Доклады Академии наук СССР
1971. Том 200, № 2

УДК 517.11 *МАТЕМАТИКА*

В. А. ИЛЮШКИН

О МНОГОЗНАЧНОСТИ ГРАММАТИЧЕСКОГО АНАЛИЗА КОНТЕКСТНО-СВОБОДНЫХ ЯЗЫКОВ

(Представлено академиком П. С. Новиковым 11 III 1971)

Как показал Парих в [1], существует такой контекстно-свободный язык, что грамматический анализ этого языка в любой описывающей его контекстно-свободной грамматике обязательно неоднозначен.

В дополнение к этому результату Н. Хомский в [2] заметил, что язык, построенный Парихом, может быть описан контекстно-свободной грамматикой, в которой каждое предложение данного языка имеет не больше двух различных структурных описаний, или C-маркеров. Тем самым было установлено, что степень неоднозначности грамматического анализа построенного Парихом языка равна 2. В связи с этим результатом Н. Хомский в [2] поставил проблему, состоящую в том, чтобы выяснить, существуют ли контекстно-свободные языки более высокой (быть может, неограниченной) степени неоднозначности грамматического анализа. Теоремы, приводимые ниже, дают положительное решение этой проблемы.

В данной заметке используются без разъяснений понятия контекстно-свободной грамматики и контекстно-свободного языка, необходимые сведения о которых можно найти в [2]. Без разъяснений используется также понятие вхождения, заимствованное из [3]. Термин «натуральное число» употребляется для обозначения целых положительных чисел.

Пусть заданы контекстно-свободная грамматика

$$G = \{V_N, S, V_T, R\}$$

и символ $*$, не входящий в ее алфавит. Положим $V = V_N \cup V_T$.

Определение 1. Пусть P, Q — слова в V, $A \to \omega$ — правило G, j — натуральное число. Пусть существуют слова S, T в V такие, что

$$P = SAT, \quad Q = S\omega T, \quad [S^{\partial} + 1 = j.$$

Тогда пара $(A \to \omega, j)$ называется *непосредственным выводом* Q из P в G.

Пусть далее слова S_1, S_2, S_3, T_1, T_2, T_3 в V таковы, что $S_1S_2S_3 = S$, $T_1T_2T_3 = T$.

Тогда вхождение $S_1 * S_2 * S_3\omega T$ называется *непосредственным потомком вхождения* $S_1 * S_2 * S_3AT$ *относительно* $(A \to \omega, j)$, вхождение $S_1S_2 * S_3\omega T_1 * T_2T_3$ называется *непосредственным потомком вхождения* $S_1S_2 * S_3AT_1 * T_2T_3$ *относительно* $(A \to \omega, j)$, вхождение $S\omega T_1 * T_2 * T_3$ называется *непосредственным потомком вхождения* $SAT_1 * T_2 * T_3$ *относительно* $(A \to \omega, j)$.

Определение 2. Пусть P, Q — слова в V, S — последовательность слов S_0, $S_1, \ldots, S_n$ в V, D — последовательность $d_1, \ldots, d_n$ такие, что $S_0 = P$, $S_n = Q$, и при любом i $(1 \leqslant i \leqslant n)$ d_i есть непосредственный вывод S_i из S_{i-1} в G.

Тогда D называется *выводом* Q *из* P в G, S называется *следом* D.

Пусть далее α_0, $\alpha_1, \ldots, \alpha_n$ — вхождения в S_0, $S_1, \ldots, S_n$ соответственно такие, что при любом i $(1 \leqslant i \leqslant n)$ α_i есть непосредственный потомок α_{i-1} относительно d_i.

Тогда α_n называется потомком α_0 относительно D.

Определение 3. Пусть Q — слово в V_T, β — вхождение в Q, $D = d_1, \ldots, d_n$ — вывод Q из $\hat{S}$ в G, $S = S_0, S_1, \ldots, S_n$ — след D. Пусть существует вхождение α символа A $(A \in V_N)$ в S_m $(0 \leqslant m < n)$ такое, что β есть потомок α относительно $d_{m+1}, \ldots, d_n$.

Тогда β называется составляющей Q типа A (или просто составляющей Q) относительно D.

Приведенные определения есть несколько измененные и уточненные (на основе точного понятия вхождения) варианты соответствующих формулировок ([1]).

Определение 4. Пусть Q — слово в V_T, D_1 и D_2 — выводы Q из $\hat{S}$ в G. Если при этом любая составляющая α слова Q относительно D_1 есть составляющая Q относительно D_2, и обратно, то D_1 и D_2 называются слабо эквивалентными. Если же любая составляющая α слова Q относительно D_1 есть составляющая Q того же типа относительно D_2, и обратно, то D_1 и D_2 называются эквивалентными.

Это определение было подсказано работами ([1]) и ([2]), где можно найти соображения в пользу указанных вариантов определения эквивалентности (по отношению к грамматическому анализу) выводов.

Теорема 1. *Для любого натурального числа k осуществим контекстно-свободный язык L_k такой, что*

1) *существует контекстно-свободная грамматика G языка L_k, в которой каждое предложение данного языка имеет не больше 2^k попарно неэквивалентных выводов из начального символа G;*

2) *для любой контекстно-свободной грамматики G языка L_k осуществимы предложение W данного языка и 2^k выводов W в G из начального символа G, никакие два из которых не являются слабо эквивалентными.*

Для любого k язык L_k, упоминаемый в теореме, есть множество всех слов Q в алфавите $\{a, b\}$, для которых существуют натуральные числа m_i, n_i, p_i, q_i, где $i = 1, \ldots, k$, удовлетворяющие условиям

$$Q = a^{m_1}b^{n_1}a^{p_1}b^{q_1} \ldots a^{m_k}b^{n_k}a^{p_k}b^{q_k},$$

$$m_i = p_i \quad \text{или} \quad n_i = q_i, \quad \text{где} \quad i = 1, \ldots, k.$$

Теорема 2. *Существует такой контекстно-свободный язык L, что для любой описывающей его контекстно-свободной грамматики G и любого натурального числа k осуществимы предложение W данного языка и k выводов W в G из начального символа G, никакие два из которых не являются слабо эквивалентными.*

Язык L, упоминаемый в теореме, есть объединение всех языков L_k, т. е. $L = \bigcup\limits_k L_k$.

Поскольку эквивалентные выводы являются в то же время слабо эквивалентными, из теорем 1, 2 можно вывести следствия, формулировки которых получаются соответственно из формулировок упомянутых теорем заменой слов «слабо эквивалентными» на «эквивалентными».

В заключение автор выражает благодарность А. А. Маркову за внимание к данной работе и полезные советы.

Московский физико-технический институт
г. Долгопрудный Моск. обл.

Поступило
1 II 1971

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] R. Parikh, J. Assoc. Comp. Mach., 13, № 4, 570 (1966). [2] Н. Хомский, Кибернетич. сборн., 2, 170, 196 (1966). [3] А. А. Марков, Тр. Матем. инст. им. В. А. Стеклова АН СССР, 42, 26 (1954).