§ 5. Основная теорема LR анализа

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

Опр. Пункт $\left[A \to \beta_1 \bullet \beta_2\right]$ называется допустимым для активного префикса $\alpha\beta_1$ некоторой r-формы, если существует правый вывод $S' \Rightarrow^* \alpha A\mathrm{w} \Rightarrow \alpha\beta_1\beta_2\mathrm{w} \Rightarrow^* \mathrm{uw}$.

Теорема. Пусть G – зафиксированная расширенная грамматика. $\mathtt{I}_G$ – ее автомат пунктов. Пункт q = $\left[A \to \beta_1 \bullet \beta_2\right]$ является допустимым для активного префикса γ некоторой r-формы $\Leftrightarrow$ когда в $\mathtt{I}_G$ существует путь из q_0 в q, помеченный цепочкой γ.

Доказательство:

Будем называть «базисными» пунктами те пункты, где $\beta_1 \neq \varepsilon$ (точка не в начале), и пункт $[S' \to \bullet S]$.

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

$$S' \Rightarrow^{*} \mathrm{w}\,.$$

(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 $\gamma\alpha$ не на первом шаге.

$S' \Rightarrow S \Rightarrow^{*} \gamma\alpha \Rightarrow^{*} \mathrm{w}$.

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

(от противного) Предположим $\gamma\alpha = \gamma\alpha'(основа)\alpha''$.

$S' \Rightarrow S \Rightarrow^{*} \gamma\alpha'(основа)\alpha'' \Rightarrow^{*} \mathrm{w}$ равносильно

$S' \Rightarrow S \Rightarrow^{*} \gamma\alpha' B\alpha'' \Rightarrow \gamma\alpha'(основа)\alpha'' \Rightarrow^{*} \mathrm{w}$

Тогда $\gamma\alpha'$B – префикс r-формы, стоящей ближе к S' .

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

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

$$\Rightarrow\ S' \Rightarrow^* \gamma' A\mathrm{u} \Rightarrow \gamma'\beta_1\beta_2\mathrm{u} = \gamma\alpha \Rightarrow^* \mathrm{w}.$$

Тогда пункт $\left[A \to \beta_1 \bullet \beta_2\right]$ – базисный, допустим для γ.

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

$$S' \Rightarrow S = \gamma\alpha \Rightarrow^* \mathrm{w}\,.$$

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

либо γ = S, $\alpha = \varepsilon$; это значит $[S' \to S\bullet]$ допустим для γ.

Лемма 1 доказана.

Лемма 2. Пункт q = $\left[B \to \bullet\beta\right]$ допустим для активного префикса $\gamma \Leftrightarrow$ состояние q достижимо по ε-переходам автомата $\mathtt{I}_G$ из некоторого базисного пункта, допустимого для γ.

Доказательство:

$\Leftarrow$) Пусть $\left[A \to \beta_1 \bullet B\beta_2\right]$ допустим для $\gamma = \gamma'\beta_1$, пункт не обязательно базисный, и существует правило $B \to \beta$.

Построим вывод

$$S' \Rightarrow^* \gamma' A\mathrm{u} \Rightarrow \gamma'\beta_1 B\beta_2\mathrm{u} = \gamma B\beta_2\mathrm{u} \Rightarrow^* \gamma B\mathrm{vw} \Rightarrow \gamma\beta\mathrm{vw} \Rightarrow^* \mathrm{w}.$$

(β_2 может содержать нетерминалы, v содержит только терминалы)

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

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

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

($\Leftarrow$) доказана.

$\Rightarrow$) Дано: существует $S' \Rightarrow^* \gamma B\mathrm{v} \Rightarrow \gamma\beta\mathrm{v} \Rightarrow^* \mathrm{w}$.
Рассмотрим первое появление γ.
По лемме 1 существует базисный пункт $[A \to \beta_1 \bullet \beta_2]$, допустимый для γ.
Более того, γ появляется сразу после применения правила $A \to \beta_1\beta_2$.

Тогда $S' \Rightarrow^* \gamma' A\mathrm{u} \Rightarrow \gamma'\beta_1\beta_2\mathrm{u} = \gamma\beta_2\mathrm{u} \Rightarrow^* \gamma B\mathrm{v} \Rightarrow \gamma\beta\mathrm{v} \Rightarrow^* \mathrm{w}$.
(возможно $\gamma\beta_2$u = γBv)

Тогда из β_2 выводима цепочка с префиксом Bv (т.е. начинающаяся с нетерминала).
Следовательно, β_2 начинается с нетерминала. Обозначим его C.

Если $\gamma\beta_2$u = γBv, то C = B.
Если $\gamma\beta_2$u ≠ γBv, то существует вывод

$$C \Rightarrow B_1\alpha_1 \Rightarrow^* B_1\mathrm{u}_1 \Rightarrow B_2\alpha_2\mathrm{u}_1 \Rightarrow^* B_2\mathrm{u}_2 \Rightarrow \ldots \Rightarrow^* B_k\mathrm{u}_k = B\mathrm{u}_k\,.$$

В таком выводе применялись правила:

$C \to B_1\alpha_1$

$B_1 \to B_2\alpha_2$

…

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

Тогда в автомате пунктов имеются ε-переходы:

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

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

…

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

Доказали, что пункт q = $[B \to \bullet \beta]$ достижим из $[A \to \beta_1 \bullet \beta_2]$ по ε-переходам.

Лемма 2 доказана.

(продолжение доказательства теоремы)
$\Leftarrow$) Дано: существует путь из q_0 в q, помеченный γ.
(индукция по длине цепочки γ)

Б.И. $|\gamma| = 0$, т.е. $\gamma = \varepsilon$.
Пункт $[S' \to \bullet S]$ = 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
$\delta\left(\left[A \to \bar{\beta}_1 \bullet X\beta_2\right], X\right) = \left[A \to \bar{\beta}_1 X \bullet \beta_2\right]$.

После базисного перехода возможны несколько ε-переходов.
Путь от q_0 до $\left[A \to \bar{\beta}_1 \bullet X\beta_2\right]$ 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
$S' \Rightarrow^* \gamma' A\mathrm{u} \Rightarrow \gamma'\bar{\beta}_1 X\beta_2\mathrm{u} \Rightarrow^* \mathrm{w}$.

По определению, базисный пункт $\left[A \to \bar{\beta}_1 X \bullet \beta_2\right]$ допустим для
$\gamma = \gamma'\bar{\beta}_1 X$.

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

($\Leftarrow$) доказана.

$\Rightarrow$) Дано: q, допустим для γ.
(индукция по длине цепочки γ)

Б.И. Пусть q = $[A \to \beta_1 \bullet \beta_2]$ допустим для $\gamma = \varepsilon$.
Тогда $\beta_1 = \varepsilon$, т.к. $\gamma = \alpha\beta_1$.

Единственный **базисный** пункт, допустимый для ε, это $[S' \to \bullet S]$ = q_0.

По лемме 2 q достижима из q_0 по ε-переходам.

Ш.И. $\gamma = \bar{\gamma}$ X.

По лемме 1 существует базисный пункт, допустимый для γ,

это $[A \to \beta_1 X \bullet \beta_2] = \tilde{q}$.

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

$$S' \Rightarrow^* \gamma' A\mathrm{u} \Rightarrow \gamma' \beta_1 X \beta_2 \mathrm{u} \Rightarrow^* \mathrm{w} .$$

Пункт $[A \to \beta_1 \bullet X \beta_2]$ 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 $\delta([A \to \beta_1 \bullet X\beta_2], X) = [A \to \beta_1 X \bullet \beta_2]$. 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.

($\Rightarrow$) доказана.
Теорема доказана.

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

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