International Journal of Foundations of Computer Science

INVOLUTION BORDERED WORDS

LILA KARI
Department of Computer Science,
University of Western Ontario,
London, ON N6A5B7

and

KALPANA MAHALINGAM
Department of Computer Science,
University of Western Ontario,
London, ON N6A5B7

ABSTRACT

In this paper we study a generalization of the classical notions of bordered and unbordered words. A nonempty word is called bordered if it has a proper prefix which is also a suffix of that word. A nonempty word is called unbordered if it is not bordered. We extend the notion of bordered and unbordered words to incorporate the notion of an involution function. (An involution function θ is such that θ^2 is the identity function.) We show that the set of all θ-bordered words is regular, when θ is an antimorphic involution and the set of all θ-bordered words is context sensitive when θ is a morphic involution. We study the properties of involution bordered and unbordered words and also the relation between involution bordered and unbordered words and certain type of involution codes. *

Keywords: Bordered words, DNA words, Watson-Crick Involution, DNA codes.

1. Introduction

The study of combinatorial properties of strings of symbols from a finite alphabet set is profoundly connected to numerous fields. In particular periodicity and borderedness are two basic word properties that play a role in many areas including string searching algorithms [5, 7, 8], data compression [9, 29] and in the study of coding properties of sets of words [1, 28, 30] and sequence assembly [26] in computational biology. A word u is called bordered if it has a proper prefix which is also its suffix. A word which is not bordered is called unbordered. Unbordered words have also been called dipolar words in [27], non-overlapping words and d-primitive words in [30] and d-minimal words in [31]. There are several classical results about bordered words. Several properties of bordered and unbordered words have been presented in [27, 30]. An authoritative text on the study of combinatorial properties of strings would be [25]. The relationship between the length of a word and the maximal length of its unbordered factors have been investigated in [12]. Factorizations of primitive words have been discussed in [17]. In [13], the authors define the border correlation function, which specifies the bordered conjugates (u is a conjugate of w if $uv = vw$ for some $v \in \Sigma^*$) of a given word w of length n and use it to study the relationship between unbordered conjugates and critical points. In [14], the authors

Figure 1: The word is θ-bordered which forms a hairpin with no sticky ends.

estimate the number of words that have a unique border. In [6], the author characterizes the biinfinite words in terms of their unbordered factors. A shorter proof was presented in [16]. A proof of the extended version of the Duval-Conjecture [10] which states that "Let u and v be words such that $u \neq v$, $|u| = |v| = n$ and u unbordered. Then uv contains an unbordered word of length atleast $n+1$" was given in [15]. The study of unbordered partial words was discussed in [2] and [3]. In [4], the authors have discussed the equations on partial words. The relation between monogenic expansion closed languages and unbordered words has been discussed in [27].

The stimulus for recent work on combinatorics of finite words is the study of molecules such as DNA that play a crucial role in molecular biology and biomolecular computation. Finding repeats or duplicated oligo nucleotides present as a string within the genome is an active research area in genomics. In [11], the authors have developed a computer program that identifies the periodic distribution of unique words. In this paper we study a generalization of the classical notions of bordered and unbordered words motivated by DNA based computing. We use an antimorphic involution map θ to formalize the notion of Watson-Crick complementarity of the DNA strands. We extend the study of bordered and unbordered words to θ-bordered and θ-unbordered words where θ is either a morphism or an antimorphism. The study of θ-unbordered words was initiated in [19] and [21] for an involution map θ. (An involution θ is such that θ^2 is identity). A word u is called θ-bordered if v is a proper prefix of u and $\theta(v)$ is a proper suffix of u. A word u is called θ-unbordered if u is not θ-bordered. A particular type of θ-bordered word as described in Fig.1, has non-overlapping θ-borders and such words form the well known hairpin structure. The words that avoid the hairpin structure were called θ-hairpin-free words in [19]. Another type of θ-bordered word has overlapping θ-borders such that the complement of a prefix of the word appears as a suffix of the word (See Fig.2).

In this paper we extend the properties of bordered and unbordered words [30] to θ-bordered and θ-unbordered words for θ either a morphic or an antimorphic involution. We begin the paper by reviewing basic concepts on words and introducing the definition of θ-bordered and θ-unbordered words. We define a relation $<_d^\theta$ such that $v <_d^\theta u$ iff v is a θ-border of u and also show that for an antimorphic involution the relation $<_d^\theta$ is transitive. In Section 3, we give a characterization of the set of all θ-bordered words when θ is an antimorphic involution and show that the set of all θ-unbordered words is a dense set. We also provide necessary and sufficient conditions for a word u to be θ-unbordered. In Section 4, we study the closure property of the set of all θ-unbordered words with respect to the catenation operation. In Section 5, we show that the set of all θ-bordered words is regular for an antimorphic involution θ and the set of all θ-bordered words is context-sensitive for a morphic involution θ. We discuss the relation between involution codes and

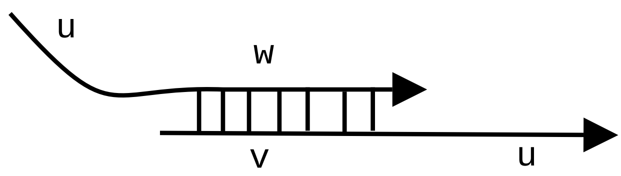

Figure 2: The word u has v as its prefix and $\theta(v) = w$ as its suffix and they overlap within u.

the sets of all θ-bordered and θ-unbordered words for a morphic or an antimorphic involution θ in Section 6. (For more on involution codes we refer the reader to [18, 19, 20, 21, 22].)

2. Basic concepts and properties

An alphabet Σ is a finite non-empty set of symbols. A word u over Σ is a finite sequence of symbols in Σ. We denote by Σ^* the set of all words over Σ, including the empty word λ and by Σ^+ the set of all non-empty words over Σ. We note that with the concatenation operation on words, Σ^* is the free monoid and Σ^+ is the free semigroup generated by Σ. For a word $w \in \Sigma^*$, the length of w is the number of non empty symbols in w and is denoted by $|w|$. Throughout the paper we assume that for an alphabet Σ, $|\Sigma| \geq 2$. In the following we review some known concepts. For a word w, the set of its proper prefixes, proper suffixes and proper subwords are defined as follows.

$$\begin{aligned}
\mathrm{PPref}(w) &= \{u \in \Sigma^+ \mid \exists v \in \Sigma^+, uv = w\}. \\
\mathrm{PSuff}(w) &= \{u \in \Sigma^+ \mid \exists v \in \Sigma^+, vu = w\}. \\
\mathrm{PSub}(w) &= \{u \in \Sigma^+ \mid \exists v_1, v_2 \in \Sigma^*, v_1 v_2 \neq \lambda, v_1 u v_2 = w\}.
\end{aligned}$$

Note that $\mathrm{Pref}(w) = \{u \in \Sigma^+ | \exists v \in \Sigma^*, w = uv\}$ and $\mathrm{Suff}(w) = \{u \in \Sigma^+ | \exists v \in \Sigma^*, w = vu\}$.

We also recall some partial orders, the notion of bordered and unbordered words and their relation to certain partial orders in the following. For more on these relations and bordered words we refer the reader to [30].

Definition 1

1. (*Prefix order*). *For $v, w \in \Sigma^*$, $w \leq_p v$ iff $v \in w\Sigma^*$.*
2. (*Suffix order*). *For $v, w \in \Sigma^*$, $w \leq_s v$ iff $v \in \Sigma^* w$.*
3. (*Division order*). *Define $\leq_d = \leq_p \cap \leq_s$.*
4. *For $u \in \Sigma^*$, $v \in \Sigma^*$ is said to be a border of u if $v \leq_d u$.*
5. *For $w, v \in \Sigma^*$, $w <_p v$ iff $v \in w\Sigma^+$.*
6. *For $w, v \in \Sigma^*$, $w <_s v$ iff $v \in \Sigma^+ w$.*
7. *$<_d = <_p \cap <_s$.*
8. *For $u \in \Sigma^*$, $v \in \Sigma^*$ is said to be a proper border of u if $v <_d u$.*
9. *For $u \in \Sigma^+$, define $L_d(u) = \{v | v \in \Sigma^*, v <_d u\}$.*
10. *$\nu(u) = |L_d(u)|$.*
11. *$D(i) = \{u | u \in \Sigma^+, \nu(u) = i\}$.*

12. *A word $u \in \Sigma^+$ is bordered if there exists $v \in \Sigma^+$ such that $v <_d u$, i.e., $u = vx = yv$ for some $x, y \in \Sigma^+$.*

13. *A non-empty word which is not bordered is called unbordered.*

Bordered words were initially called overlapped words and unbordered words were called non-overlapping words. Note that $D(1)$ is the set of all unbordered words.

Similar to the above definition, we define relations that involves either a morphic or an antimorphic involution θ. For properties of bordered and unbordered words we refer the reader to [30].

Definition 2 *Let θ be either a morphic or antimorphic involution on Σ^*.*

1. *For $v, w \in \Sigma^*$, $w \leq_p^\theta v$ iff $v \in \theta(w)\Sigma^*$.*

2. *For $v, w \in \Sigma^*$, $w \leq_s^\theta v$ iff $v \in \Sigma^*\theta(w)$.*

3. *$\leq_d^\theta = \leq_p \cap \leq_s^\theta$.*

4. *For $u \in \Sigma^*$, $v \in \Sigma^*$ is said to be a θ-border of u if $v \leq_d^\theta u$, i.e., $u = vx = y\theta(v)$.*

5. *For $w, v \in \Sigma^*$, $w <_p^\theta v$ iff $v \in \theta(w)\Sigma^+$.*

6. *For $w, v \in \Sigma^*$, $w <_s^\theta v$ iff $v \in \Sigma^+\theta(w)$.*

7. *$<_d^\theta = <_p \cap <_s^\theta$.*

8. *For $u \in \Sigma^*$, $v \in \Sigma^*$ is said to be a proper θ-border of u if $v <_d^\theta u$.*

9. *For $u \in \Sigma^+$, define $L_d^\theta(u) = \{v : v \in \Sigma^*, v <_d^\theta u\}$.*

10. *$\nu_\theta(u) = |L_d^\theta(u)|$.*

11. *$D_\theta(i) = \{u | u \in \Sigma^+, \nu_\theta(u) = i\}$.*

12. *A word $u \in \Sigma^+$ is said to be θ-bordered if there exists $v \in \Sigma^+$ such that $v <_d^\theta u$, i.e., $u = vx = y\theta(v)$ for some $x, y \in \Sigma^+$.*

13. *A non-empty word which is not θ-bordered is called θ-unbordered.*

Note that we call a word u to be θ-bordered if it has non empty θ-border .i.e., if it has a proper θ-border. Also note that the empty word λ is a θ-border of any word in Σ^+.

Example 2.1 *Let $u = ababab a$ be a word over the alphabet set $\{a, b\}$ and let θ be a morphic involution such that $\theta(a) = b$ and $\theta(b) = a$. Then $L_d^\theta(u) = \{\lambda, ab, abab, ababab\}$ and $\nu_\theta(u) = 4$, hence $u \in D_\theta(4)$.*

Based on the above definition we have the following observations.

Lemma 1 *Let θ be either morphic or an antimorphic involution.*

1. *$D_\theta(1)$ is the set of all θ-unbordered words.*

2. *A θ-bordered word $x \in \Sigma^+$ has length greater than or equal to 2.*

3. *For all $a \in \Sigma$, a is θ-unbordered.*

4. *For all $u \in \Sigma^+$ such that $u \neq \theta(u)$, $L_d^\theta(u) = \{v | v \in \Sigma^*, v \leq_d^\theta u\}$.*

5. *For all $a \in \Sigma$ such that $a \neq \theta(a)$, $a^+ \subseteq D_\theta(1)$.*

Recall that an involution is a map θ on Σ^* such that θ^2 is the identity map.

Lemma 2 *Let $u \in \Sigma^+$. Then for a morphic involution θ, $\theta(L_d^\theta(u)) = L_d^\theta(\theta(u))$ and when θ is an antimorphic involution we have, $L_d^\theta(u) = L_d^\theta(\theta(u))$.*

Proof. Let θ be a morphic involution and let $v \in L_d^\theta(u)$ which implies $u = vx = y\theta(v)$ for some $x, y \in \Sigma^+$ and hence $\theta(u) = \theta(v)\theta(x) = \theta(y)\theta(\theta(v))$ which implies $\theta(v) \in L_d^\theta(\theta(u))$. Thus $\theta(L_d^\theta(u)) \subseteq L_d^\theta(\theta(u))$. Similarly let $v \in L_d^\theta(\theta(u))$ which implies $\theta(u) = vx = y\theta(v)$ for some $x, y \in \Sigma^+$ and $u = \theta(v)\theta(x) = \theta(y)v$ which implies $\theta(v) \in L_d^\theta(u)$ and hence $v \in \theta(L_d^\theta(u))$. Thus $\theta(L_d^\theta(u)) = L_d^\theta(\theta(u))$.
Let θ be an antimorphic involution and let $v \in L_d^\theta(u)$, then $u = vx = y\theta(v)$ for some $x, y \in \Sigma^+$ which imply that $\theta(u) = \theta(x)\theta(v) = v\theta(y)$. Thus $v \in L_d^\theta(\theta(u))$. Similarly we can show that $L_d^\theta(\theta(u)) \subseteq L_d^\theta(u)$. Hence $L_d^\theta(u) = L_d^\theta(\theta(u))$. $\square$

Using the following lemma we show that the relation $<_d^\theta$ is transitive for an antimorphic involution θ.

Lemma 3 *Let $u \in \Sigma^*$ and $v, w \in \Sigma^+$ such that $u <_d^\theta w$ and $w <_d^\theta v$. Then for a morphic involution θ, we have $u <_d v$ and for an antimorphic involution θ, we have $u <_d^\theta v$.*

Proof. When θ is a morphic involution, $u <_d^\theta w$ and $w <_d^\theta v$ imply that $w = ux = y\theta(u)$ and $v = w\alpha = \beta\theta(w)$ for some $x, y, \alpha, \beta \in \Sigma^+$ which implies $v = ux\alpha = \beta\theta(y\theta(u))$ and hence $v = ux\alpha = \beta\theta(y)u$ which implies $u <_d v$.
When θ is an antimorphic involution, $u <_d^\theta w$ and $w <_d^\theta v$ imply that $w = ux = y\theta(u)$ and $v = w\alpha = \beta\theta(w)$ for some $x, y, \alpha, \beta \in \Sigma^+$ and hence $v = ux\alpha = \beta\theta(ux)$ which implies $v = ux\alpha = \beta\theta(x)\theta(u)$ implying that $u <_d^\theta v$. $\square$

Corollary 1 *If θ is an antimorphic involution, the relation $<_d^\theta$ is transitive.*

Lemma 4 *Let u, v, w be such that $u, v \in \Sigma^+$, $u \neq v$ and $u <_d^\theta w$, $v <_d^\theta w$. If θ is a morphic involution, then either $v <_d u$ or $u <_d v$. If θ is an antimorphic involution, then either $v <_p u$ or $u <_p v$.*

Proof. Let θ be a morphic involution and $u <_d^\theta w$, $v <_d^\theta w$ which imply that $w = ux = y\theta(u)$, $w = v\alpha = \beta\theta(v)$ for some $x, y, \alpha, \beta \in \Sigma^+$. If $|u| > |v|$, then $u = vp$ and $\theta(u) = q\theta(v)$ for some $p, q \in \Sigma^+$. Thus $u = \theta(q)v$ implying that $u = vp = \theta(q)v$ which implies $v <_d u$. If $|u| < |v|$ then $v = up$ and $\theta(v) = q\theta(u)$ for some $p, q \in \Sigma^+$ which imply that $v = \theta(q)u$. Therefore $v = up = \theta(q)u$ and hence $u <_d v$.
Let θ be an antimorphic involution and $u <_d^\theta w$, $v <_d^\theta w$ which imply that $w = ux = y\theta(u)$ and $w = v\alpha = \beta\theta(v)$ for some $x, y, \alpha, \beta \in \Sigma^+$. If $|u| > |v|$ then $u = vp$ and $\theta(u) = q\theta(v)$ for some $p, q \in \Sigma^+$ and hence $u = v\theta(q)$ which implies $v <_p u$. Similarly if $|v| > |u|$, we can show that $u <_p v$. $\square$

Corollary 2 *Let u, v, w be such that $u, v \in \Sigma^+$, $u \neq v$ and $u <_d^\theta w$, $v <_d^\theta w$. Then for an antimorphic involution θ, either $\theta(v) <_s \theta(u)$ or $\theta(u) <_s \theta(v)$.*

Corollary 3 *Let $u \in \Sigma^+$. Then*

1. *For a morphic involution θ, $L_d^\theta(u)$ is a totally ordered set with $<_d$.*

2. *For an antimorphic involution θ, $L_d^\theta(u)$ is a totally ordered set with $<_p$ and $\theta(L_d^\theta(u))$ is a totally ordered set with $<_s$.*

Lemma 5 *Let θ be a morphic involution. Then for all θ-unbordered words x, y such that $x \neq y$, $xy \neq \theta(y)x$.*

Proof. Let x, y be two θ-unbordered words, i.e., $x, y \in D_\theta(1)$. Note that both x and y are non empty as $D_\theta(i) \subseteq \Sigma^+$. Suppose $xy = \theta(y)x$ then we have the following cases to consider. If $|x| = |y|$ then $x = \theta(y)$ and $y = x$ a contradiction to our assumption that $x \neq y$. If $|x| > |y|$ then there exists $p \in \Sigma^+$ such that $x = \theta(y)p$ and $x = py$ which imply that $x = \theta(y)p = p\theta(\theta(y))$ since θ is an involution, which is a contradiction since x is θ-unbordered. If $|x| < |y|$ then there exists $q \in \Sigma^+$ such that $\theta(y) = xq$ and $y = qx$ which imply that $y = qx = \theta(x)\theta(q)$ since θ is a morphic involution, which is a contradiction since y is θ-unbordered. Thus $xy \neq \theta(y)x$. □

3. θ-bordered words

In the next result we give a characterization of the set of all θ-bordered words when θ is an antimorphic involution. We use this characterization to show several properties of the set of all θ-bordered and θ-unbordered words for an antimorphic involution θ.

Lemma 6 *Let θ be an antimorphic involution. Then $x \in \Sigma^+$ is θ-bordered iff $x = ay\theta(a)$ for some $a \in \Sigma$ and $y \in \Sigma^*$.*

Proof. If x is θ-bordered then $x = p\alpha = \beta\theta(p)$ for some $p, \alpha, \beta \in \Sigma^+$. Let $p = ar$ for some $a \in \Sigma$ and $r \in \Sigma^*$. Then $\theta(p) = \theta(r)\theta(a)$ and since $\alpha \in \Sigma^+$, we have $\alpha = s\theta(a)$ for some $s \in \Sigma^*$. Thus there exists $y \in \Sigma^*$ such that $x = ay\theta(a)$. The converse is obvious. □

We recall that a language or a set $X \subseteq \Sigma^*$ is said to be dense if for all $u \in \Sigma^*$, $X \cap \Sigma^* u \Sigma^* \neq \emptyset$. We use the above lemma to show that $D_\theta(1)$ is a dense set.

Corollary 4 *Let θ be an antimorphic involution on Σ^*. Then*

1. *$u \in D_\theta(1)$ iff $\theta(u) \in D_\theta(1)$.*
2. *If Σ is such that there exists $a, b \in \Sigma$ with $\theta(a) \neq b$ then $D_\theta(1)$ is a dense set.*
3. *Let $a, b \in \Sigma$ such that $\theta(a) = b$ then for all $u \in \Sigma^+$ either ua is θ-unbordered or ub is θ-unbordered.*
4. *If $uwv \in D_\theta(1)$ for some $u, v \in \Sigma^+$ and $w \in \Sigma^*$ then $uv \in D_\theta(1)$.*
5. *For all $a, b \in \Sigma$ such that $a \neq \theta(b)$, $a\Sigma^* b \subseteq D_\theta(1)$.*
6. *Let $u \in \Sigma^+$ be θ-bordered and x be the shortest θ-border of u, then x is θ-unbordered.*

Proof. We only prove the first two statements. The rest of them follow from Lemma 6. Let θ be an antimorphic involution on Σ^*.

1. Let $u \in D_\theta(1)$ and suppose $\theta(u) \notin D_\theta(1)$ then we have $\theta(u) = a\alpha\theta(a)$ for some $a \in \Sigma$ which imply that $u = a\theta(\alpha)\theta(a)$ and hence $u \notin D_\theta(1)$ a contradiction. The converse is similar.
2. Choose $a, b \in \Sigma$ such that $a \neq \theta(b)$ then for all $w \in \Sigma^*$ there exists $a, b \in \Sigma^*$ such that $awb \in D_\theta(1)$ which implies that $D_\theta(1)$ is a dense set.

□

Statement 6 in the above corollary does not hold true when θ is a morphism. For example let $\Sigma = \{a, b\}$ and θ be a morphism such that $\theta(a) = b$ and $\theta(b) = a$. Take $u = ababa$. The shortest θ-border of u is $x = ab$. But $x = ab = a.b = a.\theta(a)$ which is θ-bordered.

It was shown in [30] that when θ is identity and if x is the shortest border of u, then for all other borders $y \neq x$ of u, y is bordered. But this is not true when θ is an antimorphism, as shown by the following example.

Example 3.1 *Let $\Sigma = \{a, b, c\}$ and θ be antimorphism that maps $a \mapsto b$, $b \mapsto a$ and $c \mapsto c$. Then for $u = acacb$, we have $x = a$ to be the shortest θ-border of u. Also $y = ac$ is a θ-border of u as $\theta(ac) = cb$, but y is θ-unbordered.*

The following lemma relates the set of all prefixes and suffixes of a word with the set of all prefixes and suffixes of the set of all words obtained by concatenating the word with itself. We use the lemma to show some closure properties of the set of all θ-bordered and θ-unbordered words.

Lemma 7 *Let θ be a morphism or an antimorphism of Σ^* and let $u, v \in \Sigma^*$. Then $\theta(Pref(u)) \cap Suff(v) = \emptyset$ iff $\theta(Pref(u^+)) \cap Suff(v^+) = \emptyset$.*

Proof. "$\Rightarrow$" Assume that $\theta(Pref(u)) \cap Suff(v) = \emptyset$ and we need to show that $\theta(Pref(u^+)) \cap Suff(v^+) = \emptyset$. Suppose there exists $x \in \theta(Pref(u^+)) \cap Suff(v^+)$ then $x = \theta(u^k u_1) = v_2 v^l$ where $u_1 \in Pref(u)$ and $v_2 \in Suff(v)$. When θ is a morphism, we have $x = \theta(u^k)\theta(u_1) = v_2 v^l$ which implies that either $\theta(u_1)$ is a suffix of v or $\theta(u_1) = v' v^r$ for some $v' \in Suff(v)$ which imply that $\theta(u_1') = v'$ for some $u_1' \in Pref(u_1)$. Both cases lead to a contradiction since $\theta(Pref(u)) \cap Suff(v) = \emptyset$. The converse is obvious.
The case when θ is an antimorphism can be proved similarly. □

In the next lemma we give a necessary and sufficient condition for a word to be θ-unbordered. Note that it is clear from Lemma 6 that a word u is θ-unbordered for an antimorphic involution θ iff $u = ayb$ such that $a \neq \theta(b)$. The following lemma provides a much weaker characterization of θ-unbordered words. However this characterization can be used in proving certain closure properties of θ-unbordered words.

Lemma 8 *Let θ be an antimorphic involution on Σ^*. Then for all $u \in \Sigma^+$ such that $|u| \geq 2$, u is θ-unbordered iff $\theta(Pref(u)) \cap Suff(u) = \emptyset$.*

Proof. Let u be θ-unbordered. Suppose there exists $x \in \theta(Pref(u)) \cap Suff(u)$ then $x = \theta(u_1) = u''$ for some $u = u_1 u_2 = u' u''$ which imply that $u = u_1 u_2 = u'\theta(u_1)$. Then we have the following cases. If $u_2, u' \in \Sigma^+$ then $u \notin D_\theta(1)$ which is a contradiction since u is θ-unbordered. If $u_2 = u' = \lambda$ then $u = \theta(u)$ and $u = av$ for some $a \in \Sigma$ and $v \in \Sigma^+$ since $|u| \geq 2$ which imply that $u = av = \theta(v)\theta(a) = \theta(u)$ which is a contradiction since u is θ-unbordered. Hence $\theta(Pref(u)) \cap Suff(u) = \emptyset$. Conversely assume that $\theta(Pref(u)) \cap Suff(u) = \emptyset$ and suppose u is θ-bordered then there exists $y \in \Sigma^*$ and $a \in \Sigma$ such that $u = ay\theta(a)$ which is a contradiction since $\theta(a) \in \theta(Pref(u)) \cap Suff(u)$. □

Corollary 5 *Let θ be an antimorphic involution on Σ^* and let $u \in \Sigma^+$ such that $|u| \geq 2$. Then u is θ-unbordered iff $u^+ \subseteq D_\theta(1)$.*

Proof. Follows from Lemma 8 and Lemma 7. □

Lemma 9 *Let θ be a morphic involution on Σ^*. Then for all $u \in \Sigma^+$ such that $|u| \geq 2$ and $u \neq \theta(u)$, u is θ-unbordered iff $\theta(Pref(u)) \cap Suff(u) = \emptyset$.*

Proof. Let $u \in D_\theta(1)$ such that $|u| \geq 2$ and $u \neq \theta(u)$. Suppose there exists an $x \in \theta(Pref(u)) \cap Suff(u)$ then we have the following cases. If $x = \theta(u)$ then $x = u \in Suff(u)$ which implies that $u = \theta(u)$ which is a contradiction. If $x = \theta(u_1)$ for some $u_1, u_2 \in \Sigma^+$ such that $u = u_1 u_2$ and $u = u_1 u_2 = u'\theta(u_1)$ since $x \in Suff(u)$ which is a contradiction since u is θ-unbordered. □

Corollary 6 *Let θ be a morphic involution on Σ^* and let $u \in \Sigma^+$ such that $|u| \geq 2$ and $u \neq \theta(u)$. Then u is θ-unbordered iff $u^+ \subseteq D_\theta(1)$.*

Proof. Follows from Lemma 9 and 7. □

In view of Lemma 8 and Lemma 9 we have the following observation. The proof of the following lemma is similar to that of the above two lemmas and hence we omit the proof.

Lemma 10 *Let θ be either a morphic or an antimorphic involution. Then for $u \in \Sigma^+$ such that $|u| \geq 2$, u is θ-unbordered iff $\theta(PPref(u)) \cap PSuff(u) = \emptyset$.*

4. Closure properties of the set of all θ-unbordered words

In the next proposition we give a necessary and sufficient condition for the set of all θ-unbordered words to be closed under concatenation.

Proposition 1 *Let θ be either a morphic or an antimorphic involution and let $u, v \in \Sigma^+$ be θ-unbordered. Then uv is θ-unbordered iff $\theta(Pref(u)) \cap Suff(v) = \emptyset$.*

Proof. Assume that for $u, v \in \Sigma^+$ such that $|uv| \geq 2$, $\theta(Pref(u)) \cap Suff(v) = \emptyset$ and suppose uv is not θ-unbordered.
Then for an antimorphic involution θ, we have by Lemma 6, $uv = ay\theta(a)$ for some $a \in \Sigma$ and $y \in \Sigma^*$. Then $a \in Pref(u)$ and $\theta(a) \in Suff(v)$ which implies that $\theta(a) \in \theta(Pref(u)) \cap Suff(v)$ which is a contradiction. Hence uv is θ-unbordered.
When θ is a morphism, then there exists $x \in \Sigma^+$ such that $uv = x\alpha = \beta\theta(x)$ for some $\alpha, \beta \in \Sigma^+$. We have the following cases:
(i) $|x| \leq |u|$ and $|\theta(x)| \leq |v|$
(ii) $|x| \leq |u|$ and $|\theta(x)| > |v|$
(iii) $|x| > |u|$ and $|\theta(x)| \leq |v|$
(iv) $|x| > |u|$ and $|\theta(x)| > |v|$

Note that case(i) implies that $x \in \mathrm{Pref}(u)$ and $\theta(x) \in \mathrm{Suff}(v)$ which immediately leads to a contradiction since $x \in Pref(u)$ and $\theta(x) \in \theta(Pref(u)) \cap Suff(v)$.

Case(ii) implies that $x \in \mathrm{Pref}(u)$, $\theta(x) \in \mathrm{Suff}(uv)$ and $\theta(x) \notin \mathrm{Suff}(u)$ and hence $x = u_1$ for some $u_1 \in \Sigma^+$ and $u_2 \in \Sigma^*$ such that $u = u_1u_2$ and $\theta(x) \in Suff(v)$ implies that $\theta(x) = u''v$ for some $u' \in \Sigma^+$ and $u'' \in \Sigma^*$ such that $u = u'u''$. Thus $x = \theta(u'')\theta(v) = u_1$ which imply that $\theta(u'') \in Pref(u)$ and $u = \theta(u'')y = u'u''$ with $y, u' \in \Sigma^+$ since $v \in \Sigma^+$, which is a contradiction since u is θ-unbordered.

Case(iii) implies that $x \in \mathrm{Pref}(uv)$, $\theta(x) \in \mathrm{Suff}(v)$ and $x \notin \mathrm{Pref}(u)$ and hence $x = uv_1$ for some $v_1 \in \Sigma^+$ and $v = v_1v_2$ with $v_2 \in \Sigma^+$ and $\theta(x) \in Suff(v)$ implies that $\theta(x) = v''$ for some $v'' \in \Sigma^+$, $v' \in \Sigma^*$ with $v = v'v''$. Thus for $x = uv_1$, $\theta(x) = \theta(u)\theta(v_1) = v''$ which implies that $v = v_1v_2 = y\theta(v_1)$ with $v_2, y \in \Sigma^+$ since $u \in \Sigma^+$ which is a contradiction since v is θ-unbordered.

Case(iv) implies that $x \in \mathrm{Pref}(uv)$ and $\theta(x) \in \mathrm{Suff}(uv)$ but none of the above hold. $x \in Pref(uv)$ implies that $x = uv_1$ for some $v_1, v_2 \in \Sigma^+$ with $v = v_1v_2$ and $\theta(x) \in Suff(uv)$ implies that $\theta(x) = u_2v$ for some $u_1, u_2 \in \Sigma^+$ with $u = u_1u_2$. Thus for $x = uv_1$, $\theta(x) = \theta(u)\theta(v_1) = u_2v$. If $u = u'u''$ then $\theta(u)\theta(v_1) = \theta(u')\theta(u'')\theta(v_1) = u_2v$ such that $\theta(u') = u_2$ which imply that $u = u'u'' = u_1\theta(u')$ with $u', u'', u_1 \in \Sigma^+$ which is a contradiction since u is θ-unbordered. Hence uv is θ-unbordered.

Conversely for u, v both θ-unbordered and $|uv| \geq 2$, assume that uv is also θ-unbordered. Suppose there exists $x \in \theta(Pref(u)) \cap Suff(v)$ such that $x = \theta(u_1) = v_2$ for $u = u_1u_2$ and $v = v_1v_2$ with $u_1, v_2 \in \Sigma^+$ and $u_2, v_1 \in \Sigma^*$. Then $uv = u_1u_2v_1v_2 = u_1u_2v_1\theta(u_1)$ which is a contradiction since uv is θ-unbordered. Hence $\theta(Pref(u)) \cap Suff(v) = \emptyset$. □

Lemma 11 *Let θ be either a morphic or an antimorphic involution on Σ^* and let $u, v \in \Sigma^+$ with both u and v θ-unbordered and non $\theta(u) \neq u$, $\theta(v) \neq v$. Then the following are equivalent.*

1. *uv is θ-unbordered.*
2. *The set of all words in u^+v^+ is θ-unbordered.*
3. *$\theta(Pref(u)) \cap Suff(v) = \emptyset$.*

4. *For all $x \in (uv)^+$, x is θ-unbordered.*

Proof. Note that from Proposition 1 it is clear that $1 \Leftrightarrow 3$. From Lemma 7 and Proposition 1 it is clear that $1 \Leftrightarrow 2$. Note that from Lemma 8 $uv \in D_\theta(1)$ iff $\theta(Pref(uv)) \cap Suff(uv) = \emptyset$. Also from Lemma 7 $\theta(Pref(uv)) \cap Suff(uv) = \emptyset$ iff $\theta(Pref((uv)^+)) \cap Suff((uv)^+) = \emptyset$. Hence from Proposition 1 $\theta(Pref((uv)^+)) \cap Suff((uv)^+) = \emptyset$ iff $(uv)^+ \subseteq D_\theta(1)$. Hence $1 \Leftrightarrow 4$. □

We use the following result from [23] to prove the next result.

Lemma 12 *([23]) Let u and w be such that $uv = \theta(v)w$ for some $v \in \Sigma^*$. Then for a morphic involution θ there exists $x, y \in \Sigma^*$ such that $u = xy$ and one of the following hold*

1. *If $|u| > |v|$ then $w = y\theta(x)$ and $v = (\theta(x)\theta(y)xy)^i\theta(x)$ for $i \geq 0$.*

2. *If $|u| < |v|$ then $w = \theta(y)x$ and $v = (\theta(x)\theta(y)xy)^i\theta(x)\theta(y)x$ for $i \geq 0$.*

Proposition 2 *Let $x_1, x_2 \in \Sigma^+$ and θ be either a morphic or an antimorphic involution. If x_1x_2 is θ-unbordered, then for any $k > 1$, $x_1x_2^k$ is θ-unbordered.*

Proof. We first consider the case when θ is an antimorphism. Suppose that, for some $k > 1$, $x_1x_2^k$ is θ-bordered, then from Lemma 6, there exists $a \in \Sigma$ and $y \in \Sigma^*$, $x_1x_2^k = ay\theta(a)$. Since both $x_1, x_2 \in \Sigma^+$ we have $x_1x_2 = ax\theta(a)$ for some $x \in \Sigma^*$ which is a contradiction since x_1x_2 is θ-unbordered. Hence $x_1x_2^k$ is θ-unbordered. We shall prove by induction on k the case when θ is morphism.

Base Case: Let $k = 2$. Suppose $x_1x_2^2$ is θ-bordered. Then there exists $x, y, u \in \Sigma^+$ such that $x_1x_2^2 = ux = y\theta(u)$. We have several cases:

Case 1 Let $|u| \leq |x_1|$ then we have $x_1 = u\alpha$ for some $\alpha \in \Sigma^*$.

- If $|\theta(u)| \leq |x_2|$ then $x_2 = \beta\theta(u)$ for some $\beta \in \Sigma^*$ and $x_1x_2 = u\alpha\beta\theta(u)$ with $u \in \Sigma^+$, which is a contradiction since x_1x_2 is θ-unbordered.

- If $|x_2| < |\theta(u)| \leq |x_2^2|$ then $\theta(u) = \beta_1x_2$ for some $x_2 = \beta\beta_1$ with $\beta_1 \in \Sigma^+$. Thus $u = \theta(\beta_1)\theta(x_2)$ and $x_1x_2 = u\alpha x_2 = u\alpha\beta\beta_1 = \theta(\beta_1)\theta(x_2)\alpha\beta\beta_1$, which is a contradiction since x_1x_2 is θ-unbordered.

- If $|\theta(u)| > |x_2^2|$ then $\theta(u) = \beta_1x_2^2$ with $x_1 = \beta\beta_1$ and $\beta_1 \in \Sigma^+$. Thus $u = \theta(\beta_1)\theta(x_2^2)$ and $x_1 = u\alpha = \beta\beta_1$ which implies that $x_1x_2 = \theta(\beta_1)\theta(x_2^2)\alpha x_2 = \beta\beta_1x_2$ which imply that $x_1x_2 = \theta(\beta_1x_2)\theta(x_2)\alpha x_2 = \beta(\beta_1x_2)$ which is a contradiction since x_1x_2 is θ-unbordered.

Case 2 Let $|x_1| \leq |u| \leq |x_1x_2|$ then we have $u\alpha = x_1x_2$ for some $\alpha \in \Sigma^*$.

- If $|\theta(u)| \leq |x_2|$ then $\beta_1\theta(u) = x_2$ which implies $x_1x_2 = u\alpha = x_1\beta_1\theta(u)$ a contradiction.

- If $|x_2| \leq |\theta(u)| \leq |x_2x_2|$ then $x_1x_2 = u\alpha$ and $\theta(u) = \beta_1x_2$ for $x_2 = \beta\beta_1$. As θ is a morphism, $x_1x_2 = u\alpha = \theta(\beta_1)\theta(x_2)\alpha$ which imply that $x_1x_2 = x_1\beta\beta_1 = \theta(\beta_1)\theta(x_2)\alpha$ a contradiction.

- If $|x_2x_2| \leq |\theta(u)| \leq |x_1x_2x_2|$, then $x_1x_2 = ua$ and $\theta(u) = s_1x_2x_2$ for $x_1 = ss_1$. Then we have $x_1x_2 = u\alpha = \theta(s_1)\theta(x_2)\theta(x_2)\alpha$ and hence $x_1x_2 = ss_1x_2 = \theta(s_1)\theta(x_2)\theta(x_2)\alpha$ a contradiction.

Case 3 Let $|x_1x_2| < |u| < |x_1x_2x_2|$. If $|x_2| \leq |\theta(u)| \leq |x_2x_2|$ then we have $u = x_1x_2\beta$ with $x_2 = \beta\beta_1$ and $\theta(u) = s_1x_2$ for $x_2 = ss_1$. Then we have $u = x_1x_2\beta = \theta(s_1)\theta(x_2)$.

Note that $|x_1\beta| = |s_1|$ hence $\theta(s_1) = x_1r$, $x_2 = rp$ and $\theta(x_2) = p\beta$ which implies $x_1x_2\beta = \theta(s_1)p\beta$ which imply that $x_1x_2 = x_1ss_1 = \theta(s_1)p$ a contradiction. If $|x_2x_2| \leq |\theta(u)| \leq |x_1x_2x_2|$ then $\theta(u) = s_1x_2x_2$ and $u = x_1x_2\beta$ for $x_1 = ss_1$ and $x_2 = \beta\beta_1$ with $s, s_1, \beta, \beta_1 \in \Sigma^+$. Then $u = x_1x_2\beta = \theta(s_1)\theta(x_2)\theta(x_2)$ which implies that $u = x_1\beta\beta_1\beta = \theta(s_1)\theta(x_2)\theta(x_2)$ and by the length argument we have $\theta(x_2) = \beta_1\beta$ and hence $x_2 = \beta\beta_1 = \theta(\beta_1)\theta(\beta)$ or $\beta_1\beta = \theta(\beta)\theta(\beta_1)$. Thus $x_1\beta = \theta(s_1)\theta(x_2)$ which implies that $x_1x_2 = ss_1\theta(\beta_1)\theta(\beta) = \theta(s_1)\beta_1\beta\beta_1$ which is a contradiction since x_1x_2 is θ-unbordered. Hence we have $x_1x_2^2 \in D_\theta(1)$.
Induction Step Assume $x_1x_2^k \in D_\theta(1)$. Suppose $x_1x_2^{k+1} \notin D_\theta(1)$, then we have $x_1x_2^{k+1} = ux = y\theta(u)$ for some $x, y \in \Sigma^+$.
Case 1: Let u be such that $|x_1x_2^k| < |\theta(u)| < |x_1x_2^{k+1}|$ then $\theta(u) = \alpha_1x_2^{k+1}$ for some $\alpha_1 \in \Sigma^+$ such that $x_1 = \alpha\alpha_1$. If $|x_1x_2^k| < |u| < |x_1x_2^{k+1}|$, then $u = x_1x_2^k\beta$ for some $\beta \in \Sigma^+$ such that $x_2 = \beta\beta_1$. Hence $u = \theta(\alpha_1)\theta(x_2^k)\theta(x_2) = \theta(\alpha_1)\theta(x_2^k)\theta(\beta)\theta(\beta_1) = x_1x_2\beta\beta_1\beta$. Thus $x_1x_2^{k-1}\beta = \theta(\alpha_1)\theta(x_2^k)$ and hence $x_1x_2^k = \theta(\alpha_1)\theta(x_2^k)\beta_1 = \alpha\alpha_1x_2^k$ which is a contradiction since $x_1x_2^k$ is θ-unbordered. If $|u| \leq |x_1x_2^k|$ then $u = x_1x_2^i\beta$ for some $i < k$ and $x_2 = \beta\beta_1$ for some $\beta, \beta_1 \in \Sigma^*$. Thus $u = \theta(\alpha_1)\theta(x_2^{k+1})$ which implies that $x_1x_2^i\beta = \theta(\alpha_1)\theta(x_2^k)\theta(\beta)\theta(\beta_1)$ and hence $x_1x_2^{i-1}\beta = \theta(\alpha_1)\theta(x_2^k)$. Therefore $x_1x_2^k = \theta(\alpha_1)\theta(x_2^k)\beta_1x_2^{k-i} = \alpha\alpha_1x_2^k$ a contradiction since $x_1x_2^k$ is θ-unbordered.
Case 2: Let u be such that $|\theta(u)| \leq |x_2^{k+1}|$. Then $\theta(u) = \beta_1x_2^i$ with $x_2 = \beta\beta_1$ and $i \leq k$ and $\beta, \beta_1 \in \Sigma^*$. If $|x_1x_2^k| < |u| < |x_1x_2^{k+1}|$ then $u = x_1x_2^k\alpha$ with $x_2 = \alpha\alpha_1$ and $\alpha_1 \in \Sigma^+$. Hence $u = \theta(\beta_1)\theta(x_2^{i-1})\theta(x_2) = x_1x_2^{k-1}\alpha\alpha_1\alpha$ which implies that $x_1x_2^{k-1}\alpha = \theta(\beta_1)\theta(x_2^{i-1})$. Therefore $x_1x_2^k = \theta(\beta_1)\theta(x_2^{i-1})\alpha_1 = x_1x_2^{k-1}\beta\beta_1$ a contradiction. If $|u| \leq |x_1x_2^k|$ then $u = x_1x_2^j\alpha$ with $x_2 = \alpha\alpha_1$, $\alpha_1 \in \Sigma^*$ and $j < k$. Thus $x_1x_2 = x_1x_2^j\alpha\alpha_1x_2^{k-j-1} = \theta(\beta_1)\theta(x_2^i)\alpha_1x_2^{k-j-1}$ which implies that $x_1x_2^k = x_1x_2^{k-1}\beta\beta_1 = \theta(\beta_1)\theta(x_2^i)\alpha_1x_2^{k-j-1}$ a contradiction since $x_1x_2^k$ is θ-unbordered. Hence $x_1x_2^k \in D_\theta(1)$ for all $k > 1$. □

The proof of the next proposition is similar to that of the previous one and hence we omit the proof.

Proposition 3 *Let $x_1, x_2 \in \Sigma^+$ and θ be either morphic or an antimorphic involution. If x_1x_2 is θ-unbordered, then for any $k > 1$, $x_1^kx_2$ is θ-unbordered.*

Proposition 4 *Let θ be an antimorphic involution and let v be θ-unbordered. Then for all $v_p \in PPref(v)$ and $v_s \in PSuff(v)$, v_puv_s is θ-unbordered for all $u \in \Sigma^*$.*

Proof. Let $x \in v_p\Sigma^*v_s$ such that x is θ-bordered. Then there exists $a \in \Sigma$ and $y \in \Sigma^*$ such that $x = ay\theta(a)$ which implies that $a \in Pref(v_p)$ and $\theta(a) \in Suff(v_s)$. Thus there exists $z \in \Sigma^*$ such that $v = az\theta(a)$ which is a contradiction since v is θ-unbordered. Hence x is also θ-unbordered. □

Note that the above lemma does not hold when θ is a morphic involution. For example, let $\Sigma = \{a, b\}$ such that $\theta(a) = b$ and θ is a morphism. Note that $aa, b \in D_\theta(1)$ but $aba = (ab)a = a\theta(ab)$ and hence $aba \notin D_\theta(1)$.

Proposition 5 *Let θ be a morphic or an antimorphic involution and v be θ-unbordered.*
1. If $u = v_0v_1...v_{n-1}$ for some $v_i \in PPref(v)$, then $uv \in D_\theta(1)$.
2. If $u = v_0v_1...v_{n-1}$ for some $v_i \in PSuff(v)$, then $vu \in D_\theta(1)$.

Proof. We prove the first case (the second one is similar to the first case). The case when θ is an antimorphic involution follows directly from Proposition 4. We only consider the case when θ is a morphism. Let $v \in D_\theta(1)$ such that $|v| \geq 2$ and let $u = v_0v_1...v_{n-1}$ for some $v_i \in$ PPref(v). Suppose uv is θ-bordered, then there exists $x, \alpha, \beta \in \Sigma^+$ such that $uv = x\alpha = \beta\theta(x)$.

1. If $|x| > |u|$ then there exists $v', v'' \in \Sigma^+$ such that $v = v'v''$ and $x = uv'$ then

we have $uv = uv'v'' = \beta\theta(uv') = \beta\theta(u)\theta(v')$ which implies that $v = v'v'' = r\theta(v')$ for some $r \in \Sigma^+$ a contradiction since v is θ-unbordered.

2. If $|x| \leq |u|$ then there exists α_1, α_2 such that $\alpha_1 \in \Sigma^*$ and $\alpha_2 \in \Sigma^+$ and $u = x\alpha_1$, $v = \alpha_2$. If $|x| < |v_0|$ then there exists $p \in \Sigma^+$ such that $v_0 = xp$ which implies that $x \in \mathrm{PPref}(v)$ and hence $v = xr = s\theta(x)$ for some $r, s \in \Sigma^+$ which is a contradiction. If $|x| \geq |v_0|$ then there exists $p_1, p \in \Sigma^*$ such that $x = v_0pp_1$ and $p = v_1..v_k$ for some k and $p_1 \in \mathrm{PPref}(v_{k+1})$ with $|p_1| < |v|$. Hence $uv = x\alpha = \beta\theta(x) = v_0pp_1\alpha = \beta\theta(v_0)\theta(p)\theta(p_1)$ which implies that $v = p_1r = s\theta(p_1)$ a contradiction.

□

5. Classification of the set of all θ-bordered words

In this section we show that the set of all θ-bordered words is regular when θ is an antimorphic involution and context sensitive when θ is a morphic involution. In the next proposition we use Lemma 7 and show that the set of all θ-unbordered words is indeed a regular language when θ is an antimorphic involution.

Proposition 6 *When θ is an antimorphic involution on Σ^*, $D_\theta(1)$ is a regular language.*

Proof. Note that for all $a \in \Sigma$, a is θ-unbordered and from Lemma 7, we have $D_\theta(1) = \Sigma \cup Y$ where $Y = \bigcup_{a,b\in\Sigma} a\Sigma^*b$ such that $\theta(a) \neq b$. Since Σ is finite, Y is regular and hence $D_\theta(1)$ is regular. □

In the next proposition we find an example of θ, which is a morphic involution but not the identity function and an alphabet Σ such that the set of all θ-bordered words over Σ is not context free and hence not regular.

Proposition 7 *If θ is a morphic involution over an alphabet Σ, such that θ is not identity, the set of all θ-bordered words over Σ is not context free.*

Proof. Let $a, b \in \Sigma$ such that $a \neq b$ and $\theta(a) = b$. Then $\theta(b) = a$ holds because θ is an involution map. Denote by L the set of all θ-bordered words over Σ. We will prove, by contradiction, that L is not context-free.

Indeed, assume L were context-free. Let n be the constant defined by the Pumping Lemma for context-free languages. Choose the word $z_1 = a^{n+1}b^{n+1}a^{n+1}$, which is clearly θ-bordered. By the pumping lemma, there is a decomposition $z_1 = \alpha xvy\beta$ such that $|xvy| \leq n$, $|xy| \geq 1$, and for all $i \geq 0$, $z_i = \alpha x^ivy^i\beta \in L$. Note that any θ-border w_i of z_i has the property $w_i = au$ for some $u \in \Sigma^*$ because z_i begins with a for any $i \geq 0$.

We will consider first the case where xvy is a subword of $a^{n+1}b^{n+1}$ of z_1. In this case, $\theta(w_i) = b\Sigma^*a^{n+1}$ for any $i \geq 0$ because z_i has the suffix a^{n+1}. Consequently, $w_i \in a\Sigma^*b^{n+1}$. If neither x nor y contains any bs, that is, xvy is in the prefix a^{n+1} of z_1, $z_i = a^mb^{n+1}a^{n+1}$ for $i \geq 2$, where $m > n+1$. Considering the form of w_i mentioned above, $w_i = a^mb^{n+1}$. This further implies $\theta(w_i) = b^ma^{n+1}$, which is a contradiction since z_i does not contain m consecutive bs. Consequently, x or y must include at least one letter b. However, in this case z_0 has at most n letters b which contradicts the fact that z_0 has $w_0 = aub^{n+1}$ for $u \in \Sigma^*$ as its θ-border.

By virtue of the symmetric form of z_1, it is clear that the second case, where xvy occurs as a subword of b^na^n of z_1, leads to the same contradiction.

These two cases cover all possible decompositions, and they all lead to contradictions. Consequently, our assumption was false and L is not context-free. □

Note that in [24], it was shown that for a morphic involution θ, for all θ-bordered words v, either $v = ur\theta(u)$ for some $r, u \in \Sigma^*$ or $v = (xy\theta(x)\theta(y))^* xy\theta(x)\theta(y)x$ for some $x, y \in \Sigma^*$. In the next proposition we construct a grammar that generates all such θ-bordered words.

Proposition 8 *Let θ be a morphic involution on Σ^*. Then the set of all θ-bordered words is context sensitive i.e., $\Sigma^* \setminus D_\theta(1)$ is context sensitive.*

Proof. Let $\Sigma = \{a_1, a_2, ..., a_n\}$ be a finite alphabet set and take $G = (V_N, V_T, X_0, \Sigma)$ where $V_N = \{X, X_0, X_1, X_2, X_3, Y_i, Z, Z_1, P, Q\}$ where $1 \leq i \leq n$ and $V_T \subseteq \Sigma^*$. Define the productions of G for all $a_i \in \Sigma$ to be

$$X_0 \to ZX_1X_2X_3XZ_1 \tag{1}$$

$$X_1X_2 \to a_iX_1Y_i \tag{2}$$

$$Y_iX_3 \to X_2\theta(a_i)X_3, Y_ia_j \to a_jY_i, a_iX_2 \to X_2a_i \tag{3}$$

$$Y_iX_3 \to PX_2\theta(a_i)X_3Q, a_iPX_2 \to PX_2a_i, Y_ia_j \to a_jY_i \tag{4}$$

$$X_1PX_2 \to a_iX_1PX_2 \tag{5}$$

$$a_jX \to Xa_j, Y_iXZ_1 \to a_iXZ_1, ZXa_i \to a_iZY_iX, Y_iXa_j \to a_jY_iX \tag{6}$$

$$X_1X_2 \to \lambda, X_3 \to \lambda \tag{7}$$

$$QXZ_1 \to \lambda, X_1PX_2 \to \lambda, Z \to \lambda \tag{8}$$

$$XZ_1 \to \lambda \tag{9}$$

Consider derivations D from $ZRX_1X_2\theta(R)X_3XZ_1$ leading to a terminal word (after an application of the initial rule 1 and $R = \lambda$). If the rule in 2 is used then we can either use rules 3 or rules 4. If rule 2 is used then we eventually end up with $ZuX_1X_2\theta(u)XZ_1$. Then we can either use rules in 6 and 7 which results in the word $(uv\theta(u)\theta(v))^*uv\theta(u)\theta(v)u$ for $u, v \in \Sigma^*$ or use rules in 2 and 4 which results in word of the type $ur\theta(u)$ for $r, u \in \Sigma^*$. If D begins with an application of rule 2 and the first rule in 3 then the only possibility is to continue the derivation to the word $Zra_iX_1Y_i\theta(r)X_3XZ_1 \to Zra_iX_1\theta(r)Y_iX_3XZ_1$ which leads to $Zra_iX_1X_2\theta(r)\theta(a_i)X_3XZ_1$. Here we have two choices, either we continue to apply rules in 2 or apply rules in 7 and get $Zra_i\theta(r)\theta(a_i)XZ_1$ and we can apply rules in 6 which will lead to $ZXra_i\theta(r)\theta(a_i)Z_1$ and the only possibility to continue the derivation is to apply the rule $ZXa_i \to a_iZY_iX$ in 6 and we get the word $a_jZY_jXr_2a_i\theta(r)\theta(a_i)Z_1$ which leads to $a_jZr_2a_i\theta(r)\theta(a_i)Y_jXZ_1$ and hence $a_jZr_2a_i\theta(r)\theta(a_i)a_jXZ_1$. Continuing to apply the rules in 6 we end up with the word of type $(uv\theta(u)\theta(v)) * uv\theta(u)\theta(v)u$. If D begins with an application of rule 2 and the first rule in 4, then it will lead to the word $Zra_iX_1PX_2\theta(r)\theta(a_i)X_3QXZ_1$. Then we can either apply rules in 8 to get words of type $u\theta(u)$ or apply the rule in 5 to get words of type $us\theta(u)$ for $s \in \Sigma^*$. Hence $L(G) = \{xs\theta(x), (uv\theta(u)\theta(v))^iu$ for $i \geq 1$ and $u, v, s, x \in \Sigma^*\}$. Note that $L(G) = \Sigma^* \setminus D_\theta(1)$. $\Box$

Proposition 9 *Given $v \in \Sigma^+$ it is decidable whether $v \in D_\theta(1)$ or not.*

Proof. Follows immediately from the decidability of membership for context sensitive and regular languages. $\Box$

Note that for an antimorphic involution θ and for $u \in D_\theta(i)$ for some $i \geq 2$ with $L_d^\theta(u) = \{\lambda <_p u_1 <_p u_2 <_p ... <_p u_{i-1}\}$ we have $u_1 \in D_\theta(1)$.

Proposition 10 *Let $u \in D_\theta(1)$. If $v <_d^\theta u^i$ then either $v = \lambda$ or $u = \theta(u)$ and $v = u^j$ for $1 \leq j < i$.*

Proof. Let $v <_d^\theta u^i$ for some $u \in D_\theta(1)$. If $v \neq \lambda$, $u^i = v\alpha = \beta\theta(v)$, for $\alpha, \beta, v \in \Sigma^+$, then $v = u^j r_1$ and $\theta(v) = s_2 u^j$ for $u = r_1 r_2 = s_1 s_2$ and $0 \leq j < i$. We only prove the statement when θ is a morphic involution. The case when θ is an antimorphic involution is similar. If $v = u^j r_1$, then $\theta(v) = \theta(u^j)\theta(r_1) = s_2 u^j$. If $r_1.s_2 \in \Sigma^+$, then $u = r_1 r_2 = p\theta(r_1)$. If $r_2 \neq \lambda$ then $u \notin D_\theta(1)$ which is a contradiction. If $r_2 = \lambda$ then $p = \lambda$ and $u = r_1 = \theta(r_1)$ which implies that $u = \theta(u)$ and $v = u^{j+1} = \theta(v)$. If $r_1 = \lambda$ then $v = u^j = \theta(v)$ and $u = \theta(u)$. □

The following lemma provides for a given $u \in \Sigma^*$, the number of θ-borders of u. We recall that $u \in \Sigma^*$ is said to be primitive if $u = v^i$ for some $v \in \Sigma^+$, $i \geq 1$, then $i = 1$ and the set of all primitive words over Σ is denoted by Q.

Lemma 13 *Let $u \in Q$ such that $u = \theta(u)$ and $j \geq 1$. Then,*

1. *For a morphic involution θ, $\nu_d^\theta(u^j) = \nu_d^\theta(u) + j - 1$.*
2. *For an antimorphic involution θ, $\nu_d^\theta(u^j) = |u|^j = j.|u|$.*

Proof. Let θ be a morphic involution and $u \in L_{pal}^\theta$, .i.e., $u = \theta(u)$. For $u = a_1 a_2 ... a_n$, $\theta(u) = \theta(a_1)...\theta(a_n)$, $a_i \in \Sigma$ which implies $a_i = \theta(a_i)$ for all i. Hence θ is identity on Σ and thus $\nu_d(u) = \nu_d^\theta(u)$. It was shown in [30] that $\nu_d(u^j) = \nu_d^\theta(u) + j - 1$. Hence $\nu_d(u^j) = \nu_d^\theta(u^j) = \nu_d^\theta(u) + j - 1$.

Let θ be an antimorphic involution and $u = \theta(u)$. If $u = a_1 ... a_n$ then $\theta(u) = \theta(a_n)...\theta(a_1)$ and since $u = \theta(u)$ we have $a_i = \theta(a_{n+1-i})$. Hence $\nu_d^\theta(u) = |u|$ since $L_d^\theta(u) = \{\lambda, a_1, a_1 a_2, ..., a_1 a_2 ... a_{n-1}\}$. Note that for all $j \geq 1$, $u^j = \theta(u^j)$. Hence $\nu_d^\theta(u^j) = |u^j| = j.|u|$. □

6. Relations to involution codes

Involution codes were introduced in [18] in the process of designing DNA strands with certain properties. Several properties of involution codes that avoid various types of unwanted hybridizations have been discussed in [18, 19, 21, 22]. In this section we discuss the relations between certain involution codes and the set of all words that are unbordered with respect to the involution map θ. We begin the section with the review of definitions of some involution codes defined in [19, 20, 21].

Definition 3 *Let $\theta : \Sigma^* \to \Sigma^*$ be a morphic or antimorphic involution and $X \subseteq \Sigma^+$.*

1. *The set X is called* θ-infix *if $\Sigma^*\theta(X)\Sigma^+ \cap X = \emptyset$ and $\Sigma^+\theta(X)\Sigma^* \cap X = \emptyset$.*
2. *The set X is called* θ-comma-free *if $X^2 \cap \Sigma^+\theta(X)\Sigma^+ = \emptyset$.*
3. *The set X is called* θ-intercode *if $X^{m+1} \cap \Sigma^+\theta(X^m)\Sigma^+ = \emptyset$, $m \geq 1$. The integer m is called the index of X.*
4. *The set X is called* n-θ-comma-free *if every n element subset of X is θ-comma-free.*
5. *The set X is called* n-θ-intercode *of index m if every n element subset of X is a θ-intercode of index m.*
6. *The set X is called* θ-overlap-free *if $PPref(X) \cap PSuff(\theta(X)) = \emptyset$ and $PPref(\theta(X)) \cap PSuff(X) = \emptyset$.*

7. *The set X is called* θ-sticky-free *if $wx, y\theta(w) \in X$ then $xy = \lambda$.*

8. *The set X is called* θ-strict *if $X \cap \theta(X) = \emptyset$.*

We recall the following definition. Let $\mathcal{R}$ be a binary relation on Σ^*. A language L is $\mathcal{R}$-independent if for any $u, v \in L$, $u\mathcal{R}\sqsubseteq$ implies $u = v$. In the following propositions we show that some of the involution sets are independent with respect to the binary relation $<_d^\theta$, where θ is either a morphic or an antimorphic involution.

Proposition 11 *If $X \subseteq \Sigma^*$ is θ-infix (θ-comma-free) then the set X is independent with respect to $<_d^\theta$.*

Proof. Suppose there exists $u, v \in X$ such that $v = ux = y\theta(u)$ for some $x, y \in \Sigma^+$ which implies X is not θ-infix and hence not θ-comma-free since $\theta(u)$ is a suffix of v. Hence X is independent with respect to $<_d^\theta$. □

Proposition 12 *If $X \subseteq \Sigma^*$ is θ-sticky-free then X is independent with respect to $<_d^\theta$.*

Proof. Let $u, v \in X$ such that $v = ux = y\theta(u)$ for some $x, y \in \Sigma^+$. Then $ux, y\theta(u) \in X$ but $x \neq y \neq \lambda$ which is a contradiction since X is θ sticky-free. □

Proposition 13 *Let θ be a morphic involution. If $X \subseteq \Sigma^*$ is strictly θ-overlap-free then X is independent with respect to $<_d^\theta$.*

Proof. Since X is θ-overlap-free we have $PPref(X) \cap PSuff(\theta(X)) = \emptyset$ and $PSuff(X) \cap PPref(\theta(X)) = \emptyset$. Suppose for $u, v \in X$ we have $v = ux = y\theta(u)$, for some $x, y \in \Sigma^+$ then $\theta(v) = \theta(u)\theta(x)$ and $\theta(v) = \theta(y)u \Rightarrow u \in PPref(X) \cap PSuff(\theta(X))$ and $\theta(u) \in PSuff(X) \cap PPref(\theta(X))$ a contradiction. □

Proposition 14 *Let θ be morphic involution and let $L_{(n)}$ be a set of all θ-unbordered words such that for all $x, y \in L_{(n)}$, $|x| = |y| = n$ and $xy \in D_\theta(1)$. Then $L_{(n)}$ is θ-comma-free.*

Proof. Note that from Proposition 1 for all $x, y \in D_\theta(1)$, $xy \in D_\theta(1)$ iff $\theta(\text{Pref}(x)) \cap \text{Suff}(y) = \emptyset$. Suppose $L_{(n)}$ is not θ-comma-free then there exists $x, y, z \in L_{(n)}$ such that $xy = \alpha\theta(z)\beta$ for some $\alpha, \beta \in \Sigma^+$. Then we have $\theta(z) = x_2y_1$ where $x = x_1x_2$ and $y = y_1y_2$ with both $x_2, y_1 \in \Sigma^*$. The case when $\theta(z) = x$ or $\theta(z) = y$ implies that $zx = z\theta(z)$ or $zy = z\theta(z)$ which is a contradiction since zx and zy are θ-unbordered. The case when $\theta(z) = x_2y_1$ with $x_2, y_1 \in \Sigma^+$ implies that $x_2 \in \theta(Pref(z))$ and thus $zx = \theta(x_2)z_2x_1x_2$ which is a contradiction since $zx \in D_\theta(1)$. Similar contradiction arises when $y_1 \in \theta(Suff(z))$. Hence $L_{(n)}$ is θ-comma-free. □

Corollary 7 *Let θ be a morphic involution. Let $L_{(n)}$ be as defined in Proposition 14. Then $L_{(n)}$ is a θ-intercode of index m for all $m \geq 1$.*

Proof. Obvious, since every θ-comma-free is also a θ-intercode of index m for all $m \geq 1$. □

Note that the set $L_{(n)}$ defined in Proposition 14 is not unique. For example, let $\Sigma = \{a, b, c, d\}$ and θ be a morphic involution such that $\theta(a) = b$ and $\theta(c) = d$. Then $L_{(2)} = \{aa, cc, ac, ca\}$ or $\{bc, bb, cc, cb\}$ or $\{ad, da, dd, aa\}$ or $\{bd, bb, db, dd\}$.
The above proposition does not hold when θ is an antimorphic involution. Let $\Sigma = \{a, b, c, d\}$ and θ be an antimorphic involution such that $a \mapsto b$, $c \mapsto d$ and viceversa. Note that $aaba, cbbc, adba \in L_{(4)}$, but $aa(bacb)bc = aa\theta(adba)bc$ which implies that $L_{(4)}$ is not θ-comma-free.

Proposition 15 *Let θ be a morphic or an antimorphic involution such that θ is not identity. Then $L \subseteq \Sigma^+$ is θ-strict and θ-sticky-free if and only if $L \subseteq D_\theta(1)$ and $L^2 \subseteq D_\theta(1)$.*

Proof. Assume that L is θ-strict and θ-sticky-free. We need to show that both $L, L^2 \subseteq D_\theta(1)$. Note that since L is θ-sticky-free for all $wx, y\theta(w) \in L$ we have $xy = \lambda$ and since L is θ-strict we have $L \cap \theta(L) = \emptyset$. Thus for all $u, v \in L$ we have $\theta(Pref(u)) \cap Suff(v) = \emptyset$. Hence from Lemma 8, 9 and Proposition 1 we have $L, L^2 \subseteq D_\theta(1)$.

Conversely, assume that $L, L^2 \subseteq D_\theta(1)$. We need to show that L is θ-strict and L is θ-sticky-free. Suppose L is not θ-strict. Then there exist $u, v \in L$ such that $u = \theta(v)$. This implies that $vu = \theta(u)u \notin D_\theta(1)$ a contradiction since $L^2 \subseteq D_\theta(1)$. Suppose L is not θ-sticky-free. Then there exist $wx, y\theta(w) \in L$ with $xy \neq \lambda$, which implies that $wxy\theta(w) \in L^2$ but $wxy\theta(w) \notin D_\theta(1)$ a contradiction. Hence L is both θ-strict and θ-sticky-free. □

The following results follow from Corollary 11.

Corollary 8 *Let L be θ-strict and θ-sticky-free. Then $L^+ \subseteq D_\theta(1)$.*

Corollary 9 *Let $L_1, L_2 \subseteq \Sigma^+$ be θ-strict and θ-sticky-free. Then $L_1L_2 \subseteq D_\theta(1)$ iff $L_1^+L_2^+ \subseteq D_\theta(1)$.*

Acknowledgment

Research supported by NSERC and Canada Research Chair grants for Lila Kari. The authors thank Shinnosuke Seki (Ph.D. student) for his valuable comments and for the proof of Proposition 7.

References

1. J.Berstel, D.Perrin, *Theory of Codes,* Academis Press, Inc. Orlando Florida, (1985).
2. F.Blanchet-Sadri, *Periodicity on partial words*, Computers and Mathematics with Applications 47 (2004), 71-82.
3. F.Blanchet-Sadri, C.D.Davis, J.Dodge and M.Moorefield, *Unbordered partial words*, March 24 (2003).
4. F.Blanchet-Sadri and Rebecca V.Lewis, *Equations on partial words*, 31st International symosium on mathematical Foundations of Computer Science, LNCS 4162, (2006), 167-178.
5. R.S.Boyer and J.S.Moore, *A fast string searching algorithm* Communication of the ACM 20 (1977), 762-772.
6. J.C.Costa, *Biinfinite words with maximal recurrent unbordered factors*, Theoretical Computer Science 290(3) (2003) 2053-2061.
7. M.Crochemore and D.Perrin, *Two-way string matching* Journal of Assciation of Computer Mathematics 38 (1991), 651-675.
8. M.Crochmore and W.Rytter, *Jewels of stringology* World Scientific (2003).
9. M.Crochmore, F.Mignosi, A.Restivo and S.Salemi, *Text compression using antidictionaries* LNCS 1644 Springer-Verlag, Berlin (1999), 261-270.
10. J.P.Duval, *Relationship between the period of a finite word and the length of its unbordered segments*, Discrete Mathematics 40 (1998), 31-44.
11. A. Fadiel, K. Eichenbaum, A. Hamza. 'Genomemark': detecting word periodicity in biological sequences. *Journal of Biomolecular Structure and Dynamics*, 23-4 (2006), 457-64.
12. T.Harju and D.Nowotka, *Periodicity and unbordered words*, TUCS Technical Report 523, April (2003).
13. T.Harju and D.Nowotka, *Border correlation of binary words with a fixed weight*, TUCS Technical Report 546, August (2003).

14. T.Harju and D.Nowotka, *Counting border and primitive words*, TUCS Technical Report 630, November (2004).

15. T.Harju and D.Nowotka, *Periodicity and unbordered words: A proof of the extended Duval conjecture*, LNCS 2996 (2004) 294-394.

16. T.Harju, A.Lepsito and D.Nowotka, *A characterization of periodicity of biinfinite words*, Theoretical Computer Science 347 (2005), 419-422.

17. T.Harju and D.Nowotka, *On unique factorizations of primitive words*, Theoretical Computer Science 356 (2006), 186-189.

18. S.Hussini, L.Kari, S.Konstantinidis, *Coding properties of DNA languages*, Proceedings of the 7th International Meeting on DNA Based Computers (N.Jonoska, N.C.Seeman editors) Springer LNCS 2340 (2002), 57-69.

19. N.Jonoska , K.Mahalingam and J.Chen, *Involution codes: with application to DNA coded languages*, Natural Computing, Vol 4-2(2005), 141-162.

20. N.Jonoska, L.Kari and K.Mahalingam, *Involution solid and join codes*, Developments in Language Theory: 10th International Conference, Oscar H.Ibarra, Zhe Dang editors LNCS 4036 (2006), 192-202.

21. L.Kari, S.Konstantinidis, E.Losseva and G.Wozniak, *Sticky-free and overhang-free DNA languages*, Acta Informatica 40 (2003), 119-157.

22. L.Kari, S.Konstantinidis, E.Losseva, P.Sosik and G.Thierrin, *Hairpin structures in DNA words*, Proceedings of the 11th International Meeting on DNA Computing, A.Carbone, M.Daley, L.Kari, I.McQuilan, N.Pierce editors LNCS 3892 (2005), 158-170.

23. L.Kari and K.Mahalingam, *Involution conjugate and commutative words*, Submitted.

24. E.Losseva, *Optimal methods of encoding information for DNA computing* , Doctoral Dissertation in Computer Science, University of Western Ontario, London, Canada (2005).

25. M.Lothaire, *Combinatorics of Words*, Cambridge University Press, Cambridge UK (1997).

26. D.Margaritis and S.Skiena, *Reconstructing strings from substrings in rounds*, Proceedings of the 36th Annual Symposium on Foundations of Computer Science (1995), 613-620.

27. H.J.Shyr, G.Thierrin and S.S.Yu, *Monogenic e-closed languages and dipolar words*, Discrete Mathematics 126 (1994), 339-348.

28. H.J.Shyr, *Free Monoids and Languages*, Hon Min Book Company (2001).

29. J.A.Storer, *Data compression: methods and theory*, Computer Science Press, Rockville, MD(1998).

30. S.S.Yu, *Languages and Codes*, Lecture Notes, Department of Computer Science, National Chung-Hsing University, Taichung, Taiwan 402, (2005).

31. S.Yu, *A characterization of local regular languages* Discrete Mathematics 184 (1998), 195-203.