Open problems about regular languages, 35 years later

Dedicated to Janusz A. Brzozowski for his 80th birthday

Jean-Éric Pin *
IRIF, CNRS and Université Paris-Diderot,
Case 7014, 75205 Paris Cedex 13, France.
Jean-Eric.Pin@irif.fr

October 2016

Abstract

In December 1979, Janusz A. Brzozowski presented a selection of six open problems about regular languages and mentioned two other problems in the conclusion of his article. These problems have been the source of some of the greatest breakthroughs in automata theory over the past 35 years. This survey article summarizes the state of the art on these questions and the hopes for the next 35 years.

Thirty-five years ago, at a conference in Santa Barbara in December 1979, Janusz A. Brzozowski [8] presented a selection of six open problems about regular languages and mentioned two other topics in the conclusion of his article. These six open problems were, in order, *star height*, *restricted star height*, *group complexity*, *star removal*, *regularity of non-counting classes* and *optimality of prefix codes*. The two other topics were the *limitedness problem* and the *dot-depth hierarchy*.

These problems proved to be very influential in the development of automata theory and were the source of critical breakthroughs. The aim of this paper is to survey these results, to describe their impact on current research and to outline some hopes for the next thirty-five years. Due to the lack of space, the dot-depth hierarchy is treated in a separate article [61].

1 Terminology, notation and background

This goal of this section is to fix notation and terminology. We define in particular the notions of syntactic monoid, class of languages, variety, profinite word, profinite identity, semiring and weighted automaton.

In the sequel, A denotes a finite alphabet and 1 denotes the empty word. A semigroup S *divides* a semigroup T if S is a quotient of a subsemigroup of T.

*The author was funded from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 670624).

1.1 Quotients and syntactic monoid

Given a language L of A^* and two words x and y, the *left quotient* of L by x is the language

$$x^{-1}L = \{u \in A^* \mid xu \in L\}.$$

Similarly, the *right quotient* of L by y is the language

$$Ly^{-1} = \{u \in A^* \mid uy \in L\}.$$

Finally, we set

$$x^{-1}Ly^{-1} = \{u \in A^* \mid xuy \in L\}.$$

Observe that $(x^{-1}L)y^{-1} = x^{-1}Ly^{-1} = x^{-1}(Ly^{-1})$.

The *syntactic congruence* of a language L of A^* is the equivalence relation $\sim_L$ defined by $u \sim_L v$ if and only if, for every $x, y \in A^*$,

$$xuy \in L \iff xvy \in L,$$

or, equivalently, if

$$u \in x^{-1}Ly^{-1} \iff v \in x^{-1}Ly^{-1}.$$

The *syntactic monoid* of L is the quotient $M(L)$ of A^* by $\sim_L$ and the natural morphism $\eta : A^* \to A^*/{\sim_L}$ is called the *syntactic morphism* of L.

For instance, the syntactic monoid of the language $(ab)^*$ is the monoid $M = \{1, a, b, ab, ba, 0\}$ presented by the relations $a^2 = b^2 = 0$, $aba = a$, $bab = b$ and $0a = 0b = 0 = a0 = b0$.

1.2 Classes of languages

Many properties of languages, such as being regular, finite, commutative, star-free, etc., are defined without any explicit reference to an alphabet. However, these properties do not define a set of languages, unless the alphabet is specified. The notion of class of languages is a convenient way to avoid this problem.

A *class of languages* $\mathcal{C}$ associates with each finite alphabet A a set $\mathcal{C}(A)$ of regular languages[1] of A^*. A class $\mathcal{C}$ of languages is said to be closed under some operation, such as union, intersection, complement, quotients, product, star, etc., if, for each alphabet A, the set of languages $\mathcal{C}(A)$ is closed under this operation. Similarly, $\mathcal{C}$ is said to be *closed under Boolean operations* if, for each A, $\mathcal{C}(A)$ is a Boolean algebra of languages.

A monoid morphism $\varphi : A^* \to B^*$ is said to be

(1) *length-preserving* if $|\varphi(u)| = |u|$ for all $u \in A^*$, or equivalently, if $\varphi(A) \subseteq B$;

(2) *length-increasing* if $|\varphi(u)| \geqslant |u|$ for all $u \in A^*$, or equivalently, if $\varphi(A) \subseteq B^+$;

(3) *length-decreasing* if $|\varphi(u)| \leqslant |u|$ for all $u \in A^*$, or, equivalently, if $\varphi(A) \subseteq B \cup \{1\}$.

A class of languages $\mathcal{C}$ is *closed under inverses of morphisms* if for each morphism $\varphi : A^* \to B^*$, the condition $L \in \mathcal{C}(B)$ implies $\varphi^{-1}(L) \in \mathcal{C}(A)$. Closure under inverses of length-increasing, length-decreasing or length-preserving morphisms is defined in the same way.

[1] This definition can be extended to nonregular languages, but we are only interested in regular languages in this paper.

1.3 Varieties of languages and varieties of finite monoids

Varieties constitute important examples of classes of languages. A *variety of languages* is a class of languages closed under Boolean operations, left and right quotients and inverses of morphisms. Star-free languages, which will be defined in Section 2, form the most emblematic example of a variety of languages.

Closure under inverses of morphisms can be relaxed by requiring only closure under inverses of length-preserving, length-increasing or length-decreasing morphisms, leading to the notions of *lp-variety*, *li-variety* and *ld-variety*, respectively. These notions were introduced independently by Esik [27] and Straubing [85].

A *variety of finite monoids* is a class of finite monoids closed under taking submonoids, quotients and finite products. We refer the reader to the books [1, 26, 63] for more details on varieties of monoids.

Eilenberg's variety theorem [26] states that there is a bijective correspondence between varieties of languages and varieties of finite monoids.

Theorem 1.1 *Let* $\mathbf{V}$ *be a variety of finite monoids. For each alphabet* A*, let* $\mathcal{V}(A)$ *be the set of all languages of* A^* *whose syntactic monoid is in* $\mathbf{V}$*. Then* $\mathcal{V}$ *is a variety of languages. Moreover, the correspondence* $\mathbf{V} \to \mathcal{V}$ *is a bijection between varieties of finite monoids and varieties of languages.*

Let us just mention for the record that a generalization of Eilenberg's variety theorem also holds for *lp*-varieties, *li*-varieties and *ld*-varieties. See [27, 10, 49, 65, 66, 85] for more details.

1.4 Profinite words

A finite monoid M *separates* two words u and v of A^* if there is a monoid morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$. We set

$$r(u, v) = \min\{\mathrm{Card}(M) \mid M \text{ is a finite monoid that separates } u \text{ and } v\}$$

and $d(u, v) = 2^{-r(u,v)}$, with the usual conventions $\min \emptyset = +\infty$ and $2^{-\infty} = 0$. Then d is a *metric* on A^* and the completion of A^* for this metric is denoted by $\widehat{A^*}$. The (concatenation) product on A^* can be extended by continuity to $\widehat{A^*}$, making $\widehat{A^*}$ a compact topological monoid, called the *free profinite monoid*. Its elements are called *profinite words*.

This abstract definition does not make it easy to really understand what a profinite word is. Actually, although $\widehat{A^*}$ is known to be uncountable if A is nonempty, it is difficult to exhibit "concrete" examples of profinite words, other than words of A^*. One such example can be obtained as a consequence of a standard result of semigroup theory:

> *Given an element* s *of a compact semigroup* S*, the closed subsemigroup of* S *generated by* s *contains a unique idempotent, usually denoted by* s^ω*.*

In particular, if u is a profinite word, then u^ω is also a profinite word. For instance, if $A = \{x, y\}$, then $(xy)^\omega$ and $((xy)^\omega yx(yxy)^\omega)^\omega$ are examples of profinite words.

1.5 Profinite identities

Let M be a finite monoid. Let us equip M with the discrete metric d, defined by $d(x,x) = 0$ and $d(x,y) = 1$ if $x \neq y$. Then every morphism $\varphi : A^* \to M$ is uniformly continuous since $d(x,y) < 2^{-|M|}$ implies $\varphi(x) = \varphi(y)$. Thus φ admits a unique continuous extension $\widehat{\varphi} : \widehat{A^*} \to M$.

Let u, v be two profinite words on some alphabet B. We say that a finite monoid M *satisfies the profinite identity* $u = v$ if the equality $\widehat{\varphi}(u) = \widehat{\varphi}(v)$ holds for all morphisms $\varphi : B^* \to M$. For instance, a monoid is commutative if and only if it satisfies the identity $xy = yx$. It is *aperiodic* if and only if it satisfies the identity $x^\omega x = x^\omega$.

Reiterman's theorem [72] states that a class of finite monoids is a variety if and only if it can be defined by a set of profinite identities. Since varieties of languages are in bijection with varieties of finite monoids, one can also define varieties of languages by profinite identities. This was made precise by Gehrke, Grigorieff and the author [28] as follows.

Let L be a regular language of A^* and let $\eta : A^* \to M$ be its syntactic morphism. Then η admits a unique continuous extension $\widehat{\eta} : \widehat{A^*} \to M$. In the same way, every monoid morphism from B^* to A^* admits a unique continuous extension $\widehat{\varphi} : \widehat{B^*} \to \widehat{A^*}$. Given two profinite words u and v on the alphabet B, we say that L *satisfies the profinite identity* $u = v$ if the equality $\widehat{\eta}(\widehat{\varphi}(u)) = \widehat{\eta}(\widehat{\varphi}(v))$ holds for all morphisms $\varphi : B^* \to A^*$. Reiterman's theorem can now be transposed to varieties of languages:

> *A class of languages is a variety of languages if and only if it can be defined by a set of profinite identities.*

A similar characterization was also proved [28] for *lp*-varieties, *li*-varieties and *ld*-varieties. We just state this result for *ld*-varieties, the other cases being similar. Let us say that a regular language *satisfies the profinite ld-identity* $u = v$ if the equality $\widehat{\eta}(\widehat{\varphi}(u)) = \widehat{\eta}(\widehat{\varphi}(v))$ holds for all length-decreasing morphisms $\varphi : B^* \to A^*$. Then one can state

> *A class of languages is a ld-variety of languages if and only if it can be defined by a set of profinite ld-identities.*

We refer the reader to [2] for a detailed study of profinite identities defining varieties of finite monoids and to the survey [65] for profinite equations on languages.

1.6 Semirings

A *semiring* is a set K equipped with two binary operations, written additively and multiplicatively, and two elements 0 and 1, satisfying the following conditions:

(1) K is a commutative monoid for the addition with identity 0,

(2) K is a monoid for the multiplication with identity 1,

(3) Multiplication is distributive over addition: for all $s, t_1, t_2 \in K$, $s(t_1 + t_2) = st_1 + st_2$ and $(t_1 + t_2)s = t_1 s + t_2 s$,

(4) for all $s \in K$, $0s = s0 = 0$.

Examples of semirings include $(\mathbb{N}, +, \times)$, $(\mathbb{Z}, +, \times)$ and the *tropical semirings* $(\mathbb{N} \cup \{+\infty\}, \min, +)$ and $(\mathbb{N} \cup \{-\infty\}, \max, +)$. The set of languages over A^* also forms a semiring with union as addition and concatenation product as multiplication. Consequently, we adopt the algebraic notation which consists of writing $+$ for union, 0 for the empty language and u for the language $\{u\}$, when u is a word. Thus, for instance, $1 + ab + baa + bb$ denotes the language $\{1, ab, baa, bb\}$.

Let K be a semiring. We let $K\langle\langle A\rangle\rangle$ (respectively $K[[A]]$) denote the semiring of *formal power series* in noncommutative (respectively commutative) variables in A with coefficients in K. We let also $K\langle A\rangle$ (respectively $K[A]$) denote the semiring of *polynomials* in noncommutative (respectively commutative) variables in A with coefficients in K.

1.7 Weighted automata

Let K be a semiring. A *K-weighted automaton* (or *K-transducer*) is a quintuple $\mathcal{A} = (Q, A, E, I, F)$, where Q (resp., I, F) is the *set of states* (resp., *initial* and *final* states) and A is the *input alphabet*. The set of *transitions* E is a subset of $Q \times A \times K \times Q$. A transition (q, a, x, q') is also written as $q \xrightarrow{a|x} q'$. A path

$$q_0 \xrightarrow{a_1|x_1} q_1 \xrightarrow{a_2|x_2} q_2 \quad \cdots \quad q_{n-1} \xrightarrow{a_n|x_n} q_n$$

is successful if $q_0 \in I$ and $q_n \in F$. The *output* of this path is the product $x_1 x_2 \cdots x_n$.

The function $\tau : A^* \to K$ realized by $\mathcal{A}$ is defined as follows. Given a word $u \in A^*$, $\tau(u)$ is the sum of the outputs of all successful paths of label u. If there is no successful path of label u the output is 0. Note that addition and product refer to the operations of the semiring K. Thus, if K is the tropical semiring $(\mathbb{N} \cup \{+\infty\}, \min, +)$, the output of a path is the *sum* $x_1 + x_2 + \cdots + x_n$ and $\tau(u)$ is the *minimum* of the outputs of all successful paths of label u.

Automata with outputs in the tropical semiring are sometimes called *automata with distance* [32, 34, 35, 36] or *distance automata* [44, 43, 45].

2 Star height

Extended regular expressions on the alphabet A are defined recursively as follows:

(1) 0, 1 and a, for each $a \in A$, are regular expressions.

(2) If E and F are extended regular expressions, then $(E + F)$, (EF) and $(E)^c$ and $(E)^*$ are regular expressions.

The *value* of an extended regular expression E is the language of A^* obtained from E by interpreting $(E)^c$ as the complement of E and the other operators as union, concatenation and star.

The *star height* of an extended regular expression is the maximum nested depth of stars in the expression. For instance

$$\Bigl(\bigl(a(ba)^*b^c\bigr)^* + \bigl(b(aa)^*b + a\bigr)^* + \bigl(b(a+b)^*bb\bigr)^c\Bigr)^*$$

is an extended regular expression of star height 3. More formally, the *star height* $h(E)$ of an extended regular expression E is defined recursively by

(1) $h(0) = 0$, $h(1) = 0$ and, for each $a \in A$, $h(a) = 0$.

(2) If E and F are regular expressions, then $h((E)^c) = h(E)$, $h((E+F)) = h((EF)) = \max(h(E), h(F))$ and $h((E)^*) = h(E) + 1$.

The *star height* $h(L)$ of a regular language L is the minimum of the star heights of the extended regular expressions representing L. In other words

$$h(L) = \min\{h(E) \mid E \text{ is an extended regular expression of value } L\}.$$

Note that allowing intersection would not change the star height of a language since $K \cap L = (K^c + L^c)^c$. The *star height problem* is the following question:

Problem 1 *Is there an algorithm which computes the star height of a given regular language?*

A language of star height 0 is said to be *star-free*. For instance, the language $A^* = \emptyset^c$ is star-free and so is the language $(ab)^*$ on the alphabet $A = \{a, b\}$ since

$$(ab)^* = \big(bA^* \cup A^*a \cup A^*aaA^* \cup A^*bbA^*\big)^c = \big(b\emptyset^c \cup \emptyset^c a \cup \emptyset^c aa\emptyset^c \cup \emptyset^c bb\emptyset^c\big)^c.$$

Star-free languages were characterized by Schützenberger [77] in 1965.

Theorem 2.1 *A language is star-free if and only if its syntactic monoid is aperiodic.*

Theorem 2.1 gives an algorithm to decide whether a language is star-free. For instance, the language $(aa)^*$ is not star-free, and hence has star height 1, since its syntactic monoid is the cyclic group of order 2, a non-aperiodic monoid.

Theorem 2.1 also suggests that languages of star height $\leqslant n$ form a variety of languages, and can therefore be characterized by a property of their syntactic monoid. This hypothesis, explicitly mentioned in Henneman's thesis [37] in 1971, has never been invalidated. However, the author [62] proved in 1978 that, for any finite monoid M, there is a finite language F such that M divides the syntactic monoid of F^*, with the following consequence:

Theorem 2.1 *If the languages of star-height $\leqslant 1$ form a variety of languages, then all regular languages have star-height $\leqslant 1$.*

Proposition 2.1 appears to kill the algebraic approach, unless every language has star height 0 or 1. However, Straubing, Thérien and the author [67] proved in 1992 a weaker property: the languages of star-height $\leqslant n$ form a *ld-variety*.

Theorem 2.2 *For each n, the class of all languages of star-height $\leqslant n$ is closed under Boolean operations, quotients and inverse of length-decreasing morphisms.*

As we have seen, ld-varieties of languages can be defined by profinite *ld-identities*. Therefore, it would suffice to find a single nontrivial ld-identity satisfied by all languages of star-height $\leqslant 1$ to prove the existence of a language of star-height > 1. Unfortunately, no such identity is known and we still have no example of a language of star height 2. Several past candidates have been turned down

[67, 87, 75], but languages of the form $\pi^{-1}(1)$, where π is a surjective morphism from A^* onto a sufficiently complicated group (say the symmetric group S_5) might be reasonable candidates.

On the other hand, if you believe that all languages have star-height $\leqslant 1$, it "just" remains to prove that the languages of star-height $\leqslant 1$ are closed under inverses of morphisms. One difficulty is that even languages of the form $\varphi^{-1}(F^*)$, where φ is a morphism and F is finite, can be extremely complicated. In particular, we do not know whether the ld-variety generated by the languages of the form F^*, where F is finite, is the variety of all regular languages. A very partial result has been obtained in this direction. Daviaud and Paperman [18] gave profinite equations characterizing the closure under Boolean operations and quotients of the set of languages of the form u^*, where u is a word. However, finding a characterization of the ld-variety generated by these languages is still an open problem and moreover, there is still a giant step to pass from u^* to F^*.

An interesting logical approach was proposed by Lippert and Thomas [53]. The idea was to consider star-free expressions with an additional constant L, where L is a fixed language. However, Lippert and Thomas proved that the equivalence between star-free expressions and first-order logic [59] fails to extend to this setting: the relativized star-free expressions are strictly weaker than the corresponding first-order formulas.

3 Limitedness problem and restricted star height

This section gathers two problems, the limitedness problem and the restricted star height problem. As we will see, the former one can be viewed as a warmup problem for the latter one.

3.1 Limitedness problem

The limitedness problem is not only interesting on its own, but the tools introduced to solve it ultimately led to an elegant solution to the restricted star-height problem. They proved to be very influential in automata theory and they are still the topic of very active research.

The *Kleene star* of a language L is defined as the infinite union

$$L^* = 1 + L + L^2 + \cdots = \sum_{n \geqslant 0} L^n.$$

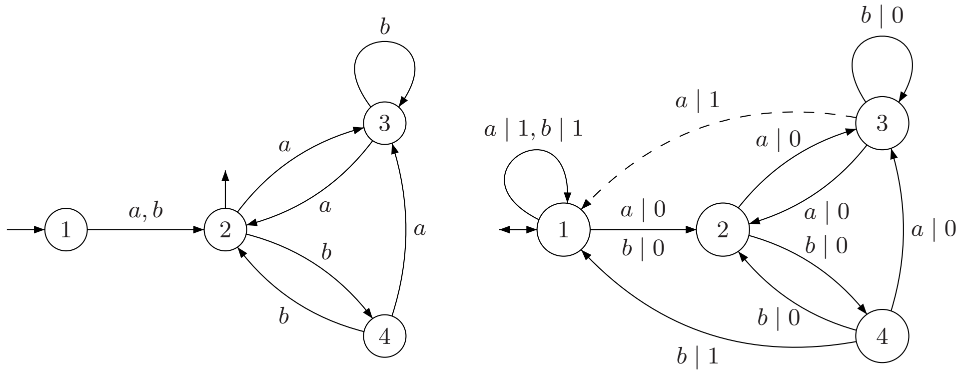
The question arises whether this infinite sum can be truncated. A language is said to be *k-limited* if

$$L^* = 1 + L + \cdots + L^k = (1 + L)^k.$$

It is easy to see that L is k-limited if and only if $1 + L$ is k-limited. Moreover, if L contains the empty word, then L is k-limited if and only if $L^* = L^k$.

A language is *limited* (or has the *finite power property*) if it is k-limited for some $k > 0$. According to Simon [80], the *limitedness problem* was first proposed by Brzozowski in 1966 during the seventh SWAT (now FOCS) Conference. It can be stated as follows:

Problem 2 *Decide whether a given regular language is limited or not.*



The minimal automaton $\mathcal{A}$ of L. An automaton $\mathcal{B}$ for L^*, with weights added.

Figure 3.1: A 4-limited language: $L^* = (1 + L)^4$.

After some preliminary work by Linna [52] in 1973, the problem was solved independently by Hashiguchi [31] and Simon [79] in the late seventies, and an elegant semigroup solution was proposed by Kirsten [41] in 2002. Hashiguchi's solution worked directly on the minimal deterministic automaton recognizing L. Simon went another way and came up with a very simple reduction to a problem on weighted automata. Given the considerable influence of this method on subsequent research, it is worth explaining his idea. Given a regular language L, consider the function $f_L : A^* \to \mathbb{N} \cup \{+\infty\}$ defined by

$$f_L(u) = \begin{cases} \min\{n \mid u \in L^n\} & \text{if } u \in L^*; \\ +\infty & \text{otherwise.} \end{cases}$$

It is clear that L is k-limited if and only if the range of f_L is contained in $\{0, \cdots, k\} \cup \{+\infty\}$. Consequently, L is limited if and only if f_L has a finite range. Now, f_L can be computed by a weighted automaton with output in the tropical semiring $(\mathbb{N} \cup \{+\infty\}, \min, +)$. Let us describe this construction on the example proposed by Simon [80] in his 1988 survey.

Let $A = \{a, b\}$ and $L = (a + b)(b^2 + (a + ba)b^*a)^*$. We start with a standard automaton $\mathcal{A}$ for L, represented on the left hand side of Figure 3.1. Next we build an automaton $\mathcal{B}$ accepting L^* by the standard construction and convert it to a weighted automaton by adding a 0-output to each transition of $\mathcal{A}$ and a 1-output to each new transition in $\mathcal{B}$ (the dashed transitions in Figure 3.1). Intuitively, there is a fee of 1 each time a path reaches state 1 and hence this weighted automaton realizes f_L. Thus the limitedness problem can be reduced to the *finite range problem* for weighted automata.

Problem 3 *Decide whether or not the behaviour of a given weighted automaton has a finite range.*

For the limitedness problem, it suffices to solve the finite range problem for weighted automata over the tropical semiring. But historically, the finite range problem was first studied for the semiring $(\mathbb{N}, +, \times)$ by Mandel and Simon [56] in 1977 and for fields by Jacob [39] in 1978. The first solution for the tropical semiring was given by Simon [79] in 1978. Successive improvements were proposed by Hashiguchi [32, 35] (1982 and 1990), Leung [50] (1987) and Simon [80, 81] (1988 and 1994). See the survey [64] for more details.

Theorem 3.1 *The finite range problem for the tropical semiring* $(\mathbb{N}\cup\{+\infty\}, \min, +)$ *is decidable.*

The complexity of the problem was also analyzed. Hashiguchi [36] showed in 2000 that every n-state distance automaton is either unlimited or limited by $2^{C(n)}$, where $C(n) = 4n^3 + n\ln(n+2) + n \leqslant 4n^3 + n^2 + 2n$. In 2004, Leung and Podolskiy [51] improved this bound to $C(n) = 3n^3 + n\ln n + n - 1$ and proved that limitedness of distance automata is decidable in PSPACE. The problem is in fact PSPACE-complete, as shown by Kirsten [44].

The limitedness problem is undecidable for context-free languages, see Hugues and Selkow [38]. Kirsten and Richomme [42, 46] also investigated the limitedness problem in trace monoids.

3.2 Restricted star height

Regular expressions on the alphabet A are defined recursively as follows:

(1) 0, 1 and a, for each $a \in A$ are regular expressions.

(2) if E and F are regular expressions, then $E + F$, $(E)(F)$ and $(E)^*$ are regular expressions.

The notion of *restricted star height* of a regular expression is the maximum nested depth of stars in the expression. For instance

$$\Big(\big(a(ba)^*b\big)^* + \big(b(aa)^*b + c\big)^*\Big)^*$$

is a regular expression of height 3. More formally, the *restricted star height* $h(E)$ of a regular expression E is defined recursively by

(1) $h(0) = 0$, $h(1) = 0$ and, for each $a \in A$, $h(a) = 0$.

(2) If E and F are regular expressions, then $h(E+F) = h(EF) = \max(h(E), h(F))$ and $h((E)^*) = h(E) + 1$.

The *value* $v(E)$ of a regular expression E is the language of A^* represented by E. Formally, v is a function from the set of regular expressions to the set of regular languages of A^* defined by

(1) $v(0) = 0$, $v(1) = 1$ and $v(a) = a$ for each $a \in A$,

(2) if E and F are two regular expressions, $v(E \cup F) = v(E) \cup v(F)$, $v(EF) = v(E)(F)$ and $v(E^*) = v(E)^*$.

The *restricted star height* $h(L)$ of a regular language L is the minimum of the restricted star heights of the regular expressions which represent it. In other words

$$h(L) = \min\{h(E) \mid E \text{ is a regular expression such that } v(E) = L\}.$$

The star height problem was raised by L.C. Eggan [25] in 1963:

Problem 4 *Is there an algorithm which computes the restricted star height of a given regular language?*

Dejean and Schützenberger [20] first proved in 1966 that for each $n \geqslant 0$, there exists a language of restricted star height n. Only a few partial results [11, 12, 13, 14, 37] were known until Hashiguchi [33] proved in 1982 that restricted star height one is decidable. A few years later, in 1988, Hashiguchi [34] succeeded to prove the general case: restricted star height is decidable. However, Hashiguchi's proof is hard to read and yields an algorithm of non-elementary complexity (cf. Lombardy's thesis [54], Annexe B and examples by Lombardy and Sakarovitch [55]). It took 25 years to obtain Hashiguchi's first solution to Eggan's problem but it took another 23 years until Kirsten [45] found a simplified proof in 2005. Just like for the limitedness problem, the idea of this proof is to reduce the restricted star-height problem to finding upper bounds for the function computed by a new kind of automata, the nested-distance desert automata. The resulting algorithm has a complexity in double exponential space.

Nested-distance desert automata are a particular case of *hierarchical cost automata*. A hierarchical cost automaton is a nondeterministic finite automaton equipped with a totally ordered finite set of counters, initially set to zero. Transitions can increment or reset a given counter, but then all counters of smaller rank have to be reset. A cost automaton is defined in a similar way, but the counters are not ordered and thus transitions can only increment or reset a given counter. These notions have been widely studied in the recent years, notably by Kirsten [44, 43], Bojanczyk and Colcombet [5, 6, 16, 17]. Very recently, Bojanczyk [5] proved that the limitedness problem for cost automata reduces to solving Gale-Stewart games with ω-regular winning conditions, which leads to an entirely new proof of the decidability of the restricted star-height problem.

Even with the progress realised in the recent years, the complexity of the algorithms seemed to exclude the possibility of any practical computation. However, Fijalkow, Gimbert, Kelmendi and Kuperberg addressed the challenge by writing a C++-programme computing the star-height of regular languages accepted by (small) automata, as part of their package `ACME++` (*Automata with Counters, Monoids and Equivalence*), freely available at http://www.liafa.univ-paris-diderot.fr/~nath/?pag

Let me conclude these two sections on star-height by suggesting another problem. Let us define an *intermediate regular expression* as an extended regular expression allowing union and intersection but not complement. Intermediate regular expressions are clearly more general than regular expressions but less general than extended regular expressions. The *intermediate star-height* of a language is the minimum of the star heights of the intermediate regular expressions representing the language. The intermediate version of Problems 1 and 4 can now be stated as follows:

Problem 5 *Are there languages of arbitrary intermediate star-height? Is there an algorithm which computes the intermediate star height of a given regular language?*

4 Group complexity

In this section, all semigroups and groups are supposed to be finite.

The Krohn-Rhodes theorem states that every semigroup S divides a finite alternating wreath product of groups and aperiodic semigroups.

Theorem 4.1 (Krohn-Rhodes 1966) *Every semigroup S divides a wreath product of the form*

$$A_0 \circ G_1 \circ A_1 \cdots A_{n-1} \circ G_n \circ A_n, \tag{$*$}$$

where $A_0, A_1, \ldots, A_n$ are aperiodic semigroups and $G_1, \ldots, G_n$ are groups.

The *group complexity* of S is the smallest possible integer n over all decompositions of type ($*$). Thus aperiodic semigroups have group complexity 0 and nontrivial groups have group complexity 1. However, the following problem is still open:

Problem 6 *Is there an algorithm to compute the group complexity of a semigroup, given its multiplication table?*

This question generated intense research and several important tools of semigroup theory, like the *Rhodes expansion* and *pointlike sets* were introduced in connection with this problem. As a result, the group complexity of many semigroups has been computed. However, as of today, there is no known algorithm to decide whether a semigroup has group complexity 1 and the only known results regarding decidability, due to Karnofsky and Rhodes [40], date back to 1982.

Theorem 4.2 *One can decide whether a semigroup divides a wreath product of the form $G \circ A$. One can decide whether a semigroup divides a wreath product of the form $A \circ G$.*

The book by J. Rhodes and B. Steinberg [74] is by far the most important reference on these questions. It contains a thorough presentation of the numerous tools introduced to attack Problem 6 as well as a detailed survey of the existing partial results up to 2009. Several authors, including Almeida, Auinger, Henckell, Margolis, Rhodes, Steinberg and Volkov enriched the literature since then and the reader is invited to look at the articles of these authors to follow recent progress.

Problem 6 is of course deeply related to the study of the wreath product and have strong connection with language theory. A key reason is that the syntactic semigroup of the composition of two sequential functions divides the wreath product of the syntactic semigroups of the two functions. Cohen and Brzozowski [15] and Meyer [60] used a refinement of the Krohn-Rhodes theorem for aperiodic semigroups to obtain an alternative proof of Schützenberger's characterization of star-free languages. Straubing [82] took further advantage of wreath product decompositions to characterize various classes of regular languages and to state its influential wreath product principle [84], which was later generalized in several ways [68, 10].

It is interesting to see that in the opposite direction, major progress on wreath product decompositions came from language theory. For instance, the characterization of locally testable languages by Brzozowski-Simon [9] and McNaughton [58], Knast's description of dot-depth one languages [47, 48, 86] and Straubing's study of concatenation hierarchies [83] opened the way to Tilson's delay theorem [88].

5 Star removal

Star removal is the only problem of the list that remained untouched, although it is certainly a fascinating question. The absence of references makes it hazardous to evaluate its difficulty, but I hope it will attract more attention in the future. Here is the problem.

Let K be a regular language. Then the equation $K = XK$ has a maximal solution L^*. Then $K = L^*K$ and one can show that the equation in R:

$$K = L^*R$$

has a minimal solution $R = K - (L^* - 1)K$.

Iterating this process on R, we get a decomposition

$$K = L_1^* L_2^* \cdots L_k^* R_k,$$

where R_k is the minimal solution of $K = L_1^* L_2^* \cdots L_k^* R$.

Problem 7 *Does this process terminate (i.e., $L_k^* = 1$ at some point)?*

6 Regularity of non-counting classes

A language L of A^* is said to be *noncounting of order n* if for all $x, y, u \in A^*$,

$$xu^n y \in L \iff xu^{n+1} y \in L.$$

Let $\sim_n$ be the smallest congruence on A^* satisfying $x^n \sim_n x^{n+1}$ for all $x \in A^*$ and let $\mu : A^* \to A^*/{\sim_n}$ be the natural morphism. The problem of the regularity of non-counting classes can be stated formally as follows:

Problem 8 *Is $\mu^{-1}(m)$ a regular language for every $m \in A^*/{\sim_n}$?*

An extended version of the problem was studied by McCammond [57].

Problem 9 *Let $\sim_{n,m}$ be the smallest congruence on A^* satisfying $x^n \sim x^{n+m}$ for all $x \in A^*$. Are the congruence classes regular?*

The problem is not yet entirely solved but generated intense research. The results by de Luca and Varricchio [19] (1990), McCammond [57] (1991), Guba [29, 30] (1993), Do Lago [21, 22, 23, 24] (1996, 1998, 2001, 2002) can be summarized in the following theorems:

Theorem 6.1 *Problem 9 has a positive answer for $n \geqslant 3$ and $m > 0$.*

Theorem 6.2 *Problem 9 has a negative answer for $n = 2$ and $m > 1$.*

For $n = 2$, $m = 1$ ($x^3 = x^2$), the problem is still open, but a partial result is known. Let us say that a word is *overlap-free* if it contains no factor of the form $xyxyx$ for any $x \in A^+$ and $y \in A^*$. If it contains no proper factor of the form above, then the word is said to be *almost overlap-free*.

Theorem 6.3 (Plyushchenko and Shur 2011 [69, 71, 70]) *For $n = 2$ and $m = 1$, the congruence class of a word containing an overlap-free or an almost overlap-free word is a regular language.*

The regularity of noncounting classes is also reminiscent of Burnside's celebrated problem, posed by Burnside in 1902. Burnside asked whether a k-generated group satisfying the identity $x^n = 1$ is necessarily finite. In 1968, Novikov and Adian disproved the conjecture for every odd n larger than 4381, a bound that was later reduced to 665 by Adian. Ivanov also disproved the conjecture for each even n divisible by 2^9 and larger or equal to 2^{48}. The problem has been solved positively for $k = 1$ and for $k > 1$ and $n = 2$, 3, 4 and 6, but is still open for $n = 5$ and $k > 1$.

7 Optimality of prefix codes

Recall that a language X of A^+ is a *code* if the condition

$$x_1 \cdots x_n = x'_1 \cdots x'_m \quad (\text{where } x_i, x'_i \in X)$$

implies $n = m$ and $x_i = x'_i$ for $i = 1, \ldots, n$. It is a prefix code if any two distinct words in X are incomparable for the prefix order.

A language of A^* can be identified with an element of $\mathbb{Z}\langle\langle A \rangle\rangle$, the set of formal series in noncommutative variables in A and coefficients in $\mathbb{Z}$. Let $\alpha : \mathbb{Z}\langle\langle A \rangle\rangle \to \mathbb{Z}\langle A \rangle$ be the natural morphism mapping a series in noncommutative variables onto its commutative version. For instance, if $X = ba + abab + baab + bbab$, then $\alpha(X) = ab + 2a^2b^2 + ab^3$.

A language X is *commutatively prefix* if $\alpha(X) = \alpha(P)$ for some prefix code P. In other words, X is commutatively prefix if there exists a bijection from X to some prefix code mapping every word of X to one of its anagrams.

A nontrivial result, Theorem 14.6.4 in the book of Berstel, Perrin and Reutenauer [4], states that a language X is commutatively prefix if and only if the series $(1 - \alpha(X))/(1 - \alpha(A))$ has nonnegative coefficients. Schützenberger [76] proposed in 1965 the following conjecture:

Conjecture 1 *Every code is commutatively prefix.*

This conjecture generated intense research and was proved in some particular cases, but a counterexample was ultimately found by Peter Shor [78] in 1983. The code

$$X = \{ba, ba^7, ba^{13}, ba^{14}, a^3b, a^3ba^2, a^3ba^4, a^3ba^6, a^8b, \\ a^8ba^2, a^8ba^4, a^8ba^6, a^{11}b, a^{11}ba^2, a^{11}ba^4\}$$

is not commutatively prefix. Following the discovery of this counterexample, Perrin suggested a weaker version of Schützenberger's conjecture. A code is said to be *maximal* if it is not properly contained in any other code.

Problem 10 *Is every finite maximal code commutatively prefix?*

This problem is closely related to a question on optimal encodings [3]. A monoid morphism $\gamma : B^* \to A^*$ is a (prefix) *encoding* if $\gamma(B)$ is a (prefix) code. Let p be a probability on B, representing for instance the frequency of the letters of B. For instance, if B is the usual latin alphabet, $p(a)$ could be the frequency of each letter in written English. Suppose also that each letter a of A has a cost $c(a)$, which, in practice, is often interpreted as the time to send the symbol a. The cost of a word $a_1 a_2 \cdots a_n$ is then defined as the sum $c(a_1) + \cdots + c(a_n)$. The *average weighted cost* of γ is the quantity

$$W(\gamma) = \sum_{b \in B} p(b)c(b)$$

and the *optimal encoding problem* is to find, given A, B, p and c, an encoding γ such that $W(\gamma)$ is minimal. Thus a positive solution to Problem 10 would imply that an optimal encoding can always chosen to be prefix.

Interestingly, Problem 10 is also strongly related to a problem on formal power series. Let $X \subseteq A^+$. A pair (P, S) of subsets of A^* is called a *positive factorization* for X if each word w factorizes uniquely into $w = sxp$ with $p \in P$, $s \in S$, $x \in X$. In terms of formal power series in $\mathbb{N}\langle A \rangle$, this means that

$$A^* = SX^*P \quad \text{or equivalently} \quad 1 - X = P(1 - A)S.$$

This condition implies that X is commutatively prefix. Moreover, if P and S are finite, then X is a finite maximal code. These results motivated the following conjecture, known as the Factorization Conjecture.

Conjecture 2 *For any finite maximal code X over A, there exist two polynomials $P, S \in \mathbb{N}\langle A \rangle$ such that $1 - X = P(1 - A)S$.*

A positive answer to the Factorization Conjecture would also solve positively Problem 10. Both questions are still open, but in 1985, Reutenauer [73] proved the following weaker version of the Factorization Conjecture.

Theorem 7.1 *For any finite maximal code X over A, there exist two polynomials $P, S \in \mathbb{Z}\langle A \rangle$ such that $1 - X = P(1 - A)S$.*

Reutenauer's theorem gives strong evidence that the Factorization Conjecture might be true. For a complete discussion, the reader is referred to the book of Berstel, Perrin and Reutenauer [4] and to the survey papers of Bruyère and Latteux [7] and of Béal, Berstel, Marcus, Perrin, Reutenauer and Siegel [3].

8 Conclusion

Janusz A. Brzozowski really has excellent taste! The challenging problems he selected 35 years ago fostered intense studies and are still at the heart of current research. Only two of them, the limitedness problem and the restricted star height problem, have been completely solved. One of them, the regularity of non-counting classes, is almost solved. Significant progress has been done on group complexity and on optimality of prefix codes. Only little progress is to be reported on star height and the star removal problem remained untouched.

The amount of new ideas created or expanded to solve these questions, mixing algebra, logic and automata theory are cause for optimism and one can hope for a complete solution of some of Brzozowski's open problems within the next 35 years.

Acknowledgments

I would like to thank Jeffrey Shallit for his kind invitation to the Brzozowski's conference and for his useful comments on this paper. The author was funded from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 670624).

References

[1] J. Almeida, *Finite semigroups and universal algebra. Series in Algebra*, vol. 3, World Scientific, Singapore, 1994.

[2] J. Almeida and P. Weil, Relatively free profinite monoids: an introduction and examples, in *NATO Advanced Study Institute Semigroups, Formal Languages and Groups*, J. Fountain (ed.), vol. 466, pp. 73–117, Kluwer Academic Publishers, 1995.

[3] M.-P. Béal, J. Berstel, B. H. Marcus, D. Perrin, C. Reutenauer and P. H. Siegel, Variable-length codes and finite automata, in *Selected topics in information and coding theory*, I. Woungang (ed.), pp. 505–584, *Ser. Coding Theory Cryptol.* vol. 7, World Sci. Publ., Hackensack, NJ, 2010.

[4] J. Berstel, D. Perrin and C. Reutenauer, *Codes and Automata*, *Encyclopedia of Mathematics and its Applications* vol. 129, Cambridge University Press, 2009. 634 pages.

[5] M. Bojańczyk, Star Height via Games, in *Logic in Computer Science (LICS), 2015 30th Annual ACM/IEEE Symposium on*, pp. 214–219, IEEE Computer Society, 2015.

[6] M. Bojańczyk and T. Colcombet, Bounds in w-Regularity, in *21th IEEE Symposium on Logic in Computer Science (LICS 2006), 12-15 August 2006, Seattle, WA, USA, Proceedings*, pp. 285–296, IEEE Computer Society, 2006.

[7] V. Bruyère and M. Latteux, Variable-length maximal codes, in *Automata, languages and programming (Paderborn, 1996)*, pp. 24–47, *Lect. Notes Comp. Sci.* vol. 1099, Springer, Berlin, 1996.

[8] J. A. Brzozowski, Open problems about regular languages, in *Formal language theory. Perspectives and open problems. (Proceedings of a symposium held in Santa Barbara, California, December 10-14, 1979).*, R. V. Book (ed.), pp. 23–47, New York Etc.: Academic Press, a Subsidiary of Harcourt Brace Jovanovich, Publishers. XIII, 454 p., 1980.

[9] J. A. Brzozowski and I. Simon, Characterizations of locally testable events, *Discrete Math.* **4** (1973), 243–271.

[10] L. Chaubard, J.-É. Pin and H. Straubing, Actions, Wreath Products of $\mathcal{C}$-varieties and Concatenation Product, *Theoret. Comput. Sci.* **356** (2006), 73–89.

[11] R. S. Cohen, *Cycle rank of transition graphs and the star height of regular events*, PhD thesis, University of Ottawa, Ottawa, 1968.

[12] R. S. Cohen, Star height of certain families of regular events, *J. Comput. System Sci.* **4** (1970), 281–297.

[13] R. S. Cohen, Techniques for establishing star height of regular sets, *Math. Systems Theory* **5** (1971), 97–114.

[14] R. S. Cohen and J. A. Brzozowski, General properties of star height of regular events, *J. Comput. System Sci.* **4** (1970), 260–280.

[15] R. S. Cohen and J. A. Brzozowski, Dot-Depth of Star-Free Events, *J. Comput. Syst. Sci.* **5**,1 (1971), 1–16.

[16] T. Colcombet, The theory of stabilisation monoids and regular cost functions, in *Automata, languages and programming. Part II*, Berlin, 2009, pp. 139–150, *Lecture Notes in Comput. Sci.* vol. 5556, Springer.

[17] T. Colcombet, Regular cost functions, Part I: Logic and algebra over words, *Log. Methods Comput. Sci.* **9**,3 (2013), 3:3, 47.

[18] L. Daviaud and C. Paperman, Classes of Languages Generated by the Kleene Star of a Word, in *Mathematical Foundations of Computer Science 2015 - 40th International Symposium, MFCS 2015, Milan, Italy, August 24-28, 2015, Proceedings, Part I*, G. F. Italiano, G. Pighizzini and D. Sannella (eds.), pp. 167–178, *Lect. Notes Comp. Sci.* vol. 9234, Springer, 2015.

[19] A. De Luca and S. Varricchio, On noncounting regular classes, in *Automata, languages and programming (Coventry, 1990)*, pp. 74–87, *Lect. Notes Comp. Sci.* vol. 443, Springer, New York, 1990.

[20] F. Dejean and M. P. Schützenberger, On a question of Eggan, *Information and Control* **9** (1966), 23–25.

[21] A. P. Do Lago, On the Burnside semigroups $x^n - x^{n+m}$, *Internat. J. Algebra Comput.* **6**,2 (1996), 179–227.

[22] A. P. Do Lago, Maximal groups in free Burnside semigroups, in *LATIN'98: theoretical informatics (Campinas, 1998)*, pp. 65–75, *Lect. Notes Comp. Sci.* vol. 1380, Springer, Berlin, 1998.

[23] A. P. Do Lago, Local groups in free groupoids satisfying certain monoid identities, *Internat. J. Algebra Comput.* **1–2** (2002), 357–369. International Conference on Geometric and Combinatorial Methods in Group Theory and Semigroup Theory (Lincoln, NE, 2000).

[24] A. P. Do Lago and I. Simon, Free Burnside semigroups, *Theor. Inform. Appl.* **35**,6 (2001), 579–595 (2002). A tribute to Aldo de Luca.

[25] L. C. Eggan, Transition graphs and the star height of regular events, *Michigan Math. J.* **10** (1963), 385–397.

[26] S. Eilenberg, *Automata, languages, and machines. Vol. B*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976. With two chapters ("Depth decomposition theorem" and "Complexity of semigroups and morphisms") by Bret Tilson, Pure and Applied Mathematics, Vol. 59.

[27] Z. Ésik, Extended temporal logic on finite words and wreath products of monoids with distinguished generators, in *Developments in language theory. 6th international conference, DLT 2002, Kyoto, Japan, September 18-21*, M. E. A. Ito (ed.), Berlin, 2002, pp. 43–58, *Lect. Notes Comp. Sci.* n° 2450, Springer.

[28] M. Gehrke, S. Grigorieff and J.-É. Pin, Duality and equational theory of regular languages, in *ICALP 2008, Part II*, L. Aceto and al. (eds.), Berlin, 2008, pp. 246–257, *Lect. Notes Comp. Sci.* vol. 5126, Springer.

[29] V. S. Guba, The word problem for the relatively free semigroup satisfying $T^m = T^{m+n}$ with $m \geqslant 4$ or $m = 3,\ n = 1$, *Internat. J. Algebra Comput.* **3**,2 (1993), 125–140.

[30] V. S. Guba, The word problem for the relatively free semigroup satisfying $T^m = T^{m+n}$ with $m \geqslant 3$, *Internat. J. Algebra Comput.* **3**,3 (1993), 335–347.

[31] K. Hashiguchi, A decision procedure for the order of regular events, *Theoret. Comput. Sci.* **8** (1979), 69–72.

[32] K. Hashiguchi, Limitedness theorem on finite automata with distance functions, *J. Comput. System Sci.* **24** (1982), 233–244.

[33] K. Hashiguchi, Regular languages of star height one, *Inform. and Control* **53** (1982), 199–210.

[34] K. Hashiguchi, Algorithms for determining relative star height and star height, *Inform. and Comput.* **78**,2 (1988), 124–169.

[35] K. Hashiguchi, Improved limitedness theorems on finite automata with distance functions, *Theoret. Comput. Sci.* **72**,1 (1990), 27–38.

[36] K. Hashiguchi, New upper bounds to the limitedness of distance automata, *Theoret. Comput. Sci.* **233**,1-2 (2000), 19–32.

[37] W. Henneman, *Algebraic theory of automata*, PhD thesis, Massachusetts Institute of Technology, 1971.

[38] C. E. Hughes and S. M. Selkow, The finite power property for context-free languages, *Theoret. Comput. Sci.* **15**,1 (1981), 111–114.

[39] G. Jacob, La finitude des représentations linéaires de semi-groupes est décidable, *J. Algebra* **52** (1978), 437–459.

[40] J. Karnofsky and J. Rhodes, Decidability of complexity one-half for finite semigroups, *Semigroup Forum* **24**,1 (1982), 55–66.

[41] D. Kirsten, The finite power problem revisited, *Inf. Process. Lett.* **84**,6 (2002), 291–294.

[42] D. Kirsten, The star problem and the finite power property in trace monoids: reductions beyond C4, *Inform. and Comput.* **176**,1 (2002), 22–36.

[43] D. Kirsten, Desert automata and the finite substitution problem (extended abstract), in *STACS 2004*, pp. 305–316, *Lect. Notes Comp. Sci.* vol. 2996, Springer, Berlin, 2004.

[44] D. Kirsten, Distance desert automata and the star height one problem (extended abstract), in *Foundations of software science and computation structures*, pp. 257–272, *Lect. Notes Comp. Sci.* vol. 2987, Springer, Berlin, 2004.

[45] D. Kirsten, Distance desert automata and the star height problem, *Theoret. Informatics Appl.* **39**,3 (2005), 455–509.

[46] D. Kirsten and G. Richomme, Decidability equivalence between the star problem and the finite power problem in trace monoids, *Theory Comput. Syst.* **34**,3 (2001), 193–227.

[47] R. Knast, A semigroup characterization of dot-depth one languages, *RAIRO Inform. Théor.* **17**,4 (1983), 321–330.

[48] R. Knast, Some theorems on graph congruences, *RAIRO Inform. Théor.* **17**,4 (1983), 331–342.

[49] M. Kunc, Equational description of pseudovarieties of homomorphisms, *Theoretical Informatics and Applications* **37** (2003), 243–254.

[50] H. Leung, *An algebraic method for solving decision problems in finite automata theory*, PhD thesis, Department of Computer Science, The Pennsylvania State University, 1987.

[51] H. Leung and V. Podolskiy, The limitedness problem on distance automata: Hashiguchi's method revisited, *Theoret. Comput. Sci.* **310**,1-3 (2004), 147–158.

[52] M. Linna, Finite power property of regular languages, in *Automata, languages and programming (Proc. Sympos., Rocquencourt, 1972)*, Nivat (ed.), pp. 87–98, North Holland, Amsterdam, 1973.

[53] D. Lippert and W. Thomas, Relativized star-free expressions, first-order logic, and a concatenation game, in *Semigroups, theory and applications (Oberwolfach, 1986)*, pp. 194–204, *Lecture Notes in Math.* vol. 1320, Springer, Berlin, 1988.

[54] S. Lombardy, *Approche structurelle de quelques problèmes de la théorie des automates*, PhD thesis, École nationale sup ?erieure des télécommunications, Paris, 2001.

[55] S. Lombardy and J. Sakarovitch, Star height of reversible languages and universal automata, in *LATIN 2002: Theoretical informatics (Cancun)*, pp. 76–90, *Lecture Notes in Comput. Sci.* vol. 2286, Springer, Berlin, 2002.

[56] A. Mandel and I. Simon, On Finite Semigroups of Matrices, *Theoret. Comput. Sci.* **5** (1977), 101–111.

[57] J. McCammond, The solution to the word problem for the relatively free semigroups satisfying $T^a = T^{a+b}$ with $a \geqslant 6$, *Internat. J. Algebra Comput.* **1**,1 (1991), 1–32.

[58] R. McNaughton, Algebraic decision procedures for local testability, *Math. Systems Theory* **8**,1 (1974), 60–76.

[59] R. McNaughton and S. Papert, *Counter-free automata*, The M.I.T. Press, Cambridge, Mass.-London, 1971. With an appendix by William Henneman, M.I.T. Research Monograph, No. 65.

[60] A. R. Meyer, A note on star-free events, *J. Assoc. Comput. Mach.* **16** (1969), 220–225.

[61] J.-É. Pin, The Brzozowski hierarchy, 45 years later. To appear.

[62] J.-É. Pin, Sur le monoïde de L^* lorsque L est un langage fini, *Theoret. Comput. Sci.* **7** (1978), 211–215.

[63] J.-É. Pin, *Varieties of formal languages*, North Oxford, London and Plenum, New-York, 1986. (Traduction de Variétés de langages formels).

[64] J.-É. Pin, Tropical Semirings, in *Idempotency*, J. Gunawardena (ed.), pp. 50–69, Cambridge University Press, 1998.

[65] J.-É. Pin, Equational descriptions of languages, *Int. J. Found. Comput. S.* **23** (2012), 1227–1240.

[66] J.-E. Pin and H. Straubing, Some results on $\mathcal{C}$-varieties, *Theoret. Informatics Appl.* **39** (2005), 239–262.

[67] J.-É. Pin, H. Straubing and D. Thérien, Some results on the generalized star-height problem, *Information and Computation* **101** (1992), 219–250.

[68] J.-É. Pin and P. Weil, The wreath product principle for ordered semigroups, *Communications in Algebra* **30** (2002), 5677–5713.

[69] A. N. Plyushchenko, On the word problem in free Burnside semigroups with the identity $x^2 = x^3$, *Izv. Vyssh. Uchebn. Zaved. Mat.* **11** (2011), 89–93.

[70] A. N. Plyushchenko and A. M. Shur, Almost overlap-free words and the word problem for the free Burnside semigroup satisfying $x^2 = x^3$, *Internat. J. Algebra Comput.* **21**,6 (2011), 973–1006.

[71] A. N. Plyushchenko and A. M. Shur, On Brzozowski's conjecture for the free Burnside semigroup satisfying $x^2 = x^3$, in *Developments in language theory*, pp. 362–373, *Lect. Notes Comp. Sci.* vol. 6795, Springer, Heidelberg, 2011.

[72] J. Reiterman, The Birkhoff theorem for finite algebras, *Algebra Universalis* **14**,1 (1982), 1–10.

[73] C. Reutenauer, Noncommutative factorization of variable-length codes, *J. Pure Appl. Algebra* **36**,2 (1985), 167–186.

[74] J. Rhodes and B. Steinberg, *The q-theory of finite semigroups*, *Springer Monographs in Mathematics*, Springer, New York, 2009.

[75] J. M. Robson, More languages of generalised star height 1, *Theoret. Comput. Sci.* **106**,2 (1992), 327–335.

[76] M.-P. Schützenberger, Codes à longueur variable. Lecture held in 1965, at a seminar in Royan (also published in École de printemps "Théorie des codes", 1979, p. 247-271), 1965.

[77] M.-P. Schützenberger, On finite monoids having only trivial subgroups, *Information and Control* **8** (1965), 190–194.

[78] P. W. Shor, A counterexample to the triangle conjecture, *J. Combin. Theory Ser. A* **38**,1 (1985), 110–112.

[79] I. Simon, Limited subsets of a free monoid, in *Proc. 19th Annual Symposium on Foundations of Computer Science*, pp. 143–150, IEEE, Piscataway, N.J., 1978.

[80] I. Simon, Recognizable sets with multiplicities in the tropical semiring, in *Mathematical foundations of computer science, 1988 (Carlsbad, 1988)*, pp. 107–120, *Lect. Notes Comp. Sci.* vol. 324, Springer, Berlin, 1988.

[81] I. Simon, On semigroups of matrices over the tropical semiring, *RAIRO Inform. Théor. Appl.* **28**,3-4 (1994), 277–294.

[82] H. Straubing, Families of recognizable sets corresponding to certain varieties of finite monoids, *J. Pure Appl. Algebra* **15**,3 (1979), 305–318.

[83] H. Straubing, Finite semigroup varieties of the form $V * D$, *J. Pure Appl. Algebra* **36**,1 (1985), 53–94.

[84] H. Straubing, The wreath product and its applications, in *Formal properties of finite automata and applications (Ramatuelle, 1988)*, pp. 15–24, *Lect. Notes Comp. Sci.* vol. 386, Springer, Berlin, 1989.

[85] H. Straubing, On logical descriptions of regular languages, in *LATIN 2002*, Berlin, 2002, pp. 528–538, *Lect. Notes Comp. Sci.* n° 2286, Springer.

[86] D. Thérien, Catégories et langages de dot-depth un, *RAIRO Inform. Théor. Appl.* **22**,4 (1988), 437–445.

[87] W. Thomas, Remark on the star-height-problem, *Theoret. Comput. Sci.* **13**,2 (1981), 231–237.

[88] B. Tilson, Categories as algebra: an essential ingredient in the theory of monoids, *J. Pure Appl. Algebra* **48**,1-2 (1987), 83–198.