Глава 3

LR(*k*)-ГРАММАТИКИ И ТРАНСЛЯИИ

§ 3.1. Синтаксический анализ типа “снизу—вверх”

В предыдущей главе был рассмотрен класс $LL(k)$-грамматик, наибольший подкласс КС-грамматик, естественным образом детерминированно анализируемых в технике “сверху—вниз”, которая предполагает последовательное построение сентенциальных форм левостороннего вывода, начиная с начальной формы S, причем S — начальный нетерминал грамматики, и заканчивая конечной формой — анализируемой терминальной цепочкой. 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 $wA\alpha$ — текущая сентенциальная форма, где w — закрытая часть, $A\alpha$ — открытая часть и x — анализируемая цепочка, то правило определяется по нетерминалу A, множеству допустимых локальных правых контекстов $L = \mathrm{FIRST}_k^G(\alpha)$ и аванцепочке $u = \mathrm{FIRST}_k(y)$, причем $x = wy$. Адекватный тип анализатора — k-предсказывающий алгоритм анализа, выполняя эти шаги, воспроизводит открытые части сентенциальных форм в своем магазине.

В альтернативной терминологии такого рода анализатор выстраивает дерево вывода анализируемой цепочки, начиная с корня и на каждом шаге пристраивая очередное поддерево к крайнему левому нетерминальному листу строящегося дерева (рис. 3.1).

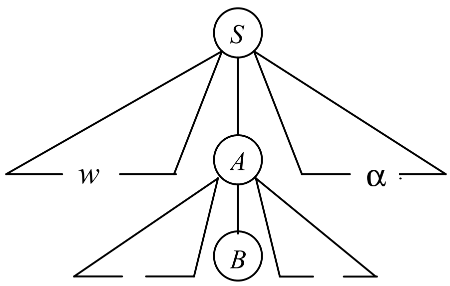

Рис. 3.1.

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

$$S = \alpha_0 \xRightarrow[rm]{p_1} \alpha_1 \xRightarrow[rm]{p_2} \ldots \xRightarrow[rm]{p_n} \alpha_n = x$$

— правосторонний вывод терминальной цепочки x в некоторой КС-граммати-ке. Индекс p_i ($i = 1, 2, \ldots, n$) 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 крайний правый нетерминал. Последовательность номеров правил $\pi^R = p_n \ldots \quad \ldots p_2\, p_1$ называется *правосторонним анализом 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: $\alpha_{i-1} = \gamma B z \Rightarrow \gamma\beta A y z = \alpha A w = \alpha_i$ посредством правила вида $B \rightarrow \beta A y$, где $\alpha = \gamma\beta$, $w = yz$. Как всюду в этом пособии мы обозначаем прописными буквами латинского алфавита нетерминалы, строчными латинскими буквами из конца алфавита — терминальные цепочки, а греческими буквами — цепочки над терминальными и нетерминальными символами.

Размещение сентенциальной формы в восходящем анализаторе показано в табл. 3.1. Восходящий анализатор располагает текущую сентенциальную форму α_i в магазине и на входной ленте таким образом, что в магазине располагается открытая ее часть, а закрытая представлена еще не прочитанной частью входной цепочки, которая начинается с текущего входного символа и простирается до конца этой цепочки.

Табл. 3.1

№	Магазин	Вход
1	$\alpha_i = \alpha A$	w
2	$\gamma\beta A$	yz
3	$\gamma\underline{\beta A y}$	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 она подчеркнута в строке 3.

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, где в магазине находится левая граница основы и по какому правилу (к какому нетерминалу) свернуть ее. Таким образом он воспроизводит предыдущую сентенциальную форму правостороннего вывода анализируемой цепочки. Если задача решается детерминированным анализатором, то свойства КС-грамматики, в которой производится анализ, должны гарантировать упомянутую однозначность,

Процесс заканчивается, когда в магазине остается один символ S, а входная цепочка прочитана вся.

Замечание 3.1. Если $S \underset{rm}{\overset{*}{\Rightarrow}} \alpha Aw$, 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 $\alpha_1 B\alpha_2 Aw$ и из нее 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 $V_N = \{S\}$, $V_T = \{a, b\}$, $P = \{(1)\ S \to SaSb, (2)\ S \to \varepsilon\}$.

Рассмотрим правосторонний вывод в этой грамматике:

$$S \underset{rm}{\overset{(1)}{\Rightarrow}} SaSb \underset{rm}{\overset{(1)}{\Rightarrow}} SaSaSbb \underset{rm}{\overset{(2)}{\Rightarrow}} SaSabb \underset{rm}{\overset{(2)}{\Rightarrow}} Saabb \underset{rm}{\overset{(2)}{\Rightarrow}} aabb.$$

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

Последовательность конфигураций восходящего анализатора во время разбора этой цепочки показана в табл. 3.2.

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	$*SaSaSb*	*b*	reduce 1
8	$*SaS*	*b*	shift
9	$*SaSb*		reduce 1
10	$*S*		

"Дно" магазина отмечено маркером $. Исходная конфигурация характеризуется тем, что магазин пуст (маркер "дна" не считается), непросмотренная часть входа — вся цепочка *aabb*. Первое действие — свертка 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. 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 путем заглядывания на k символов, следующих за основой. Именно: (1) производить сдвиг или свертку? (2) если делать свертку, 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 аванцепочки.

§ 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 $V_N' = V_N \cup \{S'\}$; $S' \notin V_N$; $V_T' = V_T$; $P' = P \cup \{S' \to S\}$.

***Определение* 3.2.** Пусть $G' = (V_N', V_T', P', S')$ — пополненная грамматика для контекстно-свободной грамматики G. Грамматика G является *LR*(*k*)-*грамматикой*, если для любых двух правосторонних выводов вида

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

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

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

Говоря неформально, если согласно первому выводу β — основа, сворачиваемая в нетерминал A, то и во втором выводе β должна быть основой, сворачиваемой в нетерминал A.

Из этого определения следует, что если имеется правовыводимая цепочка $\alpha_i = \alpha\beta w$, где β — основа, полученная из A, и если $\alpha\beta = X_1X_2\ldots X_r$, то

1) зная первые символы $X_1X_2\ldots X_j$ 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 $\alpha\beta = X_1 X_2 \ldots X_r$ и не более k символов цепочки w, мы можем быть уверены, что именно β является основой, сворачиваемой в нетерминал A;

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

Использование в определении $LR(k)$-грамматики пополненной грамматики существенно для однозначного определения конца анализа. Действительно, если грамматика использует S в правых частях правил, то свертка основы в S не может служить сигналом приема входной цепочки. Свертка же в S'в пополненной грамматике служит таким сигналом, поскольку S' нигде, кроме начальной сентенциальной формы, 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) $S \to Sa$,

3) $S \to a$.

Если игнорировать 0-е правило, то, не заглядывая в правый контекст основы Sa, можно сказать, что она должна сворачиваться в S. Аналогично основа a безусловно должна сворачиваться в S. Создается впечатление, что данная грамматика без 0-го правила есть $LR(0)$-грамматика.

С учетом же 0-го правила имеем:

1) $S' \underset{rm}{\Rightarrow} S$,

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

Сопоставив эти два вывода с образцом определения (образец приведен в рамке),

> Если существуют правосторонние выводы
>
> $\left.\begin{array}{l} 1)\ S' \overset{*}{\underset{rm}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta w, \\ 2)\ S' \overset{*}{\underset{rm}{\Rightarrow}} \gamma B x \underset{rm}{\Rightarrow} \alpha\beta y \end{array}\right\}$ в которых $\mathrm{FIRST}_k(w) = \mathrm{FIRST}_k(y)$,
>
> то должно быть $\alpha A y = \gamma B x$.

получаем $\alpha = \varepsilon$, $\beta = S$, $w = \varepsilon$, $A = S'$, $\gamma = \varepsilon$, $B = S$, $x = \varepsilon$, $y = a$, и при $k = 0$ $\mathrm{FIRST}_0(w) = \mathrm{FIRST}_0(\varepsilon) = \varepsilon$, $\mathrm{FIRST}_0(y) = \mathrm{FIRST}_0(a) = \varepsilon$, $\alpha A y = S'a$, $\gamma B x = S$, а так как $S'a \neq S$, то требование $\alpha A y = \gamma B x$ не выполняется.

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

Лемма 3.1. *Пусть* $G = (V_N, V_T, P, S)$ — $LR(\mathrm{k})$-*грамматика и* G' — пополненная грамматика для G. Если существуют правосторонние выводы в G' вида

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

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

в которых $\mathrm{FIRST}_k(w) = \mathrm{FIRST}_k(y)$, *то* $\gamma = \alpha$, $B = A$, $x = y$, $\beta' = \beta$.

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

Если $\gamma\beta' x = \alpha\beta y$, то в силу первых трех равенств имеем $\gamma\beta' x = \alpha\beta' y = \alpha\beta y$. Следовательно, $\beta' = \beta$. Что и требовалось доказать.

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, причем правый контекст для него всегда пуст. Очевидно также, что любая сентенциальная форма имеет вид $a^i C$, $a^i b$, $a^i D$, $a^i c$, где $i \geq 0$.

Пополненная грамматика содержит еще одно правило:

(0) $S' \to S$.

В роли нетерминала A могут быть при сопоставлении с образцом определения $LR(k)$-грамматик нетерминалы S' либо C, либо D. Отметим, что нетерминал S в роли нетерминала A нас не интересует, поскольку не существует двух разных правосентенциальных форм, в которых участвовал бы нетерминал S. В любом случае в роли цепочек w и x будет выступать пустая цепочка ε из-за праволинейности грамматики G.

Принимая все 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 правосторонних 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. Легко убедиться в том, что любая из указанных возможных сентенциальных форм рассматриваемой грамматики обладает этим свойством, и потому $LR(0)$-условие выполняется и, следовательно, G — $LR(0)$-грамматика.

Очевидно, что данная грамматика G не $LL(k)$-грамматика ни при каком k.

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 леволинейная грамматика порождает тот же самый язык, что и грамматика предыдущего примера, и она не является $LR(k)$-грамматикой ни при каком k. Действительно, рассмотрим, например, два таких правосторонних вывода:

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

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

Здесь цепочки $a^i b$ и $a^i c$ являются правыми контекстами для пустой основы, которая в одном случае сворачивается в нетерминал A, а в другом — в нетерми-

нетерминал B. В какой нетерминал сворачивать 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 нее на сколько угодно большое расстояние (в зависимости от выбора i). Следовательно, каким бы большим ни было k, всегда найдется такое i, что $\mathrm{FIRST}_k(a^i b) = \mathrm{FIRST}_k(a^i c)$, но $A \neq B$.

***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 дает возможности однозначно определить, в какой нетерминал следует ее сворачивать.

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

$$1)\, S \to AB,\ 2)\, A \to a,\ 3)\, B \to CD,\ 4)\, B \to aE,\ 5)\, C \to ab,\ 6)\, D \to bb,\ 7)\, E \to bba.$$

В этой грамматике рассмотрим два правосторонних вывода:

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

$$2)\ S' \underset{rm}{\Rightarrow} S \underset{rm}{\Rightarrow} AB \underset{rm}{\Rightarrow} AaE \underset{rm}{\Rightarrow} \overbrace{Aa\,\underline{b}}^{\alpha\beta}\overbrace{\underline{ba}}^{y}.$$

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

§ 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:

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

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

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.

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

Вход: $G = (V_N, V_T, P, S)$ — $LR(k)$-грамматика; $\mathcal{T}$ — множество $LR(k)$-таблиц для G; $T_0 \in \mathcal{T}$ — начальная $LR(k)$-таблица; $x \in V_T^*$ — входная цепочка.

Выход: π^R — правосторонний анализ x.

Метод.

$LR(k)$-Анализатор реализует классический механизм "сдвиг", описанный в параграфе §3.1. Его действия будем описывать в терминах конфигураций, понимая под конфигурацией тройку (α, w, y), где $\alpha \in (V_N \cup V_T \cup \mathcal{T})^*$ — 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:

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 $u = \text{FIRST}_k(w)$.

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

1.1.1. $g(a) = T'$. Тогда $LR(k)$-анализатор совершает следующее 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. $w = \varepsilon$. В этом случае $u = \varepsilon$ и $f(u) = f(\varepsilon) = \text{error}$, так как сдвиг из пустой цепочки в магазин невозможен. Анализатор сигнализирует об ошибке и останавливается.

2: свертка. Пусть текущая конфигурация есть $(\gamma TX_1T_1X_2T_2\ldots X_mT_m, w, \pi)$, где $T, T_1, T_2, \ldots, T_m \in \mathcal{T}$ — некоторые $LR(k)$-таблицы; $T_m = (f_m, g_m)$, $f_m(u) = \text{reduce } i$, $A \to \alpha$ — i-е правило из множества правил P, где $\alpha = X_1X_2\ldots X_m$ — основа, $u = \text{FIRST}_k(w)$, и пусть $T = (f, g)$.

2.1. $g(A) = T'$, где $T' \in \mathcal{T}$ — некоторая $LR(k)$-таблица. В этом случае анализатор совершает переход

$$(\gamma T\underbrace{X_1T_1X_2\ldots X_mT_m}_{\text{Свертка в } A}, w, \pi) \vdash (\gamma TA, w, \pi i) \vdash (\gamma TAT', w, \pi i),$$

воспроизводящий свертку.

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; $u = \text{FIRST}_k(w)$, $T = (f, g)$ и $f(u) = \text{error}$. Анализатор сигнализирует об ошибке и останавливается.

4: прием. Пусть текущая конфигурация есть $(T_0ST, \varepsilon, \pi^R)$, $T = (f, g)$, $f(\varepsilon) = \text{accept}$. Анализатор сигнализирует о приеме цепочки x и останавливается. Выходная цепочка π^R представляет правосторонний анализ цепочки x.

П р и м е р 3.6. Обратимся еще раз к грамматике из примера 3.1. Как мы увидим далее, она — $LR(1)$-грамматика. Она имеет следующие правила: 1) $S \to SaSb$, 2) $S \to \varepsilon$. Управляющая таблица $LR(1)$-анализатора, построенная для нее, имеет вид, представленный табл. 3.3, где пустые клетки соответствуют значениям error.

Табл. 3.3

$LR(1)$-таблицы	$f(u)$			$g(X)$		
	a	b	ε	S	a	b
T_0	reduce 2		reduce 2	T_1		
T_1	shift		accept		T_2	
T_2	reduce 2	reduce 2		T_3		
T_3	shift	shift			T_4	T_5
T_4	reduce 2	reduce 2		T_6		
T_5	reduce 1		reduce 1			
T_6	shift	shift			T_4	T_7
T_7	reduce 1	reduce 1				

Рассмотрим действия этого анализатора на входной цепочке *aabb*:

$$(T_0, aabb, \varepsilon) \vdash (T_0ST_1, aabb, 2) \vdash (T_0ST_1aT_2, abb, 2) \vdash (T_0ST_1aT_2ST_3, abb, 22) \vdash$$
$$\vdash (T_0ST_1aT_2ST_3aT_4, bb, 22) \vdash (T_0ST_1aT_2ST_3aT_4ST_6, bb, 222) \vdash$$
$$\vdash (T_0ST_1aT_2ST_3aT_4ST_6bT_7, b, 222) \vdash (T_0ST_1aT_2ST_3, b, 2221) \vdash$$
$$\vdash (T_0ST_1aT_2ST_3bT_5, \varepsilon, 2221) \vdash (T_0ST_1, \varepsilon, 22211).$$

Итак, цепочка *aabb* принимается, и $\pi^R = 22211$ — ее правосторонний анализ.

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

***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*.

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

Пример 3.7. Обратимся еще 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 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 $[S \to C., \varepsilon]$.

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

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

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

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

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

Во всех случаях правый контекст основы пуст.

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

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

2) $S' \underset{rm}{\overset{*}{\Rightarrow}} \gamma B x \underset{rm}{\Rightarrow} \gamma\delta x = \alpha\beta y$, *такие, что* $x, y, w \in V_T^*$ *и*

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

б) $\gamma B x \neq \alpha A y$,

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

Доказательство. Если G — не $LR(k)$-грамматика, то условия а) и б) выполняются как прямое следствие из определения $LR(k)$-грамматик. Условие в) неформально означает, что правая граница основы в последней сентенциальной форме второго вывода удалена от ее начала, по крайней мере, не менее, чем удалена основа в последней сентенциальной форме в первом выводе. Это условие не столь очевидно. Простой обмен ролями этих двух выводов ничего не дает, так как этим приемом мы добъемся 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 трем условиям.

Поскольку $\gamma\delta x = \alpha\beta y$, 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 ее 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.

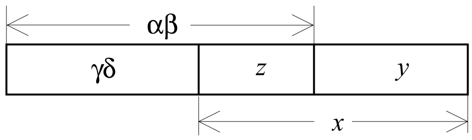

Рис. 3.2.

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

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

Второй вывод разметим по образцу первого с учетом равенства $x = zy$, а первый разметим по образцу второго с учетом равенства (3.1):

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

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

Эти два вывода обладают следующими особенностями:

а′) $\mathrm{FIRST}_k([w]) = \mathrm{FIRST}_k([y])$, так как $[w] = zy$, $[y] = zw$ и

$\mathrm{FIRST}_k(w) = \mathrm{FIRST}_k(y)$;

б′) $[\gamma][B][x] \neq [\alpha][A][y]$, так как $[\gamma] = \alpha$, $[B] = A$, $[x] = w$, $[\alpha]=\gamma$, $[A]=B$, $[y]=zw$, 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 трем условиям а′), б′) и в′). Что и требовалось доказать.

Введем функцию $\mathrm{EFF}_k^G(\alpha)$, где $\alpha \in (V_{\mathrm{N}} \cup V_{\mathrm{T}})^*$, необходимую для построения $LR(k)$-анализатора. Она будет помогать при определении, является ли ε-це-почка основой для данной сентенциальной формы, подлежащей свертке.

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

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

Говоря неформально, функция $\mathrm{EFF}_k^G(\alpha)$ ничем не отличается от функции $\mathrm{FIRST}_k^G(\alpha)$, если α начинается с терминала, а если α начинается с нетерминала, то при правостороннем выводе именно он является последним нетерминалом, который замещается терминальной цепочкой на последнем шаге вывода. Функция EFF_k^G отличается от функции FIRST_k^G тем, что в первую не включаются префиксы терминальных цепочек, если эти цепочки получены посредством ε-правила, примененного на последнем шаге вывода.

Пример 3.8. Рассмотрим контекстно-свободную 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. 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 $S \underset{rm}{\Rightarrow} AB \overset{*}{\underset{rm}{\Rightarrow}} Aw$, где $w \in V_T^*$ — некоторая терминальная цепочка. Далее возможны следующие продолжения:

$$S \underset{rm}{\Rightarrow} AB \overset{*}{\underset{rm}{\Rightarrow}} Aw \underset{rm}{\Rightarrow} \begin{cases} \begin{cases} \underset{rm}{\Rightarrow} Cbaw \begin{cases} \underset{rm}{\Rightarrow} cbaw, \\ \underset{rm}{\Rightarrow} baw \text{ — недопустимо (ε-правило);} \end{cases} \\ \underset{rm}{\Rightarrow} Caw \begin{cases} \underset{rm}{\Rightarrow} caw, \\ \underset{rm}{\Rightarrow} aw \text{ — недопустимо (ε-правило);} \end{cases} \end{cases} \\ w \text{ — недопустимо (ε-правило).} \end{cases}$$

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

Теорема 3.1. *Чтобы контекстно-свободная грамматика $G = (V_N, V_T, P, S)$ была $LR(k)$-грамматикой, необходимо и достаточно, чтобы выполнялось следующее условие: если $[A \to \beta.,\ u]$ — $LR(k)$-ситуация, допустимая для активного префикса $\alpha\beta$ правосентенциальной формы $\alpha\beta w$ пополненной грамматики G', где $u = \mathrm{FIRST}_k(w)$, то не существует никакой другой $LR(k)$-ситуации $[A_1 \to \beta_1.\beta_2,\ v]$, которая отличается от данной и допустима для активного префикса $\alpha\beta$ при условии, что $u \in \mathrm{EFF}_k^G(\beta_2 v)$ (в частности, и при $\beta_2 = \varepsilon$).*

Доказательство.

Необходимость. Предположим, что G — $LR(k)$-грамматика. Докажем, что тогда условие выполняется.

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

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

Рассуждая от противного, предположим, что $[A_1 \to \beta_1.\beta_2, v]$ — *другая* $LR(k)$-ситуация, допустимая для активного префикса $\alpha\beta$, причем $u \in \mathrm{EFF}_k^G(\beta_2 v)$.

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

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

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

Кроме того, выполняется условие

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

Условие 3) $u \in \mathrm{EFF}_k^G(\beta_2 v)$ означает согласно определению EFF_k^G, что существует вывод вида $\beta_2 v \underset{rm}{\overset{*}{\Rightarrow}} uz$, в котором $|u| = k$ и $z \in V_T^*$ или $|u| < k$ и $z = \varepsilon$, причем если предпоследняя сентенциальная форма в этом выводе начинается с нетерминала, то он заменяется непустой терминальной цепочкой.

Рассмотрим три возможных варианта состава цепочки β_2: 1) $\beta_2 = \varepsilon$; 2) $\beta_2 \in V_T^+$; 3) β_2 содержит нетерминалы.

<u>Вариант 1</u>: $\beta_2 = \varepsilon$. В этом случае из условия 3) следует, что $u \in \mathrm{EFF}_k^G(v)$ и, следовательно, $u = v$. Соответственно вывод 2) фактически имеет вид

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

Отсюда $\alpha_1\beta_1 = \alpha\beta$ и $x = y$; соответственно $\mathrm{FIRST}_k(y) = \mathrm{FIRST}_k(x) = v = u$.

Так как $\mathrm{FIRST}_k(w) = u$, то имеем два правосторонних вывода: 1) и 2′), в которых $\mathrm{FIRST}_k(w) = \mathrm{FIRST}_k(y)$.

По предположению $[A \to \beta., u] \neq [A_1 \to \beta_1.\beta_2, v]$, причем $u = v$. 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

$$\alpha_1 A_1 x = \alpha A y, \tag{3.4}$$

в котором $x = y$. При $A \neq A_1$ это невозможно. Если же $A = A_1$, то имеем $\alpha_1 \neq \alpha$, поскольку в этом случае $\beta \neq \beta_1$ при том, что $\alpha_1\beta_1 = \alpha\beta$. Итак, условие (3.4) не соблюдается, и G — не $LR(k)$-грамматика вопреки первоначальному предположению. Полученное противоречие доказывает необходимость сформулированного в теореме условия в данном частном случае.

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

2″) $S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha_1 A_1 x \underset{rm}{\Rightarrow} \alpha_1\beta_1\beta_2 x = \alpha_1\beta_1 zx = \alpha\beta y$, так что $\alpha_1\beta_1 = \alpha\beta$, $y = zx$ и, кроме того, предполагается, что

3″) $u \in \mathrm{EFF}_k^G(\beta_2 x) = \mathrm{EFF}_k^G(zx) = \mathrm{EFF}_k^G(y) = \mathrm{FIRST}_k(y)$, т.е. $u = \mathrm{FIRST}_k(y)$.

Итак, в выводах 1) и 2) $\mathrm{FIRST}_k(w) = \mathrm{FIRST}_k(y)$.

Чтобы грамматика G была $LR(k)$-грамматикой, должно быть $\alpha_1 A_1 x = \alpha A y$ или, что то же самое, $\alpha_1 A_1 x = \alpha A z x$, 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.

<u>Вариант 3</u>: цепочка β_2 не пуста и содержит, по крайней мере, один нетерминал. Пусть B — самый левый из них, т.е. $\beta_2 = u_1 B\delta$, где $u_1 \in V_\mathrm{T}^*$, $B \in V_\mathrm{N}$, $\delta \in (V_\mathrm{N} \cup V_\mathrm{T})^*$. Имеем:

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

2) $S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha_1 A_1 x \underset{rm}{\Rightarrow} \alpha_1\beta_1\beta_2 x = \alpha\beta\beta_2 x$, где $\alpha_1\beta_1 = \alpha\beta$, $\mathrm{FIRST}_k(x) = v$.

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

В этом варианте из условия 3) следует, что существует правосторонний вывод $\beta_2 v = u_1 B\delta v \underset{rm}{\overset{*}{\Rightarrow}} u_1 B u_3 v \underset{rm}{\overset{*}{\Rightarrow}} u_1 y_1 B_1 y_2 u_3 v \underset{rm}{\Rightarrow} u_1 y_1 u_2 y_2 u_3 v$, $u = \mathrm{FIRST}_k(u_1 y_1 u_2 y_2 u_3 v)$. Здесь на последнем шаге используется правило $B_1 \to u_2$, где $u_2 \in V_\mathrm{T}^*$ с гарантией, что $u_1 y_1 u_2 \neq \varepsilon$. Действительно, если $u_1 y_1 = \varepsilon$, то в соответствии с определением EFF_k^G должно быть $u_2 \neq \varepsilon$.

Продолжим вывод 2), учитывая, что $\beta_2 = u_1 B\delta \underset{rm}{\overset{*}{\Rightarrow}} u_1 B u_3 \underset{rm}{\Rightarrow} u_1 y_1 u_2 y_2 u_3$:

2′) $S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha_1 A_1 x \underset{rm}{\Rightarrow} \alpha_1\beta_1\beta_2 x = \alpha\beta u_1 B\delta x \underset{rm}{\overset{*}{\Rightarrow}} \alpha\beta u_1 B u_3 x \underset{rm}{\overset{*}{\Rightarrow}} \alpha\beta u_1 y_1 B_1 y_2 u_3 x \underset{rm}{\Rightarrow}$

$\underset{rm}{\Rightarrow} \alpha\beta u_1 y_1 u_2 y_2 u_3 x = \alpha\beta y$, где $y = u_1 y_1 u_2 y_2 u_3 x$, $v = \mathrm{FIRST}_k(x)$.

Вычислим $\mathrm{FIRST}_k(y)$:

$$\mathrm{FIRST}_k(y) = \mathrm{FIRST}_k(u_1 y_1 u_2 y_2 u_3 x) = \mathrm{FIRST}_k(u_1 y_1 u_2 y_2 u_3 v) = u.$$

Итак, $\mathrm{FIRST}_k(w) = u$, $\mathrm{FIRST}_k(y) = u$, т.е. $\mathrm{FIRST}_k(w) = \mathrm{FIRST}_k(y)$, и остается проверить, выполняется ли $LR(k)$-условие $\alpha\beta u_1 y_1 B_1 y_2 u_3 x = \alpha A u_1 y_1 u_2 y_2 u_3 x$. Для выполнения этого равенства необходимо, по крайней мере, чтобы $y_2 u_3 x = u_1 y_1 u_2 y_2 u_3 x$, т.е. $u_1 y_1 u_2 = \varepsilon$, чего, как мы выяснили, нет ($u_1 y_1 u_2 \neq \varepsilon$).

Таким образом, $LR(k)$-условие нарушено, и G — не $LR(k)$-грамматика вопреки исходному предположению.

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. Согласно лемме 3.2 существуют два вывода вида

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

2) $S' \underset{rm}{\overset{*}{\Rightarrow}} \gamma B x \underset{rm}{\Rightarrow} \gamma\delta x = \alpha\beta y$, такие, что $x, y, w \in V_{\text{т}}^*$ и

а) $\text{FIRST}_k(w) = \text{FIRST}_k(y)$,

б) $\gamma B x \neq \alpha A y$,

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 нем явно начальный участок, на котором получается *последняя* цепочка, причем ее открытая часть не длиннее $|\alpha\beta| + 1$:

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

здесь $|\alpha_1 A_1| \leq |\alpha\beta| + 1$ или, что то же самое, $|\alpha_1| \leq |\alpha\beta| \leq |\gamma\delta|$, $A_1 \to \beta_1\beta_2 \in P$. Цепочка $\beta_1\beta_2$ — основа сентенциальной формы $\alpha_1\beta_1\beta_2 y_1$, причем β_1 — ее префикс такой длины, что выполняется равенство $|\alpha_1\beta_1| = |\alpha\beta|$.

Отметим, что активный префикс длиной $|\alpha\beta|$, 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 $|\alpha\beta|$. Очевидно, что в выводе 2′) $\beta_2 y_1 \underset{rm}{\overset{*}{\Rightarrow}} y$ и $\alpha_1\beta_1 = \alpha\beta$. Таким образом, вывод 2′) можно переписать в следующем виде:

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

Из факта существования вывода 1) следует, что $LR(k)$-ситуация $[A \to \beta., u]$, где $u = \text{FIRST}_k(w)$, допустима для активного префикса $\alpha\beta$ правосентенциальной формы $\alpha\beta w$.

Аналогично из факта существования вывода 2″) следует, что $LR(k)$-ситуа-ция $[A_1 \to \beta_1.\beta_2, v]$, где $v = \text{FIRST}_k(y_1)$, допустима для активного префикса $\alpha\beta$ правосентенциальной формы $\alpha\beta\beta_2 y_1$.

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

$$u = \text{FIRST}_k(w) = \text{FIRST}_k(y) \in \text{FIRST}_k^G(\beta_2 y_1) = \text{FIRST}_k^G(\beta_2) \oplus_k \text{FIRST}_k^G(y_1) =$$
$$= \text{FIRST}_k^G(\beta_2) \oplus_k \{v\} = \text{FIRST}_k^G(\beta_2) \oplus_k \text{FIRST}_k^G(v) = \text{FIRST}_k^G(\beta_2 v).$$

Остается показать, что $u \in \text{EFF}_k^G(\beta_2 v)$. Действительно, если бы $u \notin \text{EFF}_k^G(\beta_2 v)$, то только потому, что цепочка β_2 начилась бы с нетерминала, который на последнем шаге вывода $\beta_2 y_1 \underset{rm}{\overset{*}{\Rightarrow}} y$ замещался бы ε-цепочкой. Сопоставим исходное представление вывода 2) с его же представлением в виде 2″):

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

и

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

Последним замещаемым нетерминалом в этом выводе является B, причем эта-то последняя замена и дает 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 $B \rightarrow \varepsilon$, то вывод 2″) можно переписать следующим образом:

$$S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha_1 A_1 y_1 \underset{rm}{\Rightarrow} \alpha_1\beta_1\beta_2 y_1 = \alpha_1\beta_1 A_2\alpha_2 y_1 \underset{rm}{\overset{*}{\Rightarrow}} \alpha_1\beta_1 A_2 y_2 y_1 \underset{rm}{\overset{*}{\Rightarrow}} \alpha_1\beta_1 A_m\alpha_m y_{m-1}\dots y_2 y_1 \underset{rm}{\overset{*}{\Rightarrow}}$$
$$\underset{rm}{\overset{*}{\Rightarrow}} \alpha_1\beta_1 A_m y_m y_{m-1}\dots y_2 y_1 = \alpha_1\beta_1 B y_m y_{m-1}\dots y_2 y_1 \underset{rm}{\Rightarrow} \alpha_1\beta_1 y_m y_{m-1}\dots y_2 y_1 = \alpha\beta y,$$

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

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 $\alpha\beta$: $[A \rightarrow \beta., u]$ и $[A_1 \rightarrow \beta_1.\beta_2, v]$ 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, то должно быть $[A \rightarrow \beta., u] = [A_1 \rightarrow \beta_1.\beta_2, v]$. 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 учетом этого вывод 2″) можем переписать так:

2″) $S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha A y_1 \underset{rm}{\Rightarrow} \alpha\beta y_1 = \alpha\beta y$, 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. Следовательно, G — $LR(k)$-грамматика. Достаточность доказана. Теорема также доказана.

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

$$V_k^G(\gamma) = \{[A \rightarrow \beta_1.\beta_2,\ u] \mid \exists A \rightarrow \beta_1\beta_2 \in P; \exists S' \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta_1\beta_2 w, \gamma = \alpha\beta_1,$$
$$u = \mathrm{FIRST}_k(w)\}.$$

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

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

Вход: $G = (V_N, V_T, P, S)$ — контекстно-свободная грамматика, $\gamma = X_1X_2\dots X_m$, $X_i \in V_N \cup V_T$, $i = 1, 2, \dots, m$; $m \geq 0$.

Выход: множество $V_k^G(\gamma)$.

Метод.

Будем строить последовательность множеств $V_k^G(\varepsilon)$, $V_k^G(X_1)$, $V_k^G(X_1X_2)$,…, $V_k^G(X_1X_2...X_m)$

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 $B\in V_\text{N}$, и существует $B\to\beta\in P$, то $[B\to.\beta,\ v]$, где $v\in \text{FIRST}_k^G(\alpha u)$, тоже включается в множество $V_k^G(\varepsilon)$.

в) Шаг б) повторяется до тех пор, пока во множестве $V_k^G(\varepsilon)$ не будут просмотрены все имеющиеся в нем $LR(k)$-ситуации. Замыкание завершается за конечное число шагов, так как множества P и V_T^{*k} конечны.

2. Строится следующий элемент последовательности. Пусть множества $V_k^G(X_1X_2...X_i)$, где $0 \le i < m$, уже построены. Покажем, как построить следующее множество $V_k^G(X_1X_2...X_{i+1})$.

а) Инициализация: если $[A \to \beta_1.X_{i+1}\beta_2,\ u]\in V_k^G(X_1X_2...X_i)$, то $LR(k)$-ситуация $[A\to\beta_1X_{i+1}.\beta_2,\ u]$ включается в множество $V_k^G(X_1X_2...X_{i+1})$.

б) Замыкание $V_k^G(X_1X_2...X_{i+1})$: если $\{[A\to\beta_1.B\beta_2,\ u]\in V_k^G(X_1X_2...X_{i+1})$, где $B\in V_\text{N}$, и существует $B\to\beta\in P$, то $[B\to.\beta,\ v]$, где $v\in \text{FIRST}_k^G(\beta_2 u)$, тоже включается в множество $V_k^G(X_1X_2...X_{i+1})$.

в) Шаг б) повторяется до тех пор, пока во множестве $V_k^G(X_1X_2...X_{i+1})$ не будут просмотрены все имеющиеся в нем $LR(k)$-ситуации. Замыкание завершается за конечное число шагов, так как множества P и V_T^{*k} конечны.

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

4. Процесс завершается при $i = m$; $V_k^G(\gamma) = V_k^G(X_1X_2...X_m)$.

Замечание 3.2. Алгоритм 3.2 не требует использования пополненной грамматики.

***Определение* 3.9.** Пусть $G = (V_\text{N}, V_\text{T}, P, S)$ — контекстно-свободная грамматика. На множестве $LR(k)$-ситуаций в этой грамматике определим функцию: $\text{GOTO}(\mathscr{A}, X) = \mathscr{A}'$, где $\mathscr{A} = V_k^G(\gamma)$ — некоторое множество $LR(k)$-ситуаций, допустимых для активного префикса $\gamma\in (V_\text{N}\cup V_\text{T})^*$; $X\in V_\text{N}\cup V_\text{T}$; $\mathscr{A}' = V_k^G(\gamma X)$.

Очевидно, что эта функция строится попутно с построением множеств $V_k^G(\gamma)$ на шаге 2 алгоритма 3.2: если множество $V_k^G(X_1X_2...X_i)$ уже построено, то

$$V_k^G(X_1X_2...X_iX_{i+1}) = \text{GOTO}(V_k^G(X_1X_2...X_i), X_{i+1}).$$

Остается ввести лишь обозначения: $\gamma = X_1X_2...X_i$, $X = X_{i+1}$, $\mathscr{A} = V_k^G(X_1X_2...X_i)$, $\mathscr{A}' = V_k^G(X_1X_2...X_iX_{i+1})$, чтобы увидеть, как это делается.

Замечание 3.3. Важно отметить, что результат функции GOTO($\mathscr{A}, X$) = $\mathscr{A}'$, где $\mathscr{A}$ = $V_k^G(X_1X_2...X_i)$, зависит не от $X_1X_2...X_i$, а от $V_k^G(X_1X_2...X_i)$.

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

множество $V_1^G(\varepsilon)$ пополняется ситуациями $[S \to .SaSb, a]$, $[S \to ., a]$.

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

множество $V_1^G(Sa)$ пополняется ситуациями $[S \to .SaSb, a]$ и $[S \to ., a]$.

Здесь шаг б) замыкания выполнялся дважды.

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

Теорема 3.2. Алгоритм 3.2 правильно вычисляет $V_k^G(\gamma)$, где $\gamma \in (V_N \cup V_T)^*$ — активный префикс правосентенциальной 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 $S \underset{\text{rm}}{\overset{*}{\Rightarrow}} \alpha A w \underset{\text{rm}}{\Rightarrow} \alpha\beta_1\beta_2 w$, в котором $\alpha\beta_1 = \gamma$ (γ — активный префикс правосентенциальной формы $\alpha\beta_1\beta_2 w$), а $\text{FIRST}_k(w) = u$ (цепочка u есть правый контекст основы $\beta_1\beta_2$ в данной сентенциальной форме $\alpha\beta_1\beta_2 w$).

I. Докажем сначала, что если $\gamma = \alpha\beta_1$ — активный префикс сентенциальной формы $\alpha\beta_1\beta_2 w$ в выводе $S \underset{\text{rm}}{\overset{*}{\Rightarrow}} \alpha A w \underset{\text{rm}}{\Rightarrow} \alpha\beta_1\beta_2 w$, то $LR(k)$-ситуация $[A \to \beta_1.\beta_2, u]$, где $u = \text{FIRST}_k(w)$, содержится в множестве $V_k^G(\gamma)$.

Индукция по $l = |\gamma|$.

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

Имеем вывод $S \underset{\text{rm}}{\overset{*}{\Rightarrow}} \alpha A w \underset{\text{rm}}{\Rightarrow} \alpha\beta_1\beta_2 w$, $\gamma = \alpha\beta_1 = \varepsilon$. Фактически $\alpha = \beta_1 = \varepsilon$, и вывод имеет вид $S \underset{\text{rm}}{\overset{*}{\Rightarrow}} Aw \underset{\text{rm}}{\Rightarrow} \beta_2 w$, где на последнем шаге применялось правило $A \to \beta_2$. Во всех деталях он мог бы быть только таким:

$$S \underset{\text{rm}}{\Rightarrow} A_1\alpha_1 \underset{\text{rm}}{\overset{*}{\Rightarrow}} A_1 w_1 \underset{\text{rm}}{\Rightarrow} A_2\alpha_2 w_1 \underset{\text{rm}}{\overset{*}{\Rightarrow}} A_2 w_2 w_1 \underset{\text{rm}}{\Rightarrow} \ldots \underset{\text{rm}}{\Rightarrow} A_m\alpha_m w_{m-1}\ldots w_1 \underset{\text{rm}}{\overset{*}{\Rightarrow}}$$
$$\underset{\text{rm}}{\overset{*}{\Rightarrow}} A_m w_m w_{m-1}\ldots w_1 = Aw \underset{\text{rm}}{\Rightarrow} \beta_2 w.$$

Следовательно, $w = w_m w_{m-1}\ldots w_1$, $A = A_m$ 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, $\text{FIRST}_k(w_i) \in \text{FIRST}_k^G(\alpha_i)$, поскольку $\alpha_i \underset{\text{rm}}{\overset{*}{\Rightarrow}} w_i$, $i = 1, 2, \ldots, m$.

Согласно алгоритму 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 $v_1 \in \text{FIRST}_k^G(\alpha_1\varepsilon)$,

в частности для $v_1 = \text{FIRST}_k(w_1)$;

$[A_2 \to .A_3\alpha_3, v_2] \in V_k^G(\varepsilon)$ для любой $v_2 \in \text{FIRST}_k^G(\alpha_2 v_1)$,

в частности для $v_2 = \text{FIRST}_k(w_2 v_1) = \text{FIRST}_k(w_2 w_1)$;

…

$[A_{m-1} \to A_m\alpha_m, v_{m-1}] \in V_k^G(\varepsilon)$ для любой $v_{m-1} \in \text{FIRST}_k^G(\alpha_{m-1}v_{m-2})$,

в частности для $v_{m-1} = \text{FIRST}_k(w_{m-1}v_{m-2}) = \text{FIRST}_k(w_{m-1}\ldots w_1)$.

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 $v_m \in \text{FIRST}_k^G(\alpha_m v_{m-1})$,

в частности для $v_m = \text{FIRST}_k(w_m v_{m-1}) = \text{FIRST}_k(w_m w_{m-1}\ldots w_1)$.

Принимая во внимание, что $w = w_m w_{m-1}\ldots w_1$ и что $u = \text{FIRST}_k(w)$, 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:

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

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

Случай 1: $\beta_1 \neq \varepsilon$.

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 $\beta_1' \in (V_N \cup V_T)^*$. В этом случае имеем

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

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

Шаг 2а алгоритма 3.2 дает $[A \rightarrow \beta_1' X.\beta_2,\ u] = [A \rightarrow \beta_1.\beta_2,\ u] \in V_k^G(\gamma' X) = V_k^G(\gamma)$, что и требовалось доказать.

Случай 2: $\beta_1 = \varepsilon$.

Имеем $S \underset{rm}{\overset{*}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta_1\beta_2 w = \alpha\beta_2 w$, причем на последнем шаге вывода применено правило $A \rightarrow \beta_2 \in P$, а $\gamma = \alpha = \gamma' X$ и $|\gamma| = n + 1$, т.е. $\gamma' \in (V_N \cup V_T)^*$, $|\gamma'| = n$, $X \in V_N \cup V_T$. Здесь $\mathrm{FIRST}_k(w) = u$. Надо показать, что $LR(k)$-ситуация $[A \rightarrow .\beta_2,\ u] \in V_k^G(\gamma)$.

Рассмотрим подробнее этот вывод, чтобы показать, как впервые появляется символ X, завершающий цепочку α, и как образуется правосентенциальная форма $\alpha A w$. В общем случае он имеет следующий вид:

$$\begin{array}{lll}
S & \underset{rm}{\overset{*}{\Rightarrow}} \alpha_1 A_1 w_1 & \\
 & \underset{rm}{\Rightarrow} \alpha_1\alpha_2' X A_2 \delta_2 w_1 & (\exists\, A_1 \rightarrow \alpha_2' X A_2 \delta_2 \in P,\ \alpha_1\alpha_2' = \gamma'), \\
 & \underset{rm}{\overset{*}{\Rightarrow}} \gamma' X A_2 w_2 w_1 & (\exists\, \delta_2 \underset{rm}{\overset{*}{\Rightarrow}} w_2), \\
 & \ldots & \\
 & \underset{rm}{\Rightarrow} \gamma' X A_{m-1}\delta_{m-1} w_{m-2} \ldots w_1 & (\exists\, A_{m-2} \rightarrow A_{m-1}\delta_{m-1} \in P), \\
 & \underset{rm}{\overset{*}{\Rightarrow}} \gamma' X A_{m-1} w_{m-1} w_{m-2} \ldots w_1 & (\exists\, \delta_{m-1} \underset{rm}{\overset{*}{\Rightarrow}} w_{m-1}), \\
 & \underset{rm}{\Rightarrow} \gamma' X A_m \delta_m w_{m-1} w_{m-2} \ldots w_1 & (\exists\, A_{m-1} \rightarrow A_m\delta_m \in P), \\
 & \underset{rm}{\overset{*}{\Rightarrow}} \gamma' X A_m w_m w_{m-1} w_{m-2} \ldots w_1 & (\exists\, \delta_m \underset{rm}{\overset{*}{\Rightarrow}} w_m), \\
 & = \gamma' X A w & (A_m = A,\ w_m w_{m-1} \ldots w_1 = w), \\
 & \underset{rm}{\Rightarrow} \gamma' X \beta_2 w = \gamma\beta_2 w. & (\exists\, A \rightarrow \beta_2 \in P,\ \gamma = \gamma' X).
\end{array}$$

Отметим, что в сентенциальной форме $\alpha_1\alpha_2' X A_2 \delta_2 w_1$ 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 $[A \rightarrow .\beta_2, u]$ не была бы допустима для префикса γ.

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

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 $LR(k)$-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 $\delta_2 \overset{*}{\underset{rm}{\Rightarrow}} w_2$, то согласно шагу 2б $[A_2 \to .A_3\delta_3, v_2] \in V_k^G(\gamma)$, где $v_2 = \mathrm{FIRST}_k(w_2v_1) = \mathrm{FIRST}_k(w_2w_1)$.

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

Наконец, согласно шагу 2б, поскольку $A_m = A$ и существуют правило $A \to \beta_2$ и вывод $\delta_m \overset{*}{\underset{rm}{\Rightarrow}} w_m$, заключаем, что $LR(k)$-ситуация $[A \to .\beta_2,\ v_m] \in V_k^G(\gamma)$, где $v_m = \mathrm{FIRST}_k(w_m v_{m-1}) = \mathrm{FIRST}_k(w_m w_{m-1} \ldots w_1) = \mathrm{FIRST}_k(w) = u$.

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

II. Докажем теперь, что если $[A \to \beta_1.\beta_2, u] \in V_k^G(\gamma)$, то существует правосторонний вывод вида $S \overset{*}{\underset{rm}{\Rightarrow}} \alpha A w \underset{rm}{\Rightarrow} \alpha\beta_1\beta_2 w$, в котором $\alpha\beta_1 = \gamma$ есть активный префикс правосентенциальной формы $\alpha\beta_1\beta_2 w$, а $\mathrm{FIRST}_k(w) = u$ 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 $S \overset{*}{\underset{rm}{\Rightarrow}} Aw \underset{rm}{\Rightarrow} \beta_2 w$, в котором на последнем шаге применяется правило $A \to \beta_2$, а $\mathrm{FIRST}_k(w) = u$.

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 алгоритма и правилу $S \to \alpha_1 \in P$;

$\alpha_1 = A_1\delta_1, \exists\, A_1 \to \alpha_2 \in P$, и шаг 1б дает $[A_1 \to .\alpha_2, v_1] \in V_k^G(\varepsilon)$,

где $v_1 = \mathrm{FIRST}_k(\delta_1)$, т.е. $\exists\ \delta_1 \overset{*}{\underset{rm}{\Rightarrow}} w_1$ (любой вывод терминальной цепочки можно перестроить в правосторонний) и $v_1 = \mathrm{FIRST}_k(w_1)$;

$\alpha_2 = A_2\delta_2, \exists\, A_2 \to \alpha_3 \in P$, и шаг 1б дает $[A_2 \to .\alpha_3, v_2] \in V_k^G(\varepsilon)$,

где $v_2 = \mathrm{FIRST}_k(\delta_2 v_1)$, т.е. $\exists\ \delta_2 \overset{*}{\underset{rm}{\Rightarrow}} w_2$, и $v_2 = \mathrm{FIRST}_k(w_2v_1) = \mathrm{FIRST}_k(w_2w_1)$;

…

$\alpha_m = A_m\delta_m, \exists\, A_m \to \alpha_{m+1} \in P$, и шаг 1($b$) дает $[A_m \to .\alpha_{m+1}, v_m] \in V_k^G(\varepsilon)$, где $v_m = \text{FIRST}_k\,(\delta_m v_{m-1})$; т.е. $\exists\ \delta_m \overset{*}{\underset{rm}{\Rightarrow}} w_m$ и $v_m = \text{FIRST}_k\,(w_m v_{m-1}) =$

$= \text{FIRST}_k\,(w_m w_{m-1} \ldots w_1) = \text{FIRST}_k\,(w) = u,\ A_m = A,\ \alpha_{m+1} = \beta_2,$

$w_m w_{m-1} \ldots w_1 = w.$

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

$$S \underset{rm}{\Rightarrow} \alpha_1 = A_1\delta_1 \overset{*}{\underset{rm}{\Rightarrow}} A_1 w_1 \underset{rm}{\Rightarrow} \alpha_2 w_1 = A_2\delta_2 w_1 \overset{*}{\underset{rm}{\Rightarrow}} A_2 w_2 w_1 \underset{rm}{\Rightarrow} \ldots \underset{rm}{\Rightarrow} A_m\delta_m w_{m-1} \ldots w_2 w_1 \overset{*}{\underset{rm}{\Rightarrow}}$$
$$\overset{*}{\underset{rm}{\Rightarrow}} A_m w_m w_{m-1} \ldots w_2 w_1 \underset{rm}{\Rightarrow} \alpha_{m+1} w_m w_{m-1} \ldots w_2 w_1 = \beta_2 w.$$

Итак, существует вывод $S \overset{*}{\underset{rm}{\Rightarrow}} Aw \underset{rm}{\Rightarrow} \beta_2 w$ и $\text{FIRST}_k\,(w) = \text{FIRST}_k\,(w_m \ldots w_1) = u$. 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 $|\gamma'| = n$, $X \in V_\text{N} \cup V_\text{T}$.

Согласно алгоритму 3.2 $LR(k)$-ситуация $[A \to \beta_1.\beta_2,\ u] \in V_k^G(\gamma)$ 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 построения множества $V_k^G(\gamma)$ 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

$$S \overset{*}{\underset{rm}{\Rightarrow}} \alpha_0 A_1 w_1 \underset{rm}{\Rightarrow} \underbrace{\alpha_0\rlap{$\overbrace{}^{\beta}$}\alpha_1}_{\gamma} X\underbrace{\alpha_2}_{\beta} w_1 = \underbrace{\gamma'}_{\gamma} X\beta_2 w_1 = \gamma\beta_2 w_1, \text{ причем } \text{FIRST}_k\,(w_1) = v_1.$$

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 данная $LR(k)$-ситуация $[A \to \beta_1.\beta_2,\ u]$ получается посредством нескольких шагов 2б, производящих последовательность $LR(k)$-ситуаций следующим образом:

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

где $v_2 \in \text{FIRST}_k\,(\delta_2 v_1)$, т.е. $\exists\ \delta_2 \overset{*}{\underset{rm}{\Rightarrow}} w_2$ и $v_2 = \text{FIRST}_k\,(w_2 v_1)$;

$\alpha_3 = A_3\delta_3$, $\exists\ A_3 \to \alpha_4 \in P$, и шаг 2б дает $[A_3 \to .\alpha_4,\ v_3] \in V_k^G(\gamma)$, где $v_3 \in \mathrm{FIRST}_k(\delta_3 v_2)$, т.е. $\exists\ \delta_3 \underset{\mathrm{rm}}{\overset{*}{\Rightarrow}} w_3$ и $v_3 = \mathrm{FIRST}_k(w_3 v_2) =$

$= \mathrm{FIRST}_k(w_3 w_2 v_1)$;

…

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

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

где $v_m \in \mathrm{FIRST}_k(\delta_m v_{m-1})$, т.е. $\exists\ \delta_m \underset{\mathrm{rm}}{\overset{*}{\Rightarrow}} w_m$ и $v_m = \mathrm{FIRST}_k(w_m v_m) =$

$= \mathrm{FIRST}_k(w_m \ldots w_2 v_1)$.

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 $A \to \beta_2$.

Извлечем теперь полезные следствия из вышеизложенной истории.

Из п.1 согласно индукционной гипотезе следует существование вывода вида

$S \underset{\mathrm{rm}}{\overset{*}{\Rightarrow}} \alpha_0 A_1 w_1 \underset{\mathrm{rm}}{\Rightarrow} \alpha_0\alpha_1 X\alpha_2 w_1 = \gamma' X\alpha_2 w_1 = \gamma A_2\delta_2 w_1$, в котором $\mathrm{FIRST}_k(w_1) = v_1$.

Благодаря п.3 этот вывод может быть продолжен следующим образом:

$$S \underset{\mathrm{rm}}{\overset{*}{\Rightarrow}} \gamma A_2\delta_2 w_1 \underset{\mathrm{rm}}{\overset{*}{\Rightarrow}} \gamma A_2 w_2 w_1 \underset{\mathrm{rm}}{\Rightarrow} \gamma\alpha_3 w_2 w_1 = \gamma A_3\delta_3 w_2 w_1 \underset{\mathrm{rm}}{\overset{*}{\Rightarrow}} \gamma A_3 w_3 w_2 w_1 \underset{\mathrm{rm}}{\overset{*}{\Rightarrow}} \gamma A_m w_m \ldots w_1 =$$

$$= \gamma A w \underset{\mathrm{rm}}{\Rightarrow} \gamma\beta_2 w,\ w = w_m \ldots w_1 \text{ и } \mathrm{FIRST}_k(w) = v_m = u.$$

Согласно определению, учитывая вышеприведенный вывод, заключаем, что $LR(k)$-ситуация $[A \to \beta_1.\beta_2,\ u] \in V_k^G(\gamma)$ допустима 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: $G = (V_\mathrm{N}, V_\mathrm{T}, P, S)$ — контекстно-свободная грамматика.

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

Метод.

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

2. Пусть $\mathscr{A} \in \mathscr{S}$ и $\mathscr{A}$ — не помечено. Пометить его и построить множество $\mathscr{A}' = \mathrm{GOTO}(\mathscr{A}, X)$ 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. Момент, когда в $\mathscr{S}$ все множества 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 $\mathscr{S}$ 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($\mathscr{A}$, S') никогда не потребуется вычислять, так как оно всегда пусто[18].

Пример 3.10. Рассмотрим еще 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:

$\mathscr{A}_0 = V_1^G(\varepsilon) = \{[S' \to .S, \varepsilon], [S \to .SaSb, \varepsilon \mid a], [S \to ., \varepsilon \mid a]\}$ (см. пример 3.9).

Затем строим GOTO($\mathscr{A}_0, X$), где $X \in \{S, a, b\}$:

$\mathscr{A}_1 = \mathrm{GOTO}(\mathscr{A}_0, S) = V_1^G(S) = \{[S' \to S., \varepsilon], [S \to S.aSb, \varepsilon \mid a]\}$ (см. пример 3.9);

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

Теперь вычисляем GOTO($\mathscr{A}_1, X$), где $X \in \{S, a, b\}$:

$\mathrm{GOTO}(\mathscr{A}_1, S) = \varnothing$;

$\mathscr{A}_2 = \mathrm{GOTO}(\mathscr{A}_1, a) = \mathrm{GOTO}(V_1^G(S), a) = V_1^G(Sa) =$

$= \{[S \to Sa.Sb, \varepsilon \mid a], [S \to .SaSb, a \mid b], [S \to ., a \mid b]\}$ (см. пример 3.9);

$\mathrm{GOTO}(\mathscr{A}_1, b) = \mathrm{GOTO}(V_1^G(S), b) = \varnothing$.

Теперь вычисляем GOTO($\mathscr{A}_2, X$), где $X \in \{S, a, b\}$:

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

$\mathrm{GOTO}(\mathscr{A}_2, a) = \mathrm{GOTO}(V_1^G(Sa), a) = \varnothing$;

$\mathrm{GOTO}(\mathscr{A}_2, b) = \mathrm{GOTO}(V_1^G(Sa), b) = \varnothing$.

Теперь вычисляем GOTO($\mathscr{A}_3, X$), где $X \in \{S, a, b\}$:

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

$\mathscr{A}_4 = \mathrm{GOTO}(\mathscr{A}_3, a) = \{[S \to Sa.Sb, a \mid b], [S \to .SaSb, a \mid b], [S \to ., a \mid b]\}$;

$\mathscr{A}_5 = \mathrm{GOTO}(\mathscr{A}_3, b) = \{[S \to SaSb., \varepsilon \mid a\}$.

Теперь вычисляем GOTO($\mathscr{A}_4, X$), где $X \in \{S, a, b\}$:

$\mathscr{A}_6 = \mathrm{GOTO}(\mathscr{A}_4, S) = \{[S \to SaS.b, a \mid b], [S \to S.aSb, a \mid b]\}$.

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

Теперь вычисляем GOTO($\mathscr{A}_5, X$), где $X \in \{S, a, b\}$:

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

Теперь вычисляем GOTO ($\mathscr{A}_6, X$), где $X \in \{S, a, b\}$:

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

---

[18] Так как S' не встречается в правых частях правил.

$\text{GOTO}(\mathscr{A}_6, a) = \{\,[S \to Sa.Sb,\ a \mid b],\ S \to .SaSb,\ a|b],\ [S \to .,\ a|b]\} = \mathscr{A}_4;$

$\mathscr{A}_7 = \text{GOTO}(\mathscr{A}_6, b) = \{\,[S \to SaSb.,\ a \mid b]\}.$

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

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

Таб. 3.4

$\mathscr{A}$	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 ней соответствуют неопределенным значениям. Заметим, что множество $\text{GOTO}(\mathscr{A}, X)$ всегда пусто, если в каждой $LR(k)$-ситуации из $\mathscr{A}$ 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 определенной последовательности, т.е. $\mathscr{S} = \bigcup_{i=0}^{m} \mathscr{S}_i$, причем 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 $\mathscr{A} = \text{GOTO}(\mathscr{A}', X) = V_k^G(\gamma' X) \in \mathscr{S}_{i+1}$, где $X \in V_N \cup V_T$.

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 $X \in V_\mathrm{N} \cup V_\mathrm{T}$, такие, что $\mathcal{A} = \mathrm{GOTO}(\mathcal{A}', X) = V_k^G(\gamma' X) = V_k^G(\gamma)$, где $\gamma = \gamma' X$.

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

II. Докажем теперь, что если $\gamma \in (V_\mathrm{N} \cup V_\mathrm{T})^*$ — активный префикс грамматики G и $\mathcal{A} = V_k^G(\gamma)$, то $\mathcal{A} \in \mathcal{S}$.

Индукция по $l = |\gamma|$.

База. Пусть $l = 0$, $\gamma = \varepsilon$. Имеем $\mathcal{A} = V_k^G(\varepsilon) = \mathcal{A}_0$. Согласно алгоритму 3.3 $\mathcal{A}_0$ включается в множество $\mathcal{S}$ на первом же шаге.

Индукционная гипотеза. Предположим, что утверждение выполняется для всех γ: $l = |\gamma| \le n$ ($n \ge 0$).

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

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

Следовательно, γ' — тоже активный префикс грамматики G и $|\gamma'| = n$. Согласно индукционной гипотезе $V_k^G(\gamma') = \mathcal{A}' \in \mathcal{S}$. Кроме того, в соответствии с алгоритму 3.3 $\mathcal{A} = \mathrm{GOTO}(\mathcal{A}', X) = \mathrm{GOTO}(V_k^G(\gamma'), X) = V_k^G(\gamma' X) = V_k^G(\gamma)$ включается в множество $\mathcal{S}$.

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 $k \ge 0$.

***Определение* 3.11.** Множество $\mathcal{A} = V_k^G(\gamma) \in \mathcal{S}$ назовем *непротиворечивым*, если оно не содержит двух разных $LR(k)$-ситуаций вида $[A \to \beta., u]$ и $[B \to \beta_1.\beta_2, v]$ (цепочка β_2 может быть пустой), таких, что $u \in \mathrm{EFF}_k^G(\beta_2 v)$.

Алгоритм 3.4: $LR(k)$-тестирование.

Вход: $G = (V_N, V_T, P, S)$ — контекстно-свободная грамматика, $k \geq 0$.

Выход: “Да”, если G — $LR(k)$-грамматика. “Нет” в противном случае.

Метод.

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

2. Каждое $\mathscr{A} \in \mathscr{S}$ проверим на непротиворечивость. Если окажется, что рассматриваемое множество $\mathscr{A}$ 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).

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 $\mathscr{A}_0$, $\mathscr{A}_1$, $\mathscr{A}_2$, $\mathscr{A}_4$, $\mathscr{A}_5$, $\mathscr{A}_7$, построенных в предыдущем примере (новая нумерация).

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

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

Вывод: множество $\mathscr{A}_0$ непротиворечиво.

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

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

Вывод: множество $\mathscr{A}_1$ непротиворечиво.

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

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

Вывод: множество $\mathscr{A}_2$ непротиворечиво.

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

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

Вывод: множество $\mathscr{A}_4$ непротиворечиво.

Проверка $\mathscr{A}_5 = \{[S \to SaSb., \varepsilon \mid a]\}$ — множество непротиворечиво.

Проверка $\mathscr{A}_7 = \{[S \to SaSb., a \mid b]\}$ — множество непротиворечиво.

Заметим, что $\mathrm{EFF}_1^G(Sb) = \mathrm{EFF}_1^G(SaSb) = \varnothing$ потому, что из цепочек, начинающихся на нетерминал S, выводимы терминальные цепочки только благодаря ε-правилу, используемому на последнем шаге. Таким образом, получаемые цепочки не участвуют в образовании EFF_1^G.

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

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 $G = (V_N, V_T, P, S)$ — контекстно-свободная грамматика и $\mathscr{S}$ — каноническое множество $LR(k)$-ситуаций для грамматики G. Для каждого множества $\mathscr{A} \in \mathscr{S}$ определим $LR(k)$-*таблицу* $T(\mathscr{A}) = (f, g)$, *состоящую из пары функций*:

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

g: $V_N \cup V_T \to \{T(\mathscr{A}) \mid \mathscr{A} \in \mathscr{S}\} \cup \{\text{error}\}$.

Функция f определяется следующим образом:

а) $f(u)$ = shift, если существует $[A \to \beta_1.\beta_2, v] \in \mathscr{A}$, $\beta_2 \neq \varepsilon$ и $u \in \mathrm{EFF}_k^G(\beta_2 v)$;

б) $f(u)$ = reduce i, $[A \to \beta., u] \in \mathscr{A}$ и $A \to \beta$ есть i-е правило из множества правил P, $i \geq 1$[19];

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:

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

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

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*.

---

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

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

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

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

Метод.

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

2. Взять $\mathcal{T} = \{T(\mathcal{A}) \mid \mathcal{A} \in \mathcal{S}\}$ в качестве множества $LR(k)$-таблиц.

3. Положить $T_0 = T(\mathcal{A}_0)$, где $\mathcal{A}_0 = V_k^G(\varepsilon)$.

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, приведенные в этом примере.

Поскольку $\mathcal{A}_0 = \{[S' \to .S, \varepsilon], [S \to .SaSb, \varepsilon \mid a], [S \to ., \varepsilon \mid a]\}$, то $T_0 = (f_0, g_0)$ имеет следующий табличный вид:

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

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

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

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

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

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

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

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

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

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

$f_5(u)$			$g_5(X)$		
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:

$f_6(u)$			$g_6(X)$		
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:

$f_7(u)$			$g_7(X)$		
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.

Итак, $LR(k)$-анализаторы обладают следующими свойствами:

1. Диагностичностью. Простой индукцией по числу шагов анализатора можно показать, что каждая $LR(k)$-таблица, находящаяся в его магазине, соответствует цепочке символов, расположенной слева от таблицы. Именно: если в цепочке, о которой идет речь, отбросить символы $LR(k)$-таблиц, то получим активный префикс грамматики, которому эта $LR(k)$-таблица соответствует. Поэтому $LR(k)$-анализатор сообщает об ошибке при первой же возможности в ходе чтения входной цепочки слева направо.

2. Пусть $T_j = (f_j, g_j)$. Если $f_j(u) = \text{shift}$ и анализатор находится в конфигурации $(T_0X_1T_1X_2...X_jT_j, x, \pi)$, то найдется $LR(k)$-ситуация $[B \to \beta_1.\beta_2, v]$, $\beta_2 \neq \varepsilon$, допустимая для активного префикса $X_1X_2...X_j$, такая, что $u \in \mathrm{EFF}_k^G(\beta_2 v)$, где $u = \mathrm{FIRST}_k(x)$. Поэтому, если $S' \overset{*}{\underset{rm}{\Rightarrow}} X_1X_2\ldots X_j uy$ при некоторой цепочке $y \in V_T^*$, то по теореме 3.1 правый конец основы цепочки $X_1X_2...X_j uy$ должен быть где-то справа от символа X_j.

3. Если в конфигурации, указанной в п.2, $f_j(u) = \text{reduce } i$ и $A \to Y_1Y_2\ldots Y_p$ — правило с номером i, то цепочка $X_{j-p+1}\ldots X_{j-1}X_j$ в магазине должна совпадать с $Y_1Y_2\ldots Y_p$, так как множество ситуаций, по которому построена таблица T_j, содержит ситуацию $[A \to Y_1Y_2\ldots Y_p., u]$. Таким образом, на шаге свертки не обязательно рассматривать символы верхней части магазина, надо просто выбросить из магазина $2p$ символов грамматики и $LR(k)$-таблиц.

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 $[S' \to S., \varepsilon]$.

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

Теорема 3.5. *Пусть* $G = (V_{\mathrm{N}}, V_{\mathrm{T}}, P, S)$ — $LR(k)$-*грамматика и пусть* $\mathfrak{A}$ — $LR(k)$-*анализатор, построенный по грамматике* G *согласно алгоритму* 3.5. *Утверждается, что вывод* $S \underset{\mathrm{rm}}{\overset{\pi}{\Longrightarrow}} w$ *существует тогда и только тогда, когда* $(T_0, w, \varepsilon) \vdash_{\mathfrak{A}}^{*} (T_0 S\, T, \varepsilon, \pi^R)$, где $T = (f, g) \in \mathcal{T}$, *причем* $f(\varepsilon) = \mathrm{accept}$.

Доказательство.

I. Индукцией по $l = |\pi|$ покажем теперь, что если $S \underset{\mathrm{rm}}{\overset{\pi}{\Longrightarrow}} \alpha x$, где $\alpha \in (V_{\mathrm{N}} \cup V_{\mathrm{T}})^*$, $x \in V_{\mathrm{T}}^*$, то $(T_0 X_1 T_1 X_2 \ldots X_m T_m,\ x,\ \varepsilon) \vdash_{\mathfrak{A}}^{*} (T_0 S T,\ \varepsilon,\ \pi^R)$, где $X_1 X_2 \ldots X_m = \alpha$, $T_0 = T(\mathcal{A}_0)$, $\mathcal{A}_0 = V_k^G(\varepsilon)$, а $T_i = T(\mathcal{A}_i)$, $\mathcal{A}_i = V_k^G(X_1 X_2 \ldots X_i)$, $i = 1, 2, \ldots, m$.

База. Пусть $l = 1$, т.е. $\pi = i$. Имеем $S \underset{\mathrm{rm}}{\overset{(i)}{\Longrightarrow}} \alpha x$. Следовательно, существует правило $(i)\ S \to \alpha x \in P$.

Пусть $x = a_1 a_2 \ldots a_p$, $a_j \in V_{\mathrm{T}}$, $j = 1, 2, \ldots, p$. Согласно определению активных префиксов и допустимых для них $LR(k)$-ситуаций имеем

$[S \to \alpha.a_1 a_2 \ldots a_p, \varepsilon] \in \mathcal{A}_m = V_k^G(\alpha) = V_k^G(X_1 X_2 \ldots X_m)$,

$[S \to \alpha a_1.a_2 \ldots a_p, \varepsilon] \in \mathcal{A}_{m+1} = V_k^G(\alpha a_1)$,

…

$[S \to \alpha a_1 a_2 \ldots a_{p-1}.a_p, \varepsilon] \in \mathcal{A}_{m+p-1} = V_k^G(\alpha a_1 a_2 \ldots a_{p-1})$,

$[S \to \alpha a_1 a_2 \ldots a_p., \varepsilon] = [S \to \alpha x., \varepsilon] \in \mathcal{A}_{m+p} = V_k^G(\alpha a_1 a_2 \ldots a_p) = V_k^G(\alpha x)$.

Тогда согласно алгоритму построения $LR(k)$-таблиц $T_{m+j} = T(\mathcal{A}_{m+j}) = (f_{m+j}, g_{m+j})$, $f_{m+j-1}(u_j) = \mathrm{shift}$ для $u_j \in \mathrm{EFF}_k^G(a_j a_{j+1} \ldots a_p \varepsilon) = \mathrm{FIRST}_k(a_j a_{j+1} \ldots a_p)$; $g_{m+j-1}(a_j) = T_{m+j}$ $(j = 1, 2, \ldots, p)$; а $T_{m+p} = (f_{m+p}, g_{m+p})$ и $f_{m+p}(\varepsilon) = \mathrm{reduce}\ i$. Поэтому

$$(T_0 X_1 T_1 X_2 \ldots X_m T_m, x, \varepsilon) = (T_0 X_1 T_1 X_2 \ldots X_m T_m, a_1 a_2 \ldots a_p, \varepsilon) \vdash_{\mathfrak{A}}^{*}$$
$$\vdash_{\mathfrak{A}}^{*} (T_0 X_1 T_1 X_2 \ldots X_m T_m a_1 T_{m+1} a_2 T_{m+2} \ldots a_p T_{m+p}, \varepsilon, \varepsilon) \vdash_{\mathfrak{A}} (T_0 S T, \varepsilon, i).$$

Здесь сначала выполняется p раз сдвиг, затем один раз свертка по i-му правилу; $X_1 X_2 \ldots X_m a_1 a_2 \ldots a_p = \alpha x$, $T = T(V_k^G(S)) = (f, g)$, причем очевидно, что $[S' \to S., \varepsilon] \in V_k^G(S)$ и потому согласно алгоритму 3.5 $f(\varepsilon) = \mathrm{accept}$. База доказана.

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 $S \underset{\mathrm{rm}}{\overset{\pi'}{\Longrightarrow}} \alpha' A w \underset{\mathrm{rm}}{\overset{(i)}{\Longrightarrow}} \alpha' \beta w = \alpha x$. Очевидно, что x заканчивается цепочкой w, т.е. $x = zw$. Здесь $\alpha', \alpha \in (V_{\mathrm{N}} \cup V_{\mathrm{T}})^*$, $w, x, z \in V_{\mathrm{T}}^*$, $\pi = \pi' i$, $A \to \beta \in P$ — (i)-е правило. Кроме того, $\alpha' \beta w = \alpha x = \alpha z w$, и потому $\alpha' \beta = \alpha z$.

Рассмотрим исходную конфигурацию

$$(\underbrace{\overbrace{T_0X_1T_1X_2\ldots X_jT_j}^{\alpha}\overbrace{X_{j+1}T_{j+1}\ldots X_mT_m}^{\beta'}}_{\alpha},\ \overbrace{zw}^{x},\ \varepsilon).$$

В ней $\alpha' = X_1X_2...X_j$, $\beta' = X_{j+1}...X_m$, $\beta = \beta' z$, $\alpha = \alpha'\beta'$, $x = zw$. Поэтому имеющийся вывод представим в виде $S \overset{\pi'}{\underset{rm}{\Longrightarrow}} \alpha' Aw \overset{(i)}{\underset{rm}{\Longrightarrow}} \alpha'\beta w = \alpha'\beta' zw$.

Поскольку $A \to \beta = \beta' z \in P$, то $\alpha'\beta'$ — активный префикс правосентенциальной формы $\alpha'\beta' zw$, причем $V_k^G(\alpha'\beta') = V_k^G(\alpha) = \mathscr{A}_m$, и потому $[A \to \beta'.z,\ u] \in \mathscr{A}_m$, где $u = \mathrm{FIRST}_k(w)$. Пусть $z = a_1a_2...a_p$. Согласно алгоритму построения $LR(k)$-таблиц имеем

$T_m = T(\mathscr{A}_m) = (f_m, g_m)$,

$f_m(v_1) = \text{shift}$ для $v_1 \in \mathrm{EFF}_k^G(zu) = \mathrm{FIRST}_k(a_1a_2...a_pw) = \mathrm{FIRST}_k(x)$,

$g_m(a_1) = T_{m+1}$, где $T_{m+1} = T(\mathscr{A}_{m+1})$, $\mathscr{A}_{m+1} = V_k^G(\alpha'\beta' a_1)$;

$T_{m+1} = (f_{m+1}, g_{m+1})$,

$f_{m+1}(v_2) = \text{shift}$ для $v_2 \in \mathrm{FIRST}_k(a_2...a_pu) = \mathrm{FIRST}_k(a_2...a_pw)$,

$g_{m+1}(a_2) = T_{m+2}$, где $T_{m+2} = T(\mathscr{A}_{m+2})$, $\mathscr{A}_{m+2} = V_k^G(\alpha'\beta' a_1a_2)$;

...

$T_{m+p-1} = T(\mathscr{A}_{m+p-1}) = (f_{m+p-1}, g_{m+p-1})$, $\mathscr{A}_{m+p-1} = V_k^G(\alpha'\beta' a_1a_2...a_{p-1})$,

$f_{m+p-1}(v_p) = \text{shift}$ для $v_p \in \mathrm{EFF}_k^G(a_pu) = \mathrm{FIRST}_k(a_pu) = \mathrm{FIRST}_k(a_pw)$,

$g_{m+p-1}(a_p) = T_{m+p} = T(\mathscr{A}_{m+p})$, где $\mathscr{A}_{m+p} = V_k^G(\alpha'\beta' a_1a_2...a_{p-1}a_p) = V_k^G(\alpha'\beta' z) =$
$= V_k^G(\alpha'\beta)$.

$T_{m+p} = (f_{m+p}, g_{m+p})$, $f_{m+p}(u) = \text{reduce } i$, $(i)\ A \to \beta \in P$, $u = \mathrm{FIRST}_k(w)$.

Используя существующие $LR(k)$-таблицы, канонический $LR(k)$-анализатор совершает следующие движения, начиная с исходной конфигурации:

$(T_0X_1T_1X_2...X_mT_m,\ x,\ \varepsilon) = (T_0X_1T_1X_2...X_mT_m,\ zw,\ \varepsilon) =$
$= (T_0X_1T_1X_2...X_mT_m,\ a_1a_2...a_pw,\ \varepsilon) \vdash_{\mathfrak{A}}$
$\vdash_{\mathfrak{A}} (T_0X_1T_1X_2...X_mT_ma_1T_{m+1},\ a_2...a_pw,\ \varepsilon) \vdash_{\mathfrak{A}} ...$
$... \vdash_{\mathfrak{A}} (T_0X_1T_1X_2...X_mT_ma_1T_{m+1}a_2...a_pT_{m+p},\ w,\ \varepsilon) =$
$= (T_0X_1T_1X_2...X_jT_jX_{j+1}T_{j+1}...X_mT_ma_1T_{m+1}a_2...a_pT_{m+p},\ w,\ \varepsilon) \vdash_{\mathfrak{A}}$
$\vdash_{\mathfrak{A}} (T_0X_1T_1X_2...X_jT_jA\,T'_{j+1},\ w,\ i) \vdash_{\mathfrak{A}}^{*} (T_0ST,\ \varepsilon,\ i\pi'^R) = (T_0ST,\ \varepsilon,\ \pi^R)$.

Последний переход обоснован индукционной гипотезой с учетом того, что $T'_{j+1} = T(\mathscr{A}'_{j+1}), \mathscr{A}'_{j+1} = V_k^G(\alpha' A)$, $\alpha' = X_1X_2...X_j$. Вспомогательное утверждение I доказано.

В частности, если $\alpha = \varepsilon$, т.е. $S \xRightarrow[\text{rm}]{\pi} x$, то $(T_0, x, \varepsilon) \vdash_{\mathfrak{A}}^{*} (T_0ST, \varepsilon, \pi^R)$. Достигнутая конфигурация является принимающей по тем же причинам, что и в базовом случае.

II. Индукцией по l — числу движений $LR(k)$-анализатора — покажем, что если $(T_0, w, \varepsilon) \vdash_{\mathfrak{A}}^{l} (T_0X_1T_1X_2...X_mT_m, x, \pi^R)$, то $\alpha x \xRightarrow[\text{rm}]{\pi} w$, где $\alpha = X_1X_2...X_m$, $X_j \in V_N \cup V_T$, $j = 1, 2, \ldots, m$; $x, w \in V_T^*$. Другими словами, покажем, что магазинная цепочка (без учета $LR(k)$-таблиц) и конкатенированная с ней входная цепочка представляют правосентенциальную форму, из которой выводится анализируемое предложение посредством последовательности правил π.

База. Пусть $l = 1$: имеется одно движение.

<u>Случай 1</u>: shift-движение.

Пусть $(T_0, w, \varepsilon) = (T_0, X_1x, \varepsilon) \vdash_{\mathfrak{A}} (T_0X_1T, x, \varepsilon)$. Здесь $w = X_1x$, где $X_1 \in V_T$, $x \in V_T^*$. Первый символ цепочки w переносится в магазин, на вершине оказывается некоторая $LR(k)$-таблица T, в выходную цепочку ничего не пишется. Следует показать, что существует вывод $X_1x \xRightarrow[\text{rm}]{*} w$ без использования каких-либо правил грамматики. Но это возможно лишь в случае, когда $w = X_1x$, что как раз и имеет место.

<u>Случай 2</u>: reduce i -движение.

Это движение предполагает свертку какой-то цепочки в верхней части магазина в нетерминал по правилу номер i. Поскольку в начальной конфигурации (T_0, w, ε) магазин пуст (T_0 не считается), то основа магазинной цепочки тоже пуста. Именно она и сворачивается к некоторому нетерминалу. Формально это движение обусловлено тем, что $T_0 = (f_0, g_0)$, $f_0(u) = \text{reduce } i$ для $u = \text{FIRST}_k(w)$.

Наряду с этим согласно алгоритму построения $LR(k)$-анализатора $T_0 = T(\mathscr{A}_0)$, $\mathscr{A}_0 = V_k^G(\varepsilon)$, $[A \to \varepsilon., u] \in \mathscr{A}_0$, что и обусловливает движение:

$$(T_0, w, \varepsilon) \vdash_{\mathfrak{A}} (T_0AT, w, \varepsilon), \text{ где } T = g_0(A).$$

Остается проверить, что $Aw \xRightarrow[\text{rm}]{(i)} w$. Но это очевидно, поскольку данное движение обусловлено существованием правила $A \to \varepsilon \in P$, номер которого есть i. 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 $(T_0, w, \varepsilon) \vdash_{\mathfrak{A}}^{n} (T_0X_1T_1X_2...X_mT_m, x', \pi'^R)$. По индукционной гипотезе немедленно получаем $X_1X_2...X_mx' \xRightarrow[\text{rm}]{\pi'} w$. Затем совершается последнее, $(n + 1)$-е, движение.

Случай 1: shift-движение.

Это движение происходит следующим образом:

$(T_0X_1T_1X_2\dots X_mT_m,\ x',\ \pi'^R) =$

$= (T_0X_1T_1X_2\dots X_mT_m,\ X_{m+1}x,\ \pi'^R) \vdash_{\mathfrak{A}} (T_0X_1T_1X_2\dots X_mT_mX_{m+1}\ T_{m+1},\ x,\ \pi'^R),$

т.е. X_{m+1} переносится в магазин, в выходную цепочку ничего не пишется. Остается лишь убедиться в том, что $X_1X_2\dots X_mX_{m+1}x \overset{\pi'}{\underset{rm}{\Rightarrow}} w$. Так как $x' = X_{m+1}x$, то $X_1X_2\dots X_mX_{m+1}x = X_1X_2\dots X_mx' \overset{\pi'}{\underset{rm}{\Rightarrow}} w$ по индукционной гипотезе.

Случай 2: reduce i -движение.

Имеем конфигурацию, достигнутую за первые n движений: $(T_0X_1T_1X_2\dots X_mT_m,$ $x', \pi'^R)$. Далее совершается последнее движение: свертка верхней части магазина по i-у правилу. Оно происходит благодаря тому, что $T_m = (f_m,\ g_m)$, $f_m(u') =$ reduce i для $u' = \mathrm{FIRST}_k\,(x')$.

Согласно алгоритму построения анализатора $T_m = T(\mathcal{A}_m)$, $\mathcal{A}_m = V_k^G\,(X_1X_2\dots X_m)$, и существуют $LR(k)$-ситуация $[A \to Y_1Y_2\dots Y_p.,\ u'] \in \mathcal{A}_m$ и правило $A \to Y_1Y_2\dots Y_p$ под номером i, такое, что $Y_1Y_2\dots Y_p = X_{m-p+1}\dots X_{m-1}X_m$. Имеем:

$$(T_0X_1T_1X_2\dots X_mT_m, x', \pi'^R) =$$

$$= (T_0X_1T_1X_2\dots X_{m-p}T_{m-p}Y_1T_{m-p+1}Y_2T_{m-p+2}\dots Y_pT_mx', \pi'^R) \vdash_{\mathfrak{A}}$$

$$\vdash_{\mathfrak{A}} (T_0X_1T_1X_2\dots X_{m-p}T_{m-p}A\,T'_{m-p+1},\ x',\ \pi'^R i),$$

где $T'_{m-p+1} = g_{m-p}(A)$, если $T_{m-p} = (f_{m-p}, g_{m-p})$.

Остается убедиться лишь в том, что $X_1X_2\dots X_{m-p}Ax' \overset{i\pi'}{\underset{rm}{\Rightarrow}} w$. Действительно,

$$X_1X_2\dots X_{m-p}Ax' \overset{i}{\underset{rm}{\Rightarrow}} X_1X_2\dots X_{m-p}Y_1Y_2\dots Y_px' =$$

$$= \mathrm{X}_1\mathrm{X}_2\dots\mathrm{X}_{m-p}\mathrm{X}_{m-p+1}\dots\mathrm{X}_{m-1}\mathrm{X}_m\mathrm{x}' \overset{\pi'}{\underset{rm}{\Rightarrow}} \mathrm{w},$$

причем первый шаг вывода выполняется при помощи упомянутого правила, а завершение вывода обеспечено индукционной гипотезой. Вспомогательное утверждение II доказано.

В частности, при $\alpha = S$ и $x = \varepsilon$ заключаем, что если $(T_0,\ w,\ \varepsilon) \vdash_{\mathfrak{A}}^{*} (T_0ST,\ \varepsilon,\ \pi^R)$, 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 ведется от противного. Пусть G — $LR(k)$-грамматика, но она не является синтаксически однозначной. Это значит, что существуют два разных правосторонних вывода одной и той же терминальной цепочки:

1) $S = \alpha_0 \underset{rm}{\Rightarrow} \alpha_1 \underset{rm}{\Rightarrow} \dots \underset{rm}{\Rightarrow} \alpha_m = w$,

2) $S = \beta_0 \underset{rm}{\Rightarrow} \beta_1 \underset{rm}{\Rightarrow} \dots \underset{rm}{\Rightarrow} \beta_n = w,\ w \in V_T^*$,

и существует $i > 0$ $(i < m,\ i < n)$, такое, что $\alpha_{m-i} \neq \beta_{n-i}$, но $\alpha_{m-j} = \beta_{n-j}$ при $j < i$.

В более детальном представлении эти выводы имеют вид

$$1')\ S \overset{*}{\underset{rm}{\Rightarrow}} \underbrace{\alpha A x}_{\alpha_{m-i}} \underset{rm}{\Rightarrow} \alpha\beta x \overset{*}{\underset{rm}{\Rightarrow}} w,$$

$$2')\ S \overset{*}{\underset{rm}{\Rightarrow}} \underbrace{\gamma B y}_{\beta_{n-i}} \underset{rm}{\Rightarrow} \gamma\delta y = \underbrace{\gamma\beta_1}_{\alpha\beta}\overbrace{\beta_1\underbrace{\beta_2 y}_{x}}^{\delta} = \alpha\beta x \overset{*}{\underset{rm}{\Rightarrow}} w, \text{ где } \delta = \beta_1\beta_2,\ \gamma\beta_1 = \alpha\beta,\ \beta_2 y = x,\ \beta_2 \in V_T^*,$$

причем 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 $u = \mathrm{FIRST}_k(x)$.

Аналогично из 2′) следует, что $[B \to \beta_1.\beta_2, v] \in V_k^G(\alpha\beta)$, $v = \mathrm{FIRST}_k(y)$. При этом

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

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, что противоречит исходному предположению о том, что G — $LR(k)$-грамматика. Следова-тельно 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 свертка (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 $LR(k)$-грам-матикой, что противоречило бы первоначальному предположению.

Итак, общее число движений канонического $LR(k)$-анализатора $N \leq n + c \times \times n = n \times (c + 1) = O(n)$, где $c \leq \#\mathcal{T}$; 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 ее 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. Но к сожалению, существуют такие простые семантически однозначные схемы, которые задают трансляции, не реализуемые детерминированными магазинными преобразователями.

Пример 3.13. Пусть имеется схема синтаксически управляемой трансляции

$$T = \big(\{S\},\ \{a, b\},\ \{a, b\},\ \{1)\ S \to Sa, aSa,\ \ 2)\ S \to Sb, bSb,\ \ 3)\ S \to \varepsilon, \varepsilon\}\big).$$

Очевидно, что ее входная грамматика — $LR(1)$. Действительно, каноническая система множеств допустимых $LR(1)$-ситуаций для этой грамматики $\mathscr{S} = \{\mathscr{A}_0, \mathscr{A}_1, \mathscr{A}_2, \mathscr{A}_3\}$ — не противоречива, ибо

$\mathscr{A}_0 = \{[S' \to .S, \varepsilon], [S \to .Sa, \varepsilon|a|b], [S \to .Sb, \varepsilon|a|b], [S \to ., \varepsilon|a|b]\}$,

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

$\mathscr{A}_2 = \mathrm{GOTO}(\mathscr{A}_1, a) = \{[S \to Sa., \varepsilon|a|b]\}$,

$\mathscr{A}_3 = \mathrm{GOTO}(\mathscr{A}_1, b) = \{[S \to Sb., \varepsilon|a|b]\}$.

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

Построенный канонический $LR(1)$-анализатор для входной цепочки bba выдает правосторонний анализ $\pi^R(bba) = 3221$, т.е.

$$(T_0, bba, \varepsilon) \vdash_{\mathfrak{A}} (T_0ST_1, bba, 3) \vdash_{\mathfrak{A}} (T_0ST_1bT_3, ba, 3) \vdash_{\mathfrak{A}} (T_0ST_1, ba, 32) \vdash_{\mathfrak{A}}$$
$$\vdash_{\mathfrak{A}} (T_0ST_1bT_3, a, 322) \vdash_{\mathfrak{A}} (T_0ST_1, a, 322) \vdash_{\mathfrak{A}} (T_0ST_1aT_2, \varepsilon, 3221) \vdash_{\mathfrak{A}} (T_0ST_1, \varepsilon, 3221).$$

Данная схема задает трансляцию

$$\tau(T) = \{(w, w^Rw) \mid w \in \{a, b\}^*\}.$$

В частности, имеем вывод

$$(S, S) \overset{(1)}{\underset{rm}{\Rightarrow}} (Sa, aSa) \overset{(2)}{\underset{rm}{\Rightarrow}} (Sba, abSba) \overset{(2)}{\underset{rm}{\Rightarrow}} (Sbba, abbSbba) \overset{(3)}{\underset{rm}{\Rightarrow}} (bba, abbbba).$$

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, вы-

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 ее символы из магазина. Далее требуется на выходе сгенерировать цепочку w, но в магазине, пустом к этому времени, нет для этого никакой информации. Где еще, 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. 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 устанавлено, 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 ее 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 выдает 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 упражнения читателю.

Пример 3.14. Пусть имеется 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 схеме:

$$(S, S) \underset{rm}{\overset{(1)}{\Rightarrow}} (SaSb, SSc) \underset{rm}{\overset{(1)}{\Rightarrow}} (SaSaSbb, SSScc) \underset{rm}{\overset{(2)}{\Rightarrow}} (SaSabb, SScc) \underset{rm}{\overset{(2)}{\Rightarrow}}$$
$$\underset{rm}{\overset{(2)}{\Rightarrow}} (Saabb, Scc) \underset{rm}{\overset{(2)}{\Rightarrow}} (aabb, cc).$$

Руководствуясь табл. 3.3, $LR(1)$-транслятор совершает следующие движения:

$$(T_0, aabb, \varepsilon) \vdash_{\mathfrak{A}} (T_0ST_1, aabb, \varepsilon) \vdash_{\mathfrak{A}} (T_0ST_1aT_2, abb, \varepsilon) \vdash_{\mathfrak{A}} (T_0ST_1aT_2ST_3, abb, \varepsilon) \vdash_{\mathfrak{A}}$$
$$\vdash_{\mathfrak{A}} (T_0ST_1aT_2ST_3aT_4, bb, \varepsilon) \vdash_{\mathfrak{A}} (T_0ST_1aT_2ST_3aT_4ST_6, bb, \varepsilon) \vdash_{\mathfrak{A}}$$
$$\vdash_{\mathfrak{A}} (T_0ST_1aT_2ST_3aT_4ST_6bT_7, b, \varepsilon) \vdash_{\mathfrak{A}} (T_0ST_1aT_2ST_3, b, c) \vdash_{\mathfrak{A}}$$
$$\vdash_{\mathfrak{A}} (T_0ST_1aT_2ST_3bT_5, \varepsilon, c) \vdash_{\mathfrak{A}} (T_0ST_1, \varepsilon, cc).$$

§ 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.

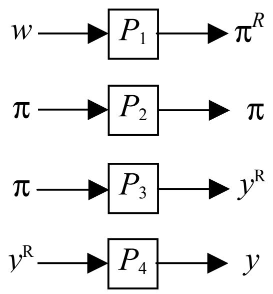

Рис. 3.3.

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 четырехуровневую схему перевода (рис. 3.3).

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 ее из магазина, выдавая на выход. Получается цепочка π — последовательность номеров правил правостороннего вывода входной цепочки w. На следующем этапе она используется для порождения соответствующей инвертированной выходной цепочки правосторонним выводом в выходной грамматике схемы T.

Входом для третьего уровня служит цепочка π. Выход — перевод, определяемый простой sdts $T' = (N, \Sigma', \Delta, R', S)$, где R' содержит правило вида

$A \to iB_mB_{m-1}\ldots B_1, y_mB_my_{m-1}B_{m-1}\ldots y_1B_1y_0$

тогда и только тогда, когда $A \to x_0B_1x_1...B_mx_m$, $y_0B_1\, y_1...B_my_m$ — правило из R, а правило $A \to x_0B_1\,x_1...B_mx_m$ 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, ее 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*(*k*)*-грамматиками.*

***Определение* 3.17.** *Ядром* $LR(k)$-ситуации $[A \to \beta_1.\beta_2,\ u]$ назовем $A \to \beta_1.\beta_2$. 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

$$\mathscr{S}' = \{\mathscr{A}' \mid \mathscr{A}' = \bigcup_{\mathscr{B} \in \mathscr{S}} \{\mathscr{B} \mid \text{CORE}(\mathscr{B}) = \text{CORE}(\mathscr{A})\},\ \mathscr{A} \in \mathscr{S}\}.$$

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 — $LALR(k)$-грам-матика. Число множеств, полученных при слиянии, разве лишь уменьшится. Соответственно уменьшится число $LR(k)$-таблиц. Последние строятся обычным образом по объединенным множествам $LR(k)$-ситуаций. Очевидно, что корректность $LALR(k)$-анализатора, использующего таким образом полученные $LR(k)$-таблицы, не нуждается в доказательстве.

Пример 3.15. Проверим, является ли рассмотренная ранее $LR(1)$-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 нее была построена в примере 3.10:

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

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

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

$\mathcal{A}_3=\{[S\to SaS.b, \varepsilon \mid a], [S\to S.aSb, a \mid b]\}$;

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

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

$\mathcal{A}_6 = \{[S\to SaS.b, a \mid b], [S\to S.aSb, a \mid b]\}$;

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

Ясно, что в приведенной системе можно слить множества $\mathcal{A}_2$ и $\mathcal{A}_4$:

$\mathcal{A}_{2\,4} = \mathcal{A}_2 \cup \mathcal{A}_4 = \{[S\to Sa.Sb, \varepsilon \mid a \mid b], [S\to .SaSb, a \mid b], [S\to ., a \mid b]\}$,

множества $\mathcal{A}_3$ и $\mathcal{A}_6$:

$\mathcal{A}_{3\,6} = \mathcal{A}_3 \cup \mathcal{A}_6 = \{[S\to SaS.b, \varepsilon \mid a \mid b], [S\to S.aSb, a \mid b]\}$,

а также множества $\mathcal{A}_5$ и $\mathcal{A}_7$:

$\mathcal{A}_{5\,7} = \mathcal{A}_5 \cup \mathcal{A}_7 = \{[S\to SaSb., \varepsilon \mid a \mid b\}$.

Полученные множества $\mathcal{A}_{2\,4}$, $\mathcal{A}_{3\,6}$ и $\mathcal{A}_{5\,7}$ — не противоречивы. Системе объединенных множеств $LR(k)$-ситуаций соответствует управляющая таблица (табл. 3.6).

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_{24}	T_{57}
T_{57}	reduce 1	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 для рассматриваемой грамматики обнаруживает ошибку в цепочке abb, достигнув пятой конфигурации:

$$(T_0, abb, \varepsilon) \vdash (T_0ST_1, abb, 2) \vdash (T_0ST_1aT_2, bb, 2) \vdash (T_0ST_1aT_2ST_3, bb, 2\,2) \vdash$$
$$\vdash (T_0ST_1aT_2ST_3bT_5, b, 2\,2),$$

а $LALR(1)$-анализатор — на шестой:

$$(T_0, abb, \varepsilon) \vdash (T_0ST_1, abb, 2) \vdash (T_0ST_1aT_{24}, bb, 2) \vdash (T_0ST_1aT_{24}ST_{36}, bb, 2\,2) \vdash$$
$$\vdash (T_0ST_1aT_{24}ST_{36}bT_{57}, b, 2\,2) \vdash (T_0ST_1, b, 2\,2\,1).$$