Marcus External Contextual Grammars with Choice of the Languages of Primitive and Generalized Primitive Words. An Alternative Proof: To the honor Professor Masami Ito on his 70-th birthday (Algebraic Systems and Theoretical Computer Science)

Author(s)
Domosi, Pal; Fazekas, Szilard

Citation
数理解析研究所講究録 数学的解析・数理科学 (2012), 1809: 110-118

Issue Date
2012-09

URL
http://hdl.handle.net/2433/194467

Type
Departmental Bulletin Paper

Textversion
publisher

Kyoto University
Marcus External Contextual Grammars with Choice of the Languages of Primitive and Generalized Primitive Words. 
An Alternative Proof

To the honor Professor Masami Ito on his 70-th birthday

Pál Dömösi, * and Szilárd Fazekas†,‡

Abstract

In this paper we unify some well-known results describing an alternative proof that all of the languages of primitive, quasi-primitive, and hyper-primitive words are Marcus external contextual languages with choice.

Keywords: Formal languages, Marcus contextual languages, combinatorics of words and languages.

1 Preliminaries

Marcus contextual grammars were introduced and intensively studied by S. Marcus and his students (see [11, 13]). The word is primitive if it is not a

---

*College of Nyíregyháza, Institute of Mathematics and Informatics, H-4400 Nyíregyháza, Sóstói út 31/B, Hungary, e-mail: domosi@nyf.hu
†College of Nyíregyháza, Institute of Mathematics and Informatics, H-4400 Nyíregyháza, Sóstói út 31/B, Hungary, e-mail: domosi@nyf.hu
‡The authors are grateful to JSPS (Japanese Society for Promotion of Science), Kyoto Sangyo University, and Nyíregyháza College for their constant support. The first author was also supported by Czech Ministry of Education, Youth and Sport, and Hungarian National Development Agency No CZ-1/2009.
power of its proper prefix. The quasi-primitivity and hyper-primitivity are natural extensions of this concept. The relation of the language of primitive words to the Marcus contextual languages was studied first in [4]. On the line of this research, the relation of the language of quasi-primitive and hyper-primitive words and their certain further generalizations was described in [2] and [6]. In this paper we unify some of the results in [4, 2, 6] describing an alternative proof that all of the languages of primitive, quasi-primitive, and hyper-primitive words are Marcus external contextual languages with choice.

All notion and notations not defined here we refer to [3]. A word (over \( \Sigma \)) is a finite sequence of elements of some finite non-empty set \( \Sigma \). We call the set \( \Sigma \) an alphabet, the elements of \( \Sigma \) letters. If \( u \) and \( v \) are words over an alphabet \( \Sigma \), then their catenation \( uv \) is also a word over \( \Sigma \). Especially, for every word \( u \) over \( \Sigma \), \( u\lambda = \lambda u = u \), where \( \lambda \) denotes the empty word. Given a word \( u \), we define \( u^0 = \lambda \), \( u^n = u^{n-1}u \), \( n > 0 \), \( u^* = \{u^n : n \geq 0\} \) and \( u^+ = u^* \setminus \{\lambda\} \).

For every triplet \( u, v, w \) of words we say that \( u \) is a prefix, \( w \) is a suffix, and \( v \) is a subword of \( uvw \). If \( u(v, w) \) is nonempty then we speak about proper prefix (proper subword, proper suffix). A word \( z \) is called overlapping or bordered if there are \( u, v, w \in \Sigma^+ \) with \( z = uvw \).

The length \( |w| \) of a word \( w \) is the number of letters in \( w \), where each letter is counted as many times as it occurs. Thus \( |\lambda| = 0 \). By the free monoid \( \Sigma^* \) generated by \( \Sigma \) we mean the set of all words (including the empty word \( \lambda \)) having catenation as multiplication. We set \( \Sigma^+ = \Sigma^* \setminus \{\lambda\} \), where the subsemigroup \( \Sigma^+ \) of \( \Sigma^* \) is said to be the free semigroup generated by \( \Sigma \). Subsets of \( \Sigma^* \) are referred to as languages over \( \Sigma \).

A primitive word (over \( \Sigma \), or actually over an arbitrary alphabet) is a nonempty word not of the form \( w^m \) for any nonempty word \( w \) and integer \( m \geq 2 \). The set of all primitive words over \( \Sigma \) will be denoted by \( Q(\Sigma) \), or simply by \( Q \) if \( \Sigma \) is understood. \( Q \) has received special interest: \( Q \) and \( \Sigma^+ \setminus Q \) play an important role in the algebraic theory of codes and formal languages (see [7, 8, 9, 14]). If \( u \in \Sigma^+ \) can not be written into the form \( u = v^n v', n \geq 2 \) such that \( u, v \in \Sigma^+ \) and \( v' \) is a prefix of \( u \) then we say that \( u \) is strongly-primitive.

We say that a word \( u \in \Sigma^+ \) is covered by the word \( v \in \Sigma^+ \) if for every \( u', u'' \in \Sigma^*, a \in \Sigma \) with \( u = u'au'' \) there are \( v_1, v_2, v_3, v_4 \in \Sigma^* \) with \( u = v_1v_2av_3v_4, v = v_2av_3, u' = v_1v_2, u'' = v_3v_4 \).

A word \( u \in \Sigma^+ \) is called hyper-primitive if it can not be covered by any of its proper subwords.
$u \in \Sigma^+$ is super strongly primitive if $u \neq v^n v'$, $n \geq 2$ such that $v$ has a suffix $v''$ for which $v''v'$ is a prefix of $u$.

$u$ is called strongly hyper-primitive if $u \neq wv'$, where $w$ is covered by $v$, which is one of its proper subwords, and $v'$ is a prefix of $v$.

Finally, $u$ is hyper hyper-primitive if $u \neq wv'$, where $w$ is covered by $v$, which is one of its proper subwords, and $w$ has a suffix $v''$ such that $v''v'$ is a prefix of $v$.

Denote, in order, $SQ(\Sigma), HQ(\Sigma), SSQ(\Sigma), SHQ(\Sigma), HHQ(\Sigma)$, or, if $\Sigma$ is understood, then $SQ, HQ, SSQ, SHQ, HHQ$ the language of all strongly primitive, hyper primitive, super strongly primitive, strongly hyper-primitive, and hyper hyper primitive words (over $\Sigma$).

Moreover, denote by $|H|$ the cardinality of $H$ for every set $H$.

A (Marcus) contextual grammar with choice is a structure $G = (V, A, C, \varphi)$, where $V$ is an alphabet, $A$ is a finite language over $V$, $C$ is a finite subset of $V^* \times V^*$, and $\varphi : V^* \rightarrow 2^C$. If $\varphi(x) = C$ holds for every $x \in V^*$ then we say that $G$ is a (Marcus) contextual grammar without choice and then we omit $\varphi$ sometimes.

We define two relations on $V^*$ as usual: for any $x \in V^*$, we write

$x \Rightarrow_{ex} y$ if and only if $y = u xv$, for a context $(u, v)$ in $\varphi(x),$

$x \Rightarrow_{int} y$ if and only if $x = x_1x_2x_3, y = x_1ux_2vx_3$ for any $(u, v) \in \varphi(x_2)$.

Denote $\Rightarrow_{ex}, \Rightarrow_{int}$ the reflexive and transitive closure of these relations and let $L_{\alpha}(G) = \{x \in V^* : w \Rightarrow_{\alpha} x, w \in A\}$ 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 $G$. Now let $G = (V, A, \varphi)$, where $V$ is an alphabet, $A$ is a finite language over $V$, $C$ is a finite subset of $V^* \times V^*$, and $\varphi : V^* \times V^* \times V^* \rightarrow 2^C$.

Define the relation $\Rightarrow$ on $V^*$ such that $x \Rightarrow y$ for some $x, y \in V^*$ if and only if $x = x_1x_2x_3, y = x_1ux_2vx_3, x_1, x_2, x_3 \in V^*, (u, v) \in \varphi(x_1, x_2, x_3)$.

Moreover, let $\Rightarrow$ denote the reflexive and transitive closure of $\Rightarrow$. Thus $L(G)$ is defined to be a (Marcus) total contextual grammar (with or without choice) generated by $G$. If $\varphi(x_1, x_2, x_3) = C$ holds for every $x_1, x_2, x_3 \in V^*$ then we say that $G$ is a (Marcus) total contextual grammar without choice and sometimes we omit $\varphi$ having this property.

---

1Observe that the definition of $\varphi$ is not the same as before.
The following statement is a unified form of some results in [2, 4, 6]. It has been formulated by [6].

**Theorem 1** [2, 4, 6] The languages $Q, SQ, \text{ and } HQ$ are external contextual languages with choice. This is not true for the sets $SSQ, SHQ, \text{ and } HHQ$, furthermore, none of the sets $Q, SQ, HQ, SSQ, SHQ, \text{ and } HHQ$ is an external contextual language without choice or an internal contextual language with or without choice.

We shall use the following results.

**Theorem 2** [5] Let $u, v \in \Sigma^+, s, t \geq 1$, with $s \neq t$. If $\sqrt{u} \neq \sqrt{v}$ and $uv^s \notin Q$, then $uv^t \in Q$.

**Theorem 3** [1] Let $u, v \in Q, u^m = v^kw, k, m \geq 2$ for some prefix $w$ of $v$. Then $u = v$ and $w \in \{u, \lambda\}$.

**Theorem 4** [14][Borwein Lemma] Let $u \in \Sigma^+, u \notin a^+, a \in \Sigma$. Then at least one of $ua, u$ must be primitive.

**Theorem 5** [10] If $uv = vq, u \in \Sigma^+, v, q \in \Sigma^*$, then $u = wz, v = (wz)^kw, q = zw$ for some $w \in \Sigma^*, z \in \Sigma^+$ and $k \geq 0$.

We shall use the following two widely known consequences of Theorem 5.

**Proposition 6** For every bordered word $z \in \Sigma^+$ there exists a nonempty word $u \in \Sigma^+$ and a (not necessarily nonempty) word $v \in \Sigma^*$ having $z = uvu$.

**Theorem 7** [10] Let $u, v \in \Sigma^+$ with $uv = vu$. There exists $w \in \Sigma^+$ with $u, v \in w^+$.

---

2This statement can also be derived directly from [5].
2 Results

Next we show alternative proofs of some known results.

**Theorem 8** [2, 6] Let \( V \) be an alphabet with \( |V| \geq 2 \). If \( awb \in SQ \) where \( w \in V^* \) and \( a, b \in V \), then \( aw \in SQ \) or \( wb \in SQ \).

**Proof:** Suppose the contrary. Then \( aw, wb \in SPer \), i.e., there are \( u, v \in V^+ \), positive integers \( m, n \geq 2 \) such that \( u' \) is a prefix of \( u \), \( v' \) is a prefix of \( v \), and \( u^m u' = aw, v^n v' = wb \).

Then \( u = aw_1 w_2 \) and \( u' \in \{ \lambda, aw_1 \} \) for some \( w_1, w_2 \in V^* \). Similarly, \( v = w_3 b w_4 \) and \( v' \in \{ \lambda, w_3 b \} \) for an appropriate pair \( w_3, w_4 \in V^* \). Thus we can write \( w = (w_1 w_2 a)^m w_1 = (w_3 b w_4)^n w_3 \). By the symmetricity we may assume \( |w_1| \leq |w_3| \). Thus \( (w_1 w_2 a)^m = (w_3 b w_4)^n w' \) for some prefix \( w' \) of \( w_3 \). Applying Theorem 3, \( \sqrt{w_1 w_2 a} = \sqrt{w_3 b w_4} \). Therefore, \( w_3 b w_4 = w_3 w'' a \) for some \( w'' \in V^* \). Hence \( awb = a(w_3 b w'' a)^n w_3 b = (aw_3 b w'')^n a w_3 b \notin SQ \), a contradiction.

We can get the same conclusion if \( w = (w_1 w_2 a)^m w_1 = (w_3 b w_4)^n w_3 \) and \( n > 2 \) (or \( w = (w_1 w_2 a)^{-1} w_1 w_2 = (w_3 b w_4)^n w_3 \) and \( m > 2 \)). Thus let \( w = (w_1 w_2 a)^m w_1 w_2 = (w_3 b w_4)^n w_3 \). By the symmetricity we may assume \( |w_1| \leq |w_3| \). Thus \( (w_1 w_2 a)^m = (w_3 b w_4)^n w' \) with \( m \geq 2 \). Applying again Theorem 3, \( \sqrt{w_1 w_2 a} = \sqrt{w_3 b w_4} \). Therefore, \( awb = a w_3 b w_4 b w_3 w_4 b = (aw_3 b w_4)^2 a \notin SQ \), a contradiction.

We can derive the impossibility of \( w = w_1 w_2 a w_1 w_2 = (w_3 b w_4)^n w_3 \) and \( n \geq 2 \) in the same way.

The rest of the cases is the equality \( w_1 w_2 a w_1 w_2 = (w_3 b w_4)^n w_3 \). But then \( |w_1 w_2| = |w_3 w_4| \) which implies \( w_1 w_2 a = w_3 b \), i.e., \( a = b \). Then \( awb = a w_3 b w_4 b w_3 w_4 b = (aw_3 b w_4)^2 a \notin SQ \), a contradiction again. \(\square\)

**Lemma 9** If a word \( w \) can be covered by a word \( va \), with \( v \in \Sigma^* \), \( a \in \Sigma \), then \( vb \) is not a subword of \( w \), for any \( b \in \Sigma \), \( b \neq a \).

**Proof:** Consider a covering of \( w \) by \( va \). We will assume that \( vb \) can occur in \( w \) and show that it leads to a contradiction.

There are two possibilities for \( vb \) to occur in \( w \):

**Case 1.** \( vb \) is a proper subword (not only pre- or suffix) of \( v'v \), where \( v' \) is a prefix of \( v \): in this case \( vb \) is neither a prefix nor a suffix of \( v'v \) because

\( a = b \) is possible.
va \neq vb. Thus v has two different borders, i.e. by Proposition 6, \( v = x_1ux_1 \) and \( v = x_2y'x_2 \). Without loss of generality we can assume \( |x_2| < |x_1| \). Then \( x_1 \) itself is bordered, hence, applying Proposition 6 again, \( x_1 = xyx \), for some \( x, y \). This gives us \( v = xyyuxyx \) and because \( v \) overlaps twice with itself (by \( xyx \) and also by \( x \)), \( v = xyyuxyx = xuyx, \) for some \( z \), but then \( x \) is a suffix of \( z \) and immediately before it is \( y \), so \( xyyuxyx = xuyx \). Simplifying gives us \( yxy = uxy \), hence

\[ xuyx = xuyuxyx = xuyxux with v = xuyuxyx, \]  

taking away the first \( x \), we get \( uxyuxyx = yxyuxyx \), so \( uxy = (yx)^2 \). Therefore, by Theorem 7, \( uxy, (yx)^2 \in z^+ \) for some \( z \in \Sigma^+ \). From here applying (1), \( v = xz^k \), where \( z \) is a primitive word and \( k \geq 3 \). Moreover, since \( x \) is a suffix of \( v \), we get \( x = z'z^j \), with \( z' \) a suffix of \( z \) and \( j < k \), so \( z = z''z' \) and \( v = z'(z''z')^{j+k} \), with \( z''z' \) primitive, therefore \( z'(z''z')^{j+k}b \) would have to be a proper subword of either \( z'(z''z')^{j+k}a \) or \( z'(z''z')^i \), with \( i > j + k \). In both cases the first letter of \( z'' \) would have to be at the same time \( a \) and \( b \), contradiction.

**Case 2.** \( vb \) is a proper subword of \( vav \). In this case \( v \) from \( va \) overlaps the first \( v \) in \( vav \) with a part \( u_1 \) and the second with \( u_2 \), that is, \( v = u_1au_2 \) and \( v = u_2bu_1 \). If \( |u_1| = |u_2| \), we instantly get \( a = b \), contradiction. Without loss of generality \( |u_1| < |u_2| \), and then \( u_1 \) is a border of \( u_2 \) so, applying Theorem 5, for some \( x \in \Sigma^*, y \in \Sigma^+ \) we have \( u_1((xy)^i)x = x(yx)^j \) and \( u_2((xy)^j)x = x(yx)^j \), with \( 1 \leq i < j \). This gives \( v = x(yx)^iax(yx)^j = x(yx)^jbx(yx)^i \). Taking away \( x(yx)^i \) from both sides we get \( ax(yx)^{j-i} = x(yx)^{j-b} \). By this equality, \( x \neq \lambda \) implies \( ax = xc \) and \( dx = xb \) for some \( c, d \in \Sigma \). Hence we could get \( x \in a^+ \cap b^+ \), a contradiction. Therefore, \( x = \lambda \). Then \( ay^{j-i} = y^{j-i}b \) with \( a \neq b \) and \( i < j \). (By \( a \neq b \), \( i = j \) would be impossible even if we would not suppose before \( i < j \).) By this connection, \( y \neq \lambda \) implies \( ay = yc \) and \( dy = yb \) for some \( c, d \in \Sigma \). Then \( y \in a^+ \cap b^+ \), which is impossible unless \( a = b \).

**Theorem 10** [6] For any word \( w \) and (not necessarily distinct) letters \( a, b \in \Sigma \), if \( aw, wb \notin HQ \), then \( awb \notin HQ \).

**Proof:** If \( aw \notin HQ \), then there is some hyper-primitive \( av \) which covers \( aw \). Similarly, there is some hyper-primitive \( ub \) which covers \( wb \). Without loss
of generality, we can assume $|v| \leq |u|$. Then, $u$ is a suffix of $v$, therefore wherever there is an occurrence of $v$ in the string, it ends in $u$. Now, Lemma 9 tells us that if $ub$ covers $wb$, and $c \neq b$, then $uc$ is not a subword of $wb$.

There are two cases.

Case 1. $a \neq b$. Whenever $v$ appears in the string $wb$, it should be followed by $b$. From here, we get that $avb$ covers $awb$, so $awb \notin HQ$.

Case 2. $a = b$. Whenever $v$ appears in the string $wa$, it should be followed by $a$. From here, we get that $ava$ covers $awa$, so $awa \notin HQ$. □

**Corollary 11** Let $V$ be an alphabet with $|V| \geq 2$. If $awbc \in XQ$, where $XQ \in \{Q, SQ, HQ\}$, $w \in V^*$ and $a, b, c \in V^4$, then one of $aw, awb, wbc$ is in $XQ$.

**Proof:** If $XQ = Q$ and $awb \notin a^+$, then Theorem 4 implies that one of $aw, awb$ should be in $Q$. If $XQ = Q$ and $awb \in a^+$, then $awbc \in XQ$ implies $c \neq a$. In this case, $wbc \in a^+c$ with $a \neq c$, for which $wbc \in Q$ obviously holds. If $XQ \in \{SQ\}$ then by Theorem 8, if $XQ \in \{SQ, HQ\}$ then by Theorem 10 we have that one of $awb, wbc$ should be in $XQ$. □

On the basis of Lemma 11, similarly to Theorem 12 published by [6], next we show an alternative (and unified) proof of the next statement which is a union of three previous results.

**Theorem 12** [2, 4, 6] All of the languages $Q, SQ, HQ$ are Marcus external contextual languages with choice.

**Proof:** Let $G = (V, A, C, \varphi)$ be be an external Marcus contextual grammar with choice defined by $A = V$, $C = \{((\lambda, \lambda), (\lambda, a), (\lambda, ab), (a, \lambda) : a, b \in V\}$, moreover, let for every $w \in V^*$, $z \in \varphi(w)$ with

$$z = \begin{cases} 
\{(\lambda, \lambda)\} & \text{if } |V| = 1, \\
\{(a, \lambda)\} & \text{if } a \in V \text{ and } aw \in XQ, \\
\{(\lambda, a)\} & \text{if } a \in V \text{ and } wa \in XQ, \\
\{(\lambda, ab)\} & \text{if } a, b \in V \text{ and } wab \in XQ. 
\end{cases} \tag{5}$$

Moreover, let $XQ \in \{Q, SQ, HQ\}$. Obviously, the proposition holds true for $|V| = 1$. Hence we assume $|V| \geq 2$. By the definition of the grammar $G$, it is obvious that $L_{ex}(G) \subseteq SQ$. Now we prove that $SQ \subseteq L_{ex}(G)$

\footnote{a, b, c are not necessarily distinct}
by induction. By definition, $V \cap SQ = V(= A)$ and $V^2 \cap SQ = \{ab \mid a, b \in V, a \neq b\}$. Similarly, $V \cap V^2 = \{abc \mid a, b, c \in V, a \neq b, a \neq c, b \neq c\} \cup \{aab, abb \mid a, b \in V, a \neq b\}$. Moreover, by our construction, $a, b \in V$ and $a \neq b$ imply $a \Rightarrow_{ex} ab$. Thus we have $(V \cup V^2) \cap Q \subseteq L_{ex}(G)$. Similarly, by our construction, $a, b, c \in V$ and $a \neq b, a \neq c, b \neq c$ imply $ab \Rightarrow_{ex} abc$ and $a, b \in V$ and $a \neq b$ imply $ab \Rightarrow_{ex} abb$ and $ab \Rightarrow_{ex} aab$. Now, assume that $(V \cup V^2 \cup \cdots \cup V^n) \cap XQ \subseteq L_{ex}(G)$ for some $n \geq 3$. Let $u \in V^{n+1} \cap XQ$ and let $u = awbc \in XQ$ where $a, b, c \in V$. (Note that $a, b, c$ are not necessarily distinct.) Corollary 11 states that, by this condition, one of $aw, awb, wbc$ in $XQ$. Hence, either $aw \in XQ$ with $aw \Rightarrow_{ex} awbc$ or $awb \in XQ$ with $awb \Rightarrow_{ex} awbc$, or $wbc \in XQ$ with $wbc \Rightarrow_{ex} awbc$.

\[\square\]

References


