**Доклады Академии наук СССР**
**1974. Том 216, № 5**

УДК 51.801

*МАТЕМАТИКА*

Э. Д. СТОЦКИЙ

# ОБ АЛГОРИТМИЧЕСКИХ СВОЙСТВАХ УСЛОВНЫХ ГРАММАТИК С РАССЕЯННЫМ КОНТЕКСТОМ

*(Представлено академиком П. С. Новиковым 3 XII 1973)*

В заметке (1) введено понятие условной грамматики с рассеянным контекстом, которое является естественным обобщением понятия условной грамматики (2). Там же показано, что грамматики типа $\mathscr{E}+\mathscr{E}\mathscr{L}+\widetilde{\mathscr{N}}$ и типа $\mathscr{E}+\mathscr{E}\mathscr{R}+\widetilde{\mathscr{N}}$ порождают классы языков, строго содержащие класс бесконтекстных языков и в то же время существенно более узкие, чем класс языков типа 1 по Н. Хомскому. Основной целью данной заметки является доказательство алгоритмической разрешимости проблем непустоты и конечности языка, порождаемого условной грамматикой типа $\mathscr{E}+\mathscr{E}\mathscr{L}+\widetilde{\mathscr{N}}$ или типа $\mathscr{E}+\mathscr{E}\mathscr{R}+\widetilde{\mathscr{N}}$.

Теорема 1. *В классах грамматик типа $\mathscr{E}+\mathscr{E}\mathscr{L}+\widetilde{\mathscr{N}}$ и типа $\mathscr{E}+\mathscr{E}\mathscr{R}+\widetilde{\mathscr{N}}$ алгоритмически разрешима проблема непустоты порождаемого языка.*

Доказательство использует конструкции, в которых обобщаются идеи М. В. Ломковской (3). Приводим основные рассуждения для грамматики $\Gamma=(V, W, S, R)$ типа $\mathscr{E}+\mathscr{E}\mathscr{L}+\widetilde{\mathscr{N}}$.

Будем говорить, что слово $\alpha$ терминализуемо в грамматике $\Gamma$, если найдется такое слово $\beta$ в терминальном алфавите $V$, $\beta \mho V$, что $\beta$ выводимо * из $\alpha$ по правилам грамматики $\Gamma$, $\alpha \vDash \beta$. В частности, если $\alpha \mho V$, то $\beta \eqcirc \alpha$. Через $\mathrm{Pr}(\alpha)$ обозначается слово, полученное из слова $\alpha$ вычеркиванием всех вхождений букв нетерминального алфавита $W$. Длина слова $\alpha$ обозначается $l(\alpha)$.

Лемма 1. *Пусть дана грамматика $\Gamma=(V, W, S, R)$ типа $\mathscr{E}+\mathscr{E}\mathscr{L}+\widetilde{\mathscr{N}}$ и даны слова $\alpha$ и $\beta$ такие, что $\mathrm{Pr}(\alpha) \succcurlyeq_{LS} \mathrm{Pr}(\beta)$. Тогда, если $\alpha$ терминализуемо в грамматике $\Gamma$, то и $\beta$ также терминализуемо.*

Докажем лемму в предположениях, что $\alpha \mho W$, $\beta \mho W$ и $l(\beta)=1+l(\alpha)$. Общий случай рассматривается сходным образом.

Итак, пусть $\alpha \eqcirc \xi_1 B \xi_2 \xi_3$ и $\beta \eqcirc \xi_1 B \xi_2 B \xi_3$. По условию имеется вывод $\alpha \eqcirc \alpha_0 \vdash \alpha_1 \vdash \ldots \vdash \alpha_s$, где $\alpha_s \mho V$. Представим слова $\alpha_i$ в виде $\alpha_i \eqcirc \xi_1^{(i)} B^{(i)} \xi_2^{(i)} \xi_3^{(i)}$, где индекс $i$ у слова $\chi^{(i)}$ означает, что рассматривается потомок слова $\chi$ в $i$-м слове вывода. Построим последовательность слов $\beta \eqcirc \beta_0 \vDash \beta_1 \vDash \ldots \vDash \beta_s$ такую, что $\beta_s \mho V$.

1) Если на шаге вывода $\alpha_0 \vdash \alpha_1$ происходит применение правила $r$ к подслову $\xi_1$, $\xi_2$ или $\xi_3$ слова $\alpha_0$, то выполняем подстановку того же правила $r$ к соответствующему подслову $\xi_1$, $\xi_2$ или $\xi_3$ слова $\beta_0$. Результат подстановки и будет $\beta_1$. Если же на шаге $\alpha_0 \vdash \alpha_1$ заменяется вхождение $\xi_1 * B * \xi_2 \xi_3$, то применим сначала правило $r$ к вхождению $\xi_1 B \xi_2 * B * \xi_3$ и получим слово $\xi_1 B \xi_2 B^{(1)} \xi_3$, а затем применим правило $r$ к вхождению $\xi_1 * B * \xi_2 B^{(1)} \xi_3$ и получим слово $\xi_1^{(1)} B^{(1)} \xi_2^{(1)} B^{(1)} \xi_3^{(1)}$, которое и примем в качестве $\beta_1$. По определению грамматики типа $\mathscr{E}+\mathscr{E}\mathscr{L}+\widetilde{\mathscr{N}}$ указанные подстановки выполнимы.

---

* Отношение выводимости в один шаг обозначается $\vdash$, а его транзитивное замыкание обозначается $\vDash$.

2) Допустим, что уже построена последовательность $\beta_0 \vDash \beta_1 \vDash \ldots \vDash \beta_i$ и на шаге вывода $\alpha_i \vdash \alpha_{i+1}$ правило $r$ применяется к подслову $\xi_1^{(i)}$, $\xi_2^{(i)}$ или $\xi_3^{(i)}$ слова $\alpha_i$. В этом случае выполним подстановку того же правила $r$ к соответствующему подслову $\xi_1^{(i)}$ $\xi_2^{(i)}$ или $\xi_3^{(i)}$ слова $\beta_i$. Результат подстановки и есть $\beta_{i+1}$. Если же на шаге $\alpha_i \vdash \alpha_{i+1}$ правило $r$ применяется к некоторому вхождению $\xi_1^{(i)}\varphi_1 * A * \varphi_2\xi_2^{(i)}\xi_3^{(i)}$, $\varphi_1 A \varphi_2 \overline{\overline{\cdot}} B^{(i)}$, то применим сначала правило $r$ к вхождению $\xi_1^{(i)} B^{(i)} \xi_2^{(i)} \varphi_1 * A * \varphi_2 \xi_3^{(i)}$ и получим слово $\xi_1^{(i)} B^{(i)} \xi_2^{(i)} B^{(i+1)} \xi_3^{i)}$, а затем применим то же правило к вхождению $\xi_1^{(i)} \varphi_1 * A * \varphi_2 \xi_2^{(i)} B^{(i+1)} \xi_3^{(i)}$ и полученное слово примем в качестве $\beta_{i+1}$. Легко видеть, что условие применимости правила $r$ при таком образе действия не нарушается. Лемма доказана.

Следуя ([5]), определим для грамматики $\Gamma = (V, W, S, R)$ несколько словарных операций над языками:

а) нахождение множества «сплошного вверх»

$$\nu(Q) = \{w \mid w \mho (V \cup W) \& \exists v \in Q (v \succcurlyeq_{LS} w)\};$$

б) операция «шаг назад согласно грамматике Г»

$$\gamma(Q) = Q \cup \{w \mid w \mho (V \cup W) \& \exists v \in Q (w \vdash v)\};$$

в) нахождение списка минимальных слов языка $Q$

$$\mu(Q) = \{w \mid w \in Q \& \forall v (v \succcurlyeq_{LS} w \& v \neq w \supset v \notin Q)\}.$$

Лемма 2. *Операции $\nu$ и $\gamma$ переводят автоматный язык в автоматный, причем $\nu(Q)$ и $\gamma(Q)$ могут быть найдены эффективно, коль скоро язык $Q$ задан автоматной грамматикой.*

Доказательство леммы ввиду его простоты опускаем.

Лемма 3. *Если язык $Q$ задан автоматной грамматикой, то список $\mu(Q)$ можно указать эффективно.*

Доказательство леммы. Упорядочим все слова в алфавите $V \cup W$ по возрастанию их длин: $x_1, x_2, \ldots$ Если $x_1 \in Q$, то поместим его в предварительный список, который обозначим $F_1$. Если $x_1 \notin Q$, то в качестве $F_1$ берем пустой список. Далее действуем по индукции. Если список $F_{i-1}$ построен и $x_i \in Q$, причем не существует $y \in F_{i-1}$ такого, что $y \succcurlyeq_{LS} x_i$, то добавляем $x_i$ к $F_{i-1}$ и обозначаем полученный список через $F_i$. Если $x_i \notin Q$ или же найдется $y \in F_{i-1}$ такое, что $y \succcurlyeq_{LS} x_i$, то полагаем $F_i = F_{i-1}$. Затем проверяем, имеет ли место вложение автоматных языков $Q \subseteq \nu(F_i)$. В случае утвердительного ответа процедура заканчивается. Ввиду того, что отношение $\succcurlyeq_{LS}$ является кёниговым ([1]), указанный процесс должен завершиться.

Вернемся к доказательству теоремы. Строим последовательность $L_0$, $L_1, \ldots$ автоматных языков в алфавите $V \cup W$. В качестве языка $L_0$ берем множество всех слов в алфавите $V$. Остальные языки задаем соотношением $L_{i+1} = \nu(\mu(\gamma(L_i)))$. Ясно, что $L_i \subseteq L_{i+1}$. На каждом переходе от $i$ к $i+1$ выполняем проверку $L_i = L_{i+1}$. Используя леммы, можно показать, что $\exists_j (L_j = L_{j+1})$. Указанное $L_j$ представляет собой множество всех терминализуемых в грамматике Г слов. Остается выполнить проверку $S \in L_j$. Теорема доказана.

Замечание 1. В доказательстве теоремы не используется тот факт, что подстановки $A \to \omega$ имеют $\omega \neq \Lambda$. Отсюда следует, что теорема верна и для условных грамматик, содержащих подстановки вида $A \to \Lambda$.

З а м е ч а н и е 2. Ввиду замкнутости классов $\mathfrak{L}(\mathscr{E}+\mathscr{E}\mathscr{L})$ и $\mathfrak{L}(\mathscr{E}+\mathscr{E}\mathscr{R})$ относительно операций пересечения с автоматными языками, из теоремы 1 следует, что имеют место строгие включения $\mathfrak{L}(\mathscr{E}+\mathscr{E}\mathscr{L})\subset\mathfrak{L}(SCG)$ и $\mathfrak{L}(\mathscr{E}+\mathscr{E}\mathscr{R})\subset\mathfrak{L}(SCG)$, где $\mathfrak{L}(SCG)$ — класс языков, введенный в (4).

Т е о р е м а 2. *В классах грамматик типа $\mathscr{E}+\mathscr{E}\mathscr{L}+\widetilde{\mathscr{N}}$ и типа $\mathscr{E}+\mathscr{E}\mathscr{R}+\widetilde{\mathscr{N}}$ алгоритмически разрешима проблема конечности порождаемого языка.*

П р и в о д и м н а б р о с о к д о к а з а т е л ь с т в а для случая $\mathscr{E}+\mathscr{E}\mathscr{L}+\widetilde{\mathscr{N}}$. Не умаляя общности, полагаем, что в заданной грамматике $\Gamma=(V, W, S, R)$ нет правил с подстановками вида $A\to A$. По теореме 1 можно предположить, что $L(\Gamma)\neq\varnothing$.

Строим усеченное дерево выводимости $\Delta_y(\Gamma)$ грамматики Г. Вершины дерева помечаем словами в алфавите $W$.

Корень дерева $w^{(0)}$ пометим словом $m^{(0)}\rightleftharpoons S$ и назовем его вершиной нулевого яруса.

Рассмотрим всевозможные слова $\{\alpha\}$, удовлетворяющие условию $S\vdash\alpha$. По теореме 1 можно считать, что из списка $\{\alpha\}$ удалены нетерминализуемые слова. Пусть операция $\mathrm{Pr}(\alpha)$, примененная к словам из $\{\alpha\}$, дает слова $m_1^{(1)}, \ldots, m_k^{(1)}$, отличные одно от другого. Добавим теперь к построенной части дерева (к корню) вершины $w_1^{(1)}, \ldots, w_k^{(1)}$ и пометим их словами $m_1^{(1)}, \ldots, m_k^{(1)}$ соответственно. Присоединим эти вершины ребрами к $w^{(0)}$, ориентируя каждое ребро «от» $w^{(0)}$. Если какое-нибудь слово $m_i^{(1)}$ удовлетворяет условию $S\succcurlyeq_{LS}m_i^{(1)}$, то пометим соответствующую вершину $w_i^{(1)}$ дополнительной пометкой «тупик».

Дальнейшее построение основано на индукции. Пусть построена вершина $w^{(n)}$ $n$-го яруса с пометкой $m^{(n)}$, причем $m^{(n)}\neq\Lambda$ и $w^{(n)}$ не имеет дополнительной пометки «тупик». Рассмотрим все такие терминализуемые слова $\alpha$, что $m^{(n)}\vdash\alpha$. Применение операции $\mathrm{Pr}(\alpha)$ к этим словам дает конечный список $m_1^{(n+1)}, \ldots, m_s^{(n+1)}$ попарно различных слов. Добавим к построенной части дерева вершины $w_1^{(n+1)}, \ldots, w_s^{(n+1)}$ и пометим их словами $m_1^{(n+1)}, \ldots, m_s^{(n+1)}$. Присоединим эти вершины к вершине $w^{(n)}$ ребрами, ориентированными «от» $w^{(n)}$. Если какое-нибудь слово $m_i^{(n+1)}$ удовлетворяет условию $m\succcurlyeq_{LS}m_i^{(n+1)}$, где пометка $m$ принадлежит пути, ведущему из $w^{(0)}$ в $w^{(n)}$, то помечаем $w_i^{(n+1)}$ дополнительной пометкой «тупик».

Поскольку отношение $\succcurlyeq_{LS}$ является кёниговым, процедура построения дерева $\Delta_y(\Gamma)$ закончится через конечное число шагов.

Л е м м а 4. *Язык, порождаемый грамматикой* Г *типа $\mathscr{E}+\mathscr{E}\mathscr{L}+\widetilde{\mathscr{N}}$, бесконечен тогда и только тогда, когда существуют такие терминализуемые слова $\alpha$ и $\beta$, что $S\models\alpha\models\beta$, $l(\alpha)<l(\beta)$ и $\mathrm{Pr}(\alpha)\succcurlyeq_{LS}\mathrm{Pr}(\beta)$.*

Д о к а з а т е л ь с т в о леммы совпадает с завершающим этапом доказательства теоремы 2. Мы опускаем доказательство необходимости условий леммы.

Заметим теперь, что в случае бесконечности языка $L(\Gamma)$ слова $\alpha$ и $\beta$, указанные в лемме 4, могут быть эффективно найдены при помощи дерева $\Delta_y(\Gamma)$. При этом $\mathrm{Pr}(\beta)$ является пометкой в одной из тупиковых вершин $w$ дерева $\Delta_y(\Gamma)$, а $\mathrm{Pr}(\alpha)$ является пометкой на пути, ведущем из $w^{(0)}$ в $w$. Если язык $L(\Gamma)$ конечен, то указанных слов $\mathrm{Pr}(\alpha)$ и $\mathrm{Pr}(\beta)$ в дереве $\Delta_y(\Gamma)$ нет.

Итак, пусть даны слова $\alpha$ и $\beta$, удовлетворяющие условиям леммы 4. Обозначим $\mathrm{Pr}(\alpha)$ через $m_1$ и $\mathrm{Pr}(\beta)$ через $m_2$. Коль скоро $m_1\rightleftharpoons m_2$, то бесконечность языка $L(\Gamma)$ очевидна. Пусть же $l(m_1)>l(m_2)$. Представим $m_1$

в виде $m_1 \doteq A_1A_2\ldots A_n$, где $A_i \in W$. Тогда $m_2$ имеет вид $m_2 \doteq A_1\xi_1A_2\xi_2\ldots$ $\ldots A_n\xi_n$, где $\xi_i \in W$ и слово $\xi_i$ не содержит вхождений буквы $A_{i+1}$, $\xi_1\ldots$ $\ldots\xi_n \neq \Lambda$. Назовем это представление левым вложением $m_1$ в $m_2$.

Рассмотрим вывод $\alpha \doteq \alpha_1 \vdash \alpha_2 \vdash \ldots \vdash \alpha_t \doteq \beta$. Применим этот вывод к левому вложению $m_1$ в $m_2$ следующим образом: все буквы, входящие в $\xi_1, \ldots, \xi_n$, будем преобразовывать по лемме 1 тем же способом, которым преобразуются самые левые вхождения одноименных букв в $m_1$. В результате получим слово $m_3'$ вида $m_3' \doteq A_1'\xi_1'A_2'\xi_2'\ldots A_n'\xi_n'$, где через $\chi'$ обозначено слово, полученное из слова $\chi$ согласно указанной процедуре. Заметим, что $A_1'\ldots A_n' \doteq \alpha_t$, $\mathrm{Pr}(A_1'\ldots A_n') \doteq m_2$ и каждое слово $\xi_i'$ может быть представлено в виде $\xi_i' \doteq B_1B_2\ldots B_{k_i}$, где $B_j \in \{A_1', \ldots, A_i'\}$, $j=1, 2, \ldots, k_i$. Следовательно, слово $m_3 \doteq \mathrm{Pr}(m_3')$ удовлетворяет условию $m_2 \succcurlyeq_{LS} m_3$. Если $l(m_3) = l(m_2)$, то приходим к рассмотренному случаю. Пусть $l(m_3) > l(m_2)$. Далее, повторяя описанный процесс для последовательных пар $(m_i, m_{i+1})$, либо получим слова $m_j$ и $m_{j+1}$ такие, что $l(m_j) = l(m_{j+1})$, либо получим бесконечную последовательность терминализуемых слов $m_1, m_2, m_3, \ldots$ такую, что $m_i \succcurlyeq_{LS} m_{i+1}$ и $l(m_{i+1}) > l(m_i)$. На этом доказательство теоремы 2 закончено.

Замечание. Проблема разрешимости конечности для грамматик типа $\mathcal{N}$ остается открытой.

Теорема 3. *Для всякой грамматики $\Gamma = (V, W, S, R)$ типа $\mathcal{E}+\mathcal{EL}$ или типа $\mathcal{E}+\mathcal{ER}$, порождающей бесконечный язык $L(\Gamma)$, можно указать такое натуральное число $m$ и такие слова $\alpha_0, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m$, что имеет место*:

1) $\beta_1\ldots\beta_m \neq \Lambda$;

2) *все слова вида $\alpha_0\beta_1{}^n\alpha_1\beta_2{}^n\ldots\beta_m{}^n\alpha_m$, $n=1, 2, \ldots$, принадлежат языку $L(\Gamma)$.*

Доказательство теоремы существенно опирается на возможность эффективно указать слова $\alpha$ и $\beta$, удовлетворяющие условиям леммы 4.

Следствие. *Язык $L = \{a^nb^nc^n \mid n \geqslant 1\}$ не принадлежит классам $\mathfrak{L}(\mathcal{EL})$ и $\mathfrak{L}(\mathcal{ER})$.*

Указанное следствие представляет интерес ввиду того, что язык $L$ принадлежит классу $\mathfrak{L}(E)$.

Всесоюзный институт научной и
технической информации
Москва

Поступило
24 XI 1973

## ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] *Э. Д. Стоцкий*, ДАН, т. 207, № 4, 796 (1972). [2] *М. В. Ломковская*, ДАН, т. 207, № 4, 781 (1972). [3] *М. В. Ломковская*, Научно-техн. информация, сер. 2, № 1, 16 (1972). [4] *S. Greibach, J. Hopcroft*, J. Comp. System Sci., № 3, 233 (1969).