

Small Grammars and Primitive Words¹

GÉZA HORVÁTH

Department of Computer Science
Institute of Mathematics and Informatics
University of Debrecen
Debrecen, Egyetem tér 1, H-4032, Hungary
e-mail: geza@math.klte.hu

Abstract

In [4] Pál Dömösi, Dirk Hauschildt, Géza Horváth and Manfred Kudlek gave all context-free grammars in Chomsky normal form with not more than three nonterminals generating only primitive words. In this paper we extend the characterization for grammars with four nonterminals generating only primitive words over three nonterminal symbols.

1 Introduction

A number of recent papers have investigated the language of all primitive words over an alphabet of at least two letters, and considered its placement to the Chomsky-hierarchy (see [1]-[7]). In [1] the authors conjectured that this language is not context-free. This conjecture is still open. To help the research on this problem, in this report we consider certain ‘small’ and ‘maximal’ context-free grammars in Chomsky normal form which generate only primitive words. These grammars are small with respect to nonterminals and maximal with respect to productions. Since a necessary condition for the generated language to contain only primitive words (over terminal symbols) is that all sentential forms are also primitive words (over nonterminals) it suffices to consider only the sentential form languages.

The motivation for studying these grammars was to hopefully deduce from the structure of such grammars some insight for a proof of the conjecture that the set of all primitive words is not context-free by showing that there are always certain primitive words missing from the language generated by the grammar.

In the paper [4] the authors gave all context-free grammars with not more than three nonterminals generating only primitive words. The one and two nonterminal case was quite simple, but for the 3 nonterminal cases a computer program is created. Using the computer program the authors found 12 different maximal skeleton candidates, (one of them is not reduced) up to symmetries. The computer program in question checked that none of these 12 skeletons generates a non-primitive word W of nonterminals with length $|W| \leq 12$. The authors gave an exact proof for that all of these 12 skeletons generate only primitive words, where the length of the words was not limited any more. The program was run for the case of 4 nonterminals, starting with length 6, and repeating the procedure for length 8, 9, 10, 12, 14, 15, 16, 18, 20, 21, 22, 24, 25, 26, and 27. The program produced 413 candidates for maximal skeletons. This number is, however, too big to give an exact proof of that each of them is a maximal skeleton indeed. In fact, it turned out that all such grammars generate infinite subsets of all primitive words over 3 nonterminal symbol.

However, we observe that a grammar which generates all primitive words over an (at least two letter) alphabet and does not generate any non-primitive word can not contain the start symbol on the right hand side of any rule. So in this paper we limited our examination for skeletons, to those whose rules do

¹This work was supported by the Hungarian National Foundation for Scientific Research (OTKA T030140) and (OTKA T038225).

not contain the start symbol on the right hand side. Using the computer program we found 30 different maximal, above mentioned skeleton candidates over 4 nonterminals (2 of them are not reduced), and we present the exact mathematical proof for all of these skeletons generating only primitive words. It turned out that all such grammars generate infinite subsets of all primitive words over 3 nonterminal symbol.

2 Preliminaries

An *alphabet* Σ is a finite, nonempty set of symbols. Elements of Σ are called *letters*. A *word* is a finite sequence of the elements of an alphabet. The set of all words over Σ is denoted by Σ^* . We put $\Sigma^+ = \Sigma^* \setminus \{\lambda\}$, where λ denotes the *empty word* having no letters. The length of a word w , denoted by $|w|$, means the number of letters in w when each letter is counted as many times as it occurs. By definition, $|\lambda| = 0$. If u and v are words over an alphabet Σ , then their *catenation* uv is also a word over Σ . For any word uvw , we say that v is a *subword* of uvw .

A *language* over Σ is a set $L \subseteq \Sigma^*$. We extend the concept of catenation to the class of languages as usual. Therefore, if L_1 and L_2 are languages, then their catenation is $L_1L_2 = \{p_1p_2 \mid p_1 \in L_1, p_2 \in L_2\}$. Let p be a word. We put $p^0 = \lambda$ and $p^n = p^{n-1}p$ ($n > 0$). Thus p^k ($k \geq 0$) is the k -th *power* of p . If there is no danger of confusion, then sometimes we identify p with the singleton set $\{p\}$. Thus we will write p^* and p^+ instead of $\{p\}^*$ and $\{p\}^+$, respectively. A nonempty word is said to be *primitive* if it is not a proper power ($k > 1$) of another word. A word is *non-primitive* if it is not primitive. Let Q_Σ denote the set of primitive words over Σ .

An (*unrestricted generative, or simply, unrestricted*) *grammar* is an ordered quadruple $G = (N, \Sigma, S, P)$ where N and Σ are disjoint alphabets, $S \in N$, and P is a finite set of ordered pairs (U, V) such that V is a word over the alphabet $N \cup \Sigma$ and U is a word over $N \cup \Sigma$ containing at least one letter of N . The elements of N are called *variables* or *nonterminals*, and those of Σ *terminals*. $N \cup \Sigma$ is the *total alphabet* and S is called the *start symbol*. Elements (U, V) of P are called *productions* and are written $U \rightarrow V$. If $U \rightarrow V \in P$ implies $U \in N$ then G is called *context-free*. Especially, G is a context-free grammar given in *Chomsky normal form* if all productions are of the form $A \rightarrow BC$ or $A \rightarrow a$, where A, B, C are variables and a is a terminal.

A word W over $N \cup \Sigma$ *derives directly* a word W' , in symbols, $W \Rightarrow W'$, if and only if there are words W_1, U, W_2, V such that $W = W_1UW_2$, $W' = W_1VW_2$ and $U \rightarrow V$ belongs to P . W *derives* W' , or in symbols, $W \stackrel{*}{\Rightarrow} W'$ if and only if there is a finite sequence of words $W_0, \dots, W_k$ ($k \geq 0$) over $N \cup \Sigma$ with $W_0 = W$, $W_k = W'$ and $W_i \Rightarrow W_{i+1}$ for $0 \leq i \leq k-1$. Thus for every $W \in (N \cup \Sigma)^*$ we have $W \stackrel{*}{\Rightarrow} W$.

The set $S(G) = \{W \mid W \in (N \cup \Sigma)^*, S \stackrel{*}{\Rightarrow} W\}$ is called the set of *sentential forms* of G . The *language* $L(G)$ *generated by* G is defined by $L(G) = S(G) \cap \Sigma^*$. $L \subseteq \Sigma^*$ is a *context-free language* if we have $L = L(G)$ for some context-free grammar G .

The grammar $G_1 = (N_1, \Sigma_1, S_1, P_1)$ is *letter-isomorphic* to another grammar $G_2 = (N_2, \Sigma_2, S_2, P_2)$ if there exists a bijective mapping $\varphi : N_1 \cup \Sigma_1 \rightarrow N_2 \cup \Sigma_2$ such that $\varphi(S_1) = S_2$, $\{\varphi(A) \mid A \in N_1\} = N_2$, $\{\varphi(a) \mid a \in \Sigma_1\} = \Sigma_2$, moreover, $\{\varphi(x_1) \dots \varphi(x_s) \rightarrow \varphi(y_1) \dots \varphi(y_t) \mid x_1 \dots x_s \rightarrow y_1 \dots y_t \in P_1\} = P_2$. In this report we will not distinguish the letter-isomorphic grammars. Throughout this report by a *grammar* $G = (N, \Sigma, S, P)$ we mean a (λ -free) context-free grammar given in Chomsky normal form.

For any terminal symbol x we consider the set $N(x) = \{X \in N \mid X \rightarrow x \in P\}$. We say that $x \in \Sigma$ is *similar* to $y \in \Sigma$ with respect to $M \subseteq N$ with $M \neq \emptyset$ if $M \subseteq N(x) \cap N(y)$. (Then we also say, in short, that x is similar to y .)

A grammar G is *reduced* if it has the following properties :

- (i.) For any pair x, y of terminal symbols, $N(x) = N(y)$ implies $x = y$.
- (ii.) For any $x \in N \cup \Sigma$, there exists a pair $W_1, W_2 \in (N \cup \Sigma)^*$ such that the word $W_1xW_2 \in S(G)$.

We shall restrict our investigations to reduced grammars.

For $X \in N$ let $\Sigma(X) = \{x \in \Sigma \mid X \rightarrow x \in P\}$ where also $\Sigma(X) = \emptyset$ is possible.

Now we define the *skeleton* of $G = (N, \Sigma, S, P)$ as $G_0 = (N, \Sigma, P_0)$ with productions $P_0 = \{A \rightarrow BC \in P \mid A, B, C \in N\}$. The set $S(G_0) = \{W \in N^+ \mid S \stackrel{*}{\Rightarrow} W\}$ is called the (*sentential form*) *language* generated by the skeleton G_0 . We also say that a skeleton G_0 is *maximal* (with respect to the primitive

words) if $S(G_0)$ contains only primitive words, and if for any $X, Y, Z \in N, X \rightarrow YZ \notin P_0$ we obtain a non-primitive word in $S(G'_0)$ with $G'_0 = (N, S, P'_0)$ and $P'_0 = P_0 \cup \{X \rightarrow YZ\}$.

Note that $L(G) \subseteq Q_\Sigma \Rightarrow S(G_0) \subseteq Q_N$. The opposite implication $S(G_0) \subseteq Q_N \Rightarrow L(G) \subseteq Q_\Sigma$ holds if $\Sigma(X) \cap \Sigma(Y) = \emptyset$ for all $X, Y \in N$ with $X \neq Y$. The proof of this can be found in [4]. With these facts we may characterize all reduced grammars by using the characterization of maximal skeletons for any fixed cardinality of nonterminals. (If $|N| > 2$ then we have to take into consideration the similarity possibilities of terminals as well.)

3 Maximal Skeletons with 4 Nonterminals

Before start investigating the 4 nonterminal cases we have to prove the following theorem:

Lemma 1 *Suppose a grammar generates all of the primitive words over an (at least two letter) alphabet, and does not generate any non-primitive word. Then the start symbol can not be contained in any (at least two letter) word generated by the grammar.*

Proof: Suppose the grammar generates the word ASB , where S is the start symbol, and A, B are any words, one of them is not empty. Since the grammar generates all primitive words, it also generates the word $X^n B A X^n$ (for an appropriate letter X and integer n), and thus $A X^n B A X^n B$. This is a contradiction because $A X^n B A X^n B$ is not primitive. $\square$

In the rest of the paper we limit our examination to skeletons, the right hand side of whose rules do not contain the start symbol.

Using an appropriate computer program we found 30 different maximal skeleton candidates, up to symmetries, with 4 nonterminals (S, X, Y, Z) , the right hand side of whose rules do not contain the start symbol.

These symmetries are σ defined by $\sigma(A) = B, \sigma(B) = A, \delta$ defined by $\delta(A) = B, \delta(B) = C, \delta(C) = A$ and π defined by $\pi(A \rightarrow BC) = A \rightarrow CB$.

The computer program in question checked that none of these 30 skeletons generates a non-primitive word W of nonterminals with length $|W| \leq 12$.

It was run in several steps, using an input list of skeletons generating some non-primitive word such that any enlarged skeleton (some productions added) could be disregarded. Another list contained only such skeletons generating no non-primitive word with $|W| \leq 12$ such that any skeleton with a subset of productions could be disregarded. Finally we got a list of 30 candidates for maximal skeletons (with respect to primitive words, and up to symmetries). The program is the same which was given in the appendix of paper [4] with minor corrections.

There exist no more skeletons with the property from above. In this section we prove that each of them is a maximal skeleton indeed.

Consider $N = \{S, X, Y, Z\}$ with the start symbol S , and (for simplicity) denote by Q the set of all primitive words having at least two letter over $\{X, Y, Z\}$. We distinguish the following 30 cases.

Case 1

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, S \rightarrow YZ, \\ S \rightarrow ZY, X \rightarrow XX, Y \rightarrow YY, Z \rightarrow ZZ\}.$$

$$S(G_0) = (X^+ \cdot Y^+) \cup (Y^+ \cdot X^+) \cup (X^+ \cdot Z^+) \cup (Z^+ \cdot X^+) \cup (Y^+ \cdot Z^+) \cup \\ (Z^+ \cdot Y^+) \subset Q.$$

This is shown in the following way:

Let $L = (X^+ Y^+) \cup (Y^+ X^+) \cup (X^+ Z^+) \cup (Z^+ X^+) \cup (Y^+ Z^+) \cup (Z^+ Y^+) \subset Q$. By induction on $W \in L$, namely $XY \in L, YX \in L, XZ \in L, ZX \in L, YZ \in L, ZY \in L$, and any application of a production from P_0 on some $W \in L$ yields again some $W' \in L$, which implies $S(G_0) \subseteq L$.

On the other hand, any $W \in L$ can be derived from S . $\{XY, YX, XZ, ZX, YZ, ZY\} \subset S(G_0)$ is obvious. $X^m Y^n \in S(G_0)$ by $S \Rightarrow XY \xrightarrow{m-1} X^m Y \xrightarrow{n-1} X^m Y^n$ with productions $\{S \rightarrow XY, X \rightarrow XX, Y \rightarrow YY\}$. $Y^m X^n \in S(G_0)$ by $S \Rightarrow YX \xrightarrow{m-1} Y^m X \xrightarrow{n-1} Y^m X^n$ with productions $\{S \rightarrow YX, X \rightarrow XX, Y \rightarrow YY\}$, and so on. This implies $L \subseteq S(G_0)$.

Note that all productions have to be applied.

$S(G_0) \subseteq Q$ is obvious.

$XYX \notin S(G_0)$ implies $S(G_0) \subset Q$.

□

Case 2

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, S \rightarrow YZ, \\ S \rightarrow ZY, X \rightarrow XX, Y \rightarrow YY, Z \rightarrow XY, Z \rightarrow YX\}.$$

$$S(G_0) = (X^+ \cdot Y^+ \cdot X^*) \cup (Y^+ \cdot X^+ \cdot Y^*) \cup (X^+ \cdot \{Z\}) \cup (\{Z\} \cdot X^+) \cup \\ (Y^+ \cdot \{Z\}) \cup (\{Z\} \cdot Y^+) \subset Q.$$

The proof is similar to case 1, for $S(G_0) \subseteq L$ is shown by induction that any application of a production yields again an element from L , and for $L \subseteq S(G_0)$ all productions have to be applied.

$S(G_0) \subset Q$, because $XYZ \notin S(G_0)$.

□

Case 3

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, S \rightarrow YZ, S \rightarrow ZY, \\ X \rightarrow XX, Y \rightarrow YY, Z \rightarrow XY, Z \rightarrow XZ, Z \rightarrow ZY\}.$$

$$S(G_0) = (Y^* \cdot X^* \cdot \{Z, \lambda\} \cdot Y^*) \cup (X^* \cdot \{Z, \lambda\} \cdot Y^* \cdot X^*) \setminus \{\lambda, X^+, Y^+, Z\} \subset Q.$$

Again, the proof is similar to case 1, to show that $S(G_0) \subseteq L$, and for $L \subseteq S(G_0)$ all productions have to be applied.

$XYZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 4

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, S \rightarrow YZ, S \rightarrow ZY, \\ X \rightarrow XX, Y \rightarrow YY, Z \rightarrow XZ, Z \rightarrow ZX, Z \rightarrow YZ, Z \rightarrow ZY\}.$$

$$S(G_0) = (X^+ \cdot Y^+) \cup (Y^+ \cdot X^+) \cup (\{X, Y\}^* \cdot \{Z\} \cdot \{X, Y\}^*) \setminus \{Z\} \subset Q.$$

For $L \subseteq S(G_0)$ all productions have to be applied.

$S(G_0) \subseteq Q$, because all words $W \in S(G_0)$ contains one occurrence of the letter Z , or are of the form X^+Y^+ or Y^+X^+ , and $S(G_0) \subset Q$, since $XYX \notin S(G_0)$.

□

Case 5

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, S \rightarrow YZ, \\ S \rightarrow ZY, X \rightarrow XX, X \rightarrow YZ, X \rightarrow ZY\}.$$

Let $n(Y), n(Z)$ the number of Y, Z letters in the generated word. It can be show by induction that $n(Y) \pm 1 = n(Z)$ holds for all $W \in S(G_0)$, $|W| \geq 3$. Assume that $W = U^k$, $k > 1$, and let $n_U(Y), n_U(Z)$ the number of letters Y and Z in the word U . Now $n(Y) = kn_U(Y)$, $n(Z) = kn_U(Z)$, so $kn_U(Y) \pm 1 = kn_U(Z)$, $k > 1$, which is a contradiction.

$S(G_0) \subset Q$, since $XYX \notin S(G_0)$.

□

Case 6

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, S \rightarrow YZ, \\ S \rightarrow ZY, X \rightarrow XX, X \rightarrow YZ, Y \rightarrow YX, Z \rightarrow XZ\}.$$

We consider 3 subcases, based on the productions involving S :

- Start generating with one of the productions $\{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX\}$. In this case the $n(Y) \pm 1 = n(Z)$ equation holds.
- Now start with the production $S \rightarrow YZ$. We receive a skeleton which is letter-isomorphic to the case 9 of the 3 nonterminals which was proved in [4]. (The $X \rightarrow XX$ production is unnecessary.)

- (c) Finally start with the production $S \rightarrow ZY$. Suppose there exists a non-primitive word W , which is generated by the skeleton started with $S \rightarrow ZY$. In this case – since non-primitive words are closed under cyclic permutation, – there exists a non-primitive word W' , which is generated by the skeleton started with the $S \rightarrow YZ$ production. This contradicts (b).

$XYX \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 7

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, S \rightarrow YZ, \\ S \rightarrow ZY, X \rightarrow XX, Y \rightarrow XY, Y \rightarrow YX, Y \rightarrow YZ, \\ Y \rightarrow ZY, Z \rightarrow XZ, Z \rightarrow ZX\}.$$

In this case every $W \in S(G_0)$ contains either:

- (a) exactly one Y , when we start generating with $\{S \rightarrow XY, S \rightarrow YX, S \rightarrow YZ, S \rightarrow ZY\}$, or
 (b) exactly one Z , when we start generating with $\{S \rightarrow XZ, S \rightarrow ZX\}$.

$XYX \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 8

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, S \rightarrow YZ, \\ S \rightarrow ZY, X \rightarrow YZ, X \rightarrow ZY, Y \rightarrow XZ, Y \rightarrow ZX\}.$$

- (a) Start with one of the productions $\{S \rightarrow XZ, S \rightarrow ZX, S \rightarrow YZ, S \rightarrow ZY\}$. Then any generated word contains exactly one X or exactly one Y .
 (b) Start with one of the productions $\{S \rightarrow XY, S \rightarrow YX\}$. Then $n(X) + n(Y) = 2$ holds for every generated word, and if $n(X) = 2$ or $n(Y) = 2$, then $n(Z)$ is odd.

$XYX \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 9

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, X \rightarrow XX, \\ Y \rightarrow YY, Y \rightarrow ZZ, Z \rightarrow YY, Z \rightarrow ZZ, Y \rightarrow YZ, \\ Y \rightarrow ZY, Z \rightarrow YZ, Z \rightarrow ZY\}.$$

$$S(G_0) = (X^+ \cdot \{Y, Z\}^+) \cup (\{Y, Z\}^+ \cdot X^+) \subset Q.$$

The proof is similar to case 1. To show $L \subseteq S(G_0)$, only productions $\{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, X \rightarrow XX, Y \rightarrow YY, Z \rightarrow ZZ, Y \rightarrow YZ, Y \rightarrow ZY\}$ have to be used.

$YZ \notin S(G_0)$ implies $S(G_0) \subset Q$.

□

Case 10

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, X \rightarrow XX, Y \rightarrow YY, \\ Z \rightarrow YY, Z \rightarrow XY, Z \rightarrow YX, Z \rightarrow XZ, Z \rightarrow ZX\}.$$

$$S(G_0) = (X^+ \cdot \{Z, Y^+\} \cdot X^+) \setminus \{Z, Y^+\} \subset Q.$$

Again, the proof is similar to case 1. To show $L \subseteq S(G_0)$, only productions $\{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, X \rightarrow XX, Y \rightarrow YY, Z \rightarrow XY, Z \rightarrow XZ\}$ have to be applied.

$YZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 11

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, X \rightarrow XX, \\ Y \rightarrow YY, Z \rightarrow YY, Z \rightarrow XY, Z \rightarrow XZ, Z \rightarrow ZY\}.$$

$$S(G_0) = (X^* \cdot \{Z, \lambda\} \cdot Y^* \cdot X^*) \setminus \{\lambda, X^+, Y^+, ZY^*\} \subset Q.$$

Similar to case 1 again. For $L \subseteq S(G_0)$ all productions have to be applied except $Z \rightarrow YY$.
 $YZ \notin S(G_0)$, so $S(G_0) \subset Q$. □

Case 12

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, X \rightarrow XX, \\ Y \rightarrow YY, Y \rightarrow YZ, Z \rightarrow XZ\}.$$

- (a) If we start with one of the productions $\{S \rightarrow XZ, S \rightarrow ZX\}$, any generated word contains exactly one Z .
- (b) If we start with the production $S \rightarrow YX$, any generated W word has a form $Y\{Y, X^*Z\}^*X^+$. if $W = U^k$, $k \geq 2$, then $U = YU'X$ for some U' , and $U = ZU'X$ or $U = XU'X$, which is a contradiction.
- (c) Start with the production $S \rightarrow XY$. In this case – since non-primitive words are closed under cyclic permutation, and starting with $S \rightarrow YX$ we receive only primitive words, – all of the generated words are primitive.

$YZ \notin S(G_0)$ implies $S(G_0) \subset Q$. □

Case 13

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, X \rightarrow XX, X \rightarrow YZ, \\ X \rightarrow ZY, Y \rightarrow XY, Y \rightarrow YX, Z \rightarrow XZ, Z \rightarrow ZX\}.$$

Let $n(Y), n(Z)$ the number of Y, Z letters in the generated word. It can be show by induction that $n(Y) \pm 1 = n(Z)$ holds for all $W \in S(G_0)$. Let $W = U^k$, $k > 1$, and let $n_U(Y), n_U(Z)$ the number of letters Y and Z in the word U . Then $n(Y) = kn_U(Y)$, $n(Z) = kn_U(Z)$, so $kn_U(Y) \pm 1 = kn_U(Z)$, $k > 1$, which is a contradiction.

$YZ \notin S(G_0)$, so $S(G_0) \subset Q$. □

Case 14

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, X \rightarrow XX, \\ X \rightarrow YZ, Y \rightarrow XY, Y \rightarrow XZ\}.$$

- (a) If we start with the production $S \rightarrow XY$, any generated word W has the form $\{X, Y\}V\{XZ^i, YZ^j\}$ for some V , where $i \geq 1$, i odd, $j \geq 2$, j even. In W every letter Y , every word XZ^i , $i \geq 1$, i odd, and every word YZ^j , $j \geq 2$, j even, is followed by Z . Now, if $W = U^k$, $k \geq 2$, then $U = \{X, Y\}U'$ for some U' , and $U = ZU'$, which is a contradiction.
- (b) Start with the production $S \rightarrow YX$. In this case – since non-primitive words are closed under cyclic permutation, and starting with $S \rightarrow XY$, we receive only primitive words, – all of the generated words are primitive.
- (c) If we start with the production $S \rightarrow XZ$, the generated word W has the form XZ^+ or YZ^+ or $\{X, Y\}V\{XZ^i, YZ^j\}$ for some V , where $i \geq 1$, i odd, $j \geq 2$, j even. In W every word XZ^i , $i \geq 1$, i odd, and every word YZ^j , $j \geq 2$, j even is followed by Z . Now, if $W = U^k$, $k \geq 2$, then $U = \{X, Y\}U'$ for some U' , and $U = ZU'$, which is a contradiction.
- (d) Start with the production $S \rightarrow ZX$. In this case – since non-primitive words are closed under cyclic permutation, and starting with $S \rightarrow XZ$, we receive only primitive words, – all of the generated words are primitive.

$YZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 15

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, X \rightarrow XX, \\ Y \rightarrow XY, Y \rightarrow YX, Y \rightarrow XZ, Y \rightarrow ZX, Z \rightarrow XY, \\ Z \rightarrow YX, Z \rightarrow XZ, Z \rightarrow ZX\}.$$

The generated word contains exactly one Y , or exactly one Z . This can be proved by simple induction.
 $YZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 16

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, Y \rightarrow YY, Y \rightarrow ZZ, \\ Z \rightarrow YY, Z \rightarrow ZZ, X \rightarrow XY, X \rightarrow YX, X \rightarrow XZ, \\ X \rightarrow ZX, Y \rightarrow YZ, Y \rightarrow ZY, Z \rightarrow YZ, Z \rightarrow ZY\}.$$

The generated word contains exactly one X . This can be proved by simple induction.
 $YZ \notin S(G_0)$ implies $S(G_0) \subset Q$.

□

Case 17

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, Y \rightarrow YY, \\ Y \rightarrow ZZ, Y \rightarrow XZ\}.$$

- (a) Start with one of the productions $\{S \rightarrow XZ, S \rightarrow ZX\}$. The generated language is $\{XZ, ZX\}$.
- (b) Start with the $S \rightarrow YX$ production. Any generated word W has the form $\{X, Y, Z^i\}V\{X\}$ for some V , where $i \geq 2$, i even. In W every X is followed by an odd number of Z . If $W = U^k$, $k \geq 2$, then $U = \{X, Y, Z^i\}U'$ for some U' , where i even, and $U = Z^iU'$, where i odd. This is a contradiction.
- (c) Start with the production $S \rightarrow XY$. In this case – since non-primitive words are closed under cyclic permutation, and starting with $S \rightarrow YX$, we receive only primitive words, – all of the generated words are primitive.

$YZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 18

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, Y \rightarrow YY, \\ Z \rightarrow ZZ, Y \rightarrow XZ, Y \rightarrow YZ\}.$$

- (a) Start with one of the productions $\{S \rightarrow XZ, S \rightarrow ZX\}$. The generated language is $(XZ^+)^+ \cup (Z^+X) \subset Q$.
- (b) Start with the production $S \rightarrow YX$, any generated word W has a form $\{X, Y\}V\{X\}$ for some V . In W every letter X is followed by a Z . Now, if $W = U^k$, $k \geq 2$, then $U = \{X, Y\}U'$ for some U' , and $U = ZU'$, which is a contradiction.
- (c) Start with the production $S \rightarrow XY$. In this case all of the generated words are primitive, since starting with $S \rightarrow YX$, we receive only primitive words.

$YZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 19

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, S \rightarrow XZ, S \rightarrow ZX, Y \rightarrow ZZ, Z \rightarrow XY\}.$$

Let $n(X), n(Y), n(Z)$ the number of letters X, Y, Z in the generated word.

- (a) First start with one of the productions $\{S \rightarrow XY, S \rightarrow YX\}$. It can be shown by induction that $2n(Y) + n(Z) - n(X) = 1$ holds for all $W \in S(G_0)$. Let $W = U^k$, $k > 1$, and let $n_U(X), n_U(Y), n_U(Z)$ the number of letters X, Y and Z in the word U . Then $n(X) = kn_U(X)$, $n(Y) = kn_U(Y)$, $n(Z) = kn_U(Z)$, so $k(2n_U(Y) + n_U(Z) - n_U(X)) = 1$, which is a contradiction.
- (b) Now start with the production $S \rightarrow XZ$. Here $2n(Y) + n(Z) = n(X)$ holds for all $W \in S(G_0)$. By induction it follows that for any proper prefix of any $W \in S(G_0)$: $2n(Y) + n(Z) < n(X)$. Let $W = U^k$, $k > 1$, and let $n_U(X), n_U(Y), n_U(Z)$ the number of letters X, Y and Z in the word U . Then $n(X) = kn_U(X)$, $n(Y) = kn_U(Y)$, $n(Z) = kn_U(Z)$, and – since the U is proper prefix of W – $k(2n_U(Y) + n_U(Z)) < kn_U(X)$, which is contradicting $2n(Y) + n(Z) = n(X)$.
- (c) Start with the production $S \rightarrow ZX$. In this case all of the generated words are primitive, since starting with $S \rightarrow XZ$, we receive only primitive words.

$YZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 20

$P_0 = \{S \rightarrow XY, S \rightarrow YX, X \rightarrow XX, X \rightarrow ZZ, X \rightarrow YZ, Y \rightarrow XY\}$.

- (a) If we start with the production $S \rightarrow XY$, any generated word W has the form $\{X, Y, Z^i\}V\{Y\}$ for some V , where $i \geq 2$, i even. In W every letter Y is followed by an odd number of occurrences of Z . Now, if $W = U^k$, $k \geq 2$, then $U = \{X, Y, Z^i\}U'$ for some U' , where i is even, and $U = Z^iU'$, where i is odd. This is a contradiction.
- (b) Start with the production $S \rightarrow YX$. In this case all of the generated words are primitive, since starting with $S \rightarrow XY$, we receive only primitive words.

$XZ \notin S(G_0)$ implies $S(G_0) \subset Q$.

□

Case 21

$P_0 = \{S \rightarrow XY, S \rightarrow YX, X \rightarrow XX, Y \rightarrow YY, Z \rightarrow ZZ, X \rightarrow XZ, X \rightarrow ZX, Y \rightarrow YZ, Y \rightarrow ZY\}$.

$S(G_0) = (\{X, Z\}^* \cdot X \cdot Z^* \cdot \{Y, Z\}^* \cdot Y \cdot Z^*) \cup (\{Y, Z\}^* \cdot Y \cdot Z^* \cdot \{X, Z\}^* \cdot X \cdot Z^*) \subset Q$.

Similar to case 1 again. For $L \subseteq S(G_0)$ all productions have to be applied except $Z \rightarrow ZZ$.

Let the function h be the following: $h(S) = S$, $h(X) = X$, $h(Y) = Y$, $h(Z) = \lambda$. Let $L_1 = \{h(W) \mid W \in S(G_0)\}$. $L_1 = (X^+Y^+) \cup (Y^+X^+)$.

Suppose there exists a non-primitive word $W \in S(G_0)$. Then there exists a non-primitive word $h(W)$ in L_1 . This is a contradiction.

$XZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 22

$P_0 = \{S \rightarrow XY, S \rightarrow YX, X \rightarrow XX, Y \rightarrow YY, Z \rightarrow ZZ, X \rightarrow XZ, Y \rightarrow ZY, Z \rightarrow ZX, Z \rightarrow YZ\}$.

- (a) If we start with the production $S \rightarrow XY$, any generated word W has the form $\{X\}V\{Y\}$ for some V . In W every letter Y is followed by Y or Z . Now, if $W = U^k$, $k \geq 2$, then $U = XU'$ for some U' , and $U = \{Y, Z\}U'$. This is a contradiction.
- (b) Start with the production $S \rightarrow YX$. In this case all of the generated words are primitive, since starting with $S \rightarrow XY$ we receive only primitive words.

$XZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 23

$P_0 = \{S \rightarrow XY, S \rightarrow YX, X \rightarrow XX, Z \rightarrow XX, Z \rightarrow YY, Z \rightarrow ZZ, X \rightarrow XZ, X \rightarrow ZX, Z \rightarrow XZ, Z \rightarrow ZX\}$.

- (a) Start with the production $S \rightarrow XY$. Then any generated word W has the form $\{X, Z, Y^i\}V\{Y^j\}$ for some V , where $i \geq 2$, i even, $j \geq 1$, j odd. In W every odd number of occurrences of Y is followed by odd number of occurrences of Y . (Except the last Y -block.) If $W = U^k$, $k \geq 2$, then $U = \{X, Z, Y^i\}U'$ for some U' , where $i \geq 2$, i even, and $U = Y^jU'$, where $j \geq 1$, j odd. This is a contradiction.
- (b) Start with the production $S \rightarrow YX$. In this case all of the generated words are primitive, since starting with $S \rightarrow XY$ we receive only primitive words.

$XZ \notin S(G_0)$, implies $S(G_0) \subset Q$.

□

Case 24

$P_0 = \{S \rightarrow XY, S \rightarrow YX, X \rightarrow XX, Z \rightarrow YY, Z \rightarrow ZZ, X \rightarrow XZ, X \rightarrow ZX, Y \rightarrow YZ, Y \rightarrow ZY\}$.

- (a) Let the function h be the following: $h(S) = S$, $h(X) = X$, $h(Y) = Y$, $h(Z) = \lambda$. Let $L_1 = \{h(W) \mid W \in S(G_0)\}$.

Start with the production $S \rightarrow XY$. Let $W \in S(G_0)$. In this case the word $h(W)$ has a form $\{X, Y^i\}V\{Y^j\}$ for some V , where $i \geq 2$, i even, $j \geq 1$, j odd. In $h(W)$ every odd number of occurrences of Y is followed by odd number of occurrences of Y . (Except the last Y -block.) If $h(W) = U^k$, $k \geq 2$, then $U = \{X, Y^i\}U'$ for some U' , where $i \geq 2$, i even, and $U = Y^jU'$, where $j \geq 1$, j odd. This is the proof that L_1 contains only primitive words.

Suppose that exists a non-primitive word $W \in S(G_0)$. Then exists a non-primitive word $h(W)$ in L_1 . This is a contradiction.

- (b) Start with the production $S \rightarrow YX$. In this case all of the generated words are primitive, since starting with $S \rightarrow XY$, we receive only primitive words.

$XZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 25

$P_0 = \{S \rightarrow XY, S \rightarrow YX, X \rightarrow XX, Z \rightarrow ZZ, X \rightarrow XZ, X \rightarrow YZ, Y \rightarrow XY, Z \rightarrow ZX\}$.

- (a) If we start with the production $S \rightarrow XY$, any generated word W has the form $\{X, Y\}V\{Y\}$ for some V . In W every letter Y is followed by Z . Now, if $W = U^k$, $k \geq 2$, then $U = \{X, Y\}U'$ for some U' , and $U = ZU'$. This is a contradiction.
- (b) Start with the production $S \rightarrow YX$. In this case all of the generated words are primitive, since starting with $S \rightarrow XY$ we receive only primitive words.

$XZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 26

$P_0 = \{S \rightarrow XY, S \rightarrow YX, X \rightarrow ZZ, Z \rightarrow XY, Z \rightarrow YX\}$.

Let $n(X), n(Y), n(Z)$ the number of letters X, Y, Z in the generated word. It can be show by induction that $2n(X) + n(Z) - n(Y) = 1$ holds for all $W \in S(G_0)$. Let $W = U^k$, $k > 1$, and let $n_U(X), n_U(Y), n_U(Z)$

the number of letters X , Y and Z in the word U . Then $n(X) = kn_U(X)$, $n(Y) = kn_U(Y)$, $n(Z) = kn_U(Z)$, so $k(2n_U(Y) + n_U(Z) - n_U(X)) = 1$, which is a contradiction.

$XZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 27

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, Z \rightarrow XX, Z \rightarrow YY, Z \rightarrow ZZ, \\ X \rightarrow XZ, X \rightarrow ZX, Y \rightarrow YZ, Y \rightarrow ZY\}.$$

Let the function h be the following: $h(S) = S$, $h(X) = X$, $h(Y) = Y$, $h(Z) = \lambda$. Let $L_1 = \{h(W) \mid W \in S(G_0)\}$. Every $W \in S(G_0)$ contains X and Y , so if there exists a non-primitive word $W \in S(G_0)$, then there exists a non-primitive word $W' \in L_1$ too. So to prove that $S(G_0)$ does not contain any non-primitive word, it is enough to prove that L_1 does not contain any non-primitive words.

Every $W' \in L_1$ has the following form:

1. $\{XY, YX\} \subset L_1$,
2. If $W' \in L_1$, then we can insert XX or YY into W' before or after any letter, and we receive another word from L_1 .

Let $V = U^k$, $k > 1$ be a non-primitive word over the alphabet $\{X, Y\}$. Remove all of the occurrences of XX and YY from U , and repeat until there exists no XX or YY in the word. The received U' word has the following form:

$$\{\{XY\}^i \cdot \{X, \lambda\}, \{YX\}^i \cdot \{Y, \lambda\} \mid i \geq 0\}.$$

Now remove all of the occurrences of XX and YY from U'^k , and repeat until there exists no XX or YY in the word. Denote the received word by V' . Now if $|U'|$ was even, then V' has the form:

$$\{\{XY\}^{i+k}, \{YX\}^{i+k} \mid i \geq 0, k > 1\},$$

and if $|U'|$ was odd, then V' has the form:

$$\{U', \lambda\}.$$

Suppose there exists $W' \in L_1$, $W' = U^k$, $k > 1$ non-primitive word. In this case the W' contains odd number of X and odd number of Y , k is odd, and $|U|$ is even. Execute the above mentioned algorithm on W' . The received word W'' has the form $\{\{XY\}^{i+k}, \{YX\}^{i+k} \mid i \geq 0, k > 1\}$, because $|U'|$ is even. However we started from XY or YX , and these words do not have the form $\{\{XY\}^{i+k}, \{YX\}^{i+k} \mid i \geq 0, k > 1\}$. This is a contradiction.

$XZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 28

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, X \rightarrow YZ, Y \rightarrow ZX, Z \rightarrow YX\}.$$

- (a) Let us start with the production $S \rightarrow YX$. We receive a skeleton which is letter-isomorphic to the case 11 of the 3 nonterminals which was proved in [4].
- (b) Start with the production $S \rightarrow XY$. In this case all of the generated words are primitive, since starting with $S \rightarrow YX$, we receive only primitive words.

$XZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Finally we get the following non-reduced skeletons:

Case 29

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, X \rightarrow XX, Y \rightarrow YY, Z \rightarrow XX, \\ Z \rightarrow YY, Z \rightarrow ZZ, Z \rightarrow XY, Z \rightarrow YX, Z \rightarrow XZ, \\ Z \rightarrow ZX, Z \rightarrow YZ, Z \rightarrow ZY\}.$$

$$S(G_0) = (X^+ \cdot Y^+) \cup (Y^+ \cdot X^+) \subset Q.$$

This is a non-reduced skeleton, since there does not exist $W \in S(G_0)$ which contains Z .

The proof is similar to case 1. To show $L \subseteq S(G_0)$, only productions $\{S \rightarrow XY, S \rightarrow YX, X \rightarrow XX, Y \rightarrow YY\}$ have to be used.

$XZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

Case 30

$$P_0 = \{S \rightarrow XY, S \rightarrow YX, X \rightarrow XX, Z \rightarrow XX, Z \rightarrow YY, \\ Z \rightarrow ZZ, Y \rightarrow XY, Y \rightarrow YX, Z \rightarrow XY, Z \rightarrow YX, \\ Z \rightarrow XZ, Z \rightarrow ZX, Z \rightarrow YZ, Z \rightarrow ZY\}.$$

$$S(G_0) = (X^* \cdot Y \cdot X^*) \setminus \{Y\} \subset Q.$$

This is a non-reduced skeleton, since there does not exist $W \in S(G_0)$ which contains Z .

The proof is similar to case 1. To show $L \subseteq S(G_0)$, only productions $\{S \rightarrow XY, S \rightarrow YX, Y \rightarrow XY, Y \rightarrow YX\}$ have to be used.

$XZ \notin S(G_0)$, so $S(G_0) \subset Q$.

□

References

1. Dömösi, P., Horváth, S., Ito, M., Formal languages and primitive words, a.) *Proc. First Conf. on Scientific Communication, Univ. Oradea, Romania, 1991*; b.) *Publ. Math. (Debrecen)*, 42(1993), 315-321.
2. Dömösi, P., Horváth, S., Ito, M., Kászonyi, L., Katsura, M., Some combinatorial properties of words, and the Chomsky hierarchy, *Proc. 2nd Int. Coll. Words, Languages and Combinatorics, Kyoto, Japan, 25-28 Aug., 1992*, ed.: M. Ito and H. Jürgensen, World Scientific Publishers, Singapore, 105-123, 1994.
3. Ito, M., Katsura, M., Shyr, H. J., Yu, S. S., Automata accepting primitive words, *Semigroup Forum*, 37 (1988), 45-52.
4. Dömösi, P., Horváth, G., Kudlek, M., Hauschildt, D., Some Results on Small Context-free Grammars Generating Primitive Words, a.) *Technical Report FBI-HH-B-187/96, University of Hamburg, 1996*; b.) *Publ. Math. Debrecen* 54, 667-686, 1999.
5. Horváth, S., Strong interchangeability, nonlinearity and related properties of primitive words, *manuscript, Budapest, August, 1994*. *Report FBI-HH 183/96, FB Informatik, Universität Hamburg, 1996*.
6. Dömösi, P., Horváth, S., Ito, M., Kászonyi, L., Katsura, M., Formal languages consisting of primitive words, *Proc. Conf. FCT'93*, ed.: Z. Ésik, Springer LNCS 710, 194-203, 1993.
7. Petersen, H., The ambiguity of primitive words, *Proc. STACS'94, Springer LNCS 775*, 679-690, 1994.