Journal of Mathematics and Informatics
Vol. 1, 2013-14, 52-59
ISSN: 2349-0632 (P), 2349-0640 (online)
Published on 8 May 2014
www.researchmathsci.org

Journal of
Mathematics and Informatics

Factors of words under an involution

C. Annal Deva Priya Darshini[1], ***V. Rajkumar Dare***[1], ***Ibrahim Venkat***[2] **and** ***K.G. Subramanian***[2]

[1]Department of Mathematics, Madras Christian College, Chennai-600059, India
E-mail: annaldevapriyadarshini@gmail.com;rajkumardare@yahoo.com

[2]School of Computer sciences, UniversitiSains Malaysia, 11800 Penang, Malaysia
E-mail:Ibrahim@cs.usm.my; kgsmani1948@yahoo.com

Received 22 February 2014; accepted 7 May 2014

Abstract. Generalizing the notion of Watson-Crick complementarity which is central in DNA computing, involutively bordered words have been introduced and investigated in the literature. In this paper, motivated by these studies, we introduce a general notion of involutive factors of words and obtain several combinatorial properties of such words under a morphic or anti-morphic involution. Propertiesof the complementary notion of involutive factor-free words are also obtained.

Keywords: Factor of a word, Involution

AMS Mathematics Subject Classification (2010): 68R15

1. Introduction

Combinatorics on words [6], which is classified as a branch of Discrete Mathematics, is an area of study and research on general properties of words which are finite or infinite sequences of symbols. A finite word or simply, a word is a finite sequence of symbols. Words play a central role in many branches of study, including DNA computing [1, 2]. A DNA strand which is the basic unit of every living cell is a sequence of nucleotides adenine, guanine, cytosine, thymine denoted by the symbols A, G, C, T with these nucleotides binding to each other with A to T and C to G. A DNA strand can thus be viewed as a word over the four letter DNA alphabet $\{A, T, C, G\}$ and the relation between the nucleotides can be expressed in terms of an involution mapping, also known as the Watson Crick involution and the ends of the DNA strand can "bind" with each other if they are images of each other with respect to this involution. The involutively bordered words were introduced by Kari et al. [3] as a generalization of the classical notions of bordered and unbordered words, in the context of DNA computing. In fact, Kari et al. [3] point out that theoretical properties of bioinformation are studied in their work by investigating models of DNA encoded information based on formal language theoretic and combinatorial properties of words. Also, combinatorial properties of words have been investigated in other worksrelated to DNA computing (See, for example [4, 5, 7]). Here, motivated by these studies, we introduce the concept of θ–factor or involutive

factor of a finite word which extends the notion of involutive border considered in [3] and obtain several combinatorial properties of such factors.
The paper is organized as follows: Section 2 introduces and investigates involutive factors of words. Section 3deals with the complementary notion of involutive factor free words. Section 4 gives the concluding remarks.

2. Involutive factors of a word

We first recall certain basic notions [4]. An alphabet Σ is a finite set of symbols. A finite word or simply a word *w* over Σ is a finite sequence of symbols of Σ. For example, *abaab* is a word over $\Sigma = \{a,b\}$. The set of all words over Σ is denoted by Σ^* and this includes the empty word λ, which has no symbols. In fact Σ^* is a free monoid over Σ under the operation $\circ$ of string concatenation defined by $u \circ v = a_1 \cdots a_n b_1 \cdots b_m$, for words $u = a_1 \cdots a_n$, $v = b_1 \cdots b_m$, $a_i, b_j \in \Sigma$, $1 \le i \le n$, $1 \le j \le m$. We write $u \circ v$ as *uv*. Let $\Sigma^+ = \Sigma^* - \{\lambda\}$. A language *L* over Σ is a subset of Σ^*. For $w \in \Sigma^+$, *alph(w)* is the set of symbols of Σ that appear in *w*. For example, if $\Sigma = \{a, b, c\}$, *w* = *abaab*, then*alph(w)* = *{a,b}*. The length of a word *w* denoted by $|w|$, is the number of symbols in *w*, counting repetitions. A word *w* is a palindrome if it is the same read from the left to the right or the right to the left. For example, the word *abba* over $\{a, b\}$ is a palindrome.

For a word $w \in \Sigma^+$, the word $u \in \Sigma^+$ is a factor of *w* if there are words $\alpha, \beta \in \Sigma^*$ such that $w = \alpha u \beta$. If $\alpha = \lambda$ then *u* is called a prefix of *w* and if $\beta = \lambda$ then *u* is called a suffix of *w*. The set of all factors of *w* is denoted by *F(w)*. A mapping $\theta : \Sigma^* \to \Sigma^*$ satisfying the property that $\theta(xy) = \theta(x)\theta(y)$ is a morphism on Σ^*. If $\theta(xy) = \theta(y)\theta(x)$, then θ is an antimorphism on Σ^*. θ is an involution on Σ^*, if $\theta(\theta(x)) = x$, for all $x \in \Sigma^*$.

We now recall notions of bordered and unbordered words. For properties and other results on involutively bordered words we refer to [3]. Throughout the paper, Σ is considered an alphabet unless stated otherwise.

Definition 2.1. Let θ be a morphic or an antimorphic involution on Σ^*. A word $u \in \Sigma^+$ is said to be θ - bordered if there exists a $v \in \Sigma^+$ such that $u = vx = y\theta(v)$ for $x, y \in \Sigma^+$. In this case, we say that *v* is a θ - border of *u*. Otherwise, the word *u* is called θ - unbordered.

We now define the concept of involutive factor or θ –factor of a finite word.

Definition 2.2. Let θ be a morphic or an antimorphic involution on Σ^*. A word $u \in \Sigma^+$ is a θ- factor or an involutive factor of *w* if *u* and $\theta(u) \in F(w)$ .i.e. both *u* and its image

$\theta(u)$ are factors of w. We also say that w is a word with θ-factors. The wordw can then be written as $w = xuy = x'\theta(u)y', x \neq x', x, y, x', y' \in \Sigma^*$.

Remark: If v is a θ - border of a word u, then v is also a θ- factor of u.

Example 2.1. Let $\Sigma = \{a, t, c, g\}$ be an alphabet. Let θ be a mapping on Σ^* defined by $\theta(a) = T, \theta(t) = A, \theta(c) = g, \theta(g) = c$.

(i) If θ is a morphic involution then the factor $v = acga$ of the wordu = *t**acga**atcg**tgct**tt*is a θ - factor since $\theta(v) = tgct$is also a factor of u

(ii) If θ is an antimorphic involution on Σ^*, then the factor v =*ggta* of w = *cagtgac**taccggta**c* is a θ - factor since $\theta(v) = \theta(ggta) = tacc$ is also a factor of w .

We note that for a word u = *tctcccctt*overΣ, no symbol of u has its θ - image also in u, when θ is a morphic involution on Σ defined as above. We call u an involutive factor free or θ-factor free word.

Definition 2.3. Let θbe a morphic or an antimorphic involution on Σ^* . For $w \in \Sigma^*$, we denote by $L_f^\theta(w)$, the set of all θ – factors of w i.e.
$L_f^\theta(w) = \{v \in \Sigma^* / v, \theta(v) \in F(w)\}$.

Example 2.2. Let $\Sigma = \{a, b\}$ and w = *abaabaaab*be a word onΣ^*. Let θ be a morphic involution on Σ^* defined by $\theta(a) = b, \theta(b) = a$. Then the only θ - factors of *w*are a, b, ab, ba. $L_f^\theta(w) = \{a, b, ab, ba\}$.

The following result is a consequence of the definitions.

Theorem 2.1. Let θ be a morphic or an antimorphic involution on Σ^* . Then

(i) if $u \in \Sigma^*$, then for every $v \in L_f^\theta(u)$, we have $F(v) \subseteq L_f^\theta(u)$,

(ii) for a word $u \in \Sigma^*$, we have $F(u) \cap \theta(F(u)) = L_f^\theta(u)$,

(iii) ifu_1, u_2 are any two words over Σ then $L_f^\theta(u_1) \subseteq L_f^\theta(u_2)$, if $u_1$is a factor of u_2 and in particular, u_1 is a prefix or suffix of u_2.

Proof: In order to prove (i), let $x \in F(v)$. Since $v \in L_f^\theta(u)$, we have $v, \theta(v) \in F(u)$. Now $x \in F(u)$ as x is a factor of v and v is a factor of u. Also $\theta(x)$ is a factor of $\theta(v)$ and so $\theta(x) \in F(u)$. Since both $x, \theta(x) \in F(u)$ we have $x \in L_f^\theta(u)$. This proves $F(v) \subseteq L_f^\theta(u)$.

To prove (ii), let $x \in F(u) \cap \theta(F(u))$. Then, by definition of $F(u)$, $u = \alpha x \beta, \alpha, \beta \in \Sigma^*$ and $x = \theta(v)$ for some $v \in F(u)$. Also $u = \alpha' v \beta'$ for some $\alpha', \beta' \in \Sigma^*$. Since the mapping θ is an involution, we have $\theta(x) = \theta(\theta(v)) = v$. Hence $u = \alpha' \theta(x) \beta'$. This implies that $\theta(x) \in F(u)$ and so $x \in L_f^\theta(u)$. This proves that

$F(u) \cap \theta(F(u)) \subseteq L_f^\theta(u)$. Conversely, let $x \in L_f^\theta(u)$. Then $x, \theta(x) \in F(u)$ so that $u = \alpha x \beta; \alpha, \beta \in \Sigma^*; u = \alpha' \theta(x) \beta'; \alpha', \beta' \in \Sigma^*$. Now, $\theta(x) \in F(u)$ implies that x = $\theta(\theta(x)) \in \theta(F(u))$. Since x is arbitrary, $L_f^\theta(u) \subseteq F(u) \cap \theta(F(u))$.

Finally, to prove (iii), let v be aθ – factor of the word u_1. Then $v, \theta(v) \in F(u_1)$. Since u_1 is a factor of u_2, we have $F(u_1) \subseteq F(u_2)$. Therefore $v, \theta(v) \in F(u_2)$.Hence the proof.

Theorem 2.2. Let θ be a morphic or an antimorphic involution on Σ^*. Let $u \in \Sigma^*$. Then $\theta(L_f^\theta(u)) = L_f^\theta(u) = L_f^\theta(\theta(u))$.

Proof: We first prove $\theta(L_f^\theta(u)) = L_f^\theta(u)$.Let $v' \in \theta(L_f^\theta(u))$. Then $v' = \theta(v)$ for some $v \in L_f^\theta(u)$ so that both v and $v' \in F(u)$ by definition of $L_f^\theta(u)$. Also $\theta(v') = \theta(\theta(v)) = v$ and so $\theta(v') \in F(u)$. Hence $v' \in L_f^\theta(u)$. This proves that $\theta(L_f^\theta(u)) \subseteq L_f^\theta(u)$. Conversely, let $v \in L_f^\theta(u)$. This implies that both v and $\theta(v) \in F(u)$. Since θ is an involution, setting $v' = \theta(v)$ we find that $\theta(v') = v$. Hence v' and $\theta(v')$ $\in F(u)$. This means that $v' \in L_f^\theta(u)$ so that $v = \theta(v') \in \theta(L_f^\theta(u))$. Hence, $L_f^\theta(u) \subseteq \theta(L_f^\theta(u))$. Thus $\theta(L_f^\theta(u)) = L_f^\theta(u)$.

To prove $L_f^\theta(u) = L_f^\theta(\theta(u))$, let $x \in L_f^\theta(u)$. Then $x, \theta(x) \in F(u)$, so that

$u = \alpha_1 x \beta_1, u = \alpha_2 \theta(x) \beta_2$, for some $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \Sigma^*$.

If θ is a morphic involution, then $\theta(u) = \theta(\alpha_1)\theta(x)\theta(\beta_1), \theta(u) = \theta(\alpha_2) x \theta(\beta_2)$. If θ is an antimorphic involution, then $\theta(u) = \theta(\beta_1)\theta(x)\theta(\alpha_1), \theta(u) = \theta(\beta_2) x \theta(\alpha_2)$. In either case, $x, \theta(x) \in F(\theta(u))$. Hence $x \in L_f^\theta(\theta(u))$. Thus $L_f^\theta(u) \subseteq L_f^\theta(\theta(u))$. Let $y \in L_f^\theta(\theta(u))$. Then $y, \theta(y) \in F(\theta(u))$. This implies $\theta(y), \theta(\theta(y)) \in F(\theta(\theta(u)))$. i.e. $y, \theta(y) \in F(u)$. Hence $y \in L_f^\theta(u)$. This proves $L_f^\theta(\theta(u)) \subseteq L_f^\theta(u)$. Hence the proof.

Definition 2.4. Let L be a language over an alphabet Σ. Let θ be a morphic or an antimorphic involution. Then we define $L_f^\theta(L) = \cup_{u \in L} L_f^\theta(u)$.

Example 2.3. Let $L = \{(ab)^n \mid n \geq 1\}$ be a language over Σ = $\{a, b\}$. Let θ given by $\theta(a)=b$, $\theta(b)=a$ be a morphic involution on Σ^*. Then $L_f^\theta(ab) = \{a, b\}$, $L_f^\theta(abab) = \{a, b, ab, ba, aba, bab\}$ and so on. Thus $L_f^\theta(L) = \cup_{u \in L} L_f^\theta(u)$ $= \{(ab)^n \mid n \geq 1\} \cup \{(ba)^n \mid n \geq 1\} \cup \{(ab)^n a \mid n \geq 0\} \cup \{(ba)^n b \mid n \geq 0\}$.

We give some properties of this language $L_f^\theta(L)$.

Theorem 2.3. Let θbe a morphic or an antimorphic involution on $\sum^*$. Let $L \subseteq \Sigma^*$ be a language. Then

(i) for every $v \in L_f^\theta(L)$, the image word $\theta(v) \in L_f^\theta(L)$,

(ii) for any two languages L_1, L_2 with $L_1 \subseteq L_2$, we have $L_f^\theta(L_1) \subseteq L_f^\theta(L_2)$.

Proof:

(i) Let $v \in L_f^\theta(L)$. Then $v \in L_\theta^f(u)$, for some $u \in L$, by definition of $L_f^\theta(L)$. This implies $v, \theta(v) \in F(u)$. Let $\theta(v) = w$. Then $v = \theta(w)$, since θ is an involution. This implies that $w, \theta(w) \in F(u)$ and hence $w \in L_\theta^f(u)$. Thus $w = \theta(v) \in L_\theta^f(L)$

(ii) Let $L_1 \subseteq L_2$. To prove $L_f^\theta(L_1) \subseteq L_f^\theta(L_2)$, let $v \in L_f^\theta(L_1)$. Then $v \in L_f^\theta(u)$, for some $u \in L_1$. But $L_1 \subseteq L_2$ means $u \in L_2$. Therefore, $v \in L_f^\theta(u)$ for $u \in L_2$. Since v is arbitrary, we can conclude that $L_f^\theta(L_1) \subseteq L_f^\theta(L_2)$.

3. Involutive factor free words

In this section we obtain properties of the complementary notion of θ - factor free or involutive factor free words.

Definition 3.1. Let θ be a morphic or an antimorphic involution on $\sum^*$. A word $w \in \Sigma^*$ is θ - factor free or involutive factor free if none of its factors is a θ – factor of *w*.

Example 3.1. Consider the word *w = acccccaaa* over $\Sigma = \{a, t, c, g\}$. Let θ be a morphic involution on $\sum^*$ defined by $\theta(a) = t, \theta(t) = a, \theta(c) = g, \theta(g) = c$. No factor *u* of*w* is a θ – factor of *w*, as $\theta(w)$ is not a factor of *w*.

Theorem 3.1. If $w \in \Sigma^*$ is a θ - factor free word, then $\theta(w) \in \Sigma^*$ is also θ - factor free, for a morphic or an anti-morphic involution on $\sum$.

Proof: Let $w \in \Sigma^*$ be a θ - factor free word. Then for any factor $u \in F(w)$ we have $\theta(u) \notin F(w)$. Then w can be written as $w = \alpha u \beta$, for some $\alpha, \beta \in \Sigma^*$. But $w \neq \alpha' \theta(u) \beta'$, for any $\alpha', \beta' \in \Sigma^*$. Assume that $\theta(w)$ has a θ –factor *u*. Then $u, \theta(u) \in F(\theta(w))$ .Then $\theta(w) = \alpha_1 u \beta_1$, for some $\alpha_1, \beta_1 \in \Sigma^*$ and $w = \theta(\theta(w))$ $= \theta(\alpha_1)\theta(u)\theta(\beta_1)$, if θ is a morphic involution. This implies $\theta(u)$ is a factor of *w*, a contradiction. Hence $\theta(w)$ is also θ - factor free. A similar argument holds if θ is an anti-morphic involution.

Theorem 3.2. Let $\Sigma = \{a, b\}$ andθ given by $\theta(a) = b$, $\theta(b) = a$ be a morphic or an antimorphic involution. Then for any word $u \in \Sigma^+$ which is θ - factor free, either *ua* is θ - factor free or *ub* is θ - factor free.

Proof: Let $u \in \Sigma^+$ be a θ - factor free word. Then *alph(u)* cannot contain both *a* and *b* since *u* is θ-factor free. If *alph (u) = {a}* then *ua* also does not contain *b* and so is θ - factor free but if *alph (u) = {b}* then, likewise, *ub* is θ - factor free.

We now give a characterization of θ - factor free words.

Theorem 3.3. A word $u \in \Sigma^+$ is θ - factor free if and only if for every $a \in$ *alph(u)*, we have $\theta(a) \notin alph(u)$.

Proof: Assume *u* is θ-factor free. This means that none of the factors of *u* is a θ – factor. Then every *a* in *alph(u)* is a factor of *u*. Hence $\theta(a) \notin alph(u)$. Conversely, let $\theta(a) \notin alph(u)$, for every $a \in alph(u)$. To show that $u \in \Sigma^+$ is θ - factor free. Suppose *u* has a θ – factor *v*. Then *v* as well as $\theta(v)$ are factors of *u*. This implies that $a \in alph(v) \subseteq alph(u)$ and $\theta(a) \in alph(\theta(v)) \subseteq alph(u)$. Therefore, for some $a \in \Sigma$ such that $a \in alph(v)$, the image $\theta(a) \in alph(u)$ which is a contradiction to our assumption. Hence the proof.

Corollary 3.1. i)If $w \in \Sigma^*$ is a θ - factor free word, then w^+ and $\theta(w^+)$are also θ - factor free, for a morphic or an anti-morphic involution on $\sum$ where w^+ is the product of *w* with itself certain number of times.
This follows from Theorem 3.3 and the fact that $alph(w^+) = alph(w)$.
ii) A word $u \in \Sigma^+$ is θ - factor free if and only if $L_f^{\theta}(u) = \phi$.

This again follows from Theorem 3.3. In fact, for any factor *v* of *u*, $\theta(v)$ cannot be a factor of *u* if *u* is θ - factor free and this means $L_f^{\theta}(u)$ is empty and the converse is similar.

We now examine the problem of whether the concatenation of two θ - factor free words is also θ - factor free.

Theorem 3.4. Let θ be a morphic or antimorphic involution on Σ^*. Let $u, v \in \Sigma^+$ be θ - factor free. Then $uv \in \Sigma^+$ is θ - factor free if and only if $\theta(alph(u)) \cap alph(v) = \phi$.

Proof: Assume that $w = uv \in \Sigma^+$ is θ - factor free. Then by Theorem 3.3, for every $a \in alph(w)$, we have $\theta(a) \notin alph(w)$. In order to show that $\theta(alph(u)) \cap alph(v) = \varphi$, assume the contrary *i.e.,* $\theta(alph(u)) \cap alph(v) \neq \phi$. Then there exists an element $a \in \Sigma$ such that $a \in \theta(alph(u)) \cap alph(v)$. Now, $a \in \theta(alph(u))$ implies that $\theta(a) \in alph(u)$. But also $a \in alph(v)$ which implies $a \in alph(w)$ and $\theta(a) \in alph(u)$ implies $\theta(a) \in alph(w)$ which is a contradiction. Hence, $\theta(alph(u)) \cap alph(v) = \phi$.
Conversely, let $\theta(alph(u)) \cap alph(v) = \phi$. To prove that *uv* is θ - factor free. Suppose *uv*hasa θ – factor. Then $uv = w$ is such that there exists $a \in \Sigma$ with $a \in alph(w)$ as well as $\theta(a) \in alph(w)$. Both $a, \theta(a) \in alph(u)$ cannot happen as *u* is θ - factor free.

Likewise, both $a, \theta(a) \in alph(v)$ cannot also happen as v is θ - factor free. On the other hand, if $a \in alph(u), \theta(a) \in alph(v)$ then $\theta(a) \in \theta(alph(u)) \cap alph(v)$ which is a contradiction. If $a \in alph(v), \theta(a) \in alph(u)$ a similar contradiction arises. This proves the Theorem.

Corollary 3.2. Let θ be a morphic or antimorphic involution on Σ^*. Let $u, v \in \Sigma^+$ be θ - factor freewith $\theta(alph(u)) \cap alph(v) = \phi$. Then $u^+v^+, (uv)^+$, are θ - factor free.

Proof: By Corollary 3.1 and Theorem 3.4, we have $u^+v^+, (uv)^+$ are θ - factor free.

4. Concluding remarks

A preliminary version of part of this paper was presented in theSixth International Conference on Bio-Inspired Computing: Theories and Applications, in 2011 [8]. The originality of the contribution in this paper lies in introducing a general notion of involutive factor or θ - factor of a word and in obtaining several combinatorial properties of such words. The complementary notion of θ - factor free words is also examined and their properties are derived. We can also consider a notion of involutive factorization of a word. A factorization of a word $w \in \Sigma^+$ is $w = u_1u_2\ldots u_n$ where the $u_i \in \Sigma^*$ for $i = 1, 2, \ldots, n$. This factorization need not be unique and in general, a word w can have more than one factorization. If θ is a morphic or an antimorphic involution on Σ, then a factorization of $w \in \Sigma^+$, given by $w = u_1u_2\ldots u_n$ where the $u_i \in \Sigma^+$, for $i = 1, 2, \ldots, n$, can be called a θ - factorization or involutive factorization of w if and only if $w = \theta(u_{i_1})\theta(u_{i_2})\ldots\theta(u_{i_m})$,where $i_j \in \{1, 2, \ldots, n\}$for j = 1, 2, …, m. It will be of interest to study properties of such factorizations, which is for future work.

Acknowledgement: The authors thank the referees for their useful comments which helped to improve the presentation of the paper.

REFERENCES

1. L.M. Adleman, Computing with DNA, *Scientific American*, August 1998, 34-41.
2. N. Jonoska, Computing with Biomolecules: Trends and Challenges, *XVI Tarragona Seminar on Formal Syntax and Semantics*, 27(2003)
3. L. Kari, K. Mahalingam, Involutively bordered words, *Intern. J. Found. Comp. Sci.*,18 (2007) 1089 – 1106.
4. L. Kari, K. Mahalingam, Watson-crick bordered words and their syntactic monoid, *Intern. J. Found. Comp. Sci.*,19 (2008) 1163-1179.
5. L. Kari, S. Seki, On pseudoknot-bordered words and their properties, *J. Comp. System Sci.*, 75 (2009) 113-121.
6. M.Lothaire, *Combinatorics on words*, Addison Wesley Publishing Company, 1983.
7. K. Mahalingam and K.G. Subramanian, Palindromic completion of a word, *Proc. IEEE Fifth Int. Conf. on Bio-inspired Computing: Theories and Applications (BIC-TA)*, (2010) 1459-1465.

8. V. Rajkumar Dare, C. Annal Deva Priya Darshini, Bordered factors of a finite word, *Proceedings of the 2011 Sixth International Conference on Bio-Inspired Computing: Theories and Applications* (2011) 163 - 166.