θ - Bordered Infinite Words

C. Annal Deva Priya Darshini[1], V. Rajkumar Dare[2]

Dept. of Mathematics, Madras Christian College, Tambaram, Chennai – 600 059.

{[1]annaldevapriyadarshini@gmail.com; [2]rajkumardare@yahoo.com}

Abstract. In this paper, we define θ - bordered infinite word and unbordered words and study their properties. We give a characterization of θ - bordered infinite words for an antimorphic involution θ. We show that the limit language of the set of all θ - bordered words is a ω - regular language for an antimorphic involution θ.

Keywords: θ - bordered infinite word, θ - unbordered infinite word, θ – bordered infinite language, limit language.

1 Introduction

The concept of using DNA for computation as studied by Adleman [4] has opened a wide area of research. The application of formal language theory in molecular biology has solved many problems in language theory. The DNA strand has been treated as a finite word over the four letter alphabet {A, T, C, G} and the relations between the alphabets is modeled as a morphic or antimorphic involution on the set of alphabets. A general overview of this field can be seen in [3].

The study of finite bordered words is useful in the context of DNA computations since DNA strands are finite in nature. This paper is a theoretical study of the generalisation to the case of infinite words. In this paper, we introduce involutively bordered infinite words. By definition, a finite word u is θ - bordered, for a morphic or antimorphic involution θ if there exists $v \in \Sigma^+$ that is a proper prefix of u while $\theta(v)$ is a proper suffix of u. A word u is called θ - unbordered if it is not bordered. Thus, θ - bordered words are defined with the help of prefixes and suffixes. However, the ω - word does not have a suffix and hence we cannot find its border on the right. This problem has been overcome by using the concept of limit of a language. The infinite θ - bordered word is constructed as the limit of an increasing sequence of θ - bordered words which are prefixes of the infinite word. This definition carries the properties of θ - bordered words to infinite words in a natural way.

J. C. Bansal et al. (eds.), *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012),* Advances in Intelligent Systems and Computing 202, DOI: 10.1007/978-81-322-1041-2_31, © Springer India 2013

The paper is organised as follows: Section 2 gives the preliminaries and definitions. Section 3 gives the definition and the characterization of θ - bordered and unbordered infinite words. The language of θ - bordered infinite words is proved to be the limit language of the involutively bordered words. Hence we are able to show that the set of all θ - bordered infinite words is a ω - regular set for an antimorphic involution θ.

2 Preliminaries and Basic Definitions

In this section, we introduce some basic definitions used in this paper. Let Σ be a finite alphabet and Σ^* be the collection of all words over Σ including the empty word λ. Let $\Sigma^+ = \Sigma^* - \{\lambda\}$. For $w \in \Sigma^+$, *alph(w)* is the elements of Σ found in w. The length of a word w is denoted by $|w|$.

A mapping $\theta : \Sigma^* \rightarrow \Sigma^*$ such 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^*. An involution θ is a mapping such that $\theta(\theta(x)) = x$ for all $x \in \Sigma^*$. We recall the definition of involutively bordered and unbordered words proposed in [1].

Definition 2.1. [1]

Let θ be a morphic or 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 some $x, y \in \Sigma^+$. A nonempty word which is not θ - bordered is called θ - unbordered.

3 θ- Bordered Infinite Words and their Properties

In this section we introduce the θ - bordered infinite word as a generalization of a θ - bordered word. Bordered words are also called as overlapping words and they were studied as words with a prefix which is also a suffix. The extension of this was involutively bordered words where the prefix is the θ - image of the suffix. We define θ - bordered infinite word as a generalization of this word to the infinite case. We make use of prefixes alone to define θ - bordered infinite words.

Definition 3.1.

Let $x \in \Sigma^\omega$ and *x[n]* be the prefix of x of length n. Let θ - be a morphic or antimorphic involution on Σ. We say that x is θ - bordered infinite word if for each $i \in \mathrm{N}, \exists n_i > i$ such that *x[n_i]* is a θ - bordered word.

We give some simple examples of infinite words that are θ - bordered.

Example 3.1.

Consider the word $w = at^{\omega}$ on $\Sigma = \{a,t\}$ over which a morphic involution is defined as $\theta(a) = t, \theta(t) = a$. w is θ - bordered infinite word since the prefixes $w[n] = at^{(n-1)}$ are θ - bordered as n tends to infinity.

Example 3.2.

The word $w = (ab)^{\omega}$ on $\Sigma = \{a,b\}$ over which an antimorphic involution is defined as $\theta(a) = b, \theta(b) = a$ is θ - bordered infinite word since the prefixes $w[2n] = (ab)^n$ are θ - bordered as n tends to infinity.

Example 3.3.

The infinite word $w = abaabaaaba^4ba^5b\ldots$ on $\Sigma = \{a,b\}$ with an antimorphic involution defined as $\theta(a) = b, \theta(b) = a$ is θ - bordered as the prefixes $w[n_i] = abaab...b^{n}{}_{i}$ are θ - bordered as i tends to infinity.

The examples show that a θ - bordered infinite word is the limit of an increasing sequence of finite θ - bordered words. Hence the properties of θ - bordered words carry over to the limit word.

Example 3.4.

Consider the word $w = ac^{\omega}$ on $\Sigma = \{a,t,c,g\}$ over which a morphic involution is defined as $\theta(a) = t, \theta(c) = g$. w is θ - unbordered infinite word since the prefixes $w[n] = ac^{(n-1)}$ are θ - unbordered as n tends to infinity.

We observe that $u \in \Sigma^{\omega}$ is θ - unbordered infinite word if $\exists i \in \mathrm{N} : \forall n_i > i$ $u[n_i]$ is a θ - unbordered word.

Now, we can study the properties of θ - bordered infinite words in two directions:

(i) we can examine whether the properties of θ - bordered words extend to θ - bordered infinite words,

(ii) we can study how the properties of words are affected by the involution.

Moreover, we are especially interested in studying unbordered words.

We state some results for involutively bordered words and then give its extension.

Lemma 3.1. [2]

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

Lemma 3.2. [2]

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

The two results can be extended to ω - words to give a characterization of θ - unbordered infinite words.

Proposition 3.1.

Let θ - be a morphic or antimorphic involution on Σ (not the identity mapping). Then $w \in \Sigma^\omega$ is θ - unbordered infinite word if and only if there is a prefix $u \in \Sigma^*$ of w such that $\theta(Pref(u)) \cap Suff(u) \neq \phi$ and for every other prefix $v \in \Sigma^*$ of w with $|v| > |u|$ we have $\theta(Pref(v)) \cap Suff(v) = \phi$

Proof.

Let $w \in \Sigma^\omega$ be θ - unbordered infinite word. Then $\exists i \in \mathrm{N} : \forall n_i > i \quad w[n_i]$ is a θ - unbordered word. We form the required prefixes u and v of w as follows: Take $v = w[n_{i+1}]$. The word v is θ - unbordered word. By Lemma 3.1 and 3.2 we have $\theta(Pref(v)) \cap Suff(v) = \phi$. In fact, the words $v_1 = w[n_{i+1}]$, $v_2 = w[n_{i+2}]$, ... all unbordered. To construct u we search for a bordered prefix of $w[n_i]$. If we find a prefix which is bordered we take that as u. Then $\theta(Pref(u)) \cap Suff(u) \neq \phi$. If not, u is the empty word.

Conversely, let $\theta(Pref(w[n_i])) \cap Suff(w[n_i]) \neq \phi$ for some prefix $w[n_i]$ of $w \in \Sigma^\omega$ and $\theta(Pref(w[n_k])) \cap Suff(w[n_k]) = \phi$ for all $k > n_i$. This implies $w[n_i]$ is a θ - bordered word and all words of length greater than $w[n_i]$ are θ - unbordered. For this $w[n_{i+1}]$ we have $\exists i+1 \in \mathrm{N} : \forall n_k > i+1$, $w[n_k]$ is a θ - unbordered word. Hence the result follows.

We give a characterization of θ - bordered infinite words for an antimorphic involution θ.

Lemma 3.3. [2]

Any $x \in \Sigma^*$ is θ - bordered for an antimorphic involution θ if and only if $x = ay\theta(a)$ for some $a \in \Sigma$ $y \in \Sigma^*$.

Proposition 3.2.

Let θ be an antimorphic involution on Σ^{ω}. Then u is θ - bordered infinite word if and only if $u = ay$; $y \in \Sigma^{\omega}$ and $\theta(a)$ is repeated infinitely often in y, $a \in \Sigma$.

Proof.

Let u be a θ - bordered infinite word. By definition, for each $i \in \mathrm{N}, \exists n_i > i$ such that $u[n_i]$ is a θ - bordered word. By lemma 3.3, $u[n_1] = ay\theta(a)$ for $a \in \Sigma \quad y \in \Sigma^*$. Now, $u[n_2]$ is also θ - bordered word and $n_2 > n_1$. Therefore, $u[n_2] = ty'\theta(t)$ for $t \in \Sigma \quad y' \in \Sigma^*$. But $u[n_1]$ is a prefix of $u[n_2]$. Therefore, $u[n_2] = u[n_1]\, r$ for some $r \in \Sigma^+$. That is,

$$ty'\theta(t) = ay\theta(a)r \tag{1}$$

where $a \in \Sigma \quad y \in \Sigma^* \; t \in \Sigma \quad y' \in \Sigma^*$. Now, $a \in \Sigma, \; t \in \Sigma$ implies that $a = t$. Then we must have $\theta(a) = \theta(t)$. Substituting this relation in (1) we have, $ay'\theta(a) = ay\theta(a)r$ This implies $r = r'\theta(a)$where $r' \in \Sigma^*$. Thus, $u[n_2] = ay\theta(a)\, r'\theta(a)$. Similarly, $u[n_3] = ay\theta(a)\, r'\theta(a)\, r''\theta(a)$. Thus, we have $u = ay\theta(a)\, r'\theta(a)\, r''\theta(a)\, r'''\theta(a)...$ where $\theta(a)$ is repeated infinitely often. Hence we write u as $u = ay$ where $a \in \Sigma$ and $\theta(a)$ is repeated infinitely often in y.

We can form sets of θ - bordered infinite words or ω - languages of θ - bordered words. The set of all θ - bordered infinite words form a language which we denote by B_{θ}^{ω}. Subsets of B_{θ}^{ω} are θ - bordered ω - languages. For example, $L = \{a^* b^{\omega}\}$ is a θ - bordered ω - language on $\Sigma = \{a, b\}$ for an antimorphic involution θ defined as $\theta(a) = b, \theta(b) = a$.

Recall that a language L is θ-stable if $\theta(L) \subseteq L$. We prove that B_{θ}^{ω} is stable.

Proposition 3.3.

B_{θ}^{ω} is θ-stable for a morphic or antimorhic involution θ.

Proof.

Let $x \in B_{\theta}^{\omega}$. Then for each $i \in \mathrm{N}, \exists n_i > i$ such that $x[n_i]$ is a θ - bordered word. That is, $x[n_i] = t\alpha = \beta\theta(t)$ for $t, \alpha, \beta \in \Sigma^+$. Now, $\theta(x[n_i]) = \theta(t\alpha) = \theta(\beta\theta(t))$. For a morphic involution θ, we have $\theta(t)\theta(\alpha) = \theta(\beta)\theta(\theta(t)) = \theta(\beta)t$. This implies that $\theta(x[n_i])$ is a θ-bordered word.

If θ is a antimorphic involution, we have $\theta(t\alpha) = \theta(\alpha)\theta(t) = \theta(\theta(t))\theta(\beta) = t\theta(\beta)$. This implies that $\theta(x[n_i])$ is a θ-bordered word.

Thus, for each $i \in \mathrm{N}, \exists n_i > i$ such that θ $(x[n_i])$ is a θ - bordered word. $\theta(x) \in B_\theta^\omega$. Since x is arbitrary, we have $\theta(B_\theta^\omega) \subseteq B_\theta^\omega$. Therefore, B_θ^ω is θ_ stable.

We recall that a language $L \subseteq \Sigma^\omega$ is the limit language of $L_1 \subseteq \Sigma^*$ if for each $x \in L$, there is a sequence of words $x_1 < x_2 < x_3 < \ldots < x_i \ldots$ in $L_1 \subseteq \Sigma^*$ such that $x_1 < x_2 < x_3 < \ldots < x_i \rightarrow x$ in $L \subseteq \Sigma^\omega$.

Next, we give a result which shows that B_θ^ω is the limit language of a finite language of θ - bordered words. This gives the limit language all the properties of the finite language.

Proposition 3.4.

Let B_θ be the set of all θ - bordered words in for an antimorphic involution θ. Let B_θ^ω be the set of all θ-bordered infinite words over Σ^ω. Then lim B_θ = B_θ^ω.

Proof.

Let $x \in \lim B_\theta$. Then there exists an increasing sequence $\{x_n\}$ of elements in B_θ such that $x_1 < x_2 < x_3 < \ldots < x_i \rightarrow x$. Since all the x_i*'s* are θ - bordered. x is θ - bordered infinite word. Hence $x \in B_\theta^\omega$ which implies that for each i there exists n_i such that $x[n_i] \in B_\theta$ implies $x[n_1] < x[n_2] < x[n_3] < \ldots$ $x \in \lim B_\theta$. Thus, lim B_θ = B_θ^ω.

By the above result, the language of θ - bordered infinite words is the limit language of θ - bordered words. The limit language possesses the properties of the finite language. We can generalise the properties of θ - bordered words to θ - bordered infinite words.

It has been proved in [2] that B_θ is regular for an antimorphic involution θ. Since the limit language of a regular language is regular, B_θ^ω is ω - regular. We state this as a theorem.

Theorem 3.1.

The limit language B_θ^ω of the set of all θ - bordered words B_θ is a ω - regular language for an antimorphic involution θ.

4 Conclusion

Thus the generalization of θ - bordered words to the infinite case has been defined and studied. The main result is that the limit language of the set of θ - bordered words is also ω - regular as in the finite case.

References

[1] Kari, L., Mahalingam, K.: Involutively bordered words. International Journal of Foundations of Computer Science. 18, 1089–1106 (2007).

[2] Kari, L., Mahalingam, K.: Watson Crick bordered words and their syntactic monoid. International Journal of Foundations of Computer Science. 19(5) (2008) 1163 – 1179 (2008).

[3] Jonoska, N.: Computing with Biomolecules: Trends and Challenges. XVI Tarragona Seminar on Formal Syntax and Semantics. 27 (2003).

[4] Adleman, L.M.: Computing with DNA. Scientific American (1998).