Upper Bound on Syntactic Complexity of Suffix-Free Languages

Marek Szykuła

University of Wrocław, Poland

Joint work with Janusz Brzozowski

University of Waterloo, Canada

DCFS, 25.06.2015

Abstract

We study syntactic complexity of the class of suffix-free languages:

- A language L is suffix-free if $w = uv \in L$, where u is non-empty, implies that the suffix $v \notin L$.

Contribution

- Brzozowski, Li, and Ye (TCS 2012) conjectured that:
  - For $n \geq 6$, the syntactic complexity of suffix-free languages is $(n-1)^{n-2} + n - 2$.
  - For $n \geq 6$, the syntactic complexity of bifix-free languages is $(n-1)^{n-3} + (n-2)^{n-3} + (n-3)2^{n-3}$.
  - For $n \geq 6$, the syntactic complexity of factor-free languages is $(n-1)^{n-3} + (n-3)2^{n-3} + 1$.
- We prove the first conjecture (about suffix-free languages).

Left quotient

The (left) quotient of a regular language L by a word w is

$$w^{-1}L = \{x \in \Sigma^* \mid wx \in L\}.$$

Analogously, for a state q of a minimal DFA $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ recognizing L:

$$L_q = \{x \in \Sigma^* \mid qt_x \in F\},$$

where t_x is the transformation induced by word x.

So L_q is the set of words taking q to an accepting state.

Left quotient

The (left) quotient of a regular language L by a word w is

$$w^{-1}L = \{x \in \Sigma^* \mid wx \in L\}.$$

Analogously, for a state q of a minimal DFA $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ recognizing L:

$$L_q = \{x \in \Sigma^* \mid qt_x \in F\},$$

where t_x is the transformation induced by word x.

So L_q is the set of words taking q to an accepting state.

Left quotient

The (left) quotient of a regular language L by a word w is

$$w^{-1}L = \{x \in \Sigma^* \mid wx \in L\}.$$

Analogously, for a state q of a minimal DFA $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ recognizing L:

$$L_q = \{x \in \Sigma^* \mid qt_x \in F\},$$

where t_x is the transformation induced by word x.

So L_q is the set of words taking q to an accepting state.

State complexity

Nerode right congruence on Σ^*

For a regular language L and words $x, y \in \Sigma^*$:

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

State complexity

The state complexity or quotient complexity $\kappa(L)$ of a regular language L is:

- The number of equivalence classes of $\sim_L$.
- The number of left quotients of L.
- The number of states in a minimal DFA recognizing L.

Syntactic complexity

Myhill congruence

For a regular language L and words $x, y \in \Sigma^*$:

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

Syntactic complexity

The syntactic complexity $\sigma(L)$ of a regular language L is:

- $|\Sigma^+/\approx_L|$ – the number of equivalence classes of $\approx_L$.
- The size of the syntactic semigroup of L.
- The size of the transition semigroup of a minimal DFA recognizing L.

Syntactic complexity of a class of languages

The syntactic complexity of a class of languages is:

- The size of the largest syntactic semigroups of languages in that class.
- Expressed as a function of the state complexities $n = \kappa(L)$ of the languages.

In other words

- Suppose we have an n-state minimal DFA recognizing some language from the given class.
- We ask how many transformations (at most) can be in the transition semigroup of the DFA.

Syntactic complexity of a class of languages

The syntactic complexity of a class of languages is:

- The size of the largest syntactic semigroups of languages in that class.
- Expressed as a function of the state complexities $n = \kappa(L)$ of the languages.

In other words

- Suppose we have an *n*-state minimal DFA recognizing some language from the given class.
- We ask how many transformations (at most) can be in the transition semigroup of the DFA.

Proposition

- $n - 1 \leq \sigma(L) \leq n^n$
- *The bounds are tight for* $n > 1$ *in the class of all regular languages.*

Previous results

- Gomes, Howie 1992: (partially) monotonic semigroups.
- Krawetz, Lawrence, Shallit 2003: unary and binary alphabets.
- Holzer, König 2004: unary and binary alphabets.
- Brzozowski, Ye 2010: ideal and closed languages.
- Beaudry, Holzer 2011: semigroups of reversible DFAs.
- Brzozowski, Liu 2012: finite, cofinite, definite, reverse definite languages.
- Brzozowski, Li, Ye 2012: prefix-, suffix-, bifix-, factor-free languages.
- Iván, Nagy-György 2013: (generalized) definite languages.
- Brzozowski, Li 2013: $\mathcal{J}$-trivial and $\mathcal{R}$-trivial languages.
- Brzozowski, Li, Liu 2013: aperiodic, nearly monotonic semigroups.
- Brzozowski, Szykuła 2014: aperiodic semigroups.
- Brzozowski, Szykuła 2014: left and two-sided ideal languages.

Suffix-free languages

Suffix-free languages

- A language L is suffix-free if $w = uv \in L$, where u is non-empty, implies that the suffix $v \notin L$.

Basic properties

- L has the empty quotient $\emptyset$.
- For $w, x \in \Sigma^+$, if $w^{-1}L \neq \emptyset$ then $w^{-1}L \neq (xw)^{-1}L$.
- For $w \in \Sigma^*$, the chain of quotients $w^{-1}L, (ww)^{-1}L, \ldots$ ends in $\emptyset$.

Syntactic complexity of suffix-free languages

Brzozowski, Li, Ye 2012

- Two maximal semigroups were introduced: $\mathbf{W}^{\leq 5}(n)$ and $\mathbf{W}^{\geq 6}(n)$.
- For $n \leq 5$ the largest transition semigroup is $\mathbf{W}^{\leq 5}(n)$.
- For $n = 6$ a largest transition semigroup is $\mathbf{W}^{\geq 6}(n)$.
- It was conjectured that $\mathbf{W}^{\geq 6}(n)$ is also largest for all $n \geq 7$.
- The size of $\mathbf{W}^{\geq 6}(n)$ is $(n-1)^{n-2} + (n-2)$.

Here

- We have proved that $\mathbf{W}^{\geq 6}(n)$ is the unique largest transition semigroup for $n \geq 7$.

Syntactic complexity of suffix-free languages

Brzozowski, Li, Ye 2012

- Two maximal semigroups were introduced: $\mathbf{W}^{\leq 5}(n)$ and $\mathbf{W}^{\geq 6}(n)$.
- For $n \leq 5$ the largest transition semigroup is $\mathbf{W}^{\leq 5}(n)$.
- For $n = 6$ a largest transition semigroup is $\mathbf{W}^{\geq 6}(n)$.
- It was conjectured that $\mathbf{W}^{\geq 6}(n)$ is also largest for all $n \geq 7$.
- The size of $\mathbf{W}^{\geq 6}(n)$ is $(n-1)^{n-2} + (n-2)$.

Here

- We have proved that $\mathbf{W}^{\geq 6}(n)$ is the unique largest transition semigroup for $n \geq 7$.

Properties of DFAs of suffix-free languages

Let $\mathcal{D} = (Q, \Sigma, \delta, 0, F)$ be a minimal DFA of a suffix-free language.

- $Q = \{0, 1, \dots, n-1\}$.
- 0 is the start state.
- $n-1$ is the empty state.

Let T_n be the transition semigroup of $\mathcal{D}$.

Basic property

For any transformation $t \in T_n$ and any state $q \in Q \setminus \{0\}$, we have $0t \neq qt$ unless $0t = qt = n-1$.

Properties of DFAs of suffix-free languages

Let $\mathcal{D} = (Q, \Sigma, \delta, 0, F)$ be a minimal DFA of a suffix-free language.

- $Q = \{0, 1, \ldots, n-1\}$.
- 0 is the start state.
- $n-1$ is the empty state.

Let T_n be the transition semigroup of $\mathcal{D}$.

Basic property

For any transformation $t \in T_n$ and any state $q \in Q \setminus \{0\}$, we have $0t \neq qt$ unless $0t = qt = n-1$.

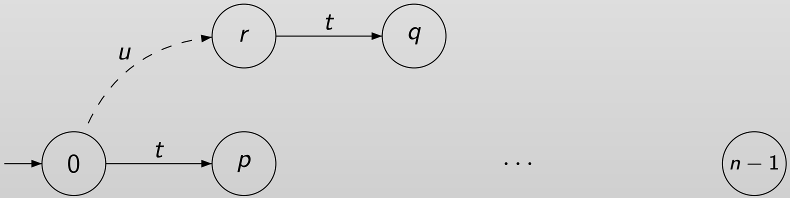
Colliding pairs of states

Definition

Two states $p, q \in Q$ are colliding in T_n, if there is a transformation $t \in T_n$ such that:

- $0t = p$,
- $rt = q$ for some state $r \in Q \setminus \{0, n-1\}$.

No transformation s can map a pair of colliding states to a single state, except to $n-1$.



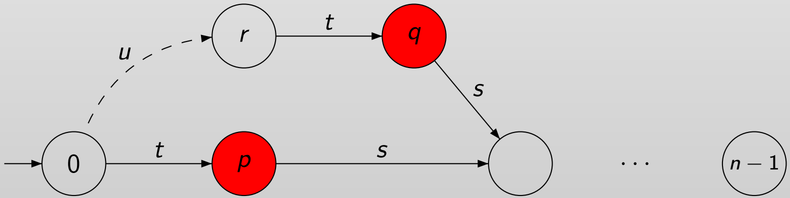
Colliding pairs of states

Definition

Two states $p, q \in Q$ are colliding in T_n, if there is a transformation $t \in T_n$ such that:

- $0t = p$,
- $rt = q$ for some state $r \in Q \setminus \{0, n-1\}$.

No transformation s can map a pair of colliding states to a single state, except to $n-1$.



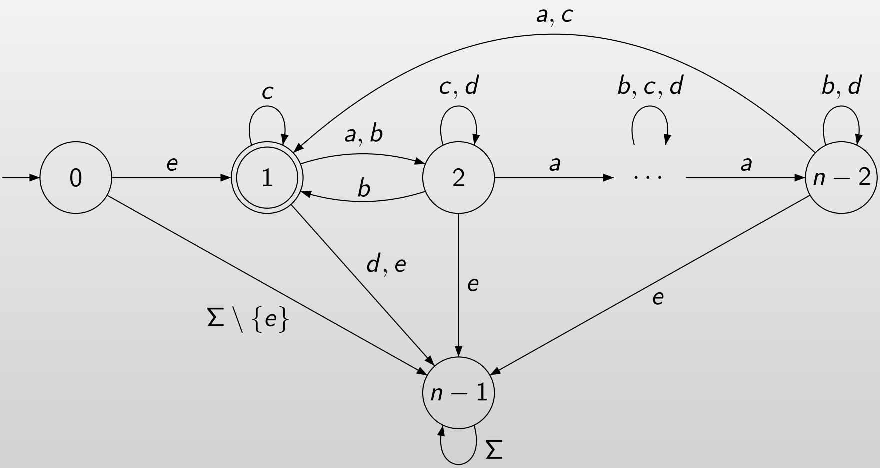
Colliding states

Remark

In $\mathbf{W}^{\leq 5}(n)$ all pairs $p, q \in Q \setminus \{0, n-1\}$ are colliding.

In $\mathbf{W}^{\geq 6}(n)$ there are no colliding pairs.

Witness DFA with the transition semigroup $\mathbf{W}^{\geq 6}(n)$ generated by 5 letters.



Let $S_n = \mathbf{W}^{\geq 6}(n)$ be the transition semigroup of the witness

The transition semigroup S_n

S_n contains:

- All transformations that map 0 to $n-1$.
- All transformations that map 0 to a state in $\{1, \dots, n-2\}$, and all other states to $n-1$.

These are $(n-1)^{n-2} + (n-2)$ transformations in total.

Upper bound

- We assume $n \geq 7$.
- S_n is the transition semigroup of the witness.
- T_n is the transition semigroup of an arbitrary DFA of a suffix-free language.
- We show that $|T_n| \leq |S_n| = (n-1)^{n-2} + (n-2)$.

Idea

- It is possible that $T_n \not\subseteq S_n$.
- We construct an injective function:

$$\varphi \colon T_n \to S_n$$

- (Injective: for every $t \neq t'$ we have $\varphi(t) \neq \varphi(t')$.)

Upper bound

- We assume $n \geq 7$.
- S_n is the transition semigroup of the witness.
- T_n is the transition semigroup of an arbitrary DFA of a suffix-free language.
- We show that $|T_n| \leq |S_n| = (n-1)^{n-2} + (n-2)$.

Idea

- It is possible that $T_n \not\subseteq S_n$.
- We construct an injective function:

$$\varphi \colon T_n \to S_n$$

- (Injective: for every $t \neq t'$ we have $\varphi(t) \neq \varphi(t')$.)

Upper bound

- We assume $n \geq 7$.
- S_n is the transition semigroup of the witness.
- T_n is the transition semigroup of an arbitrary DFA of a suffix-free language.
- We show that $|T_n| \leq |S_n| = (n-1)^{n-2} + (n-2)$.

Idea

- It is possible that $T_n \not\subseteq S_n$.
- We construct an injective function:

$$\varphi\colon T_n \to S_n$$

- (Injective: for every $t \neq t'$ we have $\varphi(t) \neq \varphi(t')$.)

Case 1

If $t \in S_n$, then let $\varphi(t) = t$.

This is obviously injective, and $\varphi(t) \in T_n$.

From now, for $t \notin S_n$ we need to assign a transformation $\varphi(t) \notin T_n$.

Case 1

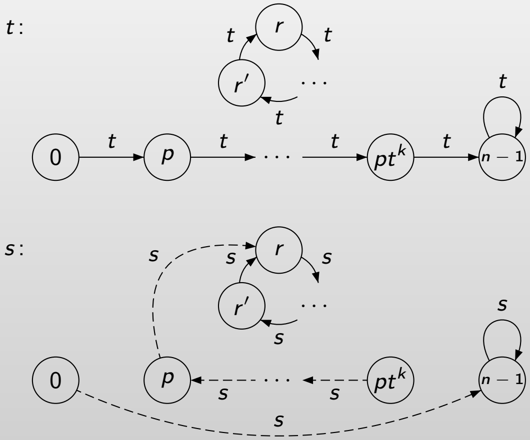
If $t \in S_n$, then let $\varphi(t) = t$.

This is obviously injective, and $\varphi(t) \in T_n$.

From now, for $t \notin S_n$ we need to assign a transformation $\varphi(t) \notin T_n$.

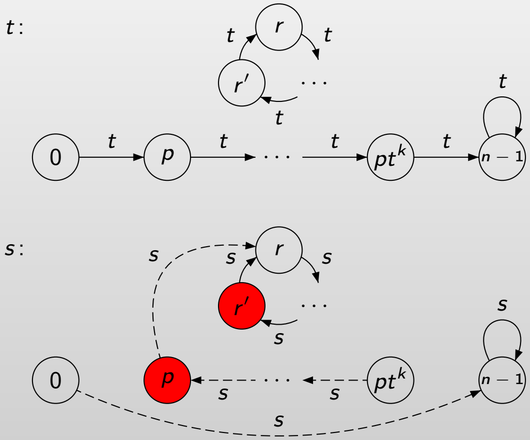
Case 2

If $t \notin S_n$ and t has a cycle, then we reverse the chain $p, pt, \dots$ and map p to a minimal state that appear in cycles.



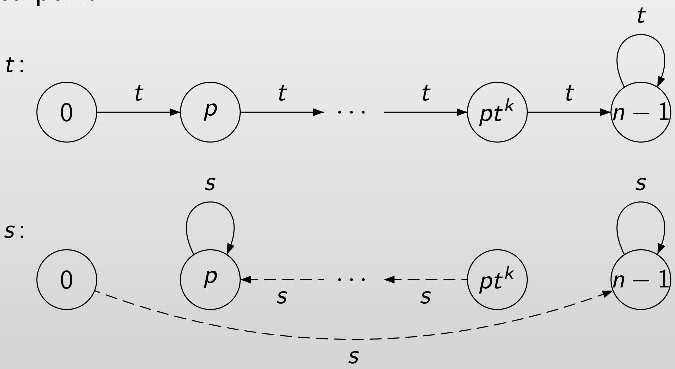
Case 2

If $t \notin S_n$ and t has a cycle, then we reverse the chain $p, pt, \ldots$ and map p to a minimal state that appear in cycles.



Case 3

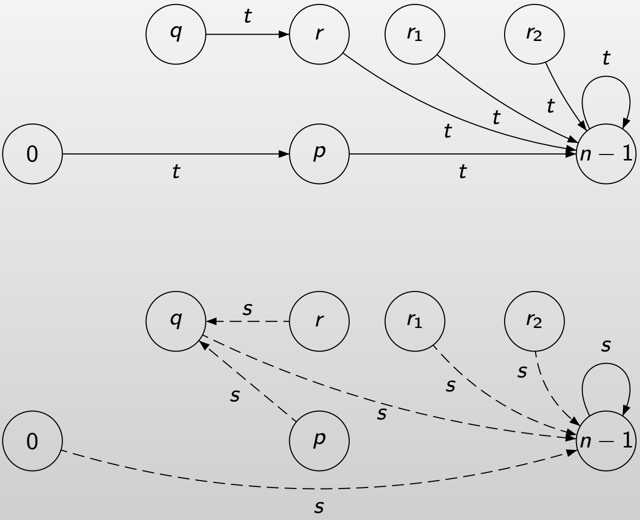
If $t \notin S_n$ and t has no cycles, then we reverse the chain and make p a fixed point.



Case 11 with $p < r$

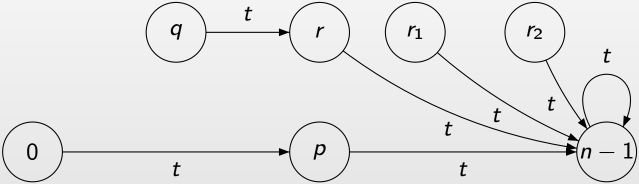
t :

s :

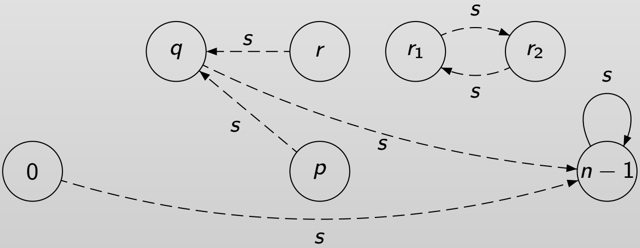


Case 11 with $p > r$

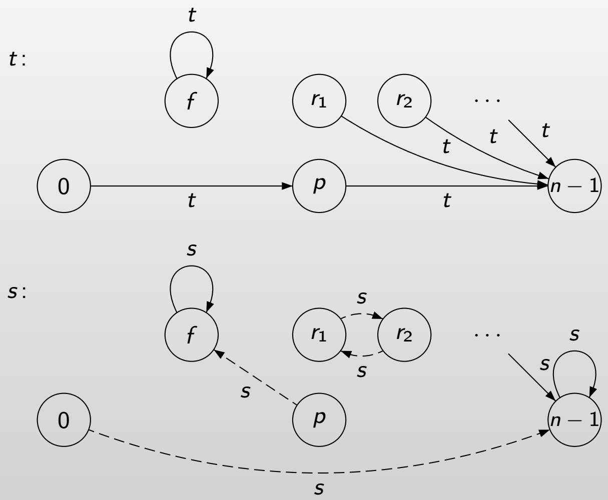
t:



s:



The last case 12



We have 12 cases

Case 1 $t \in S_n$.

Case 2 t has a cycle.

Case 3 $pt \neq n-1$.

Case 4 There is a fixed point $r \in Q \setminus \{0, n-1\}$ with in-degree ≥ 2.

Case 5 There is r with in-degree ≥ 1 that is not a fixed point and $rt \neq n-1$.

Case 6 There is $r \in Q \setminus \{0, n-1\}$ with in-degree ≥ 2.

Case 7 There are $q_1, q_2 \in Q \setminus \{0, n-1\}$ that are not fixed points and satisfy $q_1 t \neq n-1$ and $q_2 t \neq n-1$.

Case 8 There are two fixed points r_1 and r_2 in $Q \setminus \{0, n-1\}$ with in-degree 1.

Case 9 There is $q \in Q \setminus \{0, n-1\}$ that is not a fixed point and satisfies $qt \neq n-1$, $p < qt$, and a fixed point $f \neq n-1$.

Case 10 There is $q \in Q \setminus \{0, n-1\}$ that is not a fixed point and satisfies $qt \neq n-1$, and a fixed point $f \neq n-1$.

Case 11 There is $q \in Q \setminus \{0, n-1\}$ that is not a fixed point and satisfies $qt \neq n-1$.

Case 12 Any other transformation.

In each case we assume that t does not belong to the previous cases.

Summary

The 12 cases cover all possibilities for t.

Function φ is injective and $\varphi(T_n) \subseteq S_n$, and $|T_n| = |\varphi(T_n)| \leq |S_n|$.

Uniqueness of maximality

Theorem

For $n \geq 7$, the transition semigroup S_n of the witness is the only one reaching the upper bound.

Theorem

Five letters are necessary to generate S_n.
So five letters are needed to reach the upper bound.

Future work

Other problems

The technique of injective functions can be applied to similar problems.

- Earlier, we solved the problems of syntactic complexity of left ideals (5 subcases) and two-sided ideals (8 subcases) (DLT 2014).
- The problems of upper bounds for syntactic complexity of bifix- and factor-free languages remain open.

Thank you!

Future work

Other problems

The technique of injective functions can be applied to similar problems.

- Earlier, we solved the problems of syntactic complexity of left ideals (5 subcases) and two-sided ideals (8 subcases) (DLT 2014).
- The problems of upper bounds for syntactic complexity of bifix- and factor-free languages remain open.

Thank you!