On a conjecture of Schnoebelen

Antonio Cano Gómez[1] and Jean-Éric Pin[2]

[1] Departamento de Sistemas Informáticos y Computación, Universidad Politénica de Valencia, Camino de Vera s/n, P.O. Box: 22012, E-46020 - Valencia, `acano@dsic.upv.es`,

[2] LIAFA, Université Paris VII and CNRS, Case 7014, 2 Place Jussieu, 75251 Paris Cedex 05, France. Work supported by INTAS project 1224, `Jean-Eric.Pin@liafa.jussieu.fr`, `http://liafa.jussieu.fr/~jep`

Abstract. The notion of sequential and parallel decomposition of a language over a set of languages was introduced by Schnoebelen. A language is decomposable if it belongs to a finite set of languages $\mathcal{S}$ such that each member of $\mathcal{S}$ admits a sequential and parallel decomposition over $\mathcal{S}$. We disprove a conjecture of Schnoebelen concerning decomposable languages and establish some new properties of these languages.

1 Introduction

The shuffle product is a standard tool for modeling process algebras [2]. This motivates the study of "robust" classes of recognizable languages which are closed under shuffle product.

In [6], Schnoebelen introduced the notion of *sequential and parallel decomposition* (a precise definition is given in Section 3). A language is decomposable if it belongs to a finite set of languages $\mathcal{S}$ such that each member of $\mathcal{S}$ admits a sequential and parallel decomposition over $\mathcal{S}$. Schnoebelen proved that the class of decomposable languages is a quite robust class of rational languages, since it is closed under finite union, product and shuffle.

It is a challenging question to find an effective characterization of this class of languages. In [6], Schnoebelen proved that a finite union of products of commutative languages are decomposable and conjectured there are no other decomposable languages.

In this paper, we disprove this conjecture and establish some new properties of decomposable languages: they are closed under left and right quotients, but are not closed under intersection. They are also closed under inverse length preserving morphisms, but not under inverse morphisms. Finally, we propose a new conjecture for the description of decomposable languages.

2 Preliminaries

In this section, we recall a few important results about recognizable and rational sets. Let M be a monoid. A subset P of M is *recognizable* if there exists a finite monoid F, and a monoid morphism $\varphi : M \to F$ such that $P = \varphi^{-1}(\varphi(P))$.

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 $\mathcal{R}$.

Recognizable sets are closed under boolean operations, quotients and inverse morphisms. Rational sets are closed under rational operations (union, product and star) and under morphisms. Kleene theorem states that for every finite alphabet A, $\mathrm{Rec}(A^*) = \mathrm{Rat}(A^*)$. We conclude by three standard results.

Theorem 1. (McKnight) *The intersection of a rational set and of a recognizable set is rational.*

Theorem 2. (Mezei) *Let $M_1, \ldots, M_n$ be monoids. A subset of $M_1 \times \cdots \times M_n$ is recognizable if and only if it is a finite union of subsets of the form $R_1 \times \cdots \times R_n$, where $R_i \in \mathrm{Rec}(M_i)$.*

Proposition 1. *Let $A_1, \ldots, A_n$ be finite alphabets. Then $\mathrm{Rec}(A_1^* \times A_2^* \times \cdots \times A_n^*)$ is closed under product.*

A *substitution* from A^* into a monoid M is a monoid morphism from A^* into $\mathcal{P}(M)$, the monoid of subsets of M.

3 Decompositions of languages

Schnoebelen [6] introduced the following definitions. Consider the transductions τ and σ from A^* into $A^* \times A^*$ defined as follows:

$$\begin{aligned}\tau(w) &= \{(u, v) \in A^* \times A^* \mid w = uv\}\\ \sigma(w) &= \{(u, v) \in A^* \times A^* \mid w \in u \mathbin{\text{Ш}} v\}\end{aligned}$$

Therefore, if K is a language

$$\begin{aligned}\tau(K) &= \{(u, v) \in A^* \times A^* \mid uv \in K\}\\ \sigma(K) &= \{(u, v) \in A^* \times A^* \mid (u \mathbin{\text{Ш}} v) \cap K \neq \emptyset\}\end{aligned}$$

Proposition 2. *The transduction σ is a substitution.*

Proof. We claim that $\sigma(x_1x_2) = \sigma(x_1)\sigma(x_2)$. First, if $(u_1, v_1) \in \sigma(x_1)$ and $(u_2, v_2) \in \sigma(x_2)$, then $x_1 \in u_1 \mathbin{\text{Ш}} v_1$ and $x_2 \in u_2 \mathbin{\text{Ш}} v_2$. It follows that $x_1x_2 \in u_1u_2 \mathbin{\text{Ш}} u_2v_2$ and thus $(u_1u_2, v_1v_2) \in \sigma(x_1x_2)$. Conversely, if $(u, v) \in \sigma(x_1x_2)$, then $x_1x_2 \in u \mathbin{\text{Ш}} v$. Therefore, u and v can be decomposed as $u = u_1u_2$, $v = v_1v_2$ in such a way that $x_1 \in u_1 \mathbin{\text{Ш}} v_1$ and $x_2 \in u_2 \mathbin{\text{Ш}} v_2$. It follows that $(u, v) \in \sigma(x_1)\sigma(x_2)$.

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 $L, R \in \mathcal{S}$. Sequential decompositions of recognizable languages where considered by Conway [4]. Actually, Conway proved that every recognizable language K admits a *maximal* sequential decomposition

$$\tau(K) = \bigcup_{1 \le i \le n} L_i \times R_i$$

in the following sense. If $LR \subseteq K$, then for some i, $L \subseteq L_i$ and $R \subseteq R_i$.

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.

Theorem 3. *Let K be a language of A^*. The following conditions are equivalent:*

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

Proof. (1) implies (3). Let K be a rational language and let $\mathcal{A}$ be the minimal automaton of K. For each state p, q of $\mathcal{A}$, let $K_{p,q}$ be the language accepted by $\mathcal{A}$ with p as initial state and q as unique final state. Let $\mathcal{S}$ be the set of finite unions of languages of the form $K_{p,q}$. Since $\mathcal{A}$ is finite, $\mathcal{S}$ is a finite set. We claim that

$$\tau(K_{p,q}) = \bigcup_{r \in Q} K_{p,r} \times K_{r,q} \tag{1}$$

Indeed, if, for some state r, $u \in K_{p,r}$ and $v \in K_{r,q}$, then $p \cdot u = r$ and $r \cdot v = q$ in $\mathcal{A}$. It follows that $uv \in K_{p,q}$ and $(u, v) \in \tau(K_{p,q})$. Conversely, let $(u, v) \in \tau(K_{p,q})$ and let $r = p \cdot u$. Since $uv \in K_{p,q}$, $p \cdot uv = q$, and thus $r \cdot v = q$, whence $v \in K_{r,q}$. Thus $(u, v) \in K_{p,r} \times K_{r,q}$.
It follows from (1) that each language of $\mathcal{S}$ admits a sequential decomposition on $\mathcal{S}$. Thus $\mathcal{S}$ is a sequential system and K itself is sequentially decomposable.

(3) implies (1) (Arnold, Carton). Let $\mathcal{S}$ be a sequential system containing K, and let $\sim_{\mathcal{S}}$ be the equivalence on A^* defined by $u \sim_{\mathcal{S}} v$ if, for every $F \in \mathcal{S}$, $u \in F$ is equivalent to $v \in F$. Clearly, $\sim_{\mathcal{S}}$ is an equivalence of finite index, which saturates K by definition. Therefore, it suffices to show that $\sim_{\mathcal{S}}$ is a congruence. Suppose that $u \sim_{\mathcal{S}} v$ and let F be a language of $\mathcal{S}$. Let $w \in A^*$ and suppose that $uw \in F$. Since F is sequentially decomposable over $\mathcal{S}$, there exist two languages $L, R \in \mathcal{S}$ such that $u \in L$, $w \in R$ and $LR \subset F$. It follows that $v \in L$, since

$u \sim_{\mathcal{S}} v$, and thus $vw \in F$. Similarly, $vw \in F$ implies $uw \in F$ and thus $uw \sim_{\mathcal{S}} vw$. A dual argument would show that $wu \sim_{\mathcal{S}} wv$, which concludes the proof.

(2) implies (1). Suppose that $\tau(K)$ is recognizable. Observing that

$$\tau(K) \cap \{1\} \times A^* = \{1\} \times K$$

it follows by McKnight's theorem, that $\{1\} \times K$ is rational. Now $K = \pi(\{1\} \times K)$, where π denotes the second projection from $A^* \times A^*$ onto A^*. It follows that K is rational.

(3) implies (2). If K is sequentially decomposable, it belongs to some sequential system $\mathcal{S}$. By definition, each element of $\mathcal{S}$ is sequentially decomposable and thus is rational, since (3) implies (1). Furthermore, $\tau(K)$ is a finite union of languages of the form $L \times R$, where $L, R \in \mathcal{S}$. It follows, by Mezei's theorem, that $\tau(K)$ is recognizable.

It follows that every decomposable language is rational. Another important consequence is the following:

Proposition 3. *For each decomposable language K, $\sigma(K)$ is recognizable.*

Proof. Indeed, if $\mathcal{S}$ is a parallel and sequential system for K, every language of $\mathcal{S}$ is decomposable and hence rational, by Theorem 3. Since is sequentially decomposable over $\mathcal{S}$, $\sigma(K)$ is a finite union of languages of the form $L \times R$, where $L, R \in \mathcal{S}$. It follows, by Mezei's theorem, that $\sigma(K)$ is recognizable.

We shall see later that the converse does not hold. That is, there exist a rational, indecomposable language K such that $\sigma(K)$ is recognizable. However, Proposition 3 remains a powerful tool for proving that a language is not decomposable. The most important example was already given in [6].

Proposition 4. (Schnoebelen) *The set $\sigma((ab)^*)$ is not recognizable. In particular, the language $(ab)^*$ is not decomposable.*

Proof. Let $K = (ab)^*$. Suppose that

$$\sigma(K) = \bigcup_{1 \leq i \leq n} L_i \times R_i$$

Then, for every $k \geq 0$, $(a^k, b^k) \in \sigma(K)$ since $(ab)^k \in a^k \text{ Ш } b^k$. It follows that one of the blocks $L_i \times R_i$ contains two distinct pairs (a^r, b^r) and (a^s, b^s) with $r \neq s$. It follows that (a^r, b^s) also belongs to this block, and hence $a^r \text{ Ш } b^s$ should contain a word of $(ab)^*$, a contradiction.

4 Closure properties of decomposable languages

The following closure properties were established in [6], where the closure under shuffle is credited to Arnold. For the convenience of the reader, we give a self-contained proof.

Proposition 5. *Decomposable languages are closed under finite union, product and shuffle.*

Proof. Let K and K' be decomposable languages. Let $\mathcal{S}$ (resp. $\mathcal{S}'$) be a sequential and parallel system for K (resp. K'). Then $\mathcal{S} \cup \mathcal{S}' \cup \{K \cup K'\}$ is a sequential and parallel system for $K \cup K'$. Let us show that

$$\mathcal{R} = \mathcal{S} \cup \mathcal{S}' \cup \{XX' \mid X \in \mathcal{S} \text{ and } X' \in \mathcal{S}'\}$$

is a sequential and parallel system for KK'. Assume that

$$\sigma(K) = \bigcup_{1 \leq i \leq n} L_i \times R_i \quad \text{and} \quad \sigma(K') = \bigcup_{1 \leq j \leq n'} L'_j \times R'_j$$

Since σ is a substitution by Proposition 2,

$$\sigma(KK') = \bigcup_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n'}} L_i L'_j \times R_i R'_j$$

Furthermore, if

$$\tau(K) = \bigcup_{1 \leq i \leq n} L_i \times R_i \quad \text{and} \quad \tau(K') = \bigcup_{1 \leq j \leq n'} L'_j \times R'_j$$

then

$$\tau(KK') = \bigcup_{1 \leq i \leq n} (L_i \times R_i K') \cup \bigcup_{1 \leq j \leq n} (KL'_i \times R'_i)$$

It follows that KK' is decomposable. Finally, the system

$$\mathcal{T} = \{X \text{ Ш } X' \mid X \in \mathcal{S} \text{ and } X' \in \mathcal{S}'\}$$

is a sequential and parallel system for K Ш K'. Indeed, we have, with the previous notation,

$$\tau(K \text{ Ш } K') = \bigcup_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n'}} (L_i \text{ Ш } L'_j) \times (R_i \text{ Ш } R'_j)$$

$$\sigma(K \text{ Ш } K') = \bigcup_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n'}} (L_i \text{ Ш } R'_i) \times (L_j \text{ Ш } R'_j)$$

Therefore K Ш K' is decomposable.

We establish some new closure properties. We first consider inverse morphisms. We shall see later (Proposition 19) that decomposable languages are not closed under inverse morphisms. However, closure under inverse morphisms holds for a restricted class of morphisms. Recall that a morphism $\varphi : A^* \to B^*$ is *length preserving* (or *litteral*, or *strictly alphabetic*) if each letter of A is mapped onto a letter of B.

Lemma 1. *Let $\varphi : A^* \to B^*$ be a morphism and let K be a language of B^*. If*

$$\tau(K) = \bigcup_{1 \le i \le n} L_i \times R_i$$

then

$$\tau(\varphi^{-1}(K)) = \bigcup_{1 \le i \le n} \varphi^{-1}(L_i) \times \varphi^{-1}(R_i)$$

Furthermore, if φ is length preserving, and if

$$\sigma(K) = \bigcup_{1 \le i \le n} L_i \times R_i$$

then

$$\sigma(\varphi^{-1}(K)) = \bigcup_{1 \le i \le n} \varphi^{-1}(L_i) \times \varphi^{-1}(R_i)$$

Proof. The first result follows from the following sequence of equivalences

$$\begin{aligned}
(u, v) \in \tau(\varphi^{-1}(K)) &\iff uv \in \varphi^{-1}(K) \\
&\iff \varphi(uv) \in K \\
&\iff \varphi(u)\varphi(v) \in K \\
&\iff \text{ for some } i \in \{1, .., n\},\ (\varphi(u), \varphi(v)) \in L_i \times R_i \\
&\iff \text{ for some } i \in \{1, .., n\},\ (u, v) \in \varphi^{-1}(L_i) \times \varphi^{-1}(R_i)
\end{aligned}$$

If φ is length preserving, then a word w belongs to u Ш v if and only if $\varphi(w)$ belongs to $\varphi(u)$ Ш $\varphi(v)$. Therefore, the following sequence of equivalences holds:

$$\begin{aligned}
(u, v) \in \sigma(\varphi^{-1}(K)) &\iff (u \text{ Ш } v) \cap \varphi^{-1}(K) \neq \emptyset \\
&\iff (\varphi(u) \text{ Ш } \varphi(v)) \cap K \neq \emptyset \\
&\iff (\varphi(u), \varphi(v)) \in \sigma(K) \\
&\iff \text{ for some } i \in \{1, .., n\},\ (\varphi(u), \varphi(v)) \in L_i \times R_i \\
&\iff \text{ for some } i \in \{1, .., n\},\ (u, v) \in \varphi^{-1}(L_i) \times \varphi^{-1}(R_i)
\end{aligned}$$

This proves the second part of the lemma.

Proposition 6. *Decomposable languages are closed under inverse of length preserving morphisms.*

Proof. Let $\varphi : A^* \to B^*$ be a length preserving (or letter to letter) morphism. Let K be a decomposable language of B^* and let $\mathcal{S}$ be a sequential and parallel system for K. By Lemma 1, the set $\{\varphi^{-1}(L) \mid L \in \mathcal{S}\}$ is a sequential and parallel system for $\varphi^{-1}(K)$.

Proposition 7. *Decomposable languages are closed under left and right quotients.*

Proof. Let $\mathcal{S}$ be a sequential and parallel system for L and let $a \in A$. Let

$$\tau(L) = \bigcup_{1 \le i \le n} L_i \times R_i$$

Observing that $\{a\} \times a^{-1}L \subseteq \tau(L)$, we set

$$J = \{j \mid a \in L_j \text{ and } a^{-1}L \cap R_j \neq \emptyset\}$$

We claim that $a^{-1}L = \bigcup_{j \in J} R_j$. Indeed, if $u \in a^{-1}L$, then $au \in L$ and $(a, u) \in \tau(L)$. Therefore, $(a, u) \in L_j \times R_j$ for some j. But since $a \in L_j$ and $u \in a^{-1}L \cap R_j$, j belongs to J. Thus $u \in \bigcup_{j \in J} R_j$. Conversely, if $u \in R_j$ for some $j \in J$, then $a \in L_j$ and hence $(a, u) \in \tau(L)$, whence $au \in L$ and $u \in a^{-1}L$.

It follows from the claim that $a^{-1}L$ is a finite union of elements of $\mathcal{S}$. Since the elements of $\mathcal{S}$ are decomposable, $a^{-1}L$ is decomposable.

A symmetrical argument would show that La^{-1} is decomposable.

It is tempting to conjecture that a language is decomposable if and only if $\sigma(L)$ is recognizable. However, Proposition 7 can be used to give a counterexample. Let $L = (ab)^* \cup bA^* \cup aaA^*$. Then

$$\sigma(L) = (aA^* \times aA^*) \cup (bA^* \times A^*) \cup (aaA^* \times A^*) \\ \cup (A^* \times aaA^*) \cup (1 \times L) \cup (L \times 1)$$

However, L is not decomposable, since $(ab)^{-1}L = (ab)^*$.

The next result was proved by in [6].

Proposition 8. (Schnoebelen) *Every commutative rational language is decomposable.*

Proof. Let L be a commutative rational language. We claim that $\tau(L) = \sigma(L)$. Indeed, if $uv \in L$, then $u \text{ Ш } v \cap L \neq \emptyset$. Conversely, if $u \text{ Ш } v$ meets L, there exist two facorizations $u = u_0u_1 \cdots u_k$ and $v = v_1v_2 \cdots v_k$ such that $u_0v_1u_1 \cdots v_ku_k \in L$. It follows, since L is commutative, that $uv \in L$. It follows from the claim that any sequential decomposition of L is also a parallel decomposition. Now, by Theorem 3, L has a sequential decomposition. Therefore, L is decomposable.

Recall that the *polynomial closure* of a class of languages $\mathcal{C}$ is the class of languages which are union of products of languages of $\mathcal{C}$. Proposition 5 shows that decomposable languages are closed under polynomial closure. In [6], Schnoebelen proposed the following conjecture.

Conjecture 1. A language is decomposable if and only if it belongs to the polynomial closure of the commutative languages.

Denote by Pol($\mathcal{C}om$) the *polynomial closure* of the commutative languages. This class contains in particular the finite and the cofinite languages. One direction of the conjecture is an immediate consequence of Propositions 8 and 5.

Theorem 4. (Schnoebelen) *Every language of Pol($\mathcal{C}om$) is decomposable.*

We shall give a counterexample to Schnoebelen's conjecture in the next section. Let us first give a last closure property of the class of decomposable languages, due to Arnold. It is based on the following property.

Proposition 9. *Let K and L be languages. Then $\tau(K \cap L) = \tau(K) \cap \tau(L)$. Furthermore, if one of the languages is commutative, $\sigma(K \cap L) = \sigma(K) \cap \sigma(L)$.*

Proof. The first formula follows from the following sequence of equivalent statements:

$$\begin{aligned}(u,v) \in \tau(K \cap L) &\iff uv \in K \cap L \iff uv \in K \text{ and } uv \in L \\ &\iff (u,v) \in \tau(L) \text{ and } (u,v) \in \tau(K)\end{aligned}$$

The inclusion $\sigma(K \cap L) \subseteq \sigma(K) \cap \sigma(L)$ is trivial. Assume that L is commutative and let $(u,v) \in \sigma(K) \cap \sigma(L)$. Since L is commutative, $\sigma(L) = \tau(L)$ and hence $uv \in L$. Furthermore there exist a word w of K such that $w = u_0 v_1 u_1 \cdots v_n u_n$ for some factorizations $u = u_0 u_1 \cdots u_n$ and $v = v_1 v_2 \cdots v_n$ of u and v. Now uv and w are commutatively equivalent, and thus $w \in K \cap L$. It follows that (u,v) belongs to $\sigma(K \cap L)$ and hence $\sigma(K \cap L) = \sigma(K) \cap \sigma(L)$.

Corollary 1. (Arnold) *The intersection of a decomposable language with a commutative recognizable language is decomposable*

Proposition 9 can be used to give non-trivial examples of indecomposable language.

Proposition 10. *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 Proposition 7. The intersection of this language with the recognizable commutative language $\{a,b\}^*$ is equal to $(ab)^*$, and thus by Corollary 1, $(ab)^*$ should also be decomposable. But this contradicts Proposition 4 and thus L is not decomposable.

We shall see later that decomposable languages are not closed under intersection.

5 Schnoebelen's conjecture

In this section, we disprove Schoebelen's conjecture by giving an example of a decomposable language which is not a finite union of products of commutative languages.

An algebraic characterization of the languages of Pol($\mathcal{C}om$) was given in [5], but a much weaker property will suffice for our purpose: Every group in the syntactic monoid of a language of Pol($\mathcal{C}om$) is commutative.

Let u and v be two words of A. A word $u = a_1 a_2 \cdots a_k$ is said to be a *subword* of v if v can be factorized as $v = v_0 a_1 v_1 \cdots a_k v_k$ where $v_0, v_1, \ldots, v_k \in A^*$. Following Eilenberg [3], we set

$$\binom{v}{u} = \mathrm{Card}\{(v_0, v_1, \ldots, v_k) \in A^* \times A^* \times \cdots \times A^* \mid v_0 a_1 v_1 \cdots a_k v_k = v\}$$

Thus $\binom{v}{u}$ is the number of distinct ways to write u as a subword of v. For instance, $\binom{aabbaa}{aba} = 8$ and $\binom{a^n}{a^m} = \binom{n}{m}$. The basic properties of these *binomial coefficients* are summarized by the following formulae

(1) For every word $u \in A^*$, $\binom{u}{1} = 1$.

(2) For every non-empty word $u \in A^*$, $\binom{1}{u} = 0$.

(3) If $w = uv$, then $\binom{w}{x} = \sum_{x_1 x_2 = x} \binom{u}{x_1}\binom{v}{x_2}$.

Let $A = \{a, b\}$ be an alphabet, and let K be the language of words x over the alphabet A such that $\binom{x}{ab} \equiv 1 \mod 2$. The minimal automaton of this language is represented in Figure 1.

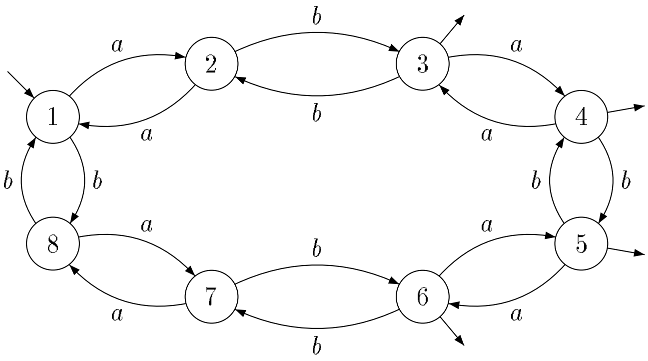

Fig. 1. An automaton for M.

The syntactic monoid of K is a non-abelian group with eight elements. Therefore K does not belong to Pol($\mathcal{C}om$). We now prove that K is decomposable.
For $i, j, k \in \{0, 1\}$ and $c \in A$, let us set

$$M_k^{i,j} = \left\{x \in A^* \mid |x|_a \equiv i \bmod 2,\ |x|_b \equiv j \bmod 2 \text{ and } \binom{x}{ab} \equiv k \bmod 2\right\}$$

$$M^{i,j} = \left\{x \in A^* \mid |x|_a \equiv i \bmod 2,\ |x|_b \equiv j \bmod 2\right\}$$

$$M_c^{i,j} = M^{i,j} \cap A^* c A^*$$

Let $\mathcal{F}$ be the set of languages that are finite union of languages of the form $M_k^{i,j}$, $M_c^{i,j}$ or $\{1\}$. Observe that, since

$$M^{i,j} = M_0^{i,j} \cup M_1^{i,j}$$

$M^{i,j}$ belongs to $\mathcal{F}$. We claim that $\mathcal{F}$ is a sequential and parallel system. We first show it is a sequential system by proving that the languages $M_k^{i,j}$, $M_c^{i,j}$ or $\{1\}$ have a sequential decomposition over $\mathcal{F}$. This is the purpose of the next proposition.

Proposition 11.

(a) $\tau(1) = \{1\} \times \{1\}$

(b) *For* $i, j, k \in \{0, 1\}$,

$$\tau(M_h^{i,j}) = \bigcup_{\substack{k+m \equiv i \bmod 2 \\ \ell+n \equiv j \bmod 2 \\ h \equiv p+q+kn \bmod 2}} (M_p^{k,\ell} \times M_q^{m,n}) \tag{2}$$

(c) *For* $i, j \in \{0, 1\}$, *and any* $c \in A$,

$$\tau(M_c^{i,j}) = \bigcup_{\substack{k+m \equiv i \bmod 2 \\ \ell+n \equiv j \bmod 2}} (M_c^{k,\ell} \times M^{m,n}) \bigcup_{\substack{k+m \equiv i \bmod 2 \\ \ell+n \equiv j \bmod 2}} (M^{m,n} \times M_c^{k,\ell}) \tag{3}$$

Proof. (a) is trivial.
(b) Let $(u, v) \in \tau(M_h^{i,j})$, that is, $uv \in M_h^{i,j}$. Define $k, \ell, m, n, p, q \in \{0, 1\}$ by the conditions $|u|_a \equiv k \bmod 2$, $|u|_b \equiv \ell \bmod 2$, $\binom{u}{ab} \equiv p \bmod 2$, $|v|_a \equiv m \bmod 2$, $|v|_b \equiv n \bmod 2$ and $\binom{v}{ab} \equiv q \bmod 2$. By definition, $u \in M_p^{k,\ell}$ and $v \in M_q^{m,n}$. Furthermore, since $|uv|_a \equiv i \bmod 2$ and $|uv|_b \equiv j \bmod 2$, $k + m \equiv i \bmod 2$ and $\ell + n \equiv j \bmod 2$. Finally, since $\binom{u}{ab} \equiv h \bmod 2$, the formula

$$\binom{uv}{ab} \equiv \binom{u}{ab} + \binom{v}{ab} + |u|_a|v|_b \equiv p + q + kn \mod 2 \tag{4}$$

shows that $p + q + kn \equiv h \bmod 2$. This proves that (u, v) belongs to the right hand side of (2).

In the opposite direction, let $u \in M_p^{k,\ell}$ and $v \in M_q^{m,n}$, with $k + m \equiv i \bmod 2$, $\ell + n \equiv j \bmod 2$ and $h \equiv p + q + kn \bmod 2$. Then $|uv|_a \equiv k + m \equiv i \bmod 2$, $|uv|_b \equiv \ell + n \equiv j \bmod 2$, and, by Formula (4), $\binom{u}{ab} \equiv p + q + kn \equiv h \bmod 2$.

(c) Let $(u, v) \in \tau(M_c^{i,j})$, that is, $uv \in M_c^{i,j}$. Define $k, \ell, m, n \in \{0, 1\}$ by the conditions $|u|_a \equiv k \bmod 2$, $|u|_b \equiv \ell \bmod 2$, $|v|_a \equiv m \bmod 2$ and $|v|_b \equiv n \bmod 2$. Since $|uv|_a \equiv i \bmod 2$ and $|uv|_b \equiv j \bmod 2$, the relations $k + m \equiv i \bmod 2$ and $\ell + n \equiv j \bmod 2$ hold. Furthermore, since $uv \in M_c^{i,j}$, the letter c occurs at least once in uv. Thus it occurs in u or in v and hence $(u, v) \in (M_c^{k,\ell} \times M^{m,n}) \cup (M^{m,n} \times M_c^{k,\ell})$.

In the opposite direction, let $u \in M_c^{k,\ell}$ and $v \in M^{m,n}$, with $k + m \equiv i \bmod 2$ and $\ell + n \equiv j \bmod 2$ (the proof would be similar for $u \in M^{k,\ell}$ and $v \in M_c^{m,n}$). Then $|uv|_a \equiv k + m \equiv i \bmod 2$, $|uv|_b \equiv \ell + n \equiv j \bmod 2$, and since $|u|_c > 0$, $|uv|_c > 0$ and $uv \in M_c^{i,j}$, that is, $(u, v) \in \tau(M_c^{i,j})$.

We now prove that $\mathcal{F}$ is a parallel system. The proof relies on a simple, but useful observation.

Lemma 2. *For any words $x, y \in A^*$, $\binom{xaby}{ab} = \binom{xbay}{ab} + 1$.*

Proof. We have on the one hand

$$\binom{xaby}{ab} = \binom{y}{ab} + \binom{xa}{a}\binom{by}{b} + \binom{x}{ab} = \binom{y}{ab} + (|x|_a + 1)(|y|_b + 1) + \binom{x}{ab}$$

and, on the other hand

$$\binom{xbay}{ab} = \binom{ay}{ab} + \binom{xb}{a}\binom{ay}{b} + \binom{xb}{ab} = \binom{y}{ab} + |y|_b + |x|_a|y|_b + |x|_a + \binom{x}{ab}$$

which proves the lemma.

Proposition 12.

(a) $\sigma(1) = \{1\} \times \{1\}$

(b)

$$\textit{For any } i, j \in \{0,1\}, \sigma(M_0^{i,j}) = \bigcup_{\substack{k+m \equiv i \bmod 2 \\ \ell+n \equiv j \bmod 2}} (M^{k,\ell} \times M^{m,n}) \tag{5}$$

(c) *For any $i, j \in \{0,1\}$,*

$$\sigma(M_1^{i,j}) = \bigcup_{\substack{k+m \equiv i \bmod 2 \\ \ell+n \equiv j \bmod 2}} \Big((M_a^{k,\ell} \times M_b^{m,n}) \cup (M_b^{k,\ell} \times M_a^{m,n})\Big)$$
$$\cup (\{1\} \times M_1^{i,j}) \cup (M_1^{i,j} \times \{1\}) \tag{6}$$

(d) *For any $i, j \in \{0,1\}$, for any $c \in A$, $\sigma(M_c^{i,j}) = \tau(M_c^{i,j})$.*

Proof. (a) is trivial.
(b) Let $(u, v) \in \sigma(M_0^{i,j})$, and let $w \in (u \text{ Ш } v) \cap M^{i,j}$. Define $k, \ell, m, n \in \{0,1\}$ by the conditions $|u|_a \equiv k \bmod 2$, $|u|_b \equiv \ell \bmod 2$, $|v|_a \equiv m \bmod 2$ and $|v|_b \equiv n \bmod 2$. Then $(u, v) \in M^{k,\ell} \times M^{m,n}$. Furthermore, since $|w|_a \equiv i \bmod 2$ and $|w|_b \equiv j \bmod 2$, the relations $k + m \equiv i \bmod 2$ and $\ell + n \equiv j \bmod 2$ hold. Thus (u, v) belongs to the right hand side of (5).

Conversely, let $(u, v) \in M^{k,\ell} \times M^{m,n}$ with $k + m \equiv i \bmod 2$ and $\ell + n \equiv j \bmod 2$. First, if $(u, v) \in (A^*aA^* \times A^*bA^*) \cup (A^*bA^* \times A^*aA^*)$, then $\{u, v\} = \{(x_1ay_1, x_2by_2)\}$ for some words $x_1, x_2, y_1, y_2 \in A^*$. Setting $x = x_1x_2$ and $y = y_1y_2$, the set $u \text{ Ш } v$ contains the words $xaby$ and $xbay$. Note that $|xaby|_a = |xbay|_a = |uv|_a \equiv i \mod 2$ and $|xaby|_b = |xbay|_b = |uv|_b \equiv i \mod 2$. It follows by Lemma 2 that one of these words is in $M_0^{i,j}$ and thus $(u, v) \in \sigma(M_0^{i,j})$. If now, $(u, v) \notin (A^*aA^* \times A^*bA^*) \cup (A^*bA^* \times A^*aA^*)$, then $(u, v) \in (a^* \times a^*) \cup (b^* \times b^*)$ and thus $\binom{uv}{ab} = 0$. Therefore $uv \in M_0^{i,j}$ and $(u, v) \in \sigma(M_0^{i,j})$.
(c) The proof of is quite similar to that of (b). Let $(u, v) \in \sigma(M_1^{i,j})$, and let $w \in (u \text{ Ш } v) \cap M^{i,j}$. Since $\binom{w}{ab} \equiv 1 \bmod 2$, $|w|_a > 0$ and $|w|_b > 0$.
(d) holds, since $M_c^{i,j}$ is a commutative language.

Theorem 5. *The language K is decomposable.*

Proof. Indeed, K belongs to the parallel and sequential system $\mathcal{F}$, since $K = \bigcup_{0 \le i,j \le 1} M_1^{i,j}$.

6 Non decomposable languages

In this section, we give several examples of non decomposable languages. We first generalize Proposition 4 to a larger class of languages.

Proposition 13. *Let u and v be two words of A^* such that $uv \neq vu$. Then $\sigma((uv)^*)$ is not recognizable. In particular, the language $(uv)^*$ is not recognizable.*

Proof. We first remind the reader with a classical result of combinatorics on words (see [1]): two words commute if and only if they are powers of the same word. Next we claim that $\sigma((uv)^*)$ is recognizable if and only if $\sigma((vu)^*)$ is recognizable. Indeed, $(vu)^* = \{1\} \cup v(uv)^*u$ and since σ is a substitution,

$$\sigma((vu)^*) = \sigma(1) \cup \sigma(v)\sigma((uv)^*)\sigma(u)$$

Now $\sigma(1)$, $\sigma(u)$ and $\sigma(v)$ are finite and, by Mezei's theorem, belong to $\mathrm{Rec}(A^* \times A^*)$. Furthermore, $\mathrm{Rec}(A^* \times A^*)$ is closed under union and product. It follows that if $\sigma((uv)^*)$ is recognizable, so is $\sigma((vu)^*)$. The converse also holds by duality, which proves the claim.

We can now assume, without loss of generality, that $|u| \leq |v|$. Let $K = (uv)^*$. If $\sigma(K)$ is recognizable, it can be written as a finite union of blocks $L \times R$, where $L, R \in \mathrm{Rec}(A^*)$. For each $n \geq 0$, $(uv)^n \in u^n \text{ Ш } v^n$, and hence $(u^n, v^n) \in \sigma(K)$. Therefore, there exists a block $L \times R$ such that the set

$$S = \{n \in \mathbb{N} \mid (u^n, v^n) \in L \times R\}$$

is infinite. Let $s = \min S$. By assumption, we have $u^s \in L$ and $v^n \in R$ for each $n \in S$ and thus $(u^s \text{ Ш } v^n) \cap (uv)^* \neq \emptyset$.

Let us fix $n \geq 4s|u|$ and let k be such that $(uv)^k \in u^s \text{ Ш } v^n$. There exist two factorizations $u^s = x_0 \cdots x_{\ell+1}$ and $v^n = y_1 \cdots y_\ell$ such that $x_0, x_{\ell+1} \in A^*$, $x_1, y_1, \ldots, x_\ell, y_\ell \in A^+$ and $x_0 y_1 x_1 \cdots x_\ell y_\ell x_{\ell+1} = (uv)^k$. It follows, by the choice of n

$$|y_1 \cdots y_\ell| = n|v| \geq 4s|u||v|$$

Now, since $\ell \leq |x_1 \cdots x_\ell| \leq s|u|$, one of the words y_i has length $\geq 4|v|$. Now, since y_i is a factor of $(uv)^k$ and since $|u| \leq |v|$, y_i contains vuv as a factor. At the same time, y_i is a factor of v^n of length $\geq 4|v|$, that is, a factor of v^5. It follows that vuv is a factor of v^5. Let us write v as the power of some primitive word, say $v = t^p$. Then tut is a factor of vuv, which is in turn a factor of t^{5p}. Now, a primitive word cannot have any conjugate, which means that if $t^{5p} = xtuty$, then x and y are necessarily powers of t. But this forces u itself to be a power of t, a contradiction, since u and v are not powers of the same word. Thus $\sigma(K)$ is not recognizable.

7 Inverse morphisms

In this section, we show that decomposable languages are not closed under inverse morphisms.

Let L be the language defined over the alphabet $A = \{a, b\}$ by the following regular expression:

$$L = (aab)^* \cup A^* b(aa)^* abA^*$$

We claim that this language is decomposable. We now show that L is decomposable by constructing, step by step, a parallel and sequential system containing L. First, consider the automaton $\mathcal{A}$ represented in Figure 2.

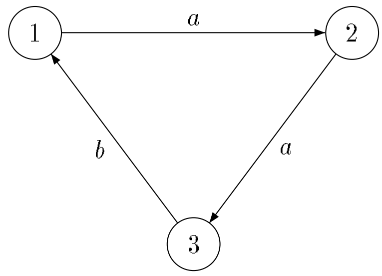

Fig. 2. The automaton $\mathcal{A}$.

Let $K_{i,j}$ be the language accepted by $\mathcal{A}$ with i as initial state and j as the only final state. These languages are $K_{1,1} = (aab)^*$, $K_{1,2} = (aab)^*a$, $K_{1,3} = (aab)^*aa$, $K_{2,1} = (aba)^*ab$, $K_{2,2} = (aba)^*$, $K_{2,3} = (aba)^*a$, $K_{3,1} = (baa)^*b$, $K_{3,2} = (baa)^*ba$ and $K_{3,3} = (baa)^*$. In the next proposition, we compute the image by τ of these languages.

Proposition 14. *For* $1 \leq i, j \leq 3$,

$$\tau(K_{i,j}) = \bigcup_{1 \leq k \leq 3} (K_{i,k} \times K_{k,j})$$

Proof. If $(u, v) \in \tau(K_{i,j})$, then $uv \in K_{i,j}$, and thus $i \cdot uv = j$ in $\mathcal{A}$. Setting $k = i \cdot u$, we have $k \cdot v = j$ and thus $u \in K_{i,k}$ and $v \in K_{k,j}$. In the opposite direction, if $u \in K_{i,k}$ and $v \in K_{k,j}$, then $i \cdot u = k$ and $k \cdot v = j$, whence $i \cdot uv = j$ and $uv \in K_{i,j}$.

Let $L_0 = A^* b(aa)^* abA^*$. We first compute the image of L_0 under σ and τ.

Proposition 15.

$$\tau(L_0) = (A^* \times L_0) \cup (L_0 \times A^*) \cup (A^* b(aa)^* \times a(aa)^* bA^*) \cup (A^* b(aa)^* a \times (aa)^* bA^*)$$

Proof. Let $(u, v) \in \tau(L_0)$, that is, $uv \in L_0$. First suppose that $u \in L_0$ or $v \in L_0$, this implies that $(u, v) \in (A^* \times L_0)$ or $(u, v) \in (L_0 \times A^*)$. Otherwise we can write uv as $u_1 b(aa)^n abv_1$ where $n \geq 0$, u_1 is a prefix of u and v_1 is a suffix of v, and then $(u, v) \in (A^*b(aa)^* \times a(aa)^*bA^*)$ or $(u, v) \in (A^*b(aa)^*a \times (aa)^*bA^*)$. In the opposite direction, if $u \in L_0$ its clear that for any $v \in A^*$, $uv, vu \in L_0$. Now suppose that $u = u_1 b(aa)^n$ and $v = a(aa)^m bv_1$ (resp. $u = u_1(aa)^n a$ and $v = (aa)^m bv_1$) for some $n, m \geq 0$, then $uv \in L_0$.

Proposition 16.

$$\begin{aligned}\sigma(L_0) = \ & (L_0 \times A^*) \cup (A^* \times L_0) \cup (A^*aA^*bA^* \times A^*bA^*) \\ & \cup (A^*bA^*aA^* \times A^*bA^*) \cup (A^*bA^*bA^* \times A^*aA^*) \\ & \cup (A^*bA^* \times A^*aA^*bA^*) \cup (A^*bA^* \times A^*bA^*aA^*) \\ & \cup (A^*aA^* \times A^*bA^*bA^*)\end{aligned}$$

Proof. If $(u, v) \in \sigma(L_0)$, then by definition $(u \text{ Ш } v) \cap L_0 \neq \emptyset$. Let us take some $w \in (u \text{ Ш } v) \cap L_0$. By definition of L_0, w can be decomposed as $w = w_1 w_2 w_3$, where $w_1, w_3 \in A^*$ and $w_2 = ba^n b$ with n odd. Since $w \in u \text{ Ш } v$, w_2 can be written as $w_2 = u_1 v_1 \cdots u_n v_n$ with $u_i, v_i \in A^*$ for $1 \leq i \leq n$, $u = u_0 u_1 \cdots u_n u_{n+1}$ and $v = v_0 v_1 \cdots v_n v_{n+1}$ for some $u_0, v_0, \ldots, u_{n+1}, v_{n+1} \in A^*$. Let us consider the words $\overline{u} = u_1 \cdots u_n$ and $\overline{v} = v_1 \cdots v_n$. Since $w_2 = ba^n b$, the words $\overline{u}$ and $\overline{v}$ can take the following values, where $n = n_1 + n_2$:

(1) $\overline{u} = ba^{n_1}b$ and $\overline{v} = a^{n_2}$
(2) $\overline{u} = a^{n_1}$ and $\overline{v} = ba^{n_2}b$
(3) $\overline{u} = ba^{n_1}$ and $\overline{v} = a^{n_2}b$
(4) $\overline{u} = a^{n_1}b$ and $\overline{v} = ba^{n_2}$

Let R be the right hand side of the equality to be proved. We show that $\sigma(L_0)$ is a subset of R by considering the four cases separately.
Case (1). If n_1 is odd, then $u \in L_0$, and so $(u, v) \in (L_0 \times A^*)$, otherwise, if n_1 is even, and $n_1 + n_2$ is odd, we have $n_2 > 0$, and then $(u, v) \in A^*bA^*bA^* \times A^*aA^*$.
Case (2). A proof similiar to that of case (1) shows that (u, v) belongs either to $A^* \times L_0$ or to $A^*aA^* \times A^*bA^*bA^*$.
Case (3). If $n_1 > 0$, $(u, v) \in A^*bA^*aA^* \times A^*bA^*$. Otherwise, since $n_1 + n_2$ is odd, $n_2 > 0$ and then $(u, v) \in A^*bA^* \times A^*aA^*bA^*$.
Case (4). A proof similiar to that of case (3) shows that (u, v) belongs either to $A^*bA^* \times A^*bA^*aA^*$ or to $A^*aA^*bA^* \times A^*bA^*$.

We now prove that R is a subset of $\sigma(L_0)$. If $(u, v) \in (L_0 \times A^*) \cup (A^* \times L_0)$, it is clear that $(u \text{ Ш } v) \cap L_0 \neq \emptyset$. Suppose that $(u, v) \in (A^*aA^*bA^* \times A^*bA^*)$. Observing that $A^*aA^*bA^* = A^*abA^*$, the word u can be decomposed as $u = u_1 abu_2$ with $u_1, u_2 \in A^*$, and v can be decomposed as $v = v_1 bv_2$ with $v_1, v_2 \in A^*$. Now $v_1 u_1 babv_2 \in (u \text{ Ш } v) \cap L_0$ and thus $(u, v) \in \sigma(L_0)$. The proof is similar if (u, v) belongs to $A^*bA^*aA^* \times A^*bA^*$, $A^*bA^* \times A^*bA^*aA^*$ or $A^*bA^* \times A^*aA^*bA^*$.

Finally, if $(u, v) \in A^*bA^*bA^* \times A^*aA^*$, then u can be decomposed as $u = u_1 ba^n bu_2$ with $n \geq 0$, $u_1, u_2 \in A^*$, and v can be decomposed as $v = v_1 av_2$ with $v_1, v_2 \in A^*$. If n is even, u and v can be shuffled as $u_1 v_1 ba^{(n+1)} bv_2 u_2$, a word of L_0, otherwise, u and v can be shuffled as $u_1v_1ba^nbav_2u_2$, another word of L_0. Thus $(u,v) \in \sigma(L_0)$ in both cases. The proof for $A^*aA^* \times A^*bA^*bA^*$ is similar.

For $1 \leq i,j \leq 3$, define the languages $L_{i,j}$ as the union of L_0 with the language $K_{i,j}$. We first compute the image of these languages under τ.

Proposition 17. *For $1 \leq i,j \leq 3$,*

$$\tau(L_{i,j}) = \tau(L_0) \cup \bigcup_{1 \leq k \leq 3} (L_{i,k} \times L_{k,j})$$

Proof. Let R be the right handside of the relation to be proved. Since $L_{i,j} = L_0 \cup K_{i,j}$, $\tau(L_{i,j}) = \tau(L_0) \cup \tau(K_{i,j})$. Now, by Proposition 14, $\tau(K_{i,j}) = \bigcup_{1 \leq k \leq 3}(K_{i,k} \times K_{k,j})$, and since each $K_{i,j}$ is a subset of $L_{i,j}$, $\tau(K_{i,j})$ is a subset of R. It follows that $\tau(L_{i,j})$ is a subset of R.

We now prove the opposite inclusion. Since L_0 is a subset of $L_{i,j}$, $\tau(L_0)$ is a subset of $\tau(L_{i,j})$. Furthermore, if $(u,v) \in L_{i,k} \times L_{k,j}$ for some k, two possibilities arise. First, if $u \in L_0$ or $v \in L_0$ then $(u,v) \in \tau(L_0)$ and hence $(u,v) \in \tau(L_{i,j})$. Otherwise, $u \in K_{i,k}$, $v \in K_{k,j}$, $uv \in K_{i,j}$ and thus $(u,v) \in \tau(L_{i,j})$.

In order to compute the image of the languages $L_{i,j}$ under σ, we introduce the languages

$$M_{i,j} = K_{i,j} \cap a^*(b \cup 1)a^*$$

These languages are clearly finite and thus decomposable, and they belong to some parallel and sequential systems $\mathcal{S}(M_{i,j})$.

Proposition 18. *For $1 \leq i,j \leq 3$,*

$$\sigma(L_{i,j}) = \sigma(L_0) \cup \sigma(M_{i,j}) \cup (L_{i,j} \times \{1\}) \cup (\{1\} \times L_{i,j})$$

Proof. If $(u,v) \in \sigma(L_{i,j})$, then $(u \text{ Ш } v) \cap L_{i,j} \neq \emptyset$. First, if $u = 1$ or $v = 1$ then $(u,v) \in (L_{i,j} \times \{1\}) \cup (\{1\} \times L_{i,j})$. Now, suppose that u and v are non-empty words.

If $|u|_b + |v|_b < 2$ then necessarily $(u,v) \in \sigma(M_{i,j})$. Suppose that $|u|_b + |b|_v \geq 2$. We claim that $(u,v) \in (A^*aA^*bA^* \times A^*bA^*) \cup (A^*bA^*aA^* \times A^*bA^*) \cup (A^*bA^*bA^* \times A^*aA^*) \cup (A^*bA^* \times A^*aA^*bA^*) \cup (A^*bA^* \times A^*bA^*aA^*) \cup (A^*aA^* \times A^*bA^*bA^*) \subseteq \sigma(L_0)$. Since u and v are both non-empty, they contain at least one letter. If u and v are equal to b, $(u \text{ Ш } v) \cap L_0 = \emptyset$, and so u or v contains the letter a and since $|u|_b + |v|_b < 2$ the result holds. Conversely, if $(u,v) \in \sigma(L_0) \cup \sigma(M_{i,j}) \cup (L_{i,j} \times \{1\}) \cup (\{1\} \times L_{i,j})$ by the definition of $M_{i,j}$, Proposition 16 and the definition of $L_{i,j}$, $(u,v) \in \sigma(L_{i,j})$.

We are now ready to show that L is decomposable.

Theorem 6. *The language L is decomposable.*

Proof. By Theorem 4, all languages which are products of commutative languages are decomposable. In particular, the following languages are decomposable:

$$A^*b(aa)^*, (aa)^*abA^*, A^*b(aa)^*a, (aa)^*bA^*, A^*aA^*bA^*,\\ A^*bA^*aA^*, A^*bA^*bA^*, A^*aA^*, A^*bA^*$$

Let us define a system $\mathcal{S}$ consisting of the unions of the following languages: L_0, $\{1\}$, A^*, $L_{i,j}$, for $1 \leq i, j \leq 3$, the languages of the systems associated with the languages in display and the languages of $\mathcal{S}(M_{i,j})$ for $1 \leq i, j \leq 3$. Note that L belongs to $\mathcal{S}$ since $L = L_{1,1}$.

It remains to show that for any language L' of $\mathcal{S}$, $\tau(L')$ and $\sigma(L')$ can be written as the union of languages $(M \times N)$ where M and N belong to $\mathcal{S}$. For L_0, the result follows from Propositions 15 and 16. For $L_{i,j}$, the result follows from Propositions 17 and 18. Clearly, $\tau(\{1\}) = \sigma(\{1\}) = (\{1\} \times \{1\})$ and $\sigma(A^*) = \tau(A^*) = (A^* \times A^*)$. For any language of the systems described above, its image under τ and σ can be obtained from the languages of these systems, because they are parallel and sequential systems. Finally, since for any language N_1, N_2, the formulas $\tau(N_1 \cup N_2) = \tau(N_1) \cup \tau(N_2)$ and $\sigma(N_1 \cup N_2) = \sigma(N_1) \cup \sigma(N_2)$ hold, the system $\mathcal{S}$ is a parallel and sequential system.

Proposition 19. *Decomposable languages are not closed under inverse morphism.*

Proof. Let $A = \{a, b\}$ and let $\varphi : A^* \to A^*$ be the morphism of monoids defined by $\psi(a) = aa$ and $\psi(b) = b$. If $L = (aab)^* \cup A^*b(aa)^*abA^*$, then $\varphi^{-1}(L) = (ab)^*$. Now L is decomposable by Theorem 6 but, by Proposition 4, $\varphi^{-1}(L)$ is not.

Corollary 2. *Decomposable languages do not form a positive variety of languages.*

8 Intersection

Let L_1 and L_2 be the languages defined over the alphabet $A = \{a, b\}$ by the following regular expressions:

$$L_1 = (ab)^+ \cup (ab)^*bA^*$$
$$L_2 = (ab)^+ \cup (ab)^*aaA^*$$

One can show (proof omitted) that these two languages are decomposable but that their intersection is not decomposable. Therefore, decomposable languages are not closed under intersection.

9 More on group languages

Recall that a *group language* is a language recognized by a finite group. Equivalently, this is a language whose minimal automaton is a permutation automaton. The aim of this section is to prove the following result

Theorem 7. *If L is a group language, then $\sigma(L)$ is recognizable.*

The proof relies on several lemmas of independent interest. Let G be a finite group, let $\pi : A^* \to G$ be a surjective morphism.

Lemma 3. *Let $g_1, g_2, \ldots, g_n$ be a sequence of elements of G, with $n \geq |G|$. Then there exist two indices i, j, with $i \leq j$ such that $g_i \cdots g_j = 1$.*

Proof. Consider the sequence $g_1, g_1g_2, \ldots, g_1g_2 \cdots g_n$. Either one of these elements is equal to 1, or, since $n > |G|-1$, two of them are equal, say $g_1 \cdots g_{i-1} = g_1 \cdots g_j$. In this case, $g_i \cdots g_j = 1$.

Let $L = \pi^{-1}(1)$. The next lemma states that sufficiently long words contain a factor in L. More precisely

Lemma 4. *Every word of A^* of length $\geq |G|$ contains a non-empty factor in L.*

Proof. Let $a_1 \cdots a_n$ be a word of length $n \geq |G|$. By Lemma 3, there exist two indices i, j, with $i \leq j$ such that $\pi(a_i) \cdots \pi(a_j) = 1$. It follows that $\pi(a_i \cdots a_j) = 1$ and hence $a_i \cdots a_j \in L$.

Lemma 5. *Let x be a word of L and let u and v be two words. Then $uv \in L$ if and only if $uxv \in L$.*

Proof. Indeed, if $x \in L$, then $\pi(x) = 1$. It follows that

$$\pi(uxv) = \pi(u)\pi(x)\pi(v) = \pi(u)\pi(v) = \pi(uv)$$

Therefore $\pi(uxv) = 1$ if and only if $\pi(uv) = 1$, that is if and only if $uxv \in L$.

Lemma 6. *Let $a_1, \ldots, a_r$ be letters, let x be a word of L and let u and v be two words. If $uv \in La_1La_2L \cdots La_rL$, then $uxv \in La_1La_2L \cdots La_rL$.*

Proof. Indeed, if $uv \in La_1La_2L \cdots La_rL$, then there exist an index I and two words $x', x'' \in A^*$ such that $u \in La_1L \cdots La_ix'$ and $v \in x''a_{i+1}L \cdots La_rL$. It follows immediately that $uxv \in La_1La_2L \cdots La_rL$.

Proposition 20. *If the language L is decomposable, then every language recognized by π is decomposable.*

Proof. Let P be a subset of G. Since decomposable languages are closed under union and since

$$\pi^{-1}(P) = \bigcup_{g \in P} \pi^{-1}(g)$$

it suffices to prove that each language $\pi^{-1}(g)$ is decomposable. Now, let g be a fixed element of G and let u be a word such that $\pi(u) = g^{-1}$. We claim that $\pi^{-1}(g) = Lu^{-1}$. Indeed if $x \in \pi^{-1}(g)$, then $\pi(xu) = \pi(x)\pi(u) = gg^{-1} = 1$. It follows that $xu \in L$ and $x \in Lu^{-1}$. Conversely, if $x \in Lu^{-1}$, then $xu \in L$, that is, $\pi(xu) = 1$. Therefore $\pi(x) = \pi(u)^{-1} = g$ and $x \in \pi^{-1}(g)$, which proves the claim. Now, Proposition 7 shows that decomposable languages are closed under quotients, and thus $\pi^{-1}(g)$ is decomposable.

Theorem 7 now follows from the next proposition, which gives an explicit formula for $\sigma(L)$.

Proposition 21. *The following formula holds, with $N = 2|G|^4$:*

$$\sigma(L) = \bigcup_{\substack{r,s\leq N \\ (a_1\cdots a_r \text{Ш} b_1\cdots b_s)\cap L\neq\emptyset}} (La_1La_2L\cdots La_rL) \text{ Ш } (Lb_1Lb_2L\cdots Lb_sL)$$

Proof. Let K be the right member of the equality to be proved. Let $a_1\cdots a_r$ and $b_1\cdots b_s$ be two words such that $r, s\leq N$ and

$$(a_1\cdots a_r \text{ Ш } b_1\cdots b_s)\cap L\neq\emptyset$$

Then there exists in L a word $w = c_1\cdots c_{r+s}$, a partition (I, J) of $\{1,\cdots, r+s\}$ and two bijections $\alpha : \{1,\cdots, r\}\to I$ and $\beta : \{1,\cdots, s\}\to J$ such that, for $1\leq i\leq r+s$,

$$c_i = \begin{cases} a_{\alpha^{-1}(i)} & \text{if } i\in I \\ b_{\beta^{-1}(i)} & \text{if } j\in I \end{cases}$$

Suppose that $(u, v)\in(La_1La_2L\cdots La_rL)$ Ш $(Lb_1Lb_2L\cdots Lb_sL)$. Then $u = u_0a_1u_1\cdots a_ru_r$ and $v = v_0b_1v_1\cdots b_sv_s$ for some words $u_0, u_1, \ldots, u_r, v_0, \ldots, v_r\in L$. Let, for $1\leq i\leq r+s$,

$$x_i = \begin{cases} c_iu_{\alpha^{-1}(i)} & \text{if } i\in I \\ c_iv_{\beta^{-1}(i)} & \text{if } i\in J \end{cases}$$

and let $x = u_0x_1\cdots x_{r+s}$. For instance, if $w = a_1a_2b_1a_3b_2b_3a_4$, then $x = u_0a_1u_1a_2u_2b_1v_1a_3u_3b_2v_2b_3v_3a_4u_4$. Then x belongs to u Ш v par construction. Furthermore, since for $1\leq i\leq r$, $\pi(u_i) = 1$ and for $1\leq i\leq s$, $\pi(v_i) = 1$, we have, for $1\leq i\leq r+s$, $\pi(x_i) = \pi(c_i)$. It follows that $\pi(x) = \pi(c_1\cdots c_{r+s}) = \pi(w)$ and hence $\pi(x) = 1$, since $w\in L$. Therefore $x\in(u$ Ш $v)\cap L$ and $(u, v)\in\sigma(L)$.

In the opposite direction, consider a pair $(u, v)\in\sigma(L)$. We prove by induction on $|u|+|v|$ that $(u, v)\in K$. First assume that $|u|+|v|\leq N$. Let $u = a_1\cdots a_r$ and $v = b_1\cdots b_s$ with $r, s\leq N$. Since $1\in L$, one has $u\in La_1La_2L\cdots La_rL$ and $v\in Lb_1Lb_2L\cdots Lb_sL$. Furthermore, $(a_1\cdots a_r$ Ш $b_1\cdots b_s)\cap L\neq\emptyset$, since $(u, v)\in\sigma(L)$. Therefore $(u, v)\in K$.

We may now assume that $|u|+|v| > N$. By assumption, there are two factorizations $u = u_1u_2\cdots u_n$ and $v = v_1v_2\cdots v_n$, with $u_1\in A^*$, v_1, u_2, v_2, $\ldots$, v_{n-1}, u_n in A^+, $v_n\in A^*$ such that $u_1v_1\cdots u_nv_n\in L$.

Suppose first that one of the words $u_1,\ldots, u_n, v_1,\ldots, v_n$ has length $\geq|G|$. We may assume for instance, that this word is some u_i (a symmetrical argument works for v_i). Then by Lemma 4, this word contains a non-empty factor in L, that is $u_i = u_i'xu_i''$ with $x\in L\cap A^+$. It follows by Lemma 5 that $u_i'u_i''\in L$. Let $u' = u_1\cdots u_{i-1}u_i'$ and $u'' = u_i''u_{i+1}\cdots u_n$. Then $u'u''$ is the word obtained from u by deleting the factor x in u_i. By Lemma 5, the word $u_1v_1\cdots u_{i-1}v_{i-1}u_i'u_i''v_i\cdots u_nv_n$ belongs to L and thus $(u'u'', v)\in\sigma(L)$.

It follows by the induction hypothesis that $(u'u'', v) \in K$. Therefore, there exist two words $a_1 \cdots a_r$ and $b_1 \cdots b_s$ such that $(a_1 \cdots a_r \text{ Ш } b_1 \cdots b_s) \cap L \neq \emptyset$, $u'u'' \in La_1La_2L \cdots La_rL$ and $v \in Lb_1Lb_2L \cdots Lb_sL$. Now by Lemma 6, $u'xu'' \in La_1La_2L \cdots La_rL$, and hence $(u, v) \in K$.

Suppose now that the length of each u_i and v_i is strictly smaller that $|G|$. Since $|u_1v_1 \cdots u_nv_n| = |u| + |v| \geq 2|G|^4$, we have $n \geq |G|^3$. By Lemma 3, applied to the group G^3, there exist two indices i and j, with $i \leq j$ such that $(\pi(u_i \cdots u_j), \pi(u_iv_i \cdots u_jv_j), \pi(v_i \cdots v_j)) = (1, 1, 1)$, which means that $u_i \cdots u_j$, $u_iv_i \cdots u_jv_j$ and $v_i \cdots v_j$ are in L. Now, since $u_1v_1 \cdots u_nv_n \in L$, it follows by Lemma 5 that $u_1v_1 \cdots u_{i-1}v_{i-1}u_{j+1}v_{j+1} \cdots u_nv_n \in L$. Therefore

$$(u_1 \cdots u_{i-1}u_{j+1} \cdots u_n, v_1 \cdots v_{i-1}v_{j+1} \cdots v_n) \in \sigma(L)$$

and by the induction hypothesis, $(u_1 \cdots u_{i-1}u_{j+1} \cdots u_n, v_1 \cdots v_{i-1}v_{j+1} \cdots v_n) \in K$. It follows by Lemma 6 that $(u, v) \in K$. Therefore $\sigma(L) = K$.

10 Conclusion

We have given a counterexample to the original Schnoebelen's Conjecture, and we have studied the closure properties of the class of decomposable languages. This class is closed under union, product, shuffle, inverse of length preserving morphisms and residuals, but it is not closed under intersection and inverse morphisms.

However the main problem remains open: is it decidable whether a given recognizable language is decomposable?

An example of a non-decomposable language L such that $\sigma(L)$ is recognizable was given in Section 4. Nevertheless, we think that decomposable languages might be characterized by some recognizability property of σ, and we propose the following precise formulation:

Conjecture 2. Let K be a recognizable language and let $\bigcup_{1 \leq i \leq n} L_i \times R_i$ be the Conway canonical form of $\tau(K)$. Then L is decomposable if and only if all the sets $\sigma(L_i)$ and $\sigma(R_i)$ are recognizable.

If this conjecture is true, Theorem 7 implies that every group language is decomposable. Actually, the conjecture would imply more generally that every language of the polynomial closure of the group languages (languages that are finite union of products of group languages) is decomposable. These languages form a positive variety of languages and admit a nice algebraic characterization: in their ordered syntactic monoid, every idempotent e satisfies $e \leq 1$. They are also the rational open sets of the pro-group topology [7]. Proving directly that these sets are decomposable seems to be an accessible challenge and a good test for our conjecture.

However it would not be a characterization of the class of decomposable languages, since decomposable languages do not form a positive variety.

References

1. Lothaire, *Combinatorics on Words*, *Encyclopedia of Mathematics and its Applications* vol. 17, aw, reading, 1983.
2. J. Baeten and W. Weijland, *Process algebra*, *Cambridge Tract in Theoretical Computer Science* vol. 18, Cambridge University Press, Cambridge UK, 1990.
3. S. Eilenberg, *Automata, Languages and Machines*, vol. B, Academic Press, New York, 1976.
4. J. H. Conway, *Regular Algebra and Finite Machines*, Chapman and Hall, London, 1971.
5. J.-E. Pin and P. Weil, Polynomial closure and unambiguous product, *Theory Comput. Systems* **30** (1997), 1–39.
6. P. Schnoebelen, Decomposable regular languages and the shuffle operator, *EATCS Bull.* ,67 (1999), 283–289.
7. J.-E. Pin, Polynomial closure of group languages and open sets of the Hall topology, *Theoretical Comput. Sci.* **169** (1996), 185–200.