

# Syntactic Complexities of Nine Subclasses of Regular Languages

by

Baiyu Li

A thesis  
presented to the University of Waterloo  
in fulfillment of the  
thesis requirement for the degree of  
Master of Mathematics  
in  
Computer Science

Waterloo, Ontario, Canada, 2012

© Baiyu Li 2012

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Abstract

The syntactic complexity of a regular language is the cardinality of its syntactic semi-group. The syntactic complexity of a subclass of the class of regular languages is the maximal syntactic complexity of languages in that class, taken as a function of the state complexity  $n$  of these languages. We study the syntactic complexity of suffix-, bifix-, and factor-free regular languages, star-free languages including three subclasses, and  $\mathcal{R}$ - and  $\mathcal{J}$ -trivial regular languages.

We found upper bounds on the syntactic complexities of these classes of languages. For  $\mathcal{R}$ - and  $\mathcal{J}$ -trivial regular languages, the upper bounds are  $n!$  and  $\lfloor e(n-1)! \rfloor$ , respectively, and they are tight for  $n \geq 1$ . Let  $C_k^n$  be the binomial coefficient “ $n$  choose  $k$ ”. For monotonic languages, the tight upper bound is  $C_n^{2n-1}$ . We also found tight upper bounds for partially monotonic and nearly monotonic languages. For the other classes of languages, we found tight upper bounds for languages with small state complexities, and we exhibited languages with maximal known syntactic complexities. We conjecture these lower bounds to be tight upper bounds for these languages.

We also observed that, for some subclasses  $\mathcal{C}$  of regular languages, the upper bound on state complexity of the reversal operation on languages in  $\mathcal{C}$  can be met by languages in  $\mathcal{C}$  with maximal syntactic complexity. For  $\mathcal{R}$ - and  $\mathcal{J}$ -trivial regular languages, we also determined tight upper bounds on the state complexity of the reversal operation.

Part of the work presented in this thesis has appeared in the papers [6, 12].

## Acknowledgements

First, I thank my supervisor Janusz Brzozowski, who shared his great ideas with me and brought me many challenging problems ever since I was a third year undergraduate student. Over the years he guided me on many topics we studied together, and he introduced to me the problem of syntactic complexity that became the topic of this thesis. I enjoyed numerous discussions we had, and I am very grateful for his valuable suggestions on my writings and presentations.

I also thank my co-supervisor Jeffrey O. Shallit for his great teaching in CS365 and CS662, which are the most exciting courses I have ever taken. He also organized a weekly seminar on Languages and Automata Theory at the University of Waterloo, from which I learned ideas from many different subfields of theoretical computer science and number theory.

I am grateful to professor Jonathan Buss and professor Kevin Hare for reading this thesis. I thank all my teachers at University of Waterloo for their excellent courses.

I thank my office mate, Navid Vafaei, for his kindness in sharing cookies with me.

Lastly, I thank my parents and grandparents for their endless love and support. My study would not be possible without their dedication.

The research in this thesis was partly supported by Ontario Graduate Scholarship, the President's scholarship of University of Waterloo, the Natural Sciences and Engineering Research Council of Canada grant No. OGP0000871, and financial support from the David R. Cheriton School of Computer Science at the University of Waterloo.

## **Dedication**

To my parents and grandparents.

# Table of Contents

List of Tables	viii
List of Figures	ix
<b>1 Introduction</b>	<b>1</b>
1.1 Suffix-, Bifix-, and Factor-Free Regular Languages . . . . .	3
1.2 Star-Free Languages and Three Subclasses . . . . .	3
1.3 $\mathcal{R}$ -Trivial and $\mathcal{J}$ -Trivial Regular Languages . . . . .	4
1.4 Organization . . . . .	5
<b>2 Preliminaries</b>	<b>6</b>
2.1 Transformations . . . . .	6
2.2 Syntactic Complexity of Regular Languages . . . . .	7
<b>3 Syntactic Complexity of Suffix-, Bifix-, and Factor-Free Languages</b>	<b>10</b>
3.1 Prefix-Free Regular Languages . . . . .	10
3.2 Suffix-Free Regular Languages . . . . .	11
3.3 Bifix-Free Regular Languages . . . . .	21
3.4 Factor-Free Regular Languages . . . . .	27
3.5 Summary . . . . .	31

<b>4</b>	<b>Syntactic Complexity of Star-Free Languages</b>	<b>34</b>
4.1	Aperiodic Transformations . . . . .	34
4.2	Monotonicity in Transformations, Automata and Languages . . . . .	37
4.2.1	Monotonic Transformations, DFA's and Languages . . . . .	37
4.2.2	Monotonic Partial Transformations and IDFA's . . . . .	38
4.2.3	Nearly Monotonic Transformations and DFA's . . . . .	41
4.3	Summary . . . . .	43
<b>5</b>	<b>Syntactic Complexity of <math>\mathcal{R}</math>- and <math>\mathcal{J}</math>-Trivial Languages</b>	<b>45</b>
5.1	$\mathcal{R}$ -Trivial Regular Languages . . . . .	45
5.2	$\mathcal{J}$ -Trivial Regular Languages . . . . .	47
<b>6</b>	<b>Quotient Complexity of Reverse Languages</b>	<b>59</b>
<b>7</b>	<b>Conclusions</b>	<b>64</b>
	<b>Permissions Page</b>	<b>66</b>
	<b>References</b>	<b>76</b>

# List of Tables

3.1	The number $S_m(h)$ of labeled rooted trees with $m$ nodes and height at most $h$ . . . . .	14
3.2	Syntactic complexities of prefix- and suffix-free regular languages. . . . .	32
3.3	Syntactic complexities of bifix- and factor-free regular languages. . . . .	33
4.1	Syntactic complexity of star-free languages. . . . .	44
5.1	Number of transformations in $\mathcal{E}(\pi_Z)$ for each $Z \in \mathcal{P}_4(Q)$ . . . . .	54



# List of Figures

3.1	Transition graph of some $t \in \mathbf{B}_{\text{sf}}(15)$ with principal sequence 1, 2, 3, 4, 5, 15.	14
3.2	Three types of trees of the form $T_t(3)$ , where $\{i_1, i_2, i_3\} = \{8, 9, 10\}$ .	15
4.1	Forests and transformations.	35
4.2	Conflict graph for $n = 3$ .	36
4.3	Partially monotonic automata.	39
4.4	Partially monotonic DFA that is monotonic and has an empty state.	40
5.1	DFA $\mathcal{A}_4$ with $\kappa(L(\mathcal{A}_4)) = 4$ and $\sigma(L(\mathcal{A}_4)) = 24$	48
5.2	Nondecreasing transformations $t = [1, 3, 3, 5, 6, 6]$ and $s = [4, 3, 3, 6, 6, 6]$ .	50
5.3	DFA $\mathcal{B}$ with two inputs $a$ and $b$ , where $t_a = [2, 2, 4, 4]$ and $t_b = [3, 2, 4, 4]$ .	51
5.4	Transition graphs of $t$ , $t'$ , and $t_{Z''}$ .	57
6.1	NFA $\mathcal{N}'_6$ of $L_6'^R$ with quotient complexity $\kappa(L_6'^R) = 17$	60
6.2	NFA $\mathcal{N}''_7$ of $L_7''^R$ with quotient complexity $\kappa(L_7''^R) = 18$	61
6.3	NFA $\mathcal{N}_5 = \mathcal{A}_5^R$ for $n = 5$ accepting $L_5^R$ .	62

# Chapter 1

## Introduction

The class of regular languages plays a fundamental role in theoretical computer science. It is exactly the class of languages accepted by deterministic / nondeterministic finite automata (DFA's / NFA's), which are the simplest, but still powerful, computational models. There are many applications of regular languages and finite automata in computer systems, especially in various information processing algorithms and programming languages. Operations on finite automata are often needed to accomplish a computational task. Beside computational complexity measures, the amount of resources needed to describe finite automata resulting from various operations has become crucial in recent years due to the massive usage of large finite automata. At the same time, the development of a number of software packages for manipulating regular languages and finite automata, such as AMoRE, FAdo, Grail, and GAP, made it much easier to explore problems on descriptonal complexity of regular languages.

One of the widely considered descriptonal complexity measures for regular languages is the so-called *state complexity*, which is the number of states in the minimal DFA accepting a regular language. An equivalent notion is *quotient complexity* [3], which is the number of distinct quotients of a regular language. State complexity of regular languages and regular operations has been studied quite extensively. In 1959, Rabin and Scott [39] proved the  $2^n$  upper bound on the number of states in the minimal DFA obtained from an  $n$ -state NFA. This bound was proved to be tight by Mirkin [33] in 1966. In 1970, Maslov [31] stated without proof the tight upper bounds on state complexity of some regular operations. State complexity became a very active topic in the 1990s. For surveys of this topic and lists of references we refer the reader to [53, 4].

State complexity of operations on subclasses of regular languages has also attracted

many researchers. Subclasses of regular languages are important not only because they are theoretically interesting, but also because some applications may not require the full class of regular languages, but only a special subclass. For example, from the theoretical point of view, it is very interesting that the state complexity of unary languages is closely related to Jacobsthal's function from number theory [37]. In practice, special subclasses of regular languages are used in coding theory and its applications, and more will be said about this later. Also, the class of ideal regular languages, studied in [10], is very useful in pattern matching and its applications.

Another descriptonal complexity measure for regular languages is *syntactic complexity*, which is the cardinality of the syntactic semigroup of a language. A construct similar to the syntactic semigroup is the syntactic monoid of a language. The syntactic semigroup and monoid are among the central objects in algebraic automata theory, and they are used to bring deep connections between algebra and formal language theory. Examples include the work of Krohn and Rhodes [29], Schützenberger [44], and Simon [49]. The syntactic semigroup of a regular language is isomorphic to the semigroup of all transformations performed by the minimal DFA of that language. So it is natural to consider the relation between the syntactic complexity and the state complexity of a language. By the *syntactic complexity of a subclass of regular languages*, we mean the maximal syntactic complexity of languages in that class, taken as a function of the state complexity of these languages.

In contrast to state complexity, syntactic complexity has not received much attention. In 1970 Maslov [31] stated without proof that  $n^n$  was a tight upper bound on the number of transformations performed by a DFA of  $n$  states. In 2003–2004, Holzer and König [23], and Krawetz, Lawrence and Shallit [28] studied the syntactic complexity of unary and binary languages. Brzozowski and Ye [8] examined the syntactic complexity of ideal and closed regular languages in 2010; they found a tight upper bound on the syntactic complexity of prefix-free regular languages [6] in 2012. A language is *definite* (*reverse-definite*) if it can be decided whether a word  $w$  belongs to the language simply by examining the suffix (prefix) of  $w$  of some fixed length. Recently Brzozowski and Liu [13] studied the syntactic complexity of finite/cofinite, definite, and reverse definite languages.

It was pointed out in [8] that syntactic complexity can be very different for regular languages with the same state complexity. Thus, for a fixed  $n$ , languages with state complexity  $n$  may possibly be distinguished by their syntactic complexities. Another interesting discovery [4] is that the state complexity upper bounds for most regular operations and many of their combinations can be reached simultaneously by a family of regular languages with maximal syntactic complexities. An earlier theorem in [43] shows that the upper bound on the reversal operation on regular languages is reached by languages with maximal syntactic complexities.

In this thesis, we extend recent studies in syntactic complexity to more subclasses of regular languages, namely suffix-, bifix-, and factor-free regular languages, star-free languages including monotonic, partially monotonic, and nearly monotonic languages, and  $\mathcal{R}$ - and  $\mathcal{J}$ -trivial regular languages. In the following, we first give an overview of these subclasses. We then describe the structure of this thesis at the end of this chapter.

## 1.1 Suffix-, Bifix-, and Factor-Free Regular Languages

A *proper* prefix (suffix, factor) of a word  $w$  is a prefix (respectively, suffix, factor) of  $w$  other than  $w$  itself. A language is *prefix-free* (respectively, *suffix-free*, *factor-free*) if it does not contain any pair of words such that one is a proper prefix (respectively, suffix, factor) of the other. It is *bifix-free* if it is both prefix- and suffix-free. Nontrivial prefix-, suffix-, bifix-, and factor-free languages are also known as prefix, suffix, bifix, and infix codes [2, 46], respectively, and have many applications in areas such as cryptography, data compression and information processing.

In this thesis we consider only regular prefix-, suffix-, bifix-, and factor-free languages. With regard to state complexity, Han, Salomaa and Wood [21] examined prefix-free regular languages in 2009, and Han and Salomaa [20] studied suffix-free regular languages in the same year. Bifix- and factor-free regular languages were studied by Brzozowski, Jirásková, Li, and Smith [11] in 2011.

## 1.2 Star-Free Languages and Three Subclasses

*Star-free* languages are the smallest class containing the finite languages and closed under boolean operations and concatenation. In 1965, Schützenberger [44] proved that a language is star-free if and only if its syntactic monoid is *group-free*, that is, has only trivial subgroups. An equivalent condition is that the minimal DFA accepting a star-free language is *permutation-free*, that is, has only trivial permutations (cycles of length 1). Such DFA's are called *aperiodic*, and this is the term we use. Star-free languages were studied in detail in 1971 by McNaughton and Papert [32].

*Monotonic* (*partially monotonic*) languages are regular languages whose minimal DFA's can perform only monotonic (partially monotonic) transformations with respect to a fixed total order. A language is called *nearly monotonic* if its minimal DFA can perform only partially monotonic or constant transformations with respect to a fixed total order. The

precise definitions of monotonic and partially monotonic transformations will be given later in the corresponding chapter. The syntactic semigroups of monotonic (partially monotonic) languages are subsets of the semigroup of monotonic (partially monotonic) transformations, which were studied in [24, 19]. In 2005, Ananichev and Volkov [1] studied monotonic DFA's for the problem of reset words for synchronizing automata.

It has been shown in 2011 by Brzozowski and Liu [7] that boolean operations, concatenation, star, and reversal in the class of star-free languages meet all the state complexity upper bounds of regular languages, with very few exceptions. Monotonic languages were used as witnesses to these upper bounds. Also, Kutrib, Holzer, and Meckel [22] proved in 2011 that in most cases exactly the same tight state complexity bounds are reached by operations on aperiodic NFA's as on general NFA's.

### 1.3 $\mathcal{R}$ -Trivial and $\mathcal{J}$ -Trivial Regular Languages

The well-known Green equivalence relations  $\mathcal{L}, \mathcal{R}, \mathcal{J}, \mathcal{H}$  on semigroups, treated in [25] for example, are very important in semigroup theory. The definition of the Green equivalence relations will be given later. If  $\rho$  is a Green equivalence relation on a monoid  $M$ , then  $M$  is  $\rho$ -trivial if and only if every  $\rho$ -equivalence class contains only one element. A language is  $\rho$ -trivial if and only if its syntactic monoid is  $\rho$ -trivial. In this thesis we consider only regular  $\rho$ -trivial languages. Note that  $\mathcal{H}$ -trivial regular languages are exactly the star-free languages, and  $\mathcal{L}$ -,  $\mathcal{R}$ -, and  $\mathcal{J}$ -trivial languages are all subclasses of star-free languages.

A language  $L \subseteq \Sigma^*$  is *piecewise-testable* if it is a finite boolean combination of languages of the form  $\Sigma^*a_1 \cdots \Sigma^*a_l\Sigma^*$ , where  $a_i \in \Sigma$ . Simon [49] proved in 1975 that a language is piecewise-testable if and only if it is  $\mathcal{J}$ -trivial. A *biautomaton* is a finite automaton which can read the input word alternatively from the left and from the right. In 2011 Klíma and Polák [27] showed that a language is piecewise-testable if and only if it is accepted by an acyclic biautomaton.

In 1979 Brzozowski and Fich [5] proved that a regular language is  $\mathcal{R}$ -trivial if and only if its minimal DFA is *partially ordered*, that is, has an acyclic graph representation. They also showed that  $\mathcal{R}$ -trivial regular languages are exactly the languages of a finite boolean combination of languages  $\Sigma_1^*a_1 \cdots \Sigma_l^*a_l\Sigma^*$ , where  $a_i \in \Sigma$  and  $\Sigma_i \subseteq \Sigma \setminus \{a_i\}$ .

## 1.4 Organization

We first state basic definitions and facts about syntactic complexity in Chapter 2. In Chapters 3 to 5 we find upper and lower bounds on syntactic complexities of suffix-, bifix-, and factor-free regular languages, star-free languages including three subclasses, and  $\mathcal{R}$ - and  $\mathcal{J}$ -trivial regular languages. We also exhibit witness languages to lower bounds on the syntactic complexities in their classes.

It is very difficult to obtain tight upper bounds on the syntactic complexities of some languages. For example, although tight upper bounds for right ideals and prefix-closed regular languages are easy to derive, our knowledge on tight upper bounds for left ideals and suffix-closed regular languages is limited to small cases [8]. Our upper bounds for monotonic, partially monotonic, nearly monotonic, and  $\mathcal{R}$ - and  $\mathcal{J}$ -trivial regular languages are tight. For other classes of languages, we prove tight upper bounds for some small cases, and we present conjectures on tight upper bounds for general cases.

In Chapter 6 we show that, for most language classes that we study, the upper bounds on the state / quotient complexity of reversal of these languages can be met by our languages with largest syntactic complexities. Chapter 7 concludes the thesis and discusses future work.

# Chapter 2

## Preliminaries

In this chapter we state basic definitions and facts that will be helpful for our study of syntactic complexity of regular languages.

### 2.1 Transformations

A *partial transformation* of a set  $Q$  is a partial mapping of  $Q$  into itself. We consider partial transformations of finite sets only, and we assume without loss of generality that  $Q = \{1, 2, \dots, n\}$ . Let  $t$  be a partial transformation of  $Q$ . If  $t$  is defined for  $i \in Q$ , then  $it$  is the image of  $i$  under  $t$ ; otherwise it is undefined and we write  $it = \square$ . For convenience, we let  $\square t = \square$ . If  $X$  is a subset of  $Q$ , then  $Xt = \{it \mid i \in X\}$ . The *composition* of two partial transformations  $t_1$  and  $t_2$  of  $Q$  is a partial transformation  $t_1 \circ t_2$  such that  $i(t_1 \circ t_2) = (it_1)t_2$  for all  $i \in Q$ . We usually drop the composition operator “ $\circ$ ” and write  $t_1 t_2$  for short.

An arbitrary partial transformation can be written in the form

$$t = \begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ i_1 & i_2 & \cdots & i_{n-1} & i_n \end{pmatrix},$$

where  $i_k = kt \in Q \cup \{\square\}$ , for  $k \in Q$ . We also use the notation  $t = [i_1, i_2, \dots, i_{n-1}, i_n]$  for the partial transformation  $t$  above. The *domain* of  $t$  is the set  $\text{dom}(t) = \{k \in Q \mid kt \neq \square\}$ . The *range* of  $t$  is the set  $\text{rng}(t) = Qt$ . The *rank* of  $t$ , denoted by  $\text{rank}(t)$ , is the cardinality of  $\text{rng}(t)$ , i.e.,  $\text{rank}(t) = |\text{rng}(t)|$ . The binary relation  $\omega_t$  on  $Q \times Q$  is defined as follows: For any  $i, j \in Q$ ,  $i \omega_t j$  if and only if  $i, j \in Q$  and  $it^k = jt^l$  for some  $k, l \geq 0$ . Such a relation is an equivalence relation, and each equivalence class is called an *orbit* of  $t$ . For any  $i \in Q$ ,

the orbit of  $t$  containing  $i$  is denoted by  $\omega_t(i)$ . The set of all orbits of  $t$  is denoted by  $\Omega(t)$ . Clearly,  $\Omega(t)$  is a partition of  $Q$ .

A (full) transformation  $t$  of a set  $Q$  is a partial transformation such that  $\text{dom}(t) = Q$ . The *identity* transformation maps each element to itself, that is,  $it = i$  for  $i = 1, \dots, n$ . A transformation  $t$  is a *cycle* of length  $k \geq 2$  if there exist pairwise distinct elements  $i_1, \dots, i_k$  such that  $i_1t = i_2, i_2t = i_3, \dots, i_{k-1}t = i_k, i_kt = i_1$ , and  $jt = j$  for all  $j \notin \{i_1, \dots, i_k\}$ . Such a cycle is denoted by  $(i_1, i_2, \dots, i_k)$ . For  $i < j$ , a *transposition* is the cycle  $(i, j)$ . A *singular* transformation, denoted by  $\binom{i}{j}$ , has  $it = j$  and  $ht = h$  for all  $h \neq i$ . A *constant* transformation, denoted by  $\binom{Q}{j}$ , has  $it = j$  for all  $i$ .

The set of all full transformations of a set  $Q$ , denoted by  $\mathcal{T}_Q$ , is a finite semigroup, in fact, a monoid. We refer the reader to the book of Ganyushkin and Mazorchuk [17] for a detailed discussion of finite transformation semigroups. In 1935 Piccard [36] proved that three transformations of  $Q$  are sufficient to generate the monoid  $\mathcal{T}_Q$ . In the same year, Eilenberg showed that fewer than three generators are not possible, as reported by Sierpiński [47]. Dénes [16] (apparently unaware of the earlier work) studied more general generators in 1968; we use his formulation:

**Theorem 2.1** (Transformations). *The complete transformation monoid  $\mathcal{T}_Q$  of size  $n^n$  can be generated by any cyclic permutation of  $n$  elements together with any transposition and any singular transformation. In particular,  $\mathcal{T}_Q$  can be generated by  $a = (1, 2, \dots, n)$ ,  $b = (1, 2)$  and  $c = \binom{n}{1}$ .*

A *permutation* of  $Q$  is a mapping of  $Q$  onto itself. In other words, a permutation  $\pi$  of  $Q$  is a transformation where  $\text{rng}(\pi) = Q$ . The set of all permutations of a set  $Q$  of  $n$  elements is a group, denoted by  $\mathfrak{S}_Q$  and called the *symmetric group* of degree  $n$ . It is well-known that two generators are sufficient to generate the symmetric group of degree  $n$ .

**Theorem 2.2** (Permutations). *The symmetric group  $\mathfrak{S}_Q$  of size  $n!$  can be generated by any cyclic permutation of  $n$  elements together with any transposition. In particular,  $\mathfrak{S}_Q$  can be generated by  $a = (1, 2, \dots, n)$  and  $b = (1, 2)$ .*

## 2.2 Syntactic Complexity of Regular Languages

For general definitions and facts about regular languages, we refer the reader to the handbook chapter by Yu [52]. Let  $\Sigma$  be a finite non-empty alphabet,  $\Sigma^*$  the free monoid generated by  $\Sigma$ , and  $\Sigma^+$  the free semigroup generated by  $\Sigma$ . A *word*  $w$  is an element of  $\Sigma^*$ ,



and its length is denoted by  $|w|$ . The empty word is denoted by  $\varepsilon$ . A *language over*  $\Sigma$  is a subset of  $\Sigma^*$ .

The *left quotient*, or simply *quotient*, of a language  $L$  by a word  $w$  is the language  $L_w = \{x \in \Sigma^* \mid wx \in L\}$ . The *quotient complexity* of  $L$ , denoted by  $\kappa(L)$ , is the number of distinct quotients of  $L$ . The *Nerode right congruence* [35]  $\sim_L$  of any language  $L \subseteq \Sigma^*$  is defined as follows: For all  $x, y \in \Sigma^*$ ,

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

Clearly,  $L_x = L_y$  if and only if  $x \sim_L y$ . Thus each equivalence class of this right congruence corresponds to a distinct quotient of  $L$ .

The *Myhill congruence* [34]  $\approx_L$  of any language  $L \subseteq \Sigma^*$  is defined as follows: For all  $x, y \in \Sigma^*$ ,

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

This congruence is also known as the *syntactic congruence* of  $L$ . The quotient set  $\Sigma^+ / \approx_L$  of equivalence classes of the relation  $\approx_L$  is a semigroup called the *syntactic semigroup* of  $L$ , and  $\Sigma^* / \approx_L$  is the *syntactic monoid* of  $L$ . The *syntactic complexity*  $\sigma(L)$  of  $L$  is the cardinality of its syntactic semigroup. The *monoid complexity*  $\mu(L)$  of  $L$  is the cardinality of its syntactic monoid. If the equivalence class containing  $\varepsilon$  is a singleton in the syntactic monoid, then  $\sigma(L) = \mu(L) - 1$ ; otherwise,  $\sigma(L) = \mu(L)$ .

*Regular languages* are the smallest class of languages containing finite languages and are closed under concatenation, union, and star operations. Both Nerode and Myhill congruences are important for regular languages because of the following famous equivalent conditions on any language  $L$ :

1.  $L$  is a regular language;
2. The Nerode right congruence  $\sim_L$  of  $L$  is of finite index;
3. The Myhill congruence  $\approx_L$  of  $L$  is of finite index.

In other words,  $L$  is regular if and only if  $L$  has a finite number of quotients, and if and only if the syntactic semigroup of  $L$  is finite.

A *deterministic finite automaton* (DFA) over  $\Sigma$  is a quintuple  $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$ , where  $Q$  is a finite, non-empty set of *states*,  $\delta : Q \times \Sigma \rightarrow Q$  is the *transition function*,  $q_1 \in Q$  is the *initial state*, and  $F \subseteq Q$  is the set of *final states*. We extend  $\delta$  to  $Q \times \Sigma^*$  in the usual way. The DFA  $\mathcal{A}$  accepts a word  $w \in \Sigma^*$  if  $\delta(q_1, w) \in F$ . The set of all words

accepted by  $\mathcal{A}$  is  $L(\mathcal{A})$ . If  $q$  is a state of  $\mathcal{A}$ , then the language  $L_q$  of  $q$  is the language accepted by the DFA  $(Q, \Sigma, \delta, q, F)$ . Two states  $p$  and  $q$  of  $\mathcal{A}$  are *equivalent* if  $L_p = L_q$ . A state  $q$  of  $\mathcal{A}$  is *empty* if  $L_q = \emptyset$ . Regular languages are exactly the languages accepted by DFA's. The *state complexity* of a regular language is the number of states in the minimal DFA accepting that language.

If  $L$  is regular, then we define the *quotient DFA* of  $L$  as  $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$ , where  $Q = \{L_w \mid w \in \Sigma^*\}$  is the set of quotients of  $L$ ,  $\delta(L_w, a) = L_{wa}$ ,  $q_1 = L_\varepsilon = L$ , and  $F = \{L_w \mid \varepsilon \in L_w\}$ . Such a DFA is the minimal DFA accepting  $L$ . Hence, for regular languages, quotient complexity and state complexity are the same.

An *incomplete deterministic finite automaton (IDFA)* is a quintuple  $\mathcal{I} = (Q, \Sigma, \delta, q_1, F)$ , where  $Q, \Sigma, q_1$  and  $F$  are as in a DFA, and  $\delta$  is a partial function such that, for any  $p, q \in Q$ ,  $a \in \Sigma$ , either  $\delta(q, a) = p$  for some  $p \in Q$  or  $\delta(q, a)$  is undefined. An IDFA is *minimal* if no two of its states are equivalent. Every DFA is also an IDFA.

The *quotient IDFA* of  $L$  is the quotient DFA of  $L$  after the empty state (if present) and all transitions incident to it are removed. The quotient IDFA is minimal. If a regular language  $L$  has quotient IDFA  $\mathcal{I}$ , then the DFA  $\mathcal{A}$  obtained by adding the empty quotient to  $\mathcal{I}$ , if necessary, is the quotient DFA of  $L$ . Conversely, if  $L$  has quotient DFA  $\mathcal{A}$ , then the IDFA  $\mathcal{I}$  obtained from  $\mathcal{A}$  by removing the empty quotient, if present, is the quotient IDFA of  $L$ . The two automata  $\mathcal{A}$  and  $\mathcal{I}$  are equivalent, in the sense that they accept the same language.

Let  $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$  be a DFA. For each word  $w \in \Sigma^+$ , the transition function for  $w$  defines a transformation  $t_w$  of  $Q$  by the word  $w$ : for all  $i \in Q$ ,  $it_w \stackrel{\text{def}}{=} \delta(i, w)$ . The set  $T_{\mathcal{A}}$  of all such transformations by non-empty words forms a subsemigroup of  $\mathcal{T}_Q$ , called the *transition semigroup* of  $\mathcal{A}$  [38]. Conversely, we can use a set  $\{t_a \mid a \in \Sigma\}$  of transformations to define  $\delta$ , and so the DFA  $\mathcal{A}$ . When the context is clear we simply write  $a = t$ , where  $t$  is a transformation of  $Q$ , to mean that the transformation performed by  $a \in \Sigma$  is  $t$ .

If  $\mathcal{A}$  is the quotient DFA of  $L$ , then  $T_{\mathcal{A}}$  is isomorphic to the syntactic semigroup  $T_L$  of  $L$  [32], and we represent elements of  $T_L$  by transformations in  $T_{\mathcal{A}}$ .

For any IDFA  $\mathcal{I}$ , each word  $w \in \Sigma^*$  performs a partial transformation of  $Q$ . The set of all such partial transformations is the *transition semigroup* of  $\mathcal{I}$ . If  $\mathcal{I}$  is the quotient IDFA of a language  $L$ , this semigroup is isomorphic to the transition semigroup of the quotient DFA of  $L$ , and hence also to the syntactic semigroup of  $L$ .

## Chapter 3

# Syntactic Complexity of Suffix-, Bifix-, and Factor-Free Languages

If  $w = uxv$  for some  $u, x, v \in \Sigma^*$ , then  $u$  is a *prefix* of  $w$ ,  $v$  is a *suffix* of  $w$ , and  $x$  is a *factor* of  $w$ . Both  $u$  and  $v$  are also factors of  $w$ . A *proper* prefix (suffix, factor) of  $w$  is a prefix (suffix, factor) of  $w$  other than  $w$ . A language is *prefix-free* (respectively, *suffix-free*, *factor-free*) if it does not contain any pair of words such that one is a proper prefix (respectively, suffix, factor) of the other. It is *bifix-free* if it is both prefix- and suffix-free.

In this chapter, we study the syntactic complexity of suffix-, bifix-, and factor-free regular languages. The syntactic complexity of prefix-free regular languages was studied by Brzozowski and Ye, and their results were published in [6]. For completeness, and because of the close relations to other language classes in this chapter, we first review the syntactic complexity of prefix-free regular languages in Section 3.1. Then we continue with suffix-, bifix-, and factor-free regular languages in Sections 3.2 to 3.4. We summarize our results in Section 3.5.

### 3.1 Prefix-Free Regular Languages

We start with the case when  $|\Sigma| = 1$ . For unary languages, the concepts prefix-, suffix-, bifix-, and factor-free, coincide. Suppose  $L$  is a unary prefix-free regular language with quotient complexity  $\kappa(L) = n$ . When  $n = 1$ , the only prefix-free language is  $L = \emptyset$  with  $\sigma(L) = 1$ . For  $n \geq 2$ , a prefix-free language  $L$  must be a singleton,  $L = \{a^{n-2}\}$ . The syntactic semigroup  $T_L$  of  $L$  consists of  $n - 1$  transformations  $t_w$  by words  $w = a^i$ , where  $1 \leq i \leq n - 1$ . Thus we have

**Proposition 3.1** (Unary Prefix-Free Regular Languages). *If  $L$  is a unary prefix-, suffix-, bifix-, or factor-free regular language with  $\kappa(L) = n \geq 2$ , then  $\sigma(L) = n - 1$ .*

Note that the tight upper bound on the syntactic complexity of regular unary languages [23] is  $n$ .

We assume that  $|\Sigma| \geq 2$  in the following. To simplify notation we write  $\varepsilon$  for the language  $\{\varepsilon\}$ . For any regular language  $L$ , a quotient  $L_w$  is *final* if  $\varepsilon \in L_w$ . Recall that a regular language  $L$  is prefix-free if and only if it has exactly one final quotient, and that quotient is  $\varepsilon$  [21]. The following upper bound was shown in [6]:

**Theorem 3.2** (Prefix-Free Regular Languages). *If  $L$  is regular and prefix-free with  $\kappa(L) = n \geq 2$ , then  $\sigma(L) \leq n^{n-2}$ . Moreover, this bound is tight for  $n = 2$  if  $|\Sigma| \geq 1$ , for  $n = 3$  if  $|\Sigma| \geq 2$ , for  $n = 4$  if  $|\Sigma| \geq 4$ , and for  $n \geq 5$  if  $|\Sigma| \geq n + 1$ .*

## 3.2 Suffix-Free Regular Languages

For any regular language  $L$ , a quotient  $L_w$  is *uniquely reachable* [3] if  $L_w = L_x$  implies that  $w = x$ . It is known from [20] that, if  $L$  is a suffix-free regular language, then  $L = L_\varepsilon$  is uniquely reachable by  $\varepsilon$ , and  $L$  has the empty quotient. Without loss of generality, we assume that 1 is the initial state, and  $n$  is the empty state in the quotient DFA of  $L$ . We will show that the cardinality of  $\mathbf{B}_{\text{sf}}(n)$ , defined below, is an upper bound (**B** for “bound”) on the syntactic complexity of suffix-free regular languages with quotient complexity  $n$ . For  $n \geq 2$ , let

$$\mathbf{B}_{\text{sf}}(n) = \{t \in \mathcal{T}_Q \mid 1 \notin \text{rng}(t), nt = n, \text{ and for all } j \geq 1, \\ 1t^j = n \text{ or } (1t^j \neq it^j \forall i, 1 < i < n)\}.$$

**Proposition 3.3.** *If  $L$  is a regular language with quotient DFA  $\mathcal{A}_n = (Q, \Sigma, \delta, 1, F)$  and syntactic semigroup  $T_L$ , then the following hold:*

1. *If  $L$  is suffix-free, then  $T_L$  is a subset of  $\mathbf{B}_{\text{sf}}(n)$ .*
2. *If  $L$  has the empty quotient, only one final quotient, and  $T_L \subseteq \mathbf{B}_{\text{sf}}(n)$ , then  $L$  is suffix-free.*

*Proof.* 1. Let  $L$  be suffix-free, and let  $\mathcal{A}_n$  be its quotient DFA. Consider an arbitrary  $t \in T_L$ . Since the quotient  $L$  is uniquely reachable,  $it \neq 1$  for all  $i \in Q$ . Since the quotient corresponding to state  $n$  is empty,  $nt = n$ . Since  $L$  is suffix-free, for any two quotients  $L_w$  and  $L_{uw}$ , where  $u, w \in \Sigma^+$  and  $L_w \neq \emptyset$ , we must have  $L_w \cap L_{uw} = \emptyset$ , and so  $L_w \neq L_{uw}$ . This means that, for any  $j \geq 1$ , if  $1t^j \neq n$ , then  $1t^j \neq it^j$  for all  $i$ ,  $1 < i < n$ . So  $t \in \mathbf{B}_{\text{sf}}(n)$ , and  $T_L \subseteq \mathbf{B}_{\text{sf}}(n)$ .

2. Assume that  $T_L \subseteq \mathbf{B}_{\text{sf}}(n)$ , and let  $f$  be the only final state. If  $L$  is not suffix-free, then there exist non-empty words  $u$  and  $v$  such that  $v, uv \in L$ . Let  $t_u$  and  $t_v$  be the transformations by  $u$  and  $v$ , and let  $i = 1t_u$ ; then  $i \neq 1$ . Assume without loss of generality that  $n$  is the empty state. Then  $f \neq n$ , and we have  $1t_v = f = 1t_{uv} = 1t_u t_v = it_v$ , which contradicts the fact that  $t_v \in \mathbf{B}_{\text{sf}}(n)$ . Therefore  $L$  is suffix-free.  $\square$

Let  $\text{b}_{\text{sf}}(n) = |\mathbf{B}_{\text{sf}}(n)|$ . We now prove that  $\text{b}_{\text{sf}}(n)$  is an upper bound on the syntactic complexity of suffix-free regular languages.

With each transformation  $t$  of  $Q$ , we associate a directed graph  $G_t$ , where  $Q$  is the set of nodes, and  $(i, j) \in Q \times Q$  is a directed edge from  $i$  to  $j$  if  $it = j$ . We call such a graph  $G_t$  the *transition graph* of  $t$ . For each node  $i$ , there is exactly one edge leaving  $i$  in  $G_t$ . Consider the infinite sequence  $i, it, it^2, \dots$  for any  $i \in Q$ . Since  $Q$  is finite, there exists least  $j \geq 0$  such that  $it^{j+1} = it^{j'}$  for some  $j' \leq j$ . Then the finite sequence  $\mathfrak{s}_t(i) = i, it, \dots, it^j$  contains all the distinct elements of the above infinite sequence, and it induces a directed path  $P_t(i) = i, it, \dots, it^j$  from  $i$  to  $it^j$  in  $G_t$ . In particular, if  $n \in \mathfrak{s}_t(1)$ , and  $nt = n$ , then we call  $\mathfrak{s}_t(1)$  the *principal sequence* of  $t$ , and  $P_t(1)$ , the *principal path* of  $G_t$ .

**Proposition 3.4.** *There exists a principal sequence for every transformation  $t \in \mathbf{B}_{\text{sf}}(n)$ .*

*Proof.* Suppose  $t \in \mathbf{B}_{\text{sf}}(n)$  and  $\mathfrak{s}_t(1) = 1, 1t, \dots, 1t^j$ . If  $t$  does not have a principal sequence, then  $n \notin \mathfrak{s}_t(1)$ , and  $1t^{j+1} = 1t^{j'} \neq n$  for some  $j' \leq j$ . Let  $i = 1t^{j+1-j'}$ ; then  $i \neq 1$  and  $1t^{j'} = it^{j'}$ , violating the last property of  $\mathbf{B}_{\text{sf}}(n)$ . Therefore there is a principal sequence for every  $t \in \mathbf{B}_{\text{sf}}(n)$ .  $\square$

Fix a transformation  $t \in \mathbf{B}_{\text{sf}}(n)$ . Let  $i \in Q$  be such that  $i \notin \mathfrak{s}_t(1)$ . If the sequence  $\mathfrak{s}_t(i)$  does not contain any element of the principal sequence  $\mathfrak{s}_t(1)$  other than  $n$ , then we say that  $\mathfrak{s}_t(i)$  has *no principal connection*. Otherwise, there exists least  $j \geq 1$  such that  $1t^j \neq n$  and  $1t^j = it^{j'} \in \mathfrak{s}_t(i)$  for some  $j' \geq 1$ , and we say that  $\mathfrak{s}_t(i)$  has a *principal connection* at  $1t^j$ . If  $j' < j$ , the principal connection is *short*; otherwise, it is *long*.

**Lemma 3.5.** *For all  $t \in \mathbf{B}_{\text{sf}}(n)$  and  $i \notin \mathfrak{s}_t(1)$ , the sequence  $\mathfrak{s}_t(i)$  has no long principal connection.*

*Proof.* Let  $t$  be any transformation in  $\mathbf{B}_{\text{sf}}(n)$ . Suppose for some  $i \notin \mathfrak{s}_t(1)$ , the sequence  $\mathfrak{s}_t(i)$  has a long principal connection at  $1t^j = it^{j'} \neq n$ , where  $j \leq j'$ . Hence  $it^{j'-j} \neq n$ , and  $1t^j = (it^{j'-j})t^j$ , which is a contradiction. Therefore, for all  $i \notin \mathfrak{s}_t(1)$ ,  $\mathfrak{s}_t(i)$  has no long principal connection.  $\square$

To calculate the cardinality of  $\mathbf{B}_{\text{sf}}(n)$ , we need the following observation.

**Lemma 3.6.** *For all  $t \in \mathbf{B}_{\text{sf}}(n)$  and  $i \notin \mathfrak{s}_t(1)$ , if  $\mathfrak{s}_t(i)$  has a principal connection, then there is no cycle incident to the path  $P_t(i)$  in the transition graph  $G_t$ .*

*Proof.* This observation can be derived from Theorem 1.2.9 of [17]. However, our proof is shorter. Pick any  $i \notin \mathfrak{s}_t(1)$  such that  $\mathfrak{s}_t(i)$  has a principal connection at  $1t^j = it^{j'}$  for some  $i, j$  and  $j'$ . Then the sequence  $\mathfrak{s}_t(i)$  contains  $n$ , and the path  $P_t(i)$  does not contain any cycle. Suppose  $C$  is a cycle which includes node  $x = it^k \in P_t(i)$ . Since there is only one outgoing edge for each node in  $G_t$ , the cycle  $C$  must be oriented and must contain a node  $x' \notin P_t(i)$  such that  $(x', x)$  is an edge in  $C$ . Then the next node in the cycle must be  $it^{k+1}$  since there is only one outgoing edge from  $x$ . But then  $x'$  can never be reached from  $P_t(i)$ , and so no such cycle can exist.  $\square$

By Lemma 3.6, for any  $1t^j \in \mathfrak{s}_t(1)$ , where  $j \geq 1$ , the union of directed paths from various nodes  $i$  to  $1t^j$ , if  $i \notin \mathfrak{s}_t(1)$  and  $\mathfrak{s}_t(i)$  has a principal connection at  $1t^j$ , forms a labeled tree  $T_t(j)$  rooted at  $1t^j$ . Suppose there are  $r_j + 1$  nodes in  $T_t(j)$  for each  $j$ , and suppose there are  $r$  elements of  $Q$  that are not in the principal sequence  $\mathfrak{s}_t(1)$  nor in any tree  $T_t(j)$ , for some  $r_j, r \geq 0$ . Note that  $1t^j$  is the only node in  $T_t(j)$  that is also in the principal sequence  $\mathfrak{s}_t(1)$ . Each tree  $T_t(j)$  has height at most  $j - 1$ ; otherwise, some  $i \in T_t(j)$  has a long principal connection. In particular, tree  $T_t(1)$  has height 1; so it is trivial with only one node  $1t$ . Then  $r_1 = 0$ , and we need consider trees  $T_t(j)$  only for  $j \geq 2$ . Let  $S_m(h)$  be the number of labeled rooted trees with  $m$  nodes and height at most  $h$ . This number can be found in the paper of Riordan [40]; the calculation is somewhat complex, and we refer the reader to [40] for details. For convenience, we include the values of  $S_m(h)$  for small values of  $m$  and  $h$  in Table 3.1, where the row number is  $h$  and the column number is  $m$ .

Since each of the  $m$  nodes can be the root, there are  $S'_m(h) = \frac{S_m(h)}{m}$  labeled trees rooted at a fixed node and having  $m$  nodes and height at most  $h$ . The following is an example of trees  $T_t(j)$  in transformations  $t \in \mathbf{B}_{\text{sf}}(n)$ .

Table 3.1: The number  $S_m(h)$  of labeled rooted trees with  $m$  nodes and height at most  $h$ .

$h/m$	1	2	3	4	5	6	7
0	1	0	0	0	0	0	0
1	1	2	3	4	5	6	7
2	1	2	9	40	205	1176	7399
3	1	2	9	64	505	4536	46249
4	1	2	9	64	625	7056	89929
5	1	2	9	64	625	7776	112609
6	1	2	9	64	625	7776	117649

*Example 3.7.* Let  $n = 15$ . Consider any transformation  $t \in \mathbf{B}_{\text{sf}}(15)$  with principal sequence  $\mathfrak{s}_t(1) = 1, 2, 3, 4, 5, 15$ . There are 9 elements of  $Q$  that are not in  $\mathfrak{s}_t(1)$ , and some of them are in the trees  $T_t(j)$  for  $2 \leq j \leq 4$ . Consider the cases where  $r_2 = 2$ ,  $r_3 = 3$ ,  $r_4 = 1$ , and  $r = 3$ . Figure 3.1 shows one such transformation  $t$ .

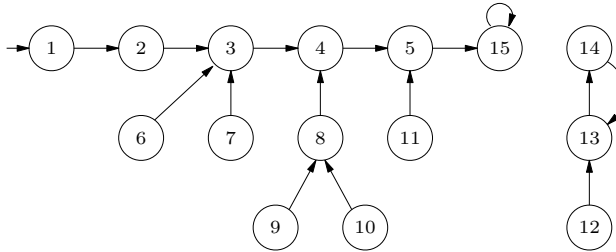


Figure 3.1: Transition graph of some  $t \in \mathbf{B}_{\text{sf}}(15)$  with principal sequence 1, 2, 3, 4, 5, 15.

For  $j = 2$ , the tree  $T_t(2)$  has height at most 1, and there are  $S'_{r_2+1}(1) = \frac{S_{r_2+1}(1)}{r_2+1} = \frac{3}{3} = 1$  possible  $T_t(2)$ . For  $j = 3$ , there are  $S'_{r_3+1}(2) = \frac{S_{r_3+1}(2)}{r_3+1} = 10$  possible  $T_t(3)$ , which are of one of the three types shown in Figure 3.2. Among the 10 possible  $T_t(3)$ , one is of type (a), three are of type (b), and six are of type (c). For  $j = 4$ , there are  $S'_{r_4+1}(3) = \frac{S_{r_4+1}(3)}{r_4+1} = 1$  possible  $T_t(4)$ . ■

Let  $C_k^n$  be the binomial coefficient, and let  $C_{k_1, \dots, k_m}^n$  be the multinomial coefficient. Then we have the following lemma:

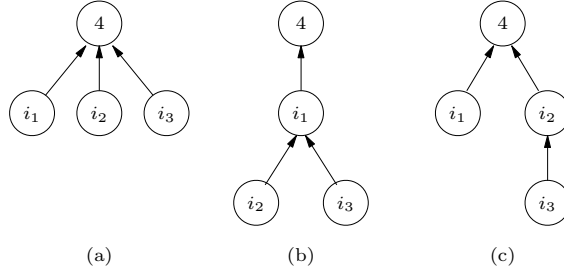


Figure 3.2: Three types of trees of the form  $T_t(3)$ , where  $\{i_1, i_2, i_3\} = \{8, 9, 10\}$ .

**Lemma 3.8.** *For  $n \geq 2$ , we have*

$$b_{\text{sf}}(n) = \sum_{k=0}^{n-2} C_k^{n-2} k! \sum_{\substack{r_2 + \dots + r_k + r \\ = n-k-2}} C_{r_2, \dots, r_k, r}^{n-k-2} (r+1)^r \prod_{j=2}^k S'_{r_j+1}(j-1). \quad (3.1)$$

*Proof.* Let  $t$  be any transformation in  $\mathbf{B}_{\text{sf}}(n)$ . Suppose  $\mathfrak{s}_t(1) = 1, 1t, \dots, 1t^k, n$  for some  $k$ ,  $0 \leq k \leq n-2$ . There are  $C_k^{n-2} k!$  different principal sequences  $\mathfrak{s}_t(1)$ . Now, fix  $\mathfrak{s}_t(1)$ . Suppose  $n-k-2 = r_2 + \dots + r_k + r$ , where, for  $2 \leq j \leq k$ , tree  $T_t(j)$  contains  $r_j + 1$  nodes, for some  $r_j \geq 0$ . There are  $C_{r_2, \dots, r_k, r}^{n-k-2}$  different tuples  $(r_2, \dots, r_k, r)$ . Each tree  $T_t(j)$  has height at most  $j-1$ , and it is rooted at  $1t^j$ . There are  $S'_{r_j+1}(j-1) = \frac{S_{r_j+1}(j-1)}{r_j+1}$  different trees  $T_t(j)$ . Let  $E$  be the set of the remaining  $r$  elements  $x$  of  $Q$  that are not in any tree  $T_t(j)$  nor in the principal sequence  $\mathfrak{s}_t(1)$ . The image  $xt$  can only be chosen from  $E \cup \{n\}$ . There are  $(r+1)^r$  different mappings of  $E$ . Altogether we have the desired formula.  $\square$

From Proposition 3.3 and Lemma 3.8 we have

**Proposition 3.9.** *For  $n \geq 2$ , if  $L$  is a suffix-free language with quotient complexity  $n$ , then its syntactic complexity  $\sigma(L)$  satisfies  $\sigma(L) \leq b_{\text{sf}}(n)$ , where  $b_{\text{sf}}(n)$  is the cardinality of  $\mathbf{B}_{\text{sf}}(n)$ , and it is given by Equation (3.1).*

Note that  $\mathbf{B}_{\text{sf}}(n)$  is not a semigroup for  $n \geq 4$  because  $s_1 = [2, 3, n, \dots, n, n]$ ,  $s_2 = [n, 3, 3, \dots, 3, n] \in \mathbf{B}_{\text{sf}}(n)$ , but  $s_1 s_2 = [3, 3, n, \dots, n, n] \notin \mathbf{B}_{\text{sf}}(n)$ . Hence, although  $b_{\text{sf}}(n)$  is an upper bound on the syntactic complexity of suffix-free regular languages, that bound is not tight. Our objective is to find the largest subset of  $\mathbf{B}_{\text{sf}}(n)$  that is a semigroup.



For  $n \geq 2$ , let

$$\mathbf{W}_{\text{sf}}^{\leq 5}(n) = \{t \in \mathbf{B}_{\text{sf}}(n) \mid \text{for all } i, j \in Q \text{ where } i \neq j, \\ \text{we have } it = jt = n \text{ or } it \neq jt\},$$

where  $\mathbf{W}$  stands for “witness”, and the superscript  $\leq 5$  will be explained in Theorem 3.14.

**Proposition 3.10.** *For  $n \geq 2$ ,  $\mathbf{W}_{\text{sf}}^{\leq 5}(n)$  is a semigroup contained in  $\mathbf{B}_{\text{sf}}(n)$ , and its cardinality is*

$$w_{\text{sf}}^{\leq 5}(n) = |\mathbf{W}_{\text{sf}}^{\leq 5}(n)| = \sum_{k=1}^{n-1} C_k^{n-1} (n-1-k)! C_{n-1-k}^{n-2}.$$

*Proof.* We know that any  $t$  is in  $\mathbf{W}_{\text{sf}}^{\leq 5}(n)$  if and only if the following hold:

1.  $it \neq 1$  for all  $i \in Q$ , and  $nt = n$ ;
2. for all  $i, j \in Q$ , such that  $i \neq j$ , either  $it = jt = n$  or  $it \neq jt$ .

Clearly  $\mathbf{W}_{\text{sf}}^{\leq 5}(n) \subseteq \mathbf{B}_{\text{sf}}(n)$ . For any transformations  $t_1, t_2 \in \mathbf{W}_{\text{sf}}^{\leq 5}(n)$ , consider the composition  $t_1 t_2$ . Since  $1 \notin \text{rng}(t_2)$ , we have  $1 \notin \text{rng}(t_1 t_2)$ . We also have  $nt_1 t_2 = nt_2 = n$ . Pick any  $i, j \in Q$  such that  $i \neq j$ . Suppose  $it_1 t_2 \neq n$  or  $jt_1 t_2 \neq n$ . If  $it_1 t_2 = jt_1 t_2$ , then  $it_1 = jt_1$  and thus  $i = j$ , a contradiction. Hence  $t_1 t_2 \in \mathbf{W}_{\text{sf}}^{\leq 5}(n)$ , and  $\mathbf{W}_{\text{sf}}^{\leq 5}(n)$  is a semigroup contained in  $\mathbf{B}_{\text{sf}}(n)$ .

Let  $t \in \mathbf{W}_{\text{sf}}^{\leq 5}(n)$  be any transformation. Note that  $nt = n$  is fixed. Let  $Q' = Q \setminus \{n\}$ , and  $Q'' = Q \setminus \{1, n\}$ . Suppose  $k$  elements in  $Q'$  are mapped to  $n$  by  $t$ , where  $0 \leq k \leq n-1$ ; then there are  $C_k^{n-1}$  choices of these elements. For the set  $D$  of the remaining  $n-1-k$  elements, which must be mapped by  $t$  to pairwise distinct elements of  $Q''$ , there are  $C_{n-1-k}^{n-2} (n-1-k)!$  choices for the mapping  $t|_D$ . When  $k=0$ , there is no such  $t$  since  $|Dt| = n-1 > n-2 = |Q''|$ . Altogether, the cardinality of  $\mathbf{W}_{\text{sf}}^{\leq 5}(n)$  is  $|\mathbf{W}_{\text{sf}}^{\leq 5}(n)| = \sum_{k=1}^{n-1} C_k^{n-1} (n-1-k)! C_{n-1-k}^{n-2}$ .  $\square$

*Remark 3.11.* A *partial injective transformation* of a set  $Q$  is a partial injective mapping of  $Q$  into itself. The set of all such transformations of  $Q$  is a semigroup, usually called the *symmetric inverse semigroup* [17] and denoted by  $\mathcal{IS}_Q$ . Let  $Q' = Q \setminus \{n\}$ . The number  $w_{\text{sf}}^{\leq 5}(n)$  coincides with the number of *nilpotents* in  $\mathcal{IS}_{Q'}$ , which are the transformations  $t \in \mathcal{IS}_{Q'}$  such that  $\text{dom}(t^k) = \emptyset$  for some  $k \geq 1$ . Riordan [41] reported that  $w_{\text{sf}}^{\leq 5}(n)$  has the asymptotic approximation

$$w_{\text{sf}}^{\leq 5}(n) \sim \frac{1}{\sqrt{2e}} (n-1)^{n-\frac{4}{5}} e^{-(n-1)+2\sqrt{n-1}}.$$

We now construct a generating set  $\mathbf{G}_{\text{sf}}^{\leq 5}(n)$  ( $\mathbf{G}$  for “generators”) of size  $n$  for  $\mathbf{W}_{\text{sf}}^{\leq 5}(n)$ , which will show that there exist DFA’s accepting suffix-free regular languages with quotient complexity  $n$  and syntactic complexity  $w_{\text{sf}}^{\leq 5}(n)$ .

**Proposition 3.12.** *When  $n \geq 2$ , the semigroup  $\mathbf{W}_{\text{sf}}^{\leq 5}(n)$  is generated by the following set  $\mathbf{G}_{\text{sf}}^{\leq 5}(n)$  of transformations of  $Q$ :*

$$\mathbf{G}_{\text{sf}}^{\leq 5}(2) = \{a_1\}, \text{ where } a_1 = [2, 2];$$

$$\mathbf{G}_{\text{sf}}^{\leq 5}(3) = \{a_1, a_2\}, \text{ where } a_1 = [3, 2, 3] \text{ and } a_2 = [2, 3, 3];$$

and for  $n \geq 4$ ,  $\mathbf{G}_{\text{sf}}^{\leq 5}(n) = \{a_0, \dots, a_{n-1}\}$ , where

- $a_0 = \binom{1}{n}(2, 3)$ ,
- $a_1 = \binom{1}{n}(2, 3, \dots, n-1)$ ,
- For  $2 \leq i \leq n-1$ ,  $ja_i = j+1$  for  $j = 1, \dots, i-1$ ,  $ia_i = n$ , and  $ja_i = j$  for  $j = i+1, \dots, n$ .

For  $n = 4$ ,  $a_0$  and  $a_1$  coincide, and three transformations suffice.

*Proof.* We have  $\mathbf{G}_{\text{sf}}^{\leq 5}(n) \subseteq \mathbf{W}_{\text{sf}}^{\leq 5}(n)$ , and so  $\langle \mathbf{G}_{\text{sf}}^{\leq 5}(n) \rangle$ , the semigroup generated by  $\mathbf{G}_{\text{sf}}^{\leq 5}(n)$ , is a subset of  $\mathbf{W}_{\text{sf}}^{\leq 5}(n)$ . We now show that  $\mathbf{W}_{\text{sf}}^{\leq 5}(n) \subseteq \langle \mathbf{G}_{\text{sf}}^{\leq 5}(n) \rangle$ .

It is easy to verify the cases for  $n = 2, 3$ . Assume  $n \geq 4$ . Pick any  $t$  in  $\mathbf{W}_{\text{sf}}^{\leq 5}(n)$ . Note that  $nt = n$  is fixed. Let  $Q' = Q \setminus \{n\}$ ,  $E_t = \{j \in Q' \mid jt = n\}$ ,  $D_t = Q' \setminus E_t$ , and  $Q'' = Q \setminus \{1, n\}$ . Then  $D_t t \subseteq Q''$ , and  $|E_t| \geq 1$ , since  $|Q''| < |Q'|$ . We prove by induction on  $|E_t|$  that  $t \in \langle \mathbf{G}_{\text{sf}}^{\leq 5}(n) \rangle$ .

First, note that  $\langle a_0, a_1 \rangle$ , the semigroup generated by  $\{a_0, a_1\}$ , is isomorphic to the symmetric group  $\mathfrak{S}_{Q''}$  by Theorem 2.2. Consider  $E_t = \{i\}$  for some  $i \in Q'$ . Then  $ia_i = it = n$ . Moreover, since  $D_t a_i, D_t t \subseteq Q''$ , there exists  $\pi \in \langle a_0, a_1 \rangle$  such that  $(ja_i)\pi = jt$  for all  $j \in D_t$ . Then  $t = a_i \pi \in \langle \mathbf{G}_{\text{sf}}^{\leq 5}(n) \rangle$ .

Assume that any transformation  $t \in \mathbf{W}_{\text{sf}}^{\leq 5}(n)$  with  $|E_t| < k$  can be generated by  $\mathbf{G}_{\text{sf}}^{\leq 5}(n)$ , where  $1 < k < n-1$ . Consider  $t \in \mathbf{W}_{\text{sf}}^{\leq 5}(n)$  with  $|E_t| = k$ . Suppose  $E_t = \{e_1, \dots, e_{k-1}, e_k\}$ . Let  $s \in \mathbf{W}_{\text{sf}}^{\leq 5}(n)$  be such that  $E_s = \{e_1, \dots, e_{k-1}\}$ . By assumption,  $s$  can be generated by  $\mathbf{G}_{\text{sf}}^{\leq 5}(n)$ . Let  $i = e_k s$ ; then  $i \in Q''$ , and  $e_j(sa_i) = n$  for all  $1 \leq j \leq k$ . Moreover, we have  $D_t(sa_i) \subseteq Q''$ . Thus, there exists  $\pi \in \langle a_0, a_1 \rangle$  such that, for all  $d \in D_t$ ,  $d(sa_i\pi) = dt$ . Altogether, for all  $e_j \in E_t$ , we have  $e_j(sa_i\pi) = e_j t = n$ , for all  $d \in D_t$ ,  $d(sa_i\pi) = dt$ , and  $n(sa_i\pi) = nt = n$ . Thus  $t = sa_i\pi$ , and  $t \in \langle \mathbf{G}_{\text{sf}}^{\leq 5}(n) \rangle$ .

Therefore  $\mathbf{W}_{\text{sf}}^{\leq 5}(n) = \langle \mathbf{G}_{\text{sf}}^{\leq 5}(n) \rangle$ . □

**Theorem 3.13.** For  $n \geq 2$ , let  $\mathcal{A}_n = (Q, \Sigma, \delta, 1, \{n-1\})$  be a DFA with alphabet  $\Sigma$ , where each  $a \in \Sigma$  defines a distinct transformation in  $\mathbf{G}_{\text{sf}}^{\leq 5}(n)$  as in Proposition 3.12. Then  $L = L(\mathcal{A}_n)$  has quotient complexity  $\kappa(L) = n$ , and syntactic complexity  $\sigma(L) = w_{\text{sf}}^{\leq 5}(n)$ . Moreover,  $L$  is suffix-free.

*Proof.* The cases for  $n = 2, 3$  are easy to verify. Assume  $n \geq 4$ . First we show that all the states of  $\mathcal{A}_n$  are reachable: 1 is the initial state, state  $n$  is reached by  $a_1$ , and for  $2 \leq i \leq n-1$ , state  $i$  is reached by  $a_i^{i-1}$ . For  $1 \leq i \leq n-1$ , the word  $a_{n-1}^{n-1-i}$  is accepted only by state  $i$ . Also  $n$  is the empty state. Thus all the states of  $\mathcal{A}_n$  are distinct, and  $\kappa(L) = n$ .

By Proposition 3.12, the syntactic semigroup of  $L$  is  $\mathbf{W}_{\text{sf}}^{\leq 5}(n)$ ; hence  $\sigma(L) = |\mathbf{W}_{\text{sf}}^{\leq 5}(n)| = w_{\text{sf}}^{\leq 5}(n)$ . By Proposition 3.3,  $L$  is suffix-free.  $\square$

As shown in Table 3.2 on p. 32, the size of  $\Sigma$  cannot be decreased for  $n \leq 5$ .

**Theorem 3.14.** For  $2 \leq n \leq 5$ , if a suffix-free regular language  $L$  has quotient complexity  $\kappa(L) = n$ , then its syntactic complexity satisfies  $\sigma(L) \leq w_{\text{sf}}^{\leq 5}(n)$ , and this is a tight upper bound.

*Proof.* By Proposition 3.3, the syntactic semigroup of a suffix-free regular language  $L$  is contained in  $\mathbf{B}_{\text{sf}}(n)$ . For  $n \in \{2, 3\}$ ,  $w_{\text{sf}}^{\leq 5}(n) = b_{\text{sf}}(n)$ . So  $w_{\text{sf}}^{\leq 5}(n)$  is an upper bound, and it is met by the language  $L = \varepsilon$  for  $n = 2$  and by  $L = ab^*$  for  $n = 3$ . For  $n = 4$ , we have  $|\mathbf{B}_{\text{sf}}(4)| = 15$  and  $|\mathbf{W}_{\text{sf}}^{\leq 5}(4)| = 13$ . Two transformations,  $s_1 = [4, 2, 2, 4]$  and  $s_2 = [4, 3, 3, 4]$ , in  $\mathbf{B}_{\text{sf}}(4)$  are such that  $s_1$  conflicts with  $t_1 = [3, 2, 4, 4] \in \mathbf{W}_{\text{sf}}^{\leq 5}(4)$  (because  $t_1 s_1 = [2, 2, 4, 4] \notin \mathbf{B}_{\text{sf}}(4)$ ), and  $s_2$  conflicts with  $t_2 = [2, 3, 4, 4]$  (because  $t_2 s_2 = [3, 3, 4, 4] \notin \mathbf{B}_{\text{sf}}(4)$ ). Thus  $\sigma(L) \leq 13$ . Let  $L$  be the language accepted by the DFA  $\mathcal{A}_4$  in Theorem 3.13; then  $\kappa(L) = 4$  and  $\sigma(L) = 13$ . So the bound is tight.

For  $n = 5$ , we have  $|\mathbf{B}_{\text{sf}}(5)| = 115$  and  $|\mathbf{W}_{\text{sf}}^{\leq 5}(5)| = 73$ . Suppose  $\mathbf{B}_{\text{sf}}(5) \setminus \mathbf{W}_{\text{sf}}^{\leq 5}(5) = \{s_1, \dots, s_{42}\}$ . For each  $s_i$ , we enumerated transformations in  $\mathbf{W}_{\text{sf}}^{\leq 5}(5)$  using *GAP* and found a unique  $t_i \in \mathbf{W}_{\text{sf}}^{\leq 5}(5)$  such that the semigroup  $\langle t_i, s_i \rangle$  is not contained in  $\mathbf{B}_{\text{sf}}(5)$ . Thus at most one transformation in each pair  $\{t_i, s_i\}$  can appear in the syntactic semigroup of  $L$ . So we reduce the upper bound to 73. By Theorem 3.13, this bound is tight.  $\square$

When  $n \geq 6$ , the semigroup  $\mathbf{W}_{\text{sf}}^{\leq 5}(n)$  is no longer the largest semigroup contained in  $\mathbf{B}_{\text{sf}}(n)$ ; hence the upper bound in Theorem 3.14 does not apply. In the following, we define and study another semigroup  $\mathbf{W}_{\text{sf}}^{\geq 6}(n)$ , which is larger than  $\mathbf{W}_{\text{sf}}^{\leq 5}(n)$  and is also contained in  $\mathbf{B}_{\text{sf}}(n)$ . For  $n \geq 2$ , let

$$\mathbf{W}_{\text{sf}}^{\geq 6}(n) = \{t \in \mathbf{B}_{\text{sf}}(n) \mid 1t = n \text{ or } it = n \ \forall i, 2 \leq i \leq n-1\}.$$

**Proposition 3.15.** For  $n \geq 2$ , the set  $\mathbf{W}_{\text{sf}}^{\geq 6}(n)$  is a semigroup contained in  $\mathbf{B}_{\text{sf}}(n)$ , and its cardinality is

$$w_{\text{sf}}^{\geq 6}(n) = |\mathbf{W}_{\text{sf}}^{\geq 6}(n)| = (n-1)^{n-2} + (n-2).$$

*Proof.* Pick any  $t_1, t_2$  in  $\mathbf{W}_{\text{sf}}^{\geq 6}(n)$ . If  $1t_1 = n$ , then  $1(t_1t_2) = n$  and  $t_1t_2 \in \mathbf{W}_{\text{sf}}^{\geq 6}(n)$ . If  $1t_1 \neq n$ , then, for all  $i \in \{2, \dots, n-1\}$ ,  $it_1 = n$  and  $i(t_1t_2) = n$ ; so  $t_1t_2 \in \mathbf{W}_{\text{sf}}^{\geq 6}(n)$  as well. Hence  $\mathbf{W}_{\text{sf}}^{\geq 6}(n)$  is a semigroup contained in  $\mathbf{B}_{\text{sf}}(n)$ .

For any  $t \in \mathbf{W}_{\text{sf}}^{\geq 6}(n)$ ,  $nt = n$  is fixed. There are two possible cases:

1.  $1t = n$ : For each  $i \in \{2, \dots, n-1\}$ ,  $it$  can be chosen from  $\{2, \dots, n\}$ . Then there are  $(n-1)^{n-2}$  different  $t$ 's in this case.
2.  $1t \neq n$ : Now  $1t$  can be chosen from  $\{2, \dots, n-1\}$ . For each  $i \in \{2, \dots, n-1\}$ ,  $it = n$  is fixed. Then, for any  $t' \in \mathbf{W}_{\text{sf}}^{\geq 6}(n)$  such that  $1t' \neq n$ ,  $t$  differs from  $t'$  if and only if  $1t \neq 1t'$ . So there are  $n-2$  different  $t$ 's in this case.

Therefore  $w_{\text{sf}}^{\geq 6}(n) = (n-1)^{n-2} + (n-2)$ . □

When  $n \geq 6$ , one verifies that  $w_{\text{sf}}^{\geq 6}(n) > w_{\text{sf}}^{\leq 5}(n)$ . Hence  $\mathbf{W}_{\text{sf}}^{\geq 6}(n)$  is a larger semigroup than  $\mathbf{W}_{\text{sf}}^{\leq 5}(n)$ . Table 3.2 on p. 32 contains values of  $w_{\text{sf}}^{\leq 5}(n)$  and  $w_{\text{sf}}^{\geq 6}(n)$  for small  $n$ 's. For  $n \in \{2, 3\}$ , we have  $\mathbf{W}_{\text{sf}}^{\geq 6}(n) = \mathbf{W}_{\text{sf}}^{\leq 5}(n)$ . From now on, we are only interested in larger values of  $n$ .

**Proposition 3.16.** For  $n \geq 4$ , the semigroup  $\mathbf{W}_{\text{sf}}^{\geq 6}(n)$  is generated by the set  $\mathbf{G}_{\text{sf}}^{\geq 6}(n) = \{a_1, a_2, a_3, b_1, \dots, b_{n-2}, c\}$  of transformations of  $Q$ , where

- $a_1 = \binom{1}{n}(2, \dots, n-1)$ ,  $a_2 = \binom{1}{n}(2, 3)$ ,  $a_3 = \binom{1}{n} \binom{n-1}{2}$ ;
- For  $1 \leq i \leq n-2$ ,  $b_i = \binom{1}{n} \binom{i+1}{n}$ ;
- $c = \binom{Q \setminus \{1\}}{n} \binom{1}{2} = [2, n, \dots, n]$ .

For  $n = 4$ ,  $a_1$  and  $a_2$  coincide, and five transformations suffice.

*Proof.* Clearly  $\mathbf{G}_{\text{sf}}^{\geq 6}(n) \subseteq \mathbf{W}_{\text{sf}}^{\geq 6}(n)$ , and  $\langle \mathbf{G}_{\text{sf}}^{\geq 6}(n) \rangle \subseteq \mathbf{W}_{\text{sf}}^{\geq 6}(n)$ . We show in the following that  $\mathbf{W}_{\text{sf}}^{\geq 6}(n) \subseteq \langle \mathbf{G}_{\text{sf}}^{\geq 6}(n) \rangle$ .

Let  $Q' = \{2, \dots, n-1\}$ . By Theorem 2.1,  $a_1, a_2$  and  $a_3$  together generate the semigroup

$$\mathbf{Y} = \{t \in \mathbf{W}_{\text{sf}}^{\geq 6}(n) \mid \text{for all } i \in Q', it \in Q'\},$$

which is isomorphic to  $\mathcal{T}_{Q'}$  and is contained in  $\mathbf{W}_{\text{sf}}^{\geq 6}(n)$ . Next, consider any  $t \in \mathbf{W}_{\text{sf}}^{\geq 6}(n) \setminus \mathbf{Y}$ . We have two cases:

1.  $1t = n$ : Let  $E_t = \{i \in Q' \mid it = n\}$ . Since  $t \notin \mathbf{Y}$ ,  $E_t \neq \emptyset$ . Suppose  $E_t = \{i_1, \dots, i_k\}$ , for some  $1 \leq k \leq n-2$ . Then there exists  $t' \in \mathbf{Y}$  such that, for all  $i \notin E_t$ ,  $it' = it$ . Let  $s = b_{i_1-1} \cdots b_{i_k-1}$ . Note that  $E_t s = \{n\}$ , and, for all  $i \notin E_t$ ,  $i(t's) = (it')s = it$ . So  $t = t's \in \langle \mathbf{G}_{\text{sf}}^{\geq 6}(n) \rangle$ .
2.  $1t \neq n$ : If  $1t = 2$ , then  $t = c$ . Otherwise,  $1t \in \{3, \dots, n-1\} \subseteq Q'$ , and we know from the above case that there exists  $t' \in \mathbf{G}_{\text{sf}}^{\geq 6}(n)$  such that  $2t' = 1t$ . Then  $1(ct') = 1t$ , and  $i(ct') = (ic)t' = n = it$ , for all  $i \in Q'$ . Hence  $t = ct' \in \langle \mathbf{G}_{\text{sf}}^{\geq 6}(n) \rangle$ .

Therefore  $\langle a_1, a_2, a_3, b_1, \dots, b_{n-2}, c \rangle = \mathbf{W}_{\text{sf}}^{\geq 6}(n)$ . □

**Theorem 3.17.** *For  $n \geq 4$ , let  $\mathcal{A}'_n = (Q, \Sigma, \delta, 1, \{2\})$  be a DFA with alphabet  $\Sigma = \{a_1, a_3, b_1, b_2, c\}$  if  $n = 4$  or  $\Sigma = \{a_1, a_2, a_3, b_1, \dots, b_{n-2}, c\}$  if  $n \geq 5$ , where each letter defines a transformation as in Proposition 3.16. Then  $L' = L(\mathcal{A}'_n)$  is suffix-free with quotient complexity  $\kappa(L') = n$  and syntactic complexity  $\sigma(L') = w_{\text{sf}}^{\geq 6}(n)$ .*

*Proof.* First we show that  $\kappa(L') = n$ . From the initial state, we can reach state 2 by  $c$  and state  $n$  by  $a_1$ . From state 2 we can reach state  $i$ ,  $3 \leq i \leq n-1$ , by  $a_1^{i-1}$ . So all the states in  $Q$  are reachable. The word  $c$  is accepted only by state 1. For  $2 \leq i \leq n-1$ , the word  $a_1^{n-i}$  is accepted only by state  $i$ . State  $n$  is the empty state, which rejects all words. Thus all the states in  $Q$  are distinct.

By Proposition 3.16, the syntactic semigroup of  $L'$  is  $\mathbf{W}_{\text{sf}}^{\geq 6}(n)$ , and  $\sigma(L') = w_{\text{sf}}^{\geq 6}(n)$ . Also  $L'$  is suffix-free by Proposition 3.3. □

**Theorem 3.18.** *If  $L$  is a suffix-free regular language with  $\kappa(L) = 6$ , then  $\sigma(L) \leq w_{\text{sf}}^{\geq 6}(6) = 629$  and this is a tight bound.*

*Proof.* Note that  $|\mathbf{B}_{\text{sf}}(6)| = 1169$  and  $|\mathbf{W}_{\text{sf}}^{\geq 6}(6)| = 629$ . Suppose  $\{s_1, \dots, s_{540}\} = \mathbf{B}_{\text{sf}}(6) \setminus \mathbf{W}_{\text{sf}}^{\geq 6}(6)$ . For each  $i$ , we enumerated transformations in  $\mathbf{W}_{\text{sf}}^{\geq 6}(6)$  using *GAP* and found a unique  $t_i \in \mathbf{W}_{\text{sf}}^{\geq 6}(6)$  such that  $\langle t_i, s_i \rangle \not\subseteq \mathbf{B}_{\text{sf}}(6)$ . As in the proof of Theorem 3.14, for each  $i$ , at most one transformation in  $\{t_i, s_i\}$  can appear in the syntactic semigroup of  $L$ . Then we can reduce the upper bound to 629. This bound is met by the language  $L'$  in Theorem 3.17; so it is tight. □

We know that the upper bound on the syntactic complexity of suffix-free regular languages is achieved by the largest semigroup contained in  $\mathbf{B}_{\text{sf}}(n)$ . We conjecture that  $\mathbf{W}_{\text{sf}}^{\geq 6}(n)$  is such a semigroup also for  $n \geq 7$ .

**Conjecture 3.19** (Suffix-Free Regular Languages). *If  $L$  is a suffix-free regular language with  $\kappa(L) = n \geq 7$ , then  $\sigma(L) \leq w_{\text{sf}}^{\geq 6}(n)$ .*

### 3.3 Bifix-Free Regular Languages

Let  $L$  be a regular bifix-free language with  $\kappa(L) = n$ . From Sections 3.1 and 3.2 we have:

1.  $L$  has  $\varepsilon$  as a quotient, and this is the only final quotient;
2.  $L$  has  $\emptyset$  as a quotient;
3.  $L$  as a quotient is uniquely reachable.

Let  $\mathcal{A}$  be the quotient DFA of  $L$ , with  $Q$  as the set of states. We assume that 1 is the initial state,  $n - 1$  corresponds to the quotient  $\varepsilon$ , and  $n$  is the empty state. For  $n \geq 2$ , consider the set

$$\mathbf{B}_{\text{bf}}(n) = \{t \in \mathbf{B}_{\text{sf}}(n) \mid (n - 1)t = n\}.$$

The following is an observation similar to Proposition 3.3.

**Proposition 3.20.** *If  $L$  is a regular language with quotient complexity  $n$  and syntactic semigroup  $T_L$ , then the following hold:*

1. If  $L$  is bifix-free, then  $T_L$  is a subset of  $\mathbf{B}_{\text{bf}}(n)$ .
2. If  $\varepsilon$  is the only final quotient of  $L$ , and  $T_L \subseteq \mathbf{B}_{\text{bf}}(n)$ , then  $L$  is bifix-free.

*Proof.* 1. Since  $L$  is suffix-free,  $T_L \subseteq \mathbf{B}_{\text{sf}}(n)$ . Since  $L$  is also prefix-free, it has  $\varepsilon$  and  $\emptyset$  as quotients. By assumption,  $n - 1 \in Q$  corresponds to the quotient  $\varepsilon$ . Thus for any  $t \in T_L$ ,  $(n - 1)t = n$ , and so  $T_L \subseteq \mathbf{B}_{\text{bf}}(n)$ .

2. Since  $\varepsilon$  is the only final quotient of  $L$ ,  $L$  is prefix-free, and  $L$  has the empty quotient. Since  $T_L \subseteq \mathbf{B}_{\text{bf}}(n) \subseteq \mathbf{B}_{\text{sf}}(n)$ ,  $L$  is suffix-free by Proposition 3.3. Therefore  $L$  is bifix-free.  $\square$

**Lemma 3.21.** *We have  $|\mathbf{B}_{\text{bf}}(2)| = 1$ , and for  $n \geq 3$ ,  $|\mathbf{B}_{\text{bf}}(n)| = M_n + N_n$ , where*

$$M_n = \sum_{k=1}^{n-2} C_{k-1}^{n-3} (k-1)! \sum_{\substack{r_2+\dots+r_k+r \\ =n-k-2}} C_{r_2,\dots,r_k,r}^{n-k-2} (r+1)^r \prod_{j=2}^k S'_{r_j+1}(j-1), \quad (3.2)$$

$$N_n = \sum_{k=0}^{n-3} C_k^{n-3} k! \sum_{\substack{r_2+\dots+r_k+r \\ =n-k-3}} C_{r_2,\dots,r_k,r}^{n-k-3} (r+2)^r \prod_{j=2}^k S'_{r_j+1}(j-1). \quad (3.3)$$

*Proof.* It is easy to see that  $\mathbf{B}_{\text{sf}}(2) = \{[2, 2]\}$ . Assume  $n \geq 3$ . Let  $t$  be any transformation in  $\mathbf{B}_{\text{bf}}(n)$ . Suppose  $\mathfrak{s}_t(1) = 1, 1t, \dots, 1t^k, n$ , where  $0 \leq k \leq n - 2$ . For  $2 \leq j \leq k$ , suppose tree  $T_t(j)$  contains  $r_j + 1$  nodes, for some  $r_j \geq 0$ ; then there are  $S'_{r_j+1}(j - 1)$  different trees  $T_t(j)$ . Let  $E$  be the set of elements of  $Q$  that are not in any tree  $T_t(j)$  nor in the principal sequence  $\mathfrak{s}_t(1)$ . Then there are two cases:

1.  $n - 1 \in \mathfrak{s}_t(1)$ : Since  $(n - 1)t = n$ , we must have  $1t^k = n - 1$ , and  $k \geq 1$ . So there are  $C_{k-1}^{n-3}(k - 1)!$  different  $\mathfrak{s}_t(1)$ . Let  $r = |E| = (n - k - 2) - (r_2 + \dots + r_k)$ . Then there are  $C_{r_2, \dots, r_k, r}^{n-k-2}$  tuples  $(r_2, \dots, r_k, r)$ . For any  $x \in E$ , its image  $xt$  can be chosen from  $E \cup \{n\}$ . Then the number of transformations  $t$  in this case is  $M_n$ .
2.  $n - 1 \notin \mathfrak{s}_t(1)$ : Then  $k \leq n - 3$ , and there are  $C_k^{n-3}k!$  different  $\mathfrak{s}_t(1)$ . Note that  $n - 1 \in E$ , and  $(n - 1)t = n$  is fixed. Let  $r = |E \setminus \{n - 1\}| = (n - k - 3) - (r_2 + \dots + r_k)$ . Then there are  $C_{r_2, \dots, r_k, r}^{n-k-3}$  tuples  $(r_2, \dots, r_k, r)$ . For any  $x \in E \setminus \{n - 1\}$ ,  $xt$  can be chosen from  $E \cup \{n\}$ . Thus the number of transformations  $t$  in this case is  $N_n$ .

Altogether we have the desired formula. □

Let  $\text{b}_{\text{bf}}(n) = |\mathbf{B}_{\text{bf}}(n)|$ . From Proposition 3.20 and Lemma 3.21 we have

**Proposition 3.22.** *For  $n \geq 2$ , if  $L$  is a bifix-free regular language with quotient complexity  $n$ , then its syntactic complexity  $\sigma(L)$  satisfies  $\sigma(L) \leq \text{b}_{\text{bf}}(n)$ , where  $\text{b}_{\text{bf}}(n)$  is the cardinality of  $\mathbf{B}_{\text{bf}}(n)$  as in Lemma 3.21.*

When  $2 \leq n \leq 4$ , the set  $\mathbf{B}_{\text{bf}}(n)$  is a semigroup. But for  $n \geq 5$ , it is not a semigroup because  $s_1 = [2, 3, n, \dots, n, n]$ ,  $s_2 = [n, 3, 3, n, \dots, n, n] \in \mathbf{B}_{\text{bf}}(n)$  while  $s_1 s_2 = [3, 3, n, \dots, n, n] \notin \mathbf{B}_{\text{bf}}(n)$ . Hence  $\text{b}_{\text{bf}}(n)$  is not a tight upper bound on the syntactic complexity of bifix-free regular languages in general. We look for a large semigroup contained in  $\mathbf{B}_{\text{bf}}(n)$  that can be the syntactic semigroup of a bifix-free regular language. For  $n \geq 2$ , let

$$\mathbf{W}_{\text{bf}}^{\leq 5}(n) = \{t \in \mathbf{B}_{\text{bf}}(n) \mid \text{for all } i, j \in Q \text{ where } i \neq j, \\ \text{we have } it = jt = n \text{ or } it \neq jt\}.$$

(The reason for using the superscript  $\leq 5$  will be made clear in Theorem 3.27.)

**Proposition 3.23.** *For  $n \geq 2$ ,  $\mathbf{W}_{\text{bf}}^{\leq 5}(n)$  is a semigroup contained in  $\mathbf{B}_{\text{bf}}(n)$  with cardinality*

$$w_{\text{bf}}^{\leq 5}(n) = |\mathbf{W}_{\text{bf}}^{\leq 5}(n)| = \sum_{k=0}^{n-2} (C_k^{n-2})^2 (n - 2 - k)!$$

*Proof.* First, note that  $\mathbf{W}_{\text{bf}}^{\leq 5}(n) = \mathbf{W}_{\text{sf}}^{\leq 5}(n) \cap \mathbf{B}_{\text{bf}}(n)$ , and that  $\mathbf{W}_{\text{sf}}^{\leq 5}(n)$  is a semigroup contained in  $\mathbf{B}_{\text{sf}}(n)$  by Proposition 3.10. For any  $t_1, t_2 \in \mathbf{W}_{\text{bf}}^{\leq 5}(n)$ , we have  $t_1 t_2 \in \mathbf{W}_{\text{sf}}^{\leq 5}(n)$ , and  $(n-1)t_1 t_2 = nt_2 = n$ ; so  $t_1 t_2 \in \mathbf{B}_{\text{bf}}(n)$ . Then  $t_1 t_2 \in \mathbf{W}_{\text{bf}}^{\leq 5}(n)$ , and  $\mathbf{W}_{\text{bf}}^{\leq 5}(n)$  is a semigroup contained in  $\mathbf{B}_{\text{bf}}(n)$ .

Pick any  $t \in \mathbf{W}_{\text{bf}}^{\leq 5}(n)$ . Note that  $(n-1)t = n$  and  $nt = n$  are fixed, and  $1 \notin \text{rng}(t)$ . Let  $Q' = Q \setminus \{n-1, n\}$ ,  $E = \{i \in Q' \mid it = n\}$ , and  $D = Q' \setminus E$ . Suppose  $|E| = k$ , where  $0 \leq k \leq n-2$ ; then there are  $C_k^{n-2}$  choices of  $E$ . Elements of  $D$  are mapped to pairwise different elements of  $Q \setminus \{1, n\}$ ; then there are  $C_{n-2-k}^{n-2} (n-2-k)!$  different mappings  $t|_D$ . Altogether, we have  $|\mathbf{W}_{\text{bf}}^{\leq 5}(n)| = \sum_{k=0}^{n-2} (C_k^{n-2})^2 (n-2-k)!$   $\square$

*Remark 3.24.* Assume  $n \geq 3$ , and let  $Q' = Q \setminus \{n-1, n\}$ . Then the semigroup  $\mathbf{W}_{\text{bf}}^{\leq 5}(n)$  is isomorphic to the symmetric inverse semigroup  $\mathcal{IS}_{Q'}$ ; so  $w_{\text{bf}}^{\leq 5}(n) = |\mathcal{IS}_{Q'}|$ . Janson and Mazorchuk [26] showed that, for large  $n$ , the number  $w_{\text{bf}}^{\leq 5}(n)$  is asymptotically

$$w_{\text{bf}}^{\leq 5}(n) \sim \frac{1}{\sqrt{2e}} e^{2\sqrt{n-2}-n+2} (n-2)^{n-\frac{7}{4}}.$$

**Proposition 3.25.** *For  $n \geq 2$ , let  $Q' = Q \setminus \{n-1, n\}$  and  $Q'' = Q \setminus \{1, n\}$ . Then the semigroup  $\mathbf{W}_{\text{bf}}^{\leq 5}(n)$  is generated by*

$$\mathbf{G}_{\text{bf}}^{\leq 5}(n) = \{t \in \mathbf{W}_{\text{bf}}^{\leq 5}(n) \mid Q't = Q'' \text{ and } it \neq jt \text{ for all } i, j \in Q'\}.$$

*Proof.* The case for  $n = 2$  is trivial since  $\mathbf{G}_{\text{bf}}^{\leq 5}(2) = \mathbf{W}_{\text{bf}}^{\leq 5}(2)$ . Assume  $n \geq 3$ . We want to show that  $\mathbf{W}_{\text{bf}}^{\leq 5}(n) = \langle \mathbf{G}_{\text{bf}}^{\leq 5}(n) \rangle$ . Since  $\mathbf{G}_{\text{bf}}^{\leq 5}(n) \subseteq \mathbf{W}_{\text{bf}}^{\leq 5}(n)$ , we have  $\langle \mathbf{G}_{\text{bf}}^{\leq 5}(n) \rangle \subseteq \mathbf{W}_{\text{bf}}^{\leq 5}(n)$ . Let  $t \in \mathbf{W}_{\text{bf}}^{\leq 5}(n)$ . By definition,  $(n-1)t = nt = n$ . Let  $E_t = \{i \in Q' \mid it = n\}$ . If  $E_t = \emptyset$ , then  $t \in \mathbf{G}_{\text{bf}}^{\leq 5}(n)$ ; otherwise, there exists  $x \in Q''$  such that  $x \notin \text{rng}(t)$ . We prove by induction on  $|E_t|$  that  $t \in \langle \mathbf{G}_{\text{bf}}^{\leq 5}(n) \rangle$ .

First note that, for all  $t \in \mathbf{G}_{\text{bf}}^{\leq 5}(n)$ ,  $t|_{Q'}$  is an injective mapping from  $Q'$  to  $Q''$ . Consider  $E_t = \{i\}$  for some  $i \in Q'$ . Since  $|E_t| = 1$ ,  $\text{rng}(t) \cup \{x\} = Q''$ . Let  $t_1, t_2 \in \mathbf{G}_{\text{bf}}^{\leq 5}(n)$  be defined by

1.  $jt_1 = j+1$  for  $j = 1, \dots, i-1$ ,  $it_1 = n-1$ ,  $jt_1 = j$  for  $j = i+1, \dots, n-2$ ,
2.  $1t_2 = x$ ,  $jt_2 = (j-1)t$  for  $j = 2, \dots, i$ ,  $jt_2 = jt$  for  $j = i+1, \dots, n-2$ .

Then  $t_1 t_2 = t$ , and  $t \in \langle \mathbf{G}_{\text{bf}}^{\leq 5}(n) \rangle$ .

Assume that any transformation  $t \in \mathbf{W}_{\text{bf}}^{\leq 5}(n)$  with  $|E_t| < k$  can be generated by  $\mathbf{G}_{\text{bf}}^{\leq 5}(n)$ , where  $1 < k < n-2$ . Consider  $t \in \mathbf{W}_{\text{bf}}^{\leq 5}(n)$  with  $|E_t| = k$ . Suppose  $E_t = \{e_1, \dots, e_{k-1}, e_k\}$ ,



and let  $D_t = Q' \setminus E_t = \{d_1, \dots, d_l\}$ , where  $l = n - 2 - k$ . By assumption, all  $s \in \mathbf{W}_{\text{bf}}^{\leq 5}(n)$  with  $|E_s| = k - 1$  can be generated by  $\mathbf{G}_{\text{bf}}^{\leq 5}(n)$ . Let  $s$  be such that  $E_s = \{1, \dots, k - 1\}$ ; then  $1s = \dots = (k - 1)s = n$ . In addition, let  $ks = x$ , and let  $(k + j)s = d_j t$  for  $j = 1, \dots, l$ . Let  $t' \in \mathbf{G}_{\text{bf}}^{\leq 5}(n)$  be such that  $e_j t' = j$  for  $j = 1, \dots, k - 1$ ,  $kt' = n - 1$ , and  $d_j t' = k + j$  for  $j = 1, \dots, l$ . Then  $t's = t$ , and  $t \in \langle \mathbf{G}_{\text{bf}}^{\leq 5}(n) \rangle$ . Therefore,  $\mathbf{W}_{\text{bf}}^{\leq 5}(n) = \langle \mathbf{G}_{\text{bf}}^{\leq 5}(n) \rangle$ .  $\square$

**Theorem 3.26.** *For  $n \geq 2$ , let  $\mathcal{A}_n = (Q, \Sigma, \delta, 1, \{n - 1\})$  be a DFA with alphabet  $\Sigma$  of size  $(n - 2)!$ , where each  $a \in \Sigma$  defines a distinct transformation  $t_a \in \mathbf{G}_{\text{bf}}^{\leq 5}(n)$ . Then  $L = L(\mathcal{A}_n)$  has quotient complexity  $\kappa(L) = n$ , and syntactic complexity  $\sigma(L) = w_{\text{bf}}^{\leq 5}(n)$ . Moreover,  $L$  is bifix-free.*

*Proof.* The case for  $n = 2$  is easy to verify. Assume  $n \geq 3$ . We first show that all the states of  $\mathcal{A}_n$  are reachable. Note that there exists  $a \in \Sigma$  such that  $t_a = [2, \dots, n - 1, n, n] \in \mathbf{G}_{\text{bf}}^{\leq 5}(n)$ . State  $1 \in Q$  is the initial state, and  $a^{i-1}$  reaches state  $i \in Q$  for  $i = 2, \dots, n$ . Furthermore, for  $1 \leq i \leq n - 1$ , state  $i$  accepts  $a^{n-1-i}$ , while for  $j \neq i$ , state  $j$  rejects it. Also,  $n$  is the empty state. Thus all the states of  $\mathcal{A}_n$  are distinct, and  $\kappa(L) = n$ .

By Proposition 3.25, the syntactic semigroup of  $L$  is  $\mathbf{W}_{\text{bf}}^{\leq 5}(n)$ . So the syntactic complexity of  $L$  is  $\sigma(L) = w_{\text{bf}}^{\leq 5}(n)$ . By Proposition 3.20,  $L$  is bifix-free.  $\square$

**Theorem 3.27.** *For  $2 \leq n \leq 5$ , if a bifix-free regular language  $L$  has quotient complexity  $\kappa(L) = n$ , then  $\sigma(L) \leq w_{\text{bf}}^{\leq 5}(n)$ , and this bound is tight.*

*Proof.* We know by Proposition 3.20 that the upper bound on the syntactic complexity of bifix-free regular languages is reached by the largest semigroup contained in  $\mathbf{B}_{\text{bf}}(n)$ . Since  $w_{\text{bf}}^{\leq 5}(n) = b_{\text{bf}}(n)$  for  $n = 2, 3$ , and  $4$ ,  $w_{\text{bf}}^{\leq 5}(n)$  is an upper bound, and it is tight by Theorem 3.26.

For  $n = 5$ , we have  $b_{\text{bf}}(5) = |\mathbf{B}_{\text{bf}}(5)| = 41$ , and  $w_{\text{bf}}^{\leq 5}(5) = |\mathbf{W}_{\text{bf}}^{\leq 5}(5)| = 34$ . Let  $\mathbf{B}_{\text{bf}}(5) \setminus \mathbf{W}_{\text{bf}}^{\leq 5}(5) = \{\tau_1, \dots, \tau_7\}$ . We found for each  $\tau_i$  a unique  $t_i \in \mathbf{W}_{\text{bf}}^{\leq 5}(5)$  such that the semigroup  $\langle \tau_i, t_i \rangle$  is not a subset of  $\mathbf{B}_{\text{bf}}(5)$ :

$$\begin{array}{ll}
\tau_1 = [2, 4, 4, 5, 5], & t_1 = [3, 4, 2, 5, 5]; \\
\tau_2 = [3, 4, 4, 5, 5], & t_2 = [3, 5, 2, 5, 5]; \\
\tau_3 = [4, 2, 2, 5, 5], & t_3 = [2, 4, 3, 5, 5]; \\
\tau_4 = [4, 3, 3, 5, 5], & t_4 = [2, 5, 3, 5, 5]; \\
\tau_5 = [5, 2, 2, 5, 5], & t_5 = [3, 2, 4, 5, 5]; \\
\tau_6 = [5, 3, 3, 5, 5], & t_6 = [2, 3, 4, 5, 5]; \\
\tau_7 = [5, 4, 4, 5, 5], & t_7 = [3, 2, 5, 5, 5].
\end{array}$$

Since  $\langle \tau_i, t_i \rangle \subseteq T_L$ , if both  $\tau_i$  and  $t_i$  are in  $T_L$ , then  $T_L \not\subseteq \mathbf{B}_{\text{bf}}(5)$ , and  $L$  is not bifix-free by Proposition 3.20. Thus, for  $1 \leq i \leq 7$ , at most one of  $\tau_i$  and  $t_i$  can appear in  $T_L$ , and  $|T_L| \leq 34$ . Since  $|\mathbf{W}_{\text{bf}}^{\leq 5}(5)| = 34$  and  $\mathbf{W}_{\text{bf}}^{\leq 5}(5)$  is a semigroup, we have  $\sigma(L) \leq 34 = w_{\text{bf}}^{\leq 5}(5)$  as the upper bound for  $n = 5$ . This bound is reached by the DFA  $\mathcal{A}_5$  in Theorem 3.26.  $\square$

When  $n \geq 6$ , the semigroup  $\mathbf{W}_{\text{bf}}^{\leq 5}(n)$  is no longer the largest semigroup contained in  $\mathbf{B}_{\text{bf}}(n)$ , and the upper bound in Theorem 3.27 does not apply. We find another large semigroup  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  suitable for bifix-free regular languages. For  $n \geq 3$ , let

$$\begin{aligned} \mathbf{U}_n^1 &= \{t \in \mathbf{B}_{\text{bf}}(n) \mid 1t = n\}, \\ \mathbf{U}_n^2 &= \{t \in \mathbf{B}_{\text{bf}}(n) \mid 1t = n - 1\}, \\ \mathbf{U}_n^3 &= \{t \in \mathbf{B}_{\text{bf}}(n) \mid 1t \notin \{n, n - 1\}, \text{ and } it \in \{n - 1, n\} \text{ for all } i \neq 1\}, \end{aligned}$$

and let  $\mathbf{W}_{\text{bf}}^{\geq 6}(n) = \mathbf{U}_n^1 \cup \mathbf{U}_n^2 \cup \mathbf{U}_n^3$ . Note that  $\mathbf{U}_3^3 = \emptyset$ .

**Proposition 3.28.** *For  $n \geq 3$ ,  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  is a semigroup contained in  $\mathbf{B}_{\text{bf}}(n)$  with cardinality*

$$w_{\text{bf}}^{\geq 6}(n) = |\mathbf{W}_{\text{bf}}^{\geq 6}(n)| = (n - 1)^{n-3} + (n - 2)^{n-3} + (n - 3)2^{n-3}.$$

*Proof.* First we show that  $\mathbf{U}_n^1$  is a semigroup. For any  $t_1, t'_1 \in \mathbf{U}_n^1$ , since  $1(t_1 t'_1) = (1t_1)t'_1 = nt'_1 = n$ , we have  $t_1 t'_1 \in \mathbf{U}_n^1$ . Next, let  $t_2 \in \mathbf{U}_n^2$  and  $t \in \mathbf{U}_n^1 \cup \mathbf{U}_n^2$ . If  $t \in \mathbf{U}_n^1$ , then  $1(t_2 t) = (n - 1)t = n$  and  $1(tt_2) = nt_2 = n$ ; so  $t_2 t, tt_2 \in \mathbf{U}_n^1$ . If  $t \in \mathbf{U}_n^2$ , then  $1(t_2 t) = (n - 1)t = n$  and  $1(tt_2) = (n - 1)t_2 = n$ ; so  $t_2 t, tt_2 \in \mathbf{U}_n^1$  as well. Thus  $\mathbf{U}_n^1 \cup \mathbf{U}_n^2$  is also a semigroup. For any  $t_3 \in \mathbf{U}_n^3$  and  $t' \in \mathbf{W}_{\text{bf}}^{\geq 6}(n)$ , since  $it_3 \in \{n - 1, n\}$  for all  $i \neq 1$ , and  $(n - 1)t' = nt' = n$ , we have  $i(t_3 t') = n$ , and  $t_3 t' \in \mathbf{W}_{\text{bf}}^{\geq 6}(n)$ . Also  $1(t' t_3) = (1t')t_3 \in \{n - 1, n\}$ , so  $t' t_3 \in \mathbf{U}_n^1 \cup \mathbf{U}_n^2$ . Hence  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  is a semigroup contained in  $\mathbf{B}_{\text{bf}}(n)$ .

Note that  $\mathbf{U}_n^1$ ,  $\mathbf{U}_n^2$ , and  $\mathbf{U}_n^3$  are pairwise disjoint. For any  $t \in \mathbf{W}_{\text{bf}}^{\geq 6}(n)$ , there are three cases:

1.  $t \in \mathbf{U}_n^1$ : For any  $i \notin \{1, n - 1, n\}$ ,  $it$  can be chosen from  $Q \setminus \{1\}$ . Then  $|\mathbf{U}_n^1| = (n - 1)^{n-3}$ ;
2.  $t \in \mathbf{U}_n^2$ : For any  $i \notin \{1, n - 1, n\}$ ,  $it$  can be chosen from  $Q \setminus \{1, n - 1\}$ . Then  $|\mathbf{U}_n^2| = (n - 2)^{n-3}$ ;
3.  $t \in \mathbf{U}_n^3$ : Now,  $1t$  can be chosen from  $Q \setminus \{1, n - 1, n\}$ . For any  $i \notin \{1, n - 1, n\}$ ,  $it$  has two choices:  $it = n - 1$  or  $n$ . Then  $|\mathbf{U}_n^3| = (n - 3)2^{n-3}$ .

Therefore we have  $|\mathbf{W}_{\text{bf}}^{\geq 6}(n)| = (n - 1)^{n-3} + (n - 2)^{n-3} + (n - 3)2^{n-3}$ .  $\square$

Table 3.3 on p. 33 contains values of  $w_{\text{bf}}^{\leq 5}(n)$  and  $w_{\text{bf}}^{\geq 6}(n)$  for small  $n$ 's. When  $n \in \{3, 4\}$ , we have  $\mathbf{W}_{\text{bf}}^{\geq 6}(n) = \mathbf{W}_{\text{bf}}^{\leq 5}(n)$ , and these cases were already discussed. So we are only interested in larger values of  $n$ . When  $n \geq 6$ , one verifies that  $w_{\text{bf}}^{\geq 6}(n) > w_{\text{bf}}^{\leq 5}(n)$ ; hence  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  is larger than  $\mathbf{W}_{\text{bf}}^{\leq 5}(n)$ .

**Proposition 3.29.** *For  $n \geq 5$ , the semigroup  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$  is generated by*

$$\mathbf{G}_{\text{bf}}^{\geq 6}(n) = \{a_1, a_2, a_3, b_1, \dots, b_{n-3}, c_1, \dots, c_m, d_1, \dots, d_l\},$$

where  $m = (n-2)^{n-3} - 1$  and  $l = (n-3)(2^{n-3} - 1)$ , and

- $a_1 = \binom{1}{n} \binom{n-1}{n} (2, \dots, n-2)$ ,  $a_2 = \binom{1}{n} \binom{n-1}{n} (2, 3)$ ,  $a_3 = \binom{1}{n} \binom{n-1}{n} \binom{n-2}{2}$ ;
- For  $1 \leq i \leq n-3$ ,  $b_i = \binom{1}{n} \binom{n-1}{n} \binom{i+1}{n-1}$ ;
- Each  $c_i$  defines a distinct transformation in  $\mathbf{U}_n^2$  other than  $[n-1, n, \dots, n, n]$ ;
- Each  $d_i$  defines a distinct transformation in  $\mathbf{U}_n^3$  other than  $[j, n, \dots, n, n]$  for all  $j \in \{2, \dots, n-2\}$ .

For  $n = 5$ ,  $a_1$  and  $a_2$  coincide, and 18 transformations suffice.

*Proof.* Since  $\mathbf{G}_{\text{bf}}^{\geq 6}(n) \subseteq \mathbf{W}_{\text{bf}}^{\geq 6}(n)$ , we have  $\langle \mathbf{G}_{\text{bf}}^{\geq 6}(n) \rangle \subseteq \mathbf{W}_{\text{bf}}^{\geq 6}(n)$ . It remains to be shown that  $\mathbf{W}_{\text{bf}}^{\geq 6}(n) \subseteq \langle \mathbf{G}_{\text{bf}}^{\geq 6}(n) \rangle$ . Let  $Q' = Q \setminus \{1, n-1, n\}$ .

1. First consider  $\mathbf{U}_n^1$ . By Theorem 2.1,  $a_1, a_2$  and  $a_3$  together generate the semigroup

$$\mathbf{Y}' = \{t \in \mathbf{U}_n^1 \mid \text{for all } i \in Q', it \in Q'\},$$

which is contained in  $\mathbf{U}_n^1$ . For any  $t \in \mathbf{U}_n^1 \setminus \mathbf{Y}'$ , let  $E_t = \{i \in Q \mid it = n-1\}$ ; then  $E_t \neq \emptyset$ . Suppose  $E_t = \{i_1, \dots, i_k\}$ , where  $1 \leq k \leq n-3$ . Then there exists  $t' \in \mathbf{Y}'$  such that, for all  $i \notin E_t$ ,  $it' = it$ . Let  $s = b_{i_1-1} \cdots b_{i_k-1}$ . Note that  $E_t s = \{n-1\}$ , and, for all  $i \notin E_t$ ,  $i(t's) = (it')s = it$ . So  $t's = t$ , and  $\langle a_1, a_2, a_3, b_1, \dots, b_{n-3} \rangle = \mathbf{U}_n^1$ .

2. Next, the transformations that are in  $\mathbf{U}_n^2 \cup \mathbf{U}_n^3$  but not in  $\mathbf{G}_{\text{bf}}^{\geq 6}(n)$  are  $t_i = [i, n, \dots, n, n]$ , where  $2 \leq i \leq n-1$ . Note that  $d = \binom{1}{2} \binom{n-1}{n} \binom{Q'}{n-1} \in \mathbf{G}_{\text{bf}}^{\geq 6}(n)$ , and, for each  $i \in \{2, \dots, n-1\}$ ,  $s_i = \binom{1}{n} \binom{n-1}{n} \binom{2}{i} \in \mathbf{U}_n^1$ . Then  $t_i = ds_i \in \langle \mathbf{G}_{\text{bf}}^{\geq 6}(n) \rangle$ , and  $\mathbf{U}_n^2 \cup \mathbf{U}_n^3 \subseteq \langle \mathbf{G}_{\text{bf}}^{\geq 6}(n) \rangle$ .

Therefore  $\mathbf{W}_{\text{bf}}^{\geq 6}(n) = \langle \mathbf{G}_{\text{bf}}^{\geq 6}(n) \rangle$ . □

**Theorem 3.30.** For  $n \geq 5$ , let  $\mathcal{A}'_n = (Q, \Sigma, \delta, 1, \{n-1\})$  be a DFA with alphabet  $\Sigma$  of size 18 if  $n = 5$  or  $(n-2)^{n-3} + (n-3)2^{n-3} + 2$  if  $n \geq 6$ , where each letter defines a transformation as in Proposition 3.29. Then  $L' = L(\mathcal{A}'_n)$  has quotient complexity  $\kappa(L') = n$ , and syntactic complexity  $\sigma(L') = w_{\text{bf}}^{\geq 6}(n)$ . Moreover,  $L'$  is bifix-free.

*Proof.* First, for all  $i \in Q \setminus \{1\}$ , there exists  $a \in \Sigma$  such that  $t_a = [i, n, \dots, n, n] \in \mathbf{G}_{\text{bf}}^{\geq 6}(n)$ , and state  $i$  is reachable by  $a$ . So all the states in  $Q$  are reachable. Next, there exist  $b, c \in \Sigma$  such that  $t_b = [n-1, n, \dots, n, n] \in \mathbf{G}_{\text{bf}}^{\geq 6}(n)$  and  $t_c = [n, 3, 4, \dots, n-1, n, n] \in \mathbf{G}_{\text{bf}}^{\geq 6}(n)$ . The initial state accepts  $b$ , while all other states reject it. For  $2 \leq i \leq n-2$ , state  $i$  accepts  $c^{n-i-1}$ , while all other states reject it. Also, state  $n-1$  is the only final state, and state  $n$  is the empty state. Then all the states in  $Q$  are distinct, and  $\kappa(L') = n$ .

By Proposition 3.29, the syntactic semigroup of  $L'$  is  $\mathbf{W}_{\text{bf}}^{\geq 6}(n)$ ; so  $\sigma(L') = w_{\text{bf}}^{\geq 6}(n)$ . By Proposition 3.20,  $L'$  is bifix-free.  $\square$

**Theorem 3.31.** If  $L$  is a bifix-free regular language with  $\kappa(L) = 6$ , then  $\sigma(L) \leq w_{\text{bf}}^{\geq 6}(6) = 213$  and this is a tight bound.

*Proof.* Since  $|\mathbf{B}_{\text{bf}}(6)| = 339$  and  $|\mathbf{W}_{\text{bf}}^{\leq 5}(6)| = 213$ , there are 126 transformations  $\tau_1, \dots, \tau_{126}$  in  $\mathbf{B}_{\text{bf}}(6) \setminus \mathbf{W}_{\text{bf}}^{\leq 5}(6)$ . For each  $\tau_i$ , we enumerated transformations in  $\mathbf{W}_{\text{bf}}^{\geq 6}(6)$  using *GAP* and found a unique  $t_i \in \mathbf{W}_{\text{bf}}^{\leq 5}(6)$  such that  $\langle t_i, \tau_i \rangle \not\subseteq \mathbf{B}_{\text{bf}}(6)$ . Thus, for each  $i$ , at most one of  $t_i$  and  $\tau_i$  can appear in the syntactic semigroup  $T_L$  of  $L$ . So we lower the bound to  $\sigma(L) \leq 213$ . This bound is reached by the DFA  $\mathcal{A}'_6$  in Theorem 3.30; so it is a tight upper bound for  $n = 6$ .  $\square$

**Conjecture 3.32** (Bifix-Free Regular Languages). If  $L$  is a bifix-free regular language with  $\kappa(L) = n \geq 7$ , then  $\sigma(L) \leq w_{\text{bf}}^{\geq 6}(n)$ .

## 3.4 Factor-Free Regular Languages

Let  $L$  be a factor-free regular language with  $\kappa(L) = n$ . Since factor-free regular languages are also bifix-free,  $L$  as a quotient is uniquely reachable,  $\varepsilon$  is the only final quotient of  $L$ , and  $L$  also has the empty quotient. As in Section 3.3, we assume that  $Q$  is the set of states of quotient DFA of  $L$ , in which 1 is the initial state, and states  $n-1$  and  $n$  correspond to the quotients  $\varepsilon$  and  $\emptyset$ , respectively. For  $n \geq 2$ , let

$$\mathbf{B}_{\text{ff}}(n) = \{t \in \mathbf{B}_{\text{bf}}(n) \mid \text{for all } j \geq 1, 1t^j = n-1 \Rightarrow it^j = n \ \forall i, 1 < i < n-1\}.$$

We first have the following observation:

**Proposition 3.33.** *If  $L$  is a regular language with quotient complexity  $n$  and syntactic semigroup  $T_L$ , then the following hold:*

1. *If  $L$  is factor-free, then  $T_L$  is a subset of  $\mathbf{B}_{\text{ff}}(n)$ .*
2. *If  $\varepsilon$  is the only final quotient of  $L$ , and  $T_L \subseteq \mathbf{B}_{\text{ff}}(n)$ , then  $L$  is factor-free.*

*Proof.* 1. Assume  $L$  is factor-free. Then  $L$  is bifix-free, and  $T_L \subseteq \mathbf{B}_{\text{bf}}(n)$  by Proposition 3.20. For any transformation  $t_w \in T_L$  performed by some non-empty word  $w$ , if  $1t_w^j = n - 1$  for some  $j \geq 1$ , then  $w^j \in L$ . If we also have  $it_w^j \neq n$  for some  $i \in Q \setminus \{1\}$ , then  $i \notin \{n - 1, n\}$  as  $(n - 1)t = nt = n$  for all  $t \in \mathbf{B}_{\text{ff}}(n)$ . Thus there exist non-empty words  $u$  and  $v$  such that state  $i$  is reachable by  $u$ , and state  $it_w^j$  accepts  $v$ . So  $uw^jv \in L$ , which is a contradiction. Hence  $T_L \subseteq \mathbf{B}_{\text{ff}}(n)$ .

2. Since  $\varepsilon$  is the only final state and  $\mathbf{B}_{\text{ff}}(n) \subseteq \mathbf{B}_{\text{bf}}(n)$ ,  $L$  is bifix-free by Proposition 3.20. If  $L$  is not factor-free, then there exist non-empty words  $u, v$  and  $w$  such that  $w, u w v \in L$ . Thus  $1t_w = n - 1$ , and  $1t_{u w v} = 1(t_u t_w t_v) = n - 1$ . Since  $L$  is bifix-free,  $1t_u \neq 1$  and  $nt_v = n$ ; thus  $(1t_u)t_w \neq n$ , which contradicts the assumption that  $t_w \in T_L \subseteq \mathbf{B}_{\text{ff}}(n)$ . Therefore  $L$  is factor-free.  $\square$

The properties of suffix- and bifix-free regular languages still apply to factor-free regular languages. Moreover, we have

**Lemma 3.34.** *For all  $t \in \mathbf{B}_{\text{ff}}(n)$  and  $i \notin \mathfrak{s}_t(1)$ , if  $n - 1 \in \mathfrak{s}_t(1)$ , then  $n \in \mathfrak{s}_t(i)$ .*

*Proof.* Suppose  $n - 1 = 1t^k \in \mathfrak{s}_t(1)$  for some  $k \geq 1$ . If  $n \notin \mathfrak{s}_t(i)$ , then for all  $j \geq 1$ ,  $it^j \neq n$ . In particular,  $it^k \neq n$ , which contradicts the definition of  $\mathbf{B}_{\text{ff}}(n)$ . Therefore  $n \in \mathfrak{s}_t(i)$ .  $\square$

**Lemma 3.35.** *We have  $|\mathbf{B}_{\text{ff}}(2)| = 1$ , and for  $n \geq 3$ ,  $|\mathbf{B}_{\text{ff}}(n)| = N_n + O_n$ , where*

$$O_n = 1 + \sum_{k=2}^{n-2} C_{k-1}^{n-3} (k-1)! \sum_{\substack{r_2 + \dots + r_k + r \\ = n - k - 2}} C_{r_2, \dots, r_k, r}^{n-k-2} S'_{r+1}(k) \prod_{j=2}^k S'_{r_j+1}(j-1),$$

and  $N_n$  as given in Equation (3.3).

*Proof.* First we have  $\mathbf{B}_{\text{ff}}(2) = \{[2, 2]\}$  and  $|\mathbf{B}_{\text{ff}}(2)| = 1$ . Assume  $n \geq 3$ . Let  $t \in \mathbf{B}_{\text{ff}}(n)$  be any transformation. Suppose  $\mathfrak{s}_t(1) = 1, 1t, \dots, 1t^k, n$ , where  $0 \leq k \leq n - 2$ . Then there are two cases:

1.  $n - 1 \in \mathfrak{s}_t(1)$ . Since  $(n - 1)t = n$ , we have  $n - 1 = 1t^k$ , and  $k \geq 1$ . If  $k = 1$ , then  $1t = n - 1$ , and  $it = n$  for all  $i \neq 1$ ; such a  $t$  is unique. Consider  $k \geq 2$ . There are  $C_{k-1}^{n-3}(k - 1)!$  different  $\mathfrak{s}_t(1)$ . For  $2 \leq j \leq k$ , suppose there are  $r_j + 1$  nodes in tree  $T_t(j)$ ; then there are  $S'_{r_j+1}(j - 1)$  such trees. Let  $E$  be the set of elements  $x$  that are not in any tree  $T_t(j)$  nor in  $\mathfrak{s}_t(1)$ , and let  $r = |E| = (n - k - 2) - (r_2 + \dots + r_k)$ . By Lemma 3.34,  $n \in \mathfrak{s}_t(x)$  for all  $x \in E$ . Then the union of paths  $P_t(x)$  for all  $x \in E$  form a labeled tree rooted at  $n$  with height at most  $k$ , and there are  $S'_{r+1}(k)$  such trees. Thus the number of transformations in this case is  $O_n$ .
2.  $n - 1 \notin \mathfrak{s}_t(1)$ . Now, for all  $j \geq 1$ ,  $1t^j \neq n - 1$ . Then  $t \in \mathbf{B}_{\text{bf}}(n)$ . As in the proof of Lemma 3.21, the number of transformations in this case is  $N_n$ .

Altogether we have the desired formula. □

Let  $\text{b}_{\text{ff}}(n) = |\mathbf{B}_{\text{ff}}(n)|$ . From Proposition 3.33 and Lemma 3.35 we have

**Proposition 3.36.** *For  $n \geq 2$ , if  $L$  is a factor-free regular language with quotient complexity  $n$ , then its syntactic complexity  $\sigma(L)$  satisfies  $\sigma(L) \leq \text{b}_{\text{ff}}(n)$ , where  $\text{b}_{\text{ff}}(n)$  is the cardinality of  $\mathbf{B}_{\text{ff}}(n)$  as in Lemma 3.35.*

The tight upper bound on the syntactic complexity of factor-free regular languages is reached by the largest semigroup contained in  $\mathbf{B}_{\text{ff}}(n)$ . When  $2 \leq n \leq 4$ ,  $\mathbf{B}_{\text{ff}}(n)$  is a semigroup. The languages  $L_2 = \varepsilon$ ,  $L_3 = a$  over alphabet  $\{a, b\}$ , and  $L_4 = ab^*a$  have syntactic complexities  $1 = \text{b}_{\text{ff}}(2)$ ,  $2 = \text{b}_{\text{ff}}(3)$ , and  $6 = \text{b}_{\text{ff}}(4)$ , respectively. So  $\text{b}_{\text{ff}}(n)$  is a tight upper bound for  $n \in \{2, 3, 4\}$ . However, the set  $\mathbf{B}_{\text{ff}}(n)$  is not a semigroup for  $n \geq 5$ , because  $s_1 = [2, 3, \dots, n - 1, n, n]$ ,  $s_2 = \binom{n-1}{n} \binom{2}{n-1} \binom{1}{n} = [n, n - 1, 3, \dots, n - 2, n, n] \in \mathbf{B}_{\text{ff}}(n)$  but  $s_1 s_2 = [n - 1, 3, \dots, n - 2, n, n, n] \notin \mathbf{B}_{\text{ff}}(n)$ .

Next, we find a large semigroup that can be the syntactic semigroup of a factor-free regular language. For  $n \geq 3$ , let  $t_0 = \binom{Q \setminus \{1\}}{n} \binom{1}{n-1} = [n - 1, n, \dots, n]$ , and let  $\mathbf{W}_{\text{ff}}(n) = \mathbf{U}_n^1 \cup \{t_0\} \cup \mathbf{U}_n^3$ .

**Proposition 3.37.** *For  $n \geq 3$ ,  $\mathbf{W}_{\text{ff}}(n)$  is a semigroup contained in  $\mathbf{B}_{\text{ff}}(n)$  with cardinality*

$$\text{w}_{\text{ff}}(n) = |\mathbf{W}_{\text{ff}}(n)| = (n - 1)^{n-3} + (n - 3)2^{n-3} + 1.$$

*Proof.* As we have shown in the proof of Proposition 3.28,  $\mathbf{U}_n^1$  is a semigroup. For any  $t \in \mathbf{U}_n^1 \cup \{t_0\}$ , since  $t_0 \in \mathbf{U}_n^2$ , we have  $tt_0, t_0t \in \mathbf{U}_n^1$ ; so  $\mathbf{U}_n^1 \cup \{t_0\}$  is also a semigroup. We also know that, for any  $t_3 \in \mathbf{U}_n^3$  and  $t' \in \mathbf{W}_{\text{ff}}(n)$ , since  $\mathbf{W}_{\text{ff}}(n) \subseteq \mathbf{W}_{\text{bf}}^{\geq 6}(n)$ ,  $i(t_3 t') = n$  for

all  $i \neq 1$ ; so  $t_3 t' \in \mathbf{W}_{\text{ff}}(n)$ . If  $t' \in \mathbf{U}_n^1 \cup \{t_0\}$ , then  $1t't_3 = n$  and  $t't_3 \in \mathbf{U}_n^1$ ; otherwise,  $t' \in \mathbf{U}_n^3$ , and  $t't_3 = t_2$  or  $\binom{Q}{n} \in \mathbf{U}_n^1$ . Hence  $\mathbf{W}_{\text{ff}}(n)$  is a semigroup.

For any  $t \in \mathbf{U}_n^1$ , since  $1t = n$ , we have  $t \in \mathbf{B}_{\text{ff}}(n)$ . For any  $t \in \mathbf{U}_n^3$ ,  $1t \neq n - 1$ , and  $it^2 = n$  for all  $i \in \{2, \dots, n\}$ ; then  $t \in \mathbf{B}_{\text{ff}}(n)$  as well. Clearly  $t_0 \in \mathbf{B}_{\text{ff}}(n)$ . Hence  $\mathbf{W}_{\text{ff}}(n)$  is contained in  $\mathbf{B}_{\text{ff}}(n)$ .

We know that  $|\mathbf{U}_n^1| = (n - 1)^{n-3}$  and  $|\mathbf{U}_n^3| = (n - 3)2^{n-3}$ . Therefore  $|\mathbf{W}_{\text{ff}}(n)| = (n - 1)^{n-3} + (n - 3)2^{n-3} + 1$ .  $\square$

For  $n \in \{3, 4\}$ , we have  $\mathbf{W}_{\text{ff}}(n) = \mathbf{B}_{\text{ff}}(n)$ . So we are interested in larger values of  $n$  in the rest of this section.

**Proposition 3.38.** *For  $n \geq 5$ , the semigroup  $\mathbf{W}_{\text{ff}}(n)$  is generated by*

$$\mathbf{G}_{\text{ff}}(n) = \{a_1, a_2, a_3, b_1, \dots, b_{n-3}, c_1, \dots, c_m\},$$

where  $m = (n - 3)(2^{n-3} - 1)$ , and

- $a_1 = \binom{1}{n} \binom{n-1}{n} (2, \dots, n - 2)$ ,  $a_2 = \binom{1}{n} \binom{n-1}{n} (2, 3)$ ,  $a_3 = \binom{1}{n} \binom{n-1}{n} \binom{n-2}{2}$ ;
- For  $1 \leq i \leq n - 3$ ,  $b_i = \binom{1}{n} \binom{n-1}{n} \binom{i+1}{n-1}$ ;
- Each  $c_i$  defines a distinct transformation in  $\mathbf{U}_n^3$  other than  $[j, n, \dots, n, n]$  for all  $j \in \{2, \dots, n - 2\}$ .

For  $n = 5$ ,  $a_1$  and  $a_2$  coincide, and 10 transformations suffice.

*Proof.* We know that  $\mathbf{U}_n^1$  is generated by  $\{a_1, a_2, a_3, b_1, \dots, b_{n-3}\}$ , by the proof of Proposition 3.29. Also, the transformations that are in  $\{t_0\} \cup \mathbf{U}_n^3$  but not in  $\mathbf{G}_{\text{ff}}(n)$  are  $t_j = [j, n, \dots, n, n]$ , where  $j \in \{2, \dots, n - 1\}$ . Let  $Q' = Q \setminus \{1, n - 1, n\}$ . Each  $t_j$  is a composition of  $d = \binom{n-1}{n} \binom{Q'}{n-1} \binom{1}{2} \in \mathbf{G}_{\text{bf}}^{\geq 6}(n)$  and  $s_j = \binom{1}{n} \binom{n-1}{n} \binom{2}{j} \in \mathbf{U}_n^1$ . Therefore  $\langle \mathbf{G}_{\text{ff}}(n) \rangle = \mathbf{W}_{\text{ff}}(n)$ .  $\square$

**Theorem 3.39.** *For  $n \geq 5$ , let  $\mathcal{A}_n = (Q, \Sigma, \delta, 1, \{n - 1\})$  be a DFA with alphabet  $\Sigma$  of size 10 if  $n = 5$  or  $(n - 3)2^{n-3} + 3$  if  $n \geq 6$ , where each letter defines a transformation as in Proposition 3.38. Then  $L = L(\mathcal{A}_n)$  has quotient complexity  $\kappa(L) = n$ , and syntactic complexity  $\sigma(L) = w_{\text{ff}}(n)$ . Moreover,  $L$  is factor-free.*

*Proof.* Since  $\mathbf{G}_{\text{ff}}(n) \subseteq \mathbf{G}_{\text{bf}}^{\geq 6}(n)$ , the DFA  $\mathcal{A}_n$  can be obtained from the DFA  $\mathcal{A}'_n$  of Theorem 3.30 by restricting the alphabet. The words used to show that all the states of  $\mathcal{A}'_n$  are reachable and distinct still exist in  $\mathcal{A}_n$ . Then we have  $\kappa(L) = n$ . By Proposition 3.38, the syntactic semigroup of  $L$  is  $\mathbf{W}_{\text{ff}}(n)$ ; so  $\sigma(L) = \text{w}_{\text{ff}}(n)$ . By Proposition 3.33,  $L$  is factor-free.  $\square$

**Theorem 3.40.** *For  $n \in \{5, 6\}$ , if  $L$  is a factor-free regular language with  $\kappa(L) = n$ , then  $\sigma(L) \leq \text{w}_{\text{ff}}(n)$  and this is a tight upper bound.*

*Proof.* For  $n = 5$ ,  $|\mathbf{B}_{\text{ff}}(5)| = 31$ , and  $|\mathbf{W}_{\text{ff}}(5)| = 25$ . There are 6 transformations  $\tau_1, \dots, \tau_6$  in  $\mathbf{B}_{\text{ff}}(5) \setminus \mathbf{W}_{\text{ff}}(5)$ . For each  $\tau_i$ ,  $1 \leq i \leq 6$ , we found a unique  $t_i \in \mathbf{W}_{\text{ff}}(5)$  such that  $\langle t_i, \tau_i \rangle \not\subseteq \mathbf{B}_{\text{ff}}(5)$ :

$$\begin{array}{ll} \tau_1 = [2, 3, 4, 5, 5], & t_1 = [5, 2, 2, 5, 5], \\ \tau_2 = [2, 3, 5, 5, 5], & t_2 = [5, 4, 2, 5, 5], \\ \tau_3 = [2, 5, 3, 5, 5], & t_3 = [5, 3, 3, 5, 5], \\ \tau_4 = [3, 2, 5, 5, 5], & t_4 = [5, 2, 4, 5, 5], \\ \tau_5 = [3, 4, 2, 5, 5], & t_5 = [5, 3, 2, 5, 5], \\ \tau_6 = [3, 5, 2, 5, 5], & t_6 = [5, 3, 4, 5, 5]. \end{array}$$

For each  $1 \leq i \leq 6$ , at most one of  $t_i$  and  $\tau_i$  can appear in the syntactic semigroup  $T_L$  of a factor-free regular language  $L$ . Then  $\sigma(L) = |T_L| \leq 25$ . By Theorem 3.39, this upper bound is tight for  $n = 5$ .

For  $n = 6$ ,  $|\mathbf{B}_{\text{ff}}(6)| = 246$ , and  $|\mathbf{W}_{\text{ff}}(6)| = 150$ . There are 96 transformations  $\tau_1, \dots, \tau_{96}$  in  $\mathbf{B}_{\text{ff}}(6) \setminus \mathbf{W}_{\text{ff}}(6)$ . For each  $\tau_i$ ,  $1 \leq i \leq 96$ , we enumerated the transformations in  $\mathbf{W}_{\text{ff}}(6)$  using *GAP* and found a unique  $t_i \in \mathbf{W}_{\text{ff}}(6)$  such that  $\langle t_i, \tau_i \rangle \not\subseteq \mathbf{B}_{\text{ff}}(6)$ . Thus 150 is a tight upper bound for  $n = 6$ .  $\square$

**Conjecture 3.41** (Factor-Free Regular Languages). *If  $L$  is a factor-free regular language with  $\kappa(L) = n$ , where  $n \geq 7$ , then  $\sigma(L) \leq \text{w}_{\text{ff}}(n)$ .*

## 3.5 Summary

We summarize our results on suffix-, bifix-, and factor-free regular languages in Tables 3.2 and 3.3. The upper bound of prefix-free regular languages is also included. Each cell of Table 3.2 shows the syntactic complexity bounds of prefix- and suffix-free regular languages, in that order, with a particular alphabet size. Table 3.3 is structured similarly for bifix- and factor-free regular languages. The figures in bold type are tight bounds verified by *GAP*. To



Table 3.2: Syntactic complexities of prefix- and suffix-free regular languages.

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
$ \Sigma  = 1$	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
$ \Sigma  = 2$	*	<b>3/3</b>	<b>11/11</b>	<b>49/49</b>	?
$ \Sigma  = 3$	*	*	<b>14/13</b>	<b>95/61</b>	?
$ \Sigma  = 4$	*	*	<b>16/*</b>	<b>110/67</b>	?
$ \Sigma  = 5$	*	*	*	<b>119/73</b>	?
$ \Sigma  = 6$	*	*	*	<b>125/ *</b>	? /501
$ \Sigma  = 7$	*	*	*	*	<b>1296/ ?</b>
$ \Sigma  = 8$	*	*	*	*	* / <b>629</b>
...					
$n^{n-2}$	<b>1</b>	<b>3</b>	<b>16</b>	<b>125</b>	<b>1296</b>
$w_{\text{sf}}^{\leq 5}(n)$	<b>1</b>	<b>3</b>	<b>13</b>	<b>73</b>	501
$w_{\text{sf}}^{\geq 6}(n)$	<b>1</b>	<b>3</b>	11	67	<b>629</b>
$b_{\text{sf}}(n)$	<b>1</b>	<b>3</b>	15	115	1169

compute the bounds for suffix-, bifix-, and factor-free languages, we enumerated semigroups generated by elements of  $\mathbf{B}_{\text{sf}}(n)$ ,  $\mathbf{B}_{\text{bf}}(n)$ , and  $\mathbf{B}_{\text{ff}}(n)$  that are contained in  $\mathbf{B}_{\text{sf}}(n)$ ,  $\mathbf{B}_{\text{bf}}(n)$ , and  $\mathbf{B}_{\text{ff}}(n)$ , respectively, and recorded the largest ones. By Propositions 3.3, 3.20, 3.33, we obtained the desired bounds from the enumeration. The asterisk  $*$  indicates that the bound is already tight for a smaller alphabet. In Table 3.2, the last four rows include the tight upper bound  $n^{n-2}$  for prefix-free languages,  $w_{\text{sf}}^{\leq 5}(n)$ , which is a tight upper bound for  $2 \leq n \leq 5$  for suffix-free languages, conjectured upper bound  $w_{\text{sf}}^{\geq 6}(n)$  for suffix-free languages, and a weaker upper bound  $b_{\text{sf}}(n)$  for suffix-free languages. In Table 3.3, the last four rows include  $w_{\text{bf}}^{\leq 5}(n)$ , which is a tight upper bound for bifix-free languages for  $2 \leq n \leq 5$ , conjectured upper bounds  $w_{\text{bf}}^{\geq 6}(n)$  for bifix-free languages and  $w_{\text{ff}}(n)$  for factor-free languages, and weaker upper bounds  $b_{\text{bf}}(n)$  for bifix-free languages and  $b_{\text{ff}}(n)$  for factor-free languages.

Table 3.3: Syntactic complexities of bifix- and factor-free regular languages.

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
$ \Sigma  = 1$	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
$ \Sigma  = 2$	*	*	<b>7/6</b>	<b>20/12</b>	?
$ \Sigma  = 3$	*	*	*	<b>31/16</b>	?
$ \Sigma  = 4$	*	*	*	<b>32/19</b>	?
$ \Sigma  = 5$	*	*	*	<b>33/20</b>	?
$ \Sigma  = 6$	*	*	*	<b>34/ ?</b>	?
...					
$w_{\text{bf}}^{\leq 5}(n)$	<b>1</b>	<b>2</b>	<b>7</b>	<b>34</b>	209
$w_{\text{bf}}^{\geq 6}(n)$	<b>1</b>	<b>2</b>	<b>7</b>	33	<b>213</b>
$w_{\text{ff}}(n)$	<b>1</b>	<b>2</b>	<b>6</b>	<b>25</b>	<b>150</b>
$b_{\text{bf}}(n)/b_{\text{ff}}(n)$	<b>1/1</b>	<b>2/2</b>	<b>7/6</b>	41/31	339/246

# Chapter 4

## Syntactic Complexity of Star-Free Languages

Recall that regular languages can be defined as the smallest class containing the finite languages and closed under union, concatenation, and star operations. *Star-free* languages are the smallest class containing the finite languages and closed under boolean operations (union, intersection, and complementation) and concatenation. An equivalent condition is that a language is star-free if and only if its minimal DFA can perform only *aperiodic transformations* [32]. Such DFA's are called *aperiodic*. In Section 4.1, we obtain a weak upper bound on the syntactic complexity of star-free languages by studying aperiodic transformations, and we discuss some small cases. In Section 4.2, we study three subclasses of star-free languages: monotonic, partially monotonic, and nearly monotonic languages, and we show tight upper bounds on their syntactic complexities. We summarize our results about star-free languages in Section 4.3.

### 4.1 Aperiodic Transformations

A transformation is *aperiodic* if it contains no cycles of length greater than 1. A semigroup  $T$  of transformations is *aperiodic* if and only if it contains only aperiodic transformations. Thus a language  $L$  with quotient DFA  $\mathcal{A}$  is star-free if and only if every transformation in  $T_{\mathcal{A}}$  is aperiodic.

Let  $A_n$  be the set of all aperiodic transformations of  $Q$ . Each aperiodic transformation can be characterized by a forest of labeled rooted trees as follows. Consider, for example,

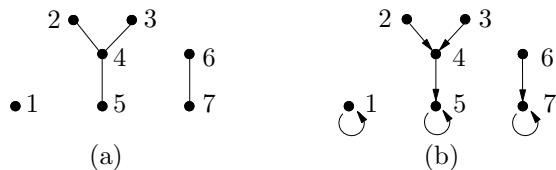


Figure 4.1: Forests and transformations.

the forest of Figure 4.1 (a), where the roots are at the bottom. Convert this forest into a directed graph by adding a direction from each child to its parent and a self-loop to each root, as shown in Figure 4.1 (b). This directed graph defines the transformation  $[1, 4, 4, 5, 5, 7, 7]$  and such a transformation is aperiodic since the directed graph has no cycles of length greater than one. Thus there is a one-to-one correspondence between aperiodic transformations of a set of  $n$  elements and forests with  $n$  nodes.

**Proposition 4.1.** *There are  $(n + 1)^{n-1}$  aperiodic transformations of a set of  $n \geq 1$  elements.*

*Proof.* By Cayley's theorem [14, 45], there are  $(n + 1)^{n-1}$  labeled unrooted trees with  $n + 1$  nodes. If we fix one node, say node  $n + 1$ , in each of these trees to be the root, then we have  $(n + 1)^{n-1}$  labeled trees rooted at  $n + 1$ . Let  $T$  be any one of these trees, and let  $v_1, \dots, v_m$  be the parents of  $n + 1$  in  $T$ . By removing the root  $n + 1$  from each such rooted tree, we get a labeled forest  $F$  with  $n$  nodes formed by  $m$  rooted trees, where  $v_1, \dots, v_m$  are the roots. The forest  $F$  is unique since  $T$  is a unique tree rooted at  $n + 1$ . Then we get a unique aperiodic transformation of  $\{1, \dots, n\}$  by adding self-loops on  $v_1, \dots, v_m$ .

All labeled directed forests with  $n$  nodes can be obtained uniquely from some rooted tree with  $n + 1$  nodes by deleting the root. Hence there are  $(n + 1)^{n-1}$  labeled forests with  $n$  nodes, and that many aperiodic transformations of  $Q$ .  $\square$

Since the quotient DFA of a star-free language can perform only aperiodic transformations, we have

**Corollary 4.2.** *For  $n \geq 1$ , the syntactic complexity  $\sigma(L)$  of a star-free language  $L$  with  $n$  quotients satisfies  $\sigma(L) \leq (n + 1)^{n-1}$ .*

The bound of Corollary 4.2 is our first upper bound on the syntactic complexity of a star-free language with  $n$  quotients, but this bound is not tight in general because the set  $A_n$  is not a semigroup for  $n \geq 3$ . For example, if  $a = [1, 3, 1]$  and  $b = [2, 2, 1]$ , then

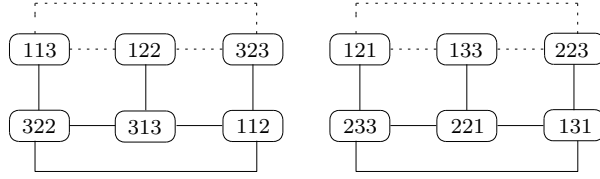


Figure 4.2: Conflict graph for  $n = 3$ .

$ab = [2, 1, 2]$ , which contains the cycle  $(1, 2)$ . Hence our task is to find the size of the largest semigroup contained in  $A_n$ .

First, let us consider small values of  $n$ :

1. If  $n = 1$ , the only two languages,  $\emptyset$  and  $\Sigma^*$ , are both star-free, since  $\Sigma^* = \overline{\emptyset}$ . Here  $\sigma(L) = 1$ , for both languages, the bound  $2^0 = 1$  of Corollary 4.2 holds and it is tight.
2. If  $n = 2$ ,  $|A_2| = 3$ . The only unary languages are  $\varepsilon$  and  $\bar{\varepsilon} = aa^*$ , and  $\sigma(L) = 1$  for both. For  $\Sigma = \{a, b\}$ , one verifies that  $\sigma(L) \leq 2$ , and  $\Sigma^*a\Sigma^*$  meets this bound. If  $\Sigma = \{a, b, c\}$ , then  $L = \Sigma^*a\Sigma^*b\Sigma^*$  has  $\sigma(L) = 3$ .

In summary, for  $n = 1$  and  $2$ , the bound of Corollary 4.2 is tight for  $|\Sigma| = 1$  and  $|\Sigma| = 3$ , respectively.

We say that two aperiodic transformations  $a$  and  $b$  *conflict* if  $ab$  or  $ba$  contains a cycle; then  $(a, b)$  is called a *conflicting pair*. When  $n = 3$ ,  $|A_3| = 4^2 = 16$ . The transformations  $a_0 = [1, 2, 3]$ ,  $a_1 = [1, 1, 1]$ ,  $a_2 = [2, 2, 2]$ ,  $a_3 = [3, 3, 3]$  cannot create any conflict. Hence we consider only the remaining 12 transformations.

Let  $b_1 = [1, 1, 3]$ ,  $b_2 = [1, 2, 1]$ ,  $b_3 = [1, 2, 2]$ ,  $b_4 = [1, 3, 3]$ ,  $b_5 = [2, 2, 3]$ , and  $b_6 = [3, 2, 3]$ . Each of them has only one conflict. There are also two *conflicting triples*  $(b_1, b_3, b_6)$  and  $(b_2, b_4, b_5)$ , since  $b_1b_3b_6$  and  $b_2b_4b_5$  both contains a cycle. Figure 4.2 shows the conflict graph of these 12 transformations, where normal lines indicate conflicting pairs, and dotted lines indicate conflicting triples. To save space we use three digits to represent each transformation, for example, 112 stands for the transformation  $[1, 1, 2]$ , and  $(112)(113) = 111$ . We can choose at most two inputs from each triple and at most one from each conflicting pair. So there are at most 6 conflict-free transformations from the 12, for example,  $b_1, b_3, b_4, b_5, c_1 = [1, 1, 2]$ ,  $c_2 = [2, 3, 3]$ . Adding  $a_0, a_1, a_2$  and  $a_3$ , we get a total of at most 10. The inputs  $a_0, b_4, b_5, c_1$  are conflict-free and generate precisely these 10 transformations. Hence  $\sigma(L) \leq 10$  for any star-free language  $L$  with  $\kappa(L) = n = 3$ , and this bound is tight.

## 4.2 Monotonicity in Transformations, Automata and Languages

We now study syntactic semigroups of languages accepted by monotonic and related automata.

### 4.2.1 Monotonic Transformations, DFA's and Languages

We have shown that the tight upper bound for  $n = 3$  is 10, and it turns out that this bound is met by a monotonic language (defined below). This provides one reason to study monotonic automata and languages. A second reason is the fact that all the tight upper bounds on the quotient complexity of operations on star-free languages are met by monotonic languages [7].

A transformation  $t$  of  $Q$  is *monotonic* if there exists a total order  $\leq$  on  $Q$  such that, for all  $p, q \in Q$ ,  $p \leq q$  implies  $pt \leq qt$ . From now on we assume that  $\leq$  is the usual order on integers, and that  $p < q$  means that  $p \leq q$  and  $p \neq q$ .

Let  $M_Q$  be the set of all monotonic transformations of  $Q$ . In the following, we restate slightly the result of Gomes and Howie [19, 24] for our purposes, since the work in [19] does not consider the identity transformation to be monotonic.

**Theorem 4.3** (Gomes and Howie). *When  $n \geq 1$ , the set  $M_Q$  is an aperiodic semigroup of cardinality*

$$|M_Q| = f(n) = \sum_{k=1}^n C_{k-1}^{n-1} C_k^n = C_n^{2n-1},$$

and it is generated by the set  $H = \{a, b_1, \dots, b_{n-1}, c\}$ , where, for  $1 \leq i \leq n-1$ ,

1.  $1a = 1$ ,  $ja = j - 1$  for  $2 \leq j \leq n$ ;
2.  $ib_i = i + 1$ , and  $jb_i = j$  for all  $j \neq i$ ;
3.  $\exists pt \ c$  is the identity transformation.

Moreover, for  $n = 1$ ,  $a$  and  $c$  coincide and the cardinality of the generating set cannot be reduced for  $n \geq 2$ .

*Remark 4.4.* By Stirling's approximation,  $f(n) = |M_Q|$  grows asymptotically like  $4^n / \sqrt{\pi n}$  as  $n \rightarrow \infty$ .

*Example 4.5.* For  $n = 1$  there is only one transformation  $a = c = [1]$  and it is monotonic. For  $n = 2$ , the three generators are  $a = [1, 1]$ ,  $b_1 = [2, 2]$  and  $c = [1, 2]$ , and  $M_Q$  consists of these three transformations. For  $n = 3$ , the four generators  $a = [1, 1, 2]$ ,  $b_1 = [2, 2, 3]$ ,  $b_2 = [1, 3, 3]$ , and  $c = [1, 2, 3]$  generate all ten monotonic transformations. ■

Now we turn to DFA's whose inputs perform monotonic transformations. A DFA is *monotonic* [1] if all transformations in its transition semigroup are monotonic with respect to some fixed total order. Every monotonic DFA is aperiodic because monotonic transformations are aperiodic. A regular language is *monotonic* if its quotient DFA is monotonic.

Let us now define a DFA having as inputs the generators of  $M_Q$ :

**Definition 4.6.** For  $n \geq 1$ , let  $\mathcal{A}_n = (Q, \Sigma, \delta, 1, \{1\})$  be the DFA in which  $Q = \{1, \dots, n\}$ ,  $\Sigma = \{a, b_1, \dots, b_{n-1}, c\}$ , and each letter in  $\Sigma$  performs the transformation defined in Theorem 4.3.

DFA  $\mathcal{A}_n$  is minimal, since state 1 is the only accepting state, and for  $2 \leq i \leq n$  only state  $i$  accepts  $a^{i-1}$ . From Theorem 4.3 we have

**Corollary 4.7.** For  $n \geq 1$ , the syntactic complexity  $\sigma(L)$  of any monotonic language  $L$  with  $n$  quotients satisfies  $\sigma(L) \leq f(n) = C_n^{2n-1}$ . Moreover, this bound is met by the language  $L(\mathcal{A}_n)$  of Definition 4.6, and, when  $n \geq 2$ , it cannot be met by any monotonic language over an alphabet having fewer than  $n + 1$  letters.

## 4.2.2 Monotonic Partial Transformations and IDFA's

As we shall see, for  $n \geq 4$  the maximal syntactic complexity cannot be reached by monotonic languages; hence we continue our search for larger semigroups of aperiodic transformations. In this subsection, we extend the concept of monotonicity from full transformations to partial transformations, and hence define a new subclass of star-free languages. The upper bound of syntactic complexity of languages in this subclass is above that of monotonic languages for  $n \geq 4$ .

A partial transformation  $t$  of  $Q$  is *monotonic* if there exists a total order  $\leq$  on  $Q$  such that, for all  $p, q \in \text{dom}(t)$ ,  $p \leq q$  implies  $pt \leq qt$ . As before, we assume that the total order on  $Q$  is the usual order on integers. Let  $PM_Q$  be the set of all monotonic partial transformations of  $Q$  with respect to such an order. Gomes and Howie [19] showed the following result, again restated slightly:

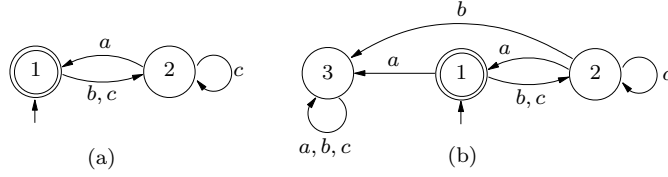


Figure 4.3: Partially monotonic automata: (a) IDFA; (b) DFA.

**Theorem 4.8** (Gomes and Howie). *When  $n \geq 1$ , the set  $PM_Q$  is an aperiodic semigroup of cardinality*

$$|PM_Q| = g(n) = \sum_{k=0}^n C_k^n C_k^{n+k-1},$$

and it is generated by the set  $I = \{a, b_1, \dots, b_{n-1}, c_1, \dots, c_{n-1}, d\}$ , where, for  $1 \leq i \leq n-1$ ,

1.  $1a = \square$ , and  $ja = j-1$  for  $j = 2, \dots, n$ ;
2.  $ib_i = i+1$ ,  $(i+1)b_i = \square$ , and  $jb_i = j$  for  $j = 1, \dots, i-1, i+2, \dots, n$ ;
3.  $ic_i = i+1$ , and  $jc_i = j$  for all  $j \neq i$ ;
4.  $d$  is the identity transformation.

Moreover, the cardinality of the generating set cannot be reduced.

*Example 4.9.* For  $n = 1$ , the two monotonic partial transformations are  $a = [\square]$ , and  $d = [1]$ . For  $n = 2$ , the eight monotonic partial transformations are generated by  $a = [\square, 1]$ ,  $b_1 = [2, \square]$ ,  $c_1 = [2, 2]$ , and  $d = [1, 2]$ . For  $n = 3$ , the 38 monotonic partial transformations are generated by  $a = [\square, 1, 2]$ ,  $b_1 = [2, \square, 3]$ ,  $b_2 = [1, 3, \square]$ ,  $c_1 = [2, 2, 3]$ ,  $c_2 = [1, 3, 3]$  and  $d = [1, 2, 3]$ .

Partial transformations correspond to IDFA's. For example,  $a = [\square, 1]$ ,  $b = [2, \square]$  and  $c = [2, 2]$  correspond to the transitions of the IDFA of Figure 4.3 (a). ■

Laradji and Umar [30] proved the following asymptotic approximation:

*Remark 4.10.* For large  $n$ ,  $g(n) = |PM_Q| \sim 2^{-3/4}(\sqrt{2} + 1)^{2n+1}/\sqrt{\pi n}$ .



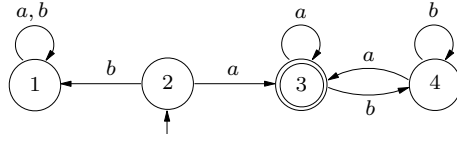


Figure 4.4: Partially monotonic DFA that is monotonic and has an empty state.

An IDFA is *monotonic* if all partial transformations in its transition semigroup are monotonic with respect to some fixed total order. A quotient DFA is *partially monotonic* if its corresponding quotient IDFA is monotonic. A regular language is *partially monotonic* if its quotient DFA is partially monotonic. Note that monotonic languages are also partially monotonic.

*Example 4.11.* If we complete the transformations in Figure 4.3 (a) by replacing the undefined entry  $\square$  by a new empty (or “sink”) state 3, as usual, we obtain the DFA of Figure 4.3 (b). That DFA is not monotonic, because  $1 < 2$  implies  $2 < 3$  under input  $b$  and  $3 < 2$  under  $ab$ . A contradiction is also obtained if we assume that  $2 < 1$ . However, this DFA is partially monotonic, since its corresponding IDFA, shown in Figure 4.3 (a), is monotonic.

The DFA of Figure 4.4 is monotonic for the order shown. It has an empty state, and is also partially monotonic for the same order. ■

Consider any partially monotonic language  $L$  with quotient complexity  $n$ . If its quotient DFA  $\mathcal{A}$  does not have the empty quotient, then  $L$  is monotonic; otherwise, its quotient IDFA  $\mathcal{I}$  has  $n - 1$  states, and the transition semigroup of  $\mathcal{I}$  is a subset of  $PM_{Q'}$ , where  $Q' = \{1, \dots, n - 1\}$ . Hence we consider the following semigroup  $CM_Q$  of *monotonic completed transformations* of  $Q$ . Start with the semigroup  $PM_{Q'}$ . Convert all  $t \in PM_{Q'}$  to full transformations by adding  $n$  to  $\text{dom}(t)$  and letting  $it = n$  for all  $i \in Q \setminus \text{dom}(t)$ . Such a conversion provides a one-to-one correspondence between  $PM_{Q'}$  and  $CM_Q$ . For  $n \geq 2$ , let  $e(n) = g(n - 1)$ . Then semigroups  $CM_Q$  and  $PM_{Q'}$  are isomorphic, and  $e(n) = |CM_Q|$ .

**Definition 4.12.** For  $n \geq 1$ , let  $\mathcal{B}_n = (Q, \Sigma, \delta, 1, \{1\})$  be the DFA in which  $Q = \{1, \dots, n\}$ ,  $\Sigma = \{a, b_1, \dots, b_{n-2}, c_1, \dots, c_{n-2}, d\}$ , and each letter in  $\Sigma$  defines a transformation such that, for  $1 \leq i \leq n - 2$ ,

1.  $1a = na = n$ , and  $ja = j - 1$  for  $j = 2, \dots, n - 1$ ;
2.  $ib_i = i + 1$ ,  $(i + 1)b_i = n$ , and  $jb_i = j$  for  $j = 1, \dots, i - 1, i + 2, \dots, n$ ;

3.  $ic_i = i + 1$ , and  $jc_i = j$  for all  $j \neq i$ ;

4.  $d$  is the identity transformation.

We know that monotonic languages are also partially monotonic. As shown in Table 4.1,  $|M_Q| = f(n) > e(n) = |CM_Q|$  for  $n \leq 3$ . On the other hand, one verifies that  $e(n) > f(n)$  when  $n \geq 4$ . By Corollary 4.7 and Theorem 4.8, we have

**Corollary 4.13.** *The syntactic complexity of a partially monotonic language  $L$  with  $n$  quotients satisfies  $\sigma(L) \leq f(n)$  for  $n \leq 3$ , and  $\sigma(L) \leq e(n)$  for  $n \geq 4$ . Moreover, when  $n \geq 4$ , this bound is met by  $L(\mathcal{B}_n)$  of Definition 4.12, and it cannot be met by any partially monotonic language over an alphabet having fewer than  $2n - 2$  letters.*

Table 4.1 contains these upper bounds for small values of  $n$ . By Remark 4.10, the upper bound  $e(n)$  is asymptotically  $2^{-3/4}(\sqrt{2} + 1)^{2n-1} / \sqrt{\pi(n-1)}$ .

### 4.2.3 Nearly Monotonic Transformations and DFA's

In this section we develop an even larger aperiodic semigroup based on partially monotonic languages.

Let  $K_Q$  be the set of all constant transformations of  $Q$ , and let  $NM_Q = CM_Q \cup K_Q$ . We shall call the transformations in  $NM_Q$  *nearly monotonic* with respect to the usual order on integers.

**Theorem 4.14.** *When  $n \geq 2$ , the set  $NM_Q$  of all nearly monotonic transformations of a set  $Q$  of  $n$  elements is an aperiodic semigroup of cardinality*

$$|NM_Q| = h(n) = e(n) + (n - 1) = \sum_{k=0}^{n-1} C_k^{n-1} C_k^{n+k-2} + (n - 1),$$

and it is generated by the set  $J = \{a, b_1, \dots, b_{n-2}, c_1, \dots, c_{n-2}, d, e\}$  of  $2n - 1$  transformations of  $Q$ , where  $e$  is the constant transformation  $\binom{Q}{1}$ , and all other transformations are as in Definition 4.12. Moreover, the cardinality of the generating set cannot be reduced.

*Proof.* Pick any  $t_1, t_2 \in NM_Q$ . If  $t_1, t_2 \in CM_Q$ , then  $t_1 t_2, t_2 t_1 \in CM_Q$ . Otherwise  $t_1 \in K_Q$  or  $t_2 \in K_Q$ , and both  $t_1 t_2, t_2 t_1$  are constant transformations. Hence  $t_1 t_2, t_2 t_1 \in NM_Q$  and  $NM_Q$  is a semigroup. Since constant transformations are aperiodic and  $CM_Q$  is aperiodic,  $NM_Q$  is also aperiodic.

If  $X$  is a set of transformations, let  $\langle X \rangle$  denote the semigroup generated by  $X$ . Since  $J \subseteq NM_Q$ ,  $\langle J \rangle \subseteq NM_Q$ . Let  $I' = J \setminus \{e\}$ , and  $Q' = Q \setminus \{n\}$ . Then  $PM_{Q'} \simeq CM_Q = \langle I' \rangle$ . For any  $t = \binom{Q}{j} \in K_Q$ , where  $j \in Q$ , since  $s_j = \binom{Q}{j} \binom{n}{n} \in CM_Q \subseteq \langle J \rangle$ , we have that  $t = es_j \in \langle J \rangle$ . So  $NM_Q = \langle J \rangle$ . Note that  $\binom{Q}{i} \in CM_Q$  if and only if  $i = n$ . Thus  $h(n) = |NM_Q| = |PM_{Q'}| + (n - 1) = e(n) + (n - 1)$ .

Since the cardinality of  $I'$  cannot be reduced, and  $e \notin \langle I' \rangle$ , also the cardinality of  $J$  cannot be reduced.  $\square$

*Example 4.15.* For  $n = 2$ , the three nearly monotonic transformations are  $a = [2, 2]$ ,  $d = [1, 2]$  and  $e = [1, 1]$ . For  $n = 3$ , the ten nearly monotonic transformations are generated by  $a = [3, 1, 3]$ ,  $b_1 = [2, 3, 3]$ ,  $c_1 = [2, 2, 3]$ ,  $d = [1, 2, 3]$ , and  $e = [1, 1, 1]$ .  $\blacksquare$

An input  $a \in \Sigma$  is *constant* if it performs a constant transformation of  $Q$ . Let  $\mathcal{A}$  be a DFA with alphabet  $\Sigma$ ; then  $\mathcal{A}$  is *nearly monotonic* if, after removing constant inputs, the resulting DFA  $\mathcal{A}'$  is partially monotonic. A regular language is *nearly monotonic* if its quotient DFA is nearly monotonic.

**Definition 4.16.** For  $n \geq 2$ , let  $\mathcal{C}_n = (Q, \Sigma, \delta, 1, \{1\})$  be a DFA, where  $Q = \{1, \dots, n\}$ ,  $\Sigma = \{a, b_1, \dots, b_{n-2}, c_1, \dots, c_{n-2}, d, e\}$ , and each letter in  $\Sigma$  performs the transformation defined in Theorem 4.14 and Definition 4.12.

Theorem 4.14 now leads us to the following result:

**Theorem 4.17.** For  $n \geq 2$ , if  $L$  is a nearly monotonic language  $L$  with  $n$  quotients, then  $\sigma(L) \leq h(n) = \sum_{k=0}^{n-1} C_k^{n-1} C_k^{n+k-2} + (n - 1)$ . Moreover, this bound is met by the language  $L(\mathcal{C}_n)$  of Definition 4.16, and cannot be met by any nearly monotonic language over an alphabet having fewer than  $2n - 1$  letters.

*Proof.* State 1 is reached by  $\varepsilon$ . For  $2 \leq i \leq n - 1$ , state  $i$  is reached by  $w_i = b_1 \cdots b_{i-1}$ . State  $n$  is reached by  $w_{n-1} b_{n-2}$ . Thus all states are reachable. For  $1 \leq i \leq n - 1$ , the word  $a^{i-1}$  is only accepted by state  $i$ . Also, state  $n$  rejects  $a^i$  for all  $i \geq 0$ . So all  $n$  states are distinguishable, and  $\mathcal{C}_n$  is minimal. Thus  $L$  has  $n$  quotients. The syntactic semigroup of  $L$  is generated by  $J$ ; so  $L$  has syntactic complexity  $\sigma(L) = h(n) = \sum_{k=0}^{n-1} C_k^{n-1} C_k^{n+k-2} + (n - 1)$ , and it is star-free.  $\square$

As shown earlier,  $e(n) > f(n)$  for  $n \geq 4$ . Since  $h(n) = e(n) + (n - 1)$ , and  $h(n) = f(n)$  for  $n \in \{2, 3\}$ , as shown in Table 4.1, we have that  $h(n) \geq f(n)$  for  $n \geq 2$ , and the maximal syntactic complexity of nearly monotonic languages is at least that of both monotonic and partially monotonic languages.

Although we cannot prove that  $NM_Q$  is the largest semigroup of aperiodic transformations, we can show that no transformation can be added to  $NM_Q$ .

A set  $\mathbb{S} = \{T_1, T_2, \dots, T_k\}$  of transformation semigroups is a *chain* if  $T_1 \subset T_2 \subset \dots \subset T_k$ . Semigroup  $T_k$  is the largest in  $\mathbb{S}$ , and we denote it by  $\max(\mathbb{S}) = T_k$ . The following result shows that the syntactic semigroup  $T_{L(\mathcal{C}_n)} = T_{\mathcal{C}_n}$  of  $L(\mathcal{C}_n)$  in Definition 4.16 is a local maximum among aperiodic subsemigroups of  $\mathcal{T}_Q$ .

**Proposition 4.18.** *Let  $\mathbb{S}$  be a chain of aperiodic subsemigroups of  $\mathcal{T}_Q$ . If  $T_{\mathcal{C}_n} \in \mathbb{S}$ , then  $T_{\mathcal{C}_n} = \max(\mathbb{S})$ .*

*Proof.* Suppose  $\max(\mathbb{S}) = T_k$  for some aperiodic subsemigroup  $T_k$  of  $\mathcal{T}_Q$ , and  $T_k \neq T_{\mathcal{C}_n}$ . Then there exist  $t \in T_k$  such that  $t \notin T_{\mathcal{C}_n}$ , and  $i, j \in Q$  such that  $i < j \neq n$  but  $it > jt$ , and  $it, jt \neq n$ . Let  $\tau \in \mathcal{T}_Q$  be such that  $(jt)\tau = i$ ,  $(it)\tau = j$ , and  $h\tau = n$  for all  $h \neq i, j$ ; then  $\tau \in T_{\mathcal{C}_n}$ . Let  $\lambda \in \mathcal{T}_Q$  be such that  $i\lambda = i$ ,  $j\lambda = j$ , and  $h\lambda = n$  for all  $h \neq i, j$ ; then also  $\lambda \in T_{\mathcal{C}_n}$ . Since  $T_k = \max(\mathbb{S})$ ,  $T_{\mathcal{C}_n} \subset T_k$  and  $\tau, \lambda \in T_k$ . Then  $s = \lambda t \tau$  is also in  $T_k$ . However,  $is = i(\lambda t \tau) = j$ ,  $js = j(\lambda t \tau) = i$ , and  $hs = n$  for all  $h \neq i, j$ ; then  $s = (i, j) \binom{P}{n}$ , where  $P = Q \setminus \{i, j\}$ , is not aperiodic, a contradiction. Therefore  $T_{\mathcal{C}_n} = \max(\mathbb{S})$ .  $\square$

**Conjecture 4.19.** *The syntactic complexity of a star-free language  $L$  with  $\kappa(L) = n \geq 4$  satisfies  $\sigma(L) \leq h(n)$ .*

### 4.3 Summary

Our results on star-free languages are summarized in Table 4.1. Let  $Q = \{1, \dots, n\}$ , and  $Q' = Q \setminus \{n\}$ . The figures in bold type are tight bounds verified using *GAP* [18], by enumerating aperiodic subsemigroups of  $\mathcal{T}_Q$ . The asterisk  $*$  indicates that the bound is already tight for a smaller alphabet. The last four rows show the values of  $f(n) = |M_Q|$ ,  $e(n) = |CM_Q| = g(n-1) = |PM_{Q'}|$ ,  $h(n) = |NM_Q|$ , and the weak upper bound  $(n+1)^{n-1}$ .

Table 4.1: Syntactic complexity of star-free languages.

$ \Sigma  / n$	1	2	3	4	5	6
1	<b>1</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>5</b>	<b>6</b>
2	*	<b>2</b>	<b>7</b>	<b>19</b>	<b>62</b>	?
3	*	<b>3</b>	<b>9</b>	<b>31</b>	?	?
4	*	*	<b>10</b>	<b>34</b>	?	?
5	*	*	*	37	125	?
...	...	...	...	...	...	...
$f(n) =  M_Q $	<b>1</b>	<b>3</b>	<b>10</b>	35	126	462
$e(n) =  CM_Q  = g(n-1) =  PM_{Q'} $	—	2	8	38	192	1002
$h(n) =  NM_Q  = e(n) + (n-1)$	—	<b>3</b>	<b>10</b>	41	196	1007
$(n+1)^{n-1}$	1	3	16	125	1296	16807

# Chapter 5

## Syntactic Complexity of $\mathcal{R}$ - and $\mathcal{J}$ -Trivial Languages

Recall that the Green equivalence relations on any monoid  $M$  are defined as follows: For any  $s, t \in M$ ,

$$\begin{aligned} s \mathcal{L} t &\Leftrightarrow Ms = Mt, \\ s \mathcal{R} t &\Leftrightarrow sM = tM, \\ s \mathcal{J} t &\Leftrightarrow MsM = MtM, \\ s \mathcal{H} t &\Leftrightarrow s \mathcal{L} t \text{ and } s \mathcal{R} t. \end{aligned}$$

We say that  $M$  is  $\rho$ -trivial, where  $\rho \in \{\mathcal{R}, \mathcal{J}, \mathcal{L}, \mathcal{H}\}$ , if and only if  $(s, t) \in \rho$  implies  $s = t$  for all  $s, t \in M$ . A language  $L$  is  $\rho$ -trivial if and only if its syntactic monoid is  $\rho$ -trivial. In this chapter we consider only regular  $\rho$ -trivial languages.

### 5.1 $\mathcal{R}$ -Trivial Regular Languages

Given DFA  $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$ , we can define the *reachability relation*  $\rightarrow$  as follows. For all  $p, q \in Q$ ,  $p \rightarrow q$  if and only if  $\delta(p, w) = q$  for some  $w \in \Sigma^*$ . We say that  $\mathcal{A}$  is *partially ordered* [5] if the relation  $\rightarrow$  is a partial order on  $Q$ .

Consider the natural order  $<$  on  $Q$ . A transformation  $t$  of  $Q$  is *nondecreasing* if  $p \leq pt$  for all  $p \in Q$ . The set  $\mathcal{F}_Q$  of all nondecreasing transformations of  $Q$  is a semigroup, since the composition of two nondecreasing transformations is again nondecreasing. It was shown

in [5] that a language  $L$  is  $\mathcal{R}$ -trivial if and only if its quotient DFA is partially ordered. Hence, equivalently,  $L$  is an  $\mathcal{R}$ -trivial language if and only if its syntactic semigroup contains only nondecreasing transformations.

A transformation  $t$  of  $Q$  is an *idempotent* if  $t^2 = t$ . It is known [17] that the semigroup  $\mathcal{F}_Q$  can be generated by the following set

$$\mathcal{GF}_Q = \{\mathbf{1}_Q\} \cup \{t \in \mathcal{F}_Q \mid t^2 = t \text{ and } \text{rank}(t) = n - 1\}.$$

For any transformation  $t$  of  $Q$ , let  $\text{Fix}(t) = \{i \in Q \mid it = i\}$ . Then

**Lemma 5.1.** *For any  $t \in \mathcal{GF}_Q$ ,  $\text{rng}(t) = \text{Fix}(t)$ .*

*Proof.* Pick arbitrary  $t \in \mathcal{GF}_Q$ . The claim holds trivially for  $\mathbf{1}_Q$ . Assume  $t \neq \mathbf{1}_Q$ . Clearly  $\text{Fix}(t) \subseteq \text{rng}(t)$ . Suppose there exists  $i \in \text{rng}(t)$  but  $it \neq i$ . Then  $jt = i$  for some  $j \in Q$ , and  $j \neq i$ . However, since  $jt^2 = it \neq i = jt$ ,  $t$  is not an idempotent, which is a contradiction. Therefore  $\text{rng}(t) = \text{Fix}(t)$ .  $\square$

If  $n = 1$ , then  $\mathcal{F}_Q$  contains only the identity transformation  $\mathbf{1}_Q$ , and  $\mathcal{GF}_Q = \mathcal{F}_Q = \{\mathbf{1}_Q\}$ . So  $|\mathcal{GF}_Q| = |\mathcal{F}_Q| = 1$ . If  $n \geq 2$ , then we have

**Lemma 5.2.** *For  $n \geq 2$ ,  $|\mathcal{GF}_Q| = 1 + C_2^n$ .*

*Proof.* Pick  $t \in \mathcal{GF}_Q$  such that  $t \neq \mathbf{1}_Q$ . Then  $\text{rank}(t) = n - 1$ , and, by Lemma 5.1,  $|\text{Fix}(t)| = n - 1$ . There is only one element  $i \in Q \setminus \text{Fix}(t)$ , and  $i < it$ . Note that  $t$  is fully determined by the pair  $(i, it)$ . Hence there are  $C_2^n$  different  $t$ . Together with the identity  $\mathbf{1}_Q$ , the cardinality of  $\mathcal{GF}_Q$  is  $1 + C_2^n$ .  $\square$

**Lemma 5.3.** *If  $G \subseteq \mathcal{F}_Q$  and  $G$  generates  $\mathcal{F}_Q$ , then  $\mathcal{GF}_Q \subseteq G$ .*

*Proof.* Suppose there exists  $t \in \mathcal{GF}_Q$  such that  $t \notin G$ . Since  $G$  generates  $\mathcal{F}_Q$ ,  $t$  can be written as  $t = g_1 \cdots g_k$  for some  $g_1, \dots, g_k \in G$ , where  $k \geq 2$ . Then  $\text{rng}(g_1) \supseteq \cdots \supseteq \text{rng}(g_k) \supseteq \text{rng}(t)$ . Note that  $\mathbf{1}_Q$  is the only element in  $\mathcal{F}_Q$  with range  $Q$ ; so if  $t = \mathbf{1}_Q$ , then  $g_1 = \cdots = g_k = \mathbf{1}_Q$ , a contradiction.

Assume  $t \neq \mathbf{1}_Q$ . Then  $\text{rank}(t) = n - 1$ , and  $\text{rng}(g_1) = \cdots = \text{rng}(g_k) = \text{rng}(t)$ . Since each  $g_i$  is nondecreasing, for all  $p \in \text{Fix}(t)$ , we must have  $p \in \text{Fix}(g_i)$  as well; so  $\text{Fix}(t) \subseteq \text{Fix}(g_i)$ . Moreover, since  $\text{Fix}(g_i) \subseteq \text{rng}(g_i) = \text{rng}(t)$  and  $\text{rng}(t) = \text{Fix}(t)$  by Lemma 5.1,  $\text{Fix}(g_i) = \text{Fix}(t) = \text{rng}(t)$ . Now, let  $q$  be the unique element in  $Q \setminus \text{Fix}(t)$ . Then  $q \notin \text{Fix}(g_1)$ , and  $qg_1 \in \text{Fix}(g_2) = \cdots = \text{Fix}(g_k)$ . So  $q(g_1 \cdots g_k) = qg_1$ . However, since  $t = g_1 \cdots g_k$ ,  $q(g_1 \cdots g_k) = qt$  and  $qg_1 = qt$ . Hence  $g_1 = t$ , and we get a contradiction again. Therefore  $\mathcal{GF}_Q \subseteq G$ .  $\square$

Consequently,  $\mathcal{GF}_Q$  is the unique minimal generator of  $\mathcal{F}_Q$ . So we obtain

**Theorem 5.4.** *If  $L \subseteq \Sigma^*$  is a regular  $\mathcal{R}$ -trivial language of quotient complexity  $\kappa(L) = n \geq 1$ , then its syntactic complexity  $\sigma(L)$  satisfies  $\sigma(L) \leq n!$ , and this bound is tight if  $|\Sigma| = 1$  for  $n = 1$  and  $|\Sigma| \geq 1 + C_2^n$  for  $n \geq 2$ .*

*Proof.* Let  $\mathcal{A}$  be the quotient DFA of  $L$ , and let  $T_L$  be its syntactic semigroup. Then  $T_L$  is a subset of  $\mathcal{F}_Q$ . Pick an arbitrary  $t \in \mathcal{F}_Q$ . For each  $p \in Q$ , since  $p \leq pt$ ,  $pt$  can be chosen from  $\{p, p+1, \dots, n\}$ . Hence there are exactly  $n!$  transformations in  $\mathcal{F}_Q$ , and  $\sigma(L) \leq n!$ .

When  $n = 1$ , the only regular languages are  $\varepsilon$  or  $\emptyset$ , and they both are  $\mathcal{R}$ -trivial. To see the bound is tight for  $n \geq 2$ , let  $\mathcal{A}_n = (Q, \Sigma, \delta, 1, \{n\})$  be the DFA with alphabet  $\Sigma$  of size  $1 + C_2^n$  and set of states  $Q = \{1, \dots, n\}$ , where each  $a \in \Sigma$  defines a distinct transformation in  $\mathcal{GF}_Q$ . For each  $p \in Q$ , since  $\mathcal{GF}_Q$  generates  $\mathcal{F}_Q$  and  $t_p = [p, n, \dots, n] \in \mathcal{F}_Q$ ,  $t_p = e_1 \cdots e_k$  for some  $e_1, \dots, e_k \in \mathcal{GF}_Q$ , where  $k$  depends on  $p$ . Then there exist  $a_1, \dots, a_k \in \Sigma$  such that each  $a_i$  performs  $e_i$  and state  $p$  is reached by  $w = a_1 \dots a_k$ . Moreover, since  $t = [2, 3, \dots, n, n] \in \mathcal{F}_Q$ , there exist  $b_1, \dots, b_l \in \Sigma$  such that the word  $u = b_1 \dots b_l$  performs  $t$ . So state  $p \in Q$  can be distinguished from other states by the word  $u^{n-p}$ . Hence  $L = L(\mathcal{A}_n)$  has quotient complexity  $\kappa(L) = n$ . The syntactic monoid of  $L$  is  $\mathcal{F}_Q$ , and so  $\sigma(L) = n!$ . By Lemma 5.3, the alphabet of  $\mathcal{A}_n$  is minimal.  $\square$

*Example 5.5.* When  $n = 4$ , there are  $4! = 24$  nondecreasing transformations of  $Q = \{1, 2, 3, 4\}$ . Among them, there are 11 transformations with rank  $n - 1 = 3$ . The following 6 transformations from the 11 are idempotents:

$$\begin{aligned} e_1 &= [1, 2, 4, 4], & e_2 &= [1, 3, 3, 4] \\ e_3 &= [1, 4, 3, 4], & e_4 &= [2, 2, 3, 4] \\ e_5 &= [3, 2, 3, 4], & e_6 &= [4, 2, 3, 4] \end{aligned}$$

Together with the identity transformation  $\mathbf{1}_Q$ , we have the generating set  $\mathcal{GF}_Q$  for  $\mathcal{F}_Q$  with 7 transformations. We can then define the DFA  $\mathcal{A}_4$  with 7 inputs as in the proof of Theorem 5.4;  $\mathcal{A}_4$  is shown in Figure 5.1. The quotient complexity of  $L = L(\mathcal{A}_4)$  is 4, and the syntactic complexity of  $L$  is 24.  $\blacksquare$

## 5.2 $\mathcal{J}$ -Trivial Regular Languages

We first recall some facts from universal algebra. Let  $Q$  be a non-empty finite set with  $n$  elements, and assume without loss of generality that  $Q = \{1, 2, \dots, n\}$ . There is a linear



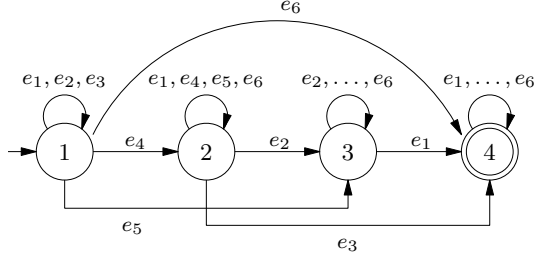


Figure 5.1: DFA  $\mathcal{A}_4$  with  $\kappa(L(\mathcal{A}_4)) = 4$  and  $\sigma(L(\mathcal{A}_4)) = 24$ ; the input performing the identity transformation is not shown.

order on  $Q$ , namely the natural order  $<$  on integers. If  $X$  is a non-empty subset of  $Q$ , then the maximal element in  $X$  is denoted by  $\max(X)$ . A *partition*  $\pi$  of  $Q$  is a collection  $\pi = \{X_1, X_2, \dots, X_m\}$  of non-empty subsets of  $Q$  such that

1.  $Q = X_1 \cup X_2 \cup \dots \cup X_m$ , and
2.  $X_i \cap X_j = \emptyset$  for all  $1 \leq i < j \leq m$ .

We call each subset  $X_i$  a *block* in  $\pi$ . For any partition  $\pi$  of  $Q$ , let  $\text{Max}(\pi) = \{\max(X) \mid X \in \pi\}$ . The set of all partitions of  $Q$  is denoted by  $\Pi_Q$ . We can define a partial order  $\preceq$  on  $\Pi_Q$  such that, for any  $\pi_1, \pi_2 \in \Pi_Q$ ,  $\pi_1 \preceq \pi_2$  if and only if each block of  $\pi_1$  is contained in some block of  $\pi_2$ . We say  $\pi_1$  *refines*  $\pi_2$  if  $\pi_1 \preceq \pi_2$ . Then  $(\Pi_Q, \preceq)$  forms a poset. Furthermore,  $(\Pi_Q, \preceq)$  is a finite lattice; for any  $\pi_1, \pi_2 \in \Pi_Q$ , their *meet*  $\pi_1 \wedge \pi_2$  is the  $\preceq$ -largest partition that refines both  $\pi_1$  and  $\pi_2$ , and their *join*  $\pi_1 \vee \pi_2$  is the  $\preceq$ -smallest partition that is refined by both  $\pi_1$  and  $\pi_2$ . From now on, we simply refer to the lattice  $(\Pi_Q, \preceq)$  as  $\Pi_Q$ .

For any  $m \geq 1$ , we can define an equivalence relation  $\leftrightarrow_m$  on  $\Sigma^*$  as follows. For any  $u, v \in \Sigma^*$ ,  $u \leftrightarrow_m v$  if and only if for every  $x \in \Sigma^*$  with  $|x| \leq m$ ,

$$x \text{ is a subword of } u \Leftrightarrow x \text{ is a subword of } v.$$

Let  $L$  be any language over  $\Sigma$ . Then  $L$  is *piecewise-testable* if there exists  $m \geq 1$  such that, for every  $u, v \in \Sigma^*$ ,  $u \leftrightarrow_m v$  implies that  $u \in L \Leftrightarrow v \in L$ . Let  $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$  be a DFA. If  $\Gamma$  is a subset of  $\Sigma$ , a *component* of  $\mathcal{A}$  restricted to  $\Gamma$  is a minimal subset  $P$  of  $Q$  such that, for all  $p \in Q$  and  $w \in \Gamma^*$ ,  $\delta(p, w) \in P$  if and only if  $p \in P$ . A state  $q$  of  $\mathcal{A}$  is *maximal* if  $\delta(q, a) = q$  for all  $a \in \Sigma$ . Simon [49] proved the following characterization of piecewise-testable languages.

**Theorem 5.6** (Simon). *Let  $L$  be a regular language over  $\Sigma$ , let  $\mathcal{A}$  be its quotient DFA, and let  $T_L$  be its syntactic monoid. Then the following are equivalent:*

1.  $L$  is piecewise-testable;
2.  $\mathcal{A}$  is partially ordered, and for every non-empty subset  $\Gamma$  of  $\Sigma$ , each component of  $\mathcal{A}$  restricted to  $\Gamma$  has exactly one maximal state;
3.  $T_L$  is  $\mathcal{J}$ -trivial.

Consequently, a regular language is piecewise-testable if and only if it is  $\mathcal{J}$ -trivial. The following theorem is due to Saito [42]. It is another characterization of  $\mathcal{J}$ -trivial monoids.

**Theorem 5.7** (Saito). *Let  $S$  be a monoid of transformations of  $Q$ . Then the following are equivalent:*

1.  $S$  is  $\mathcal{J}$ -trivial;
2.  $S$  is a subset of  $\mathcal{F}_Q$  and  $\Omega(ts) = \Omega(t) \vee \Omega(s)$  for all  $t, s \in S$ .

Let  $L$  be a regular  $\mathcal{J}$ -trivial language with quotient DFA  $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$  and syntactic monoid  $T_L$ . Since  $T_L$  is a subset of  $\mathcal{F}_Q$ , to get an upper bound on the syntactic complexity of  $L$ , we find an upper bound on the cardinality of  $\mathcal{J}$ -trivial submonoids of  $\mathcal{F}_Q$ .

**Lemma 5.8.** *If  $t, s \in \mathcal{F}_Q$ , then*

1.  $\text{Fix}(t) = \text{Max}(\Omega(t))$ ;
2.  $\Omega(t) \preceq \Omega(s)$  implies that  $\text{Fix}(t) \supseteq \text{Fix}(s)$ , where the equality holds if and only if  $\Omega(t) = \Omega(s)$ ;

*Proof.* 1. First, for each  $j \in \text{Max}(\Omega(t))$ , since  $t \in \mathcal{F}_Q$ , we have  $jt = j$ , and  $j \in \text{Fix}(t)$ . So  $\text{Max}(\Omega(t)) \subseteq \text{Fix}(t)$ . On the other hand, if there exists  $j \in \text{Fix}(t) \setminus \text{Max}(\Omega(t))$ , then  $jt = j$ , and  $j < \max(\omega_t(j))$ . Let  $i = \max(\omega_t(j))$ ; then for any  $k, l \geq 0$ ,  $jt^k = j < i = it^l$ . So  $i \notin \omega_t(j)$ , which is a contradiction. Hence  $\text{Fix}(t) = \text{Max}(\Omega(t))$ .

2. Assume  $\Omega(t) \preceq \Omega(s)$ . By definition, we have  $\text{Max}(\Omega(t)) \supseteq \text{Max}(\Omega(s))$ . Then, by 1,  $\text{Fix}(t) \supseteq \text{Fix}(s)$ . Furthermore,  $\Omega(t) = \Omega(s)$  if and only if  $\text{Max}(\Omega(t)) = \text{Max}(\Omega(s))$ , and if and only if  $\text{Fix}(t) = \text{Fix}(s)$ .  $\square$

*Example 5.9.* Consider nondecreasing  $t = [1, 3, 3, 5, 6, 6]$ , as shown in Figure 5.2 (a). The orbit set  $\Omega(t)$  has three blocks:  $\{1\}$ ,  $\{2, 3\}$ , and  $\{4, 5, 6\}$ . Note that  $\text{Fix}(t) = \{1, 3, 6\} = \text{Max}(\Omega(t))$ , as expected.

In addition, let  $s = [4, 3, 3, 6, 6, 6]$  be another nondecreasing transformation, as shown in Figure 5.2 (b). The orbit set  $\Omega(s)$  has two blocks:  $\{1, 4, 5, 6\}$  and  $\{2, 3\}$ . Note that  $\Omega(t) \prec \Omega(s)$  and  $\text{Fix}(t) \supset \text{Fix}(s)$ . ■

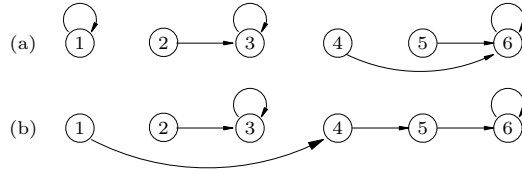


Figure 5.2: Nondecreasing transformations  $t = [1, 3, 3, 5, 6, 6]$  and  $s = [4, 3, 3, 6, 6, 6]$ .

Define the transformation  $t_{\max} = [2, 3, \dots, n, n]$ . The subscript “max” is chosen because  $\Omega(t_{\max}) = \{Q\}$  is the maximum element in the lattice  $\Pi_Q$ . Clearly  $t_{\max} \in \mathcal{F}_Q$  and  $\text{Fix}(t_{\max}) = \{n\}$ . For any submonoid  $S$  of  $\mathcal{F}_Q$ , let  $S[t_{\max}]$  be the smallest monoid containing  $t_{\max}$  and all elements of  $S$ .

**Lemma 5.10.** *Let  $S$  be a  $\mathcal{J}$ -trivial submonoid of  $\mathcal{F}_Q$ . Then*

1.  $S[t_{\max}]$  is  $\mathcal{J}$ -trivial.
2. Let  $\mathcal{A} = (Q, \Sigma, \delta, 1, \{n\})$  be the DFA in which each  $a \in \Sigma$  defines a distinct transformation in  $S[t_{\max}]$ . Then  $\mathcal{A}$  is minimal.

*Proof.* 1. By Theorem 5.7, it is sufficient to prove that for any  $t \in S$ ,  $\Omega(t) \vee \Omega(t_{\max}) = \Omega(tt_{\max})$  and  $\Omega(t_{\max}) \vee \Omega(t) = \Omega(t_{\max}t)$ . Since  $\Omega(t_{\max}) = \{Q\}$ , we have  $\Omega(t) \vee \Omega(t_{\max}) = \Omega(tt_{\max}) \vee \Omega(t) = \{Q\}$ . On the other hand, since  $S \subseteq \mathcal{F}_Q$  and  $t_{\max} \in \mathcal{F}_Q$ , both  $tt_{\max}$  and  $t_{\max}t$  are nondecreasing as well. Suppose  $i \in \text{Fix}(tt_{\max})$ ; then  $i(tt_{\max}) = (it)_{\max} = i$ . Since  $t_{\max}$  is nondecreasing,  $it \leq i$ ; and since  $t$  is also nondecreasing,  $i \leq it$ . Hence  $it = i$ , and  $it_{\max} = i$ , which implies that  $i \in \text{Fix}(t_{\max})$  and  $i = n$ . Then  $\text{Fix}(tt_{\max}) = \{n\}$  and  $\Omega(tt_{\max}) = \{Q\}$ . Similarly,  $\text{Fix}(t_{\max}t) = \{n\}$  and  $\Omega(t_{\max}t) = \{Q\}$ . Therefore  $S[t_{\max}]$  is also  $\mathcal{J}$ -trivial.

2. Suppose  $a_0 \in \Sigma$  performs the transformation  $t_{\max}$ . Each state  $p \in Q$  can be reached from the initial state 1 by the word  $u = a_0^{p-1}$ , and  $p$  accepts the word  $v = a_0^{n-p}$ , while all other states reject  $v$ . So  $\mathcal{A}$  is minimal. □

For any  $\mathcal{J}$ -trivial submonoid  $S$  of  $\mathcal{F}_Q$ , we denote by  $\mathcal{A}(S, t_{\max})$  the DFA in Lemma 5.10. Then  $\mathcal{A}(S, t_{\max})$  is the quotient DFA of some regular  $\mathcal{J}$ -trivial language  $L$ . Next, we have

**Lemma 5.11.** *Let  $S$  be a  $\mathcal{J}$ -trivial submonoid of  $\mathcal{F}_Q$ . For any  $t, s \in S$ , if  $\text{Fix}(t) = \text{Fix}(s)$ , then  $\Omega(t) = \Omega(s)$ .*

*Proof.* Pick any  $t, s \in S$  such that  $\text{Fix}(t) = \text{Fix}(s)$ . If  $t = s$ , then it is trivial that  $\Omega(t) = \Omega(s)$ . Assume  $t \neq s$ , and  $\Omega(t) \neq \Omega(s)$ . By Part 2 of Lemma 5.8, we have  $\Omega(t) \not\leq \Omega(s)$  and  $\Omega(s) \not\leq \Omega(t)$ . Then there exists  $i \in Q$  such that  $\omega_t(i) \neq \omega_s(i)$ . Suppose  $p = \max(\omega_t(i))$  and  $q = \max(\omega_s(i))$ ; then  $p, q \in \text{Fix}(t) = \text{Fix}(s)$ , and  $p \neq q$ . Consider the DFA  $\mathcal{A}(S, t_{\max})$  with alphabet  $\Sigma$ , and suppose  $a \in \Sigma$  performs  $t$  and  $b \in \Sigma$  performs  $s$ . Let  $\mathcal{B}$  be the DFA  $\mathcal{A}(S, t_{\max})$  restricted to  $\{a, b\}$ . Since  $p \in \omega_t(i)$  and  $q \in \omega_s(i)$ , then  $p, q$  are in the same component  $P$  of  $\mathcal{B}$ . However,  $p$  and  $q$  are two distinct maximal states in  $P$ , which contradicts Theorem 5.6. Therefore  $\Omega(t) = \Omega(s)$ .  $\square$

*Example 5.12.* To illustrate one usage of Lemma 5.11, we consider two nondecreasing transformations  $t = [2, 2, 4, 4]$  and  $s = [3, 2, 4, 4]$ . They have the same set of fixed points  $\text{Fix}(t) = \text{Fix}(s) = \{2, 4\}$ . However,  $\Omega(t) = \{\{1, 2\}, \{3, 4\}\}$  and  $\Omega(s) = \{\{2\}, \{1, 3, 4\}\}$ . By Lemma 5.11,  $t$  and  $s$  cannot appear together in a  $\mathcal{J}$ -trivial monoid. Indeed, consider any minimal DFA  $\mathcal{A}$  having at least two inputs  $a, b$  such that  $a$  performs  $t$  and  $b$  performs  $s$ . The DFA  $\mathcal{B}$  of  $\mathcal{A}$  restricted to the alphabet  $\{a, b\}$  is shown in Figure 5.3. There is only one component in  $\mathcal{B}$ , but there are two maximal states 2 and 4. By Theorem 5.6, the syntactic monoid of  $\mathcal{A}$  is not  $\mathcal{J}$ -trivial.  $\blacksquare$

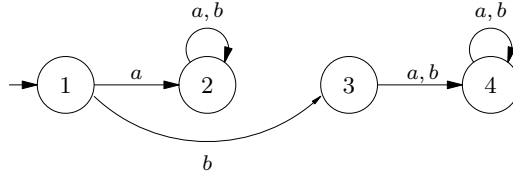


Figure 5.3: DFA  $\mathcal{B}$  with two inputs  $a$  and  $b$ , where  $t_a = [2, 2, 4, 4]$  and  $t_b = [3, 2, 4, 4]$ .

For any partition  $\pi$  of  $Q$ , define  $\mathcal{E}(\pi) = \{t \in \mathcal{F}_Q \mid \Omega(t) = \pi\}$ . Then

**Lemma 5.13.** *If  $\pi$  is a partition of  $Q$  with  $r$  blocks, where  $1 \leq r \leq n$ , then  $|\mathcal{E}(\pi)| \leq (n-r)!$ .*

*Proof.* Suppose  $\pi = \{X_1, \dots, X_r\}$ , and  $|X_i| = k_i$  for each  $i$ ,  $1 \leq i \leq r$ . Without loss of generality, we can rearrange subsets  $X_i$ 's such that  $k_1 \leq \dots \leq k_r$ . Let  $t \in \mathcal{E}(\pi)$  be any

transformation. Then  $t \in \mathcal{F}_Q$ , and hence  $\text{Fix}(t) = \text{Max}(\Omega(t)) = \text{Max}(\pi)$ . Consider each block  $X_i$ , and suppose  $X_i = \{j_1, \dots, j_{k_i}\}$  such that  $j_1 < \dots < j_{k_i}$ . Since  $j_{k_i} = \max(X_i)$ , then  $j_{k_i} \in \text{Fix}(t)$  and  $j_{k_i}t = j_{k_i}$ . On the other hand, if  $1 \leq l < k_i$ , then  $j_l \notin \text{Max}(\pi)$ , and since  $t \in \mathcal{F}_Q$ , we have  $j_l t > j_l$ ; since  $j_l t \in \omega_t(j_l) = X_i$ ,  $j_l t \in \{j_{l+1}, \dots, j_{k_i}\}$ . So there are  $(k_i - 1)!$  different  $t|_{X_i}$ , and there are  $\prod_{i=1}^r (k_i - 1)!$  different transformations  $t$  in  $\mathcal{E}(\pi)$ .

Clearly, if  $r = 1$ , then  $k_r = n$  and  $|\mathcal{E}(\pi)| = (n - 1)!$ . Assume  $r \geq 2$ . Note that  $k_i \geq 1$  for all  $i$ ,  $1 \leq i \leq r$ , and  $\sum_{i=1}^r k_i = n$ . If  $k_1 = \dots = k_{r-1} = 1$ , then  $k_r = n - r + 1$ , and  $|\mathcal{E}(\pi)| = (k_r - 1)! \prod_{i=1}^{r-1} 0! = (n - r)!$ . Otherwise, let  $h$  be the smallest index such that  $k_h > 1$ . Then

$$\begin{aligned} \prod_{i=1}^r (k_i - 1)! &= \prod_{i=1}^{h-1} 0! \prod_{i=h}^r (k_i - 1)! \\ &= (k_r - 1)! \prod_{i=h}^{r-1} (k_i - 1)! \end{aligned}$$

Since  $(k_h - 1)! < (k_h - 1)^{k_h - 1} \leq (k_r - 1)^{k_h - 1} < (k_r + k_h - 2) \cdots k_r$ :

$$< (k_r + k_h - 2)! \prod_{i=h+1}^{r-1} (k_i - 1)!$$

Similarly, we have that

$$\begin{aligned} &< (k_r + k_h + \dots + k_{r-1} - (r - h + 1))! \\ &= (n - r)! \end{aligned}$$

Therefore  $|\mathcal{E}(\pi)| \leq (n - r)!$ . □

*Example 5.14.* Suppose  $n = 10$ ,  $r = 3$ , and consider the partition  $\pi = \{X_1, X_2, X_3\}$ , where  $X_1 = \{1, 2, 5\}$ ,  $X_2 = \{3, 7\}$ , and  $X_3 = \{4, 6, 8, 9, 10\}$ . Then  $k_1 = |X_1| = 3$ ,  $k_2 = |X_2| = 2$ , and  $k_3 = |X_3| = 5$ . Let  $t \in \mathcal{E}(\pi)$  be an arbitrary transformation; then  $\text{Fix}(t) = \{5, 7, 10\}$ . For any  $i \in X_1$ , if  $i = 1$ , then  $it$  could be 2 or 5; otherwise  $i = 2$  or 5, and  $it$  must be 5. So there are  $(k_1 - 1)! = 2!$  different  $t|_{X_1}$ . Similarly, there are  $(k_2 - 1)! = 1!$  different  $t|_{X_2}$  and  $(k_3 - 1)! = 4!$  different  $t|_{X_3}$ . Hence we have  $|\mathcal{E}(\pi)| = 2!1!4! = 48$ .

Consider another partition  $\pi' = \{X'_1, X'_2, X'_3\}$  with three blocks, where  $X'_1 = \{5\}$ ,  $X'_2 = \{7\}$ , and  $X'_3 = \{1, 2, 3, 4, 6, 8, 9, 10\}$ . Then  $k_1 = |X'_1| = 1$ ,  $k_2 = |X'_2| = 1$ , and  $k_3 = |X'_3| = 8$ . We have that  $\text{Max}(\pi') = \text{Max}(\pi) = \{5, 7, 10\}$ . Then, for any  $t \in \mathcal{E}(\pi')$ ,

$\text{Fix}(t) = \{5, 7, 10\}$  as well. Since  $k_1 = k_2 = 1$ , both  $t|_{X_1}$  and  $t|_{X_2}$  are unique. There are  $(k_3 - 1)! = 7!$  different  $t|_{X_3}$ . Together we have  $|\mathcal{E}(\pi')| = 1!1!7! = (10 - 3)! = 5040$ , which is the upper bound in Lemma 5.13 for  $n = 10$  and  $r = 3$ .  $\blacksquare$

Note that, for any  $t \in \mathcal{F}_Q$ , we have  $n \in \text{Fix}(t)$ . Let  $\mathcal{P}_n(Q)$  be the set of all subsets  $Z$  of  $Q$  such that  $n \in Z$ . Then we obtain the following upper bound.

**Proposition 5.15.** *If  $S$  is a  $\mathcal{J}$ -trivial submonoid of  $\mathcal{F}_Q$ , then*

$$|S| \leq \sum_{r=1}^n C_{r-1}^{n-1} (n-r)! = \lfloor e(n-1)! \rfloor.$$

*Proof.* Assume  $S$  is a  $\mathcal{J}$ -trivial submonoid of  $\mathcal{F}_Q$ . For any  $Z \in \mathcal{P}_n(Q)$ , let  $S_Z = \{t \in S \mid \text{Fix}(t) = Z\}$ . Then  $S = \bigcup_{Z \in \mathcal{P}_n(Q)} S_Z$ , and for any  $Z_1, Z_2 \in \mathcal{P}_n(Q)$  with  $Z_1 \neq Z_2$ ,  $S_{Z_1} \cap S_{Z_2} = \emptyset$ .

Pick any  $Z \in \mathcal{P}_n(Q)$ . By Lemma 5.11, for any  $t, s \in S_Z$ , since  $\text{Fix}(t) = \text{Fix}(s) = Z$ , we have  $\Omega(t) = \Omega(s)$ . Let  $\pi \in \Pi_Q$  denote such a partition  $\Omega(t)$  of  $Q$ . Suppose  $r = |Z|$ . Since  $n \in Z$ , we have  $r \geq 1$ ; and clearly  $r \leq n$ . Note that  $S_Z \subseteq \mathcal{E}(\pi)$ . By Lemma 5.13,  $|S_Z| \leq |\mathcal{E}(\pi)| = (n-r)!$ . Since there are  $C_{r-1}^{n-1}$  different  $Z$ , we have that

$$\begin{aligned} |S| &= \sum_{Z \in \mathcal{P}_n(Q)} |S_Z| \leq \sum_{r=1}^n C_{r-1}^{n-1} (n-r)! \\ &= \sum_{r=1}^n \frac{(n-1)!}{(r-1)!} \\ &= \lfloor e(n-1)! \rfloor. \end{aligned}$$

The last equality is due to a well-known combinatorics identity.  $\square$

The above upper bound is met by the following monoid  $\mathcal{S}_n$ . For any  $Z \in \mathcal{P}_n(Q)$ , suppose  $Z = \{j_1, \dots, j_r\}$  such that  $j_1 < \dots < j_r$ ; then we define partition  $\pi_Z = \{Q\}$  if  $Z = \{n\}$ , and  $\pi_Z = \{\{j_1\}, \dots, \{j_{r-1}\}, Q \setminus \{j_1, \dots, j_{r-1}\}\}$  otherwise. Let

$$\mathcal{S}_n = \bigcup_{Z \in \mathcal{P}_n(Q)} \mathcal{E}(\pi_Z).$$

*Example 5.16.* Suppose  $n = 4$ ; then  $|\mathcal{P}_4(Q)| = 2^3 = 8$ . First consider  $Z = \{1, 3, 4\} \in \mathcal{P}_4(Q)$ . By definition,  $\pi_Z = \{\{1\}, \{3\}, \{2, 4\}\}$ . There is only one transformation  $t_1 =$

$[1, 4, 3, 4]$  in  $\mathcal{E}(\pi_Z)$ . If  $Z' = \{3, 4\}$ , then  $\pi_{Z'} = \{\{3\}, \{1, 2, 4\}\}$ . There are two transformations  $t_2 = [2, 4, 3, 4]$  and  $t_3 = [4, 4, 3, 4]$  in  $\mathcal{E}(\pi_{Z'})$ . Table 5.1 summaries the number of transformations in  $\mathcal{E}(\pi_Z)$  for each  $Z \in \mathcal{P}_4(Q)$ . Note that the set  $\mathcal{S}_4$  contains 16 transformations in total. ■

Table 5.1: Number of transformations in  $\mathcal{E}(\pi_Z)$  for each  $Z \in \mathcal{P}_4(Q)$ .

$Z$	Blocks of $\pi_Z$	$ \mathcal{E}(\pi_Z) $
$\{1, 2, 3, 4\}$	$\{1\}, \{2\}, \{3\}, \{4\}$	1
$\{1, 2, 4\}$	$\{1\}, \{2\}, \{3, 4\}$	1
$\{1, 3, 4\}$	$\{1\}, \{3\}, \{2, 4\}$	1
$\{2, 3, 4\}$	$\{2\}, \{3\}, \{1, 4\}$	1
$\{1, 4\}$	$\{1\}, \{2, 3, 4\}$	2
$\{2, 4\}$	$\{2\}, \{1, 3, 4\}$	2
$\{3, 4\}$	$\{3\}, \{1, 2, 4\}$	2
$\{4\}$	$\{1, 2, 3, 4\}$	6

**Proposition 5.17.** *The set  $\mathcal{S}_n$  is a  $\mathcal{J}$ -trivial submonoid of  $\mathcal{F}_Q$  with cardinality*

$$g(n) = |\mathcal{S}_n| = \sum_{r=1}^n C_{r-1}^{n-1} (n-r)! = \lfloor e(n-1)! \rfloor. \quad (5.1)$$

*Proof.* First we prove the following claim:

**Claim:** For any  $t, s \in \mathcal{S}_n$ ,  $\Omega(ts) = \pi_Z$  for some  $Z \in \mathcal{P}_n(Q)$ .

Let  $t \in \mathcal{E}(\pi_{Z_1})$  and  $s \in \mathcal{E}(\pi_{Z_2})$  for some  $Z_1, Z_2 \in \mathcal{P}_n(Q)$ . Suppose  $\Omega(ts) \neq \pi_Z$  for any  $Z \in \mathcal{P}_n(Q)$ . Then there exists a block  $X_0 \in \Omega(ts)$  such that  $n \notin X_0$  and  $|X_0| \geq 2$ . Suppose  $i \in X_0$  with  $i \neq \max(X_0)$ . We must have  $i \in \omega_t(n)$  or  $it \in \omega_s(n)$ ; otherwise  $it = i$  and  $(it)s = i$  and so  $i = \max(X_0)$ . However, in either case, there exists large  $m$  such that  $it^m = n$  or  $(it)s^m = n$ , respectively. Then  $n \in \omega_{ts}(i) = X_0$ , a contradiction. So the claim holds. ■

By the claim, for any  $t, s \in \mathcal{S}_n$ , since  $\Omega(ts) = \pi_Z$  for some  $Z \in \mathcal{P}_n(Q)$ ,  $ts \in \mathcal{E}(\pi_Z) \subseteq \mathcal{S}_n$ . Hence  $\mathcal{S}_n$  is a submonoid of  $\mathcal{F}_Q$ .

Next we show that  $\mathcal{S}_n$  is  $\mathcal{J}$ -trivial. Pick any  $t, s \in \mathcal{S}_n$ , and suppose  $t \in \mathcal{E}(\pi_{Z_1})$  and  $s \in \mathcal{E}(\pi_{Z_2})$  for some  $Z_1, Z_2 \in \mathcal{P}_n(Q)$ . Suppose  $\text{Max}(Z_1) \cap \text{Max}(Z_2) = \{j_1, \dots, j_r\}$ , for some  $r \geq 0$ . Then we have  $Z_1 \vee Z_2 = \{\{j_1\}, \dots, \{j_r\}, X\}$ , where  $X = Q \setminus \{j_1, \dots, j_r\}$  and  $n \in X$ . On the other hand, by the claim,  $\Omega(ts) = \{\{p_1\}, \dots, \{p_k\}, Y\}$ , where  $Y = Q \setminus \{p_1, \dots, p_k\}$ . Note that, since  $\mathcal{E}(\pi_{Z_1}), \mathcal{E}(\pi_{Z_2}) \subseteq \mathcal{F}_Q$ ,  $\text{Max}(\Omega(ts)) = \text{Fix}(ts) = \text{Fix}(t) \cap \text{Fix}(s) = \text{Max}(Z_1) \cap \text{Max}(Z_2)$ . Then  $r = k$  and  $\{j_1, \dots, j_r\} = \{p_1, \dots, p_k\}$ . Hence  $\Omega(t) \vee \Omega(s) = Z_1 \vee Z_2 = \Omega(ts)$ . By Theorem 5.7,  $\mathcal{S}_n$  is  $\mathcal{J}$ -trivial.

For any  $Z \in \mathcal{P}_n(Q)$  with  $|Z| = r$ , where  $1 \leq r \leq n$ , suppose  $\pi_Z = \{X_1, \dots, X_r\}$  with  $k_i = |X_i| = 1$  for  $1 \leq i < r$ , and  $k_r = |X_r|$ . By Lemma 5.13,  $|\mathcal{E}(\pi_Z)| = (n-r)!$ . Since  $n \in Z$  is fixed, there are  $C_{r-1}^{n-1}$  different  $Z$ . Therefore  $|\mathcal{S}_n| = \sum_{r=1}^n C_{r-1}^{n-1} (n-r)! = \lfloor e(n-1)! \rfloor$ .  $\square$

Let  $t$  be any transformation of  $Q$ . An orbit  $X$  of  $t$  is *trivial* if it contains just one element of  $Q$ ; otherwise it is *non-trivial*. Hence any transformation  $t \in \mathcal{S}_n$  has only one non-trivial orbit. We now define a generating set of the monoid  $\mathcal{S}_n$ .

**Definition 5.18.** Suppose  $n \geq 1$ . For any  $Z \in \mathcal{P}_n(Q)$ , if  $Z = Q$ , then let  $t_Z = \mathbf{1}_Q$ . Otherwise, let  $h_Z = \max(Q \setminus Z)$ , and let  $t_Z$  be a transformation of  $Q$  defined by: For all  $i \in Q$ ,

$$it \stackrel{\text{def}}{=} \begin{cases} i & \text{if } i \in Z, \\ n & \text{if } i = h_Z, \\ h_Z & \text{otherwise.} \end{cases}$$

Let  $\mathcal{GS}_n = \{t_Z \mid Z \in \mathcal{P}_n(Q)\}$ .

*Example 5.19.* Suppose  $n = 5$ . As the first example, consider  $Z = \{1, 3, 4, 5\}$ . Then  $h_Z = \max(Q \setminus Z) = 2$ , and  $t_Z = [1, 5, 3, 4, 5]$ . If  $Z' = \{4, 5\}$ , then  $h_{Z'} = 4$  and  $t_{Z'} = [3, 3, 5, 4, 5]$ . If  $Z'' = \{5\}$ , then  $h_{Z''} = 4$  and  $t_{Z''} = [4, 4, 4, 5, 5]$ . The set  $\mathcal{GS}_5$  contains the following 16 transformations:

$$\begin{array}{lll} t_1 = [1, 2, 3, 4, 5], & t_2 = [1, 2, 3, 5, 5], & t_3 = [1, 2, 4, 5, 5], \\ t_4 = [1, 2, 5, 4, 5], & t_5 = [1, 3, 5, 4, 5], & t_6 = [1, 4, 3, 5, 5], \\ t_7 = [1, 4, 4, 5, 5], & t_8 = [1, 5, 3, 4, 5], & t_9 = [2, 5, 3, 4, 5], \\ t_{10} = [3, 2, 5, 4, 5], & t_{11} = [3, 3, 5, 4, 5], & t_{12} = [4, 2, 3, 5, 5], \\ t_{13} = [4, 2, 4, 5, 5], & t_{14} = [4, 4, 3, 5, 5], & t_{15} = [4, 4, 4, 5, 5], \\ t_{16} = [5, 2, 3, 4, 5]. \end{array}$$

■

**Proposition 5.20.** For  $n \geq 1$ , the monoid  $\mathcal{S}_n$  can be generated by the set  $\mathcal{GS}_n$  of  $2^{n-1}$  transformations of  $Q$ .



*Proof.* First, for any  $t_Z \in \mathcal{GS}_n$ , where  $Z \in \mathcal{P}_n(Q)$ , we have  $\Omega(t_Z) = \pi_Z$ ; hence  $t_Z \in \mathcal{E}(\pi_Z) \subseteq \mathcal{S}_n$ . So  $\mathcal{GS}_n \subseteq \mathcal{S}_n$  and  $\langle \mathcal{GS}_n \rangle \subseteq \mathcal{S}_n$ .

Fix arbitrary  $Z \in \mathcal{P}_n(Q)$ , and suppose  $U = Q \setminus Z$ . Note that  $n \in Z$ . Let  $Y$  be the block in  $\pi_Z$  such that  $n \in Y$ . For any  $t \in \mathcal{E}(\pi_Z)$ , we have  $\text{Fix}(t) = Z$ . Furthermore, if  $i \in Q \setminus Y$ , then  $i \in \text{Fix}(t)$  and  $it = i$ . We prove by induction on  $|U|$  that  $\mathcal{E}(\pi_Z) \subseteq \langle \mathcal{GS}_n \rangle$ .

1.  $U = \emptyset$ : Then  $\pi_Z = \{\{1\}, \dots, \{n\}\}$ , and  $\mathcal{E}(\pi_Z) = \{\mathbf{1}_Q\}$ . Note that  $\mathbf{1}_Q \in \mathcal{GS}_n$ . So  $\mathcal{E}(\pi_Z) \subseteq \langle \mathcal{GS}_n \rangle$ .
2.  $U = \{h\}$  for some  $h \neq n$ : Then  $Y = \{h, n\}$ . For any  $t \in \mathcal{E}(\pi_Z)$ , since  $\text{Fix}(t) = Z$  and  $h \notin Z$ , we have  $ht > h$ . Since  $Y$  is an orbit of  $t$ , we have  $ht = n$ , and  $t = \binom{h}{n}$ . Note that  $t_Z = \binom{h}{n} = t$ . So  $\mathcal{E}(\pi_Z) \subseteq \langle \mathcal{GS}_n \rangle$ .
3.  $U = \{h_1, h_2\}$  for some  $h_1 < h_2 < n$ : Then  $Y = \{h_1, h_2, n\}$ . Note that  $t_Z = \binom{h_2}{n} \binom{h_1}{h_2}$ . For any  $t \in \mathcal{E}(\pi_Z)$ , since  $h_1 < h_2$ , and  $Y$  is an orbit of  $t$ , we have  $h_2t = n$  and  $h_1t \in \{h_2, n\}$ . If  $h_1t = h_2$ , then  $t = t_Z$ ; otherwise,  $t = t_Z^2$ . So  $\mathcal{E}(\pi_Z) \subseteq \langle \mathcal{GS}_n \rangle$ .
4.  $U = \{h_1, \dots, h_l\}$  for some  $h_1 < \dots < h_l < n$ , where  $l \geq 3$ : Assume that, for any  $Z' \in \mathcal{P}_n(Q)$  with  $|Q \setminus Z'| < l$ , we have  $\mathcal{E}(\pi_{Z'}) \subseteq \langle \mathcal{GS}_n \rangle$ . Now  $Y = \{h_1, \dots, h_l, n\}$ , and  $t_Z = \binom{h_l}{n} \binom{h_{l-1}}{h_l} \dots \binom{h_1}{h_l}$ . For any  $t \in \mathcal{E}(\pi_Z)$ , since  $Y$  is an orbit of  $t$  and  $Q \setminus Y \subseteq \text{Fix}(t)$ , we have

$$t = \binom{h_l}{p_l} \binom{h_{l-1}}{p_{l-1}} \dots \binom{h_1}{p_1},$$

where  $p_l = n$ , and  $p_i \in \{h_{i+1}, \dots, h_l, n\}$  for  $i = 2, \dots, l-1$ . We have three cases:

- (a)  $p_1 = \dots = p_l = n$ : Then  $t = t_Z^2$ , and  $t \in \langle \mathcal{GS}_n \rangle$ .
- (b)  $p_1 = \dots = p_{l-1} = h_l$ : Then  $t = t_Z$ , and  $t \in \langle \mathcal{GS}_n \rangle$  as well.
- (c) Otherwise, there exists some  $h_i$ , where  $1 \leq i < l$ , such that  $p_i = h_i t \notin \{h_l, n\}$ . Let  $h_r$  be the smallest such  $h_i$ , and let  $Y' = Y \setminus \{h_r\}$ . Then  $h_r \notin \text{rng}(t)$ , and  $p_r = h_r t \in Y' \setminus \{h_l, n\}$ . Now, let

$$t' = \binom{h_l}{n} \dots \binom{h_{r+1}}{p_{r+1}} \binom{h_{r-1}}{p_{r-1}} \dots \binom{h_1}{p_1}$$

and  $Z' = \text{Fix}(t')$ . Then  $Y'$  is an orbit of  $t'$ , and  $Z' = \text{Fix}(t) \cup \{h_r\}$ ; so  $t' \in \mathcal{E}(\pi_{Z'})$ . By assumption, since  $|Q \setminus Z'| = l-1 < l$ , we have  $t' \in \langle \mathcal{GS}_n \rangle$ . As the last step,

let  $Z'' = \{h_r, p_r\}$ . Since  $p_r = h_r t > h_r$ , we have  $t_{Z''} = \binom{p_r}{n} \binom{h_r}{p_r} \in \mathcal{GS}_n$ . Note that  $h_r \notin \text{rng}(t')$  and  $p_r \notin \text{Fix}(t')$ . So

$$\begin{aligned}
t' t_{Z''} &= \binom{h_l}{n} \cdots \binom{h_{r+1}}{p_{r+1}} \binom{h_{r-1}}{p_{r-1}} \cdots \binom{h_1}{p_1} \circ \binom{p_r}{n} \binom{h_r}{p_r} \\
&= \binom{h_l}{n} \cdots \binom{h_{r+1}}{p_{r+1}} \binom{h_{r-1}}{p_{r-1}} \cdots \binom{h_1}{p_1} \circ \binom{h_r}{p_r} \\
&= \binom{h_l}{n} \cdots \binom{h_{r+1}}{p_{r+1}} \binom{h_r}{p_r} \binom{h_{r-1}}{p_{r-1}} \cdots \binom{h_1}{p_1} \\
&= t.
\end{aligned}$$

Thus  $t \in \langle \mathcal{GS}_n \rangle$ .

By induction we have  $\mathcal{S}_n = \bigcup_{Z \in \mathcal{P}_n(Q)} \mathcal{E}(\pi_Z) \subseteq \langle \mathcal{GS}_n \rangle$ . Therefore  $\mathcal{S}_n = \langle \mathcal{GS}_n \rangle$ . Since there are  $2^{n-1}$  different  $Z \in \mathcal{P}_n(Q)$ , there are  $2^{n-1}$  transformations in  $\mathcal{GS}_n$ .  $\square$

*Example 5.21.* Suppose  $n = 5$ . The list of all transformations in  $\mathcal{GS}_5$  is shown in Example 5.19. Consider  $Z = \{3, 5\} \in \mathcal{P}_5(Q)$ , and  $t = [2, 4, 3, 5, 5] \in \mathcal{E}(\pi_Z)$ . The transition graph of  $t$  is shown in Figure 5.4 (a). As in Proposition 5.20, we have  $Y = \{1, 2, 4, 5\}$ , and  $U = \{1, 2, 4\}$ . To show that  $t \in \langle \mathcal{GS}_5 \rangle$ , we find  $h_r = 1$ . Then  $t' = \binom{4}{5} \binom{2}{4} = [1, 4, 3, 5, 5]$ , and  $Z' = \{1, 3, 5\}$ . We assume that  $t' \in \langle \mathcal{GS}_5 \rangle$ ; in fact,  $t' = t_{Z'}$  in this example. We also need  $Z'' = \{1, 1t\} = \{1, 2\}$ , and  $t_{Z''} = \binom{2}{5} \binom{1}{2} = [2, 5, 3, 4, 5]$ . The transition graphs of  $t'$  and  $t_{Z''}$  are shown in Figure 5.4 (a) and (b), respectively. One can verify that  $t = t' t_{Z''}$ , and hence  $t \in \langle \mathcal{GS}_5 \rangle$ .  $\blacksquare$

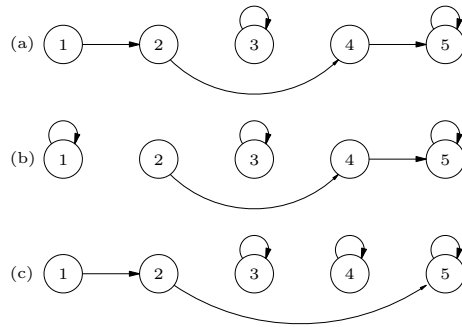


Figure 5.4: Transition graphs of  $t = [2, 4, 3, 5, 5]$ ,  $t' = [1, 4, 3, 5, 5]$ , and  $t_{Z''} = [2, 5, 3, 4, 5]$ .

Now, by Propositions 5.15, 5.17, and 5.20, we have

**Theorem 5.22.** *Let  $L \subseteq \Sigma^*$  be a regular language with quotient complexity  $n \geq 1$  and  $\mathcal{J}$ -trivial syntactic monoid. Then its syntactic complexity  $\sigma(L)$  satisfies  $\sigma(L) \leq g(n) = \lfloor e(n-1)! \rfloor$ , and this bound is tight if  $|\Sigma| \geq 2^{n-1}$ .*

*Remark 5.23.* It was shown by Saito [42] that, if  $S$  is a  $\mathcal{J}$ -trivial submonoid of  $\mathcal{F}_Q$ , then  $\Omega(S) = \{\Omega(t) \mid t \in S\} \subseteq \Pi_Q$  forms a  $\vee$ -semilattice such that  $\text{Max}(\Omega(t) \vee \Omega(s)) = \text{Fix}(t) \cap \text{Fix}(s)$ , called a  $\mathcal{J}$ - $\vee$ -semilattice. Let  $\mathcal{P}_\vee(\Pi_Q)$  be the set of all  $\mathcal{J}$ - $\vee$ -semilattices that are subsets of  $\Pi_Q$ . A maximal  $\mathcal{J}$ -trivial submonoid  $S$  of  $\mathcal{F}_Q$  corresponds to an maximal element  $P$  in  $\mathcal{P}_\vee(\Pi_Q)$ , with respect to set inclusion, such that  $S = \bigcup_{\pi \in P} \mathcal{E}(\pi)$ .  $P \in \mathcal{P}_\vee(\Pi_Q)$  is called *full* if  $\{\text{Max}(\pi) \mid \pi \in P\} = \mathcal{P}_n(Q)$ , which is an maximal element in  $\mathcal{P}_\vee(\Pi_Q)$  with respect to set inclusion. The monoid  $\mathcal{S}_n$  then corresponds to a full  $\mathcal{J}$ - $\vee$ -semilattice, and hence it is maximal. Saito described all maximal  $\mathcal{J}$ -trivial submonoid of  $\mathcal{F}_Q$  and those corresponding to full  $\mathcal{J}$ - $\vee$ -semilattices. However, here we consider the  $\mathcal{J}$ -trivial submonoid of  $\mathcal{F}_Q$  with maximum cardinality.

*Remark 5.24.* The number  $\lfloor e(n-1)! \rfloor$  also appears in the paper of Brzozowski and Liu [13] as a lower bound and the conjectured upper bound for the syntactic complexity of definite languages. However, the semigroup  $B_n$  with this cardinality in [13] for definite languages is not isomorphic to  $\mathcal{S}_n$ , since  $B_n$  is not  $\mathcal{J}$ -trivial.

## Chapter 6

# Quotient Complexity of Reverse Languages

In this chapter we consider *nondeterministic finite automata* (NFA's). A NFA  $\mathcal{N}$  is a quintuple  $\mathcal{N} = (Q, \Sigma, \delta, I, F)$ , where  $Q$ ,  $\Sigma$ , and  $F$  are as in a DFA,  $\delta : Q \times \Sigma \rightarrow 2^Q$  is the nondeterministic transition function, and  $I$  is the set of initial states. For any word  $w \in \Sigma^*$ , the *reverse* of  $w$  is defined inductively as follows:  $w^R = \varepsilon$  if  $w = \varepsilon$ , and  $w^R = u^R a$  if  $w = au$  for some  $a \in \Sigma$  and  $u \in \Sigma^*$ . The *reverse* of any language  $L$  is the language  $L^R = \{w^R \mid w \in L\}$ . For any finite automaton (DFA or NFA)  $\mathcal{M}$ , we let  $\mathcal{M}^R$  denote the NFA obtained by reversing all the transitions of  $\mathcal{M}$  and exchanging the roles of initial and final states, and by  $\mathcal{M}^D$ , the DFA obtained by applying the subset construction to  $\mathcal{M}$ . Then  $L(\mathcal{M}^R) = (L(\mathcal{M}))^R$ , and  $L(\mathcal{M}^D) = L(\mathcal{M})$ . To simplify our proofs, we use an observation from [9] that, for any NFA  $\mathcal{N}$  without empty states, if the automaton  $\mathcal{N}^R$  is deterministic, then the DFA  $\mathcal{N}^D$  is minimal if only reachable subsets are included in the subset construction.

For any regular language  $L$  with quotient complexity  $\kappa(L) = n \geq 2$ , the upper bound on the quotient complexity of the reverse language  $L^R$  is  $2^n$  [39], and this bound is tight [33]. In 2004, Salomaa, Wood, and Yu showed that if a regular language  $L$  has quotient complexity  $n \geq 2$  and syntactic complexity  $n^n$ , then its reverse language  $L^R$  has quotient complexity  $2^n$ . As shown in [8], for certain regular languages with maximal syntactic complexity in their subclasses, the reverse languages have maximal quotient complexity. We now show that the similar statement holds true for some subclasses of regular languages studied in previous chapters.

First, consider suffix-, bifix-, and factor-free regular languages. Although we do not

have tight upper bounds on their syntactic complexities, for languages in these classes with the maximal known syntactic complexities, their reverse languages reach the upper bounds on the quotient complexities of the reversal operation.

**Theorem 6.1.** *For  $n \geq 4$ , the reverse of the suffix-free regular language accepted by the DFA  $\mathcal{A}'_n$  of Theorem 3.17 restricted to  $\{a_1, a_2, a_3, c\}$  has  $2^{n-2} + 1$  quotients, which is the maximum possible for a suffix-free regular language.*

*Proof.* Let  $\mathcal{C}_n$  be the DFA  $\mathcal{A}'_n$  restricted to the alphabet  $\{a_1, a_2, a_3, c\}$ . Since  $L(\mathcal{A}'_n)$  is suffix-free, so is  $L'_n = L(\mathcal{C}_n)$ . Let  $\mathcal{N}'_n$  be the NFA obtained from  $\mathcal{C}_n^R$  by removing unreachable states. Figure 6.1 shows the NFA  $\mathcal{N}'_6$ .

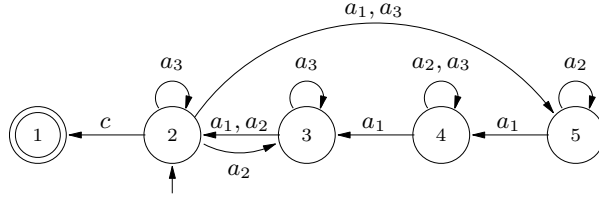


Figure 6.1: NFA  $\mathcal{N}'_6$  of  $L'_6$  with quotient complexity  $\kappa(L'_6) = 17$ ; empty state omitted.

Applying the subset construction to  $\mathcal{N}'_n$ , we get a DFA  $\mathcal{N}'_n{}^D$ . Its initial state is a singleton set  $\{2\}$ . From the initial state, we can reach state  $\{2, 3, \dots, i\}$  by  $(a_3 a_1^{n-3})^{i-2}$ , where  $3 \leq i \leq n-1$ . Then the state  $\{2, 3, \dots, n-1\}$  is reached from  $\{2\}$  by  $(a_3 a_1^{n-3})^{n-3}$ . Assume that any set  $S$  of cardinality  $l$  can be reached, where  $2 \leq l \leq n-2$ . If  $j \in S$ , then we can reach  $S' = S \setminus \{j\}$  from  $S$  by  $a_1^{j-1} a_3 a_1^{n-j-1}$ . So all the non-empty subsets of  $\{2, 3, \dots, n-1\}$  can be reached. We can also reach the singleton set  $\{1\}$  from  $\{2\}$  by  $c$ , and, from there, the empty state by  $c$  again. Hence  $\mathcal{N}'_n{}^D$  has  $2^{n-2} + 1$  reachable states.

Since the automaton  $\mathcal{N}'_n{}^R$ , the reverse of  $\mathcal{N}'_n$ , is a subset of  $\mathcal{C}_n$ , it is deterministic; hence  $\mathcal{N}'_n{}^D$  is minimal. Then the quotient complexity of  $L'_n$  is  $2^{n-2} + 1$ , which meets the upper bound for reversal of suffix-free regular languages [20].  $\square$

**Theorem 6.2.** *For  $n \geq 5$ , the reverse of the factor-free regular language accepted by the DFA  $\mathcal{A}_n$  of Theorem 3.39 restricted to the alphabet  $\{a_1, a_2, a_3, c\}$ , where  $c = [2, n-1, n, \dots, n, n] \in \mathbf{G}_{\text{ff}}(n)$ , has  $2^{n-3} + 2$  quotients, which is the maximum possible for a bifactor-free regular language.*

*Proof.* Let  $\mathcal{D}_n$  be the DFA  $\mathcal{A}_n$  restricted to the alphabet  $\{a_1, a_2, a_3, c\}$ ; then  $L_n'' = L(\mathcal{D}_n)$  is factor-free. Let  $\mathcal{N}_n''$  be the NFA obtained from  $\mathcal{D}_n^R$  by removing unreachable states. An example of  $\mathcal{N}_n''$  is shown in Figure 6.2.

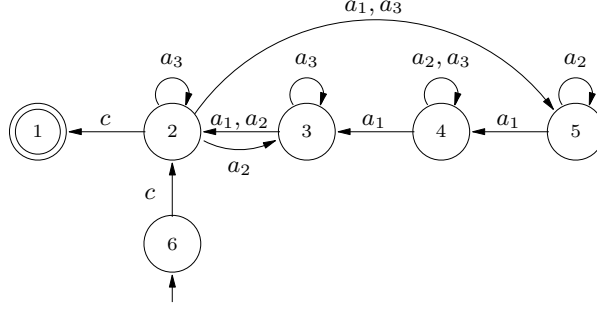


Figure 6.2: NFA  $\mathcal{N}_7''$  of  $L_7^{''R}$  with quotient complexity  $\kappa(L_7^{''R}) = 18$ ; empty state omitted.

Note that  $\mathcal{N}_n''$  can be obtained from the NFA  $\mathcal{N}_{n-1}'$  in Theorem 6.1 by adding a new state  $n - 1$ , which is the only initial state in  $\mathcal{N}_n''$ , and the transition from  $\{n - 1\}$  to  $\{2\}$  under input  $c$ . We know that all non-empty subsets of  $\{2, 3, \dots, n - 2\}$  are reachable from  $\{2\}$ . The final state  $\{1\}$  is also reachable from  $\{2\}$ . From the initial state  $n - 1$ , we reach the empty state under input  $a_1$ . Then  $\mathcal{N}_n^{''D}$  has  $2^{n-3} + 2$  reachable states.

Since  $\mathcal{N}_n^{''R}$  is a subset of  $\mathcal{D}_n$  and it is deterministic, the DFA  $\mathcal{N}_n^{''D}$  is minimal. Therefore  $\kappa(L_n^{''R}) = 2^{n-3} + 2$ , and it reaches the upper bound for reversal of both bifix- and factor-free regular languages with quotient complexity  $n$  [11].  $\square$

We prove in the following the tight upper bounds on the quotient complexities of the reverse of  $\mathcal{R}$ - and  $\mathcal{J}$ -trivial regular languages. These bounds can be reached by languages with maximal syntactic complexities in their classes.

**Theorem 6.3.** *For  $n \geq 2$ , if  $L$  is a regular  $\mathcal{R}$ -trivial language with quotient complexity  $\kappa(L) = n$ , then  $\kappa(L^R) \leq 2^{n-1}$ . Moreover, this bound can be met if  $L$  is over an alphabet of size  $n$ .*

*Proof.* Let  $\mathcal{A} = (Q, \Sigma, \delta, q_1, F)$  be the quotient DFA of  $L$ , where  $|Q| = n$ . Suppose  $L$  is  $\mathcal{R}$ -trivial; then  $\mathcal{A}$  must be partially ordered. Assume  $q_1 = 1$ , and upon reordering elements in  $Q$ , let the reachability relation  $\rightarrow$  as a partial order be *compatible* with the natural order  $<$  on  $Q$ ; that is, for all  $p, q \in Q$ ,  $p \rightarrow q$  implies  $p < q$ . Then for all  $a \in \Sigma$ , we must have  $\delta(n, a) = n$ . Let  $\mathcal{N} = \mathcal{A}^R$  be the NFA accepting  $L^R$ . There are two cases:

1.  $n \in F$ : Then  $n$  is in the set of initial states of  $\mathcal{N}$ . Since  $\delta(n, a) = n$  for all  $a \in \Sigma$ , for any reachable set  $P$  of states of  $\mathcal{N}$ , we must have  $n \in P$ . There are  $2^{n-1}$  subsets of  $Q$  containing  $n$ . So there are at most  $2^{n-1}$  reachable set of states of  $\mathcal{N}$ .
2.  $n \notin F$ : Now, since  $\delta(n, a) = n$ , the state  $n$  is not reachable in  $\mathcal{N}$ . There are  $2^{n-1}$  subsets of  $Q$  not containing  $n$ . So there are at most  $2^{n-1}$  reachable set of states of  $\mathcal{N}$  as well.

Therefore  $\kappa(L^R) \leq 2^{n-1}$ .

To see the bound is tight, consider the DFA  $\mathcal{A}_n = (Q, \Sigma, \delta, 1, \{n\})$ , where  $Q = \{1, \dots, n\}$ ,  $\Sigma = \{a_1, \dots, a_n\}$ , and the transitions are defined such that:

- For  $1 \leq i \leq n - 2$ ,  $\delta(i, a_i) = n$ , and  $\delta(p, a_i) = p$  if  $p \neq i$ ;
- $\delta(p, a_{n-1}) = p + 1$  if  $p < n$ , and  $\delta(n, a_{n-1}) = n$ ;
- $\delta(1, a_n) = 2$ , and  $\delta(p, a_n) = p$  if  $p \neq 1$ .

Let  $L_n = L(\mathcal{A}_n)$ , and let  $\mathcal{N}_n = \mathcal{A}_n^R$  be an NFA accepting  $L_n^R$ . The NFA  $\mathcal{N}_5$  is shown in Figure 6.3. Note that there is no unreachable state in  $\mathcal{A}_n^R$ . Clearly  $\mathcal{A}_n$  is partially ordered; so  $L_n$  is  $\mathcal{R}$ -trivial.

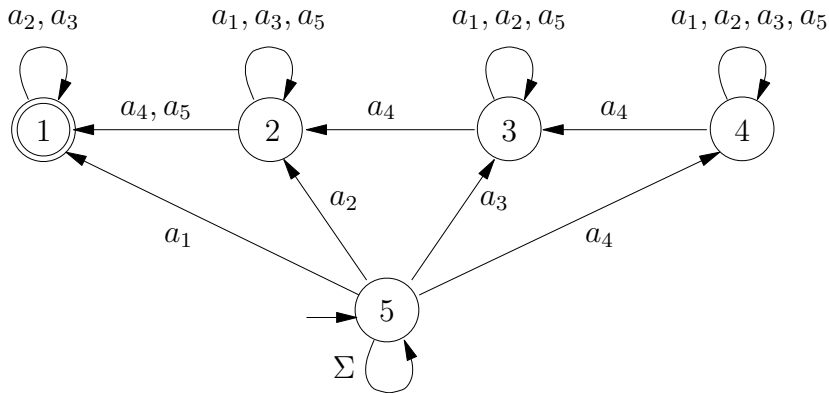


Figure 6.3: NFA  $\mathcal{N}_5 = \mathcal{A}_5^R$  for  $n = 5$  accepting  $L_5^R$ .

We now show that all  $2^{n-1}$  subsets  $P$  of  $Q$  containing  $n$  is reachable in  $\mathcal{N}_n$ . First,  $I = \{n\}$  is the initial subset of  $\mathcal{N}_n$ . If  $P = \{i, n\}$ , where  $1 \leq i \leq n - 1$ , then  $P$  is reached

by  $a_i$ . Suppose  $P = \{i_1, \dots, i_k, n\}$ , where  $1 \leq i_1 < \dots < i_k < n$  and  $k \geq 2$ . Note that, for any  $1 < p < n$ , we have  $\delta(p, a_q) = p$  for all  $q < p$ . Then  $P$  can be reached by the word  $a_{i_k} \cdots a_{i_1}$ . Hence  $\mathcal{N}_n^D$  has  $2^{n-1}$  reachable states.

Since there is no unreachable state in  $\mathcal{A}_n^R$ , and  $\mathcal{N}_n^R = \mathcal{A}_n$  is deterministic, the DFA  $\mathcal{N}_n^D$  is minimal. Then the quotient complexity of  $L_n^R$  is  $2^{n-1}$ , which is the tight upper bound for reversal of  $\mathcal{R}$ -trivial regular languages.  $\square$

**Theorem 6.4.** *For  $n \geq 2$ , if  $L$  is a regular  $\mathcal{J}$ -trivial language with quotient complexity  $\kappa(L) = n$ , then  $\kappa(L^R) \leq 2^{n-1}$ . Moreover, this bound can be met if  $L$  is over an alphabet of size  $2^{n-1} - 1$ .*

*Proof.* Since any  $\mathcal{J}$ -trivial regular language is also  $\mathcal{R}$ -trivial, the upper bound  $2^{n-1}$  in Theorem 6.3 also holds for  $\mathcal{J}$ -trivial regular languages.

To see that the bound is tight, consider the DFA  $\mathcal{B}_n = (Q, \Sigma, \delta, 1, \{n\})$  such that  $Q = \{1, \dots, n\}$ ,  $|\Sigma| = 2^{n-1} - 1$ , and each  $a \in \Sigma$  defines a distinct transformation in  $\mathcal{GS}_n$  other than  $\mathbf{1}_Q$ . Let  $L_n = L(\mathcal{B}_n)$ . By Proposition 5.20, the transition semigroup  $\mathcal{T}_{\mathcal{B}_n}$  of  $\mathcal{B}_n$  contains all transformations in  $\mathcal{S}_n$  other than  $\mathbf{1}_Q$ .

Let  $\mathcal{N}_n = \mathcal{B}_n^R$  be an NFA accepting  $L_n^R$ , which contains no unreachable state. Let  $P$  be any subset of  $Q$  containing  $n$ . If  $P = \{n\}$ , then it is the initial set of states of  $\mathcal{N}_n$ . Otherwise, suppose  $P = \{p_1, \dots, p_k, n\}$  for some  $p_i \in Q$  and  $1 \leq k \leq n - 1$ . Let  $t = \binom{p_1}{n} \cdots \binom{p_k}{n}$  be a transformation of  $Q$ , and let  $R = Q \setminus P = \{q_1, \dots, q_l\}$ , where  $k+l+1 = n$ . Then  $\text{Fix}(t) = R \cup \{n\}$ , and  $\Omega(t) = \{\{q_1\}, \dots, \{q_l\}, R \cup \{n\}\}$ . So  $t \in \mathcal{S}_n$ . Clearly  $t \neq \mathbf{1}_Q$  as  $k \geq 1$ . Thus  $t \in \mathcal{T}_{\mathcal{B}_n}$ , and there exists  $w \in \Sigma^*$  such that  $w$  performs the transformation  $t$ , i.e.,  $t_w = t$ . This means that, for any  $p \in Q$ ,  $\delta(p, w) = n$  if and only if  $p \in P$ . Hence we can reach the set  $P$  of states of  $\mathcal{N}_n$  from the initial set of states by the word  $w$ . Since there are  $2^{n-1}$  distinct subsets  $P$  of  $Q$  containing  $n$ , there are  $2^{n-1}$  reachable states in  $\mathcal{N}_n^D$ .

Note that there is no unreachable state in  $\mathcal{B}_n^R$ . Then the DFA  $\mathcal{N}_n^D$  is minimal, and  $\kappa(L_n^R) = 2^{n-1}$ . This shows that the upper bound  $2^{n-1}$  is tight for reversal of  $\mathcal{J}$ -trivial regular languages.  $\square$



# Chapter 7

## Conclusions

We have presented our results on syntactic complexity of several subclasses of regular languages, namely suffix-, bifix-, and factor-free regular languages, star-free languages and three subclasses, and  $\mathcal{R}$ - and  $\mathcal{J}$ -trivial regular languages. We found upper bounds on the syntactic complexities of these classes of languages. For monotonic, partially monotonic, and partially monotonic languages, and for  $\mathcal{R}$ - and  $\mathcal{J}$ -trivial regular languages, the upper bounds are tight. For the other classes of languages, we found tight upper bounds for languages with small quotient complexities, and we also found lower bounds. We conjecture these lower bounds to be tight upper bounds for these languages.

Once again we have demonstrated that upper bounds on syntactic complexity are different for different subclasses of regular languages. In particular, this is true for classes where one class is contained in another. Hence syntactic complexity may be able to distinguish different subclasses of regular languages with the same state / quotient complexity, and it is a useful complexity measure in addition to state / quotient complexity.

We also observed that, for some subclasses  $\mathcal{C}$  of regular languages, the upper bound on quotient complexity of the reversal operation on languages in  $\mathcal{C}$  can be met by languages in  $\mathcal{C}$  with maximal syntactic complexity. So far we were not able to generalize this observation.

We hope that our results can stimulate more studies in syntactic complexity. For possible future directions, one should certainly try to prove tight upper bounds on syntactic complexity of suffix-, bifix-, and factor-free regular languages, and star-free languages. For related classes of languages, the problem of syntactic complexity of subword-free and  $\mathcal{L}$ -trivial regular languages is still open. One can also consider languages in the various hierarchies of star-free languages, for example, the dot-depth hierarchy [15], the Straubing-Thérien hierarchy [50, 51], and depth-one hierarchy [48]. Within the depth-one hierarchy,

definite and reverse definite languages were considered by Brzozowski and Liu [13], and piecewise-testable languages were considered here. One might also consider the syntactic complexity of the languages obtained from regular operations as a function of the state / quotient complexities of the operands.

# Permissions Page

# ELSEVIER LICENSE TERMS AND CONDITIONS

Jul 31, 2012

---

This is a License Agreement between Baiyu Li ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Baiyu Li
Customer address	University of Waterloo Waterloo, ON N2L 3G1
License number	2953120502817
License date	Jul 20, 2012
Licensed content publisher	Elsevier
Licensed content publication	Theoretical Computer Science
Licensed content title	Syntactic complexity of prefix-, suffix-, bifix-, and factor-free regular languages
Licensed content author	Janusz Brzozowski,Baiyu Li,Yuli Ye
Licensed content date	31 August 2012
Licensed content volume number	449
Licensed content issue number	None
Number of pages	17
Start Page	37

End Page	53
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	None
Title of your thesis/dissertation	Syntactic Complexities of Nine Subclasses of Regular Languages
Expected completion date	Jul 2012
Estimated size (number of pages)	79
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
<b>Total</b>	<b>0.00 USD</b>
<b>Terms and Conditions</b>	

#### INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

#### GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of

article, Pages No., Copyright (Year), with permission from Elsevier.”

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at [permissions@elsevier.com](mailto:permissions@elsevier.com))

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be

responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

#### LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation:** This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Website:** The following terms and conditions apply to electronic reserve and author websites:

**Electronic reserve:** If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:

This license was made in connection with a course,

This permission is granted for 1 year only. You may obtain a license for future website posting,

All content posted to the web site must maintain the copyright information line on the bottom of each image,

A hyper-text must be included to the Homepage of the journal from which you are licensing at

<http://www.sciencedirect.com/science/journal/xxxx> or the Elsevier homepage for books at

<http://www.elsevier.com> , and

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

All content posted to the web site must maintain the copyright information line on the bottom of each image, and the permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version. A hyper-text must be included to the Homepage of the journal from which you are licensing at

<http://www.sciencedirect.com/science/journal/xxxx> . As part of our normal production process, you will

receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect

([www.sciencedirect.com](http://www.sciencedirect.com)). That e-mail will include the article's Digital Object Identifier (DOI). This

number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses:

Authors are permitted to place a brief summary of their work online only.

A hyper-text must be included to the Elsevier homepage at <http://www.elsevier.com> . All content

posted to the web site must maintain the copyright information line on the bottom of each image. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxx>. or for books to the Elsevier homepage at <http://www.elsevier.com>

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other Conditions**:

v1.6

**If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK500822508.**

**Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.**

**Make Payment To:  
Copyright Clearance Center  
Dept 001  
P.O. Box 843006  
Boston, MA 02284-3006**

**For suggestions or comments regarding this order, contact RightsLink Customer Support: [customercare@copyright.com](mailto:customercare@copyright.com) or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.**

**Gratis licenses (referencing \$0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.**

---

---



# SPRINGER LICENSE TERMS AND CONDITIONS

Jul 31, 2012

---

This is a License Agreement between Baiyu Li ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

License Number	2953130977196
License date	Jul 20, 2012
Licensed content publisher	Springer
Licensed content publication	Springer eBook
Licensed content title	Syntactic Complexities of Some Classes of Star-Free Languages
Licensed content author	Janusz Brzozowski
Licensed content date	Jul 9, 2012
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	3
Author of this Springer article	Yes and you are a contributor of the new work
Order reference number	None
Title of your thesis / dissertation	Syntactic Complexities of Nine Subclasses of Regular Languages
Expected completion date	Jul 2012
Estimated size(pages)	79
<b>Total</b>	<b>0.00 USD</b>

## Terms and Conditions

### Introduction

The publisher for this copyrighted material is Springer Science + Business Media. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

### Limited License

With reference to your request to reprint in your thesis material on which Springer Science and Business Media control the copyright, permission is granted, free of charge, for the use indicated in your enquiry.

Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process.

This License includes use in an electronic form, provided its password protected or on the university's intranet or repository, including UMI (according to the definition at the Sherpa website: <http://www.sherpa.ac.uk/romeo/>). For any other electronic use, please contact Springer at ([permissions.dordrecht@springer.com](mailto:permissions.dordrecht@springer.com) or [permissions.heidelberg@springer.com](mailto:permissions.heidelberg@springer.com)).

The material can only be used for the purpose of defending your thesis, and with a maximum of 100 extra copies in paper.

Although Springer holds copyright to the material and is entitled to negotiate on rights, this license is only valid, provided permission is also obtained from the (co) author (address is given with the article/chapter) and provided it concerns original material which does not carry references to other sources (if material in question appears with credit to another source, authorization from that source is required as well).

Permission free of charge on this occasion does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

### Altering/Modifying Material: Not Permitted

You may not alter or modify the material in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s) and/or Springer Science + Business Media. (Please contact Springer at ([permissions.dordrecht@springer.com](mailto:permissions.dordrecht@springer.com) or [permissions.heidelberg@springer.com](mailto:permissions.heidelberg@springer.com)))

### Reservation of Rights

Springer Science + Business Media reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

### Copyright Notice:Disclaimer

You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Springer and the original publisher /journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure number(s), original copyright notice) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media"

### Warranties: None

Example 1: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

#### Indemnity

You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

#### No Transfer of License

This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

#### No Amendment Except in Writing

This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

#### Objection to Contrary Terms

Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

#### Jurisdiction

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut' (Netherlands Institute of Arbitration). **OR:**

**All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.**

**Other terms and conditions:**

**v1.3**

**If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK500822560.**

**Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.**

**Make Payment To:**

**Copyright Clearance Center**

**Dept 001**

P.O. Box 843006  
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support:  
[customercare@copyright.com](mailto:customercare@copyright.com) or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing \$0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

---

---

# References

- [1] Ananichev, D., Volkov, M.: Synchronizing generalized monotonic automata. *Theoret. Comput. Sci.* **330**(1) (2005) 3–13
- [2] Berstel, J., Perrin, D., Reutenauer, C.: *Codes and Automata*. Cambridge University Press (2009)
- [3] Brzozowski, J.: Quotient complexity of regular languages. *J. Autom. Lang. Comb.* **15**(1/2) (2010) 71–89
- [4] Brzozowski, J.: In search of most complex regular languages. In Moreira, N., Reis, R., eds.: *17th International Conference on Implementation and Application of Automata (CIAA)*. Volume 7381 of LNCS, Springer (2012) 5–24
- [5] Brzozowski, J., Fich, F.E.: Languages of  $\mathcal{R}$ -trivial monoids. *J. Comput. System Sci.* **20**(1) (1980) 32 – 49
- [6] Brzozowski, J., Li, B., Ye, Y.: Syntactic complexity of prefix-, suffix-, bifix-, and factor-free regular languages. *Theoret. Comput. Sci.* **449** (2012) 37 – 53
- [7] Brzozowski, J., Liu, B.: Quotient complexity of star-free languages. In Dömösi, P., Szabolcs, I., eds.: *13th International Conference on Automata and Formal Languages (AFL)*. Institute of Mathematics and Informatics, College of Nyíregyháza, Debrecen, Hungary (2011) 138–152
- [8] Brzozowski, J., Ye, Y.: Syntactic complexity of ideal and closed languages. In Mauri, G., Leporati, A., eds.: *15th International Conference on Developments in Language Theory, DLT 2011*. Volume 6795 of LNCS, Springer Berlin / Heidelberg (2011) 117–128

- [9] Brzozowski, J.: Canonical regular expressions and minimal state graphs for definite events. In: *Mathematical theory of Automata*. Volume 12 of MRI Symposia Series. Polytechnic Press, Polytechnic Institute of Brooklyn, N.Y. (1962) 529–561
- [10] Brzozowski, J., Jirásková, G., Li, B.: Quotient complexity of ideal languages. In López-Ortiz, A., ed.: *Proceedings of the 9th Latin American Theoretical Informatics Symposium (LATIN)*. Volume 6034 of LNCS, Springer (2010) 208–211 (Full paper at <http://arxiv.org/abs/0908.2083>).
- [11] Brzozowski, J., Jirásková, G., Li, B., Smith, J.: Quotient complexity of bifix-, factor-, and subword-free regular languages. In Dömösi, P., Iván, S., eds.: *Automata and Formal Languages 13th International Conference AFL 2011*, College of Nyíregyháza, Debrecen, Hungary (2011) 123–137
- [12] Brzozowski, J., Li, B.: Syntactic complexities of some classes of star-free languages. In Kutrib, M., Moreira, N., Reis, R., eds.: *14th International Workshop on Descriptive Complexity of Formal Systems (DCFS)*. Volume 7386 of LNCS, Springer (2012) 117–129
- [13] Brzozowski, J., Liu, D.: Syntactic complexity of finite/cofinite, definite, and reverse definite languages. <http://arxiv.org/abs/1103.2986> (March 2012)
- [14] Cayley, A.: A theorem on trees. *Quart. J. Math.* **23** (1889) 376–378
- [15] Cohen, R.S., Brzozowski, J.: Dot-depth of star-free events. *J. Comput. System Sci.* **5**(1) (1971) 1 – 16
- [16] Dénes, J.: On transformations, transformation semigroups and graphs. In Erdős, P., Katona, G., eds.: *Theory of Graphs. Proceedings of the Colloquium on Graph Theory held at Tihany, 1966*, Akadémiai Kiado (1968) 65–75
- [17] Ganyushkin, O., Mazorchuk, V.: *Classical Finite Transformation Semigroups: An Introduction*. Springer (2009)
- [18] GAP-Group: GAP - Groups, Algorithms, Programming - a System for Computational Discrete Algebra. (2010) <http://www.gap-system.org/>.
- [19] Gomes, G., Howie, J.: On the ranks of certain semigroups of order-preserving transformations. *Semigroup Forum* **45** (1992) 272–282
- [20] Han, Y.S., Salomaa, K.: State complexity of basic operations on suffix-free regular languages. *Theoret. Comput. Sci.* **410**(27-29) (2009) 2537–2548

- [21] Han, Y.S., Salomaa, K., Wood, D.: Operational state complexity of prefix-free regular languages. In Ésik, Z., Fülöp, Z., eds.: Automata, Formal Languages, and Related Topics, Inst. of Informatics, University of Szeged, Hungary (2009) 99–115
- [22] Holzer, M., Kutrib, M., Meckel, K.: Nondeterministic state complexity of star-free languages. In Bouchou-Markhoff, B., Caron, P., Champarnaud, J.M., Maurel, D., eds.: 16th International Conference on Implementation and Application of Automata (CIAA). Volume 6807 of LNCS. Springer (2011) 178–189
- [23] Holzer, M., König, B.: On deterministic finite automata and syntactic monoid size. Theoret. Comput. Sci. **327**(3) (2004) 319–347
- [24] Howie, J.M.: Products of idempotents in certain semigroups of transformations. Proc. Edinburgh Math. Soc. **17**(2) (1971) 223–236
- [25] Howie, J.M.: Fundamentals of Semigroup Theory (London Mathematical Society Monographs New Series). Oxford University Press, USA (February 1996)
- [26] Janson, S., Mazorchuk, V.: Some remarks on the combinatorics of  $\mathcal{IS}_n$ . Semigroup Forum **70** (2005) 391–405
- [27] Klíma, O., Polák, L.: On biautomata. In Freund, R., Holzer, M., Mereghetti, C., Otto, F., Palano, B., eds.: Third Workshop on Non-Classical Models for Automata and Applications - NCMA 2011, Milan, Italy, July 18 – July 19, 2011. Proceedings. Volume 282, Austrian Computer Society (2011) 153–164
- [28] Krawetz, B., Lawrence, J., Shallit, J.: State complexity and the monoid of transformations of a finite set. <http://arxiv.org/abs/math/0306416v1> (2003)
- [29] Krohn, K., Rhodes, J.: Algebraic theory of machines. i. prime decomposition theorem for finite semigroups and machines. Trans. Amer. Math. Soc. **116** (1965) 450 – 464
- [30] Laradji, A., Umar, A.: Asymptotic results for semigroups of order-preserving partial transformations. Comm. Alg. **34**(3) (2006) 1071–1075
- [31] Maslov, A.N.: Estimates of the number of states of finite automata. Dokl. Akad. Nauk SSSR **194** (1970) 1266–1268 (Russian) English translation: Soviet Math. Dokl. **11** (1970), 1373–1375.
- [32] McNaughton, R., Papert, S.A.: Counter-Free Automata. Volume 65 of M.I.T. Research Monographs. The MIT Press (1971)

- [33] Mirkin, B.G.: On dual automata. *Cybernet. Systems Anal.* **2** (1966) 6–9
- [34] Myhill, J.: Finite automata and the representation of events. Wright Air Development Center Technical Report **57–624** (1957)
- [35] Nerode, A.: Linear automaton transformations. *Proc. Amer. Math. Soc.* **9** (1958) 541–544
- [36] Piccard, S.: Sur les fonctions définies dans les ensembles finis quelconques. *Fund. Math.* **24** (1935) 298–301
- [37] Pighizzini, G., Shallit, J.: Unary language operations, state complexity and Jacobsthal’s function. *Int. J. Found. Comput. Sci.* **13**(1) (2002) 145–159
- [38] Pin, J.E.: Syntactic semigroups. In Rozenberg, G., Salomaa, A., eds.: *Handbook of Formal Languages*, vol. 1: Word, Language, Grammar. Springer (1997) 679–746
- [39] Rabin, M.O., Scott, D.: Finite automata and their decision problems. *IBM J. Res. Dev.* **3**(2) (April 1959) 114–125
- [40] Riordan, J.: The enumeration of trees by height and diameter. *IBM J. Res. Dev.* **4** (November 1960) 473–478
- [41] Riordan, J.: Forests of labeled trees. *J. Combin. Theory* **5**(1) (1968) 90 – 103
- [42] Saito, T.:  $\mathcal{J}$ -trivial subsemigroups of finite full transformation semigroups. *Semigroup Forum* **57** (1998) 60–68
- [43] Salomaa, A., Wood, D., Yu, S.: On the state complexity of reversals of regular languages. *Theoret. Comput. Sci.* **320**(23) (2004) 315 – 329
- [44] Schützenberger, M.: On finite monoids having only trivial subgroups. *Inform. and Control* **8** (1965) 190–194
- [45] Shor, P.W.: A new proof of Cayley’s formula for counting labeled trees. *J. Combin. Theory Ser. A* **71**(1) (1995) 154–158
- [46] Shyr, H.J.: *Free Monoids and Languages*. Third edn. Hon Min Book Co, Taiwan (2001)
- [47] Sierpiński, W.: Sur les suites infinies de fonctions définies dans les ensembles quelconques. *Fund. Math.* **24** (1935) 209–212



- [48] Simon, I.: Hierarchies of Events With Dot-Depth One. PhD thesis, Dept. of Applied Analysis & Computer Science, University of Waterloo, Waterloo, Ont., Canada (1972)
- [49] Simon, I.: Piecewise testable events. In: Proceedings of the 2nd GI Conference on Automata Theory and Formal Languages, London, UK, Springer-Verlag (1975) 214–222
- [50] Straubing, H.: A generalization of the Schützenberger product of finite monoids. *Theoret. Comput. Sci.* **13**(2) (1981) 137 – 150
- [51] Thérien, D.: Classification of finite monoids: the language approach. *Theoret. Comput. Sci.* **14**(2) (1981) 195 – 208
- [52] Yu, S.: Regular languages. In Rozenberg, G., Salomaa, A., eds.: *Handbook of Formal Languages*, vol. 1: Word, Language, Grammar. Springer (1997) 41–110
- [53] Yu, S.: State complexity of regular languages. *J. Autom. Lang. Comb.* **6** (2001) 221–234