Some Periodicity of Words and Marcus Contextual Grammars

Pál Dömösi
University of Debrecen
e-mail: domosi@inf.unideb.hu

Géza Horváth
University of Debrecen
e-mail: geza@inf.unideb.hu

Masami Ito
Kyoto Sangyo University
e-mail: ito@ksuvx0.kyoto-su.ac.jp

Kayoko Shikishima-Tsuji
Tenri University
e-mail: tsuji@sta.tenri-u.ac.jp

Abstract

In this paper, first we define a periodic (semi-periodic, quasi-periodic) word and then we define a primitive (strongly primitive, hyper primitive) word. After we define several Marcus contextual grammars, we show that the set of all primitive (strongly primitive, hyper primitive) words can be generated by some Marcus contextual grammar.

1 Introduction

Let X^* denote the free monoid generated by a nonempty finite alphabet X and let $X^+ = X^* \setminus \{\epsilon\}$ where ϵ denotes the empty word of X^*. For the sake of simplicity, if $X = \{a\}$, then we write a^+ and a^* instead of $\{a\}^+$ and $\{a\}^*$, respectively. Let $L \subseteq X^*$. Then L is called a *language* over X. By $|L|$, we denote the cardinality of L. If $L \subseteq X^*$, then L^+ denotes the set of all concatenations of words in L and $L^* = L^+ \cup \{\epsilon\}$. In particular, if $L = \{w\}$, then we write w^+ and w^* instead of $\{w\}^+$ and $\{w\}^*$, respectively. Let $u \in X^*$. Then u is called a *word* over X.

Definition 1.1 A word $u \in X^+$ is said to be *periodic* if u can be represented as $u = v^n, v \in X^+, n \geq 2$. If u is not periodic, then it is said to be *primitive*. By Q we denote the set of all primitive words.

Remark 1.1 Fig. 1.1 indicates that u is a periodic word.

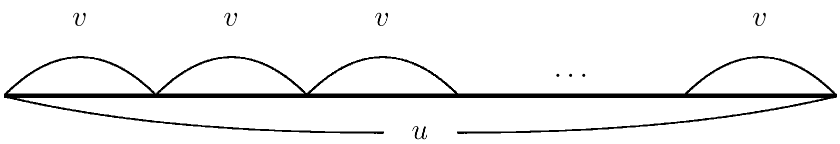

Fig. 1.1

Definition 1.2 A word $u \in X^+$ is said to be *semi-periodic* if u can be represented as $u = v^n v', v \in X^+, n \geq 2$ and $v' \in Pr(v)$ where $Pr(v)$ denotes the set of all prefixes of v. If u is not semi-periodic, then it is said to be *strongly primitive.* By SQ we denote the set of all strongly primitive words.

Remark 1.2 Fig. 1.2 indicates that u is a semi-periodic word.

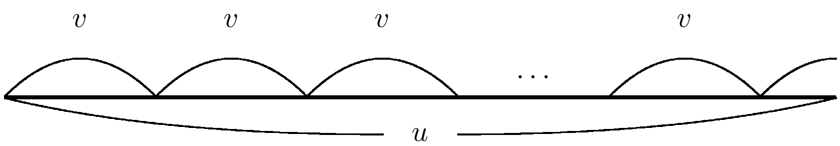

Fig. 1.2

Definition 1.3 A word $u \in X^+$ is said to be *quasi-periodic* if a letter in any position in u can be covered by some $v \in X^+$ with $|v| < |u|$. More precisely, if $u = wax, w, x \in X^*$ and $a \in X$, then $v \in Suf(w)aPr(x)$ where $Suf(w)$ denotes the set of all suffixes of w. If u is not quasi-periodic, then it is said to be *hyper primitive.* By HQ we denote the set of all hyper primitive words.

Remark 1.3 Fig. 1.3 indicates that u is a qusi-periodic word.

Then we have the following inclusion relations.

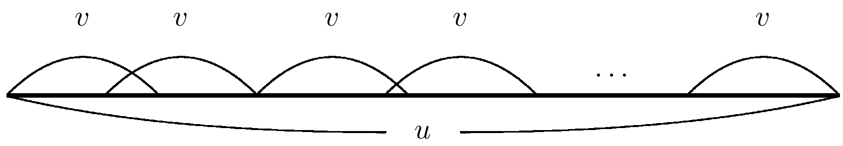

Fig. 1.3

Fact 1.1 $HQ \subset SQ \subset Q$.

Proof That $HQ \subseteq SQ \subseteq Q$ is obvious. Now consider the following example. Let $X = \{a, b, \ldots\}$. Then $ababa \in Q \setminus SQ$ and $aabaaabaaba \in SQ \setminus HQ$. Thus $HQ \neq SQ \neq Q$. Therefore, every inclusion is proper.

2 Marcus Contextual Grammars

We begin this section by the following definition.

Definition 2.1 *(Marcus) contextual grammar with choice* is a structure $G = (X, A, C, \varphi)$ where X is an alphabet, A is a finite subset of X^*, i.e. the set of axioms, C is a finite subset of $X^* \times X^*$, i.e. the set of contexts, and $\varphi : X^* \to 2^C$ is the choice function. If $\varphi(x) = C$ holds for every $x \in X^*$ then we say that G is a *(Marcus) contextual grammar without choice*. In this case, we write $G = (X, A, C)$ instead of writing $G = (X, A, C, \varphi)$.

Definition 2.2 We define two relations on X^*: for any $x \in X^*$, we write $x \Rightarrow_{ex} y$ if and only if $y = uxv$ for a context $(u, v) \in \varphi(x), x \Rightarrow_{in} y$ if and only if $x = x_1x_2x_3, y = x_1ux_2vx_3$ for some $(u, v) \in \varphi(x_2)$. By $\Rightarrow_{ex}^*$ and $\Rightarrow_{in}^*$, we denote the reflexive and transitive closure of each relation and let $L_\alpha(G) = \{x \in X^* \mid w \Rightarrow_\alpha^* x, w \in A\}$ for $\alpha \in \{ex, in\}$. Then $L_{ex}(G)$ is the *(Marcus) external contextual language (with or without choice) generated by* G, and similarly, $L_{in}(G)$ is the *(Marcus) internal contextual language (with or without choice) generated by* G.

Example 2.1 Let $X = \{a, b\}$ and let $G = (X, A, C, \varphi)$ be a Marcus contextual grammar where $A = \{a\}, C = \{(\epsilon, \epsilon), (\epsilon, a), (\epsilon, b)\}, \varphi(\epsilon) =$

$\{(\epsilon, \epsilon)\}, \varphi(ua) = \{(\epsilon, b)\}$ for $u \in X^*$ and $\varphi(ub) = \{(\epsilon, a)\}$ for $u \in X^*$. Then $L_{ex}(G) = a(ab)^* \cup a(ba)^+ b$ and $L_{in}(G) = aX^* \cup X^* a^2 X^*$.

The following example shows that the classes of languages generated by Marcus contextual grammaes have no relation with the Chomsky language classes.

Example 2.2 Let $|X| \geq 2$ and let $w = a_1 a_2 a_3 \cdots$ be an ω-word over X where $a_i \in X$ for any $i \geq 1$. Let $G = (X, A, C, \varphi)$ be be a Marcus contextual grammar where $A = \{a_1\}, C = \{(\epsilon, \epsilon\}, (\epsilon, a) \mid a \in X\}, \varphi(\epsilon) = \{\epsilon, \epsilon), \varphi(a_1 a_2 a_3 \cdots a_i) = \{\epsilon, a_{i+1}\}$ and $\varphi(u) = \emptyset$ if u is not a prefix of w. Then $L_{ex}(G) = \{a_1, a_1 a_2, a_1 a_2 a_3, \cdots\}$. Hence, there exists a Marcus contextual grammar generating a language which is not recursively enumerable.

As for more details on Marcus contextual grammars and languages, see [3].

3 Set of Primitive Words

In this section, we deal with the set of all primitive words. First we provide the following three lemmas. The proofs of the lemmas are based on the results in [2] and [5].

Lemma 3.1 *For any $u \in X^+$, there exist unique $q \in Q$ and $i \geq 1$ such that $u = q^i$.*

Lemma 3.2 *Let $i \geq 1$, let $u, v \in X^*$ and let $uv \in \{q^i \mid q \in Q\}$. Then $vu \in \{q^i \mid q \in Q\}$.*

Lemma 3.3 *Let X be an alphabet with $|X| \geq 2$. If $w, wa \notin Q$ where $w \in X^+$ and $a \in X$, then $w \in a^+$.*

Using the above lemmas, we can prove the following. The proof can be seen in [1].

Proposition 3.1 *The language Q is a Marcus external contextual language with choice.*

However, in the case of $|X| \geq 2$ we can prove that the other types of Marcus contextual grammars cannot generate Q.

4 Set of Strongly Primitive Words

In this section, we deal with the set of all primitive words. First we provide the following three lemmas. All results in this section can be seen in [1].

Lemma 4.1 *Let X be an alphabet with $|X| \geq 2$. If $awb \in SQ$ where $w \in X^*$ and $a, b \in X$, then $aw \in SQ$ or $wb \in SQ$.*

Using the above lemma, we can prove the following.

Proposition 4.1 *The language SQ is a Marcus external contextual language with choice.*

However, we can prove that the other types of Marcus contextual grammars cannot generate SQ.

5 Set of Hyper Primitive Words

In this section, first we characterize a quasi-periodic word.

Definition 5.1 Let $u \in X^+$ be a quasi-periodic word and let any letter in u be covered by a word v. Then we denote $u = v \otimes v \otimes \cdots \otimes v$.

Remark 5.1 Fig. 5.1 indicates that $u = v \otimes v \otimes \cdots \otimes v$.

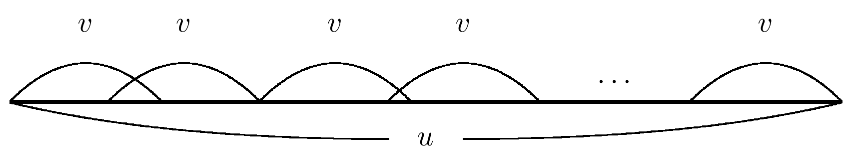

Fig. 5.1

The following lemma is fundamental (see [3]).

Lemma 5.1 *Let $u \in X^+$ and let $u = xv = vy$ for some $x, y, v \in X^+$. Then there exist $\alpha, \beta \in X^*$ and $n \geq 1$ such that $\alpha \neq \epsilon$ and $u = (\alpha\beta)^n\alpha$.*

Lemma 5.2 *Let $x, u, v \in X^+$. If $u = xv = vy$ and $|v| \geq |u|/2$, then $u \notin HQ$.*

Proof By Lemma 5.1, there exist $\alpha, \beta \in X^*$ with $\alpha\beta \neq \epsilon$ and $n \geq 2$ such that $x = \alpha\beta, v = (\alpha\beta)^{n-1}\alpha$ and $y = \beta\alpha$, i.e. $u = (\alpha\beta)^n\alpha$. In this case, $u = \alpha\beta\alpha \otimes \alpha\beta\alpha \cdots \otimes \alpha\beta\alpha$. Thus $\alpha\beta\alpha$ covers u and $u \notin HQ$.

Proposition 5.1 *Let $u \in X^+$. Then there exists a hyper primitive word $v \in HQ$ such that $u = v \otimes v \otimes \cdots \otimes v$. In this representation, v and each position of v are uniquely determined*

Proof Let $u = v \otimes v \otimes \cdots \otimes = w \otimes w \otimes \cdots \otimes w$ where $v, w \in HQ$. If $|v| < |w|(|w| < |v|)$, then $w(v)$ is covered by $v(w)$. This contradicts the assumption that $v, w \in HQ$. Thus $|v| = |w|$ and $v = w$. This means that v is uniquely determined.

Now suppose there exist two distinct representations for $u = v \otimes v \otimes \cdots \otimes v$. Then there exists some position of u such that $v \otimes v = xv$ where $x, y \in X^+$ and $|v| \geq 1/2|u|$. By Lemma 5.2, $v \notin HQ$, a contradiction. Hence each position of v is uniquely determined as well.

Now we show the following lemma.

Lemma 5.3 *Let X be an alphabet with $|X| \geq 2$. If $aw \notin HQ$ and $wb \notin HQ$ where $w \in X^*$ and $a, b \in X$, then $awb \notin HQ$.*

Proof Assume that $aw \notin HQ$ and $wb \notin HQ$. Then $aw \in v \otimes v \otimes \cdots \otimes v$ and $wb \in u \otimes u \otimes \cdots \otimes u$ where $u, v \in HQ$ (see Fig. 5.2). We can assume $|u| \leq |v|$. Notice that the proof can be carried out symmetrically for the case $|v| \leq |u|$. Hence $u = u'b$ and $vb \in X^*u$ for some $u' \in X^*$ (see Fig. 5.3). We prove that the first letter after v in every position in Fig. 5.2 becomes b. Then $awb = vb \otimes vb \otimes \cdots \otimes vb$ and $awb \notin HQ$. To prove this, we consider the case in Fig. 5.4. In the figure, $vv' \in v \otimes v$. Since $|xy| \leq |u|$, $|x| \leq |u|/2$ or $|y| \leq |u|/2$. In the former case, if $x \neq \epsilon$, then $u = xu'' = u''y'$ for some $u'', y' \in X^+$. By Lemma 5.2, this contradicts the assumption that $u \in HQ$. In the latter case, if $y \neq \epsilon$, then $u = yu'' = u''y''$ where $u'', y'' \in X^+$. This contradicts the assumption that $u \in HQ$ again. Thus $x = \epsilon$ or $y = \epsilon$. Since $u = u'b$, the first letter after v must be b. This comletes the proof of the lemma.

Now we are ready to prove the following theorem.

Theorem 5.1 *The language HQ is a Marcus external contextual language with choice.*

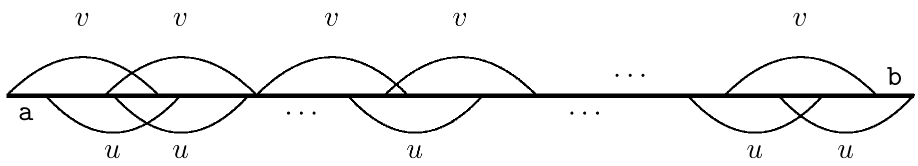

Fig. 5.2

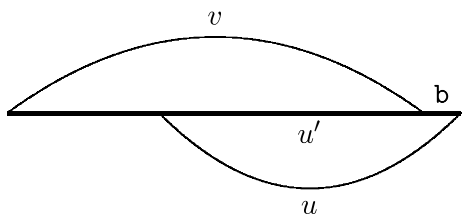

Fig. 5.3

Proof Notice that the theorem holds true for $|X| = 1$. Hence we assume that $|X| \geq 2$. Define $G = (X, A, C, \varphi)$ in the following way: Let $A = X$ and let $C = \{(\alpha, \beta) : \alpha\beta \in X^+, |\alpha\beta| = 1\}$, Moreover, let for every $w \in X^*$, $\varphi(w) = \{(\alpha, \beta) : (\alpha, \beta) \in C, \alpha w \beta \in HQ\}$. By the above definition of the grammar G, it is easy to see that $L_{ex}(G) \subseteq HQ$. Now we prove that $HQ \subseteq L_{ex}(G)$ by induction. First, we have $(X \cup X^2) \cap HQ \subseteq L_{ex}(G)$. Now, assume that $(X \cup X^2 \cup \cdots \cup X^n) \cap HQ \subseteq L_{ex}(G)$ for some $n \geq 2$. Let $u \in X^{n+1} \cap HQ$ and let $u = awb$ where $a, b \in X$. By Lemma 5.3, we have $aw \in HQ$ or $wb \in HQ$. Notice that, in this case, $aw \Rightarrow_{ex} awb = u$ or $wb \Rightarrow_{ex} awb = u$. Since $aw \in HQ$ or $wb \in HQ$, $u = wab \in L_{ex}(G)$. Consequently, $u \in L_{ex}(G)$, i.e. $HQ \subseteq L_{ex}(G)$. This completes the proof of the theorem.

However, the other types of Marcus contextual grammars cannot generate HQ.

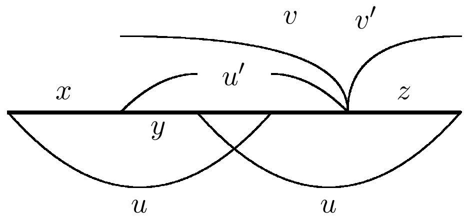

Fig. 5.4

Theorem 5.2 *The language HQ of all hyper primitive words over an alphabet X with* $|X| \geq 2$ *is not an internal contextual language with choice.*

Proof Suppose that there exists a $G = (X, A, C, \varphi)$ with $HQ = L_{in}(G)$. Then there exist $u, v, w \in X^*$ such that $uv \in X^+$ and $(u, v) \in \varphi(w)$. Let $a, b \in X$ with $a \neq b$. Then it is obvious that $a^{|uwv|}b^{|uwv|}wa^{|uwv|}b^{|uwv|}uwv \in HQ$ and $a^{|uwv|}b^{|uwv|}wa^{|uwv|}b^{|uwv|}uwv \Rightarrow_{in} (a^{|uwv|}b^{|uwv|}uwv)^2$. However, this contradicts the assumption that $HQ = L_{in}(G)$. Thus the statement of theorem must hold true.

By the above proof argument, we have the following.

Corollary 5.1 *The language HQ of all hyper primitive words over an alphabet X with* $|X| \geq 2$ *is not an internal contextual language without choice.*

Theorem 5.3 *The language HQ of all primitive words over an alphabet X with* $|X| \geq 2$ *is not an external contextual language without choice.*

Proof Assume that $G = (X, A, C)$ with $Q = L_{ex}(G)$. Then there exists $(u, v) \in C$ such that $(u, v) \neq (\epsilon, \epsilon)$ and $uv \notin a^+$ for some $a \in X$. It is obvious that $a^{|uv|}vua^{|uv|} \in HQ$. Moreover, $a^{|uv|}vua^{|uv|} \Rightarrow_{ex} (ua^{|uv|}v)^2 \notin HQ$. This contradicts the assumption that $HQ = L_{ex}(G)$. Thus the statement of the theorem must hold true.

References

[1] P. D´om´osi, M. Ito and S. Marcus, Marcus contextual languages consisting of primitive words, to appear in Discrete Mathematics.

[2] N.J. Fine and H.S. Wilf, Uniqueness theorems for periodic functions, Proceedings of the American Mathematical Society **16** (1965), 109-114.

[3] M. Lothaire, Combinatorics on Words, Addison-Wesley, Reading, 1983.

[4] Gh. Păun, Marcus Contextual Grammars, Kluwer, Dordrecht, Boston, London, 1997.

[5] H.J. Shyr, Free monoids and languages, Ho Min Book Company, Taichung, Taiwan, 1991.