# A presentation of a finitely generated submonoid of invertible endomorphisms of the free monoid

Christian Choffrut* Štěpán Holub†‡

November 9, 2015

**Abstract**

An endomorphism of the free monoid $A^*$ is invertible if it is injective and extends to an automorphism of the free group generated by $A$. A simple example: the endomorphism that leaves all generators $A$ invariant except one, say $a$, which is mapped to $ba$ for some other generator $b$. We give a monoid presentation for the submonoid generated by all such endomorphisms when $a$ and $b$ are taken arbitrarily. These left translations are a special case of Nielsen positive transformations: "left" because the mutiplicative constant acts on the left and "positive" because this constant belongs to the free monoid, not the free group.

## 1 Introduction

Given a finite set $A$, denote by $A^*$ and $F(A)$ respectively the free monoid and the free group generated by $A$. Let $k$ be the cardinality of $A$. Taking advantage of the natural embedding of $A^*$ in $F(A)$ and following [10], we say that an endomorphism of $A^*$ is *invertible* if it extends to an automorphism of $F(A)$ or seen differently as an automorphism of $F(A)$ which preserves the free monoid. Little seems to be known so far: it is not finitely generated provided $k > 2$, cf. [18] and independently [11], and a nice and fairly precise characterization is given in the case $k = 3$, see [16], but the authors do not give any monoid presentation for this case.

The automorphism group of the free group has been much more studied. Nielsen's publication ninety years ago can be reformulated as saying that the group is generated by two types of automorphisms which are traditionally

*LIAFA, CNRS and Université Paris 7 Denis Diderot, France. Email cc@liafa.univ-paris-diderot.fr.

†Department of Algebra, Faculty of Mathematics and Physics, Charles University in Prague. Email: holub@karlin.mff.cuni.cz

‡Supported by the Czech Science Foundation grant number 13-01832S

called *Nielsen transformations*, cf. [9]. The first type performs a permutation on the generators and their inverses subject to natural restrictions. The second is associated with any distinct generators $a, b \in A$: it assigns to $b$ one of the four different elements $ab$, $ba$, $ab^{-1}$ or $b^{-1}a$ and leaves all other generators invariant. The group presentation of Nielsen was simplified by J. McCool in [8] and later a presentation for the so-called special subgroup of automorphisms was given by S. M. Gersten in [2]. Nielsen's result is focused on an algorithm for testing whether an endomorphism of the free group is actually an automorphism, i.e., it solves both the characterization of the group of automorphisms of $F(A)$ and the membership problem. For this latter problem a more recent presentation can be read in [15].

The Nielsen transformations of the type $b \mapsto ab$ or $b \mapsto ba$ denoted $\mathcal{N}$ and $\mathcal{N}^\rho$ respectively map $A^*$ into itself and are invertible. They generate the two submonoids $\langle \mathcal{N} \rangle$ and $\langle \mathcal{N}^\rho \rangle$ of all invertible endomorphisms. Here we study the subsemigroup of invertible endomorphisms generated by $\mathcal{N}$ along with those automorphisms of the free monoid $\Omega_A$ which induce permutations on $A$. Our main contribution is to give a monoid presentation. We also show that the monoid has a decidable membership problem.

A few words on previous works in the domain might be of interest. In [17], the monoid of invertible endomorphisms over the binary alphabet has been shown to be generated by three endomorphisms: the two Nielsen transformations and the transposition exchanging the two letters. The presentation (with infinitely many relations) for this simplest case was given in [13]. For at least three letters, the monoid of invertible endomorphisms is not finitely generated [11, 18]. In the case of precisely three letters, however, it is generated by the union of $\mathcal{N}$, $\mathcal{N}^\rho$ and $\Omega_A$ along with the cyclic shifts thereof, [16]. Invertible substitutions over two letters were originally studied as Sturmian morphisms, that is, morphisms preserving Sturmian words, [13]. Generalizing this approach, Justin and Pirillo introduced the so-called *standard Episturmian* morphisms which happen to be invertible endomorphisms generating a proper submonoid of $\langle \mathcal{N} \cup \Omega_A \rangle$. The *Episturmian* morphisms were obtained by adding the "right" version of the standard Episturmian morphisms and thus the submonoid $\mathcal{E}$ they generate is properly included in the monoid generated by the union of $\mathcal{N}$, $\mathcal{N}^\rho$ and $\Omega_A$. Richomme [11] gives a monoid presentation for $\mathcal{E}$ which generalizes the above mentioned presentation of Sturmian morphisms by Séébold.

Our motivation was not originally the study of a class of endomorphisms of the free monoid. Rather, our curiosity started with a revisit of the very old issue on the combinatorial problem of solving word equations in the free monoid, see the Chapter "Equations in words" in [6] where the problem of describing the set of solutions of a fixed equation is surveyed. Actually, apart from the case of an equation with at most three unknowns, no such description is available. For three unknowns, Hmelevskii [3] showed that

all solutions can be described by finitely many expressions, the so-called *parametric words*, containing two types of parameters: word and integer parameters. He also gave a conterexample ruining the possibility of extending the result to equations with more than three unknowns. Solving an equation in words such as $xy = zt$, where $x, y, z, t$ are unknowns, consists of determining all morphsims $\phi$ from $\{x, y, z, t\}^*$ into an arbitrary free monoid $A^*$ such that $\phi(xy) = \phi(zt)$ holds. With this particular example the morphism $x\phi = ab, y\phi = cde, z\phi = abc, t\phi = de$ is a solution with $A = \{a, b, c, d, e\}$. Lentin showed that all solutions of a given equation can be factored through an up to a renaming unique minimal solution which he called principal, see [5]. In the present equation this solution associates $x$ with $x$, $zy$ with $y$, $xz$ with $z$ and $t$ with $t$. It turns out that it is a composition of one-sided Nielsen automorphisms and of endomorphisms which identify letters. In the Chapter entitled "Equations in words" of the handbook [6], this composition is made unique by imposing restrictions on the way endomorphisms occur. The natural question was then to study the consequences of relaxing these restrictions, said differently, to investigate in which different ways the same solution can be expressed as a product of transformations in $\mathcal{N}$. A final word of caution: the equations we are dealing with contain word variables only and no constants. Makanin initiated the very rich field of determining whether or not such an equation has a solution, which is trivial with constant free equations. This latter problem is completely different a thorough account of which can be found in [7].

Now we briefly comment on the organization of the paper. In section 2 we survey the various submonoids of the automorphism group of the free group which leave the free monoid invariant and which were introduced for different purposes in the literature placing thus our investigation in the string of publications on the subject. Section 3 is the core of the work and is dedicated to the presentations of the monoid generated by $\mathcal{N}$, and then to the monoid generated by $\mathcal{N}$ along with the permutations of the letters. Section 4 shows that the membership problem is solvable in quadratic time and that the word problem is polynomially decidable. We conjecture that these estimations can be refined by use of suitable data structures. An intriguing question is that of finding a presentation of the full monoid of invertible endomorphisms.

## 2 Preliminaries

The free monoid generated by the (in our context finite) set $A$ is denoted by $A^*$ and its identity by 1. Let $A^{-1}$ be a disjoint copy of $A$. The one-to-one correspondence $x \leftrightarrow x^{-1}$ between $A$ and $A^{-1}$ defines an involution on $A \cup A^{-1}$. The free group $F(A)$ generated by $A$ is the quotient of $(A \cup A^{-1})^*$ by the relations of the form $aa^{-1} = a^{-1}a = 1$. The elements in $A$ are

called *positive* and those in $A^{-1}$ are called *negative*. Since this paper is concerned with composition of morphisms, we make the convention that the composition is performed from left to right. Thus we write the arguments of the functions to the left of the function symbol. The composition of two mappings $f$ and $g$ is written as $f \cdot g$ or simply as $fg$.

Let $\mathcal{I}(A^*)$, or simply $\mathcal{I}$ when $A$ is understood, be the family of invertible endomorphisms of $A^*$, which is better introduced by making a detour through the group $\mathrm{Aut}(F(A))$ of automorphisms of $F(A)$. Because $A^*$ is naturally embedded in $F(A)$, $\mathcal{I}$ can be identified with the submonoid of those elements $h \in \mathrm{Aut}(F(A))$ preserving $A^*$, i.e., $h \in \mathrm{Aut}(F(A))$ belongs to $\mathcal{I}$ if and only if it satisfies $h(A^*) \subseteq A^*$.

The group $\mathrm{Aut}(F(A))$ has been shown to be (finitely) generated by two types of mappings. First, by morphisms $\sigma$ inducing a permutations over $A \cup A^{-1}$ and consistent with the sign, namely satisfying $\sigma(a^{-1}) = \sigma(a)^{-1}$. The set of such automorphisms form a subgroup of $\mathrm{Aut}(F(A))$. The second type are automorphisms $N_{ab}$ defined by the rules

$$N_{ab} \quad \begin{cases} b \mapsto ab \\ c \mapsto c, \quad \text{if } c \neq b, b^{-1}\,. \end{cases} \tag{1}$$

where $a, b \in A \cup A^{-1}$, $a \neq b, b^{-1}$. When $a$ and $b$ are positive, the restriction of $N_{ab}$ to $A^*$ defines an endomorphism of $A^*$ which belongs to $\mathcal{I}$. Similarly, again with the same hypothesis that $a$ and $b$ are positive, the automorphism $N_{a^{-1}b^{-1}}$ belongs to $\mathcal{I}$ since it associates $ba$ to $b$. We denote it by $N^{\rho}_{ba}$. The superscript $\rho$ is meant to suggest that $N^{\rho}_{ba}$ is the reverse of $N_{ab}$ in the following sense. For all $a_1, b_1, \ldots, a_n, b_n \in A$ and each $w \in A^*$ we have

$$(wN_{a_1,b_1} \cdots N_{a_n,b_n})^{\rho} = w^{\rho} N^{\rho}_{b_1,a_1} \cdots N^{\rho}_{b_n,a_n}$$

where $v^{\rho}$ denotes the *reverse* or *mirror image* of $v$ ($w^{\rho} = w$ if $w$ is the empty word and $(wa)^{\rho} = aw^{\rho}$ if $a \in A$ and $w \in A^*$).

The *standard Episturmian* morphisms were introduced in [4] and are defined by the rules

$$E_a \quad \begin{cases} b \mapsto ba, \quad \text{if } b \neq a, \\ a \mapsto a\,. \end{cases} \tag{2}$$

We draw the attention of the reader to the subtle difference between $E_a$ and $N^{\rho}_{ba}$: in the former case the image of all letters except that of $a$ is followed by $a$ while in the latter case the only letter whose image is followed by $a$ is $b$. Furthermore, $E_a$ is generated by the endomorphisms $N^{\rho}{}_{ba}$. Indeed, if $\{a_1, \ldots, a_n\}$ is an enumeration of the letters in $A$, then we have

$$E_{a_1} = N^{\rho}_{a_2 a_1} \cdot N^{\rho}_{a_3 a_1} \cdots N^{\rho}_{a_n a_1}$$

The converse does not hold, [11, page 31]. In the same way that the endomorphisms $N_{ba}^{\rho}$ is the reverse of $N_{ba}$, the morphism $E_a^{\rho}$ is obtained by reversing $E_a$, to wit

$$E_a^{\rho} \quad \begin{cases} b \mapsto ab, & \text{if } b \neq a, \\ a \mapsto a\,. \end{cases} \tag{3}$$

We review the main families of invertible endomorphisms encountered in the literature. Given a subset of morphisms $\mathcal{M}$ we denote by $\langle \mathcal{M} \rangle$ the monoid they generate. Also, $\Omega_A$ is the group of automorphisms of $A^*$, i.e., the morphisms obtained by extending any permutation of $A$ to $A^*$.

**Definition 2.1.** *We set the following notations.*
*1.* $\mathcal{I}$ *is the monoid of invertible endomorphisms of* $A^*$.
*2.* $\mathcal{N} = \{N_{ab} \mid a, b \in A\}$.
*3.* $\mathcal{N}^{\rho} = \{N_{ab}^{\rho} \mid a, b \in A\}$.
*4.* $\mathcal{E} = \{E_a \mid a \in A\}$.
*5.* $\mathcal{E}^{\rho} = \{E_a^{\rho} \mid a \in A\}$.

If $|A| \geq 3$, the proper inclusions between the different submonoids are illustrated in the diagram 2. The proper inclusion $\langle \mathcal{N} \cup \mathcal{N}^{\rho} \cup \Omega_A \rangle \subsetneq \mathcal{I}$ follows from the fact that $\mathcal{I}$ is not finitely generated, [18, Th. 2.2.], [11, Th. 10.4], and the proper inclusion $\langle \mathcal{E} \cup \mathcal{E}^{\rho} \cup \Omega_A \rangle \subsetneq \langle \mathcal{N} \cup \mathcal{N}^{\rho} \cup \Omega_A \rangle$ is mentioned above. In [11, Th. 7.1], a presentation of $\langle \mathcal{E} \cup \mathcal{E}^{\rho} \cup \Omega_A \rangle$ is given. If the cardinality of $A$ is three, then $\mathcal{I}$ is equal to $\langle \mathcal{E} \cup \mathcal{E}^{\rho} \cup \Omega_A \rangle$ up to conjugation by words from $A^*$, see [16]. In analogy with the automorphsims of the free associative algebra, cf. [14], the elements in $\langle \mathcal{N} \cup \mathcal{N}^{\rho} \cup \Omega_A \rangle$ are known as the *tame* automorphisms, see [1].

*Example* 2.2. Consider the substitutions

$$\phi_n : \quad a \mapsto ab,\ b \mapsto acb,\ c \mapsto ac^n\,.$$

with $n \geq 2$ over three letters. Since $\{ab, acb, ac^n\}$ is a bifix code, the substitution $\phi_n$ is not tame. Indeed, in the image of a tame automorphism at least one word is a prefix or a suffix of some other according to whether the first generator applied is in $\mathcal{N}$ or in $\mathcal{N}^{\rho}$ respectively. On the other hand, the substitution

$$\phi_n' : \quad a \mapsto ba,\ b \mapsto cba,\ c \mapsto c^n a,$$

which results from composing $\phi_n$ with the inner automorphism $x \mapsto a^{-1}xa$, satisfies $\phi_n' = N_{ba}^{\rho} N_{bc}^{n} N_{ac}^{\rho} \pi$, where $\pi$ denotes the permutation $a \mapsto b, b \mapsto c, c \mapsto a$. This illustrates the above mentioned result that each invertible substitution over three letters can be written as a product of a tame automorphism and an inner automorphism. In [18], it is also shown that $\phi_n$ is indecomposable for each $n$, showing that $\mathcal{I}$ is not finitely generated for $|A| \geq 3$.

For $|A| \geq 4$, inner automorphisms along with tame automorphisms are no longer enough to generate the whole $\mathcal{I}$. To see this, we work out the example given in [16] after Remark 3.2. Consider the substitution

$$\phi : \quad a \mapsto ab,\ b \mapsto acb,\ c \mapsto ac^2,\ d \mapsto d\,, \tag{4}$$

which is a trivial extension of the above $\phi_2$ to the alphabet of four letters. As above for $\phi_n$, we can see that $\phi$ is not tame. The subgroup of inner automorphisms is a normal subgroup of the group of automorphisms. In particular, for all inner automorphisms $\iota_w$, defined by $x \mapsto w^{-1}xw$, and all automorphisms $\theta$, we have

$$\iota_w\theta = \theta\iota_{w\theta}\,.$$

Thus, if the automorphism $\phi$ were a product of tame and inner automorphisms, it could be written as $\phi = \psi\iota_u$ for some tame automorphism $\psi$. Then $d\phi = d\psi\iota_u = u^{-1}(d\psi)u = d$ which easily implies $u = d^k$ for some integer $k$. But then $ab = a\phi = d^{-k}(a\psi)d^k$. Considering $d^kab = (a\psi)d^k$ we obtain that $k = 0$ and $\phi$ is tame, a contradiction.

**Remark 2.3.** *Observe that we have $\langle \mathcal{N} \cup \Omega_A\rangle = \langle \{N_{ab}\} \cup \Omega_A\rangle$ for arbitrary $a \neq b \in A$, since for each $\sigma \in \Omega_A$ we have $N_{\sigma(a)\sigma(b)} = \sigma^{-1}N_{ab}\sigma$.*

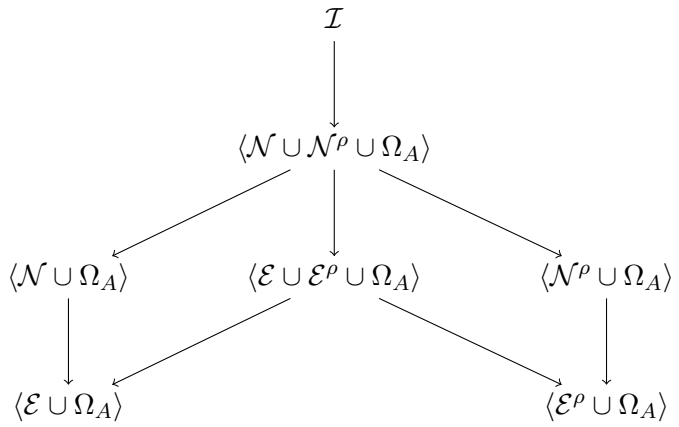


Figure 1: Proper inclusions between families of invertible endomorphisms

We wrap up the preliminaries by observing that the literature provides us with all the material to obtain "for free" the solution in case of a binary alphabet where the above picture is greatly simplified since it reduces to a hierarchy of three monoids. Indeed, from one hand it is proved in [17, Th. 1] that $\mathcal{I} = \langle \mathcal{N} \cup \mathcal{N}^\rho \cup \Omega_A\rangle = \langle N_{ab}, N_{ba}^\rho, X\rangle$, where $X$ is the transposition $a \leftrightarrow b$. On the other hand by definition the sets $\mathcal{E}$ and $\mathcal{N}$ coincide and

$\langle N_{ab}, N_{ba}^{\rho}, X\rangle$ is the monoid of Sturmian morphisms of which Séébold gave a presentation with the following equations, [13]

$$XX = Id$$
$$N_{ab}X\left(N_{ba}^{\rho}\right)^{k}XN_{ba}^{\rho} = N_{ba}^{\rho}XN_{ab}^{k}XN_{ab}, \quad k \geq 0\,.$$

Consequently, the above is the monoid presentation of the invertible morphisms for a binary alphabet and it is the only case when a presentation of $\mathcal{I}$ is known. Concerning $\langle N_{ab}, X\rangle$ it is an easy exercise to prove that it has a presentation with the following two equations

$$XX = Id \quad N_{ab}X = XN_{ba}.$$

From now on we assume that $A$ has at least three elements.

# 3 Monoid presentation

A monoid presentation is concerned with the different ways a given element of a monoid can be written as a product of generators and how it is possible or not to convert one factorization into another by applying a set of rewrite rules.

We start with a modest question: given an endomorphism $\theta \in \langle\mathcal{N}\rangle$, i.e., a composition of elements in $\mathcal{N}$, can we say something about the possible candidates to be the last element of the composition ?

## 3.1 Tracking the possible last elementary transformations

A morphism $N_{ab}$ is called an *elementary transformation*. We say that it is a *b-transformation* to emphasize the fact that $b$ is the only letter which is not invariant under the mapping.

*Example* 3.1. The images of $\{1, 2, 3, 4\}$ by the product of six transformations $N_{31} \cdot N_{14} \cdot N_{21} \cdot N_{13} \cdot N_{24} \cdot N_{43}$ in $\langle\mathcal{N}\rangle$ is

$$1 \mapsto 14321,\ 2 \mapsto 2,\ 3 \mapsto 143,\ 4 \mapsto 2124\,.$$

In this paragraph we show how to determine the possible last elementary transformation of a product of transformations.

Assume that $\theta \in \langle\mathcal{N}\rangle$ is a product of elementary transformations the last of which is $N_{ab}$. Then it is clear that all occurrences of $b$ in all $x\theta$ are preceded by $a$. We are going to show that the opposite implication holds as well. More precisely, we will show that if for some $\theta \in \langle\mathcal{N}\rangle$ each occurrence of each letter $b$ in each $x\theta$ is preceded by $a$ then $\theta$ can be written as a product of elementary transformations the last of them being $N_{ab}$ (Lemma 3.8).

With this idea in mind, we define for a $\theta \in \langle \mathcal{N} \rangle$ the mapping $P_\theta : A \to \mathcal{P}(A \cup \{\#\})$ as follows: for all $b \in A \cup \{\#\}$ we have

$$b \in aP_\theta \Leftrightarrow \begin{cases} \exists c \in A : \ c\theta \in A^* baA^* \\ \text{ or } b = \# \text{ and } \exists c \in A : \ c\theta \in aA^* \, . \end{cases}$$

The relationship between $P_\theta$ and the expression of $\theta$ in terms of elementary transformations is established by the following rules, which allow to compute $P_\theta$ easily:

$$\begin{aligned} cP_\theta &= \{\#\} \text{ for all } c \text{ if } \theta \text{ is identity,} \\ cP_{\theta \cdot N_{ab}} &= \begin{cases} \{a\} & \text{if } c = b, \\ aP_\theta \cup bP_\theta & \text{if } c = a, \\ cP_\theta & \text{otherwise .} \end{cases} \end{aligned} \tag{5}$$

*Example* 3.2. (Continued) For the transformation of Example 3.1 we get

| | | $N_{31}$ | $N_{14}$ | $N_{21}$ | $N_{13}$ | $N_{24}$ | $N_{43}$ |
|---|---|---|---|---|---|---|---|
| 1 | # | 3 | #, 3 | 2 | #, 2 | #, 2 | #, 2 |
| 2 | # | # | # | #, 3 | #, 3 | #, 1, 3 | #, 1, 3 |
| 3 | # | # | # | # | 1 | 1 | 4 |
| 4 | # | # | 1 | 1 | 1 | 2 | 1, 2 |

For future reference, we formulate the following direct consequence of the definition of $P_\theta$.

**Corollary 3.3.** *Let $\theta \in \langle \mathcal{N} \rangle$ and let $\phi_1, \ldots, \phi_m$ be elementary transformations none of which is a b-transformation. Pose*

$$\theta_i = \theta \cdot \phi_1 \cdot \cdots \cdot \phi_i, \ i = 0, \ldots, m$$

*Then*

$$bP_{\theta_0} \subseteq \cdots \subseteq bP_{\theta_i} \subseteq \cdots bP_{\theta_m} \, .$$

*In particular, if b has a unique predecessor in $P_{\theta_m}$, then it also has a unique predecessor in $\theta$.*

This observation gives us the following piece of information about a composition of elementary transformations.

**Lemma 3.4.** *Consider*

$$\theta = \phi_1 \cdot \cdots \cdot \phi_k \text{ with } \phi_i \in \mathcal{N}, i = 1, \ldots, k \, , \tag{6}$$

*and assume that $bP_\theta = \{a\}$. Then the following holds if $a \neq \#$*
*1. there exists some b-transformation in the sequence $\phi_1, \ldots, \phi_k$;*
*2. the last occurrence of a b-transformation in the sequence is $N_{ab}$.*
*Moreover, $bP_\theta = \{\#\}$ if and only if $b\theta = b$.*

*Proof.* It is straightforward that $bP_\theta = \{\#\}$ and $b\theta = b$ hold if and only if there is no $b$-transformation in the sequence $\phi_1, \ldots, \phi_k$.

Assume now that the last occurrence of a $b$-transformation is $\phi_\ell = N_{cb}$ for some $c \in A$ and some $1 \leq \ell \leq k$. Then $bP_{\phi_1 \cdots \phi_\ell} = \{c\}$. By Corollary 3.3, $bP_{\phi_1 \cdots \phi_\ell}$ is included in $bP_\theta$, and

$$\{c\} = bP_{\phi_1 \cdots \phi_\ell} \subseteq bP_\theta = \{a\}$$

implies $c = a$. □

Observe that Lemma 3.4 does not imply that if $N_{ab}$ is the last occurrence of a $b$-transformation in (6) then all occurrences of $b$ are preceded by $a$, see Example 3.2.

The following is trivial and will be used in our proofs.

**Lemma 3.5.** *For each $\theta \in \langle \mathcal{N} \rangle$, we have*

$$\forall a \in A \quad a\theta \in A^* a\,.$$

The following definition is useful when proving claims by induction.

**Definition 3.6.** *The* size *of an endomorphisms $\theta : A^* \to A^*$ is the integer $\sum_{a \in A} (|a\theta| - 1)$ and is denoted by $|\theta|$.*

Since we are working with invertible endomorphisms, we have $|\theta| \geq 0$ and $|\theta| = 0$ if and only if $\theta \in \Omega_A$.

## 3.2 A monoid presentation for $\langle \mathcal{N} \rangle$

We recall that a *monoid presentation* of a monoid $M$ is given by a set of *generators* $A$ and a set of *relations* $R \subseteq A^* \times A^*$ such that $M$ is isomorphic to $A^*/_{\sim_R}$, the quotient of the $A^*$ by the monoid congruence $\sim_R$ generated by the relation $R$. This presentation is written $\langle A; R \rangle$. We shall adopt the notation $u = v$ for each $(u, v) \in R$. For example, the free abelian monoid with two generators has the presentation is $\langle a, b; ab = ba \rangle$ and the free group with two generators has the monoid presentation $\langle a, b, \bar{a}, \bar{b}; a\bar{a} = 1, \bar{a}a = 1, b\bar{b} = 1, \bar{b}b = 1 \rangle$.

Here we give the following presentation of the monoid $\langle \mathcal{N} \rangle$.

**Theorem 3.7.** *A presentation for $\langle \mathcal{N} \rangle$ is given by the set of generators $\mathcal{N}$ and the following set $R$ of relations for all $a, b, c, d \in A$*

$$N_{ab}N_{cd} = N_{cd}N_{ab} \qquad \text{if } \{a, b\} \cap \{c, d\} = \emptyset \text{ or } a = c\,, \tag{7}$$

$$N_{ab}N_{ca} = N_{ca}N_{cb}N_{ab} \qquad \text{if } b \neq c\,. \tag{8}$$

*Proof.* We use the following notation. For $\phi_i, \psi_j \in \mathcal{N}$, $i = 1, \ldots, k$ and $j = 1, \ldots, \ell$ we write

$$\phi_1 \cdot \cdots \cdot \phi_k \underset{R}{=} \psi_1 \cdot \cdots \cdot \psi_\ell \tag{9}$$

if the two handsides are equal in $\mathcal{N}^*/_{\sim_R}$. Sometimes we shall add the number of the relation as a superscript in order to make the derivation explicit. We want to show that $\phi_1 \cdot \cdots \cdot \phi_k = \psi_1 \cdot \cdots \cdot \psi_\ell$ if and only if $\phi_1 \cdot \cdots \cdot \phi_k \underset{R}{=} \psi_1 \cdot \cdots \cdot \psi_\ell$.

It is routine to verify that the equations (7) and (8) are satisfied in $\langle \mathcal{N} \rangle$ which proves the "if" part. It remains to prove the converse. The main ingredient is to show that the two rules (7) and (8) allow us to shift some elementary transformations to the left.

**Lemma 3.8.** *Let $\theta \in \langle \mathcal{N} \rangle$ be such that $bP_\theta = \{a\}$ for some $a, b \in A$. If $\theta = \phi_1 \cdots \phi_k$, $\phi_i \in \mathcal{N}$, $i = 1, \ldots, k$, then*

$$\phi_1 \cdots \phi_k \underset{R}{=} \psi_1 \cdots \psi_\ell \cdot N_{ab} \tag{10}$$

*for some $\psi_i \in \mathcal{N}$, $i = 1, \ldots, \ell$.*

*Proof.* We proceed by induction on the length $|\theta|$. If $|\theta| = 0$, then $\theta$ is identity, and the claim is trivial. Let $|\theta| > 0$.

By Lemma 3.4, there is some $p \in \{1, \ldots, k\}$ such that $\phi_p = N_{ab}$ and no $\phi_{p+1}, \ldots, \phi_k$ is a $b$-transformation. Let $m$ be the smallest integer such that

$$\phi_1 \cdots \phi_k \underset{R}{=} \psi_1 \cdots \psi_\ell \cdot N_{ab} \cdot \psi'_1 \cdots \psi'_m,$$

for some $\psi'_1, \cdots, \psi'_m \in \mathcal{N}$, none of them being a $b$-transformation, and some $\psi_1, \cdots, \psi_\ell \in \mathcal{N}$. We want to show that $m = 0$. Suppose the contrary, and let $\psi'_1 = N_{cd}$, where $d \neq b$.
Case 1. If $\{c, d\} \cap \{a, b\} = \emptyset$ or $c = a$, then

$$\psi_1 \cdots \psi_\ell \cdot N_{ab} N_{cd} \cdot \psi'_2 \cdots \psi'_m \overset{\text{relation(7)}}{\underset{R}{=}} \psi_1 \cdots \psi_\ell \cdot N_{cd} N_{ab} \cdot \psi'_2 \cdots \psi'_m \,,$$

a contradiction with the minimality of $m$.
Case 2. If $d = a$ and $c \neq b$, then

$$\psi_1 \cdots \psi_\ell \cdot N_{ab} N_{ca} \psi'_2 \cdots \psi'_m \overset{\text{relation(8)}}{\underset{R}{=}} \psi_1 \cdots \psi_\ell \cdot N_{ca} N_{cb} N_{ab} \cdot \psi'_2 \cdots \psi'_m \,,$$

again a contradiction.
Case 3. Let finally $c = b$. Let $\alpha$ denote the element $\psi_1 \cdots \psi_\ell$ of $\langle \mathcal{N} \rangle$. Since $\theta = \alpha \cdot N_{ab} \cdot N_{bd} \cdot \psi'_2 \cdots \psi'_m$, Corollary 3.3 yields

$$\{a\} = bP_{\alpha N_{ab}} \subseteq bP_{\alpha N_{ab} N_{bd}} \subseteq bP_\theta = \{a\},$$

which implies $bP_{\alpha N_{ab} N_{bd}} = \{a\}$. From (5), we have

$$bP_{\alpha N_{ab} N_{bd}} = bP_{\alpha N_{ab}} \cup dP_{\alpha N_{ab}},$$

and therefore $dP_{\alpha N_{ab}} = \{a\}$. This, in particular, rules out the case $d = a$, since, in a finite word, not every occurrence of $a$ can be preceded by another occurrence of $a$. Therefore $d \notin \{a, b\}$, and we have $dP_{\alpha N_{ab}} = dP_\alpha = \{a\}$ from (5). Since $|\alpha| < |\theta|$, the induction assumption implies that

$$\psi_1 \cdots \psi_\ell \underset{R}{=} \psi''_1 \cdots \psi''_{\ell'} \cdot N_{ad}$$

for some $\psi''_1, \ldots, \psi''_{\ell'} \in \mathcal{N}$, i.e., $\alpha = \beta \cdot N_{ad}$ for some $\beta \in \langle \mathcal{N} \rangle$. Therefore,

$$\begin{aligned}
\psi_1 \cdots \psi_\ell \cdot N_{ab} \cdot N_{bd} \cdot \psi'_2 \cdots \psi'_m &\underset{R}{=} \psi''_1 \cdots \psi''_{\ell'} \cdot N_{ad} \cdot N_{ab} \cdot N_{bd} \cdot \psi'_2 \cdots \psi'_m \\
&\overset{\text{relation(7)}}{\underset{R}{=}} \psi''_1 \cdots \psi''_{\ell'} \cdot N_{ab} \cdot N_{ad} \cdot N_{bd} \cdot \psi'_2 \cdots \psi'_m \\
&\overset{\text{relation(8)}}{\underset{R}{=}} \psi''_1 \cdots \psi''_{\ell'} \cdot N_{bd} \cdot N_{ab} \cdot \psi'_2 \cdots \psi'_m \,,
\end{aligned}$$

a contradiction. This concludes the proof of Lemma 3.8. □

Now we return to the proof of the theorem and show by induction on $|\theta|$ that for any two factorizations of $\theta$, say $\phi_1 \phi_2 \cdots \phi_k$ and $\psi_1 \cdot \psi_2 \cdots \psi_\ell$ we have

$$\phi_1 \cdot \phi_2 \cdots \phi_k \underset{R}{=} \psi_1 \cdot \psi_2 \cdots \psi_\ell \,.$$

The claim is trivial if $|\theta| = 0$. Let $|\theta| > 0$ and pose $\psi_\ell = N_{ab}$. By Lemma 3.8 applied to $\phi_1 \cdot \cdots \cdot \phi_k$, we obtain

$$\phi_1 \cdot \cdots \cdot \phi_k \underset{R}{=} \chi_1 \cdot \cdots \cdot \chi_h \cdot N_{ab} \,,$$

for some $\chi_1, \ldots, \chi_h \in \mathcal{N}$. Because $\langle \mathcal{N} \rangle$ is cancellative, $\chi_1 \cdot \chi_2 \cdots \chi_h$ and $\psi_1 \cdot \psi_2 \cdots \psi_{\ell-1}$ are two factorizations of the same endomorphism $\theta'$ with $|\theta'| < |\theta|$. By induction, this implies

$$\chi_1 \cdot \cdots \cdot \chi_h \underset{R}{=} \psi_1 \cdot \cdots \cdot \psi_{\ell-1}, \quad \text{thus } \chi_1 \cdot \cdots \cdot \chi_h \cdot \psi_\ell \underset{R}{=} \psi_1 \cdot \cdots \cdot \psi_\ell \,.$$

We conclude by observing that

$$\phi_1 \cdot \cdots \cdot \phi_k \underset{R}{=} \chi_1 \cdot \cdots \cdot \chi_h \cdot \psi_\ell \underset{R}{=} \psi_1 \cdot \cdots \cdot \psi_\ell \,.$$

□

### 3.3 Adding the permutations of the letters

Recall that $\Omega_A$ is the group of automorphisms of the free monoid, i.e., automorphisms generated by a permutation on $A$. Here we give a presentation for $\langle \mathcal{N} \cup \Omega_A \rangle$. Let $E_{ab}$ be the automorphism exchanging the letters $a$ and $b$.

**Theorem 3.9.** *The monoid $\langle \mathcal{N} \cup \Omega_A \rangle$ has the presentation whose generator set is*

$$\{N_{ab} \mid a, b \in A\} \cup \{E_{ab} \mid a, b \in A, a \neq b\}$$

*and whose set of relations $S$ is given by* (7), (8), *the relations*

$$\sigma N_{ab} = N_{\sigma(a)\sigma(b)}\, \sigma \tag{11}$$

*for each $\sigma \in \{E_{ab} \mid a, b \in A, a \neq b\}$, and some set of relations for the presentation of the symmetric group over $A$.*

*Proof.* Similarly to the notation in Theorem 3.7, we set

$$\phi_1 \cdot \cdots \cdot \phi_k \underset{S}{=} \psi_1 \cdot \cdots \cdot \psi_\ell \quad \text{with } \phi_i, \psi_j \in \mathcal{N} \cup \Omega_A \tag{12}$$

if the two handsides are equal in $(\mathcal{N} \cup \Omega_A)^*/_{\sim_R}$. Since it is clear that $\langle \mathcal{N} \cup \Omega_A \rangle$ satisfies the equalities of the statement, it remains to prove that whenever $\phi_1 \cdot \cdots \cdot \phi_k$ and $\psi_1 \cdot \cdots \cdot \psi_\ell$, with $\phi_i, \psi_j \in \mathcal{N} \cup \Omega_A$, are two factorizations of the same element $\theta \in \langle \mathcal{N} \cup \Omega_A \rangle$, then

$$\phi_1 \cdot \cdots \cdot \phi_k \underset{S}{=} \psi_1 \cdot \cdots \cdot \psi_\ell \,.$$

We prove it by induction on $|\theta|$. First, observe that by using repeatedly the relations (11), it is possible to shift all transpositions to the left of each factorization so that we may assume that the two factorizations are of the form

$$\theta = \alpha_1 \cdots \alpha_p \cdot \phi_1 \cdot \cdots \cdot \phi_k = \beta_1 \cdots \beta_r \cdot \psi_1 \cdot \cdots \cdot \psi_\ell \tag{13}$$

where the $\alpha$'s and the $\beta$'s are transpositions and the $\phi$'s and $\psi$'s are elementary transformations.

We want to prove

$$\alpha_1 \cdots \alpha_p \cdot \phi_1 \cdot \cdots \cdot \phi_k \underset{S}{=} \beta_1 \cdots \beta_r \cdot \psi_1 \cdot \cdots \cdot \psi_\ell \,. \tag{14}$$

Because of Lemma 3.5, $\alpha_1 \cdots \alpha_p$ and $\beta_1 \cdots \beta_r$ are equal to the permutation $\gamma$ satisfying, for all $a, b \in A$,

$$a\gamma = b \Leftrightarrow a\theta \in A^* b$$

Then $\alpha_1 \cdots \alpha_p \underset{S}{=} \beta_1 \cdots \beta_r$ by using the presentation of the symmetric group. Since $\langle \mathcal{N} \cup \Omega_A \rangle$, is cancelative we have $\phi_1 \cdots \phi_k = \psi_1 \cdots \psi_\ell$, and we may conclude by Theorem 3.7. □

### 3.4 Stability of $\langle \mathcal{N} \rangle$

If we turn our attention from the last generator to the first one, we obtain the following counterpart of Lemma 3.8.

**Lemma 3.10.** *Let $\theta \in \langle \mathcal{N} \rangle$ be such that $a\theta$ is a prefix of $b\theta$ for some $a, b \in A$. Then $\theta = N_{ab} \cdot \phi$ for some $\phi \in \langle \mathcal{N} \rangle$.*

*Proof.* We proceed by induction on $|\theta|$. The claim is trivial for $|\theta| = 0$. Let $|\theta| > 0$ and let $a\theta$ be a prefix of $b\theta$, written $a\theta < b\theta$. Let $\theta = N_{cd} \cdot \chi$ with $\chi \in \langle \mathcal{N} \rangle$ and $N_{cd} \neq N_{ab}$.

Case 1: If $d \notin \{a, b\}$, then $a\theta = a\chi < b\chi = b\theta$ and by induction we obtain $\chi = N_{ab} \cdot \psi$ with $\psi \in \langle \mathcal{N} \rangle$.

Case 1a: If $c \neq b$, then

$$\theta = N_{cd}N_{ab} \cdot \psi \overset{\text{relation}(7)}{=} N_{ab}N_{cd} \cdot \psi,$$

and we are through.

Case 1b: Similarly, if $c = b$, then

$$\theta = N_{bd}N_{ab} \cdot \psi \overset{\text{relation}(8)}{=} N_{ab}N_{ad}N_{bd} \cdot \psi\,.$$

Case 2: Let $d = a$. Then $c \neq b$ since $a\theta < b\theta$. Note that $c\chi = c\theta < a\theta < b\theta = b\chi$. By induction, we have that $\chi = N_{cb} \cdot \psi$ for some $\psi \in \langle \mathcal{N} \rangle$. Then $a\theta = (c\psi)(a\psi) < b\theta = (c\psi)(b\psi)$ which implies $a\psi < b\psi$. By induction, we have $\psi = N_{ab} \cdot \psi'$ for some $\psi' \in \langle \mathcal{N} \rangle$. Then

$$\theta = N_{ca}N_{cb}N_{ab} \cdot \psi' \overset{\text{relation}(8)}{=} N_{ab}N_{ca} \cdot \psi'\,.$$

Case 3: Let, finally, $d = b$ (and $a \neq c$). Then $a\theta = a\chi < b\theta = (c\chi)(b\chi)$. Therefore $a\chi$ and $c\chi$ are prefix-comparable.

Case 3a: If $a\chi < c\chi$, then, by induction, $\chi = N_{ac} \cdot \psi$ for some $\psi \in \langle \mathcal{N} \rangle$. Therefore

$$\theta = N_{cb}N_{ac} \cdot \psi \overset{\text{relation}(8)}{=} N_{ac}N_{ab}N_{cb} \cdot \psi \overset{\text{relation}(7)}{=} N_{ab}N_{ac}N_{cb} \cdot \psi\,.$$

Case 3b: If $c\chi < a\chi$, then $\chi = N_{ca} \cdot \psi$ for some $\psi \in \langle \mathcal{N} \rangle$. As above, $a\theta = (c\psi)(a\psi) < b\theta = (c\psi)(b\psi)$ implies that $\psi = N_{ab} \cdot \psi'$ for some $\psi' \in \langle \mathcal{N} \rangle$, and

$$\theta = N_{cb}N_{ca}N_{ab} \cdot \psi' \overset{\text{relation}(7)}{=} N_{ca}N_{cb}N_{ab} \cdot \psi' \overset{(8)}{=} N_{ab}N_{ca} \cdot \psi'\,.$$

□

We can now formulate the following "stability" property of $\langle \mathcal{N} \rangle$.

**Theorem 3.11.** *Let $\theta, \phi \in \langle\mathcal{N}\rangle$.*

1. *If $\theta \cdot \phi^{-1} \in \mathcal{I}$, then $\theta \cdot \phi^{-1} \in \langle\mathcal{N}\rangle$.*
2. *If $\phi^{-1} \cdot \theta \in \mathcal{I}$, then $\phi^{-1} \cdot \theta \in \langle\mathcal{N}\rangle$.*

*Proof.* It suffices to consider the basic case $\phi = N_{ab}$. Set $\psi = \theta \cdot \phi^{-1}$. Then $\theta = \psi \cdot N_{ab} \in \langle\mathcal{N}\rangle$. Since $\psi \in \mathcal{I}$, we have $bP_\theta = \{a\}$, and Lemma 3.8 implies that $\theta = \chi \cdot N_{ab}$ with $\chi = \theta \cdot \phi^{-1} \in \langle\mathcal{N}\rangle$.

The proof of the second claim is analogous, using Lemma 3.10. □

# 4 Complexity remarks

We conclude by evoking two natural complexity results. In both cases we leave open a more precise estimate of the complexity.

**Proposition 4.1.** *The following Membership problem is decidable in $O(|\phi|)^2$.*
*Instance: a morphism $\phi : A^* \to A^*$ given by the images $a\phi$ for all $a \in A$.*
*Question: $\phi \in \langle\mathcal{N}\rangle$ resp. $\langle\mathcal{N} \cup \Omega_A\rangle$?*

*Proof.* Consider the morphism $\lambda_\phi : A^* \to A^*$ which assigns to each letter the last letter of its image by $\phi$, i.e., defined for all $a, b \in A$ by

$$a\lambda_\phi = b \text{ if } a\phi \in A^*b\,.$$

If this mapping is not in $\Omega_A$, i.e., if it does not permute $A$, then $\phi$ does not belong to $\langle\mathcal{N}\rangle$ nor $\langle\mathcal{N} \cup \Omega_A\rangle$. If $\lambda_\phi \in \Omega$, then $\phi$ belongs to $\langle\mathcal{N} \cup \Omega_A\rangle$ if and only if $\lambda_\phi^{-1}\phi$ belongs to $\langle\mathcal{N}\rangle$. It thus suffices to consider the question $\phi \stackrel{?}{\in} \langle\mathcal{N}\rangle$.

Theorem 3.11 yields that for any $\theta : A^* \to A^*$ we have

$$\theta \in \langle\mathcal{N}\rangle \Leftrightarrow N_{ab} \cdot \theta \in \langle\mathcal{N}\rangle\ .$$

This leads to the following procedure, starting with $X = \{a\phi \mid a \in A\} \subseteq A^*$:

$$\begin{aligned} &\textbf{while } \exists\, x, y \in X : x < y \textbf{ do} \\ &\qquad X := X \setminus \{y\} \cup \{x^{-1}y\} \end{aligned}$$

If the procedure ends with $X = A$, then $\phi$ belongs to $\langle\mathcal{N}\rangle$, otherwise it does not. The number of while-loops is bounded by $|\phi|$. Testing whether or not a word of a subset $X$ is a prefix of another word of the subset can again be achieved in $|\phi|$ by constructing the prefix-tree of $\{a\phi \mid a \in A\}$. □

**Proposition 4.2.** *There exists a polynomial algorithm deciding the following problem*
*Instance: two words from $\mathcal{N}^*$*
*Question: are these two words equivalent in $=_R$ ?*

*Proof.* It is proved in [12] that the word problem in the automorphism group of the free group is polynomially decidable. This covers the word problem for the monoid $\langle \mathcal{N} \cup \Omega_A \rangle$. □

## References


[1] Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Maximal bifix decoding. *Discrete Mathematics*, 338(5):725–742, 2015.

[2] Steven .M. Gersten. A presentation for the special automorphism group of a free group. *J. Pure Appl. Alg*, 33:269–279, 1984.

[3] Yuri I. Hmelevskii. *Equations in free semigroups*, volume 107. Proc. Steklov Inst. of Math., 1971.

[4] Jacques Justin and Giuseppe Pirillo. Episturmian words and Episturmian morphisms. *Theor. Comput. Sci.*, 276(1-2):281–313, 2002.

[5] André Lentin. *Equations dans les monoïdes libres*. Gauthier-Villars, Paris, 1972.

[6] M. Lothaire. *Combinatorics on Words*, volume 17 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley, Gian-Carlo Rota edition, 1983.

[7] M. Lothaire. *Makanin's Algorithm*, volume 20 of *Encyclopedia of Mathematics and its Applications*. Cambridge Journals, 2002.

[8] James McCool. A presentation for the automorphism group of a free group of finite rank. *J. London Math.Soc.*, 8:259–266, 1974.

[9] Jakob Nielsen. The isomorphismengruppe des freien Gruppen. *Math. Ann.*, 91:169–209, 1924.

[10] Jacques Peyrière, Zhi-Xiong Wen, and Zhi-Ying Wen. On the dynamic behaviors of the iterations of the trace map associated with substitutive sequences. *Nonlinear problems in Engineering and Sciences*, pages 259–266, 1992.

[11] Gwénaël Richomme. Conjugacy and Episturmian morphisms. *Theor. Comput. Sci.*, 302(1-3):1–34, 2003.

[12] Saul Schleimer. Polynomial time word problems, 2008. arxiv.org/pdf/math/0608563.

[13] Patrice Séébold. Fibonacci morphisms and sturmian words. *Theoretical Computer Science*, 88(2):365 – 384, 1991.

[14] Ivan P. Shestakov and Ualbai U. Umirbaev. The tame and the wild automorphisms of polynomial rings in three variables. *Jounal of the Amarican Mathematical Society*, 17(1):197–227, 2004.

[15] John R. Stallings. Topology of finite graphs. *Invent. Math.*, 71(3):551–565, 1983.

[16] Bo Tan, Zhi-Xiong Wen, and Yiping Zhang. The structure of invertible substitutions on a three-letter alphabet. *Adv. Applied Math.*, 32:736–753, 2004.

[17] Zhi-Xiong Wen and Zhi-Ying Wen. Local isomorphism of the invertible substitutions. *C. R. Acad. Sci. Paris Sér. I Math.*, 318:299–304, 1994.

[18] Zhi-Xiong Wen and Yiping Zhang. Some remarks on invertible substitutions on three letter alphabet. *Chinese Science Bulletin*, 44(19):1755–1760, 1999.