Часть 3

$LR(k)$-ГРАММАТИКИ И ТРАНСЛЯЦИИ

§ 3.1. Синтаксический анализ типа “снизу—вверх”

В предыдущей части был рассмотрен класс $LL(k)$-грамматик, наибольший подкласс КС-грамматик, естественным образом детерминированно анализируемых ***“сверху-вниз”***. Анализ заключается в последовательном построении сентенциальных форм ***левостороннего вывода***, начиная с ***начальной формы*** (S), и заканчивая конечной формой — ***анализируемой терминальной цепочкой*** (x).

Один шаг этого процесса состоит в определении того правила, правая часть которого должна использоваться для замены ***крайнего левого нетерминала*** в текущей сентенциальной форме, чтобы получить следующую сентенциальную форму.

Именно, если

$$S \underset{lm}{\overset{*}{\Rightarrow}} wA\alpha \underset{lm}{\Rightarrow} w\beta\alpha \underset{lm}{\overset{*}{\Rightarrow}} wy = x,$$

где $wA\alpha$ — текущая сентенциальная форма,

$A\alpha$ — открытая её часть и

x — анализируемая цепочка,

то используемое правило $A \to \beta$ определяется по нетерминалу A, множеству допустимых локальных правых контекстов $L = \mathrm{FIRST}_k^G(\alpha)$ и аванцепочке $u \in \mathrm{FIRST}_k^G(y)$.

k-предсказывающий алгоритм анализа, воспроизводит открытые части сентенциальных форм в своём магазине.

Простая модификация такого анализатора, выстраивает дерево вывода анализируемой цепочки (x), начиная с корня (S), на каждом шаге пристраивая к крайнему левому нетерминальному листу (A) очередное поддерево, представляющее применённое правило (рис.) очередное поддерево, представляющее применённое правило (рис. 3) очередное поддерево, представляющее применённое правило (рис. 3.1).

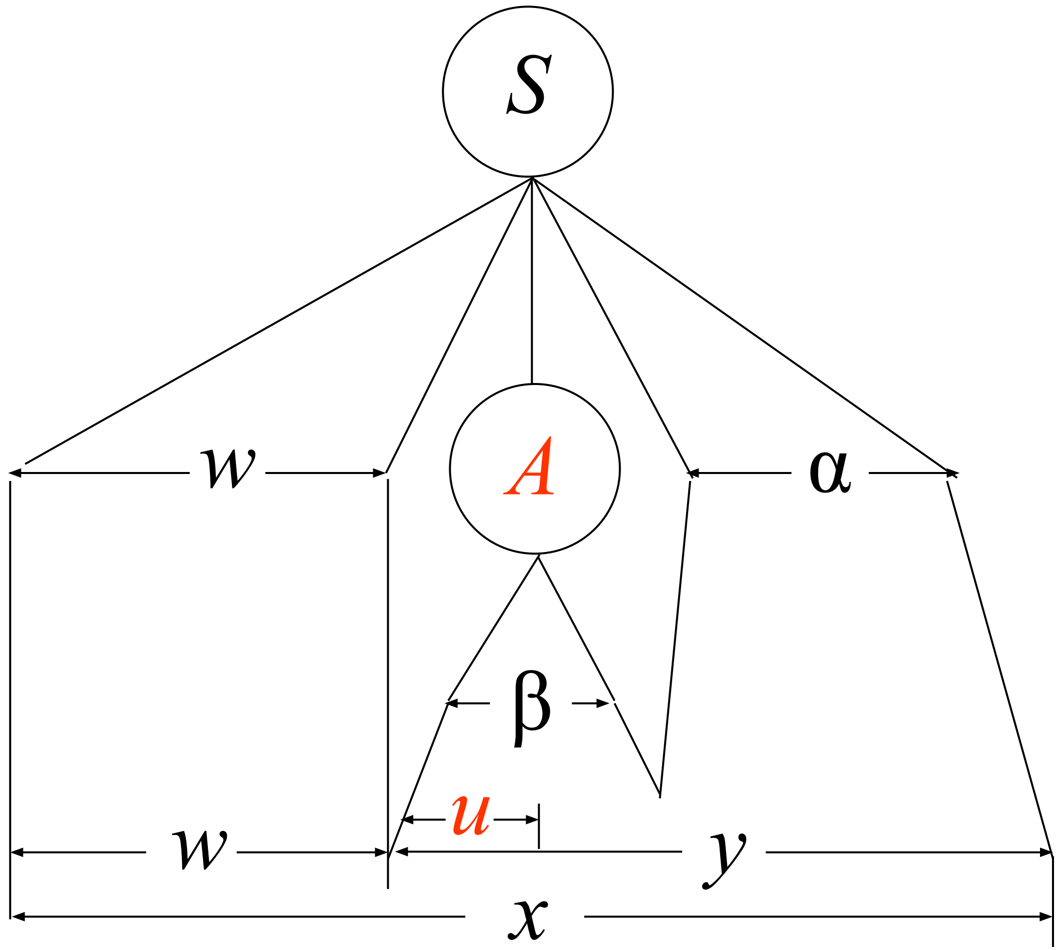

Рис. 3.1. Построение вывода “сверху-вниз”.

В противоположность этому подходу техника анализа “снизу—вверх” основана на воспроизведении сентенциальных форм ***правостороннего*** вывода, начиная с последней — ***анализируемой цепочки*** и заканчивая первой — ***начальным нетерминалом*** грамматики.

Именно, пусть

$$S = \alpha_0 \underset{rm}{\overset{p_1}{\Longrightarrow}} \alpha_1 \underset{rm}{\overset{p_2}{\Longrightarrow}} \ldots \underset{rm}{\overset{p_n}{\Longrightarrow}} \alpha_n = x$$

— правосторонний вывод терминальной цепочки x в некоторой КС-грамматике.

Индекс p_i (i = 1, 2,…, n) над стрелкой означает, что на данном шаге нетерминал текущей сентенциальной формы α_{i-1} замещается правой частью правила номер p_i.

Индекс *rm* (*right-most*) под стрелкой означает, что замещается крайнее правое вхождение нетерминала.

Последовательность номеров правил

$$\pi^R = p_n \ldots p_2 p_1$$

называется ***правосторонним анализом*** терминальной цепочки x.

Задача анализатора типа “снизу-вверх”, называемого также ***восходящим анализатором***, состоит в том, чтобы найти правосторонний анализ данной входной цепочки x в данной КС-грамматике G.

Как было сказано, исходная сентенциальная форма, с которой анализатор начинает работу, есть $\alpha_n = x$.

Далее он должен строить последующие сентенциальные формы и заканчивать свою работу тогда, когда будет построена последняя сентенциальная форма $\alpha_0 = S$.

Рассмотрим один шаг работы такого анализатора.

Пусть $\alpha_i = \alpha A w$ — текущая сентенциальная форма правостороннего вывода.

Эта форма получается из предыдущей:

$$\alpha_{i-1} = \gamma B z \Rightarrow \underline{\gamma}\beta A \underline{y} z = \underline{\alpha} A \underline{w} = \alpha_i$$

посредством правила вида

$$B \rightarrow \beta A y, \text{ где } \alpha = \gamma\beta,\ w = yz.$$

Здесь, как всегда ,

$$A, B \in V_{\mathrm{N}};\ w, y, z \in V_{\mathrm{T}}^{*};\ \alpha, \beta, \gamma \in V^{*},$$

где $V = V_{\mathrm{N}} \cup V_{\mathrm{T}}$.

Восходящий анализатор располагает текущую сентенциальную форму α_i в магазине и на входной ленте таким образом, что в магазине располагается открытая её часть, а закрытая представлена ещё непрочитанной частью входной цепочки, которая начинается с текущего входного символа и простирается до конца этой цепочки (*см*. табл. 3.1).

Табл. 3.1

№	Магазин	Вход
1	$\alpha_i = \alpha A$	w
2	$\alpha_i = \gamma\beta A$	yz
3	$\alpha_i = \gamma\underline{\beta A y}$ $B \to \beta A y$	z
4	$\alpha_{i-1} = \gamma B$	z

В строке 1 табл. 3.1 представлено расположение текущей сентенциальной формы α_i, в строке 2 — то же самое, но более детально. Предполагается, что вершина магазина расположена справа, а текущий входной символ — слева.

Анализатор посимвольно сдвигает часть входной цепочки в магазин пока не достигнет правой границы цепочки, составляющей правую часть правила, при помощи которого данная сентенциальная форма α_i была получена из предыдущей α_{i-1}.

В строке 3 представлено размещение сентенциальной формы после сдвига в магазин части входа — цепочки y.

Далее анализатор сворачивает часть цепочки, примыкающую к вершине магазина и совпадающую с правой частью упомянутого правила, в нетерминал левой части этого правила.

В строке 4 приведен результат свертки цепочки βAy, располагавшейся на предыдущем шаге в верхней части магазина, в нетерминальный символ B посредством правила

$$B \to \beta Ay.$$

Цепочка, подлежащая свертке, называется ***основой***: в таблице 3.1 в строке 3 она подчёркнута и выделена красным цветом.

Итак, один шаг работы анализатора типа “снизу-вверх” состоит в последовательном ***сдвиге*** символов из входной цепочки в магазин до тех пор, пока не достигается ***правая граница основы***, а затем у него должна быть возможность ***однозначно*** определить,

(1) ***где*** в магазине находится ***левая граница основы*** и

(2) ***по какому правилу*** (к какому нетерминалу) ***свернуть основу***.

Таким образом он воспроизводит предыдущую сентенциальную форму правостороннего вывода анализируемой цепочки.

Ret
2Ret
24

Если задача правостороннего анализа решается описанным выше механизмом детерминированно, то свойства КС-грамматики, в которой производится анализ, должны гарантировать упомянутую выше однозначность.

Процесс заканчивается, когда в магазине остается один символ S, а входная цепочка прочитана вся (на входе ничего нет).

Замечание 3.1. Если $S \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w$, то основа β не может быть в пределах α.

Действительно, если бы $\alpha = \alpha_1 \beta \alpha_2$, то предыдущая сентенциальная форма имела бы вид $\alpha_1 B \alpha_2 A w$ и из неё текущая получалась бы заменой нетерминала B на β. Но символ B не является крайним правым нетерминалом, что противоречило бы предположению о том, что мы имеем дело с правосторонним выводом. Однако основа могла бы быть в пределах цепочки w или даже цепочки Aw.

Пример 3.1. Пусть $G = (V_N, V_T, P, S)$ — контекстно-свободная грамматика, в которой $V_N = \{S\}$, $V_T = \{a, b\}$,

$$P = \{(1)\ S \to SaSb,\ (2)\ S \to \varepsilon\}.$$

Ret
1Ret
122

Рассмотрим правосторонний вывод в этой грамматике:

$$S \underset{rm}{\overset{(1)}{\Rightarrow}} \underline{SaSb} \underset{rm}{\overset{(1)}{\Rightarrow}} Sa\underline{SaSb}b \underset{rm}{\overset{(2)}{\Rightarrow}} SaSa\underline{\varepsilon}bb \underset{rm}{\overset{(2)}{\Rightarrow}} Sa\underline{\varepsilon}abb \underset{rm}{\overset{(2)}{\Rightarrow}} \underline{\varepsilon}aabb.$$

Здесь $\pi^R = 22211$ — правосторонний анализ цепочки $x = aabb$.

Ret
1Ret
162

Табл. 3.2

№	Магазин	Вход	Действие
1	$	*aabb*	reduce 2
2	$*S*	*aabb*	shift
3	$*Sa*	*abb*	reduce 2
4	$*SaS*	*abb*	shift
5	$*SaSa*	*bb*	reduce 2
6	$*SaSaS*	*bb*	shift
7	$*Sa<u>SaSb</u>*	*b*	reduce 1
8	$*SaS*	*b*	shift
10	$*<u>SaSb</u>*		reduce 1
	$*S*		

“Дно” магазина отмечено маркером $. ***Исходная конфигурация*** характеризуется тем, что магазин пуст (маркер “дна” не считается), непросмотренная часть входа — вся цепочка *aabb*.

Первое действие — свёртка пустой основы на вершине магазина по правилу 2. Это приводит к конфигурации, показанной в строке 2.

Следующее действие по команде shift — сдвиг: текущий символ *a* перемещается со входа на вершину магазина. Положение вершины магазина тоже изменяется, как и текущий входной символ. Эта конфигурация представлена в строке 3.

Дальнейшие действия “сдвиг—свертка” в конце концов приводят к ***заключительной конфигурации***: в магазине — начальный нетерминал грамматики, и вся входная цепочка просмотрена.

Правый анализ ‘*aabb*’:
$\pi^R = 22211$
≡ последовательность свёрток

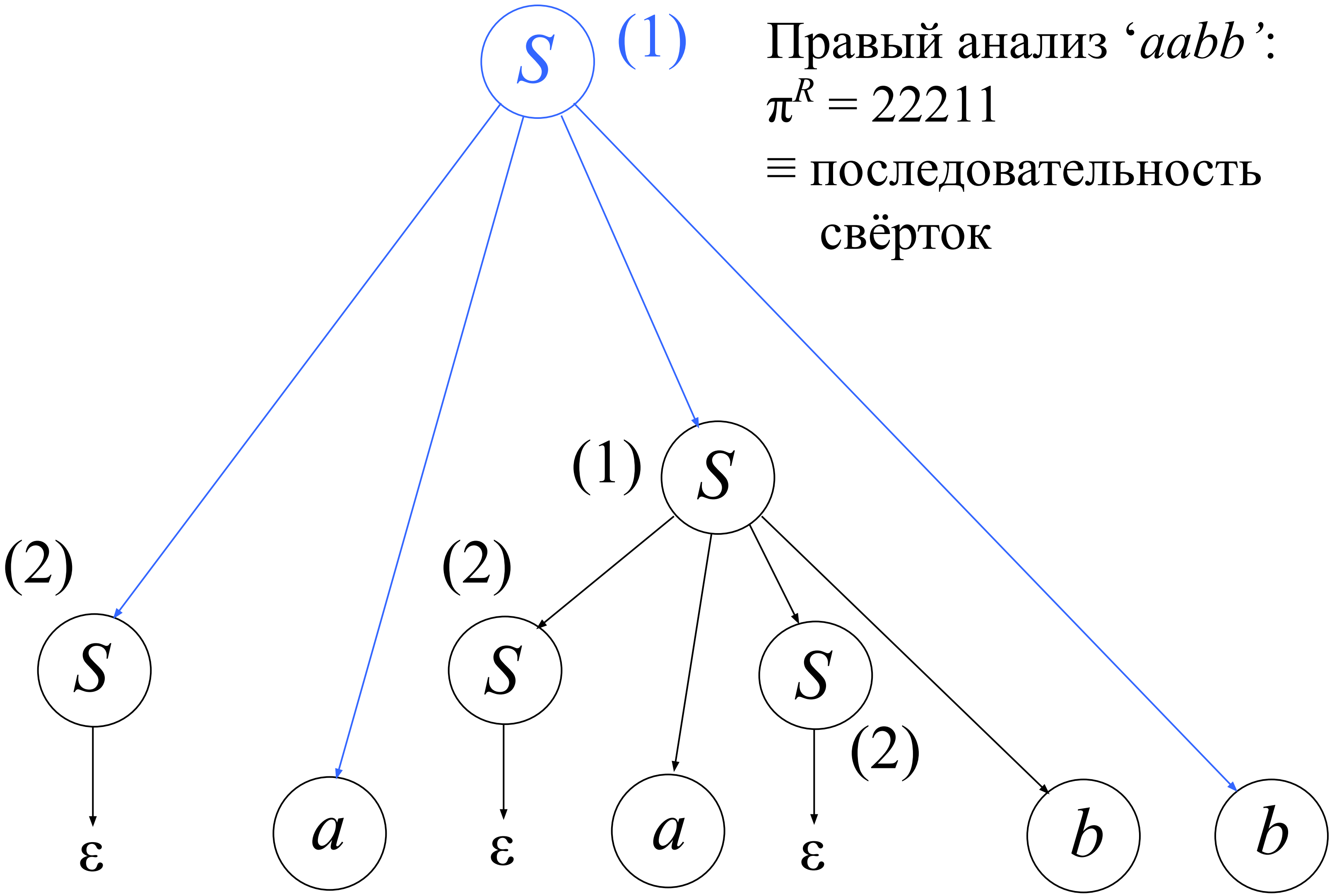

Рис. 3.2. Построение дерева вывода “снизу-вверх”.

Номера правил при командах свертки (reduce) образуют правосторонний анализ входной цепочки *aabb*.

Не все КС-грамматики поддаются правостороннему анализу посредством детерминированного механизма типа “*перенос– свертка*”. Мы рассмотрим здесь класс КС-грамматик, которые позволяют однозначно разрешать упомянутые проблемы путём заглядывания на *k* символов, следующих за основой (*аванцепочка*).

Именно:

(1) производить сдвиг или свертку?

(2) если делать свёртку, то по какому правилу?

(3) когда закончить процесс?

Этот класс грамматик, открытый Д. Кнутом, называется ***LR*(*k*)-грамматиками**.

В этом названии ***L*** обозначает направление просмотра входной цепочки *слева направо*, ***R*** — результатом является *правосторонний анализ*, ***k*** — *предельная длина правого контекста основы*.

§3.2. $LR(k)$-Грамматики

В этом параграфе мы дадим строгое определение $LR(k)$-грамматик и опишем характерные их свойства.

Определение 3.1. Пусть $G = (V_N, V_T, P, S)$ — контекстно-свободная грамматика.

Пополненной грамматикой, полученной из G, назовем грамматику

$$G' = (V'_N, V'_T, P', S'),$$

где $V'_N = V_N \cup \{S'\}$; $S' \notin V'_N \cup V'_T$;

$$V'_T = V_T; \ P' = P \cup \{S' \to S\}.$$

***Определение* 3.2.** Пусть $G' = (V'_{\mathrm{N}}, V'_{\mathrm{T}}, P', S')$ — пополненная грамматика для КС-грамматики G. Грамматика G является *$LR(k)$-грамматикой*, если для любых двух правосторонних выводов вида

1) $S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta w$,

2) $S' \underset{rm}{\overset{*}{\Rightarrow}} \gamma B x \underset{rm}{\Rightarrow} \alpha\beta y$,

в которых, $\mathrm{FIRST}_k^G(w) = \mathrm{FIRST}_k^G(y)$, должно быть $\alpha A y = \gamma B x$.

Другими словами, $\alpha = \gamma$, $A = B$, $y = x$.

Ret Ret 38 72 99 98 104
3Ret 3
36

Иначе говоря, если согласно выводу 1) β — основа сентенциальной формы $\alpha\beta w$, сворачиваемая в нетерминал A по правилу вида $A \rightarrow \beta$, то и выводе 2) β должна быть основой сентенциальной формы $\alpha\beta y$, сворачиваемой *по тому же самому правилу* (*инвариант* правосторонних выводов в определе-нии $LR(k)$-грамматик).

И поскольку в выводе 2) в результате свёрки основы β в цепочке $\alpha\beta y$ в нетерминал A получает-ся цепочка αAy, которая должна быть равна пре-дыдущей сентенциальной форме γBx, то $\alpha Ay = \gamma Bx$. Это равенство возможно лишь при $\alpha = \gamma$, $A = B$, $x = y$, поскольку цепочки γ и x наследуются от предыду-щей формы γBx.

Из этого определения следует, что если имеется право-выводимая цепочка $\alpha_i = \alpha\beta w$, где β — основа, полученная по правилу $A \rightarrow \beta$, и если $\alpha\beta = X_1 X_2 \ldots X_j \ldots X_r$, $X_p \in V^*$ ($p = 1, 2, \ldots, r$), то

1) зная первые символы $X_1 X_2 \ldots X_j$ цепочки $\alpha\beta$ и не более, чем k следующих символов цепочки $X_{j+1} X_{j+2} \ldots X_r w$, мы можем быть уверены, что правый конец основы не будет достигнут до тех пор, пока $j \neq r$;

2) зная цепочку $\alpha\beta = X_1X_2\ldots X_r$ и не более k символов цепочки w, мы можем быть уверены, что именно β является основой, сворачиваемой в нетерминал по правилу $A \to \beta$;

3) если $\alpha_{i-1} = S'$, можно сигнализировать о выводимости исходной терминальной цепочки из S' и, следовательно, из S.

Более того, последовательность номеров правил, использованных при свёрках, есть правосторонний анализ предложения языка.

Использование *пополненной* $LR(k)$-грамматики при анализе существенно для однозначного установления момента его конца.

Действительно, если грамматика использует S в правых частях правил, то свёртка основы к S не может служить сигналом приёма входной цепочки. Свёртка же к S' в пополненной грамматике служит таким сигналом, поскольку нигде, кроме начальной сентенциальной формы, S' не встречается.

Существенность использования пополненной грамматики в определении $LR(k)$-грамматик продемонстрируем на следующем конкретном примере.

Пример 3.2. Пусть пополненная грамматика имеет следующие правила:
0) $S' \to S$, 1) $S \to Sa$, 2) $S \to a$.

Пример 3.2.

Если игнорировать 0-е правило, то, не заглядывая в правый контекст основы Sa, можно сказать, что она должна сворачиваться в S благодаря правилу 1.

Аналогично основа a безусловно должна сворачиваться в S благодаря правилу 2.

Создается впечатление, что данная грамматика без 0-го правила есть $LR(0)$-грамматика.

С учётом же 0-го правила имеем:

(1) $S' \underset{rm}{\Rightarrow} S$, $\qquad$ (0) $S' \to S$

(2) $S' \underset{rm}{\Rightarrow} S \underset{rm}{\Rightarrow} Sa$.

По определению $LR(0)$-грамматики: если в выводе (1) основа S сворачивается в нетерминал S' по правилу (0), то с учётом вывода (2) должно быть $S'a = S$, чего нет!

Итак, данная грамматика не является $LR(0)$-грамматикой.

Лемма 3.1. *Пусть* $G = (V_N, V_T, P, S)$ — *LR*(*k*)-*грамматика и* G' — *пополненная грамматика для* G.

Если существуют правосторонние выводы в G' *вида*

$$1)\ S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta w,$$

$$2)\ S' \underset{rm}{\overset{*}{\Rightarrow}} \gamma B \underset{rm}{\Rightarrow} \gamma\beta' x = \alpha\beta y,$$

в которых $\mathrm{FIRST}_k^G(w) = \mathrm{FIRST}_k^G(y)$,

то $\gamma = \alpha,\ B = A,\ x = y,\ \beta' = \beta$.

47

Доказательство. Заметим, что первые три равенства непосредственно следуют из определения 3.2 и остается установить только, что $\beta' = \beta$.

Если $\beta' x = \alpha\beta y$, то, поскольку $\gamma = \alpha$ и $x = y$, имеем $\gamma\beta' x = \alpha\beta' y = \alpha\beta y$.

Следовательно, $\beta' = \beta$. Что и требовалось доказать.

Рассмотрим несколько примеров, иллюстрирующих свойства КС-грамматик, от которых зависят их принадлежность к классу *LR*.

Пример 3.3. Пусть грамматика G содержит следующие правила:

1) $S \to C \mid D$; 2) $C \to aC \mid b$; 3) $D \to aD \mid c$.

Спрашивается, является ли она $LR(0)$-грамматикой?

Отметим прежде всего, что грамматика G — *праволинейна*. Это значит, что любая сентенциальная форма содержит не более одного нетерминала, причём его правый контекст всегда пуст.

Очевидно также, что любая сентенциальная форма имеет один из следующих видов

$$a^iC,\quad a^ib,\quad a^iD,\quad a^ic,\quad \text{где } i \geq 0.$$

Пополненная грамматика содержит ещё одно правило: (0) $S' \rightarrow S$.

При сопоставлении с образцами выводов в определении $LR(k)$-грамматик в роли нетерминала A могут быть нетерминалы S', C или D.

Отметим, что нетерминал S в роли нетерминала A нас не интересует, поскольку не существует двух разных право-сентенциальных форм, в которых участвовал бы нетерминал S.

В любом случае в роли цепочек w, x и y выступает только пустая цепочка ε из-за праволинейности грамматики G.

Принимая всё это во внимание, проверим, отвечает ли данная грамматика определению $LR(0)$-грамматики. В данном конкретном случае $LR(0)$-условие состоит в том, что если существуют два право-сторонних вывода вида

1) $S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha A \underset{rm}{\Rightarrow} \alpha\beta$,

2) $S' \underset{rm}{\overset{*}{\Rightarrow}} \gamma B \underset{rm}{\Rightarrow} \alpha\beta$,

то должно быть $B = A$ и $\gamma = \alpha$.

Другими словами, любая сентенциальная форма должна быть выводима единственным способом. Что это именно так, можно убедится непосредственно.

Условие $B = A$ и $\gamma = \alpha$ выполняется тривиальным образом, поскольку не существует двух разных выводов одной и той же сентенциальной формы.

Итак, $LR(0)$-условие выполняется и, следовательно, G — $LR(0)$-грамматика.

Заметим, что G — не LL-грамматика.

Пример 3.4. Рассмотрим грамматику G с правилами:

1) $S \to Ab \mid Bc$; 2) $A \to Aa \mid \varepsilon$; 3) $B \to Ba \mid \varepsilon$.

Эта лево-линейная грамматика порождает тот же самый язык, что и грамматика предыдущего примера, но она *не является* $LR(k)$-грамматикой ни при каком $k \geq 0$.

44

Действительно, рассмотрим, например, два таких правосторонних вывода в расширенной грамматики:

$$(1)\ S' \underset{rm}{\Rightarrow} S \underset{rm}{\Rightarrow} Ab \overset{*}{\underset{rm}{\Rightarrow}} Aa^ib \Rightarrow a^ib,$$

$$(2)\ S' \underset{rm}{\Rightarrow} S \overset{*}{\underset{rm}{\Rightarrow}} Bc \overset{*}{\underset{rm}{\Rightarrow}} Ba^ic \Rightarrow a^ic.$$

Здесь цепочки a^ib и a^ic являются правыми контекстами для пустой основы, которая в одном случае сворачивается в нетерминал A, а в другом — в нетерминал B.

В какой нетерминал сворачивать пустую основу, можно определить лишь по последнему символу (если он — b, то сворачивать в нетерминал A, если он — c, то сворачивать в нетерминал B), который может отстоять от этой основы на сколько угодно большое расстояние (в зависимости от выбора i). Следовательно, каким бы большим ни было k, всегда найдется такое i, что $\mathrm{FIRST}_k^G(a^i b) = \mathrm{FIRST}_k^G(a^i c)$,

но при этом $A \neq B$.

Определение* 3.3.** Грамматики, в которых существует несколько разных правил, отличающихся только нетерминалами в левой части, называются ***необратимыми.

В примере 3.4 мы имели дело с необратимой грамматикой.

Причина, по которой данная грамматика не LR, в том, что правый контекст основы, каким бы длинным он ни был, не даёт возможности однозначно определить, в какой нетерминал следует её сворачивать.

Пример 3.5. Рассмотрим грамматику, иллюстрирующую другую причину, по которой она не $LR(1)$: невозможность однозначно определить, что является основой в право-выводимой сентенциальной форме:

1) $S \rightarrow AB$, 2) $A \rightarrow a$, 3) $B \rightarrow CD$,

4) $B \rightarrow aE$, 5) $C \rightarrow ab$, 6) $D \rightarrow bb$,

7) $E \rightarrow bba$.

В этой грамматике рассмотрим два правосторонних вывода:

$C \to ab$

$$(1)\ S' \underset{rm}{\overset{(0)}{\Rightarrow}} S \underset{rm}{\overset{(1)}{\Rightarrow}} AB \underset{rm}{\overset{(3)}{\Rightarrow}} ACD \underset{rm}{\overset{(6)}{\Rightarrow}} ACbb \underset{rm}{\overset{(5)}{\Rightarrow}} \overbrace{A\underline{ab}}^{\alpha\beta}\ \overbrace{bb}^{w},$$

$$(2)\ S' \underset{rm}{\overset{(0)}{\Rightarrow}} S \underset{rm}{\overset{(1)}{\Rightarrow}} AB \underset{rm}{\overset{(4)}{\Rightarrow}} AaE\overbrace{\ }^{x=\varepsilon} \underset{rm}{\overset{(7)}{\Rightarrow}} \overbrace{Aa\ \underline{b}}^{\alpha\beta}\overbrace{\underline{ba}}^{y}.$$

$\beta = ab$

Здесь $\alpha\beta = Aab$, $w = bb$, $\beta = ab$, $x = \varepsilon$, $y = ba$. И хотя $\mathrm{FIRST}_1^G(w) = \mathrm{FIRST}_1^G(y) = \{b\}$, оказывается, что $x \neq y$, а это является нарушением условия $LR(1)$ (см. лемму 3.1).

$ACbb \neq AaE$!!!

§ 3.3. *LR*(*k*)-Анализатор

Аналогично тому, как для *LL*(*k*)-грамматик адекватным типом анализаторов является *k*-предсказывающий алгоритм анализа, поведение которого диктуется *LL*(*k*)-таблицами, для *LR*(*k*)-грамматик адекватным механизмом анализа является *LR*(*k*)-анализатор, управляемый *LR*(*k*)-таблицами.

Эти *LR*(*k*)-таблицы являются строчками управляющей таблицы *LR*(*k*)-анализатора.

$LR(k)$-Таблица состоит из двух подтаблиц, представляющих следующие функции:

$$f : V_{\mathrm{T}}^{k*} \to \{\text{shift, reduce } i\text{, accept, error}\},$$

$$g : V_{\mathrm{N}} \cup V_{\mathrm{T}} \to \mathcal{T} \cup \{\text{error}\},$$

где V_{T} — входной алфавит анализатора (терминалы грамматики); V_{N} — нетерминалы грамматики; $\mathcal{T}$ — множество $LR(k)$-таблиц для G, оно строится по пополненной грамматике G' для $LR(k)$-грамматики G.

Подтаблица f по аванцепочке определяет одно из трёх действий над необработанной частью входной цепочки: *сдвиг*, *свёрка*, *приём* (счастливый конец анализа), или сигнализирует об *ошибке* в ней.

Подтаблица g по символу грамматики, определяет, какой $LR(k)$-таблицей следует руководствоваться на следующем такте работы анализатора. Она помещается на вершину магазина.

Алгоритм 3.1: действия $LR(k)$-анализатора.

<u>*Вход*</u>: $G = (V_\mathrm{N}, V_\mathrm{T}, P, S)$ — $LR(k)$-грамматика;

$\mathscr{T}$ — множество $LR(k)$-таблиц для G;

$T_0 \in \mathscr{T}$ — начальная $LR(k)$-таблица;

$x \in V_\mathrm{T}^*$ — входная цепочка.

<u>*Выход*</u>: π^R — правосторонний анализ x.

Напомним, что π — правосторонний вывод цепочки x.

205 222

Метод.

LR(*k*)-Анализатор реализует классический механизм “сдвиг-свёртка”, описанный в параграфе §3.1. Его действия будем описывать в терминах конфигураций, понимая под конфигурацией тройку (α, w, y), где $\alpha \in (V_N \cup V_T \cup \mathscr{T})^*$ — магазинная цепочка; $w \in V_T^*$ — непросмотренная часть входной цепочки; y — выходная цепочка, состоящая из номеров правил грамматики G.

Начальная конфигурация есть (T_0, x, ε).

Далее алгоритм действует согласно следующему описанию в зависимости от того, какая $LR(k)$-таблица, находится не вершине магазина.

1: ***Сдвиг.***

Пусть текущая конфигурация есть

$$(\gamma T, w, \pi),$$

где $T \in \mathcal{T}$ — некоторая $LR(k)$-таблица,

$T = (f, g)$ и пусть

$f(u) = \text{shift}$ для $u \in \text{FIRST}_k^G(w)$.

1.1. $w \neq \varepsilon$, $w = aw'$, где $a \in V_T$, $w' \in V_T^*$.

1.1.1. $g(a) = T'$, $T' \in \mathsf{T}$.

Анализатор совершает движение:

$(\gamma T, w, \pi) = (\gamma T, aw', \pi) \vdash (\gamma TaT', w', \pi)$,

воспроизводящее сдвиг.

1.1.2. $g(a) = \text{error}$.

Анализатор сигнализирует об ошибке и останавливается.

1.2. $w = \varepsilon$, $u = \varepsilon$, $f(u) = \text{error}$.

Сдвигать нечего. Анализатор сигнализирует об ошибке и останавливается.

2: ***Свертка.***

Пусть текущая конфигурация есть

$$(\gamma T X_1 T_1 X_2 T_2 \ldots X_m T_m, w, \pi),$$

где $T, T_1, T_2, \ldots, T_m \in \mathcal{T}$ — некоторые $LR(k)$-таблицы; и пусть $T = (f, g)$, $T_m = (f_m, g_m)$,

$u \in \mathrm{FIRST}_k^G(w)$,

$f_m(u) =$ reduce i, $A \to \alpha$ — i-е правило из множества правил P,

$$\alpha = X_1 X_2 \ldots X_m \text{ — основа.}$$

2.1. $g(A) = T'$,

где $T' \in \mathrm{T}$ — некоторая $LR(k)$- таблица.

Анализатор совершает переход

$$(\gamma T \underbrace{X_1 T_1 X_2 \ldots X_m T_m}_{\text{свёртка в } A}, w, \pi) \vdash (\gamma T A, w, \pi i) \vdash$$

$$\vdash (\gamma T A T', w, \pi i),$$

воспроизводящий свертку в магазине и переход к следующей $LR(k)$-таблице T'.

2.2. $g(A) = \text{error}$.

Анализатор сигнализирует об ошибке и останавливается.

3: ***Ошибка***.

Пусть текущая конфигурация есть

$$(\gamma T, w, \pi),$$

где $T \in \mathcal{T}$ — некоторая $LR(k)$-таблица;

$$u \in \mathrm{FIRST}_k^G(w),$$

$$T = (f, g) \text{ и } f(u) = \text{error}.$$

Анализатор сигнализирует об ошибке и останавливается.

4: ***Приём***.

Пусть текущая конфигурация есть

$$(T_0ST, \varepsilon, \pi^R),$$

$T = (f, g),\ f(\varepsilon) = \text{accept}.$

Анализатор сигнализирует о приёме цепочки x и останавливается.

Выходная цепочка π^R представляет правосторонний анализ цепочки x.

Заметим, что структура магазинной цепочки всегда имеет вид $T_0(XT)^*$, где T_0, T — $LR(k)$-таблицы, а $X \in V_N \cup V_T$.

Пример 3.6. Обратимся ещё раз к грамматике из примера 3.1. Как мы увидим далее (*см*. примеры 3.10. примеры 3.10 и 3.11), она — $LR(1)$-грамматика.

Она имеет следующие правила:

$$1)\ S \to SaSb, \quad 2)\ S \to \varepsilon.$$

Управляющая таблица $LR(1)$-анализатора, построенная по ней, имеет вид, представленный табл. 3.3, где пустые клетки соответствуют значениям error, а целые представляют номера правил свёртки.

Табл. 3.3

$LR(1)$-таблицы	$f(u)$			$g(X)$			
	a	b	ε	S	a	b	
T_0	2		2	T_1			171
T_1	shift		accept		T_2		189
T_2	2	2		T_3			215
T_3	shift	shift			T_4	T_5	239
T_4	2	2		T_6			242 275
T_5	1 (‘c’)		1 (‘c’)				2276
T_6	shift	shift			T_4	T_7	227
T_7	1 (‘c’)	1 (‘c’)					9 303

Пример 3.6.

Рассмотрим действия этого анализатора на входной цепочке *aabb*:

После сдвига или свёрки на вершину магазина выкладывается $LR(k)$-табличка, управляющая следующим шагом анализа.

$(T_0\varepsilon, aabb, \varepsilon) \vdash$

$(T_0 S T_1, aabb, 2) \vdash$

$(T_0 S T_1 a T_2\, \varepsilon, abb, 2) \vdash$

$(T_0 S T_1 a T_2 S T_3, abb, 22) \vdash$

$(T_0 S T_1 a T_2 S T_3 a T_4\, \varepsilon, bb, 22) \vdash$

$(T_0 S T_1 a T_2 S T_3 a T_4 S T_6, bb, 222) \vdash$

$(T_0 S T_1 a T_2 \cancel{S T_3 a T_4 S T_6 b T_7}, b, 222) \vdash$

$(T_0 S T_1 a T_2 S T_3, b, 2221) \vdash$

$(T_0 \cancel{S T_1 a T_2 S T_3 b T_5}, \varepsilon, 2221) \vdash$

$(T_0 S T_1, \varepsilon, 22211).$

Итак, цепочка $aabb$ принимается, и $\pi^R = 22211$ — её правосторонний анализ, а $\pi = 11222$ — её правосторонний вывод.

§ 3.4. Свойства $LR(k)$-грамматик

Рассмотрим некоторые следствия из определения $LR(k)$-грамматик, подводящие нас к механизму анализа.

***Определение* 3.4.** Пусть $G = (V_N, V_T, P, S)$ — контекстно-свободная грамматика и

$$S \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta w \qquad (\beta \text{ — основа})$$

— некоторый правосторонний вывод в грамматике G, где $\alpha, \beta \in (V_N \cup V_T)^*$, $w \in V_T^*$.

Активным префиксом сентенциальной формы $\alpha\beta w$ называется любая начальная часть (префикс) цепочки $\alpha\beta$, включая, в частности, ε и всю эту цепочку $\alpha\beta$.

***Определение* 3.5.** Пусть $G = (V_N, V_T, P, S)$ — контекстно-свободная грамматика и $A \to \beta_1\beta_2 \in P$.

Композицию $[A \to \beta_1.\beta_2, u]$, где $u \in V_T^{k*}$, назовем *LR(k)-ситуацией*.

Здесь $\beta_1, \beta_2 \in (V_N \cup V_T)^*$, то есть позиция точки в правой части правила грамматики может выбираться произвольно.

В частности, при

$\beta_1 = \varepsilon$ — точка перед основой,

$\beta_2 = \varepsilon$ — точка за основой,

$\beta_1\beta_2 = \varepsilon$ — точка представляет пустую основу.

***Определение* 3.6.** Пусть $G = (V_N, V_T, P, S)$ — контекстно-свободная грамматика и

$$S \overset{*}{\underset{rm}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta w$$

— правосторонний вывод в грамматике G, где $\beta = \beta_1\beta_2$; $\alpha, \beta_1, \beta_2 \in (V_N \cup V_T)^*$; $w \in V_T^*$.

Назовём $[A \to \beta_1.\beta_2, u]$, где $u \in \mathrm{FIRST}_k^G(w)$, *$LR(k)$-ситуацией, допустимой для активного префикса* $\alpha\beta$.

88 137
8

$$S \overset{*}{\underset{rm}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta w = \overbrace{a_1 a_2 \ldots a_m}^{\alpha_0 = \alpha}\overbrace{b_1 b_2 \ldots b_n}^{\beta}\overbrace{\underbrace{c_1 c_2 \ldots c_k}_{u \in \mathrm{FIRST}_k^G(w)} \ldots c_p}^{w}$$

$$\underbrace{a_1 a_2 \ldots a_m b_1}_{\alpha_1}$$

$$\underbrace{a_1 a_2 \ldots a_m b_1 b_2}_{\alpha_2} \quad \ldots$$

$$\underbrace{a_1 a_2 \ldots a_m b_1 b_2 \ldots b_n}_{\alpha_n}$$

$\alpha = \alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_n$ — активные префиксы для ситуации $[A \to \beta_1.\beta_2, u]$, $\beta = \beta_1\beta_2$.

Позиция точки может быть после

α_0 $(\beta_1 = \varepsilon, \beta_2 = \beta)$,

α_1 $(\beta_1 = \alpha b_1)$,

α_2 $(\beta_1 = \alpha b_1 b_2)$,

...

α_n $(\beta_1 = \beta, \beta_2 = \varepsilon)$.

Пример 3.7. Обратимся ещё раз к $LR(0)$-грамматике из примера 3.3, которая содержит следующие правила:

1) $S \to C \mid D$, 2) $C \to aC \mid b$, 3) $D \to aD \mid c$.

Рассмотрим правосторонний вывод $S \underset{rm}{\Rightarrow} C$.

В право-сентенциальной форме C основой является C. Эта форма имеет два активных префикса: ε и C.

Для активного префикса ε допустима $LR(0)$-ситуация $[S \to .C, \varepsilon]$, а для активного префикса C — $LR(0)$-ситуация $[S \to C., \varepsilon]$.

Пример 3.7.

Рассмотрим выводы, дающие активный префикс *aaaa*, и четыре *LR*(0)-ситуации, допустимые для него:

$$\boxed{C \to aC}$$

$$(1)\ S \underset{rm}{\overset{*}{\Rightarrow}} aaaC \underset{rm}{\Rightarrow} \overbrace{aaa}^{\alpha}\ \overbrace{\underbrace{a}_{\beta_1}\underbrace{C}_{\beta_2}}^{\beta},\ \alpha\beta_1 = aaaa,\ [C \to a.C,\ \varepsilon];$$

$$\boxed{C \to b}$$

$$(2)\ S \underset{rm}{\overset{*}{\Rightarrow}} aaaaC \underset{rm}{\Rightarrow} \overbrace{aaaa}^{\alpha}\ \overbrace{\underbrace{\ }_{\beta_1}\underbrace{b}_{\beta_2}}^{\beta},\ \alpha\beta_1 = aaaa,\ [C \to .b,\ \varepsilon];$$

Пример 3.7.

$\boxed{D \to aD}$

$$(3)\ S \underset{rm}{\overset{*}{\Rightarrow}} aaaD \underset{rm}{\overset{*}{\Rightarrow}} \overbrace{aaa}^{\alpha}\ \overbrace{\underbrace{a}_{\beta_1}\underbrace{D}_{\beta_2}}^{\beta},\ \alpha\beta_1{=}aaaa,\ [D \to a.D,\ \varepsilon];$$

$\boxed{D \to c}$

$$(4)\ S \underset{rm}{\overset{*}{\Rightarrow}} aaaaD \underset{rm}{\overset{*}{\Rightarrow}} \overbrace{aaaa}^{\alpha}\ \overbrace{\underbrace{\ }_{\beta_1}\underbrace{c}_{\beta_2}}^{\beta},\ \alpha\beta_1{=}aaaa,\ [C \to .c,\ \varepsilon];$$

Во всех случаях правый контекст основы (β) пуст ($k = 0$).

Лемма 3.2. *Пусть* $G = (V_N, V_T, P, S)$ — ***не*** *LR(k)-грамматика. Тогда существуют два правосторонних вывода в пополненной грамматике:*

$$1)\ S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha \underline{\beta} w,$$

$$2)\ S' \underset{rm}{\overset{*}{\Rightarrow}} \gamma B x \underset{rm}{\Rightarrow} \gamma \underline{\delta} x = \alpha\beta y,$$

такие, что $x, y, w \in V_T^*$ *и*

а) $\mathrm{FIRST}_k^G(w) = \mathrm{FIRST}_k^G(y)$,

б) $\gamma B x \neq \alpha A y$,

в) $|\gamma\delta| \geq |\alpha\beta|$.

98

Доказательство. Если G — ***не*** $LR(k)$-грамматика, то условия а) и б) выполняются как *отрицание* $LR(k)$-условия из определения $LR(k)$-грамматик.

Условие в) неформально означает, что правая граница основы δ в выводе 2) удалена от начала сентенциальной формы $\gamma\delta x$, по крайней мере, не менее, чем удалена основа β в выводе 1) от начала сентенциальной формы $\alpha\beta w$.

Условие в) не столь очевидно. Простой обмен ролями этих двух выводов ничего не даёт, так как этим приёмом мы добьёмся только выполнения условий а) и в), но не очевидно, что при этом будет выполнено условие б).

Предположим, что выводы 1) и 2) удовлетворяют условиям а) и б), но условие в) не выполнено, т. е. что $|\gamma\delta| < |\alpha\beta|$.

Покажем, что тогда найдется другая пара выводов, которые удовлетворяют всем трём условиям.

Поскольку в выводе 2) $\gamma\delta x = \alpha\beta y$, то $|\,\gamma\delta x\,| = |\,\alpha\beta y\,|$, и, учитывая, что $|\,\gamma\delta\,| < |\,\alpha\beta\,|$, заключаем: $|\,x\,| > |\,y\,|$, т. е. $x = zy$ при некотором $z \in V_{\mathrm{T}}^{+}$, $|\,z\,| > 0$.

Заметим, что цепочка z является префиксом цепочки x, а не её окончанием, так как именно цепочка y является окончанием всей сентенциальной формы $\alpha\beta y$. Два разных разбиения одной и той же сентенциальной формы $\gamma\delta x = \alpha\beta y$ представлено на рис. 3.2.

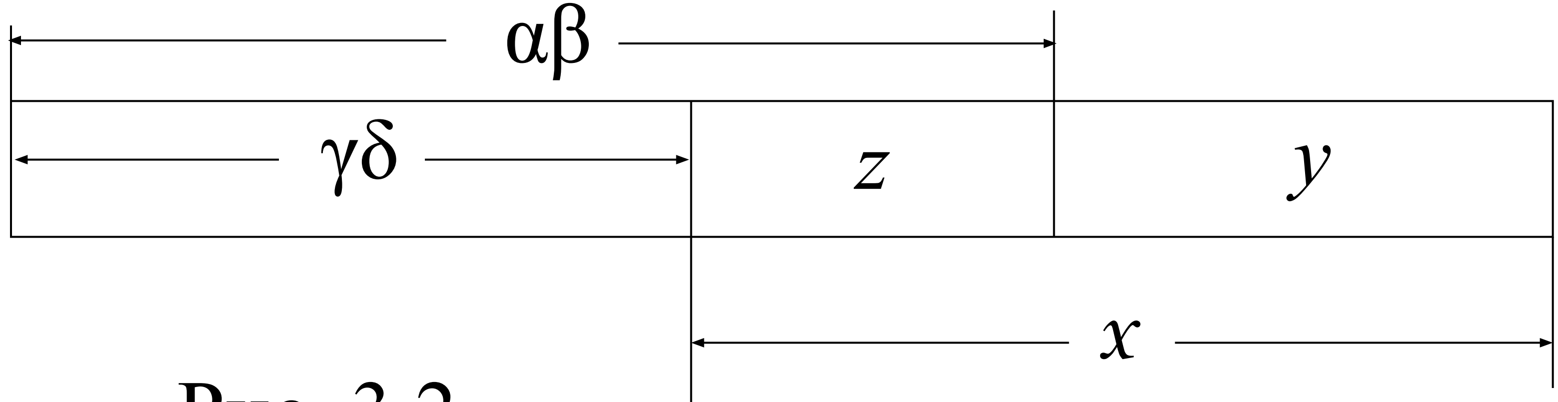

Рис. 3.2.

Условие $\gamma\delta x = \alpha\beta y$ можно переписать как $\gamma\delta z y = \alpha\beta y$, и потому

$$\gamma\delta z = \alpha\beta. \tag{3.1}$$

Это видно и на рис. 3.2.

Вывод 2) разметим по образцу первого с учётом равенства $x = zy$, а вывод 1) разметим по образцу второго с учётом равенства (3.1):

$$\alpha\beta = \gamma\delta z \ (3.1)$$

$$(1') \ S' \overset{*}{\underset{rm}{\Rightarrow}} \overbrace{\gamma}^{[\alpha]} \ \overbrace{B}^{[A]} \ \overbrace{\underbrace{zy}_{x}}^{[w]} \underset{rm}{\Rightarrow} \overbrace{\gamma}^{[\alpha]} \ \overbrace{\delta}^{[\beta]} \ \overbrace{\underbrace{zy}_{x}}^{[w]},$$

вывод 2)

$$(2') \ S' \overset{*}{\underset{rm}{\Rightarrow}} \overbrace{\alpha}^{[\gamma]} \ \overbrace{A}^{[B]} \ \overbrace{w}^{[x]} \underset{rm}{\Rightarrow} \overbrace{\alpha}^{[\gamma]} \ \overbrace{\beta}^{[\delta]} \ \overbrace{w}^{[x]} = \overbrace{\gamma\delta}^{[\alpha\beta]} \ \overbrace{zw}^{[y]}.$$

вывод 1)

Эти два вывода обладают следующими особенностями:

а′) $\mathrm{FIRST}_k^G([w]) = \mathrm{FIRST}_k^G([y])$,
так как $[w] = zy$, $[y] = zw$ при том, что $\mathrm{FIRST}_k^G(w) = \mathrm{FIRST}_k^G(y)$;

б′) $[\gamma][B][x] \neq [\alpha][A][y]$, т. к. $[\gamma] = \alpha$, $[B] = A$, $[x] = w$, $[\alpha] = \gamma$, $[A] = B$, $[y] = zw$, поскольку $[\gamma][B][x] = \alpha Aw$ и $[\alpha][A][y] = \gamma Bzw$, а цепочки αAw и γBzw не могут быть равными, так как их терминальные окончания w и zw не равны между собой, ибо $|\,z\,| > 0$.

Наконец, выполняется условие

в$'$) $\|[\gamma\delta]\| > \|[\alpha\beta]\|$, ибо $[\gamma\delta] = \alpha\beta$, $[\alpha\beta] = \gamma\delta$ и $|\gamma\delta| < |\alpha\beta|$ по предположению.

Итак, исходная пара правосторонних выводов 1) и 2), которые обменялись ролями, представлены в требуемом виде $(1')$ и $(2')$, которые удовлетворяют всем трём условиям а$'$), б$'$) и в$'$).

Что и требовалось доказать.

Введём функцию $\mathrm{EFF}_k^G(\alpha)$, где $\alpha \in (V_N \cup V_T)^*$, необходимую для построения $LR(k)$-анализатора. Она будет помогать при определении, является ли ε-цепочка основой для данной право-сентенциальной формы, подлежащей свёртке.

***Определение* 3.7.** Пусть $G = (V_\mathrm{N}, V_\mathrm{T}, P, S)$ — cfg и $\alpha \in (V_\mathrm{N} \cup V_\mathrm{T})^*$. Положим

$$\mathrm{EFF}_k^G(\alpha) = \begin{cases} \mathrm{FIRST}_k^G(\alpha), \text{ если } \alpha \text{ начинается на терминал, а иначе} \\ \{ w \in V_\mathrm{T}^{*k} \mid \exists \alpha \overset{*}{\underset{rm}{\Rightarrow}} \beta \underset{rm}{\Rightarrow} wx, \\ \quad \beta \neq Awx, A \in V_\mathrm{N}, \mid w \mid = k \\ \quad \text{или} \mid w \mid < k \text{ и } x = \varepsilon \}, \end{cases}$$

где

RetRe 88 104
t 8Ret
83

Иначе говоря, могут представиться следующие случаи:

(1) $\alpha = c\gamma$, где $c \in V_T$, $\gamma \in (V_N \cup V_T)^*$,

$\mathrm{FIRST}_k^G(c\gamma) \subseteq \mathrm{EFF}_k^G(\alpha)$;

(2) $\alpha = Awx$, где $A \in V_N$, $w, x \in V_T^*$,

$$Awx \underset{rm}{\Rightarrow} \beta wx,\ A \to \beta,$$

а) $\beta \neq \varepsilon$: $\mathrm{FIRST}_k^G(\beta wx) \subseteq \mathrm{EFF}_k^G(\alpha)$;

б) $\beta = \varepsilon$: $\mathrm{FIRST}_k^G(\beta wx) \not\subseteq \mathrm{EFF}_k^G(\alpha)$.

Ясно, что в силу данного определения

$$\mathrm{EFF}_k^G(\varepsilon) = \{\varepsilon\}.$$

104

Функция $\mathrm{EFF}_k^G(\alpha)$ отличается от функции $\mathrm{FIRST}_k^G(\alpha)$ тем, что значение $\mathrm{EFF}_k^G(\alpha)$ *не включает* префиксы терминальных цепочек, выводимых из α, только в случае 2a), тогда как значение $\mathrm{FIRST}_k^G(\alpha)$ включает *все* такие префиксы *без исключения*.

Иными словами, в значение $\mathrm{EFF}_k^G(\alpha)$ входят те терминальные цепочки, выводимые из α правосторонним выводом, в котором последний шаг не использует ε-порождение.

Пример 3.8. Рассмотрим КС-грамматику со следующими правилами:

1) $S \to AB$, 2) $A \to Ba \mid \varepsilon$,

3) $B \to Cb \mid C$, 4) $C \to c \mid \varepsilon$.

Вычислим функцию $\mathrm{EFF}_2^G(S)$. Поскольку аргумент начинается на нетерминал, то согласно определению 3.7 необходимо построить всевозможные правосторонние выводы, начинающиеся с нетерминала S и дающие терминальные цепочки, в которых на последнем шаге не применяется ε-правило.

Пример 3.8.

В искомое множество нужно включить префиксы этих терминальных цепочек длиной 2 символа, а если они короче, то включить их целиком.

Любой вывод имеет единственное начало:

$$S \underset{rm}{\Rightarrow} AB.$$

Любое продолжение даст результат вида:

$$S \underset{rm}{\Rightarrow} AB \overset{*}{\underset{rm}{\Rightarrow}} Aw, \qquad w \in V_{\text{Т}}^{*}$$

где w — некоторая терминальная цепочка.

Далее возможны следующие продолжения:

$Aw \underset{rm}{\Rightarrow}$
- $Baw \underset{rm}{\Rightarrow}$
  - $\underset{rm}{\Rightarrow} Cbaw$
    - $\underset{rm}{\Rightarrow} cbaw$
    - $\underset{rm}{\Rightarrow} baw$ — недопустимо (ε-правило)
  - $\underset{rm}{\Rightarrow} Caw$
    - $\underset{rm}{\Rightarrow} caw$
    - $\underset{rm}{\Rightarrow} aw$ — недопустимо (ε-правило)
- w — недопустимо (ε-правило).

Таким образом, функция

$\mathrm{EFF}_2^G(S)$ = $\{ca,\ cb\}$, тогда как

$\mathrm{FIRST}_2^G(S)$ = $\{\varepsilon, a, b, c, ab, ac, ba, ca, cb\}$.

Действительно,

$S \Rightarrow AB \Rightarrow B \Rightarrow C \Rightarrow \varepsilon$

$S \Rightarrow AB \Rightarrow BaB \Rightarrow CaB \Rightarrow aB \Rightarrow aC \Rightarrow a$

$S \Rightarrow AB \Rightarrow B \Rightarrow Bb \Rightarrow Cb \Rightarrow b$

$S \Rightarrow AB \Rightarrow B \Rightarrow C \Rightarrow c$

…

Теорема 3.1. *Чтобы cfg* $G = (V_N, V_T, P, S)$ *была* $LR(k)$*-грамматикой, необходимо и достаточно, чтобы выполнялось следующее условие:*

если $[A \to \beta., u]$ — $LR(k)$*-ситуация, допустимая для активного префикса* $\alpha\beta$ *расширенной грамматики* G',

то не существует никакой другой $LR(k)$*-ситуации* $[A_1 \to \beta_1.\beta_2, v]$ *для того же активного префикса при условии, что* $u \in \mathrm{EFF}_k^G(\beta_2 v)$.

108 199 221 229 251 254 263

<u>*Доказательство*</u>.

Необходимость. Предположим, что G — $LR(k)$-грамматика, но существуют две ситуации, о которых говорится в условии.

По определению 3.6 $[A \to \beta., u]$ — $LR(k)$-ситуация, допустимая для активного префикса $\alpha\beta$ право-сентенциальной формы $\alpha\beta w$ пополненной грамматики G', если суще-ствует правосторонний вывод вида

1) $S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta w$ и $u \in \mathrm{FIRST}_k^G(w)$.

Ret 90 91 92 94 95 96

Пусть $[A_1 \to \beta_1.\beta_2,\ v]$ — *другая* $LR(k)$-ситуация, допустимая для активного префикса $\alpha\beta$.

Согласно определению $LR(k)$-ситуации, допустимой для активного префикса $\alpha\beta$, существует вывод вида

$$2)\ S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha_1 A_1 x \underset{rm}{\Rightarrow} \overbrace{\alpha_1 \beta_1}^{\alpha\beta} \underline{\underline{\beta_1 \beta_2}} x \underset{rm}{\overset{*}{\Rightarrow}} \alpha\beta y,$$

в котором применено правило $A_1 \to \beta_1\beta_2$, и $\alpha_1\beta_1 = \alpha\beta$, $\beta_2 x \underset{rm}{\overset{*}{\Rightarrow}} y, v \in \mathrm{FIRST}_k^G(x)$.

Кроме того, выполняется условие

$$\boxed{3)\ u \in \mathrm{EFF}_k^G(\beta_2 v).}$$

91

Рассмотрим три возможных варианта состава цепочки β_2:

(1) $\beta_2 = \varepsilon$;

(2) $\beta_2 \in V_{\mathrm{T}}{}^{+}$;

(3) β_2 содержит нетерминалы.

Вариант 1: $\beta_2 = \varepsilon$.

Условие 3) даёт $u \in \mathrm{EFF}_k^G(\beta_2 v) = \mathrm{EFF}_k^G(v)$, и, учитывая, что $v \in V_\mathrm{T}^*, |\, v \,| \le k$, получаем согласно определению функции EFF_k^G, равенство $u = v$.

Соответственно вывод 2) фактически имеет вид 2′) $S' \overset{*}{\underset{rm}{\Rightarrow}} \alpha_1 A_1 x \underset{rm}{\Rightarrow} \alpha_1 \beta_1 x = \alpha\beta y$.

Отсюда $\alpha_1\beta_1 = \alpha\beta$ и $x = y$; соответственно

$$\mathrm{FIRST}_k^G(y) = \mathrm{FIRST}_k^G(x) = \{v\},$$

$$\mathrm{FIRST}_k^G(w) = \{u\}.$$

99

3

Последнее равенство означает то же самое, что $u \in \mathrm{FIRST}_k^G(w)$.

Итак, имеем два право-сторонних вывода:

1) $S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta w$,

2′) $S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha_1 A_1 x \underset{rm}{\Rightarrow} \alpha_1\beta_1 x = \alpha\beta y$,

в которых $\mathrm{FIRST}_k^G(w) = \mathrm{FIRST}_k^G(y)$.

Теорема 3.1. (необходимость)

По предположению

$$[A \to \beta., u] \neq [A_1 \to \beta_1.\beta_2, v],$$

причём, как показано, $uu == v$.

Следовательно, либо $A \neq A_1$,

либо $\beta \neq \beta_1$ (ведь $\beta_2 = \varepsilon$).

Но так как G — $LR(k)$-грамматика, то ***должно*** выполняться равенство (*см*. определение 3.2)

$$\boxed{\alpha_1 A_1 x = \alpha A y,} \qquad (*)$$

в котором, как было показано ранее, $xx == y$.

96

При $A \neq A_1$ равенство (*) невозможно, а при $A = A_1$ и $\beta \neq \beta_1$ из того, что $\alpha_1\beta_1 = \alpha\beta$ заключаем, что $\alpha_1 \neq \alpha$.

В последнем случае условие (*) имеет вид: $\alpha_1 Ax = \alpha Ax$ (ведь $y = x$) и при $\alpha_1 \neq \alpha$ выполнятся не может.

Итак, $LR(k)$-условие (*) не выполняется, и согласно определению G — ***не*** $LR(k)$-грамматика вопреки первоначальному предположению.

Это противоречие доказывает необходимость условия теоремы при варианте 1.

<u>*Вариант 2*</u>: $\beta_2 = z$, $z \in V_T^+$ (β_2 — непустая терминальная цепочка). В этом случае вывод 2) имеет вид

$$2'')\ S' \overset{*}{\underset{rm}{\Rightarrow}} \alpha_1 A_1 x \underset{rm}{\Rightarrow} \alpha_1\beta_1\beta_2 x = \alpha_1\beta_1 z x = \alpha\beta y,$$

в котором $\alpha_1\beta_1 = \alpha\beta$, $y = zx$ и, кроме того, предполагается, что

$$3''')\ u \in \mathrm{EFF}_k^G(\beta_2 x) = \mathrm{EFF}_k^G(zx) = \mathrm{EFF}_k^G(y),$$

т. е. $u \in \mathrm{FIRST}_k^G(y)$, поскольку в этом случае $\mathrm{EFF}_k^G(y) = \mathrm{FIRST}_k^G(y)$ (ведь $y \in V_T^+$).

Напомним, что, кроме того, $u = \mathrm{FIRST}_k^G(w)$ (*см*. <u>вывод 1</u>).

Чтобы грамматика G была $LR(k)$-грамматикой, должно быть $\alpha_1 A_1 x = \alpha A y$ (*) или, что то же самое, $\alpha_1 A_1 x = \alpha A z x$ (ведь $y = zx$), но это невозможно при $z \neq \varepsilon$. Получается, что G — ***не*** $LR(k)$-грамматика, а это противоречит исходному предположению.

Данное противоречие доказывает неправомерность предположения о существовании двух разных $LR(k)$-ситуаций, о которых шла речь по варианту 2.

Вариант 3: цепочка β_2 не пуста и содержит, по крайней мере, один нетерминал. Поскольку нас интересуют только цепочки u, которые участвуют в условии

$$3)\ u \in \mathrm{EFF}_k^G(\beta_2 v),$$

то необходимо рассматривать выводы вида

$$4)\ \beta_2 \underset{rm}{\overset{*}{\Rightarrow}} u_1 B u_3 \underset{rm}{\Rightarrow} u_1 u_2 u_3,\ B \to u_2 \in P,$$

в которых $u_2 \neq \varepsilon$, если $u_1 = \varepsilon$, то есть $u_1 u_2 \neq \varepsilon$.

Итак, имеем два вывода

1) $S' \overset{*}{\underset{rm}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta w$,

и 2) с учётом вывода 4)

$$2')\ S' \overset{*}{\underset{rm}{\Rightarrow}} \alpha_1 A_1 x \underset{rm}{\Rightarrow} \overbrace{\alpha_1\beta_1}^{\alpha\beta}\underline{\underline{\beta_1\beta_2}} x \overset{*}{\underset{rm}{\Rightarrow}} \overbrace{\alpha_1\beta_1}^{\alpha\beta} u_1 B u_3 x \underset{rm}{\Rightarrow}$$

$$\underset{rm}{\Rightarrow} \overbrace{\alpha_1\beta_1}^{\alpha\beta} u_1 \underline{\underline{u_2}} u_3 x = \alpha\beta u_1 u_2 u_3 x.$$

G — $LR(k)$-грамматика. Поэтому должно быть $\alpha A u_1 u_2\, u_3 x = \alpha_1\beta_1 u_1 B\, u_3 x$, или $\alpha \underline{\underline{A u_1 u_2}} = \alpha_1\beta_1 u_1 B = \alpha\underline{\underline{\beta u_1 B}}$ и $A u_1 u_2 = \beta u_1 B$.

Последнее равенство возможно лишь при $u_1 u_2 = \varepsilon$ и $\beta = \varepsilon$, чего нет по варианту 3! — Противоречие!

Рассмотренные варианты состава цепочки β_2 исчерпывающе доказывают необходимость сформулированного условия.

Достаточность. Рассуждая от противного, предположим, что условие теоремы выполнено, но G — ***не*** *LR*(*k*)-грамматика.

Тогда согласно лемме 3.2 существуют два вывода в пополненной грамматике вида

1) $S' \overset{*}{\underset{rm}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta w,$

2) $S' \overset{*}{\underset{rm}{\Rightarrow}} \gamma B x \underset{rm}{\Rightarrow} \gamma\delta x = \alpha\beta y,$ 101

такие, что $x, y, w \in V_{\mathrm{T}}^{\;*}$ и при этом

а) $\mathrm{FIRST}_k^G(w) = \mathrm{FIRST}_k^G(y),$ 103

б) $\gamma B x \neq \alpha A y,$ 110

в) $|\, \gamma\delta \,| \geq |\, \alpha\beta \,|.$

Не ограничивая общности рассуждений, можно считать, что $\alpha\beta$ — одна из самых коротких цепочек, удовлетворяющих описанным условиям.

Представим вывод 2) иначе, выделив в нём явно начальный участок, на котором получается *последняя* цепочка с открытой частью не длиннее $|\,\alpha\beta\,| + 1$:

$$2')\ S' \overset{*}{\underset{rm}{\Rightarrow}} \alpha_1 A_1 x \underset{rm}{\Rightarrow} \alpha_1\beta_1\beta_2 y_1 \overset{*}{\underset{rm}{\Rightarrow}} \alpha_1\beta_1 y = \alpha\beta y.$$ 104

Здесь $|\,\alpha_1 A_1\,| \le |\,\alpha\beta\,| + 1$ или, что то же самое, $|\,\alpha_1\,| \le |\,\alpha\beta\,| \le |\,\gamma\delta\,|$, $A_1 \to \beta_1\beta_2 \in P$.

Цепочка $\beta_1\beta_2$ — основа сентенциальной формы $\alpha_1\beta_1\beta_2\,y_1$, причём β_1 — её префикс такой длины, что выполняется равенство $|\,\alpha_1\beta_1\,| = |\,\alpha\beta\,|$.

Отметим, что активный префикс длины $|\,\alpha\beta\,|$, для которого допустима хотя бы какая-нибудь $LR(k)$-ситуация, не может получиться из сентенциальной формы, открытая часть которой длиннее $|\,\alpha\beta\,| + 1$.

Действительно, если, например, $|\alpha_1 A_1| > |\alpha\beta| + 1$, т. е. $|\alpha_1| > |\alpha\beta|$, то крайний левый символ основы $\beta_1\beta_2$ находился бы в позиции, по меньшей мере, $|\alpha\beta| + 2$, и эта основа не имела бы никакого касательства к префиксу длиной $|\alpha\beta|$ (*см*. рис. 3.3.).

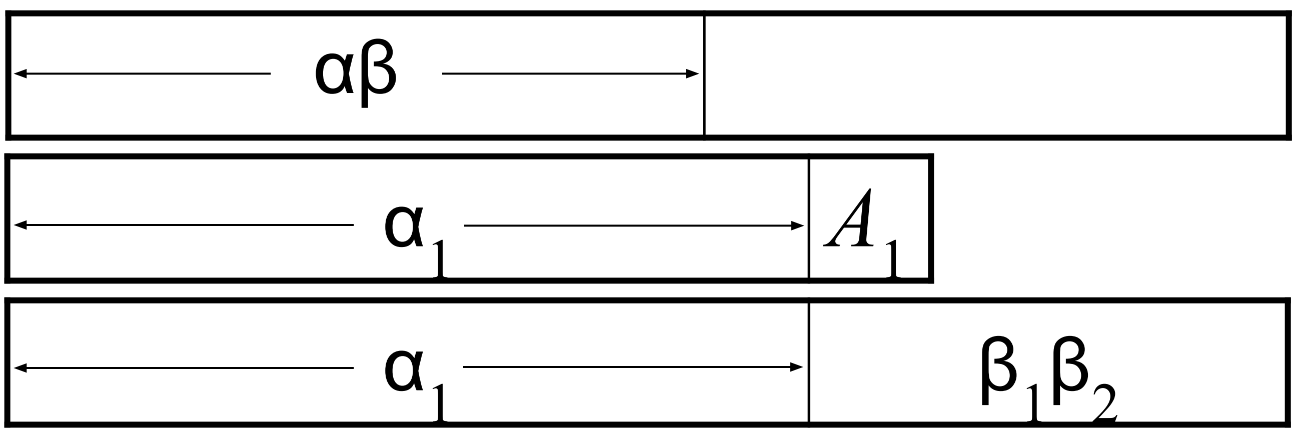

Рис. 3.3.

Из вывода 2′) следует, что $\boxed{\beta_2 y_1 \overset{*}{\underset{rm}{\Rightarrow}} y.}$

Поскольку вывод 2′) правосторонний и $\alpha_1\beta_1 y = \alpha\beta y,$ то $\alpha_1\beta_1 = \alpha\beta,$ и вывод 2′) фактически имеет вид

$$2'')\ S' \overset{*}{\underset{rm}{\Rightarrow}} \alpha_1 A_1 y_1 \underset{rm}{\Rightarrow} \alpha_1\beta_1\beta_2 y_1 = \alpha\beta\beta_2 y_1 \overset{*}{\underset{rm}{\Rightarrow}} \alpha\beta y.$$

105

Равенство а) $\mathrm{FIRST}_k^G(w) = \mathrm{FIRST}_k^G(y)$ означает, что оба множества состоят из одной терминальной цепочки, скажем u, то есть

$$\mathrm{FIRST}_k^G(w) = \mathrm{FIRST}_k^G(y) = \{u\},$$

или, что тоже самое,

$$u \in \mathrm{FIRST}_k^G(w) \text{ и } u \in \mathrm{FIRST}_k^G(y).$$

Из факта существования вывода 1) следует, что $LR(k)$-ситуация $[A \to \beta., u]$, где $u \in \mathrm{FIRST}_k^G(w)$, допустима для активного префикса $\alpha\beta$ право-сентенциальной формы $\alpha\beta w$.

Аналогично из факта существования вывода 2″) следует, что $LR(k)$-ситуация $[A_1 \to \beta_1.\beta_2, v]$, где $v \in \mathrm{FIRST}_k^G(y_1)$, допусти-ма для активного префикса $\alpha\beta$ право-сентенциальной формы $\alpha\beta\beta_2 y_1$.

Учитывая условие а), вывод $\beta_2 y_1 \overset{*}{\underset{\text{rm}}{\Rightarrow}} y$ и $v \in \mathrm{FIRST}_k^G(y_1)$, заключаем, что

$$u \in \mathrm{FIRST}_k^G(w) =$$

$$= \mathrm{FIRST}_k^G(y) \subseteq \mathrm{FIRST}_k^G(\beta_2 y_1) =$$

$$= \mathrm{FIRST}_k^G(\beta_2) \oplus_k \mathrm{FIRST}_k^G(y_1) =$$

$$= \mathrm{FIRST}_k^G(\beta_2) \oplus_k \{v\} =$$

$$= \mathrm{FIRST}_k^G(\beta_2) \oplus_k \mathrm{FIRST}_k^G(v) = \mathrm{FIRST}_k^G(\beta_2 v).$$

Остаётся показать, что $u \in \mathrm{EFF}_k^G(\beta_2 v)$.

Действительно, если бы $u \notin \mathrm{EFF}_k^G(\beta_2 v)$, то только потому, что цепочка β_2 началась бы с нетерминала, который на последнем шаге вывода $\beta_2 y_1 \overset{*}{\underset{\mathrm{rm}}{\Rightarrow}} y$ замещался бы ε-цепочкой.

Сопоставим исходное представление вывода 2) $S' \overset{*}{\underset{rm}{\Rightarrow}} \gamma Bx \underset{rm}{\Rightarrow} \gamma\delta x = \alpha\beta y$

с его же представлением в виде

$$2'')\ S' \overset{*}{\underset{\mathrm{rm}}{\Rightarrow}} \alpha_1 A_1 y_1 \underset{\mathrm{rm}}{\Rightarrow} \alpha_1\beta_1\beta_2 y_1 = \alpha\beta\beta_2 y_1 \overset{*}{\underset{\mathrm{rm}}{\Rightarrow}} \alpha\beta y.$$

Последним замещаемым нетерминалом в этом выводе является B, причём эта-то последняя замена и даёт цепочку $\alpha\beta y$.

На завершающем участке этого вывода используется $\beta_2 y_1 \underset{rm}{\overset{*}{\Rightarrow}} y$, так что, если $\beta_2 = A_2\alpha_2$ и последнее используемое правило есть $B \to \varepsilon$, то вывод 2″) можно переписать следующим образом:

$$2'')\ S' \overset{*}{\underset{rm}{\Rightarrow}} \alpha_1 A_1 y_1 \underset{rm}{\Rightarrow} \alpha_1 \beta_1 \beta_2 y_1 =$$

$$= \alpha_1 \beta_1 A_2 \alpha_2 y_1 \overset{*}{\underset{rm}{\Rightarrow}} \alpha_1 \beta_1 A_2 y_2 y_1 \overset{*}{\underset{rm}{\Rightarrow}}$$

$$\overset{*}{\underset{rm}{\Rightarrow}} \alpha_1 \beta_1 A_m \alpha_m y_{m-1} \ldots y_2 y_1 \overset{*}{\underset{rm}{\Rightarrow}}$$

$$\overset{*}{\underset{rm}{\Rightarrow}} \alpha_1 \beta_1 A_m y_m y_{m-1} \ldots y_2 y_1 =$$

$$= \alpha_1 \beta_1 B y_m y_{m-1} \ldots y_2 y_1 \underset{rm}{\Rightarrow}$$

$$\underset{rm}{\Rightarrow} \alpha_1 \beta_1 y_m y_{m-1} \ldots y_2 y_1 = \alpha\beta y,$$

где $\alpha_1\beta_1 = \alpha\beta$, $A_m = B$, $y_m y_{m-1} \ldots y_2 y_1 = y$.

Отметим, что $|\ \alpha_1\beta_1 B\ | = |\ \alpha\beta\ | + 1$, но это противоречит предположению, что $\alpha_1 A_1 y_1$ — *последняя* цепочка в этом выводе, открытая часть которой имеет длину, не превосходящую величину $|\ \alpha\beta\ | + 1$.

Следовательно, $u \in \mathrm{EFF}_k^G(\beta_2 v)$.

Мы нашли две $LR(k)$-ситуации, допустимые для одного и того же активного префикса $\alpha\beta$: $[A \to \beta.,\ u]$ и $[A_1 \to \beta_1.\beta_2,\ v]$ при том, что $u \in \mathrm{EFF}_k^G(\beta_2 v)$.

Поскольку с самого начала предполагалось, что не существует двух таких *разных* $LR(k)$-ситуаций для активного префикса $\alpha\beta$, то должно быть $[A \to \beta., u] = [A_1 \to \beta_1.\beta_2, v]$.

Из этого равенства следует: $A = A_1$, $\beta = \beta_1$, $\beta_2 = \varepsilon$.

Кроме того, поскольку $\alpha_1\beta_1 = \alpha\beta$, а $\beta_1 = \beta$, то $\alpha_1 = \alpha$. С учётом этого вывод можем переписать так:

$$2'')\ S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha_1 A y_1 \underset{rm}{\Rightarrow} \alpha\beta y_1 = \alpha\beta y,$$

откуда заключаем, что $y_1 = y$.

Не забывая, что это другой вид того же самого вывода 2), заменим цепочку β на A, и получим цепочку $\alpha\beta y$, которая совпадает с предыдущей сентенциальной формой в этом выводе. Это означает нарушение условия б) $\gamma Bx \neq \alpha Ay$.

Данное противоречие — следствие *неправомерного* допущения, что G — не $LR(k)$-грамматика при выполнении условия теоремы.

Следовательно, G — $LR(k)$-грамматика.

Достаточность и вместе с этим и теорема доказаны.

***Определение* 3.8.** Пусть $G = (V_N, V_T, P, S)$ — контекстно-свободная грамматика и $\gamma \in (V_N \cup V_T)^*$ — некоторый её активный префикс. Определим множество $V_k^G(\gamma)$ как множество всех $LR(k)$-ситуаций, допустимых для γ:

$$V_k^G(\gamma) = \{[A \to \beta_1.\beta_2, u] \mid \exists A \to \beta_1\beta_2 \in P;$$

$$\exists\, S' \overset{*}{\underset{rm}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta_1\beta_2 w,\ \gamma = \alpha\beta_1,$$

$$u \in \mathrm{FIRST}_k^G(w)\}.$$

209

Множество $\mathscr{S} = \{\, \mathscr{A} \mid \mathscr{A} = V_k^G(\gamma) \,\}$, где γ — активный префикс G назовем *системой множеств* $LR(k)$*-ситуаций для грамматики* G.

Алгоритм 3.2: вычисление множества $V_k^G(\gamma)$.

Вход: $G = (V_{\text{N}}, V_{\text{T}}, P, S)$ — КС-грамматика,

$\gamma = X_1 X_2 \ldots X_m$, $X_i \in V_{\text{N}} \cup V_{\text{T}}$,

$i = 1, 2, \ldots, m$; $m \geq 0$.

Выход: множество $V_k^G(\gamma)$.

123 139 144 150 210

Метод.

Будем строить последовательность множеств

$$V_k^G(\varepsilon),\ V_k^G(X_1),\ V_k^G(X_1X_2),\ \ldots,\ V_k^G(X_1X_2\ldots X_m).$$

1. Строится множество $V_k^G(\varepsilon)$.

а) Инициализация :

$$V_k^G(\varepsilon) = \{[S \to .\alpha\ ,\ \varepsilon] \mid \exists\, S \to \alpha \in P\}.$$

б) Замыкание $V_k^G(\varepsilon)$, если $[A \to .B\alpha,\ u] \in V_k^G(\varepsilon)$, где $B \in V_{\mathrm{N}}$, и существует $B \to \beta \in P$, то *LR*-ситуация $[B \to .\beta,\ v]$, где $v \in \mathrm{FIRST}_k^G(\alpha u)$, тоже включается в множество $V_k^G(\varepsilon)$.

121 130 145 146 147 148 156 193 147 150 210 222

в) Шаг 1б) повторяется до тех пор, пока во множестве $V_k^G(\varepsilon)$ не будут просмотрены все имеющиеся в нём $LR(k)$-ситуации.

Замыкание завершается за конечное число шагов, так как множества P и $V_\text{Т}^{*k}$ конечны.

2. Строится следующий элемент последовательности.

Пусть множества $V_k^G(X_1 X_2 \ldots X_i)$, где $0 \le i < m$, уже построены. Покажем, как построить следующее множество $V_k^G(X_1 X_2 \ldots X_i X_{i+1})$.

120

а) Инициализация:

если $[A \to \beta_1 . X_{i+1}\beta_2, u] \in V_k^G(X_1 X_2 \ldots X_i)$, то $LR(k)$-ситуация $[A \to \beta_1 X_{i+1} . \beta_2, u]$ включается в множество $V_k^G(X_1 X_2 \ldots X_{i+1})$.

б) Замыкание $V_k^G(X_1 X_2 \ldots X_{i+1})$:

если $\{[A \to \beta_1 . B\beta_2, u] \in V_k^G(X_1 X_2 \ldots X_{i+1})$, где $B \in V_N$, и существует $B \to \beta \in P$, то $[B \to .\beta, v]$, где $v \in \mathrm{FIRST}_k^G(\beta_2 u)$, тоже включается в множество $V_k^G(X_1 X_2 \ldots X_{i+1})$.

Заметим, что именно это значение v наследуется ситуациями на базе ***всех*** правил с B в левых частях.

124 125 115 113 141 141 142 143 151 153 154 155 145 141 146 147 148 149

в) Шаг 2б) повторяется до тех пор, пока во множестве $V_k^G(X_1 X_2 \ldots X_{i+1})$ не будут просмотрены все имеющиеся в нём $LR(k)$-ситуации.

Замыкание завершается за конечное число шагов, так как множества P и $V_{\text{т}}^{*k}$ конечны.

3. Шаг 2 повторять до тех пор, пока $i < m$.

4. Процесс завершается при $i = m$;

$$V_k^G(\gamma) = V_k^G(X_1 X_2 \ldots X_m).$$

Замечание 3.2. Алгоритм 3.2 не требует использования пополненной грамматики.

115
6

***Определение* 3.9.** Пусть $G = (V_N, V_T, P, S)$ — контекстно-свободная грамматика.

На множестве $LR(k)$-ситуаций в этой грамматике определим функцию:

$$\text{GOTO}\ (\mathscr{A}, X) = \mathscr{A}', \text{ где } \mathscr{A} = V_k^G(\gamma)$$ 159

— некоторое множество $LR(k)$-ситуаций, допустимых для активного префикса $\gamma \in (V_N \cup V_T)^*$; $X \in V_N \cup V_T$; $\mathscr{A}' = V_k^G(\gamma X)$.

Очевидно, что эта функция строится попутно с построением множеств $V_k^G(\gamma)$ на шаге 2 алгоритма 3.2:

Функция GOTO ($\mathscr{A}$, X)

если множество $V_k^G(X_1 X_2 \ldots X_i)$ уже построено, то

$$V_k^G(X_1 X_2 \ldots X_i X_{i+1}) = \text{GOTO}\ (V_k^G(X_1 X_2 \ldots X_i), X_{i+1}).$$

Остается ввести лишь обозначения:

$$\gamma = X_1 X_2 \ldots X_i,\ X = X_{i+1},\ \mathscr{A} = V_k^G(X_1 X_2 \ldots X_i),$$
$$\mathscr{A}' = V_k^G(X_1 X_2 \ldots X_i X_{i+1}),$$

чтобы увидеть, как это делается.

Замечание 3.3. Важно отметить, что результат функции GOTO ($\mathscr{A}$, X) = $\mathscr{A}'$, где $\mathscr{A} = V_k^G(X_1 X_2 \ldots X_i)$. зависит не от $X_1 X_2 \ldots X_i$, а от $V_k^G(X_1 X_2 \ldots X_i)$.

Пример 3.9. Рассмотрим пополненную грамматику примера 3.1, содержащую правила: 0) $S' \to S$, 1) $S \to SaSb$, 2) $S \to \varepsilon$.

Построим множества

$$V_1^G(\varepsilon),\ V_1^G(S),\ V_1^G(Sa).$$

1: построение множества $V_1^G(\varepsilon)$:

а) $V_1^G(\varepsilon) = \{[S' \to .S, \varepsilon]\}$.

б) Множество $V_1^G(\varepsilon)$ пополняется ситуациями:

$$[S \to .SaSb, \varepsilon], [S \to ., \varepsilon];$$

и ещё $[S \to .SaSb, a], [S \to ., a]$.

Ret

1Ret

162

Другие шаги алгоритма 3.2 никаких других элементов в множество $V_1^G(\varepsilon)$ не добавляют. Окончательно получаем

$$V_1^G(\varepsilon) = \{[S' \to .S, \varepsilon], [S \to .SaSb, \varepsilon], [S \to ., \varepsilon], [S \to .SaSb, a], [S \to ., a]\}.$$

В сокращенных обозначениях то же самое принято записывать следующим образом:

$$V_1^G(\varepsilon) = \{[S' \to .S, \varepsilon], [S \to .SaSb, \varepsilon \mid a], [S \to ., \varepsilon \mid a]\}.$$

2: построение множества $V_1^G(S)$.

а) $V_1^G(S) = \{[S' \to S., \varepsilon], [S \to S.aSb, \varepsilon \mid a]\}$.

Так как точка ни в одной из этих ситуаций не стоит перед нетерминалом, то шаг б) не выполняется ни разу.

Попутно мы вычислили

$$\mathrm{GOTO}\,(V_1^G(\varepsilon), S) = V_1^G(S).$$

3: построение множества $V_1^G(Sa)$.

а) $V_1^G(Sa) = \{[S \to Sa.Sb, \varepsilon \mid a]\}$.

б) Множество $V_1^G(Sa)$ пополняется ситуациями

$$[S \to .SaSb,\ b] \text{ и } [S \to .,\ b];$$

и ещё $[S \to .SaSb,\ a]$ и $[S \to .,\ a]$.

Здесь шаг б) замыкания выполнялся дважды.

Итак,

$$V_1^G(Sa) = \{[S \to Sa.Sb,\ \varepsilon \mid a],\ [S \to .SaSb,\ a \mid b],\ [S \to .,\ a \mid b]\} \text{ и}$$

$$\text{GOTO}\ (V_1^G(S),\ a) = V_1^G(Sa).$$

Теорема 3.2. Алгоритм 3.2 правильно вычисляет $V_k^G(\gamma)$, где $\gamma \in (V_N \cup V_T)^*$ — актив-ный префикс право-сентенциальной формы в грамматике G.

Доказательство. Фактически требуется доказать, что $LR(k)$-ситуация

$$[A \to \beta_1.\beta_2, u] \in V_k^G(\gamma)$$

тогда и только тогда, когда существует правосторонний вывод вида

171 117 172 239
76 8

$$S \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta_1\beta_2 w,$$

в котором $\alpha\beta_1 = \gamma$ (γ — активный префикс право-сентенциальной формы $\alpha\beta_1\beta_2 w$), а

$$u \in \mathrm{FIRST}_k^G(w)$$

(цепочка w есть правый контекст основы $\beta_1\beta_2$ в данной сентенциальной форме $\alpha\beta_1\beta_2 w$).

239

Индукция по $l = |\gamma|$.

База. Пусть $l = 0$, т. е. $\gamma = \varepsilon$.

Имеем вывод

$$S \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta_1\beta_2 w,\ \gamma = \alpha\beta_1 = \varepsilon.$$

Фактически $\alpha = \beta_1 = \varepsilon$, и вывод имеет вид

$$S \underset{rm}{\overset{*}{\Rightarrow}} A w \underset{rm}{\Rightarrow} \beta_2 w,$$

где на последнем шаге применялось правило $A \to \beta_2$.

Во всех деталях этот вывод мог бы быть только таким:

$$S \underset{rm}{\Rightarrow} A_1\alpha_1 \overset{*}{\underset{rm}{\Rightarrow}} A_1 w_1 \underset{rm}{\Rightarrow} A_2\alpha_2 w_1 \underset{rm}{\Rightarrow} A_2 w_2 w_1 \underset{rm}{\Rightarrow} \ldots$$

$$\underset{rm}{\Rightarrow} A_m\alpha_m w_{m-1}\ldots w_2 w_1 \overset{*}{\underset{rm}{\Rightarrow}} A_m \underbrace{w_m w_{m-1}\ldots w_2 w_1}_{w} =$$

$$= Aw = \beta_2 w.$$

Следовательно, $w = w_m w_{m-1}\ldots w_2 w_1$, $A = A_m$ и существуют правила

$S \to A_1\alpha_1$, $A_1 \to A_2\alpha_2, \ldots, A_{m-1} \to A_m\alpha_m = A\alpha_m$, $A \to \beta_2$.

Кроме того, $\mathrm{FIRST}_k^G(w_i) \subseteq \mathrm{FIRST}_k^G(\alpha_i)$, поскольку $\alpha_i \overset{*}{\underset{rm}{\Rightarrow}} w_i$, $i = 1, 2, \ldots, m$.

Согласно шагу **1** алгоритма 3.2

$$[S \to .A_1\alpha_1, \varepsilon] \in V_k^G(\varepsilon),$$

$[A_1 \to .A_2\alpha_2, v_1] \in V_k^G(\varepsilon)$ для любой

$v_1 \in \mathrm{FIRST}_k^G(\alpha_1\varepsilon)$, в частности для

$$v_1 \in \mathrm{FIRST}_k^G(w_1);$$

$[A_2 \to .A_3\alpha_3, v_2] \in V_k^G(\varepsilon)$ для любой

$v_2 \in \mathrm{FIRST}_k^G(\alpha_2 v_1)$, в частности для

$$v_2 \in \mathrm{FIRST}_k^G(w_2 v_1) = \mathrm{FIRST}_k^G(w_2 w_1);$$

…

$[A_{m-1} \to A_m\alpha_m, v_{m-1}] \in V_k^G(\varepsilon)$ для любой $v_{m-1} \in \mathrm{FIRST}_k^G(\alpha_{m-1}v_{m-2})$, в частности для $v_{m-1} \in \mathrm{FIRST}_k^G(w_{m-1}v_{m-2}) = \mathrm{FIRST}_k^G(w_{m-1}...w_1)$.

Наконец, поскольку $A_m = A$ и имеется правило $A \to \beta_2$, $[A \to .\beta_2, v_m] \in V_k^G(\varepsilon)$ для любой

$v_m \in \mathrm{FIRST}_k^G(\alpha_m v_{m-1})$, в частности для

$v_m \in \mathrm{FIRST}_k^G(w_m v_{m-1}) = \mathrm{FIRST}_k^G(w_m ...w_1)$.

Принимая во внимание, что $w = w_m w_{m-1} \ldots w_1$ и что $u \in \mathrm{FIRST}_k^G(w)$, заключаем, что $v_m = u$ и $[A \to .\beta_2, u] \in V_k^G(\varepsilon)$.

База доказана.

Индукционная гипотеза. Предположим, что утверждение выполняется для всех активных префиксов γ, таких, что $|\gamma| \le n$ $(n \ge 0)$.

Индукционный переход. Покажем, что утверждение выполняется для γ, таких, что $|\gamma| \le n + 1$.

Пусть $\gamma = \gamma' X$, где $|\gamma'| = n$, а $X \in V_N \cup V_T$. Поскольку γ — активный префикс, то существует вывод такой сентенциальной формы, где γ участвует в этой роли:

$$S \overset{*}{\underset{rm}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta_1\beta_2 w = \gamma\beta_2 w = \gamma' X \beta_2 w,$$

здесь на последнем шаге применялось правило $A \to \beta_1\beta_2$, и $\alpha\beta_1 = \gamma = \gamma' X$.

<u>Случай 1</u>: $\beta_1 \neq \varepsilon$.

Поскольку $\alpha\beta_1 = \gamma' X$ и $\beta_1 \neq \varepsilon$, то именно β_1 заканчивается символом X, т. е. $\beta_1 = \beta_1' X$ при некоторой $\beta_1' \in (V_{\mathrm{N}} \cup V_{\mathrm{T}})^*$.

В этом случае имеем

$$S \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta_1\beta_2 w = \alpha\beta_1' X \beta_2 w = \gamma' X \beta_2 w,$$

где $\gamma' = \alpha\beta_1'$, $\;|\,\gamma'\,| = n$.

В соответствии с индукционным предположением, поскольку γ' является активным префиксом последней сентенциальной формы и $|\gamma'| = n$, $[A \to \beta_1'.X\beta_2, u] \in V_k^G(\gamma')$, где $u \in \mathrm{FIRST}_k^G(w)$, то

Шаг 2а алгоритма 3.2 даёт

$$[A \to \beta_1' X.\beta_2, u] =$$

$$[A \to \beta_1.\beta_2, u] \in V_k^G(\gamma' X) = V_k^G(\gamma),$$

т. е. $[A \to \beta_1.\beta_2, u] \in V_k^G(\gamma)$.

Случай 1 доказан.

Случай 2: $\beta_1 = \varepsilon$.

Имеем

$$S \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta_1\beta_2 w = \alpha\beta_2 w,$$

причём на последнем шаге вывода применено правило $A \to \beta_2 \in P$, а $\gamma = \alpha = \gamma' X$ и $|\gamma| = n + 1$, т. е. $\gamma' \in (V_N \cup V_T)^*$, $|\gamma'| = n$, $X \in V_N \cup V_T$ или $\mathrm{FIRST}_k^G(w) = \{u\}$

$u \in \mathrm{FIRST}_k^G(w)$.

Согласно определению 3.6, $[A \to .\beta_2, u]$ — $LR(k)$-ситуация, допустимая для активного префикса γ.

Надо показать, что $LR(k)$-ситуация

$$[A \to .\beta_2, u] \in V_k^G(\gamma),$$

где $V_k^G(\gamma)$ — множество вычисленное посредством алгоритма 3.2.

Рассмотрим подробнее этот вывод, чтобы показать, как впервые появляется символ X, завершающий цепочку α, и как образуется право-сентенциальная форма $\alpha A w$.

В общем случае он имеет следующий вид:

$$S \underset{rm}{\overset{*}{\Rightarrow}} \alpha_1 A_1 w_1 \qquad (\exists\ A_1 \to \alpha_2' X A_2 \delta_2 \in P,),$$

$$\underset{rm}{\Rightarrow} \underbrace{\overbrace{\alpha_1\alpha_2'}^{\gamma'} X}_{\gamma} A_2 \delta_2 w_1 \qquad (\delta_2 \underset{rm}{\overset{*}{\Rightarrow}} w_2,\ \alpha_1\alpha_2' = \gamma'),$$

$$\underset{rm}{\overset{*}{\Rightarrow}} \gamma' X A_2 w_2 w_1 \qquad (\exists\ A_2 \to A_3 \delta_1 \in P),$$

$$\ldots \qquad (\exists\ A_{m-2} \to A_{m-1} \delta_{m-1} \in P),$$

$$\underset{rm}{\Rightarrow} \gamma' X A_{m-1} \delta_{m-1} w_{m-2} \ldots w_1 \qquad (\delta_{m-1} \underset{rm}{\overset{*}{\Rightarrow}} w_{m-1}),$$

$$\overset{*}{\underset{rm}{\Rightarrow}} \gamma' X A_{m-1} w_{m-1} w_{m-2} \ldots w_1$$

$(\exists\ A_{m-1} \to A_m \delta_m \in P)$,

$$\underset{rm}{\Rightarrow} \gamma' X A_m \delta_m w_{m-1} w_{m-2} \ldots w_1$$

$(\delta_m \overset{*}{\underset{rm}{\Rightarrow}} w_m)$,

$$\overset{*}{\underset{rm}{\Rightarrow}} \gamma' X A_m w_m w_{m-1} w_{m-2} \ldots w_1$$

$(A_m = A,\ w_m w_{m-1} \ldots w_1 = w)$,

$$= \gamma' X A w$$

$(\exists\ A \to \beta_2 \in P,\ \gamma = \gamma' X)$.

$$\underset{rm}{\Rightarrow} \gamma' X \beta_2 w = \gamma \beta_2 w.$$

Отметим, что в сентенциальной форме $\alpha_1\alpha_2'XA_2\delta_2w_1$ за префиксом $\gamma = \alpha_1\alpha_2'X$ может следовать только нетерминал, ибо иначе основа β_2 ***не*** могла бы появиться в рассматриваемом выводе непосредственно за этим префиксом, и $LR(k)$-ситуация $[A \to .\beta_2, u]$ не была бы допустима для префикса γ.

По определению $[A_1 \to \alpha_2'.XA_2\delta_2, v_1]$, где $v_1 \in \mathrm{FIRST}_k^G(w_1)$, есть $LR(k)$-ситуация, допустимая для активного префикса γ'.

Если $\gamma = \alpha\beta_1$& $S \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta_1\beta_2 w$, то $[A \to \beta_1.\beta_2, u] \in V_k^G(\gamma)$

Поскольку $|\gamma'| = n$, то в соответствии с индукционной гипотезой можно утверждать, что $[A_1 \to \alpha_2'.XA_2\delta_2, v_1] \in V_k^G(\gamma')$,

а тогда согласно шагу 2а алгоритма 3.2 $LR(k)$-ситуация

$$[A_1 \to \alpha_2'X.A_2\delta_2, v_1] \in V_k^G(\gamma'X) = V_k^G(\gamma).$$

Поскольку существуют правило $A_2 \to A_3\delta_3$ и вывод $\delta_2 \underset{rm}{\overset{*}{\Rightarrow}} w_2$, то согласно шагу 2б

$[A_2 . A_3\delta_3, v_2] \in V_k^G(\gamma)$, где

$$v_2 \in \mathrm{FIRST}_k^G(w_2 v_1) = \mathrm{FIRST}_k^G(w_2 w_1).$$

Рассуждая далее аналогичным образом, приходим к выводу, что

$[A_{m-1} \to .\, A_m \delta_m, v_{m-1}] \in V_k^G(\gamma)$, где

$$v_{m-1} \in \mathrm{FIRST}_k^G(w_{m-1} v_{m-2}) = \mathrm{FIRST}_k^G(w_{m-1} \ldots w_1).$$

Наконец, согласно шагу 2б, поскольку $A_m = A$, и существуют правило $A \to \beta_2$ и вывод $\delta_m \underset{rm}{\overset{*}{\Rightarrow}} w_m$, то $LR(k)$-ситуация

$[A \to .\beta_2, v_m] \in V_k^G(\gamma)$, где

$$v_m = \mathrm{FIRST}_k^G(w_m v_{m-1}) = \mathrm{FIRST}_k^G(w_m w_{m-1} \ldots w_1) =$$
$$= \mathrm{FIRST}_k^G(w) = u.$$

Итак, $[A \to .\beta_2, u] \in V_k^G(\gamma)$.

Случай 2 и утверждение I доказаны .

II. Докажем теперь, что

если $[A \to \beta_1.\beta_2, u] \in V_k^G(\gamma)$,

где $V_k^G(\gamma)$ —результат алгоритма 3.2, то существует право-сторонний вывод вида

$$S \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta_1\beta_2 w,$$

в котором $\alpha\beta_1 = \gamma$ есть активный префикс право-сентенциальной формы $\alpha\beta_1\beta_2 w$, а $u \in \mathrm{FIRST}_k^G(w)$ есть правый контекст основы $\beta_1\beta_2$ в данной сентенциальной форме $\alpha\beta_1\beta_2 w$.

Индукция по $l = |\gamma|$.

База. Пусть $l = 0$, т. е. $\gamma = \varepsilon$.

В этом случае $\alpha\beta_1 = \gamma = \varepsilon$, следовательно, $\alpha = \varepsilon$, $\beta_1 = \varepsilon$, и надо доказать существование вывода вида $S \underset{rm}{\overset{*}{\Rightarrow}} Aw \underset{rm}{\Rightarrow} \beta_2 w$, в котором на последнем шаге применяется правило $A \to \beta_2$, а $u \in \mathrm{FIRST}_k^G(w)$.

Имеем $[A \to .\beta_2, u] \in V_k^G(\varepsilon)$.

Все $LR(k)$-ситуации из множества $V_k^G(\varepsilon)$ согласно алгоритму 3.2 получаются на шаге 1а или 1б.

В общем случае история попадания данной $LR(k)$-ситуации в множество $V_k^G(\varepsilon)$ такова:

$$[S \to .\alpha_1, \varepsilon] \in V_k^G(\varepsilon)$$

благодаря шагу 1а и правилу $S \to \alpha_1 \in P$;

$\alpha_1 = A_1\delta_1$, $\exists\; A_1 \to \alpha_2 \in P$, и шаг, и шаг 1, и шаг 1б, и шаг 1б даёт

$$[A_1 \to .\alpha_2, v_1] \in V_k^G(\varepsilon),$$

где $v_1 \in \mathrm{FIRST}_k^G(\delta_1)$, и если $\delta_1 \overset{*}{\underset{rm}{\Rightarrow}} w_1$,

то $v_1 \in \mathrm{FIRST}_k^G(w_1)$;

Далее $\alpha_2 = A_2\delta_2$, $\exists\ A_2 \to \alpha_3 \in P$, и

шагшаг 1шаг 1б даёт $[A_2 \to .\alpha_3, v_2] \in V_k^G(\varepsilon)$,

где $v_2 \in \mathrm{FIRST}_k^G(\delta_2 v_1)$, и если $\delta_2 \overset{*}{\underset{rm}{\Rightarrow}} w_2$,

то $v_2 \in \mathrm{FIRST}_k^G(w_2 v_1) = \mathrm{FIRST}_k^G(w_2 w_1)$;

…

Наконец, $\alpha_m = A_m\delta_m$, $\exists\ A_m \to \alpha_{m+1} \in P$, и шагшаг 1шаг 1б даёт $[A_m \to .\alpha_{m+1}, v_m] \in V_k^G(\varepsilon)$,

и если $\delta_m \overset{*}{\underset{rm}{\Rightarrow}} w_m$, то

$$v_m \in \mathrm{FIRST}_k^G(\delta_m v_{m-1}) = \mathrm{FIRST}_k^G(w_m w_{m-1} \ldots w_1) =$$

$$= \mathrm{FIRST}_k^G(w) = \{u\}.$$

При этом

$A_m = A$, $\alpha_{m+1} = \beta_2$, $w_m w_{m-1} \ldots w_1 = w$.

Используя существующие правила и упомянутые частичные выводы, построим требуемый вывод:

$$S \underset{rm}{\Rightarrow} \alpha_1 = A_1\delta_1 \underset{rm}{\overset{*}{\Rightarrow}} A_1 w_1 \underset{rm}{\Rightarrow} \alpha_2 w_1 = A_2\delta_2 w_1 \underset{rm}{\overset{*}{\Rightarrow}}$$

$$\underset{rm}{\overset{*}{\Rightarrow}} A_2 w_2 w_1 \underset{rm}{\Rightarrow} \ldots \underset{rm}{\Rightarrow} A_m\delta_m w_{m-1}\ldots w_2 w_1 \underset{rm}{\overset{*}{\Rightarrow}}$$

$$\underset{rm}{\overset{*}{\Rightarrow}} A_m w_m w_{m-1}\ldots w_2 w_1 \underset{rm}{\Rightarrow} \alpha_{m+1} w_m w_{m-1}\ldots w_2 w_1 =$$

$$= \beta_2 w.$$

Итак, существует вывод $S \underset{rm}{\overset{*}{\Rightarrow}} A w \underset{rm}{\Rightarrow} \beta_2 w$ и

$$u \in \mathrm{FIRST}_k^G(w) = \mathrm{FIRST}_k^G(w_m \ldots w_1).$$

База доказана.

Индукционная гипотеза. Предположим, что утверждение выполняется для всех γ, длина которых не превосходит n ($n \geq 0$).

Индукционный переход. Покажем, что утверждение выполняется для $\gamma = \gamma' X$, где $|\gamma'| = n$, $X \in V_{\mathrm{N}} \cup V_{\mathrm{T}}$.

Согласно алгоритму 3.2 $LR(k)$-ситуация

$$[A \rightarrow \beta_1.\beta_2, u] \in V_k^G(\gamma)$$

в общем случае может быть получена следующим образом:

1. Существует $LR(k)$-ситуация

$$[A_1 \to \alpha_1 .\, X\alpha_2, v_1] \in V_k^G(\gamma'),$$

допустимая для активного префикса γ'.

2. Посредством шага 2а алгоритма 3.2 получается $LR(k)$-ситуация

$$[A_1 \to \alpha_1 X.\alpha_2, v_1] \in V_k^G(\gamma' X) = V_k^G(\gamma).$$

<u>Случай 1</u>: $[A_1 \to \alpha_1 X.\alpha_2, v_1] = [A \to \beta_1.\beta_2, u]$, т. е. рассматриваемая $LR(k)$-ситуация получена на этом шаге алгоритма.

Из того, что $[A_1 \to \alpha_1 . X\alpha_2, v_1] \in V_k^G(\gamma')$, в соответствии с индукционным предположением следует существование вывода вида

$$S \xRightarrow[rm]{*} \alpha_0 A_1 w_1 \xRightarrow[rm]{} \rlap{$\underbrace{}_{\gamma'}$}\alpha_0\overbrace{\alpha_1 X \alpha_2}^{\beta} w_1 = \underbrace{\gamma' X}_{\gamma}\beta_2 w_1 = \gamma\beta_2 w_1,$$

причём $u \in \mathrm{FIRST}_k^G(w_1) = \{v_1\}$,

$$A = A_1,\ \beta = \beta_1\beta_2 = \alpha_1 X\alpha_2,\ \beta_1 = \alpha_1 X,\ \beta_2 = \alpha_2.$$

Другими словами, $[A \to \beta_1.\beta_2, u]$ — $LR(k)$-ситуация, допустимая для активного префикса γ.

<u>Случай 2:</u> данная $LR(k)$-ситуация получена посредством замыкания

$$[A_1 \to \alpha_1 X.\alpha_2,\, v_1] \in V_k^G(\gamma).$$

Это значит, что согласно алгоритму 3.2 данная $LR(k)$-ситуация $[A \to \beta_1.\beta_2,\ u]$ получается посредством нескольких шагов <u>2б</u>, производящих последовательность $LR(k)$-ситуаций следующим образом:

Пусть $\alpha_2 = A_2\delta_2$, $\exists\ A_2 \to \alpha_3 \in P$, и шаг, и шаг 2, и шаг 2б, и шаг 2б даёт $[A_2 \to .\alpha_3, v_2] \in V_k^G(\gamma)$,

где $v_2 \in \mathrm{FIRST}_k^G(\delta_2 v_1)$, и если $\delta_2 \underset{rm}{\overset{*}{\Rightarrow}} w_2$,

то $v_2 \in \mathrm{FIRST}_k^G(w_2 v_1)$;

Аналогично, если $\alpha_3 = A_3\delta_3$, $\exists\ A_3 \to \alpha_4 \in P$, и шаг, и шаг 2, и шаг 2б даёт $[A_3 \to .\alpha_4, v_3] \in V_k^G(\gamma)$,

где $v_3 \in \mathrm{FIRST}_k^G(\delta_3 v_2)$, и если $\delta_3 \underset{rm}{\overset{*}{\Rightarrow}} w_3$,

то $v_3 \in \mathrm{FIRST}_k^G(w_3 v_2) = \mathrm{FIRST}_k^G(w_3 w_2 v_1)$;

…

Наконец, если $\alpha_m = A_m\delta_m$, $\exists\; A_m \to \alpha_{m+1} \in P$, и шаг, и шаг 2, и шаг 2б даёт

$$[A_m \to .\alpha_{m+1}, v_m] = [A \to \beta_1.\beta_2, u] \in V_k^G(\gamma),$$

где $v_m \in \mathrm{FIRST}_k^G(\delta_m v_{m-1})$, и если $\delta_m \underset{rm}{\overset{*}{\Rightarrow}} w_m$, то $v_m \in \mathrm{FIRST}_k^G(w_m v_{m-1}) = \mathrm{FIRST}_k^G(w_m \ldots w_2 v_1)$.

Кроме того, из равенства двух $LR(k)$-ситуаций следует $A_m = A$, $\beta_1 = \varepsilon$, $\alpha_{m+1} = \beta_2$, $v_m = u$; также существует правило $A \to \beta_2$.

Извлечём теперь полезные следствия из вышеизложенной истории.

Из п.1 алгоритма 3.2 согласно индукционной гипотезе следует существование вывода вида

$$S \underset{rm}{\overset{*}{\Rightarrow}} \alpha_0 A_1 w_1 \underset{rm}{\Rightarrow} \alpha_0 \alpha_1 X \alpha_2 w_1 = \gamma' X \alpha_2 w_1 = \gamma A_1 \delta_2 w_1,$$

в котором $\mathrm{FIRST}_k^G(w_1) = \{v_1\}$.

Благодаря п.2**2** в алгоритма 3.2 этот вывод может быть продолжен следующим образом:

$$\overset{*}{\underset{rm}{\Rightarrow}} \gamma A_1 w_2 w_1 \underset{rm}{\Rightarrow} \gamma \alpha_3 w_2 w_1 = \gamma A_3 \delta_3 w_2 w_1 \overset{*}{\underset{rm}{\Rightarrow}}$$

$$\overset{*}{\underset{rm}{\Rightarrow}} \gamma A_3 w_3 w_2 w_1 \overset{*}{\underset{rm}{\Rightarrow}} \gamma A_m w_m \ldots w_2 w_1 = \gamma A w \underset{rm}{\Rightarrow}$$

$$\underset{rm}{\Rightarrow} \gamma\beta_2 w,\ \ w = w_m \ldots w_1 \quad \mathrm{FIRST}_k^G(w) = \{v_1\} = \{u\}.$$

Учитывая вышеприведённый вывод, заключаем, что согласно определению $[A \to \beta_1.\beta_2, u]$ есть $LR(k)$-ситуация, допусти-мая для активного префикса γ.

Утверждение II доказано.

Из рассуждений I и II следует справедливость теоремы.

Алгоритм 3.3: построение системы множеств допустимых $LR(k)$-ситуаций для данной КС-грамматики.

Вход: $G = (V_N, V_T, P, S)$ — контекстно-свободная грамматика.

Выход: $\mathscr{S} = \{\mathscr{A} \mid \mathscr{A} = V_k^G(\gamma)$, где γ —активный префикс $G\}$.

Метод. Вначале система $\mathscr{S}$ пуста.

1. Построить множество $\mathscr{A}_0 = V_k^G(\varepsilon)$, и поместить его в качестве элемента $\mathscr{S}$ как непомеченное множество.

174 176 179 173 210

2. Пусть $\mathscr{A} \in \mathscr{S}$ и $\mathscr{A}$ — не помечено. Пометить $\mathscr{A}$ и построить множество $\mathscr{A}' =$ GOTO ($\mathscr{A}$, X) для всех $X \in V_{\mathrm{N}} \cup V_{\mathrm{T}}$. Если $\mathscr{A}' \neq \varnothing$ и $\mathscr{A}' \notin \mathscr{S}$, то включить $\mathscr{A}'$ в $\mathscr{S}$ как непомеченное множество.

3. Повторять шаг 2 до тех пор, пока все множества $LR(k)$-ситуаций в $\mathscr{S}$ не будут помечены.

177 181 17175

Момент, когда в все элементы множества $\mathcal{S}$ окажутся помеченными, обязательно наступит, так как число правил грамматики G конечно, число позиций в них конечно, число терминальных цепочек, длина которых не превосходит k, конечно и, соответственно, число $LR(k)$-ситуаций, допустимых для грамматики G, тоже конечно.

4. Полученное множество $\mathcal{S}$ является искомым.

***Определение* 3.10.** Если G — контекстно-свободная грамматика, то систему множеств допустимых $LR(k)$-ситуаций для пополненной грамматики G' будем называть *канонической системой множеств* $LR(k)$*-ситуаций* для грамматики G.

Замечание 3.4. Множество GOTO ($\mathcal{A}$, S') никогда не потребуется вычислять, так как оно всегда пусто[1].

---

[1] S' не встречается в правых частях правил.

Пример 3.10. Рассмотрим ещё раз пополненную грамматику из примера 3.1, содержащую следующие правила:

0) $S' \to S$, 1) $S \to SaSb$, 2) $S \to \varepsilon$, и

проиллюстрируем на ней алгоритм 3.3.

Как положено, начинаем с построения $V_1^G(\varepsilon)$ (*см*. пример. пример 3.9).

57 115 118 206 228 229

Пример 3.10. Каноническое множество $LR(1)$-ситуаций

Посторение $V_1^G(\varepsilon)$:

$$[S' \to .S\varepsilon,\ \varepsilon] \quad \mathrm{FIRST}_1^G(\varepsilon)\oplus_1\{\varepsilon\} = \{\varepsilon\}$$

$$[S \to .SaSb,\ \varepsilon] \quad [S \to .,\ \varepsilon] \quad \mathrm{FIRST}_1^G(aSb)\oplus_1\{\varepsilon\} = \{a\}$$

$$[S \to .SaSb,\ a] \quad [S \to .,\ a]$$

$$\mathscr{A}_0 = \{[S' \to .S,\ \varepsilon],$$
$$[S \to .SaSb,\ \varepsilon], [S \to .,\ \varepsilon],$$
$$[S \to .SaSb,\ a], [S \to .,\ a]\};$$

Переходы из $\mathscr{A}_0$:

Продвинуть точку в следующую позицию в каждой $LR(k)$-ситуации из $\mathscr{A}_0$.

$$\mathscr{A}_1 = \mathrm{GOTO}(\mathscr{A}_0, S) =$$
$$= \{[S' \to S., \varepsilon],$$
$$[S \to S.aSb, \varepsilon],$$
$$[S \to S.aSb, a]\};$$

$$\mathrm{GOTO}(\mathscr{A}_0, a) = \mathrm{GOTO}(\mathscr{A}_0, b) = \varnothing;$$

Поскольку за позиционной точкой в каждой ситуации не следует нетерминал, то замыкание не требуется.

Переходы из $\mathcal{A}_1$:

$$\mathcal{A}_2 = \mathrm{GOTO}(\mathcal{A}_1, a) =$$

$$= \{(1)\ [S \rightarrow Sa.Sb,\ \varepsilon], (2)\ [S \rightarrow Sa.Sb,\ a], \quad \text{Сдвиг}$$

$$[S \rightarrow .SaSb,\ b], [S \rightarrow .,\ b],$$

$$[S \rightarrow .SaSb,\ a], [S \rightarrow .,\ a]\ \}; \quad \text{Замыкание (1)}$$

Замыкание ситуации (2) ничего нового не даёт!

$\mathrm{GOTO}(\mathcal{A}_1, b) = \varnothing$; — точки перед '$b$' в $\mathcal{A}_1$ нет.

Переходы из $\mathcal{A}_2$:

$$\mathcal{A}_3 = \mathrm{GOTO}(\mathcal{A}_2, S) =$$
$$= \{[S \to SaS.b, \varepsilon], [S \to SaS.b, a],$$
$$[S \to S.aSb, b], [S \to S.aSb, a]\};$$

Поскольку за позиционной точкой в каждой ситуации из $\mathcal{A}_3$ не следует нетерминал, то замыкание не требуется.

$$\mathrm{GOTO}(\mathcal{A}_2, a) = \mathrm{GOTO}(\mathcal{A}_2, b) = \varnothing;$$

Пример 3.10. Каноническое множество $LR(1)$-ситуаций

Переходы из $\mathcal{A}_3$:

$$\mathrm{GOTO}(\mathcal{A}_3, S) = \varnothing;$$

$$\mathcal{A}_4 = \mathrm{GOTO}(\mathcal{A}_3, a) =$$

$$= \{[S \to Sa.Sb, b], [S \to Sa.Sb, a], \quad \text{Сдвиг}$$

$$[S \to .SaSb, b], [S \to ., b],$$

$$[S \to .SaSb, a], [S \to ., a]\}; \quad \text{Замыкание}$$

$$\mathcal{A}_5 = \mathrm{GOTO}(\mathcal{A}_3, b) =$$

$$= \{[S \to SaSb., \varepsilon], [S \to SaSb., a]\}; \quad \text{Сдвиг}$$

Поскольку за позиционной точкой в каждой ситуации из $\mathcal{A}_5$ не следует нетерминал, то замыкание не требуется.

Переходы из $\mathscr{A}_4$:

$$\mathscr{A}_6 = \mathrm{GOTO}(\mathscr{A}_4, S) =$$

$$= \{[S \to SaS.b,\, b], [S \to SaS.b,\, a],$$

$$[S \to S.aSb,\, b], [S \to S.aSb,\, a]\};$$

Поскольку за позиционной точкой в каждой ситуации из $\mathscr{A}_6$ не следует нетерминал, то замыкание не требуется.

$$\mathrm{GOTO}(\mathscr{A}_4, a) = \mathrm{GOTO}(\mathscr{A}_4, b) = \varnothing;$$

Переходы из $\mathcal{A}_5$:

$$\mathrm{GOTO}(\mathcal{A}_5, S) = \mathrm{GOTO}(\mathcal{A}_5, a) = \mathrm{GOTO}(\mathcal{A}_5, b) = \varnothing;$$

Переходы из $\mathcal{A}_6$:

$$\mathrm{GOTO}(\mathcal{A}_6, S) = \varnothing;$$

$$\mathrm{GOTO}(\mathcal{A}_6, a) =$$

$$= \{[S \to Sa.Sb, b], [S \to Sa.Sb, a], \quad \text{Сдвиг}$$

$$[S \to .SaSb, b], [S \to ., b],$$

$$[S \to .SaSb, a], [S \to ., a]\} = \mathcal{A}_4; \quad \text{Замыкание}$$

$$\mathcal{A}_7 = \mathrm{GOTO}(\mathcal{A}_6, b) =$$

$$= \{[S \to SaSb., b], [S \to SaSb., a]\};$$

Переходы из $\mathcal{A}_7$:

$$\mathrm{GOTO}(\mathcal{A}_7, S) = \mathrm{GOTO}(\mathcal{A}_7, a) = \mathrm{GOTO}(\mathcal{A}_7, b) = \varnothing.$$

Таким образом $\{\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5, \mathcal{A}_6, \mathcal{A}_7\}$ — каноническая система множеств $LR(1)$-ситуаций для грамматики G.

Табл. 3.4 представляет функцию $\mathrm{GOTO}(\mathcal{A}, X)$ для грамматики G. Заметим, что с точностью до обозначений она совпадает с частью $g(X)$ табл. 3.3.

Таб. 3.4

206

$\mathscr{A}$	X		
	S	a	b
$\mathscr{A}_0$	$\mathscr{A}_1$		
$\mathscr{A}_1$		$\mathscr{A}_2$	
$\mathscr{A}_2$	$\mathscr{A}_3$		
$\mathscr{A}_3$		$\mathscr{A}_4$	$\mathscr{A}_5$
$\mathscr{A}_4$	$\mathscr{A}_6$		
$\mathscr{A}_5$			
$\mathscr{A}_6$		$\mathscr{A}_4$	$\mathscr{A}_7$
$\mathscr{A}_7$			

Пустые клетки в Таб. 3.4 соответствуют неопределённым значениям.

Заметим, что множество GOTO ($\mathscr{A}$, X) всегда пусто, если в каждой $LR(k)$-ситуации из множества $\mathscr{A}$ позиционная точка расположена на конце правила. Примерами таких множеств здесь служат $\mathscr{A}_5$ и $\mathscr{A}_7$.

Теорема 3.3. *Пусть* $G = (V_N, V_T, P, S)$ *— контекстно-свободная грамматика.*

Множество LR(k)-ситуаций $\mathscr{A} \in \mathscr{S}$ *тогда и только тогда, когда существует активный префикс* $\gamma \in (V_N \cup V_T)^*$, *такой, что* $V_k^G(\gamma) = \mathscr{A}$.

Доказательство.

I. Докажем сначала, что если $\mathscr{A} \in \mathscr{S}$, то существует активный префикс $\gamma \in (V_N \cup V_T)^*$, такой, что $V_k^G(\gamma) = \mathscr{A}$.

Согласно алгоритму 3.3 элементы множества $\mathscr{S}$ образуются в определённой последовательности, т. е.

$$\mathscr{S} = \boxtimes_{i=0}^{m} \mathscr{S}_i ,$$

причём сначала строится множество

$$\mathscr{S}_0 = \{ \mathscr{A}_0 \mid \mathscr{A}_0 = V_k^G(\varepsilon) \},$$

а элементы из множества $\mathscr{S}_{i+1}$ строятся по элементам множества $\mathscr{S}_i$ следующим образом:

если $\mathscr{A}' = V_k^G(\gamma') \in \mathscr{S}_i$, то

$$\mathscr{A} = \mathrm{GOTO}\,(\mathscr{A}', X) = V_k^G(\gamma' X) \in \mathscr{S}_{i+1},$$

где $X \in V_{\mathrm{N}} \cup V_{\mathrm{T}}$.

Доказательство проведем индукцией по номеру i множества $\mathscr{S}_i$, которому принадлежит элемент $\mathscr{A}$.

База. Пусть $i = 0$. Имеем $\mathcal{A} \in \mathcal{S}_0$.

Согласно алгоритму 3.3 $\mathcal{A}_0 = V_k^G(\varepsilon)$,
и по теореме 3.2 о корректности алгоритма построения функции $V_k^G(\gamma)$, цепочка $\gamma = \varepsilon$ — как раз такой активный префикс грамматики G, что

$$V_k^G(\varepsilon) = \mathcal{A}_0 = \mathcal{A} \in \mathcal{S}_0 \subseteq \mathcal{S}.$$

Индукционная гипотеза. Предположим, что утверждение выполняется для всех $i \le n$ $(n < m)$.

Индукционный переход. Покажем, что аналогичное утверждение выполняется для $i = n + 1$.

Пусть $\mathcal{A} \in \mathcal{S}_{n+1}$. В соответствии с алгоритмом 3.3 существуют $\mathcal{A}' \in \mathcal{S}_n$ и $X \in V_\text{N} \cup V_\text{T}$, такие, что

$$\mathcal{A} = \text{GOTO}(\mathcal{A}', X) = V_k^G(\gamma' X) = V_k^G(\gamma),$$

где $\gamma = \gamma' X$.

Согласно теореме 3.2 цепочка γ является активным префиксом грамматики G, таким, что $\mathscr{A} = V_k^G(\gamma)$.

Утверждение I доказано.

II. Докажем теперь, что если $\gamma \in (V_N \cup V_T)^*$ — активный префикс грамматики G и

$\mathscr{A} = V_k^G(\gamma)$, то $\mathscr{A} \in \mathscr{S}$.

Индукция по $l = |\gamma|$.

База. Пусть $l = 0$, $\gamma = \varepsilon$.

Имеем $\mathscr{A} = V_k^G(\varepsilon) = \mathscr{A}_0$.

Согласно алгоритму 3.3 $\mathscr{A}_0$ включается в множество $\mathscr{S}$ на первом же шаге.

Индукционная гипотеза. Предположим, что утверждение выполняется для всех γ:

$$|\gamma| = l,\ l \le n\ (n \ge 0).$$

Индукционный переход. Покажем, что такое утверждение выполняется для $l = n + 1$.

Пусть $\gamma = \gamma' X$, где $|\gamma'| = n$, $X \in V_\mathrm{N} \cup V_\mathrm{T}$ и γ — активный префикс грамматики G.

Это значит, что существует правосторонний вывод вида

$$S \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta_1\beta_2 w = \gamma\beta_2 w = \gamma' X \beta_2 w,$$

где $\alpha\beta_1 = \gamma = \gamma' X$.

Следовательно, γ' — тоже активный префикс грамматики G и $|\gamma'| = n$. Согласно индукционной гипотезе

$$V_k^G(\gamma') = \mathscr{A}' \in \mathscr{S}$$

Кроме того, в соответствии с алгоритмом 3.3

$$\mathscr{A} = \text{GOTO}\,(\mathscr{A}', X) = \text{GOTO}\,(V_k^G(\gamma'), X) =$$
$$= V_k^G(\gamma' X) = V_k^G(\gamma)$$

включается в множество $\mathscr{S}$

Утверждение II доказано. Из рассуждений I и II следует утверждение теоремы.

§ 3.5. Тестирование *LR*(*k*)-грамматик

Здесь мы рассмотрим метод проверки, является ли данная cfg $LR(k)$-грамматикой при заданном значении $k \geq 0$.

***Определение* 3.11.** Множество

$$\mathscr{A} = V_k^G(\gamma) \in \mathscr{S}$$

назовем *непротиворечивым*, если оно ***не*** содержит двух разных $LR(k)$-ситуаций вида $[A \to \beta., u]$ и $[B \to \beta_1.\beta_2, v]$ (не исключается случай $\beta_2 = \varepsilon$), таких, что $u \in \mathrm{EFF}_k^G(\beta_2 v)$.

Алгоритм 3.4: $LR(k)$-тестирование.

Ret Ret 199

Вход: $G = (V_N, V_T, P, S)$ — cfg, $k \geq 0$.

Выход: "Да", если G — $LR(k)$-грамматика.

"Нет" — в противном случае.

Метод.

1. Построим $\mathcal{S}$ — каноническую систему множеств $LR(k)$-ситуаций, допустимых для грамматики G.

2. Каждое $\mathcal{A} \in \mathcal{S}$ проверим на непротиворечивость. Если окажется, что рассматриваемое множество $\mathcal{A}$ противоречиво, то алгоритм заканчивается с ответом "Нет".

3. Если алгоритм не закончился на шаге 2, то он выдает ответ "Да" и завершается.

Замечание 3.5. При тестировании достаточно просматривать лишь множества $\mathscr{A} \in \mathscr{S}$, в которых имеется хотя бы одна $LR(k)$-ситуация вида $[A \to \beta., u]$ (с точкой в конце правила). В случае отсутствия таковых данное множество *априори не противоречиво*.

В то же время, если $[A \to \beta., u]$, $[A' \to \beta'., u] \in \mathscr{A}$ при $A \neq A'$ или $\beta \neq \beta'$, то $\mathscr{A}$ *априори противоречиво*.

185

Пример 3.11. Протестируем грамматику примера 3.10, содержащую следующие правила: 0) $S' \to S$, 1) $S \to SaSb$, 2) $S \to \varepsilon$.

В соответствии с замечанием 3.5 необходимо и достаточно проверить лишь непротиворечивость множеств $\mathcal{A}_0$, $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_4$, $\mathcal{A}_5$, $\mathcal{A}_7$, построенных в примере 3.10. Множества $\mathcal{A}_3$ и $\mathcal{A}_6$ нас не интересуют, как не содержащие *LR*(*k*)-ситуаций вида $[A \to \beta., u]$.

Проверка

$\mathscr{A}_0 = \{[S' \to .S, \varepsilon], [S \to .SaSb, \varepsilon \mid a], [S \to ., \varepsilon \mid a]\}$:

$u \in \{\varepsilon, a\}$, $u \notin \mathrm{EFF}_1^G(S\{\varepsilon\}) = \varnothing$;

$u \notin \mathrm{EFF}_1^G(SaSb\{\varepsilon, a\}) = \varnothing$;

Множество $\mathscr{A}_0$ непротиворечиво.

Заметим, что

$$\mathrm{EFF}_1^G(S\{\varepsilon\}) = \mathrm{EFF}_1^G(SaSb\{\varepsilon, a\}) = \varnothing$$

потому, что из цепочек, начинающихся на нетерминал S, выводимы терминальные цепочки только благодаря ε-правилу, используемому на последнем шаге.

И в самом деле, существуют только два вывода, дающих терминальные цепочки:

$$S \underset{rm}{\overset{(2)}{\Rightarrow}} \varepsilon;$$

$$SaSb \underset{rm}{\overset{(2)}{\Rightarrow}} Sab \underset{rm}{\overset{(2)}{\Rightarrow}} ab,$$

причём в каждом из них последнее используемое правило есть ε-порождение.

Поэтому $\mathrm{EFF}_1^G(S) = \mathrm{EFF}_1^G(SaSb) = \varnothing$,

и конкатенация пустого множества с чем угодно даёт пустое множество.

Проверка

$\mathscr{A}_1 = \{[S' \to S., \varepsilon], [S \to S.aSb, \varepsilon \mid a]\}:$

$u = \varepsilon,\ u \notin \mathrm{EFF}_1^G(aSb\{\varepsilon, a\}) = \{a\};$

Множество $\mathscr{A}_1$ непротиворечиво.

Пояснение.

Здесь цепочка aSb в качестве аргумента функции EFF_1^G начинается на терминал, и в этом случае $\mathrm{EFF}_1^G(aSb) = \mathrm{FIRST}_1^G(aSb) = \{a\}$.

Проверка

$\mathcal{A}_2 = \{[S \to Sa.Sb,\ \varepsilon \mid a],\ [S \to .SaSb,\ a \mid b],\ [S \to .,\ a \mid b]\}$:

$$u \in \{a, b\},\quad \begin{array}{l} u \notin \mathrm{EFF}_1^G(Sb\{\varepsilon, a\}) = \varnothing; \\ u \notin \mathrm{EFF}_1^G(SaSb\{a, b\}) = \varnothing; \end{array}$$

Множество $\mathcal{A}_2$ непротиворечиво.

Заметим, что

$$\mathrm{EFF}_1^G(Sb\{\varepsilon, a\}) = \mathrm{EFF}_1^G(SaSb\{a, b\}) = \varnothing$$

потому, что из цепочек, начинающихся на нетерминал S, выводимы терминальные цепочки только благодаря ε-правилу, используемому на последнем шаге.

Проверка

$\mathcal{A}_4 = \{[S \to Sa.Sb,\ a \mid b], [S \to .SaSb,\ a \mid b], [S \to .,\ a \mid b]\}$:

$u \in \{a, b\},$ $u \notin \mathrm{EFF}_1^G(Sb\{a,b\}) = \varnothing;$

$u \notin \mathrm{EFF}_1^G(SaSb\{a,b\}) = \varnothing;$

Множество $\mathcal{A}_4$ непротиворечиво.

Заметим, что

$$\mathrm{EFF}_1^G(Sb\{a,b\}) = \mathrm{EFF}_1^G(SaSb\{a,b\}) = \varnothing$$

потому, что из цепочек, начинающихся на нетерминал S, выводимы терминальные цепочки только благодаря ε-правилу, используемому на последнем шаге.

Проверка

$\mathcal{A}_5 \;= \{[S \to SaSb., \varepsilon \mid a]\}$

— множество непротиворечиво.

Проверка

$\mathcal{A}_7 \;= \{[S \to SaSb., a \mid b]\}$

— множество непротиворечиво.

Эти два множества непротиворечивы, так как их сравнивать не с чем.

Поскольку противоречивых множеств не найдено, то рассматриваемая грамматика — $LR(1)$-грамматика.

Пример II-3.11 (а)

Дана КС-грамматика

$G = (\{S', S\}, \{a\},$
$\{(0)\ S' \to S,\ (1)\ S \to aS,\ (2)\ S \to \varepsilon\}, S').$

Является ли G — $LR(1)$-грамматикой?

***Решение*:**

$$\mathscr{A}_0 = \{[S' \to .S, \varepsilon],\ [S \to .aS, \varepsilon],\ [S \to .\varepsilon, \varepsilon]\}.$$

Построение $\mathscr{A}_0 = V_1^G(\varepsilon)$:

Инициализация:

$V_1^G(\varepsilon) = \{[S' \to .S, \varepsilon]\}.$

Замыкание:

а) Добавить $[S \to .\alpha, \varepsilon]$.

б) Добавить $[A \to .\beta, \varepsilon]$ для всех A: $\alpha = A\alpha'$.

в) Повторять (б) для $\beta = B\beta'$.

<u>Проверка</u> $\mathcal{A}_0$ по (1):

$[S \to .\varepsilon, \varepsilon] \in \mathcal{A}_0$

(1) $\{[S' \to .S, \varepsilon] \in \mathcal{A}_0$

Условие противоречия по (1):

$$\boxed{\varepsilon \in \mathrm{EFF}_1^G(S) \oplus_1 \{\varepsilon\}.}$$

Так как $\mathrm{EFF}_1^G(S) \oplus_1 \{\varepsilon\} = \varnothing \oplus_1 \{\varepsilon\} = \varnothing$,

и $\varepsilon \notin \varnothing$, то противоречия по (1) нет.

Поясним, что любой вывод вида $S \underset{rm}{\overset{*}{\Rightarrow}} x$, где V_T^G заканчивается ε-правилом.

Пример II-3.11 (а)

<u>Проверка</u> $\mathcal{A}_0$ по (2):

$[S \to .\varepsilon, \varepsilon] \in \mathcal{A}_0$

(2) $[S \to .aS, \varepsilon] \in \mathcal{A}_0$

Условие противоречия по (2):

$$\boxed{\varepsilon \in \mathrm{EFF}_1^G(aS) \oplus_1 \{\varepsilon\}}$$

Так как

$$\mathrm{EFF}_1^G(aS) \oplus_1 \{\varepsilon\} = \mathrm{FIRST}_1^G(aS) \oplus_1 \{\varepsilon\} =$$

$$= \{a\} \oplus_1 \{\varepsilon\} = \{a\},$$

и $\varepsilon \notin \{a\}$, то противоречия по (2) нет.

непротиворечиво!

Вычисление
GOTO ($\mathcal{A}_0$, X),
$X \in (V_T \cup V_N)$.

Построение $\mathcal{A}'$ = GOTO ($\mathcal{A}$, X), где $X \in (V_T \cup V_N)$:
Пусть $[A \to \beta_1 . X\beta_2, u] \in \mathcal{A}$.
Инициализация: $\mathcal{A}' = [A \to \beta_1 X.\beta_2, u]$.
Замыкание: Если $X \in V_N$, то
$\mathcal{A}' = \mathcal{A}' \cup [A \to \beta_1 X.\beta_2, v]$, где $v \in \mathrm{FIRST}_k^G(\beta_2 u)$,
причём v наследуется всеми ситуациями с правилами вида $X \to \beta$: $[X \to .\beta, v]$.

$[S' \to \underbrace{\varepsilon}_{\beta_1} \underbrace{.S}_{X} \underbrace{\varepsilon}_{\beta_2}, \underbrace{\varepsilon}_{u}] \in \mathcal{A}_0$

GOTO ($\mathcal{A}_0$, S) = $\{[S' \to S., \varepsilon]\}$ = $\mathcal{A}_1$. Замыкание не требуется.

$[S \to \underbrace{\varepsilon}_{\beta_1} \underbrace{.a}_{X} \underbrace{S}_{\beta_2}, \underbrace{\varepsilon}_{u}] \in \mathcal{A}_0$,

GOTO ($\mathcal{A}_0$, a) = $\{[S \to a.S, \varepsilon]$, (замыкание требуется)
$[S \to .aS, \varepsilon \mid a]\} = \mathcal{A}_2$.

$[S \to .\varepsilon, \varepsilon] \in \mathcal{A}_0$. Не существует $X \notin (V_T \cup V_N)$, и ничего делать не надо.

Проверка непротиворечивости $\mathscr{A}_1$:

Поскольку в $\mathscr{A}_1$ только одна ситуация, то это множество непротиворечиво.

Проверка непротиворечивости $\mathscr{A}_2$:

Поскольку в $\mathscr{A}_2$ нет ситуации с точкой в конце правой части правила, то это множество непротиворечиво.

Вычисление GOTO для $\mathscr{A}_1$ даёт

GOTO ($\mathscr{A}_1$, S) = ∅, GOTO ($\mathscr{A}_1$, a) = ∅.

Это значит, что в этом случае порождаются пустые множества ситуаций, которые не включаются в $\mathscr{S}$.

Вычисление GOTO для $\mathscr{A}_2$ даёт

GOTO ($\mathscr{A}_2$, a) = $\{[S \to a.S, \varepsilon \mid a]$, замыкание даёт:
$[S \to .aS, \varepsilon \mid a], [S \to ., \varepsilon \mid a]\} = \mathscr{A}_3$.

GOTO ($\mathscr{A}_2$, S) = $\{[S \to aS., \varepsilon]\} = \mathscr{A}_4$.

Пример II-3.11 (а)

<u>Проверка</u> $\mathcal{A}_3$ по (1):

$[S \to .,\ \varepsilon \mid a] \in \mathcal{A}_3$

(1) $[S \to a.S,\ \varepsilon \mid a] \in \mathcal{A}_3$

$$\boxed{\begin{array}{ll} [A \to \beta.,\ u] & \underline{\text{Противоречие, если}} \\ [B \to \beta_1.\beta_2,\ v] & u \in \mathrm{EFF}_k^G(\beta_2 v) \end{array}}$$

Условие противоречия по (1):

$$\boxed{\{\varepsilon|a\} \cap (\mathrm{EFF}_1^G(S) \oplus_1 \{\varepsilon|a\}) \neq \varnothing.}$$

$\mathrm{EFF}_1^G(S) = \varnothing$, т. к. любой правосторонний вывод вида $S \overset{*}{\underset{rm}{\Rightarrow}} x$, где $x \in V_{\mathrm{T}}^*$, заканчивается ε-правилом.

ϕ

Поэтому $\{\varepsilon|a\} \cap (\mathrm{EFF}_1^G(S) \oplus_1 \{\varepsilon|a\}) = \varnothing$

и противоречия пока не обнаружено!

<u>**Проверка**</u> $\mathscr{A}_3$ по (2):

$$
\boxed{
\begin{array}{l}
[A \to \beta., u] \\
[B \to \beta_1.\beta_2, v]
\end{array}
\;\Bigg]\;
\begin{array}{l}
\underline{\text{Противоречие, если}} \\
u \in \mathrm{EFF}_k^G(\beta_2 v)
\end{array}
}
$$

$[S \to ., \varepsilon \mid a] \in \mathscr{A}_3$

(2) $[S \to .aS, \varepsilon \mid a] \in \mathscr{A}_3$

$\mathscr{A}_3$

Условие противоречия по (2):

$$\boxed{\{\varepsilon \mid a\} \cap (\mathrm{EFF}_1^G(aS) \oplus_1 \{\varepsilon \mid a\}) \neq \varnothing.}$$

$$\mathrm{EFF}_1^G(aS) \oplus_1 \{\varepsilon \mid a\} = \mathrm{FIRST}_1^G(aS) \oplus_1 \{\varepsilon \mid a\}) = \{a\}.$$

$\{\varepsilon \mid a\} \cap \{a\} = \{a\} \neq \varnothing$ — противоречие!

— противоречиво!

Итак, грамматика G — не $LR(1)$.

Виной тому правая рекурсивность (?).

Проверка $\mathscr{A}_4$:

Поскольку в $\mathscr{A}_4$ только одна ситуация, то это множество непротиворечиво.

Если попытаться построить $LR(k)$-таблицу, соответствующую $\mathscr{A}_3$, то функция $f_3(u)$ — *неоднозначна*: для $u \in \{\varepsilon, a\}$ она даёт свёртку по правилу 2 или сдвиг.

Теорема 3.4. *Алгоритм 3.4 дает правильный ответ, т. е. является правильным алгоритмом тестирования.*

Доказательство. Утверждение теоремы является прямым следствием теоремы 3.1, на которой и основывается алгоритм 3.4.

§ 3.6. Канонические *LR*(*k*)-анализаторы

***Определение* 3.12.** Пусть $G = (V_{\mathrm{N}}, V_{\mathrm{T}}, P, S)$ — контекстно-свободная грамматика и $\mathcal{S}$ — каноническое множество $LR(k)$-ситуаций для грамматики G.

Для каждого множества $\mathcal{A} \in \mathcal{S}$ определим $LR(k)$-*таблицу* $T(\mathcal{A}) = (f, g)$, *состоящую из пары функций*:

$$f : V_{\mathrm{T}}^{*k} \to \{\text{shift, reduce } i, \text{ accept, error}\},$$

$$g : V_{\mathrm{N}} \cup V_{\mathrm{T}} \to \{T(\mathcal{A}) \mid \mathcal{A} \in \mathcal{S}\} \cup \{\text{error}\}.$$

229

Функция f определяется следующим образом:

а) $f(u)$ = shift,

если существует $[A \to \beta_1.\beta_2, v] \in \mathcal{A}$,

$\beta_2 \neq \varepsilon$ и $u \in \mathrm{EFF}_k^G(\beta_2 v)$.

б) $f(u)$ = reduce i,

если существует $[A \to \beta., u] \in \mathcal{A}$ и $A \to \beta$ есть i-е правило из множества правил P, $i \geq 1$[1];

---

[1] Напомним, что под нулевым номером числится правило $S' \to S$, пополняющее данную грамматику G.

в) $f(u)$ = accept, если $[S' \to S., \varepsilon] \in \mathscr{A}$ и $u = \varepsilon$;
г) $f(u)$ = error — в остальных случаях.

Если G — $LR(k)$-грамматика, то никаких неоднозначностей по пунктам а) и б) не может быть.

Функция g определяется следующим образом:

$$g(X) = \begin{cases} T(\mathscr{A}'), \text{ где } \mathscr{A}' = \mathrm{GOTO}(\mathscr{A}, X), \\ \qquad\qquad \text{если } \mathscr{A}' \neq \varnothing; \\ \text{error}, \quad \text{если } \mathscr{A}' = \varnothing. \end{cases}$$

Ret 210

***Определение* 3.13.** Будем говорить, что $LR(k)$-*таблица* $T(\mathscr{A})$ *соответствует активному префиксу* γ, если $\mathscr{A} = V_k^G(\gamma)$.

***Определение* 3.14.** *Канонической системой* $LR(k)$-*таблиц для* cfg G назовем пару $(\mathscr{T}, T_0)$, где $\mathscr{T}$ — множество $LR(k)$-таблиц, соответствующих канонической системе множеств $LR(k)$-ситуаций для G, а $T_0 = T(\mathscr{A}_0)$, где $\mathscr{A}_0 = V_k^G(\varepsilon)$.

Алгоритм 3.5: построение канонического $LR(k)$-анализатора.

Вход: $G = (V_{\mathrm{N}}, V_{\mathrm{T}}, P, S)$ — $LR(k)$-грамматика, $k \geq 0$.

Выход: $(\mathscr{T}, T_0)$ — каноническая система $LR(k)$-таблиц для грамматики G.

Метод.

1. Построить каноническую систему множеств $LR(k)$-ситуаций $\mathscr{S}$ для грамматики G.
2. Взять $\mathscr{T} = \{T(\mathscr{A}) \mid \mathscr{A} \in \mathscr{S}\}$ в качестве множества $LR(k)$-таблиц.
3. Положить $T_0 = T(\mathscr{A}_0)$, где $\mathscr{A}_0 = V_k^G(\varepsilon)$.

Обычно $LR(k)$-анализатор представляется управляющей таблицей, каждая строка которой является $LR(k)$-таблицей.

***Определение* 3.15.** Описанный ранее алгоритм 3.1 (*см.* § 3.3), если он использует каноническую систему $LR(k)$-таблиц, назовем *каноническим* $LR(k)$-*алгоритмом разбора* или просто *каноническим* $LR(k)$-*анализатором*.

Пример 3.12. Построим каноническую систему $LR(k)$-таблиц для грамматики из примера 3.10, содержащей следующие правила: 0) $S' \to S$, 1) $S \to SaSb$, 2) $S \to \varepsilon$, учитывая результаты построения системы множеств $LR(k)$-ситуаций и функции GOTO, приведённые в этом примере (см. Табл. 3.4).

Поскольку

$\mathscr{A}_0 = \{[S' \to .S, \varepsilon], [S \to .SaSb, \varepsilon \mid a],$

$[S \to ., \varepsilon \mid a]\}$, то $T_0 = (f_0, g_0)$ имеет следующий табличный вид:

Табл. T_0

$f_0(u)$			$g_0(X)$		
a	b	ε	S	a	b
reduce 2	error	reduce 2	T_1	error	error

Поскольку $\mathscr{A}_1 = \{[S' \to S., \varepsilon], [S \to S.aSb, \varepsilon \mid a]\}$, то $T_1 = (f_1, g_1)$ имеет следующий табличный вид:

Табл. T_1

$f_1(u)$			$g_1(X)$		
a	b	ε	S	a	b
shift	error	accept	error	T_2	error

Поскольку $\mathscr{A}_2 = \{[S \to Sa.Sb, \varepsilon \mid a], [S \to .SaSb, a \mid b],$ $[S \to ., a \mid b]\}$, то $T_2 = (f_2, g_2)$ имеет следующий табличный вид:

Табл. T_2

$f_2(u)$			$g_2(X)$		
a	b	ε	S	a	b
reduce 2	reduce 2	error	T_3	error	error

Поскольку $\mathscr{A}_3 = \{[S \to SaS.b, \varepsilon \mid a], [S \to S.aSb, a \mid b]\}$, то $T_3 = (f_3, g_3)$ имеет следующий табличный вид:

Табл. T_3

$f_3(u)$			$g_3(X)$		
a	b	ε	S	a	b
shift	shift	error	error	T_4	T_5

Поскольку
$\mathscr{A}_4 = \{[S \to Sa.Sb,\ a \mid b], [S \to .SaSb,\ a \mid b],$
$[S \to .,\ a \mid b]\}$, то $T_4 = (f_4, g_4)$ имеет следующий табличный вид:

Табл. T_4

$f_4(u)$			$g_4(X)$		
a	b	ε	S	a	b
reduce 2	reduce 2	error	T_6	error	error

Поскольку
$\mathcal{A}_5 = \{[S \rightarrow SaSb., \varepsilon \mid a\}$, то $T_5 = (f_5, g_5)$ имеет следующий табличный вид:

Табл. T_5

$f_5(u)$			$g_5(X)$		
a	b	ε	S	a	b
reduce 1	error	reduce 1	error	error	error

Поскольку
$\mathscr{A}_6 = \{[S \to SaS.b,\ a \mid b],\ [S \to S.aSb,\ a \mid b]\}$,
то $T_6 = (f_6, g_6)$ имеет следующий табличный вид:

Табл. T_6

$f_6(u)$			$g_6(X)$		
a	b	ε	S	a	b
shift	shift	error	error	T_4	T_7

Поскольку $\mathscr{A}_7 = \{[S \to SaSb., a \mid b]\}$, то $T_7 = (f_7, g_7)$ имеет следующий табличный вид:

Табл. T_7

$f_7(u)$			$g_7(X)$		
a	b	ε	S	a	b
reduce 1	reduce 1	error	error	error	error

Все эти $LR(k)$-таблицы сведены в управляющую таблицу 3.1 канонического $LR(k)$-анализатора, которая была приведена в примере 3.6.

Канонический $LR(k)$-анализатор обладает следующими свойствами:
1. Простая индукция по числу шагов анализатора показывает, что каждая $LR(k)$-таблица, находящаяся в его магазине, *соответствует* цепочке символов, расположенной слева от неё (имена $LR(k)$-таблиц игнорируются) .

Поэтому, как только k символов необработанной части входной цепочки оказываются такими, что нет суффикса, который вместе с обработанной частью давал бы цепочку, принадлежащую $L(G)$, анализатор сразу сообщает об ошибке.

В каждый момент его работы цепочка символов грамматики, хранящаяся в магазине, должна быть активным префиксом грамматики.

Поэтому $LR(k)$-анализатор сообщает об ошибке при первой же возможности в ходе считывания входной цепочки слева направо.

2. Пусть $T_j = (f_j, g_j)$. Если $f_j(u) =$ shift и анализатор находится в конфигурации $(T_0X_1T_1X_2 \ldots X_jT_j, x, \pi)$, то найдется $LR(k)$-ситуация $[B \to \beta_1.\beta_2, v]$, $\beta_2 \neq \varepsilon$, допустимая для активного префикса $X_1X_2 \ldots X_j$, такая, что

$$u \in \mathrm{EFF}_k^G(\beta_2 v), \text{где } u \in \mathrm{FIRST}_k^G(x).$$

Поэтому, если $S' \underset{rm}{\overset{*}{\Rightarrow}} X_1X_2 \ldots X_j uy$ при некоторой цепочке $y \in V_\mathrm{T}^*$, то по теореме 3.1 правый конец основы цепочки $X_1 X_2 \ldots X_j uy$ должен быть где-то справа от символа X_j.

3. Если в конфигурации, указанной в п.2, $f_j(u)$ = reduce i и $A \to Y_1 Y_2 \ldots Y_p$ — правило с номером i, то цепочка $X_{j-p+1} \ldots X_{j-1} X_j$ в магазине должна совпадать с $Y_1 Y_2 \ldots Y_p$, так как множество ситуаций, по которому построена таблица T_j, содержит ситуацию $[A \to Y_1 Y_2 \ldots Y_p., u]$. Таким образом, на шаге свёртки необязательно рассматривать символы верхней части магазина, надо просто выбросить $2p$ символов грамматики и $LR(k)$-таблиц с вершины магазина.

4. Если $f_j(u) = \text{accept}$, то $u = \varepsilon$. Содержимое магазина в этот момент имеет вид: T_0ST_1, где T_1 — $LR(k)$-таблица, соответствующая множеству $LR(k)$-ситуаций, в которое входит ситуация $[S' \to S., \varepsilon]$.

5. Можно построить детерминированный магазинный преобразователь (dpdt), реализующий канонический $LR(k)$-алгоритм разбора.

Действительно, так как аванцепочку можно хранить в конечной памяти управляющего устройства dpdt, то ясно, как построить расширенный dpdt, реализующий алгоритм 3.1.

§ 3.7. Корректность $LR(k)$-анализаторов

Теорема 3.5. Канонический $LR(k)$-*алгоритм правильно находит правый разбор входной цепочки, если он существует, и объявляет об ошибке в противном случае.*

Доказательство.

I. Индукцией по числу шагов вывода $n = | \pi |$ докажем вспомогательное утверждение:

257

если $S \underset{rm}{\overset{\pi}{\Rightarrow}} \alpha A x$,
то $(T_0X_1T_1X_2 \ \dots \ X_mT_m, \ x, \ \varepsilon) \ \overset{*}{\underset{\mathfrak{P}}{\vdash}} \ (T_0ST, \ \varepsilon, \ \pi^R)$,
при том, что $\alpha = X_1X_2 \dots X_m$, $X_i \in V_{\mathrm{N}} \cup V_{\mathrm{T}}$,
$A \in V_{\mathrm{N}}, x \in V_{\mathrm{T}}^*$,
T_0, T, $T_i \in \mathscr{T}$ ($i = 1, 2, \dots, m$),
$T_0 = T(\mathscr{A}_0)$, $\mathscr{A}_0 = V_k^G(\varepsilon)$,
$T = (f, g) \in \mathscr{T}$, и $f(\varepsilon) = \text{accept}$.

215

Напомним, что каждая $LR(k)$-таблица $T \in \mathcal{T}$ соответствует множеству $LR(k)$-ситуаций $\mathcal{A} = V_k^G(\gamma)$, допустимых для активных префиксов γ. Причём такой префикс γ представлен цепочкой символов грамматики, предшествующих таблице в магазине анализатора.

Например, в конфигурации

$$(T_0 X_1 T_1 X_2 \ldots X_m T_m, x, \varepsilon)$$

таблица T_m соответствует префиксу $X_1 X_2 \ldots X_m$, а в конфигурации $(T_0 S T, \varepsilon, \pi^R)$ таблица T — префиксу S, а таблица T_0 — префиксу ε.

Таких множеств $\mathcal{A}$ для данной $LR(k)$-грамматики конечное число. Они и составляют каноническую систему множеств $LR(k)$-ситуаций $\mathcal{S}$, допустимых для грамматики G.

Напомним также, что множества $\mathcal{A} \in \mathcal{S}$ вычисляются итеративно по формуле $\mathcal{A}_i = \text{GOTO}\ (\mathcal{A}_{i-1}, X)$, начиная с $\mathcal{A}_0 = V_k^G(\varepsilon)$.

Здесь $X \in V_{\text{N}} \cup V_{\text{T}}$.

База: $n = 1$.

Имеем $S \overset{(i)}{\underset{rm}{\Rightarrow}} w, w \in V_{\text{Т}}^{*}, S \to w \in P$.

Пусть $w = a_1a_2 \dots a_m$, где $j \in V_{\text{Т}}, j = 1, 2, \dots, m$.

При $m = 0$ имеет место равенство $w = \varepsilon$.

Очевидно, что при анализе w анализатор руководствуется $LR(k)$-таблицами $\{T_p\}_{p=0}^{m}$:

$T_0 = T(\mathscr{A}_0)$, $\mathscr{A}_0 = V_k^G(\varepsilon)$,

$\mathscr{A}_p = \text{GOTO}\,(\mathscr{A}_{p-1}, a_p)$,

$T_p = T(\mathscr{A}_p) = (f_p, g_p)$,

причём $f_p\,(a_{p+1}) = \text{shift}$ при $p < m$, а

$f_m\,(\varepsilon) = \text{reduce } i$.

Рассмотрим движения анализатора

$$(T_0, w, \varepsilon) \vdash^*_{\mathfrak{A}} (T_0 \underbrace{a_1 T_1 a_2 T_2 \dots a_m T_m}_{a_1 a_2 \dots a_m \text{ — основа по правилу } (i)\ S \to w}, \varepsilon, \varepsilon) \vdash_{\mathfrak{A}}$$

$$\vdash_{\mathfrak{A}} (T_0ST, \varepsilon, i).$$

Здесь $T = g_0(S)$, $T = (f, g)$, $f(\varepsilon) = \text{accept}$.

В случае $w = \varepsilon$ используются только таблицы T_0 и T.

Действительно, согласно определению 3. Действительно, согласно определению 3.12, имея в виду всевозможные разбиения цепочки $w = \beta_1.\beta_2$, начиная с $\beta_1 = \varepsilon$, $\beta_2 = w$, и заканчивая разбиением с $\beta_1 = w$, $\beta_2 = \varepsilon$, можно утверждать, что ситуации

$$[S \rightarrow \beta_1.\beta_2, v] \in \quad , \quad \mathcal{A}$$

задействованные в множествах $\mathcal{A}$, соответствуют активным префиксам

$\alpha_0 = \varepsilon$, $\alpha_1 = a_1$, $\alpha_2 = a_1 a_2$, ... , $\alpha_m = a_1 a_2 \ldots a_m$, и потому обладают требуемыми свойствами (*см.* п.п. а). п.п. а) и б) определения

Пояснение. В рассматриваемых случаях:

а) $[S \to \beta_1.\beta_2, v] \in \mathcal{A}$, где

$\beta_2 \neq \varepsilon$, $u \in \mathrm{EFF}_k^G(\beta_2 v) = \mathrm{FIRST}_k^G(\beta_2 v)$, потому что аргументы этих функций — цепочки из V_T^+.

б) $f_m(u) = \text{reduce } i$, $u = \varepsilon$, поскольку существует $[S \to w., u] \in \mathcal{A}_m$ и $S \to w$ есть i-е правило $i \geq 1$.

Наконец, после единственной свёрки таблица $T = g_0(S)$, где $T = (f, g)$, и $f(\varepsilon) = \text{accept}$.

Индукционная гипотеза.

Предположим, что утверждение I выполняется для всех $l \le n$ ($n \ge 1$).

Индукционный переход. Покажем, что утверждение I выполняется для $l = n + 1$.

Имеем $S \underset{rm}{\overset{\pi'}{\Rightarrow}} \alpha' A w \underset{rm}{\overset{(i)}{\Rightarrow}} \alpha' \beta w = \alpha x$.

Очевидно, что x заканчивается цепочкой w: $x = zw$. Здесь α', $\alpha \in (V_{\mathrm{N}} \cup V_{\mathrm{T}})^*$, $w, x, z \in V_{\mathrm{T}}{}^*$, $\pi = \pi' i$, $A \to \beta \in P$ — i-е правило, $i > 0$. Кроме того, $\alpha' \beta w = \alpha x = \alpha z w$, и потому $\alpha' \beta = \alpha z$.

Рассмотрим исходную конфигурацию

$$\left(T_0\overbrace{\overbrace{X_1T_1X_2\ldots X_jT_j}^{\alpha' = X_1\ldots X_j}\overbrace{X_{j+1}T_{j+1}\ldots X_mT_m}^{\beta' = X_{j+1}\ldots X_m}}^{\alpha}\underbrace{}_{(i)\ A\to\beta,\ \beta=\beta' z}, \overbrace{zw}^{x}, \varepsilon\right).$$

В ней $\alpha' = X_1X_2 \ldots X_j$, $\beta' = X_{j+1} \ldots X_m$, $\beta = \beta' z$, $\alpha = \alpha'\beta'$, $x = zw$. Поэтому имеющийся вывод представим в виде

$$S \underset{rm}{\overset{\pi'}{\Rightarrow}} \alpha' A w \underset{rm}{\overset{(i)}{\Rightarrow}} \alpha'\beta w = \alpha'\underline{\beta' z}w.$$

Поскольку $A \to \beta = \beta' z \in P$, то $\alpha'\beta'$ — активный префикс право-сентенциальной формы $\alpha'\beta' zw$, причём $V_k^G(\alpha'\beta') = V_k^G(\alpha) = \mathcal{A}_m$, и поэтому $[A \to \beta'. z, u] \in \mathcal{A}_m$, где $u \in \mathrm{FIRST}_k^G(w)$.

Пусть $z = a_1a_2 \dots a_p$. Согласно алгоритму построения $LR(k)$-таблиц имеем

$T_m = T(\mathcal{A}_m) = (f_m, g_m)$,

$f_m(v_1) = \text{shift}$ для $v_1 \in \mathrm{EFF}_k^G(zu) =$

$$= \mathrm{FIRST}_k^G(a_1a_2 \dots a_p w) = \mathrm{FIRST}_k^G(x),$$

$g_m(a_1) = T_{m+1}$, где $T_{m+1} = T(\mathcal{A}_{m+1})$, $\mathcal{A}_{m+1} = V_k^G(\alpha'\beta' a_1)$;

$T_{m+1} = (f_{m+1}, g_{m+1})$,

$f_{m+1}(v_2) = \text{shift}$ для $v_2 = \mathrm{FIRST}_k^G(a_2 \ldots a_p u) =$

$= \mathrm{FIRST}_k^G(a_2 \ldots a_p w)$,

$g_{m+1}(a_2) = T_{m+2}$, где $T_{m+2} = T(\mathcal{A}_{m+2})$,

$\mathcal{A}_{m+2} = V_k^G(\alpha'\beta' a_1 a_2)$;

…

$T_{m+p-1} = T(\mathcal{A}_{m+p-1}) = (f_{m+p-1}, g_{m+p-1})$,

$\mathcal{A}_{m+p-1} = V_k^G(\alpha'\beta' a_1 a_2 \ldots a_{p-1})$;

$f_{m+p-1}(v_p) =$ shift для $v_p \in \mathrm{EFF}_k^G(a_pu) = \mathrm{FIRST}_k^G(a_pu) = \mathrm{FIRST}_k^G(a_pw)$,

$g_{m+p-1}(a_p) = T_{m+p} = T(\mathcal{A}_{m+p})$,

где $\mathcal{A}_{m+p} = V_k^G(\alpha'\beta'a_1a_2 \ldots a_{p-1}a_p) =$

$= V_k^G(\alpha'\beta'z) = V_k^G(\alpha'\beta)$.

$T_{m+p} = (f_{m+p}, g_{m+p})$, $f_{m+p}(u) =$ reduce i,

$(i)\, A \to \beta \in P, u \in \mathrm{FIRST}_k^G(w)$.

Используя существующие $LR(k)$-таблицы, канонический $LR(k)$-анализатор совершает следующие движения, начиная с исходной конфигурации:

$$(T_0X_1T_1X_2 \dots X_mT_m, x, \varepsilon) =$$
$$= (T_0X_1T_1X_2 \dots X_mT_m, zw, \varepsilon) =$$
$$= (T_0X_1T_1X_2 \dots X_mT_m, a_1a_2 \dots a_pw, \varepsilon) \vdash_{\mathfrak{A}}$$
$$\vdash_{\mathfrak{A}} (T_0X_1T_1X_2 \dots X_mT_ma_1T_{m+1}, a_2 \dots a_pw, \varepsilon) \vdash^*_{\mathfrak{A}}$$
$$\vdash^*_{\mathfrak{A}} (T_0X_1T_1X_2 \dots X_mT_ma_1T_{m+1}a_2T_{m+2} \dots a_pT_{m+p}, w, \varepsilon) =$$
$$= (T_0X_1T_1X_2 \dots X_jT_jX_{j+1}T_{j+1} \dots$$
$$X_mT_ma_1T_{m+1}a_2T_{m+2} \dots \quad \dots a_pT_{m+p}, w, \varepsilon) \vdash_{\mathfrak{A}}$$
$$\vdash_{\mathfrak{A}} (T_0X_1T_1X_2 \dots X_jT_jAT'_{j+1}, w, i) \vdash^*_{\mathfrak{A}}$$
$$\vdash^*_{\mathfrak{A}} (T_0ST, \varepsilon, i\pi'^R) = (T_0ST, \varepsilon, \pi^R).$$

Последний переход обоснован индукционной гипотезой с учётом того, что

$$T'_{j+1} = T(\qquad), T(\qquad) = V_k^G(\alpha' A),$$

$$\alpha' = X_1 X_2 \ldots X_j.$$

Вспомогательное утверждение I доказано.

В частности, если $\alpha = \varepsilon$, т. е. $S \underset{rm}{\overset{*}{\Rightarrow}} x$,

то $(T_0, x, \varepsilon) \vdash^{*}_{\mathcal{P}} (T_0 ST, \varepsilon, \pi^R)$.

Достигнутая конфигурация является принимающей по тем же причинам, что и в базовом случае.

II. Индукцией по l — числу свёрток канонического анализатора — покажем, что если

$$(T_0, w, \varepsilon) \vdash^{l}_{\mathfrak{A}} (T_0 X_1 T_1 X_2 \ldots X_m T_m, x, \pi^R),$$

то $\alpha x \overset{\pi}{\underset{rm}{\Rightarrow}} w$, где $\alpha = X_1 X_2 \ldots X_m$, $X_j \in V_N \cup V_T$, $j = 1, 2, \ldots, m$; $x, w \in V_T^*$.

Заметим, что в этом переходе участвует хотя бы одна свёртка ($\pi^R \neq \varepsilon$), и сколько то сдвигов, может быть ноль, перед каждой из них.

База. Пусть $l = 1$.

Пусть $(T_0, w, \varepsilon) = (T_0, a_1 a_2 \ldots a_m, \varepsilon)$

$(T_0 a_1 T_1 a_2 T_2 \ldots a_m T_m, \varepsilon, i)$.

Здесь $w = a_1 a_2 \ldots a_m, a_p \in V_{\text{Т}}, p = 1, 2, \ldots, m$.

Случай 1. $w \neq \varepsilon$.

Действия анализатора, с учётом способа его построения, могли происходить только так:

$[S' \to .S, \varepsilon] \in \mathcal{A}_0$, выполняется замыкание,

$\mathcal{A}_0 = V_k^G(\varepsilon)$, $T_0 = T(\mathcal{A}_0) = (f_0, g_0)$, $f_0(a_1) = \text{shift}$, существует множество $LR(k)$-ситуаций $\mathcal{A}_1 = \mathrm{GOTO}\,(\mathcal{A}_0, a_1)$ и таблица $T_1 = T(\mathcal{A}_1)$, $T_1 = (f_1, g_1), f_1(a_2) = \text{shift}, g_1(a_2) = T_2$, существует множество $LR(k)$-ситуаций $\mathcal{A}_2 = \mathrm{GOTO}\,(\mathcal{A}_1, a_2)$, $T_2 = T(\mathcal{A}_2)$, ... существует множество $LR(k)$-ситуаций $\mathcal{A}_m = \mathrm{GOTO}\,(\mathcal{A}_{m-1}, a_m)$, $T_m = T(\mathcal{A}_m)$, $T_m = (f_m, g_m), f_m(\varepsilon) = \text{reduce } i, g_0(S) = T = (f, g)$, $[S \to a_1 a_2 \ldots a_m\bullet, \varepsilon] \in \mathcal{A}_0 = V_k^G(\varepsilon)$, $f(\varepsilon) = \text{accept}$.

Последнее означает, что $S \to a_1 a_2 \ldots a_m$ есть i-е правило грамматики.

Заметим, в случае, когда не ни одного сдвига $w = \varepsilon$ $(m = 0)$, то необходимо, чтобы $T_0 = T(\mathcal{A}_0)$, $T_0 = (f_0, g_0)$, $f_0(\varepsilon) =$ reduce i, $\mathcal{A}_1 = \mathrm{GOTO}\,(\mathcal{A}_0, S)$, $T_1 = T(\mathcal{A}_1)$, $T = (f, g)$, $f(\varepsilon) =$ accept, при том, что $g_0(S) = T_1$ и $[S \to \varepsilon., \varepsilon] \in \mathcal{A}_0 = V_k^G(\varepsilon)$.

Последнее означает, что $S \to \varepsilon$ есть i-е правило грамматики.

В любом случае существует требуемый вывод $S \overset{(i)}{\underset{rm}{\Rightarrow}} w.$ База доказана.

Индукционная гипотеза. Предположим, что утверждение II выполняется для всех $l \le n$ $(n \ge 1)$.

Индукционный переход. Покажем, что утверждение II выполняется для $l = n + 1$.

Пусть первые n движений дают

$$(T_0, w, \varepsilon) \overset{n}{\underset{\mathcal{A}}{\vdash}} (T_0X_1T_1X_2 \dots X_mT_m, x', \pi'^R).$$

По индукционной гипотезе немедленно получаем $X_1X_2 \dots X_mx' \overset{\pi'}{\underset{rm}{\Rightarrow}} w$.

Затем совершается последнее, $(n + 1)$-е, движение.

Случай 1: shift-движение.

Это движение происходит следующим образом:

$$(T_0X_1T_1X_2 \dots X_mT_m, x', \pi'^R) =$$
$$= (T_0X_1T_1X_2 \dots X_mT_m, X_{m+1}x, \pi'^R) \underset{\mathfrak{A}}{\vdash}$$
$$\underset{\mathfrak{A}}{\vdash} (T_0X_1T_1X_2 \dots X_mT_mX_{m+1}T_{m+1}, x, \pi'^R),$$

т. е. X_{m+1} переносится в магазин, в выходную цепочку ничего не пишется.

Остается лишь убедиться в том, что

$$X_1 X_2 \dots X_m \xRightarrow[rm]{\pi'} w.$$

$X_{m+1}x$

Так как $x' = X_{m+1}x$, то

$$X_1 X_2 \dots X_m X_{m+1} x = X_1 X_2 \dots X_m x' \xRightarrow[rm]{\pi'} w$$

по индукционной гипотезе.

<u>Случай 2</u>: reduce i -движение.

Имеем конфигурацию, достигнутую за первые n движений: $(T_0 X_1 T_1 X_2 \dots X_m T_m, x', \pi'^R)$.

Далее совершается последнее движение: свёртка верхней части магазина по i-му правилу. Оно происходит благодаря тому, что $T_m = (f_m, g_m)$, $f_m(u') =$ reduce i для

u' Î $\mathrm{FIRST}_k^G(x')$.

Согласно алгоритму построения анализатора $T_m = T(\mathsf{A}_m)$, $\mathsf{A}_m = V_k^G(X_1 X_2 \ldots X_m)$,
и существуют $LR(k)$-ситуация

$$[A \to Y_1 Y_2 \ldots Y_p \bullet, u'] \in \mathsf{A}_m$$

и правило $A \to Y_1 Y_2 \ldots Y_p$ под номером i, такое, что $Y_1 Y_2 \ldots Y_p = X_{m-p+1} \ldots X_{m-1} X_m$.

Имеем:

$$(T_0 X_1 T_1 X_2 \ldots X_m T_m, x', \pi'^R) =$$

$$= (T_0 X_1 T_1 X_2 \ldots X_{m-p} T_{m-p} Y_1 T_{m-p+1} Y_2 T_{m-p+2} \ldots$$

$$\ldots Y_p T_m x', \pi'^R) \vdash_{\mathfrak{A}}$$

$$\vdash_{\mathfrak{A}} (T_0 X_1 T_1 X_2 \ldots X_{m-p} T_{m-p} A T'_{m-p+1}, x', \pi'^R i),$$

где

$$T'_{m-p+1} = g_{m-p}(A), \text{ если } T_{m-p} = (f_{m-p}, g_{m-p}).$$

Остается убедиться лишь в том, что

$$X_1 X_2 \ldots X_{m-p} A x' \overset{i\pi'}{\underset{rm}{\Rightarrow}} w.$$

Действительно,

$$X_1 X_2 \ldots X_{m-p} A\, x' \overset{i}{\underset{rm}{\Rightarrow}} X_1 X_2 \ldots X_{m-p} Y_1 Y_2 \ldots Y_p\, x' =$$

$$= X_1 X_2 \ldots X_{m-p} X_{m-p+1} \ldots X_{m-1} X_m\, x' \overset{\pi'}{\underset{rm}{\Rightarrow}} w,$$

причём первый шаг этого вывода выполняется при помощи упомянутого правила номер i, а его завершение обеспечено индукционной гипотезой.

Вспомогательное утверждение II доказано.

В частности, при $\alpha = S$ и $x = \varepsilon$ заключаем, что если $(T_0, w, \varepsilon) \overset{*}{\underset{\mathcal{P}}{\vdash}} (T_0 S T, \varepsilon, \pi^R)$, где последняя конфигурация — *принимающая*, то $S \overset{\pi}{\underset{rm}{\Rightarrow}} w$.

Из рассуждений I и II следует утверждение теоремы.

Теорема 3.6. *Если* $G = (V_N, V_T, P, S)$ — $LR(k)$*-грамматика, то грамматика* G — *синтаксически однозначна.*

Доказательство ведётся от противного.

Пусть G — $LR(k)$-грамматика, но она не является синтаксически однозначной. Это значит, что существуют два *разных* право-сторонних вывода одной и той же терминальной цепочки:

Ret 2Ret
257

1) $S = \alpha_0 \underset{rm}{\Rightarrow} \alpha_1 \underset{rm}{\Rightarrow} \ldots \underset{rm}{\Rightarrow} \alpha_m = w$,

2) $S = \beta_0 \underset{rm}{\Rightarrow} \beta_1 \underset{rm}{\Rightarrow} \ldots \underset{rm}{\Rightarrow} \beta_n = w$, $w \in V_{\text{Т}}^*$.

Покажем, что тогда в этой грамматике нарушается $LR(k)$-условие (*см*. теор. 3.2.).

Найдём наименьшее i, при котором $\alpha_{m-i} \neq \beta_{n-i}$. С этой целью будем двигаться синхронно от последних сентенциальных форм $\alpha_m = \beta_n = w$ к началу этих выводов, приращивая i на 1 каждый раз, пока цепочки $\alpha_{m-i} \neq \beta_{n-i}$. Последнее значение i, полученное таким образом, является искомым.

Пусть такое $i \geq 1$ найдено. Тогда эти выводы фактически имеют вид:

$$1')\ S \xRightarrow[rm]{*} \underbrace{\alpha A x}_{\alpha_{m-i}} \xRightarrow[m]{} \alpha\beta x \xRightarrow[rm]{*} w,$$

$$2')\ S \xRightarrow[rm]{*} \underbrace{\gamma B y}_{\beta_{n-i}} \xRightarrow[rm]{} \gamma\delta y = \overbrace{\underbrace{\gamma\beta_1}_{\alpha\beta}\underbrace{\beta_2\, y}_{x}}^{\delta} = \alpha\beta x \xRightarrow[rm]{*} w,$$

где $\delta = \beta_1\beta_2$, $\gamma\beta_1 = \alpha\beta$, $\beta_2 y = x$, $\beta_2 \in V_\text{Т}^*$, причём так как $\alpha_{m-i} \neq \beta_{n-i}$, то $\alpha \neq \gamma \vee A \neq B \vee x \neq y$.

Из вывода 1′) очевидно, что $[A \to \beta., u] \in V_k^G(\alpha\beta)$, где $u \in \mathrm{FIRST}_k^G(x)$.

Аналогично из 2′) следует, что $[B \to \beta_1.\beta_2, v] \in V_k^G(\alpha\beta)$.

При этом

$$u \in \mathrm{FIRST}_k^G(x) = \mathrm{FIRST}_k^G(\beta_2 y) =$$

$$= \mathrm{FIRST}_k^G(\beta_2) \oplus_k \mathrm{FIRST}_k^G(y) =$$

$$= \mathrm{FIRST}_k(\beta_2) \oplus_k \{v\} = \mathrm{FIRST}_k(\beta_2 v) =$$

$$= \mathrm{EFF}_k^G(\beta_2 v).$$

Последнее равенство имеет место, так как $\beta_2 \in V_T{}^*$. Существование этих двух $LR(k)$-ситуаций, допустимых для одного и того же активного префикса $\alpha\beta$, означает нарушение необходимого и достаточного $LR(k)$-условия (*см.* теорему 3.1), что противоречит исходному предположению о том, что G — $LR(k)$-грамматика.

Следовательно теорема доказана.

Теорема 3.7. *Пусть* $G = (V_N, V_T, P, S)$ — $LR(k)$-*грамматика. Канонический* $LR(k)$-*анализатор, построенный по* G, *выполняя разбор цепочки* $x \in L(G)$, *совершает* $O(n)$ *движений, где* $n = | x |$.

Доказательство. Разбирая цепочку $x \in L(G)$, $LR(k)$-анализатор выполняет чередующиеся движения типа сдвиг (shift) и свёртка (reduce). Любое из этих движений на вершине магазина устанавливает некоторую $LR(k)$-таблицу.

Очевидно, что число сдвигов в точности равно n. Каждому сдвигу может предшествовать не более $\#\mathcal{T}$ сверток, где $\mathcal{T}$ — каноническое множество $LR(k)$-таблиц (состояний) анализатора. В противном случае какая-то из $LR(k)$-таблиц появилась бы повторно. При неизменной непросмотренной части входной цепочки это означало бы зацикливание анализатора.

Тогда вследствие теоремы 3.5 существовало бы как угодно много правосторонних выводов сколь угодно большой длины одной и той же цепочки w, что означало бы неоднозначность грамматики G.

На основании теоремы 3.6 грамматика G не являлась бы $LR(k)$-грамматикой, что противоречило бы первоначальному предположению.

Итак, общее число движений канонического $LR(k)$-анализатора

$$N \le n + c \times n = n \times (c + 1) = O(n),$$

где $c \le \#\mathcal{T}$; c — константа, зависящая от грамматики.

Что и требовалось доказать.

Замечание 3.6. $LR(k)$-анализатор на ошибочных цепочках “зациклиться” не может. Цепочка — ошибочная, если для некоторого её префикса не существует продолжения, дающего цепочку из $L(G)$.

Действительно, если бы анализатор зациклился, прочитав только часть входной цепочки, то, как мы только что выяснили, это означало бы, что грамматика G не есть $LR(k)$-грамматика.

§ 3.8. Простые постфиксные синтаксически управляемые *LR*-трансляции

Мы знаем, что простые семантически однозначные схемы синтаксически управляемых трансляций с входными $LL(k)$-грамматиками определяют трансляции, реализуемые детерминированными магазинными преобразователями.

Аналогичную ситуацию интересно рассмотреть в отношении схем с $LR(k)$-грамматиками в качестве входных. К сожалению, существуют такие простые семантически однозначные схемы, которые задают трансляции, не реализуемые детерминированными магазинными преобразователями.

Пример 3.13 (не постфиксной схемы).

Пусть имеется схема синтаксически управляемой трансляции
$T = (\{S\}, \{a, b\}, \{a, b\},$
$\{1)\ S \to Sa, aSa,\ 2)\ S \to Sb, bSb,\ 3)\ S \to \varepsilon, \varepsilon\})$.

Очевидно, что её входная грамматика — $LR(1)$. Действительно, каноническая система множеств допустимых $LR(1)$-ситуаций для этой грамматики $\mathcal{S} = \{\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}$ — не противоречива, ибо

Ret 267

$$\mathscr{A}_0 = \{[S' \to .S, \varepsilon], [S \to .Sa, \varepsilon|a|b],$$
$$[S \to .Sb, \varepsilon|a|b], [S \to ., \varepsilon|a|b]\},$$

$$\mathscr{A}_1 = \text{GOTO}(\text{A}_0, S) = \{[S' \to S., \varepsilon],$$
$$[S \to S.a, \varepsilon|a|b], [S \to S.b, \varepsilon|a|b]\},$$

$$\mathscr{A}_2 = \text{GOTO}(\text{A}_1, a) = \{[S \to Sa., \varepsilon|a|b]\},$$
$$\mathscr{A}_3 = \text{GOTO}(\text{A}_1, b) = \{[S \to Sb., \varepsilon|a|b]\},$$

и все $LR(1)$-ситуации в множествах $\mathscr{A}_0$ и $\mathscr{A}_1$ различны, а $\mathscr{A}_2$ и $\mathscr{A}_3$ имеют лишь одну $LR(1)$-ситуацию при соотвествующей аван-цепочке (*см*. теорему 3.1).

Этому множеству соответствует управляющая таблица $LR(1)$-анализатора (табл. 3.5).

Табл. 3.5

T	$f(u)$			$g(X)$		
	a	b	ε	S	a	b
T_0	reduce 3	reduce 3	reduce 3	T_1	error	error
T_1	shift	shift	accept	error	T_2	T_3
T_2	reduce 1	reduce 1	reduce 1	error	error	error
T_3	reduce 2	reduce 2	reduce 2	error	error	error

Пример 3.13 (не постфиксной схемы)

Построенный канонический $LR(1)$-анализатор для входной цепочки bba выдает правосторонний анализ $\pi^R(bba) = 3221$, т. к.

$$(T_0, bba, \varepsilon) \vdash_{\mathfrak{A}} (T_0ST_1, bba, 3) \vdash_{\mathfrak{A}}$$

$$\vdash_{\mathfrak{A}} (T_0ST_1bT_3, ba, 3) \vdash_{\mathfrak{A}} (T_0ST_1, ba, 32) \vdash_{\mathfrak{A}}$$

$$\vdash_{\mathfrak{A}} (T_0ST_1bT_3, a, 322) \vdash_{\mathfrak{A}} (T_0ST_1, a, 322) \vdash_{\mathfrak{A}}$$

$$\vdash_{\mathfrak{A}} (T_0ST_1aT_2, \varepsilon, 3221) \vdash_{\mathfrak{A}} (T_0ST_1, \varepsilon, 3221).$$

Данная схема задает трансляцию

$$\tau(T) = \{(w, w^R w) \mid w \in \{a, b\}^*\}.$$

В частности, имеем вывод

$$(S, S) \underset{rm}{\overset{(1)}{\Rightarrow}} (Sa, aSa) \underset{rm}{\overset{(2)}{\Rightarrow}} (Sba, abSba) \underset{rm}{\overset{(2)}{\Rightarrow}}$$

$$\underset{rm}{\overset{(2)}{\Rightarrow}} (Sbba, abbSbba) \underset{rm}{\overset{(3)}{\Rightarrow}} (bba, abbbba).$$

Пример 3.13 (не постфиксной схемы)

То, что в начале и в конце выходной цепочки должна быть порождена буква a, определяется лишь в момент, когда сканирование входной цепочки заканчивается и выясняется, что правилом (1) порождается буква a. Следовательно, выдача на выход может начаться только после того, как вся входная цепочка прочитана.

Естественный способ получить на выходе цепочку w^R — запомнить w в магазине, а затем выдать цепочку w^R на выход, выбирая её символы из магазина.

Далее требуется на выходе сгенерировать цепочку w, но в магазине, пустом к этому времени, нет для этого никакой информации.

Где ещё, помимо магазина, могла бы быть информация для восстановления цепочки w? — Только в состояниях управления детерминированного магазинного преобразователя (dpdt).

Но и там невозможно сохранить информацию о всей входной цепочке, так как она может быть сколь угодно большой длины, а число состояний dpdt всегда конечно.

Короче говоря, dpdt, который мог бы реализовать описанную трансляцию, не существует. Однако, если простая синтаксически управляемая трансляция с входной грамматикой класса $LR(k)$ не требует, чтобы выходная цепочка порождалась *до того*, как установлено, какое правило применяется, то соответствующая трансляция может быть реализована посредством dpdt.

Это приводит нас к понятию *постфиксной схемы синтаксически управляемой трансляции*.

Определение **3.16.** $T = (N, \Sigma, \Delta, R, S)$ называется *постфиксной схемой синтаксически управляемой трансляции*, если каждое её правило имеет вид $A \to \alpha, \beta$, где $\beta \in N^*\Delta^*$.

Теорема 3.8. *Пусть* $T = (N, \Sigma, \Delta, R, S)$ — *простая семантически однозначная постфиксная схема синтаксически управляемой трансляции с входной* $LR(k)$*-грамматикой.*

Существует детерминированный магазинный преобразователь P*, такой, что* $\tau(P) = \{(x\$, y) \mid (x, y) \in \tau(T)\}$.

Доказательство. По входной грамматике схемы T можно построить канонический $LR(k)$-анализатор, а затем моделировать его работу посредством dpdt P, накапливающего аванцепочку в состояниях и воспроизводящего действия shift и reduce i. При этом вместо выдачи на выходную ленту номера правила i он выдаёт выходные символы, входящие в состав семантической цепочки этого правила.

В момент принятия входной цепочки dpdt P переходит в конечное состояние.

Именно:
если правило с номером i есть $A \to \alpha, \beta z$, где $\beta \in N^*$, $z \in \Delta^*$, то dpdt P выдает цепочку z на выход.

Технические детали построения dpdt P и доказательство его адекватности sdts T оставляем в качестве самостоятельного упражнения.

Пример 3.14. Пусть имеется простая схема T с правилами
0) $S' \to S, S$; 1) $S \to SaSb, SSc$; 2) $S \to \varepsilon, \varepsilon$.

Входную грамматику этой схемы, являющуюся $LR(1)$-грамматикой во всех деталях мы обсуждали ранее. По ней была построена управляющая таблица адекватного канонического $LR(1)$-анализатора.

Эта же таблица может быть использована $LR(1)$-транслятором, который отличается от анализатора только тем, что вместо номера правила пишет на выходную ленту выходные символы из семантической цепочки этого правила.

Пусть имеется следующий вывод в данной схеме:

$$(S, S) \underset{rm}{\overset{(1)}{\Longrightarrow}} (SaSb, SSc) \underset{rm}{\overset{(1)}{\Longrightarrow}} (SaSaSbb, SSScc) \underset{rm}{\overset{(2)}{\Longrightarrow}}$$

$$\underset{rm}{\overset{(2)}{\Longrightarrow}} (SaSabb, SScc) \underset{rm}{\overset{(2)}{\Longrightarrow}} (Saabb, Scc) \underset{rm}{\overset{(2)}{\Longrightarrow}} (aabb, cc).$$

Руководствуясь табл. 3.3, $LR(1)$-транслятор совершает следующие движения:

$$(T_0, aabb, \varepsilon) \vdash_{\mathfrak{T}} (T_0ST_1, aabb, \varepsilon) \vdash_{\mathfrak{T}}$$

$$\vdash_{\mathfrak{T}} (T_0ST_1aT_2, abb, \varepsilon) \vdash_{\mathfrak{T}} (T_0ST_1aT_2ST_3, abb, \varepsilon) \vdash_{\mathfrak{T}}$$

$$\vdash_{\mathfrak{T}} (T_0ST_1aT_2ST_3aT_4, bb, \varepsilon) \vdash_{\mathfrak{T}}$$

$$\vdash_{\mathfrak{T}} (T_0ST_1aT_2ST3aT_4ST_6, bb, \varepsilon) \vdash_{\mathfrak{T}}$$

$$\vdash_{\mathfrak{T}} (T_0ST_1aT_2\boxed{ST_3aT_4ST_6bT_7}, b, \varepsilon) \vdash_{\mathfrak{T}}$$

$$\vdash_{\mathfrak{T}} (T_0ST_1aT_2ST_3, b, c) \vdash_{\mathfrak{T}}$$

Свёртка по правилу (1).

$$\vdash_{\mathfrak{T}} (T_0\boxed{ST_1aT_2ST_3bT_5}, \varepsilon, c) \vdash_{\mathfrak{T}}$$

$$\vdash_{\mathfrak{T}} (T_0ST_1, \varepsilon, cc).$$

Каждый раз, когда происходит свёртка по правилу 1 схемы, на выход посылается символ ‘c’.

В таблице 3.3 $LR(1)$-анализатора для входной грамматики схемы вставлены действия $LR(1)$-транслятора, приуроченные к упомянутым свёрткам.

§ 3.9. Простые непостфиксные синтаксически управляемые *LR*-трансляции

Предположим, что имеется простая, но ***не*** постфиксная sdts, входная грамматика которой есть $LR(k)$-грамматика.

Как реализовать такой перевод?

Один из возможных методов состоит в использовании многопросмотровой схемы перевода на базе нескольких dpdt.

Пусть $T = (N, \Sigma, \Delta, R, S)$ — простая семантически однозначная sdts с входной $LR(k)$-грамматикой G.

Для реализации трансляции, задаваемой схемой T, можно построить четырёх-уровневую схему перевода (рис. 3.3).

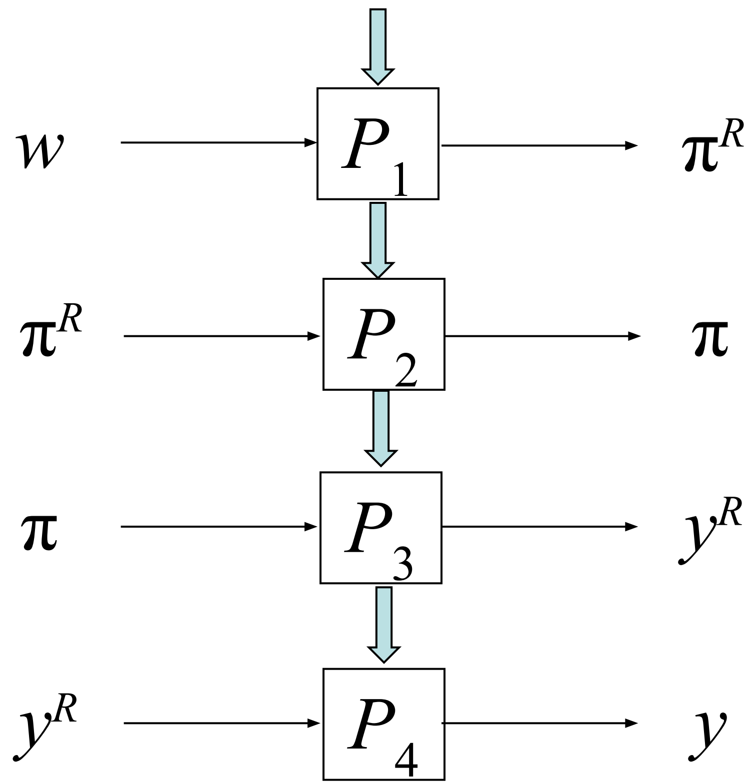

Рис. 3.3. Четырёхуровневая схема перевода.

Первый уровень занимает dpdt P_1. Его входом служит входная цепочка $w\$$, а выходом π^R — правосторонний анализ цепочки w.

На втором уровне dpdt P_2 обращает цепочку π^R. Для этого ему достаточно поместить всю цепочку π^R в магазин типа last-in-first-out и прочитать её из магазина, выдавая на выход. Получается цепочка π — последовательность номеров правил правостороннего вывода входной цепочки w.

На следующем 3-м этапе цепочка π используется для порождения соответствующей инвертированной выходной цепочки y^R правосторонним выводом в выходной грамматике схемы T.

Именно, получив на вход π — правосторонний анализ w, dpdt P_3 реализует перевод, определяемый простой sdts

$$T' = (N, \Sigma', \Delta, R', S),$$

где R' содержит правило вида

$$A \to iB_m B_{m-1} \ldots B_1, y_m B_m y_{m-1} B_{m-1} \ldots y_1 B_1 y_0$$

тогда и только тогда, когда

$$A \to x_0 B_1 x_1 \ldots B_m x_m, y_0 B_1 y_1 \ldots B_m y_m$$

— правило из R, а правило

$$A \to x_0 B_1 x_1 \ldots B_m x_m$$

есть правило номер i входной $LR(k)$-грамматики.

Нетрудно доказать, что $(\pi, y^R) \in \tau(T')$ тогда и только тогда, когда

$$(S, S) \overset{\pi}{\underset{rm}{\Rightarrow}} (w, y).$$

Схема T' — это простая sdts, основанная на $LL(1)$-грамматике, и, следовательно, её можно реализовать посредством dpdt P_3.

На четвертом уровне dpdt P_4 просто обращает цепочку y^R — выход P_3, записывая его в магазин типа first-in-last-out, а затем выдавая цепочку y из магазина на свой выход.

Число основных операций, выполняемых на каждом уровне, пропорционально длине цепочки w.

Таким образом, можно сформулировать следующий результат:

Теорема 3.9. *Трансляция, задаваемая простой семантически однозначной схемой синтаксически управляемой трансляции с входной LR(k)-грамматикой, может быть реализована за время, пропорциональное длине входной цепочки.*

Доказательство представляет собой формализацию вышеизложенного.

§ 3.10. *LALR*(*k*)-Грамматики

На практике часто используются частные подклассы *LR*(*k*)-грамматик, анализаторы для которых имеют более компактные управляющие таблицы по сравнению с таблицами канонического *LR*(*k*)-анализатора.

Здесь мы определим один из таких подклассов грамматик, называемых *LALR-грамматиками.*

***Определение* 3.17.** *Ядром LR-ситуации* $[A \to \beta_1.\beta_2, u]$ назовём $A \to \beta_1.\beta_2$, то есть правило грамматики с позицией в нём, отмеченной точкой.

Определим функцию CORE($\mathscr{A}$), где $\mathscr{A}$ — некоторое множество $LR(k)$-ситуаций, как *множество ядер*, входящих в состав $LR(k)$-ситуаций из $\mathscr{A}$.

***Определение* 3.18.** Пусть G — контекстно-свободная грамматика, $\mathscr{S}$ — каноническая система множеств $LR(k)$-ситуаций для грамматики G и

$$\mathscr{S}' = \left\{ \mathscr{A}' \;\middle|\; \begin{array}{l} \forall \mathscr{A} \in \mathscr{S}: \\ \mathscr{A}' = \bigcup\limits_{\mathscr{B} \in \mathscr{S}} \{\mathscr{B} \mid \mathrm{CORE}(\mathscr{B}) = \mathrm{CORE}(\mathscr{A})\} \end{array} \right\}.$$

Если каждое множество $\mathscr{A}' \in \mathscr{S}'$ — непротиворечиво, то G называется *LALR*(*k*)-*грамматикой*.

Другими словами, если слить все множества $LR(k)$-ситуаций с одинаковыми наборами ядер в одно множество и окажется, что все полученные таким образом множества $LR(k)$-ситуаций непротиворечивы, то G — $LALR(k)$-*грамматика.*

Число множеств, полученных при таком слиянии, разве лишь уменьшится. Соответственно уменьшится и число $LR(k)$-таблиц. Последние строятся обычным образом по объединённым множествам $LR(k)$-ситуаций.

Очевидно, что корректность $LALR(k)$-анализатора, использующего таким образом полученные таблицы, не нуждается в доказательстве.

Пример 3.15. Проверим, является ли рассмотренная раннее $LR(1)$-грамматика с правилами 0) $S' \to S$, 1) $S \to SaSb$, 2) $S \to \varepsilon$ $LALR(1)$- грамматикой.

Каноническая система множеств $LR(1)$-ситуаций для неё была построена в примере 3.10.

Именно, $\mathscr{S} = \{\mathscr{A}_0, \mathscr{A}_1, \mathscr{A}_2, \mathscr{A}_3, \mathscr{A}_4, \mathscr{A}_5, \mathscr{A}_6, \mathscr{A}_7\}$, где

$$\mathscr{A}_0 = \{[S' \to .S, \varepsilon], [S \to .SaSb, \varepsilon \mid a], [S \to ., \varepsilon \mid a]\};$$

$$\mathscr{A}_1 = \{[S' \to S.,\varepsilon],[S \to S.aSb, \varepsilon \mid a]\};$$

$$\mathscr{A}_2 = \{[S \to Sa.Sb, \varepsilon \mid a], [S \to .SaSb, a \mid b], [S \to ., a \mid b]\};$$

$$\mathscr{A}_3 = \{[S \to SaS.b, \varepsilon \mid a], [S \to S.aSb, a \mid b]\};$$

$$\mathscr{A}_4 = \{[S \to Sa.Sb, a \mid b], [S \to .SaSb, a \mid b], [S \to ., a \mid b]\};$$

$$\mathscr{A}_5 = \{[S \to SaSb., \varepsilon \mid a\};$$

$$\mathscr{A}_6 = \{[S \to SaS.b, a \mid b], [S \to S.aSb, a \mid b]\};$$

$$\mathscr{A}_7 = \{[S \to SaSb., a \mid b]\}.$$

Ясно, что в приведённой системе можно слить множества $\mathscr{A}_2$ и $\mathscr{A}_4$:

$$\mathscr{A}_{24} = \{[S \to Sa.Sb, \varepsilon \mid a \mid b],$$
$$[S \to .SaSb, a \mid b],$$
$$[S \to ., a \mid b]\},$$

множества $\mathscr{A}_3$ и $\mathscr{A}_6$:

$$\mathscr{A}_{36} = \{[S \to SaS.b, \varepsilon \mid a \mid b],$$
$$[S \to S.aSb, a \mid b]\},$$

а также множества $\mathscr{A}_5$ и $\mathscr{A}_7$:

$$\mathscr{A}_{57} = \{[S \to SaSb., \varepsilon \mid a \mid b]\}.$$

Полученные множества $\mathscr{A}_{24}$ $\mathscr{A}_{36}$, $\mathscr{A}_{57}$ и —не противоречивы.

Системе объединённых множеств $LR(1)$-ситуаций соответствует управляющая таблица:

Табл. 3.6

$LR(1)$-таблицы	$f(u)$			$g(X)$		
	a	b	ε	S	a	b
T_0	reduce 2		reduce 2	T_1		
T_1	shift		accept		T_{24}	
T_{24}	reduce 2	reduce 2		T_{36}		
T_{36}	shift	shift			T_4	T_{57}
T_{57}	reduce 1		reduce 1			

Отметим, что анализатор, использующий *LALR*(*k*)-таблицы, может чуть запаздывать с обнаружением ошибки по отношению к анализатору, использующему каноническое множество *LR*(*k*)-таблиц.

Например, канонический *LR*(1)-анализатор $\mathfrak{A}$ для рассматриваемой грамматики обнаруживает ошибку в цепочке *abb*, достигнув пятой конфигурации (*см*. табл. 3.3):

$$(T_0, abb, \varepsilon) \vdash_{\mathfrak{A}} (T_0ST_1, abb, 2) \vdash_{\mathfrak{A}} (T_0ST_1aT_2, bb, 2) \vdash_{\mathfrak{A}}$$

$$\vdash_{\mathfrak{A}} (T_0ST_1aT_2ST_3, bb, 22) \vdash_{\mathfrak{A}} (T_0ST_1aT_2ST_3bT_5, b, 22),$$

а $LALR(1)$-анализатор $\mathcal{Q}'$ —на шестой:

$$(T_0, abb, \varepsilon) \vdash_{\mathcal{Q}'} (T_0ST_1, abb, 2) \vdash_{\mathcal{Q}'} (T_0ST_1aT_{24}, bb, 2) \vdash_{\mathcal{Q}'}$$

$$\vdash_{\mathcal{Q}'} (T_0ST_1aT_2ST_{36}, bb, 22) \vdash_{\mathcal{Q}'}$$

$$\vdash_{\mathcal{Q}'} (T_0ST_1aT_{24}ST_{36}bT_{57}, b, 22) \vdash_{\mathcal{Q}'}$$

$$\vdash_{\mathcal{Q}'} (T_0ST_1, b, 221).$$