

В. А. ИЛЮШКИН

## О СЛОЖНОСТИ ГРАММАТИЧЕСКОГО ОПИСАНИЯ КОНТЕКСТНО-СВОБОДНЫХ ЯЗЫКОВ

(Представлено академиком П. С. Новиковым 27 X 1971)

1. В работах по математической лингвистике принято оценивать сложность языка  $L$  по сложности его распознавания на машине Тьюринга  $M$ . По числу ячеек ленты, используемых машиной  $M$ , оценивается так называемая емкостная сложность языка  $L$ , а по числу элементарных шагов, совершаемых машиной  $M$ , оценивается так называемая временная сложность языка  $L$ . И емкостная, и временная сложности языка  $L$  являются по этой оценке функциями натурального аргумента — длины входа. Точные формулировки соответствующих определений и относящиеся к ним результаты можно найти в <sup>(2)</sup> и <sup>(3)</sup>.

Сложность языка  $L$  можно оценивать также по сложности описывающей его грамматики  $G$ , например, по числу нетерминальных (вспомогательных) символов и числу правил этой грамматики. Ниже приводятся три результата о сложности контекстно-свободных языков, оцениваемой именно таким образом.

2. При дальнейшем изложении предполагаются известными понятия контекстно-свободной грамматики и контекстно-свободного языка, необходимые сведения о которых можно найти в <sup>(4)</sup>. Предполагаются известными также понятия алгорифма (нормального алгорифма) и массовой (нормальной массовой) проблемы, подробно разъясняемые в <sup>(5)</sup>. Термин «натуральное число» употребляется для обозначения целых положительных чисел.

3. Введем обозначения  $n(G)$  и  $r(G)$  соответственно для числа нетерминальных символов и числа правил контекстно-свободной (поскольку другие не рассматриваются) грамматики  $G$ .

**Теорема 1.** Для любой пары натуральных чисел  $k, l$  таких, что  $l \geq 2k - 1$ , существует контекстно-свободный язык  $L_{k,l}$  в алфавите  $\{a, b\}$  такой, что

1) существует контекстно-свободная грамматика  $G$  языка  $L_{k,l}$  такая, что  $n(G) = k, r(G) = l$ ;

2) какова бы ни была контекстно-свободная грамматика  $G$  языка  $L_{k,l}$ ,  $n(G) \geq k, r(G) \geq l$ .

Опишем построение языка  $L_{k,l}$ . Предварительно введем некоторые определения и обозначения. Определим последовательность натуральных чисел  $m_1, m_2, \dots$  следующим образом ( $i$  — натуральное число):  $m_1 = 1$ ,  $m_{i+1} = (m_i)^i + 1$ . Для любого натурального числа  $i$  положим  $M_i = \{a^{m_1}, \dots, a^{m_i}\}$ .  $M_0$  будем использовать для обозначения пустого множества. На словах алфавита  $\{a, b\}$  определим функцию  $s$  следующим образом ( $x$  — слово в  $\{a, b\}$ ):

$$\begin{aligned} s(a) &= s(b) = 1, \quad s(xaa) = s(xa), \quad s(xbb) = s(xb), \\ s(xab) &= s(xa) + 1, \quad s(xba) = s(xb) + 1. \end{aligned}$$

Наконец, условимся посредством  $\tilde{x}$  обозначать обращение слова  $x$ . Пусть теперь  $k, l$  — натуральные числа такие, что  $l \geq 2k - 1$ . Положим  $m =$

$= l - 2k + 1$  и отметим, что  $m \geq 0$ . Если  $k = 1$ , то полагаем  $L_{k,l} = M_l$ . Если  $k > 1$  и  $k$  четно, то полагаем  $L_{k,l} = \{y \mid y \in M_m \text{ или существует начинающееся с символа } a \text{ слово } x \text{ в } \{a, b\} \text{ такое, что } s(x) = k - 1 \text{ и } y = b\tilde{x}bxb\}\}$ . Если  $k > 1$  и  $k$  нечетно, то полагаем  $L_{k,l} = \{y \mid y \in M_m \text{ или существует начинающееся с символа } a \text{ слово } x \text{ в } \{a, b\} \text{ такое, что } s(x) = k - 1 \text{ и } y = a\tilde{x}bxa\}\}$ .

Отметим, что ограничение  $l \geq 2k - 1$ , налагаемое при формулировке теоремы 1 на  $k$  и  $l$ , не может быть ни снято, ни ослаблено. Так, если заменить в формулировке теоремы 1 неравенство  $l \geq 2k - 1$  на неравенство  $l \geq 2k - 2$ , то полученное утверждение будет неверным.

В качестве следствия из теоремы 1 отметим также, что при  $k = 1$  множество  $L_{k,l}$  оказывается довольно простым — состоящим из  $l$  слов однобуквенного алфавита.

Теорема 1 есть своеобразное объединение двух результатов Грушки <sup>(1)</sup>, которые можно сформулировать следующим образом:

*Для любого натурального числа  $k$  существует контекстно-свободный язык в  $\{a, b\}$ , описываемый контекстно-свободной грамматикой с  $k$  нетерминальными символами ( $k$  правилами) и не описываемый никакой контекстно-свободной грамматикой с меньшим числом нетерминальных символов (правил).*

**Теорема 2.** *Невозможен алгоритм, выясняющий для любого контекстно-свободного языка в алфавите  $\{a, b\}$  возможность его описания контекстно-свободной грамматикой  $G$  такой, что  $n(G) = 1, r(G) = 3$ .*

Формулировку теоремы 2 можно сделать более точной, если договориться о стандартном способе кодирования контекстно-свободных грамматик, терминальный алфавит которых есть  $\{a, b\}$ , словами некоторого фиксированного алфавита, например, некоторого фиксированного расширения И алфавита  $\{a, b\}$ . Если полагать такой способ заданным, то теорему 2 можно формулировать следующим образом.

*Невозможен нормальный алгоритм над алфавитом И, применимый к любому слову в этом алфавите, являющемуся кодом некоторой контекстно-свободной грамматики  $G$ , и перерабатывающий это слово в пустое в том и только в том случае, если существует эквивалентная  $G$  контекстно-свободная грамматика  $G'$  такая, что  $n(G') = 1, r(G') = 3$ .*

**Теорема 3.** *Для любого натурального числа  $k$  существуют однозначный линейный контекстно-свободный язык  $L_k$  в алфавите  $\{a, b\}$  и описывающая его линейная контекстно-свободная грамматика  $G_k$  такие, что, какова бы ни была однозначная контекстно-свободная грамматика  $G$  языка  $L_k$ ,*

$$n(G) \geq n(G_k) + k, \quad r(G) \geq r(G_k) + k.$$

Упоминаемый в теореме 3 язык  $L_k$  есть множество всех слов  $Q$  в алфавите  $\{a, b\}$ , для которых существуют натуральные числа  $m_i, n_i, i = 1, \dots, k$ , удовлетворяющие условиям

$$Q = \begin{cases} a^{m_1}b^{m_2}a^{m_3} \dots b^{m_k}ab^{n_1}a^{n_2}b^{n_3} \dots a^{n_k}, & \text{если } k \text{ четно,} \\ a^{m_1}b^{m_2}a^{m_3} \dots a^{m_k}ba^{n_1}b^{n_2}a^{n_3} \dots a^{n_k}, & \text{если } k \text{ нечетно,} \end{cases}$$

$$1 \leq \frac{m_i}{n_{k+1-i}} \leq 2, \quad i = 1, \dots, k.$$

Автор выражает глубокую признательность А. А. Маркову за внимание к данной работе.

Московский  
энергетический институт

Поступило  
25 VII 1971

#### ЦИТИРОВАННАЯ ЛИТЕРАТУРА

<sup>1</sup> Gruska, Information and Control, 14, 164 (1969). <sup>2</sup> Дж. Хартманис, Р. Е. Стирнз, Кибернетич. сборн., 6, 1969, стр. 165. <sup>3</sup> А. В. Ахоу, Дж. Д. Улман, Киберн. сборн., 6, 1969, стр. 165. <sup>4</sup> Н. Хомский, Кибернетич. сборн., 2, 1966, стр. 169. <sup>5</sup> А. А. Марков, Тр. Матем. инст. им. В. А. Стеклова АН СССР, 42 (1954).