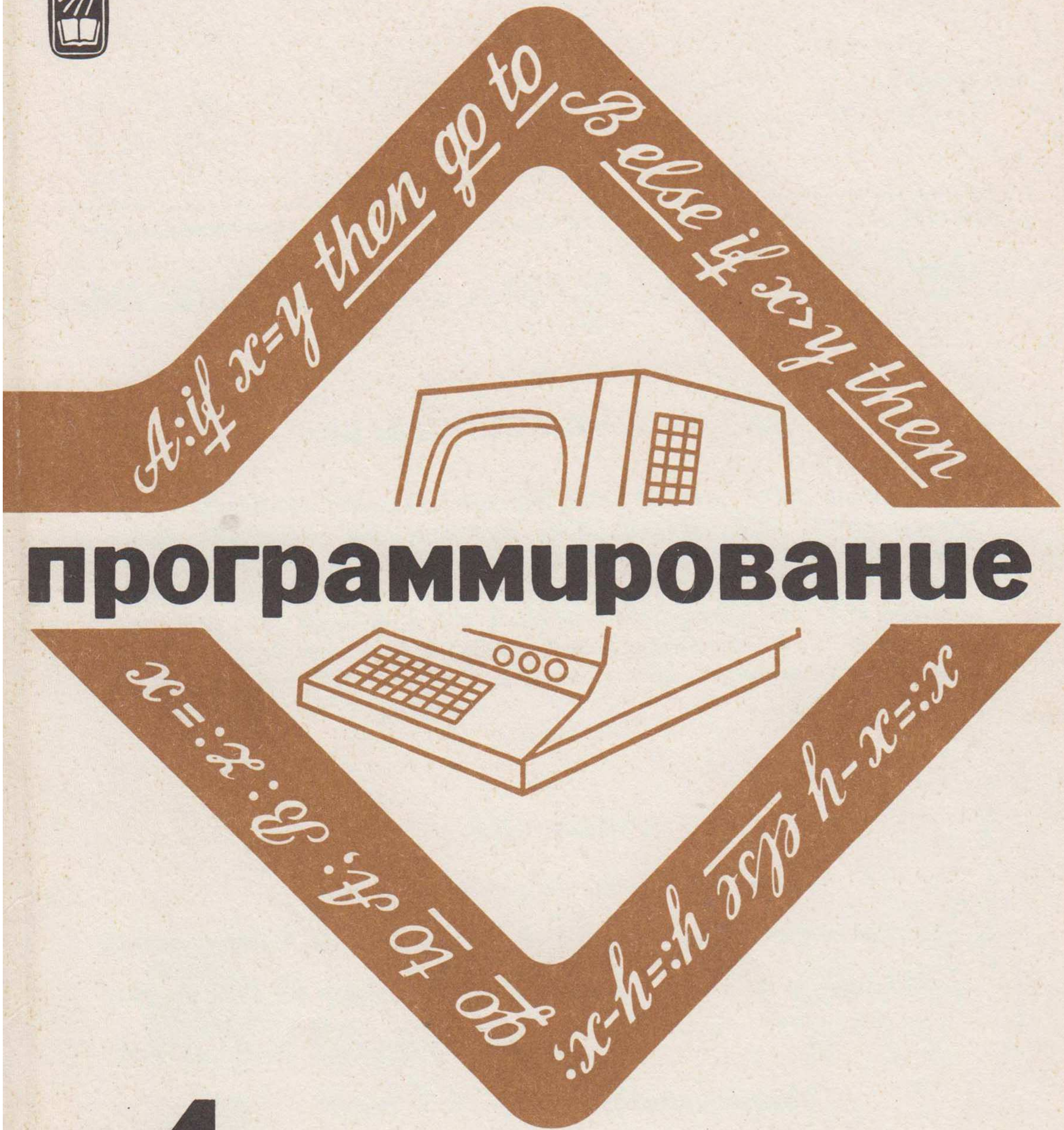
Российская академия наук

ISSN 0132-3474




# программирование

4

1993

УДК 519.685.3

© 1993 г. В. М. Курочкин

# КРИТЕРИЙ LR(1)-ГРАММАТИКИ[1]

**Даются необходимые и достаточные условия для $LR(1)$-грамматики, проверка которых по сложности и трудоемкости не превосходит проверки грамматики на свойство $LALR(1)$.**

## Введение

Наиболее логичным и широкоохватывающим подходом для быстрого однопросмотрового синтаксического анализа является $LR(k)$-анализ. К сожалению, целый ряд трудностей, как нам кажется, чисто технического характера привели к тому, что на практике $LR$-анализ применяется с различного рода упрощениями и ограничениями. В значительной степени этим обусловлено появление таких понятий, как $SLR$- и $LALR$-грамматики, и построение анализаторов, базирующихся на таких грамматиках (см. [1]). Используемый при этом метод построения анализатора неприменим, вообще говоря, к любой $LR(k)$-грамматике. Следует при этом отметить, что общепризнана практическая нецелесообразность применения анализа с заглядыванием на несколько символов вперед, т.е. всегда считают, что $k \leq 1$. В то же время, преобразование грамматики к виду $LR(0)$ связано, как правило, с существенной ее переработкой, приданием ей неестественного вида и рассогласованием семантических свойств грамматических правил, что отражается на последующем процессе трансляции. Поэтому и мы ограничимся случаем $k = 1$. Ниже в этой статье рассматривается лишь один вопрос: проверка грамматики на свойство $LR(1)$. Сформулирован и доказан очень простой критерий (необходимые и достаточные условия) того, что КС-грамматика является $LR(1)$-грамматикой.

## Основные определения и используемые обозначения и соглашения

Пусть $G(V,T,P,S)$ — КС-грамматика ($V$ — весь алфавит, $T$ — терминальный алфавит, $T \subset V, P$ — совокупность бесконтекстных правил вывода, $S$ — аксиома, $S \in V$). Будем предполагать дополнительно, что $G$ порождает непродолжаемый язык $L(G)$, т.е. если $x \in L(G)$ и $y \neq \varepsilon$ (пустое слово), то $xy \notin L(G)$. (Если это не так, то, не меняя основных свойств грамматики $G$ и языка $L(G)$, легко добиться непродолжаемости, например, рассмотрев вместо $L(G)$ язык $L(G)\$$, где новый символ $\$ \notin V$. В дальнейшем используются следующие обозначения:

$a, b$ — символы алфавита $T$;
$A, S$ — символы алфавита $V - T$;
$\alpha, \beta \ldots$ — символы алфавита $V$;
$u, v, w, x, y$ — слова из $T^*$;
$\varphi, \psi, \chi$ — слова из $V^*$;
$\varepsilon$ — пустое слово;
$A_p \rightarrow \psi_p$ или $A_p \rightarrow \alpha_{p1}\alpha_{p2}\ldots\alpha_{pl_p}$ — правило $p$ из $P$;

[1]Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект 93-12-573).

$\varphi \overset{*}{\Rightarrow} \psi$ — вывод из $\varphi$ слова $\psi$ за несколько ($\geq 0$) шагов;
$\varphi \underset{p}{\Rightarrow} \psi$ — вывод из $\varphi$ слова $\psi$ за один шаг путем применения правила $p$;
$C_\varphi$ — состояние анализа (или "ситуация") для слова $\varphi = \alpha_1 \ldots \alpha_n$:
$C_\varphi = \{(p, i, a) \mid$ существует вывод вида

$$S \overset{*}{\Rightarrow} \alpha_1 \ldots \alpha_{n-i} A_p a u \underset{p}{\Rightarrow} \alpha_1 \ldots \alpha_{n-i} \ldots \alpha_n \alpha_{pi+1} \ldots \alpha_{pl_p} a u, \text{ где } a \in T \text{ или } au = \varepsilon\}.$$

Все рассматриваемые выводы правосторонние (если не оговорено явно противное).

## Критерий LR(1)

Сформулируем и докажем теперь критерий $LR(1)$. Сначала повторим (см. [1]) определение $LR(1)$-грамматики.

Грамматика $G$ называется $LR(1)$-грамматикой, если из существования двух правосторонних выводов

(1) $$S \overset{*}{\Rightarrow} \psi A_p a u \underset{p}{\Rightarrow} \psi \psi_p a u = \varphi a u, a \in T \text{ или } au = \varepsilon,$$

(2) $$S \overset{*}{\Rightarrow} \varphi a v$$

следует, что второй вывод имеет вид

(3) $$S \overset{*}{\Rightarrow} \psi A_p a v \underset{p}{\Rightarrow} \psi \psi_p a v = \varphi a v.$$

Приводимые здесь определения $LR(1)$ и $C_\varphi$ очевидным образом получаются из тех, которые даются в [1] при $k = 1$. В [1] приводится и критерий $LR(k)$, однако даже для $k = 1$ его формулировка и проверка требуют выполнения довольно сложных действий. Приводимый здесь критерий во многих отношениях проще и, насколько нам известно, нигде до сих пор не встречался.

**Критерий LR(1).** Грамматика является $LR(1)$-грамматикой тогда и только тогда, когда все состояния анализа $C_\varphi$ удовлетворяют следующим двум условиям:

(A) в множестве $C_\varphi$ нет двух троек вида $(p, l_p, a)$ и $(q, l_q, a)$ для $p \neq q$;
(B) в множестве $C_\varphi$ нет двух троек вида $(p, l_p, a)$ и $(q, j, b)$ для $a = \alpha_{qj+1}$.

*Необходимость.* Если не выполняется (A), то, по определению $C_\varphi$, существуют выводы

$$S \overset{*}{\Rightarrow} \alpha_1 \ldots \alpha_{n-l_p} A_p a u \underset{p}{\Rightarrow} \alpha_1 \ldots \alpha_n a u = \varphi a u \text{ и}$$
$$S \overset{*}{\Rightarrow} \alpha_1 \ldots \alpha_{n-l_q} A_q a v \underset{q}{\Rightarrow} \alpha_1 \ldots \alpha_n a v = \varphi a v,$$

причем на последнем шаге применяются разные правила, ибо $p \neq q$, что противоречит определению $LR(1)$. Если не выполняется (B), то существуют выводы

$$S \overset{*}{\Rightarrow} \alpha_1 \ldots \alpha_{n-l_p} A_p a u \underset{p}{\Rightarrow} \alpha_1 \ldots \alpha_n a u = \varphi a u \text{ и}$$
$$S \overset{*}{\Rightarrow} \alpha_1 \ldots \alpha_{n-j} A_q b w \underset{q}{\Rightarrow} \alpha_1 \ldots \alpha_n \alpha_{qj+1} \ldots \alpha_{ql_q} b w = \varphi a \psi b w \overset{*}{\Rightarrow} \varphi a v,$$

и последний шаг во втором выводе состоит в применении какого-либо правила не в том месте, где в первом (независимо от того, разворачивается $\psi$ или нет), что опять противоречит $LR(1)$.

*Достаточность.* Пусть грамматика удовлетворяет (A) и (B) и не является $LR(1)$. Пусть вывод (2) имеет вид

$$S \overset{*}{\Rightarrow} \chi A_r y \underset{r}{\Rightarrow} \chi \psi_r y = \varphi a v = \psi \psi_p a v.$$

Будем считать, что правый конец подслова $\psi_r$ находится не левее правого конца подслова $\psi_p$ (иначе можно поменять местами выводы (1) и (2)). Если правые концы этих подслов совпадают, т.е. находятся перед символом $a$, то $C_\varphi$ содержит $(p, l_p, a)$ и $(r, l_r, a)$, а тогда $p = r$ и вывод (2) имеет вид (3). Если правый конец подслова $\psi_r$ расположен правее, то выделим в выводе (2) тот шаг, когда появляется символ $a$. Пусть это будет применение правила $q : A_q \to \psi_q$ и символ $a$ является $j+1$-м символом слова $\psi_q = \alpha_{q1} \dots \alpha_{ql_q}, \alpha_{qj+1} = a, j \geq 0$. Вывод (2) тогда имеет вид

(4) $$S \overset{*}{\Rightarrow} \chi' A_q b y \underset{r}{\Rightarrow} \chi' \alpha_{q1} \dots \alpha_{qj} a \alpha_{qj+2} \dots \alpha_{ql_q} b y \overset{*}{\Rightarrow} \varphi a v.$$

Так как последним в этом выводе применяется правило $r$, вставляющее слово $\psi_r$ не левее символа $a$, то либо правило $r$ совпадает с $q$ (и является последним в выводе (4)), либо оно применяется после $q$ в последней части вывода (4), которая, следовательно, состоит в развертывании подслова $\alpha_{qj+2} \dots \alpha_{ql_q}$. В любом случае начало слова $\chi' \alpha_{q1} \dots \alpha_{qj}$ до символа $a$ не меняется и поэтому равно $\varphi$. Таким образом начало вывода (4) имеет вид

$$S \overset{*}{\Rightarrow} \chi' A_q b y \underset{q}{\Rightarrow} \chi' \alpha_{q1} \dots \alpha_{qj} a \chi'' b y = \varphi a \chi'' b y$$

и поэтому $C_\varphi$ содержит $(p, l_p, a)$ и $(q, j, b)$, и $\alpha_{qj+1} = a$, что противоречит (B). Достаточность также доказана.

Легко видеть, что коль скоро для грамматики $G$ построены множества (состояния анализа) $C_\varphi$, дальнейшая работа по проверке приведенного критерия заключается лишь в просмотре множеств троек $C_\varphi$, сопоставлении их с правилами грамматики и не требует никаких более сложных (например, поиска тех или иных выводов) действий.

## Список литературы

1. *Ахо А., Ульман Дж.* Теория синтаксического анализа, перевода и компиляции. М. Мир, 1978.

Поступила в редакцию 28.04.93

**V. M. Kurochkin**

**A CRITERION FOR LR(1)-GRAMMAR**

Necessary and sufficient conditions for a grammar to be $LR(1)$ are given. The complexity of checking them does not exceed the complexity of checking for the $LALR(1)$-property.

**Курочкин Владимир Михайлович**, окончил Московский государственный университет им. М.В.Ломоносова в 1946 г., кандидат физ.-мат. наук (1949). В настоящее время старший научный сотрудник ВЦ РАН. Научные интересы — языки программирования, компиляторы.

Адрес автора: 117421 Москва ул. Новаторов д.40, корп.2, кв.49.