Kybernetika

Karel Čulík

Semantics and translation of grammars and ALGOL-like languages

Kybernetika, Vol. 1 (1965), No. 1, (47)--49

Persistent URL: `http://dml.cz/dmlcz/124848`

KYBERNETIKA ČÍSLO 1, ROČNÍK 1/1965

Semantics and Translation of Grammars and ALGOL-like Languages

KAREL ČULÍK

A notion of translation is defined in usual mathematical way as a mapping of the expressions under which their meanings are invariant. Further this mapping is described in certain constructive way.

The *language* L is determined by the set of *expressions* E and by a *semantics* S. E contains certain strings of *basic symbols* choosen from V_T and S is a function which joins a meaning $S(x)$ to each expression $x \in E$.

The *translation* from the language L into the language L' is the function T such that $T(x) \in E'$ and $S(x) = S'(T(x))$ for each $x \in E$. The problem si how to determine the translation T in some economical and constructive way (not as the set of couples x, $T(x)$ for all $x \in E$). This may be done if we use suitable grammars generating the given languages.

The set E is *generated* by the *contex-free grammar* $G = \langle V_T, V_N, \mathfrak{R} \rangle$ if $E = \bigcup_{x \in V_N} G(x)$, where $G(x)$ denotes the set of all expressions $e \in E$ such that there is a *derivation* in G beginning with x and ending with e. V_N contains the matalinguistic variables and $\mathfrak{R}$ syntactical definitions (or rules) $\mathfrak{w} = a_0 ::= a_1 a_2 \dots a_n$, where $a_0 \in V_N$ and $a_i \in V_N \cup V_T$ for each $i = 1, 2, \dots, n$.

In regard to the translation T the set V_T is divided in two parts V_A and V_P. V_A contains *auxiliary symbols* (as brackets, space, coma and other separators) and V_P the *proper symbols* (the symbols in V_P usually are again the strings over an alphabet A_P). Each rule $\mathfrak{w} \in \mathfrak{R}$ can be unically expressed in the *standard form* $a_0 ::= b_0 a_1 b_1 \dots \dots a_k b_k$, where b_i are either zero-strings or strings over V_A and $a_i \in V_P \cup V_N$; a_i is said to be the ith *standard symbol* of $\mathfrak{w}$.

Further we define an extension $\bar{S}$ of S as follows: $\bar{S}(x) = S(x)$ for $x \in E$ and $\bar{S}(x) = \bigcup_{y \in G(x)} S(y)$ for $x \in V_N$.

The grammar G of the language L(i.e. G generates E of L) is said to be *well translatable* into the grammar G' of L' if there is a mapping Φ of $\mathfrak{R}$ into $\mathfrak{R}'$ and another mapping τ of V_P into V'_P and of V_N into V'_N satisfying the following conditions:

(1) if $a_0 ::= b_0a_1b_1 \dots a_nb_n$ and $c_0 ::= d_0c_1d_1 \dots c_md_m$ are the standard forms of the rules $\mathfrak{w}$ and $\Phi(\mathfrak{w})$ resp. then $m = n$ and there is a permutation π of the set $\{1, 2, \dots, n\}$ such that $c_{\pi(i)} = \tau(a_i)$ for each $i = 1, 2, \dots, n$ and

(2) if $x_i \in G(a_i)$ and $y_i \in G'(c_i)$ and $S(x_i) = S'(y_{\pi(i)})$ for all $i = 1, 2, \dots, n$ then $S(b_0x_1b_1 \dots x_nb_n) = S'(d_0y_1d_1 \dots y_nd_n)$.

If G is well translatable into G' we obtain the expression $T(x) \in E'$ from the expression $x \in E$ using three algorithms (A), (B) and (C) as follows:

(*A*) *Analysis*

1. We begin with $x_0 = x$ and we make the 1st step (x_0 does not contain any symbol from V_N).

2. In the ith step we deal with x_{i-1} and we take a rule $\mathfrak{w} \in \mathfrak{R}$ (it is necessary to search it) such that $x_{i-1} = vb_0a_1b_1 \dots a_kb_kw$, where v a w are some suitable strings and $a_0 ::= b_0a_1b_1 \dots a_kb_k$ is the standard form of $\mathfrak{w}$. We construct $x_i = va_0w$ and we put $r(a_0) = i$, where the integer $r(a_0)$ is called the *rang of metalinguistic variable* a_0 in x_i. The rangs of metalinguistic varaibles contained in v or w in x_{i-1} remains unchanded also in x_i. We store the statements "the ith rule is" as the pair $(i, \mathfrak{w})$ and "hth rule was used to the jth standard symbol of the ith rule" as the triple (h, j, i) if $a_j \in V_N$ and $r(a_j) = h$ in x_{i-1}.

3. We stop if $x_n \in V_N$.

The result of analysis is a set P of pairs $(i, \mathfrak{w})$ and a set Q of triples (h, j, i). It is clear that by P and Q is determined the phrase marker of x in G.

(B) *Translation of phrase marker*

We construct $P' = \{(i, \Phi(\mathfrak{w})); (i, \mathfrak{w}) \in P\}$ and $Q' = \{(h, \pi_i(j), i); (h, j, i) \in Q\}$, where π_i is the permutation belonging in Φ to the ith rule (i.e. to the rule $\mathfrak{w}$ such that $(i, \mathfrak{w}) \in P$). It is clear that by P' and Q' is determined a phrase marker in G' of the required expression $x' = T(x)$.

(C) *Synthesis*

1. We begin with $x'_n = \tau(x_n)$, where x_n is the last string in our analysis, we define $r(x'_n) = n$ and we make the 1st step.

2. In the ith step we deal with x'_{n-i+1} and we take the rule $\mathfrak{w}'$ such that $(i, \mathfrak{w}') \in P'$. We concern $c_0 \in V'_N$ such that $x'_{n-i+1} = vc_0w$ for some strings v and w and $r(c_0) = i$. We construct $x'_{n-1} = vd_0c_1d_1 \dots c_kd_kw$, where $c_0 ::= d_0c_1d_1 \dots c_kd_k$ is the standard

form of $\mathfrak{w}'$ and we define $r(c_j) = h$ in x'_{n-1} if $(h, j, i) \in Q'$. The rangs of other metalinguistic variables remain unchanged again.

3. We stop with x'_0 and put $x' = x'_0$.

(Received June 25th, 1964.)

VÝTAH

Sémantika a překlad gramatik a jazyků podobných ALGOLu

KAREL ČULÍK

Běžným matematickým způsobem je zaveden pojem překladu z jednoho jazyka do druhého jako zobrazení, které výrazům prvního jazyka přiřazuje takové výrazy druhého, které mají stejný význam jako výrazy výchozí. Toto zobrazení je dále popsáno jistým konstruktivním způsobem a lze je rozdělit do tří hlavních částí: analýza daného výrazu, jejímž výsledkem je příslušný frázový ukazatel; překlad symbolů a pravidel, jehož výsledkem je odpovídající frázový ukazatel v druhém jazyku a nakonec syntéza výsledného výrazu.

Dr. Karel Čulík, CSc., Matematický ústav ČSAV, Žitná 25, Praha 1.