Kybernetika

Eduard Kostolanský

Comments on effective and unambiguous context-free languages

Kybernetika, Vol. 5 (1969), No. 3, (181)--185

Persistent URL: `http://dml.cz/dmlcz/125843`

KYBERNETIKA ČÍSLO 3, ROČNÍK 5/1969

Comments on Effective and Unambiguous Context-free Languages

EDUARD KOSTOLANSKÝ

The paper defines the concept of effective language and grammar. The sufficient condition is found for the effectiveness of the language and comments are given as to unambiguous and effective languages.

EFFECTIVE GRAMMARS

The symbols and concepts currently used in the theory of formal languages as given e.g. in paper [1] will be applied. Let $G = (V, V_T, P, S)$ be a context-free grammar and $L(G)$ be a language generated by this grammar. Grammar and language are further conceived to be context-free.

V^∞ is the set of all strings over alphabet V. If $\alpha \in V^\infty$, then $y \subset \alpha$ if y is a substring of α.

$\mathscr{L}(G) = \{\alpha \mid \alpha \in V^\infty \,\&\, S^* \mathscr{L} \alpha\}$ i.e. $\mathscr{L}(G)$ is the set of all those strings $\alpha \in V^\infty$ (i.e. not only the terminal ones), which can be generated by grammar G.

$P = \{P_1 \to X_1, P_2 \to X_2, \ldots, P_n \to X_n\}$ i.e. the set of rules of grammar G is formed by rules $P_i \to X_i$, $i = 1, 2, \ldots, n$.

For our considerations it appears to be appropriate to formulate the definition of the language generated by a grammar G as follows.

If $P_i \to X_i$, $1 \leqq i \leqq n$ are rules of grammar G then $\bar{P} = \{X_i \to P_i, 1 \leqq i \leqq n\}$ and

Definition 1. $\alpha \in L(G)$ if a) there exists a sequence of strings

$$(1) \qquad p_1, p_2, \ldots, p_n$$

such that $p_1 = \alpha$, $p_n = S$ and for each $1 \leqq i < n(\geqq 2)$ is $p_i = x_1 w x_2$, $p_{i+1} = x_1 v x_2$ and $w \to v$ is a rule $\in \bar{P}$;

$$\text{b)} \qquad \alpha \in V_T^\infty .$$

The sequence (1) is called *a reduction sequence* for α and the process of forming this sequence *the reduction* of α. Let $\alpha \in L(G)$ and let p_n, $n = 1, 2, \ldots, k$ be formed similarly as in the reduction sequence but $\alpha_k \notin \mathscr{L}(G)$.

Then the sequence $p_1, p_2, \ldots, p_k$ is *a degenerated reduction sequence* for α.

Definition 2. $L(G)$ is *effective* if for no $\alpha \in L(G)$ exists degenerated reduction sequence.

The grammar that generates the effective language is called *an effective grammar.*

We introduce some concepts here.

Let $\alpha \in L(G)$ and let $x \subset \alpha$ (x is substring of α). x has the *index* j if it is obtained by using the rule $P_j \to X_j$, $1 \leqq j \leqq n$.

For $\alpha \in \mathscr{L}(G)$ the following set of strings is defined:

Definition 3. Set $\mathscr{L}\, G(\alpha)$ is formed by strings with following properties:

1. $y \subset \alpha$;
2. $y = x_1 x_2 \ldots x_l$, $x_i \in V$, $1 \leqq i \leqq l$;
3. There exists at least one i, $1 \leqq i \leqq l$ for which index x_i differs from k (see following 4) or y contains the own substring which is the right hand side of another rule of the grammar G.
4. $y = X_k$, (y is the right hand side $k - th$ rule of grammar G).

Thus the set $\mathscr{L}\, G(\alpha)$ is formed by substrings y of string α with the following properties. Every y can be obtained by using of at least two rules and it is simultaneously the right hand side of a certain rule $P_k \to X_k$ or y as the right hand side of a certain rule contains substring z which is the right hand side of another rule.

Example 1. $G = (V = S, a;\ V_T = a;\ P = S \to aS,\ S \to a;\ S)$ and $L(G) = \{a^n, n \geqq 1\}$.

$$\alpha = aaS \quad \text{belongs to} \quad \mathscr{L}(G)\,.\ \mathscr{L}\, G(\alpha) = \{aS\}\,,$$

for, aS is the right hand side of the rule $S \to aS$ and it contains simultaneously substring a which is the right hand side of the rule $S \to a$.

Example 2. $G = (V = S, R, b, a;\ V_T = a, b;\ P = S \to aRa,\ R \to bSb,\ R \to ba;\ S)$ and $L(G) = \{(ab)^n\, aa(ba)^{n-1}, n \geqq 1\}$.

$$\alpha = abSba \quad \text{belongs to} \quad \mathscr{L}(G)\,.\ \mathscr{L}\, G(\alpha) = \{ba\}\,,$$

for, ba is obtained by two rules $S \to aRa$ and $R \to bSb$ i.e. index b differs from index a, and simultaneously ba is the right hand side of the rule $R \to ba$.

Strings for which $\mathscr{L}\, G(\alpha) \neq \emptyset$ will form the set $\overline{LG}$.

Definition 4. $\overline{LG} = \{\beta \mid \beta \in \mathscr{L}\,(G)\,\&\, \mathscr{L}\, G(\beta) \neq \emptyset\}$.

Prove the following theorem:

Theorem 1. *Let* $\overline{LG} = \emptyset$. *Then* $L(G)$ *is effective.*

Proof. If $\overline{LG} = \emptyset$ that means that each string p_i from the reduction sequence for an arbitrary string $\alpha \in L(G)$ has the following properties:

There exists no string y such that

1. $y(=x_1, \ldots, x_z) \subset p_i$,
2. $y = X_k$,
3. index of x_j, $1 \leqq j \leqq z$ is different from k for at least one j or y contains the own substring w and $N \to w$ is the rule of grammar G.

That means that with the reduction there always occurs a substitution of the right side of a certain rule by the left side of this rule so that final result is the sequence of strings $p_1 = \alpha, p_2, \ldots, p_l = S$ i.e. always a reduction sequence and thus $L(G)$ is an effective language.

The reverse statement of theorem 1 does not hold.

Theorem 2. *There exist an effective languages in which* $\overline{LG} \neq 0$.

Proof. Consider the language generated by the grammar

$$G = (V = S, L, C, a, 1;\ V_T = a, 1;\ P = S \to SS,\ S \to SC,\ S \to L,\ L \to a,\ C \to 1;\ S) \tag{2}$$

forming strings of type $a^{i1}1^{i2}a^{i3}1^{i4} \ldots$ where $i_1 \geqq 1$, $il \geqq 0$ for $l \geqq 1$. This language is effective, for, any reduction in which string S^n $(n \geqq 1)$ is gained, being further reduced to $\alpha = S$. Hereby $\overline{LG} \neq 0$, for, to $\overline{LG}$ belongs e.g. string $\beta = SSC$.

An example of the effective language is $L(G) = \{a^n b^n, n \geqq 1\}$, which is generated with the grammar

$$G = (V = a, b, S;\ V_T = a, b;\ P = S \to ab,\ S \to aSb;\ S).$$

2. NOTES ON THE RELATION OF LANGUAGE EFFECTIVITY AND UNAMBIGUITY

It is said that language $L(G)$ generated by grammar G is ambiguous, if there exists $\alpha \in L(G)$, which grammar G generates in two different ways, i.e. if assigns to string α two different structural descriptions [2].

As to the relation between unambiguous and effective languages the following question can be raised:

Is the class of effective languages equal to the class of unambiguous languages or in what relation are these classes?

Theorem 3. *There exist effective languages that are not unambiguous and not every unambiguous language is effective.*

The following examples prove theorem 3.

Consider grammar (2) generating the effective language but not the unambiguous one, for e.g. two different structural descriptions can be assigned to the string $aa1$ (Fig. 1).

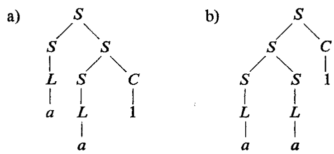

Fig. 1.

Reversely, the language generated by grammar

$$G = (V = S, A, B, D, E, a, b, c, d, e; V_T = a, b, c, d, e;$$
$$P = S \rightarrow Bd, S \rightarrow aE, B \rightarrow Ac, E \rightarrow bD, A \rightarrow ab, D \rightarrow ce; S)$$

is an unambiguous but not an effective one. E.g. in reducing string *abce* the degenerated reduction sequence, *abce, Ace, Be* can be formed.

(Received February 22nd, 1968.)

REFERENCES

[1] Bar Hiller Y., Perles M., Shamir E.: On formal properties of simple phrase structure grammars. Zeitschrift Für Phonetik, Sprachwissenschaft und Kommunikationsforschung, *14* (1961), 143—172.

[2] Chomsky N.: Formal properties of grammars. Handbook of Mathematical Psychology. Vol. 2, 1963, 328—418.

[3] Chomsky N.: On certain formal properties of grammars, Information and Control *2* (1959), 137—167.

[4] Ginsburg S.: The Mathematical Theory of Context Free Languaes. McGraw-Hill, 1966.

[5] Gruska J.: On Structural Unambiguity of Formal Languages, Czechoslovak Math. J. *15* (1965), 283—294.

[6] Irons E.: The Structure and Use of Syntax Directed Compiler. Annual Review in Automatic Programming V. *3* (1963).

Poznámky o efektívnych a jednoznačných gramatikách a jazykoch

EDUARD KOSTOLANSKÝ

V práci je definovaný pojem efektívnej gramatiky a jazyka. Sú hľadané postačujúce podmienky pre efektívnosť jazyka a skúmaný vzťah medzi efektívnymi a jednoznačnými jazykmi. Je ukázané, že existujú efektívne jazyky, ktoré nie sú jednoznačné a že nie každý jednoznačný jazyk je efektívny.

Eduard Kostolanský, prom. mat., Ústav technickej kybernetiky SAV, Dúbravská cesta, Bratislava.