Shuffle product of regular languages: results and open problems

Jean-Éric Pin[1]

[1]IRIF, CNRS and Université de Paris Cité

CAI, October 2022, Thessaloniki (on line)

Table of Contents

Syntactic monoid

A language L is recognized by a monoid M if there is a monoid morphism $h : A^* \to M$ and a subset P of M such that $L = h^{-1}(P)$.

The syntactic congruence of L is the equivalence relation $\sim_L$ on A^* defined as follows: $u \sim_L v$ iff, for every $x, y \in A^*$, xuy and xvy are either both in L or both outside of L. The syntactic monoid of L is the quotient monoid $A^*/\sim_L$.

The syntactic monoid of a regular language is also the transition monoid of its minimal automaton.

Shuffle product

The shuffle of two words u and v of A^* is the set $u \sqcup\!\sqcup v$ of words of A^* of the form $u_1v_1 \cdots u_nv_n$, with $n \geqslant 0$, $u_1 \cdots u_n = u$, $v_1 \cdots v_n = v$.

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

The shuffle of two languages K and L of is the language

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

The shuffle product is commutative, associative and distributes over union.

Renaming

A renaming $\varphi\colon A^* \to B^*$ is a length-preserving morphism. 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 set $\mathcal{P}(M)$ of subsets of a monoid M is a monoid under the product of subsets defined by

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

Proposition

Let $\varphi\colon A^* \to B^*$ be a renaming. If L is recognized by M, then $\varphi(L)$ is recognized by $\mathcal{P}(M)$.

Shuffle and recognition

Proposition

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 $\mathcal{P}(M_1 \times M_2)$.

Thus power monoids are related to renaming and shuffle. We will try to make this statement more precise in the following slides.

Part I

Varieties of languages

Varieties of languages

A variety of languages is a class of regular languages closed under union, intersection, complement, left and right quotients and inverses of morphisms.

Eilenberg gave a bijection between varieties of languages and varieties of finite monoids.

A variety of finite monoids is a class of finite monoids closed under taking submonoids, quotients and finite direct products.

For instance, star-free languages correspond to aperiodic monoids.

Varieties of languages closed under renaming (1)

Given 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 $M \in \mathbf{V}$. $\mathbf{V}$ is a fixed point of the operator $\mathbf{P}$ if $\mathbf{P}(\mathbf{V}) = \mathbf{V}$

Proposition (Reutenauer 1979, Straubing 1979)

A variety of languages is closed under renaming iff the corresponding variety of finite monoids is a fixed point of the operator $\mathbf{P}$.

Consequence. Every variety of languages closed under renaming is also closed under shuffle.

Varieties of languages closed under renaming (2)

Let $[u]$ denote the commutative closure of a word u. For instance, $[aab] = \{aab, aba, baa\}$. A language L is commutative if, for every word $u \in L$, $[u]$ is contained in L.

(1) A variety of commutative languages is closed under shuffle iff it corresponds to a variety of finite commutative monoids whose groups belong to a given variety of finite commutative groups.

(2) The unique non-commutative variety of languages closed under renaming is the variety of all regular languages (Pin 1980).

Varieties of languages closed under shuffle (1)

An early question (Perrot 1978). Which varieties of languages are closed under shuffle?

(1) The smallest variety of languages closed under shuffle is the variety of commutative star-free languages.

(2) A variety of commutative languages is closed under shuffle iff it corresponds to a variety of finite commutative monoids whose groups belong to a given variety of finite commutative groups.

Example: commutative monoids whose groups are commutative p-groups.

Varieties of languages closed under shuffle (2)

Perrot (1978) conjectured that no other variety of languages (except the regular languages) is closed under shuffle.

Theorem (Ésik-Simon 1998)

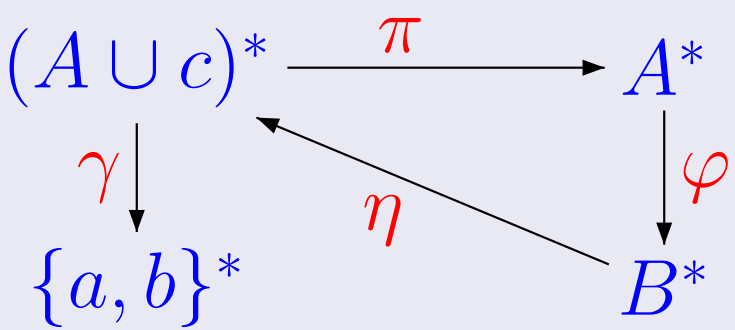
The unique non-commutative variety of languages closed under shuffle is the variety of all regular languages.

Simulating a renaming through a shuffle

Let $\varphi\colon A^* \to B^*$ be a surjective renaming and let c be a new letter.

Proposition (Ésik-Simon 1998)

There exist monoid morphisms π, γ and η



such that

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

Consequence

It follows that if a variety of languages contains the language $(ab)^*$ and is closed under shuffle, then it is also closed under renaming, a key argument in the proof of the theorem of Ésik and Simon.

Corollary

A variety of languages is closed under shuffle iff it is closed under renaming.

Part II

Positive varieties

A positive variety of languages is a class of regular languages closed under union, intersection, left and right quotients and inverses of morphisms.

There is a bijection between varieties of languages and varieties of finite ordered monoids (Pin 1995).

Ordered monoids

An ordered monoid is a monoid M equipped with a stable partial order $\leqslant$ on M:

$$x \leqslant y \implies xz \leqslant yz \text{ and } zx \leqslant zy$$

Morphisms of ordered monoids are order-preserving.

A variety of finite ordered monoids is a class of finite ordered monoids closed under taking ordered submonoids, quotients and finite direct products.

Recognition by ordered monoids

A subset U of an ordered monoid $(M, \leqslant)$ is an upper set if $s \in U$ and $s \leqslant t$ imply $t \in U$.

An ordered monoid $(M, \leqslant)$ recognizes a language L of A^* if there exist a morphism $\varphi \colon A^* \to M$ and an upper set U of M such that $L = \varphi^{-1}(U)$.

Lower set monoids

A subset P of an ordered monoid $(M, \leqslant)$ is a lower set if $s \in P$ and $t \leqslant s$ imply $t \in P$.

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.

Varieties of finite ordered monoids

Given 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 $M \in \mathbf{V}$. Then $\mathbf{V}$ is a fixed point of the operator $\mathbf{P}^{\downarrow}$ if $\mathbf{P}^{\downarrow}\mathbf{V} = \mathbf{V}$.

Proposition (Polák 2002, Cano-Gómez, Pin 2004)

A positive variety of languages is closed under renaming iff the corresponding variety of finite ordered monoids is a fixed point of the operator $\mathbf{P}^{\downarrow}$.

Comparison of the operators $\mathbf{P}$ and $\mathbf{P}^{\downarrow}$

Theorem (Margolis, Pin 1984)

The operator $\mathbf{P}$ *satisfies* $\mathbf{P}^3 = \mathbf{P}^4$.

Theorem (Cano-Gómez, Pin 2012)

The operator $\mathbf{P}^{\downarrow}$ *is idempotent.*

Positive varieties closed under shuffle

Corollary

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.

But contrary to the case of varieties, this corollary only gives a sufficient condition for a positive variety of languages to be closed under shuffle.

Maximal positive variety closed under shuffle

The unique maximal proper variety of languages closed under shuffle is the variety of commutative languages. For positive varieties, one gets

Theorem (Cano-Gómez, Pin 2004)

There is a largest proper positive variety of languages closed under shuffle. It is the largest positive variety of languages $\mathcal{W}$ not containing the language $(ab)^*$.

Let $\mathbf{W}$ be the variety of finite ordered monoids corresponding to $\mathcal{W}$.

Maximal positive variety closed under shuffle

In a finite monoid, every element x has a unique idempotent power, denoted by x^ω.

Theorem (Cano-Gómez, Pin 2004)

A finite ordered monoid M belongs to $\mathbf{W}$ iff, 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, $(stzst)^\omega \leqslant st$.

Corollary

$\mathbf{W}$ (and hence $\mathcal{W}$) is decidable.

Robustness of $\mathcal{W}$

Theorem

The positive variety $\mathcal{W}$ is closed under product, shuffle and renaming.

It can also be defined as the largest proper positive variety of languages satisfying one of (1), (2) or (3):

(1) not containing the language $(ab)^*$;

(2) closed under shuffle;

(3) closed under renaming;

Open problems

Problem

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

Problem

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

This is the case for $\mathcal{W}$, but the general case is unknown.

Fixed points of the operator $\mathbf{P}^{\downarrow}$

An in-depth study can be found in [Almeida, Cano-Gómez, Klíma, Pin 2015].

Proposition

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

The article ACKP15 gives six independent basic types of such fixed points, from which many more may be constructed using intersection. It is conjectured that all fixed points of $\mathbf{P}^{\downarrow}$ can be obtained in this way.

Part III

Intermixed languages

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

Intermixed languages

The class 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.

Problem

Can one decide whether a given regular language is intermixed?

This problem is still widely open, and only partial results are known.

Star-free languages

Star-free languages are intermixed by definition, but are not closed under shuffle: $(abb)^* \sqcup\!\sqcup a^*$ is not star-free since $\left((abb)^* \sqcup\!\sqcup a^*\right) \cap (ab)^* = (abab)^*$.

Problem (Castiglione and Restivo 2012)

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

Theorem

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

Properties of intermixed languages

Proposition

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

In particular, intermixed languages do not form a variety of languages. However, an algebraic approach is still possible, using ld-varieties (Ésik, Straubing).

A property of intermixed languages

Theorem (Berstel, Boasson, Carton, Pin, Restivo 2010)

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

$$x^{\omega+1} = x^\omega \text{ and } (x^\omega y^\omega)^{\omega+1} = (x^\omega y^\omega)^\omega.$$

Consequence: the languages $(aa)^*$ and $(a^+b^+a^+b^+)^*$ are not intermixed.

Consequences

Corollary

Intermixed languages form a proper subclass of the class of regular languages.

Unfortunately, we do not know whether the equations

$$x^{\omega+1} = x^{\omega} \text{ and } (x^{\omega}y^{\omega})^{\omega+1} = (x^{\omega}y^{\omega})^{\omega}.$$

characterise the intermixed languages and hence the decidability of this class remains open.

Conclusion

(1) The varieties closed under shuffle are completely classified.

(2) Positive varieties closed under shuffle are reasonably well-known but a better description of the positive variety $\mathcal{W}$ is still missing.

(3) The decidability of the class of intermixed languages is still unknown.

Part IV

Sequential and parallel decompositions

Two transductions

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

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

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 $L, R \in \mathcal{S}$.

An example

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

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

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

K has a sequential and parallel decomposition over

$$\mathcal{S} = \{\{1\}, \{a\}, \{b\}, \{c\}, \{ab\}, \{ac\}, \{bc\}, \{abc\}\}.$$

Decomposable languages

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.

Problem (Schnoebelen 1999)

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

Decomposable languages (2)

Theorem (CGP 2003, Arnold, Carton, Schnoebelen)

Let K be a language. TFCAE:

(1) K *is regular,*

(2) $\tau(K)$ *is recognizable,*

(3) K *is sequentially decomposable.*

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

Example [Schnoebelen] $(ab)^*$ is not decomposable, since $\sigma((ab)^*)$ is not recognizable.

Closure properties

Theorem (CGP 2003, Schnoebelen 1999)

The class of decomposable languages is closed under union, product, shuffle, left and right quotients and inverses of length preserving morphisms. It is not closed under intersection, complementation and inverses of morphisms.

Proposition (Schnoebelen 1999)

Commutative regular languages are decomposable.

Some consequences

Proposition (Arnold)

The intersection of a decomposable language with a commutative regular language is decomposable.

Let $\mathrm{Pol}(\mathcal{C}om)$ be the polynomial closure of the class of commutative regular languages ($=$ finite unions of products of commutative regular languages).

Proposition (Schnoebelen 1999)

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

A conjecture

A group language is a regular language recognized by a finite group.

Let $\mathrm{Pol}(\mathcal{G})$ be the polynomial closure of the class of group languages, that is, the finite unions of languages of the form $L_0 a_1 L_1 \cdots a_n L_n$, where each L_i is a group language and the a_i's are letters.

Conjecture

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

It would suffice to prove that every group language is decomposable.

Toward the conjecture

Proposition

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

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

(2) *The following formula holds*

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

The bound $|G|^4$ is probably not optimal. Would $|G|$ be the optimal bound?