

§ 5. Основная теорема LR анализа

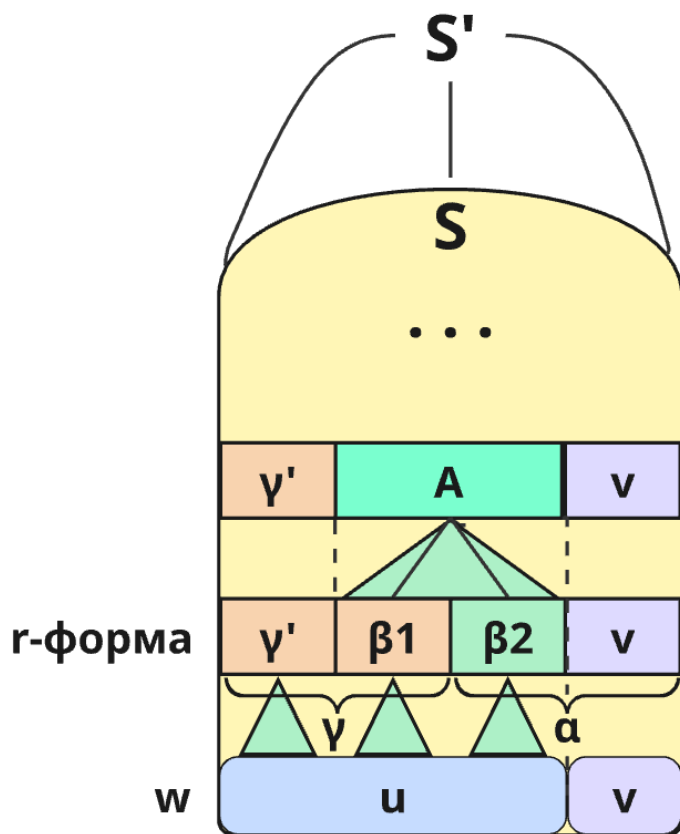
Вернемся к LR(0)-пунктам и ε -НКА пунктам.

Опр. Пункт $[A \rightarrow \beta_1 \bullet \beta_2]$ называется **допустимым** для активного префикса $\gamma = \gamma' \beta_1$ некоторой r -формы $\gamma \alpha$, если существует правый вывод

$$S' \Rightarrow^* \gamma' A v \Rightarrow \gamma' \beta_1 \beta_2 v \Rightarrow^* uv \quad (\text{см. рис. 1}).$$

($\gamma', \beta_1, \beta_2$ может содержать нетерминалы, u, v содержит только терминалы)

Рис. 1



Теорема (основная теорема LR-анализа). Пусть G – зафиксированная расширенная грамматика. I_G – ее автомат пунктов. Пункт $q = [A \rightarrow \beta_1 \bullet \beta_2]$ является допустимым для активного префикса γ некоторой r -формы $\Leftrightarrow$ в автомате пунктов I_G есть путь из q_0 в q , помеченный цепочкой γ .

Доказательство:

Будем называть «**базисными**» пунктами те пункты $q = [A \rightarrow \beta_1 \bullet \beta_2]$, где $\beta_1 \neq \varepsilon$ (точка не в начале) (β_2 может быть равным ε), и пункт $[S' \rightarrow \bullet S]$.

Будем называть «**базисными**» переходами – переходы с пометками, не равными ε .

Замечание: В базисные пункты ведут базисные переходы, и только они.

Лемма 1. Пусть γ – активный префикс. Тогда найдется базисный пункт, допустимый для γ .

Доказательство леммы 1:

Пусть γ – активный префикс некоторой r -формы, принадлежащей выводу

$$S' \Rightarrow^* w.$$

(**Напоминание:** γ не выходит за правый конец основы этой r -формы).

Пусть $\gamma\alpha$ – первая (ближайшая к S') r -форма, где γ такой активный префикс.

Случай 1) Пусть $\gamma\alpha$ r -форма, встретившаяся не на первом шаге.

$$S' \Rightarrow S \Rightarrow^* \gamma\alpha \Rightarrow^* w.$$

Основа r -формы $\gamma\alpha$ не может целиком лежать в α .

От противного: предположим $\gamma\alpha = \gamma\alpha'\beta\alpha''$, где правило. Тогда

$$S' \Rightarrow S \Rightarrow^* \gamma\alpha'\beta\alpha'' \Rightarrow^* w \text{ равносильно}$$

$$S' \Rightarrow S \Rightarrow^* \gamma\alpha'\beta\alpha'' \Rightarrow \gamma\alpha'\beta\alpha'' \Rightarrow^* w$$

Следовательно, $\gamma\alpha\beta$ – префикс r -формы, стоящей ближе к S' (рис. 2).

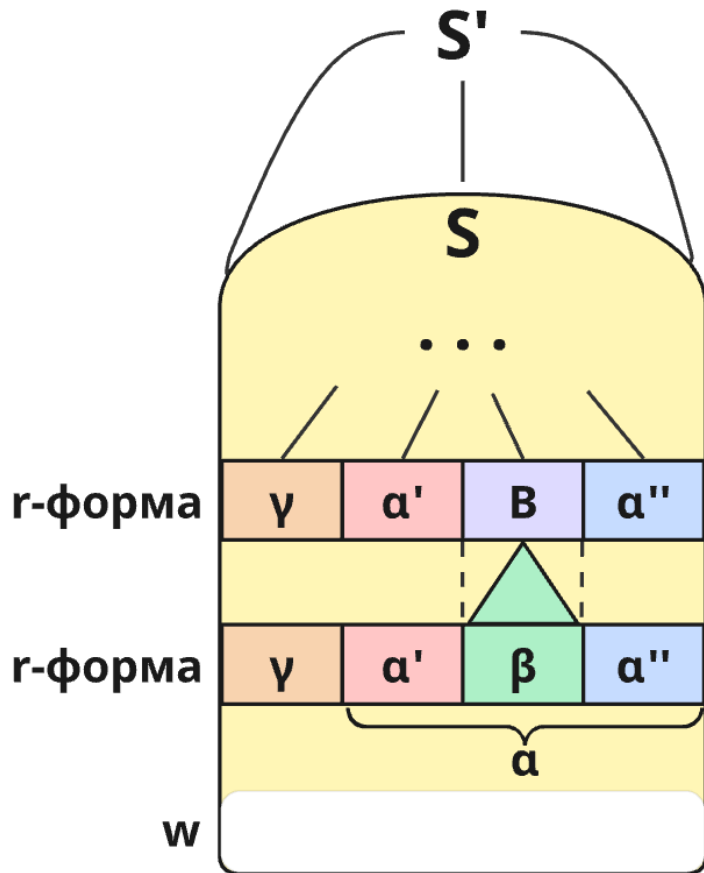


Рис. 2

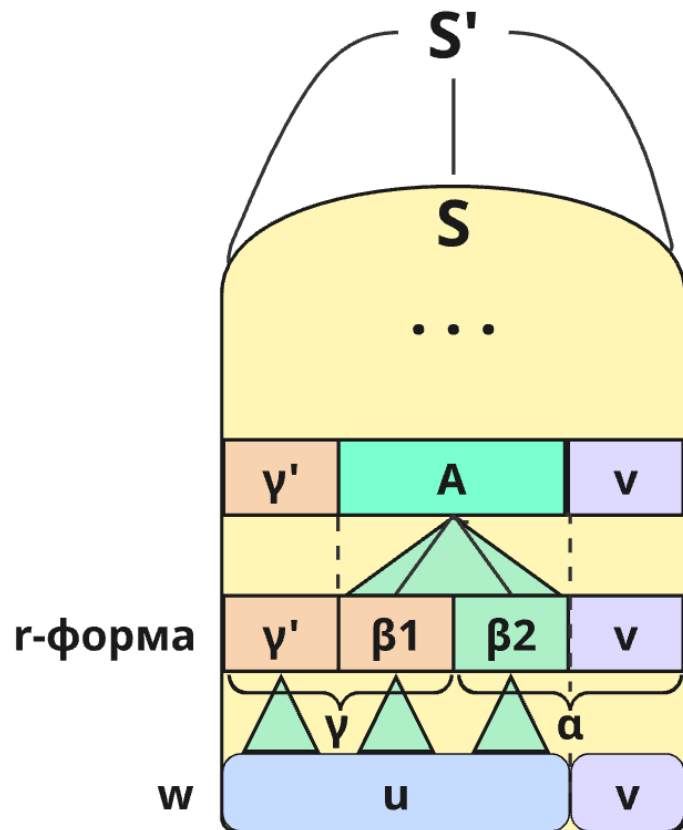
Таким образом, основа r -формы $\gamma\alpha$ имеет вид $\beta_1\beta_2$, где $\gamma = \gamma'\beta_1$,
 $\alpha = \beta_2v$, $\beta_1 \neq \varepsilon$, $v \in \Sigma^*$.

$\Rightarrow$ существует правило $A \rightarrow \beta_1\beta_2$.

$\Rightarrow S' \Rightarrow^* \gamma'Av \Rightarrow \gamma'\beta_1\beta_2v = \gamma\alpha \Rightarrow^* w$.

Тогда пункт $[A \rightarrow \beta_1 \bullet \beta_2]$ – базисный, допустим для γ (см. рис. 3).

Рис. 3



Случай 2) Пусть $\gamma\alpha$ появилась на первом шаге.

$$S' \Rightarrow S = \gamma\alpha \Rightarrow^* w.$$

Т.е. либо $\gamma = \varepsilon$, $\alpha = S$; это значит $[S' \rightarrow \bullet S]$ допустим для γ ,
либо $\gamma = S$, $\alpha = \varepsilon$; это значит $[S' \rightarrow S \bullet]$ допустим для γ .

Лемма 1 доказана.

Лемма 2. Пункт $q = [B \rightarrow \bullet \beta]$ допустим для активного префикса $\gamma \Leftrightarrow$ состояние q достижимо по ε -переходам $\delta([A \rightarrow \beta_1 \bullet B\beta_2], \varepsilon) = [B \rightarrow \bullet \beta]$ автомата $\mathbf{I}_G$ из некоторого базисного пункта $[A \rightarrow \beta_1 \bullet B\beta_2]$ допустимого для γ .

Доказательство:

$\Leftarrow$) Пусть $[A \rightarrow \beta_1 \bullet B\beta_2]$ - некоторый пункт, допустимый для $\gamma = \gamma'\beta_1$, причем пункт не обязательно базисный (т.е. β_1 не обязательно не равно ε), и существует правило $B \rightarrow \beta$, т.е. в автомате пунктов $\mathbf{I}_G$ есть ε -переход $\delta([A \rightarrow \beta_1 \bullet B\beta_2], \varepsilon) = [B \rightarrow \bullet \beta]$.

Поскольку пункт $[A \rightarrow \beta_1 \bullet B\beta_2]$ - допустимый, существует вывод

$$S' \Rightarrow^* \gamma' A v \Rightarrow \gamma' \beta_1 B \beta_2 v = \gamma B \beta_2 v \Rightarrow^* uv$$

Восстановим вывод полностью

$$S' \Rightarrow^* \gamma' A v \Rightarrow \gamma' \beta_1 B \beta_2 v = \gamma B \beta_2 v \Rightarrow^*$$

$$\Rightarrow^* \gamma B u_2 v \Rightarrow^* \gamma \beta u_2 v \Rightarrow^* u_1 u_2 v = uv = w \text{ (см. рис. 4).}$$

(β_2 может содержать нетерминалы, v содержит только терминалы, поскольку вывод правый, нетерминалы из β_2 заменяются раньше нетерминала B)

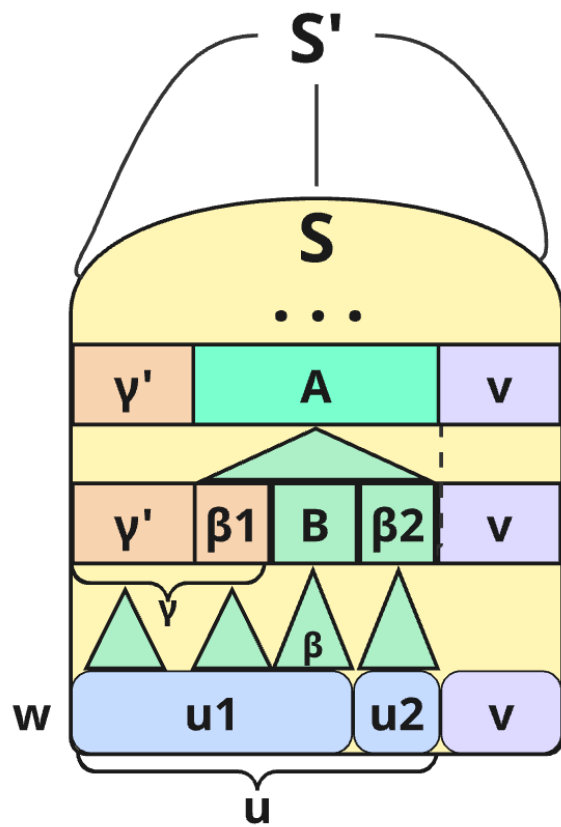


Рис. 4

Таким образом,

$$S' \Rightarrow^* \gamma B \beta_2 v \Rightarrow^* \gamma B u_2 v \Rightarrow^* \gamma \beta u_2 v \Rightarrow^* u_1 u_2 v = uv = w$$

Это значит, что по определению пункт $[B \rightarrow \bullet \beta]$ допустим для γ .

Мы показали, что если в автомате пунктов существует ε -переход $\delta([A \rightarrow \beta_1 \bullet B \beta_2], \varepsilon) = [B \rightarrow \bullet \beta]$, то второй пункт также допустим для γ .

Допустимость наследуется по ε -переходу $\Rightarrow$ она наследуется по любому пути из ε -переходов.

($\Leftarrow$) доказано.

$\Rightarrow$) Пусть пункт $q = [B \rightarrow \bullet \beta]$ допустим для активного префикса γ , т.е. существует вывод $S' \Rightarrow^* \gamma Bu \Rightarrow \gamma \beta u \Rightarrow^* w$.
Рассмотрим первое такое появление γ .

По **лемме 1** пункту $q = [B \rightarrow \bullet \beta]$ соответствует базисный пункт $[A \rightarrow \beta_1 \bullet \beta_2]$, допустимый для γ , где γ появляется сразу после применения правила $A \rightarrow \beta_1 \beta_2$:

$$S' \Rightarrow^* \gamma' Au \Rightarrow \gamma' \beta_1 \beta_2 u = \gamma \beta_2 u.$$

Тогда $S' \Rightarrow^* \gamma' Au \Rightarrow \gamma' \beta_1 \beta_2 u = \gamma \beta_2 u \Rightarrow^* \gamma Bu \Rightarrow \gamma \beta u \Rightarrow^* w$.

Здесь u, v, w – цепочки терминалов.

Так u – цепочка из терминалов, то из β_2 правым выводом выводима цепочка с префиксом B (т.е. начинающаяся с нетерминала).

Следовательно, β_2 начинается с нетерминала.

Обозначим его C . Тогда $\beta_2 = C\alpha_0$.

Если $\gamma\beta_2u = \gamma Bu$, то $C = B$.

Если $\gamma\beta_2u \neq \gamma Bu$, то существует (**правый!**) вывод

$$\beta_2 = C\alpha_0 \Rightarrow B_1\alpha_1u_0 \Rightarrow^* B_1u_1 \Rightarrow B_2\alpha_2u_1 \Rightarrow^* B_2u_2 \Rightarrow \dots \Rightarrow^* B_ku_k = Bu_k.$$

Здесь $u_1, u_2, \dots, u_k$ – цепочки терминалов, и

$$\alpha_0 \Rightarrow^* u_0, \alpha_1u_0 \Rightarrow^* u_1, \alpha_2u_1 \Rightarrow^* u_2, \alpha_3u_2 \Rightarrow^* u_3, \dots, \alpha_{k-1}u_{k-2} \Rightarrow^* u_{k-1},$$

$$\alpha_ku_{k-1} \Rightarrow^* u_k \text{ (см. рис.5)}$$

В таком выводе применялись правила:

$$C \rightarrow B_1\alpha_1$$

$$B_1 \rightarrow B_2\alpha_2$$

...

$$B_{k-1} \rightarrow B\alpha_k.$$

Следовательно, в автомате пунктов I_G имеются ε -переходы:

$$\delta([A \rightarrow \beta_1 \bullet \beta_2], \varepsilon) = [C \rightarrow \bullet B_1\alpha_1] \quad (\beta_2 = C\alpha_0)$$

$$\delta([C \rightarrow \bullet B_1\alpha_1], \varepsilon) = [B_1 \rightarrow \bullet B_2\alpha_2]$$

...

$$\delta([B_{k-1} \rightarrow \bullet B\alpha_k], \varepsilon) = [B \rightarrow \bullet \beta]$$

Таким образом, мы доказали, что пункт $q = [B \rightarrow \bullet \beta]$ достижим из $[A \rightarrow \beta_1 \bullet \beta_2]$ по ε -переходам.

Лемма 2 доказана.

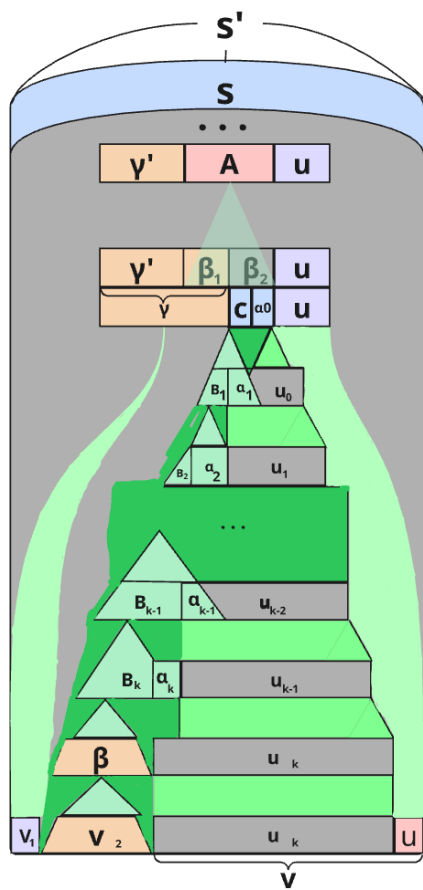


Рис. 5

Продолжение доказательства теоремы.

$\Leftarrow$) Дано: существует путь из q_0 в q , помеченный γ .

Докажем индукцией по длине $|\gamma|$, что пункт $q = [A \rightarrow \beta_1 \bullet \beta_2]$ является допустимым для активного префикса γ

Б.И. $|\gamma| = 0$, т.е. $\gamma = \varepsilon$.

Пункт $[S' \rightarrow \bullet S] = q_0$ допустим для активного префикса ε по определению.

По **лемме 2**, любой пункт (в том числе базисный), достижимый из q_0 по ε -переходам, допустим для ε .

Ш.И. $|\gamma| \geq 1$, т.е. $\gamma = \bar{\gamma} X$.

В пути из q_0 в q последний базисный переход идет по $\delta([A \rightarrow \bar{\beta}_1 \bullet X\beta_2], X) = [A \rightarrow \bar{\beta}_1 X \bullet \beta_2]$.

После базисного перехода возможны несколько ε -переходов.

Путь от q_0 до $[A \rightarrow \bar{\beta}_1 \bullet X\beta_2]$ помечен $\bar{\gamma}$, длина которого строго меньше длины γ . По предположению индукции, это пункт допустим для $\bar{\gamma}$.

Следовательно, $\bar{\gamma} = \gamma' \bar{\beta}_1$ и существует вывод

$$S' \Rightarrow^* \gamma' A v \Rightarrow \gamma' \underbrace{\bar{\beta}_1}_{\beta_1} \underbrace{X \beta_2}_{\beta_2} v \Rightarrow^* w$$

$$(S' \Rightarrow^* \gamma' A v \Rightarrow \gamma' \beta_1 \beta_2 v \Rightarrow^* w)$$

По определению, базисный пункт $[A \rightarrow \bar{\beta}_1 X \bullet \beta_2]$ также допустим для $\gamma = \gamma' \bar{\beta}_1 X$:

$$S' \Rightarrow^* \gamma' A v \Rightarrow \gamma' \underbrace{\bar{\beta}_1 X}_{\beta_1} \underbrace{\beta_2}_{\beta_2} v \Rightarrow^* w$$

$$(S' \Rightarrow^* \gamma' A v \Rightarrow \gamma' \beta_1 \beta_2 v \Rightarrow^* w)$$

Конец пути до q проходит по ε -переходам. Следовательно, q допустим для γ (по лемме 2).

($\Leftarrow$) доказано.

$\Rightarrow$) Дано: пункт $q = [A \rightarrow \beta_1 \bullet \beta_2]$ допустим для γ .

Докажем индукцией по длине цепочки γ , что в автомате пунктов I_G есть путь из q_0 в q , помеченный цепочкой γ .

Б.И. Пусть $q = [A \rightarrow \beta_1 \bullet \beta_2]$ допустим для $\gamma = \varepsilon$.

Тогда $\beta_1 = \varepsilon$, т.к. $\gamma = \gamma' \beta_1$.

Единственный **базисный** пункт, допустимый для ε ($\beta_1 = \varepsilon$), это пункт $[S' \rightarrow \bullet S] = q_0$.

По **лемме 2** состояние q достижимо из q_0 по ε -переходам.

Ш.И. $\gamma = \bar{\gamma} X$.

По **лемме 1** существует базисный пункт, допустимый для $\gamma = \gamma' \bar{\beta}_1$,
это пункт $[A \rightarrow \beta_1 X \bullet \beta_2] = \tilde{q}$:

$$S' \Rightarrow^* \gamma' A v \Rightarrow \gamma' \underbrace{\beta_1}_{\beta_1} X \underbrace{\beta_2}_{\beta_2} v \Rightarrow^* w$$

$$(S' \Rightarrow^* \gamma' A v \Rightarrow \gamma' \beta_1 \beta_2 v \Rightarrow^* w)$$

Следовательно, $\gamma = \gamma' \beta_1 X$ и существует вывод

$$S' \Rightarrow^* \gamma' A v \Rightarrow \gamma' \beta_1 X \beta_2 v \Rightarrow^* w.$$

Пункт $[A \rightarrow \beta_1 \bullet X \beta_2]$ допустим для $\bar{\gamma} = \gamma' \beta_1$. Тогда по предположению индукции, существует путь q_0 в этот пункт, помеченный $\bar{\gamma}$.

В автомате существует переход $\delta([A \rightarrow \beta_1 \bullet X\beta_2], X) = [A \rightarrow \beta_1 X \bullet \beta_2]$.

Достроим по этому переходу путь из q_0 в $\tilde{q}$, помеченный γ .

Для всех базисных пунктов вида $\tilde{q}$ существует путь из q_0 в $\tilde{q}$, помеченный γ .

Для небазисных пунктов, по лемме 2, существует путь из базисного пункта по ε -переходам.

($\Rightarrow$) доказано.

Теорема доказана.

Следствие 1. Язык, распознаваемый автоматом пунктов грамматики G , совпадает с языком всех активных префиксов грамматики G .

Это просто переформулировка теоремы.

Следствие 2. Состояние в LR(0)-автомате, достижимое из q_0 по пути, помеченному γ , совпадает с множеством пунктов, допустимых для активного префикса γ .

Доказательство следует из теоремы и того факта, что LR(0)-автомат — ДКА, построенное по ε -НКА пунктов I_G .