North-Holland

# RATIONAL LANGUAGES AND THE BURNSIDE PROBLEM

Antonio RESTIVO
*Istituto di Matematica dell' Università di Palermo, C.N.R., Palermo 90133, Italy*

Christophe REUTENAUER
*Institut de Programmation, LITP, CNRS, 75230 Paris Cedex 05, France*

**Abstract.** The problem of finding regularity conditions for languages is, via the syntactic monoid, closely related to the classical Burnside problem. This survey paper presents several results and conjectures in this direction as well as on related subjects, including bounded languages, pumping, square-free words, commutativity, and rational power series.

## 1. Introduction

It is well known to language theoretists that such different concepts as pumping, commutative image, boundedness, square-free words, and regularity conditions are connected, even if no theory exists which actually explains these connections—as far as we know at present. In this paper we shall discuss some results where these concepts are involved. The leitmotiv will be to characterise the class of regular languages.

This problem is closely related to the extended Burnside problem: is every torsional (or periodic) and finitely generated semigroup finite? As we shall see, this torsion property of semigroups has an equivalent property for languages, periodicity, which is related to pumping. So, we shall call *Burnside problem for languages* the problem to find characterisations of regularity (or rationality) which involve pumping-like conditions (Burnside would hopefully forgive us).

This paper is intended to be a survey paper, extending the ideas of an earlier paper [46]. However, we cannot even hope to be complete on the subject; moreover, we shall only briefly discuss the Burnside problem for semigroups, to show its connections with regularity conditions for languages and not for itself: to be complete on the Burnside problem (groups and semi-groups) would lead to writing a book in several volumes! On the other hand, there will be only one real proof: the proof of our characterisation of rationality via the transposition property (in fact, we shall generalise this result through the notion of 'property $\sigma$'). We have added a section on rational power series, because these are closely related to regular languages, both by the results obtained and the languages which they define (supports). In Appendix A we recall some classical definitions on words and languages.

## 2. Burnside problem

By Kleene's theorem, a language $L \subset A^*$ is *regular* (we shall also say *rational*) if and only if it is *recognizable* (by a finite automaton). This last condition is equivalent to: the syntactic monoid of $L$ is finite (see, e.g., [17, Vol. A, Proposition III.10.1] or [28, Proposition 6.1.8]).

Recall that the *syntactic congruence* of a language $L$, written

$$x \equiv y \mod L,$$

is the least congruence of $A^*$ (that is, an equivalence relation on $A^*$ which is compatible with the product on $A^*$) such that $L$ is a union of equivalence classes of this congruence.

Equivalently, it is easy to show that $x \equiv y \mod L$ holds if and only if

$$\forall u, v \in A^*: \quad uxv \in L \Leftrightarrow uyv \in L. \tag{1}$$

This means that $x$ and $y$ have the same contexts in $L$: every time you see an $x$ in a word in $L$, you may replace it by a $y$, and vice versa. This explains the terminology 'syntactic', which refers to its origin in linguistics. For instance, in a very rough linguistic model of the French language (say), any verb may be replaced by any other in any correct sentence: verbs are syntactically equivalent.

Now, the syntactic monoid of $L$ is the monoid which is the quotient of $A^*$ by the syntactic congruence of $L$. So, rationality of a language is reduced to the finiteness of a certain monoid. This naturally leads to the (generalised) *Burnside problem*, which asks if a given monoid having some finiteness properties is always finite.

Before going into details, let us present a characterisation of rationality, due to Ehrenfeucht, Haussler and Rozenberg [15], which generalises the one with the syntactic congruence.

**Definition 2.1** ([15]). A *monotone well quasi-order* on $A^*$ is a binary relation $\leqslant$ on $A^*$ which is reflexive, transitive, compatible with the concatenation of words (i.e., $u_1 \leqslant u_2$ and $v_1 \leqslant v_2$ implies $u_1v_1 \leqslant u_2v_2$) and such that, for any $L \subset A^*$, the set of minimal elements of $L$ is finite.

**Theorem 2.2** ([15, Theorem 3.3]). *A language $L \subset A^*$ is rational if and only if $L$ is upwards closed for some monotone well quasi-order $\leqslant$ on $A^*$ (i.e., $x \in L$ and $x \leqslant y$ implies $y \in L$).*

We now come back to the Burnside problem, which we briefly discuss before returning to languages. In 1902, Burnside [13] posed the following problem: given a finitely generated group $G$ such that each $x$ in $G$ satisfies $x^n = 1$ ($n$ depending only on $G$), is $G$ finite?

The answer is "yes" for $n = 2$ ($G$ is then commutative), $n = 3$ (Burnside [13]), $n = 4$ (Sanov, see [1] for reference), and $n = 6$ (Hall [21]).

But Adjan [2] showed that, in general, the answer is negative (see [42] for another proof). In fact, the story of the answer to the Burnside problem is a rather long one and mentions the names of many mathematicians such as Novikov [40] and Britton [11]. The known proofs are very long and controversial still. So, we can only advise the interested reader to check these proofs himself.

**Remark 2.3.** Kaplansky [26, p. 101] states the *strong Burnside problem* as follows: if $G$ is a finitely generated *torsion* group (i.e. every $x$ in $G$ generates a finite subgroup), is $G$ finite? This was answered negatively, before Adjan's proof, by Golod and Shafarevitch (see [24, Chapter 8]).

For monoids, the situation is much more easy: as noted by Morse and Hedlund [39], the existence of an infinite number of square-free words in $A^*$ ($|A| \geqslant 3$) implies that the monoid quotient $M$ of $A^* \cup 0$ by all the relations $xx = 0$ is infinite; and in $M$, every element $x$ satisfies $x^2 = x^3$. So, the answer to the Burnside problem for monoids (and for semigroups) is negative.

**Remark 2.4.** In [12, Corollary 5.7], Brzozowski, Culik, II and Gabrielan show that if $|A| = 2$, the monoid quotient of $A^*$ by all the relations $x^2 = x^3$ $(x \in A^*)$ is infinite, although there are only finitely many square-free words in $A^*$.

An interesting connection between the Burnside problem for groups and that for monoids was given by Green and Rees [20].

**Theorem 2.5** ([20], see also [28, Chapter 10]). *The following conditions are equivalent for $n \geqslant 2$:*

(i) *If $G$ is a finitely generated group where each element satisfies $x^{n-1} = 1$, then $G$ is finite.*

(ii) *If $M$ is a finitely generated monoid where each element satisfies $x^n = x$, then $M$ is finite.*

A particular case is the case $n = 2$, that is, each finitely generated idempotent monoid is finite. This means that the quotient monoid of $A^*$ by all the relations $xx = x$ $(x \in A^*)$ is finite; compare with the above counterexample of Morse and Hedlund $(A^* \cup 0/xx = 0)$! (for a proof of the finiteness of idempotent semigroups, see [17, Vol. B, Proposition IX.7.1] or [33, Theorem 2.4.1]; see also [62] for a generalisation).

As the answer to the Burnside problem was found to be negative (and even before that), many people have studied groups or monoids under special assumptions, in which case they showed that the answer is positive. We only give a few references on this topic. Let us say that a monoid is *periodic* (or torsional) if any element of it generates a finite submonoid. If was shown by Schur [[57] that any finitely generated periodic group of matrices over the set of complex numbers is finite; this result was extended to arbitrary fields by Kaplansky [26, Theorem G, p. 105], to monoids by McNaughton and Zalcstein [38], and to submonoids of pi-rings by Straubing [64]. For decidability questions on this topic, see [25] and [36] or [7, Chapter 6].

Other positive results on the Burnside problem for semigroups are given by Simon [62] ('strong periodicity'), De Luca and Restivo [34] ('iteration on the right', which looks like pumping), and Restivo and Reutenauer [47] ('permutation property').

We now come back to languages. It is an easy consequence of (1) that the syntactic monoid of a language $L$ is periodic if and only if, for any word $x$, there exist, $k$, $p$, $k \neq p$, such that, for any words $u$, $v$,

$$ux^kv \in L \Leftrightarrow ux^pv \in L. \tag{2}$$

We say that, in this case, the language is *periodic*. Note that if relation (2) holds (with $p > k$), then the set $ux^k(x^{p-k})^*v$ is contained in $L$ or in its complement. This shows that periodicity is close to pumping: $x$ is an iterating factor of $L$ or of its complement (see Appendix A).

Examples of periodic languages are the set of words with squares and also the set of square-free words. Note that periodicity, if true for $L$, is also true for its complement. The same will hold for the transposition and cancellation properties, and for weak commutativity.

Furthermore, the set of palindromes is a nonperiodic language (consider the words $a^nba^n$).

The alphabet being finite, we want to know under which conditions a periodic language $L$ is rational. This could be called the *Burnside problem for languages*. Now, periodicity is not enough, as the following example shows:

$$L = \{\text{square-free words}, |A| \geqslant 3\}$$

(square-free words always serve as counterexample, in these topics). There are some cases where the answer is easily seen to be positive: the cases where $L$ is *commutative* or *bounded* (see Appendix A). Indeed, if $L$ is commutative and periodic, then its syntactic monoid is commutative and periodic, hence finite; thus, $L$ is rational.

When $L$ is bounded and periodic, then $L$ is rational, as shown in [32, Proposition 6].

We now introduce a property of languages which generalises commutativity and which will give an answer to the Burnside problem for languages.

**Definition 2.6.** Let $\sigma$ be a permutation of $\{1, \ldots, k\}$ different from the identity. We say that a language $L$ has *property* $\sigma$ if there exists an $m \geqslant k$ such that, whenever a word $w$ is written $w = ux_1 \ldots x_mv$, there exist $i_1, i_2, \ldots, i_{k+1}$ $(i_1 < i_2 < \cdots < i_{k+1})$ such that, setting

$$y_j = x_{i_j}x_{i_{j+1}} \ldots x_{i_{j+1}-1} \quad (1 \leqslant j \leqslant k),$$

$$u' = x_1x_2 \ldots x_{i_1-1},$$

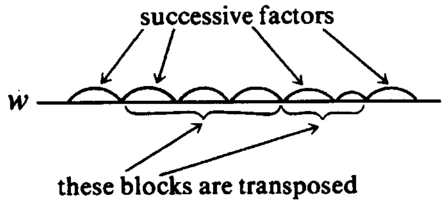
$$v' = x_{i_{k+1}} \ldots x_m,$$

one has

$$w = uu'y_1 \dots y_k v'v \in L \Leftrightarrow uu'y_{\sigma(1)} \dots y_{\sigma(k)} v'v \in L.$$

(The word on the left is $w$ and that on the right is obtained from $x$ by applying the permutation $\sigma$ to the consecutive blocks $y_1, \dots, y_k$ of $x$'s in $w$.)

**Remark 2.7.** When $k = 2$, that is, $\sigma$ is permutation (12), we obtain the ***transposition property*** of [45, 46]: A language $L$ has this property if, for some $m$, every time one distinguishes $m$ successive factors in some word $w$, then there exist two consecutive blocks of these factors which, when transposed, give a word $w'$ such that $w \in L \Leftrightarrow w' \in L$:



Note that property (12) (i.e., the transposition property), when $m = 2$, is equivalent to the commutativity of $L$.

**Remark 2.8.** Call a property of languages *syntactic* if whenever two languages $L$, $L'$ have the same syntactic monoid and if $L$ has the property, then $L'$ also has this property. In this sense, commutativity is a syntactic property ($L$ commutative $\Leftrightarrow$ its syntactic monoid is commutative), periodicity is syntactic, whereas property $\sigma$ is not syntactic for $m \geqslant 3$. This stems from the fact that the $i$'s in the definition depend on the 'contexts' $u$, $v$. We point out that when a characterisation of rationality involves only properties of languages which are syntactic, then this result is rather a result on semigroups than on languages; but when the property involved is not syntactic, then it is not possible to pass directly through the syntactic monoid to prove the rationality of the language; the dependence on contexts has to be surpassed by very strong 'forcing' arguments, like Ramsey's theorem, which we shall use in the proof of our main theorem.

A rational language always has property $\sigma$: consider the loops in a finite automaton accepting it. Furthermore, there are languages which do not have the property: for example, the set of palindromes; consider a palindrome of the form

$$\underbrace{ab}_{x_1}\underbrace{a^2b}_{x_2}\dots\underbrace{a^mb}_{x_m}a^mba^{m-1}\dots ba.$$

No permutation of the $x$'s other than the identity keeps this word a palindrome. We can now prove a characterisation of rationality.

**Theorem 2.9.** *Let $\sigma \neq \mathrm{id}$ be a permutation of $\{1, \ldots, k\}$. A language is rational if and only if it is periodic and has the property $\sigma$.*

Before proceeding to the proof of this theorem, we give a definition. Let $A$ be totally ordered and order the words of equal length lexicographically (from left to right).

**Definition 2.10** (Shirshov [60]; see also [33, p. 144] and [54, p. 205]). A word $w$ is *n-divided* if it admits a factorization $w = ux_1 \ldots x_n v$ such that for any permutation $\alpha$ of $\{1, \ldots, n\}$, $\alpha \neq \mathrm{id}$, one has

$$w < ux_{\alpha(1)} \ldots x_{\alpha(n)} v.$$

We need the following two theorems.

**Theorem 2.11** (Shirshov [60]). *Given a totally ordered alphabet $A$, integers $p$ and $n$ with $p \geqslant 2n$, there exists an integer $N = N(A, p, n)$ such that any word of length at least $N$ in $A^*$ is n-divided or contains a p-th power of a nonempty word of length $\leqslant n-1$.*

**Theorem 2.12** (Ramsey [22, Theorem 1.7.1]). *For each set $X$ and each integer $k$, denote by $X[k+1]$ the set of subsets of $X$ of cardinality $k+1$. Then, for each $m \geqslant 1$ there exists and integer $n(m)$ such that for each set $X$ with $\mathrm{card}(X) \geqslant n(m)$ and each partition $X[k+1] = I \cup J$ there exists some subset $Y$ of $X$ with $|Y| = m$ such that*

$$\textit{either } Y[k+1] \subset I \textit{ or } Y[k+1] \subset J. \tag{3}$$

**Proof of Theorem 2.9.** (i) Let $k \geqslant 2$ and let $\sigma$ be a permutation of $\{1, \ldots, k\}$, $\sigma \neq \mathrm{id}$. Let $m \geqslant k$. Denote by $\mathscr{L}_{m,p}$ the set of languages over the alphabet $A$ such that:

- $L$ has property $\sigma$ for the integer $m$,
- for any word $x$ of length smaller than $n(m)$ (the integer of Ramsey's Theorem), one hase

$$\forall u, v \in A^*: \quad ux^p v \in L \Leftrightarrow ux^{p+p!} v \in L. \tag{4}$$

(ii) Let $L$ be a periodic language which has the property $\sigma$ for $m$. We show that $L \in \mathscr{L}_{m,p}$ for some $p$ with $p+p! \geqslant 2n(m)$. Indeed, let $W$ be the set of words of length smaller than $n(m)$. For each $x$ in $W$, as $L$ is periodic, there exist $i_x$, $j_x$ with $j_x \geqslant 1$ such that

$$\forall u, v \in A^*: \quad ux^{i_x} v \in L \Leftrightarrow ux^{i_x + j_x} v \in L.$$

Let

$$i = \sup\{i_x \,|\, x \in W\}, \qquad j = \sup\{j_x \,|\, x \in W\}.$$

Choose $p$ such that $p \geqslant i$, $p \geqslant j$, $p+p! \geqslant 2n(m)$. Then, if $|x| \leqslant n(m)$, i.e., $x \in W$, one has, $\forall u, v \in A^*$:

$$\begin{aligned} ux^p v \in L &\Leftrightarrow ux^{p-i_x}x^{i_x}v \in L \\ &\Leftrightarrow ux^{p-i_x}x^{i_x+j_x}v \in L \\ &\Leftrightarrow ux^{p-i_x+j_x}x^{i_x}v \in L \\ &\Leftrightarrow ux^{p-i_x+j_x}x^{i_x+j_x}v \in L \\ &\Leftrightarrow ux^{p-i_x+2j_x}x^{i_x}v \in L \\ &\;\vdots \\ &\Leftrightarrow ux^{p-i_x+p!}x^{i_x}v \in L \quad (\text{because } j_x \text{ divides } p!) \\ &\Leftrightarrow ux^{p+p!}v \in L. \end{aligned}$$

Hence, $L$ is in $\mathscr{L}_{m,p}$ with $p+p! \geqslant 2n(m)$.

(iii) By (ii), we know that any periodic language having property $\sigma$ is in some $\mathscr{L}_{m,p}$ with $p+p! \geqslant 2n(m)$. It will thus suffice to show that any such $\mathscr{L}_{m,p}$ is finite. Indeed, if $L$ is in $\mathscr{L}_{m,p}$, then so is

$$a^{-1}L = \{w \mid aw \in L\}$$

(as may easily be verified), so one may apply Nerode's criterion to conclude that any language in $\mathscr{L}_{m,p}$ is rational [17, Vol. A, Theorem III.8.1].

(iv) Let $p+p! \geqslant 2n(m)$, $\mathscr{L} = \mathscr{L}_{m,p}$ and $n = n(m)$. Furthermore, let $N = N(A, p+p!, n)$ be the integer of Shirshov's theorem. Let $L$, $L'$ in $\mathscr{L}$ be such that

$$\forall w \in A^*, |w| < N: \quad w \in L \Leftrightarrow w \in L'$$

(i.e., $L$ and $L'$ are equal when restricted to the words of length $< N$). We show that then one has $L = L'$; this will ensure that $\mathscr{L}$ is finite and will conclude the proof.

(v) We order $A^*$, first by length, then lexicographically: that is, $w < w'$ if either $|w| < |w'|$ or $|w| = |w'|$ and $w$ is smaller than $w'$ in the lexicographic order. This order is a well ordering and has a smallest element, the empty word. So, we may use induction with respect to this order.

(vi) We show by induction that, for any word $w$, $w \in L \Leftrightarrow w \in L'$. We already know that $w \in L \Leftrightarrow w \in L'$ if $|w| < N$. Let $|w| \geqslant N$. Suppose $w$ contains the $(p+p!)$th power of some nonempty word $x$: $w = ux^{p+p!}v$, with $|x| \leqslant n-1$. As $L$, $L'$ are both in $\mathscr{L} = \mathscr{L}_{m,p}$, we have, by (4),

$$ux^{p+p!}v \in L \Leftrightarrow ux^p v \in L \quad \text{and} \quad ux^{p+p!}v \in L' \Leftrightarrow ux^p v \in L'.$$

Now, $ux^p v$ is smaller than $w$ (with regard to length), thus, by the induction hypothesis, we have

$$ux^p v \in L \Leftrightarrow ux^p v \in L'.$$

Taking these equivalences together we obtain

$$w \in L \Leftrightarrow w \in L'.$$

(vii) Suppose now that $w$ contains no $(p+p!)$th power of a nonempty word of length at most $n-1$. Then, by Shirshov's theorem, $w$ is $n$-divided:

$$w = ux_1 \ldots x_n v. \tag{5}$$

Let $X = \{1, 2, \ldots, n\}$. Define a subset $I$ of $X[k+1]$ by $\{i_1, i_2, \ldots, i_{k+1}\}$ with $i_1 < i_2 < \cdots < i_{k+1}$ is in $I$ if, setting

$$y_j = x_{i_j} x_{i_j+1} \ldots x_{i_{j+1}-1} \quad (1 \leq j \leq k),$$

$$u' = x_1 x_2 \ldots x_{i_1-1},$$

$$v' = x_{i_{k+1}} \ldots x_m,$$

one has $uu'y_{\sigma(1)}y_{\sigma(2)} \ldots y_{\sigma(k)}v'v \in L$.

A very important remark is that, because $w$ is $n$-divided by (5), one has $w > uu'y_{\sigma(1)}y_{\sigma(2)} \ldots y_{\sigma(k)}v'v$; hence, by induction, $I$ remains unchanged if in its definition $L$ is replaced by $L'$.

Let $J = X[k+1] \backslash I$. Then, by Ramsey's theorem, there exists an $Y \subset X$ with $|Y| = m$ and such that either $Y[k+1] \subset I$ or $Y[k+1] \subset J$. This means that $w$ may be written as

$$w = u''z_1 \ldots z_m v''$$

(each $z_i$ is a block of $x$'s, the indices of the $x$'s being taken in $Y$) such that either for any $i_1, \ldots, i_{k+1}$, $i_1 < i_2 < \cdots < i_{k+1}$, setting

$$t_j = z_{i_j} z_{i_j+1} \ldots z_{i_{j+1}-1} \quad (1 \leq j \leq k),$$

$$u''' = z_1 \ldots z_{i_1-1},$$

$$v''' = z_{i_{k+1} \ldots z_m},$$

one has

$$u''u'''t_{\sigma(1)}t_{\sigma(2)} \ldots t_{\sigma(k)}v'''v'' \in L, \tag{6}$$

or for any $i_1, \ldots, i_{k+1}$, $i_1 < i_2 < \cdots < i_{k+1}$ setting the same $t$'s, $u'''$, and $v'''$ as above, one has

$$u''u'''t_{\sigma(1)}t_{\sigma(2)} \ldots t_{\sigma(k)}v'''v'' \notin L. \tag{7}$$

But $L$ has property $\sigma$, hence, if $w \in L$, there are some $i_1, \ldots, i_{k+1}$ such that (6) holds; in this case, $Y[k+1] \subset I$. And if $w \notin L$, because of property $\sigma$, there are some $i_1, \ldots, i_{k+1}$ such that (7) holds; in this case, $Y[k+1] \subset J$. Thus, $w \in L \Leftrightarrow Y[k+1] \subset I$.

Because of the (important) remark above, the same holds with $L'$ in place of $L$.

Hence,

$$w \Leftrightarrow Y[k+1] \subset I \Leftrightarrow w \in L'.$$

This concludes the proof. □

Similar to property $\sigma$ is the *permutation property*: a language $L$ has this property if, for some $m$ (depending only on $L$), for any words $u, x_1, \dots, x_m, v$ there exists a permutation $\alpha$ of $\{1, .., m\}$, $\alpha \neq \mathrm{id}$, such that

$$ux_1 \dots x_m v \in L \Leftrightarrow ux_{\alpha()1)} \dots x_{\alpha(m)} v \in L.$$

This is just commutativity for $m = 2$. As before, the language of palindromes does not have the permutation property.

**Problem 2.13.** Does Theorem 2.9 still hold with the permutation property instead of property $\sigma$?

**Remark 2.14.** If, in the permutation property, permutation $\alpha$ depends on $x_1, \dots, x_m$ only (and not on the contexts $u, v$), then the property becomes syntactic and the answer to the problem is positive; indeed the syntactic monoid then has the property considered in [47] and, thus, is finite.

The previous theorem has a formal analogy with a nice result due to Ehrenfeucht, Parikh and Rozenberg [16].

**Definition 2.15** ([16]). A language $L$ has the *cancellation property* if there exists an $n \geqslant 1$ such that for any words $u, x_1, \dots, x_n, v$ there exist $i, j$, $1 \leqslant i < j \leqslant n+1$, such that

$$ux_1 \dots x_n v \in L \Leftrightarrow ux_1 \dots x_{i-1} x_j \dots x_n v \in L$$

(the word on the right is obtained by cancelling $x_i \dots x_{j-1}$ in the word on the left). This means that if a word $w$ is written $w = ux_1 \dots x_n v$, then, by cancelling a block of $x$'s, one obtains a word $w'$ with $w \in L \Leftrightarrow w' \in L$.

As for property $\sigma$, the cancellation property is not a syntactic property (except in the trivial case $n = 1$, where $L = \emptyset$ or $A^*$).

**Theorem 2.16** ([16]). *A language is rational if and only if it has the cancellation property.*

In fact Ehrenfeucht et al. [16] deal with pumping, but it is clear that the rationality is derived from the cancellation, and not from the pumping. Note that the classical pumping lemma says something like $ux^*v \subset L$. This may be written in two parts: (i) $uv \in L$ (cancellation of $x$), and (ii) $ux^+v \subset L$ (pumping of $x$).

The proof of Theorem 2.16 uses Ramsey's theorem (Theorem 2.12), as does the proof of Theorem 2.9 (but in fact, from it we were inspired to use Ramsey's theorem). For an application of Theorem 2.16, see Section 4.

Before closing this section, we give some results which involve rationality and context-freeness.

A conjecture of Thierrin (cited in [55, p. XV]) stated that every periodic contex-free language is rational (i.e., the Burnside problem for syntactic monoids of context-free languages); this was recently disproved by Main, Bucher and Haussler [35]. They also disproved a conjecture of Boasson, which also gave a characterisation of rationality: recall that an *iterative pair* (respectively a *strong iterative pair*) of a language $L$ is a five-tuple $(u, x, v, y, w)$ of words such that $ux^nvy^nw \in L$ for any $n \geqslant 1$ (respectively any $n \geqslant 0$).

This is a central notion in the theory of context-free languages, as is well known (see, e.g., [5]). The conjecture of Boasson was that if a context-free language $L$, for some $k \geqslant 1$, has the following property:

($P_k$) for each strong iterative pair $(u, x, v, y, w)$ of $L$, one has $u(x^k)^+v(y^k)^+w \subset L$, then $L$ is rational. So this conjecture is false; however, the statement corresponding to $k=1$ was proved by Ehrenfeucht, Haussler and Rozenberg [15] (and before, in a weaker form, by Boasson [9]). To disprove the two conjectures, the authors show that the set of left factors of the infinite word of Thue and Morse (see [33, Section 2.2]) is the complement of a context-free language. This was recently extended by Berstel [6] to any infinite word generated by a morphism.

Giving a positive answer to a conjecture of Latteux (in his thesis), Kortelainen [27] has recently proved the following result.

**Theorem 2.17** ([27]). *Any commutative quasi-rational language is rational.*

A weaker form of this (with 'linear' replacing 'quasi-rational') was proved earlier in [15, Corollary 6.4]. Similar to the previous result is the following, due to Latteux and Rozenberg [31].

**Theorem 2.18** ([31, Theorem 5]). *Every commutative one-counter context-free language is rational.*

We finish this section with a nice result due to Latteux, which gives another characterisation of rationality (because, as is well known, the shuffle of a context-free and a rational language is always context-free).

**Theorem 2.19** ([29, Proposition IV.4]). *If the shuffle of two context-free languages on disjoint alphabets is context-free, then one of them is rational.*

A similar result, but involving E0L and context-free languages, was proved by Engelfriet and Rozenberg [18].

## 3. Bounded languages

Bounded languages (for a definition, see Appendix A) constitute a very special class of languages, which are in some sense rather sets of $n$-tuples of integers than words. In [45], the present authors have given a characterisation of bounded languages, which is a language-theoretic version of a result of Shirshov on pi-algebras.

**Theorem 3.1** ([45]). *A language is bounded if and only if for some $n$, it contains no $n$-divided word.*

This rather abstract characterisation, however, is interesting, because you do not have to exhibit the words $u_1, \ldots, u_q$ such that

$$L \subset u_1^* \ldots u_q^*.$$

**Remark 3.2.** The statement of Shirshov [60, Theorem 1] is: given a finitely generated pi-algebra $\mathscr{A}$ over a commutative ring $R$, and given set $x_1, \ldots, x_p$ of generators of $\mathscr{A}$, there exist noncommutative monomials $u_1, \ldots, u_q$ in the $x$'s such that $\mathscr{A}$ is generated, as an $R$-module, by the monomials $u_1^{i_1} \ldots u_q^{i_q}$, $i_1, \ldots, i_q \geqslant 0$.

We may derive a nice result connecting bounded languages, pumping, square-free words, $n$-divided words and growth.

**Theorem 3.3** ([44, 45, 37]). *For a rational language $L$, the following conditions are equivalent*:

(i) *$L$ is not bounded.*

(ii) *$L$ has infinitely many primitive iterating factors.*

(iii) *For some $p$, there are arbitrary long words in $L$ without $p$-th power.*

(vi) *For any $n$, $L$ contains an $n$-divided word.*

(v) *The population function of $L$ is not polynomially bounded.*

(See Appendix A for the corresponding definitions.) A similar result holds for context-free languages (see [10, 32, 45]).

It is an immediate consequence of the above result ((i)$\Leftrightarrow$(iii)) and of Shirshov's theorem (Theorem 2.11) that a rational language without $n$-divided words (for some $n$) is bounded. This raises the question whether the language

$$L_n = \{w \in A^*, \ w \text{ is not } n\text{-divided}\}$$

is rational. This is not the case, as shown in [45, p. 211]. However, we do not know whether $L_n$ is context-free or not.

**Problem 3.4.** Does the above result extend to supports? (See Section 4 for the definition of supports.)

By a well-known theorem of Parikh, the commutative image of any context-free language is a *semi-linear set* (or a rational subset of the monoid $\mathbb{N}^{|A|}$). This result admits the following generalisation. A language is called *Parikh-bounded* if it contains some bounded language having the same commutative image (the reader should convince himself that not all languages have this property; take, for instance the complement of the Goldstine language [19]).

Blattner, Latteux and Leguy have proved the following result.

**Theorem 3.5** ([8, 30]). *Any context-free language is Parikh-bounded.*

This implies Parikh's theorem, once it has been proved that the commutative image of any bounded context-free language is semi-linear.

In [45, Theorem 5.1] a sufficient condition for a language to be Parikh-bounded was given.

**Theorem 3.6** ([45, Theorem 5.1]). *A language is Parikh-bounded if it has the weak permutation property, that is, if for some n, for any words u, $x_1, \ldots, x_n$, v, there exists a permutation $\sigma \neq \mathrm{id}$ such that*

$$ux_1 \ldots x_n v \in L \;\Rightarrow\; ux_{\sigma(1)} \ldots x_{\sigma(n)} v \in L.$$

**Remark 3.7.** The weak permutation property is obtained from the permutation property by replacing $\Leftrightarrow$ by $\Rightarrow$.

Unfortunately, context-free languages do not have this property in general (palindromes!). However, this result applies to rational languages and supports (see Section 4).

We conclude this section with a related topic.

**Conjecture 3.8** (Fliess). *Let $L \subset \{a_1, \ldots, a_n\}^*$ be a commutative language such that, for any permutation $\sigma$, $L \cap a^*_{\sigma(1)} \ldots a^*_{\sigma(n)}$ is context-free. Then, L is context-free.*

This was proved in a particular case by Beauquier, Blattner and Latteux [4].

## 4. Formal power series

Analogous to the Burnside problem is the Kurosh problem for algebras: *given a finitely generated algebra where each element is algebraic over the ground field (i.e., each element generates a finite-dimensional subalgebra), is this algebra of finite dimension over this field?*

In this general statement, the answer is negative as shown by Golod and Shafarevitch (see [24, Chapter 8]). However, the answer is positive when the algebra satisfies a *polynomial identity*: recall that an algebra $A$ over a field $R$ satisfies a polynomial identity if there exists a noncommutative polynomial $P(x_1, \ldots, x_q) \neq 0$ over $R$ vanishing whenever $x_1, \ldots, x_q$ are substituted by elements of $A$; the particular case $P = x_1x_2 - x_2x_1$ means that $A$ is commutative.

The positive answer to the Kurosh problem in the case of pi-algebras was proved by Shirshov in 1957; it was not outside Russia until the beginning of the 70s (see [54, p. 339] for details).

This theorem of Shirshov allows to characterise rational formal power series. We shall not go into details, but give the criterion only, thereby referring the interested reader to [7, 50, 56]. In [50], the following result was shown.

**Theorem 4.1** ([50, Proposition II.3.2]). *Formal power series $S = \sum_{w \in A^*} (S, w)w$ with coefficients in the commutative ring $R$ is rational if and only if*:

(i) *the syntactic algebra of $S$ satisfies a polynomial identity,*

(ii) *for any word $x$, there is a common linear recurrence relation over $R$ satisfied by all the sequences $((S, ux^nv)_{n \in \mathbb{N}}$ $(u, v \in A^*)$.*

Note that condition (i) is the analogue of property $\sigma$ of Theorem 2.9, while condition (ii) is the analogue of periodicity; however, there is no simple implication between the two results, although they have a common root: Shirshov's theorem (Theorem 2.11).

This result applies in particular to languages because, as shown by Schützenberger [58, Theorem IV.4] *a language is regular if and only if its characteristic series is rational* (see also [56, Theorem II.5.2]). A particular case of this is the following: as in [52], we say that a language $L$ is *weakly commutative* if, for some $n$, for any words $u, x_1, \ldots, x_n, v$, one has

$$|\{\sigma \text{ even} \mid ux_{\sigma(1)} \ldots x_{\sigma(n)}v \in L\}| = |\{\sigma \text{ odd} \mid ux_{\sigma(1)} \ldots x_{\sigma(n)}v \in L\}|. \tag{8}$$

In fact, this condition means exactly that the syntactic algebra of (the characteristic series of) $L$ satisfies the *standard identity* $S_n = 0$, where

$$S_n = \sum_{\sigma \in \mathfrak{S}_n} (-1)^\sigma x_{\sigma(1)} \ldots x_{\sigma(n)}$$

(the alternating sum). Hence, we have the following criterion.

**Theorem 4.2** ([52]). *A language is rational if and only if it is weakly commutative and periodic.*

The direct part of this result is also interesting because it states that, for any regular language $L$, the rather curious condition (8) holds (for some $n$ depending on $L$), that is, there are as many even permutations $\sigma$ of $\{1, \ldots, n\}$ with $ux_{\sigma(1)} \ldots x_{\sigma(n)}v \in L$ as odd permutations, for any choice of $u$, $v$, and $x$'s. This may be proved directly by elementary methods (see [52]).

The Burnside problem intervenes also for rational power series. In fact, when the ring of coefficients is a field, we have the following result.

**Theorem 4.3** (see [7, Corollary 2, p. 119]). *A rational power series is of finite image (i.e., has only finitely many distinct coefficients) if and only if for any words $u$, $x$, $v$ the set*

$$\{(S, ux^n v) \mid n \in \mathbb{N}\}$$

*is finite.*

This result (of which we do not know if it extends to rings) is in fact a reformulation of the theorem of McNaughton and Zalcstein on the Burnside problems for matrix semigroups (See Section 2). Furthermore, deciding if $S$ is of finite image is reduced to deciding if a certain matrix semigroup is finite, so it is decidable by a result of Jacob [25] and Mandel-Simon [36] (see [7, Corollary 4, p. 120]). In fact, almost all of the results of this paragraph are already implicitly given by Schützenberger [59] (whose article is difficult to read!): there he gives characterisations of the growth of rational power series over $\mathbb{Q}$ and of matrix semigroups (including growth zero = finiteness); his methods are used in [7, Chapter 6].

A related result of Simon [61] asserts that it is decidable if a regular language is limited: he reduces this problem to the decidability of the finiteness of a certain semigroup of matrices over some special semiring. For related results, see [14] and [7, Chapter 6]. A direct proof was given by Hashigushi [23].

From the results in Section 2 we deduce results on supports: the *support* of some rational power series $S$ is the language $L$, given as follows:

$$L = \{w \in A^* \mid (S, w) \neq 0\}.$$

It is well known that each regular language is a support. The family of supports is closed by all the usual operations (when $R$ is a subsemiring of $\mathbb{R}$), except complementation. The lack of this latter closure property, however, is completely explained by the following rationality criterion.

**Theorem 4.4** ([8]). *A language is rational if and only if both this language and its complement are supports.*

Here we assume that $R$ is a field. We do not know if it still holds when $R$ is a (commutative) ring. This result is, as far we know, impossible to prove directly, but it is a simple consequence of the criterion of Ehrenfeucht, Parikh and Rozenberg (which shows its strength). The above result is a particular case of the following conjecture.

**Conjecture 4.5** ([48]). *If two supports are disjoint languages, then they are rationally separated (i.e., there is a regular language containing one and not intersecting the other).*

This has some analogy with the 'density theorem' in algebraic geometry: if $P(x_1, \ldots, x_n)$, $Q(x_1, \ldots, x_n)$ are two commutative polynomials, let the series $S$, $T$ in $\mathbb{R}\langle\langle a_1, \ldots, a_n\rangle\rangle$ be defined by $(S, w) = P(x_1, \ldots, x_n)$ and $(T, w) = Q(x_1, \ldots, x_n)$, where $x_i$ is the number of occurrences of $a_i$ in the word $w$; then, $S$ and $T$ are rational. If their supports are disjoint, then, by the density theorem, $P = 0$ or $Q = 0$, one the two supports is empty: this is a special case of the above conjecture.

A partial positive answer to the above conjecture is given in [53], in the case of a one-letter alphabet (if $R$ is of charcteristic $\neq 0$; in characteristic 0, the stronger theorem of Skolem, Mahler and Lech holds (see [7, Theorem 4.4.1]).

We have seen that if a language has the weak permutation property, then it is Parikh-bounded. Thus we have the following result.

**Theorem 4.6** ([45]). *Each support is Parikh-bounded.*

We have proved this theorem in [45] when $R$ is a commutative ring. But it easily extends to commutative semirings, because, for such an $R$ and for any $n \times n$ matrices $m_1, \ldots, m_{2n}$ over $R$, one has

$$\sum_{\sigma \in \mathfrak{S}_{2n}, \sigma \text{ even}} m_{\sigma(1)} \cdots m_{\sigma(2n)} = \sum_{\sigma \text{ odd}} m_{\sigma(1)} \cdots m_{\sigma(2n)}$$

by the Amitsur–Levitzki theorem (see [5, Theorem 1.4.1]). Hence, one may conclude as in [45] that Theorem 4.6 holds.

Note that commutativity is really needed; by extending a construction of Sontag [63], it is possible to show the existence of a noncommutative *ring* $R$ such that *any* language is the support of some rational power series over $R$.

We conclude this paper by mentioning the following (hard to prove) conjecture, which asks for rationality.

**Conjecture 4.7** ([49, 51]). *Let $S$ be a formal power series with integer coefficients and $p$ a prime number; if $\sum_w p^{(S,w)} w$ is rational, then so are $S$ and $\sum_w p^{-(S,w)} w$.*

This is a particular case of a more general conjecture, involving unambiguous rational operations (see [49, 51] for details and solutions in particular cases; the one-letter case is a theorem of Pólya [43]).

## Appendix A. Definitions on words and languages

A *square-free word* is a word $w$ which has no factor of the form $xx$, for some nonempty word $x$.

An *iterating factor* of a language $L$ is a word $x$ such that for some words $u$ and $v$, the intersection

$ux^*v \cap L$ is infinite.

A *palindrome* is a word which is equal to its reverse image.

A language $L$ is *commutative* if whenever two words $w$, $w'$ have the same commutative image (i.e., for each letter, they have the same number of occurrences of this letter), then $w \in L$ is equivalent to $w' \in L$.

A language $L$ is *bounded* if for some words $u_1, \ldots, u_q$ one has

$$L \subset u_1^* \ldots u_q^*.$$

A word $w$ is *primitive* if whenever it is written $w = z^n$, we have that $n = 1$.

The *population function* $\delta : \mathbb{N} \to \mathbb{N}$ of a language $L$ is defined by

$$\delta(n) = |\{w \in L \mid |w| \leq n\}|.$$

It is *polynomially bounded* if it is bounded by some polynomial function.

A language $L$ is *limited* if for some $n$ one has

$$L^* = 1 \cup L \cup \cdots \cup L^n.$$

## Acknowledgment

Discussions with many people proved to be helpful. In particular, we want to thank R. McNaughton and B. Rozoy for several comments, J. Beauquier, L. Boasson and M. Latteux for advices on context-free languages, P.E. Schupp for a history of the original Burnside problem, and J. Sakarovitch for star-height problems.

## References

[1] S.I. Adjan, Burnside groups of odd exponent and irreducible systems of group identities, in: W.W. Boone, F.B. Cannonito and R.C. Lyndon, eds., *Word Problems* (North-Holland, Amsterdam, 1973) 19-38.

[2] S.I. Adjan, *The Burnside Problem and Identities in Groups* (Springer, Berlin, 1979).

[3] J.M. Autebert, J. Beauquier, L. Boasson and M. Latteux, Very small families of algebraic nonrational languages, in: R. Book, ed., *Formal Language Theory, Perspectives and Open Problems* (Academic Press, New York, 1980) 89-108.

[4] J. Beauquier, M. Blattner and M. Latteux, On commutative context-free languages, to appear.

[5] J. Berstel, *Transductions and Context-free Languages* (Teubner, Stuttgart, 1979).

[6] J. Berstel, Each iterated morphism yields a co-CFL, to appear.

[7] J. Berstel and C. Reutenauer, *Les Séries Rationnelles et Leurs Langages* (Masson, Paris, 1984).

[8] M. Blattner and M. Latteux, Parikh-bounded languages, in: *Lecture Notes in Computer Science* **115** (Springer, Berlin, 1981) 316-323.

[9] L. Boasson, Un critère de rationalité des langages algébriques, in: M. Nivat, ed. *Proc. 1st Internat. Coll. on Automata, Languages and Programming* (North-Holland, Amsterdam, 1973) 359-365.

[10] L. Boasson and A. Restivo, Une caractérisation des langages algébriques bornés, *RAIRO Inform.* **11** (1977) 203-205.

[11] J.L. Britton, The existence of infinite Burnside groups, in: W.W. Boone, F.B. Cannonito and R.C. Lyndon, eds., *Word Problems* (North-Holland, Amsterdam, 1973) 67-348.

[12] J.A. Brzozowski, K. Culik, II and A. Gabrielan, Classification of non-counting events, *J. Comput. System Sci.* **5** (1971) 41-53.

[13] W. Burnside, On an unsettled question in the theory of discontinuous groups, *Quart. J. Pure Appl. Math.* **33** (1902) 230-238.

[14] C. Choffrut, Sur les transductions reconnaissables, *RAIRO Inform.* **12** (1978) 203-218.

[15] A. Ehrenfeucht, D. Haussler and G. Rozenberg, On regularity of context-free languages, *Theoret. Comput. Sci.* **27** (1983) 311-332.

[16] A. Ehrenfeucht, R. Parikh and G. Rozenberg, Pumping lemmas for regular sets, *SIAM J. Comput.* **10** (1981) 536-541.

[17] S. Eilenberg, *Automata, Languages and Machines, Vols. A, B* (Academic Press, New York, 1974).

[18] J. Engelfriet and G. Rozenberg, A translational theorem for the class of E0L languages, *Inform. and Control* **50** (1981) 175-183.

[19] J. Goldstine, Substitution and bounded languages, *J. Comput. System Sci.* **6** (1972) 9-29.

[20] J.A. Green and D. Rees, On semigroups in which $x^r = x$, *Proc. Cambridge Philos. Soc.* **48** (1952) 35-40.

[21] M. Hall, Jr., Solution of the Burnside problem for exponent 6, *Proc. Nat. Acad. Sci. USA* **43** (1957) 751-753.

[22] M. Harrison, *Introduction to Formal Language Theory* (Addison-Wesley, Reading, MA, 1978).

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

[24] I.N. Herstein, Non commutative rings, *Carus Mathematical Monographs* (Wiley, New York, 1968).

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

[26] I. Kaplansky, *Fields and Rings* (University of Chicago Press, 1965).

[27] J. Kortelainen, Every commutative quasi-rational language is regular, to appear.

[28] G. Lallement, *Semigroups and Combinatorial Applications* (Wiley, New York, 1979).

[29] M. Latteux, Cônes rationnels commutatifs, *J. Comput. System Sci.* **18** (1979) 307-333.

[30] M. Latteux and J. Leguy, Une propriéte de la famille GRE, *Fundamentals of Computer Science* (Akademie Verlag, Berlin, 1979) 255-261.

[31] M. Latteux and G. Rozenberg, Commutative one-counter languages are regular, *J. Comput. System Sci.*, to appear.

[32] M. Latteux and G. Thierrin, On bounded context-free languages, *Elek. Inf. Kyb.* **20** (1984) 3-8.

[33] M. Lothaire, *Combinatorics on Words* (Addison-Wesley, Reading, MA, 1983).

[34] A. de Luca and A. Restivo, A finiteness condition for finitely generated semigroups, *Semigroup Forum* **28** (1984) 123-134.

[35] M.G. Main, W. Bucher and D. Haussler, Applications of an infinite co-CFL, *Proc. 12th ICALP*, 1985.

[36] A. Mandel and I. Simon, On finite semigroups of matrices, *Theoret. Comput. Sci.* **5** (1977) 101-111.

[37] H.A. Maurer and M. Nivat, Rational bijection of rational sets, *Acta Inform.* **13** (1980) 365-378.

[38] R. McNaughton and I. Zalcstein, The Burnside problem for semigroups, *J. Algebra* **34** (1975) 292-299.

[39] M. Morse and G. Hedlund, Unending chess, symbolic dynamics and a problem in semigroups, *Duke Math. J.* **11** (1944) 1-7.

[40] P.S. Novikov, *Dokl. Akad. Nauk SSSR* **127** (1959) 749-752 (in Russian).

[41] P.S. Novikov and S.I. Adjan, On infinite periodic groups, *Izv. Akad. Nauk SSSR, Ser. Mat.* **32** (1968), 212-214, 251-324, 709-731 (in Russian); translated in: *Math. USSR Izv.* **2** (1968) 209-236, 241-479, 665-685.

[42] A.U. Ol'shanskii, The Novikov-Adjan theorem, *Mat. Sb.(N.S.)* **118** (160) (1982) 203-235, 287 (in Russian); see also: *Math. Reviews* **83**'(1983).

[43] G. Pólya, Arithmetische Eigenschaften der Reihenentwicklungen rationaler Funktionen, *J. Reine Angew. Math.* **151** (1921).

[44] A. Restivo, Mots sans répétitions et langages rationnels bornés, *RAIRO Inform. Théor.* **11** (1977) 197-202.

[45] A. Restivo and C. Reutenauer, Some applications of a theorem of Shirshov to language theory, *Inform. and Control* **57** (1983) 205-213.

[46] A. Restivo and C. Reutenauer, Cancellation pumping and permutation in formal languages, in: *11th ICALP*, Lecture Notes in Computer Science **172** (Springer, Berlin, 1984) 414-422.

[47] A. Restivo and C. Reutenauer, On the Burnside problem for semigroups, *J. Algebra* **89** (1984) 102-104.

[48] A. Restivo and C. Reutenauer, On cancellation properties of languages which are support of rational power series, *J. Comput. System. Sci.* **29** (1984) 153-159.

[49] C. Reutenauer, Sur les séries de Polya en variable noncommutatives, *Fundamentals of Computation Theory* **2** (Akademie Verlag, Berlin, 1979) 391-396.

[50] C. Reutenauer, Séries formelles et algèbres syntactiques, *J. Algebra* **66** (1980) 448-483.

[51] C. Reutenauer, Séries rationnelles et algèbres syntactiques, Thèse de Doctorat d'Etat, Université Paris 6, 1980.

[52] C. Reutenauer, A new characterization of the regular languages, *Lecture Notes in Computer Science* **115** (Springer, Berlin, 1981) 177-183.

[53] C. Reutenauer, Sur les éléments inversibles de l'algèbre de Hadamard des séries rationnelles, *Bull. Soc. Math. France* **110** (1982) 225-232.

[54] L.H. Rowen, *Polynomial Identities in Ring Theory* (Academic Press, New York, 1980).

[55] J. Sakarovitch, Monoïdes syntactiques et langages algébriques, Thèse 3ème cycle, Université Paris 7, 1976.

[56] A. Salomaa and M. Soittola, *Automata Theoretic Aspects of Formal Power Series* (Springer, Berlin, 1978).

[57] I. Schur, Über Gruppen periodischer Substitutionen, *Sitzungsber. Preuss. Akad. Wiss. Math.-Natur. Kl.* (1911) 619-627.

[58] M.P. Schützenberger, On the definition of a family of automata, *Inform. and Control* **4** (1961) 245-270.

[59] M.P. Schützenberger, Finite counting automata, *Inform. Control* **5** (1962) 91-107.

[60] A.I. Shirshov, On rings with identity relations, *Mat. Sb.* **43** (85) (1957) 277-283 (in Russian).

[61] I. Simon, Limited subsets of a free monoid, *Proc. 19th Ann. Symp. on Foundations of Computer Science* (1978) 143-150.

[62] I. Simon, Conditions de finitude pour des semi-groupes, *C.R. Acad. Sci. Paris Sér. A* **290** (1980) 1081-1082.

[63] E. Sontag, On some questions of rationality and decidability, *J. Comput. System Sci.* **11** (1975) 375-385.

[64] H. Straubing, The Burnside problem for semigroups of matrices, in: L.J. Cummings, ed., *Combinatorics on Words, Progress and Perspectives* (Academic Press, New York, 1983) 279-295.