 410(27-29), 2537–2548 (2009)
10. Pin, J.E.: Syntactic semigroups. In: Handbook of Formal Languages, vol. 1: Word, Language, Grammar, pp. 679–746. Springer, New York, NY, USA (1997)
11. Thierrin, G.: Convex languages. In: Nivat, M. (ed.) Automata, Languages and Programming, pp. 481–492. North-Holland (1973)
12. Yu, S.: State complexity of regular languages. J. Autom. Lang. Comb. 6, 221–234 (2001)