

# Aperiodic languages and generalizations\*

Lila Kari and Gabriel Thierrin  
Department of Mathematics  
University of Western Ontario  
London, Ontario, N6A 5B7  
Canada

June 18, 2010

## Abstract

For every integer  $k \geq 0$ , a class of languages, called  $k$ -aperiodic languages, is defined, generalizing the class of aperiodic languages. Properties of these classes are investigated, in particular the operations under which they are preserved. A characterization of the syntactic monoid of a  $k$ -aperiodic language is also obtained.

## 1 Introduction

Let  $X$  be a finite alphabet and let  $X^*$  be the free monoid generated by  $X$ . Any subset of  $X^*$  is called a language.

A language  $L$  over  $X$  is called *aperiodic* or *noncounting* ([1]) if there exists an integer  $n \geq 0$  such that for all  $x, y, z \in X^*$ ,  $xy^n z \in L$  if and only if  $xy^{n+1}z \in L$ .

The language  $L$  is called *left aperiodic* or *left-noncounting* ([4]) if there exists an integer  $n \geq 0$  such that for all  $y, z \in X^*$ ,  $y^n z \in L$  if and only if  $y^{n+1}z \in L$ .

The integer  $n$ , that is dependent on  $L$ , is called the *order* of the aperiodic or left-aperiodic language  $L$ . From the definition we see that every aperiodic language over  $X$  is a left-aperiodic language over  $X$  (see [6]).

Aperiodic and left-aperiodic languages differ by the fact that in the first case there is no restriction on the prefix  $x$  and that in the second case  $x$  is reduced to the empty word. It is therefore natural to consider cases that are situated between these two extremes. This is done in the following by putting conditions on the length of the prefix  $x$ .

---

\*This research was supported by Grant OGP0007877 of the Natural Sciences and Engineering Research Council of Canada

Let  $k$  be a non-negative integer. A language  $L$  is called *k-aperiodic* if there exists an integer  $n \geq 0$  such that for all  $x, y, z \in X^*$  with  $|x| \leq k$ ,  $xy^n z \in L$ , if and only if  $xy^{n+1}z \in L$ .

Section 2 contains definitions and some preliminary results. It is shown for example that there exists an infinite hierarchy of *k-aperiodic* languages. A family of right congruences is defined and used to characterize *k-aperiodic* languages as unions of classes of these right congruences.

Section 3 deals with closure properties of aperiodic and *k-aperiodic* languages under several operations such as catenation, boolean operations, morphism, insertion and deletion.

The notion of *right syntactic congruence* of a *k-aperiodic* language is defined and investigated in Section 4. Based on this notion, the syntactic monoid of a *k-aperiodic* language is defined, and a complete characterization of monoids which are syntactic monoids of *k-aperiodic* languages is obtained.

## 2 Definitions and preliminary results

Throughout this paper  $X$  will denote a finite alphabet with  $\text{card}(X) \geq 2$ , and  $X^*$  will be the free monoid generated by  $X$ . The elements of  $X^*$  will be called words and any subset of  $X^*$  will be called a language. For a word  $u \in X^*$ ,  $|u|$  will denote the length of  $u$ . The empty word will be denoted by 1.

**Definition 2.1** *A language  $L$  is called  $k$ -aperiodic if there exists an integer  $n \geq 0$  such that for all  $x, y, z \in X^*$  with  $|x| \leq k$ ,  $xy^n z \in L$  if and only if  $xy^{n+1}z \in L$ .*

Left-aperiodic languages are 0-aperiodic languages and aperiodic languages are *k-aperiodic* languages for all  $k \geq 0$ .

The number  $n$  is called the *order* of the *k-aperiodic* language  $L$ . Since all the numbers  $n' \geq n$  can also be considered as the order of  $L$ , the order of  $L$  is in fact a subset of natural numbers.

Let  $AP$  and  $AP_k$  be respectively the family of the aperiodic and the family of *k-aperiodic* languages. Clearly

$$AP \subseteq \dots \subseteq AP_{k+1} \subseteq AP_k \subseteq \dots \subseteq AP_1 \subseteq AP_0$$

**Proposition 2.1** *The preceding hierarchy is strict, that is, we have:*

$$AP \subset \dots \subset AP_{k+1} \subset AP_k \subset \dots \subset AP_1 \subset AP_0$$

*Proof.* Let  $L_k = \{a^{k+1}b^{2n} \mid n \geq 1\}$ . Then  $L_k \in AP_k$  and  $L_k \notin AP_{k+1}$ .

Indeed, we will show that if  $m > k + 1$ , then  $L_k$  is *k-aperiodic* of order  $m$ .

If  $xy^m z$  is a word in  $L_k$ , with  $|x| \leq k$ , then necessarily  $x = a^i$ ,  $i \leq k$ ,  $y = a^j b^p$ ,  $j \geq 1$ . As no word of the form  $a^i (a^j b^p)^m b^{2n-pm}$ ,  $m > k + 1$ , belongs to  $L_k$ , it follows that  $L_k$  is *k-aperiodic*.

Suppose now that  $L_k$  is  $(k + 1)$ -aperiodic of order  $m \geq 0$ . This implies, in particular, that if the word  $xy^mz \in L_k$ , with  $x = a^{k+1}$ ,  $y = b$ ,  $z = b^{2n-m}$  then also the word  $xy^{m+1}z = a^{k+1}b^{2n+1}$  belongs to  $L_k$ - a contradiction. Consequently,  $L_k$  is not  $(k + 1)$ -aperiodic.  $\square$

The  $k$ -aperiodic languages can also be defined by using binary relations.

For words  $u, v \in X^*$ , consider the following three conditions, where  $k$  is a fixed non negative integer:

- (i)  $u = v$ ;
- (ii)  $u = xy^n z$  and  $v = xy^{n+1} z$  for some  $x, z \in X^*$ ,  $y \in X^+$ ,  $|x| \leq k$ ;
- (iii)  $u = xy^{n+1} z$  and  $v = xy^n z$  for some  $x, z \in X^*$ ,  $y \in X^+$ ,  $|x| \leq k$ .

For  $u, v \in X^*$  define the relation  $u \leftrightarrow_{k,n} v$  if and only if one of the above three conditions hold and let  $\sim_{k,n}$  denote the transitive closure of  $u \leftrightarrow_{k,n} v$ . Clearly, for each  $k$  and  $n$ ,  $\sim_{k,n}$  is an equivalence relation that is right compatible and hence a right congruence. The relation  $\sim_n$  defined by  $u \sim_n v$  if and only if  $u \sim_{k,n} v$  for every  $k$  is a congruence relation (see [1]).

**Proposition 2.2** *A language  $L$  is  $k$ -aperiodic if and only if it is a union of right congruence classes of  $\sim_{k,n}$  for some  $n$ .*

*Proof.* Let  $L$  be a  $k$ -aperiodic language. In order to prove that  $L$  is a union of classes of  $\sim_{k,n}$  we have to show that if  $L$  contains a word  $u$ , it contains also all the words  $v$  belonging to the same class as  $u$ . This follows immediately as  $u \in L$  and  $u \sim_{k,n} v$  imply  $v \in L$ .

On the other hand, if  $L$  is a union of classes of  $\sim_{k,n}$  then, from the definition of  $\sim_{k,n}$  it easily follows that  $L$  is  $k$ -aperiodic.  $\square$

### 3 Operations with aperiodic languages

In this section we investigate closure properties of the family of aperiodic ( $k$ -aperiodic, left aperiodic) languages under basic operations such as boolean operations, catenation, right/left quotient, insertion and deletion.

The study reveals intriguing differences between these apparently similar language families. It turns out that the family of aperiodic languages is closed under all the above listed operations, while for example, the family of left or  $k$ -aperiodic languages is not closed under deletion and left quotient. On the other hand, as the conditions imposed on left and  $k$ -aperiodic languages affect only the prefix of the words, these families are closed under right quotient.

It is still an open question whether or not the insertion of two left aperiodic ( $k$ -aperiodic) languages remains left aperiodic ( $k$ -aperiodic).

**Proposition 3.1** *If  $A, B$  are  $k$ -aperiodic languages over  $X$ , then  $AB$  is  $k$ -aperiodic.*

*Proof.* We may assume that the orders of  $A$  and  $B$  are respectively  $n_1 \geq 1$  and  $n_2 \geq 1$ . Let  $m = n_1 + n_2 + 1$  and let  $xv^m u \in AB$  with  $|x| \leq k$ .

(1) If  $xv^m u = (xv^m u_1)(u_2)$ , where  $u_1, u_2 \in X^*$ ,  $xv^m u_1 \in A$ ,  $|x| \leq k$  and  $u_2 \in B$  then, since  $A$  is  $k$ -aperiodic of order  $n_1$  and hence also of order  $m$ , it follows that  $xv^{m+1} u_1 \in A$ , that is,  $xv^{m+1} u \in AB$ .

(2) If  $xv^m u = xv^{n_1+n_2+1} u = (xv^{n_1} w_1)(w_2 u)$  where  $w_1, w_2 \in X^*$ ,  $xv^{n_1} w_1 \in A$  and  $w_2 u \in B$ , then, since  $A$  is  $k$ -aperiodic and  $|x| \leq k$ ,  $xv^{n_1+1} w_1 \in A$ . It follows that  $xv^{m+1} u \in AB$ .

(3) If  $xv^m u = (xv^i v_1)(v_2 v^{m-i-1} u)$  where  $v_1, v_2 \in X^*$ ,  $v_1 v_2 = v$ ,  $i < n_1$ ,  $xv^i v_1 \in A$  and  $v_2 v^{m-i-1} u \in B$ , then  $m - i - 1 \geq n_2$  and

$$\begin{aligned} v_2 v^{m-i-1} u &= (v_2 v_1)(v_2 v_1) \cdots (v_2 v_1) v_2 u \\ &= (v_2 v_1)^{m-i-1} v_2 u \in B \end{aligned}$$

Since  $B$  is  $k$ -aperiodic, we have  $(v_2 v_1)^{m-i} v_2 u \in B$ . It follows that

$$xv^{m+1} u = (xv^i v_1)(v_2 v^{m-i} u) \in AB$$

(4) If  $xv^m u = (x_1)(x_2 v^m u)$  where  $x_1, x_2 \in X^*$ ,  $x = x_1 x_2$ ,  $x_1 \in A$ ,  $x_2 v^m u \in B$ . Since  $B$  is  $k$ -aperiodic and  $|x_2| \leq |x| \leq k$ , then  $x_2 v^{m+1} u \in B$  and  $xv^{m+1} u = (x_1)(x_2 v^{m+1} u) \in AB$ .  $\square$

**Proposition 3.2** *The family of  $k$ -aperiodic languages is closed under catenation and the boolean operations of union, intersection and complementation.*

*Proof.* The closure under catenation has been proved in the preceding proposition and it is easy to see the closure under union and intersection (and hence under complementation).  $\square$

Recall that a nonempty language  $L \subseteq X^+$  is called a *prefix (suffix) code* if  $L \cap LX^+ = \emptyset$  ( $L \cap X^+ L = \emptyset$ ). This means that for any  $x \in L$ ,  $xy \notin L$  ( $yx \notin L$ ) for all  $y \in L^+$ .

**Proposition 3.3** *Let  $X$  be an alphabet and let  $A, B \subseteq X^*$ . If  $AB$  is  $k$ -aperiodic of order  $n$  and  $B$  is a suffix code, then  $A$  is  $k$ -aperiodic of order  $n$ .*

*Proof.* If  $A$  is in the set  $\{\emptyset, \{1\}\}$ , then the proposition is trivially true. Hence we may assume  $A$  is not in  $\{\emptyset, \{1\}\}$ .

Assume  $AB$  is  $k$ -aperiodic of order  $n$ . Let  $xy^n z \in A$ ,  $|x| \leq k$ ,  $x, y, z \in X^*$ . Then  $xy^n z b \in AB$ , where  $b \in B$ . Since  $AB$  is  $k$ -aperiodic of order  $n$ , we have  $xy^{n+1} z b \in AB$ . Now, since  $B$  is a suffix code by assumption, we have  $xy^{n+1} z \in A$ . Similarly, we can show that  $xy^{n+1} z \in AB$ ,  $|x| \leq k$ , implies  $xy^n z \in A$ . Hence  $A$  is  $k$ -aperiodic of order  $n$ .  $\square$

**Corollary 3.1** *Let  $A \in X^*$ . If  $A$  is a suffix code then  $A^2$  is  $k$ -aperiodic if and only if  $A$  is  $k$ -aperiodic.*

*Proof.* It follows from the preceding Proposition and Proposition 3.1.  $\square$

As we will see in the following, aperiodicity is preserved even when we consider an operation which generalizes the catenation operation.

Let  $u, v$  be words over an alphabet  $X$ . The *insertion* of  $v$  into  $u$  is defined as (see [2], [3]):

$$u \leftarrow v = \{u_1vu_2 \mid u = u_1u_2, u_1, u_2 \in X^*\}.$$

The operation can be easily extended to languages,

$$L_1 \leftarrow L_2 = \bigcup_{u \in L_1, v \in L_2} (u \leftarrow v).$$

**Proposition 3.4** *The family of aperiodic languages is closed under insertion.*

*Proof.* Let  $L_1, L_2$  be aperiodic languages of order  $n_1$ , respectively  $n_2$ . Choose  $m = 2n_1 + n_2 + 1$  and let  $xy^mz \in L_1 \leftarrow L_2$ . This means that  $xy^mz = uvw$  where  $w \in L_1, w \in L_2$ .

One of the following cases can occur:

**A.**  $y^m$  is a subword of  $w$ .

Then  $w = \alpha y^m \beta \in L_2$  which implies  $\alpha y^{m+1} \beta \in L_2$ , that is,  $xy^{m+1}z \in L_1 \leftarrow L_2$ .

**B.**  $w$  is a subword of  $y^m$  and  $y^m$  (possibly) overlaps with  $u$  and  $v$ . Then at least one of the following must happen:

- the length of the overlap between  $y^m$  and  $w$  is bigger than  $|y| \times n_2$ .
- the length of the overlap between  $y^m$  and  $u$  is bigger than  $|y| \times n_1$ .
- the length of the overlap between  $y^m$  and  $v$  is bigger than  $|y| \times n_1$ .

Indeed, if none of these happens then we have

$$|y^m| = m \times |y| = \text{sum of overlaps with } u, w, v \leq |y| \times (2n_1 + n_2)$$

- a contradiction.

Now, assume that the first case happens. This means that  $w = \alpha y^{n_2} \beta$  which implies that  $\alpha y^{n_2+1} \beta \in L_2$ . This, in turn, implies that  $xy^{m+1}z \in L_1 \leftarrow L_2$ .

In the other cases, one can reason in a similar way.

**C.**  $y^m$  overlaps with  $w$  and  $v$ .

Then, one of the following cases must hold:

- the length of the overlap between  $y^m$  and  $w$  is bigger than  $|y| \times n_2$
- the length of the overlap between  $y^m$  and  $v$  is bigger than  $|y| \times n_1$ .

(Otherwise, we would have that

$$|y^m| = m \times |y| = \text{sum of overlaps with } w, v \leq |y| \times (n_2 + n_1),$$

- a contradiction.)

Assume that the first case holds. We can reason as in case B) to see that this leads to the desired conclusion.

**D.** The word  $y^m$  overlaps with  $w$  and  $u$ . It is similar to the preceding case.  $\square$

The family of aperiodic languages is closed under right/left quotient. This result will be obtained as a consequence of a more general one which states that, by deleting any language from an aperiodic one, the result is still aperiodic. Before that, we need to introduce the definition of *deletion*, which is an operation generalizing the right/left quotient of languages.

Let  $u, v$  be two words over the alphabet  $X$ . The *deletion* of  $v$  from  $u$  is defined as (see [2], [3]):

$$u \rightarrow v = \{w \in X^* \mid w = u_1u_2, u = u_1vu_2\}.$$

The operation can be extended to languages in the natural fashion.

**Proposition 3.5** *If  $L_1$  is an aperiodic language and  $L_2$  is an arbitrary language, then the deletion of  $L_2$  from  $L_1$ ,  $L_1 \rightarrow L_2$ , is an aperiodic language.*

*Proof.* Let  $L_1$  be an aperiodic language of order  $n$  and let  $m = 2n + 2$ .

Let  $xy^mz$  be a word in  $L_1 \rightarrow L_2$ . Then,  $xy^mz = uv$ , where there exists  $w \in L_2$  such that  $uwv \in L_1$ .

One of the following case holds:

**A.**  $x_1wx_2y^m \in L_1$  ( $w$  has been erased from the  $x$ -part). Then, as  $L_1$  is aperiodic,  $x_1wx_2y^{m+1}z \in L_1$  which implies  $x_1x_2y^{m+1}z \in L_1 \rightarrow L_2$ .

**B.**  $xy^mz_1wz_2 \in L_1$  – similar to A.

**C.**  $xy^i y_1 w y_2 y^j z \in L_1$ , where  $i + j + 1 = m$ ,  $y = y_1 y_2$ .

At least one of the following cases must hold:

–  $i$  is greater than  $n$

–  $j$  is greater than  $n$

(If none of these happens, then  $m = i + j + 1 \leq n + n + 1 = 2n + 1$  – a contradiction.)

Assume, for example, that  $i > n$ . Then,  $xy^{i+1}y_1wy_2y^jz \in L_1$ , which implies  $xy^{i+1}y_1y_2y^jz \in L_1 \rightarrow L_2$ , that is,  $xy^{m+1}z \in L_1 \rightarrow L_2$ .

□

**Corollary 3.2** *The family of aperiodic languages is closed under right/left quotient.*

*Proof.* It follows from the preceding proposition and by noticing that, if  $\#$  is a letter which does not belong to  $X$ , the right/left quotient can be obtained as particular cases of deletion. Indeed, we have:

$$L_1/L_2 = L_1\# \rightarrow L_2\# \text{ and } L_2 \setminus L_1 = \#L_1 \rightarrow \#L_2,$$

for all  $L_1, L_2 \subseteq X^*$ .

□

The following examples show that the families of left aperiodic and  $k$ -aperiodic languages are closed under neither deletion, nor left quotient.

**Example 1** Let  $L_1 = \{ab^{2^n} \mid n \geq 1\}$  and  $L_2 = \{a\}$ . Both languages are left aperiodic but the deletion of  $L_2$  from  $L_1$  (which in this case coincides with the left quotient  $L_2 \setminus L_1$ ),

$$L_1 \rightarrow L_2 = \{b^{2^n} \mid n \geq 1\},$$

is not a left aperiodic language.

**Example 2** Consider the  $k$ -aperiodic languages

$$L_1 = \{a^{k+1}a^m b^n c^n \mid n, m \geq 1\} \text{ and } L_2 = da^*b,$$

where  $k > 0$ . The deletion of  $L_2$  from  $L_1$  is

$$L_1 \rightarrow L_2 = \{d^k b^{n-1} c^n \mid n \geq 1\},$$

which is not a  $k$ -aperiodic language.

**Example 3** Let  $k \geq 0$  and consider the languages

$$L_1 = \{a^{k+1}b^{2^n} \mid n \geq 1\}, \quad L_2 = \{a\}.$$

According to Proposition 2.1 the language  $L_1$  is  $k$ -aperiodic and  $L_2$  is  $k$ -aperiodic being finite. However, the left quotient  $L_2 \setminus L_1 = \{a^k b^{2^n} \mid n \geq 1\}$  is not  $k$ -aperiodic.

The situation changes if we consider right quotient instead of left quotient. As the conditions for left aperiodic languages and  $k$ -aperiodic languages affect only the prefix of the words, these families are still closed under right quotient. In fact, a more general result holds.

**Proposition 3.6** *If  $L_1$  is a  $k$ -aperiodic language,  $k \geq 0$ , and  $L_2$  is an arbitrary language, then the right quotient  $L_1/L_2$  is a  $k$ -aperiodic language.*

*Proof.* Let  $L_1$  be a  $k$ -aperiodic language of order  $m$ . We shall show that  $L_1/L_2$  is  $k$ -aperiodic of the same order.

Indeed, let  $xy^m z \in L_1/L_2$ ,  $|x| \leq k$ . This implies there exists  $v \in L_2$  such that  $xy^m z v$  is in  $L_1$ . From the  $k$ -aperiodicity of  $L_1$  we deduce that  $xy^{m+1} z v \in L_1$ , which further implies  $xy^{m+1} z \in L_1/L_2$ . The other implication can be proved analogously.  $\square$

We conclude this section with the remark that the family of  $k$ -aperiodic languages is not closed under morphisms. For example, consider the language  $L = \{a^{k+1}b^{2^n} \mid n \geq 1\}$  and the morphism  $h$  defined as  $h(a) = 1$  and  $h(b) = b$ .  $L$  is  $k$ -aperiodic but  $h(L)$  is not.

## 4 Syntactic monoids of $k$ -aperiodic languages

This section is devoted to the study of connections between syntactic congruences and  $k$ -aperiodic languages. Moreover, we introduce and investigate the notion of right syntactic congruence of a  $k$ -aperiodic language. Based on this notion, a complete characterization of monoids which are syntactic monoids of  $k$ -aperiodic languages is obtained.

Let  $M$  be a monoid and let  $L \subseteq M$ . If  $u \in M$ , then:

$$u^{-1}L = \{x \in M \mid ux \in L\}.$$

The relation  $R_L$  defined by

$$u \equiv v (R_L) \Leftrightarrow u^{-1}L = v^{-1}L$$

is a right congruence called the *right principal congruence* associated to the subset  $L$ .

The right principal congruence can also be defined by:

$$u \equiv v (R_L) \Leftrightarrow \forall x, y \in M, (ux \in L \text{ iff } vx \in L).$$

The relation  $P_L$  defined by:

$$u \equiv v (P_L) \Leftrightarrow \forall x, y \in M, (xuy \in L \text{ iff } xvy \in L)$$

is a congruence called the *principal congruence* of the language  $L$ . If  $P_L$  is the identity relation, then  $L$  is called a *disjunctive subset*.

If the monoid  $M$  is the free monoid  $X^*$  and  $L$  is a language over  $X$ , then  $R_L$  and  $P_L$  are called respectively the *right syntactic congruence* and the *syntactic congruence* of the language  $L$ . The quotient monoid  $\text{syn}(L) = X^*/P_L$  is called the *syntactic monoid* of the language  $L$ .

Recall that a language  $L$  is regular iff the index of  $R_L$  is finite.

**Proposition 4.1** *A language  $L$  over  $X$  is  $k$ -aperiodic if and only if there exists  $n > 0$  such that*

$$xu^n \equiv xu^{n+1} (R_L) \quad \forall u, x \in X^*, |x| \leq k$$

*Proof.* This follows from the fact that, if  $|x| \leq k$ , then  $z \in (xu^n)^{-1}L$  if and only if  $xu^n z \in L$ , hence if and only if  $xu^{n+1} z \in L$ .  $\square$

**Proposition 4.2** *If a language  $L$  over  $X$  is  $k$ -aperiodic, then every class of  $R_L$  is a  $k$ -aperiodic language.*

*Proof.* Let  $A$  be a class modulo  $R_L$ . By Proposition 4.1, there exists  $n > 0$  such that  $xu^n \equiv xu^{n+1} (R_L)$  for all  $u, x \in X^*$  with  $|x| \leq k$ . Since  $R_L$  is a right congruence, this implies  $xu^n y \equiv xu^{n+1} y (R_L)$ . It follows then that  $xu^n y \in A$  if and only if  $xu^{n+1} y \in A$ . Hence  $A$  is  $k$ -aperiodic.  $\square$



**Proposition 4.3** *Let  $L$  be a regular language over  $X$ . Then  $L$  is  $k$ -aperiodic if and only if every class of  $R_L$  is a  $k$ -aperiodic language.*

*Proof.* If  $L$  is  $k$ -aperiodic, the aperiodicity of every class of  $R_L$  follows from Proposition 4.1.

Suppose now that every class of  $R_L$  is  $k$ -aperiodic. The language  $L$  is a union of classes  $A_i$  of  $R_L$ , and since  $L$  is regular, the number of these classes is finite, say  $m$ . Consequently, we can write  $L = \bigcup_{i=1}^m A_i$ . Since the union of a finite number of  $k$ -aperiodic languages is  $k$ -aperiodic, it follows that  $L$  itself is  $k$ -aperiodic.  $\square$

A monoid  $M$  is called an *aperiodic* or *combinatorial* monoid if there exists a positive integer  $n$  such that  $u^n = u^{n+1}$  for all  $u \in M$  ([5], [7], [8]). A finite monoid is aperiodic if and only if all its subgroups are trivial. A language  $L$  over  $X$  is aperiodic if and only if its syntactic monoid is aperiodic.

A right congruence  $R$  of  $X^*$  is called a  *$k$ -aperiodic congruence* iff there exists a positive integer  $n$  such that  $xu^n \equiv xu^{n+1} (R)$  for all  $u, x \in X^*$  with  $|x| \leq k$ .

**Lemma 4.1** *Let  $L$  be a  $k$ -aperiodic language over  $X$ . Then*

- (i) *The syntactic right congruence  $R_L$  is a  $k$ -aperiodic right congruence.*
- (ii) *If  $T$  is a congruence of  $X^*$  such that  $P_L \subseteq T \subseteq R_L$ , then  $P_L = T$ .*

*Proof.* (i) This follows immediately from Proposition 4.1.

(ii) If  $u \equiv v (T)$ , then, since  $T$  is a congruence,  $xuy \equiv xvy (T)$  for all  $x, y \in X^*$ . Since  $L$  is a union of classes of  $R_L$  and  $T \subseteq R_L$ , it follows that  $L$  is also a union of classes of  $T$ . Hence  $xuy \in L$  implies  $xvy \in L$  and vice versa. Therefore  $u \equiv v (P_L)$  which means  $T \subseteq P_L$ . Consequently, we have  $P_L = T$ .  $\square$

If  $M$  is a monoid, a right congruence  $R$  over  $M$  is called *strict* if, for every congruence  $R'$  of  $M$ , the relation  $R' \subseteq R$  implies that  $R'$  is the identity relation.

A monoid  $M$  is called a *strict  $(k, n)$ -aperiodic monoid* where  $k$  and  $n$  are positive integers if the following conditions are satisfied:

- (i)  $M$  is finitely generated;
- (ii)  $M$  contains a strict right congruence  $R$ ;
- (iii)  $M$  contains a finite set  $G$  of generators such that

$$xu^n \equiv xu^{n+1} (R)$$

for all  $u, x \in M$  with  $x = g_1 g_2 \cdots g_m$ ,  $g_i \in G$ ,  $m \leq k$ .

**Proposition 4.4** *Let  $L$  be a  $k$ -aperiodic language over the alphabet  $X$  and let  $\text{syn}(L)$  be its syntactic monoid. Then  $\text{syn}(L)$  is a strict  $(k, n)$ -aperiodic monoid for some positive integer  $n$ . Furthermore  $\text{syn}(L)$  contains a disjunctive subset  $D$  that is a union of classes of  $R$ .*

*Proof.* If  $R_L$  is the syntactic right congruence of  $L$ , then by Lemma 4.1,  $R_L$  is a  $k$ -aperiodic right congruence and  $P_L \subseteq R_L$ . Hence there exists a positive integer  $n$  such that  $xu^n \equiv xu^{n+1}(R_L)$  for every  $u, x \in X^*$  with  $|x| \leq k$ .

The right congruence  $R_L$  induces on  $\text{syn}(L) = X^*/P_L$  a right congruence  $R$  that is strict by Lemma 4.1, (ii). If  $X = \{x_1, x_2, \dots, x_r\}$  and if  $g_i = [x_i]$  is the class of  $x_i$  modulo  $P_L$ , let  $G = \{g_1, g_2, \dots, g_r\}$ . Clearly, if  $[x]$  denotes the class of  $x$  modulo  $P_L$ , we have

$$[x][u]^n \equiv [x][u]^{n+1} (R)$$

for every  $[u], [x] \in \text{syn}(L)$  such that  $[x] = g_{i_1}g_{i_2} \cdots g_{i_m}$ ,  $g_{i_j} \in G$  and  $m \leq k$ . Let  $D = \{[x] \mid x \in L\}$  be the subset consisting of all the classes  $[x]$  of  $P_L$  containing words of  $L$ . Since  $\text{syn}(L) = X^*/P_L$ ,  $D$  is a disjunctive subset of  $\text{syn}(L)$  and a union of classes of  $R$ .  $\square$

**Proposition 4.5** *Let  $M$  be a strict  $(k, n)$ -aperiodic monoid with the strict right congruence  $R$  and containing a disjunctive subset  $D$  that is a union of classes of  $R$ . Then the monoid  $M$  is isomorphic to the syntactic monoid of a  $k$ -aperiodic language over an alphabet  $X$ .*

*Proof.* The monoid  $M$ , being finitely generated, contains a finite set  $G$  of generators. Let  $X = G$ , let  $X^*$  be the free monoid generated by  $X$  and let  $\phi$  be the canonical homomorphism of  $X^*$  onto  $M$ . If  $L = \phi^{-1}(D)$ , then  $X^*/P_L$  is isomorphic to  $M$  because  $D$  is a disjunctive subset of  $M$ . Therefore  $M$  is isomorphic to the syntactic monoid  $\text{syn}(L)$  of the language  $L$ .

Since  $R$  is a strict right congruence of  $M$ ,  $xu^n \equiv xu^{n+1} (R)$  for all  $u \in M$ ,  $x = x_{i_1}x_{i_2} \cdots x_{i_m}$ ,  $x_{i_j} \in G = X$  and  $m \leq k$ . Since  $D$  is a union of classes of  $R$ , then  $R \subseteq R_D$  and  $xu^n \equiv xu^{n+1} (R_D)$  for all  $u \in M$  and  $x$  satisfying the above condition. Therefore  $yv^n \equiv yv^{n+1} (R_L)$  for all  $v \in X^*$  and  $y \in X^*$  with  $|y| \leq k$ . By Proposition 4.1, this implies that  $L$  is  $k$ -aperiodic.  $\square$

## References

- [1] J.A. Brzozowski, K. Čulik II, and A. Gabrielian. Classification of noncounting events. *J. Comp. Syst. Sci.* 5, 1971, 41-53.
- [2] L.Kari. Insertion and deletion of words: determinism and reversibility. *Proceedings of MFCS'93*, Lecture Notes in Computer Science, 711(1993), 315-327.
- [3] L.Kari. *On insertion and deletion in formal languages*, Ph.D. thesis, University of Turku, Finland, 1991.
- [4] H.J. Shyr and G. Thierrin. Left-noncounting languages. *Intern.J. Comp. & Information Sci.* 4, 1975, 95-102.

- [5] H.J. Shyr, *Free monoids and languages*, Lectures Notes, National Chung-Hsing University, Hon Min Book Company, Taichung, 1991.
- [6] R. McNaughton and S. Papert, *Counter-free automata*, MIT Press, 1971.
- [7] G. Lallement, *Semigroups and combinatorial applications*, Wiley, New York, 1979.
- [8] J.E. Pin, *Varieties of formal languages*, Foundations of Computer Science, Plenum Press, New York, 1986.