Syntactic complexity of suffix-free languages

Janusz Brzozowski[a], Marek Szykuła[b]

[a] *David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada N2L 3G1*
[b] *Institute of Computer Science, University of Wrocław, Joliot-Curie 15, PL-50-383 Wrocław, Poland*

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 $n \geqslant 6$. 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 a witness language is unique for each n.

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

1. Introduction

The *syntactic complexity* [8] of a regular language L is the size of its syntactic semigroup [14]. This semigroup is isomorphic to the transition semigroup of the quotient automaton $\mathcal{D}$, a minimal deterministic finite automaton (DFA) accepting the language. The descriptional complexity of syntactic monoids as a function of minimal DFA size for regular languages was first considered systematically in [11, 13].

The number n of states of $\mathcal{D}$ is the *state complexity* of the language [16], 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

Email addresses: brzozo@uwaterloo.ca (Janusz Brzozowski), msz@cs.uni.wroc.pl (Marek Szykuła)

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. Prefixes and suffixes of w are also factors 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, then $u = w$. A language is *bifix-free* if it 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, 15]. 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. Moreover, we reduce the alphabet size of the witness language reaching the upper bound for suffix-free languages to five letters instead of $n+2$, and prove that five is the minimal size. As well, we show that the transition semigroup of a minimal DFA accepting a witness language is unique for each n.

A much abbreviated version of these results appeared in [7].

2. Preliminaries

2.1. 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_i can be represented also as $L.w_i$, where $w_i \in \Sigma^*$ is such that $L.w_i = K_i$. The notation $K_i.w$ points out that each word $w \in \Sigma^*$ performs an action on the

set K of quotients (states of the quotient DFA), and leads a quotient (state) K_i to quotient (state) $K_i.w$.

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 $\delta_{\mathcal{D}}(K_i, w) = K_j$ if and only if $K_i.w = K_j$, and $F_{\mathcal{D}} = \{K_i \mid \varepsilon \in K_i\}$. 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 $\delta(i, w) = j$ if $\delta_{\mathcal{D}}(K_i, w) = K_j$, and F is the set of subscripts of quotients in $F_{\mathcal{D}}$. The quotient corresponding to $q \in Q$ is then $K_q = \{w \mid \delta_{\mathcal{D}}(K_q, w) \in F_{\mathcal{D}}\}$. The quotient $K_0 = L$ is the *initial* quotient. A quotient is *final* if it contains ε. A state q is *empty* (or a *sink state* or *dead state*) if its quotient K_q is empty.

The quotient DFA of L is a 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 $q_0t = q_1, q_1t = q_2, \ldots, q_{k-2}t = q_{k-1}, q_{k-1}t = q_0$. 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 $P \subseteq Q$, we say that 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 is *constant* if it maps all states to a single state q; it is denoted by $(Q \to q)$.

The binary relation ω_t on $Q \times Q$ is defined as follows: For any $i, j \in Q$, $i \,\omega_t\, j$ if and only if $it^k = jt^\ell$ for some $k, \ell \geqslant 0$. This is an equivalence relation, and each equivalence class is called an *orbit* [9] of t. For any $i \in Q$, the orbit of t containing i is denoted by $\omega_t(i)$. An orbit contains either exactly one

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 $w\colon t$. A transformation mapping i to q_i for $i = 0, \ldots, n-1$ is sometimes denoted by $[q_0, \ldots, q_{n-1}]$. By a slight abuse of notation we sometimes represent the transformation t induced by w by w itself, and write qw instead of qt.

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 [14], syntactic complexity is equal to the cardinality of the transition semigroup.

2.2. 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 $Q = \{0, \ldots, n-1\}$, and let $Q_M = \{1, \ldots, n-2\}$ (the set of *middle* states). 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 observations are well known [4, 10]:

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 $n-1$.*

Remark 1. *If $n = 1$, the only suffix-free language is the empty language $\emptyset$ and its syntactic complexity is equal to 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 equal to $n-1$.*

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

2.3. Suffix-free semigroups

Since the cases where $2 \leqslant n \leqslant 3$ were easily resolved, we assume now that $n \geqslant 4$. Also, without loss of generality, we assume that $Q = \{0, \ldots, n-1\}$. A transformation of Q is *suffix-free* if it can belong to the transition semigroup of a minimal DFA accepting a suffix-free language.

The following set of all suffix-free transformations was defined in [4]: For $n \geqslant 2$ let

$$\begin{aligned}\mathbf{B}_{\mathrm{sf}}(n) = \{t \in \mathcal{T}_Q \quad | \quad & 0 \notin Qt,\ (n-1)t = n-1,\ \text{and for all } j \geqslant 1, \\ & 0t^j = n-1 \text{ or } 0t^j \neq qt^j \ \ \forall q,\ 0 < q < n-1\}.\end{aligned}$$

An (unordered) pair $\{p, q\}$ of distinct states in Q_M 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 $r \in Q_M$. A pair of states is *focused* by a transformation u of Q if u maps both states of the pair to a single state $r \in Q_M$. We then say that the pair $\{p, q\}$ is *focused by u to state r*, or simply that the pair (p, q) is *focused*, if there is no danger of confusion. 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 $n-1$.

Suppose $\mathcal{D}$ is a minimal DFA accepting a non-empty suffix-free language L. It was shown in [4] that he transition semigroup of such a DFA is contained in $\mathbf{B}_{\mathrm{sf}}(n)$. Each transformation in $\mathbf{B}_{\mathrm{sf}}(n)$ satisfies three conditions. First, no transformation t induced by a word y can map any state q to 0, because then we would have some words x and z such that $0x = q$, $qy = 0$, $z \in L$ and $xyz \in L$. Second, since $\mathcal{D}$ must have an empty state $n-1$, that state must be mapped to itself by every transformation. Third, since state $0t^j$ collides with qt^j for all $j \geqslant 1$ if $0t^j \neq n-1$, these two states cannot be focused.

Since the set $\mathbf{B}_{\mathrm{sf}}(n)$ is not a semigroup, we look for largest semigroups contained in $\mathbf{B}_{\mathrm{sf}}(n)$. Two such semigroups were introduced in [4].

The first semigroup is defined as follows: For $n \geqslant 4$, let

$$\begin{aligned}\mathbf{T}^{\leqslant 5}(n) = \{t \in \mathbf{B}_{\mathrm{sf}}(n) \quad | \quad & \text{for all } p, q \in Q \text{ where } p \neq q, \\ & \text{we have } pt = qt = n-1 \text{ or } pt \neq qt\}.\end{aligned}$$

It was shown in [6] that for $n \geqslant 4$, semigroup $\mathbf{T}^{\leqslant 5}(n)$ is generated by the following set of transformations of Q:

- $a\colon (0 \to n-1)(1,\ldots,n-2)$,
- $b\colon (0 \to n-1)(1,2)$,
- for $1 \leqslant p \leqslant n-2$, $c_p\colon (p \to n-1)(0 \to p)$.

Proposition 1 ([6, Proposition 3])**.** *For $n \geqslant 4$, $\mathbf{T}^{\leqslant 5}(n)$ is the unique maximal semigroup of a minimal DFA $\mathcal{D}$ of a suffix-free language in which every two states from Q_M are colliding.*

Proof. Observe that every two states $p, q \in Q_M$, $p \neq q$, are colliding, because there is a transformation $t \in \mathbf{T}^{\leqslant 5}(n)$ with $0t = p$ and $qt = q$. Then for each $p, q \in Q \setminus \{n-1\}$, there is no transformation t with $pt = qt \neq n-1$, for this would violate suffix-freeness. By definition, $\mathbf{T}^{\leqslant 5}(n)$ has all other transformations that are possible for a suffix-free language, and hence is unique. □

The second semigroup from [4] was defined as follows: For $n \geqslant 4$, let

$$\mathbf{T}^{\geqslant 6}(n) = \{t \in \mathbf{B}_{\mathrm{sf}}(n) \mid 0t = n-1 \text{ or } qt = n-1 \ \ \forall\ q,\ 1 \leqslant q \leqslant n-2\}.$$

The transition semigroup $\mathbf{T}^{\geqslant 6}(n)$ has cardinality $(n-1)^{n-2} + n - 2$, which is a lower bound on the complexity of suffix-free languages established in [4] using a witness DFA with an alphabet with $n+2$ letters. Our first contribution is to simplify the witness of [4] with the transition semigroup $\mathbf{T}^{\geqslant 6}(n)$ 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]. The structure of $\mathcal{W}_n$ is illustrated in Fig. 1 for $n = 5$.

Definition 1 (Suffix-Free Witness)**.** *For $n \geqslant 4$, we 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 $a\colon (0 \to n-1)(1,\ldots,n-2)$, $b\colon (0 \to n-1)(1,2)$, $c\colon (0 \to n-1)(n-2 \to 1)$, $d\colon (\{0,1\} \to n-1)$, and $e\colon (Q\setminus\{0\} \to n-1)(0 \to 1)$. For $n = 4$, a and b coincide, and we can use $\Sigma_{\mathcal{W}} = \{b,c,d,e\}$.*

We claim that no pair of states from Q is colliding in $\mathbf{T}^{\geqslant 6}(n)$. 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 $n-1$. Hence $q = rt = n-1$, and p and q do not collide.

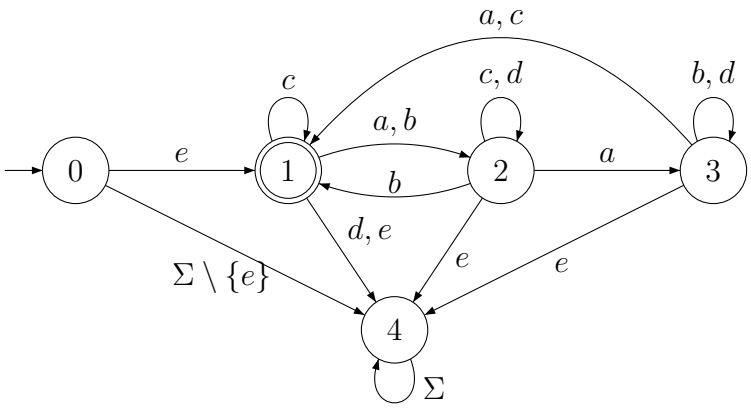

Figure 1: DFA $\mathcal{W}_5$.

Proposition 2. *For $n \geqslant 4$ the DFA $\mathcal{W}_n$ of Definition 1 is minimal, suffix-free, and its transition semigroup is $\mathbf{T}^{\geqslant 6}(n)$ with cardinality $(n-1)^{n-2}+n-2$. In particular, $\mathbf{T}^{\geqslant 6}(n)$ contains (a) all $(n-1)^{n-2}$ transformations that send 0 and $n-1$ to $n-1$ and map Q_M to $Q \setminus \{0\}$, and (b) all $n-2$ transformations that send 0 to a state in Q_M and map all the other states to $n-1$.*

Proof. $\mathcal{W}_n$ accepts a suffix-free language, since each accepted string is of the form ew with $w \in \{a, b, c, d\}$. For minimality, $n-1$ is the only empty state, e is accepted only from 0, and all states in Q_M are distinguished by a string in a^*.

Note that $\mathbf{T}^{\geqslant 6}(n)$ contains only transformations of type (a) and (b) and it contains all such transformations. The transformations induced by a, b, and c restricted to Q_M generate all transformations of the middle $n-2$ states. Together with d, they generate all transformations of Q that send 0 to $n-1$, fix $n-1$, and send a state in Q_M to $Q \setminus \{0\}$. Also, for $0 \leqslant i \leqslant n-3$, words ea^i send 0 to some $p \notin \{0, n-1\}$ and map all the other states to $n-1$. Hence the transition semigroup of $\mathcal{W}_n$ is $\mathbf{T}^{\geqslant 6}(n)$ and the proposition holds. □

Proposition 3. *For $n \geqslant 4$, $\mathbf{T}^{\geqslant 6}(n)$ is the unique maximal semigroup of a minimal DFA $\mathcal{D}$ of a suffix-free language in which no two states from $\{1, \ldots, n-2\}$ are colliding.*

Proof. Since $0t = n-1$ or $qt = n-1$ for all $q \in Q \setminus \{0\}$, no two states are colliding. By Proposition 2 all such transformations are in $\mathbf{T}^{\geqslant 6}(n)$. □

If $n = 4$ and $n = 5$, the tight upper bounds on the size of suffix-free transition semigroups are 13, and 73, respectively [4], and these bounds are

met by $\mathbf{T}^{\leqslant 5}(n)$. It was shown in [4] 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 $n \geqslant 4$, and the transition semigroup of this DFA is $\mathbf{T}^{\geqslant 6}(n)$. For $n = 4$ and $n = 5$, these bounds are 11 and 67, and hence are smaller than the sizes of $\mathbf{T}^{\leqslant 5}(4)$ and $\mathbf{T}^{\leqslant 5}(5)$. However, for $n \geqslant 6$, $(n-1)^{n-2}+n-2$ is the largest known lower bound, and it is met by $\mathbf{T}^{\geqslant 6}(n)$. The remainder of this paper is devoted to proving that $(n-1)^{n-2}+n-2$ is also an upper bound.

3. Upper bound for suffix-free languages

Our main 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 for $n \geqslant 7$.

Theorem 1 (Suffix-Free Languages)**.** *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 2 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.

Our approach is as follows: 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 $\mathbf{T}^{\geqslant 6}(n)$. We will show that there is an injective mapping $\varphi\colon T(n) \to \mathbf{T}^{\geqslant 6}(n)$, and this will prove that $|T(n)| \leqslant |\mathbf{T}^{\geqslant 6}(n)|$.

A note about terminology may be helpful to the reader. The semigroups $T(n)$ and $\mathbf{T}^{\geqslant 6}(n)$ 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 $\mathbf{T}^{\geqslant 6}(n)$. Since we are dealing with suffix-free languages, a transformation that focuses a colliding pair cannot belong to $T(n)$.

We have the cases shown below, and also summarized on page 32 in Fig. 13.

Case 1: $t \in \mathbf{T}^{\geqslant 6}(n)$.
Let $\varphi(t) = t$; so φ is injective.

Case 2: $t \notin \mathbf{T}^{\geqslant 6}(n)$, and t has a cycle.
In this case and in all the following ten cases let $p = 0t$. By Proposition 2(a),

$\mathbf{T}^{\geqslant 6}(n)$ contains all transformations that map 0 to $n-1$, so since $t \notin \mathbf{T}^{\geqslant 6}(n)$ we always have $p \neq n-1$.

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

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

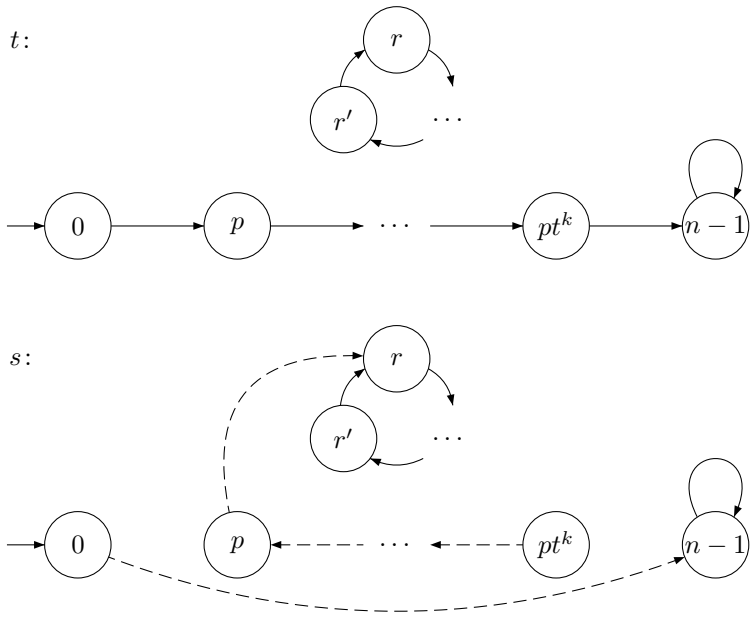

Figure 2: Case 2 in the proof of Theorem 1.

By Proposition 2, $\varphi(t) = s$ is in $\mathbf{T}^{\geqslant 6}(n)$, 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'.

Proof: 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 $q = pt^{i+1}$.

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.

(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.

We will show that s chosen as above corresponds to a unique t satisfying the conditions of this case, and we will define the inverse mapping φ^{-1} for the transformations s. From (a) there is the unique colliding pair focused to a state in a cycle. Moreover, one of its states, 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, and so φ is injective when restricted to these transformations.

Case 3: t does not fit in any of the previous cases, and $pt \neq n-1$. Let s be the transformation illustrated in Fig. 3 and defined by

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

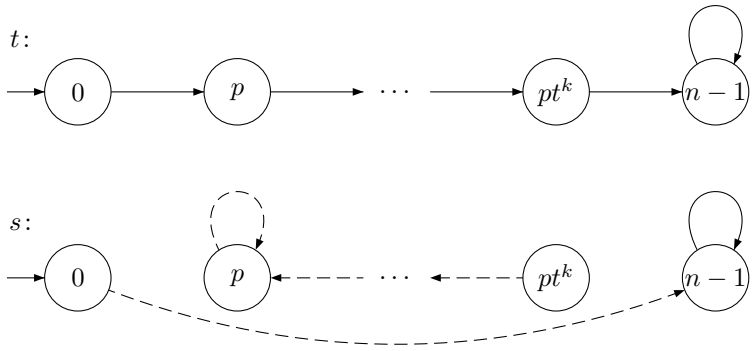

Figure 3: Case 3 in the proof of Theorem 1.

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.

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

(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 $q = pt^{i+1}$.

Proof: This follows exactly like Property (b) from Case 2.

(c) Every colliding pair focused by s consists of states from the orbit of p.

Proof: This follows exactly like Property (c) from Case 2.

(d) s does not have a cycle, but has a fixed point in Q_M with in-degree at least 2, and that fixed point 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 in s. 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, and there is a fixed point $r \in Q_M$ with in-degree $\geqslant 2$.
Let s be the transformation illustrated in Fig. 4 and defined by

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

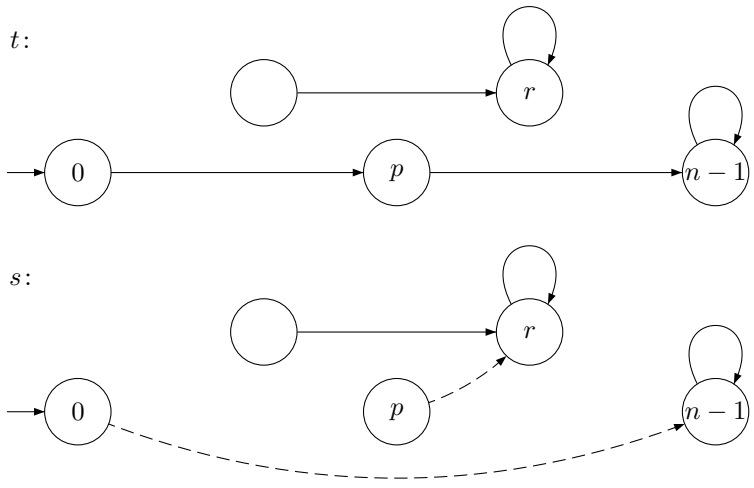

Figure 4: Case 4 in the proof of Theorem 1.

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. Moreover the fixed point has in-degree at least 3 in s.

Proof: 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 $\neq n-1$ 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 in s and p is not a fixed point in s, 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, and there is a state r with in-degree $\geqslant 1$ that is not a fixed point and satisfies $rt \neq n-1$.

Since in t there are no fixed points from Q_M with in-degree at least 2, 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 illustrated in Fig. 5 and 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 colliding pairs that are focused by s contain p, and the second state from such a pair has in-degree $\geqslant 1$ in s.

Proof: 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 $n-1$.

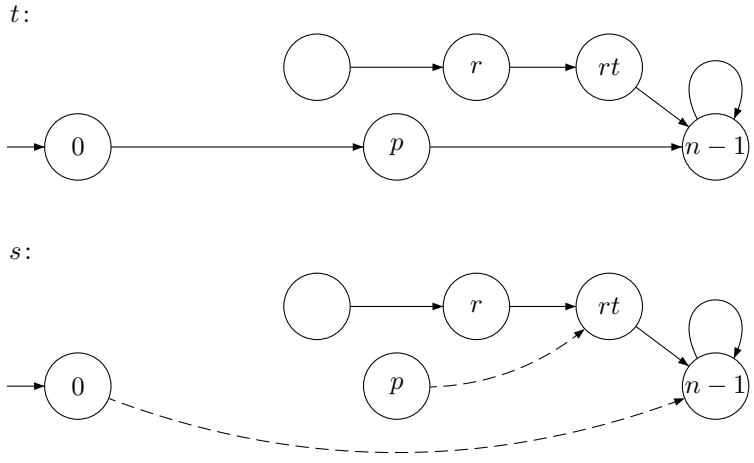

Figure 5: Case 5 in the proof of Theorem 1.

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 in Q_M with in-degree at least 2, 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 at least 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, and there is a state $r \in Q_M$ with in-degree at least 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 illustrated in Fig. 6 and defined by

$$0s = n - 1, \quad ps = q_1, \quad rs = q_1, \quad q_1 s = q_2, \quad q_2 s = n - 1,$$

$qs = qt$ for the other states $q \in Q$.

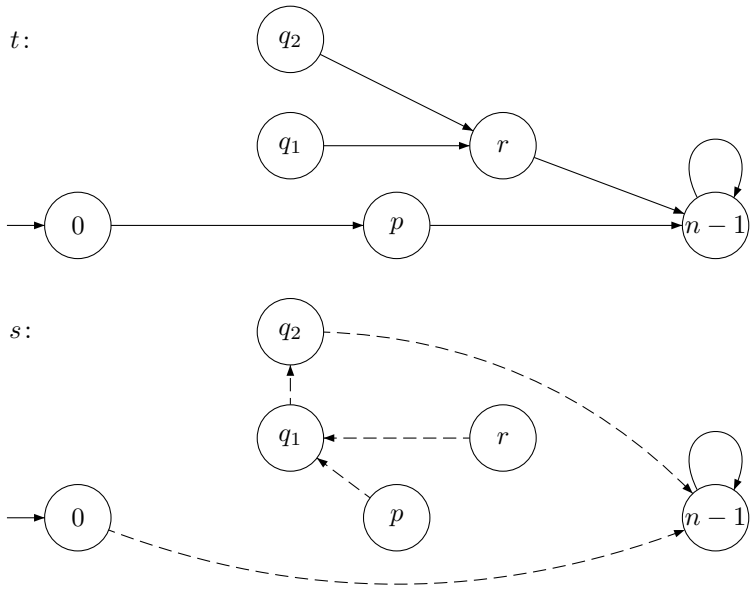

Figure 6: Case 6 in the proof of Theorem 1.

Observe that s has the following properties:

(a) There is only one colliding pair focused by s, namely $\{p, r\}$ mapped to q_1.

Proof: 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 from Q_M with in-degree at least 2 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, and there are two states $q_1, q_2 \in Q_M$ 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 in t 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 illustrated in Fig. 7 and 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 also illustrated in Fig. 7 and defined by

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

Observe that s has the following properties:

(a) There is only one colliding pair focused by s, namely the pair $\{p, r_1\}$ mapped to q_1. Both states from the pair have in-degree 0 in s.

Proof: Clearly p and r_1 collide. In (i) no state is mapped by s to q_2, and 0 and q_1 are mapped to $n-1$. In (ii) q_1 is the only state mapped to q_2 and 0 is mapped to $n-1$. 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 2 in (i), and is 4 in (ii).

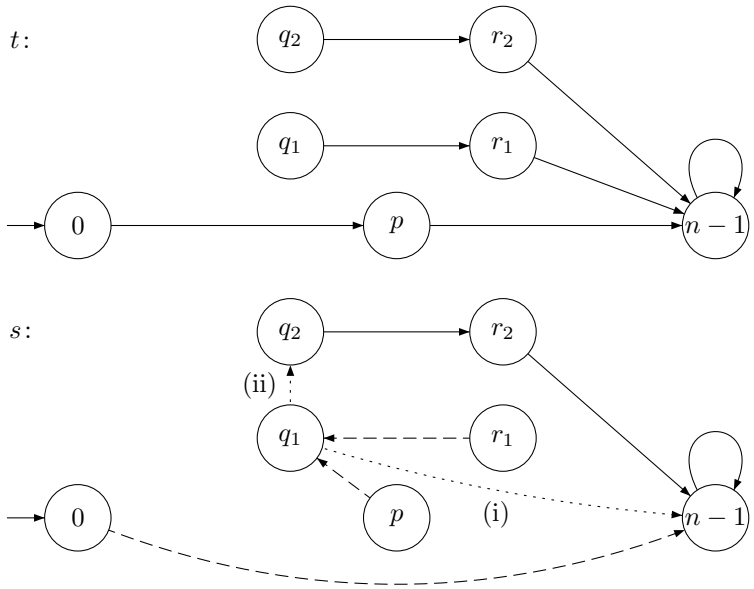

Figure 7: Case 7 in the proof of Theorem 1.

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

Since s focuses the colliding pair $\{p, r\}$, it is different from the transformations of Case 1. Also s has neither a cycle nor a fixed point with in-degree $\geqslant 2$ other than $n-1$, and so is different from the transformations of Cases 2, 3 and 4. Here the states from the colliding pair have in-degree 0, in contrast to the transformations of Case 5 (Property (a) of Case 5). Now, observe that the smallest i with $ps^i = n-1$ is 2 or 4, while in Case 6 it is 3 (Property (b) of Case 6).

From (a) we have the unique colliding pair $\{p, r_1\}$ focused to q_1. If $ps^2 = n-1$, then $p < r_1$ (i), and so p is distinguished from r_1. If $ps^2 \neq n-1$ then $ps^2 = q_2$, and $p > r_1$ (ii). Thus $0t = p$, $pt = n-1$, $r_1 t = n-1$, and $q_1 t = r_1$. Thus t is uniquely defined from s.

Case 8: t does not fit in any of the previous cases, and it has two fixed points r_1 and r_2 in Q with in-degree 1; assume that $r_1 < r_2$.

Let s be the transformation illustrated in Fig. 8 and defined by

$$0s = n-1, \quad ps = r_2, \quad r_1 s = r_2, \quad r_2 s = r_1,$$

$qs = qt$ for the other states $q \in Q$.

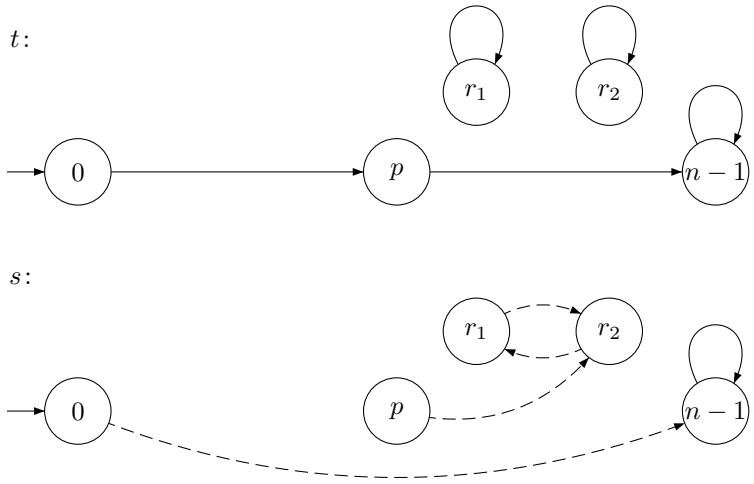

Figure 8: Case 8 in the proof of Theorem 1.

Observe that s has the following properties:

(a) $\{p, r_1\}$ is the only colliding pair that is focused by s. The state to which the pair is focused lies on a cycle of length 2 and is not the minimal state in the cycle.

Proof: This follows from the fact that r_1, r_2, and p are the only states in their orbit, only p and r_1 are mapped to a single state, and s differs from t only by the mapping of 0 and of these three states.

(b) s has a unique 2-cycle.

From (a), $s \notin T(n)$ and so s is different from the transformations of Case 1. The transformations of Case 2 contain a cycle, but (Property (a) of Case 2) the only colliding pair focused by them to a state lying on a cycle is mapped to the minimal state in cycles, in contrast to s from this case. Also, s has a cycle, and so differs from the transformations of Cases 3–7.

Again, s uniquely defines t: The orbit with the focused colliding pair contains precisely p, r_1, and r_2, and they are distinguished since $r_1 < r_2$, and r_1 and r_2 lie on a cycle whereas p does not.

Case 9: t does not fit in any of the previous cases, and there is a state $q \in Q_M$ that is not a fixed point and satisfies $qt \neq n-1$ and $p < qt$, and there is a fixed point $f \in Q_M$ with in-degree 1.

Let $r = qt$; then $rt = n-1$ because otherwise this would fit in Case 5. Here q is the only state from Q_M 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 illustrated in Fig. 9 and defined by

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

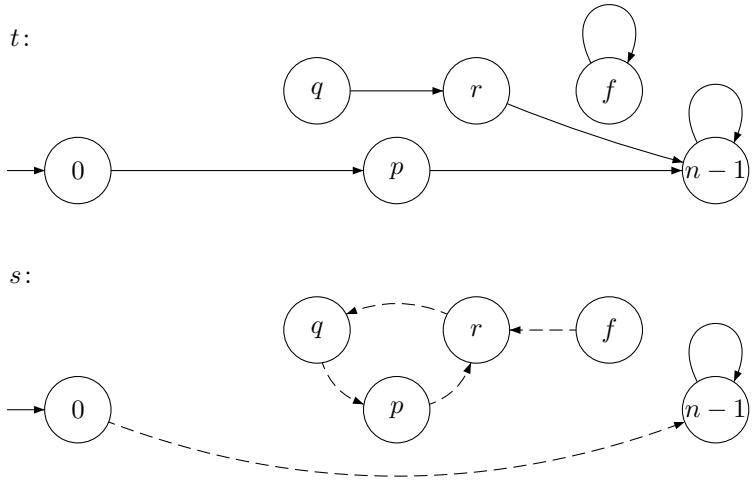

Figure 9: Case 9 in the proof of Theorem 1.

Observe that s has the following properties:

(a) $\{p, f\}$ is the only colliding pair that is focused by s, and the state to which it is focused lies on a cycle of length 3 and is not the minimal state in the cycle.

(b) s has a unique 3-cycle.

From (a), $s \notin T(n)$ and so s is different from the transformations of Case 1. The transformations of Case 2 contain a cycle, but (Property (a) of

Case 2) the only colliding pair focused by them to a state lying on a cycle is focused to the minimal state appearing in a cycle, in contrast to s from this case. Since s has a cycle, it is different from the transformations of Cases 3–7. Also, s has a unique 3-cycle in contrast with the transformations of Case 8, which have a unique 2-cycle (Property (b) of Case 8).

Again, s uniquely defines t: The orbit with the focused colliding pair contains precisely p, q, r and f, and they are uniquely determined since f is not in the 3-cycle and is mapped to r.

Case 10: t does not fit in any of the previous cases, and there is a state $q \in Q_M$ that is not a fixed point and satisfies $qt \neq n-1$, and a fixed point $f \in Q_M$.
Let $r = qt$; then $rt = n-1$ since this is not Case 5. Now, in contrast to the previous case, we have $p > r$.

Let s be the transformation illustrated in Fig. 10 and defined by

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

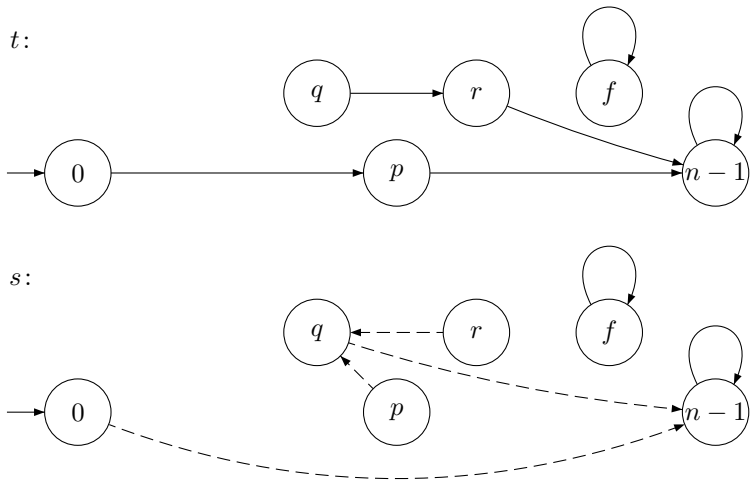

Figure 10: Case 10 in the proof of Theorem 1.

Observe that s has the following properties:

(a) $\{p, r\}$ is the only colliding pair that is focused by s.

(b) s does not have cycles, and each state $\in Q \setminus \{p, r, f\}$ is mapped to $n-1$.

(c) s has the fixed point $f \neq n-1$.

From (a), $s \notin T(n)$ and so s is different from the transformations of Case 1. Since the transformations of Cases 2, 8, and 9 contain cycles, they are different from s. Here the unique focused colliding pair is not mapped to a fixed point, in contrast with the transformations of Cases 3 and 4. Since both states from the pair have in-degree 0 in s, s is different from the transformations of Case 5. For a distinction with Case 6, observe that the smallest i such that $ps^i = n-1$ is 2, in contrast with 3 (Property (b) of Case 6). For a distinction with Case 7, observe that besides the focused colliding states p and r, there is no state q' that is not a fixed point in s and satisfies $q's \neq n-1$, whereas in the transformations of Case 7 q_2 is such a state.

Again, s uniquely defines t: Here $\{p, r\}$ is the only focused colliding pair, and p is distinguished as the larger state.

Case 11: t does not fit in any of the previous cases, and there is a state $q \in Q_M$ 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 $n-1$.

We consider the following two sub-cases:
Sub-case (i): $p < r$.
Let s be the transformation illustrated in Fig. 11 and 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 also illustrated in Fig. 11 and defined by

$$0s = n-1, \quad ps = q, \quad rs = q, \quad qs = n-1, \quad r_1s = r_2, \quad r_2s = r_1,$$
$$qs = qt \text{ for the other states } q \in Q.$$

Observe that s has the following properties:

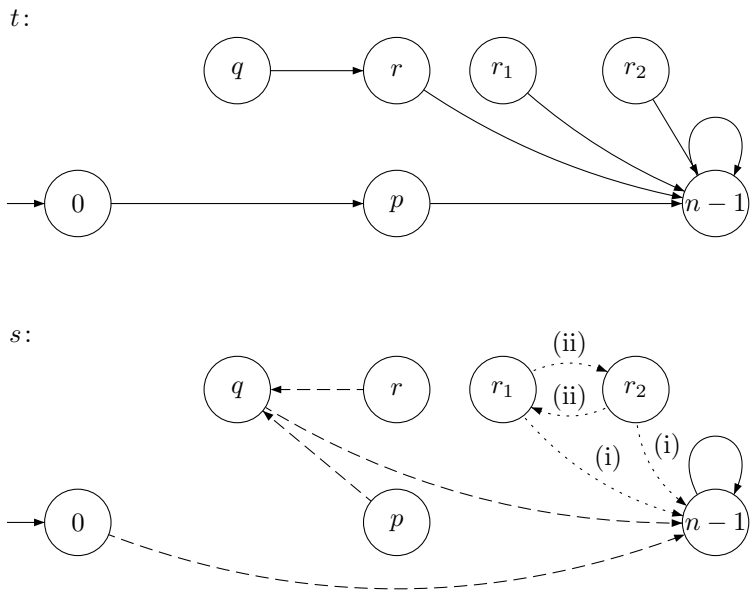

Figure 11: Case 11 in the proof of Theorem 1.

(a) $\{p, r\}$ is the only colliding pair that is focused by s, and it is focused to state q with $qs = n - 1$.

(b) s does not have any fixed points other than $n - 1$, and in (i) s does not have any cycles, whereas in (ii) s has a cycle but no colliding pair from the orbit of the cycle is focused.

From (a), $s \notin T(n)$ and so s is different from the transformations of Case 1. The transformations of Cases 2, 8, and 9 contain cycles whose orbits contain a focused colliding pair. Here s in (i) does not have a cycle, and in (ii) has a 2-cycle but the unique colliding pair is not in its orbit. Also, s does not have a fixed point from Q_M, in contrast with the transformations of Cases 3, 4, and 10. Since both states from the colliding pair have in-degree 0, s is different from the transformations of Case 5. Since the smallest i such that $ps^i = n - 1$ is 2, s is different from the transformations of Case 6, where the smallest such i is 3 (Property (b) of Case 6). For a distinction with Case 7, observe that besides the colliding states p and r, there is no state q' that is not a fixed point and satisfies $q's \neq n - 1$, whereas in the transformation of Case 7 q_2 is such a state.

Again, s uniquely defines t: Here $\{p, r\}$ is the only focused colliding pair, and if we have a 2-cycle, then p is distinguished as the smaller state; otherwise p is the larger one from the pair.

Case 12: t does not fit in any of the previous cases.
Here t must contain exactly one fixed point $f \in Q_M$, and every state from $Q_M \setminus \{f\}$ is mapped to $n-1$. If all states from Q_M would be mapped to $n-1$, then by Proposition 2, t would be in $\mathbf{T}^{\geqslant 6}(n)$ 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 illustrated in Fig. 12 and defined by

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

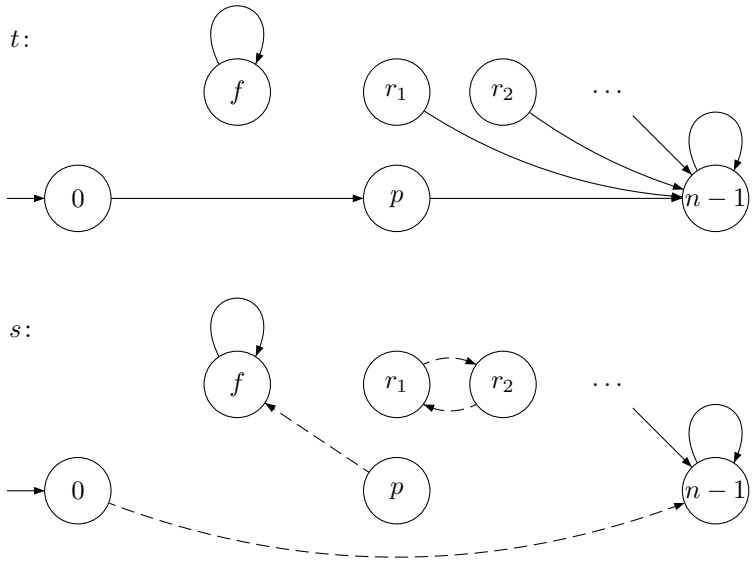

Figure 12: Case 12 in the proof of Theorem 1.

Observe that s has the following properties:

(a) $\{p, f\}$ is the only colliding pair that is focused by s, and it is focused to the fixed point f.

(b) s has a 2-cycle, but no colliding pair from the orbit of the cycle is focused.

From (a), $s \notin T(n)$ and so s is different from the transformations of Case 1. Here s has a cycle but the colliding pair is not from the orbit of the cycle, in contrast to Case 2. The transformations of Cases 3–10 do not have a cycle whose orbit has no focused colliding pairs. The transformations of Case 11 have such a cycle, but the orbit with the focused colliding pair is the orbit of fixed point $n-1$, and here it is the orbit of $f \neq n-1$.

Again, s uniquely defines t: Here $\{p, f\}$ is the only focused colliding pair, and p is distinguished from f as it is not a fixed point. Hence we can define $0t = p$, $pt = n-1$, $r_1t = n-1$, and $r_2t = n-1$. □

4. Uniqueness of maximal witness

Our third contribution is a proof that the transition semigroup $T(n)$ of a minimal 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 of states, then $T(n)$ is a subsemigroup of $\mathbf{T}^{\geqslant 6}(n)$ and $|T(n)| \leqslant (n-1)^{n-2} + n - 2$.*

Proof. Consider an arbitrary transformation $t \in T(n)$ and let $p = 0t$. If $p = n-1$, then any state other than 0 and $n-1$ can possibly be mapped by t to any one of the $n-1$ states other than 0 (0 is not possible in view of Lemma 1); hence there are at most $(n-1)^{n-2}$ such transformations. All of these transformations are in $\mathbf{T}^{\geqslant 6}(n)$ by Proposition 2.

If $p \neq n-1$, then $qt = n-1$ for any $q \neq 0$, because there are no colliding pairs. Thus $t = (Q \setminus \{0\} \to n-1)(0 \to p)$, and all of $n-2$ such transformations are in $\mathbf{T}^{\geqslant 6}(n)$ by Proposition 2. It follows that $T(n)$ is a subsemigroup of $\mathbf{T}^{\geqslant 6}(n)$ and has size at most $(n-1)^{n-2} + n - 2$. □

Lemma 3. *If $n \geqslant 7$ and $\mathcal{D}_n$ has at least one colliding pair of states, 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:

$$0s = n-1, \quad ps = r, \quad rs = r, \quad r_1s = r_2, \quad r_2s = r_3, \quad r_3s = r_1,$$
$$qs = qt \text{ for the other states } q \in Q.$$

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 $s \in \mathbf{T}^{\geqslant 6}(n)$, and so $\varphi(T(n)) \subsetneq \mathbf{T}^{\geqslant 6}(n)$. Since φ is injective, it follows that $|T(n)| < |\mathbf{T}^{\geqslant 6}(n)| = (n-1)^{n-2} + n - 2$. □

Corollary 1. *For $n \geqslant 7$, the maximal transition semigroups of minimal DFAs $(Q, \Sigma, \delta, 0, F)$ 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 $n-1$. Recall that Q_M is 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 2 and Corollary 1 all transformations of Q_M must be present, 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 2, and so $qt_a = n-1$ for any $q \in M$. So t_a cannot be a generator of a transformation of Q_M. Hence we must have $0t_a = n-1$, and also $0t_b = 0t_c = n-1$.

So far, the states in Q_M are not reachable from 0; hence there must be a letter, say e, such that $0t_e = p$ is in Q_M. This must be a transformation of type (b) from Proposition 2, and all the states of Q_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 Q_M to $n-1$. The letter e will not do, since it maps *all* states of Q_M to $n-1$. Hence we require a fifth letter, say d. □

4.1. Uniqueness of small semigroups

Here we consider maximal transition semigroups of DFAs having six or fewer states and accepting suffix-free languages. Recall that every transformation in a transition semigroup of a minimal DFA that has set Q of

cardinality n, initial state 0, and accepts a suffix-free language, must be a subset of $\mathbf{B}_{\mathrm{sf}}(n)$.

There is only one transformation in $\mathbf{B}_{\mathrm{sf}}(2)$, and three in $\mathbf{B}_{\mathrm{sf}}(3)$. Since $\mathbf{B}_{\mathrm{sf}}(2)$ and $\mathbf{B}_{\mathrm{sf}}(3)$ are semigroups, the maximal semigroups for $n = 2$ and $n = 3$ have 1 and 3 elements, respectively, and are unique. From [4] it is known that $\mathbf{T}^{\leqslant 5}(n) = \mathbf{T}^{\geqslant 6}(n) = \mathbf{B}_{\mathrm{sf}}(n)$ if $n \in \{2, 3\}$.

From [4] it is also known that $\mathbf{T}^{\leqslant 5}(n)$ is largest for $n \in \{4, 5\}$ and that $\mathbf{T}^{\geqslant 6}(n)$ is largest for $n = 6$. We now prove that $\mathbf{T}^{\leqslant 5}(4)$, $\mathbf{T}^{\leqslant 5}(5)$, and $\mathbf{T}^{\geqslant 6}(6)$ are the unique largest semigroups for those values of n.

We say that transformations t and t' in $\mathbf{B}_{\mathrm{sf}}(n)$ *conflict* if, whenever t and t' belong to the transition semigroup of a DFA $\mathcal{D}$, then one of the following conditions holds:

1. The language accepted by $\mathcal{D}$ is not suffix-free.
2. Every two states from Q_M are colliding.
3. Every two states from Q_M are focused.

Lemma 4. *In a largest transition semigroup X_n of a minimal DFA $\mathcal{D}_n = (Q, \Sigma, \delta, 0, F)$ of a suffix-free language there are no conflicting pairs of transformations, unless $X_n = \mathbf{T}^{\leqslant 5}(n)$ or $X_n = \mathbf{T}^{\geqslant 6}(n)$.*

Proof. Obviously, there are no conflicting pair of transformations because of (1). If a pair of transformations conflicts because of (2), then from Proposition 1 a largest transition semigroup must be $\mathbf{T}^{\leqslant 5}(n)$. If a pair of transformations conflicts because of (3), then no pair of states is colliding, since it is focused; from Proposition 3 a largest transition semigroup must be $\mathbf{T}^{\geqslant 6}(n)$. □

A transformation t of Q is called *semiconstant* if it maps a non-empty subset $S \subseteq Q$ to a single state $q \in Q$, and fixes $Q \setminus S$. We denote it by $(S \to q)$.

Lemma 5. *In the transition semigroup of a minimal DFA $\mathcal{D}_n = (Q, \Sigma, \delta, 0, F)$ of a suffix-free language L all semiconstant transformations $t = (S \to q)$ are such that $0 \in S$ and $q = n - 1$. Moreover, every such transformation is present if the transition semigroup is maximal.*

Proof. If $0 \notin S$, then $0t = 0$; this implies that t is not in $\mathbf{B}_{\mathrm{sf}}(n)$, contradicting our assumption that L is suffix-free.

If $0 \in S$ and $q \neq n-1$, then $0t = q$ and q is a fixed point in Q_M. Since q is non-empty, it accepts some word x; then tx and ttx are both in L contradicting that L is suffix-free.

Thus we must have $0 \in S$ and $q = n-1$. We now argue that every such transformation t must be present in a maximal semigroup by showing that t can always be added if not present. Suppose that the addition of a letter a that induces t results in a DFA that does not accept a suffix-free language. Then there must be two words v and uv in the language, such that either u or v contains a; let u and v be such that uv is a shortest such word. Suppose $v = v_1 a v_2$; since t maps some states to $n-1$ and fixes all the others and v is accepted, a must map $\delta(0, v_1)$ to itself, and hence can be removed, leaving $v' = v_1 v_2$ in L. Similarly, if $u = u_1 a u_2$ we can remove a, obtaining $u' = u_1 u_2$. Then $u'v'$ and v' are in L, and $u'v'$ is shorter than uv—a contradiction. □

Theorem 3. *For $n = 6$, the maximal transition semigroup of minimal DFAs $\mathcal{D}_n = (Q, \Sigma, \delta, 0, F)$ of suffix-free languages is $\mathbf{T}^{\geqslant 6}(6)$ and it is unique. For $n \in \{4, 5\}$, the maximal transition semigroup of minimal DFAs $\mathcal{D}_n = (Q, \Sigma, \delta, 0, F)$ of suffix-free languages is $\mathbf{T}^{\leqslant 5}(n)$ and it is unique.*

Proof. We have verified this with the help of computation. We used the idea of conflicting pairs of transformations from [4, Theorem 20], and we enumerated non-isomorphic DFAs using the approach of [12]. For $n = 6$ there are $|\mathbf{B}_{\mathrm{sf}}(6)| = 1169$ transformations that can be present in the transition semigroup of a DFA of a suffix-free language; so it is not possible to check all maximal subsets of $\mathbf{B}_{\mathrm{sf}}(6)$ in a naive way. Our method is as follows.

By a *semiautomaton* we mean a triple (Q, Σ, δ), where Q, Σ, and δ are defined as in a DFA. Say that a semiautomaton $\mathcal{D}' = (Q, \Sigma', \delta')$ is an *extension* of a semiautomaton $\mathcal{D} = (Q, \Sigma, \delta)$ if it can be obtained from $\mathcal{D}$ by adding letters, that is, if $\Sigma \subset \Sigma'$ and $\delta' \subset \delta$. The same concept can be extended to DFAs. We start from the set A_1 of all non-isomorphic unary semiautomata with n-states. Given a set A_i of semiautomata over an i-ary alphabet, using [12] we generate all of their non-isomorphic extensions A_{i+1} over an $(i+1)$-ary alphabet. To reduce the number of generated semiautomata and make the whole computation possible, we check for every semiautomaton whether there may exist an extension of it such that after adding initial and final states, the extension accepts a suffix-free language and its transition semigroup is a largest one but different from $\mathbf{T}^{\geqslant 6}$ (if $n = 6$) and $\mathbf{T}^{\leqslant 5}$ (if $n \leqslant 5$). Also, we check whether the transition semigroup of the semiautomaton is irreducibly generated by the transitions of the letters, that is,

all these transitions are required to generate the semigroup. Note that if a DFA $\mathcal{D}$ has a transition semigroup whose generating set can be reduced, then there exists a DFA $\mathcal{D}'$ over a smaller alphabet that has the same semigroup; moreover, $\mathcal{D}'$ recognizes a suffix-free language if and only if $\mathcal{D}$ does.

For every generated semiautomaton we test all selections of the initial state and the empty state, relabel them to 0 and $n-1$, and check the resulting DFAs. If all the DFAs are rejected, then we skip the semiautomaton. First, we check if the DFA accepts a suffix-free language (see [5] for testing). Then, we compute a rough bound on the maximal size of the transition semigroups that are different from $\mathbf{T}^{\geqslant 6}$ and $\mathbf{T}^{\leqslant 5}$ of extensions of the DFA. If this is smaller than the syntactic complexity, we reject the DFA. Extending the idea from [4, Theorem 20], we compute the bound in the following way:

1. We compute the transition semigroup X_n of the DFA.
2. For every transformation from $t \in B_{sf} \setminus X_n$ we check whether adding t as a generator results in a DFA that accepts a suffix-free language, and neither all pairs of states are colliding nor all are focused. Otherwise, by Lemma 4 we can omit it. Let Y_n be the set of allowed transformations. Note that $|X_n| + |Y_n|$ gives a rough bound for the size of the maximal transition semigroups.
3. We compute a matching M in the graph of conflicts of transformations induced by Y_n: Two transformations $t, t' \in Y_n$ can be matched if they conflict. We compute it by a simple greedy algorithm in $O(|Y_n|^2)$ time.
4. We finally use $|X_n| + |Y_n| - |M|$ as the bound.

Finally, by Lemma 5 we remove semiconstant transformations from the set A_1 of transformations that is used to generate non-isomorphic semiautomata. Then we always add all allowed semicontant transformations to X_n in step (1).

Using this method, for $n = 6$ we had to verify semiautomata only up to 9 letters, and the computation took a few minutes. □

5. Conclusions

We have shown that the upper bound on the syntactic complexity of suffix-free languages is $(n-1)^{n-2} + n - 2$ for $n \geqslant 6$. 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 for every n.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant No. OGP000087, and by the National Science Centre, Poland under project number 2014/15/B/ST6/00615.

References

[1] T. Ang, J. Brzozowski, Languages convex with respect to binary relations, and their closure properties, Acta Cybernet. 19 (2009) 445–464.

[2] J. Berstel, D. Perrin, C. Reutenauer, Codes and Automata, Cambridge University Press, 2009.

[3] J. Brzozowski, Quotient complexity of regular languages, J. Autom. Lang. Comb. 15 (2010) 71–89.

[4] J. Brzozowski, B. Li, Y. Ye, Syntactic complexity of prefix-, suffix-, bifix-, and factor-free regular langauges, Theoret. Comput. Sci. 449 (2012) 37–53.

[5] J. Brzozowski, J. Shallit, Z. Xu, Decision problems for convex languages, Information and Computation 209 (2011) 353–367.

[6] J. Brzozowski, M. Szykuła, Complexity of suffix-free regular languages, in: A. Kosowski, I. Walukiewicz (Eds.), FCT 2015, volume 9210 of *LNCS*, Springer, 2015, pp. 146–159. Full version at `http://arxiv.org/abs/1504.05159`.

[7] J. Brzozowski, M. Szykuła, Upper bound on syntactic complexity of suffix-free languages, in: J. Shallit (Ed.), DCFS 2015, volume 9118 of *LNCS*, Springer, 2015, pp. 33–45.

[8] J. Brzozowski, Y. Ye, Syntactic complexity of ideal and closed languages, in: G. Mauri, A. Leporati (Eds.), DLT 2011, volume 6795 of *LNCS*, Springer, 2011, pp. 117–128.

[9] O. Ganyushkin, V. Mazorchuk, Classical Finite Transformation Semigroups: An Introduction, Springer, 2009.

[10] Y.S. Han, K. Salomaa, State complexity of basic operations on suffix-free regular languages, Theoret. Comput. Sci. 410 (2009) 2537–2548.

[11] M. Holzer, B. König, On deterministic finite automata and syntactic monoid size, Theoret. Comput. Sci. 327 (2004) 319–347.

[12] A. Kisielewicz, M. Szykuła, Generating small automata and the Černý conjecture, in: S. Konstantinidis (Ed.), CIAA 2013, volume 7982 of *LNCS*, Springer, 2013, pp. 340–348.

[13] B. Krawetz, J. Lawrence, J. Shallit, State complexity and the monoid of transformations of a finite set, in: M. Domaratzki, A. Okhotin, K. Salomaa, S. Yu (Eds.), Proceedings of the Implementation and Application of Automata, (CIAA), volume 3317 of *LNCS*, Springer, 2005, pp. 213–224.

[14] J.E. Pin, Syntactic semigroups, in: Handbook of Formal Languages, vol. 1: Word, Language, Grammar, Springer, New York, NY, USA, 1997, pp. 679–746.

[15] G. Thierrin, Convex languages, in: M. Nivat (Ed.), Automata, Languages and Programming, North-Holland, 1973, pp. 481–492.

[16] S. Yu, State complexity of regular languages, J. Autom. Lang. Comb. 6 (2001) 221–234.

Appendix: List of the cases in the proof of Theorem 1

Every $t \in T(n)$ fits in the first case whose conditions are met by t.

Case 1: $t \in \mathbf{T}^{\geqslant 6}(n)$, that is, $0t = p = n-1$ or $Q_M t = \{n-1\}$.
Case 2: t has a cycle.
Case 3: $(0t)t = pt \neq n-1$.
Case 4: There is a fixed point $r \in Q_M$ with in-degree $\geqslant 2$.
Case 5: There is a state r with in-degree $\geqslant 1$ such that $rt \notin \{r, n-1\}$.
Case 6: There is a state $r \in Q_M$ with in-degree $\geqslant 2$.
Case 7: There are two states $q_1, q_2 \in Q_M$ that satisfy $q_1 t \notin \{q_1, n-1\}$ and $q_2 t \notin \{q_2, n-1\}$.
Case 8: There are two fixed points r_1 and r_2 in Q_M with in-degree 1.
Case 9: There is a state $q \in Q_M$ that satisfies $qt \notin \{q, n-1\}$ and $p < qt$, and a fixed point $f \in Q_M$.
Case 10: There is a state $q \in Q_M$ that satisfies $qt \notin \{q, n-1\}$, and a fixed point $f \in Q_M$.
Case 11: There is a state $q \in Q_M$ that satisfies $qt \notin \{q, n-1\}$.
Case 12: t is any other transformation.

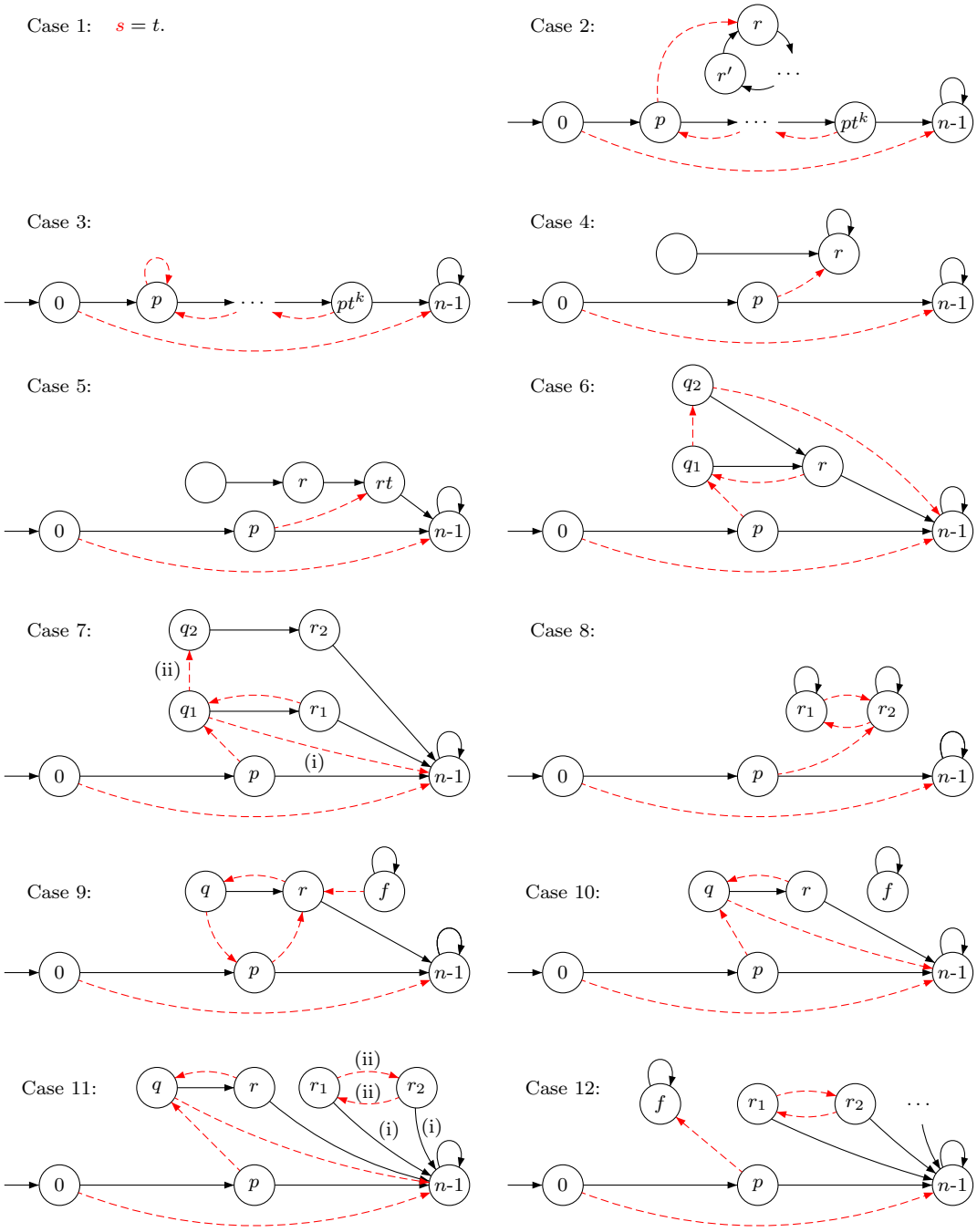

Figure 13: Map of the cases from the proof of Theorem 1. The transitions of t are represented by solid lines, and the transitions of s by dashed red lines.