

Some generalizations of periodic words *

Arturo Carpi

Aldo de Luca

Résumé

Nous considérons deux classes de mots finis, nommées *semipériodiques* et *periodic-like*, qui étendent d'une façon naturelle la classe des mots périodiques finis, car ils gardent des propriétés importantes de cette classe. La classe des mots périodiques est strictement incluse dans la classe des mots semipériodiques et cette dernière est strictement incluse dans la classe des mots *periodic-like*. Nous faisons une comparaison entre ces classes et présentons des théorèmes de base dans le cadre d'une nouvelle approche de l'analyse combinatoire des mots, récemment introduite par les auteurs, qui repose sur les notions de facteur *spécial* et *étendable*. On montre quelques applications à la périodicité des mots. En particulier, nous donnons une formulation plus générale du théorème de Fine et Wilf qui prend en compte la 'structure' d'un mot. En outre, on présente un nouveau théorème concernant une décomposition convenable d'un mot en sous-mots *periodic-like*.

Abstract

We consider two classes of finite words, called *semiperiodic* and *periodic-like*, which extend in a natural way the class of finite periodic words since they share some important properties of this latter class. The class of periodic words is properly included in the class of semiperiodic words and this latter is properly included in the class of periodic-like words. A comparison between these classes is made and some basic theorems are presented in the frame of a new approach to the combinatorial analysis of words, recently introduced by the authors, based on the notions of *special* and *extendable* factors. Some applications to periodicity of words are shown. In particular, one can give a

*The work for this paper has been supported by the Italian Ministry of Education under Project COFIN '98 "Modelli di calcolo innovativi: Metodi sintattici e combinatori".

Key words and phrases : Periodic words, semiperiodic words, periodic-like words, special factors.

more general formulation of the theorem of Fine and Wilf which takes into account the ‘structure’ of a word. Moreover, a new periodicity theorem, based on a suitable decomposition of a word in periodic-like subwords, is presented.

1 Introduction

Periodic sequences of symbols over a finite set are of fundamental importance in many different fields such as Physics, Mathematics, Computer Science, and, more recently, Biology. A recent survey paper on periodic sequence is in [14, Chap. 8].

In Computer Science the symbols are usually called *letters*, the set of symbols *alphabet* and the sequences of letters *words*.

Infinite periodic words satisfy the following important property (*harmonic property*): *any period is a multiple of the minimal period of the word*. This property is no longer true in the case of a finite word. In fact, a finite word can have periods which are not multiple of the minimal period.

A finite word w is called *periodic* if its length satisfies the inequality

$$|w| \geq 2\pi_w,$$

where π_w denotes the minimal period of w . In such a case one has that for any period p of w either p is a multiple of π_w or $p > |w| - \pi_w + 1 > |w|/2$. Thus a finite periodic word maintains the harmonic property at least for periods which are sufficiently small with respect to the length of the word (at least when they are less than half of the length of the word).

The class of finite periodic words is, however, too ‘narrow’ when one analyzes periodicities in very long sequences such as DNA, RNA, or proteins. In such cases only rarely one encounters periodic words. In fact, the periods of these words are usually larger, or also much larger, than half of the length of the word. Hence, it is important to enlarge the class of finite periodic words by considering suitable generalizations of this concept which preserve some basic properties of periodic words.

In this paper we survey and compare two important extensions of the notion of finite periodic word which have been, recently, introduced by the authors [5, 6, 7]. These extensions have been called *semiperiodic* and *periodic-like* words. The class of periodic words is properly included in the class of semiperiodic words and this latter is properly included in the class of periodic-like words.

Some other extensions of the notion of finite periodic word, such as *quasi-periodic words* have been recently proposed (see [1, 12, 15] and references therein). However, we stress that, differently from our approach, these generalizations are based on a notion of period different from the classical one.

The notions of semiperiodic and periodic-like word are very natural in the framework of a new approach to the combinatorics of finite words which has been introduced by the authors [2, 3, 4, 9]. In this approach a basic role is played by the two notions of *extendable* and *special factor* of a word.

A factor u of a word w is called *right extendable* if there exists a letter a such that ua is a factor of w . In a similar way one can define *left extendable* factors of w . The shortest unrepeated prefix (resp. suffix) of w will be denoted by h_w (resp. k_w). Observe that h_w (resp. k_w) is the shortest factor of w which is not left (resp. right) extendable in w . In the following we shall set $H_w = |h_w|$ and $K_w = |k_w|$.

A factor u of a word w is called *right special* if there exist at least two distinct letters a and b such that ua and ub are factors of w . In a similar way one can define *left special* factors of w . We shall denote by R_w (resp. L_w) the minimal natural number such that there are no right (resp. left) special factors of w of length R_w (resp. L_w).

A factor is *bispecial* if it is a right and left special factor of w . A *proper box* of w is any factor of w of the form

$$asb$$

with a and b letters and s a bispecial factors of w . A proper box of w is called *maximal* if it is not a factor of another proper box.

For instance, consider the word $w = abccbabcab$. One has that $h_w = abcc$ and $k_w = cab$, so that $H_w = 4$ and $K_w = 3$. The right special factors are ϵ , b , c , bc , and abc ; the left special factors are ϵ , a , b , c , and ab . Thus one has that $R_w = 4$ and $L_w = 3$. The bispecial factors are ϵ , b , and c and the maximal proper boxes are abc , bca , bcc , cba , and ccb .

The prefix h_w and the suffix k_w of a word w are respectively called the *initial* and the *terminal box* of w . A basic theorem proved in [3] shows that *any word is uniquely determined by the initial box, the terminal box, and the set of maximal proper boxes*.

The parameters K_w , H_w , R_w , and L_w give much information on the structure of a word w . For instance, the maximal length of a repeated factor of a non-empty word w is equal to $\max\{R_w, K_w\} - 1$ and the minimal period of w is not smaller than $|w| - \min\{H_w, K_w\} + 1$. As we shall see these parameters are related to the periodic structure of a word (cf. Propositions 1 and 6). Moreover, as proved in [3], *a word w is uniquely determined by the set of its factors up to length $1 + \max\{R_w, K_w\}$* .

In the study of semiperiodic and periodic-like words, an essential role will be played by an equivalence relation called *root-conjugacy*. The *fractional root*, or simply *root*, of a word is the prefix of the word having a length equal to its minimal period. Two words are said to be *root-conjugate* if their roots are conjugate.

In Sect. 3 we introduce the class of semiperiodic words [5]. These words may be defined in several equivalent ways. In particular, a word is semiperiodic if its minimal period is less than or equal to $|w| - R_w$. Any periodic word is semiperiodic, whereas the converse is not generally true.

In the analysis of the structure of semiperiodic words we introduce the equivalence relation \approx defined as follows: let w be a word and set $n = 1 + \min\{L_w, R_w\}$. A word v is in the relation \approx with w , i.e., $v \approx w$, if the set of factors of v of length n is equal to the set of factors of w of length n .

A first result is that a word is semiperiodic if and only if in its equivalence class (mod \approx) there is a periodic word (cf. Proposition 3). A second result (cf. Theorem 1) is that the equivalence class (mod \approx) of a semiperiodic word w coincides with the intersection of its class of root-conjugacy and the class of semiperiodic words, i.e., for a semiperiodic word w , a word v is equivalent to w (mod \approx) if and only if v is semiperiodic and root-conjugate to w .

A further result (cf. Proposition 4) is that the intersection of the equivalence class (mod \approx) of a word w which is not semiperiodic and the set of all words having the same length as w is a singleton, i.e., it contains only w .

In Sect. 4 we introduce the class of periodic-like words [6, 7]. Similarly to semiperiodic words, periodic-like words can be defined in several equivalent ways. In particular, a word w is periodic-like if its minimal period is less than or equal to $|w| - R'_w$, where R'_w is the minimal natural number such that any prefix of w of length $\geq R'_w$ is not right special. Since $R'_w \leq R_w$ one has that a semiperiodic word is periodic-like, whereas the converse is not generally true.

In the analysis of the structure of periodic-like words we introduce the equivalence relation \equiv defined as follows: two words are equivalent (mod \equiv) if they have the same set of maximal proper boxes and at least one common letter.

One can prove that the equivalence relation \equiv is coarser than \approx , i.e., $\approx \subseteq \equiv$.

Similarly to the case of semiperiodic words one can prove (cf. Proposition 9) that a word is periodic-like if and only if in its equivalence class (mod \equiv) there is a periodic word. Moreover, the equivalence class (mod \equiv) of a periodic-like word w coincides with the intersection of its class of root-conjugacy and the class of periodic-like words, i.e., for a periodic-like word w , a word v is equivalent to w (mod \equiv) if and only if v is periodic-like and root-conjugate to w (cf. Theorem 2).

In Sect. 5 we show the importance of the notions of semiperiodic and periodic-like word in the study of periodicity of words. In particular, the following result holds: if a word w has two periods $p, q \leq |w| - R'_w$, then $\gcd(p, q)$ is also a period of w (cf. Proposition 13).

By the previous proposition, one can easily derive the theorem of Fine and Wilf [11] (cf. Theorem 3) which states that if a word w has two distinct periods p and q and length $|w| \geq p + q - \gcd(p, q)$, then also $d = \gcd(p, q)$ is a period of w . It is possible to verify that, under the hypotheses of the theorem of Fine and Wilf, w is periodic and, moreover, with the only exception of a trivial case, one has $p, q < |w| - R_w$ (cf. Proposition 14).

We consider a suitable canonical decomposition of a word in periodic-like subwords and give a new periodicity theorem, showing that the minimal period of a word is equal to the sum of the minimal periods of the periodic-like subwords occurring in the canonical decomposition.

2 Preliminaries

Let A be a finite set, or *alphabet*, and A^* the set of all finite sequences of elements of A , including the empty sequence, denoted by ϵ . The elements of A are usually called *letters* and those of A^* *words*. The word ϵ is called *empty word*. We set $A^+ = A^* \setminus \{\epsilon\}$.

A word $w \in A^+$ can be written uniquely as a sequence of letters as

$$w = a_1 a_2 \cdots a_n,$$

with $a_i \in A$, $1 \leq i \leq n$, $n > 0$. The integer n is called the *length* of w and denoted $|w|$. By definition, the length of ϵ is equal to 0.

Let $w \in A^*$. The word $u \in A^*$ is a *factor* (or *subword*) of w if there exist words λ, μ such that $w = \lambda u \mu$. A factor u of w is called *proper* if $u \neq w$. If $w = u \mu$, for some word μ (resp. $w = \lambda u$, for some word λ), then u is called a *prefix* (resp. a *suffix*) of w . For any word w , we denote respectively by $F(w)$, $\text{Pref}(w)$, and $\text{Suff}(w)$

the sets of its factors, prefixes, and suffixes. Moreover, we shall denote by $\text{alph}(w)$ the set $F(w) \cap A$.

Let $u \in F(w)$. Any pair $(\lambda, \mu) \in A^* \times A^*$ such that $w = \lambda u \mu$ is called an *occurrence* of u in w . If $\lambda \neq \epsilon$ and $\mu \neq \epsilon$, then the occurrence of u is called *internal*. A factor u of w is *repeated* if it has at least two distinct occurrences in w , otherwise it is called *unrepeated*.

Let $w = a_1 a_2 \cdots a_n$ be a word, $a_i \in A, i = 1, \dots, n$. A positive integer $p \leq n$ is called a *period* of w if, for all $i, j \in [1, n]$ such that $i \equiv j \pmod{p}$ one has $a_i = a_j$. This is also equivalent to the condition that w can be factorized as

$$w = s^k s', \quad \text{with } |s| = p, s' \in \text{Pref}(s) \setminus \{s\}, \text{ and } k \geq 1.$$

For any word w , we denote by π_w its *minimal period*. We can always write w as

$$w = r^k r',$$

where $|r| = \pi_w, k \geq 1$, and $r' \in \text{Pref}(r) \setminus \{r\}$. We observe that the preceding factorization is unique. The word w is also called a *fractional power* of r of exponent $\gamma = |w|/|r|$. For this reason, we shall call r also the *fractional root* or, simply, *root* of w . For any word w we denote by r_w the root of w . A word w is called *periodic* if $|w| \geq 2\pi_w$.

The notion of period is also related to the notion of *border* of a word. A word u is called a *border* of w if it is both a proper prefix and a proper suffix of w . The longest border of the word w will be called the *maximal border* of w . If u is a border of w , one can write $w = su = ut$ with $s, t \in A^+$ so that one has (cf. [13])

$$u = s^{k-1} s' \quad \text{and} \quad w = s^k s'$$

for suitable $s' \in \text{Pref}(s) \setminus \{s\}$ and $k \geq 1$. Hence, p is a period of w if and only if w has a border of length $|w| - p$. This implies that a word w has the minimal period π_w if and only if its maximal border has length $|w| - \pi_w$.

In the study of the periods of a non-empty word w , as will appear clear in the sequel, an important role is played by the longest prefix h'_w and the longest suffix k'_w of w which are repeated factors of w . Since h_w and k_w are unrepeated factors of w , one has that

$$h_w = h'_w a \quad \text{and} \quad k_w = b k'_w$$

for suitable letters a and b . Thus, $|h'_w| = H_w - 1$ and $|k'_w| = K_w - 1$.

The following important lemma holds [3]. We report the proof for the sake of completeness.

Lemma 1. *Let w be a non-empty word. If h'_w (resp. k'_w) is not a right (resp. left) special factor of w , then $h'_w = k'_w, h'_w$ has no internal occurrences in w , and*

$$\pi_w = |w| - H_w + 1 = |w| - K_w + 1.$$

Proof. Let $h_w = h'_w a$, with $a \in A$ and suppose that h'_w is not a right special factor of w . If h'_w has an internal occurrence in w , then, since h_w is unrepeated, there exists a letter $b \neq a$ such that $h'_w b$ is a factor of w . Thus, h'_w will be a right special factor, which is a contradiction.

Since h'_w is repeated and it has no internal occurrence in w , h'_w has to be also a suffix of w . Thus, h'_w is a suffix of k'_w . If h'_w is a proper suffix of k'_w , then h'_w would have an internal occurrence in w , which has been excluded. Therefore, $h'_w = k'_w$ so that h'_w is the maximal border of w . Hence, $\pi_w = |w| - |h'_w|$. ■

Two words $w, v \in A^*$ are said to be *conjugate* if there exist $\lambda, \mu \in A^*$ such that $w = \lambda\mu$ and $v = \mu\lambda$. As is well known, conjugacy is an equivalence relation in A^* [13].

Two words w and v of A^* are *root-conjugate* if their roots r_w and r_v are conjugate. One easily verifies that root-conjugacy is an equivalence relation.

For instance, the words $w = abbababb$ and $v = bbababbabab$ are root-conjugate. Indeed, their roots $r_w = abbab$ and $r_v = bbaba$ are conjugate. The words $w = abbab$ and $v = bbabb$ are root-conjugate since $r_w = abb$ and $r_v = bba$ are conjugate. We observe that w and v are not conjugate. The last example shows that two words, even of the same length, can be root-conjugate and not conjugate. On the contrary, two words can be conjugate and not root-conjugate, as, for instance, $w = abbab$ and $v = babab$.

In conclusion of this section, we recall some basic relations concerning the parameters R_w, L_w, H_w, K_w , the length $|w|$, and the minimal period π_w of a given word w [5, 9].

Lemma 2. *Let w be a word. Then, one has*

$$\max\{R_w, K_w\} = \max\{L_w, H_w\}, \quad (1)$$

$$R_w \geq \min\{L_w, H_w\} \quad \text{and} \quad L_w \geq \min\{R_w, K_w\}, \quad (2)$$

$$\pi_w \geq |w| - \min\{K_w, H_w\} + 1 \geq \max\{R_w, L_w\} + 1. \quad (3)$$

3 Semiperiodic words

In this section, we introduce the important notion of semiperiodic word. Several equivalent characterizations of this concept will be given in Proposition 1 (see [5]). Roughly speaking, a semiperiodic word w is a word having a sufficiently ‘small’ minimal period, namely not larger than $|w| - R_w$. Any periodic word is semiperiodic, whereas the converse is not generally true.

Proposition 1. *Let w be a word. The following conditions are equivalent:*

1. $R_w < H_w$,
2. w has a period $p \leq |w| - R_w$,
3. $L_w < K_w$,
4. w has a period $p \leq |w| - L_w$,
5. $R_w = L_w < H_w = K_w$.

Proof. We shall prove only the equivalence of Conditions 1, 2, and 5. Indeed, the equivalence of Conditions 3, 4, and 5 can be proved symmetrically.

1. \Rightarrow 2. Since $R_w < H_w$ and $|h'_w| = H_w - 1 \geq R_w$, one has that h'_w is not a right special factor of w , so that by Lemma 1 one has

$$\pi_w = |w| - H_w + 1.$$

By the hypothesis $R_w < H_w$, it follows that $\pi_w \leq |w| - R_w$.

2. \Rightarrow 5. The word w has a border of length $\geq R_w$. Since a border is a repeated prefix and a repeated suffix as well, one has $H_w > R_w$ and $K_w > R_w$. Let us prove now that $R_w = L_w$. By Eq. (2), one has $L_w \geq \min\{R_w, K_w\} = R_w$ and $R_w \geq \min\{L_w, H_w\}$. Since $R_w < H_w$, one derives $R_w = L_w$. By Eq. (1) it follows that $H_w = \max\{L_w, H_w\} = \max\{R_w, K_w\} = K_w$.

5. \Rightarrow 1. Trivial. ■

A word satisfying any of the Conditions 1–5 of the previous lemma will be called *semiperiodic*.

For a semiperiodic word w , since $R_w < H_w$, one has that h'_w is not a right special factor of w , so that, by Lemma 1, the minimal period of w is $\pi_w = |w| - H_w + 1 = |w| - K_w + 1$.

Any periodic word w is semiperiodic. Indeed, if w is periodic, then $|w| \geq 2\pi_w$. Since, by Eq. (3), $\pi_w > R_w$ it follows that $|w| > \pi_w + R_w$, so that w is semiperiodic. The converse, in general, is not true, as shown by the following example. Let $w = abaab$. One has $H_w = 3$, $R_w = 2$, $|w| = 5$, and the minimal period of w is 3.

An important class of semiperiodic words is the class *PER* of all words having two periods p and q which are coprimes and such that $|w| = p + q - 2$ [10]. It has been proved [9] that if $w \in PER$, then $R_w < K_w$. Since any $w \in PER$ is a palindrome, one has $K_w = H_w$ so that w is semiperiodic. The class *PER* contains words which are not periodic such as, for instance, *aaaabaaaa* which has length 9, periods 5 and 6, with 5 equal to the minimal period.

Let us introduce in A^* the relation \approx defined as follows. For any $w, v \in A^*$,

$$w \approx v \quad \text{if} \quad F(w) \cap A^n = F(v) \cap A^n, \quad \text{with} \quad n = 1 + \min\{R_w, L_w\}.$$

The relation \approx is an equivalence. Indeed, this is a straightforward consequence of the following lemma.

Lemma 3. *For all $w, v \in A^*$, if $w \approx v$, then $\min\{L_w, R_w\} = \min\{L_v, R_v\}$.*

Proof. Let $n = 1 + \min\{R_w, L_w\}$ and $m = 1 + \min\{R_v, L_v\}$. Either all the factors of w of length $n - 1$ are not right special or all the factors of w of length $n - 1$ are not left special. Since $F(w) \cap A^n = F(v) \cap A^n$, it follows that the same property holds for the factors of v of length $n - 1$, so that either $R_v \leq n - 1$ or $L_v \leq n - 1$ and, therefore, $\min\{L_v, R_v\} \leq n - 1 = \min\{L_w, R_w\}$. Hence, $m \leq n$. Consequently, $F(w) \cap A^m = F(v) \cap A^m$, i.e., $v \approx w$. Thus, by using the preceding argument, it follows $n \leq m$. Therefore, $\min\{L_w, R_w\} = \min\{L_v, R_v\}$. ■

The following proposition shows that the equivalence \approx saturates the class of semiperiodic words.

Proposition 2. *For all $w, v \in A^*$, if w is semiperiodic and $w \approx v$, then v is semiperiodic.*

Proof. Let $n = 1 + \min\{L_w, R_w\}$. Since w is semiperiodic one has, from Condition 5 of Proposition 1, $R_w = L_w < H_w = K_w$. From this, it follows that all the factors of w of length $n - 1 = R_w$ are right and left extendable in w . Since $w \approx v$, the same property holds for the factors of v of length $n - 1$. Thus, in view of Lemma 3, one has

$$H_v, K_v \geq n = 1 + R_w = 1 + \min\{L_v, R_v\}.$$

From this one derives that either $R_v < H_v$ or $L_v < K_v$. In both cases, from Proposition 1 it follows that v is semiperiodic. ■

Proposition 3. *Let w be a word. The following conditions are equivalent:*

1. w is semiperiodic,
2. there exists a periodic word v , with the same root of w , such that $v \approx w$,
3. there exists a periodic word v such that $v \approx w$.

Proof. 1. \Rightarrow 2. Let w be a semiperiodic word and set $n = 1 + \min\{R_w, L_w\}$. By Proposition 1, $R_w = L_w = n - 1$. Let α be the border of w of maximal length. We can write

$$w = \alpha s = t \alpha,$$

where $|w| - |\alpha| = |s| = |t| = \pi_w$. Since w is semiperiodic, one has $\pi_w \leq |w| - R_w$, that implies $|\alpha| \geq R_w = n - 1$. Let us set

$$v = t \alpha s.$$

The word v has the border w , so that it has a period $|v| - |w| = |s| = \pi_w$. Thus, $\pi_v \leq \pi_w$. Since w is a factor of v , we have $\pi_w \leq \pi_v$ which implies $\pi_v = \pi_w$. Thus, the root of v is equal to t which is also the root of w . The word v is periodic. Indeed,

$$|v| \geq |ts| = 2\pi_w = 2\pi_v.$$

Finally, any factor of w of length n is trivially a factor of v . Conversely, since $|\alpha| \geq n - 1$, any factor of v of length n has to occur either in the prefix $t\alpha$ or in the suffix αs of v , so that it is a factor of w . This proves that $v \approx w$.

2. \Rightarrow 3. Trivial.

3. \Rightarrow 1. By Proposition 2. ■

We observe that, from the preceding proposition, semiperiodic words can be characterized as the words w which can be ‘prolonged’ (i.e., are factors of) in periodic words, without adding new factors of length $1 + \min\{R_w, L_w\}$. For instance, the word $abaab$ can be prolonged in the periodic word $abaaba$ and $abaab \approx abaaba$, so that $abaab$ is semiperiodic.

Theorem 1. *Let w be a semiperiodic word. One has $w \approx v$ if and only if v is semiperiodic and root-conjugate to w .*

Proof. Let us suppose that w is semiperiodic and let v be a word such that $w \approx v$. By Proposition 2 one has that v is semiperiodic, too. As we shall see in the next section, a semiperiodic word is periodic-like so that, by Proposition 8 and Theorem 2, one derives that v is root-conjugate to w .

Conversely, let v be a semiperiodic word which is a root-conjugate of a semiperiodic word w . In order to prove that $w \approx v$, in view of Proposition 3, we may suppose that v and w are periodic. As v and w are root-conjugate, $\pi_w = |r_w| = |r_v| = \pi_v$. Set $m = \pi_v = \pi_w$. Since v and w are periodic, the sets $F(v) \cap A^m$ and $F(w) \cap A^m$ are equal to the conjugacy classes of r_w and r_v , respectively. Since v and w are root-conjugate, $F(v) \cap A^m = F(w) \cap A^m$. By Eq. (3), $m = \pi_w \geq 1 + \min\{R_w, L_w\}$. Thus, setting $n = 1 + \min\{R_w, L_w\}$, one has $F(v) \cap A^n = F(w) \cap A^n$, that is $w \approx v$. ■

Example 1. Let $w = abccdab$ and $v = bccdabccd$. These words are semiperiodic, $R_w = L_w = 2$, and w and v have the same set of factors of length 3, i.e., $w \approx v$. Thus, w and v are root-conjugate. In fact, $r_w = abccd$ and $r_v = bccda$.

The preceding theorem gives a complete description of the equivalence classes (mod \approx) of semiperiodic words. *Such classes are the intersections of the root-conjugacy classes with the set of semiperiodic words.* The following proposition, proved in [5], gives information on the structure of the equivalence classes (mod \approx) of the words which are not semiperiodic.

Proposition 4. *Let w be a word which is not semiperiodic. If v is a word such that $w \approx v$ and $|w| = |v|$, then $w = v$.*

By Proposition 2, no word in the equivalence class (mod \approx) of a non-semiperiodic word w is semiperiodic. Moreover, by the preceding proposition, the lengths of any pair of words in such a class are distinct.

A straightforward consequence of Theorem 1 and Proposition 4 is the following proposition showing that the root-conjugacy class of any word w is uniquely determined by the length of w and the set of factors of w of length $1 + \min\{R_w, L_w\}$.

Proposition 5. *Let w and v be two words having the same length. If $w \approx v$, then w is root-conjugate to v .*

4 Periodic-like words

In the previous section, we have given several characterizations of the class of semiperiodic words in terms of the parameters R_w , L_w , H_w , K_w , and π_w . Now we introduce two new parameters R'_w and L'_w by means of which one can define a larger class of words which are called *periodic-like* [6, 7]. As we shall see, the parameters R'_w and L'_w will play, in the case of periodic-like words, a role corresponding to that of R_w and L_w in the case of semiperiodic words.

Let w be a word. We denote by R'_w the minimal natural number such that no prefix of w of length $\geq R'_w$ is a right special factor of w . In a similar way, we denote by L'_w the minimal natural number such that no suffix of w of length $\geq L'_w$ is a left special factor of w . Notice that

$$R'_w \leq \min\{H_w, R_w\} \quad \text{and} \quad L'_w \leq \min\{K_w, L_w\}. \tag{4}$$

For instance, let $w = abbbab$. One has $H_w = K_w = L_w = R_w = 3$, $R'_w = 1$, and $L'_w = 2$.

Proposition 6. *Let w be a non-empty word. The following conditions are equivalent.*

1. $R'_w < H_w$,
2. w has a period $p \leq |w| - R'_w$,
3. $L'_w < K_w$,
4. w has a period $p \leq |w| - L'_w$.

Proof. 1. \Rightarrow 2. If $R'_w < H_w$, then $|h'_w| \geq R'_w$ so that h'_w is not right special in w . By Lemma 1, $\pi_w = |w| - H_w + 1 \leq |w| - R'_w$.

2. \Rightarrow 1. Condition 2 implies that $\pi_w \leq |w| - R'_w$. By Eq. (3), $\pi_w \geq |w| - H_w + 1$. Thus, $R'_w < H_w$.

1. \Rightarrow 3. If $R'_w \leq H_w$, then $|h'_w| \geq R'_w$ so that h'_w is not right special in w . By Lemma 1, $h'_w = k'_w$ and h'_w has no internal occurrences in w . Hence, k'_w is not left special. Since also any suffix of w of length $\geq K_w$ is unrepeated and, therefore, it is not left special, it follows that $L'_w < K_w$.

The proofs of the implications 3. \Rightarrow 4., 4. \Rightarrow 3., and 3. \Rightarrow 1. are obtained by symmetric arguments. \blacksquare

A non-empty word w is called *periodic-like* if it satisfies any of the equivalent Conditions 1–4 of the previous proposition.

Observe that if a word w is not periodic-like, then from Eq. (4) one has $R'_w = H_w$ and $L'_w = K_w$.

If w is semiperiodic, then $R'_w \leq R_w < H_w$. Thus, by Proposition 6, any semiperiodic word is periodic-like. However, the converse is not generally true as shown by the following example.

Example 2. Let w be the word $w = ab^n a$, with $n \geq 2$. In such a case, one has $R_w = n \geq 2$, $H_w = 2$, $R'_w = L'_w = 1$, and $\pi_w = n + 1$ so that w is periodic-like but it is not semiperiodic.

The following proposition [6, 7] gives some further characterizations of periodic-like words.

Proposition 7. *Let w be a non-empty word. The following conditions are equivalent.*

1. w is periodic-like,
2. h'_w is not right special in w ,
3. h'_w has no internal occurrences in w ,
4. k'_w is not left special in w ,
5. k'_w has no internal occurrences in w ,

- 6. the maximal border of w has no internal occurrences,
- 7. w has a border which has no internal occurrences.

For a periodic-like word w , since h'_w is not a right special factor of w , by Lemma 1 the minimal period of w is

$$\pi_w = |w| - H_w + 1 = |w| - K_w + 1. \tag{5}$$

For any word w of A^* , we denote by \mathcal{B}_w the set of the maximal proper boxes of w . Let us introduce in A^* the relation \equiv defined as follows. For any $w, v \in A^+$,

$$w \equiv v \quad \text{if} \quad \mathcal{B}_w = \mathcal{B}_v \quad \text{and} \quad \text{alph}(w) \cap \text{alph}(v) \neq \emptyset.$$

Moreover, we set $\epsilon \equiv \epsilon$.

Lemma 4. *The relation \equiv is an equivalence relation.*

Proof. From the definition of \equiv , one has only to verify that the transitive property holds. Suppose that $w \equiv v$ and $v \equiv u$, with $u, v, w \in A^+$. Then $\mathcal{B}_w = \mathcal{B}_v = \mathcal{B}_u$. If $\mathcal{B}_w \neq \emptyset$, then $\text{alph}(w) \cap \text{alph}(u) \neq \emptyset$ and, therefore, $w \equiv u$. If $\mathcal{B}_w = \emptyset$, then, as one easily verifies, w, v , and u are powers of single letters so that since $\text{alph}(w) \cap \text{alph}(v) \neq \emptyset$ and $\text{alph}(v) \cap \text{alph}(u) \neq \emptyset$, one has $\text{alph}(w) = \text{alph}(v) = \text{alph}(u)$ and, again, $w \equiv u$. ■

Proposition 8. *For any $w, v \in A^*$, if $w \approx v$, then $w \equiv v$.*

Proof. If $w \approx v$, then $F(w) \cap A^n = F(v) \cap A^n$ with $n = 1 + \min\{R_w, L_w\}$. Since $n \geq 1$, one has $\text{alph}(w) \cap \text{alph}(v) \neq \emptyset$.

Let us prove now that $\mathcal{B}_w = \mathcal{B}_v$. It is sufficient to prove that any proper box of w is a proper box of v and *vice versa*. Let asb be a proper box of w , with a and b letters and s a bispecial factor of w . Since s is bispecial, there exist letters c and d such that $sc, ds \in F(w)$, $c \neq b$, $d \neq a$. Moreover, $|cs| = |sd| < |asb| \leq n$ so that $asb, cs, sd \in F(v)$. This implies that s is bispecial in v and asb is a proper box of v . Conversely, in a symmetric way, one can prove that any proper box of v is a proper box of w . ■

The following lemma, whose proof is in [5], gives a useful relation between the sets of factors of two words which are in the relation \equiv .

Lemma 5. *For any $w, v \in A^*$ such that $w \equiv v$, one has*

$$F(v) \subseteq F(w) \cup A^+h_wA^* \cup A^*k_wA^+.$$

Proposition 9. *Let w be a word. The following conditions are equivalent:*

- 1. w is periodic-like,
- 2. there exists a periodic word v , with the same root of w , such that $v \equiv w$,
- 3. there exists a periodic word v such that $v \equiv w$.

Proof. 1. \Rightarrow 2. If w is periodic, then there is nothing to prove. Let us then suppose that w is not periodic. Since w is periodic-like, by Eq. (5) we can write:

$$w = h'_w u h'_w, \quad \text{with } u \in A^+.$$

Let us set

$$v = h'_w u h'_w u = wu.$$

By Eq. (5), one has $\pi_w = |h'_w u|$. Moreover, v has the period π_w so that $\pi_v \leq \pi_w$ and $\pi_w \leq \pi_v$ since w is a factor of v . Thus, $\pi_w = \pi_v$ that implies $r_w = r_v = h'_w u$.

Since w is a prefix of v , any proper box of w is, trivially, a proper box of v . Let us prove the converse.

Let $\alpha = asb$ be a proper box of v , with $a, b \in A$ and s a bispecial factor of v . We prove that $h'_w \notin F(s)$. Indeed, by Proposition 7, h'_w has no internal occurrence in $w = h'_w u h'_w$ and, therefore, it has exactly two occurrences in v , one of which is initial, whereas s has at least two non-initial occurrences in v because s is a left special factor of v .

Since s is bispecial in v , there exist letters $c \neq b$ and $d \neq a$ such that $sc, ds \in F(v)$. As h'_w is not a factor of s , the words $\alpha = asb$, sc , and ds occur either in the prefix $h'_w u h'_w$ of v or in the suffix $h'_w u$ of v . Thus $\alpha, sc, ds \in F(w)$. This proves that α is a proper box of w . Thus $v \equiv w$.

2. \Rightarrow 3. Trivial.

3. \Rightarrow 1. Since v is periodic, by Proposition 1, $H_v = K_v > R_v$. Assume that w is not periodic-like. Then h'_w is a right special factor of w by virtue of Proposition 7. This implies that there exist letters a and b , $a \neq b$, such that

$$h_w = h'_w a \quad \text{and} \quad h'_w b \in F(w).$$

If $H_w \geq H_v$, there exists a suffix t of h'_w such that

$$ta, tb \in F(w), \quad \text{and} \quad |ta| = |tb| = H_v = K_v.$$

By Lemma 5 one derives that $ta, tb \in F(v)$. Thus t is right special in v , so that $R_v \geq |t| + 1 = H_v$, which is a contradiction.

If, on the contrary, $H_w < H_v$, one has $|h'_w a| = |h'_w b| < H_v = K_v$ and, therefore, by Lemma 5,

$$h'_w a, h'_w b \in F(v).$$

Denote by s the longest word such that $sh'_w a, sh'_w b \in F(v)$. Since sh'_w is right special in v , one has

$$|sh'_w a| = |sh'_w b| \leq R_v < H_v.$$

Thus, $sh'_w a$ and $sh'_w b$ are left extendable in v , i.e., there exist letters c and d such that

$$csh'_w a, dsh'_w b \in F(v).$$

Moreover, $c \neq d$ by the maximality of $|s|$. We conclude that $csh'_w a$ is a proper box of v and, hence, a factor of w . This yields a contradiction, for $h'_w a = h_w$ is not left extendable in w . This proves that w is periodic-like. ■

The preceding proposition shows that periodic-like words are exactly the words which can be ‘prolonged’ in a periodic word, without changing the set of maximal proper boxes.

The following proposition shows that the equivalence \equiv saturates the class of periodic-like words.

Proposition 10. *For all $w, v \in A^*$, if w is periodic-like and $w \equiv v$, then v is periodic-like.*

Proof. Since w is periodic-like, by Proposition 9 there exists a periodic word u such that $u \equiv w$. Thus, $u \equiv v$ and therefore, again by the preceding proposition, v is periodic-like. ■

Theorem 2. *Let w be a periodic-like word. One has $w \equiv v$ if and only if v is periodic-like and root-conjugate to w .*

Proof. Let w be a periodic-like word such that $w \equiv v$. By the preceding proposition, v is periodic-like. We prove that w and v are root-conjugate. By Proposition 9 we can assume, with no loss of generality, that w and v are periodic.

First, let us suppose that $\pi_v \geq \pi_w$. Since w is periodic, i.e., $|w| \geq 2\pi_w$, by Eq. (5) one derives $H_w = K_w \geq 1 + \pi_w$. Thus, by Lemma 5, all the factors of v of length $1 + \pi_w$ are factors of w . From this, one derives that v has the period π_w . Indeed, write $v = a_1 \cdots a_n$, with $a_i \in A$, $1 \leq i \leq n$. For $1 \leq i \leq n - \pi_w$ one has $a_i = a_{i+\pi_w}$ since $a_i \cdots a_{i+\pi_w}$ is a factor of w . Thus, $\pi_v \leq \pi_w$. We conclude that $\pi_w = \pi_v$. If one supposes that $\pi_v \leq \pi_w$, one obtains again $\pi_w = \pi_v$, by a symmetric argument. Hence, in any case, $\pi_w = \pi_v$.

Moreover, r_v is a factor of w of length π_w and, therefore, it is conjugate to r_w . We conclude that w and v are root-conjugate.

Conversely, suppose that v and w are periodic-like and root-conjugate. By Proposition 9 we can assume, with no loss of generality, that w and v are periodic. As v and w are root-conjugate, $\pi_w = |r_w| = |r_v| = \pi_v$. Set $m = \pi_v = \pi_w$. Since v and w are periodic, the sets $F(v) \cap A^m$ and $F(w) \cap A^m$ are equal to the conjugacy class of r_w and r_v , respectively. Since v and w are root-conjugate, then $F(v) \cap A^m = F(w) \cap A^m$. Since $m \geq n = 1 + \min\{R_w, L_w\}$ it follows that $F(v) \cap A^n = F(w) \cap A^n$. Hence, $w \approx v$ and consequently, by Proposition 8, $w \equiv v$. ■

The preceding theorem gives a complete description of the equivalence classes (mod \equiv) of periodic-like words. *Such classes are the intersections of the root-conjugacy classes with the set of periodic-like words.* Thus, the root-conjugacy class of a periodic-like word w is uniquely determined by $\text{alph}(w)$ and the set \mathcal{B}_w of its maximal proper boxes.

The following example shows that the hypothesis that w is periodic-like is essential. In fact, there are words which are not periodic-like, whose equivalence classes (mod \equiv) are not included in any root-conjugacy class.

Example 3. Consider the set of words $a^p b^q$, with $p, q \geq 2$. These words are not periodic-like and they have the same set of maximal proper boxes, given by $\{aa, ab, bb\}$. However, any two distinct words of this set are not root-conjugate. One can easily verify that the set $\{a^p b^q \mid p, q \geq 2\}$ is a class of the equivalence \equiv .

The following proposition shows that the equivalences \approx and \equiv coincide on semiperiodic words.

Proposition 11. *Let w and v be semiperiodic words. Then,*

$$w \equiv v \quad \text{if and only if} \quad w \approx v.$$

Proof. By Theorem 2, if $w \equiv v$, then w and v are root-conjugate. Consequently, by Theorem 1, one has $w \approx v$.

Conversely, if $w \approx v$, then one has $w \equiv v$, by Proposition 8. ■

The following proposition shows that the equivalence class (mod \approx) of a periodic-like word which is not semiperiodic is a singleton.

Proposition 12. *Let w be a periodic-like word which is not semiperiodic. If $w \approx v$, then $w = v$.*

Proof. Since w is periodic-like, by Eq. (5) one has $H_w = K_w$. Moreover, since w is not semiperiodic, by Proposition 1 one has

$$H_w = K_w \leq \min\{L_w, R_w\}.$$

Hence, by the previous equation and Eq. (1), one has

$$R_w = \max\{R_w, K_w\} = \max\{L_w, H_w\} = L_w = \min\{L_w, R_w\}.$$

Thus $\min\{L_w, R_w\} = \max\{R_w, K_w\}$. Now, set $n = 1 + \max\{R_w, K_w\}$. Since $w \approx v$, w and v have the same set of factors up to length n . As proved in [3], a word w is uniquely determined by its set of factors up to length $1 + \max\{R_w, K_w\}$. Hence, $w = v$. ■

5 Periodicity theorems

In this section, by using the notion of periodic-like word, we give an improvement of the periodicity theorem of Fine and Wilf in the case of finite words.

Moreover, we consider a canonical decomposition of a word in periodic-like subwords. We give a new periodicity theorem [7], showing that the minimal period of a word is equal to the sum of the minimal periods of the periodic-like subwords occurring in the canonical decomposition.

The following lemma [6, 7] gives a condition assuring that a period of a prefix of a word can be ‘extended’ to the entire word.

Lemma 6. *If a word w has a prefix of period p and length $p + R'_w$, then p is a period of w .*

Proposition 13. *If a word w has two periods $p, q \leq |w| - R'_w$, then w has also the period $d = \gcd(p, q)$.*

Proof. We can assume, with no loss of generality, that $p < q$. Set $w = a_1 a_2 \cdots a_n$, with $a_i \in A$, $1 \leq i \leq n$. For $1 \leq i \leq R'_w$, one has $i + q \leq |w| = n$ and

$$a_i = a_{i+q} = a_{i+q-p}.$$

Thus, $a_1 a_2 \cdots a_{R'_w+q-p}$ has the period $q - p$. By the preceding proposition, $q - p$ is a period of w . By making induction on $\max\{p, q\}$, w has the period $\gcd(p, q - p) = d$. ■

Corollary 1. *Let w be a word. For any period p of w either p is a multiple of the minimal period π_w or $p > |w| - R'_w$.*

Proof. If $p \leq |w| - R'_w$, then, by Proposition 13, w has the period $\gcd(p, \pi_w)$. Thus, $\gcd(p, \pi_w) = \pi_w$ so that p is a multiple of π_w . ■

The preceding corollary shows that a periodic word satisfies the harmonic property (cf. Sect. 1) for all periods which are less than or equal to $|w| - R'_w$.

By Proposition 13 one easily derives [6, 7] the theorem of Fine and Wilf for finite words.

Theorem 3. *Let w be a word having two periods p and q and length*

$$|w| \geq p + q - \gcd(p, q).$$

Then w has the period $\gcd(p, q)$.

Proof. It is well known that one can always reduce himself to consider only the case when $\gcd(p, q) = 1$ (see e.g. [13]). Since $p, q \geq R'_w + 1$, one has

$$|w| \geq p + q - 1 \geq q + R'_w \quad \text{and} \quad |w| \geq p + q - 1 \geq p + R'_w.$$

This implies that $p, q \leq |w| - R'_w$, so that the conclusion follows from Proposition 13. ■

The following proposition shows that, under the hypotheses of the previous theorem, w is periodic and, with the only exception of a trivial case, one has $p, q < |w| - R_w$.

Proposition 14. *Let w be a word having two periods p and q , with $p < q$, and length*

$$|w| \geq p + q - \gcd(p, q).$$

Then w is periodic and $p < |w| - R_w$. If, moreover, q is not a multiple of p , then $q < |w| - R_w$.

Proof. By the theorem of Fine and Wilf, $d = \gcd(p, q)$ is a period of w . Moreover, $q - d \geq p$, so that $|w| \geq p + q - d \geq 2p$, i.e., w is periodic. By Eq. (3), $p \geq R_w + 1$, so that

$$|w| \geq p + q - d \geq R_w + 1 + p.$$

Thus, $p < |w| - R_w$.

Let us now suppose that q is not a multiple of p . In such a case, $p - d \geq d$, so that

$$|w| \geq p + q - d \geq q + d.$$

Since d is a period of w , one has $d \geq R_w + 1$ and $|w| \geq q + R_w + 1$, that implies $q < |w| - R_w$. ■

We remark that the lower bound for the length of a word in the theorem of Fine and Wilf is optimal if one does not make any further hypothesis on the ‘structure’ of the word. Indeed, for any pair of integers p and q one can always construct a word w of length $|w| = p + q - d - 1$, with $d = \gcd(p, q)$, and such that w has the periods p and q but not the period d . However, by using Proposition 13, one can show [7] that this lower bound can decrease if one makes a restriction on the value of R'_w . For instance, if a word w of length 16 has periods 9 and 11 and $R'_w \leq 5$, then w has period 1, i.e., it is a power of a letter. Notice that $16 < 9 + 11 - \gcd(9, 11)$.

We shall now relate the minimal period π_w of a non-empty word w with the minimal periods of the elements of a suitable decomposition of w called *(left) canonical decomposition of the word w in periodic-like subwords* [7].

This canonical decomposition is obtained by the following inductive procedure.

We make induction on the length of w . The word h'_w occurs at least twice in w . We shall consider now all the occurrences of h'_w in w . Formally, we can write

$$w = \lambda_i h'_w \mu_i, \quad i = 1, \dots, n,$$

with

$$|\lambda_1| < |\lambda_2| < \dots < |\lambda_n|,$$

where $n \geq 2$ and (λ_i, μ_i) , $i = 1, \dots, n$, denote all the distinct occurrences of h'_w in w . Note that $\lambda_1 = \epsilon$, since h'_w is a prefix of w .

For $i = 1, \dots, n - 1$, one has $\lambda_i h'_w \mu_i = \lambda_{i+1} h'_w \mu_{i+1}$ so that for suitable $\alpha_i, \beta_i \in A^+$,

$$\lambda_{i+1} = \lambda_i \alpha_i, \quad \mu_i = \beta_i \mu_{i+1}, \quad \text{and} \quad \alpha_i h'_w = h'_w \beta_i.$$

For $i = 1, \dots, n - 1$, we set

$$w_i = \alpha_i h'_w = h'_w \beta_i.$$

If $\mu_n = \epsilon$ (and, in particular, if $|w| = 1$ which gives the base of the induction), then the canonical decomposition of w in periodic-like subwords is

$$(w_1, \dots, w_{n-1}).$$

Let us then suppose that $\mu_n \neq \epsilon$ and consider the word

$$w_n = h'_w \mu_n.$$

Since $|w_n| < |w|$, by making induction on the length of w , we can assume that the canonical decomposition of w_n in periodic-like subwords is defined. Let us denote it by (w'_1, \dots, w'_m) , $m \geq 1$. Then the canonical decomposition of w in periodic-like subwords is

$$(w_1, \dots, w_{n-1}, w'_1, \dots, w'_m).$$

The following noteworthy theorem, proved in [7], holds.

Theorem 4. *Let w be a word and (w_1, w_2, \dots, w_n) be its canonical decomposition in periodic-like subwords. Then*

$$\pi_w = \sum_{i=1}^n \pi_{w_i}.$$

Example 4. Let w be the word:

$$w = \underline{abccb} \underline{abab} \underline{abccb} \underline{baabccb} \underline{aaabccabab}$$

where the occurrences of $h'_w = abcc$ are underlined. The first elements of the canonical decomposition in periodic-like subwords of w are

$$w_1 = \underline{abccb} \underline{abab} \underline{abccb} \quad \text{and} \quad w_2 = \underline{abccb} \underline{baabccb}.$$

Since $\mu_3 = baaabccabab \neq \epsilon$, we have to compute the canonical decomposition in periodic-like subwords of $h'_w \mu_3 = \underline{abccb} \underline{aaabccabab}$. We obtain

$$w'_1 = \underline{abccb} \underline{aaabcc}$$

and $\mu'_2 = abab$. Decomposing $\underline{abccab} \underline{ab}$, one obtains

$$w''_1 = \underline{abccab} \quad \text{and} \quad w''_2 = \underline{ab} \underline{ab}.$$

Thus, the canonical decomposition of w is $(w_1, w_2, w'_1, w''_1, w''_2)$. One has $\pi_{w_1} = 8$, $\pi_{w_2} = 7$, $\pi_{w'_1} = 7$, $\pi_{w''_1} = 3$, $\pi_{w''_2} = 2$, and $\pi_w = 27$.

The following proposition [6] shows that if a word has a periodic-like factor having the same minimal period, then the word has to be periodic-like.

Proposition 15. *Let u be a periodic-like factor of a word w . If $\pi_u = \pi_w$, then w is periodic-like.*

Some noteworthy consequences of the preceding proposition concerning the ‘local’ periodicity of a word and the critical point theorem are shown in [7].

6 Concluding remarks

We have considered two suitable generalizations of the notion of periodic word, namely semiperiodic and periodic-like words. The definition of these classes was based on the relations between the minimal period π_w of a word w and the parameters R_w , L_w , H_w , K_w , R'_w , and L'_w .

However, these two notions are naturally linked to the equivalences \approx and \equiv . Indeed, Proposition 3 characterizes semiperiodic words as the words w which can be ‘prolonged’ in a periodic word, without adding new factors of length $1 + \min\{R_w, L_w\}$. Similarly, Proposition 9 shows that periodic-like words are exactly the words which can be ‘prolonged’ in a periodic word, without changing the set of maximal proper boxes.

A complete description of the equivalence classes (mod \approx) and (mod \equiv) of periodic-like words was given by Theorems 1 and 2 and Proposition 12: the equivalence \approx coincides with root-conjugacy on the class of semiperiodic words and is reduced to identity on the class of periodic-like words which are not semiperiodic, while \equiv coincides with the root-conjugacy on the entire class of periodic-like words.

Despite the uniqueness conditions of Proposition 4, we do not have, up to now, a complete description of the equivalence classes (mod \approx) and (mod \equiv) of words

which are not periodic-like. We remark that for the equivalence \equiv there is no result corresponding to Propositions 4 and 5, which hold for the equivalence \approx . Indeed, the words

$$a^p b^q, \quad \text{with } p, q \geq 2,$$

considered in Example 3 are not periodic-like and two distinct words of this kind are in the same class (mod \equiv) but they are not root-conjugate. Moreover, there are distinct words of this kind having the same length such as, e.g., $aabbb$ and $aaabb$. Another example is given by the two words $w = abcadcadc$ and $v = abcabcadc$ which are not periodic-like. One has $|w| = |v|$ and $\mathcal{B}_w = \mathcal{B}_v = \{ab, ad, bc, ca, dc\}$, so that $w \equiv v$. However, they are not root-conjugate.

As we have seen, periodic-like words w satisfy the condition

$$\pi_w = |w| - H_w + 1 = |w| - K_w + 1,$$

so that, in view of Eq. (3), they can be considered as ‘words of minimal period’. However, there are words satisfying the previous equation, which are not periodic-like. For instance, for the word $w = abbabaab$ one has $H_w = K_w = R_w = L_w = 3$, $|w| = 8$, and $\pi_w = 6$.

One can also consider the more restricted class of the words w for which

$$\pi_w = \max\{R_w, L_w\} + 1. \quad (6)$$

As proved in [9], this class includes the words of the set *PER* which are semiperiodic words (cf. Sect. 3). However, it includes also words which are not periodic-like such as $w = aaaba$.

If Eq. (6) is satisfied, then from Eq. (3), one obtains that $|w| = \max\{R_w, L_w\} + \min\{K_w, H_w\}$ and, from Eq. (1), one easily derives that

$$|w| = R_w + K_w = L_w + H_w.$$

The words satisfying the preceding relation have been studied in [9] and called *trapezoidal words* (see also [8]). A remarkable example of trapezoidal words is given by the finite factors of *Sturmian words* (cf. [14, Chap. 2]).

References

- [1] A. Apostolico and A. Ehrenfeucht, Efficient detection of quasiperiodicities in strings, *Theoretical Computer Science* **119** (1993) 247–265.
- [2] A. Carpi and A. de Luca, Words and Repeated Factors, *Séminaire Lotharingien de Combinatoire* **B421** (1998) 24 pp. Also in: D. Foata and G. Han (Eds.), *The Andrews Festschrift*, Springer-Verlag (Berlin, to appear) pp. 231–251.
- [3] A. Carpi and A. de Luca, Words and Special Factors, Preprint 98/33, Dipartimento di Matematica dell’Università di Roma “La Sapienza” (1998), *Theoretical Computer Science*, to appear.
- [4] A. Carpi and A. de Luca, Repetitions and Boxes in Words and Pictures, in: J. Karhumäki, H. Maurer, G. Păun, and G. Rozenberg, (Eds.), *Jewels are Forever*, Springer-Verlag (Berlin, 1999), pp. 295–306.

- [5] A. Carpi and A. de Luca, Semiperiodic words and root-conjugacy, Preprint 10/2000, Dipartimento di Matematica dell'Università di Roma "La Sapienza" (2000), *Theoretical Computer Science*, to appear.
- [6] A. Carpi and A. de Luca, Periodic-like words, in: M. Nielsen and B. Rovan (Eds.) *Mathematical Foundations of Computer Science 2000*, Lecture Notes in Computer Science, vol. 1893, Springer-Verlag (Berlin, 2000) pp. 264–274.
- [7] A. Carpi and A. de Luca, *Periodic-like words, periodicity, and boxes*, Preprint 30/2000, Dipartimento di Matematica dell'Università di Roma "La Sapienza" (2000).
- [8] F. D'Alessandro, Trapezoidal words, *Theoretical Computer Science*, to appear.
- [9] A. de Luca, On the Combinatorics of Finite Words, *Theoretical Computer Science* **218** (1999) 13-39.
- [10] A. de Luca and F. Mignosi, Some Combinatorial properties of Sturmian words, *Theoretical Computer Science* **136** (1994) 361–385.
- [11] N. J. Fine and H. S. Wilf, Uniqueness Theorem for Periodic Functions, *Proc. Amer. Math. Soc.* **16** (1965) 109–114.
- [12] C. S. Iliopoulos and L. Mouchard, Quasiperiodicity and string covering, *Theoretical Computer Science* **218** (1999) 205–216.
- [13] M. Lothaire, *Combinatorics on Words*, Addison-Wesley (Reading, MA, 1983).
- [14] M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press (Cambridge, to appear).
- [15] M. Régnier and L. Mouchard, Periods and quasiperiods characterization, in: R. Giancarlo and D. Sankoff (Eds.), *Combinatorial Pattern Matching*, Lecture Notes in Computer Science, vol. 1848, Springer-Verlag (Berlin, 2000), pp. 388–396.

Arturo Carpi
Istituto di Cibernetica del CNR
via Toiano 6, 80072 Arco Felice (NA), Italy
arturo@arturo.cib.na.cnr.it

Aldo de Luca
Dipartimento di Matematica dell'Università di Roma "La Sapienza"
piazzale Aldo Moro 2, 00185 Roma, Italy and
Centro Interdisciplinare "B. Segre", Accademia dei Lincei
via della Lungara 10, 00100 Roma, Italy
deluca@mat.uniroma1.it