Karel Čulík

On equivalent and similar grammars of Algol-like languages (Preliminary communication)

Commentationes Mathematicae Universitatis Carolinae, Vol. 5 (1964), No. 2, 93--95

Persistent URL: http://dml.cz/dmlcz/104962

Commentationes Mathematicae Universitatis Carolinae
5, 2 (1964)

ON EQUIVALENT AND SIMILAR GRAMMARS OF ALGOL-LIKE LANGUAGES

(Preliminary communication)

K. ČULÍK, Praha

Let $G = \langle T, N, \mathcal{R}, S \rangle$ be a context-free grammar, i.e. T and N are terminal and nonterminal vocabularies resp., $S \in N$ and in $\mathcal{R}$ are the rules $(a_0, a_1, a_2, \ldots, a_n)$, where $a_0 \in N$, $a_i \in T \cup N$ for each $1 \leq i \leq n$, $n \geq 1$ ($a_1, a_2 \ldots a_n$ is said to be strong over $T \cup N$). E.g. in ALGOL 60 T and N are sets of basic symbols and metalinguistic variables resp., $S = \langle$programm$\rangle$ and $\mathcal{R}$ contains elementary syntactic definitions $x ::= y\, z\, n$ (the metasymbol | meaning "or" is omitted). Let L be the language generated by G and let $\mathcal{P}$ be the set of all phrase markers of elements in L. The phrase markers were introduced by N. Chomsky and in [1] are defined as double graphs the vertices of which are labelled by symbols of $T \cup N$.

If we identify two nonterminal symbols x and y, i.e. if we substitute x instead of y in all places in all rules of $\mathcal{R}$ and if we omit y from N, we get a new grammar G^*. It is easy to see that $L^* \supset L$ and the mapping Φ of $\mathcal{P}$ into $\mathcal{P}^*$ is determined by the mentioned substitution. If Φ is a mapping onto $\mathcal{P}^*$ then x and y are said to be interchangeable in G.E.g. if to each $(a_0, a_1 a_2 \ldots a_n) \in \mathcal{R}$ where $a_0 = x$ (or $a_0 = y$) exists another rule $(b_0, b_1 b_2, \ldots b_m) \in \mathcal{R}$ such that $m = n$, $b_0 = y$ (or $b_0 = x$) and for each i, $1 \leq i \leq n$ either $a_i = b_i$ or $a_i, b_i \in \{x, y\}$, then x and y are interchangeable.

The homomorphism of a grammar G_1 onto a grammar G_2 is a mapping φ of $T_1 \cup N_1$ onto $T_2 \cup N_2$ such that

1) φ is an one-to-one mapping of T_1 onto T_2,

2) $(a_0, a_1 a_2 \dots a_n) \in \mathcal{R}_1$ implies $(\varphi(a_0), \varphi(a_1)\varphi(a_2)\dots \dots \varphi(a_n)) \in \mathcal{R}_2$ and

3) φ induces the mapping Φ of $\mathcal{R}_1$ onto $\mathcal{R}_2$. Two grammars are said to be equivalent if it is possible to map them homomorphically onto the same grammar.

Another way how to change the grammars are extensions and reductions. A grammar G_p is an extension of the grammar G_0 if there are grammars $G_1, G_2, \dots, G_{p-1}$ such that the following condition holds for each i, $1 \leq i \leq p$: there exists a rule $(a_0, a_1 a_2 \dots a_n) \in \mathcal{R}_{i-1}$, a symbol $b \notin \bigcup_{t=0}^{i-1} (N_t \cup T_t)$, an index j, $1 \leq j \leq n$ and an integer $k \geq 0$, $1 \leq j + k \leq n$ such that $T_i = T_{i-1}$, $N_i = N_{i-1} \cup \{b\}$, $\mathcal{R}_i = (\mathcal{R}_{i-1} - \{(a_0, a_1 a_2 \dots a_n)\}) \cup \{(a_0, a_1 \dots a_{j-1}\, b\, a_{j+k+1} \dots a_n)$ $(b, a_j a_{j+1} \dots a_{j+k})\}$ and $S_i = S_{i-1}$. E.g. it may be $\mathcal{R}_0 = \{(S, b\,v\,t), (S, b\,w)\}$ and $\mathcal{R}_2 = \{(S, b\,y), (y, w), (S, b\,x), (x, v\,t)\}$. In this case x and y are interchangeable, but they do not satisfy the above mentioned sufficient condition.

The composition + of two sets of rules $\mathcal{T}_1$ and $\mathcal{T}_2$ is defined as follows: $\mathcal{T}_1 + \mathcal{T}_2 = \{(a_0, x_0 y_1 x_1 \dots y_n x_n) ; x_i$ and y_j are some strings such that there are $(a_0, x_0 b_1 x_1 \dots \dots b_n x_n) \in \mathcal{T}_1$ and $(b_j, y_j) \in \mathcal{T}_2$ for some symbols b_j and for each j, $1 \leq j \leq n\}$.

A nonterminal symbol x of the grammar G is said to be reductible if there is no rule in $\mathcal{R}$ of the form (x, p), where p is a string containing x. The symbol x is reductible if and only if in $\mathcal{T}_1 + \mathcal{T}_2$ is no rule containing x, where $\mathcal{T}_1$ and $\mathcal{T}_2$ are the sets of all rules in $\mathcal{R}$ containing x in their right and left side resp.

Let x be a nonterminal reductible symbol in G . It is natural to construct a new grammar G^* as follows: $N^* = N - \{x\}$ and $\mathcal{R}^* = (\mathcal{R} \cup (\mathcal{T}_1 + \mathcal{T}_2)) - (\mathcal{T}_1 \cup \mathcal{T}_2)$. G^* is said to be direct reduction of G . A grammar G_p is called reduction of the grammar G_0 if there are $G_1, G_2, \ldots G_{p-1}$ such that G_i is direct reduction of G_{i-1} for each i , $1 \leq i \leq p$. Some simple examples of the reduction in ALGOL 60 are shown in [2].

Now two grammars are said to be strong or weak similar if they have equivalent extensions or reductions resp.

If x and y are interchangeable in G and G^* is direct reduction of G with the reduced symbol z, $x \neq z \neq y$, then x and y are interchangeable in G^* again. If two grammars are strong similar then they are weak similar too, but not conversely. E.g. $\mathcal{R}_1 = \{(S, b\ z), (b, x\ y)\}$ and $\mathcal{R}_2 = \{(S, x\ c), (c\ y\ z)\}$ are weak similar because $\mathcal{R} = \{(S, x\ y\ z)\}$ is their common reduction, but there are evidently no equivalent extensions of them. There are some lattice properties of the greatest extensions and smallest reductior in regard to the equivalence relation among the grammars.

R e f e r e n c e s :

[1] K. ČULÍK, Applications of graph theory to mathematical logic and linguistics, Theory of graphs and it applications

[2] K. ČULÍK, Formal structure of ALGOL and simplication of its description, 75-82, Symbolic languages in data processing (Roma 1962), Gordon-Breach, N.Y. 1963.