The syntactic monoid of hairpin-free languages

Lila Kari · Kalpana Mahalingam · Gabriel Thierrin

Abstract The study of hairpin-free words has been initiated in the context of DNA computing. DNA strands that, theoretically speaking, are finite strings over the alphabet $\{A, G, C, T\}$ are used in DNA computing to encode information. Due to the fact that A is complementary to T and G to C, DNA single strands that are complementary can bind to each other or to themselves in either intended or unintended ways. One of the structures that is usually undesirable for biocomputation, since it makes the affected DNA string unavailable for future interactions, is the hairpin: if some subsequences of a DNA single string are complementary to each other, the string will bind to itself forming a hairpin-like structure. This paper continues the theoretical study of hairpin-free languages. We study algebraic properties of hairpin-free words and hairpins. We also give a complete characterization of the syntactic monoid of the language consisting of all hairpin-free words over a given alphabet and illustrate it with an example using the DNA alphabet.

1 Introduction

The topic of this paper is the study of the hairpin-free words and languages, mainly through their syntactic monoid. The object of this study, hairpin words and languages, stems from the practical requirements of DNA computing experiments. DNA strands can be viewed as finite strings over the alphabet $\{A, G, C, T\}$ and are used in DNA computing to encode information. Due to the fact that A is complementary to T

L. Kari · K. Mahalingam (✉)
Department of Computer Science, University of Western Ontario,
London, ON, Canada N6A 5B7
e-mail: kalpana@csd.uwo.ca

L. Kari
e-mail: lila@csd.uwo.ca

G. Thierrin
Department of Mathematics, University of Western Ontario,
London, ON, Canada N6A 5B7

Fig. 1 A simple hairpin loop

$$\begin{array}{l}\text{GTCAGCGATAG}\\ \phantom{\text{G}}|\ |\ |\ |\ |\ |\ |\ |\ |\ |\\ \text{CAGTCGCTATC}\end{array}$$

and G to C, DNA single strands that are complementary can bind to each other or to themselves in either intended or unintended ways. One of the structures that is usually undesirable for biocomputation, since it makes the affected DNA string unavailable for future interactions, is the hairpin: if some subsequences of a DNA single string are complementary to each other, the string will bind to itself forming a hairpin-like structure. An example of a DNA hairpin structure is shown in Fig. 1. Algebraic properties of languages that avoid DNA sequences undesirable for DNA based computations, such as those that can form hairpins, have been extensively studied in [6,10,11]. The notion of a hairpin structure was formalized and its coding properties as well as relations between hairpin-free languages and other types of codes have been discussed in [6,15]. Certain algebraic properties of hairpin-free languages were discussed in [11]. Effective methods for the design of long hairpin-free DNA words were given in [11]. In this paper we study the algebraic properties of the set of all hairpins and hairpin-free words through their syntactic monoid.

The syntactic monoid approach to the study of languages has been very successful. Algebraic characterizations of many classes of codes through their syntactic monoid have been extensively studied [7,16–18]. Several properties of syntactic congruences for infix, outfix and hypercodes have been presented in [5,7,9,17,23]. Additional properties of syntactic congruences for infix and outfix codes have been presented in [14]. The characterization of a syntactic monoid of a semigroup generated by a comma-free code and prefix code have been discussed in [16] and [18], respectively. The characterizations of the syntactic monoid of a comma-free code X and X^+ were discussed in [16]. The syntactic characterization of strictly locally testable languages were discussed in [13]. In [7], the author formulated a general characterization method of the syntactic monoid which applies to all classes of codes that can be defined in a certain way and hence results analogous to those of [17] can be obtained for a large variety of classes of codes. For more details on codes the reader is referred to [1,8,20,24]. In [2], the authors have introduced a special type of regular languages using the syntactic congruence.

In this paper we follow the approach from [16,17] and extend these concepts to study the algebraic properties of the set of all hairpin-free words mainly through their syntactic monoid.

The paper is organized as follows: Sect. 2 reviews basic definitions. Section 3 presents several properties of the set of all hairpins and hairpin-free words. In particular we show that the set of all hairpin-free words and its complementary set are commutative. In Sect. 4 we study the properties of the syntactic monoid of hairpin-free languages. We show that all the elements of the syntactic monoid of the language of all hairpin-free words over a given alphabet are idempotents, and proceed to use this property to show that the underlying syntactic semigroup is indeed an inverse semigroup and also that the language of all hairpin-free words over a given alphabet is locally testable. Propositions 17 and 18 give necessary and sufficient condition for a finite monoid to be the syntactic monoid of the set of all hairpin free words over a given finite alphabet. These results give thus complete characterization of the syntactic

monoid of the language of hairpin-free words over an alphabet. We also discuss the Green's relations for hairpin-free languages.

2 Definitions and basic concepts

In this section we review some basic notions. An alphabet Σ is a finite non-empty set of symbols. A word u over Σ is a finite sequence of symbols in Σ. We denote by Σ^* the set of all words over Σ, and by Σ^+ the set of all non empty words over Σ. The empty word is denoted by λ. For words u, v over Σ^* we denote by uv the concatenation u and v. We note that with the concatenation operation on words, Σ^* is the free monoid and Σ^+ is the free semigroup generated by Σ. The length of a word $u = a_1 \cdots a_n$, $a_i \in \Sigma, 1 \leq i \leq n$, is n and is denoted by $|u|$. We denote by Σ^k the set of all words in Σ of length k. A language over Σ is an arbitrary subset of Σ^*.

A mapping $\theta : \Sigma^* \mapsto \Sigma^*$ is called a morphism (antimorphism) of Σ^* if $\theta(uv) = \theta(u)\theta(v)$ (respectively $\theta(uv) = \theta(v)\theta(u)$) for all $u, v \in \Sigma^*$. A d-morphism of Σ^* is either a morphism or an antimorphism of Σ^*. If θ is bijective, then θ has an inverse θ^{-1} such that $\theta\theta^{-1} = \theta^{-1}\theta = I$, the identity mapping. An involution map θ is such that $\theta^2 = I$. An involution map θ is a particular case of a bijective map where $\theta = \theta^{-1}$. The d-morphism θ is said to be *length-preserving* if $|u| = |\theta(u)|$ for all $u \in \Sigma^*$. Since Σ is finite, this is equivalent to $\theta(\Sigma) \subseteq \Sigma$. Bijective d-morphisms, in particular involutions, are examples of length preserving d-morphisms. In this paper we concentrate on relations that are defined by length preserving d-morphisms.

The following was defined in [11] for an involution θ. An analogous definition was given in [6].

Definition 1 Let θ be a length preserving d-morphism and $k \in \mathbb{N}$.

1. A word $u \in \Sigma^*$ is called an (θ, k)-*hairpin* if there exists $x, v, y, z \in \Sigma^*$ such that $u = xvy\theta(v)z$ or $u = x\theta(v)yvz$ with $|v| \geq k$.
2. A word $u \in \Sigma^*$ is said to be an (θ, k)-*hairpin-free* if $u = xvy\theta(v)z$ or $u = x\theta(v)yvz$ where $x, v, y, z \in \Sigma^*$ implies $|v| < k$.

For a length preserving d-morphism θ of Σ and $k \in \mathbb{N}$, we denote by $hpf(\theta, k)$ the set of all (θ, k)-hairpin-free words in Σ^*. We denote by $hp(\theta, k)$ its complement, i.e., $hp(\theta, k)$ is the set of words in Σ^* having at least a hairpin, i.e., $u = xvy\theta(v)z$ or $u = x\theta(v)yvz$ with $|v| \geq k$ and $x, y, v, z \in \Sigma^*$. A language L over Σ is called $hp(\theta, k)$-free if $L \subseteq hpf(\theta, k)$. Throughout the rest of the paper θ is a length preserving d-morphism unless stated otherwise and we denote $hp(\theta, 1)$ and $hpf(\theta, 1)$ by $hp(\theta)$ and $hpf(\theta)$.

Lemma 1 *A language L is $hp(\theta, k)$-free if and only if $\Sigma^* v \Sigma^* \theta(v) \Sigma^* \cap L = \emptyset$ and $\Sigma^* \theta(v) \Sigma^* v \Sigma^* \cap L = \emptyset$ for all $v \in \Sigma^*$, $|v| \geq k$.*

Proof Immediate from the definition. □

3 Properties of hairpin-free languages

Recall that the embedding order $\leq_e$ is defined as follows:
$u \leq_e v$ if and only if:

$$u = u_1 u_2 \cdots u_n, \quad v = v_1 u_1 v_2 u_2 \cdots v_n u_n v_{n+1}$$

for some integer n and $u_i, v_j \in \Sigma^*, 1 \leq i \leq n$ and $1 \leq j \leq n+1$.

A language L is called $\leq_e$-convex if $u \leq_e w \leq_e v$, $u, v \in L$ implies $w \in L$. It is called *right* $\leq_e$-convex if $u \leq_e w$, $u \in L$ implies $w \in L$,

The following result is well known (see [3,22]):

Every $\leq_e$-convex and right $\leq_e$-convex language over a finite alphabet is regular.

In the next propositions we show that the set of all words with at least a hairpin is closed under insertion and is a right $\leq_e$-convex language and hence using the result in [3,22] we show that $hp(\theta, k)$ and $hpf(\theta, k)$ are regular. In [11], it was shown that $hp(\theta, k)$ and $hpf(\theta, k)$ are regular when θ is an involution. We will show that the results hold true for a d-morphism θ.

Lemma 2 *Let θ be a length preserving d-morphism and $k \in \mathbb{N}$, $k \geq 1$. Then every hairpin $u = x'v'y'\theta(v')z' \in hp(\theta, k)$ can be written in the form $u = xvy\theta(v)z$ with $|v| = 1$.*

Proof If $v' = v_1 \cdots v_k$, for $v_i \in \Sigma$ then $\theta(v') = \theta(v_1) \cdots \theta(v_k)$ if θ is a morphism or $\theta(v') = \theta(v_k) \cdots \theta(v_1)$ if θ is an anti-morphism. Hence when θ is a morphism u can be written as $u = x'v_1v_2 \cdots v_k y'\theta(v_1)\theta(v_2) \cdots \theta(v_k)z'$ and when θ is an antimorphism $u = x'v_1v_2 \cdots v_k y'\theta(v_k) \cdots \theta(v_2)\theta(v_1)z'$. Taking $v = v_1$, we get u in the form $u = xvy\theta(v)z$ with $|v| = 1$. $\square$

Proposition 1 *The language $hp(\theta)$ is closed under insertion, i.e., $u = u_1u_2 \in hp(\theta)$ and $w \in \Sigma^*$ implies $u_1wu_2 \in hp(\theta)$.*

Proof Since $u \in hp(\theta)$, from the previous lemma the word u can be written as $u = xvy\theta(v)z$ with $|v| = 1$. Since θ is length preserving we have $|\theta(v)| = |v| = 1$. This implies that the insertion of w in u can occur either to the left of v, between v and $\theta(v)$ or to the right of $\theta(v)$. The word u_1wu_2 can be written as $u_1wu_2 = x''vy''\theta(v)z''$ where $u = u_1u_2$. Therefore $u_1wu_2 \in hp(\theta)$. $\square$

Proposition 2 *The language $hp(\theta)$ is right $\leq_e$-convex.*

Proof Let $u \in hp(\theta)$ and suppose that $u \leq_e w$, i.e., $u = u_1u_2 \cdots u_n$ and $w = w_1u_1w_2u_2 \cdots w_nu_nw_{n+1}$. The word w is obtained from u through a finite sequence of insertions of the words w_i in u. Since $u \in hp(\theta)$, then, by Proposition 1, $w \in hp(\theta)$ and hence $hp(\theta)$ is right $\leq_e$-convex. $\square$

Corollary 1 *The languages $hp(\theta)$ and $hpf(\theta)$ are regular.*

Let $L \subseteq \Sigma^*$ be a nonempty language and let:

$$\tilde{L} = \{w \in \Sigma^* | u \leq_e w, u \in L\}$$

Hence $\tilde{L}$ is the set of all the words $w \in \Sigma^*$ that can be expressed in the form $w = x_1u_1x_2u_2 \cdots x_nu_nx_{n+1}$ with $u = u_1u_2 \cdots u_n \in L$ and $x_i \in \Sigma^*$.

Recall that a nonempty set $H \subset \Sigma^+$ is called a hypercode over Σ^* iff $x \leq_e y$ and $x, y \in H$ imply $x = y$. That is, a hypercode is an independent set with respect to the embedding order.

Proposition 3 *There exists a unique hypercode H such that $hp(\theta) = \tilde{H}$.*

Proof Let H be the set of all minimal words in $hp(\theta)$ relative to the embedding order $\leq_e$. A word u is in H if and only if $u \in H$ and $v \in hp(\theta)$ with $v \leq_e u$ implies $v = u$. The set H is a hypercode, i.e., words in H are not $\leq_e$-comparable and H is finite since every hypercode over a finite alphabet is finite. Furthermore $hp(\theta) \subseteq \tilde{H}$ by Proposition 2.

For the converse inclusion, let $w \in \tilde{H}$. Then $u \leq_e w$ for some $u \in H$. By Proposition 2, $hp(\theta)$ is right $\leq_e$-convex. Hence $w \in hp(\theta)$ and $\tilde{H} \subseteq hp(\theta)$. Therefore $hp(\theta) = \tilde{H}$. The uniqueness of H is immediate. □

Recall that for a word $w = a_1 a_2 \cdots a_n \in \Sigma^*$ for $a_i \in \Sigma$, $1 \leq i \leq n$, $\pi(w) = \{a_{s(1)} a_{s(2)} \cdots a_{s(n)} | s$ a permutation of $\{1, 2, \ldots, n\}\}$.

We also recall the following definitions.

Definition 2 A language L is said to be:

1. θ-stable if $\theta(L) \subseteq L$.
2. Commutative iff $L = \pi(L)$ which is equivalent to the condition that for all $x, u, v, y \in \Sigma^*$ if $xuvy \in L$ then $xvuy \in L$.
3. Factorial if $Sub(L) = L$ where $Sub(L) = \{x \in \Sigma^* | pxq \in L$ for some $p, q \in \Sigma^*\}$.
4. Transitive if for all $x, y \in L$ there exists $z \in \Sigma^*$ such that $xzy \in L$.
5. Prolongable if for all $x \in L$ there exists $p, q \in \Sigma^+$ such that $pxq \in L$.

Note that for a commutative language L, the complement of L denoted by $\bar{L}$ is also commutative.

Proposition 4 *The language $hpf(\theta)$ is factorial.*

Proof Note that for all $u \in hp(\theta)$, u is hairpin-free and hence every subword of u is also hairpin-free which implies $Sub(hpf(\theta)) = hpf(\theta)$. Thus $hpf(\theta)$ is factorial. □

Proposition 5 *Let θ be such that for all $a \in \Sigma$, $\theta(a) \neq a$. Then $hpf(\theta)$ is prolongable.*

Proof Let $x \in hpf(\theta)$ such that $x = x_1 x_2 \cdots x_n$ for $x_i \in \Sigma$. Take $p \in \Sigma$ such that $p = x_i$ for some i such that $1 \leq i \leq n$. Since $\theta(x_i) \neq x_i$ we have $pxp \in hp(\theta)$. Hence $hp(\theta)$ is prolongable. □

Proposition 6 *$hp(\theta)$ is prolongable and transitive.*

Proof Let $x \in hp(\theta)$. Then x contains a hairpin and hence for all $p, q \in \Sigma^+$, pxq also contains a hairpin and hence $pxq \in hp(\theta)$ which implies $hp(\theta)$ is prolongable.

Let $x, y \in hp(\theta)$. Then for all $z \in \Sigma^*$, xzy contains a hairpin and hence $xzy \in hp(\theta)$. Thus $hp(\theta)$ is transitive. □

Proposition 7 *The languages $hpf(\theta)$ and $hp(\theta)$ are commutative.*

Proof Let $w = xuvy \in hpf(\theta)$. If $xvuy \notin hpf(\theta)$, then $xvuy$ is a hairpin, i.e., $xvuy = paq\theta(a)r$ with $|a| = 1$ and $|\theta(a)| = 1$. This implies that both a and $\theta(a)$ are letters of the words x, u, v, y. We have to consider the following cases:

$a, \theta(a)$ are letters of x. Then $w = x'aq\theta(a)r'uvy$ for $x', q, r', u, v, y \in \Sigma^*$ and $w \notin hpf(\theta)$, a contradiction.

a is a letter of x and $\theta(a)$ is a letter of vu and hence of uv. Then $w = x'ax''z'\theta(a)z''y$ and $w \notin hpf(\theta)$, a contradiction.

a is not a letter of x. Then both a and $\theta(a)$ are letters of vuy and hence of uvy. This implies that w is a hairpin, a contradiction. □

Throughout the rest of the paper we consider θ to be a bijective d-morphism on Σ^* (which is a particular case of length preserving d-morphism) and denote by F_Σ the set of all $hp(\theta)$-free words over Σ, i.e., $u \in F_\Sigma$ iff u is not an hairpin. In the sequel whenever the alphabet Σ is clear from the context we will omit the subscript Σ. By Corollary 1, F is regular. In the following proposition we discuss various other properties that are satisfied by F under the new stronger hypothesis that θ is a bijective d-morphism.

Proposition 8 *The language F is θ-stable.*

Proof Let $u \in F$ and suppose that $\theta(u) \notin F$. Then $\theta(u)$ can be written as $\theta(u) = xvy\theta(v)z$ with $v \neq \lambda$.

Since θ is bijective, then there exists $w \in \Sigma^*$, $w \neq \lambda$ such that $\theta(w) = v$ and hence $w = \theta^{-1}(v)$.

If θ is a morphism, then, since θ is bijective $u = \theta^{-1}\theta(u)$ which implies $u = \theta^{-1}(x)\theta^{-1}(v)\theta^{-1}(y)v\theta^{-1}(z)$. Hence $u = \theta^{-1}(x)w\theta^{-1}(y)\theta(w)\theta^{-1}(z)$. Since $w \neq \lambda$, then u is a hairpin, i.e., $u \notin F$, a contradiction.

If θ is an anti-morphism, then $u = \theta^{-1}\theta(u) = \theta^{-1}(z)v\theta^{-1}(y)\theta^{-1}(v)\theta^{-1}(x)$. Since θ is bijective, we have $v = \theta(w)$ for some $w \in \Sigma^*$, $w \neq \lambda$. Hence $u = \theta^{-1}\theta(u) = \theta^{-1}(z)\theta(w)\theta^{-1}(y)w\theta^{-1}(x)$ which implies that u is a hairpin, a contradiction.

Therefore F is θ-stable.   □

Note that the above proposition may not be true if θ is not bijective.

4 The syntactic monoid of the set of all hairpin-free words

In the theory of codes, two types of syntactic monoids are usually considered, the syntactic monoid of the code itself and the syntactic monoid of the Kleene star of the code. In this section we concentrate on the characterizations of syntactic monoid of the set of all hairpin-free words $hpf(\theta) = F$, when θ is a bijective d-morphism. Necessary and sufficient conditions for a monoid to be the syntactic monoid of the set of all hairpin-free words are also discussed. We first review some basic concepts.

Let L be a language such that $L \subseteq \Sigma^+$. We define the context, right context and left context of a word $w \in \Sigma^*$ in L as follows:

- $C_L(w) = \{(u,v)|uwv \in L, u,v \in \Sigma^*\}$.
- $\mathcal{R}_L(w) = \{u \in \Sigma^*|wu \in L\}$.
- $\mathcal{L}_L(w) = \{u \in \Sigma^*|uw \in L\}$.

$C_L(w)$, $\mathcal{R}_L(w)$ and $\mathcal{L}_L(w)$ are called the context, right context and left context of w in L, respectively.

Definition 3 Let L be a language such that $L \subseteq \Sigma^+$.

1. The syntactic congruence of $L \subseteq \Sigma^+$ is denoted by P_L and is defined by $u \equiv v(P_L)$ iff $C_L(u) = C_L(v)$.
2. The syntactic monoid of L is the quotient monoid $M(L) = \Sigma^*/P_L$ with the operation $[x][y] = [xy]$, where for $x \in \Sigma^*$, $[x]$ denotes the P_L equivalence class of x.

Let $W(L) = \{x \in \Sigma^* | C_L(x) = \emptyset\}$, i.e., $x \in W(L)$ iff $x \notin \mathrm{Sub}(L)$. $W(L)$ is called the residue of L.

Note that if $W(L) \neq \emptyset$ then $W(L)$ represents a class for P_L and is the zero of $M(L)$.

Example 1 Let $L = ab^*$ over the alphabet set $\Sigma = \{a, b\}$. Then
$C_L(ab^*) = \{(\lambda, b^*)\}$
$C_L(b^*) = \{(ab^*, b^*)\}$
$C_L(\lambda) = \{(\lambda, ab^*), (ab^*, b^*)\}$
$W(L) = \bar{L}$
Hence $M(L) = \{0, [1], [ab^*], [b^*]\}$.

Note that for a regular language L, $M(L)$ is the transition monoid (see [19]) of the minimal deterministic finite automaton (see [1,19]) of L. The above definition of the syntactic congruence P_L can be defined for an arbitrary subset L of any semigroup S. If the syntactic congruence is the equality relation then we call the set L to be a disjunctive subset of S. If $L = \{x\}$ for some $x \in \Sigma^*$ and if P_L is the equality relation then we say that x is a disjunctive element of S. For more on syntactic monoid we refer the reader to [1,12,19].

It is a well known fact that L is a regular language if and only if $M(L)$ is finite (see [12,19]). For any set L and its syntactic monoid $M(L)$, $\eta : \Sigma^* \to M(L)$ is the natural surjective syntactic morphism defined by $x \to [x]$. Note that for any L, L is a union of P_L classes.

We have the following observations on the set of all hairpin-free words F.

Proposition 9 *Let* $\eta : \Sigma^* \mapsto \mathrm{Syn}(F)$. *Then* $\mathrm{Syn}(F)$ *has a zero such that* $\eta^{-1}(\mathrm{Syn}(F) \setminus \{0\}) = F$.

If θ *is such that* $\theta(a) \neq a$ *for all* $a \in \Sigma$ *then for all non zero elements* $[x] \in \mathrm{Syn}(F)$ *there exists* $p, q \neq \lambda$ *such that* $[pxq] \neq 0$.

Proof Follows from the fact that F is factorial and prolongable. □

Proposition 10 uP_Fv *if and only if* $\mathcal{R}_F(u) = \mathcal{R}_F(v)$ *and* $\mathcal{L}_F(u) = \mathcal{L}_F(v)$.

Proof Follows from the fact that F is commutative. □

In the following proposition we show that every non zero element of $\mathrm{Syn}(F)$ is an idempotent element.

Proposition 11 *For every* $u \in \Sigma^*$, *we have* $u\, P_F\, u^2$.

Proof The congruence P_F is equivalent to the congruence $P_{\bar{F}}$ associated to the complement $\bar{F}$ of F and $\bar{F}$ is an ideal of Σ^*. Hence we have to show that $u\, P_{\bar{F}}\, u^2$, i.e., $\mathcal{R}_{\bar{F}}(u) = \mathcal{R}_{\bar{F}}(u^2)$.

Let $x \in \mathcal{R}_{\bar{F}}(u)$ then $ux \in \bar{F}$, and since $\bar{F}$ is an ideal, $u.ux = u^2x \in \bar{F}$ which implies that $x \in \mathcal{R}_{\bar{F}}(u^2)$.

Conversely let $x \in \mathcal{R}_{\bar{F}}(u^2)$ then $u^2x \in \bar{F}$, and by Lemma 2, u^2x can be written as $u^2x = x'vy'\theta(v)z'$, with $|v| = 1$ where the lengths of the words z' and y' in this expression are minimal. Since $|v| = 1$, this implies that v is either a letter of u or x. Depending on the position of the letter v appearing in the expression $x'vy'\theta(v)z'$, we have the following cases.

v is not a letter of u. Then this implies that v is a letter of x and $ux = x''vy'\theta(v)z' \in \bar{F}$.

v is a letter of u, but $\theta(v)$ is not. Then ux can be written as $ux = x'''vy''\theta(v)z'$ and $ux \in \bar{F}$.

v and $\theta(v)$ are letters of u in that order. In this case, $u = rvs\theta(v)t$ and $ux \in \bar{F}$.

Therefore $u^2x \in \bar{F}$ implies $ux \in \bar{F}$ and hence $x \in \mathcal{R}_{\bar{F}}(u)$. $\qquad\square$

Corollary 2 *The elements of the syntactic monoid $Syn(F)$ of F are idempotent elements.*

Proof The fact that $u\ P_F\ u^2$ for any $u \in \Sigma^*$ implies that $U = U^2$ for the class U containing u. $\qquad\square$

Corollary 3 *$Syn(F)$ is a semi lattice.*

Recall that a semigroup in general is a set equipped with an internal associative operation which is usually written in a multiplicative form. A monoid is a semigroup with an identity element (usually denoted by e). If S is a semigroup, S^1 denotes the monoid equal to S if S has an identity element and to $S \cup \{e\}$ otherwise. In the latter case, the multiplication on S is extended by setting $se = es = s$ for all $s \in S$. We also recall the following definitions.

Definition 4 A semigroup S is called an inverse semigroup if each element has a unique inverse. (a' is an inverse of a iff $aa'a = a$ and $a'aa' = a'$).
An element $a \in S$ is called regular if there exists $s \in S$ such that $asa = a$.
A semigroup S is said to be regular if all elements of S are regular.

Note that the notion of regularity of semigroups is not to be confused with the notion of regular languages.

Note that since F is commutative (see Proposition 7), $Syn(F)$ is a commutative monoid, i.e., for all $[x], [y] \in Syn(F)$ we have $[x][y] = [y][x]$. Using this fact we show that the underlying semigroup is an inverse semigroup.

Proposition 12 $S = Syn(F) \setminus \{1\}$ *is an inverse semigroup.*

Proof Proposition 2.12 in [12] states that a semigroup S is an inverse semigroup iff S is regular and any two idempotents commute. Since $Syn(F)$ is commutative any two idempotent elements commute and since all elements of S are idempotent elements S is regular. Hence S is an inverse semigroup. $\qquad\square$

Proposition 13 *For all elements $g \in Syn(F)$, $gSyn(F)g$ is a semi lattice, i.e., for all $s, t \in Syn(F)$, $gsgsg = gsg$ and $gsgtg = gtgsg$, i.e., $gSyn(F)g$ is idempotent and commutative.*

Proof Since every element in $Syn(F)$ is an idempotent element and $Syn(F)$ is commutative we have $gsgsg = gssgg = gsg$ and $gsgtg = gstgg = gtsgg = gtgsg$. Hence $gSyn(F)g$ is a semi lattice for all $g \in Syn(F)$. $\qquad\square$

Recall that a language L is said to be n-locally testable if whenever u and v have the same factors of length at most n and the same prefix and suffix of length $n - 1$ and $u \in L$ then $v \in L$. The language L is locally testable if it is n-locally testable for some $n \in \mathbb{N}$.

A language L is said to be strictly locally testable if there are finite sets $P, Q, R \subseteq \Sigma^*$ such that $L = (P\Sigma^* \cap \Sigma^*Q) \setminus \Sigma^*R\Sigma^*$. We also recall a characterization of the syntactic semigroup of locally testable languages which states that (Proposition 2.1 in [13])

a recognizable subset (A language is called recognizable if there exists an algorithm that accepts a given string if and only if the string belongs to that language) L of Σ^+ is locally testable iff for all idempotents $g \in \mathrm{Syn}(L)$, $g\mathrm{Syn}(L)g$ is a semi lattice. Using Propositions 13 and Proposition 2.1 in [13], we show that the set of all hairpin-free words is locally testable.

Corollary 4 *F is locally testable.*

Proof Since $\mathrm{Syn}(F)$ is finite, F is recognizable and by Proposition 13, we have that $g\mathrm{Syn}(F)g$ is a semi lattice for all $g \in \mathrm{Syn}(F)$ and hence F is locally testable. □

It is well known result (see [13]) that if L is strictly locally testable then for all idempotent elements $g \in \mathrm{Syn}(L)$, $g\mathrm{Syn}(L)g \subseteq \{g, 0\}$. We show that F is not strictly locally testable in general with an example where the bijective map in particular is an involution.

Example 2 Let $\Delta = \{A, G, C, T\}$ be the DNA alphabet and let θ be an antimorphic involution such that $A \mapsto T$, $G \mapsto C$ and viceversa. Then
$\mathrm{Syn}(F_\Delta) = \{1, [A], [G], [C], [T], [AC], [AG], [GT], [CT], 0\}$. We have that:

- $[A]1[A] = [A]$
- $[A][A][A] = [A]$
- $[A][G][A] = [AG]$
- $[A][C][A] = [AC]$
- $[A][T][A] = 0$
- $[A][AC][A] = [AC]$
- $[A][AG][A] = [AG]$
- $[A][GT][A] = 0$
- $[A][CT][A] = 0$
- $[A]0[A] = 0$

Every element in $\mathrm{Syn}(F_\Delta)$ is an idempotent element but $[A]\mathrm{Syn}(F_\Delta)[A] = \{[A], [AC], [AG], 0\} \neq \{[A], 0\}$. Hence F_Δ is not strictly locally testable.

Proposition 14 *For all $a \in \Sigma$ such that $\theta(a) \neq a$, $[a] = \{a^n | n \geq 1\}$ is an element of $\mathrm{Syn}(F)$.*

Proof Suppose there exists $w \in [a]$ such that $w \neq a^k$ then $w = xcy$ for some $c \in \Sigma$ such that $c \neq a$ and $x, y \in \Sigma^*$. Note that from our assumption, for all $a \in \Sigma, \theta(a) \neq a$. Hence we have the following two cases:

1. $\theta(c) = a$ and since $\theta(a) \neq a$ we have $a^2 \in F$ but $ca \notin F$ since ca is a hairpin and hence $xcya = wa \notin F$ which implies that $w \notin [a]$, a contradiction.
2. $\theta(c) = d$ for some $d \neq a$, then $ad \in F$ but $cd \notin F$ since cd is an haripin and hence $xcyd = wd \notin F$ which implies that $w \notin [a]$, a contradiction.

Hence $w = a^k$ for some $k \geq 1$. □

Let M be a monoid and S be the underlying semigroup. A subsemigroup T of M is *adherent* in M if for all $x, y \in S$, $xTy \cap T \neq \emptyset$ implies $x, y \in T^1$. An idempotent $g \in M$ is adherent in M if the semigroup $\{g\}$ is adherent in M.

Proposition 15 *For all $a \in \Sigma$ such that $\theta(a) \neq a$, $g = [a]$ is adherent in $\mathrm{Syn}(F)$.*

Proof If $[x]g[y] = g$ then we need to show that $[x],[y] \in \{1,g\}$ for some $[x],[y] \in \mathrm{Syn}(F)$. From Proposition 14 we have that $g = [a] = \{a^i | i \geq 1\}$ which implies that $[x],[y] \in \{1,g\}$.   □

If θ is a mapping of Σ^* into Σ^*, a congruence R is said to be θ-compatible if uRv implies $\theta(u)R\theta(v)$. If such is the case, then the mapping θ on Σ^* can be extended to a mapping of the quotient-monoid $S = \Sigma^*/R$ in the following way. Let U be the class of R containing the word u. Define $\theta(U)$ to be the class of R containing $\theta(u)$. This mapping is well defined, i.e., it does not depend on the choice of the representative u of the class U. Indeed if $u' \in U$, then, R being θ-compatible, we have $\theta(u)R\theta(u')$ and hence $\theta(u') \in \theta(U)$.

Proposition 16 *The syntactic congruence P_F is θ-compatible.*

Proof To show that P_F is θ-compatible, we have to show that uP_Fv implies $\theta(u)P_F\theta(v)$, i.e., $C_F(u) = C_F(v)$ implies $C_F(\theta(u)) = C_F(\theta(v))$. Since θ is a bijective mapping and Σ is finite, the set $\Theta = \{\theta, \theta^2, \ldots, \theta^{k-1}, \theta^k = I\}$ generated by the powers of θ is a finite group, in particular a finite group of permutations of the set Σ. Since $\theta.\theta^{k-1} = \theta^k = I$, θ^{k-1} is the inverse of θ.

Let $(x,y) \in C_F(\theta(u))$ then $x\theta(u)y \in F$. If θ is a morphism, then since F is θ-stable, we have $\theta^n(x\theta(u)y) \in F$ and $\theta^n(x)\theta^{n+1}(u)\theta^n(y) \in F$ for $n \geq 1$. In particular, for $n = k-1$, since $\theta^k = I$, then $\theta^{k-1}(x)\theta^k(u)\theta^{k-1}(y) = \theta^{k-1}(x)u\theta^{k-1}(y) \in F$. Since $C_F(u) = C_F(v)$ we have $\theta^{k-1}(x)v\theta^{k-1}(y) \in F$. Hence $\theta(\theta^{k-1}(x)v\theta^{k-1}(y)) \in \theta(F) \subseteq F$. Since θ is a morphism, $\theta^k(x)\theta(v)\theta^k(y) = x\theta(v)y \in F$. Therefore $(x,y) \in C_F(\theta(v))$. Similarly, we have the inverse inclusion and hence $C_F(\theta(u)) = C_F(\theta(v))$ and $\theta(u)P_F\theta(v)$. The case when θ is an antimorphic involution can be proved by symmetry.   □

In the following results, using the notion of the syntactic monoid, we establish an algebraic connection between the language F of the hairpin-free words relatively to a bijective d-morphism θ over a finite alphabet Σ and a certain class of finite monoids.

Proposition 17 *Let $\mathrm{Syn}(F)$ be the syntactic monoid of F. Then:*

1. *$\mathrm{Syn}(F)$ is a finite commutative monoid with a disjunctive zero 0 and every element of $\mathrm{Syn}(F)$ is idempotent.*
2. *There exists a bijective d-morphism ψ such that the set $\mathrm{Syn}(F) \setminus \{0\}$ is stable under ψ.*
3. *If $[x][y] = 0$ for $[x],[y] \in \mathrm{Syn}(F) \setminus \{0\}$ then there exists $[p],[q] \in \mathrm{Syn}(F)$ such that $[x][y] = [p][q][\theta(q)]$.*

Proof 1. The regularity of the language F implies the finiteness of its syntactic monoid $\mathrm{Syn}(F)$. Since F is commutative, then $\mathrm{Syn}(F)$ is a commutative monoid. The complement $\bar{F}$ of F is an ideal of Σ^* and $\bar{F}$ is a class of P_F. Therefore the class $\bar{F} = 0$ is a zero element of $\mathrm{Syn}(F)$. Since $P_F = P_{\bar{F}}$, then, in $\mathrm{Syn}(F)$, the congruence $P_0 = P_{\bar{F}}$ is the equality and hence 0 is a disjunctive element of the syntactic monoid.

The last part follows from Corollary 2.

2. Since the syntactic congruence P_F is θ-compatible, a d-morphism ψ can be defined on $\mathrm{Syn}(F)$ in the following way. Let U be an element of $\mathrm{Syn}(F)$, i.e., U is a

class of P_F, and define $\psi(U)$ to be the class containing the element $\theta(u)$ where $u \in U$. This mapping is well defined because it does not depend on the choice of the representative v of the class U by virtue of the θ-compatibility of P_F. Indeed, since uP_Fv, then $\theta(u)P_F\theta(v)$ and hence $\theta(v) \in \psi(U)$. Therefore if V is the class of P_F containing v, then $\psi(U) = \psi(V)$.

It is immediate that ψ is a d-morphism, i.e., if θ is a morphism then ψ is a morphism and if θ is an antimorphism then ψ is an antimorphism.

To show that ψ is bijective, we only have to show that ψ is injective. Suppose that $\psi(U) = \psi(V)$. If $u \in U$, $v \in V$, then, from the definition of ψ, we must have $\theta(u)P_F\theta(v)$. Hence u, v are in the same class and therefore $U = V$.

The last part follows from the fact that F is θ-stable.

3. Let $[x], [y] \in \mathrm{Syn}(F) \setminus \{0\}$ and $[x][y] = 0$. Then $[xy] = 0$ which implies that $xy = \alpha a v \theta(a) \beta$ for some $\alpha, v, \beta \in \Sigma^*$, $a \in \Sigma$ and $xy \in [xy]$. Then $[xy] = [\alpha][a][v][\theta(a)][\beta] = [\alpha][v][\beta][a][\theta(a)]$ since $\mathrm{Syn}(F)$ is commutative. Hence $[x][y] = [\alpha v \beta a][\theta(a)]$. Take $[p] = [\alpha v \beta]$ and $[q] = [a]$. Thus $[xy] = [p][q][\theta(q)]$.   □

Example 3 Let $\Delta = \{A, C, G, T\}$ be the DNA alphabet and θ be the Watson–Crick complement involution. If F_Δ is the set of $hp(\theta)$-free words in Δ^*, then the classes of $\mathrm{Syn}(F_\Delta)$ are the following: $\{1, [A], [C], [G], [T], [AC], [AG], [CT], [GT], 0\}$
   where

 – $[A], [C], [G], [T]$: classes containing the letters of the alphabet.
 – $[AC], [AG], [CT], [GT]$: classes containing the words AC, AG, CT, GT.
 – 0: class of all words containing at least two complementary letters such as $ACTCA$ or $AGTGC$.

Since F is a commutative language, then $[AC]$, $[AG]$, $[CT]$ and $[GT]$ are also the classes of respectively $[CA]$, $[GA]$, $[TC]$ and $[TG]$.

These classes are the elements of the syntactic monoid $\mathrm{Syn}(F_\Delta)$ and they are idempotent elements of it. Furthermore $\bar{F}_\Delta$ is a class which is the zero element.

The product in $\mathrm{Syn}(F_\Delta)$ is easily determined. For example:

 – $[A][C] = [AC]$.
 – $[A][GT] = [AGT] = 0$, because A and T are complementary elements and hence $AT \in \bar{F}_\Delta$.
 – $[GT][AG] = [GTAG] = 0$.

Since $A \mapsto T$ and $C \mapsto G$, the classes $[A], [C], [G], [T]$ are adherent in $\mathrm{Syn}(F_\Delta)$.

Next proposition is a converse of the Proposition 17.

Proposition 18 *Let M be a monoid, with identity e satisfying the following six properties*:

1. *M is finite*
2. *M is commutative*
3. *Every element of M is an idempotent element*
4. *M has a disjunctive zero element 0*
5. *There exists a bijective d-morphism ψ such that the set $M - \{0\}$ is stable under ψ and $x\psi(x) = 0$ for all $x \in M$, $x \neq e$.*
6. *For all $x, y \in M$, if $xy = 0$ then there exists $p, q \in M$ such that $xy = pq\psi(q)$.*

Then there exists a free monoid Σ^* over a finite alphabet Σ, a bijective d-morphism θ and a language V in Σ^* such that

(i)   *V is the set of all hairpin-free words under θ*
(ii)   *The syntactic monoid $\mathrm{Syn}(V) = \Sigma^*/P_V$ is isomorphic to M.*

Proof If $M = \{x_1, x_2, \ldots, x_n\}$, then take the elements of M as the letters of an alphabet $\Sigma = \{x_1, x_2, \ldots, x_n\}$ and let Σ^* be the free monoid generated by Σ.

Let ϕ be the mapping of Σ^* onto M defined in the following way. If $u \in \Sigma$, then $\phi(u) = \psi(u) \in M$. If $u = u_1 u_2 \cdots u_k \in \Sigma^+$ with $u_i \in \Sigma$, then $\phi(u) = \psi(u_1) \ldots \psi(u_k)$. If $u = \lambda$, then $\phi(u) = e$, the identity of M. It is clear that ϕ is a homomorphism or anti-homomorphism of Σ^* onto M. The relation ρ defined as $u\rho v$, $u, v \in \Sigma^*$ iff $\phi(u) = \phi(v)$ is a congruence of Σ^* and the quotient monoid Σ^*/ρ is isomorphic to M.

Let $T = \{x | x \in \Sigma^*, \phi(x) = 0\}$. T is an ideal of Σ^*. Let P_T be the syntactic congruence of T.

We now show that $\rho = P_T$. It is immediate that $\rho \subseteq P_T$.

Let $x \equiv y\ (P_T)$, $x, y \in \Sigma^*$. If x is not equivalent to y modulo ρ, then $\phi(x) \neq \phi(y)$. Since 0 is disjunctive, then the syntactic congruence P_0 on M is the equality and we have $C_0(\phi(x)) \neq C_0(\phi(y))$. This implies the existence of $a, b \in M$ such that $a\phi(x)b = 0$ and $a\phi(y)b \neq 0$ or $a\phi(x)b \neq 0$ and $a\phi(y)b = 0$. Suppose that we have the first case. Since ϕ is bijective, then there are $r, s \in \Sigma^*$ such that $a = \phi(r)$, $b = \phi(s)$.

If θ is a homomorphism, then $\phi(rxs) = 0$ and $\phi(rys) \neq 0$. Hence $rxs \in T$ and $rys \notin T$, a contradiction because $C_T(x) = C_T(y)$. It follows then that $P_T \subseteq \rho$. If θ is an anti-morphism, we get the same conclusion.

Hence $\rho = P_T$.

Let $V = \Sigma^* \setminus T$. Since V is the complement of T, we have $P_T = P_V$. Therefore the syntactic monoid $\mathrm{Syn}(V) = P_V$ of V is isomorphic to M.

Since ψ is bijective and M is finite, then ψ is in particular a permutation of the elements of M. Since we have identified the elements of M with the alphabet Σ, we define the requested d-morphism θ of Σ^* by taking the corresponding permutation of the alphabet Σ and extending it to Σ^* in the usual way. If $u \in \Sigma^+$, $u = xy \cdots z$, $x, y, \ldots, z \in \Sigma$, then:

$\theta(u) = \theta(x)\theta(y) \cdots \theta(z)$ if ψ is a morphism
$\theta(u) = \theta(z) \cdots \theta(y)\theta(x)$ if ψ is an anti-morphism .
$\theta(\lambda) = \lambda$

It is immediate that θ is a bijective d-morphism.

Let us show now that conditions (i) and (ii) are satisfied. For (i), let $u \in V$ and suppose that u is not hairpin-free. Then $u = xwy\theta(w)z$ with $|w| = 1$. Since M is commutative and P_V is isomorphic to M, then every class and every union of classes of P_V are commutative. In particular, V being a union of classes of P_V, is a commutative language. From $u = xwy\theta(w)z \in V$ follows $w\theta(w)xyz \in V$. Since V is the complement of the ideal T, then we have $w\theta(w) \in V$. From Property 5 of the Proposition 18 it follows that $w\theta(w) \in T$. Since $V \cap T = \emptyset$, we have a contradiction. Hence V is hairpin-free. Now we show that V contains the set of all hairpin-free words. Suppose there exists $u = a_1 a_2 \cdots a_n$ for $a_i \in \Sigma$ such that u is not a hairpin and $u \in T$. Then $\phi(u) = 0$ which implies that $\phi(u) = \phi(a_1 a_2 \cdots a_n) = \psi(a_1 a_2 \cdots a_n) = 0$. Thus $\psi(a_1 a_2 \cdots a_n) = [b_1][b_2] \cdots [b_n] = 0$ and there exists $[x], [y] \in M \setminus \{0\}$ such that $[x][y] = [b_1][b_2] \cdots [b_n] = 0$ and hence by Property 6 there are $[p], [q] \in M$ such that $\psi(u) = [x][y] = [p][q][\theta(q)]$ which implies that u is a hairpin, a contradiction.

Hence $u \notin T$. Thus V contains all hairpin-free words. Condition (ii) follows by the construction.   □

Example 4 Let $M = \{1, 0, [a], [b]\}$ with $[a][b] = 0 = [b][a]$ and the involution ψ defined by $\psi([a]) = [b]$, $\psi([b]) = [a]$. Clearly in this example, the conditions of Proposition 18 are satisfied.

Let $\Sigma = \{a, b\}$. The bijective morphism θ is defined on Σ^* as $\theta([a]) = [b]$, $\theta([b]) = [a]$ and it is also an involution. Note that $T = \{x \in \Sigma^* | \theta(x) = 0\}$, i.e., $T = \{x \in \Sigma^* | x \in \Sigma^* a \Sigma^* b \Sigma^*\}$. The set $V = \Sigma^* \setminus T = \{a^*, b^*\}$ which is the set of all hairpin-free words under θ and is the set of words not containing two different letters of Σ.

We end this paper by discussing the Green's relations for the set of all hairpin-free words. We recall here the definition of Green's relations and some well known facts about some of the relations. For extensive treatments of Green's relations and the related varieties of finite monoids, we refer the reader to [4,12,19].

Definition 5 (Green's relations) Let S be a semigroup. We define on S four equivalence relations $\mathcal{R}$, $\mathcal{L}$, $\mathcal{H}$ and $\mathcal{J}$ called Green's relations:

1. $a \mathcal{R} b \Leftrightarrow aS^1 = bS^1$
2. $a \mathcal{L} b \Leftrightarrow S^1 a = S^1 b$
3. $a \mathcal{J} b \Leftrightarrow S^1 a S^1 = S^1 b S^1$
4. $a \mathcal{H} b \Leftrightarrow a \mathcal{R} b$ and $a \mathcal{L} b$

Note that the relations $\mathcal{R}$ and $\mathcal{L}$ commute, i.e., $\mathcal{RL} = \mathcal{LR}$ and $\mathcal{D} = \mathcal{RL}$. In a finite semigroup $\mathcal{D} = \mathcal{J}$. A semigroup S is $\mathcal{K}$-trivial iff $e \mathcal{K} f$ implies $e = f$ for $\mathcal{K} \in \{\mathcal{D}, \mathcal{R}, \mathcal{L}, \mathcal{J}, \mathcal{H}\}$.

Note that in a commutative semigroup, $\mathcal{H} = \mathcal{R} = \mathcal{L} = \mathcal{D} = \mathcal{J}$. Also no $\mathcal{H}$ class in a semigroup contains more than one idempotent. Since every element of $\mathrm{Syn}(F)$ is an idempotent, $\mathrm{Syn}F \setminus \{1\}$ is $\mathcal{H}$-trivial and since $\mathrm{Syn}(F)$ is commutative, $\mathrm{Syn}(F) \setminus \{1\}$ is $\mathcal{K}$-trivial for all $\mathcal{K} \in \{\mathcal{R}, \mathcal{L}, \mathcal{J}, \mathcal{D}, \mathcal{H}\}$.

Recall that a language L is said to be *n*-piecewise testable if whenever u and v have the same subwords of length atmost n and $u \in L$, then $v \in L$. The language L is piecewise testable if it is *n*-piecewise testable for some n. Star-free languages can be described by star-free expressions, i.e., using the letters of the alphabets, the constants $\emptyset$ and λ (empty word), the boolean operations and concatenation (but no star). It was proved in [21] that a language L is piecewise testable iff syntactic monoid of L is $\mathcal{J}$-trivial. Also note that piecewise testable languages are star-free. Hence F is piecewise testable and star-free.

References

1. Berstel, J., Perrin, D.: Theory of Codes. Academic, Orlando (1985)
2. Bonizzoni, P., Defelice, C., Mauri, G., Zizza, R.: Linear splicing and syntatctic monoid. Discrete Appl. Math. **15**(3), 452–470 (2006)
3. Haines, H.: On free monoids partially ordered by embedding. J. Comb. Theory **6**, 94–98 (1969)
4. Howie, J.M.: Fundamentals of Semigroup Theory. Oxford Science Publications, New York (1995)
5. Ito, M., Júrgensen, H., Shyr, H.J., Thierrin, G.: Outfix and infix codes and related classes of languages. J. Comput. Syst. Sci. **43**, 484–508 (1991)

6. Jonoska, N., Mahalingam, K., Chen, J.: Involution codes: with application to DNA coded languages. Nat. Comput. **4**(2), 141–162 (2005)
7. Jürgensen, H.: Syntactic monoid of codes. Acta Cybern. **14**, 117–133 (1999)
8. Jürgensen, H., Konstantinidis, S.: Codes. In: Rozenberg, G., Salomaa A. (eds.) Handbook of Formal Languages, vol. 1, chap. 8, pp. 511–608 (1997)
9. Jürgensen, H., Thierrin, G.: Infix codes. In: Proceedings of the 4th Hungarian Computer Science Conference, pp. 25–29 (1985)
10. Kari, L., Konstantinidis, S., Losseva, E., Wozniak, G.: Sticky-free and overhang-free DNA languages. Acta Inf. **40**, 119–157 (2003)
11. Kari, L., Konstantinidis, S., Losseva, E., Sosik, P., Thierrin, G.: Hairpin structures in DNA words. In: Proceedings of the 11th International Meeting on DNA Computing, LNCS 3892, pp. 158–170 (2005)
12. Lallement, G.: Semigroups and Combinatorial Dynamics. Wiley/Interscience, New York (1995)
13. Luca, A.D., Restivo, A.: A characterization of strictly locally testable languages and its application to subsemigroups of a free semigroup. Inf. Control **44**, 300–319 (1980)
14. Moriya, T.: Closure under composition and syntactic monoids of some codes. Report 00-04, Department of Mathematics and Computer Science, Martin-Luther-Universität Halle-Wittenberg (1999)
15. Paun, G., Rozenberg, G., Yokomori, T.: Hairpin languages. Int. J. Found. Comput. Sci. **12**(6), 849–857 (2001)
16. Petrich, M., Reis, C.M.: The syntactic monoid of the semigroup generated by a comma-free code. Proc. R. Soc. Edinb. **125**, 165–179 (1995)
17. Petrich, M., Thierrin, G.: The syntactic monoid of an infix code. Proc. Am. Math. Soc. **109**(4), 865–873 (1990)
18. Petrich, M., Reis, C.M., Thierrin, G.: The syntactic monoid of the semigroup generated by a maximal prefix code. Proc. Am. Math. Soc. **124**(3), 655–663 (1996)
19. Pin, J.E.: Varieties of Formal Languages. Plenum Press, New York (1986)
20. Shyr, H.J.: Free Monoids and Languages. Hon Min Book Company, Taichung (2001)
21. Simon, I.: Piecewise testable events. In: Automata Theory and Formal Languages, Proceedings of the 2nd GI Conference, LNCS 33, pp. 214–222. Springer, Berlin (1975)
22. Thierrin, G.: Convex languages. In: Proceedings of IRIA Symposium, pp. 481–492. North Holland, Amsterdam (1972)
23. Thierrin, G.: The syntactic monoid of a hypercode. Semigroup Forum **6**, 227–231 (1973)
24. Yu, S.S.: Languages and codes. Lecture Notes, Department of Computer Science, National Chung-Hsing University, Taichung, Taiwan (2005)