


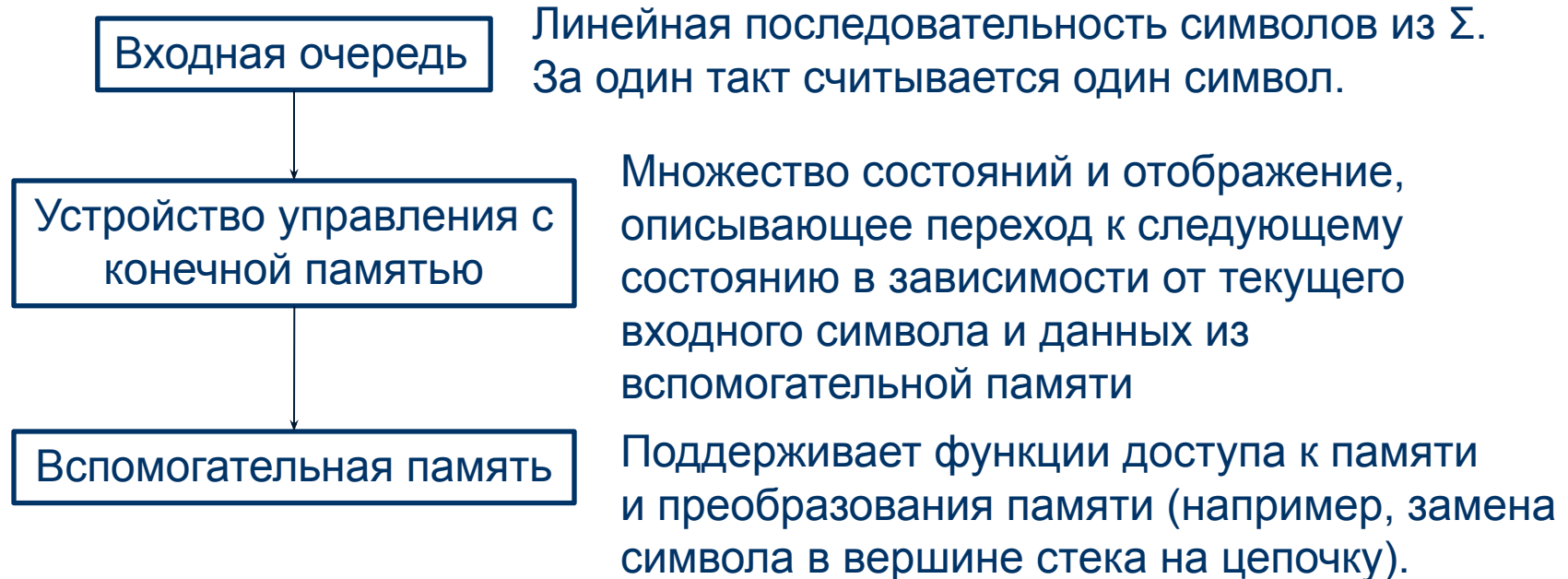
Языки программирования и методы трансляции

Лекция 8.

Распознающие автоматы для
контекстно-свободных
грамматик



Распознающие автоматы



Распознаватель допускает входную цепочку w , если, начиная с начальной конфигурации, в которой w записана во входной очереди, распознаватель может выполнить последовательность тактов, завершающуюся конечной конфигурацией.

Конечный автомат

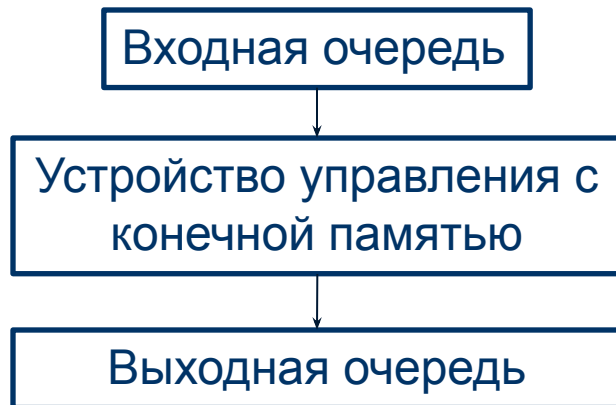
Входная очередь

Устройство управления с
конечной памятью

Недетерминированный конечный автомат –
пятерка объектов: $K=(Q, \Sigma, \delta, q_0, F)$

Q – множество состояний УУ,
 Σ – алфавит входных символов,
 $\delta: Q \times \Sigma \rightarrow P(Q)$ – функция переходов,
 q_0 – начальное состояние,
 F – множество заключительных состояний

Конечные преобразователи



Недетерминированный конечный преобразователь – шестерка объектов:
 $K=(Q, \Sigma, \Delta, \delta, q_0, F)$
 Q – множество состояний УУ,
 Σ – алфавит входных символов,
 Δ – алфавит выходных символов
 $\delta: Q \times \Sigma \rightarrow P(Q) \times \Delta$ – функция переходов,
 q_0 – начальное состояние,
 F – множество заключительных состояний

Конфигурация конечного преобразователя - тройка (q, ax, y) , где q – состояние, x – необработанная часть входной цепочки, y – построенная часть выходной цепочки.

Такт работы конечного преобразователя – переход от конфигурации (q, ax, y) к конфигурации (r, x, yz) , если $\delta(q, a)$ содержит (r, z) .

Регулярный перевод

Формализация такта КП как отношения \vdash на множестве конфигураций:

Для всех $q \in Q$, $a \in \Sigma$, $x \in \Sigma$ и $y \in \Delta$, таких, что $\delta(q, a)$ содержит (r, z) , конфигурации $(q, ax, y) \vdash (r, x, yz)$.

Транзитивное замыкание отношения \vdash обозначим \vdash^+

Рефлексивно-транзитивное замыкание отношения \vdash обозначим \vdash^*

Цепочка $y \in \Delta$ называется выходом для цепочки $x \in \Sigma$, если $(q_0, x, \varepsilon) \vdash^* (q, \varepsilon, y)$

Переводом, определяемым конечным преобразованием, называется множество пар цепочек:

$$\tau(M) = \{ (x, y) \mid (q_0, x, \varepsilon) \vdash^* (q, \varepsilon, y) \text{ для некоторого } q \in F \}$$

(регулярный, или конечный, перевод)

Пример конечного преобразователя: распознавание действительных чисел

$M = \{(q_0, q_1, q_2, q_3, q_4), \{d, .\}, \{1, 2, 3\}, \delta, q_0, \{q_4\}\}$

$\delta(q_0, d) = \{(q_1, \varepsilon)\}$

$\delta(q_0, .) = \{(q_3, \varepsilon)\}$

$\delta(q_1, d) = \{(q_1, \varepsilon)\}$

$\delta(q_1, .) = \{(q_2, \varepsilon), (q_4, 1)\}$

$\delta(q_2, d) = \{(q_2, \varepsilon), (q_4, 3)\}$

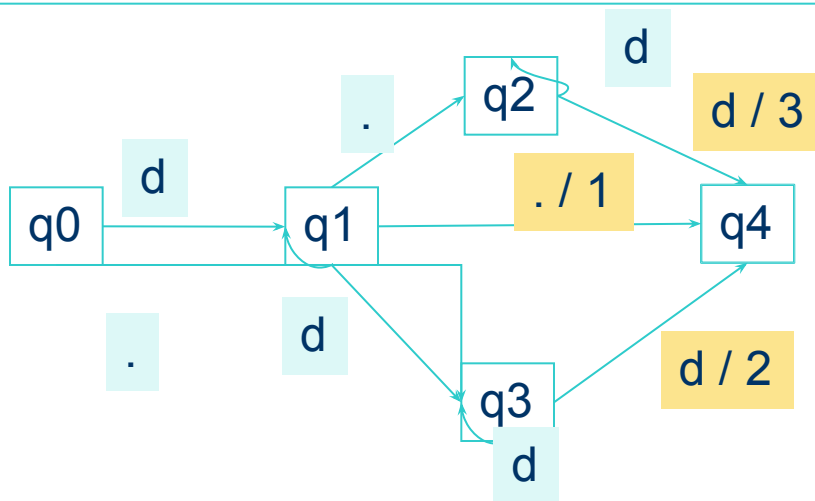
$\delta(q_3, d) = \{(q_3, \varepsilon), (q_4, 2)\}$

Выход:

1 – dd...d.

2 - .dd...d

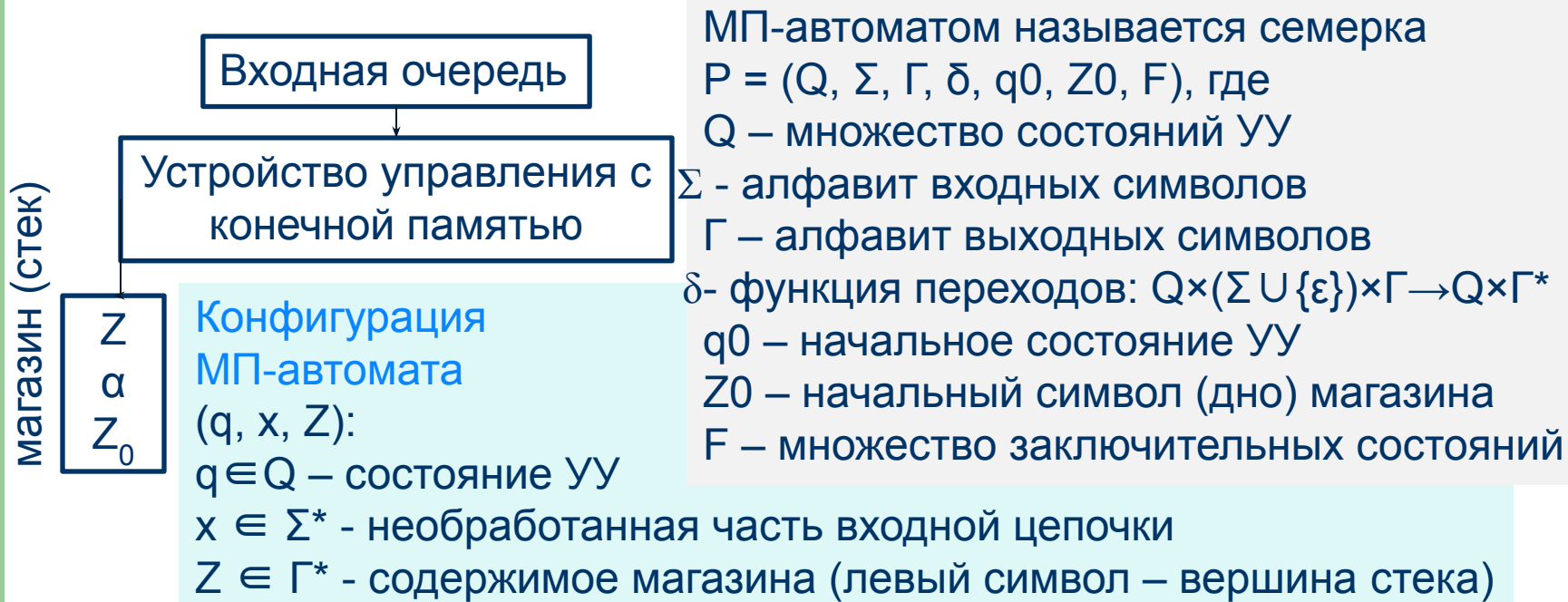
3 – dd...d.dd...d



Пример: входная цепочка 234.17

$(q_0, 234.17, \varepsilon) \vdash (q_1, 34.17, \varepsilon) \vdash$
 $\vdash (q_1, 4.17, \varepsilon) \vdash (q_1, .17, \varepsilon) \vdash (q_2, 17, \varepsilon) \vdash$
 $\vdash (q_2, 7, \varepsilon) \vdash (q_4, \varepsilon, 3)$

Автомат с магазинной памятью



Такт работы МП-автомата:

переход от конфигурации (q, aw, Z) к конфигурации (r, w, γα), если δ(q, a, Z) содержит (r, γ).

$$(q, aw, Z) \vdash (r, w, \gamma\alpha)$$

Пример: МП-автомат для распознавания симметричных цепочек

$P = (\{q_0, q_1, q_2\}, \{a, b\}, \{Z, a, b\}, \delta, q_0, Z, \{q_2\})$

Функции переходов:

- 1) $\delta(q_0, a, Z) = \{(q_0, aZ)\}$
- 2) $\delta(q_0, b, Z) = \{(q_0, bZ)\}$
- 3) $\delta(q_0, a, a) = \{(q_0, aa), (q_1, \varepsilon)\}$
- 4) $\delta(q_0, a, b) = \{(q_0, ab)\}$
- 5) $\delta(q_0, b, a) = \{(q_0, ba)\}$
- 6) $\delta(q_0, b, b) = \{(q_0, bb), (q_1, \varepsilon)\}$
- 7) $\delta(q_1, a, a) = \{(q_1, \varepsilon)\}$
- 8) $\delta(q_1, b, b) = \{(q_1, \varepsilon)\}$
- 9) $\delta(q_1, \varepsilon, Z) = \{(q_2, \varepsilon)\}$

Примеры:

$(q_0, baab, Z) \vdash (q_0, aab, bZ) \vdash$
 $\vdash (q_0, ab, abZ) \vdash (q_0, b, aabZ) \vdash$
 $\vdash (q_0, \varepsilon, baabZ) - \text{ошибка}$

$(q_0, baab, Z) \vdash (q_0, aab, bZ) \vdash$
 $\vdash (q_0, ab, abZ) \vdash (q_1, b, bZ) \vdash$
 $\vdash (q_1, \varepsilon, Z) \vdash (q_2, \varepsilon)$

Расширенный МП-автомат

Лемма:

Если $(q, w, A) \vdash^n (q', \varepsilon, \varepsilon)$, то $(q, w, A\alpha) \vdash^n (q', \varepsilon, \alpha)$ для всех $A \in \Gamma$ и

$\alpha \in \Gamma^*$
Расширенным МП-автоматом называется МП-автомат, который за один такт может заменять цепочку конечной длины в вершине магазина другой цепочкой конечной длины. Функция переходов $\delta: Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma^* \rightarrow Q \times \Gamma^*$

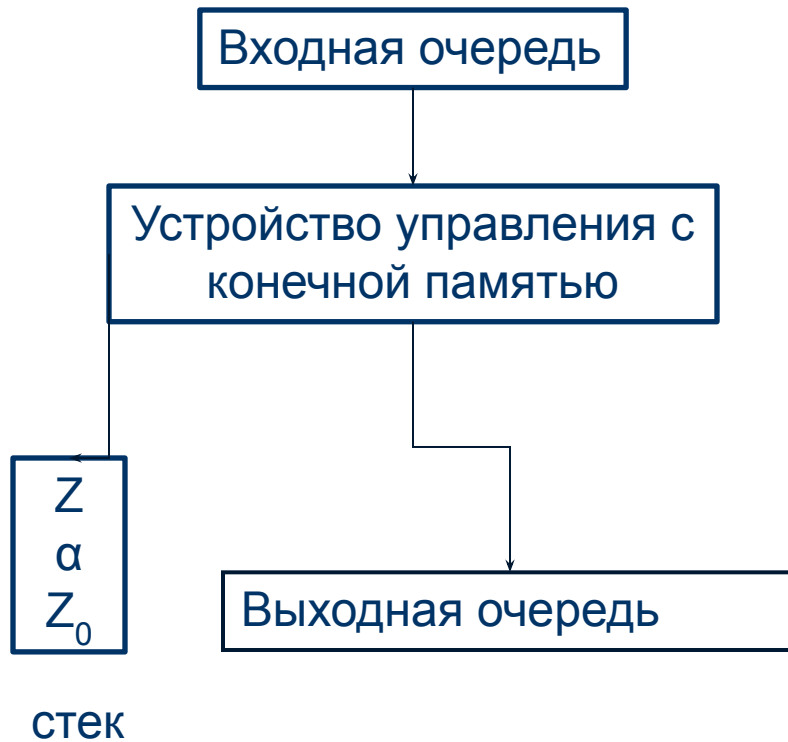
Пример расширенного МП-автомата для языка ww^R

$P = (\{q_0, q_1\}, \{a, b\}, \{S, Z, a, b\}, \delta, q_0, Z, \{q_1\})$

Функции переходов:

- 1) $\delta(q_0, a, \varepsilon) = \{(q_0, a)\}$
- 2) $\delta(q_0, b, \varepsilon) = \{(q_0, b)\}$
- 3) $\delta(q_0, \varepsilon, \varepsilon) = \{(q_0, S)\}$
- 4) $\delta(q_0, \varepsilon, aSa) = \{(q_0, S)\}$
- 5) $\delta(q_0, \varepsilon, bSb) = \{(q_0, S)\}$
- 6) $\delta(q_0, \varepsilon, SZ) = \{(q_1, \varepsilon)\}$

Преобразователи с магазинной памятью



Размеченный вывод цепочек в КС-грамматике

Предложение: $i+i*i$

Грамматика:

$$(1) E \rightarrow E+T$$

$$(2) E \rightarrow T$$

$$(3) T \rightarrow T*P$$

$$(4) T \rightarrow P$$

$$(5) P \rightarrow i$$

$$(6) P \rightarrow (E)$$

Левый вывод:

$$\begin{aligned} E &\Rightarrow (1) E+T \Rightarrow (2) T+T \Rightarrow (4) P+T \Rightarrow \\ &\Rightarrow (5) i+T \Rightarrow (3) i+T*P \Rightarrow \\ &\Rightarrow (4) i+P*P \Rightarrow (5) i+i*P \Rightarrow (5) i+i*i \end{aligned}$$

Правый вывод:

$$\begin{aligned} E &\Rightarrow (1) E+T \Rightarrow (3) E+T*P \Rightarrow (5) E+T*i \Rightarrow \\ &\Rightarrow (4) E+P*i \Rightarrow (5) E+i*i \Rightarrow (2) T+i*i \Rightarrow \\ &\Rightarrow (4) P+i*i \Rightarrow (5) i+i*i \end{aligned}$$

Дерево вывода

Упорядоченное дерево, каждый узел которого помечен символом из $\Sigma \cup N \cup \{\varepsilon\}$, называется деревом вывода в КС-грамматике $G=(N, \Sigma, P, S)$, если:

1. Корень дерева помечен основным символом грамматики S .
2. Каждый прямой потомок узла, помеченного символом A , помечен таким символом X , что грамматика содержит правило вывода $A \rightarrow X$.
3. Если узел D_i помечен символом X_i из N , то поддереву D_i должно быть деревом вывода в грамматике $G_i=(N, \Sigma, P, X_i)$.
4. Если X_i - символ из Σ , то поддереву D_i состоит из одной вершины, помеченной X_i .
5. Если корень дерева состоит из единственного потомка, помеченного ε , то этот потомок образует дерево, состоящее из единственной вершины, и в множестве правил грамматики содержится правило $S \rightarrow \varepsilon$.

Определение разбора

Цепочка для КС-грамматики *разобрана*, если известно её дерево вывода.

Пусть заданы КС-грамматика G , правила которой перенумерованы целыми числами $1, 2, \dots, p$, и цепочка терминальных символов α .

Тогда:

левым разбором цепочки α называется последовательность правил, примененных при её левом выводе из S ;

правым разбором цепочки α называется обращение последовательности правил, примененных при правом выводе цепочки из S .

Эквивалентность КС-грамматик и МП-автоматов

Лемма о нисходящем разборе

Пусть $G=(N, \Sigma, P, S)$ – КС-грамматика,

$P = (\{q\}, \Sigma, \Sigma^+N, \delta, q, S, \{q\})$ – МП-автомат с правилами перехода δ :

1. Если $A \rightarrow \alpha$ - правило вывода грамматики G , то $\delta(q, \varepsilon, A)$ содержит (q, α)
2. $\delta(q, a, a) = \{(q, \varepsilon)\}$ для всех a из Σ .

Тогда $A \Rightarrow^m w$ тогда и только тогда, когда $(q, w, A) \vdash^n (q, \varepsilon, \varepsilon)$ для некоторых m, n .

Пример: МП-автомат для скобочных выражений

Грамматика:

$G = (\{E, T, M\}, \{i, +, *, (,)\}, \{E \rightarrow E+T, E \rightarrow T, T \rightarrow T*M, T \rightarrow M, M \rightarrow (E), M \rightarrow i\}, E)$

МП-автомат:

$P = (\{q\}, \{i, +, *, (,)\}, E, T, M, \delta, q, E, \{q\})$

с правилами перехода:

1) $\delta(q, \varepsilon, E) = \{(q, E+T), \{(q, T)\}\}$

2) $\delta(q, \varepsilon, T) = \{(q, T*M), \{(q, M)\}\}$

3) $\delta(q, \varepsilon, M) = \{(q, (E)), \{(q, i)\}\}$

4) $\delta(q, a, a) = \{(q, \varepsilon)\}$ для всех a из $\{i, +, *, (,)\}$

Пример: разбор цепочки $i^*(i+i)$

$$\begin{array}{l} (q, i^*(i+i), E) \vdash (q, i^*(i+i), T) \vdash (q, i^*(i+i), T^*M) \\ \vdash (q, i^*(i+i), i^*M) \vdash (q, *(i+i), *M) \vdash (q, *(i+i), *M) \\ \vdash (q, (i+i), M) \vdash (q, (i+i), (E)) \vdash (q, i+i, E)) \vdash (q, i+i, E+T)) \\ \vdash (q, i+i, T+T)) \vdash (q, i+i, M+T)) \vdash (q, i+i, i+T)) \\ \vdash (q, +i, +T)) \vdash (q, i, T)) \vdash (q, i, M)) \vdash (q, i, i)) \\ \vdash (q,),)) \vdash (q, \varepsilon, \varepsilon) \end{array}$$

Последовательность тактов МП-автомата соответствует левому выводу цепочки $i^*(i+i)$ в грамматике G .

Восходящий синтаксический анализ (свертка слева)

Правый вывод цепочки $i+i*i$:

$$E \Rightarrow (1) E+T \Rightarrow (3) E+T*P \Rightarrow (5) E+T*i \Rightarrow (4) E+P*i \Rightarrow (5) E+i*i \Rightarrow (2) \Rightarrow T+i*i \Rightarrow (4) P+i*i \Rightarrow (5) i+i*i$$

Обращение правого вывода цепочки $i+i*i$:

$$i+i*i \leq P+i*i \leq T+i*i \leq E+i*i \leq E+P*i \leq E+T*i \leq E+T*P \leq E+T \leq E$$

Пусть G – КС-грамматика, $S \Rightarrow \alpha A w \Rightarrow \alpha \beta w \Rightarrow xw$ – правый вывод в ней.

Правовыводимую цепочку $\alpha \beta w$ можно свернуть слева к правовыводимой цепочке $\alpha A w$ с помощью правила вывода $A \rightarrow \beta$. Это вхождение цепочки β в цепочку $\alpha \beta w$ называется основой цепочки $\alpha \beta w$.

Пример: правила вывода $\{S \rightarrow A c, S \rightarrow B d, A \rightarrow a A b, A \rightarrow a b, B \rightarrow a B b b, B \rightarrow a b b\}$.

Цепочка $a a b b b b d$ – правовыводимая, $a b b$ – основа этой цепочки (т.к. $a b b$ – правая часть правила $B \rightarrow a b b$ и $a B b b d$ – правовыводимая цепочка).

Цепочка $a b$ – не основа, т.к. это правая часть правила $A \rightarrow a b$, но $a A b b b d$ – не правовыводимая цепочка.

Лемма о восходящем разборе

Пусть $G=(N, \Sigma, P, S)$ – КС-грамматика. Тогда можно построить такой расширенный МП-автомат, что $L(G)=L(P)$.

$P=({q, r}, \Sigma, \Sigma \cup N \cup \{\perp\}, \delta, q, \perp, \{r\})$

Правила перехода:

1. $\delta(q, a, \varepsilon) = \{(q, a)\}$ для всех терминальных a , т.е. символы со входной ленты переносятся в стек.
2. Если $A \rightarrow \alpha$ – правило вывода грамматики G , то $\delta(q, \varepsilon, \alpha)$ содержит (q, A) .
3. $\delta(q, \varepsilon, \perp S) = \{(r, \varepsilon)\}$.

Автомат P строит правовыводимые цепочки грамматики G , начиная с терминальной цепочки на входной ленте и заканчивая цепочкой S .