Upper Bounds on Syntactic Complexity of Left and Two-Sided Ideals*

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 two open problems concerning syntactic complexity. We prove that the cardinality of the syntactic semigroup of a left ideal or a suffix-closed language with n left quotients (that is, with state complexity n) is at most $n^{n-1} + n - 1$, and that of a two-sided ideal or a factor-closed language is at most $n^{n-2} + (n-2)2^{n-2} + 1$. Since these bounds are known to be reachable, this settles the problems.

Keywords: factor-closed, left ideal, regular language, suffix-closed, syntactic complexity, transition semigroup, two-sided ideal, upper bound.

1 Introduction

The *syntactic complexity* [4] of a regular language is the size of its syntactic semigroup [5]. The *transition semigroup* T of a deterministic finite automaton (DFA) $\mathcal{D}$ is the semigroup of transformations of the state set of $\mathcal{D}$ generated by the transformations induced by the input letters of $\mathcal{D}$. The transition semigroup of a minimal DFA of a language L is isomorphic to the syntactic semigroup of L [5]; hence syntactic complexity is equal to the cardinality of T.

The number n of states of $\mathcal{D}$ is known as the *state complexity* of the language [1,6], and it is the same as the 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 n.

A *right ideal* (respectively, *left ideal*, *two-sided ideal*) is a non-empty language L over an alphabet Σ such that $L = L\Sigma^*$ (respectively, $L = \Sigma^* L$, $L = \Sigma^* L \Sigma^*$). We are interested only in regular ideals; for reasons why they deserve to be studied see [2, Section 1]. Ideals appear in pattern matching. For example, if a *text* is a word w over some alphabet Σ, and a *pattern* is an arbitrary language L over Σ, then an occurrence of a pattern represented by L in text w is a triple (u, x, v) such that $w = uxv$ and x is in L. Searching text w for words in L is

* 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.

A.M. Shur and M.V. Volkov (Eds.): DLT 2014, LNCS 8633, pp. 13–24, 2014.

equivalent to looking for prefixes of w that belong to the language $\Sigma^* L$, which is the left ideal generated by L.

The syntactic complexity of right ideals was proved to be n^{n-1} in [4]. The syntactic complexities of left and two-sided ideals were also examined in [4], where it was shown that $n^{n-1}+n-1$ and $n^{n-2}+(n-2)2^{n-2}$, respectively, are lower bounds on these complexities, and it was conjectured that they are also upper bounds. In this paper we prove these conjectures.

If $w = uxv$ for some $u, v, x \in \Sigma^*$, then 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 *suffix-closed* (respectively, *factor-closed*) if $w \in L$ implies that every suffix (respectively, factor) of w is also in L. We are interested only in regular suffix- and factor-closed languages. Since every left (respectively, two-sided) ideal is the complement of a suffix-closed (respectively, factor-closed) language, and syntactic complexity is preserved by complementation, our theorems also apply to suffix- and factor-closed languages, but our proofs are given for left and two-sided ideals only.

2 Preliminaries

The *left quotient* or simply *quotient* of a regular language L by a word w is denoted by Lw and defined by $Lw = \{x \mid wx \in L\}$. A language is regular if and only if it has a finite number of quotients. The number of quotients of L is called its *quotient complexity*. 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 Lw_i, where $w_i \in \Sigma^*$ is such that $Lw_i = K_i$.

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.

The *quotient DFA* of a regular language L with n quotients is defined by $\mathcal{D} = (K, \Sigma, \delta, K_0, F)$, where $\delta(K_i, w) = K_j$ if and only if $K_i w = K_j$, and $F = \{K_i \mid \varepsilon \in K_i\}$. To simplify the notation, we use the set $Q = \{0, \ldots, n-1\}$ of subscripts of quotients to denote the states of $\mathcal{D}$; then $\mathcal{D}$ is denoted by $\mathcal{D} = (Q, \Sigma, \delta, 0, F)$. The quotient corresponding to $q \in Q$ is then $K_q = \{w \mid \delta(q, w) \in F\}$. 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 $a \in \Sigma$ defines 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 *identity* transformation $\mathbf{1}$ maps each element to itself. For $k \geq 2$, 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 acts like a k-cycle on 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 $k \geq 2$. A 2-cycle (q_0, q_1) is called a *transposition*. A transformation is *constant* if it maps all states to a single state q; it is denoted by $(Q \to q)$. If w is a word of Σ^*, the fact that w induces transformation t is denoted by $w\colon t$. A transformation mapping i to q_i for $i = 0, \ldots, n-1$ is sometimes denoted by $[q_0, \ldots, q_{n-1}]$.

3 Left Ideals

3.1 Basic Properties

Let $Q = \{0, \ldots, n-1\}$, let $\mathcal{D}_n = (Q, \Sigma_{\mathcal{D}}, \delta_{\mathcal{D}}, 0, F)$ be a minimal DFA, and let T_n be its transition semigroup. Consider the sequence $(0, 0t, 0t^2, \ldots)$ of states obtained by applying transformation $t \in T_n$ repeatedly, starting with the initial state. Since Q is finite, there must eventually be a repeated state, that is, there must exist i and 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. If the period is 1, t is said to be *initially aperiodic*; then the sequence is $0, 0t, \ldots, 0t^{j-1} = 0t^j$.

Lemma 1 ([4]). *If $\mathcal{D}_n$ is a DFA of a left ideal, all the transformations in T_n are initially aperiodic, and no state of $\mathcal{D}_n$ is empty.*

Remark 1 ([2]). A language $L \subseteq \Sigma^*$ is a left ideal if and only if for all $x, y \in \Sigma^*$, $Ly \subseteq Lxy$. Hence, if $Lx \neq L$, then $L \subset Lx$ for any $x \in \Sigma^+$.

It is useful to restate this observation it terms of the states of $\mathcal{D}_n$. For DFA $\mathcal{D}_n$ and states $p, q \in Q$, we write $p \prec q$ if $K_p \subset K_q$.

Remark 2. A DFA $\mathcal{D}_n$ is a minimal DFA of a left ideal if and only if for all $s, t \in T_n \cup \{\mathbf{1}\}$, $0t \preceq 0st$. If $0t \neq 0$, then $0 \prec 0t$ for any $t \in T_n$. Also, if $r \in Q$ has a t-predecessor, that is, if there exists $q \in Q$ such that $qt = r$, then $0t \preceq r$. (This follows because $q = 0s$ for some transformation s since q is reachable from 0; hence $0 \preceq q$ and $0t \preceq qt = r$.) In particular, if r appears in a cycle of t or is a fixed point of t, then $0t \preceq r$.

We consider chains of the form $K_{i_1} \subset K_{i_2} \subset \cdots \subset K_{i_h}$, where the K_{i_j} are quotients of L. If L is a left ideal, the smallest element of any maximal-length chain is always L. Alternatively, we consider chains of states starting from 0 and strictly ordered by $\prec$.

Proposition 1. *For $t \in T_n$ and $p, q \in Q$, $p \prec q$ implies $pt \preceq qt$. If $p \prec pt$, then $p \prec pt \prec \cdots \prec pt^k = pt^{k+1}$ for some $k \geq 1$. Similarly, $p \succ q$ implies $pt \succeq qt$, and $p \succ pt$ implies $p \succ pt \succ \cdots \succ pt^k = pt^{k+1}$ for some $k \geq 1$.*

It was proved in [4, Theorem 4, p. 124] that the transition semigroup of the following DFA of a left ideal meets the bound $n^{n-1} + n - 1$.

Definition 1 (Witness: Left Ideals). *For $n \geq 3$, we define the DFA $\mathcal{W}_n = (Q, \Sigma_{\mathcal{W}}, \delta_{\mathcal{W}}, 0, \{n-1\})$, where $Q = \{0, \ldots, n-1\}$, $\Sigma_{\mathcal{W}} = \{a, b, c, d, e\}$, and $\delta_{\mathcal{W}}$ is defined by $a\colon (1, \ldots, n-1)$, $b\colon (1, 2)$, $c\colon (n-1 \to 1)$, $d\colon (n-1 \to 0)$, and $e\colon (Q \to 1)$. For $n = 3$, a and b coincide, and we can use $\Sigma = \{b, c, d, e\}$.*

Remark 3. In $\mathcal{W}_n$, the transformations induced by a, b, and c restricted to $Q\backslash\{0\}$ generate all the transformations of the last $n-1$ states. Together with the transformation of d, they generate all transformations of Q that fix 0. To see this, consider any transformation t that fixes 0. If some states from $\{1,\ldots,n-1\}$ are mapped to 0 by t, we can map them first to $n-1$ and $n-1$ to one of them by the transformations of a, b, and c, and then map $n-1$ to 0 by the transformation of d. Also the words of the form ea^i for $i \in \{0,\ldots,n-2\}$ induce constant transformations $(Q \to i+1)$. Hence the transition semigroup of $\mathcal{W}_n$ contains all the constant transformations.

Example 1. One verifies that the maximal-length chains of quotients in $\mathcal{W}_n$ have length 2. On the other hand, for $n \geq 2$, let $\Sigma = \{a,b\}$ and let $L = \Sigma^* a^{n-1}$. Then L has n quotients and the maximal-length chains are of length n.

3.2 Upper Bound

Our main result of this section shows that the lower bound $n^{n-1}+n-1$ is also an upper bound. Our approach is as follows: We consider a minimal DFA $\mathcal{D}_n = (Q, \Sigma_{\mathcal{D}}, \delta_{\mathcal{D}}, 0, F)$, where $Q = \{0,\ldots,n-1\}$, of an arbitrary left ideal with n quotients and let T_n be the transition semigroup of $\mathcal{D}_n$. We also deal with the witness DFA $\mathcal{W}_n = (Q, \Sigma_{\mathcal{W}}, \delta_{\mathcal{W}}, 0, \{n-1\})$ 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 $f\colon T_n \to S_n$, and this will prove that $|T_n| \leq |S_n|$.

Remark 4. If $n = 1$, the only left ideal is Σ^* and the transition semigroup of its minimal DFA satisfies the bound $1^0 + 1 - 1 = 1$. If $n = 2$, there are only three allowed transformations, since the transposition $(0,1)$ is not initially aperiodic and so is ruled out by Lemma 1. Thus the bound $2^1 + 2 - 1 = 3$ holds.

Lemma 2. *If $n \geq 3$ and a maximal-length chain in $\mathcal{D}_n$ strictly ordered by $\prec$ has length 2, then $|T_n| \leq n^{n-1}+n-1$ and T_n is a subsemigroup of S_n.*

Proof. Consider an arbitrary transformation $t \in T_n$ and let $p = 0t$. If $p = 0$, then any state other than 0 can possibly be mapped by t to any one of the n states; hence there are at most n^{n-1} such transformations. All of these transformations are in S_n by Remark 3.

If $p \neq 0$, then $0 \prec p$. Consider any state $q \notin \{0,p\}$; by Remark 2, $p \preceq qt$. If $p \neq qt$, then $p \prec qt$. But then we have the chain $0 \prec p \prec qt$ of length 3, contradicting our assumption. Hence we must have $p = qt$, and so t is the constant transformation $t = (Q \to p)$. Since p can be any one of the $n-1$ states other than 0, we have at most $n-1$ such transformations. Since all of these transformations are in S_n by Remark 3, T_n is a subsemigroup of S_n. □

Theorem 1 (Left Ideals, Suffix-Closed Languages). *If $n \geq 3$ and L is a left ideal or a suffix-closed language with n quotients, then its syntactic complexity is less than or equal to $n^{n-1}+n-1$.*

Proof. It suffices to prove the result for left ideals. For a transformation $t \in T_n$, consider the following cases:

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

Case 2: $t \notin S_n$ and $0t^2 \neq 0t$.
Note that $t \notin S_n$ implies $0t \neq 0$ by Remark 3. Let $0t = p$. We have $p = 0t \prec 0tt = pt$ by Remark 2. Let $p \prec \cdots \prec pt^k = pt^{k+1}$ be the chain defined from p; this chain is of length at least 2. Let $f(t) = s$, where s is the transformation defined by

$$0s = 0, \quad pt^k s = p, \quad qs = qt \text{ for the other states } q \in Q.$$

Transformation s is shown in Figure 1, where the dashed transitions show how s differs from t.

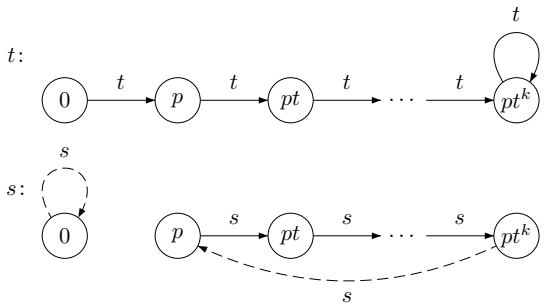

Fig. 1. Case 2 in the proof of Theorem 1

By Remark 3, $s \in S_n$. However, $s \notin T_n$, as it contains the cycle $(p, \ldots, pt^k)$ with states strictly ordered by $\prec$ in DFA $\mathcal{D}_n$, which contradicts Proposition 1. Since $s \notin T_n$, it is distinct from the transformations defined in Case 1.

In going from t to s, we have added one transition ($0s = 0$) that is a fixed point, and one ($pt^k s = p$) that is not. Since only one non-fixed-point transition has been added, there can be only one cycle in s with states strictly ordered by $\prec$. Since 0 can't appear in this cycle, p is its smallest element with respect to $\prec$.

Suppose now that $t' \neq t$ is another transformation that satisfies Case 2, that is, $0t' = p' \neq 0$ and $p't' \neq p'$; we shall show that $f(t) \neq f(t')$. Define s' for t' as s was defined for t. For a contradiction, assume $s = f(t) = f(t') = s'$.

Like s, s' contains only one cycle strictly ordered by $\prec$, and p' is its smallest element. Since we have assumed that $s = s'$, we must have $p = 0t = 0t' = p'$ and the cycles in s and s' must be identical. In particular, $pt^k t = pt^k = p(t')^k t' = p(t')^k$. For q of $Q \setminus \{0, pt^k\}$, we have $qt = qs = qs' = qt'$. Hence $t = t'$—a contradiction. Therefore $t \neq t'$ implies $f(t) \neq f(t')$.

Case 3: $t \notin S_n$ and $0t^2 = 0t$.
As before, let $0t = p$. Consider any state $q \notin \{0, p\}$; then $0 \prec q$ by Remark 2 and $0t \preceq qt$ by Proposition 1. Thus either $p \prec qt$, or $p = qt$. We consider the following sub-cases:

• **(a):** t has a cycle.

Since t has a cycle, take a state r from the cycle; then r and rt are not comparable under $\preceq$ by Proposition 1, and $p \prec r$ by Remark 2. Let $f(t) = s$, where s is the transformation shown in Figure 2 and defined by

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

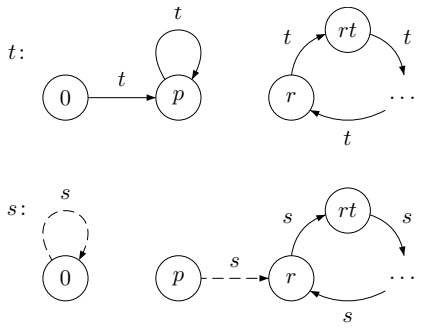

Fig. 2. Case 3(a) in the proof of Theorem 1

By Remark 3, $s \in S_n$. Suppose that $s \in T_n$; since $p \prec r$, we have $r = ps \preceq rs = rt$ by the definition of s and Proposition 1; this contradicts that r and rt are not comparable. Hence $s \notin T_n$, and so s is distinct from the transformations of Case 1.

We claim that p is not in a cycle of s; this cycle would have to be

$$p \xrightarrow{s} r \xrightarrow{s} rt \xrightarrow{s} \cdots \xrightarrow{s} rt^{k-1} \xrightarrow{s} p, \text{ that is, } p \xrightarrow{s} r \xrightarrow{t} rt \xrightarrow{t} \cdots \xrightarrow{t} rt^{k-1} \xrightarrow{t} p,$$

for some $k \geq 2$ because $r \neq p = pt$ and $rt \neq p$. Since $p \prec r$ we have $p \prec rt$; but then we have a chain $p \prec rt \prec \cdots \prec rt^k = p$, contradicting Proposition 1.

Since p is not in a cycle of s, it follows that s does not contain a cycle with states strictly ordered by $\prec$, as such a cycle would also be in t. So s is distinct from the transformations of Case 2.

We claim there is a unique state q such that (a) $0 \prec q \prec qs$, (b) $qs \not\preceq qs^2$. First we show that p satisfies these conditions: (a) holds because $ps = r$ and $p \prec r$; (b) holds because $ps = r$, $ps^2 = rt$ and r and rt are not comparable. Now suppose that q satisfies the two conditions, but $q \neq p$. Note that $qs \neq p$, because $qs = p$ implies $qs = p \prec r = qs^2$, contradicting (b). Since $q, qs \notin \{0, p\}$, we have $qt = qs \not\preceq qs^2 = qt^2$. But Proposition 1 for $q \prec qt$ implies that $qt \preceq qt^2$—a contradiction. Thus p is the only state satisfying these conditions.

If $t' \neq t$ is another transformation satisfying the conditions of this case, we define s' like s. Suppose that $s = f(t) = f(t') = s'$. Since both s and s' contain a unique state p satisfying the two conditions above, we have $0t = 0t' = p$ and $pt = pt' = p$. Since the other states are mapped by s exactly as by t and t', we have $t = t'$.

• **(b):** t has no cycles and has a fixed point $r \neq p$.
Because $0 \prec r$ by Remark 2, $0t \preceq rt$ by Proposition 1. If r is a fixed point of t, then $p = 0t \preceq rt = r$. Since $r \neq p$, we have $p \prec r$. Let $f(t) = s$, where s is the transformation shown in Figure 3 and defined by

$$0s = 0, \quad qs = 0 \text{ for each fixed point } q \neq p,$$
$$qs = qt \text{ for the other states } q \in Q.$$

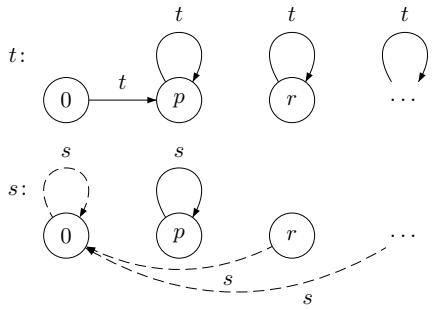

Fig. 3. Case 3(b) in the proof of Theorem 1

By Remark 3, $s \in S_n$. Suppose that $s \in T_n$; because $p \prec r$, $ps = p$, $rs = 0$, and $ps \preceq rs$ by Proposition 1, we have $p \prec 0$, which is a contradiction. Hence s is not in T_n and so is distinct from the transformations of Case 1. Also, s maps at least one state other than 0 to 0, and so is distinct from the transformations of Case 2 and also from the transformations of Case 3(a).

If $t' \neq t$ is another transformation satisfying the conditions of this case, we define s' like s. Now suppose that $s = f(t) = f(t') = s'$. There is only one fixed point of s other than 0 ($ps = p$), and only one fixed point of s' other than 0 ($p's' = p'$); hence $0t = p = p' = 0t'$. By the definition of s, for each state $q \neq 0$ such that $qs = 0$, we have $qt = q$. Similarly, for each state $q \neq 0$ such that $qs' = 0$, we have $qt' = q$. Hence t and t' agree on these states. Since the remaining states are mapped by s exactly as they are mapped by t and t', we have $t = t'$. Thus we have proved that $t \neq t'$ implies $f(t) \neq f(t')$.

• **(c):** t has no cycles, has no fixed point $r \neq p$ and there is a state r such that $p \prec r$ with $rt = p$.
Let $f(t) = s$, where s is the transformation shown in Figure 4 and defined by

$$0s = 0, \;\; ps = r, \;\; qs = 0 \text{ for each } q \succ p \text{ such that } qt = p,$$
$$qs = qt \text{ for the other states } q \in Q.$$

By Remark 3, $s \in S_n$. Suppose that $s \in T_n$; because $p \prec r$, $ps = r$, $rs = 0$, and $r = ps \preceq rs = 0$ by Proposition 1, we have $r \prec 0$—a contradiction. Hence $s \notin T_n$ and s is distinct from the transformations of Case 1.

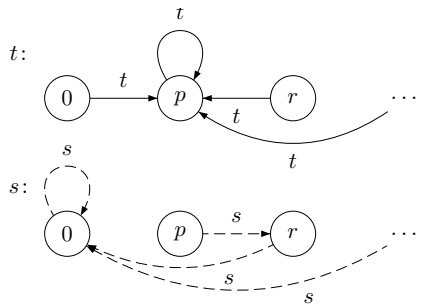

Fig. 4. Case 3(c) in the proof of Theorem 1

Because s maps at least one state other than 0 to 0 ($rs = 0$), it is distinct from the transformations of Case 2 and 3(a). Also s does not have a fixed point other than 0, while the transformations of Case 3(b) have such a fixed point.

We claim that there is a unique state q such that (a) $0 \prec q \prec qs$ and (b) $qs^2 = 0$. First we show that p satisfies these conditions. By assumption $0 \prec p \prec r$ and $rt = p$; also $rs = 0$ by the definition of s. Condition (a) holds because $0 \prec p \prec r = ps$, and (b) holds because $0 = rs = ps^2$.

Now suppose that $0 \prec q \prec qs$, $qs^2 = 0$ and $q \neq p$. Since $qs \neq 0$, we have $qs = qt$ by the definition of s. Because qt has a t-predecessor, $p \preceq qt$ by Remark 2. Also $qt = qs \neq p$, for $qs = p$ implies $0 = qs^2 = ps = r$—a contradiction. Hence $p \prec qt$. From $qt = qs$ and $q \prec qs$, we have $q \prec qt$. Since $qs^2 = 0$ we have $(qt)s = 0$ and so $(qt)t = p$, by the definition of s. By Proposition 1, from $q \prec qt$ we have $qt \preceq (qt)t = p$, contradicting $p \prec qt$. So $q = p$.

If $t' \neq t$ is another transformation satisfying the conditions of this case, we define s' like s. Suppose that $s = f(t) = t(t') = s'$. Since s and s' contain a unique state p satisfying the two conditions above, we have $0t = 0t' = p$ and $pt = pt' = p$. Then r and the states $q \succ p$ with $qt = p$ are determined by p, since they are precisely the states $q \succ p$ with $qs = 0$. Since the other states are mapped by s exactly as by t and t', we have $t = t'$, and f is again injective.

- **All Cases Are Covered**

Now we need to ensure that any transformation t fits in at least one case. It is clear that t fits in Case 1 or 2 or 3. For Case 3, it is sufficient to show that if (i) $t \notin S_n$ does not contain a fixed point $r \neq p$, and (ii) there is no state r with $p \prec r$ and $rt = p$, then t contains a cycle.

First, if there is no r such that $p \prec r$, we claim that t is the constant transformation $(Q \to p)$. Consider any state $q \in Q$ such that $qt \neq p$. Then $p \prec qt$ by Remark 2, contradicting that there is no state r such that $p \prec r$.

So let r be some state such that $p \prec r$. Consider the sequence $r, rt, rt^2, \ldots$. By Remark 2, $p \preceq rt^i$ for all $i \geq 0$. If $rt^k = p$ for some $k \geq 1$, let i be the smallest such k; we have $(rt^{i-1})t = p$, contradicting (ii). Since p is the only fixed point by (i), we have $rt^i \neq rt^{i-1}$. Since there are finitely many states, $rt^i = rt^j$ for some i and j such that $0 \leq i < j - 1$, and so the states $rt^i, rt^{i+1}, \ldots, rt^j = rt^i$ form a cycle.

We have shown that for every transformation t in T_n there is a corresponding transformation $f(t)$ in S_n, and f is injective. So $|T_n| \leq |S_n| = n^{n-1} + n - 1$. □

Next we prove that S_n is the only transition semigroup meeting the bound. It follows that minimal DFAs of left ideals with the maximal syntactic complexity have maximal-length chains of length 2.

Theorem 2. *If T_n has size $n^{n-1} + n - 1$, then $T_n = S_n$.*

Proof. Consider a maximal-length chain of states strictly ordered by $\prec$ in $\mathcal{D}_n$. If its length is 2, then by Lemma 2, T_n is a subsemigroup of S_n. Thus only $T_n = S_n$ reaches the bound in this case.

Assume now that the length of a maximal-length chain is at least 3. Then there are states p and r such that $0 \prec p \prec r$. Let $R = \{q \mid p \prec q\}$, and let $X = Q \setminus (R \cup \{0, p\})$. We shall show that there exists a transformation s that is in S_n but not in $f(T_n)$. To define s we use the constant transformation $u = (Q \to p)$ as an auxiliary transformation. Let $0s = 0$, $ps = r$, $rs = 0$ for all $r \in R$, and $qs = qu = p$ for $q \in X$; these are precisely the rules we used in Case 3(c) in the proof of Theorem 1. By Remark 3, $s \in S_n$.

It remains to be shown that there is no transformation $t \in T_n$ such that $s = f(t)$. The proof that s is different from the transformations $f(t)$ of Cases 1, 2, 3(a) and 3(b) is exactly the same as the corresponding proof in Case 3(c) following the definition of s.

It remains to verify that there is no $u' \in T_n$ in Case 3(c) such that $f(u') = s$. Suppose there is such a u'. Recall that states p and r satisfying $0 \prec p \prec r$ have been fixed by assumption. By the definition of s, state p satisfies the conditions (a) $0 \prec p \prec ps$ and (b) $ps^2 = 0$. We claim that p is the only state satisfying these conditions. Indeed, if $q \neq p$ then either $qs = 0$, $q \not\prec qs = 0$ and (a) is violated, or $qs = p$, $qs^2 = ps = r \neq 0$ and (b) is violated. This observation is used in the proof of Case 3(c) to prove the claim below.

Both u and u' satisfy the conditions of Case 3(c), except that u fails the condition $u \notin S_n$. However, that latter condition is not used in the proof that if $u \neq u'$ and u' satisfy the other conditions of Case 3(c), then $s' \neq s$, where s' is the transformation obtained from u' by the rules of s. Thus s is also different from the transformations in $f(T_n)$ from Case 3(c).

Because $s \notin f(T_n)$, $s \in S_n$ and $f(T_n) \subseteq S_n$, the bound $n^{n-1} + n - 1$ cannot be reached if the length of the maximal-length chains is not 2. □

4 Two-Sided Ideals

If a language L is a right ideal, then $L = L\Sigma^*$ and L has exactly one final quotient, namely Σ^*; hence this also holds for two-sided ideals. For $n \geq 3$, in a two-sided ideal every maximal chain is of length at least 3: it starts with L, every quotient contains L and is contained in Σ^*.

It was proved in [4, Theorem 6, p. 125] that the transition semigroup of the following DFA of a two-sided ideal meets the bound $n^{n-2} + (n-2)2^{n-2} + 1$.

Definition 2 (Witness: Two-Sided Ideals). *For $n \geq 4$, define the DFA $\mathcal{W}_n = (Q, \Sigma_{\mathcal{W}}, \delta_{\mathcal{W}}, 0, \{n-1\})$, where $a\colon (1, 2, \ldots, n-2)$, $b\colon (1, 2)$, $c\colon (n-2 \to 1)$, $d\colon (n-2 \to 0)$, for $q = 0, \ldots, n-2$, $\delta(q, e) = 1$ and $\delta(n-1, e) = n-1$, and $f\colon (1 \to n-1)$. For $n = 4$, inputs a and b coincide.*

Remark 5. If $n = 1$, the only two-sided ideal is Σ^*, its syntactic complexity is 1, and the bound above is not tight. If $n = 2$, each two-sided ideal is of the form $L = \Sigma^* \Gamma \Sigma^*$, where $\emptyset \subsetneq \Gamma \subseteq \Sigma$, its syntactic complexity is 2, and the bound is tight. If $n = 3$, there are eight transformations that are initially aperiodic and such that $(n-1)t = t$ (the property of a right-ideal transformation). We have verified that the DFA having all eight or any seven of the eight transformations is not a two-sided ideal. Hence 6 is an upper bound, and we know from [4] that the transformations $[1, 2, 2]$, $[0, 0, 2]$, and $[0, 1, 2]$ generate a 6-element semigroup. From now on we may assume that $n \geq 4$.

We consider a minimal DFA $\mathcal{D}_n = (Q, \Sigma_{\mathcal{D}}, \delta_{\mathcal{D}}, 0, \{n-1\})$, where $Q = \{0, \ldots, n-1\}$, of an arbitrary two-sided ideal with n quotients, and let T_n be the transition semigroup of $\mathcal{D}_n$. We also deal with the witness DFA $\mathcal{W}_n = (Q, \Sigma_{\mathcal{W}}, \delta_{\mathcal{W}}, 0, \{n-1\})$ of Definition 2 with transition semigroup S_n.

Remark 6. In $\mathcal{W}_n$, the transformations induced by a, b, and c restricted to $Q \setminus \{0, n-1\}$ generate all the transformations of the states $1, \ldots, n-2$. Together with the transformations of d and f, they generate all transformations of Q that fix 0 and $n-1$. For any subset $S \subseteq \{1, \ldots, n-2\}$, there is a transformation—induced by a word w_S, say—that maps S to $n-1$ and fixes $Q \setminus S$. Then the words of the form $w_S e a^i$, for $i \in \{0, \ldots, n-3\}$, induce all transformations that maps $S \cup \{n-1\}$ to $n-1$ and $Q \setminus (S \cup \{n-1\})$ to $i+1$. In $\mathcal{W}_n$, there is also the constant transformation $ef\colon (Q \to n-1)$.

Lemma 3. *If $n \geq 4$ and a maximal-length chain in $\mathcal{D}_n$ strictly ordered by $\prec$ has length 3, then $|T_n| \leq n^{n-2} + (n-2)2^{n-2} + 1$, and T_n is a subsemigroup of S_n.*

Proof. Consider an arbitrary transformation $t \in T_n$; then $(n-1)t = n-1$. If $0t = 0$, then any state not in $\{0, n-1\}$ can possibly be mapped by t to any one of the n states; hence there are at most n^{n-2} such transformations.

If $0t \neq 0$, then $0 \prec 0t$. Consider any state $q \notin \{0, 0t\}$; since $\mathcal{D}_n$ is minimal, q must be reachable from 0 by some transformation s, that is, $q = 0s$. If $0st \notin \{0t, n-1\}$, then $0t \prec 0st$ by Remark 2. But then we have the chain $0 \prec 0t \prec 0st \prec n-1$ of length 4, contradicting our assumption. Hence we must have either $0st = 0t$, or $0st = n-1$. For a fixed $0t$, a subset of the states in $Q \setminus \{0, n-1\}$ can be mapped to $0t$ and the remaining states in $Q \setminus \{0, n-1\}$ to $n-1$, thus giving 2^{n-2} transformations. Since there are $n-2$ possibilities for $0t$, we obtain the second part of the bound. Finally, all states can be mapped to $n-1$.

By Remark 6 all of the above-mentioned transformations are in S_n. □

Theorem 3 (Two-Sided Ideals, Factor-Closed Languages). *If L is a two-sided ideal or a factor-closed language with $n \geq 4$ quotients, then its syntactic complexity is less than or equal to $n^{n-2} + (n-2)2^{n-2} + 1$.*

Proof. It suffices to prove the result for two-sided ideals. As we did for left ideals, we show that $|T_n| \leq |S_n|$, by constructing an injective function $f\colon T_n \to S_n$.

We have $q \preceq n-1$ for any $q \in Q$, and $n-1$ is a fixed point of every transformation in T_n and S_n.

We omit here the detailed proof of injectivity of f. The complete proof can be found in [3].

For a transformation $t \in T_n$, consider the following cases:

Case 1: $t \in S_n$.
The proof is the same as that of Case 1 of Theorem 1.

Case 2: $t \notin S_n$, and $0t^2 \neq 0t$.
Let $0t = p \prec \cdots \prec pt^k = pt^{k+1}$ be the chain defined from p.

- **(a):** $pt^k \neq n-1$.

The proof is the same as that of Case 2 of Theorem 1.

- **(b):** $pt^k = n-1$ and $k \geq 2$.

Let $f(t) = s$, where s is the following transformation:

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

- **(c):** $pt = n-1$.

Let $P = \{0, p, n-1\}$. Since $n \geq 4$, there must be a state $r \notin P$. If $p \prec r$ for all $r \notin P$, then $n-1 = pt \preceq rt$; hence $rt = n-1$ for all such r, and $qt \in \{p, n-1\}$ for all $q \in Q$. By Remark 6, there is a transformation in S_n that maps $S \cup \{n-1\}$ to $n-1$, and $Q \setminus (S \cup \{n-1\})$ to p for any $S \subseteq \{1, \ldots, n-2\}$. Thus $t \in S_n$—a contradiction.

In view of the above, there must exist a state $r \notin P$ such that $p \not\preceq r$. By Remark 2, we have $p \preceq rt$ and of course $rt \preceq n-1$. If rt is p or $n-1$ for all $r \notin P$, we again have the situation described above, showing that $t \in S_n$. Hence there must exist an $r \notin P$ such that $p \not\preceq r$ and $p \prec rt \prec n-1$.

Let $f(t) = s$, where s is the following transformation:

$$0s = 0, \quad ps = rt, \quad (rt)s = p, \quad rs = 0,$$
$$qs = qt \text{ for the other states } q \in Q.$$

Case 3: $t \notin S_n$, $0t = p \neq 0$ and $pt = p$.

- **(a):** t has a cycle.

The proof is analogous to that of Case 3(a) in Theorem 1, but we need to ensure that s is different from the s of Cases 2(b) and 2(c).

- **(b):** t has no cycles and has a fixed point $r \notin \{p, n-1\}$.

The proof is analogous to that of Case 3(b) in Theorem 1, but we need to ensure that s is different from the s of Cases 2(b) and 2(c).

- **(c):** t has no cycles and no fixed point $r \notin \{p, n-1\}$, but has a state $r \succ p$ mapped to p.

The proof is analogous to that of Case 3(c) in Theorem 1, but we need to ensure that s is different from the s of Cases 2(b) and 2(c).

• **(d):** t has no cycles, no fixed point $r \notin \{p, n-1\}$, and no state $r \succ p$ mapped to p, but has a state r such that $p \prec r \prec n-1$, mapped to $n-1$.
Let $f(t) = s$, where s is the following transformation:

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

• **All Cases Are Covered**
We need to ensure that any transformation t fits in at least one case. It is clear that t fits in Case 1 or 2 or 3. Any transformation from Case 2 fits in Case 2(a) or 2(b) or 2(c). For Case 3, it is sufficient to show that if (i) $t \notin S_n$ does not contain a fixed point $r \notin \{p, n-1\}$, and (ii) there is no state r, $p \prec r \prec n-1$, mapped to p or $n-1$, then t has a cycle.

If there is no state r such that $p \prec r \prec n-1$, then $qt \in \{p, n-1\}$ for any $q \in Q$, since $qt \succeq p$; by Remark 6, $t \in S_n$—a contradiction.

So let r be some state such that $p \prec r \prec n-1$. Consider the sequence $r, rt, rt^2, \ldots$. By Remark 2, $p \preceq rt^i$ for all $i \geq 0$. If $rt^k \in \{p, n-1\}$ for some $k \geq 1$, then let i be the smallest such k. Then we have $(rt^{i-1})t \in p$, contradicting (ii). Since p and $n-1$ are the only fixed points by (i), we have $rt^i \neq rt^{i-1}$. Since there are finitely many states, $rt^i = rt^j$ for some i and j such that $0 \leq i < j-1$, and so the states $rt^i, rt^{i+1} \ldots, rt^j = rt^i$ form a cycle. □

Theorem 4. *If T_n has size $n^{n-2} + (n-2)2^{n-2} + 1$, then $T_n = S_n$.*

Proof. The proof is very similar to that of Theorem 2. It can be found in [3]. □

References

1. Brzozowski, J.: Quotient complexity of regular languages. J. Autom. Lang. Comb. 15(1/2), 71–89 (2010)
2. Brzozowski, J., Jirásková, G., Li, B.: Quotient complexity of ideal languages. Theoret. Comput. Sci. 470, 36–52 (2013)
3. Brzozowski, J., Szykuła, M.: Upper bounds on syntactic complexity of left and two-sided ideals (2014), `http://arxiv.org/abs/1403.2090`
4. Brzozowski, J., Ye, Y.: Syntactic complexity of ideal and closed languages. In: Mauri, G., Leporati, A. (eds.) DLT 2011. LNCS, vol. 6795, pp. 117–128. Springer, Heidelberg (2011)
5. Pin, J.E.: Syntactic semigroups. In: Handbook of Formal Languages: Word, Language, Grammar, vol. 1, pp. 679–746. Springer, New York (1997)
6. Yu, S.: State complexity of regular languages. J. Autom. Lang. Comb. 6, 221–234 (2001)