On the generative capacity of the strict global grammars

by Sorin CIOBOTARU
and Gheorghe PAUN

Universitatea Bucureşti
Secţia de studiul sistemelor
Str.Mihai Moxa 3-5 Bucureşti, s 8
Romania

In [4] Levitina introduces a new restriction in the use of the context-free (CF) rules, namely the global rules. A production rule is said to be global if in every derivation it is used to rewrite all occurrences of its left side in a sentential form. A grammar which has CF and also global rules is said a global grammar. We shall consider the grammars which have only global rules; we shall call them strict global (SG) grammars. We shall study their generative capacity by means of the parameter Rep, a parameter closely connected to the notion of index (see [1],[5]). Also, we shall analise the parameter Rep as a measure of the syntactic complexity (see [3]).

Let $G = (V_T, V_N, S, P)$ be a Chomsky CF grammar and $V = V_T \cup V_N$. For x and y in V^* (the free monoid generated by V). We put $x \underset{G}{\Longrightarrow} y$ iff $x = x_1 A x_2$, $y = x_1 Z x_2$, where $x_1, x_2 \in V^*$ and $A \longrightarrow Z$ is a rule in P. The language generated by G is the set $L(G) = \left\{ x \in V_T^*;\ S \underset{G}{\overset{*}{\Longrightarrow}} x \right\}$, where $\underset{G}{\overset{*}{\Longrightarrow}}$ is the reflexive and transitive closure of $\underset{G}{\Longrightarrow}$.

The index of a derivation $D : S = x_o \underset{G}{\Longrightarrow} x_1 \underset{G}{\Longrightarrow} \cdots \underset{G}{\Longrightarrow} x_k$ is

$$\mathrm{Ind}(D, G) = \max_{o \leq j \leq k} \sum_{i=o}^{n} A_i\ (x_j),$$

where $V_N = \left\{ S = A_o, A_1, \ldots, A_n \right\}$ and $A_i(x_j)$ is the number of the occurrences of the nonterminal A_i in x_j. For w in L(G) we put

$$\mathrm{Ind}(w, G) = \min_{D} \mathrm{Ind}(D, G),$$

where $D : S \underset{G}{\overset{*}{\Longrightarrow}} w$. The index of G is

$$\mathrm{Ind}(G) = \sup_{w \in L(G)} \mathrm{Ind}(w, G)$$

and the index of L is

$$\mathrm{Ind}(L) = \min \left\{ \mathrm{Ind}(G);\ L = L(G) \right\}.$$

A global rule is a rule which is used in the following way: $x \Rightarrow y$ by a rule $A \longrightarrow Z$ if and only if $x = x_1 A x_2 A x_3 \dots x_{n-1} A x_n$, $y = x_1 Z x_2 Z x_3 \dots x_{n-1} Z x_n$, $n \geqslant 2$ and $A(x_i) = 0$ for $1 \leqslant i \leqslant n$. A grammar having only global rules is called a SG grammar.

If R is a restriction in the use of the rules of G, then we denote by G_R the grammar G with the restriction R. If Ψ is a class of grammars we put $L_\Psi = \{L$; there exists G in Ψ such that $L = L(G)\}$. We say that the restriction R modifies in the weak sense the generative capacity of the grammars of the class Ψ if there exists G in Ψ such that $L(G_R) \neq L(G)$; we write $R(\Psi)$. We say that R modifies in the strong sense the generative capacity of the grammars of Ψ if there exists G in Ψ such that $L(G_R)$ is not in L_Ψ ; we write $R[\Psi]$. Obviously, if $R[\Psi]$, then $R(\Psi)$.

Let us denote by Sg the strict global restriction.

Proposition 1. If Lin is the set of the linear grammars, then we don't have Sg(Lin).

Proposition 2. We have Sg(C) and Sg[C] where C is the class of the CF grammars.

Proof. Let us consider the grammar with the rules $S \longrightarrow Ab\ Ab\ A$, $A \longrightarrow aA$, $A \longrightarrow a$. $L(G_{Sg}) = \{a^n b a^n b a^n;\ n \geqslant 1\}$ is not a CF language.

Proposition 3. The class of the SG languages and the class of the matrix languages are uncomparable.

Proof. The set $L_1 = \{a^n b^n c^n;\ n \geqslant 1\}$ is a matrix language but it is not a global language (see [4]). Thus it is not a SG language. The set $L_2 = \{a^{2^n};\ n \geqslant 0\}$ is SG language (it may be generated by the SG grammar with the rules $S \longrightarrow SS$, $S \longrightarrow a$) but it is not a matrix language.

According to the notations used in the above definition of the index we define the parameter Rep in the following way:

$$\mathrm{Rep}(D, G) = \max_{\substack{o \leqslant j \leqslant k \\ o \leqslant i \leqslant k}} A_i(x_j),$$

$$\mathrm{Rep}(w, G) = \min_{D} \mathrm{Rep}(D,G),$$

$$\mathrm{Rep}(G) = \sup_{w \in L(G)} \mathrm{Rep}(w,G),$$

$$\mathrm{Rep}(L) = \min \{\mathrm{Rep}(G);\ L = L(G)\}.$$

Obviously, $\mathrm{Rep}(L) = 1$ for any linear language and $\mathrm{Rep}(L) \leqslant \mathrm{Ind}(L)$ for any CF language.

Proposition 4. For any CF language L, Rep(L) is finite if and only if Ind(L) is finite.

Proposition 5. For any CF grammar G with $\mathrm{Rep}(G) = n < \infty$ there exists a grammar G' such that L(G) = L(G') and Rep(G') = 1.

Proof. For n = 1 the assertion is true. Let us consider a grammar $G = (V_N, V_T, S, P)$ such that Rep(G) = n + 1. Let U_N be the set of the symbols of V_N which establish the value of Rep(G). If $U_N = \{A_1, A_2, \dots, A_k\}$ let us consider $\bar{U}_N = \{\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k\}$ where $\bar{A}_i$ are not in V. Let Q be the set of the rules of P in which occurs at least a symbol of U_N. Let $\bar{Q}$ be the set of the rules obtained from the rules of Q by the substitution of at least an occurrence of each non-terminal of U_N which occurs in the rule by the corresponding symbol of $\bar{U}_N$. Let us consider the grammar $G'' = (V_N \cup \bar{U}_N, V_T, S, P \cup \bar{Q})$. Obviously L(G") = L(G). It may be proved that $\text{Rep}(G'') \leqslant n$. By the induction hypothesis there exists G' such that L(G") = L(G') and Rep(G') = 1.

Proposition 6. If $\Psi = \{G;\ \text{Rep}(G) = 1\}$ then we don't have $Sg(\Psi)$.

Theorem 1. Any CF language of finite index is a SG language.

The theorem results from the propositions 4, 5 and 6.

Corollary. For any language of finite index, L, and for any $n \geqslant 1$, the language $L_n = \{w^n;\ w \in L\}$ is SG.

Theorem 2. There is a CF language of infinite index which is not a SG language.

Proof. Let us consider the language L generated by the grammar with the rules $S \longrightarrow SS$, $S \longrightarrow aSb$, $S \longrightarrow cS$, $S \longrightarrow c$. Let us consider the homomorphism defined by h(a) = a, h(b) = b, $h(c) = \varepsilon$. Obviously h(L) is the Dick language on the vocabulary $\{a, b\}$. Since the class of the languages of finite index is full AFL [3], it follows that Ind(L)= $= \infty$. In what follows, by the assertion "c^k is subword in w" we understand that $w = w_1 c^k w_2$ and $w_1 \neq w_1' c$, $w_2 \neq c w_2'$. We suppose that there is a grammar $G = (V_N, V_T, S, P)$ such that $L(G_{Sg}) = L$. For A in V_n we have three cases:

i) $L_A = \{w \in V_T^* ;\ A \underset{G_{Sg}}{\overset{*}{\Longrightarrow}} w\}$ is a finite language,

ii) $L_A = L_1 L_2 L_3$ where $L_2 \subset \{c\}^*$ and L_1, L_2 are finite languages,

iii) L_A is a finite union of languages of the form $L_1 L L_2$ where L_1, L_2 are finite languages.

Let L' be the set of w in L such that any derivation of w is of the form $S \underset{G_{Sg}}{\overset{*}{\Longrightarrow}} Z \underset{G_{Sg}}{\overset{*}{\Longrightarrow}} w$, with $A(Z) \geqslant 2$ for A with the property iii).

Let be also $L'' = \{w \in L;$ if c^k and c^i are subwords in w, then $k \neq i\}$. Obviously $L' \cap L'' \neq \emptyset$. On the other hand, we have $(L' \cap L(G_{Sg})) \cap L'' = \emptyset$. Contradiction.

Open problem. Does there exist a CF language of infinite index which is a SG language ?

Following Gruska [3] a measure K of syntactic complexity is said to be nontrivial if for any $n \geq 1$ there exists a language L such that $K(L) > n$. K is said to be bounded if for any $n \geq 1$ there exists a language L such that $K(L) = n$.

Because there are languages L for which $Rep(L) = \infty$, Rep is a nontrivial measure. As a consequence of the proposition 5 it results that

Rep is not a bounded measure. For any language L we have either $Rep(L) = \infty$, or $Rep(L) = 1$. Obviously, for any $n \geq 1$ there exists a grammar G such that $Rep(G) = n$. Moreover, we have:

Proposition 7. For any CF language L with $Rep(L) = 1$, and for any $n \geq 1$ there exists a grammar G_n such that $Rep(G_n) = n$.

Following Gruska [3], for K we put $K^{-1}(L) = \{G;\ L = L(G),\ K(G) = K(L)\}$. Then, two measures K_1 and K_2 are said to be compatible if for any CF language L we have $K_1^{-1}(L) \cap K_2^{-1}(L) \neq \emptyset$.

Proposition 8. Rep and Ind are compatible, but Rep and $K \in \{Var$, Prod, Symb$\}$ (see [3]) are uncompatible.

Proof. The first assertion follows from the proposition 5 and 4. To prove the second assertion it is sufficient to find a language L such that $Rep(L) = 1$ and every grammar G for L with $K(G) = K(L)$ has $Rep(G) \geq 2$. This languages is

$$L = \{a^n\ b^n\ a^m\ b^m;\ n,\ m \geq 0\}.$$

References

1. Brainerd B. An analog of a theorem about context free languages. Information and Control 11 (1968), 561-567.
2. Ginsburg S. The mathematical theory of context free languages. Mc Graw-Hill Book Company New-York (1966).
3. Gruska J. Descriptional complexity of context free languages. Proc. of symp. and summ. school.Math. fund of computer sci. High Tatras (1973).
4. Levitina M.K. On some grammars with global productions. N.T.I. 2(1972), 32-36 (in russ.).
5. Salomaa A. On the index of a context free grammar and language. Information and Control, 14 (1968), 474-477.
6. Salomaa A. Formal languages. Academic Press New York and London (1973).