The expressive power of the shuffle product*

Jean Berstel[1], Luc Boasson[2], Olivier Carton[2],
Jean-Eric Pin[2], Antonio Restivo[3]

Jean.Berstel@univ-mlv.fr, Luc.Boasson@liafa.jussieu.fr,
Olivier.Carton@liafa.jussieu.fr, Jean-Eric.Pin@liafa.jussieu.fr,
restivo@math.unipa.it

August 7, 2010

Abstract

There is an increasing interest in the shuffle product on formal languages, mainly because it is a standard tool for modeling process algebras. It still remains a mysterious operation on regular languages.

Antonio Restivo proposed as a challenge to characterize the smallest class of languages containing the singletons and closed under Boolean operations, product and shuffle. This problem is still widely open, but we present some partial results on it. We also study some other smaller classes, including the smallest class containing the languages composed of a single word of length 2 which is closed under Boolean operations and shuffle by a letter (resp. shuffle by a letter and by the star of a letter). The proof techniques have both an algebraic and a combinatorial flavour.

Introduction

The study of classes of regular languages closed under shuffle is a difficult problem, partly motivated by its applications to the modeling of process algebras [2] and to program verification. Significant progress has been made over the last decade in the study of the shuffle operation. First, Ésik and Simon [7] have completed the classification of varieties of languages closed under shuffle. It was known [9] that the commutative varieties of languages closed under shuffle correspond to the varieties of commutative monoids whose groups belong to a given variety of commutative groups. Ésik and Simon proved that, apart from the variety of all regular languages, no other variety of languages is closed under shuffle. In particular, the variety of commutative languages is the largest proper variety of languages closed under shuffle. It was also proved that there is a largest proper positive variety of languages closed under shuffle and that this variety is decidable [3, 4].

[1]Corresponding author. Institut Gaspard-Monge, Université Paris-Est Marne-la-Vallée.
[2]LIAFA, Université Paris Diderot-Paris 7 and CNRS.
[3]Dipartimento di Matematica e Applicazioni, Università di Palermo.
*The authors acknowledge support from the AutoMathA programme of the European Science Foundation.

A few years ago, the fifth author proposed as a challenge to study the smallest class of languages $\mathcal{C}$ containing the singletons and closed under Boolean operations, product and shuffle. Let us call *intermixed* the languages of this class. We show that intermixed languages are closed under quotients, but they are not closed under inverses of morphisms. Therefore, they do not form a variety of languages and the result of Ésik and Simon cannot be applied. However, intermixed languages are closed under inverses of length-decreasing morphisms and under quotients. Consequently, they form a d-variety, in the sense of [14, 6]. This fact is interesting since, by a result of Kunc [8] (see also [12]), d-varieties can be characterised by a certain type of identities, called *d-identities*. The formal definition of d-identities, as well as all definitions and background used in this paper, are presented in Section 1.

We give in Section 2 two d-identities satisfied by all intermixed languages, namely $x^{\omega+1} = x^\omega$ and $(x^\omega y^\omega)^{\omega+1} = (x^\omega y^\omega)^\omega$. This proves the main result of this paper: intermixed languages form a proper subclass of the class of regular languages, since the language $(aa)^*$ does not satisfy the first identity. Unfortunately, we do not know whether our two identities suffice to characterise the intermixed languages and hence the decidability of this question remains open.

Our two identities give, in a sense, an upper approximation of the class of intermixed languages. In order to get lower approximations, we investigate some subclasses obtained by restricting the use of the shuffle operation. We briefly study the case of commutative languages. Then we set aside this case by considering classes containing at least one noncommutative language. In fact, for technical reasons which are partly justified by Proposition 3.1, our classes will always contain the languages of the form $\{ab\}$ where a and b are two distinct letters of the alphabet.

We first consider in Section 3 the smallest class of languages $\mathcal{C}_0$ containing the languages of the form $\{ab\}$ and closed under Boolean operations and shuffle by a letter, a very drastic restriction on the shuffle operation. These languages are called *almost star-free commutative* for the following reason: a language L belongs to $\mathcal{C}_0$ if and only if there exists a star-free commutative language C such that the symmetric difference $L \triangle C$ is finite. This rather small class is closed under inverses of length-increasing morphisms and thus forms an i-variety. We give explicitly a finite set of i-identities which characterizes this class. It follows in particular that one can decide whether a given regular language is almost star-free commutative.

Increasing the power of the shuffle operation, we next consider two classes $\mathcal{C}_1$ and $\mathcal{C}_2$. The class $\mathcal{C}_1$ is defined as the smallest Boolean algebra of languages containing $\mathcal{C}_0$ and closed under the operations $L \mapsto L \sqcup a$ (shuffle by a letter) and $L \mapsto L \sqcup a^*$ (shuffle by the star of a letter), where a is a letter. The class $\mathcal{C}_2$ is the closure of $\mathcal{C}_0$ under shuffle. We call *jumbled* the languages of $\mathcal{C}_1$ and *shuffled* the languages of $\mathcal{C}_2$. We prove that all these classes are d-varieties and that $\mathcal{C}_0$ is a proper subclass of $\mathcal{C}_1$. These results are synthesized in the tables below. The first table summarizes the definition of our four classes.

Closed under	$L \sqcup a$	$L \sqcup a,\ L \sqcup a^*$	$L \sqcup L'$	$L \sqcup L', LL'$
Boolean operations	$\mathcal{C}_0$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}$

The second table gathers the known properties of each class.

Class	Languages	Type	Known identities	Decidable
$\mathcal{C}_0$	Almost star-free commutative	i-variety	$x^{\omega+1} = x^\omega$, $x^\omega y = yx^\omega$ $x^\omega yz = x^\omega zy$	Yes
$\mathcal{C}_1$	Jumbled	d-variety		?
$\mathcal{C}_2$	Shuffled	d-variety		?
$\mathcal{C}$	Intermixed	d-variety	$x^{\omega+1} = x^\omega$ $(x^\omega y^\omega)^{\omega+1} = (x^\omega y^\omega)^\omega$	?

Let us clarify two issues concerning the identities of the fourth column of the second table. First, the type of these identities depends on the nature of the corresponding variety. In particular, the given identities for $\mathcal{C}_0$ are i-identities, while those given for $\mathcal{C}$ are d-identities. Secondly, the set of i-identities given for $\mathcal{C}_0$ is complete, that is, a language satisfies these i-identities if and only if it belongs to $\mathcal{C}_0$. In contrast, it is an open problem to know whether our given set of d-identities for $\mathcal{C}$ is complete.

We give several partial results on jumbled languages in Section 4. In particular, we show that every regular language can be written as the inverse image, under a morphism, of a jumbled language. By contrast, we have almost nothing to say about the class $\mathcal{C}_2$ of shuffled languages, which is the smallest class of languages containing $\mathcal{C}_0$ and closed under Boolean operations and shuffle. We know very little about this class, apart from the fact that it is a d-variety of languages (the proof is similar to that of Theorem 2.1). In particular, we failed to prove our conjectures that $\mathcal{C}_1$ is strictly contained in $\mathcal{C}_2$ and that $\mathcal{C}_2$ is strictly contained in $\mathcal{C}$. A possible candidate to separate $\mathcal{C}_1$ from $\mathcal{C}$ is the language A^*abbaA^*, but we have no proof that this language is not jumbled.

One possible method to find an intermixed language which is not shuffled would be to find some d-identities satisfied by all shuffled languages. Proposition 1.9, which gives *ordered* identities which are, in a sense, stable under shuffle could be a useful tool. Unfortunately, we were not able to derive from these ordered d-identities a nonordered identity stable under shuffle.

Finally, it is interesting to compare our four classes with the class of star-free languages, which is the smallest class of languages containing the singletons and closed under Boolean operations and product. Clearly every almost star-free commutative language is star-free and every star-free language is intermixed. Further, it follows from Proposition 4.10 that some jumbled languages are non star-free and hence some shuffled languages are non star-free. The question remains whether every star-free language is shuffled or even jumbled. We conjecture that the answer to these questions is negative. For instance, the language A^*abbaA^*, that we believed to be nonjumbled, is star-free.

1 Definitions and background

In this paper, A denotes a finite alphabet and A^* is the free monoid on A. The empty word is denoted by 1. We usually identify a singleton language $\{u\}$ with the word u itself.

1.1 Languages

Let L be a language over A and let u be a word. The *left quotient* of L by u is the language $u^{-1}L = \{v \in A^* \mid uv \in L\}$. The *right quotient* Lu^{-1} is defined in a symmetrical way. Given a language $L \subset A^*$, we write L^c for the complement $A^* \setminus L$ of L.

A *morphism* between two free monoids A^* and B^* is a map $\varphi : A^* \to B^*$ such that, for all $u, v \in A^*$, $\varphi(uv) = \varphi(u)\varphi(v)$. This condition implies in particular that $\varphi(1) = 1$. We say that φ is *length-preserving* (p) if, for each $u \in A^*$, the words u and $\varphi(u)$ have the same length. Equivalently, φ is length-preserving if, for each letter $a \in A$, $\varphi(a) \in B$. Similarly, φ is *length-decreasing* (d) if the image of each letter is either a letter or the empty word, and *length-increasing* (i) if the image of each letter is a nonempty word.

A *class of languages* is a correspondence $\mathcal{V}$ which associates with each alphabet A a set $\mathcal{V}(A^*)$ of regular languages over A. A *variety of languages* is a class of languages closed under Boolean operations (union, intersection and complement), left and right quotients and inverses of morphisms. The weaker notions of *p-variety* [*d-variety*, *i-variety*] are obtained by relaxing the latter condition [14, 6]: only closure under inverses of p- [d-, i-] morphisms is required.

The *shuffle product* (or simply *shuffle*) of two languages L_1 and L_2 over A is the language

$$L_1 \sqcup\!\sqcup L_2 = \{w \in A^* \mid w = u_1 v_1 \cdots u_n v_n \text{ for some words } u_1, \ldots, u_n \\ v_1, \ldots, v_n \text{ of } A^* \text{ such that } u_1 \cdots u_n \in L_1 \text{ and } v_1 \cdots v_n \in L_2\}\,.$$

The shuffle product defines a commutative and associative operation over the set of languages over A.

Two special cases of shuffle product play an important role in this paper. These are the operations $L \mapsto L \sqcup\!\sqcup a$ and $L \mapsto L \sqcup\!\sqcup a^*$ where a is a letter. The first will be referred to as *shuffle by a letter*, and the second as *shuffle by the star of a letter*.

Recall that Boolean operations commute with quotients and inverses of morphisms. There are also well known formulas for computing right and left quotients of the product (or the shuffle) of two languages. We shall use freely these standard commutation rules and two commutation rules which are specific to inverses of length-decreasing morphisms.

Proposition 1.1 *Let L_1 and L_2 be languages over A and let $\varphi : B^* \to A^*$ be a length-decreasing morphism. Then the following formulas hold:*

$$\varphi^{-1}(L_1 L_2) = \varphi^{-1}(L_1)\varphi^{-1}(L_2)\,, \tag{1}$$

$$\varphi^{-1}(L_1 \sqcup\!\sqcup L_2) = \varphi^{-1}(L_1) \sqcup\!\sqcup \varphi^{-1}(L_2)\,. \tag{2}$$

Proof. Formula (1) holds because φ is length-decreasing. Let us prove (2). Since φ^{-1} commutes with union, it suffices to establish the formula

$$\varphi^{-1}(u_1 \sqcup\!\sqcup u_2) = \varphi^{-1}(u_1) \sqcup\!\sqcup \varphi^{-1}(u_2) \tag{3}$$

when u_1 and u_2 are words of A^*.

Let $w \in \varphi^{-1}(u_1 \sqcup u_2)$. Then there exist $x_1, \ldots, x_n$, $y_1, \ldots, y_n$ such that $u_1 = x_1 \cdots x_n$ and $u_2 = y_1 \cdots y_n$ and $w \in \varphi^{-1}(x_1 y_1 \cdots x_n y_n)$. In view of Formula (1), $w \in \varphi^{-1}(x_1)\varphi^{-1}(y_1) \cdots \varphi^{-1}(x_n)\varphi^{-1}(y_n)$. Since

$$\varphi^{-1}(x_1) \cdots \varphi^{-1}(x_n) = \varphi^{-1}(x_1 \cdots x_n) = \varphi^{-1}(u_1)$$

and

$$\varphi^{-1}(y_1) \cdots \varphi^{-1}(y_n) = \varphi^{-1}(u_2)\,,$$

the word w is in $\varphi^{-1}(u_1) \sqcup \varphi^{-1}(u_2)$.

Conversely, if $w \in \varphi^{-1}(u_1) \sqcup \varphi^{-1}(u_2)$, then $w \in v_1 \sqcup v_2$ for some $v_1 \in \varphi^{-1}(u_1)$ and $v_2 \in \varphi^{-1}(u_2)$. It follows that $\varphi(w) \in \varphi(v_1) \sqcup \varphi(v_2) = u_1 \sqcup u_2$. This proves (3) and the proposition. □

1.2 Syntactic monoids and varieties

The syntactic monoid of a language is an algebraic invariant which plays a crucial role in the study of regular languages. We review its definition and basic properties in this short section.

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

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

The monoid $M = A^*/\sim_L$ is the *syntactic monoid* of L and the natural morphism $\eta : A^* \to M$ is called the *syntactic morphism* of L. The set $P = \eta(L)$ is called the *syntactic image* of L. Note that L is saturated for $\sim_L$, which means that $\eta^{-1}(P) = L$. It is a well-known fact that a language is regular if and only if its syntactic monoid is finite.

An ordered monoid is a monoid equipped with a stable partial order relation, usually denoted by $\leqslant$. The *syntactic preorder* of a language L is the relation $\leqslant_L$ over A^* defined by $u \leqslant_L v$ if and only if, for every $x, y \in A^*$,

$$xvy \in L \implies xuy \in L\,.$$

It is easy to see that $\leqslant_L$ is a partial preorder on A^*, whose associated equivalence relation is the *syntactic congruence* of L. Further, the syntactic preorder of L induces a partial order on M which makes it an ordered monoid as follows. Given $u, v \in M$, one has $u \leqslant v$ if and only if, for all $x, y \in M$,

$$xvy \in P \implies xuy \in P\,.$$

Here $P = \eta(L)$ is the syntactic image of L. The ordered monoid $(M, \leqslant)$ is called the *syntactic ordered monoid* of L. We write $\leqslant_P$ instead of $\leqslant$ when we want to emphasize the subset P of M.

For each finite semigroup S, there exists an integer n such that, for each $s \in S$, s^n is idempotent. The least integer satisfying this property is called the *exponent* of S and is often denoted by ω. By extension, the *exponent* of a regular language L of A^* is the smallest integer n such that, for all $u \in A^*$, $u^n \sim_L u^{2n}$. A finite monoid M of exponent ω is *aperiodic* if, for all $x \in M$, $x^\omega = x^{\omega+1}$.

A *variety of finite monoids* [semigroups] is a class of finite monoids [semigroups] closed under taking submonoids [subsemigroups], morphic images and finite direct products. If $\mathbf{V}$ is a variety of finite monoids, denote by $\mathcal{V}(A^*)$ the set of regular languages of A^* whose syntactic monoid belongs to $\mathbf{V}$. The correspondence $\mathbf{V} \mapsto \mathcal{V}$ associates with each variety of finite monoids a variety of languages. Conversely, to each variety of languages $\mathcal{V}$, we associate the variety of finite monoids generated by the syntactic monoids of the languages of $\mathcal{V}$. Eilenberg's variety theorem [5] states that these two correspondences define mutually inverse bijective correspondences between varieties of finite monoids and varieties of languages. For instance, Schützenberger's theorem states that star-free languages correspond to aperiodic monoids.

There is an analogous correspondence between i-varieties of languages and varieties of finite semigroups, obtained by associating to each language L of A^* the syntactic semigroup of the language $L \cap A^+$. For instance, finite or cofinite languages correspond to nilpotent semigroups.

To complete this section, let us describe the smallest nontrivial variety closed under shuffle. We denote by $[u]$ the *commutative closure* of a word u, which is the set of words commutatively equivalent to u. For instance, $[aab] = \{aab, aba, baa\}$. A language L is *commutative* if, for every word $u \in L$, $[u]$ is contained in L. Equivalently, a language is *commutative* if its syntactic monoid is commutative. A description of the class of star-free commutative languages is given in [10, Chapter 2, Proposition 3.14]. Let us give a variation of this result using the shuffle operation.

Proposition 1.2 *A language of A^* is star-free commutative if and only if it is a finite union of languages of the form $[u] \sqcup B^*$ where u is a word and B is a subset of A.*

Proof. In one direction, it suffices to observe that if F is a finite commutative language and B is a subset of A, then the syntactic monoid of $F \sqcup B^*$ is commutative and aperiodic.

Consider now a commutative star-free language L and let $\varphi : A^* \to M$ be its syntactic morphism. Our aim is to prove that L can be written as a finite union of languages of the form $[u] \sqcup B^*$. Let $P = \varphi(L)$ and let N be the exponent of M. Since $L = \bigcup_{m \in P} \varphi^{-1}(m)$, it suffices establish the result for $L = \varphi^{-1}(m)$, where m is an element of M. We claim that

$$L = \bigcup_{u \in F} [u] \sqcup B^*$$

where $B = \{a \in A \mid m\varphi(a) = m\}$ and

$$F = \{u \in A^* \mid |u| \leqslant N|A|,\ \varphi(u) = m \text{ and for all subwords } v \text{ of } u,\ \varphi(v) \neq m\}.$$

If $u \in F$ and $w \in [u] \sqcup B^*$, then $w \in u' \sqcup v$ for some $u' \in [u]$ and some $v \in B^*$. Since M is commutative, it follows that $\varphi(w) = \varphi(u)\varphi(v) = m\varphi(v) = m$. Thus $w \in L$. Conversely, let $w \in L$ and let u be a minimal subword of w in L. By construction, $\varphi(u) = m$ and for all subwords v of u, $\varphi(v) \neq m$. Further, if $|u| > N|A|$, then $|u|_a > N$ for some letter $a \in A$. Therefore, u can be written as $u_1 a u_2$ for some words u_1, u_2 such that $|u_1 u_2|_a \geqslant N$. Since M is commutative and $\varphi(a^N) = \varphi(a^{N+1})$, it follows that $\varphi(u_1 u_2) = \varphi(u)$, a contradiction with the definition of u. Thus $|u| \leqslant N|A|$ and $u \in F$.

Let v be a word such that $w \in u \sqcup v$. Since M is commutative, $\varphi(w) = \varphi(u)\varphi(v)$, that is $m = m\varphi(v)$. Since M is aperiodic and commutative, it is $\mathcal{J}$-trivial and thus $m\varphi(a) = m$ for each letter a of v. In other words, $v \in B^*$ and $w \in [u] \sqcup B^*$. □

Corollary 1.3 *The star-free commutative languages form a variety of languages, which is the smallest variety of languages closed under shuffle. It is also the smallest class of languages closed under Boolean operations and under shuffle by a letter.*

Proof. The first part of the statement is proved in [9]. Let $\mathcal{F}$ be the smallest class of languages closed under Boolean operations and the operation $L \mapsto L \sqcup a$, where a is a letter. It just remains to prove that $\mathcal{F}$ contains all star-free commutative languages.

Since $\mathcal{F}(A^*)$ is a Boolean algebra, it contains A^* and thus, for each letter $a \in A$, the language $A^*aA^* = A^* \sqcup a$. Therefore, for each subset B of A, $\mathcal{F}(A^*)$ also contains also the languages $A^*BA^* = \cup_{a\in B} A^*aA^*$ and $B^* = A^* \setminus A^*B^cA^*$. In particular, it contains the language $\{1\} = \emptyset^*$. Now observe that if $u = a_1 \cdots a_n$, then $[u] = \{1\} \sqcup a_1 \sqcup \cdots \sqcup a_n$. Thus $\mathcal{F}$ contains the languages of the form $[u]$ and by Proposition 1.2, it contains all star-free commutative languages. □

1.3 Equations and identities

The formal approach to identities requires the introduction of profinite words. The definition of those and appropriate references can be found in [1, 11]. However, the weaker notion of ω-term will suffice to state and prove the results of this paper. For the sake of completeness, let us just mention that Propositions 1.6 and 1.7 below can be readily extended to profinite words.

An *ω-term* on an alphabet A is built from the letters of A using the usual concatenation product and the unary operator $x \to x^\omega$. For instance, if $A = \{a, b, c\}$, abc, a^ω and $((ab^\omega c)^\omega ab)^\omega$ are examples of ω-terms. The symbol ω plays an abstract role similar to the star symbol in a regular expression and should not be interpreted as denoting infinite iteration. Two ω-terms can be concatenated to form their product. This product is associative and extends the usual product on words. Further, if x is an ω-term, x and x^ω commute, that is, $xx^\omega = x^\omega x$. This ω-term is often denoted by $x^{\omega+1}$, and more generally, we write $x^{\omega+n}$ for $x^n x^\omega$ or $x^\omega x^n$. Finally, $1^\omega = 1$ and for each ω-term, $x^\omega x^\omega = x^\omega$ and $(x^\omega)^\omega = x^\omega$.

Morphisms between free monoids extend to ω-terms in a natural way. For instance, if $\varphi : \{a, b, c\}^* \to \{a, b\}^*$ is defined by $\varphi(a) = ab$, $\varphi(b) = ba$ and $\varphi(c) = 1$, then $\varphi(((ab^\omega c)^\omega ab)^\omega) = ((\varphi(a)\varphi(b)^\omega c)^\omega \varphi(a)\varphi(b))^\omega = ((ab(ba)^\omega)^\omega abba)^\omega$.

Morphisms from a free monoid into a finite monoid M also extend to ω-terms in a very simple way by interpreting the symbol ω as the exponent of M. It follows that if $\varphi : A^* \to M$ is a morphism and x is an ω-term, then $\varphi(x^\omega)$ is equal to $\varphi(x)^\omega$, the unique idempotent of the subsemigroup of M generated by $\varphi(x)$.

We now consider ordered equations of the form $u \leqslant v$, where u and v are two ω-terms. This kind of equations is mainly used in Proposition 1.6. This

proposition avoids to duplicate proofs unnecessarily. Equations of the form $u = v$ are then just a shortcut for $u \leqslant v$ and $v \leqslant u$. Let L be a regular language of A^*, let $(M, \leqslant)$ be its syntactic ordered monoid and let $\eta : A^* \to M$ be its syntactic morphism. We say that L *satisfies the equation* $u \leqslant v$ if $\eta(u) \leqslant \eta(v)$.

Denote by $\mathcal{T}$ one of the following types of morphisms: all morphisms, all p-morphisms, all d-morphisms or all i-morphisms. Let now u and v be two ω-terms on the alphabet B. We say that L *satisfies the* $\mathcal{T}$*-identity* $u \leqslant v$ if, for all $\mathcal{T}$-morphisms $\gamma : B^* \to A^*$, it satisfies the equation $\gamma(u) \leqslant \gamma(v)$. As promised, we illustrate our abstract definition by two examples.

Proposition 1.4 *Let L be a regular language over A and let n be its exponent. Then L satisfies the identity* [*p-, d-, i-identity*] $x^{\omega+1} \leqslant x^\omega$ *if and only if, for every word* [*letter, word of length* $\leqslant 1$*, nonempty word*] $u \in A^*$*, one has* $u^{n+1} \leqslant_L u^n$.

Proof. Let $\gamma : B^* \to A^*$ be a morphism and let $u = \gamma(x)$. When γ ranges over the set of all morphisms [p-, d-, i-morphisms], u ranges over the set of all words [letters, words of length $\leqslant 1$, nonempty words]. Since $\gamma(x)^\omega = u^\omega$ and $\eta(u^\omega) = \eta(u)^n$, the equation $u^{\omega+1} \leqslant u^\omega$ is satisfied if and only if $\eta(u)^{n+1} \leqslant \eta(u)^n$ or, equivalently, $u^{n+1} \leqslant_L u^n$. □

The proof of the next result is similar and is therefore omitted. Both results are immediate consequences of the general definition of identities and $\mathcal{T}$-identities [1, 8, 14].

Proposition 1.5 *Let L be a regular language over A and let n be its exponent. Then L satisfies the identity* [*p-, d-, i-identity*] $(x^\omega y^\omega)^{\omega+1} = (x^\omega y^\omega)^\omega$ *if and only if, for every pair of words* [*letters, words of length* $\leqslant 1$*, nonempty words*] $(u, v) \in A^* \times A^*$*, one has* $(u^n v^n)^{n+1} \sim_L (u^n v^n)^n$.

As one can see from these examples, the symbols occurring in the equations can be considered as variables. These variables are interpreted as words of length depending on the class of morphisms $\mathcal{T}$, according to the table below.

Class of morphisms	**Identity type**	**Interpretation of variables**
all morphisms	identity	words
length preserving morphisms	p-identity	words of length 1
length increasing morphisms	i-identity	words of length $\geqslant 1$
length decreasing morphisms	d-identity	words of length $\leqslant 1$

Note that if L satisfies a $\mathcal{T}$-identity $u \leqslant v$ where u and v are ω-terms on the alphabet B, then for all ω-terms x, y, it satisfies the $\mathcal{T}$-identity $xuy \leqslant xvy$.

Proposition 1.6 *Let u be an ω-term. Then the equations $u^{\omega+1} \leqslant u^\omega$ and $u^{\omega+1} = u^\omega$ are equivalent on regular languages.*

Proof. Let L be a regular language of exponent n, let $(M, \leqslant)$ be its syntactic ordered monoid and let $\eta : A^* \to M$ be its syntactic morphism. We claim

that if L satisfies the equation $u^{\omega+1} \leqslant u^\omega$, then it also satisfies the equation $u^{\omega+1} = u^\omega$. If L satisfies $u^{\omega+1} \leqslant u^\omega$, then by induction it also satisfies the equations

$$u^{\omega+n} \leqslant \cdots \leqslant u^{\omega+1} \leqslant u^\omega .$$

Since n is the exponent of L, one gets

$$\eta(u^{\omega+n}) = \eta(u^\omega)\eta(u)^n = \eta(u)^n\eta(u)^n = \eta(u)^n = \eta(u^\omega) .$$

Now the relations

$$\eta(u^\omega) = \eta(u^{\omega+n}) \leqslant \eta(u^{\omega+1}) \leqslant \eta(u^\omega)$$

show that $\eta(u^{\omega+1}) = \eta(u^\omega)$, which proves the claim and the proposition. □

We conclude this section by proving three stability results. The first one states that the class of languages satisfying an equation of the form $u^{\omega+1} \leqslant u^\omega$ is stable under product. The second and the third results assert that the class of languages satisfying the d-identity $x^{\omega+1} \leqslant x^\omega$ (respectively $x^\omega y x^\omega y \leqslant x^\omega yy$) is stable under shuffle.

Proposition 1.7 *Let u be an ω-term of A^*. If two regular languages satisfy the equation $u^{\omega+1} \leqslant u^\omega$, then their product also satisfies this equation.*

Proof. Let L_1 and L_2 be languages of A^* satisfying the equation $u^{\omega+1} \leqslant u^\omega$ and let L be their product. Let n be the least common multiple of the exponents of the languages L_1, L_2 and L. By Proposition 1.4, it suffices to prove that for every word $u \in A^*$, $u^{n+1} \leqslant_L u^n$. Suppose that $xu^n y \in L$. Since $u^n \sim_L u^{2n}$, one has $xu^{2n}y \in L$ and thus $xu^{2n}y = u_1u_2$ for some $u_1 \in L_1$ and $u_2 \in L_2$. It follows that one of the words u_1 or u_2 contains u^n as a factor. Since the two cases are symmetrical, we may assume that $u_1 = xu^n z$ for some $z \in A^*$. It follows that $xu^{n+1}z \in L_1$, since L_1 satisfies the identity $u^{\omega+1} \leqslant u^\omega$. Thus $xu^{2n+1}y \in L$ and finally $xu^{n+1}y \in L$ since $u^{2n} \sim_L u^n$. Therefore L satisfies the equation $u^{\omega+1} \leqslant u^\omega$. □

Proposition 1.8 *If two regular languages satisfy the d-identity $x^{\omega+1} \leqslant x^\omega$, then their shuffle also satisfies this d-identity.*

Proof. Let L_1 and L_2 be languages of A^* satisfying the d-identity $x^{\omega+1} \leqslant x^\omega$ and let $L = L_1 \sqcup\!\sqcup L_2$. Let n be the least common multiple of the exponents of the languages L_1, L_2 and L. According to Proposition 1.4, it suffices to prove that, for each word u of length 0 or 1, one has $u^{n+1} \leqslant_L u^n$. The result is obvious if u is the empty word and thus we may assume that $u = a$ for some letter $a \in A$.

Suppose that $xa^n y \in L$ for some words $x, y \in A^*$. Since $a^n \sim_L a^{2n}$, one has $xa^{2n}y \in L$ and thus $xa^{2n}y \in u_1 \sqcup\!\sqcup u_2$ for some $u_1 \in L_1$ and $u_2 \in L_2$. It follows that one of the words u_1 or u_2 contains a^n as a factor. If, for instance, $u_1 = ra^n s$ for some $r, s \in A^*$, then $ra^{n+1}s \in L_1$ since L_1 satisfies the d-identity $a^{\omega+1} \leqslant a^\omega$. It follows that $xa^{2n+1}y \in L$ and finally $xa^{n+1}y \in L$ since $a^{2n} \sim_L a^n$. Thus L satisfies the d-identity $a^{\omega+1} \leqslant a^\omega$. □

Proposition 1.9 *If two regular languages satisfy the d-identity $x^\omega yx^\omega y \leqslant x^\omega yy$ (resp. $yx^\omega yx^\omega \leqslant yyx^\omega$), then their shuffle also satisfies this identity.*

Proof. By symmetry, it suffices to prove the first identity. Let L_1 and L_2 be languages satisfying this identity and let $L = L_1 \sqcup\!\sqcup L_2$. Let n be the least common multiple of the exponents of the languages L_1, L_2 and L. According to Proposition 1.4, it suffices to prove that, for all words u, v of length 0 or 1, one has $u^n vu^n v \leqslant_L u^n vv$. The result is obvious if u or v is the empty word and thus we may assume that $u = a$ and $v = b$ for some letters $a, b \in A$.

Suppose that $xa^nbby \in L$ for some words $x, y \in A^*$. Since $a^n \sim_L a^{2n}$, the word $u = xa^{2n}bby$ also belongs to L and thus $u \in u_1 \sqcup\!\sqcup u_2$ for some $u_1 \in L_1$ and $u_2 \in L_2$.

First assume that each of the words u_1 and u_2 contains exactly one of the two letters b. Since u contains at least $2n$ occurrences of a on the left of the two letters b, the word a^nb is a factor of either u_1 or u_2. Without loss of generality, we may assume that a^nb is a factor of u_1, as depicted in Figure 1. In this diagram, the letters of u_1 are represented in white, the letters of u_2 in grey and the factors of u in which letters from u_1 and u_2 may occur simultaneously are represented in light grey.

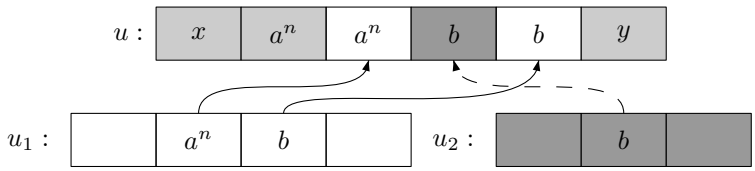

Figure 1: $u \in u_1 \sqcup\!\sqcup u_2$

Since n is a multiple of the exponent of L_1, iterating a^n in u_1 produces a word u_1' of L_1. One can also insert this new factor in u, as indicated in Figure 2, to obtain the word $u' = xa^{2n}ba^nby$ as a shuffle of u_1' and u_2. Thus, in this case, $xa^{2n}ba^nby$ belongs to L. Finally, since $a^n \sim_L a^{2n}$, $xa^nba^nby \in L$.

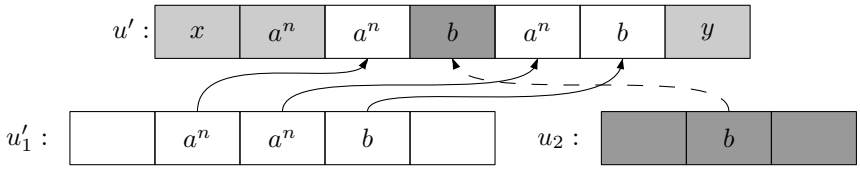

Figure 2: Inserting a^n.

Suppose now that one of the words u_1 or u_2, say u_1, contains the two occurrences of b. Out of the $2n$ occurrences of a preceding the two letters b, at least n originate from the same word. First assume that this word is u_2, as illustrated in Figure 3.

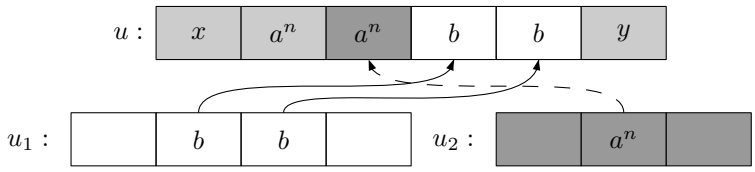

Figure 3: Case a^n in u_2

Since n is a multiple of the exponent of L_2, iterating a^n in u_2 produces a word u_2' of L_2. One can also insert this new factor in u, as indicated in Figure 4, to obtain the word $xa^{2n}ba^nby$ as a shuffle of u_1 and u_2'. Thus, in this case again, xa^nba^nby belongs to L.

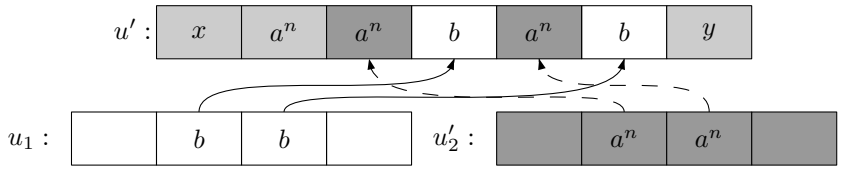

Figure 4: Inserting a^n again...

Finally, if at least n occurrences of a originate from u_1, then a^nbb is a factor of u_1, and since L_1 satisfies the equation $a^\omega ba^\omega b \leqslant a^\omega bb$, the word obtained from u_1 by replacing this factor by a^nba^nb is still in L_1. It follows, once again, that xa^nba^nby belongs to L. This exhausts all cases and concludes the proof. □

2 Intermixed languages

By definition, the class $\mathcal{C}$ of *intermixed languages* is the smallest class of languages containing the singletons $\{1\}$ and $\{a\}$, for each letter a, and closed under Boolean operations, product and shuffle. Let us show immediately that these properties entail two other closure properties.

Theorem 2.1 *Intermixed languages form a d-variety of languages.*

Proof. We proceed in four steps. After a preliminary step, we show that $\mathcal{C}$ is closed under quotients, then that it is closed under inverses of length-preserving morphisms and finally under inverses of length-decreasing morphisms.

Preliminary step. We show that, for each alphabet A, $\mathcal{C}(A^*)$ contains the languages B and B^*, for each subset B of A. The first property is obvious, since $B = \cup_{a\in B}\{a\}$. For the second one, it suffices to prove that the complement of B^* is in $\mathcal{C}(A^*)$. This complement is equal to $A^*(A \setminus B)A^*$ and since $\mathcal{C}(A^*)$ is closed under product, it belongs to $\mathcal{C}(A^*)$.

First step. Let $\mathcal{C}'$ be the class of all languages L of $\mathcal{C}$ such that, for each letter a, $a^{-1}L$ and La^{-1} are in $\mathcal{C}$. Clearly, the singletons $\{1\}$ and $\{b\}$, for each letter b, are in $\mathcal{C}'$. Further, standard commutation rules show that $\mathcal{C}'$ is closed under

Boolean operations, product and shuffle. Therefore $\mathcal{C}'$ contains $\mathcal{C}$ and thus $\mathcal{C}$ is closed under quotient by a letter. It follows by induction that $\mathcal{C}$ is closed under quotient.

Second step. Let $\mathcal{F}$ be the class defined as follows: for each alphabet A, $\mathcal{F}(A^*)$ is the class of all languages L of A^* such that, for each length-preserving morphism $\varphi : B^* \to A^*$, one has $\varphi^{-1}(L) \in \mathcal{C}(B^*)$. First, $\mathcal{F}(A^*)$ contains the singletons $\{1\}$ and $\{a\}$, for each letter $a \in A$, since $\varphi^{-1}(1) = \{1\}$ and $\varphi^{-1}(a)$ is a subset of B. Next, standard commutation rules and Proposition 1.1 show that $\mathcal{F}$ is closed under Boolean operations, product and shuffle. This shows that $\mathcal{F}$ contains $\mathcal{C}$. Thus $\mathcal{C}$ is closed under inverses of length-preserving morphisms.

Third step. Let L be a language of $\mathcal{C}(A^*)$ and let $\varphi : B^* \to A^*$ be a length-decreasing morphism. If A is empty, then $L = \{1\}$ and $\varphi^{-1}(1) = B^*$. Otherwise, let us fix a letter a of A. Setting

$$C = \{b \in B \mid \varphi(b) \neq 1\} \quad \text{and} \quad D = \{b \in B \mid \varphi(b) = 1\},$$

define a length-preserving morphism $\psi : B^* \to A^*$ by setting

$$\psi(b) = \begin{cases} \varphi(b) & \text{if } b \in C, \\ a & \text{if } b \in D. \end{cases}$$

Then the equality $\varphi^{-1}(L) = (\psi^{-1}(L) \cap C^*) \sqcup D^*$ and the previous steps show that $\varphi^{-1}(L) \in \mathcal{C}(B^*)$. Thus $\mathcal{C}$ is closed under inverses of length-decreasing morphisms. □

We now come to the main theorem of this paper.

Theorem 2.2 *Intermixed languages satisfy the d-identities $x^{\omega+1} = x^\omega$ and $(x^\omega y^\omega)^{\omega+1} = (x^\omega y^\omega)^\omega$.*

Proof. Let $\mathcal{F}$ be the class of languages satisfying the two identities of the statement. Then $\mathcal{F}$ is closed under Boolean operations and it is easy to see that it contains the singletons $\{1\}$ and $\{a\}$ for each letter a. We also know that, according to Proposition 1.6, one may replace the d-identities of the statement by $x^{\omega+1} \leqslant x^\omega$ and $(x^\omega y^\omega)^{\omega+1} \leqslant (x^\omega y^\omega)^\omega$. Finally, Proposition 1.7 shows that $\mathcal{F}$ is closed under product. Therefore, we just need to prove that $\mathcal{F}$ is closed under shuffle to conclude.

Let L_1 and L_2 be two languages of $\mathcal{F}(A^*)$ and let $L = L_1 \sqcup L_2$. Proposition 1.8 already shows that L satisfies the d-identity $x^{\omega+1} \leqslant x^\omega$. Let n be the least common multiple of the exponents of L_1, L_2 and L. By Propositions 1.5 and 1.6, we have to prove that if u and v are words of length $\leqslant 1$ of A^*, then $(u^n v^n)^{n+1} \leqslant_L (u^n v^n)^n$. The result is trivial if one of the words is empty, and we may assume that $u = a$ and $v = b$, for some (possibly equal) letters a and b.

Let $x, y \in A^*$ and suppose that $x(a^n b^n)^n y \in L$. Since L satisfies the d-identity $a^{\omega+1} = a^\omega$, one has $a^{2n-1} \sim_L a^n$ and $b^{2n-1} \sim_L b^n$. Setting $u = x(a^{2n-1}b^{2n-1})^n y$, we get $u \in L$ and thus $u \in u_1 \sqcup u_2$ for some $u_1 \in L_1$ and $u_2 \in L_2$. A factor of u of the form a^{2n-1} or b^{2n-1} will be called a *block* in the sequel. Let us say that a letter of u is *red* if it projects onto u_1 and *black* if it projects onto u_2 and that a block is *red* (resp. *black*) if it contains a majority of red (resp. black) letters.

First suppose that in u, at least two consecutive blocks have different colours. Let us assume for instance that a red block a^{2n-1} is followed by a black block b^{2n-1} (the three other cases are similar). We may also assume that the n last letters of a^{2n-1} are red and the first n letters of b^{2n-1} are black: if it is not the case, it suffices to permute a few letters a (resp. b) without changing the shuffle product. Since n is a multiple of an exponent of L_1 and L_2, replacing a^{2n} by a^n and b^{2n} by b^n within u_1 (resp. u_2) yields a word of L_1 (resp. L_2). Reshuffling these words a^{2n} and b^{2n}, we can replace the central factor a^nb^n of $a^{2n-1}b^{2n-1}$ by $a^nb^na^nb^n$. Thus we may replace in u the factor $a^{2n-1}b^{2n-1}$ by $a^{2n-1}b^na^nb^{2n-1}$, and still obtain a word of L. Since $a^{2n-1} \sim_L a^n$ and $b^{2n-1} \sim_L b^n$, one may replace $a^{2n-1}b^na^nb^{2n-1}$ by $a^nb^na^nb^n$ and the other factors $a^{2n-1}b^{2n-1}$ by a^nb^n, to obtain the word $x(a^nb^n)^{n+1}y$, which is therefore still in L. This proves the result in this case.

The only remaining possibility is that all blocks have the same colour, say red. This means that

$$u_1 = x_1a^{p_1}b^{q_1}\cdots a^{p_n}b^{q_n}y_1 \text{ and } u_2 = x_2a^{r_1}b^{s_1}\cdots a^{r_n}b^{s_n}y_2\,,$$

with $p_i, q_i \geqslant n$ and $r_i, s_i < n$. Since L_1 satisfies the equation $a^{\omega+1} = a^\omega$, the word $x_1(a^nb^n)^ny_1$ is in L_1, and since L_1 satisfies the equation

$$(a^\omega b^\omega)^{\omega+1} \leqslant (a^\omega b^\omega)^\omega\,,$$

it also contains the words $x_1(a^nb^n)^{n+1}y_1$ and $u_1' = x_1a^nb^na^{p_1}b^{q_1}\cdots a^{p_n}b^{q_n}y_1$. Reshuffling with u_2 shows that $x(a^nb^n)^{n+1}y$ is in L. □

Theorem 2.2 suffices to prove that intermixed languages form a proper subclass of the class of all regular languages.

Corollary 2.3 *The language $(aa)^*$ over the single letter alphabet $\{a\}$ is not intermixed.*

Proof. This language clearly does not satisfy the d-identity $x^{\omega+1} = x^\omega$. □

It is tempting to try to generalise the identities of Theorem 2.2 to three variables or more. The next paragraph summarizes one of our unsuccessful attempts to do so.

Let L be a regular language of A^* and let $\eta : A^* \to M$ be its syntactic morphism. For each nonnegative integer n, consider the property (P_n) defined as follows:

> *In M, the subsemigroup generated by n elements of the form $\eta(a)^\omega$, where a is a letter, is aperiodic.*

It is easy to see that L satisfies (P_2) if and only if it satisfies the d-identities $x^{\omega+1} = x^\omega$ and $(x^\omega y^\omega)^{\omega+1} = (x^\omega y^\omega)^\omega$ and thus, by Theorem 2.2, every intermixed language satisfies (P_2). This result lead us to conjecture that every intermixed language should satisfy (P_n) for all n, until our hopes were ruined by the following counterexample.

Example 2.1 Let $A = \{a, b, c\}$, $H = (a^+b)^+a^+$ and

$$L = \left(Hc^+Hc^+\right)^+ H \sqcup b^+ .$$

A computation shows that the syntactic monoid of L is the 76-element monoid M presented by

$$\begin{aligned}\langle \{a,b,c\} \mid\ & aa = a,\ cb = bc,\ cc = c,\ b^2b = b^2,\ b^2c = bc,\ ab^2a = abab,\ acac = 0,\\ & baba = abab,\ babb = bab,\ b^2ab = bab,\ b^2ac = babc,\ bcac = 0,\\ & cabb = bcab,\ cabc = 0,\ cac = 0,\ cabab = bcaba,\ cabacabac = c\rangle .\end{aligned}$$

Denote by S the subsemigroup of M generated by the three idempotents a, b^2 and c. Then S is a 44-element semigroup in which the element $x = ab^2acab^2a$ satisfies $x^3 = x$ but $x^2 \neq x$. Therefore S is not aperiodic.

We claim that L is intermixed. First, one has $L = K \sqcup b^+$ with

$$K = \left(Hc^+Hc^+\right)^+ H .$$

Next $K = (R \sqcup a^*) \setminus A^*ca^+cA^*$ where $R = \left((ab)^+ac^+(ab)^+ac^+\right)^+(ab)^+a$. The language $A^*ca^+cA^*$ is star-free. Further, one has

$$R = \left(aA^* \cap \left(\left((ab)^+ac(ab)^+ac\right)^+(ab)^+a \sqcup c^*\right) \cap A^*a\right) \setminus A^*(bc \cup cb)A^*$$

and

$$\left((ab)^+ac(ab)^+ac\right)^+ = \left((ab)^+ac\right)^+ \cap \left((acc)^* \sqcup \{a,b\}^*\right) .$$

Since the remaining pieces are star-free and hence intermixed, the claim is proved.

A weaker condition (Q_n) could also be considered:

> *In M, the minimal ideal of each subsemigroup generated by n elements of the form $\eta(a)^\omega$, where a is a letter, is aperiodic.*

It is easy to see that L satisfies the d-identities $x^{\omega+1} = x^\omega$ and $(x^\omega y^\omega)^{\omega+1} = (x^\omega y^\omega)^\omega$ if and only if it satisfies (Q_1) and (Q_2). We leave as an open problem to know whether every intermixed language satisfies (Q_n) for all n.

We conclude this section by a nontrivial example of intermixed languages. Recall that a word is *primitive* if it is not a power of another word. If u is a primitive word, then u^* is a star-free language (see for instance [13] for a more general result).

Proposition 2.4 *Let u be a primitive word of length > 1. Then for each nonnegative integer r, the language $(u^r)^*$ is intermixed.*

Proof. Let u be a primitive word of A^*. Then u contains at least two distinct letters of A. Let a be the last letter of u. Then u can be written as vba^k, where $v \in A^*$, $k > 0$ and b is a letter distinct from a.

Let $w = a^{r|u|_a - k}ba^k$. Since w contains a single b, it is primitive. Observing that $w^* \sqcup (A \setminus b)^* = \{z \in A^* \mid |z|_a \equiv 0 \bmod r|u|_a\}$, we get

$$\left(u^r\right)^* = u^* \cap \left[w^* \sqcup (A \setminus b)^*\right] .$$

Since u and w are primitive, the languages u^* and w^* are star-free and thus $\left(u^r\right)^*$ is intermixed. □

Note that the condition $|u| > 1$ is mandatory in Proposition 2.4 since the language $(aa)^*$ is not intermixed.

3 Shuffle by a letter

The operation of shuffling a language L by a letter a is the operation $L \mapsto L \sqcup a$. In this section we consider classes of languages closed under Boolean operations and under shuffle by a letter. Proposition 1.3 shows that the smallest class with these properties is the class of commutative star-free languages. We are interested in larger classes containing at least one noncommutative language. The following proposition shows that, under some reasonable conditions, it is natural to start with the language $\{ab\}$ on a two letter alphabet.

Proposition 3.1 *Let $\mathcal{V}$ be a p-variety of languages. If $\mathcal{V}$ contains the finite commutative languages and at least one noncommutative language, then $\mathcal{V}(\{a, b\}^*)$ contains the language $\{ab\}$.*

Proof. Let L be a noncommutative language of $\mathcal{V}(A^*)$. By definition, there exist two distinct letters $c, d \in A$ and two words $x, y \in A^*$ such that $xcdy \in L$ and $xdcy \notin L$. Setting $K = x^{-1}Ly^{-1}$, we get $cd \in K$ and $dc \notin K$. But since $\mathcal{V}$ is closed under quotients, $K \in \mathcal{V}(A^*)$. Furthermore, since $\mathcal{V}$ contains the finite commutative languages, $\mathcal{V}(A^*)$ contains the language $R = \{cd, dc\}$. It follows that $\{dc\} = R \setminus K$ is a language of $\mathcal{V}(A^*)$. Let now $\varphi : \{a, b\}^* \to A^*$ be the length-preserving morphism defined by $\varphi(a) = d$ and $\varphi(b) = c$. By construction, $\varphi^{-1}(\{dc\}) = \{ab\}$ and thus $\mathcal{V}(\{a, b\}^*)$ contains the language $\{ab\}$. □

Let $\mathcal{C}_0$ denote the smallest class of languages containing the languages of the form $\{ab\}$, where a, b are distinct letters, and which is closed under Boolean operations and under shuffle by a letter.

The aim of this section is to give both a combinatorial and an algebraic characterization of $\mathcal{C}_0$. Although the combinatorial characterization may appear more descriptive to the reader, the algebraic one is more powerful. It shows in particular that the class $\mathcal{C}_0$ is decidable: given a regular language, one can effectively decide whether or not it belongs to $\mathcal{C}_0$.

We first prove a combinatorial result of independent interest.

Proposition 3.2 *Let u be a word of length $\geqslant 3$. Then the language $\{u\}$ is a Boolean combination of languages of the form $v \sqcup a$, where a is a letter and v is a word of length $|u| - 1$.*

Proof. Let $n = |u| - 1$ and $E = \{(v, a) \in A^n \times A \mid u \in v \sqcup a\}$. The result will follow from the formula

$$\{u\} = \Big(\bigcap_{(v,a)\in E} v \sqcup a \Big) \setminus \Big(\bigcup_{(v,a)\in (A^n\times A)\setminus E} v \sqcup a \Big). \qquad (*)$$

Let L be the right hand side of $(*)$. It is clear that $u \in L$. Suppose that L contains another word w. Then $|w| = |u|$ and, for every $(v, a) \in A^n \times A$,

$u \in v \sqcup a$ if and only if $w \in v \sqcup a$. Let f be the longest common prefix of u and w. Assuming $u \neq w$, one can write $u = fau'$ and $w = fbw'$, for some $u', w' \in A^*$, $a, b \in A$ and $a \neq b$. We claim that f is the empty word. Otherwise, let c be a letter of f and let $f = f_1cf_2$. Let us assume that $c \neq a$ (the case $c \neq b$ would be symmetric by exchanging u and w). Then $u \in f_1f_2au' \sqcup c$ and thus $w = f_1cf_2bw' \in f_1f_2au' \sqcup c$. This means that c has to be inserted in the word f_1f_2au' to produce f_1cf_2bw'. Since $a \neq b$, this insertion cannot occur inside the prefix f_1f_2a. Therefore $f_1f_2a = f_1cf_2$, a contradiction, since $|f_1f_2a|_a > |f_1cf_2|_a$.

Thus the longest common prefix of u and w is the empty word, and by a symmetric argument, their longest common suffix is also the empty word. Let c be the first letter of u'. Then $u' = cx$ for some word $x \in A^*$. It follows that $u \in ax \sqcup c$ and thus $w \in ax \sqcup c$. Since the first letter of w is b, it means that $c = b$ and $w = bax$. It follows that x is a common suffix of u and w and thus x is the empty word. Therefore $u = ab$ and $w = ba$, a contradiction, since $|u| \geqslant 3$. □

Proposition 3.3 *The class $\mathcal{C}_0$ contains all finite languages.*

Proof. Since $\mathcal{C}_0(A^*)$ is closed under union, it suffices to prove that it contains the languages reduced to a single word u. If $|u| \leqslant 1$, the language $\{u\}$ is star-free commutative and the result follows from Proposition 1.3. If $|u| = 2$, say $u = ab$, either $a \neq b$ and the language $\{ab\}$ belongs by definition to $\mathcal{C}_0(A^*)$, or $a = b$ and the language $\{u\}$ is star-free commutative. Finally, if $|u| > 2$, Proposition 3.2 permits to conclude by induction on the length of u. □

A language L is said to be *almost star-free commutative* if there exists a star-free commutative language C such that the symmetric difference $L \triangle C$ is finite.

Theorem 3.4 *The class $\mathcal{C}_0$ is the class of almost star-free commutative languages.*

Proof. Since $\mathcal{C}_0(A^*)$ is a Boolean algebra, Propositions 1.3 and 3.3 show that $\mathcal{C}_0(A^*)$ contains the almost star-free commutative languages. Since this latter class of languages is closed under Boolean operations and contains the languages of the form $\{ab\}$, it suffices to show that it is closed under shuffle by a letter. But this property follows immediately from the formula

$$(L \sqcup a) \triangle (C \sqcup a) \subseteq (L \triangle C) \sqcup a\,,$$

which holds[1] for any languages L and C, and any letter a. □

Corollary 3.5 *The class $\mathcal{C}_0$ is an i-variety of languages.*

[1]Note that this inclusion might be strict. For instance, if $L = \{ab\}$ and $C = \{ba\}$, then $(L \sqcup a) \triangle (C \sqcup a) = \{aab, baa\}$ and $(L \triangle C) \sqcup a = \{aab, aba, baa\}$.

Proof. Let L be an almost star-free commutative language of A^*. By assumption, there exists a star-free commutative language C such that $L \triangle C$ is finite. If u is a word, then $u^{-1}C$ is star-free commutative and $u^{-1}(L \triangle C)$ is finite. It follows, since $(u^{-1}L) \triangle (u^{-1}C) = u^{-1}(L \triangle C)$, that $u^{-1}L$ is almost star-free commutative. The proof that Lu^{-1} is almost star-free commutative is dual. Thus $\mathcal{C}_0$ is closed under quotients.

Let $\varphi : B^* \to A^*$ be a length-increasing morphism. Since star-free commutative languages form a variety of languages, $\varphi^{-1}(C)$ is star-free commutative. Further, since φ is length-increasing, $\varphi^{-1}(L \triangle C)$ is finite. Finally, φ^{-1} commutes with Boolean operations and hence $\varphi^{-1}(L \triangle C) = \varphi^{-1}(L) \triangle \varphi^{-1}(C)$, which shows that $\varphi^{-1}(L)$ is almost star-free commutative. Therefore $\mathcal{C}_0$ is an i-variety of languages. □

Since $\mathcal{C}_0$ is an i-variety of languages, it corresponds to some variety of semigroups $\mathbf{V}$. Now, an almost star-free commutative language is a Boolean combination of finite languages and of star-free commutative languages. The i-variety of finite or cofinite languages corresponds to the variety of finite nilpotent semigroups **Nil** and the i-variety of star-free commutative languages corresponds to the variety of finite aperiodic commutative semigroups **Acom** [5]. It follows that $\mathbf{V}$ is the join of the varieties **Nil** and **Acom**. We are indebted to Jorge Almeida for providing us with a set of equations defining $\mathbf{V}$, which lead to the following characterization.

Theorem 3.6 *A regular language is almost star-free commutative if and only if it satisfies the i-identities $x^\omega = x^{\omega+1}$, $x^\omega y = yx^\omega$ and $x^\omega yz = x^\omega zy$.*

Proof. Let $\mathbf{V}$ be the join of the varieties **Nil** and **Acom**. As explained before, it suffices to prove that a finite semigroup belongs to $\mathbf{V}$ if and only if it satisfies the three identities $x^\omega = x^{\omega+1}$, $x^\omega y = yx^\omega$ and $x^\omega yz = x^\omega zy$. These identities are clearly satisfied by a nilpotent semigroup and by a commutative aperiodic semigroup.

Let S be a finite semigroup satisfying these identities. The identity $x^\omega = x^{\omega+1}$ says that S is aperiodic and the identity $x^\omega y = yx^\omega$ means that each idempotent of S commutes with any other element of S. These properties imply that the minimal ideal of S is a singleton and therefore S has a zero. We now prove by induction on the number of elements of S that S belongs to $\mathbf{Nil} \vee \mathbf{Acom}$.

If S has only one idempotent, then S is nilpotent and the result is trivial. Otherwise, let e be a nonzero idempotent of S. Then eS is an ideal of S, since $SeS = eSS \subseteq eS$. Observe also that if $s \in eS$, then $es = s$ since, if $s = ex$ for some $x \in S$, then $es = eex = ex = s$. Finally let us show that eS is a commutative semigroup. Let $y, z \in S$. Since $eS = S$, one has $y = ey$ and $z = ez$. Further, the identity $x^\omega yz = x^\omega zy$ gives $eyz = ezy$. Putting these relations together, we get $yz = eyz = ezy = zy$.

Denote by π the projection from S onto the Rees quotient S/eS and let $\varphi : S \to eS \times S/eS$ be the morphism defined by $\varphi(s) = (es, \pi(s))$. We claim that φ is injective. Indeed, suppose that $\varphi(s) = \varphi(t)$. The condition $\pi(s) = \pi(t)$ implies that s and t are either both in eS or both in its complement. If $s, t \in S \setminus eS$, the condition $\pi(s) = \pi(t)$ ensures that $s = t$. If $s, t \in eS$, then $es = s$ and $et = t$. Therefore $es = et$ implies $s = t$, which proves the claim. Thus S

is a subsemigroup of $eS \times S/eS$. Since 0 and e are in eS, they are identified by π and $|S/eS| < |S|$. It follows by the induction hypothesis that $S/eS \in \mathbf{V}$. Since the semigroup eS is aperiodic and commutative, it also belongs to $\mathbf{V}$ and finally S also belongs to $\mathbf{V}$. □

Corollary 3.7 *It is decidable whether a given regular language is almost star-free commutative.*

4 Jumbled languages

In this section, we consider the smallest class of languages $\mathcal{C}_1$ containing $\mathcal{C}_0$ and closed under Boolean operations and under shuffle by a letter and by the star of a letter. We call the languages of this class *jumbled languages*. We first establish some closure properties.

Proposition 4.1 *A class of languages which is closed under finite union and under shuffle by a letter and by the star of a letter is also closed under the operations $L \mapsto L \sqcup K$, where K is a star-free commutative language.*

Proof. By Proposition 1.2, every star-free commutative language is a finite union of languages of the form $[u] \sqcup B^*$ where u is a word and B is a subset of A. Now, set $u = a_1 \cdots a_n$, and $B = \{b_1, b_2, \ldots, b_k\}$, where $a_1, \ldots, a_n, b_1, \ldots, b_k$ are letters. Then $[u] = a_1 \sqcup a_2 \sqcup \cdots \sqcup a_n$ and $B^* = b_1^* \sqcup b_2^* \cdots \sqcup b_k^*$. The result now follows, since the shuffle product is associative and distributes over union. □

Proposition 4.2 *The class of jumbled languages forms a d-variety of languages.*

Proof. The proof is similar to that of Theorem 2.1. We first prove that the class of jumbled languages is closed under quotient and then that it is closed under inverses of length decreasing morphisms.

First step. Let $\mathcal{F}$ be the class of languages containing the jumbled languages L such that for all $a \in A$, $a^{-1}L, La^{-1}$ is jumbled. Since quotients commute with Boolean operations, $\mathcal{F}$ is closed under Boolean operations. Further, since $\mathcal{C}_0$ is contained in $\mathcal{C}_1$ and is closed under quotient, it is contained in $\mathcal{F}$. In particular, for each alphabet A, the languages of the form $\{a\}$, a^* and $\{ab\}$, where a and b are letters of A, are in $\mathcal{F}(A^*)$. Next we show that $\mathcal{F}$ is closed under shuffle by a letter and shuffle by the star of a letter. Suppose that $L \in \mathcal{F}(A^*)$ and let a be a letter. Then L is jumbled and thus $L \sqcup a$ and $L \sqcup a^*$ are also jumbled. For $b \in A$, the following formulas hold:

$$(L \sqcup a)b^{-1} = \begin{cases} (La^{-1} \sqcup a) \cup L & \text{if } b = a\,. \\ Lb^{-1} \sqcup a & \text{otherwise.} \end{cases} \tag{4}$$

$$(L \sqcup a^*)b^{-1} = \begin{cases} (La^{-1} \sqcup a^*) \cup (L \sqcup a^*) & \text{if } b = a\,. \\ Lb^{-1} \sqcup a^* & \text{otherwise.} \end{cases} \tag{5}$$

It follows that $(L \sqcup a)b^{-1}$ and $(L \sqcup a^*)b^{-1}$ (and by symmetry $b^{-1}(L \sqcup a)$ and $b^{-1}(L \sqcup a^*)$) are jumbled. Thus $L \sqcup a$ and $L \sqcup a^*$ are in $\mathcal{F}(A^*)$.

It follows that $\mathcal{F}$ contains $\mathcal{C}_0$ and is closed under the Boolean operations and under shuffle by a letter and by the star of a letter. In other words, $\mathcal{F}$ contains the jumbled languages. Coming back to the definition of $\mathcal{F}$, it means that the class of jumbled languages is closed under quotients.

Second step. Let $\mathcal{F}'$ be the class of all jumbled languages L of A^* such that, for each length-decreasing morphism $\varphi : B^* \to A^*$, $\varphi^{-1}(L) \in \mathcal{C}_1(B^*)$. We claim that, for each alphabet A, the languages of the form $\{a\}$, a^* and $\{ab\}$, where a and b are letters of A, belong to $\mathcal{F}'(A^*)$. First, these languages are jumbled. Let now $\varphi : B^* \to A^*$ be a length-decreasing morphism. Let B_a, B_b and C be the subsets of B consisting of the letters c such that $\varphi(c)$ is respectively equal to a, b and 1. Then

$$\varphi^{-1}(\{a\}) = C^*B_aC^* = \bigcup_{c \in B_a} c \sqcup C^*\,, \tag{6}$$

$$\varphi^{-1}(a^*) = (C^*B_a)^*C^* = B_a^* \sqcup C^*\,, \tag{7}$$

$$\varphi^{-1}(\{ab\}) = C^*B_aC^*B_bC^* = \bigcup_{c \in B_a, d \in B_b} cd \sqcup C^*\,. \tag{8}$$

By Proposition 4.1, these languages are jumbled.

Next we show that $\mathcal{F}'$ is closed under shuffle by a letter and by the star of a letter. Suppose that $L \in \mathcal{F}'(A^*)$ and let a be a letter. Then the languages L, $L \sqcup a$ and $L \sqcup a^*$ are jumbled. Let $\varphi : B^* \to A^*$ be a length-decreasing morphism. Proposition 1.1 shows that

$$\varphi^{-1}(L \sqcup a) = \varphi^{-1}(L) \sqcup \varphi^{-1}(a) \text{ and } \varphi^{-1}(L \sqcup a^*) = \varphi^{-1}(L) \sqcup \varphi^{-1}(a^*).$$

Now Formulas (6) and (7) and Proposition 4.1 show that $\varphi^{-1}(L \sqcup a)$ and $\varphi^{-1}(L \sqcup a^*)$ are in $\mathcal{C}_1(B^*)$. Thus $L \sqcup a$ and $L \sqcup a^*$ are in $\mathcal{F}'(A^*)$.

Finally, $\mathcal{F}'(A^*)$ is closed under Boolean operations, since these operations commute with inverse morphisms. It follows that the class $\mathcal{F}'$ contains the languages of the form $\{ab\}$, and is closed under the Boolean operations and shuffle by a letter and by the star of a letter. Since $\mathcal{C}_1$ is by definition the smallest class with these properties, $\mathcal{F}'$ contains $\mathcal{C}_1$. Coming back to the definition of $\mathcal{F}'$, it means that the class of jumbled languages is closed under inverses of length decreasing morphisms. □

Recall that a language is *piecewise testable* if it is a Boolean combination of languages of the form $u \sqcup A^*$, where u is a word. These languages have been characterized by I. Simon: a language is piecewise testable if and only if its syntactic monoid is $\mathcal{J}$-trivial.

Note that the class of piecewise testable languages is closed under the operation $L \mapsto L \sqcup A^*$ since, by a celebrated theorem of Higman, every language of the form $L \sqcup A^*$ can be written as $F \sqcup A^*$ for some finite language.

Proposition 4.3 *Every piecewise testable language is jumbled.*

Proof. By Proposition 3.3, $\mathcal{C}_1$ contains all languages of the form $\{u\}$, where u is a word. Therefore, by Proposition 4.1, it also contains the languages of the form $u \sqcup A^*$. □

We shall frequently use the following consequence of Proposition 4.3.

Corollary 4.4 *The languages of the form $A_0^* a_1 A_1^* a_2 \cdots a_k A_k^*$, where $A_0, \ldots, A_k$ are subsets of the alphabet and, for $1 \leqslant i \leqslant k$, $a_i \notin A_{i-1} \cup A_i$, are jumbled.*

Proof. Indeed these languages are piecewise testable. □

Proposition 4.5 *The languages of the form $a^n A^*$ and $a^n b A^*$, where $a, b \in A$, are jumbled.*

Proof. Let $B = A \setminus \{a\}$. Then $a^n B^*$ can be written as $\emptyset^* a \emptyset^* a \cdots \emptyset^* a B^*$ and thus is jumbled by Corollary 4.4. Now $a^n A^* = a^n B^* \sqcup a^*$ and thus $a^n A^*$ is also jumbled. Corollary 4.4 also shows that $a^* b (A \setminus \{b\})^*$ is jumbled. Consequently, the language $\{a, b\}^* b A^*$, which is equal to $a^* b (A \setminus \{b\})^* \sqcup b^*$, is also jumbled.

Let us show by induction on n that $a^n b A^*$ is jumbled. For $n = 0$, it follows from the previous result. Next, the formula

$$a^n b A^* = \big((a^n A^* \sqcup b) \cap \{a, b\}^* b A^*\big) \setminus \big((1 \cup a \cup \cdots \cup a^{n-1}) b A^* \cup a^{n+1} A^*\big)$$

provides the induction step. □

Proposition 4.5 already suffices to separate $\mathcal{C}_0$ from $\mathcal{C}_1$.

Corollary 4.6 *Let $A = \{a, b\}$. Then the language aA^* is jumbled but is not almost star-free commutative.*

Proof. The first part follows from Proposition 4.5. Next, the syntactic semigroup of aA^* is the semigroup $S = \{a, b\}$ defined by $a^2 = ab = a$ and $ba = bb = b$. In particular, idempotents do not commute in S and by Theorem 3.6, aA^* is not in $\mathcal{C}_0(A^*)$. □

Recall that a language L of A^* is *local* if there exist two subsets P and S of A and a subset N of A^2 such that $L \setminus \{1\} = (PA^* \cap A^* S) \setminus A^* N A^*$.

Proposition 4.7 *Every local language is jumbled.*

Proof. The language $\{1\}$ is jumbled since it is finite and the languages of the form aA^* are jumbled by Proposition 4.5. Let now a and b be two (possibly equal) letters of A and let $C = A \setminus \{a, b\}$. Then by Corollary 4.4, $C^* a b C^*$ is jumbled. Now $A^* a b A^* = C^* a b C^* \sqcup \{a, b\}^*$ and thus $A^* a b A^*$ is also jumbled. Thus every local language is jumbled. □

We now state a useful property of the jumbled languages, which is not an immediate consequence of the definition.

Proposition 4.8 *If L is a jumbled language and u is a word, the languages uL and Lu are also jumbled.*

Proof. Let L be a jumbled language. By symmetry, it suffices to prove that uL is also jumbled. Actually, it suffices to show that for each letter a, aL is jumbled.

If the empty word belongs to L, one has $L = \{1\} \cup (L \setminus \{1\})$. Therefore one may assume that L does not contain the empty word. By Theorem 2.2,

L satisfies the equation $a^\omega = a^{\omega+1}$. Let n be the smallest integer such that $a^n \sim_L a^{n+1}$.

Let $B = A \setminus \{a\}$ and, for $0 \leqslant i \leqslant n-1$, let $L_i = L \cap a^i BA^*$. Finally, let $L_n = L \cap a^n A^*$, $L_{n+1} = L \cap a^{n+1}A^*$ and $K = L \cap a^*$. By Proposition 4.5, the languages L_i, where $0 \leqslant i \leqslant n+1$, are jumbled. The language K is also jumbled. Note that

$$L = L_0 \cup L_1 \cup \cdots \cup L_n \cup K.$$

We claim that $aL_n = L_{n+1}$. Indeed, let $u \in L_n$. Then $u = a^n v$ for some $v \in A^*$ and since $a^n \sim_L a^{n+1}$ and $a^n v \in L$, one gets $au = a^{n+1}v \in L$. Therefore $au \in L_{n+1}$ and thus $aL_n \subseteq L_{n+1}$. To prove the opposite inclusion, consider a word $u \in L_{n+1}$. Then $u \in L$ and $u = a^{n+1}v$ for some $v \in A^*$. Since $a^n \sim_L a^{n+1}$, one also has $a^n v \in L$ and thus $a^n v \in L_n$. Therefore $u \in aL_n$, which proves the claim. Now, the formulas

$$\begin{aligned}
aL_0 &= (a \sqcup L_0) \setminus BA^*\,,\\
aL_1 &= (a \sqcup L_1) \setminus aBA^*\,,\\
&\vdots\\
aL_{n-1} &= (a \sqcup L_{n-1}) \setminus a^n BA^*\,,\\
aL_n &= L_{n+1}
\end{aligned}$$

show that for $0 \leqslant i \leqslant n$, the languages aL_i are jumbled. Since aK is commutative star-free, it is also jumbled. Finally, the language

$$aL = aL_0 \cup aL_1 \cup \cdots \cup aL_n \cup aK$$

is jumbled. □

Proposition 4.9 *The languages $A^*a^nA^*$ and $A^*a^nba^mA^*$, for $n, m \geqslant 0$ and $a, b \in A$, are jumbled.*

Proof. Let $B = A \setminus \{a\}$. Then $A^*a^nA^* = B^*a^nB^* \sqcup a^*$. Since $B^*a^nB^*$ is jumbled by Corollary 4.4, $A^*a^nA^*$ is also jumbled.

Let $C = A\setminus\{a,b\}$. By Corollary 4.4, the languages of the form $C^*a^kba^{n-k}C^*$ are jumbled. Now the formulas

$$\begin{aligned}
B^*a^nbB^* &= (C^*a^nbC^* \sqcup b^*) \setminus \bigcup_{0<k<n} (C^*a^kba^{n-k}C^* \sqcup b^*)\,,\\
A^*a^nbA^* &= B^*a^nbB^* \sqcup a^*
\end{aligned}$$

show that $A^*a^nbA^*$ is jumbled.

Finally, if $m > 0$, the languages of the form $B^*a^nba^mB^*$ are jumbled by Corollary 4.4. Since

$$A^*a^nba^mA^* = B^*a^nba^mB^* \sqcup a^*\,,$$

the languages of the form $A^*a^nba^mA^*$ are also jumbled. □

Proposition 4.10 *The languages $(a^n b)^*$ and $((ab)^n)^*$, for $n \geqslant 0$, are jumbled. In particular, there exist non star-free jumbled languages.*

Proof. Let $A = \{a, b\}$. By Corollary 4.4, $\mathcal{C}_1$ contains, for each k, the language $a^*ba^kba^*$. The result now follows from the previous propositions by the following sequence of relations:

$$A^*b(ab^*)^kbA^* = a^*ba^kba^* \sqcup b^*\,,$$

$$(a^nb)^* = \{1\} \cup \left(\left(a^nA^* \cap A^*b\right) \setminus \left(A^*a^{n+1}A^* \cup \bigcup_{0 \leqslant k \leqslant n-1} A^*b(ab^*)^kbA^*\right)\right),$$

$$((ab)^n)^* = \left((a^nb)^* \sqcup b^*\right) \cap (ab)^*\,.$$

Now, the word ab generates a cyclic group of order n in the syntactic monoid of $((ab)^n)^*$. Therefore, by Schützenberger's theorem, $((ab)^n)^*$ is not star-free for $n \geqslant 2$. □

Other examples of jumbled languages include the language

$$\{a, b\}^*abc\{a, b\}^* = (b^*abcb^* \sqcup a^*) \setminus (b^*acb^* \sqcup a^*)\,,$$

but it is an open problem to know whether the language A^*abbaA^* is jumbled.

Corollary 4.11 *The class of jumbled languages is not closed under inverses of morphisms.*

Proof. Let $\varphi : \{a\}^* \to \{a, b\}^*$ be the morphism defined by $\varphi(a) = ab$. Then $\varphi^{-1}((abab)^*) = (aa)^*$. Now, the language $(abab)^*$ is jumbled by Proposition 4.10, but Corollary 2.3 shows that the language $(aa)^*$ is not jumbled. □

In fact, the closure of the class of jumbled languages under inverses of morphisms is equal to the class of all regular languages. More precisely, one has the following result, the proof of which relies on an argument of [7].

Proposition 4.12 *For every regular language L over A, there exist an alphabet C, a morphism φ from A^* to C^* and a jumbled language K over C such that $L = \varphi^{-1}(K)$.*

Proof. It is a well-known fact that every regular language is the image of some local language under a length-preserving morphism. Therefore, there is an alphabet B, a length-preserving morphism $\gamma : B^* \to A^*$ and a local language R of B^* such that $L = \gamma(R)$. By Proposition 4.7, R is jumbled.

Let c be a new letter and let $C = B \cup \{c\}$. We claim that the languages of C^*

$$R_1 = R \sqcup c^*\,, \qquad R_2 = (Bc)^*$$

are jumbled. This is clear for R_1. For R_2 observe that $R_2 = \pi^{-1}((ab)^*)$, where π denotes the length-preserving morphism from C^* into $\{a, b\}^*$ mapping c to b and each letter of B to a, and apply Propositions 4.10 and 4.2 to conclude. It follows that the language

$$K = (R_1 \cap R_2) \sqcup B^*$$

is jumbled. To finish the proof, let, for each $a \in A$, u_a be a word of B^* containing exactly one occurrence of each letter in $\gamma^{-1}(a)$, and no other letter. Consider the morphism $\varphi : A^* \to C^*$ defined, for each $a \in A$, by $\varphi(a) = u_a c$. It is shown in [7] that $\gamma(R) = \varphi^{-1}(K)$. Thus $L = \varphi^{-1}(K)$. □

5 Conclusion

We introduced four classes of regular languages related to the shuffle operation: almost star-free commutative, jumbled, shuffled and intermixed languages. We completed the study of the first class and proved only partial results on the other ones. Our hope is that these incomplete results and open problems will stimulate research on the shuffle, one of the most fascinating operations on regular languages.

Acknowledgements

We would like to thank Howard Straubing and the anonymous referees for their helpful comments and suggestions.

References

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

[2] J. C. M. Baeten and W. P. Weijland, *Process algebra*, *Cambridge Tracts in Theoretical Computer Science* vol. 18, Cambridge University Press, Cambridge, 1990.

[3] A. Cano Gómez and J.-É. Pin, Shuffle on positive varieties of languages, *Theoret. Comput. Sci.* **312** (2004), 433–461.

[4] A. Cano Gómez and J.-E. Pin, A Robust Class of Regular Languages, in *Mathematical Foundations of Computer Science 2008, 33rd International Symposium, MFCS 2008, Torun, Poland, August 25-29, 2008, Proceedings*, E. Ochmanski and J. Tyszkiewicz (eds.), Berlin, 2008, pp. 36–51, *Lect. Notes Comp. Sci.* vol. 5162, Springer.

[5] S. Eilenberg, *Automata, languages, and machines. Vol. B*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976.

[6] Z. Ésik and M. Ito, Temporal logic with cyclic counting and the degree of aperiodicity of finite automata, *Acta Cybernetica* **16** (2003), 1–28.

[7] Z. Ésik and I. Simon, Modeling literal morphisms by shuffle, *Semigroup Forum* **56**,2 (1998), 225–227.

[8] M. Kunc, Equational description of pseudovarieties of homomorphisms, *Theoret. Informatics Appl.* **37** (2003), 243–254.

[9] J.-F. Perrot, Variétés de langages et operations, *Theoret. Comput. Sci.* **7** (1978), 197–210.

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

[11] J.-É. Pin, Profinite methods in automata theory, in *26th International Symposium on Theoretical Aspects of Computer Science (STACS 2009)*, S. Albers (ed.), Dagstuhl, Germany, 2009, pp. ???–???, Internationales Begegnungs- Und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

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

[13] H. Straubing, Relational morphisms and operations on recognizable sets, *RAIRO Inf. Theor.* **15** (1981), 149–159.

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