A Characterization of 2^+-free Words over a Binary Alphabet

Arto Lepistö

Department of Mathematics, University of Turku,
FIN-20500 Turku, Finland

Turku Centre for Computer Science
TUCS Technical Report No 74
December 1996

ISBN 951-650-914-2
ISSN 1239-1891

Abstract

It is shown that 2^+-repetition, i.e. a word of the form $uvuvu$ where u is a letter and v is a word, is the smallest repetition which can be avoided in infinite words over binary alphabet. Such binary words avoiding pattern $uvuvu$, finite or infinite, are called as 2^+-*free words* and those words are the main topic of this work. It is shown here that 2^+-free words over binary alphabet can be presented as words built from special kind of blocks, called Morse-blocks, with some rules. In particular, the given presentation by these blocks is unique for 2^+-free words long enough. Moreover, it is also shown that the language generated by this presentation can be described by some automaton. In fact, the corresponding presentation in blocks for finite 2^+-free words can be generated by a modified version of cross-product of two gsm's. On the other hand, for infinite 2^+-free words, one gsm is enough. Also, a little more accurate bounds for the count of 2^+-free words of a given length are given here.

I Introduction and preliminaries

Problems about repetions and avoidable patterns (cf. [BEM]), i.e. patterns which can be avoided in some infinite words, have an important role in language theory. Since Axel Thue's work (cf. [Th1] and [Th2]) there have been many papers about this topic. In this work we are going to deal with a specific pattern.

An *alphabet* is a finite set. A *word* over some alphabet is a sequence of elements (called *symbols* or *letters*) of this alphabet (cf. [Lo] and [CK]). It was proved by Thue (and independently by Morse (cf. [MH]) that the Thue-Morse sequence, which is obtained by the Thue-Morse morphism, is an infinite word over binary alphabet which does not contain any subwords of the form $uvuvu$, where u is a symbol from that alphabet and v is an arbitrary word over the same alphabet. Therefore, the pattern $uvuvu$ is the smallest repetition avoidable on binary alphabet. Such words, which does not contain this pattern, are called *overlap-free words*, *weakly square-free words* or *2^+-free words*. Our goal in this work is to characterize the 2^+-free words and obtain some results about the size of the language consisting of these words. In order to do that we shall also consider the words of the form $uvuvu$, where u and v are as previously described. Some elementary properties of this kind of words are studied in many papers, cf. [RS] and [SS].

In Section II we introduce a recursive characterization for 2^+-free words. That recursive characterization has the property that changes are applied to a whole given word. There is also another recursive way to characterize 2^+-free words. This method is introduced by Fife, cf. [Fi] and [Be]. In this characterization changes indicated by an *automaton* are applied to a suffix of given word instead of the whole word. An automaton (cf. [Sa]) is a 5-tuple $A = (Q, \Sigma, E, q_0, F)$, where

Q is a finite *state set*,
Σ is an alphabet,
$E \subseteq Q \times \Sigma \times Q$ is a set of transitions,
$q_0 \in Q$ is an initial state, and
$F \subseteq Q$ is a set of final states.

In Chapter II.5 we give a brief introduction to this Fife's characterization of infinite 2^+-free words.

In Chapters II.2 and II.3, we study the formal structure of 2^+-free words. The first task is to find some kind of deterministic mathematical mechanism by which for any given word one can check whether it is a 2^+-free word or not. By chapter II.2, one such way is to use the Thue-Morse morphism, which replaces a letter with two letters consisting that letter and its complement, recursively with some operations which modifies the ends of a given word.

In Chapter II.3 we are able to construct a mathematical structure which is related to 2^+-free words. This structure, referred to as *S-structure* contains mainly information about those specific operations modifying the left and the right ends of words. As a matter of fact we can even characterize 2^+-free words by some S-structures. Finally in this chapter we construct a structural way, based on the images of letters in alphabet produced by powers of the Thue-Morse morphism, to represent all 2^+-free words. A similar but weaker way is introduced in [RS], by Restivo and Salemi, in which the basic elements were the words $abba$, $baab$, and their images produced also by iterating the Thue-Morse morphism. Basicly, these both methods are based on facts which are more or less the same than those in [Car].

It is proved by Restivo and Salemi ([RS]) that the function giving the number of 2^+-free words having a given length can be bounded from above by a polynomial. On the other hand, Brandenburg proved in [Br] that the number of cube-free words having a given length grows exponentially. In this light the former result is remarkable.

In Chapter II.4 we obtain a result showing that the function giving the number of 2^+-free words having a given length cannot be bounded from above and below by two polynomial functions having the same exponent. This result is independently obtained also by Cassaigne in [Cas]. In his work, the method used is similar to the method used in this work. The main difference is that the mechanism used to deal with 2^+-free words in this work is smaller and it is analysed further. Also, characterization of 2^+-free words is done also by describing the block-structure of such words (cf. [RS]) instead of only describing sequences of morphisms or actions needed to obtain them (cf. [Fi], [Car] and [Cas]). We also give some upper and lower bounds for the number of 2^+-free words of a given length. These estimates are somewhat tighter than those presented in [Ko1].

After that we study infinite 2^+-free words. These results are only the applications of structural information obtained earlier in this work. For example, we give another proof for the theorem stating that there are uncountably many infinite 2^+-free words. More precisely, a subset of infinite 2^+-free words can be related to the set of all real numbers such that this correspondence is one-to-one. Another proof of this result is based on a construction due to Kakutani (see probl. 2.2.3 in [Lo]). We also characterize those finite 2^+-free words which are prefixes of some infinite 2^+-free word. Actually, every prefix of an infinite 2^+-free word can be extended uncountably many ways to an infinite 2^+-free word. Moreover, we give a structural presentation for those based on the results obtained in Chapter II.3.

II Language of 2^+-free words

II.1 Definitions

In this section we are going to introduce some notations and definitions we use in this work. Some specific notations can also be introduced in the appropriate sections.

Let A be a binary alphabet consisting of letters a and b. By *infinite words* over this alphabet we mean members in set A^ω. The *complement* of a letter with respect to this binary alphabet is denoted by a bar over the letter. The complement of a word is denoted by a bar over the word and it is the word obtained by taking complements of individual letters in the same order. And, the *length* of a given word w, i.e. number of symbols in w, is denoted by $|w|$.

A word is *weakly square-free* or 2^+*-free* if it does not contain any subword of the form $uvuvu$, where $u \in A$ and $v \in A^*$. To simplify our considerations, we define *allmost 2^+-free words*. Such a word itself is not 2^+-free, but its all proper subwords are 2^+-free. The languages of 2^+-free and allmost 2^+-free words are denoted by $\mathcal{LWSF}$ and $\mathcal{LAWSF}$, respectively. As we will soon see, the prefixes and suffixes of words will play an important role, and therefore we introduce the following notations: For words u and v we set

$$\Psi_{v-} = \{w \in \mathcal{LAWSF} \mid \exists x \in A^*, w \in \{vx, \overline{vx}\}\};$$

$$\Theta_{v-} = \{w \in \mathcal{LWSF} \mid \exists x \in A^*, w \in \{vx, \overline{vx}\}\};$$

$$\Psi_{-v} = \{w \in \mathcal{LAWSF} \mid \exists x \in A^*, w \in \{xv, \overline{xv}\}\};$$

$$\Theta_{-v} = \{w \in \mathcal{LWSF} \mid \exists x \in A^*, w \in \{xv, \overline{xv}\}\};$$

$$\Psi_{u-v} = \{w \in \mathcal{LAWSF} \mid \exists x \in A^*, w \in \{uxv, ux\overline{v}, \overline{u}xv, \overline{uxv}\}\};$$

$$\Theta_{u-v} = \{w \in \mathcal{LWSF} \mid \exists x \in A^*, w \in \{uxv, ux\overline{v}, \overline{u}xv, \overline{uxv}\}\};$$

The notations $\Psi_X(n)$ and $\Theta_X(n)$ denote the functions giving the number of allmost weakly and 2^+-free words, respectively, having the length n. Later in this paper we also consider *infinite 2^+-free words*, i.e. those words of A^ω which have a prefix of any given length and does not have any allmost 2^+-free word as a subword.

Next we define the mappings Pref_n and Suff_n:

$$\mathrm{Pref}_n \ : \ A^* \to A^n, w \mapsto u, \text{ such that } w = ux, \text{ for some } x \in A^*;$$

$$\mathrm{Suff}_n \ : \ A^* \to A^n, w \mapsto v, \text{ such that } w = xu, \text{ for some } x \in A^*.$$

The sets

$$F_l = \{aaba, bbab\}$$

and

$$F_r = \{abaa, babb\}$$

will be very closely related to the recursive characterization of the 2^+-free words.

The Thue-Morse morphism θ,

$$\theta(u) = u\overline{u}, \text{ for } u \in A,$$

has the infinite Thue-Morse word,

$$w_{TM} = \lim_{n \mapsto \infty} \theta^n(a)$$

and its complement $\overline{w}_{TM}$ as its only *fixed points*, i.e. infinite words w satisfying $w = \theta(w)$.

By $\sigma_l, \sigma_r, \rho_l$ and ρ_r, we denote the following four mappings:

$$\sigma_l \ : \ A \to A, u \mapsto \overline{u};$$

$$\sigma_r \ \text{ (identity)};$$

$$\rho_l \ : \ A \to A^2, u \mapsto \overline{uu};$$

$$\rho_r \ : \ A \to A^2, u \mapsto uu.$$

The mappings σ_l and ρ_l will be applied only to the left end of a given word, and correspondingly the mappings σ_r and ρ_r are applied to the right end.

Let $w = uw'v$, where $u, v \in A$ and $w' \in A^*$, be a word with length greater or equal than 4. Then we define some sets related to the previous four mappings:

$$R'(w) = \begin{cases} \{\theta(u), \sigma_l(u), \rho_l(u))\}, \text{ if } \mathrm{Pref}_4(w) = uu\overline{u}u, \\ \{\theta(u), \sigma_l(u)\}, \text{ otherwise;} \end{cases}$$

$$S'(w) = \begin{cases} \{\theta(v), \sigma_r(v), \rho_r(v)\}, \text{ if } \mathrm{Suff}_4(w) = v\overline{v}vv, \\ \{\theta(v), \sigma_r(v)\}, \text{ otherwise;} \end{cases}$$

$$R(w) = R'(w) \times S'(w);$$

$$S(w) = (R'(w) \setminus \{\theta(u)\}) \times (S'(w) \setminus \{\theta(v)\}) \setminus \{(\rho_l(u), \rho_r(v))\}.$$

For the structural presentation we define the sets $\mathcal{LAWSF}_{base}$ and $\mathcal{LWSF}_{base}$ which contains the words presented in Tables II.1.1 and II.1.2, respectively.

We also introduce the words a_n^+ and b_n^+, called as *Morse-blocks*, which are obtained by applying a power of the Thue-Morse morphism:

$$u_n^+ = \theta^n(u), \quad \text{where} \quad u \in A \text{ and } , n \geq 1.$$

For these words we use also the notations a_n^- and b_n^- in such a way that

$$a_n^- = \begin{cases} a_n^+, & \text{if } 2 \not| n, \\ b_n^+, & \text{if } 2|n \end{cases}$$

and

$$b_n^- = \begin{cases} b_n^+, & \text{if } 2 \not| n, \\ a_n^+, & \text{if } 2|n \end{cases}.$$

In the above equations n is an *index* of a Morse-block, and a^+, a^-, b^+ and b^- are called *stamps* of Morse-blocks. As we see from the previous equations, there is allways two possible stamps, with different upper indexes, for a given Morse-block.

$ababa$	$babab$	$aabaaba$
$bbabbab$	$abaabaa$	$babbabb$
$baabaab$		$abbabba$

Table II.1.1 The words in $\mathcal{LAWSF}_{base}$.

$a.ab.a$	$a.ab.b$	$a.ba.a$	$a.ba.b$
$b.ba.b$	$b.ba.a$	$b.ab.b$	$b.ab.a$
	$ab.ba$	$ba.ab$	
$a.ab.aa$	$a.ab.ab$	$a.ab.ba$	$a.ba.ab$
$ab.ab.b$	$ab.ba.a$	$ab.ba.b$	
$b.ba.bb$	$b.ba.ba$	$b.ba.ab$	$b.ab.ba$
$ba.ba.a$	$ba.ab.b$	$ba.ab.a$	
$aa.ba.ab$	$ab.ab.ba$	$ab.ba.ba$	$ab.ba.bb$
$ab.ba.ab$	$bb.ab.ba$	$ba.ba.ab$	$ba.ab.ab$
	$ba.ab.aa$	$ba.ab.ba$	

Table II.1.2. The words in $\mathcal{LWSF}_{base}$.

II.2 Preliminary considerations

II.2.1 Block-structure of 2^+-free words

In this section we are going to show that the 2^+-free words have a specific kind of block-structure. With that specific kind of block-structure of the 2^+-free words we can express a way to construct a given 2^+-free word from a shorter 2^+-free word.

Firstly, in order to show that every 2^+-free word has a block-structure we want, we have to consider the 2^+-free words with length less than nine. Those words are listed in the following three tables. In Tables II.2.2 and II.2.3 block-structures are also presented different blocks being separated by periods.

	a	b
	aa	ab
	bb	ba
aab	aba	abb
bba	bab	baa

Table II.2.1. The 2^+-free words with length less than 4.

In the following lemma we formulate the block-structure of 2^+-free words which we need. This lemma has also been given independently (e.g. Restivo and Salemi [RS] and Kobayashi [Ko2].)

Remark The periods in Tables II.2.2 and II.2.3 denote the bounds of s_i-blocks in Lemma II.2.1. Actually, there is several ways to divide some words in Tables II.2.1 and II.2.2 into blocks satisfying the conditions of this lemma. As an example, the word $aabab$ can be divided in two ways: $a.ab.ab$ and $aa.ba.b$.

$a.ab.a$	$a.ab.b$	$a.ba.a$	$a.ba.b$
$b.ba.b$	$b.ba.a$	$b.ab.b$	$b.ab.a$
	$ab.ba$	$ba.ab$	

$a.ab.aa$	$a.ab.ab$	$a.ab.ba$	$a.ba.ab$
$ab.ab.b$	$ab.ba.a$	$ab.ba.b$	
$b.ba.bb$	$b.ba.ba$	$b.ba.ab$	$b.ab.ba$
$ba.ba.a$	$ba.ab.b$	$ba.ab.a$	
$aa.ba.ab$	$a.ab.ab.b$	$a.ab.ba.a$	$a.ab.ba.b$
$a.ba.ab.a$	$a.ba.ab.b$	$ab.ab.ba$	$ab.ba.ba$
	$ab.ba.bb$	$ab.ba.ab$	
$bb.ab.ba$	$b.ba.ba.a$	$b.ba.ab.b$	$b.ba.ab.a$
$b.ab.ba.b$	$b.ab.ba.a$	$ba.ba.ab$	$ba.ab.ab$
	$ba.ab.aa$	$ba.ab.ba$	

Table II.2.2. The 2^+-free words with length 4,5 or 6.

$aa.ba.ab.b$	$a.ab.ab.ba$	$a.ab.ba.ab$	$a.ab.ba.bb$
$a.ab.ba.ba$	$a.ba.ab.ab$	$a.ba.ab.ba$	$ab.ab.ba.a$
$ab.ab.ba.b$	$ab.ba.ba.a$	$ab.ba.ab.a$	$ab.ba.ab.b$
$bb.ab.ba.a$	$b.ba.ba.ab$	$b.ba.ab.ba$	$b.ba.ab.aa$
$b.ba.ab.ab$	$b.ab.ba.ba$	$b.ab.ba.ab$	$ba.ba.ab.b$
$ba.ba.ab.a$	$ba.ab.ab.b$	$ba.ab.ba.b$	$ba.ab.ba.a$
$aa.ba.ab.ba$	$a.ab.ab.ba.a$	$a.ab.ab.ba.b$	$a.ab.ba.ab.a$
$a.ab.ba.ab.b$	$a.ab.ba.ba.a$	$a.ba.ab.ab.b$	$a.ba.ab.ba.a$
$a.ba.ab.ba.b$	$ab.ab.ba.ab$	$ab.ab.ba.ba$	$ab.ba.ba.ab$
$ab.ba.ab.aa$	$ab.ba.ab.ab$	$ab.ba.ab.ba$	
$bb.ab.ba.ab$	$b.ba.ba.ab.b$	$b.ba.ba.ab.a$	$b.ba.ab.ba.b$
$b.ba.ab.ba.a$	$b.ba.ab.ab.b$	$b.ab.ba.ba.a$	$b.ab.ba.ab.b$
$b.ab.ba.ab.a$	$ba.ba.ab.ba$	$ba.ba.ab.ab$	$ba.ab.ab.ba$
$ba.ab.ba.bb$	$ba.ab.ba.ba$	$ba.ab.ba.ab$	

Table II.2.3. The 2^+-free words with length 7 and 8.

Lemma II.2.1 (Factorization Lemma) *Every 2^+-free word w over the binary alphabet $A = \{a, b\}$ can be expressed in the form*

$$r_1 r_2 s_1 \ldots s_n r_3 r_4 \ , \tag{1}$$

or in the form

$$r_1r_2r_3r_4\ , \tag{2}$$

where

$$r_1r_2, r_3r_4 \in \{\varepsilon, a, b, aa, bb\} \tag{3}$$

and

$$s_i \in \{ab, ba\} \qquad (4),$$

for all $1 \leq i \leq n$*. Moreover, the above expression is unique for every* 2^+*-free word with length greater or equal than 7.*

In the proof of this lemma we will need a little more information than is delivered by the previous three tables. That additional information we need tells us that every 2^+-free word in Table II.2.3 has only one way to be divided into blocks such that the conditions at the lemma are satisfied. That can be easily verified by moving periods one step to the right in a given word from Table II.2.3 which results at least one inner block in that word with content aa or bb.

Our last remark here lists the necessary choices for the words r_1r_2 and r_3r_4. Because the block-structures are unique, for the words in Table II.2.3, we must have at least the following five alternatives for r_1r_2 and r_3r_4:

$$\varepsilon, a, b, aa, bb.$$

Proof We carry out the proof inductively with respect to the length of 2^+-free words.

In the cases $k \leq 8$ the lemma holds by Tables II.2.1, II.2.2 and II.2.3.

Next we assume that $k \geq 8$ and that every 2^+-free word having length less or equal to k can be expressed in the way stated in the lemma. So, we have to prove that every 2^+-free word having the length $k+1$ also can be expressed in the similar way.

Let w a 2^+-free word with length $k+1$. Then the length of that word is at least nine. Let t and v be the leftmost and the rightmost letters of the word w, respectively. Then we have $w = tuv$ for some word u. On the other hand, the words tu and uv are 2^+-free words and $|uv| = k = |tu|$, so there exists exactly one block-structure for both of these words. Moreover, these words have also common part u, which has length $k - 1 \geq 7$. Now, because $|u| \geq 7$, those block-structures of 2^+-free words tu and uv must coincide in the common part u. Finally, the condition (3) can be confirmed by the fact that the words tu and uv are a prefix and a suffix of the word w, respectively, and, by our assumption their block-structure satisfies the condition (4).

Thus, there exists exactly one block-structure for the word w satisfying the conditions of Lemma II.2.1 and that block-structure is the one obtained from the block-structure of the word u. So every 2^+-free word having length $k+1$ can be expressed in the form introduced in Lemma II.2.1. Now, by the induction we conclude from what we proved above that the lemma holds for every 2^+-free word.

$\square$

We have now determined a condition which holds for every 2^+-free word. While this condition is not powerfull enough to make sure that a given word is 2^+-free, it still gives us a hint what we should next be looking for.

II.2.2 Types of the 2^+-free words

Next we are going to introduce a necessary and a sufficient condition for the 2^+-freeness of a binary word. That condition will be in a recursive form and, while it does not give us very clear view about 2^+-free words, it is very useful when we are considering the number of the 2^+-free words having a given length. A more explicit form of the condition characterizing the 2^+-free words will be introduced later in this paper.

The first step in our way to formalize a recursive characterization for the 2^+-free words is to study the easiest case in Lemma II.2.1. Namely, we study those 2^+-free words satisfying Lemma II.2.1 with $r_1r_2 = \varepsilon$ and $r_3r_4 = \varepsilon$. In other words, we shall consider those words which are of the form

$$s_1s_2\ldots s_n,$$

where $s_i \in \{ab, ba\}$, for all $1 \leq i \leq n$. But, first we shall introduce the following two lemmas which we will use to clarify our considerations.

Lemma II.2.2 *Let $n > 1$ be an integer and $s_1, \ldots, s_n$ be arbitrary binary words over the alphabet A such that*

$$s_i \in \{ab, ba\}, \text{ for all } 1 \leq i \leq n.$$

Additionally, let the letters $t_1, t_2, \ldots, t_n \in A$ be chosen such that

$$s_i = \theta(t_i), \text{ for all } 1 \leq i \leq n, \tag{5}$$

and

$$t_1t_2\ldots t_n \in \mathcal{LWSF}. \tag{6}$$

Then the equations

$$ururu = s_1\ldots s_{n-1}t_n \tag{7}$$

and

$$ururu = \overline{t_1}s_2\ldots s_n \tag{7'}$$

both fails for all words $u \in A$ and $r \in A^*$.

Proof In the case $n = 1$, the length of the left side of equations (7) and (7′) is at least three and the length of the right side is one. Therefore, it is sufficient to study the cases with n greater than one.

Let us first assume that the equation (7) holds. By the equation (5) we can write this equation in the form

$$ururu = t_1\overline{t_1}t_2 \ldots t_{n-1}\overline{t_{n-1}}t_n.$$

If n is odd, we obtain the equation

$$\begin{cases} ur & = t_1\overline{t_1} \ldots t_{\frac{n+1}{2}-1}\overline{t_{\frac{n+1}{2}-1}} \\ & = t_{\frac{n+1}{2}}\overline{t_{\frac{n+1}{2}}} \ldots t_{n-1}\overline{t_{n-1}}. \end{cases}$$

By removing the complements from this equation we obtain the equation

$$t_1 \ldots t_{\frac{n+1}{2}-1} = t_{\frac{n+1}{2}} \ldots t_{n-1},$$

which yields

$$t_1t_2 \ldots t_n \notin \mathcal{LWSF},$$

because $t_1 = t_n$ by the equation (7). But this contradicts with the formula (6). Thus, we can assume that n is even. In this case we obtain from the equation (7) the following:

$$\begin{cases} ur & = t_1\overline{t_1} \ldots t_{\frac{n}{2}} \\ & = \overline{t_{\frac{n}{2}}} \ldots t_{n-1}\overline{t_{n-1}}. \end{cases}$$

From this equivality we obtain

$$t_{\frac{n}{2}+i} = \overline{t_i} = t_{\frac{n}{2}}, \text{ for all } 1 \leq i \leq \frac{n}{2} - 1.$$

Hence we conclude that

$$t_i = t_{\frac{n}{2}} = \overline{t_{\frac{n}{2}} + i}, \text{ for all } 1 \leq i \leq \frac{n}{2} - 1,$$

and therefore

$$\begin{cases} ur & = (t_1\overline{t_1})^{\frac{n-2}{2}}t_1 \\ & = (\overline{t_1}t_1)^{\frac{n-2}{2}}\overline{t_1}. \end{cases} \tag{8}$$

But this cannot be true for $n \geq 2$. Thus, the formula (7) cannot hold for $n \geq 1$.

Next we assume that the formula (7′) holds. In the case n is odd, we obtain by the equation (5) the equation

$$\begin{cases} ru & = t_2\overline{t_2}\ldots t_{\frac{n+1}{2}}\overline{t_{\frac{n+1}{2}}} \\ & = t_{\frac{n+1}{2}+1}\overline{t_{\frac{n+1}{2}}+1}\ldots t_n\overline{t_n} \end{cases}.$$

Again, by removing complements and applying the equation $t_1 = t_n$ obtained from the equation (7′), we determine that

$$\overline{t_1}\ldots\overline{t_n} \notin \mathcal{LWSF}.$$

As before, we can now assume that n is even. By applying the equation (5), we obtain

$$\begin{cases} ur & = \overline{t_1}t_2\ldots\overline{t_{\frac{n}{2}}} \\ & = t_{\frac{n}{2}+1}\overline{t_{\frac{n}{2}+1}}\ldots t_n \end{cases}.$$

From this equation we determine that

$$t_{\frac{n}{2}+i} = t_1 = t_i, \text{ for all } 1 \leq i \leq \frac{n}{2}.$$

By this result we obtain the equation (8), which cannot hold. Hence the equation (7′) cannot be satisfied when the equation (6) is assumed. Finally, we conclude that our original assumption must be wrong. This completes our proof. □

Lemma II.2.3 *Let $n > 1$ be an integer and $s_1, \ldots, s_n$ be arbitrary binary words over the alphabet A such that*

$$s_i \in \{ab, ba\}, \text{ for all } 1 \leq i \leq n.$$

Additionally, let the letters $t_1, t_2, \ldots, t_n \in A$ be chosen such that

$$s_i = \theta(t_i), \text{ for all } 1 \leq i \leq n,$$

and

$$t_1t_2\ldots t_n \in \mathcal{LWSF}.$$

Let $u \in A$ and $r \in A^*$ be chosen such that the equation

$$uus_1\ldots s_{n-1}t_n = ururu$$

holds and

$$uus_1\ldots s_{n-1} \in \mathcal{LWSF}.$$

Then $n = 3$ and

$$uus_1 \dots s_2 t_3 = uu\overline{u}uu\overline{u}u.$$

Similarly, the only possible solution for the equation

$$t_1 s_2 \dots s_n uu = ururu$$

such that

$$s_2 \dots s_n uu \in \mathcal{LWSF}$$

is the word $u\overline{u}uu\overline{u}uu$.

Proof From the equation

$$uus_1 \dots s_{n-1} t_n = ururu,$$

we can see that the word $uus_1 \dots s_{n-1}$ is a square. Now, if the integer n is odd and greater than three, we obtain the equation

$$uus_1 \dots s_{\frac{n-1}{2}-1} t_{\frac{n-1}{2}} = \overline{t_{\frac{n-1}{2}}} s_{\frac{n-1}{2}+1} \dots s_{n-1}.$$

From this equation we determine that

$$t_1 = \overline{t_{\frac{n-1}{2}+1}} = \overline{u}$$

$$\overline{t_{\frac{n-1}{2}}} = u$$

and

$$t_i = \overline{t_{\frac{n-1}{2}+i}} = \overline{\overline{t_{i+1}}} = t_{i+1}, \text{ for all } 1 \leq i \leq \frac{n-1}{2}.$$

This equation yields the equation

$$uus_1 \dots s_{n-1} = u(u\overline{u})^{\frac{n-1}{2}} u(u\overline{u})^{\frac{n-1}{2}}.$$

And, because $n > 3$, the word $u\overline{u}u\overline{u}u$ is a subword of this word. On the other hand, we assumed that

$$uus_1 \dots s_{n-1} \in \mathcal{LWSF},$$

so we have a contradiction.

In the case $n = 3$, the only possible word satisfying the conditions of this lemma is

$$uus_1 s_2 t_3 = uu\overline{u}uu\overline{u}u,$$

as we can see from the previous considerations.

On the other hand, because $s_i \notin \{aa, bb\}$ for any $i \in \{1, \dots n-2\}$, even values of n are out of the question. Therefore the only possible solution for the word $uus_1 \dots s_2 t_3$ is that given in the lemma.

The latter part of the lemma is easy to prove by taking mirror images of given words and then applying the former part of the proof. $\square$

Next we continue our study of the 2^+-free words of the form $s_1 \dots s_n$.

Lemma II.2.4 *Let n be a positive integer and $s_1, \dots, s_n$ be arbitrary binary words over the alphabet A such that*

$$s_i \in \{ab, ba\}, \text{ for all } 1 \leq i \leq n.$$

Additionally, let the letters $t_1, t_2, \dots, t_n \in A$ be chosen such that

$$s_i = \theta(t_i), \text{ for all } 1 \leq i \leq n.$$

Then the following three properties are equivalent:

$$\begin{array}{ll} i) & s_1 \dots s_n \in \mathcal{LWSF}, \\ ii) & t_1 \dots t_n \in \mathcal{LWSF}, \\ iii) & \overline{t_1} \dots \overline{t_n} \in \mathcal{LWSF}. \end{array}$$

Proof First we are going to prove that the property i) yields the property ii). In order to do that we assume that

$$s = s_1 \dots s_n \in \mathcal{LWSF},$$

$$t_1 \dots t_n \notin \mathcal{LWSF}$$

and

$$t_i \overline{t_i} = s_i, \text{ for all } 1 \leq i \leq n.$$

Then, by definition of the 2^+-freeness, there exists positive integers h and i such that

$$2h - i \leq n$$

and

$$t_{i+j} = t_{h+j},$$

where $j \in \{0, 1, \dots, h - i\}$. From this we conclude that the the word

$$s = w_1 w_2 \dots w_{2n} = t_1 \overline{t_1} t_2 \overline{t_2} \dots t_n \overline{t_n},$$

where $w_1, \dots, w_{2n} \in A$, satisfies the equation

$$w_{2i+j} = w_{2h+j},$$

for the values $j = 0, 1, \dots, 2h - 2i + 1$. On the other hand, we assumed that

$$s \in \mathcal{LWSF},$$

so we have a contradiction. Thus the property i) yields the property ii).

Next we prove that the property ii) yields the property i). First we assume that the formulas

$$s = s_1 \dots s_n \notin \mathcal{LWSF}$$

and

$$t_1 \dots t_n \in \mathcal{LWSF}$$

holds with the equation (5) above. Then, by the definition of the 2^+-freeness, at least one of the formulas

$$t_i \overline{t_i} \dots t_{h-1} \overline{t_{h-1}} t_h \notin \mathcal{LWSF}$$

and

$$\overline{t_i} t_{i+1} \overline{t_{i+1}} \dots t_h \overline{t_h} \notin \mathcal{LWSF}$$

holds for some positive integers i and h such that $i < h$ and the value $2h - 2i + 1$ is as small as possible. Then that word satisfying this specific formula is of the form $uvuvu$, where $u \in A$ and $v \in A^+$. Now, because $h - i > 0$, it is possible to apply Lemma II.2.2. But Lemma II.2.2 states that there cannot exist any word satisfying the previous conditions.

Now we have already proved that the properties i) and ii) are equivalent. On the other hand, because the words $t_1 \dots t_n$ and $\overline{t_1} \dots \overline{t_n}$ are each others complements, the properties ii) and iii) are also equivalent. And that completes the proof. □

In the Lemma II.2.4 we considered only those words which satisfies Lemma II.2.1 with restrictions $r_1 r_2 = \varepsilon$ and $r_3 r_4 = \varepsilon$. Next we are going to generalize Lemma II.2.4.

In the general case of Lemma II.2.4 we have to consider nine different situations which are:

I	$\{ab, ba\}^n$
II	$\{a, b\}\{ab, ba\}^n$
III	$\{ab, ba\}^n\{a, b\}$
IV	$\{a, b\}\{ab, ba\}^n\{a, b\}$
V	$\{aa, bb\}\{ab, ba\}^n$
VI	$\{ab, ba\}^n\{aa, bb\}$
VII	$\{a, b\}\{ab, ba\}^n\{aa, bb\}$
VIII	$\{aa, bb\}\{ab, ba\}^n\{a, b\}$
IX	$\{aa, bb\}\{ab, ba\}^n\{aa, bb\}$

We do the generalizations concerning these nine situations in several parts. In the first part we consider the cases I, II, III and IV, where the equivalence is very similar with the equivalence in Lemma II.2.4 except in the case IV where it is possible to obtain a 2^+-free word from a word which is not a 2^+-free. For the cases V,VI,VII,VIII and IX there will be some additional difficulties.

II.2.3 A recursive characterization of the 2^+-free words

In this section our goal is to settle two theorems stating a recursive way to define the language of the 2^+-free words. Actually, one of these Theorems characterize the language of the allmost 2^+-free words, which is closely related to the language of the 2^+-free words.

The following three propositions gives us a recursive method to generate an arbitrary 2^+-free word from a 2^+-free word having length less than 9.

Proposition II.2.5 *Let $n \geq 1$, u, v be letters of the alphabet A and $s_1, \ldots, s_n$ be arbitrary binary words over the alphabet A such that*

$$s_i \in \{ab, ba\}, \text{ for all } 1 \leq i \leq n.$$

Additionally, let the letters $t_1, t_2, \ldots, t_n \in A$ be chosen such that

$$s_i = \theta(t_i), \text{ for all } 1 \leq i \leq n.$$

Then in each of the following five cases the properties listed are equivalent:

Case I:

I.1) $s_1 \ldots s_n \in \mathcal{LWSF}$
I.2) $t_1 \ldots t_n \in \mathcal{LWSF}$

Case II:

II.1) $us_1 \ldots s_n \in \mathcal{LWSF}$
II.2) $\overline{u}t_1 \ldots t_n \in \mathcal{LWSF}$

Case III:

III.1) $s_1 \ldots s_n v \in \mathcal{LWSF}$
III.2) $t_1 \ldots t_n v \in \mathcal{LWSF}$

Case IVa:

IVa.1) $us_1 \ldots s_n v \in \mathcal{LWSF}$ *is not a square*
IVa.2) $\overline{u}t_1 \ldots t_n v \in \mathcal{LWSF}$

Case IVb:

IVb.1) $us_1 \ldots s_n v \in \mathcal{LWSF}$ *is a square*
IVb.2) $\overline{u}t_1 \ldots t_n v \in \mathcal{LAWSF}$.

Proof Proof of the case I:

The equivalence between properties I.1 and I.2 was proved in Lemma II.2.4.

Proof of the case II:

Firstly, in the case $n = 1$, the words us_1 and $\overline{u}t_1$ both are allways 2^+-free word. Now, because the conditions II.1 and II.2 are equivalent when $n = 1$, we assume that $n \geq 2$.

Let us assume that the property II.2 holds, i.e.

$$\overline{u}t_1 \dots t_n \in \mathcal{LWSF}.$$

Then, by the definitions of $s_1, t_1, \dots, s_n, t_n$ and the case I, we obtain

$$\overline{u}us_1 \dots s_n \in \mathcal{LWSF}.$$

And, because every subword of 2^+-free word is also a 2^+-free word, we obtain the property II.1.

Next we assume that the property II.1 and the logical complement of the property II.2 are true. Formally, we express that in the following way:

$$us_1 \dots s_n \in \mathcal{LWSF} \tag{9}$$

and

$$\overline{u}t_1 \dots t_n \notin \mathcal{LWSF}.$$

By the case I, we obtain from the equation (9) that

$$t_1 \dots t_n \in \mathcal{LWSF}.$$

From these two last equations we conclude that there exists a prefix of the word $\overline{u}t_1 \dots t_n$ of the form $uvuvu$ and

$$\overline{u}us_1 \dots s_n \notin \mathcal{LWSF}.$$

So, there exists a positive integer $h \leq n/2$ such that

$$t_1 \dots t_h = t_{h+1} \dots t_{2h},$$

where $t_h = \overline{u} = t_{2h}$. But this yields

$$us_1 \dots s_{2h} \notin \mathcal{LWSF},$$

because $s_{2h} = \theta(t_{2h}) = \overline{u}u$. On the other hand, we assumed that the equation (9) holds, so we have a contradiction. Therefore, we obtain from the equation (9) the equation

$$\overline{u}t_1 \dots t_n \in \mathcal{LWSF}.$$

So, the property II.1 yields the property II.2 by the case I.

Now we have proved the case II.

Proof of the case III:

With similar argument as in the proof of the case II, we can assume that $n \geq 2$.

The proof from the property III.2 to the property III.1 is similar with the corresponding part of the proof of the case II:

From the condition

$$t_1 \dots t_n v \in \mathcal{LWSF}$$

we obtain

$$s_1 \dots s_n v \overline{v} \in \mathcal{LWSF}$$

and this yields the property III.1.

The proof to the other direction is also quite straightforward. From the property III.1 and the formula

$$s_1 \dots s_n v \overline{v} \notin \mathcal{LWSF}$$

we conclude, as in the proof of the case II, that there exists a positive integer $h \leq n/2$ such that

$$t_{n-2h+1} \dots t_{n-h} = t_{n-h+1} \dots t_n,$$

where $t_{n-2h+1} = v = t_{n-h+1}$. But this yields

$$t_1 \dots t_n v \notin \mathcal{LWSF},$$

so we conclude that the property III.1 yields the formula

$$s_1 \dots s_n v \overline{v} \in \mathcal{LWSF}.$$

And from this we obtain the property III.2 by the case I.

This completes the proof of the case III.

Proof of the case IVa:

First we assume that the property IVa.1 and the formula

$$\overline{u} t_1 \dots t_n v \notin \mathcal{LWSF}$$

holds. By the cases II and III, we obtain from the property IVa.1 the formulas

$$\overline{u} t_1 \dots t_n \in \mathcal{LWSF}$$

and

$$t_1 \dots t_n v \in \mathcal{LWSF}.$$

Thus, the word $\overline{u}t_1 \dots t_n v$ has to be of the form $vrvrv$, for some $r \in A^+$. From this we see that $u = \overline{v}$, so we the word $s_1 \dots s_n vu$ is a square, which contradicts our assumption. Thus, the property IVa.1 yields the property IVa.2.

If the property IVa.2 holds, the case I shows that

$$vus_1 \dots s_n vu \in \mathcal{LWSF},$$

which quarantees that $us_1 \dots s_n v$ is not a square. Otherwise, the words $us_1 \dots s_n vu$ and $vus_1 \dots s_n v$ will not be 2^+-free words. So, the property IVa.1 holds.

This completes the proof of the case IVa.

Proof of the case IVb:

Let us start with the property IVb.1. Then also the word $s_1 \dots s_n vu$ is a square. By assuming that n is even and carrying similar considerations as in the proof of Lemma II.2.2, we obtain result similar to the equation (8). So, we can assume that n is odd, yielding

$$s_1 \dots s_{\frac{n-1}{2}} = s_{\frac{n+1}{2}} \dots s_n vu.$$

From this equation we see that

$$t_1 \dots t_{\frac{n-1}{2}} = t_{\frac{n+1}{2}} \dots t_n v$$

and

$$s_{\frac{n-1}{2}} = vu,$$

which yields $u = \overline{v}$, and so, the word $t_1 \dots t_n v$ is a square, as is the word $vt_1 \dots t_n$, and so $\overline{u}t_1 \dots t_n v$ is not 2^+-free word. Now, by the cases I and II, we see that if $t_1 \dots t_n v$, or $vt_1 \dots t_n$, is not a 2^+-free word, then also the word $s_1 \dots s_n v$, or $us_1 \dots t_n$, respectively, is not a 2^+-free word. On the other hand, this contradicts with the property IVb.1. So, the property IVb.1 implies the property IVb.2.

In order to prove the opposite direction we assume that

$$\overline{u}t_1 \dots t_n v \in \mathcal{LAWSF}.$$

From this assumption we conclude that $\overline{u} = v$ and $t_1 \dots t_n v \in \mathcal{LWSF}$ is a square.

Then also the words

$$s_1 \dots s_n v\overline{v} = s_1 \dots s_n vu \in \mathcal{LWSF}$$

and

$$us_1 \dots s_n v \in \mathcal{LWSF}$$

are squares, completing our proof. $\square$

In the following proposition we consider the cases V,VI. This proposition differs essentially from the previous one. In the cases V and VI we actually change some letters which are obtained from a 2^+-free word by the Thue-Morse morphism.

Proposition II.2.6 *Let $n \geq 3$, u, v be letters of the alphabet A and $s_1, \ldots, s_n$ be arbitrary binary words over the alphabet A such that*

$$s_i \in \{ab, ba\}, \text{ for all } 1 \leq i \leq n.$$

Additionally, let the letters $t_1, t_2, \ldots, t_n \in A$ be chosen such that

$$s_i = \theta(t_i), \text{ for all } 1 \leq i \leq n,$$

Then in each of the following two cases the properties listed are equivalent:

Case V:

V.1) $uus_1 \ldots s_n \in \mathcal{LWSF}$

V.2) $\overline{u}t_1 \ldots t_n \in \mathcal{LWSF}$ *and* $\mathrm{Pref}_4(\overline{u}t_1 \ldots t_n) \in F_l$

Case VI:

VI.1) $s_1 \ldots s_n vv \in \mathcal{LWSF}$

VI.2) $t_1 \ldots t_n v \in \mathcal{LWSF}$ *and* $\mathrm{Suff}_4(t_1 \ldots t_n v) \in F_r$

Proof Proof of the case V:

Let us first consider words which are not 2^+-free. In such a word there must be a subword of the form $uvuvu$, where $u \in A$ and $v \in A^*$. Thus every word, which is not a 2^+-free word, contains at least one subword having odd length, which is not a 2^+-free word. Therefore, the formula

$$uus_1 \ldots s_n \notin \mathcal{LWSF}$$

can be expressed in the form

$$\begin{cases} us_1 \ldots s_n \notin \mathcal{LWSF} \text{ or} \\ uut_1 = uuu \text{ and } n = 1 \quad \text{ or} \\ uus_1 \ldots s_{h-1}t_h = ururu \text{ and } uus_1 \ldots s_{h-1} \in \mathcal{LWSF}, \\ \qquad \text{for some } r \in A^* \text{ and } h > 1. \end{cases} \tag{10}$$

Here the word $ururu$ must be of the form $uus_1 \ldots s_{h-1}t_h$, because $|ururu|$ is odd. Now, by Lemma II.2.3 and the case II of Proposition II.2.5, the formulas in (10) implies the formulas

$$\begin{cases} \overline{u}t_1 \ldots t_n \notin \mathcal{LWSF} \quad \text{ or} \\ t_1 = u \text{ and } n = 1 \quad \text{ or} \\ t_1t_2t_3 = \overline{u}uu \text{ and } n = 3. \end{cases}$$

On the other hand, by the case II of Proposition II.2.5, the equality

$$t_1t_2t_3 = \overline{u}uu$$

yields

$$uus_1s_2t_3 = ururu, \text{ with } r = u\overline{u},$$

and

$$uus_1s_2 \in \mathcal{LWSF}.$$

Furthermore, Lemma II.2.3 guaranties that there does not exist other possible solutions for that last formula in (10). Thus,

$$t_1t_2t_3 = \overline{u}uu$$

yields

$$uus_1 \ldots s_{h-1}t_h = ururu \text{ and } uus_1 \ldots s_{h-1} \in \mathcal{LWSF},$$

for some $r \in A^*$ and $h > 1$. So, the formula

$$uus_1 \ldots s_n \notin \mathcal{LWSF}$$

is equivalent with the formula

$$\begin{cases} \overline{u}t_1 \ldots t_n \notin \mathcal{LWSF} \quad \text{or} \\ \overline{u}t_1 = \overline{u}u \text{ and } n = 1 \quad \text{or} \\ \overline{u}t_1t_2t_3 = \overline{u}\overline{u}uu \text{ and } n = 3 \end{cases}.$$

Finally, by taking logical complements of these two equivalent formulas, we obtain the equivalence between the properties V.1 and V.2.

Proof of the case VI:

As in the previous case, the formula

$$s_1 \ldots s_nvv \notin \mathcal{LWSF}$$

can be expressed in the form

$$\begin{cases} s_1 \ldots s_nv \notin \mathcal{LWSF} \quad \text{or} \\ \overline{t_n}vv = vvv \text{ and } n = 1 \quad \text{or} \\ \overline{t_h}s_{h+1} \ldots s_nvv = vrvrv \text{ and } s_{h+1} \ldots s_nvv \in \mathcal{LWSF}, \\ \qquad \text{for some } r \in A^* \text{ and } h < n. \end{cases} \tag{11}$$

By Lemma II.2.3 and the case III of Proposition II.2.5, the formula (11) can be written to the form

$$\begin{cases} t_1 \ldots t_nv \notin \mathcal{LWSF} \quad \text{or} \\ t_n = \overline{v} \text{ and } n = 1 \quad \text{or} \\ t_1t_2t_3 = \overline{v}\overline{v}v \text{ and } n = 3. \end{cases}$$

On the other hand, by the case III of Proposition II.2.5, the equality

$$t_1t_2t_3 = \overline{v}\overline{v}v$$

yields

$$\overline{t_1}s_2s_3vv = vrvrv, \text{ with } r = \overline{v}v$$

and

$$s_2s_3vv \in \mathcal{LWSF}.$$

Furthermore, Lemma II.2.3 guaranties that there does not exist other possible solutions for that last formula in (11). Thus,

$$t_1t_2t_3 = \overline{v}\overline{v}v$$

yields

$$\overline{t_h}s_{h+1}\ldots s_nvv = vrvrv \text{ and } s_{h+1}\ldots s_nvv \in \mathcal{LWSF},$$

for some $r \in A^*$ and $h < n$. As in the previous case, we obtain equivalence between the formulas

$$s_1\ldots s_nvv \notin \mathcal{LWSF}$$

and

$$\begin{cases} s_1\ldots s_nv \notin \mathcal{LWSF} & \text{or} \\ t_1v = \overline{v}v \text{ and } n = 1 & \text{or} \\ t_1t_2t_3v = \overline{v}\overline{v}vv \text{ and } n = 3 & \end{cases}.$$

So, we conclude that the properties VI.1 and VI.2 are equivalent. □

Next we are going to handle the remaining three cases not yet considered. The following proposition gives us some conditions which must be satisfied by 2^+-free words described by the above three cases. As the case IV had two subcases, the cases VII, VIII and IX have also two similar subcases.

Proposition II.2.7 *Let $n \geq 2$, u, v be letters of the alphabet A and $s_1, \ldots, s_n$ be arbitrary binary words over the alphabet A such that*

$$s_i \in \{ab, ba\}, \text{ for all } 1 \leq i \leq n.$$

Additionally, let the letters $t_1, t_2, \ldots, t_n \in A$ be chosen such that

$$s_i = \theta(t_i), \text{ for all } 1 \leq i \leq n,$$

and denote the word $\overline{u}t_1\ldots t_nv$ by w. Then in each of the following six subcases the properties listed are equivalent:

Case VIIa:

VIIa.1) $uus_1 \dots s_n v \in \mathcal{LWSF}$ *and* $us_1 \dots s_n v$ *is not a square*

VIIa.2) $w \in \mathcal{LWSF}$ *and* $\mathrm{Pref}_4(w) \in F_l$

Case VIIb:

VIIb.1) $uus_1 \dots s_n v \in \mathcal{LWSF}$ *and* $us_1 \dots s_n v$ *is a square*

VIIb.2) $w \in \mathcal{LAWSF}$ *and* $\mathrm{Pref}_4(w) \in F_l$

Case VIIIa:

VIIIa.1) $us_1 \dots s_n vv \in \mathcal{LWSF}$ *and* $us_1 \dots s_n v$ *is not a square*

VIIIa.2) $w \in \mathcal{LWSF}$ *and* $\mathrm{Suff}_4(w) \in F_r$

Case VIIIb:

VIIIb.1) $us_1 \dots s_n vv \in \mathcal{LWSF}$ *and* $us_1 \dots s_n v$ *is a square*

VIIIb.2) $w \in \mathcal{LAWSF}$ *and* $\mathrm{Suff}_4(w) \in F_r$

Case IXa:

IXa.1) $uus_1 \dots s_n vv \in \mathcal{LWSF}$ *and* $us_1 \dots s_n v$ *is not a square*

IXa.2) $w \in \mathcal{LWSF}$, $\mathrm{Pref}_4(w) \in F_l$ *and* $\mathrm{Suff}_4(w) \in F_r$

Case IXb:

IXb.1) $uus_1 \dots s_n vv \in \mathcal{LWSF}$ *and* $us_1 \dots s_n v$ *is a square*

IXb.2) $w \in \mathcal{LAWSF}$, $\mathrm{Pref}_4(w) \in F_l$ *and* $\mathrm{Suff}_4(w) \in F_r$

Proof Proof of the case VIIa:

Let us first assume that the property VIIa.1 holds. Then, by the cases IVa of Proposition II.2.5 and V of Proposition II.2.6, we obtain

$$\overline{u}t_1 \dots t_n v \in \mathcal{LWSF},$$

and

$$\overline{u}t_1 t_2 t_3 \in \{aaba, bbab\},$$

which are the conditions in the property VIIa.2.

Next we suppose that the conditions in the property VIIa.2 holds. Then, by the cases IVa and V of the previous proposition, the word $us_1 \dots s_n v \in \mathcal{LWSF}$ is not a square and the word $uus_1 \dots s_n$ is a 2^+-free word. Moreover, if we assume that the word $uus_1 \dots s_n v$ is not a 2^+-free word, the word $uus_1 \dots s_n v$ must be of the form $ururu$, where $u \in A^*$ and $urur \in \mathcal{LWSF}$. But, by Lemma II.2.3, there is no word satisfying the condition $urur \in \mathcal{LWSF}$ for the values $n \geq 3$. By this contradiction and the above consideration, the properties VIIa.1 and VIIa.2 are equivalent.

Proof of the case VIIb:

By the cases IVb of Proposition II.2.5 and V of Proposition II.2.6, the property VIIb.1 yields the property VIIb.2. So, we need only to show that the property VIIb.1 is a consequence of the property VIIb.2.

Suppose that the property VIIb.2 holds and the property VIIb.1 does not hold. Now, from the property VIIb.2, we conclude that the word $us_1 \dots s_n v \in \mathcal{LWSF}$ is a square and the word $uus_1 \dots s_n$ is a 2^+-free word by the the cases IVb of Proposition II.2.5 and V of Proposition II.2.6. Thus, because the latter condition of the property VIIb.1 is satisfied, the word $uus_1 \dots s_n v$ has to be of the form $ururu$, for some $u \in A^*$, with the condition

$$uus_1 \dots s_n \in \mathcal{LWSF}.$$

On the other hand, this cannot be true for $n \geq 3$ by Lemma II.2.3. This contradiction completes this proof.

Proofs of the cases VIIIa and VIIIb:

The proofs of these cases are similar with the proofs of the cases VIIa and VIIb. There is four differences between these proofs.

The first one is that instead of the condition

$$\overline{u}t_1t_2t_3 \in \{aaba, bbab\}$$

we use the condition

$$t_{n-2}t_{n-1}t_n v \in \{abaa, babb\}.$$

And the condition $\overline{u}t_1 \dots t_n \in \mathcal{LWSF}$ is replaced by the condition $t_1 \dots t_n v \in \mathcal{LWSF}$. The third difference is that we use the case VI rather than the case V. Finally, in the proofs of the cases VIIIa and VIIIb we conclude that the word $s_1 \dots s_n vv$ is a square instead of the word $uus_1 \dots s_n$, which cannot be true by Lemma II.2.3.

Proof of the case IXa:

Clearly, the properties IXa.1 and IXa.2 are equivalent by the cases VIIa and VIIIa.

Proof of the case IXb:

Firstly, the property IXb.1 implies the property IXb.2 by the cases VIIb and VIIIb.

The other direction is a little bit complex. From IXb.2 we obtain

$$uus_1 \dots s_n v \in \mathcal{LWSF},$$

$$us_1 \dots s_n vv \in \mathcal{LWSF}$$

and

$$us_1 \dots s_n v \text{ is not a square.}$$

Thus, we conclude that $uus_1 \dots s_n vv \in \mathcal{LAWSF} \cup \mathcal{LWSF}$. On the other hand, the length of this word is even, so we can conclude that

$$uus_1 \dots s_n vv \in \mathcal{LWSF}.$$

So, the property IXb.2 implies the property IXb.1. □

Actually, in the previous two propositions the conditions $n \geq 3$ and $n \geq 2$ could be relaxed, but it would make these propositions more complicated. Before we combine the results of the three previous propositions we shall consider the language of allmost 2^+-free words.

Now, by using Proposition II.2.5, we see that we can generate some words of the language $\mathcal{LAWSF}$ from shorter ones in the similar way than the words of the language $\mathcal{LWSF}$. This way is described in Lemma II.2.8. Actually, it turns out that there are only very few words in $\mathcal{LAWSF}$ which cannot be generated by this way.

Lemma II.2.8 *Let $n \geq 1$, u, v be letters of the alphabet A and $s_1, \dots, s_n$ be arbitrary binary words over the alphabet A such that*

$$s_i \in \{ab, ba\}, \text{ for all } 1 \leq i \leq n.$$

Additionally, let the letters $t_1, t_2, \dots, t_n \in A$ be chosen such that

$$s_i = \theta(t_i), \text{ for all } 1 \leq i \leq n,$$

Then the following three conditions are equivalent:

1) $t_1 t_2 \dots t_n \in \mathcal{LAWSF}$
2) $s_1 \dots s_{n-1} t_n \in \mathcal{LAWSF}$
3) $\overline{t_1} s_2 \dots s_n \in \mathcal{LAWSF}$.

Proof In the case $n \leq 2$ there is no word satisfying any of these conditions. So, we can assume that $n \geq 3$.

Firstly, we suppose that

$$t_1 t_2 \dots t_n \in \mathcal{LAWSF}.$$

Then, by the definition of the language $\mathcal{LAWSF}$, the equation $t_1 = t_n$ holds and the word $t_1 \dots t_{n-1}$ is a square such that the words $t_1 \dots t_{n-1}$ and $t_2 \dots t_n$ are 2^+-free. By Proposition II.2.5, the words $s_1 s_2 \dots s_{n-1}$, $\overline{t_1} s_2 \dots s_{n-1} t_n$ and $s_2 s_3 \dots s_n$ are 2^+-free words. Moreover, the words $s_1 s_2 \dots s_{n-1}$ and $s_2 s_3 \dots s_n$ are squares. Therefore, we conclude that

$$s_1 \dots s_{n-1} t_n \in \mathcal{LAWSF}$$

and

$$\overline{t_1} s_2 \dots s_n \in \mathcal{LAWSF}.$$

Thus, we have shown that the property 1 yields the properties 2 and 3.

Secondly, we suppose that

$$s_1 \dots s_{n-1} t_n \in \mathcal{LAWSF}$$

or

$$\overline{t_1} s_2 \dots s_n \in \mathcal{LAWSF}.$$

Then, by the definition of the language $\mathcal{LAWSF}$ we have a square

$$\overline{t_1} s_2 \dots s_{n-1} t_n \in \mathcal{LWSF}.$$

By the condition IVb.1 in Proposition II.2.5, we obtain the formula

$$t_1 t_2 \dots t_n \in \mathcal{LAWSF}.$$

So, we have shown that the properties 2 and 3 both yields the property 1 and, therefore, the properties 1,2 and 3 are equivalent. □

Next we consider the words in $\mathcal{LAWSF}$ which are not covered by the previous lemma. Here we assume that the words t_i and s_i are defined as in Lemma II.2.8. Then such a word $w = uvuvu$, with $|w| \geq 9$, has to be of the form

$$w = uus_1 \dots s_{n-1} t_n,$$

where $u \in A$, $v \in A^*$ and $n \geq 4$, and all subwords of this word have to be 2^+-free words. For the shorter words, i.e. $|w| < 9$ and $2 \not| |w|$, we have to check them one by one. That is because otherwise the largest inner factor

$us_1 \dots s_{n-1}$ has length less than 6. That length is needed to make sure that the subwords $uus_1 \dots s_{n-1}$ and $us_1 \dots s_{n-1}t_n$ have the same block-structure.

Furthermore, the word $t_1 \dots t_n$ is a 2^+-free word by the proposition II.2.5. Now we can use Lemma II.2.3 to find out such words w with length greater that 7. By that lemma we conclude that the word w must be one of the words in the set

$$B = \{aaa, bbb, aabaaba, abaabaa, baabaab, bbabbab, babbabb, abbabba\}.$$

By the above considerations, we obtain the following result:

Lemma II.2.9 *Let w be a arbitrary word. Then*

$$w \in \mathcal{LAWSF}$$

if and only if one of the equations

$$uw = \theta(w')$$

and

$$wu = \theta(w'),$$

are satisfied for some $u \in A$ and $w' \in \mathcal{LAWSF}$, or

$$w \in B.$$

Proof Let us assume that $w \in \mathcal{LAWSF} \setminus B$. Then we can extend this word by a letter u such that

$$wu = \theta(w') \text{ or } uw = \theta(w'),$$

for some word w'. Then, by Lemma II.2.8, w' is an allmost 2^+-free word.

The other direction follows from the fact that Thue-Morse morphism preserves 2^+-freeness. □

By the definition of the Thue-Morse morphism θ, there is a word w_1 in the set B mentioned in the previous lemma for every 2^+-free word w such that this word w can be generated from w_1 by the previous two recursive equations. Thus,

$$|w| = 2^n|w_1| - 2^{n-1} - 2^{n-2} - \dots - 1 = 2^n(|w_1| - 1) + 1,$$

for the word w which is obtained by n recursive steps starting at w_1. Moreover, by applying the previous equations in a proper order, the word w can be any subword of the word $\theta^n(w_1)$ starting from any of the first 2^n positions.

Next we take a closer look to the words in B. In that set the words

$$aabaaba, abaabaa \text{ and } baabaab,$$

are cyclic permutations of the words

$$bbabbab, babbabb \text{ and } abbabba,$$

respectively.

Now, we are ready to state the following theorem characterizing almost 2^+-free words.

Theorem II.2.10 *The language $\mathcal{LAWSF}$ consists of exactly those subwords of the words*

$$\theta^n(www),$$

where $w \in \{a, b, aab, bba\}$, which are of length $2^{n+1}|w| + 1$.

□

Next we are going to show that the condition IXb.2 in Proposition II.2.7 cannot hold for any word from $\mathcal{LAWSF}$.

Let us suppose that the condition IXb.2 holds. Because the length of the word $\overline{u}t_1 \dots t_n v$ is odd, two leftmost or rightmost letters are in the same block with respect to the block-structure of this word. But these two letters are the same, which is impossible. Therefore, there does not exist any word in $\mathcal{LAWSF}$ such that the condition IXb.2 is satisfied.

Finally, we are able to settle a recursive definition for the 2^+-free words:

Theorem II.2.11 *Let $n \geq 3$, $u, v \in A$, $t_1 \dots t_n$ be letters over the alphabet A. Moreover, we denote the word $ut_1 \dots t_n v$ by w. Then, if the word w is a 2^+-free word, the formula*

$$x\theta(t_1) \dots \theta(t_n)y \in \mathcal{LWSF}$$

holds if and only if $(x, y) \in R(w)$. Furthermore, if the word w is not a 2^+-free word, the word

$$x\theta(t_1) \dots \theta(t_n)y \in \mathcal{LWSF},$$

if and only if,

$$w \in \mathcal{LAWSF}, (x, y) \in S(w)$$

and

$$x \neq \rho_l, \text{ if } y = \rho_r.$$

Proof Let us first assume that $w \in \mathcal{LWSF}$. Now we obtain from condition $(x,y) \in R(w)$ that

$$x\theta(t_1)\ldots\theta(t_n)y \in \mathcal{LWSF}$$

by Propositions II.2.5, II.2.6 and II.2.7. To other direction, from this last formula we obtain $w \in \mathcal{LWSF} \cup \mathcal{LAWSF}$ and $(x,y) \in R(w)$. Thus, by our assumption, we conclude that the former part of this Theorem holds.

Let us next assume that $w \notin \mathcal{LWSF}$. Then Propositions II.2.5, II.2.6 and II.2.7 yields from

$$x\theta(t_1)\ldots\theta(t_n)y \in \mathcal{LWSF}$$

the formulas

$$w \in \mathcal{LAWSF} \text{ and } (x,y) \in S(w).$$

Moreover, the equations

$$x = \rho_l(u) \text{ and } y = \rho_r(v)$$

cannot hold simultaneously, because both the conditions

$$\mathrm{Pref}_4(w) \in F_l \text{ and } \mathrm{Suff}_4(w) \in F_r$$

cannot hold simultaneously by Lemma II.2.9.

To the other direction the proof is direct application of b-cases of Propositions II.2.5 and II.2.7. $\square$

As we can see from the above theorem, the Thue-Morse morphism, or its generalized version (cf. [Sé2]), plays an important role in the language of 2^+-free words.

Let us mention here two interesting results about the Thue-Morse morphism.

Theorem II.2.12 (Harju [Ha]) *Let $h : A \mapsto A$ be a morphism preserving 2^+-freeness. Then there exists a natural number k such that*

$$h = \theta^k \quad \text{or} \quad h = \overline{\theta}^k.$$

Here the morphism $\overline{\theta}$ is defined as

$$\overline{\theta}(u) = \overline{u}u.$$

Theorem II.2.13 (Séébold [Sé1]) *Let $h : A \mapsto A$ be a morphism such that h^ω is an infinite 2^+-free word. Then there exists a natural number n such that*

$$h = \theta^k \quad \text{or} \quad h = \overline{\theta}^k.$$

In the next section we study structure of 2^+-free words.

II.3 The structure of the 2^+-free words

There is already some information about the structure of 2^+-free words, for example the structural presentation introduced by Restivo and Salemi ([RS]). But these structural presentations are not powerful enough to characterize the language of 2^+-free words. These structural constructions tells only that every 2^+-free word has a specific structure, or a word having a specific structure is a 2^+-free.

In this section we are going to introduce a way to present a given 2^+-free word. Moreover, this structural presentation of 2^+-free words can be used to characterize these words, i.e. if a word can be presented by this structural presentation it is 2^+-free, and vice versa.

II.3.1 S-structures

This presentation of 2^+-free words is based on results obtained in the previous chapter. Actually, because every 2^+-free word satisfies exactly one of the conditions I.1, II.1, III.1, IVa.1, IVb.1, V.1, VI.1, VIIa.1, VIIb.1, VIIIa.1, VIIIb.1, IXa.1 and IXb.1, it is possible to define this structural presentation such that the presentation of a given 2^+-free word is unique.

In this section we introduce a definition of *S-structures*. We also fix some useful notations for sequences of the mappings σ_l, ρ_l, θ and σ_r, ρ_r, θ, and for sets containing such sequences.

For sequences of above mappings, we use the notations

$$(h_i)_{i=1}^n \ , \ h_1, h_2, \ldots, h_n \text{ and } h_1 h_2 \ldots h_n.$$

The third notation, without commas, is used only in equalities and, when denoting sets of such sequences. The sequence containing no mappings is referred as "empty sequence" and it is denoted by ε.

Next we define a structure which is related to the 2^+-free words.

Definition II.3.1 *A S-structure related to words $w \in A^*$ and $w' \in A^*$ is a quadruple*

$$\mathcal{S}_w^{w'} = (k; w'; h_{l,1}, h_{l,2}, \ldots, h_{l,k-1}; h_{r,1}, h_{r,2}, \ldots, h_{r,k-1}),$$

where the mappings $h_{l,i}$ and $h_{r,i}$ are chosen from the sets $\{\theta, \sigma_l, \rho_l\}$ and $\{\theta, \sigma_r, \rho_r\}$, respectively, such that the equation

$$w_{i+1} = h_{l,i}(u_i)\theta(w_i')h_{r,i}(v_i) = u_{i+1}w_{i+1}'v_{i+1},$$

where $u_{i+1}, v_{i+1} \in A, w_{i+1}, w_{i+1}' \in A^*$, holds for all $1 \le i \le k-1$,

$$w_1 = w' = u_1 w_1' v_1, \text{ for some } u_1, v_1 \in A, w_1, w_1' \in A^*,$$

and

$$w_k = w.$$

For short presentations of the sequences described in the above definition, we choose the notations

$$\Lambda_{(i,j)} = (\sigma_l, \ldots, \sigma_l, \theta, \ldots, \theta, \sigma_l),$$

$$\Delta(m,n) = (\sigma_r, \ldots, \sigma_r, \theta, \ldots, \theta, \sigma_r),$$

$$\Lambda'_{(i,j)} = (\sigma_l, \ldots, \sigma_l, \theta, \ldots, \theta)$$

and

$$\Delta'_{(m,n)} = (\sigma_r, \ldots, \sigma_r, \theta, \ldots, \theta),$$

where the integers $i \geq 0$ and $j \geq 0$ tells us the number of consecutive occurrences of the mappings σ_l and θ, respectively. Similarly, the integers $m \geq 0$ and $n \geq 0$ tells us the same information about the mappings σ_r and θ. Furthermore, we define also the sets

$$\Gamma_\Lambda = \{\ \Lambda_{(i,j)} \mid i \geq 0, j \geq 1\ \},$$

$$\Gamma_\Delta = \{\ \Delta_{(i,j)} \mid i \geq 0, j \geq 1\ \},$$

$$\Gamma'_\Lambda = \{\ \Lambda'_{(i,j)} \mid i \geq 0, j \geq 0, i+j > 0\ \},$$

$$\Gamma'_\Delta = \{\ \Delta'_{(i,j)} \mid i \geq 0, j \geq 0, i+j > 0\ \},$$

$$\Gamma^p_\Lambda = \{\ \Lambda_{(i,j)} \mid j \geq 2, i \geq 0\ \} \cup \{\ \Lambda_{(i,j)} \mid 2|(i-p), i \geq 0, j = 1\ \}$$

and

$$\Gamma^p_\Delta = \{\ \Delta_{(i,j)} \mid j \geq 2, i \geq 0\ \} \cup \{\ \Delta_{(i,j)} \mid 2|(i-p), i \geq 0, j = 1\ \}.$$

Let us next recall some notations. A Morse-block is a word of the form

$$u_n^+ = \theta^n(u),$$

where u is a letter. In this notation, u^+ is called as stamp and n as index. Moreover, a stamp can also contain a minus sign instead of plus. The correspondence is the following:

$u_n^+ = u_n^-$, if n is odd;
$u_n^+ = \overline{u}_n^-$, if n is even.

The following definitions concern a way to represent a 2^+-free word and an allmost 2^+-free word.

Definition II.3.2 *A Morse-block presentation of a word $w \in A^*$ is a factorization of w which contains only consecuentive Morse-blocks such that the indexes of the Morse-blocks in this word from left to right forms a sequence*

$$i_1, \ldots, i_l, j_1, \ldots, j_m, k_1 \ldots, k_n \quad ,$$

where

$$1 \leq i_1 \leq i_2 < \ldots < i_l < j_1 \leq j_2 = \ldots = j_{m-1} \geq \\ j_m > k_1 > \ldots > k_{n-1} \geq k_n \geq 1$$

and the corresponding stamps are

$$u_1, \ldots, u_l, v_1, \ldots, v_m, w_1 \ldots, w_n \quad ,$$

where the stamps u_i and $v_1, \ldots, v_{m-1}$ have + as the upper index and the stamps v_m and w_i have - as the upper index.

In the following definition we use the same notations as in the previous definition.

Definition II.3.3 *A* **(K,L,u)**-*strict Morse-block presentation $\mathcal{A}_{(K,L,u)}$ for a word $w \in A^*$ is a Morse-block presentation satisfying the following conditions:*

1) $u \in A^*$, K, L are integers and α, β binary numbers;
2) $\alpha = \alpha_0 \ldots \alpha_{j_2-1}$, where $\alpha_i = 1$, iff $i \in \{i_1, \ldots, i_{i_l}, j_1\}$;
3) $\beta = \beta_0 \ldots \beta_{j_2-1}$, where $\beta_i = 1$, iff $i \in \{j_m, k_1, \ldots, k_{i_n}\}$;
4) $m = |u|$, $\alpha = K$, $\alpha + \beta = L$;
5) $i_1 \neq i_2$, and $k_{n-1} \neq k_n$;
6)

$$u_i = \begin{cases} \overline{u_{i+1}} & , \textit{if } 2|(i_{i+1} - i_i) \textit{ and } 1 \leq i \leq l \\ u_{i+1} & , \textit{otherwise} \end{cases}, \; u_l = \begin{cases} \overline{v_1} & , \textit{if } 2|(j_1 - i_l) \\ v_1 & , \textit{otherwise} \end{cases};$$

7)

$$w_i = \begin{cases} \overline{w_{i-1}} & , \textit{if } 2|(k_i - k_{i+1}) \textit{ and } 2 \leq i \leq n \\ w_{i-1} & , \textit{otherwise} \end{cases}, \; w_1 = \begin{cases} \overline{v_m} & , \textit{if } 2|(j_m - k_1) \\ v_m & , \textit{otherwise} \end{cases};$$

8) $u = x v_2 v_3 \ldots v_{m-1} y$, where

$$x = \begin{cases} v_1 & , \textit{if } 2|(j_2 - j_1) \\ \overline{v_1} & , \textit{otherwise} \end{cases}, y = \begin{cases} v_m & , \textit{if } 2|(j_m - j_{m-1}) \\ \overline{v_m} & , \textit{otherwise} \end{cases};$$

Next we use previous notations and definitions for presentation of 2^+-free words.

II.3.2 S-structures corresponding 2^+-free words

In order to obtain our structural presentation we study the correspondence between the 2^+-free words and the certain subset of S-structures.

Let $u, v \in A$ and $w_1 = utv$ be an arbitrary 2^+-free word of length greater or equal than 7. Now, by Propositions II.2.5, II.2.6 and II.2.7, we obtain a word $w_2 = x\theta(t)y$, where x and y are given by Theorem II.2.11. Then w_2 has length approximately half from the length of the word w_1. By repeating this deduction we obtain a sequence $w_1, w_2, \ldots w_k$, where k is such an integer that

$$|w_{k-1}| \geq 7$$

and

$$|w_k| < 7.$$

As we noted above the conditions I.1, II.1, III.1, IVa.1, IVb.1, V.1, VI.1, VIIa.1, VIIb.1, VIIIa.1, VIIIb.1 and IXa.1 defines such a sequence unique for each 2^+-free word.

Next we consider the relations between consecutive words in such a sequence. For that purpose we let the sequence

$$w_1, w_2 \ldots w_k$$

be the sequence obtained for the word $w \in \mathcal{LWSF}$. Now, by Theorem II.2.10, we can assume that there exists words $u, v \in A$, $x \in R$ and $y \in S$ satisfying the equations

$$w_i = xw_i'y$$

and

$$w_{i+1} = uw_{i+1}'v$$

where $1 \leq i \leq k-1$, and moreover the words $w_i', w_{i+1}' \in A^*$ satisfy the equations

$$\theta(w_{i+1}') = w_i'.$$

With words u, v, x and y we associate mappings $h_{l,i}$ and $h_{r,i}$, where $h_{l,i} \in \{\theta, \sigma_l, \rho_l\}$, $h_{r,i} \in \{\theta, \sigma_r, \rho_r\}$ and $1 \leq i \leq k-1$, such that

$$x = h_{l,i}(u)$$

and

$$y = h_{r,i}(v).$$

Furthermore, in Propositions II.2.5, II.2.6 and II.2.7 the correspondence between a 2^+-free word and a longer 2^+-free word obtained from this word, in the sense of Theorem II.2.11, is one-to-one for the 2^+-free words having length greater than 6.

Now, from the above considerations we obtain the following lemma.

Lemma II.3.4 *For every 2^+-free word w of length greater or equal than 7 there exists a unique S-structure*

$$\mathcal{S}_w^{w'} = (k; w'; h_{l,1}, h_{l,2}, \ldots, h_{l,k-1}; h_{r,1}, h_{r,2}, \ldots, h_{r,k-1}),$$

such that

$$w' = uxv \in \mathcal{LAWSF} \cup \mathcal{LWSF}, \text{ for some } u, v \in A \text{ and } x \in A^*,$$

$$|w'| < 7 \leq |h_{l,1}(u)\theta(x)h_{r,1}(v)|, \text{ if } w' \in \mathcal{LWSF},$$

$$w_i \in \mathcal{LWSF}, \text{ for } 2 \leq i \leq k,$$

and

$$w_k = w.$$

□

From Theorem II.2.11 we conclude that the length of the word w', obtained by the previous lemma starting at the word w, satisfies the equation

$$|w'| \geq \left\lceil \frac{|h_{l,1}(u)\theta(x)h_{r,1}(v)| - 1}{2} \right\rceil.$$

So, for the length of the word w' in Lemma II.3.4 we obtain

$$|w'| \geq 4.$$

Thus, the Theorem II.2.11 guarantees that we can replace the condition

$$|w'| < 7 \leq |h_{l,1}(u)\theta(x)h_{r,1}(v)|, \text{ if } w' \in \mathcal{LWSF},$$

by the condition

$$4 \leq |w'| < 7 \leq |h_{l,1}(u)\theta(x)h_{r,1}(v)|, \text{ if } w' \in \mathcal{LWSF}_{base}.$$

Next we are going to study those sequences of mappings in the S-structures corresponding the 2^+-free words. But, first we study the effects of individual mappings θ, σ_r and ρ_r.

Let us assume that a given 2^+-free word w of length greater than 4 can be expressed in the form

$$w = u_m^+ v_n^+ t, \text{ where } 1 \leq m \leq n, t \in A^* \text{ and } u, v \in A. \tag{12}$$

Now, we apply the mapping $h \in \{\theta, \sigma_l, \rho_l\}$ to the first letter of this word while the mapping applied to the rest of the letters is the Thue-Morse morphism θ. Then we have the following three possible cases.

In the case $h = \theta$, we simply obtain the word

$$\theta(w) = u_{m+1}^+ v_{n+1}^+ \theta(t). \tag{13}$$

So, the letters u and v are not changed, and the difference between the indexes of the first two Morse-blocks are of the same parity.

If we apply the mapping $h = \sigma_l$, we obtain the word

$$\overline{u}_1^+ \ldots \overline{u}_m^+ v_{n+1}^+ \theta(t). \tag{14}$$

That is because

$$u_{m+1}^+ = u_1^+ \overline{u}_1^+ \overline{u}_2^+ \overline{u}_3^+ \ldots \overline{u}_m^+.$$

In (14) it is possible that there are some "new Morse-blocks" in the beginning of this word. More precisely in that equation, those "new Morse-blocks" are $\overline{u}_1^+, \ldots, \overline{u}_{m-1}^+$, which are all generated from the same Morse-block by the mapping σ_l. When operating on the right ends of words, these "new Morse-blocks" are the rightmost ones with the same property. As we can see from equation (14) that happens if and only if $m \geq 2$. Now, if $m = 1$, the difference between the indexes of the two leftmost Morse-blocks is increased by one, and the stamp of the first Morse-block is switched to its complement. In the case $m \geq 2$ we conclude that the mapping h has the same effect for those Morse-blocks which were originally located in the beginning of the word w.

It remains to consider the case $h = \rho_l$. First of all, the prefix of the word w must be of the form

$$u_1^+ u_1^+ \overline{u}_k^+, \text{ where } k \geq 2, \tag{15}$$

or

$$u_1^+ u_2^+ u_k^+, \text{ where } k \geq 2. \tag{16}$$

In these cases, the words obtained has the prefixes

$$\overline{u}_1^+ \overline{u}_1^+ u_2^+ \overline{u}_{k+1}^+ \tag{15'}$$

or

$$\overline{u}_1^+\overline{u}_1^+u_3^+u_{k+1}^+, \tag{16'}$$

respectively. So, in this case, we obtain allways a new Morse-block to the beginning of the word. Moreover, for the Morse-blocks originally located in the beginning of the word w the effects of the mapping h are similar with the effects in the previous case.

In the sequence obtained during a generation of a 2^+-free word, we obtain similar results for the right ends of the 2^+-free words with the following exceptions. Firstly, in this case, the mappings are θ, σ_r and ρ_r. Secondly, we have to replace the equations (12), (13), (14), (15), (16), (15′) and (16′) by the equations

$$w = tu_m^-v_n^-, \text{ where } m \geq n \geq 1, t \in A^* \text{ and } u, v \in A,$$

$$\theta(w) = \theta(t)\overline{u}_{m+1}^-\overline{v}_{n+1}^-,$$

$$\theta(t)\overline{u}_{m+1}^-v_n^- \dots v_1^-,$$

$$\overline{u}_k^-u_1^-u_1^-, \text{ where } k \geq 2,$$

$$u_{k+1}^-\overline{u}_2^-u_1^-u_1^-,$$

$$u_k^-u_2^-u_1^-, \text{ where } k \geq 2$$

and

$$\overline{u}_{k+1}^-\overline{u}_3^-u_1^-u_1^-,$$

respectively. Moreover, the effects are mainly the same as previously for beginnings of words, i.e. the difference between the indices and the stamps of the rightmost two Morse-blocks are the same as in the previous cases.

Because the words of the language $\mathcal{LAWSF}$ can be generated similarly than those in $\mathcal{LWSF}$ by Lemma II.2.9 we can use the previous considerations when we consider Morse-block presentations for these words. As we can see from the lemma below, it is quite easy to give appropriet restrictions for Morse-block presentations characterizing the language $\mathcal{LAWSF}$.

Theorem II.3.5 *Let $v \in \mathcal{LAWSF} \setminus \{aaa, bbb\}$ be obtained by Theorem II.2.9 starting at $u \in \mathcal{LAWSF}_{base}$. Then there is an unique $(K, \frac{|v|-1}{|u|-1}+1, u)$-strict Morse-block presentation for v.*

Moreover, every word v having $(K, 2^{L+1}+1, u)$-strict Morse-block presentation for some word $u \in \mathcal{LAWSF}_{base}$ and constants $K, L \geq 1$ is also an allmost 2^+-free word.

The strict Morse-block presentation described in the above lemma can be obtained by Lemma II.2.9 with the considerations about effects of mappings σ_l and σ_r.

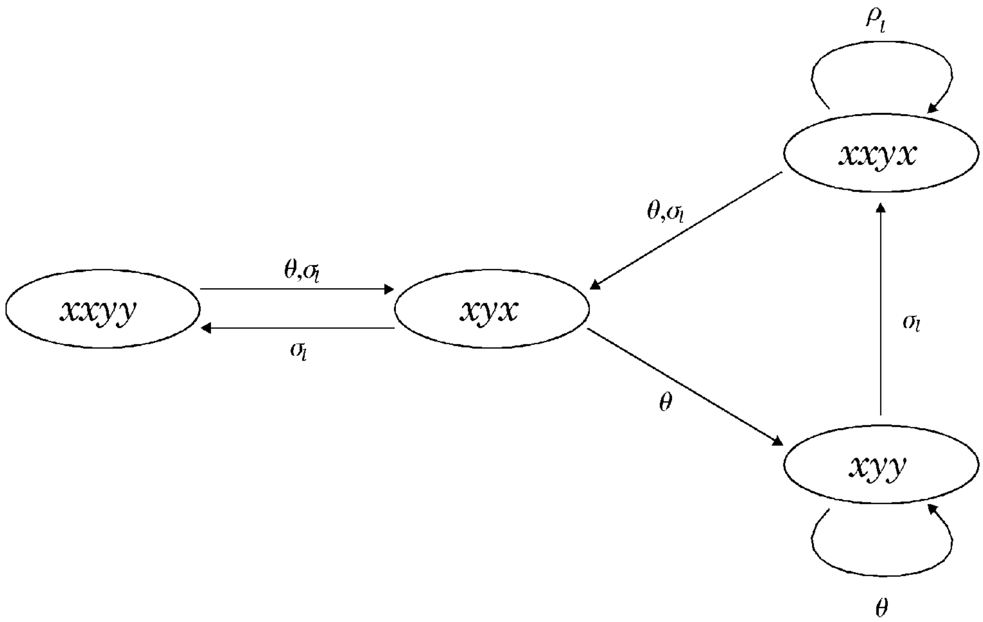

Figure II.3.1. Transformation of prefixes by θ, σ_l and ρ_l

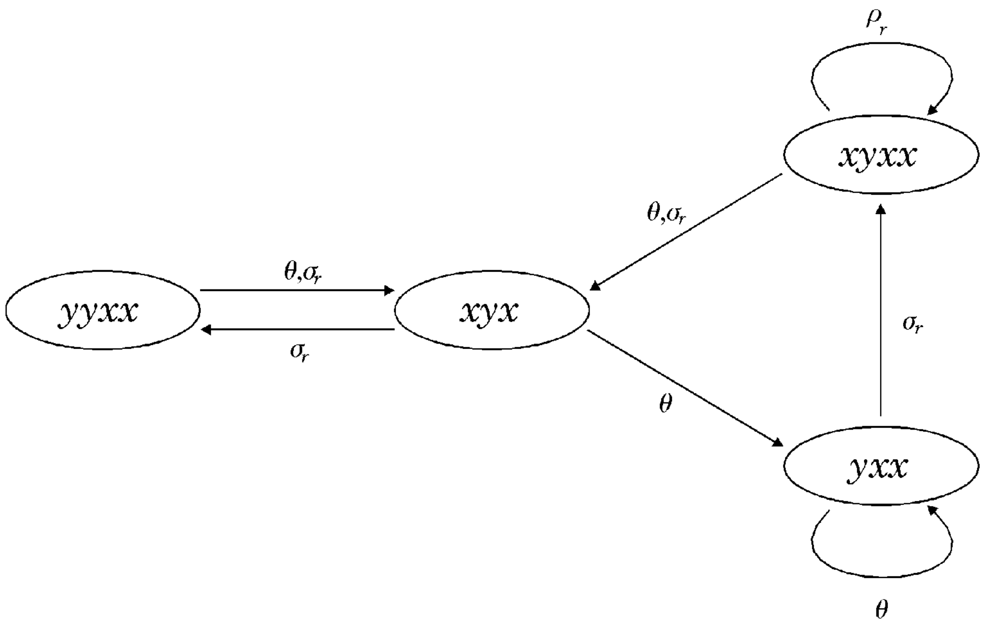

Figure II.3.2. Transformation of suffixes by θ, σ_r and ρ_r

Proof Let us assume that $v \in \mathcal{LAWSF} \setminus \{aaa, bbb\}$ and $u \in \mathcal{LAWSF}_{base}$. Then, by the consideration presented after Lemma II.2.9, $|v| = 2^i(|u|-1)+1$, where i is the number of times Lemma II.2.9 is applied to the word u in order to obtain v. Thus,

$$2^i = \frac{|v|-1}{|u|-1}.$$

At the same time, when applying Lemma II.2.9 recursively to u, we also obtain a word as a prefix of v corresponding the first symbol of u. And, now the length of this prefix is the value of K. Also, the properties from 5 to 8 listed in Definition II.3.3 are satisfied by the equations (12)-(16), ($15'$) and ($16'$).

For the words from the set $\mathcal{LAWSF}_{base}$ the uniqueness is clear. Let us now suppose that we have already proved uniqueness for all allmost 2^+-free word having length less than h, and that we have two different presentations for the word v having length h.

From the conditions in Definition II.3.3 we know that we can apply Lemma II.2.8 to the word v obtaining the word v' . On the other hand, from effects of the mappings σ_l and σ_r considered earlier, we also obtain two different presentations for this word. But, this is impossible. Thus, we obtain the uniqueness.

Let us next assume that a word $v \notin \{aaa, bbb\}$ has a required Morse-block presentation with some $u \in \mathcal{LAWSF}_{base}$. Now the conditions in Definition II.3.3 guaranties that we can use Lemma II.2.9 from shorter to the longer word. Then, by using Lemma II.2.9 recursively, we obtain the word v. Thus, $v \in \mathcal{LAWSF}$. $\square$

Next lemma shows how consecutive mappings are related in the S-structures corresponding the 2^+-free words. But, before that, we introduce two graphs which clarifies the proof of the lemma. These graphs, presented in Figures II.3.1 and II.3.2, shows how the mappings changes some spesific prefixes, or suffixes of words.

Lemma II.3.6 *Let w, w' be 2^+-free words having length greater or equal than 4 and h, g be the the sequences of the mappings in the S-structure $\mathcal{S}_w^{w'}$ satisfying Lemma II.3.4. Then*

$$h = h'h_1, \ldots, h_r, h''$$

and

$$g = g'g_1, \ldots, g_s, g''$$

such that

$$h'' \in \Gamma'_\Lambda \cup \{\varepsilon\}, \qquad g'' \in \Gamma'_\Delta \cup \{\varepsilon\},$$

$$h_i \in \Gamma_\Lambda \cup \{\rho_l\}, \qquad g_j \in \Gamma_\Delta \cup \{\rho_r\},$$

$$h' \in \Gamma_\Lambda \cup \{\rho_l, \sigma_l\rho_l\}, \qquad g' \in \Gamma_\Delta \cup \{\rho_r, \sigma_r\rho_r\},$$

for $1 \leq i \leq r-1$ *and* $1 \leq j \leq s-1$.

Moreover, the mapping ρ_l *can be preceeded only by itself, by the mapping* σ_l *or by a sequence from the set* Γ^1_Λ. *Similarly, the mapping* ρ_r *can be preceeded only by itself, by the mapping* σ_r *or by a sequence from the set* Γ^1_Δ.

Proof First we take a closer look to the two prefix conditions (15) and (16) required in the case $h = \rho_l$. By using what we already know about the effects of the mappings θ, ρ_l and σ_l, we conclude that the equation (15) can be obtained only by applying the mapping ρ_l to a 2^+-free word. Similarly, the equation (16) can be obtained only by applying the mapping σ_l to a 2^+-free word having a prefix of the form

$$\overline{u}_2^+ u_k^+ \text{ or } u_l^+, \text{ where } u \in A, k \geq 2 \text{ and } l > 2. \tag{17}$$

Now we are ready to study how new Morse-blocks are obtained. By the considerations before Theorem II.3.5 we conclude that the only ways to obtain a new Morse-block are either applying the mapping ρ_l to any 2^+-free word, or applying the mapping σ_l to any 2^+-free word having a Morse-block

$$u_k^+, \text{ where } k \geq 2 \text{ and } u \in A, \tag{18}$$

as a prefix. From the equations (12) and (14) we conclude that a new Morse-block can be generated by the mapping σ_l if and only if it is preceeded by the mapping θ. Hence we conclude that the subsequence of the mappings to be applied between appearences of new Morse-blocks, has to be

$$\Lambda_{(k,l)}, \text{ for some } k, l \geq 1, \text{ or } \rho_l,$$

when considering the beginnings of 2^+-free words. Thus, the equation for h_i holds.

The equation for h'' follows directly from the fact that only the mappings ρ_l and σ_l can produce new Morse-blocks to the beginning of the word obtained by Theorem II.2.11.

Because the sequence h' begins the sequence h, it has at least the same alternatives than the sequences h_i. On the other hand, from the equation

(17) we observe that the mapping ρ_l can also be preceeded by the mapping σ_l in the beginning of the sequence h.

By the effects of the mappings σ_l and ρ_l described earlier, we conclude that the mapping ρ_l can be preceeded only by the following three kinds of the sequences of mappings in the sequences h_i:

1) ρ_l;
2) $\Lambda_{(k,l)}$, where $k \geq 0$ and $l \geq 2$;
3) $\Lambda_{(k,1)}$, where $2 \not| k$.

Thus, Lemma II.3.6 holds for the sequence h.

Finally, we consider the mappings θ, ρ_r and σ_r. By the equations provided earlier to replace the equations (12), (13), (14), (15), (15′), (16) and (16′), we know that the behaviour of these mappings are similar with the behaviour of corresponding mappings for the beginnings of the 2^+-free words when considering the two latest Morse-blocks in the 2^+-free words. Moreover, the above conditions (17) and (18) should be replaced by the conditions

$$u_k^- \text{ or } \overline{u}_l^- u_2^-, \text{ where } u \in A, k > 2 \text{ and } l \geq 2. \tag{17'}$$

and

$$u_k^-, \text{ where } k \geq 2 \text{ and } u \in A. \tag{18'}$$

From these equations (17′) and (18′) we conclude, as for the sequence h, that the sequence g also satisfies the conditions of this lemma. □

Remark Let h be a sequence of mappings corresponding the beginnings of 2^+-free words. If we suppose that there is the mapping θ as a member in this sequence, we know from the previous considerations that the following mapping, if it exists, cannot be ρ_l. Furthermore, if there is the mapping σ_l in this sequence, the following mapping can be ρ_l. But, if this following mapping is not ρ_l, the mappings following the mentioned mapping σ_l forms a sequence from the set

$$\{\Lambda_{(k,l)} \mid k \geq 0, l \geq 0\}.$$

For the sequence corresponding the ends of 2^+-free words we obtain a similar result.

Next we introduce two partitions of the language $\mathcal{LAWSF}$. These partitions contain the following sets:

1) $w \in \Psi_{aaba-}$ if and only if one of the following conditions holds for $w = \mathcal{A}_{(k,2^{n-1}+1,u)}$:
- $k = 1$, $\mathrm{Pref}_4(u) \in \{aaba, bbab\}$ and $n = 1$;
- $k = 1$, $\mathrm{Pref}_3(u) \in \{abb, baa\}$ and $n = 2$;
- $k = 3$, $\mathrm{Pref}_2(u) \in \{ab, ba\}$, $2|n$ and $n \geq 3$;
- $k = 3$, $\mathrm{Pref}_2(u) \in \{aa, bb\}$, $2 \not| n$ and $n \geq 3$;
- $k = 3 + 4^c + 2d4^c$, for some $c \geq 1, d \geq 0$, $u \in \mathcal{LAWSF}_{base}$ and $n \geq 4$;

2) $w \in \Psi_{aabb-}$ if and only if one of the following conditions holds for $w = \mathcal{A}_{(k,2^{n-1}+1,u)}$:
- $k = 1$, $\mathrm{Pref}_3(u) \in \{aba, bab\}$ and $n = 2$;
- $k = 1 + 2 \cdot 4^c + d4^{c+1}$, for some $c \geq 1, d \geq 0$, $u \in \mathcal{LAWSF}_{base}$ and $n \geq 4$;
- $k = 3$, $\mathrm{Pref}_2(u) \in \{aa, bb\}$, $2|n$ and $n \geq 3$;
- $k = 3$, $\mathrm{Pref}_2(u) \in \{ab, ba\}$, $2 \not| n$ and $n \geq 3$;
- $k = 3 + 2 \cdot 4^c + d4^{c+1}$, for some $c \geq 1, d \geq 0$, $u \in \mathcal{LAWSF}_{base}$ and $n \geq 4$;

3) $w \in \Psi_{aba-}$ if and only if one of the following conditions holds for $w = \mathcal{A}_{(k,2^{n-1}+1,u)}$:
- $k = 1$, $\mathrm{Pref}_3(u) \in \{aba, bab\}$ and $n = 1$;
- $k = 1$, $\mathrm{Pref}_2(u) \in \{aa, bb\}$, $2|n$ and $n \geq 2$;
- $k = 1$, $\mathrm{Pref}_2(u) \in \{ab, ba\}$, $2 \not| n$ and $n \geq 2$;
- $k = 1 + 4^c + 2d4^c$, for some $c \geq 1, d \geq 0$, $u \in \mathcal{LAWSF}_{base}$ and $n \geq 4$;
- $k = 2$, $\mathrm{Pref}_2(u) \in \{aa, bb\}$, $n = 2$;
- $k = 2$, $\mathrm{Pref}_2(u) \in \{ab, ba\}$, $2 \not| n$ and $n \geq 3$;
- $k = 2$, $\mathrm{Pref}_2(u) \in \{aa, bb\}$, $2|n$ and $n \geq 3$;
- $k = 2 + 4^c + 2d4^c$, for some $c \geq 1, d \geq 0$, $u \in \mathcal{LAWSF}_{base}$ and $n \geq 4$;

4) $w \in \Psi_{abb-}$ if and only if one of the following conditions holds for $w = \mathcal{A}_{(k,2^{n-1}+1,u)}$:
- $k = 1$, $\mathrm{Pref}_3(u) \in \{abb, baa\}$ and $n = 1$;
- $k = 2$, $\mathrm{Pref}_2(u) \in \{ab, ba\}$, $n = 2$;
- $k = 2$, $\mathrm{Pref}_2(u) \in \{ab, ba\}$, $2|n$ and $n \geq 3$;
- $k = 2$, $\mathrm{Pref}_2(u) \in \{aa, bb\}$, $2 \not| n$ and $n \geq 3$;
- $k = 2 + 2 \cdot 4^c + d4^{c+1}$, for some $c \geq 1, d \geq 0$, $u \in \mathcal{LAWSF}_{base}$ and $n \geq 4$;
- $k = d4^c$, for some $c \geq 1, d \geq 1$, $u \in \mathcal{LAWSF}_{base}$ and $n \geq 3$;

5) $w \in \Psi_{-abaa}$ if and only if one of the following conditions holds for $w = \mathcal{A}_{(k,2^{n-1}+1,u)}$:
- $k = 1$, $\mathrm{Suff}_4(u) \in \{abaa, babb\}$ and $n = 1$;
- $k = 2$, $\mathrm{Suff}_3(u) \in \{aab, bba\}$ and $n = 2$;
- $k = 2^{n-1} - 2$, $\mathrm{Suff}_2(u) \in \{ab, ba\}$, $2|n$ and $n \geq 3$;
- $k = 2^{n-1} - 2$, $\mathrm{Suff}_2(u) \in \{aa, bb\}$, $2 \not| n$ and $n \geq 3$;
- $k = 2^{n-1} - 2 - 4^c - 2d4^c$, for some $c \geq 1, d \geq 0$, $u \in \mathcal{LAWSF}_{base}$ and $n \geq 4$;

6) $w \in \Psi_{-bbaa}$ if and only if one of the following conditions holds for $w = \mathcal{A}_{(k,2^{n-1}+1,u)}$:
- $k = 2$, $\mathrm{Suff}_3(u) \in \{aba, bab\}$ and $n = 2$;
- $k = 2^{n-1} - 2 \cdot 4^c - d4^{c+1}$, for some $c \geq 1, d \geq 0$, $u \in \mathcal{LAWSF}_{base}$ and $n \geq 4$;
- $k = 2^{n-1} - 2$, $\mathrm{Suff}_2(u) \in \{aa, bb\}$, $2|n$ and $n \geq 3$;
- $k = 2^{n-1} - 2$, $\mathrm{Suff}_2(u) \in \{ab, ba\}$, $2 \not| n$ and $n \geq 3$;
- $k = 2^{n-1} - 2 - 2 \cdot 4^c - d4^{c+1}$, for some $c \geq 1, d \geq 0$, $u \in \mathcal{LAWSF}_{base}$ and $n \geq 4$;

7) $w \in \Psi_{-aba}$ if and only if one of the following conditions holds for $w = \mathcal{A}_{(k,2^{n-1}+1,u)}$:
- $k = 1$, $\mathrm{Suff}_3(u) \in \{aba, bab\}$ and $n = 1$;
- $k = 2^{n-1}$, $\mathrm{Suff}_2(u) \in \{aa, bb\}$, $2|n$ and $n \geq 2$;
- $k = 2^{n-1}$, $\mathrm{Suff}_2(u) \in \{ab, ba\}$, $2 \not| n$ and $n \geq 2$;
- $k = 2^{n-1} - 4^c - 2d4^c$, for some $c \geq 1, d \geq 0$, $u \in \mathcal{LAWSF}_{base}$ and $n \geq 4$;
- $k = 1$, $\mathrm{Suff}_2(u) \in \{aa, bb\}$, $n = 2$;
- $k = 2^{n-1} - 1$, $\mathrm{Suff}_2(u) \in \{ab, ba\}$, $2 \not| n$ and $n \geq 3$;
- $k = 2^{n-1} - 1$, $\mathrm{Suff}_2(u) \in \{aa, bb\}$, $2|n$ and $n \geq 3$;
- $k = 2^{n-1} - 1 - 4^c - 2d4^c$, for some $c \geq 1, d \geq 0$, $u \in \mathcal{LAWSF}_{base}$ and $n \geq 4$;

8) $w \in \Psi_{-bba}$ if and only if one of the following conditions holds for $w = \mathcal{A}_{(k,2^{n-1}+1,u)}$:
- $k = 1$, $\mathrm{Suff}_3(u) \in \{bba, aab\}$ and $n = 1$;
- $k = 1$, $\mathrm{Suff}_2(u) \in \{ab, ba\}$, $n = 2$;
- $k = 2^{n-1} - 1$, $\mathrm{Suff}_2(u) \in \{ab, ba\}$, $2|n$ and $n \geq 3$;
- $k = 2^{n-1} - 1$, $\mathrm{Suff}_2(u) \in \{aa, bb\}$, $2 \not| n$ and $n \geq 3$;
- $k = 2^{n-1} - 1 - 2 \cdot 4^c - d4^{c+1}$, for some $c \geq 1, d \geq 0$, $u \in \mathcal{LAWSF}_{base}$ and $n \geq 4$;
- $k = 2^{n-1} - d4^c + 1$, for some $c \geq 1, d \geq 1$, $u \in \mathcal{LAWSF}_{base}$ and $n \geq 3$;

The sets described in the cases (1), (2), (3) and (4) forms a partition of the language such that these sets contains the words having the word *aaba*, *aabb*, *aba* or *abb*, respectively, as a prefix. The other four sets also form a partition. For these sets suffixes are used, instead of prefixes, as criteria.

Combining the previous results of this section, we obtain the following result:

Proposition II.3.7 *For every 2^+-free word $w \notin \{a, b, ab, aa, ba, bb, aab, bba, aba, bab, abb, baa\}$ there exists a unique S-structure*

$$\mathcal{S}_w = \{k; w_1; h_{l,1}, h_{l,2}, \ldots, h_{l,k-1}; h_{r,1}, h_{r,2}, \ldots, h_{r,k-1}\},$$

where the integer k, the word $w_1 \in \mathcal{LWSF}_{base} \cup \mathcal{LAWSF} \setminus \{aaa, bbb\}$ and the mappings $h_{l,i}$ and $h_{r,i}$, satisfies the following conditions for $1 \leq i \leq k-1$:

$$w_1 = u_1 w_1' v_1, \text{ for some } u_1, v_1 \in A,$$

$$w_{i+1} = h_{l,i}(u_i)\theta(w_i')h_{r,i}(v_i) = u_{i+1}w_{i+1}'v_{i+1}, \text{ for some } u_{i+1}, v_{i+1} \in A,$$
$$w_i \in \mathcal{LWSF}, \text{ for } 2 \leq i \leq k,$$

and

$$w_k = w.$$

Furthermore, the sequences $(h_{l,i})_{i=1}^{k-1}$ and $(h_{r,i})_{i=1}^{k-1}$ can be devided into consecutive parts $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$, respectively, such that the following conditions holds:

1)

$$w' = \begin{cases} w_1 & , \text{ if } w_1 \in \mathcal{LWSF}_{base}; \\ w_2 & , \text{ otherwise;} \end{cases}$$

2)

$$(x_1, y_1) \in \begin{cases} (h_{l,1}, h_{r,1}) \in S(w_1) & , \text{ if } w_1 \in \mathcal{LAWSF} \setminus \{aaa, bbb\}; \\ (\varepsilon, \varepsilon) & , \text{ otherwise;} \end{cases}$$

3) Sequences x_2 and y_2 satisfies the conditions

$$x_2 \in \begin{cases} \Gamma_\Lambda \cup \{\rho_l\} & , \text{ if } w' \in \Theta_{aaba-}; \\ \Gamma_\Lambda & , \text{ if } w' \in \Theta_{aba-} \cup \Theta_{aabb-}; \\ \Gamma_\Lambda \cup \{\sigma_l \rho_l\} & , \text{ if } w' \in \Theta_{abb-}; \end{cases}$$

$$y_2 \in \begin{cases} \Gamma_\Delta \cup \{\rho_r\} & , \text{ if } w' \in \Theta_{-abaa}; \\ \Gamma_\Delta & , \text{ if } w' \in \Theta_{-aba} \cup \Theta_{-bbaa}; \\ \Gamma_\Delta \cup \{\sigma_r \rho_r\} & , \text{ if } w' \in \Theta_{-bba}; \end{cases}$$

4) Sequences x_3 and y_3 satisfies the conditions

$$x_3 \in \begin{cases} \Gamma_\Lambda \cup \{\rho_l\} & \text{, if } x_2 \in \Gamma_\Lambda^1 \cup \{\rho_l\} \text{ and } w' \in \Theta_{aaba-}; \\ \Gamma_\Lambda \cup \{\rho_l\} & \text{, if } x_2 \in \Gamma_\Lambda^0 \text{ and } w' \in \Theta_{aba-}; \\ \Gamma_\Lambda \cup \{\rho_l\} & \text{, if } x_2 \in \Gamma_\Lambda^1 \text{ and } w' \in \Theta_{aabb-}; \\ \Gamma_\Lambda \cup \{\rho_l\} & \text{, if } x_2 \in \Gamma_\Lambda^0 \cup \{\sigma_l\rho_l\} \text{ and } w' \in \Theta_{abb-}; \\ \Gamma_\Lambda & \text{, otherwise;} \end{cases}$$

$$y_3 \in \begin{cases} \Gamma_\Delta \cup \{\rho_r\} & \text{, if } y_2 \in \Gamma_\Delta^1 \cup \{\rho_r\} \text{ and } w' \in \Theta_{-abaa}; \\ \Gamma_\Delta \cup \{\rho_r\} & \text{, if } y_2 \in \Gamma_\Delta^0 \text{ and } w' \in \Theta_{-aba}; \\ \Gamma_\Delta \cup \{\rho_r\} & \text{, if } y_2 \in \Gamma_\Delta^1 \text{ and } w' \in \Theta_{-bbaa}; \\ \Gamma_\Delta \cup \{\rho_r\} & \text{, if } y_2 \in \Gamma_\Delta^0 \cup \{\sigma_r\rho_r\} \text{ and } w' \in \Theta_{-bba}; \\ \Gamma_\Delta & \text{, otherwise;} \end{cases}$$

5) For $m > i > 3$ and $n > j > 3$, sequences x_i and y_j satisfies the conditions

$$x_i \in \begin{cases} \Gamma_\Lambda \cup \{\rho_l\} & \text{, if } x_{i-1} \in \Gamma_\Lambda^1 \cup \{\rho_l\}; \\ \Gamma_\Lambda & \text{, otherwise;} \end{cases}$$

$$y_j \in \begin{cases} \Gamma_\Delta \cup \{\rho_r\} & \text{, if } y_{j-1} \in \Gamma_\Delta^1 \cup \{\rho_r\}; \\ \Gamma_\Delta & \text{, otherwise;} \end{cases}$$

6) Sequences x_m and y_n satisfies the conditions

$$x_m \in \begin{cases} \Gamma_\Lambda \cup \{(\rho_l)\} \cup \Gamma'_\Lambda & \text{, if } x_{m-1} \in \Gamma_\Lambda^1 \cup \{(\rho_l)\}; \\ \Gamma_\Lambda \cup \Gamma'_\Lambda & \text{, otherwise;} \end{cases}$$

$$y_n \in \begin{cases} \Gamma_\Delta \cup \{(\rho_r)\} \cup \Gamma'_\Delta & \text{, if } y_{m-1} \in \Gamma_\Delta^1 \cup \{(\rho_r)\}; \\ \Gamma_\Delta \cup \Gamma'_\Delta & \text{, otherwise.} \end{cases}$$

Moreover, the previous conditions are necessary and sufficient for the 2^+-free words.

As we soon notice this Proposition is basicly same with our final presentation for 2^+-free words.

Proof Lemma II.3.4 states that there are unique S-structures for every 2^+-free word having length greater or equal than 7. By those few words in Table II.2.2 which are not in $\mathcal{LWSF}_{base}$, we can verify the uniqueness for all 2^+-free words obtained from the words of $\mathcal{LWSF}_{base}$. Thus, every 2^+-free word has the S-structure required.

Next we consider the conditions given in above lemma. In the case $w' \in \mathcal{LAWSF} \setminus \{aaa, bbb\}$ we obtain a word of $\mathcal{LWSF}$ as described in Table II.3.2 below. Now, the conditions 1,2,3 and 4 are obtained from Lemma II.3.6 with some simple considerations based on the graphs presented in Figures II.3.1

and II.3.2. In these graphs only the edges from state abb to $aaba$, and from bba to $abaa$, can be followed by the edge corresponding the mappings ρ_l, or ρ_r, respectively.

The parts x_2 and y_2 corresponds to the sequence of the first mappings applied to a word w' until we obtain a new Morse-block. By Lemma II.3.6 and its proof, we obtain the results concerning the words obtained by starting at word from $\mathcal{LWSF}_{base}$. For the words in $\mathcal{LAWSF} \setminus \{aaa, bbb\}$, we must consider the restrictions given by Theorem II.2.11. That theorem yields

$$h_{l,1} \in \{\sigma_l, \rho_l\}$$

and

$$h_{r,1} \in \{\sigma_r, \rho_r\}.$$

Moreover, the condition

$$h_{l,1} = \rho_l \quad \text{and} \quad h_{r,1} = \rho_r$$

cannot hold.

Finally, the last three parts are obtained from the Lemma II.3.6 and its proof with some additional considerations. As we observed above the sequences x_l and x_r must begin in a special way. Thus, we have to consider two different cases for sequences $y_{(l,i)}$ and $y_{(r,i)}$ depending from that fact if the w is 2^+-free word or not. $\square$

$$\begin{array}{ccccccc}
u \in \Theta_{aaba-} & \rightarrow & U = aaba & ; & u \in \Theta_{-abaa} & \rightarrow & V = abaa \\
u \in \Theta_{aabb-} & \rightarrow & U = aabb & ; & u \in \Theta_{-bbaa} & \rightarrow & V = bbaa \\
u \in \Theta_{abb-} & \rightarrow & U = abb & ; & u \in \Theta_{-bba} & \rightarrow & V = bba \\
u \in \Theta_{aba-} & \rightarrow & U = aba & ; & u \in \Theta_{-aba} & \rightarrow & V = aba
\end{array}$$

Table II.3.1 Starting states of Theorem II.3.8.

$$u \in \Psi_{aaba-} \overset{h=\sigma_l}{\rightarrow} U = aba \quad ; \quad u \in \Psi_{-abaa} \overset{g=\sigma_r}{\rightarrow} V = aba$$

$$u \in \Psi_{aaba-} \overset{h=\varrho_l}{\rightarrow} U = aaba \quad ; \quad u \in \Psi_{-abaa} \overset{g=\varrho_r}{\rightarrow} V = abaa$$

$$u \in \Psi_{aabb-} \overset{h=\sigma_l}{\rightarrow} U = aba \quad ; \quad u \in \Psi_{-bbaa} \overset{g=\sigma_r}{\rightarrow} V = aba$$

$$u \in \Psi_{abb-} \overset{h=\sigma_l}{\rightarrow} U = aaba \quad ; \quad u \in \Psi_{-bba} \overset{g=\sigma_r}{\rightarrow} V = abaa$$

$$u \in \Psi_{aba-} \overset{h=\sigma_l}{\rightarrow} U = aabb \quad ; \quad u \in \Psi_{-aba} \overset{g=\sigma_r}{\rightarrow} V = bbaa$$

Table II.3.2 Starting states of Theorem II.3.9.

II.3.3 The structure of the 2^+-free words

In this section we are going to state a presentation for the 2^+-free words which is based on the Morse-blocks. We begin with some further considerations about the subsequences $y_{(l,i)}$ and $y_{(r,i)}$ introduced in Theorem II.3.7.

Let $y_{(l,i)}$ be sequences, as in Theorem II.3.7, for some 2^+-free word obtained by starting at a word from $\mathcal{LWSF}_{base}$. Now we consider the effects of these sequences of mappings to the indexes in the Morse-block presentation.

First, we assume that

$$y_{(l,i)} \in \Gamma_\Lambda$$

and

$$y_{(l,i+1)} = \Lambda_{(k,l)}, \text{ for some } k \geq 0, l \geq 1,$$

for some $i > 1$. Because the sequences $y_{(l,i)}$ ends with the mapping σ_l, we know that the Morse-block presentation obtained by applying consecutively the mappings in $y_{(l,i)}$, has a prefix

$$u_1^+ u_2^+, \text{ for some } u \in A. \tag{21}$$

Hence we obtain a Morse-block presentation having prefix

$$x_1^+ x_2^+ \ldots x_l^+ u_{l+k+2}^+, \tag{22}$$

where

$$x = \begin{cases} u & ,\text{if } 2 \not| k, \\ \overline{u} & , \text{otherwise,} \end{cases}$$

after applying also the mappings in $y_{(l,i+1)}$.

Next we assume that $y_{(l,i)} = \rho_l$. In the previous considerations, the equations (22) and (23) should be replaced by the equations

$$u_1^+ u_1^+ u_2^+, \text{ for some } u \in A. \tag{23}$$

and

$$x_1^+ x_2^+ \dots x_l^+ u_{l+k+1}^+ u_{l+k+2}^+, \tag{24}$$

where

$$x = \begin{cases} u & \text{,if } 2|k, \\ \overline{u} & \text{, otherwise,} \end{cases}$$

respectively.

If we assume that

$$y_{(l,i)} \in \Gamma_\Lambda^1$$

and

$$y_{(l,i+1)} = \rho_l,$$

for some $i > 1$. Then we obtain, by applying the mappings in $y_{(l,i)}$, a Morse-block presentation having a prefix

$$u_1^+ u_1^+ u_m^+, \text{ for some } u \in A \text{ and } m > 3. \tag{25}$$

Here the restriction for the sequence $y_{(l,i)}$ is essential because otherwise the sequences cannot be consecutive. Now, by applying the mapping ρ_l, we finally obtain a Morse-block presentation havin a prefix

$$\overline{u}_1^+ \overline{u}_1^+ u_3^+ u_{m+1}^+. \tag{26}$$

As the last case we assume that

$$y_{(l,i)} = \rho_l = y_{(l,i+1)}.$$

Then the equation (23) holds for the word obtained after applying the mapping ρ_l once, and finally, after both mappings, we obtain a prefix of the form

$$\overline{u}_1^+ \overline{u}_1^+ u_2^+ \overline{u}_3^+. \tag{27}$$

By the previous considerations we are now ready to construct two directed graphs, presented in Figures II.3.3 and II.3.4, and some actions related to the edges of these graph. In Tables II.3.3, there is listed actions related to the graph of Figure II.3.3. The actions for the other graph are almost same except the mappings and labels of graph are switched to those corresponding suffixes. Let us next consider use of graph and actions conserning prefixes, i.e. use of Figure II.3.3 and Table II.3.3.

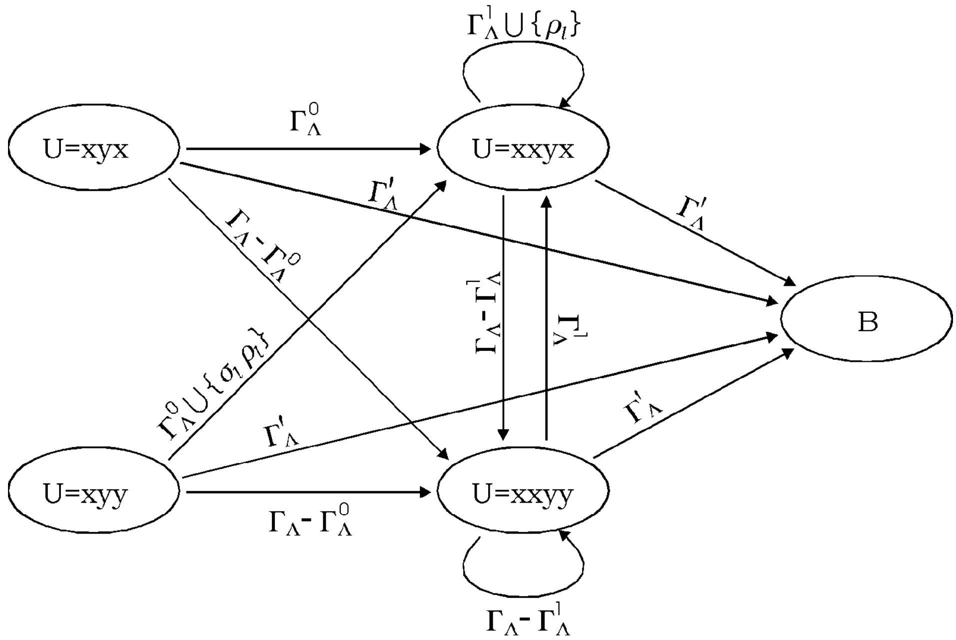

Figure II.3.3. Transition graph for prefixes

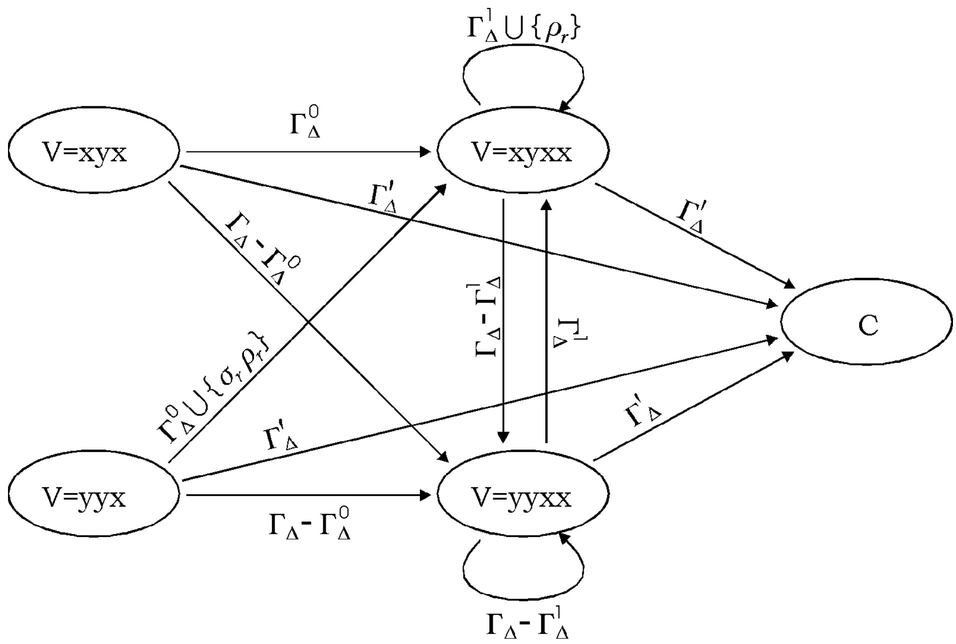

Figure II.3.4. Transition graph for suffixes

$B_2 = \{\sigma_l \rho_l\}$:

$p_1 = -2, q_1 = q_0$;

add q_1^+ to the end of the sequence $s_1, s_2, \dots$;

add p_1 to the end of the sequence $r_1, r_2, \dots$;

$B_{i+1} = \Lambda_{(k,l)}$:

$p_i = p_{i-1} - l - k - 1, q_i = \begin{cases} q_{i-1} & \text{, if } 2 \not| k, \\ \overline{q_{i-1}} & \text{, otherwise;} \end{cases}$

add $(q_i^+)^l$ to the end of the sequence $s_1, s_2, \dots$;

add $p_{i-1} - k - 1, p_{i-1} - k - 2, \dots, p_i + 1$ to the end of the sequence $r_1, r_2, \dots$;

$B_{i+1} = \{\rho_l\}$:

$p_i = p_{i-1} - 1,\ q_i = \overline{q_{i-1}}$;

add q_i^+ to the end of the sequence $s_1, s_2, \dots$;

add p_i to the end of the sequence $r_1, r_2, \dots$;

$B_m = \Lambda'_{(k,l)}$:

$p_{m-1} = p_{m-2} - k - l,\ q_{m-1} = \begin{cases} q_{m-2} & \text{, if } 2|k, \\ \overline{q_{m-2}} & \text{, otherwise;} \end{cases}$

add q_{m-1}^+ to the end of the sequence $s_1, s_2, \dots$;

add $p_{m-2} - k$ to the end of the sequence $r_1, r_2, \dots$;

$B_m \notin \Gamma'_\Lambda$:

add q_{m-1}^+ to the end of the sequence $s_1, s_2, \dots$;

add p_{m-1} to the end of the sequence $r_1, r_2, \dots$;

Table II.3.3 Actions related to the graph in Figure II.3.3.

When reading the graph in Figure II.3.3 we obtain sequences (x_i) and (a_i) for Theorems II.3.8 and II.3.9 by executing the following procedure.

Procedure A

1) At the beginning, we let $p_0 = 0$ and $q_0 = Pref_1(U)$ and we start from a state $U = X$, where X is the value of U. Moreover the length of sequences (s_i) and (r_i) is 0;

2) By Proposition II.3.7, there exists a unique division $B_2, B_3, \dots B_m$ of sequence $(h_{l,i})$, which is used to follow the labelled edges of the graph presented in Figure II.3.3. In the same order than blocks B_i occur in a given sequence, we execute corresponding actions listed in Table II.3.3;

3) Finally, sequences (x_i) and (a_i) for Theorems II.3.8 and II.3.9 are obtained by equations

3a) $x_i = s_i$;

3b) $a_i = r_i + |p_{m-1}| + 1$;

3c) $p = |p_{m-1}|$.

For the corresponding sequences related to a suffix, the corresponding procedure, say Procedure B, is similar except that both equations $h_{l,1} = \rho_l$ and $h_{r,1} = \rho_r$ cannot hold.

Actually, we can use Procedure A and B in the reverse order. In that way we can, for example, find different prefixes of infinite 2^+-free words.

Next we state the following two Theorems about Morse-block structures.

Theorem II.3.8 *Let v be a 2^+-free word having length greater or equal than 4 and $u \in \mathcal{LWSF}_{start}$ such that there exists a S-structure*

$$\mathcal{S}_v^u = (k; u; h_l; h_r).$$

Then there exits a Morse-block presentation with indexes

$$x_m \leq x_{m-1} < \ldots < x_1 \leq r_1 = r_2 = \ldots = r_{s-1} = r_s \geq \\ y_1 > y_2 > \ldots > y_{n-1} \geq y_n$$

and stamps

$$a_m^+, a_{m-1}^+, \ldots, a_1^+ b_1^+, b_2^+, \ldots, b_{s-1}^+, b_s^+, c_1^-, c_2^-, \ldots, c_{n-1}^-, c_n^-,$$

such that the conditions

1) $s = |u| - 2$;
2) $xb_1 \ldots b_s y = u$, for some $x, y \in A$;

are satisfied and other stamps and indexes can be obtained by Procedure A and B, starting at the states U and V, respectively, given by Table II.3.1. Moreover, such Morse-block presentation is unique and every word v having a Morse-block presentation, satisfying the above conditions is a 2^+-free word.

Proof Existence of a Morse-block presentation for a 2^+-free word is clear by the Theorem II.2.11 and equations (12)-(18) and (15′)-(18′).

By Proposition II.3.7 we can directly construct those two directed graphs presented in Figures II.3.3 and II.3.4. And, the two latter conditions given in the previous proposition are obvious.

Let us next consider the uniqueness. So, we assume that we have two different Morse-block presentations satisfying the conditions in previous theorem for a given word. Now, the conditions in the above theorem guarantees that there are two different S-structures, which satisfies the conditions of Proposition II.3.7, for this word. But, this is against the uniqueness of the S-structures given by Proposition II.3.7.

The last part of the previous theorem is also direct consequence of Proposition II.3.7. □

The remaining part is to give a Morse-block presentation for those 2^+-free words which are obtained from allmost 2^+-free words. In Table II.3.2 it is listed how prefixes and suffixes of words from $\mathcal{LAWSF}$ is mapped by the mappings σ_l and σ_r.

In order to clarify the presentation of the following theorem we introduce here the following conditions for indexes:

For indexes x_q, i_q, r_q, j_q and y_q there holds the formulas

$$x_m \leq x_{m-1} < \ldots < x_1 < i_1 < \ldots < i_k \leq r_1,$$

$$r_1 = r_2 = \ldots = r_s \text{ and}$$

$$r_s \geq j_l > \ldots > j_1 > y_1 > \ldots > y_{n-1} \geq y_n.$$

Theorem II.3.9 *Let v be a 2^+-free word having lenght greater or equal than 5, and $u \in \mathcal{LAWSF}_{start}$ such that there exists a S-structure*

$$\mathcal{S}_v^w = (k; w; h_l; h_r),$$

for some word $w \in \mathcal{LWSF}$. Moreover, we assume that there exists a word $w' \in \mathcal{LAWSF}$, which has (K, L, u)-strict Morse-block presentation, satisfies the equation

$$w = g_l(x)\theta(y)g_r(z),$$

where $xyz = w'$ and $(g_l, g_r) \in S(w')$. Then there exits a Morse-block presentation with indexes

$$x_m, \ldots, x_1, i_1, \ldots, i_k, r_1, \ldots, r_s, j_l, \ldots, j_1, y_1, \ldots, y_n,$$

satisfying the above condition, and stamps

$$a_m^+, \ldots, a_1^+, b_1^+, \ldots, b_k^+, c_1^+, \ldots, c_s^-, d_l^-, \ldots, d_1^-, e_1^-, \ldots, e_n^-,$$

such that the equation

$$\theta^p(Q) = (b_1^+)_{i_1} \ldots (b_k^+)_{i_k} (c_1^+)_{r_1} \ldots (c_s^-)_{r_s} (d_l^-)_{j_l} \ldots (d_1^-)_{j_1} = x'y'z',$$

where

$$Q = \begin{cases} \mathcal{A}_{(2K-2,2L-4,u)} & , \text{ if } g_l = \sigma_l \text{ and } g_r = \sigma_r \\ \mathcal{A}_{(2K-1,2L-3,u)} & , \text{ if } g_l = \rho_l \text{ and } g_r = \sigma_r \\ \mathcal{A}_{(2K-2,2L-3,u)} & , \text{ if } g_l = \sigma_l \text{ and } g_r = \rho_r \end{cases}$$

holds and baseindex p and other stamps and indexes can be obtained by Procedures A and B, starting at the states U and V, respectively, given by Table II.3.2. Moreover, such Morse-block presentation is unique and every word v having a Morse-block presentation satisfying the above conditions, for some $u \in \mathcal{LWSF}_{start}, w \in \mathcal{LWSF}$ and $w' \in \mathcal{LAWSF}$, is a 2^+-free word.

Proof The existence of such Morse-block presentation follows from Theorem II.2.11, Theorem II.3.5, Proposition II.3.7 and Table II.3.2.

The uniqueness is direct consequence from Proposition II.3.7, Table II.3.2 and Theorem II.3.5, as the last part of the previous theorem. The word Q is actually the word w without its first and last symbol. $\square$

In the following section we are going to calculate some bounds for the number of 2^+-free words and allmost 2^+-free words.

II.4 Counting 2^+-free words

Next we are going to study the number of weakly and allmost 2^+-free-words. In the second section we obtained Propositions II.2.5, II.2.6 and II.2.7 which are useful to estimate the number of 2^+-free words. By using these propositions, we managed to obtain new upper and lower bounds for the number of 2^+-free words.

Before this, we introduce notations for number of words in the sets Θ_{v-}, Θ_{-v}, Θ_{u-v}, Ψ_{v-}, Ψ_{-v} and Ψ_{u-v} having a given length n:

$$M_X(n) = |\{u \in \Theta_X \mid |u| = n\}|,$$

$$N_X(n) = |\{u \in \Psi_X \mid |u| = n\}|.$$

Next we start by introducing some equations.

II.4.1 Recursive linear functions for counting

By Propositions II.2.5, II.2.6, II.2.7 and Lemma II.2.9, we obtain easily for $n \geq 7$ the following equations:

$$M_{aa-ab}(n) = \begin{cases} M_{aa-ab}(\frac{n}{2}+1)+M_{bbab-}(\frac{n}{2})+N_{ab-aa}(\frac{n}{2}+1) & \text{, if } 2|n \\ M_{aa-ab}(\frac{n+1}{2})+M_{ab-ab}(\frac{n+1}{2})+M_{bbab-aa}(\frac{n+1}{2}) & \text{, if } 2\not|n \end{cases}; \quad (30)$$

$$M_{aa-aa}(n) = \begin{cases} M_{ab-ab}(\frac{n}{2}+1)+M_{bbab-babb}(\frac{n}{2})+N_{ab-ab}(\frac{n}{2}+1) & \text{, if } 2|n \\ 2M_{bbab-ab}(\frac{n+1}{2})+2N_{bbab-ab}(\frac{n+1}{2}) & \text{, if } 2\not|n \end{cases}; \quad (31)$$

$$M_{ab-ab}(n) = \begin{cases} 2M_{aa-ab}(\frac{n}{2})+M_{ab-ab}(\frac{n}{2})+M_{aa-aa}(\frac{n}{2})+M_{aa-aa}(\frac{n}{2}+1) & \text{, if } 2|n \\ 2M_{aa-aa}(\frac{n+1}{2})+2M_{aa-ab}(\frac{n+1}{2}) & \text{, if } 2\not|n \end{cases}; \quad (32)$$

$$M_{bbab-babb}(n) = \begin{cases} M_{bbab-babb}(\frac{n}{2})+M_{abba-abba}(\frac{n}{2}+1) & \text{, if } 2|n \\ 2M_{bbab-abba}(\frac{n+1}{2})+2N_{bbab-abba}(\frac{n+1}{2}) & \text{, if } 2\not|n \end{cases}; \quad (33)$$

$$M_{abba-abba}(n) = \begin{cases} M_{ab-ab}(\frac{n}{2}) & \text{, if } 2|n \\ 0 & \text{, if } 2\not|n \end{cases}; \quad (34)$$

$$M_{bbab-abba}(n) = \begin{cases} M_{bbab-ab}(\frac{n}{2}) & \text{, if } 2|n \\ M_{abba-ab}(\frac{n+1}{2}) & \text{, if } 2\not|n \end{cases}; \quad (35)$$

$$M_{abba-ab}(n) = \begin{cases} M_{ab-ab}(\frac{n}{2})+M_{aa-ab}(\frac{n}{2}) & \text{, if } 2|n \\ M_{aa-ab}(\frac{n+1}{2}) & \text{, if } 2\not|n \end{cases}; \quad (36)$$

$$M_{abba-aa}(n) = \begin{cases} M_{bbab-ab}(\frac{n}{2}) & \text{, if } 2|n \\ M_{ab-ab}(\frac{n+1}{2}) & \text{, if } 2\not|n \end{cases}; \quad (37)$$

$$M_{bbab-ab}(n) = \begin{cases} M_{bbab-ab}(\frac{n}{2})+M_{bbab-aa}(\frac{n}{2})+M_{abba-aa}(\frac{n}{2}+1)+N_{abba-aa}(\frac{n}{2}+1) & \text{, if } 2|n \\ M_{abba-ab}(\frac{n+1}{2})+M_{abba-aa}(\frac{n+1}{2})+M_{bbab-aa}(\frac{n+1}{2}) & \text{, if } 2\not|n \end{cases} ; \quad (38)$$

$$M_{bbab-aa}(n) = \begin{cases} M_{bbab-babb}(\frac{n}{2})+M_{abba-ab}(\frac{n}{2}+1)+N_{abba-ab}(\frac{n}{2}+1) & \text{, if } 2|n \\ M_{bbab-ab}(\frac{n+1}{2})+M_{abba-babb}(\frac{n+1}{2})+N_{bbab-ab}(\frac{n+1}{2}) & \text{, if } 2\not|n \end{cases} ; \quad (39)$$

$$N_{ab-ab}(n) = \begin{cases} 2N_{ab-aa}(\frac{n+1}{2}) & \text{, if } 2\not|n \\ 0 & \text{, if } 2|n \end{cases} ; \quad (40)$$

$$N_{ab-aa}(n) = \begin{cases} N_{ab-aa}(\frac{n+1}{2})+N_{ab-ab}(\frac{n+1}{2}) & \text{, if } 2\not|n \\ 0 & \text{, if } 2|n \end{cases} ; \quad (41)$$

$$N_{abba-aa}(n) = \begin{cases} N_{ab-ab}(\frac{n+1}{2}) & \text{, if } 2\not|n \\ 0 & \text{, if } 2|n \end{cases} ; \quad (42)$$

$$N_{abba-ab}(n) = \begin{cases} N_{ab-aa}(\frac{n+1}{2}) & \text{, if } 2\not|n \\ 0 & \text{, if } 2|n \end{cases} ; \quad (43)$$

$$N_{bbab-ab}(n) = \begin{cases} N_{abba-ab}(\frac{n+1}{2})+N_{abba-aa}(\frac{n+1}{2}) & \text{, if } 2\not|n \\ 0 & \text{, if } 2|n \end{cases} ; \quad (44)$$

$$N_{bbab-abba}(n) = \begin{cases} N_{abba-ab}(\frac{n+1}{2}) & \text{, if } 2\not|n \\ 0 & \text{, if } 2|n \end{cases} ; \quad (45)$$

$$M(n) = M_{ab-ab}(n)+2M_{ab-aa}(n)+M_{aa-aa}(n) ; \quad (46)$$

$$N(n) = N_{ab-ab}(n)+2N_{ab-aa}(n) ; \quad (47)$$

$$N_{aa-aa}(n) = 0 . \quad (48)$$

The number of above equations is reduced by the fact that if a word is 2^+-free (allmost 2^+-free) then also its mirror image is also 2^+-free (allmost 2^+-free). Thus,

$$M_X(n) = M_Y(n)$$

and

$$N_X(n) = N_Y(n),$$

where the sequence Y is the mirror image of X.

First we consider the number of allmost 2^+-free words having a given length.

Theorem II.4.1 *Let n be a non-negative integer. Then*

$$N(n)=\begin{cases}(n-1)/2 & \text{, if } n=2^{k+1}+1, \text{ for some } k\geq 1,\\ n-1 & \text{, if } n=3\cdot 2^{k+1}+1, \text{ for some } k\geq 0,\\ 2 & \text{, if } n=3,\\ 0 & \text{, otherwise.}\end{cases}$$

Proof Firstly, Theorem II.2.10 states that a word $w' \in \mathcal{LAWSF}$ is a subword of a word $\theta^k(www)$, where $w \in \{a, b, aab, bba\}$, and it has length $2^{k+1}|w|+1$, for some integer $k \geq 0$. So, it is sufficient to study only lengths in sequences

$$(2^k+1)_{k=2}^{\infty} \quad \text{and} \quad (3\cdot 2^k+1)_{k=1}^{\infty}.$$

In both of these sequences a number n is preceeded only by those numbers which are needed when calculating recursively the values of $N_{ab-aa}(n)$ and $N_{ab-ab}(n)$. Because these sequences contains only odd numbers, we can use the equations

$$N_{ab-ab}(2^{k+1}|w|+1) = 2N_{ab-aa}(2^k|w|+1)$$

and

$$N_{ab-aa}(2^{k+1}|w|+1) = N_{ab-aa}(2^k|w|+1) + N_{ab-ab}(2^k|w|+1),$$

for $k > 0$. Moreover, these equations holds also in the case $k = |w| = 1$ as we can easily verify.

By using a matrix, we can express this group of linear equations as

$$\begin{pmatrix} N_{ab-aa}(2^{k+1}|w|+1) \\ N_{ab-aa}(2^{k+1}|w|+1)\end{pmatrix} = \begin{pmatrix} 0 & 2 \\ 1 & 1\end{pmatrix}\begin{pmatrix} N_{ab-aa}(2^k|w|+1) \\ N_{ab-aa}(2^k|w|+1)\end{pmatrix}.$$

Let us denote this matrix by E. This matrix has the characteristic polynomial

$$c_E(\lambda) = \lambda^2 - \lambda - 2.$$

Then, by the Cayley-Hamilton Theorem we conclude that

$$N_{ab-ab}(2^{k+2}|w|+1) = N_{ab-ab}(2^{k+1}|w|+1) + 2N_{ab-ab}(2^k|w|+1)$$

and

$$N_{ab-aa}(2^{k+2}|w|+1) = N_{ab-aa}(2^{k+1}|w|+1) + 2N_{ab-aa}(2^k|w|+1).$$

Thus,

$$N_{ab-ab}(2^{k+1}|w|+1) = \alpha 2^k + \beta(-1)^k$$

and

$$N_{ab-aa}(2^{k+1}|w|+1) = \gamma 2^k + \delta(-1)^k,$$

for all $k \geq 0$ by some constants α, β, γ and δ. By using starting values given in Table II.4.1 we can calculate the exact values of $N(n)$ obtaining the result given in the above theorem. □

As we can see from previous theorem the function giving the number of allmost 2^+-free words having a given length is bounded from upside by linear function.

$$N_{ab-aa}(5) = 0 \quad N_{ab-aa}(9) = 2$$

$$N_{ab-ab}(5) = 2 \quad N_{ab-ab}(9) = 0$$

$$N_{ab-aa}(7) = 2 \quad N_{ab-aa}(13) = 4$$

$$N_{ab-ab}(7) = 2 \quad N_{ab-ab}(13) = 4$$

Table II.4.1 Starting values for the proof of Theorem II.2.1

In [RS] it is proved that the number of 2^+-free words having a given length can be bounded from above. Now, our next task is to show that the function giving the number of 2^+-free words having a given length cannot be bounded from both sides by two polynomial functions with same growth-factor. In order to do that we have some preliminary work to do. First we choose two number sequences which are basicly similar to those two sequences introduced previously. These two sequences to be chosen are

$$(5 \cdot 2^k + 1)_{k=1}^{\infty} \quad \text{and} \quad (\frac{11 \cdot 4^k + 1}{3})_{k=1}^{\infty}.$$

Next we assume that n has value which occurs in the former sequence. In this sequence all numbers are odd. On the other hand, we do not need values not occurring in this sequence when solving equations (30)-(39), so we can use those parts of equations (30)-(39) which are for odd values of n. Moreover, by simple number theoretical considerations, we conclude that there cannot be any 2^+-free words obtained from an allmost 2^+-free word, in the sense of Theorem II.2.11, having length in this sequence. Therefore, we do not need to consider the equations (40)-(45), (47) and (48). We can also ignore the equations (33) and (34) because they are not needed in equations (30)-(32) and (35)-(39).

A matrix F corresponding this obtained group of linear equations is

$$F = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

The corresponding charasteristic polynomial for this matrix is

$$c_F(\lambda) = \lambda^8 - \lambda^7 - 3\lambda^6 + \lambda^5 - 4\lambda^4 - 2\lambda^3 - 4\lambda + 4.$$

Next we consider the sequence

$$(\frac{11 \cdot 4^k + 1}{3})_{k=1}^{\infty}.$$

Now, when solving the equations (30)-(39) for these values of n, the values in the sequence

$$(\frac{11 \cdot 4^k + 1}{3} + 1)_{k=1}^{\infty}$$

are also needed. This is because the equations for even values of n can have two consecutive values of parameters in the right side. On the other hand, one of these values is allways odd, so we have only two consecutive values of parameters when applying equations recursively. Moreover, these values are chosen such way that we do not need to consider 2^+-free words which are generated from allmost 2^+-free words. Thus, we can drop those N_X terms from the equations (30)-(39).

Let us first define the vectors X_n and Y_n such that

$$X_n = \begin{pmatrix} M_{aa-ab}(n) \\ M_{aa-aa}(n) \\ M_{ab-ab}(n) \\ M_{bbab-babb}(n) \\ M_{abba-abba}(n) \\ M_{bbab-abba}(n) \\ M_{abba-ab}(n) \\ M_{abba-aa}(n) \\ M_{bbab-ab}(n) \\ M_{bbab-aa}(n) \end{pmatrix}$$

and

$$Y_n = (X_{n+1}, X_n)^T.$$

Next we introduce 20 equations (30′)-(39′) and (30″)-(39″), where first ten are for values $n+1$ and rest are for values n. These equations are obtained from equations (30)-(39) depending on the parity of the value of n. For example, the equation (33′) will be that part of the equation (33) which is for odd values of n if n is even, and vice versa. Thus, we obtain two possible different 20×20 matrices depending on the parity of n. These matrices shall be denoted by G and H and they satisfies the equations $Y_{2n-1} = GY_n$ and $Y_{2n} = HY_n$. From the equations (30)-(39) we conclude that these two matrices are

$$G = \begin{pmatrix} I & J \\ 0 & K \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} K & 0 \\ I & J \end{pmatrix},$$

where

$$I = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0
\end{pmatrix},$$

$$J = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

and

$$K = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Actually, the matrix F could be replaced by G in the previous considerations.

We can easily verify that the numbers in the sequence

$$(\frac{11 \cdot 4^k + 1}{3})_{k=1}^{\infty}$$

are all odd and the numbers in the sequence

$$((\frac{11 \cdot 4^k + 1}{3} + 1)/2)_{k=1}^{\infty}$$

are all even. Therefore, the equation

$$Y_{\frac{11 \cdot 4^k + 1}{3}} = (GH)^{k-1} Y_{15}.$$

holds. And, the corresponding characteristic polynomial c_{GH} for matrix GH is

$$\lambda^8(\lambda^{12} - 9\lambda^{11} + 16\lambda^{10} + 15\lambda^9 - 72\lambda^8 + 73\lambda^7 + 23\lambda^6 - \\ 139\lambda^5 + 173\lambda^4 + 4\lambda^3 - 153\lambda^2 + 102\lambda - 32).$$

II.4.2 Estimating the number of 2^+-free words

Next we are going to construct an upper and a lower bound for the function $M(n)$. We start by defining the functions

$$M_X^{max}(k) = \max\{M_X(n) | 2^k + 1 \leq n \leq 2^{k+1} + 1\}$$

and

$$M_X^{min}(k) = \min\{M_X(n) | 2^k + 1 \leq n \leq 2^{k+1} + 1\}.$$

Next we estimate the equations (30)-(48). Let us begin with construction of a lower bound for the sum

$$M^{min}(k) = M_{ab-ab}^{min}(k) + 2M_{ab-aa}^{min}(k) + M_{aa-aa}^{min}(k).$$

By the equations (30)-(32) we obtain, for $k \geq 3$,

$$M^{min}(k+1) \geq 2M^{min}(k) + 2M^{min}_{bbab-ab}(k) + 2M^{min}_{bbab-aa}(k).$$

On the other hand, we obtain the equations

$$M^{min}_{abba-ab}(k+1) + M^{min}_{abba-aa}(k+1) \geq M^{min}_{aa-ab}(k) + M^{min}_{ab-ab}(k),$$

and

$$M^{min}_{bbab-ab}(k+1) + M^{min}_{bbab-ab}(k+1) \geq$$
$$M^{min}_{bbab-ab}(k) + M^{min}_{bbab-ab}(k) + M^{min}_{abba-ab}(k) + M^{min}_{abba-aa}(k),$$

by the equations (36)-(39). Together these two formulas yields the equation

$$M^{min}_{bbab-ab}(k+1) + M^{min}_{bbab-ab}(k+1) \geq \sum_{i=1}^{k-C}(M^{min}_{aa-ab}(k-i) + M^{min}_{ab-ab}(k-i)),$$

where integer C is a constant which can be chosen freely.

Moreover, from the equations (31) and (32) we obtain

$$M^{min}_{ab-ab}(k) \geq M^{min}_{aa-aa}(k), \quad \text{for all } k \geq 3$$

yielding

$$M^{min}_{ab-aa}(k) + M^{min}_{ab-ab}(k) \geq \frac{1}{2}M^{min}(k).$$

Now, by combining the above results we conclude that

$$M^{min}(k+1) \geq 2M^{min}(k) + \sum_{i=3}^{k-2}(M^{min}(i)).$$

By fixing the value of k and marking both sides equal in the above formula, we obtain recursive equation

$$M_{MIN}(n+1) = 2M_{MIN}(n) + \sum_{i=2}^{k-3}(M_{MIN}(n-i)). \tag{49}$$

Let us assume that

$$\lambda(k) = \max\{\ \lambda \in \mathbb{R}\ \mid \lambda^{k-2} - 2\lambda^{k-3} - \sum_{i=0}^{k-5}\lambda^i = 0\ \}$$

and

$$D(\lambda) = \inf\left\{\ \frac{M^{min}(n)}{\lambda^n} \mid 3 \leq n \leq k\ \right\},$$

where $D(\lambda)$ is defined only for $\lambda > 1$. Then, by letting

$$M_{MIN}(n) = D(\lambda(k))\lambda(k)^n \leq M^{min}(n), \text{ for all } 3 \leq n \leq k,$$

we obtain the result

$$D(k)\lambda(k)^n = M_{MIN}(n) \leq M^{min}(n), \text{ for all } n \geq 3.$$

On the other hand, from the equation (49), we obtain

$$\alpha(\alpha-2)(\alpha-1) = 1 - \frac{1}{\alpha^{k-1}}.$$

Now, we are interested only about solutions of λ_k with absolute value greater than two. From the above equality we can see that, when the value of k grows, such solutions converges to the only such solution of the equation

$$\alpha(\alpha-2)(\alpha-1) = 1.$$

And this solution is

$$\alpha_1 = (\frac{1}{2} + \frac{\sqrt{69}}{18})^{\frac{1}{3}} + (\frac{1}{2} - \frac{\sqrt{69}}{18})^{\frac{1}{3}} + 1.$$

From what we have already shown we conclude that

$$M^{min}(k) \geq D(\lambda)\lambda^k, \quad \text{for all } k \geq 3 \text{ and } \lambda < \alpha_1. \tag{50}$$

Next we consider an upper bound for sum

$$M^{max}(k) = M^{max}_{ab-ab}(k) + 2M^{max}_{ab-aa}(k) + M^{max}_{aa-aa}(k).$$

From the definitions of the sets $M_X(n)$ we can straightforwardly conclude that

$$M^{max}_{bbab-babb}(k) \leq M^{max}_{bbab-aa}(k) + M^{max}_{bbab-ab}(k)$$

and

$$M^{max}_{bbab-aa}(k) + M^{max}_{bbab-ab}(k) \leq M^{max}_{aa-ab}(k) + M^{max}_{aa-aa} \leq \frac{1}{2}M^{max}(k).$$

From equations (30)-(32) we obtain

$$M^{max}(k+1) \leq 2M^{max}(k) + \frac{3}{2}M^{max}(k-1) + N^{max}(k),$$

where

$$N^{max}(k) = \max\{N(n) | 2^k + 1 \leq n \leq 2^{k+1} + 1\}.$$

Now we have an upper bound

$$M^{max}(k+1) \leq 2M^{max}(k) + \frac{3+\epsilon}{2} M^{max}(k-1), \text{ for } k > k_\epsilon,$$

for all arbitrary small positive real numbers ϵ. Marking both sides equal yields a charateristic polynomial

$$f(x) = x^2 - 2x - \frac{3+\epsilon}{2},$$

which has solutions

$$\alpha_2(\epsilon) = \frac{2+\sqrt{10+4\epsilon}}{2} \quad \text{and} \quad \alpha_3(\epsilon) = \frac{2-\sqrt{10+4\epsilon}}{2}.$$

From these solutions, α_2 has greater absolute value and, therefore, we conclude that

$$M^{max}(k) \leq E(\alpha)\alpha^k, \quad \text{for all } k \geq k(\alpha) \text{ and } \alpha < \alpha_2, \tag{51}$$

where $k(\alpha)$ and $E(\alpha)$ are properly chosen.

Now we are ready to state the following theorem:

Theorem II.4.2 *Let*

$$r = \sup\{\ t \mid \exists C > 0, M(n) \geq Cn^t, \forall n \geq 1\ \}$$

and

$$s = \inf\{\ t \mid \exists C > 0, M(n) \leq Cn^t, \forall n \geq 1\ \}.$$

Then

$$1.2170556 < r < 1.275380 < 1.332240 < s < 1.3680078.$$

Proof Let us first consider the polynomial c_F. By studying this polynomial, we observe that it has eight separate solutions from which four are not real. And, only one of the solutions of this polynomial has absolute value greater or equal than 2. This solution is $\lambda_1 \approx 2.420625$. From the equations (30)-(32) we conclude that the sequence $M(5 \cdot 2^k + 1)_{k=1}^{\infty}$ is not linearly bounded from above, so there exists some constants C and D such that

$$C\lambda_1^k \leq M(5 \cdot 2^k + 1) \leq D\lambda_1^k.$$

Thus,

$$r \leq \log_2(\lambda_1) < 1.275380.$$

Similarly, by studying the characteristic polynomial c_{GH}, we conclude that there are only one possible solution $\lambda_2 \approx 6.339992$ having absolute value greater than α^2. As before, we determine that the sequence $M(\frac{11 \cdot 4^k+1}{3})_{k=1}^{\infty}$ is not linearly bounded from above, and therefore there exists some constants C and D such that

$$C\lambda_2^k \leq M(\frac{11 \cdot 4^k + 1}{3}) \leq D\lambda_2^k.$$

This yields

$$s \geq \log_4(\lambda_2) > 1.332240.$$

Finally, from the equations (50) and (51) we conclude that

$$r \geq \log_2(\alpha_1) > 1.2170556$$

and

$$s \leq \log_2(\alpha_2) < 1.3680078.$$

This completes the proof. □

From this theorem we can easily conclude that

Theorem II.4.3 (Thue [Th]) *There are infinitely many 2^+-free words over binary alphabet.* □

Theorem II.4.2 improves the bounds given by Kobayashi ([Ko1]). These bounds were

$$r > 1.155 \quad \text{and} \quad s < 1.587.$$

The other two bounds are also obtained by Cassaigne ([Cas]). Next we discuss about infinite 2^+-free words.

II.5 Infinite 2^+-free words

In this section we introduce some structural and numerical results about infinite 2^+-free words. We also give proofs for some problems which are already solved.

II.5.1 Infinite 2^+-free words

First we state a theorem about the words which can be extended infinitely.

Theorem II.5.1 *Let w be an infinite 2^+-free word. Then this word can be expressed in the form*

$$w = (u_1)^+_{i_1}(u_2)^+_{i_2}(u_3)^+_{i_3}\dots \ , \tag{60}$$

$$w = (u_1)^+_{i_1}(u_2)^+_{i_2}\dots(u_m)^+_{i_m} w_{TM} \ , \tag{61}$$

or

$$w = (u_1)^+_{i_1}(u_2)^+_{i_2}\dots(u_m)^+_{i_m} \overline{w}_{TM} \ , \tag{61'}$$

where the sequences $(u_k)_{i_k}$ and $(i_k)_{j_k}$ can be obtained by Procedure A, a directed graph presented in Figure II.5.1 and actions described in Table II.5.1.

Proof Let us first construct that directed graph in Figure II.5.1.

First we take the directed graph from Figure II.3.3 and inverse its edges. Then we simplify this inversed graph.

Firstly, let w be an infinite 2^+-free word having an arbitrary long prefix v. On the other hand, there is also another prefix u of w such that

$$|u| >> \max\{\ |x| \mid x \in \mathcal{LAWSF} \cup \mathcal{LWSF}_{base} \setminus \{aaa, bbb\}\ \}|v|$$

. Because $u, v \in \mathcal{LWSF}$, we obtain by Theorems II.3.8 and II.3.9 and Proposition II.3.7, that, for all long enough 2^+-free words, a common sequence

$$h_1, h_2, \dots, h_m \ ,$$

ends the sequences

$$(h_{l,i})$$

given by Proposition II.3.7. Note, that in the sense of inversed graph we are interested about starting parts of paths in this graph.

Secondly, because the mappings $\theta, \sigma_l, \sigma_r, \rho_l$ and ρ_r maps the words ab, aa, $ba, bb \in \mathcal{LWSF}$ to words having length greater than 2, we can combine the states which cannot be leaved, i.e. the states corresponding the states

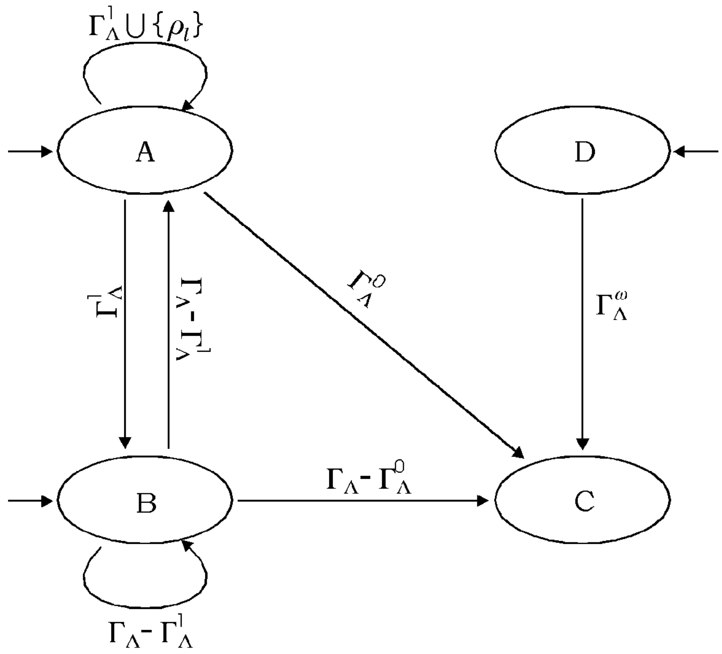

Figure II.5.1. Transition graph for infinite weakly square-free words

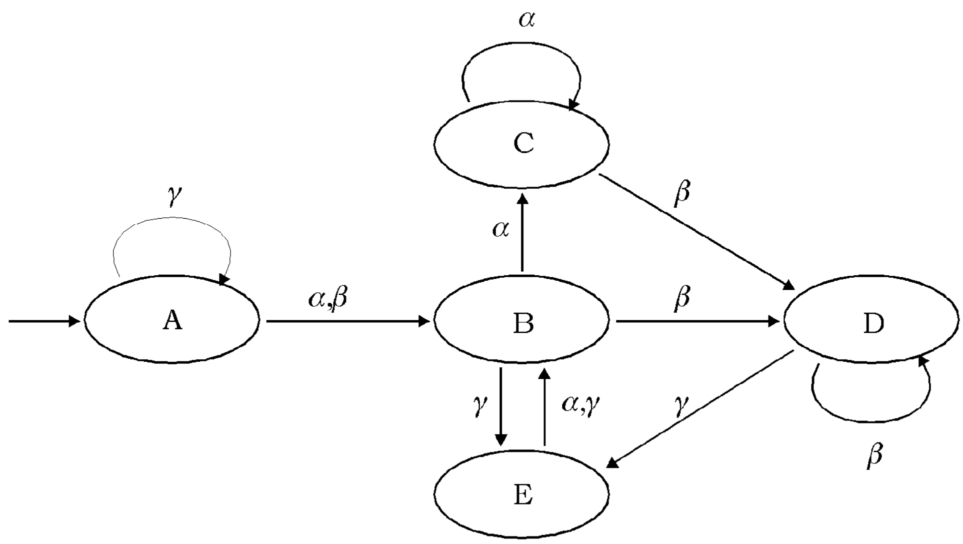

Figure II.5.2. Fife's automaton

$U = xyx$ and $U = xyy$ in Figure II.3.3. Moreover, we can also ignore the edges from the states $U = xxyx$ and $U = xxyy$ to the state B.

Finally, for infinite sequences of mappings, we have two basically different cases depending on the length of corresponding path in the inversed graph. These cases are:

1) infinitely long path in reversed graph;
2) finite path in reversed graph.

The trick is, that, in the latter case, the last edge of path in that inversed graph corresponds different sequences of mappings such that these sequences are prefixes of an infinite sequence of mappings. Thus, we obtain the graph presented in Figure II.5.1 giving us an infinite sequence of mappings corresponding a given infinite 2^+-free word.

Now, when considering paths in the graph of Figure II.5.1 corresponding infinite sequences of mappings, we have two different cases :

1) infinitely long fixed path;
2) finite fixed path.

Let P be an infinite path in this graph. Then every finite part, starting from the beginning, is a path which can be used in the reverse order to produce a 2^+-free word having same prefix by Theorem II.3.8. Thus, the directed graph in Figure II.5.1 gives us all possible infinite 2^+-free words of the form (60).

Next we assume that P is a finite path. Then the sequence of mappings corresponding the last edge in P for long prefixes is

$$\Lambda_{(k,l)}, \text{ where } \begin{cases} k \mapsto \infty & \text{, if } l \text{ is fixed,} \\ l \mapsto \infty & \text{, otherwise.} \end{cases}$$

In this case, Theorem II.3.8 states that corresponding infinite 2^+-free word has structure described in (61) and (61′).

Validity of Table II.5.1 follows directly from Table II.3.3. □

$B_{i+1} = \Lambda_{(k,l)}$:

$p_i = p_{i-1} - l - k - 1, q_i = \begin{cases} q_{i-1} & \text{, if } 2 \not| k, \\ \overline{q_{i-1}} & \text{, otherwise;} \end{cases}$

add $(q_i^+)^l$ to the end of the sequence $s_1, s_2, \ldots$;

add $p_{i-1} - k - 1, p_{i-1} - k - 2, \ldots, p_i + 1$ to the end of the sequence $r_1, r_2, \ldots$;

$B_{i+1} = \{\rho_l\}$:
  $p_i = p_{i-1} - 1$, $q_i = \overline{q_{i-1}}$;
  add q_i^+ to the end of the sequence $s_1, s_2, \ldots$;
  add p_i to the end of the sequence $r_1, r_2, \ldots$;

Table II.5.1 Actions on the graph of Figure II.5.1.

II.5.2 Some proofs

From the proof of Theorem II.5.1 we obtain, by checking graph in Figure II.5.1, a simple proof for the following two theorems (cf. Probl 2.2.3 in [Lo]).

Theorem II.5.2 (Kakutani) *There are uncountably many infinite 2^+-free words.*

Proof In the graph presented in Figure II.5.1 there are more than one edge from both nodes which can be visited infinitely many times. Thus, by combining different routes we can construct a one-to-one correspondence between real numbers in the interval $[0, 1]$ and a subset of the set of infinite 2^+-free words. □

Theorem II.5.3 *If a word is a prefix of an infinite 2^+-free word, Then there are uncountably many infinite 2^+-free words having this same prefix.*

Proof In the graph described in Figure II.5.1, the path only has a given prefix. Otherwise the proof is similar to the previous one. □

Corollary II.5.4 *Any finite 2^+-free word which occurs as a prefix of an infinite 2^+-free word is a prefix of a word of the form (60), (61) or (61′). Moreover, such words can be obtained by those Figure and Table mentioned in Theorem II.5.1.* □

From Corollary II.5.4 we conclude that we can effectively decide whether or not a finite 2^+-free word occurs as a prefix of an infinite 2^+-free word.

In the next section we are going to present briefly Fife's characterization of infinite 2^+-free words.

II.5.3 Fife's characterization of infinite 2^+-free words

Before we introduced a characterization of 2^+-free words based on Morse-blocks we constructed also a way to characterize 2^+-free words with two sequences of morphisms. An alternative way to do that characterization is to use some spesific actions as in Fife's work (cf. [Fi] and [Be]). Next are we going to present here this Fife's characterization of infinite 2^+-free words by following Berstel's work.

Let $\mathrm{Suff}_2(w) \in \{ab, ba\}$ and $x, y \in A^*$ be two words such that y is the longest Morse-block satisfying the equation

$$w = xy\overline{y}.$$

Next we define three actions $\alpha, \beta, \gamma : A^*\{ab, ba\} \to A^*\{ab, ba\}$ as follows:

$$\begin{array}{ccccccc} w \cdot \alpha & = & zy\overline{y} \cdot \alpha & = & zy\overline{y}y\overline{y} & = & = wy\overline{y}y\overline{y} \\ w \cdot \beta & = & zy\overline{y} \cdot \beta & = & zy\overline{y}y\overline{y}\overline{y}y & = & = wy\overline{y}\overline{y}y \\ w \cdot \gamma & = & zy\overline{y} \cdot \gamma & = & zy\overline{y}\overline{y}y & = & = w\overline{y}y \end{array}$$

Let us denote the set of these three actions by Q. Now, a *description* of the finite or infinite word x is a finite or infinite word $f \in Q^*$ which satisfies one of the equations

$$x = ab \cdot f, x = aab \cdot f, x = ba \cdot f \text{ and } x = bba \cdot f.$$

Now, it can be shown that every infinite 2^+-free word has a unique description (cf. Proposition 4.1. in [Be]). Now the following Theorem states the basic idea behind the Fife's characterization. For the proof see [Fi] or [Be].

Theorem II.5.5 (Fife's Theorem) *Let x be an infinite word over A. Then x is 2^+-free, if and only if,*

> *1) x starts with ab and its description is in $U = Q^\omega - Q^* I B^\omega$,*
>
> *or*
>
> *2) x starts with aab and its description is in $\{f \mid \beta f \in U\}$,*

where $I = \{\alpha, \beta\}(\gamma^2)^*\{\beta\alpha, \gamma\beta, \alpha\gamma\}$. □

Now, the set U can be recognized by an automaton with 5 states presented in Figure II.5.2.

References

[Be] J. Berstel: A Rewriting of Fife's Theorem about Overlap-free Words. In: Results and Trends in Theoretical Computer Science, J. Karhumäki, H. Maurer and G. Rozenberg (Eds.), Lecture Notes in Computer Science 812, Springer-Verlag (1994), 19-29.

[Br] F.-J. Brandenburg: Uniformly growing k-th power-free homomorphisms. Theoretical Computer Science 23 (1983), 69-82.

[Car] A. Carpi: Overlap-free words and finite automata. Theoretical Computer Science 115 (1993), 243-260.

[Cas] J. Cassaigne: Counting Overlap-Free Binary Words. In: STACS'93, Enjalbert, Finkel, Wagner (Eds.), Lecture Notes in Computer Science 665, Springer-Verlag (1993), 216-225.

[CK] K. Culik II and J. Karhumäki: Iterative devices generating infinite words. Lecture Notes in Computer Science 577, Springer-Verlag (1992), 531-544.

[Fi] E.D. Fife: Binary sequences which contain no BBb. Trans. Amer. Math. Soc. 261 (1980), 115-136.

[Ha] T. Harju: On cyclically overlap-free words in binary alphabets. In: The Book of L, G. Rozenberg and A. Salomaa (eds.), Springer-Verlag (1986), 125-130.

[Ko1] Y. Kobayashi: Enumeration of irreducible binary words. Discrete Applied Mathematics 20 (1988), 221-232.

[Ko2] Y. Kobayashi: Repetition free words. Theoretic. Comput. Sci. 44 (1986), 175-197.

[Lo] M. Lothaire: Combinatorics on Words. Addison-Wesley, Reading, Massachusetts 1981.

[MH] M. Morse and G.A. Hedlund: Unending chess, symbolic dynamics and a problem in semigroups. Duke Math. J. 11 (1944) 1-7.

[RS] A. Restivo and S. Salemi: On weakly square free words. Bulletin of EACTS 21 (1983), 49-56.

[Sa] A. Salomaa: Jewels of Formal Language Theory. Computer Science Press, (1980).

[Sé1] P. Séébold: Morphismes itérés, mot de Morse et mot de Fibonacci. C. R. Acad. Science 295 (1982), 439-441.

[Sé2] P. Séébold: Overlap-free sequences. Lecture Notes in Computer Science 192, 207-215.

[SS] R. Shelton and R. Soni: Chains and fixing blocks in irreducible sequences, Discrete Math. 54 (1985),93-99.

[Th1] A.Thue: Uber unendliche Zeichenreihen. Norske Vid. Selsk. Skr. I, Mat. Nat. KI., Kristiania 7 (1906), 1-22.

[Th2] A. Thue: Uber die gegenseitige Lage gleisher Teile gewisser Zeichernreihen. Norske Vid. Selsk. Skr. I, Mat. Nat. Kl. Kristiania 1 (1912), 1-67.

Turku Centre for Computer Science
Lemminkäisenkatu 14
FIN-20520 Turku
Finland

http://www.tucs.abo.fi

University of Turku
- **Department of Mathematical Sciences**

Åbo Akademi University
- **Department of Computer Science**
- **Institute for Advanced Management Systems Research**

Turku School of Economics and Business Administration
- **Institute of Information Systems Science**