THE PARALLEL GENERATION OF RECURSIVELY ENUMERABLE LANGUAGES USING ONLY CONTEXT-FREE PRODUCTIONS AND SIX NONTERMINALS

Ing. Petr BLATNÝ, Doctoral Degree Programme (1)
Dept. of Information Systems, FIT, BUT
E-mail: blatny@fit.vutbr.cz

Ing. Radek BIDLO, Doctoral Degree Programme (1)
Dept. of Information Systems, FIT, BUT
E-mail: bidlor@fit.vutbr.cz

Supervised by: Dr. Alexander Meduna

ABSTRACT

This paper defines the notion of an E0L grammar on a free group. The transformation of any type-0 grammar to an equivalent E0L grammar on a free group is demonstrated. Next, an algorithm reducing the number of nonterminals is introduced.

1 INTRODUCTION

In E0L grammars, the notion of a (direct) derivation has always been defined on letter monoids generated by total alphabets of these grammars. In our paper, however, this notion is introduced on free groups generated by total alphabets of these grammars.

The resulting grammars are simple and natural modification of the classic definition of E0L grammars. Moreover, their generative capacity is remarkably increased. Indeed, they generate the family of recursively enumerable languages.

2 PRELIMINARIES

We assume that the reader is familiar with formal language theory (see [1]).

For an alphabet V, $\overline{V}$ denotes the free group generated by V under the operation of concatenation. The identity element is ε and $\bar{x} \in V$ denotes the inverse element of x such that $x\bar{x} = \bar{x}x = \varepsilon$ for all $x \in V$.

An *E0L grammar* is a quadruple $G = (V, \Sigma, P, w)$, where V is a total alphabet, $\Sigma \subseteq V$ is a finite set of terminals, P is a finite set of productions of the form $a \to x$, $a \in V$, where $x \in V^*$ and $w \in V^+$ is the start string (axiom).

Let $x, y \in V^*$. Then x *directly derives* y in G, written as $x \Rightarrow_G y$, provided that $x = x_1 x_2 \ldots x_k$, $y = y_1 y_2 \ldots y_k$, $k \geq 1$, and for all i, $1 \leq i \leq k$, $x_i \to y_i \in P$. In the standard manner, $\Rightarrow_G^n$, $\Rightarrow_G^+$ and $\Rightarrow_G^*$ denote the n-fold product of $\Rightarrow_G$, $n \geq 0$, the transitive closure of $\Rightarrow_G$ and the transitive and reflexive closure of $\Rightarrow_G$ respectively.

The language of G, $L(G)$, is defined as $L(G) = \{\omega \in \Sigma^* : w \Rightarrow_G^* \omega\}$.

In the standard manner, we can introduce the relations $\Rightarrow_G$, $\Rightarrow_G^i$, $\Rightarrow_G^+$ and $\Rightarrow_G^*$, where the subscript G is usually omitted when understood.

An *E0L grammar on a free group* Γ, a **FG**(E0L) for short, is a pair $\Gamma = (G, \overline{V})$, where $G = (V, \Sigma, P, w)$ is an E0L grammar, and $\overline{V}$ is a free group generated by V under the operation of concatenation. The *direct derivation on* $\overline{V}$ is defined as follows. For every $x, y \in \overline{V}$, $x \Rightarrow_\Gamma y$ if $x = x_1 x_2 \ldots x_k$, $y = y_1 y_2 \ldots y_k$, $k \geq 1$, and for all i, $1 \leq i \leq k$, $x_i \to y_i \in P$, $x_i, y_i \in \overline{V}$. The relations $\Rightarrow_\Gamma^i$, $\Rightarrow_\Gamma^+$ and $\Rightarrow_\Gamma^*$ have the usual meaning, and they are defined on $\overline{V}$.

3 RESULTS

In this section, we introduce a construction of E0L grammars on free groups. The family of recursively enumerable languages is denoted by **RE**.

Lemma 3.1 For every phrase-structure grammar, $H = (V, \Sigma, P_H, S_H)$, $V = N \cup \Sigma$, there exists an equivalent phrase-structure grammar $G = (V \cup \{X, Y, S_G\}, \Sigma, P_G, S_G)$, such that $\{X, Y, S_G\} \cap V = \emptyset$ and each production in P_G has one of these forms:

$$AB \to AC, \quad A \to x, \quad AY \to YA,$$

$$S_G \to XS_H, \quad XY \to X, \quad X \to \varepsilon$$

where $A, B, C \in N$, $x \in N^2 \cup \Sigma$ and every sentential form of G, w, satisfies $w \in \{XY, X, \varepsilon\}V^*$.

Proof 3.1 (we only describe the construction)

Construction

Consider the grammar $H = (V, \Sigma, P_H, S_H)$, $V = N \cup \Sigma$. Without any loss of generality, assume that H satisfies the Penttonen normal form; that is, every production in P_H has one of these forms:

- $AB \to AC$, where $A, B, C \in N$
- $A \to BC$, where $A, B, C \in N$
- $A \to a$, where $A \in N$ and $a \in \Sigma \cup \{\varepsilon\}$

(see theorem 4 on page 391 in [3]). Furhtemore, assume that $\{X,Y,S_G\} \cap V = \emptyset$. Define the grammar $G = (V \cup \{X,Y,S_G\}, \Sigma, P_G, S_G)$, where P_G is constructed as follows:

- for every $A \to x \in P_H$, add $A \to x$ to P_G
- for every $AB \to AC \in P_H$, add $AB \to AC$ to P_G
- for every $A \to \varepsilon \in P_H$, add $A \to Y$ to P_G
- for every $A \in N$, add $AY \to YA$ to P_G
- add $S_G \to XS_H$, $XY \to X$ and $X \to \varepsilon$ to P_G

where $A,B,C \in N$, $x \in N^2 \cup \Sigma$

The construction of G is completed.

Theorem 3.1 **FG**(E0L)=**RE**

Proof 3.2 (we only describe the construction)

Consider any phrase-structure grammar $G = (V_G \cup \{X,Y\}, \Sigma, P_G, S)$, $V_G = N \cup \Sigma$. Without any loss of generality, assume that G satisfies the properties described in Lemma 3.1.

Construction

We construct the **FG**(E0L) grammar $G_\Gamma = (V, \Sigma, P_\Gamma, S)$, where $V = V_G \cup N_{CS} \cup \overline{N} \cup \{X,Y\}$. A new set of productions P_Γ is defined as follows:

I for every $A \to x \in P_G$, add $A \to x$ to P_Γ

II for every $AB \to AC \in P_G$,
add $A \to A\langle ABC \rangle$, $B \to \langle \overline{ABC} \rangle C$ to P_Γ and $\langle ABC \rangle, \langle \overline{ABC} \rangle$ add to N_{CS}

III for every $AY \to YA \in P_G$, add $A \to Y\langle AY \rangle$, $Y \to \langle \overline{AY} \rangle A$ to P_Γ
and $\langle AY \rangle, \langle \overline{AY} \rangle$ add to N_{CS}

IV for $XY \to X \in P_G$, add $X \to X\langle XY \rangle$, $Y \to \langle \overline{XY} \rangle$ to P_Γ
and $\langle XY \rangle, \langle \overline{XY} \rangle$ add to N_{CS}

V for $X \to \varepsilon \in P_G$, add $X \to \langle X \rangle \langle \overline{X} \rangle$ to P_Γ
and $\langle X \rangle, \langle \overline{X} \rangle$ add to N_{CS}

VI for every $Z \in V - \overline{N}$, add $Z \to Z$ to P_Γ

VII for every $Z \in \Sigma \cup N \cup \{X,Y\}$, add $\overline{Z}$ to $\overline{N}$

where $A,B,C \in N$, $x \in (N^2 \cup \{Y\} \cup \Sigma \cup \{X\}N)$.

The construction of G_Γ is completed.

Next, we introduce a construction of grammars on free groups with the reduced number of nonterminals. The family of languages generated by E0L grammars on free groups with the reduced number of nonterminals is denoted by **FG**(E0L)**R**.

Lemma 3.2 For every phrase-structure grammar, $H = (V, \Sigma, P, S)$, there exists an equivalent phrase-structure grammar, $G = (V_G, \Sigma, P_G, S)$, such that $V_G = N_G \cup \Sigma$ and each production in P_G has one of these forms:

- $AB \to CD$, where $A \neq C$ and $A, B, C, D \in N_G$
- $A \to BC$, where $A \neq B$ and $A, B, C \in N_G$
- $A \to a$, where $A \in N_G$, $a \in \Sigma \cup \{\varepsilon\}$

Proof 3.3 (we only describe the construction)

Construction

Let $H = (V, \Sigma, P, S)$ be a grammar and $N = V - \Sigma$. Without any loss of generality, assume that H satisfies the Kuroda normal form; that is, every production in P has one of these forms:

- $AB \to CD$, where $A, B, C, D \in N$
- $A \to BC$, where $A, B, C \in N$
- $A \to a$, where $A \in N$, $a \in \Sigma \cup \{\varepsilon\}$

Define the grammar $G = (V_G, \Sigma, P_G, S)$, where $V_G = N' \cup N \cup \Sigma$ and P_G is constructed as follows:

- for every $AB \to AD \in P$, add $AB \to A'D'$, $A'D' \to AD$ to P_G and add A', D' to N'
- for every $A \to AB \in P$, add $A \to A'B'$, $A'B' \to AB$ to P_G and add A', B' to N'
- add all other productions from P to P_G

A formal proof that H and G are equivalent is simple and left to the reader.

Theorem 3.2 **FG**(E0L)**R=RE**

Proof 3.4 (we only describe the construction)

Consider that $G = (V, \Sigma, P, S)$ is a phrase-structure grammar, $N = V - \Sigma = \{A_1, \ldots, A_n\}$, $A_1 = S$. Without any loss of generality, assume that G satisfies the properties described in Lemma 3.2.

Construction

We construct the **FG**(E0L)**R** grammar $\Gamma = (V_\Gamma, \Sigma, P_\Gamma, s_\Gamma)$, where $N_\Gamma = V_\Gamma - \Sigma = \{0, \overline{0}, 1, \overline{1}, 2, \overline{2}\}$. Define the injections, $h : N \to \{0, 1\}^+$ and $\overline{h} : N \to \{\overline{0}, \overline{1}\}^+$, such that for every $A_i \in N$, $h(A_i) = (i)_2 rev((i)_2)$ and $\overline{h}(A_i) = inv(h(A_i))$, where $(i)_2$ is the binary representation of i on $\lceil \log_2 |N| \rceil$ bits, for $i = 1, 2, \ldots, n$. Note that the inverses to $0, 1$ and $2 \in N_\Gamma$ are $\overline{0}, \overline{1}$ and $\overline{2} \in N_\Gamma$ respectively. The start string of Γ, s_Γ, is $h(S)2$.

The set of productions P_Γ is constructed as follows:

I for every $X \in V_\Gamma$, add $X \to X$ to P_Γ

II for every $AB \to CD \in P$, add $2 \to \overline{h}(B)\overline{2h}(A)\overline{2}2h(C)2h(D)2$ to P_Γ

III for every $A \to BC \in P$, add $2 \to \overline{h}(A)\overline{2}2h(B)2h(C)2$ to P_Γ

IV for every $A \to a \in P$, add $2 \to \overline{h}(A)a$ to P_Γ

where $A, B, C, D \in N$ and $a \in \Sigma \cup \{\varepsilon\}$.

The construction of Γ is completed.

Corollary 3.1 **FG**(E0L)=**FG**(E0L)**R**

4 CONLUSION

The inverses of the free group allow to remove the context-sensitive productions. Moreover, the encoding of nonterminal symbols reduces their number to exactly six.

REFERENCES

[1] Salomaa, A.: Formal Languages, New York: Academic Press 1973

[2] Meduna, A.: Context Free Derivations on Word Monoids, Acta informatica 27: 781-786, Springer-Verlag, 1990

[3] Penttonen, M.: One-Sided and Two-Sided Context in Fromal Grammars, Information and Control 25, 371-392, 1974

[4] Meduna, A., Kolář, D.: Homogenous Grammars with a Reduced Number of Non-Context-Free Productions, In: Information Processing Letters 81, p. 253-257, Amsterdam, NL, 2002

[5] Milne, J.S.: Group Theory, Course Notes, 2003

[6] Meduna, A.: Symbiotic E0L Systems, Acta Cybernetica, Szeged, 1992

[7] Geffert, V.: How to Generate Languages Using Only Two Pairs of Parentheses, J. Inform. Process. Cybernet. (EIK), 27, 303-15, 1991