Shuffle product of regular languages: results and open problems

Jean-Éric Pin[1]

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

Abstract. This survey paper presents known results and open problems on the shuffle product applied to regular languages. We first focus on varieties and positive varieties closed under shuffle. Next we turn to the class of intermixed languages, the smallest class of languages containing the letters and closed under Boolean operations, product and shuffle. Finally, we study Schnoebelen's sequential and parallel decompositions of languages and discuss various open questions around this notion.

Keywords: Shuffle product · Regular languages · Open problems.

1 Introduction

The shuffle product is a standard tool for modeling concurrency which has long been studied in formal language theory. A nice survey on this topic was proposed by Restivo [12] in 2015. Restivo's article is divided into two parts: the first part is devoted to language theory and the second part to combinatorics on words. Our current survey paper focuses on the shuffle product applied to regular languages and can therefore be seen as an extension of the first part of [12].

The main part of the article, Section 3, is devoted to the study of various classes of regular languages closed under shuffle. We first examine the varieties closed under shuffle, for which a complete description is known. We then turn to positive varieties of languages closed under shuffle, for which the situation is more contrasted and several questions are still open. Next we turn to the class of intermixed languages, the smallest class of languages containing the letters and closed under Boolean operations, product and shuffle. Here again, only partial results are known.

In the last part of the paper, we study Schnoebelen's sequential and parallel decompositions of languages and discuss various open questions around this notion.

2 Shuffle and recognition

2.1 Shuffle product

The *shuffle* $u \sqcup\!\sqcup v$ of two words u and v is the set of words obtained by shuffling u and v. Formally, it is the set of words of the form $u_1v_1 \cdots u_nv_n$, where the u_i's and v_i's are possibly empty words such that $u_1 \cdots u_n = u$ and $v_1 \cdots v_n = v$. For instance,

$$ab \sqcup\!\sqcup\, ba = \{abba, baab, baba, abab\}$$

This definition extends by linearity to languages: the *shuffle* of two languages L and K is the language

$$L \sqcup\!\sqcup\, K = \bigcup_{u \in L, v \in K} u \sqcup\!\sqcup\, v$$

The shuffle product is a commutative and associative operation on languages and it distributes over union.

2.2 Monoids and ordered monoids

The *algebraic approach* to the study of regular languages is based on the use of monoids to recognise languages. There are two versions, one using monoids and the other one using ordered monoids. The ordered version is more suitable for classes of languages not closed under complementation. Let us briefly recall the relevant definitions.

An *ordered monoid* is a monoid M equipped with a partial order $\leqslant$ compatible with the product on M: for all $x, y, z \in M$, if $x \leqslant y$ then $zx \leqslant zy$ and $xz \leqslant yz$. Note that the equality relation makes any monoid an ordered monoid.

Let P be a subset of M. It is a *lower set* if, for all $s, t \in P$, the conditions $s \in P$ and $t \leqslant s$ imply $t \in P$. It is an *upper set* if $s \in P$ and $s \leqslant t$ imply $t \in P$. Finally, *the lower set generated by* P is the set

$$\downarrow P = \{t \in M \mid \text{there exists } s \in P \text{ such that } t \leqslant s\}.$$

Given two ordered monoids M and N, a *morphism of ordered monoids* $\varphi \colon M \to N$ is an order-preserving monoid morphism from M to N. In particular, if $(M, \leqslant)$ is an ordered monoid, the identity on M is a morphism of ordered monoids from $(M, =)$ to $(M, \leqslant)$.

A monoid M *recognizes* a language L of A^* if there exist a morphism $\varphi \colon A^* \to M$ and a subset U of M such that $L = \varphi^{-1}(U)$. In the ordered version, the definition is the same, but U is required to be an upper set of the ordered monoid.

Let L be a language of A^*. The *syntactic preorder* of L is the preorder $\preccurlyeq_L$ defined on A^* by $u \preccurlyeq_L v$ if and only if, for every $x, y \in A^*$,

$$xuy \in L \Rightarrow xvy \in L. \quad (1)$$

The *syntactic congruence* of L is the associated equivalence relation $\sim_L$, defined by $u \sim_L v$ if and only if $u \preccurlyeq_L v$ and $v \preccurlyeq_L u$.

The *syntactic monoid* of L is the quotient monoid $\mathrm{Synt}(L)$ of A^* by $\sim_L$ and the natural homomorphism $\eta : A^* \to \mathrm{Synt}(L)$ is called the *syntactic morphism* of L. The syntactic preorder $\preccurlyeq_L$ induces a partial order $\leqslant_L$ on $\mathrm{Synt}(L)$. The resulting ordered monoid is called the *syntactic ordered monoid*[1] of L. Recall that a language is regular if and only if its syntactic monoid is finite.

[1] The syntactic ordered monoid of a language was first introduced by Schützenberger in 1956, but he apparently only made use of the syntactic monoid later on. I re-

2.3 Power monoids and lower set monoids

Let M be a monoid. Then the set $\mathcal{P}(M)$ of subsets of M is a monoid, called the *power monoid* of M, under the multiplication of subsets defined by

$$XY = \{xy \mid x \in X \text{ and } y \in Y\}$$

The ordered counterpart of this construction works as follows. Let $(M, \leqslant)$ be an ordered monoid and let $\mathcal{P}^{\downarrow}(M)$ be the set of all lower sets of M. The *product of two lower sets* X and Y is the lower set

$$XY = \{z \in M \mid \text{there exist } x \in X \text{ and } y \in Y \text{ such that } z \leqslant xy\}.$$

This operation makes $\mathcal{P}^{\downarrow}(M)$ a monoid. Furthermore, set inclusion is compatible with this product and thus $(\mathcal{P}^{\downarrow}(M), \subseteq)$ is an ordered monoid, called the *lower set monoid* of M.

The connection between the shuffle and the operator $\mathcal{P}$ was given in [10].

Proposition 1. *Let L_1 and L_2 be two languages and let M_1 and M_2 be monoids recognizing L_1 and L_2 respectively. Then $L_1 \sqcup\!\sqcup L_2$ is recognized by the monoid $\mathcal{P}(M_1 \times M_2)$.*

Since a language is regular if and only if it is recognised by a finite monoid, Proposition 1 gives an algebraic proof of the well-known fact that regular languages are closed under shuffle. The following ordered version was given in [4].[2]

Proposition 2. *Let L_1 and L_2 be two languages and let M_1 and M_2 be ordered monoids recognizing L_1 and L_2 respectively. Then $L_1 \sqcup\!\sqcup L_2$ is recognized by the ordered monoid $\mathcal{P}^{\downarrow}(M_1 \times M_2)$.*

3 Classes of languages closed under shuffle

It is a natural question to look for classes of regular languages closed under shuffle. In this section, we focus successively on varieties of languages and on positive varieties of languages. The last subsection is devoted to the class of intermixed languages and its subclasses.

3.1 Varieties of languages closed under shuffle

Following the success of the variety approach to classify regular languages (see [17] for a recent survey), Perrot proposed in 1977 to find the varieties of languages closed under shuffle. Let us first recall the definitions.

---

discovered this notion in 1995 [11], but unfortunately used the opposite order for several years, in particular in [5, 4, 6, 3], before I switched back to the original order.

[2] As explained in the first footnote, the opposite of the syntactic order was used in this paper, and consequently, upper set monoids were used in place of lower set monoids.

A *class of languages* is a correspondence $\mathcal{V}$ which associates with each alphabet A a set $\mathcal{V}(A^*)$ of languages of A^*. A class of regular languages is a *variety of languages* if it is closed under Boolean operations (that is, finite union, finite intersection, and complementation), left and right quotients and inverses of morphisms between free monoids.

A *variety of finite monoids* is a class of finite monoids closed under taking submonoids, quotients and finite direct products. If $\mathbf{V}$ is a variety of finite monoids, let $\mathcal{V}(A^*)$ denote the set of regular languages of A^* recognized by a monoid of $\mathbf{V}$. 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 states that these two correspondences define mutually inverse bijective correspondences between varieties of finite monoids and varieties of languages. For instance, Schützenberger [15] proved that star-free languages correspond to aperiodic monoids.

To start with, let us describe the smallest nontrivial variety of languages closed under shuffle. Let $[u]$ denote 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 characterisation of the star-free commutative languages related to the shuffle product is given in [2, Proposition 1.2].

Proposition 3. *A language of A^* is star-free commutative if and only if it is a finite union of languages of the form $[u] \sqcup\!\sqcup B^*$ where u is a word and B is a subset of A.*

This leads to the following result of Perrot [10].

Proposition 4. *The star-free commutative languages form a variety of languages, which is the smallest nontrivial variety of languages closed under shuffle. It is also the smallest class of languages closed under Boolean operations and under shuffle by a letter.*

Perrot actually characterised all commutative varieties of languages closed under shuffle. They correspond, via Eilenberg's correspondence, to the varieties of finite commutative monoids whose groups belong to a given variety of finite commutative groups. Perrot also conjectured that the only non-commutative variety of languages closed under shuffle was the variety of all regular languages. This conjecture remained open for twenty years, but was finally solved by Ésik and Simon [8].

Theorem 1. *The unique non-commutative variety of languages closed under shuffle is the variety of all regular languages.*

The story of the proof of Theorem 1 is worth telling. A *renaming* is a *length preserving morphism* $\varphi\colon A^* \to B^*$. This means that $|\varphi(u)| = |u|$ for each word u of A^*, or, equivalently, that each letter of A is mapped by φ to a letter of B.

The characterisation of varieties of languages closed under renaming was known for a long time. For a variety $\mathbf{V}$ of finite monoids, let $\mathbf{PV}$ denote the variety of finite monoids generated by the monoids of the form $\mathcal{P}(M)$, where $M \in \mathbf{V}$. A variety of finite monoids $\mathbf{V}$ is a *fixed point* of the operator $\mathbf{P}$ if $\mathbf{PV} = \mathbf{V}$.

Reuteunauer [13] and Straubing [16] independently proved the following result:

Proposition 5. *A variety of languages is closed under renaming if and only if the corresponding variety of finite monoids is a fixed point of the operator* $\mathbf{P}$.

It was also known that the unique non-commutative variety of languages satisfying this condition is the variety of all regular languages. Ésik and Simon managed to find an ingenious way to link renamings and the shuffle operation.

Proposition 6. *Let* $\varphi\colon A^* \to B^*$ *be a surjective renaming and let* $C = A \cup \{c\}$*, where* c *is a new letter. Then there exist monoid morphisms* $\pi : C^* \to A^*$*,* $\gamma : C^* \to \{a, b\}^*$ *and* $\eta : B^* \to C^*$ *such that*

$$\varphi(L) = \eta^{-1}\Big(\big(\pi^{-1}(L) \cap \gamma^{-1}((ab)^*)\big) ⧢ A^*\Big)$$

It follows that if a variety of languages containing the language $(ab)^*$ is closed under shuffle, then it is also closed under renaming, a key argument in the proof of Theorem 1.

3.2 Positive varieties of languages closed under shuffle

A variation of Eilenberg's variety theorem was proposed by the author in [11]. It gives two mutually inverse bijective correspondences between varieties of finite ordered monoids and positive varieties of languages.

The definition of varieties of finite ordered monoids is similar to that of varieties of finite monoids: they are classes class of finite ordered monoids closed under taking ordered submonoids, quotients and finite direct products.

A class of regular languages is a *positive variety of languages* if it is closed under finite union, finite intersection, left and right quotients and inverses of morphisms between free monoids. The difference with language varieties is that positive varieties are not necessarily closed under complementation.

It is now natural to try to describe all positive varieties closed under shuffle. As in the case of varieties, this study is related to the closure under renaming. For a variety $\mathbf{V}$ of finite ordered monoids, let $\mathbf{P}^{\downarrow}\mathbf{V}$ denote the variety of finite ordered monoids generated by the ordered monoids of the form $\mathcal{P}^{\downarrow}(M)$, where $M \in \mathbf{V}$. A variety of finite ordered monoids $\mathbf{V}$ is a *fixed point* of the operator $\mathbf{P}^{\downarrow}$ if $\mathbf{P}^{\downarrow}\mathbf{V} = \mathbf{V}$. The following result was proved in [3]:

Proposition 7. *A positive variety of languages is closed under renaming if and only if the corresponding variety of finite ordered monoids is a fixed point of the operator* $\mathbf{P}^{\downarrow}$.

Coming back to the shuffle, Proposition 2 leads to the following corollary.

Corollary 1. *If a variety of finite ordered monoids is a fixed point of the operator* $\mathbf{P}^{\downarrow}$*, then the corresponding positive variety of languages is closed under shuffle.*

Note that, contrary to Proposition 7, Corollary 1 only gives a sufficient condition for a positive variety of languages to be closed under shuffle.

The results of Section 3.1 show that the unique maximal proper variety of languages closed under shuffle is the variety of commutative languages. The counterpart of this result for positive varieties was given in [4].

Theorem 2. *There is a largest proper positive variety of languages closed under shuffle.*

Let $\mathcal{W}$ denote this positive variety and let $\mathbf{W}$ be its corresponding variety of finite ordered monoids. Their properties are summarized in the next statements, also proved in [4, 6].

Theorem 3. *The positive variety* $\mathcal{W}$ *is the largest positive variety of languages such that, for* $A = \{a, b\}$*, the language* $(ab)^*$ *does not belong to* $\mathcal{W}(A^*)$*.*

A key result is that $\mathbf{W}$, and hence $\mathcal{W}$, are decidable. A bit of semigroup theory is needed to make this statement precise.

Two elements s and t of a monoid are *mutually inverse* if $sts = s$ and $tst = t$. An *ideal* of a monoid M is a subset I of M such that $MIM \subseteq I$. It is *minimal* if, for every ideal J of M, the condition $J \subseteq I$ implies $J = \emptyset$ or $J = I$. Every finite monoid admits a unique minimal ideal. Finally, let us recall that every element s of a finite monoid has a unique idempotent power, traditionally denoted by s^ω. In the following, we will also use the notation $x^{\omega+1}$ as a shortcut for xx^ω.

Theorem 4. *A finite ordered monoid* M *belongs to* $\mathbf{W}$ *if and only if, for any pair* (s, t) *of mutually inverse elements of* M*, and any element* z *of the minimal ideal of the submonoid of* M *generated by* s *and* t*,* $(stzst)^\omega \leqslant st$*.*

Thus a regular language belongs to $\mathcal{W}$ if and only if its ordered syntactic monoid satisfies the decidable condition stated in Theorem 4. An equivalent characterisation in term of the minimal automaton of the language is given in [6].

The next theorem shows that the positive variety $\mathcal{W}$ is very robust.

Theorem 5. *The positive variety* $\mathcal{W}$ *is closed under the following operations: finite union, finite intersection, left and right quotients, product, shuffle, renaming and inverses of morphisms. It is not closed under complementation.*

The positive variety $\mathcal{W}$ can be defined alternatively as the largest proper positive variety of languages satisfying (1) (respectively (2) or (3)):

(1) not containing the language $(ab)^*$;
(2) closed under shuffle;

(3) closed under renaming;

Despite its numerous closure properties, no constructive description of $\mathcal{W}$, similar to the definition of star-free or regular languages, is known. For instance, the least positive variety of languages satisfying conditions (1)–(3) is the positive variety of polynomials of group languages, which is strictly contained in $\mathcal{W}$.

Problem 1. Find a constructive description of $\mathcal{W}$, possibly by introducing more powerful operators on languages of $\mathcal{W}$.

Let us come back to the problem of finding all positive varieties of languages closed under shuffle. The first question is to know in which case the converse of Corollary 1 holds. More precisely,

Problem 2. For which positive varieties of languages closed under shuffle is the corresponding variety of finite ordered monoids a fixed point of the operator $\mathbf{P}^{\downarrow}$?

We know this is the case for $\mathcal{W}$, but the general case is unknown. That said, an in-depth study of the fixed points of the operator $\mathbf{P}^{\downarrow}$ can be found in [1]. This paper actually covers the more general case of lower set semigroups and studies the fixed points of the operator $\mathbf{P}^{\downarrow}$ on varieties of finite ordered semigroups. An important property is the following:

Proposition 8. *Every intersection and every directed union of fixed points of $\mathbf{P}^{\downarrow}$ is also a fixed point for $\mathbf{P}^{\downarrow}$.*

The article [1] gives six independent basic types of such fixed points, from which many more may be constructed using intersection. Moreover, it is conjectured that all fixed points of $\mathbf{P}^{\downarrow}$ can be obtained in this way. The presentation of these basic types would be too technical for this survey article, but one of them is the variety $\mathbf{W}$.

3.3 Intermixed languages

In the early 2000s, Restivo proposed as a challenge to characterise the smallest class of languages containing the letters and closed under Boolean operations, product and shuffle. Let us call *intermixed* the languages of this class.

Problem 3 (Restivo). Is it decidable to know whether a given regular language is intermixed?

This problem is still widely open, and only partial results are known. To start with, the smallest class of languages containing the letters and closed under Boolean operations and product is by definition the class of star-free languages. It is not immediate to see that star-free languages are not closed under shuffle, but an example was given in [10]: the languages $(abb)^*$ and a^* are star-free, but their shuffle product is not star-free. This led Castiglione and Restivo [7] to propose the following question:

Problem 4. Determine conditions under which the shuffle of two star-free languages is star-free.

The following result, which improves on the results of [7], can be seen as a reasonable answer to this problem.

Theorem 6. *The shuffle of two star-free languages of the positive variety $\mathcal{W}$ is star-free.*

Proof. This is a consequence of the fact that the intersection of $\mathbf{W}$ and the variety of finite aperiodic monoids is a fixed point of the operator $\mathbf{P}^{\downarrow}$, a particular instance of [1, Theorem 7.4].

Here are the known closure properties of intermixed languages obtained in [2]. A morphism $\varphi\colon A^* \to B^*$ is said to be *length-decreasing* if $|\varphi(u)| \leqslant |u|$ for every word u of A^*.

Proposition 9. *The class of intermixed languages is closed under left and right quotients, Boolean operations, product and shuffle. It is also closed under inverses of length-decreasing morphisms, but it is not closed under inverses of morphisms.*

We now give an algebraic property of the syntactic morphism of intermixed languages, which is the main result of [2].

Theorem 7. *Let $\eta : A^* \to M$ be the syntactic morphism of a regular language of A^* and let $x, y \in \eta(A) \cup \{1\}$. If L is intermixed, then $x^{\omega+1} = x^\omega$ and $(x^\omega y^\omega)^{\omega+1} = (x^\omega y^\omega)^\omega$.*

Theorem 7 shows that 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 class remains open.

The reader will find in [2] several partial results on subclasses of the class of intermixed languages, but only one of these subclasses is known to be decidable. It is actually a rather small class in which the use of the shuffle is restricted to shuffling a language with a letter.

It is shown in [2] that the smallest class with these properties is the class of commutative star-free languages. Let us set aside this case by considering classes containing at least one noncommutative language. In fact, for technical reasons which are partly justified by [2, Proposition 4.1], our classes will always contain the languages of the form $\{ab\}$ where a and b are two distinct letters of the alphabet.

In summary, we consider the smallest class of languages $\mathcal{C}$ 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 following results are obtained in [2].

Proposition 10. *A language L belongs to $\mathcal{C}$ if and only if there exists a star-free commutative language C such that the symmetric difference $L \bigtriangleup C$ is finite.*

In view of this result, it is natural to call *almost star-free commutative* the languages of the class $\mathcal{C}$. These languages admit the following algebraic characterisation.

Theorem 8. *Let $\eta : A^* \to M$ be the syntactic morphism of a regular language of A^* and let $x, y, z \in \eta(A^+)$. Then L is almost star-free commutative if and only if $x^\omega = x^{\omega+1}$, $x^\omega y = yx^\omega$ and $x^\omega yz = x^\omega zy$.*

Corollary 2. *It is decidable whether a given regular language is almost star-free commutative.*

4 Sequential and parallel decompositions

We now switch to a different topic, which is still related to the shuffle product. Sequential and parallel decompositions of languages were introduced by Schnoebelen [14] for some model-cheking applications. A reminder of the notions of rational and recognizable subsets of a monoid is in order to define these decompositions properly.

Let M be a monoid. A subset P of M is *recognizable* if there exist a finite monoid F and a monoid morphism $\varphi : M \to F$ such that $P = \varphi^{-1}(\varphi(P))$. It is well known that the class $\mathrm{Rec}(M)$ of recognizable subsets of M is closed under finite union, finite intersection and complement.

The class $\mathrm{Rat}(M)$ of *rational* subsets of M is the smallest set $\mathcal{R}$ of subsets of M satisfying the following properties:

(1) For each $m \in M$, $\{m\} \in \mathcal{R}$
(2) The empty set belongs to $\mathcal{R}$, and if X, Y are in $\mathcal{R}$, then $X \cup Y$ and XY are also in $\mathcal{R}$.
(3) If $X \in \mathcal{R}$, the submonoid X^* generated by X is also in $\mathcal{R}$.

Let τ and σ be the transductions from A^* into $A^* \times A^*$ defined as follows:

$$\tau(w) = \{(u, v) \in A^* \times A^* \mid w = uv\}$$
$$\sigma(w) = \{(u, v) \in A^* \times A^* \mid w \in u \sqcup\!\sqcup v\}$$

Observe that σ is a monoid morphism from A^* into the monoid $\mathcal{P}(A^* \times A^*)$, that is, $\sigma(x_1x_2) = \sigma(x_1)\sigma(x_2)$ for all $x_1, x_2 \in A^*$.

4.1 Definitions and examples

We are now ready to give the definitions of the two types of decomposition. Let $\mathcal{S}$ be a set of languages. A language K admits a *sequential decomposition* over $\mathcal{S}$ if $\tau(K)$ is a finite union of sets of the form $L \times R$, where $L, R \in \mathcal{S}$.

A language K admits a *parallel decomposition* over $\mathcal{S}$ if $\sigma(K)$ is a finite union of sets of the form $L \times R$, where $L, R \in \mathcal{S}$.

A *sequential (resp. parallel) system* is a finite set $\mathcal{S}$ of languages such that each member of $\mathcal{S}$ admits a sequential (resp. parallel) decomposition over $\mathcal{S}$. A language is *sequentially decomposable* if it belongs to some sequential system. It is *decomposable* if it belongs to a system which is both sequential and parallel. Thus, for each decomposable language L, one can find a sequential and parallel system $\mathcal{S}(L)$ containing L.

Example 1. Let $K = \{abc\}$. Then

$$\tau(K) = (\{1\} \times \{abc\}) \cup (\{a\} \times \{bc\}) \cup (\{ab\} \times \{c\}) \cup (\{abc\} \times \{1\})$$

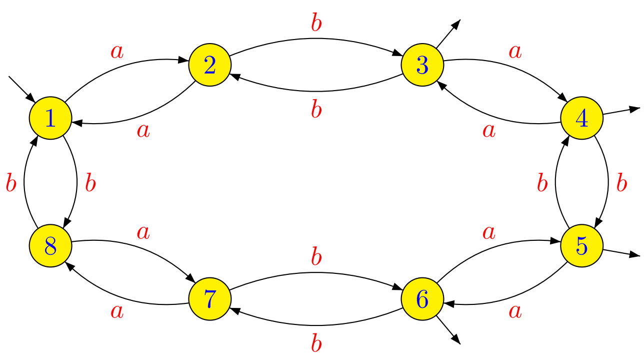
and

$$\begin{aligned}\sigma(K) = (\{1\} \times \{abc\}) &\cup (\{a\} \times \{bc\}) \\ &\cup (\{b\} \times \{ac\}) \cup (\{c\} \times \{ab\}) \cup (\{ab\} \times \{c\}) \\ &\qquad \cup (\{bc\} \times \{a\}) \cup (\{ac\} \times \{b\}) \cup (\{abc\} \times \{1\})\end{aligned}$$

One can verify that the set $\mathcal{S} = \big\{\{1\}, \{a\}, \{b\}, \{c\}, \{ab\}, \{ac\}, \{bc\}, \{abc\}\big\}$ is a sequential and parallel system, and hence K is a decomposable language.

Here is a more complex example. Recall that a word $u = a_1a_2 \cdots a_n$ (where $a_1, \ldots, a_n$ are letters) is a *subword* of a word v if v can be factored as $v = v_0a_1v_1 \cdots a_nv_n$. For instance, ab is a subword of $cacbc$. Given two words u and v, let $\binom{v}{u}$ denote the number of distinct ways to write u as a subword of v.

Example 2. Let L be the set of words of $\{a, b\}^*$ having ab as a subword an odd number of times. Its minimal automaton is represented below:



The transition monoid of L is the dihedral group D_4, a non-abelian group of order 8. For $i, j, k \in \{0, 1\}$ and $c \in A$, let

$$\begin{aligned}M_k^{i,j} &= \Big\{x \in A^* \mid |x|_a \equiv i \bmod 2,\ |x|_b \equiv j \bmod 2 \text{ and } \binom{x}{ab} \equiv k \bmod 2\Big\} \\ M^{i,j} &= \Big\{x \in A^* \mid |x|_a \equiv i \bmod 2,\ |x|_b \equiv j \bmod 2\Big\} \\ M_c^{i,j} &= M^{i,j} \cap A^*cA^*\end{aligned}$$

Let $\mathcal{F}$ be the set of finite union of languages of the form $M_k^{i,j}$, $M_c^{i,j}$ or $\{1\}$. A non-trivial verification [5] shows that $\mathcal{F}$ is a sequential and parallel system for L . Thus L is a decomposable language.

4.2 Closure properties

The following result is stated in [5], but partly relies on results from [14], where (3) implies (1) is credited to Arnold and Carton.

Theorem 9. *Let K be a language. The following conditions are equivalent:*

(1) K is regular,
(2) $\tau(K)$ is recognizable,
(3) K is sequentially decomposable.

Consequently, if K is decomposable, then K is regular and $\sigma(K)$ is recognizable.

As observed by Schnoebelen, it follows that the language $(ab)^*$ is not decomposable, since the set $\sigma((ab)^*)$ is not recognizable.

The following theorem summarises the closure properties of decomposable languages obtained in [14] and [5].

Theorem 10. *The class of decomposable languages is closed under finite union, product, shuffle and left and right quotients. It is not closed under intersection, complementation and star. It is also closed under inverses of length preserving morphisms, but not under inverses of morphisms.*

The negative parts of this theorem are obtained from the following counterexamples: the languages $(ab)^+ \cup (ab)^*bA^*$ and $(ab)^+ \cup (ab)^*aaA^*$ are decomposable but their intersection $(ab)^+$ is not. Furthermore, the language $L = (aab)^* \cup A^*b(aa)^*abA^*$ is decomposable, but if $\varphi\colon A^* \to A^*$ is the morphism defined by $\varphi(a) = aa$ and $\varphi(b) = b$, then $\varphi^{-1}(L) = (ab)^*$ is not decomposable.

4.3 Schnoebelen's problem

Schnoebelen [14] asked for a description of the class of decomposable languages, which implicitly leads to the following problem:

Problem 5. Is it decidable to know whether a regular language is decomposable?

As a first step, Schnoebelen [14] proved the following result.

Proposition 11. *Every commutative regular language is decomposable.*

Denote by $\mathrm{Pol}(\mathcal{C}om)$ the *polynomial closure* of the class of commutative regular languages, that is, the finite unions of products of commutative regular languages. Since, by Theorem 10, decomposable languages are closed under finite union and product, Proposition 11 can be improved as follows:

Theorem 11 (Schnoebelen). *Every language of* $\mathrm{Pol}(\mathcal{C}om)$ *is decomposable.*

Schnoebelen originally conjectured that a language is decomposable if and only if it belongs to Pol($\mathcal{C}om$). However, this conjecture has been refuted in [5], where it is shown that the decomposable language of Example 2 is not in Pol($\mathcal{C}om$).

Describing the class of decomposable languages seems to be a difficult question and Problem 5 is still widely open. One could hope for an algebraic approach, but decomposable languages do not form a positive variety of languages for two reasons. First, they are not closed under inverses of morphisms. This is a minor issue, since they are closed under inverses of renamings, and one could still hope to use Straubing's positive *lp*-varieties instead (see [17] for more details). However, they are also not closed under intersection, and hence we may have to rely on the conjunctive varieties defined by Klíma and Polák [9].

Even if decomposable languages are not closed under intersection, a weaker closure property still holds.

Proposition 12 (Arnold). *The intersection of a decomposable language with a commutative regular language is decomposable.*

This result can be used to give a non-trivial example of indecomposable language.

Proposition 13. *Let $A = \{a, b, c\}$. The language $(ab)^*cA^*$ is not decomposable.*

Proof. Let $L = (ab)^*cA^*$. If L is decomposable, the language

$$Lc^{-1} = (ab)^* \cup (ab)^*cA^*$$

is decomposable by Theorem 10. The intersection of this language with the commutative regular language $\{a, b\}^*$ is equal to $(ab)^*$, and thus by Proposition 12, $(ab)^*$ should also be decomposable. But we have seen this is not the case and hence L is not decomposable.

Let us conclude this section with a conjecture. A *group language* is a regular language recognized by a finite group. Let Pol($\mathcal{G}$) be the *polynomial closure* of the class of group languages, that is, the finite unions of languages of the form $L_0a_1L_1 \cdots a_nL_n$, where each L_i is a group language and the a_i's are letters. The class Pol($\mathcal{G}$) is a well studied positive variety, with a simple characterisation: a regular language belongs to Pol($\mathcal{G}$) if and only if, in its ordered syntactic monoid, the relation $1 \leqslant e$ holds for all idempotents e. We propose the following conjecture as a generalisation of Example 2:

Conjecture 1. Every language of Pol($\mathcal{G}$) is decomposable.

Since decomposable languages are closed under finite union and product, it would suffice to prove that every group language is decomposable. The following result could potentially help solve the conjecture.

Proposition 14. *Let G be a finite group, let $\pi : A^* \to G$ be a surjective morphism and let $L = \pi^{-1}(1)$.*

(1) If the language L is decomposable, then every language recognized by π is decomposable.

(2) The following formula holds

$$\sigma(L) = \bigcup_{\substack{r,s \leqslant |G|^4 \\ (a_1 \cdots a_r \sqcup\!\sqcup b_1 \cdots b_s) \cap L \neq \emptyset}} (La_1La_2L \cdots La_rL) \times (Lb_1Lb_2L \cdots Lb_sL)$$

The bound $|G|^4$ is probably not optimal. If it could be improved to $|G|$, this may lead to a parallel system containing L.

5 Conclusion

The problems presented in this article give evidence that there is still a lot to be done in the study of the shuffle product, even for regular languages. We urge the reader to try to solve them!

References

1. Almeida, J., Cano, A., Klíma, O., Pin, J.É.: Fixed points of the lower set operator. Internat. J. Algebra Comput. **25**(1-2), 259–292 (2015)
2. Berstel, J., Boasson, L., Carton, O., Pin, J.É., Restivo, A.: The expressive power of the shuffle product. Information and Computation **208**, 1258–1272 (2010)
3. Cano, A., Pin, J.É.: Upper set monoids and length preserving morphisms. J. of Pure and Applied Algebra **216**, 1178–1183 (2012)
4. Cano Gómez, A., Pin, J.É.: Shuffle on positive varieties of languages. Theoret. Comp. Sci. **312**, 433–461 (2004)
5. Cano Gómez, A., Pin, J.É.: On a conjecture of Schnoebelen. In: Ésik, Z., Fülöp, Z. (eds.) DLT 2003. pp. 35–54. No. 2710 in Lect. Notes Comp. Sci., Springer, Berlin (2003)
6. Cano Gómez, A., Pin, J.É.: A robust class of regular languages. In: Ochmanski, E., Tyszkiewicz, J. (eds.) Mathematical Foundations of Computer Science 2008, 33rd International Symposium, MFCS 2008, Torun, Poland, August 25-29, 2008, Proceedings. Lect. Notes Comp. Sci., vol. 5162, pp. 36–51. Springer, Berlin (2008)
7. Castiglione, G., Restivo, A.: On the shuffle of star-free languages. Fund. Inform. **116**(1-4), 35–44 (2012)
8. Ésik, Z., Simon, I.: Modeling literal morphisms by shuffle. Semigroup Forum **56**(2), 225–227 (1998)
9. Klíma, O., Polák, L.: On varieties of meet automata. Theoret. Comput. Sci. **407**(1-3), 278–289 (2008)
10. Perrot, J.F.: Variétés de langages et opérations. Theoret. Comp. Sci. **7**, 197–210 (1978)
11. Pin, J.É.: A variety theorem without complementation. Russian Mathematics (Izvestija vuzov.Matematika) **39**, 80–90 (1995)
12. Restivo, A.: The shuffle product: new research directions. In: Language and automata theory and applications, Lecture Notes in Comput. Sci., vol. 8977, pp. 70–81. Springer, Cham (2015)
13. Reutenauer, C.: Sur les variétés de langages et de monoïdes. In: Theoretical computer science (Fourth GI Conf., Aachen), vol. 67, pp. 260–265 (1979)

14. Schnoebelen, P.: Decomposable regular languages and the shuffle operator. EATCS Bull. (67), 283–289 (1999)
15. Schützenberger, M.P.: On finite monoids having only trivial subgroups. Information and Control **8**, 190–194 (1965)
16. Straubing, H.: Recognizable sets and power sets of finite semigroups. Semigroup Forum **18**, 331–340 (1979)
17. Straubing, H., Weil, P.: Varieties. In: Pin, J.É. (ed.) Handbook of automata theory. Volume I. Theoretical foundations, pp. 569–614. Berlin: European Mathematical Society (EMS) (2021)