THE FORMAL LANGUAGE THEORY COLUMN

BY

GIOVANNI PIGHIZZINI

Dipartimento di Informatica
Università degli Studi di Milano
20133 Milano, Italy
pighizzini@di.unimi.it
Abstract

Since the late nineties the scope of the International Conference of Descriptive Complexity of Formal Systems (DCFS) encompasses all aspects of descriptional complexity, both in theory and application. We first consider the historical development of the conference. Then we turn to some impressions from the 25 editions of the conference, which we particularly remember. In order to give a deeper inside in the field of descriptional complexity, we present some of its very basics from a general abstract perspective. Then we turn to some of the outstanding and dominating directions in the course of time. The results presented are not proved but we merely draw attention to the overall picture and some of the main ideas involved.

1 Introduction

Since the dawn of theoretical computer science the relative succinctness of different representations of (sets of) objects by formal systems have been a subject of
intensive research. An obvious choice to encode the objects is by strings over a finite alphabet. Then a set of objects is a set of strings, that is, a formal language. Formal languages can be described by several means, for example, by automata, grammars, rewriting systems, equation systems, etc. In general, such a descriptional system is a set of finite descriptors for languages. Core questions of descriptional complexity are “How succinctly (related to a size complexity measure) can a system represent a formal language in comparison with other systems?” and “What is the maximum trade-off when the representation is changed from one descriptional system to another, and can this maximum be achieved?” In the classification of automata, grammars, and related (formal) systems it turned out that the gain in economy of description heavily depends on the considered systems.

The approach to analyze the size of systems as opposed to the computational power seems to originate from Stearns [115] who studied the relative succinctness of regular languages represented by deterministic finite automata (DFAs) and deterministic pushdown automata. He showed the decidability of regularity for deterministic pushdown automata in a deep proof. The effective procedure revealed the following upper bound for the simulation. Given a deterministic pushdown automaton with \( n > 1 \) states and \( t > 1 \) stack symbols that accepts a regular language, then the number of states which is sufficient for an equivalent DFA is bounded by an expression of the order \( t^{n^3} \). Later this triple exponential upper bound has been improved by one level of exponentiation in [116]. In the levels of exponentiation it is tight, as proved in [95] by obtaining a double exponential lower bound. The precise bound is still an open problem. Probably the best-known result on descriptional complexity is the construction of a DFA that simulates a given nondeterministic finite automaton (NFA) [113]. By this so-called power-set construction, each state of the DFA is associated with a subset of NFA states. Moreover, the construction turned out to be optimal, in general. That is, the bound on the number of states necessary for the construction is tight in the sense that for an arbitrary \( n \) there is always some \( n \)-state NFA which cannot be simulated by any DFA with strictly less than \( 2^n \) states [79, 95, 97].

Let us turn to another cornerstone of descriptional complexity theory in the seminal paper by Meyer and Fischer [95]. In general, a known upper bound for the trade-off answers the question, how succinctly can a language be represented by a descriptor of one descriptional system compared with the representation by an equivalent descriptor of the other descriptional system? In [95] the sizes of finite automata and general context-free grammars for regular languages are compared. The comparison revealed a qualitatively new phenomenon. The gain in economy of description can be arbitrary, that is, there are no recursive functions serving as upper bounds for the trade-off, which is said to be non-recursive. Non-recursive trade-offs usually sprout at the wayside of the crossroads of (un)decidability, and
in many cases proving such trade-offs apparently requires ingenuity and careful constructions.

Nowadays, descriptional complexity has become a large and widespread area. On our tour on the field we first consider the historical development of the conference Descriptional Complexity of Formal Systems (DCFS). Then we turn to some impressions from the 25 editions of the conference, which we particularly remember. In order to give a deeper inside in the field of descriptional complexity, we present some of its very basics from a general abstract perspective. Our tour on the subjects covers some outstanding and dominating topics. It obviously lacks completeness and it reflects our personal view of what constitute some of the most interesting links to descriptional complexity theory. In truth there is much more to the field than can be summarized here and in the related papers [31, 39, 69, 70]. The results presented are not proved but we merely draw attention to the overall picture and some of the main ideas involved.

2 History of DCFS

In 1998, the history of DCFS started at the conference Mathematical Foundations of Computer Science (MFCS) in Brno. During a lunch break, Detlef Wotschke (1944–2019) suggested the organization of a workshop on descriptional complexity and related topics. It seems that there were two reasons for such a proposal.

Firstly, in 1997, within the organization International Federation of Information Processing (IFIP), a reestablishment of the Technical Committee TC1 Foundations of Computer Science took place, and within TC1 a Working Group WG 1.02 Descriptional Complexity was created. The chairman of this WG was Detlef Wotschke. It was natural to found a special workshop of the working group.

Secondly, the MFCS conference in Brno was accompanied by more than 10 workshops, some of them were organized as single events and some of them took place as a part of certain workshop series. Descriptional complexity was present in some of these workshops, but a special workshop on this topic was missing.

Detlef did not only come with the proposal of a workshop, he also had an idea for the place – Magdeburg (where DLT took place in 1995 and where Jürgen Dassow, the head of the group working in formal languages in Magdeburg, had a good position in the university). After some discussions, Jürgen accepted that his group will organize a workshop in Magdeburg in 1999.

In July 20–23, 1999, the workshop Descriptional Complexity of Automata, Grammars and Related Systems (DCAGRS) took place in Magdeburg. It was a

The conference series Descriptional Complexity of Formal Systems has two roots, the workshops Formal Descriptions and Software Reliability and Descriptional Complexity of Automata, Grammars and Related Systems. Here we reflect only the latter one.
terrible title, but the organizers wanted a title which describes very well the topic of the workshop. The event was successful with respect to the invited lectures (e.g. J. Gruska, Sh. Yu (1950–2012), J. Shallit) as well as to the number and quality of submissions as well as to the large number of participants.

We mention two facts where the first DCAGRS essentially differs from the later DCFS conferences. Firstly, the conference fee was only 70 euros, the average registration fee of some of the last normal DCFS conferences was 280 euros. Secondly, the program committee consisted of six person, the average of the last conferences was 24.

By the success of the first edition, it was necessary to look for a continuation. One week before the workshop in Magdeburg, there was the Workshop on Implementation of Automata in Potsdam, organized by Helmut Jürgensen (1942–2019). During the excursion of this event (a boat tour through the lakes around Potsdam) Detlef and Jürgen talked to Helmut concerning the next DCAGRS. Finally, Helmut accepted to organize it in London (Ontario) (not knowing that it will be organized in London four times almost like a biannual conference, and that he will organize it three times). The site London was chosen, because, from the beginning, there was the idea to organize the workshop alternately in Europe and North America. This idea has been followed in the sequel with only three exceptions as one can see from Figure 1.

In the following years, there was a steering committee which was looking for the persons and places of the next (two) DCFS editions. This task was not easy in some cases, but finally the committee was successful in all the years. In 2015 the international workshop DCFS became an international conference to underline
the grown importance and the history of the event. A list of all editions of DCFS is given in Figure 1.

<table>
<thead>
<tr>
<th>Time</th>
<th>Place</th>
<th>Organizing Institution / Chairman</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCAGRS 1999</td>
<td>July 20–23 Magdeburg, Germany</td>
<td>O.-v.-Guericke-Universität Magdeburg J. Dassow</td>
</tr>
<tr>
<td>DCAGRS 2000</td>
<td>July 27–29 London, Canada</td>
<td>The University of Western Ontario H. Jürgensen</td>
</tr>
<tr>
<td>DCAGRS 2001</td>
<td>July 20–22 Vienna, Austria</td>
<td>Technical University of Vienna R. Freund</td>
</tr>
<tr>
<td>DCFS 2002</td>
<td>August 21–25 London, Canada</td>
<td>The University of Western Ontario H. Jürgensen</td>
</tr>
<tr>
<td>DCFS 2003</td>
<td>July 12–14 Budapest, Hungary</td>
<td>Hungarian Academy of Sciences E. Csinhaj-Varjú</td>
</tr>
<tr>
<td>DCFS 2004</td>
<td>July 26–28 London, Canada</td>
<td>The University of Western Ontario L. Ilie</td>
</tr>
<tr>
<td>DCFS 2005</td>
<td>June 30 - July 2 Como, Italy</td>
<td>University of Milan G. Pighizzini</td>
</tr>
<tr>
<td>DCFS 2006</td>
<td>June 21–23 Las Cruces, USA</td>
<td>New Mexico State University H. Leung</td>
</tr>
<tr>
<td>DCFS 2007</td>
<td>July 20–22 Nový Smokovec, Slovakia</td>
<td>R.J. Saffarik University Košice V. Geffert</td>
</tr>
<tr>
<td>DCFS 2008</td>
<td>July 16–18 Charlottetown, Canada</td>
<td>University of Prince Edward Island C. Câmpeanu</td>
</tr>
<tr>
<td>DCFS 2009</td>
<td>July 6–9 Magdeburg, Germany</td>
<td>O-v-Guericke-Universität Magdeburg J. Dassow, B. Truthe</td>
</tr>
<tr>
<td>DCFS 2010</td>
<td>August 8–10 Saskatoon, Canada</td>
<td>University of Saskatchewan J. McQuillan</td>
</tr>
<tr>
<td>DCFS 2012</td>
<td>July 23–25 Braga, Portugal</td>
<td>Univ. of Porto and Univ. of Minho N. Moreira, R. Reis</td>
</tr>
<tr>
<td>DCFS 2013</td>
<td>July 22–25 London, Canada</td>
<td>The University of Western Ontario H. Jürgensen</td>
</tr>
<tr>
<td>DCFS 2014</td>
<td>August 5–8 Turku, Finland</td>
<td>University of Turku J. Karhoniak, A. Okhotin</td>
</tr>
<tr>
<td>DCFS 2015</td>
<td>July 25–27 Waterloo, Canada</td>
<td>University of Waterloo J. Shallit</td>
</tr>
<tr>
<td>DCFS 2016</td>
<td>July 5–8 Bucharest, Romania</td>
<td>University of Bucharest C. Câmpeanu</td>
</tr>
<tr>
<td>DCFS 2017</td>
<td>July 3–5 Milan, Italy</td>
<td>University of Milan G. Pighizzini</td>
</tr>
<tr>
<td>DCFS 2018</td>
<td>July 25–27 Halifax, Canada</td>
<td>Saint Mary’s University St. Konstantinides</td>
</tr>
<tr>
<td>DCFS 2019</td>
<td>July 17–19 Košice, Slovakia</td>
<td>Slovak Academy of Sciences G. Jirásková</td>
</tr>
<tr>
<td>DCFS 2020</td>
<td>Collected Papers</td>
<td>G. Jirásková, G. Pighizzini</td>
</tr>
<tr>
<td>DCFS 2021</td>
<td>Collected Papers</td>
<td>Y.-S.Han, S.-K. Ko</td>
</tr>
<tr>
<td>DCFS 2022</td>
<td>August, 29–31 Debrecen, Hungary</td>
<td>University of Debrecen Gy. Vaszil</td>
</tr>
<tr>
<td>DCFS 2023</td>
<td>July 4–6 Potsdam, Germany</td>
<td>University of Potsdam H. Bordihn</td>
</tr>
</tbody>
</table>

Figure 1: List of conferences.

We mention some important facts.
The initiative for DCFS came from the chairman of WG 1.02 of TC1 of IFIP, and all editions were organized by some institution and this IFIP working group together. Thus DCFS can be considered as the conference of WG 1.02. Since some years, the meetings of the Working Group take place as an evening session of DCFS.

If we compare the list of topics from 2007 (the oldest which can be found in the Web) and 2023 (the last conference) and those in between, then one notice that they are identical in appr. 75% of the items. This proves that there is a strong continuity and no following of short-lived modern directions.

In the years 2020 and 2021, DCFS conferences were planned in Vienna organized by R. Freund and in Seoul organized by Y.-S. Han and S.-K. Ko, respectively. Due to the crisis caused by the Corona virus, both conferences had to be canceled. However, there were invitations to the world for submitting papers such that proceedings could also be published in these years. Thanks go to the editors G. Jirásková/G. Pighizzini and Y.-S. Han/S.-K. Ko for their contribution to the survival of DCFS during the Corona time.

From the very beginning there was the idea that DCAGRS/DCFS should be organized with respect time and place in connection with some other conference such that e.g., only one crossing of the Atlantic Ocean is necessary to visit at least two conferences. As favorite accompanying conferences were considered Developments in Language Theory (DLT) and the International Conference Implementation and Application of Automata (CIAA, formerly Workshop on Implementation of Automata, WIA). Also this idea was realized for almost all editions (see Figure 2). Sometimes, the events were very near; for instance, in 2001, there was one day which was part of the DLT as well as of the DCFS program. Sometimes, the distance was large (in 2006, the distance between Las Cruces and Santa Barbara was 1600 km, but the Europeans had to cross the ocean only once; the four days between the two conferences could be used e.g., for a visit of the Grand Canyon almost in the middle between the towns).

The special event 50 Years of Automata Theory, that took place in 2000, was particularly remarkable. The list of speakers was very impressive. One could hear, meet and talk to all those persons which contributed by famous basic theorems as M. Rabin, D. Scott, and Sh. Greibach, introduced essential concepts as R. Mcaughton or wrote famous textbooks as J. Hopcroft and A. Salomaa (note that the mentioned names represent less than half of speakers). However, we do not know why automata theory became 50 years in 2000.

In 2015, one day before DCFS, the birthday of Janusz (John) Brzozowski (1935–2019) was celebrated in a one-day-conference.

Some remarks concerning proceedings. In the years 1999–2008, proceedings were published by the organizing institution. In the following two years, the pro-
ceedings appeared in the series *Electronic Proceedings in Theoretical Computer Science* as numbers 3 and 31, respectively. There were some attempts to publish in the LNCS series of Springer-Verlag, but only in 2011 we were successful. Starting with the thirteenth edition of DCFS, the Proceedings appeared as *Lecture Notes in Computer Science*.

Proceedings have mostly a page limit for the contributions, i.e., they do not contain often full versions. Therefore, from the very beginning, full versions of selected papers were published as special issues of some scientific journals. Thus,
one can find many full versions of a certain DCFS on a fixed place and not distributed over a lot of journals. For the first editions, the full versions appeared in *Journal of Automata, Languages, and Combinatorics* (JALC), a journal edited by the University of Magdeburg with persons in the editing staff, which also were involved in the program and organizing committees. Later the journals *Theoretical Computer Science* (TCS) from Elsevier B.V., *International Journal of Foundations of Computer Science* (IJFCS) from World Scientific Publishing Co., and *Information and Computation* (IC) from Elsevier B.V. were involved. The list in Figure 3 gives the journal in relation to the year of the conference.

The development of the number of accepted papers is shown in Figure 4. It is worth mentioning that the invited contributions are not included in the statistics. In most years there were additionally 4 invited presentations and papers. However, since in the early years the spirit of DCAGRS/DCFS was that of an intense workshop, at that times the number of invited speakers was higher, with a maximum of 8 speakers in 2004. Though from the very beginning all submitted papers were peer reviewed by at least three reviewers, respectively, the PCs had to work in the classical way without the support of a more or less professional conference managing system. Due to this fact, the information about the number of submitted and, thus, the number of rejected submissions is not available before 2011. The situation changed in 2011 when EasyChair came into play.

A pleasing fact is that the number of authors and their countries of affiliations has been at a good level from the beginning. This also shows that the interest in the topic has been maintained over the years and emphasizes once more that topics of descriptional complexity have a strong continuity and are not following short-lived directions. The development is shown in Figure 5.

Further information on the DCFS series (for instance programs, contents of the proceedings, special issues etc.) can be found on the web page

http://www.informatik.uni-giessen.de/dcfs

3 Impressions From 25 Editions of the Conference

The contents of this section consists of personal (not scientific) impressions of the first author (who did not attend all conferences such that his reflections are limited).

Mostly, the conference took place in universities or near to the universities in the towns. The exceptions were

- 2005 Como – in a theater,
Figure 4: Development of the number of papers. Invited contributions are not included in the statistics.
Figure 5: Development of the number of different authors and the number of different countries of all presentations. Invited contributions are included in the statistics.

- 2007 Nový Smokovec – in a hotel in the High Tatras,
- 2011 Limburg – in a hotel,
- 2012 Braga – in a museum.

If I should give the sites which impressed me most, then two places come to my remembering:

- Sala Bianca of Teatro Sociale in Como: It was a large room with wonderful baroque design which amazed me as I entered it for the first time as well as the rest of the workshop.
- Art in the University of Saskatoon: It seemed to me that the whole university was an exhibition of different arts and science. All buildings were full of sculptures, artistic installations and other works of arts, but also some terrariums. Moreover, already on the way from the hotel to the university, I saw a lot of sculptures, etc.

In connection with conferences, workshops etc., I visited a lot of places of interest. However, I remember especially the excursions of the DCFS conferences. There are two reasons for that: I have seen many sites which also contributed to my knowledge and were not only nice places to see, and I visited landscapes of extraordinary importance. Let me mention here:
The sites of UNESCO World Heritage
DCFS 2012 – visit of Guimarães, where Portugal was born,
DCFS 2017 – excursion to Bergamo (with the wall around the upper town),
DCFS 2018 – visit of the harbour of Lunenburg,
DCFS 2023 – excursions to the castles and gardens of Sanssouci,
DCFS 2010 – visit of Wanuskewin, an important place for the first nations in Canada (this is not really a UNESCO site but it belongs to the Canadian proposals)

DCFS 2006 – excursion to the desert White Sands National Park,
DCFS 2007 – hiking tours in the High Tatras, a UNESCO Biosphere Reserve: one tour only a short walk, one tour to the mountain hut Zamkovskeho Chata (duration appr. 2 hours), and a long tour of five hours to the Téryho Chata,
DCFS 2008 – excursion to Green Gables, a National Historic Site of Canada (it shows places related to the book Anne of Green Gables by L.M. Montgomery, which tells a story on a farmer girl in the 19th century and is popular in Japan, too).

Finally, some miscellaneous reminds:
In 2002, the reception took place as a barbecue in some park near London. As we reached the place, H. Jürgensen and some of his students started to encircle it with a red net used as a fence. Since there were enough place, their handling was surprising. The reason was that it is forbidden in Ontario to take alcoholic drinks at public places. However, if you have a fence and a special permission, it is allowed. Thus they ensured wine and beer for the barbecue reception.
In 2007, the long hiking tour through the High Tatras ended at the station for a funicular. Some participants took another way of return; they used scooters.
However, the way was very curvy and steep in some parts such that it was a little bit dangerous to take this way. Some of the participants arrived the base station without problems, some of them had a fall of one’s scooter and a painful night.

In 2010, there were a choice in the meal of the reception, lobster or something else. Because lobster is very typical for Prince Edward Island, almost all participants chose lobster. However, almost nobody had some experience in eating lobster. Therefore anybody got a bib to protect the clothes. Then everyone did his best and mostly it was done successfully.
4 Basic Concepts of Descriptive Complexity

In order to give a deeper inside in the field of descriptive complexity, we present some of its very basics from a general abstract perspective.

We denote the set of nonnegative integers by $\mathbb{N}$. Let $\Sigma^*$ denote the set of all words over a finite alphabet $\Sigma$. For the length of a word we write $|w|$. We use $\subseteq$ for inclusions and $\subset$ for strict inclusions. In general, the family of all languages accepted by a device of some type $X$ is denoted by $\mathcal{L}(X)$.

In order to be general, we first formalize the intuitive notion of a representation or description of a family of languages. A descriptional system is a collection of encodings of items where each item represents or describes a formal language. In the following, we call the items descriptors, and identify the encodings of some language representation with the representation itself. More precisely, a descriptional system $S$ is a set of finite descriptors such that each $D \in S$ describes a formal language $L(D)$. The family of languages represented (or described) by $S$ is $L(S) = \{ L(D) \mid D \in S \}$.

A complexity measure for a descriptional system $S$ is a total recursive mapping $c: S \rightarrow \mathbb{N}$. From the viewpoint that a descriptional system is a collection of encoding strings, the length of the strings is a natural measure for the size. We denote it by length.

For example, nondeterministic finite automata can be encoded over some fixed alphabet. The set of these encodings is a descriptional system $S$, and $L(S)$ is the family of regular languages.

Apart from length, examples for complexity measures for nondeterministic finite automata are the number of states and the number of transition.

Let $S_1$ and $S_2$ be descriptional systems with complexity measures $c_1$ and $c_2$, respectively. A total function $f: \mathbb{N} \rightarrow \mathbb{N}$, is said to be a lower bound for the increase in complexity when changing from a descriptor in $S_1$ to an equivalent descriptor in $S_2$, if for infinitely many $D_1 \in S_1$ with $L(D_1) \in L(S_2)$ there exists a minimal $D_2 \in S_2(L(D_1))$ such that $c_2(D_2) \geq f(c_1(D_1))$.

A total function $f: \mathbb{N} \rightarrow \mathbb{N}$ is an upper bound for the increase in complexity when changing from a descriptor in $S_1$ to an equivalent descriptor in $S_2$, if for all $D_1 \in S_1$ with $L(D_1) \in L(S_2)$, there exists a $D_2 \in S_2(L(D_1))$ such that $c_2(D_2) \leq f(c_1(D_1))$.

It may happen that the upper bound is not effectively computable. If there is no recursive upper bound, then the trade-off for changing from a description in $S_1$ to an equivalent description in $S_2$ is said to be non-recursive. Non-recursive trade-offs are independent of particular measures. That is, whenever the trade-off from one descriptional system to another is non-recursive, one can choose an arbitrarily large recursive function $f$ but the gain in economy of description eventually exceeds $f$ when changing from the former system to the latter. As an
example, we consider nondeterministic pushdown automata that are used to accept regular languages. Clearly, for any such automaton there exists an equivalent finite automaton. However, the trade-off for the conversion of the pushdown automaton into the finite automaton is non-recursive.

5 Outstanding Topics

5.1 Computational Completeness with Small Resources

In the 25th edition of DCFS, there were more than 50 papers on extensions of context-free grammars (as matrix and programmed grammars etc.), insertion-deletion systems, contextual grammars, systems of grammars and automata (as Lindenmayer systems, cooperating distributed grammar systems, parallel communicating grammar systems, etc.) The problem which is mostly discussed is the following: Let a device and a numerical parameter, which describes (partly) the size of the device, be given. Let \( L \) be the family of languages generated by such devices. Is there a constant \( c \) such that, for each \( L \in L \), there is a device \( D \) which computes \( L \) and the parameter of \( D \) is at most \( c \)? Moreover, if \( c \) exists, find the minimal one.

In the sixties and seventies, a lot of variants of context-free grammars were introduced, where the sequence of the applied rules is controlled by some mechanism. We mention very informally three mechanisms and refer to [19] for details. In a matrix grammar, sequences of rules called matrices are given, and the generation process consists of applications of matrices, i.e., rules in the order given by the matrices; in a programmed grammar, with each rule, sets of successor rules are associated; in a graph-controlled grammar, the rules are associated to nodes of a graph and the successor rule has to be taken from the successor nodes in the graph. If one allows appearance checking, i.e., there is a set \( F \) of distinguished rules, and rules of \( F \) can be overpassed if they cannot be applied, and erasing rules, all these mentioned grammars generate all recursively enumerable languages.

As numerical parameter we take the number of nonterminals.

The first result in this direction was given by Gh. Păun in [111]. He showed that each recursively enumerable language can be generated by a matrix grammar with at most six nonterminals. An improvement was only given in 2001 in [23] and [21], where it was proved that each recursively enumerable language can be generated by a programmed grammar or a graph controlled grammar with only three nonterminals. The best known bounds where given by H. Fernau, R. Freund, M. Oswald and K. Reinhardt at DCFS’05:
Theorem 1. (DCFS’05, [22])

i) For each recursively enumerable language $L$, there is a matrix grammar with at most three nonterminals which generates $L$.

ii) For each recursively enumerable language $L$, there is a programmed grammar $G$ with at most three nonterminals which generates $L$. Moreover, only two of the nonterminals are used in appearance checking mode.

iii) For each recursively enumerable language $L$, there is a graph controlled grammar with at most two nonterminals, both used in appearance checking mode, which generates $L$.

iv) The family of languages generated by graph controlled grammar with only one nonterminal used in the appearance checking mode is a proper subset of the family of all recursively enumerable languages.

We note that the results given in i) and iii) are optimal.

The operations of insertion and deletion of words are fundamental in formal language theory. They are motivated from linguistics (see contextual grammars) as well as – especially in the last 25 years – by the modeling of biological phenomena. Mostly, the insertions and deletion can only be done in a certain context, i.e., given a triple $(\alpha, w, \beta)$ of words, we can only insert $w$ in a word $x$ to obtain $y$ if $x = u \alpha \beta v$ and $y = u \alpha w \beta v$ (and analogous for deletions). In context-free insertion-deletion systems, the contexts $\alpha$ and $\beta$ are always empty. Then $w$ can be inserted at any place in $x$.

A context-free insertion-deletion system $G$ can be described as a 5-tuple $G = (V, T, I, D, A)$, where $V$ and $T$ are two alphabets with $T \subseteq V$, and $I \subseteq V^*$, $D \subseteq V^*$, and $A \subseteq V^*$ are three finite sets. The language generated by a context-free insertion-deletion system consists of all words from $T^*$ which can be obtained from $A$ by iterated applications of insertions of words from $I$ and deletions of words of $D$.

Surprisingly, M. Margenstern, Gh. Păun, J. Rogozhin, and S. Verlan proved that already context-free insertion-deletion systems, where the length of the inserted and deleted words are short, are very powerful:

Theorem 2. (DCFS’03, [88]) For each recursively enumerable language $L$, there is a context-free insertion-deletion system $G$ where all words in $I$ have a length at most three and all words in $D$ have a length at most two such that $G$ generates $L$.

Two years later, S. Verlan showed that this result is optimal:

Theorem 3. (DCFS’05, [117])

i) A context-free insertion-deletion system, where all words of $I$ and $D$ have a length at most two, generates a context-free language.

ii) A context-free insertion-deletion system, where the sets $I$ or $D$ contain only letters, generates a context-free language.
Parallel communicating grammar systems (for short PCGSs) were introduced by Gh. Păun and L. Santean (now L. Kari) in 1989 in [112]. We only give an informal description of PCGSs. A non-returning PCGS is specified as an \((n + 3)-\)tuple \(G = (N, K, T, G_1, G_2, \ldots, G_n)\), where \(V\) and \(T\) are alphabets, \(K = \{Q_1, Q_2, \ldots, Q_n\}\) is a set of \(n\) query symbols, and, for \(1 \leq i \leq n\), \(G_i = (N \cup K, T, P_i, S_i)\) is a context-free grammar with an axiom \(S_i \in N\). A configuration of \(G\) is an \(n\)-tuple of words over \(N \cup T \cup K\). We say that \((x_1, x_2, \ldots, x_n)\) derives in one step \((y_1, y_2, \ldots, y_n)\) if and only if

a) no \(x_i, 1 \leq i \leq n\), contains a query symbol, and \(x_i \Rightarrow_{G_i} y_i\) for \(1 \leq i \leq n\), or

b) if \(x_i = z_0 Q_{i_1} z_1 Q_{i_2} z_2 \ldots Q_{i_k} z_k\) with \(z_j \in (N \cup T)^*\) for \(0 \leq j \leq k\) and \(x_i\) contains no query symbol for \(1 \leq j \leq k\), then \(y_i = z_0 x_{i_1} z_1 x_{i_2} z_2 \ldots x_{i_k} z_k\) (i.e., query symbols are replaced by the corresponding sentential form); otherwise \(y_i = x_i\).

The generated language consists of all word \(x_1 \in T^*\) such that there is a configuration \((x_1, x_2, \ldots, x_n)\) which can be obtained from \((S_1, S_2, \ldots, S_n)\) by some derivation steps.

The generative power of non-returning PCGSs was open for more than 10 years. In 2000, N. Mandache proved that any recursively enumerable language can be generated by some non-returning PCGS. However, the proof allows no limitation of the number of grammars (see [87]). The first bound for the number of grammars, to generate all recursively enumerable languages was presented at DCFS’05 by Gy. Vasził, who show that eighth grammars are sufficient. An improvement was given at DCFS’09 by E. Csupai-Varjú and Gy. Vaszil. Hitherto, their bound is the best known one.

**Theorem 4.** (DCFS’09, [18]) For any recursively enumerable language \(L\), there is a non-returning PCGS with \(n\) grammars which generates \(L\).

### 5.2 State Complexity of Operations

One of the most studied topics at DCFS is operational state complexity. The topic was presented during the first DCAGRS by Sheng Yu in this invited talk State Complexity of Regular Languages [120]. Dozens of papers deal with various aspects of this field. Before we start our short tour through this topic we present its basic idea.

Let \(\circ\) be a fixed binary operation on languages from a family \(\mathcal{L}\) that is closed under the operation. Then the \(\circ\)-operation problem can be stated as follows:

- Given a language descriptor \(A\) of size \(m\) and a language descriptor \(B\) of size \(n\) such that \(L(A) \in \mathcal{L}\) and \(L(B) \in \mathcal{L}\).

- Which size is sufficient and necessary in the worst-case (in terms of \(n\) and \(m\)) for a language descriptor to describe the language \(L(A) \circ L(B)\)?
Obviously, this problem generalizes as well to \( k \)-ary language operations like, for example, complementation. In particular, if language descriptors are considered that are finite automata whose size is measured by their number of states, the notion operational state complexity is used.

**Operations on (non)deterministic finite automata**

First observations concerning basic operation problems of DFAs can be found in [90], where tight bounds for some operations are stated without proof. In [76] the tight bound of \( 2^n \) states for the DFA reversal was obtained in connection with Boolean automata. After the dawn the research direction on DFAs was revitalized in [121]. The systematic study of nondeterministic finite automata originates in [38].

In general, a method to obtain upper bounds for some \( \circ \)-operation problem is to provide an effective procedure that constructs a finite automaton accepting the result of the operation applied to some given finite automata. The number of states of the automaton constructed is an upper bound for the problem. To show that an upper bound is tight for all input automata to the procedure, a family of minimal automata must be given such that the resulting automata achieve that bound. These families are called witnesses. Naturally, a witness can also be given by a family of languages.

In [8] an automaton-independent approach, called quotient complexity, that is based on derivatives of languages is presented, which turned out to be a very useful technique for proving upper bounds for DFA operations (cf. [10, 11, 12]).

To give an impression of basic types of results, we provide the bounds for some basic operations on DFAs and NFAs accepting infinite general and unary regular languages in Table 1.

**Subregular languages**

Table 1 also reflects the distinctions between general and unary regular languages that have been made from the early beginnings. At first glance the differences between general and unary languages are not that big for NFAs while it can be exponential for DFAs. However, even if nondeterminism is available, the limitation to unary languages can have a big impact to the operational state complexity.

It turned out, that for many state complexity issues of unary languages Landau’s function

\[
F(n) = \max\{ \text{lcm}(x_1, \ldots, x_k) \mid x_1, \ldots, x_k \geq 1 \text{ and } x_1 + \cdots + x_k = n \}
\]

which gives the maximal order of the cyclic subgroups of the symmetric group on \( n \) elements, plays a crucial role. Here, \( \text{lcm} \) denotes the least common multiple.
Table 1: [31] NFA and DFA state complexities for operations on infinite languages, where \( t \) is the number of accepting states of the “left” automaton, \( \setminus \) denotes the left and \( / \) the right quotient by an arbitrary language. The tight lower bounds for union, intersection, and concatenation of unary DFAs require \( m \) and \( n \) to be relatively prime.

Since \( F \) depends on the irregular distribution of the prime numbers, we cannot expect to express \( F(n) \) as a simple function of \( n \). In [73][74] the asymptotic growth rate

\[
\lim_{n \to \infty} (\ln F(n) / \sqrt{n \cdot \ln n}) = 1
\]

was determined, which implies the (for our purposes sufficient) rough estimate

\[
F(n) \in 2^{\Theta(\sqrt{n \cdot \log n})}.
\]

The connection with the complementation operation on unary languages represented by NFAs becomes evident from Table 1. This complementation is closely related to the unary NFA by DFA simulation, which causes also a state blow-up of order \( F(n) \in 2^{\Theta(\sqrt{n \cdot \log n})} \). The proofs rely on a a normal form for unary NFAs introduced in [15]. It reads as follows.

Each \( n \)-state NFA over a unary alphabet can effectively be converted into an equivalent \( O(n^2) \)-state NFA consisting of an initial deterministic tail and some disjoint deterministic loops, where the automaton makes only a single nondeterministic decision after passing through the initial tail, which chooses one of the loops.

Apart from unary and finite languages, many other subregular language families have been considered from the viewpoint of operational state complexity. The systematic investigation of the descriptional complexity of such families origi-
nates in [5], where the state costs of determinizations are considered. The number of papers at DCFS dealing with this topic is quite huge and cannot be covered here. A comprehensive survey with valuable and detailed references is [24].

Universal witnesses

We mentioned above that the tightness of upper bounds is often shown by providing suitable witness languages. Interestingly, in [9] a witness over a ternary alphabet is obtained that shows the tightness of the DFA upper bounds for the operations union, intersection, concatenation, Kleene star, and reversal simultaneously. Therefore, it is called a universal witness. The universal witness is not always optimal with respect to the underlying alphabet size. However, from the state complexity view it can be seen as the most complex regular language. In particular, it can be used for even more. It is a witness for the maximal bounds on the number of atoms, the quotient complexity of atoms, the size of the syntactic semigroup, and about two dozen combined operations, where only a few require slightly modified versions of the universal witness. Further applications can be found in [13, 14].

Magic numbers

In connection with the well-known subset construction, a fundamental question was raised in [45]: Does there always exist a minimal $n$-state NFA whose equivalent minimal DFA has $\alpha$ states, for all $n$ and $\alpha$ satisfying $n \leq \alpha \leq 2^n$. A number $\alpha$ not satisfying this condition is called a magic number (for $n$).

It was shown in [51] that no magic numbers exist for general regular languages over a ternary alphabet. For NFAs over a two-letter alphabet it was shown that $\alpha = 2^n - 2^k$ or $2^n - 2^k - 1$, for $0 \leq k \leq n/2 - 2$ [45], and $\alpha = 2^n - k$, for $5 \leq k \leq 2n - 2$ and some coprimality condition for $k$ [46], are non-magic. In [53] it was proven that the integer $\alpha$ is non-magic, if $n \leq \alpha \leq 2 \sqrt[n]{n}$. Further non-magic numbers for a two-letter input alphabet were identified in [25] and [91].

Magic numbers for unary NFAs were studied in [26] by revising the Chrobak normal-form for NFAs. In the same paper also a brief historical summary of the magic number problem can be found. More general, magic numbers for several subregular language families were investigated in [37]. Further results on the magic number problem (in particular in relation to the operation problem on regular languages) can be found, for example, in [53, 50].
Further important sub-topics

As implied by the definition, so far, here we deal with the language operation problem in terms of worst-case complexity. However, the magic number problem can be seen as generalization. As opposed to the worst case, the range of state complexities that may result from an operation is considered. So, it is natural to look at the average case as well. In his invited talk *Size matters, but let’s have it on average* at DCFS 2023, Rogério Reis considered various facets of this exciting field (see also [7, 98]).

Turning to another sub-topic, we recall that two words over a common alphabet are said to be *Parikh-equivalent* if and only if they are equal up to a permutation of their symbols or, equivalently, for each letter the number of its occurrences in the two words is the same. This notion extends in a natural way to languages, where two languages are Parikh-equivalent when for each word in the one language there is a Parikh-equivalent word in the other and vice versa. Inspired by the famous result of Parikh that each context-free language is Parikh-equivalent to some regular language [105], Parikh-equivalence has been connected to descriptive complexity issues. For example, in [75] the operational state complexity has been considered under parikh equivalence. That is, the resulting finite automaton must accept a language that is Parikh-equivalent to the precise language only.

Finally, the operation problems have been investigated not only for the devices DFA and NFA. A bunch of further devices have been considered. We mention only a few of them exemplarily. In her invited talk *Self-Verifying Finite Automata and Descriptional Complexity* at DCFS 2016, Galina Jirásková presented various aspects of descriptional complexity, including operation problems, on self-verifying finite automata [54]. See also [47] in this respect. The operation problems for two-way DFA were investigated in [56]. Alternating and Boolean automata are the devices considered, for example, in [43, 44, 52, 55, 76, 77]. The state complexity of operations on unambiguous finite automata and their languages is the main topic in [48].

5.3 Computational Models and Descriptional Complexity

A lot of work related to computation models, their descriptional complexity and other related properties has been done. More than 150 papers presented at DCFS investigate some kind of machines. It is impossible to briefly summarize and present in a complete way all the results obtained in this context. We just give some relevant examples and pointers to the literature.

While studying a computational model, the first question concerns its computational power. In the case of devices recognizing languages, this leads to investigate the class of accepted languages. The second natural question concerns the
succinctness. In particular, when a computational model can be simulated by another one, it is quite natural to investigate the cost of such a simulation in terms of the size of the descriptions. In the Introduction we already mentioned the classical example that can used to introduce descriptional complexity, namely the simulation of NFAs automata by DFAs, given by the subset construction: each one-way nondeterministic automaton with $n$ states can be simulated by an equivalent deterministic finite automaton with $2^n$ states. Furthermore, it is well-known that in the worst case such a cost cannot be reduced [95].

During the DCFS conferences a lot of results have been presented concerning the costs of the relative succinctness of computational models.

**Finite automata**

We just mentioned the exponential cost of the simulation of NFAs by DFAs. In a paper presented at DCAGRS 1999, M. Kappes proved this cost cannot be reduced even when simulating deterministic finite automata with multiple initial states, i.e., if the only nondeterministic choice can be taken at the beginning of the computation, to choose the initial state in a given set [59]. This research was refined in [42] by giving an exact bound that keeps into account the cardinality of the set of possible initial states, besides the cardinality of the set of states of the simulated automaton.

Even for many non-trivial subclasses of regular languages (e.g., star-free languages, strictly locally testable languages) the cost of the elimination on nondeterminism remains exponential [5].

In the above mentioned results, the focus is on one-way finite automata, i.e., automata that scans the input tape from left to right. It is well-known that the computational power does not increase if the head can be moved in both directions, so obtaining two-way finite automata. A long-standing open problem related to these devices is the cost of the elimination of nondeterminism using two-way motion. This problem was formulated in 1978 by Sakoda and Sipser, who asked the costs, in terms of states, of the conversions of one-way and two-way NFAs into equivalent two-way DFAs, and it is still open [114]. For both questions the best-known upper bounds are exponential, while the lower bounds are polynomial. Several contributions related to this problem and, more in general, to two-way finite automata have been presented in DCFS conferences. We point out just few of them.

In his invited lecture at DCFS 2012, Ch. Kapoutsis presented Minicomplexity [58], a complexity theory for two-way automata which brings together several seemingly detached concepts and results.

V. Geffert and L. Isonová presented a translation of the Sakoda and Sipser question on two-way automata, to an analogous question on pebble automata [27].
The maximum length of the shortest string accepted by an $n$-state two-way finite automaton is known to be exponential in $n$. However, its exact value is not yet known. Recent contributions towards the solution of this problem have been presented at DCFS 2020 and 2023 [64, 89].

Extension of two-way automata, with restricted rewriting capabilities, have been also considered and they will be mentioned later.

**Pushdown automata**

As we just discussed, a lot of contributions related to regular languages and finite automata have been given. To represent a regular language, we could use a device from a more powerful class of machines. For instance, we could use a pushdown automaton. So the question of comparing the relative succinctness of finite automata and equivalent pushdown automata arises. This question was solved longtime ago by Meyer and Fischer by proving nonrecursive trade-offs [95].

One could ask what happens if pushdown automata in some restricted form are considered. One possible restriction is to require that the height of the pushdown store is bounded by a constant. This leads to the definition of **constant height pushdown automata**, introduced and firstly studied by Z. Bednárová, V. Geffert, C. Mereghetti, and B. Palano [29, 3]. Under this restriction, the trade-off to finite automata is recursive. In particular, the size cost of the conversion of nondeterministic constant height pushdown automata into equivalent one-way deterministic finite automata is double exponential. So, constant height pushdown automata are very interesting for their succinctness. This line of research has been recently deepened. It is well-known that it cannot be decided whether the language accepted by a pushdown automaton is regular [95]. Notice that there exists pushdown automata that accept regular languages using a non-constant amount of pushdown store. This leads to the different question of deciding for a pushdown automaton whether there exists a constant $h$ such that each string in the accepted language has an accepting computation using height at most $h$. In [110] it has been proved that also this property is undecidable. It remains undecidable when the pushdown alphabet is unary, i.e., when the machine is a one-counter automaton [109].

More in general, computations of pushdown automata have been analyzed in [6], introducing and studying **pushdown information** (roughly the properties of strings written on the pushdown store during computations of stateless pushdown automata), and in [28, 85], where the descriptional complexity of the **pushdown store language**, i.e., the set of strings that appears of the pushdown during an accepting computation, is studied. We point out that this language is regular.

A. Malcher considered **finite-turn pushdown automata**, namely pushdown au-
tomata that can switch from push to pop operations a number of times bounded by a fixed constant $k$, proving a series of interesting non-recursive trade-offs (e.g., from $k$-turns to $k+1$ turns, for each $k \geq 1$) [34].

While in the case of finite automata, having a two-way input tape does not increase the computational power, in the case of pushdown automata the situation is different. In fact, two-way pushdown automata can recognize even non-context-free languages. At the moment, it is still unknown if these devices are able to recognize all context-sensitive languages. In [86] the authors started an investigation on these devices in the case of a constant number of head reversals.

Questions related to input driven pushdown automata (also known as nested word or visibly pushdown automata) have been investigated in [106, 100, 36].

In [40] the authors consider one-time nondeterministic finite and pushdown automata. In these devices, whenever a guess is performed, it remains fixed for the rest of the computation. In the case of finite automata, the state increase to equivalent deterministic devices is bounded by an exponential function, while in the case of pushdown automata nonrecursive size trade-offs have been proved.

**Turing machines and their variants, Cellular automata**

The model of restarting automata has been the subject of many papers presented at DCFS and in other related conferences. This is a formal model for the analysis by reduction, which is used in linguistic to analyze sentences of natural languages. Roughly, this technique consists in a stepwise simplification of a given sentence in such a way that the syntactical correctness or incorrectness of the sentence is not affected. Such a process can be modeled by machines having a flexible tape where, at some point, such a simplification is performed and then the computation is restarted.

In his invited lecture On Restarting Automata with Window Size One at DCFS 2011, F. Otto presented a general overview of the most important variants of restarting automata, together with new results on restarting automata with window size one [101]. Several other contributions in this area have been presented in DCFS conferences. Among them, we address the reader to [41, 72, 102, 103].

Turing machines with restricted rewriting capabilities have been considered in several papers. At DCFS 2005, B. Durak presented worm automata. These devices are two-way finite automata equipped with a write-once track. In spite of this possibility, they still recognize only regular languages [20].

More recently, several results on limited automata have been presented. These devices are single-tape Turing machines in which the content of each tape cell can be rewritten only in the first $d$ visits, for a fixed integer $d \geq 0$. In case $d \leq 1$
these devices accept only regular languages, while for each fixed \( d > 1 \), they characterize the class of context-free languages. The conversion from nondeterministic 1-limited automata into equivalent one-way deterministic finite automata costs, in the worst case, double exponential in size. So these devices can be extremely succinct \[108\]. Other results on the descriptional complexity of limited automata have presented in \[71, 34\]. A survey on limited automata has been given in the invited lecture \textit{Limited Automata: Properties, Complexity and Variants} by G. Pighizzini at DCFS 2019 \[107\].

In his invited lecture \textit{The Descriptive Power of Sublogarithmic Resource Bounded Turing Machines} at DCFS 2007, C. Mereghetti presented a complete picture of lower bounds on space and input head reversals for deterministic, nondeterministic, and alternating Turing machines accepting nonregular languages \[92\].

In \[66\], M. Kutrib investigated multitape Turing machines having a restricted number of nondeterministic steps, proving the existence of a nondeterministic language hierarchy between real time and linear time.

Among other computational models, it is also suitable to mention \textit{cellular automata}. They have been the subject of invited lectures \textit{Cellular Automata and Descriptive Complexity} by A. Malcher at DCFS 2006 \[83\] and \textit{Linear Algebra Based Bounds for One-Dimensional Cellular Automat} by J. Kari at DCFS 2011 \[60\]. The descriptive complexity and the properties of several variants of cellular automata (e.g., one-way and two-way) have been the subject of various talks (e.g \[80, 81, 82, 68\]).

\textbf{Non classical computation modes: probabilistic and quantum}

Besides classical modes of computations, mainly based on determinism, nondeterminism, and alternation, several contributions of other modes have been presented.

At DCFS 2009, Ch. Baier gave the invited talk \textit{Probabilistic Automata over Infinite Words: Expressiveness, Efficiency, and Decidability} \[2\]. Probabilistic models have been also considered in \[96, 49\].

\textit{Quantum automata and quantum computations} have been the subject of invited talks \textit{Descriptional complexity issues in quantum computing} and \textit{Succinctness in quantum information processing} by J. Gruska at DCAGRS 1999 and DCFS 2003 \[32, 33\], \textit{Some formal tools for analyzing quantum automata} by A. Bertoni at DCFS 2005 \[4\], and \textit{Recent Developments in Quantum Algorithms and Complexity} A. Ambainis at DCFS 2014 \[1\]. Several contributions to this area have been presented \[93, 118, 119\]. In \[94\] the authors compare the succinctness of deterministic, nondeterministic, probabilistic and quantum finite automata.
Non-recursive trade-offs

A general survey on non-recursive trade-offs, with a unifying approach to the proof of them, has been given by M. Kutrib in his invited talk *The phenomenon of non-recursive trade-offs* at DCFS 2004 [67]. Further developments on this phenomenon have been obtained in [30].

In [57], Ch. Kapoutsis presented non-recursive trade-offs for multi-head two-way finite automata and multi-counter automata. We already mentioned the paper by A. Malcher with non-recursive trade-offs for related to finite-turn pushdown automata [84]. In many other papers (e.g. related to pushdown automata [86, 40, 109]) results presenting non-recursive trade-offs have been given.

Ambiguity and measures of nondeterminism

In this invited talk *Descriptional complexity of nfa of different ambiguity* at DCFS 2004, H. Leung presented relationships between descriptional complexity and ambiguity degree for nondeterministic finite automata [78]. Further results on this topic are given in [65]. The case of Büchi automata was considered in [99].

Unambiguity in automata theory was the title of the invited lecture given by Th. Colcombet at DCFS 2015 [17]: the concept of unambiguity, seen as a generalization of determinism, has been explored not only in automata on finite words, but in some extensions of them as, e.g., automata on infinite trees and tropical automata.

In [104] various measure of nondeterminism for finite automata have been investigated and compared. This idea was further explored in some subsequent papers. Among them we mention [61, 63]. More recently, also measures for alternating automata have been introduced and studied [62, 35].

References


