KYBERNETIKA — VOLUME 9 (1973), NUMBER 3

A Note on Grammars with Regular Restrictions

JAROSLAV KRÁL

A context-free ε-free grammar with regular restrictions is a context-free ε-free grammar G over which a context-free rule r or G is applicable on a string x only if $x \in \gamma(r)$ where $\gamma(r)$ is a regular set. It is known from [1] that the context-free ε-free grammars with regular restrictions are as powerful as Chomsky's grammars of the type 0. It is shown, that the same result holds for the grammars, for which the condition $x \in \gamma(r)$ is replaced by the condition $x \in \gamma$ where γ is a regular set associated with the whole grammar (i.e. independent on the rule r to be applied).

A context-free grammar with regular restrictions [1] is a quituple $G = (V, U, R, \Phi, S)$, where $G' = (V, U, R, S)$ is a context-free grammar and $\Phi = \{\gamma(r) \mid r$ is a rule of G', $\gamma(r)$ is a regular set$\}$. A context-free grammar $G = (V, U, R, \Phi, S)$ with regular restrictions is a context-free ε-free grammar with regular restrictions if $G' = (V, U, R, S)$ is a context-free ε-free grammar.

Let G be a context-free (ε-free) grammar with regular restrictions. For $x, y \in (V \cup U)^*$ we write $x \Rightarrow_G y$ if there is a rule $r = (u, v) \in R$, $x = x_1 u x_2$, $y = x_1 v x_2$ and $x \in \gamma(r)$. $\Rightarrow_G^*$ is a transitive and reflexive closure of $\Rightarrow_G$.

Friš proved in [1] and [2], that contex-free ε-free grammars with regular restrictions (ε-CFRR grammars for short) are as powerful as Chomsky type 1 (context-sensitive) grammars, i.e. to each context sensitive grammar G there is a ε-CFRR grammar G_1 such that $L(G_1) = L(G_2)$ and vice versa to each ε-CFRR grammar G_2 there is a context sensitive grammar G_1 such that $L(G_2) = L(G_1)$.

Denote $T_1 = \{A \mid A = L(G)$ for a context-sensitive grammar $G\}$, $T^{rest} = \{A \mid A = L(G)$, G is a ε-CFRR grammar$\}$. A grammar is context-free ε-free with weak regular restriction (ε-CFWRR grammar for short) if it is a ε-CFRR grammar $G = (V, U, R, \Phi, S)$ where $\Phi = \{\gamma\}$ i.e. $\gamma(r_1) = \gamma(r_2)$ for each two rules r_1, r_2 of G. Let $T'^{rest} = \{A \mid A = L(G)$, G is a ε-CFWRR grammar$\}$.

As noted above $T_1 = T^{rest}$. We shall prove the following result.

Theorem. $T_1 = T'^{rest}$.

Proof. As it obviously holds $T'^{\text{rest}} \subset T^{\text{rest}} = T_1$ it suffices to prove that $T'^{\text{rest}} \supset T_1$.

Let $G = (V, U, R, S)$ be a context sensitive grammar. Without loss of generality we can assume that all the rules $r \in R$ are of the form $r = (h_1Ah_2, h_1\omega h_2)$ where A is a nonterminal symbol.

Let $G' = (W, U, P, \Phi, S)$ be a ε-CFWRR grammar of the following properties. $W_1 = \{\uparrow_r \mid \uparrow_r \text{ is a new symbol for each } r \in R\}$, $W = W_1 \cup V$. Let further $P = P_1 \cup P_2$ where

$$P_1 = \{\bar{r} \mid \bar{r} = (A, \uparrow_r), r = (h_1Ah_2, h_1\omega h_2) \in R\},$$

$$P_2 = \{\bar{\bar{r}} \mid \bar{\bar{r}} = (\uparrow_r, \omega), r = (h_1Ah_2, h_1\omega h_2) \in R\}.$$

Finally $\Phi = \{\gamma\}$ where

$$\gamma = (V \cup U)^* \cup \bigcup_{\substack{r \in R \\ r = (h_1Ah_2, h_1\omega h_2)}} (V \cup U)^* \, h_1 \uparrow_r h_2 (V \cup U)^*$$

From this construction it follows that if $D = (w_0, w_1, \ldots, w_n)$, $w_0 \in (V \cup U)^*$, $w_n \in U^*$ is a derivation over G' then in D a rule $\bar{r}$ from P_1 is applied on w_0 (the rules from P_2 are not applicable). On w_1 the rules from P_1 and the rule $\bar{\bar{r}}$ from P_2 can be applied. If a rule q from P_1 were applied on w_1 then a string w_2' with two occurrences of symbols from W_1 would be obtained. But w_2' does not belong to γ. It must be therefore $w_2' = w_n$ which violates the assumption $w_n \in U^*$. Therefore on w_1 the rule $\bar{\bar{r}}$ must be applied. It follows $w_2 \in (U \cup V)^*$, $w_0 \Rightarrow_G w_2$. From it follows that if $(S, \ldots, \ldots, w_n)$ is a derivation over G then $n = 2j$, $S \Rightarrow_G w_2 \Rightarrow_G w_4 \ldots \Rightarrow_G w_n$ and $L(G') \subset \subset L(G)$. Because the reverse inclusion is obvious the proof is complete.

It is worth of mention that from the equality of generative powers of the type 1 grammars and the ε-CFRR grammars it does not follow that the grammars with regular restrictions (and even with context-free restriction) are not worth of study. One reason for it is that context-free grammars with regular restriction could generate non context-free languages (such as Algol 60) in a more "natural" way than context sensitive languages. For such grammars phrase markers seems to have almost no reasonable meaning. One reason for is discussed in [4]. One says that a derivation $(w_0, w_1, \ldots, w_n)$ over a Chomsky grammar G has the property H_k, $k \geqq 1$, if each w_j can be expressed in the form $w_j = w_{j1}w_{j2} \ldots w_{js_j}$ where the length $|w_{ji}|$ of w_{ji} is not greater than k and for each $h \leqq j$ and $i \leqq s_j$ there is $w_{h\Theta_h(i,j)}$ such that $w_{h\Theta_h(,ij)} \Rightarrow \Rightarrow_G^* w_{ji}$. It is clear that each derivation over a context-free grammar has the property H_1. It is shown in [4] that the set $L_k(G) = \{x \mid \text{there is a derivation } (S, \ldots, x)$ over G of the property $H_k\}$ is a context-free set for every Chomsky type 0 grammar and every $k \geqq 1$.

It follows that in the case that $L(G)$ is a set which is not context-free then to each k there is an $x \in L(G)$ such that every derivation D of x from the initial symbol contains a member m having a nonterminal substring y of the lengths greater then k. Moreover

D has the property that in the subderivation D' of x from m all the parts of y are dependent, i.e. the subderivation from arbitrary part of y cannot be separated from the subderivations in another parts of y. This fact can hardly be reflected in phrase markers, but phrase markers are fundamental for the syntactical analysis.

(Received February 3, 1972.)

REFERENCES

[1] Friš, I.: Grammars with Partial Ordering of the Rules. Information and Control *12* (1968), 412–425.

[2] Friš, I.: Correction to the article Grammars with Partial Ordering of the Rules. Information and Control *14* (1969), 5.

[3] Ginsburg, S., Spanier, E. H.: Derivation Bounded Languages. J. of Comp. and System Sci. *4* (1968), 228–251.

[4] Král, J.: A Modification of the Substitution Theorem and some Necessary and Sufficient Conditions for a Set to be Context Free. Math. Systems Theory *4* (1970), 2, 129–139.

RNDr. Jaroslav Král, CSc.; Ústav výpočtové techniky ČVUT (Institute of Computing Techniques of the Czech Technical University), Horská 3, 120 00 Praha 2. Czechoslovakia.