разрастании) цепочка $xwy = c^l$ при $0 < l \leqslant n$ (т.е. „накачиваемые“ цепочки располагаются целиком в „зоне“ символов c), то цепочка $uwv = a^n b^n c^{n-|xy|} \notin L'$, так как $|xy| \geqslant 1$ и $n - |xy| < n$, т.е. число символов c окажется меньше, чем число символов a и b. Таким образом, несмотря на то что „накачка“ цепочек x, y в данном случае возможна, нельзя их выбросить, оставаясь в пределах языка. Невозможность других вариантов расположения подцепочки xwy в цепочке языка L' доказывается точно так же, как и в предыдущих примерах. #

Итак, нужно помнить, что выполнение условий леммы о разрастании предполагает и возможность выбрасывания „накачиваемых“ цепочек — все цепочки $z_n = ux^n wy^n v$ из условия леммы при $n \geqslant 0$ должны оставаться в языке L, и если условие леммы выполняется при всех положительных n, но не выполняется при $n = 0$, то этого достаточно для признания того, что цепочка $z = uxwyv \in L$ не удовлетворяет условию леммы о разрастании.

Аналогичная ситуация имеет место, конечно, и при использовании леммы о разрастнии для регулярных языков.

8.4. Магазинные автоматы

Автоматы с магазинной памятью, или ***магазинные автоматы*** (сокращенно — ***МП-автоматы***), образуют класс *распознающих моделей* для *КС-языков* точно так же, как *конечные автоматы* являются распознающими моделями в классе *регулярных языков*.

Понятие МП-автомата является частным случаем общего интуитивного понятия *автомата*, которое мы ввели в **7.5**. Как и всякий автомат, МП-автомат — это устройство, имеющее *блок управления*, *входную ленту*, *головку* и *блок внутренней памяти* в виде ***магазина МП-автомата*** (или *стека*).

Блок управления в каждый момент времени находится в одном из конечного множества Q *состояний*.

Входная лента предполагается „полубесконечной“, т.е. она имеет начало („левый край“), но не имеет конца. Лента разделена на ячейки, пронумерованные от крайней левой ячейки натуральными числами, начиная с единицы; в каждой ячейке может быть записан символ алфавита V, называемого ***входным алфавитом МП-автомата***. В каждый момент времени читающая головка обозревает (или читает) некоторую ячейку входной ленты.

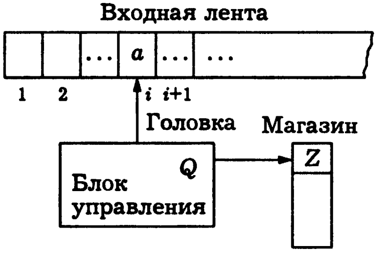

Рис. 8.28

Если в этой ячейке записан некоторый *входной символ*, т.е. символ входного алфавита, то его называют ***обозреваемым*** (в данный момент) ***входным символом*** (рис. 8.28).

Предполагается, что если на ленте записана некоторая непустая цепочка $x = x(1)x(2)\ldots x(k)$ во входном алфавите, то ее символы последовательно записаны в ячейках от первой до k-й, без пропусков, так, что символ $x(i)$ записан в i-й ячейке (рис. 8.29).

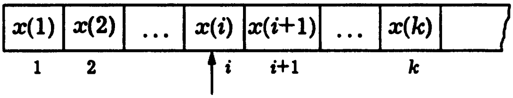

Рис. 8.29

МП-автомат может читать цепочку x, продвигая головку только в одном направлении, слева направо, по шагам, причем за каждый шаг головка переходит от обозреваемой ячейки к следующей. Если в какой-то момент времени обозревается символ $x(i)$, $1 \leqslant i \leqslant k$, записанной на ленте цепочки x, то *непрочитанной частью входной цепочки* x будет подцепочка $x(i+1)\ldots x(k)$. В частности, при $i = k$ непрочитанная часть совпадает с пустой цепочкой, и тогда говорят, что цепочка x полностью прочитана МП-автоматом (рис. 8.30).

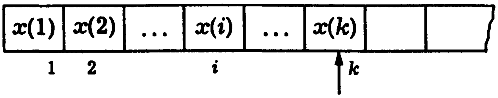

Рис. 8.30

Магазин МП-автомата устроен и работает следующим образом. Как и входная лента, магазин является полубесконечным и разделенным на пронумерованные ячейки, в каждой из которых может быть записан символ алфавита Г, называемого ***магазинным алфавитом МП-автомата***. Входной и магазинный алфавиты МП-автомата могут пересекаться и даже совпадать друг с другом. Символы алфавита Г называют ***магазинными символами***. Первую ячейку магазина называют его ***верхней ячейкой*** (иногда просто ***верхом магазина****). Символ, в данный момент записанный в верхней ячейке магазина, называют ***верхним символом магазина***.

Единственная операция, которую МП-автомат может реализовать с магазином, состоит в следующем: верхний символ Z магазина заменяется некоторой цепочкой γ магазинных символов. При этом если цепочка γ непустая, то она записывается в первых m ячейках, где $m = |\gamma|$ (*длине цепочки* γ), так, что ее первый символ становится верхним символом магазина.

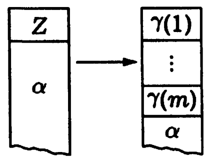

Рис. 8.31

Если до замены Z цепочкой γ под верхней ячейкой** (т.е. в ячейках, начиная со второй) была записана какая-то цепочка α, то после замены она сдвигается „в глубь" магазина и оказывается записанной уже в ячейках, начиная с $(m+1)$-й (рис. 8.31).

Если же верхний символ Z заменяется пустой цепочкой λ, то после такой замены верхним символом магазина становится

*Таким образом, в неформальном описании МП-автомата лента читается „слева направо", а магазин — „сверху вниз".

**Полагают, что, как и символы цепочек на входной ленте, символы цепочек, записанных в магазин, располагаются в последовательных ячейках „сверху вниз" без пропусков ячеек.

первый символ цепочки α (записанной под верхней ячейкой магазина), т.е. цепочка α „поднимается“ на одну ячейку.

В случае, когда α — пустая цепочка, замена верхнего символа магазина пустой цепочкой γ приводит к опустошению магазина (ни в одной из ячеек магазина не записан магазинный символ). Заметим, что, по определению, с пустым магазином МП-автомат не может производить никаких операций.

Описанная выше операция с магазином составляет основу поведения МП-автомата, его, образно говоря, „динамику“. Эта „динамика“ определяется ***системой команд МП-автомата***, которая, аналогично *системе команд конечного автомата*, определяется как конечное множество δ команд, каждая из которых записывается в виде

$$qaZ \to r\gamma, \tag{8.8}$$

где q и r — состояния из множества Q; a — входной символ или пустая цепочка; Z — магазинный символ; γ — цепочка магазинных символов (может быть и пустая).

Если в системе команд δ МП-автомата M есть команда (8.8), то M может в данный момент времени выполнить эту команду, если и только если в данный момент времени его блок управления находится в состоянии q, головка обозревает входной символ a (при условии, что $a \neq \lambda$), а верхним символом магазина является символ Z. Выполнение команды (8.8) состоит в замене верхнего символа магазина цепочкой γ (как это описано выше), переходе блока управления в состояние r (которое может и совпадать с состоянием q) и в продвижении головки на одну ячейку вправо (если a в команде (8.8) не есть пустая цепочка λ) (рис. 8.32).

Если же в команде (8.8) $a = \lambda$, то такую команду МП-автомат может выполнить всякий раз, когда его блок управления окажется в состоянии q, а верхним символом магазина будет символ Z. В этом случае выполнение команды никак не

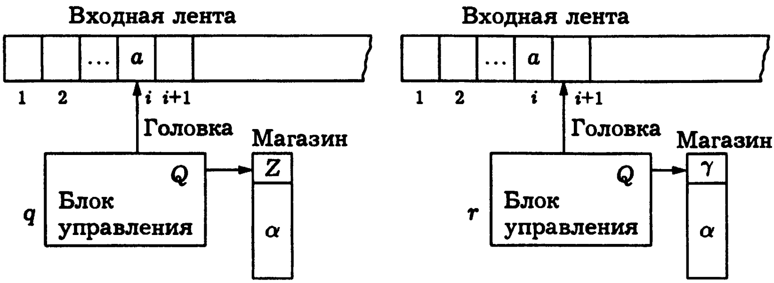

Рис. 8.32

зависит от содержимого входной ленты, а после ее выполнения головка остается на прежнем месте.

Рассмотренную выше процедуру выполнения команды вида (8.8) называют ***шагом*** (или ***тактом***) ***работы МП-автомата*** (при $a = \lambda$ такт работы называют ***λ-тактом***).

Предположим теперь, что в множестве состояний Q блока управления МП-автомата M с системой команд δ фиксировано некоторое ***начальное состояние*** q_0 и подмножество ***заключительных состояний*** F.

Пусть в какой-то начальный момент времени блок управления МП-автомата M находится в начальном состоянии q_0, на ленте записана непустая цепочка x, головка обозревает первую ячейку ленты и, следовательно, первый символ цепочки x является обозреваемым, а в магазине записан только один специальный символ Z_0, называемый ***начальным символом магазина***.

Тогда, если МП-автомат M, выполняя некоторую последовательность команд из δ, прочитывает полностью цепочку x, в результате чего блок управления переходит в некоторое заключительное состояние $q_f \in F$, говорят, что МП-автомат M допускает (распознает, воспринимает) цепочку x, которую называют ***допустимой цепочкой МП-автомата***. Множество всех допустимых цепочек МП-автомата M образует ***язык, допускаемый*** этим ***МП-автоматом***.

Далее мы докажем, что язык любого МП-автомата является КС-языком и, наоборот, любой КС-язык есть язык некоторого МП-автомата. В этом смысле мы и говорим об МП-автоматах как о „распознавателях" КС-языков: всякий МП-автомат допускает те и только те цепочки входных символов, которые порождаются некоторой *КС-грамматикой*. Но чтобы доказывать какие-либо утверждения об МП-автоматах, нужно сначала дать строгое математическое определение этого понятия, не апеллируя уже к наглядным, но не определенным формально идеям ленты, магазина, головки и т.п. Формализация изложенного выше описания проводится во многом аналогично построению определения порождающей грамматики.

Определение 8.7. МП-автомат задается упорядоченной семеркой

$$M = (Q,\ V,\ \Gamma,\ q_0,\ F,\ Z_0,\ \delta),$$

где Q — конечное множество состояний; V — алфавит, называемый входным алфавитом; Γ — алфавит, называемый магазинным алфавитом; $q_0 \in Q$ — начальное состояние; $F \subseteq Q$ — подмножество заключительных состояний; $Z_0 \in \Gamma$ — начальный магазинный символ; δ — система команд, определенная как конечное множество команд, каждая из которых записывается в виде (8.8):

$$qaZ \to r\gamma,$$

где знак „$\to$" (стрелка) — внешний символ (не принадлежащий ни одному из указанных алфавитов); $q,\ r \in Q$; $a \in V \cup \{\lambda\}$; $Z \in \Gamma$; $\gamma \in \Gamma^*$.

Замечание 8.8. Нелишне обратить внимание на то, что упоминание о ленте и магазине в приведенном выше формальном определении совершенно необязательно. Формально для определения МП-автомата достаточно задать два алфавита, конечное множество состояний, выделив в нем начальное состояние и подмножество заключительных состояний, а также систему команд. Образы ленты, магазина, да и сам термин „состояние“, являются метафорами, которые позволяют связать формальное определение с его содержательной интерпретацией, идеей некоего „устройства“ для чтения слов, с описания которого мы начали этот параграф. #

Любую цепочку в алфавите V будем называть входной цепочкой, а сами элементы входного алфавита — входными символами; точно так же любую цепочку в алфавите Γ будем называть магазинной цепочкой, а символы этого алфавита — магазинными символами. Упорядоченную тройку до знака „$\to$“ (стрелки) в записи команды называют ***левой частью команды***, а упорядоченную пару после знака „$\to$“ — ***правой частью команды***.

Пример 8.14. Рассмотрим МП-автомат

$$M_1 = (\{q_0, q_1, q_2\},\ \{0, 1\},\ \{Z, 0\},\ q_0,\ \{q_0\},\ Z, \delta_1)$$

с таким множеством команд δ_1:

$$\begin{aligned}
&q_0 0Z \to q_1\ 0Z,\\
&q_1 00 \to q_1\ 00,\\
&q_1 10 \to q_2 \lambda,\\
&q_2 10 \to q_2 \lambda,\\
&q_2 \lambda Z \to q_0 \lambda.
\end{aligned}$$

На рис. 8.33 проиллюстрировано выполнение первых двух команд. При записи системы команд этого конкретного МП-автомата мы в правых частях первых двух команд отделили магазинную цепочку от состояния пробелом для того, чтобы не возник соблазн прочитать левую и правую части этих команд одинаково. Таким приемом записи мы хотим подчеркнуть, что левая часть команды МП-автомата — упорядоченная тройка, а правая — упорядоченная пара.

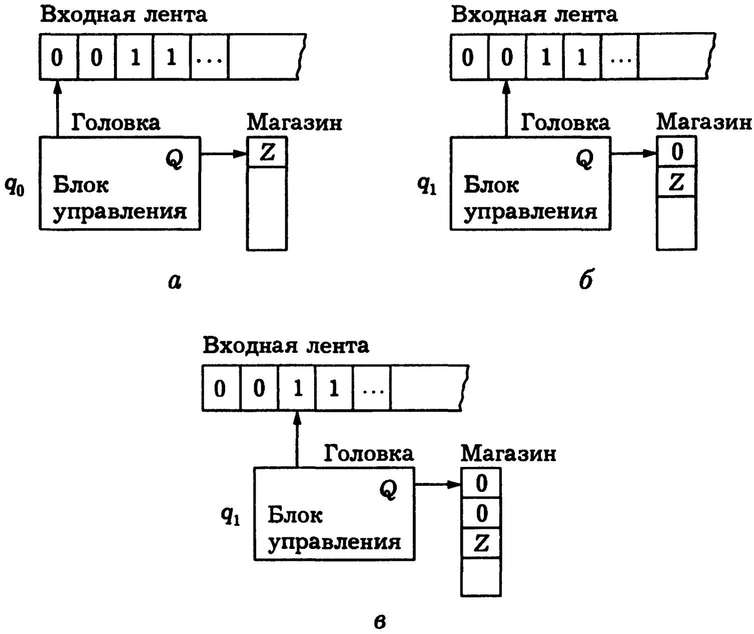

Рис. 8.33

Замечание 8.9. Систему команд МП-автомата мож́но понимать как способ определения некоторой *конечной функции*, которую называют ***функцией переходов МП-автомата*** (по аналогии с *функцией переходов конечного автомата*). Эта функция может быть определена как отображение вида

$$\delta\colon Q \times (V \cup \{\lambda\}) \times \Gamma \to \mathbf{Fin}(Q \times \Gamma^*),$$

где использовано обозначение $\mathbf{Fin}(A)$ для множества всех конечных подмножеств множества A. #

„Динамику“ МП-автомата, т.е. процесс перехода его из одного состояния в другое в соответствии с обозреваемыми символами на ленте и содержимым магазина, удобнее всего описывать в терминах *конфигураций.* Для этого общее понятие конфигурации автомата, которое на интуитивном уровне было введено в **7.5**, необходимо уточнить применительно к МП-автомату и определить также отношение выводимости на множестве конфигураций.

Определение 8.8. ***Конфигурацией МП-автомата*** M называют упорядоченную тройку вида $(q,\ x,\ \beta)$, где $q \in Q$ — состояние, $x \in V^*$ — входная цепочка, β — магазинная цепочка.

Конфигурацию $C' = (r,\ w,\ \gamma\alpha)$ называют ***непосредственно выводимой*** из конфигурации $C = (q,\ aw,\ Z\alpha)$, если в множестве команд МП-автомата есть команда (8.8). Отношение непосредственной выводимости на множестве конфигураций МП-автомата M обозначаем $\vdash_M$, используя запись $C \vdash_M C'$ (и опуская, если это не вредит пониманию, обозначение самого МП-автомата).

Итак, непосредственная выводимость

$$(q,\ aw,\ Z\alpha) \vdash_M (r,\ w,\ \gamma\alpha) \tag{8.9}$$

имеет место тогда и только тогда, когда в системе δ команд МП-автомата M есть команда (8.8), которую в этом случае называют ***применимой к конфигурации*** $(q,\ aw,\ Z\alpha)$.

Таким образом, неформально конфигурация показывает состояние блока управления, непрочитанную часть входной цепочки (первый символ этой цепочки, если она не пуста, является обозреваемым) и содержимое магазина (первый символ магазинной цепочки β, если она не пуста, есть верхний символ магазина). Если вторая компонента конфигурации — пустая цепочка, то это означает, что входная цепочка прочитана полностью. Если же пуста третья компонента, то это означает, что пуст магазин.

На рис. 8.33 изображены (в терминах приведенного выше содержательного описания) две конфигурации МП-автомата из примера 8.14, причем вторая конфигурация непосредственно выводится из первой. Понятие непосредственной выводимости, таким образом, формализует данное выше понятие такта (или шага) работы МП-автомата. Заметим, что если в (8.9) и в команде (8.8) $a = \lambda$, т.е. имеет место λ-такт, то команда (8.8) применима к любой конфигурации, у которой (говоря неформально) состояние блока управления есть q, а верхний символ магазина — Z, причем, подчеркнем это, независимо от содержимого входной ленты (т.е. для любой входной цепочки w). Если же a есть входной символ, то применимость команды к конфигурации определяется состоянием, обозреваемым символом на ленте и верхним символом магазина.

Любую конфигурацию вида $(q_0,\ x,\ Z_0)$ называют ***начальной***, а любую конфигурацию вида $(q_f,\ \lambda,\ \lambda)$, где $q_f \in F$, — ***заключительной***. Обратим внимание на то, что заключительная конфигурация — это конфигурация с пустым магазином. Множество всех заключительных конфигураций находится, таким образом, во взаимно однозначном соответствии с множеством заключительных состояний МП-автомата. Из заключительной конфигурации не может быть выведена ни одна конфигурация, так как к ней не применима ни одна команда. Конфигурацию МП-автомата, не являющуюся заключительной и к которой не применима ни одна команда, называют ***тупиковой***.

Определение 8.9. ***Выводом на множестве конфигураций МП-автомата M*** называют последовательность C_0, $C_1, \ldots, C_n, \ldots$ (конечную или бесконечную) таких его конфигураций, что для любого $i \geqslant 0$ имеет место $C_i \vdash C_{i+1}$, если C_{i+1} существует.

Если вывод конечен и C_n — его последняя конфигурация, то число n называют ***длиной вывода***. В этом случае будем говорить, что вывод $C_0,\ C_1,\ \ldots,\ C_n$ связывает конфигурацию C_0 с конфигурацией C_n.

Конфигурацию C' называют ***выводимой*** из конфигурации C, обозначая это как $C \vdash^* C'$, если существует вывод, связывающий C с C', т.е. вывод $C_0,\ C_1,\ \ldots,\ C_n$, такой, что $C_0 = C$, а $C_n = C'$.

В частности, при $n = 0$ получаем, что любая конфигурация выводится сама из себя; при $n = 1$ получаем непосредственную выводимость C' из C. В общем случае говорят, что конфигурация C' выводится из конфигурации C за n шагов, записывая при этом $C \vdash^n C'$. Желая подчеркнуть, что существует вывод ненулевой длины конфигурации C' из конфигурации C, т.е. $C \vdash^n C'$ и $n > 0$, записывают $C \vdash^+ C'$. #

Таким образом, понятие выводимости для конфигураций МП-автомата вводится аналогично таковому для грамматики. И к тому же сохраняется полная аналогия с понятием *достижимости* в *ориентированных графах*.

В терминах конфигураций может быть дано и строгое определение языка МП-автомата.

Определение 8.10. Входную цепочку x называют допустимой цепочкой МП-автомата M (см. определение 8.7), если на множестве конфигураций M существует вывод, связывающий начальную конфигурацию (q_0, x, Z_0) с заключительной конфигурацией (q_f, λ, λ), где $q_f \in F$, т.е.

$$(q_0,\ x,\ Z_0) \vdash_M^* (q_f,\ \lambda,\ \lambda).$$

Язык, допускаемый МП-автоматом M (или просто язык МП-автомата M), — это множество всех его допустимых цепочек.

МП-автоматы M_1 и M_2 называют ***эквивалентными***, если их языки совпадают, т.е. $L(M_1) = L(M_2)$.

Любой вывод на множестве конфигураций МП-автомата, связывающий начальную конфигурацию $C_0^x = (q_0, x, Z_0)$ с одной из заключительных, называют ***допускающей последовательностью конфигураций для цепочки*** x. Цепочка x принадлежит языку $L(M)$ тогда и только тогда, когда для нее существует допускающая последовательность конфигураций. Тем не менее может оказаться так, что даже в том случае, когда $x \in L(M)$, из начальной конфигурации C_0^x можно вывести отнюдь не только заключительную конфигурацию.

Пример 8.15. Пусть МП-автомат

$$M_2 = (\{q_0, q_1, q_2\},\ \{a, b\},\ \{Z, a, b\},\ q_0, \{q_2\},\ Z,\ \delta_2)$$

определен множеством команд δ_2:

$$\begin{array}{ll}
q_0aZ \to q_0aZ, & (1)\\
q_0bZ \to q_0bZ, & (2)\\
q_0aa \to q_0aa, & (3)\\
q_0aa \to q_1\lambda, & (4)\\
q_0ab \to q_0ab, & (5)\\
q_0ba \to q_0ba, & (6)\\
q_0bb \to q_0bb, & (7)\\
q_0bb \to q_1\lambda, & (8)\\
q_1aa \to q_1\lambda, & (9)\\
q_1bb \to q_1\lambda, & (10)\\
q_1\lambda Z \to q_2\lambda. & (11)
\end{array}$$

Можно доказать, что этот МП-автомат допускает зеркальный язык $\{xx^R \in \{a, b\}^*\}$, где x^R обозначает *инверсию цепочки* x.

Приведем допускающую последовательность конфигураций для цепочки $abba$:

$$(q_0,\ abba,\ Z) \vdash_{(1)} (q_0,\ bba,\ aZ) \vdash_{(6)} (q_0,\ ba,\ baZ) \vdash_{(8)}$$
$$\vdash_{(8)} (q_1,\ a,\ aZ) \vdash_{(9)} (q_1,\ \lambda,\ Z) \vdash_{(11)} (q_2,\ \lambda,\ \lambda)$$

(справа внизу под значками непосредственной выводимости подписаны номера применяемых команд).

К начальной конфигурации $(q_0,\ abba,\ Z)$ применима только команда (1). После выполнения этой команды первый символ входной цепочки будет прочитан, а в магазине вместо одного символа Z окажется цепочка aZ, т.е. символ a станет верхним символом магазина.

К полученной конфигурации $(q_0,\ bba,\ aZ)$ применима только команда (6), в результате выполнения которой будет прочитан еще один символ входной цепочки, т.е. ее второй символ b, и в магазине окажется цепочка baZ.

К третьей конфигурации записанного выше вывода применимы две команды: (7) и (8). Выбирая команду (8), мы тем самым читаем третий символ входной цепочки, а верхний символ магазина, совпавший с указанным символом входной цепочки, „выталкиваем" из магазина, после чего верхним символом магазина окажется уже символ a, т.е. возникнет конфигурация $(q_1,\ a,\ aZ)$. Заметим также, что после выполнения команды (8) МП-автомат M_2 окажется в новом состоянии q_1.

Применив к конфигурации $(q_1,\ a,\ aZ)$ единственную применимую к ней команду (9), получим конфигурацию $(q_1,\ \lambda,\ Z)$, т.е. входная цепочка прочитана полностью, а в магазине остался символ Z.

Применяя опять-таки единственно возможную для такой конфигурацию команду (11), мы убираем символ Z из магазина, тем самым опустошая его, и переходим в заключительное состояние q_2.

В то же время, если бы после третьей конфигурации было выбрано „неудачное продолжение", т.е. была бы применена команда (7), и автомат продолжал бы записывать в магазин вместо того, чтобы сравнивать магазин с лентой, то получился бы вывод

$$(q_0,\ ba,\ baZ) \vdash_{(7)} (q_0,\ a,\ bbaZ) \vdash_{(5)} (q_0,\ \lambda,\ abbaZ).$$

Последняя конфигурация в этом выводе характеризуется тем, что входная цепочка прочитана, магазин не пуст, состояние не является заключительным и не применима ни одна команда, т.е. полученная конфигурация $(q_0,\ \lambda,\ abbaZ)$ МП-автомата M_2 является тупиковой. #

Рассмотренный пример показывает, что МП-автомат может попасть в тупиковую конфигурацию, читая даже „правильную" цепочку, т.е. цепочку, принадлежащую его языку. Это связано с недетерминированностью МП-автомата, т.е. с тем, что из данной конфигурации можно непосредственно вывести, вообще говоря, более чем одну конфигурацию*. Тот факт, что входная цепочка принадлежит языку МП-автомата, означает лишь, что существует допускающая последовательность конфигураций для этой цепочки, но не всякая последовательность конфигураций, которая возникает при чтении этой цепочки, будет допускающей. Задача эффективного поиска допускающей последовательности конфигураций — одна из основных задач разработки алгоритмов синтаксического анализа на базе МП-автоматов.

Как уже отмечалось, основное свойство языков, допускаемых МП-автоматами, состоит в том, что эти языки образуют класс, совпадающий с классом КС-языков. Но перед тем как доказывать этот основной результат теории МП-автоматов, аналогичный теореме Клини для конечных автоматов, установим одно весьма полезное свойство МП-автомата, состоящее в том, что „хвост" входной цепочки и „дно" магазина не влияют на работу МП-автомата с началом входной цепочки и верхом магазина, так как входная цепочка читается строго символ за символом без возвратов и забеганий вперед, а магазин обновляется только сверху.

Приведем строгую формулировку.

Теорема 8.5. Пусть в МП-автомате

$$M = (Q,\ V,\ \Gamma,\ q_0,\ F,\ Z_0,\ \delta)$$

на множестве конфигураций для некоторых $q, r \in Q$, $x, y \in V^*$ и $\alpha, \beta \in \Gamma^*$ имеет место выводимость

$$(q,\ xy,\ \alpha) \vdash_M^* (r,\ y,\ \beta). \tag{8.10}$$

*МП-автомат называют ***детерминированным***, если из каждой его конфигурации непосредственно выводится не более чем одна конфигурация. Можно доказать, что, в отличие от конечного автомата, для МП-автомата невозможна процедура детерминизации и класс языков, допускаемых детерминированными МП-автоматами, строго включается в класс языков, допускаемых произвольными МП-автоматами.

Тогда для произвольных $z \in V^*$, $\gamma \in \Gamma^*$ имеет место выводимость

$$(q,\ xyz,\ \alpha\gamma) \vdash_M^* (r,\ yz,\ \beta\gamma). \tag{8.11}$$

Обратно, если для некоторых q, $r \in Q$, x, y, $z \in V^*$ и α, β, $\gamma \in \Gamma^*$ имеет место выводимость (8.11), то выполняется и выводимость (8.10).

◀ Индукцией по длине вывода, связывающего конфигурацию $(q,\ xy,\ \alpha)$ с конфигурацией $(r,\ y,\ \beta)$ в (8.10) докажем, что если эта выводимость имеет место, то имеет место и выводимость (8.11).

Для вывода нулевой длины утверждение тривиально.

Пусть доказываемое верно для любой длины вывода, связывающего конфигурации в (8.10), не превышающей $n-1$.

Предположим, что

$$(q,\ xy,\ \alpha) \vdash_M^n (r,\ y,\ \beta).$$

Рассмотрим первый шаг соответствующего вывода. Пусть он имеет вид

$$(q,\ ax'y,\ Z\alpha') \vdash_M (q',\ x'y,\ \sigma\alpha'),$$

где $x = ax'$, $\alpha = Z\alpha'$, $a \in V \cup \{\lambda\}$ и $Z \in \Gamma$.

Это значит, что на этом шаге применена команда

$$qaZ \to q'\sigma. \tag{8.12}$$

Согласно предположению индукции, для произвольных $z \in V^*$ и $\gamma \in \Gamma^*$ имеет место выводимость

$$(q',\ x'yz,\ \sigma\alpha'\gamma) \vdash^{n-1} (r,\ yz,\ \beta\gamma). \tag{8.13}$$

Рассмотрим конфигурацию $(q,\ xyz,\ \alpha\gamma) = (q,\ ax'yz,\ Z\alpha'\gamma)$ и применим к ней команду (8.12). Получим

$$(q,\ ax'yz,\ Z\alpha'\gamma) \vdash_M (q',\ x'yz,\ \sigma\alpha'\gamma).$$

Отсюда в силу (8.13), будем иметь

$$(q,\ ax'yz,\ Z\alpha'\gamma) \vdash (q',\ x'yz,\ \sigma\alpha'\gamma) \vdash_M^{n-1} (r,\ yz,\ \beta\gamma),$$

т.е. при выводимости (8.10) за n шагов имеет место и выводимость (8.11) также за n шагов, что и требовалось доказать.

Аналогично (но уже индукцией по длине вывода (8.11)) доказывается, что если имеет место выводимость (8.11), то имеет место и выводимость (8.10). ►

МП-автомат M называют ***эквивалентным КС-грамматике*** G, если язык, допускаемый M, совпадает с ***языком, порождаемым грамматикой*** G, т.е. если $L(M) = L(G)$.

Опишем алгоритм построения по заданной КС-грамматике эквивалентного ей МП-автомата, а также алгоритм построения по заданному МП-автомату КС-грамматики, которой он эквивалентен.

Алгоритм построения МП-автомата по КС-грамматике. Для исходной КС-грамматики $G = (V,\ N,\ S,\ P)$ определим МП-автомат $M_G = (\{q\},\ V,\ V \cup N,\ q,\ \{q\},\ S,\ \delta)$ (с единственным состоянием q), система команд δ которого строится следующим образом: для каждого правила $A \to \alpha$, принадлежащего P, в δ записывается команда $q\lambda A \to q\alpha$ и для каждого $a \in V$ — команда $qaa \to q\lambda$. Никаких других команд в системе команд δ нет.

Пример 8.16. Пусть дана грамматика

$$G = (\{a, b, c\},\ \{S\},\ S,\ P),$$

множество правил вывода P которой есть

$$S \to aSbS \,|\, aS \,|\, c.$$

Тогда эквивалентный ей МП-автомат задается следующей системой команд:

$$
\begin{aligned}
&q\lambda S \to qaSbS \,|\, qaS \,|\, qc,\\
&qaa \to q\lambda,\\
&qbb \to q\lambda,\\
&qcc \to q\lambda
\end{aligned}
$$

(мы использовали сокращенную запись нескольких команд с одинаковыми левыми частями по аналогии с тем, как мы это делаем при записи множества правил вывода грамматики).

В системе команд МП-автомата, который строится по заданной КС-грамматике так, как это описано выше, мы видим два „сорта" команд: 1) команды, „моделирующие" применение правил, исходной грамматики (в данном примере это команды первой строки); при выполнении каждой такой команды МП-автомат делает λ-такт, т.е. не продвигает головку по входной ленте; 2) команды „сравнения", согласно которым МП-автомат должен убирать верхний символ магазина всякий раз, когда он совпадает с обозреваемым символом на ленте.

Тогда, читая входную цепочку, такой МП-автомат „моделирует" левый вывод этой цепочки в исходной грамматике, применяя каждый раз, когда верхним символом магазина оказывается нетерминал, команду „первого сорта" и всякий раз, когда верхним символом магазина оказывается терминал исходной грамматики, „команду сравнения".

Прочитаем цепочку $aacbc$:

$$
\begin{aligned}
&(q,\, aacbc,\, S) \vdash (q,\, aacbc,\, aSbS) \vdash (q,\, acbc,\, SbS) \vdash\\
&\quad \vdash (q,\, acbc,\, aSbS) \vdash (q,\, cbc,\, SbS) \vdash (q,\, cbc,\, cbS) \vdash\\
&\qquad \vdash (q,\, bc,\, bS) \vdash (q,\, c,\, S) \vdash (q,\, c,\, c) \vdash (q,\, \lambda,\, \lambda).
\end{aligned}
$$

Нетрудно видеть, что этот вывод „моделирует" левый вывод в грамматике:

$$S \vdash aSbS \vdash aaSbS \vdash aacbS \vdash aacbc,$$

т.е. МП-автомат, строя допускающую последовательность конфигураций для входной цепочки, на каждом шаге, когда верхний символ магазина не совпадает с очередным символом входной цепочки, должен „угадать“ применяемое правило (выбрать в данном примере одну из трех первых команд). Неправильный выбор приведет к тупику, например:

$$(q,\ aacbc,\ S) \vdash (q,\ aacbc,\ aS) \vdash (q,\ acbc,\ S) \vdash (q,\ acbc, c). \quad \#$$

Докажем, что рассмотренный алгоритм дает МП-автомат, эквивалентный исходной КС-грамматике.

Теорема 8.6. МП-автомат M_G эквивалентен КС-грамматике G.

◀ Индукцией по длине n вывода *терминальной цепочки* x из *нетерминала* A докажем, что если $A \vdash_G^* x$, то

$$(q,\ x,\ A) \vdash_{M_G}^* (q,\ \lambda,\ \lambda).$$

Пусть $n = 1$, т.е. $A \vdash^1 x$; тогда в P есть правило $A \to x$ и, следовательно, в δ есть команда $q\lambda A \to qx$. В таком случае при $x = \lambda$ имеет место выводимость

$$(q,\ \lambda,\ A) \vdash (q,\ \lambda,\ \lambda),$$

и требуемый вывод на множестве конфигураций МП-автомата M_G построен. Если же цепочка x непустая, то тогда, расписывая ее по символам, т.е. полагая $x = x(1)\ldots x(k)$, $k \geqslant 1$, получим

$$(q,\ x(1)\ldots x(k),\ A) \vdash (q,\ x(1)\ldots x(k),\ x(1)\ldots x(k)).$$

Из последней конфигурации за k шагов посредством применения команд вида $qaa \to q\lambda$, где $a \in V$, выводится конфигурация $(q,\ \lambda,\ \lambda)$, и, таким образом, $(q,\ x,\ A) \vdash^{|x|+1} (q,\ \lambda,\ \lambda)$.

Пусть теперь доказываемое верно для любого n, не превосходящего некоторого $m - 1$ для $m \geqslant 2$, пусть $A \vdash^m x$ и первый шаг вывода длины m цепочки x из нетерминала A имеет вид $A \vdash X_1X_2\ldots X_k$, где $k \geqslant 1$ и для каждого $i = \overline{1, k}$ символ X_i есть символ объединенного алфавита грамматики*. Далее, из цепочки $X_1X_2\ldots X_k$ должна быть выведена терминальная цепочка x. Это значит, что для каждого $i = \overline{1, k}$ из символа X_i выводится какая-то подцепочка цепочки x (в частности, если этот символ является терминалом, он будет одним из символов цепочки x). Таким образом, для каждого $i = \overline{1, k}$ выполняется $X_i \vdash^{m_i} x_i$ ($m_i \geqslant 0$), и $x = x_1x_2\ldots x_k$.

Для $X_i \in N$ длина вывода m_i подцепочки x_i не может превышать $m - 1$. Следовательно, согласно предположению индукции, имеем

$$(q,\ x_i,\ X_i) \vdash^* (q,\ \lambda,\ \lambda),$$

а для $X_i \in V$ ($m_i = 0$ и, следовательно, $X_i = x_i$), согласно построению МП-автомата M_G,

$$(q,\ x_i,\ x_i) \vdash^* (q,\ \lambda,\ \lambda).$$

Тогда в силу теоремы 8.5

$$(q,\ x_1x_2\ldots x_k,\ X_1X_2\ldots X_k) \vdash^* (q,\ x_2\ldots x_k,\ X_2\ldots X_k) \vdash^* \\ \vdash^* (q,\ x_k,\ X_k) \vdash (q,\ \lambda,\ \lambda).$$

Кроме того, так как $A \vdash X_1X_2\ldots X_k$, то в P есть правило вывода $A \to X_1X_2\ldots X_k$, откуда, согласно построению МП-автомата M_G, в δ есть команда

$$q\lambda A \to qX_1X_2\ldots X_k$$

и

$$(q,\ x,\ A) \vdash (q,\ x,\ X_1X_2\ldots X_k),$$

*Цепочка $X_1X_2\ldots X_k$ не может быть пустой, так как тогда она окажется последней цепочкой рассматриваемого вывода и его длина будет равна 1, а мы предположили, что его длина равна $m > 1$.

а окончательно

$$(q,\ x,\ A) \vdash^* (q,\ \lambda,\ \lambda).$$

Следовательно, если $x \in L(G)$, то $S \vdash^* x$ и $(q,\ x,\ S) \vdash^* (q,\ \lambda,\ \lambda)$, т.е. $x \in L(M_G)$.

Итак, мы доказали, что $L(G) \subseteq L(M_G)$.

Чтобы доказать обратное включение, сначала индукцией по длине n вывода на множестве конфигураций МП-автомата M_G докажем, что

$$(q,\ x,\ A) \vdash^*_{M_G} (q,\ \lambda,\ \lambda)$$

влечет $A \vdash^*_G x$ (для произвольных $A \in N$ и $x \in V^*$).

Пусть $n = 1$, т.е.

$$(q,\ x,\ A) \vdash (q,\ \lambda,\ \lambda).$$

Согласно построению МП-автомата M_G, это может быть тогда и только тогда, когда $x \in V$ и $A = x$, т.е. $x \vdash^* x$.

Пусть доказываемое верно для любого $n \leqslant m - 1$, где $m \leqslant 2$, и

$$(q,\ x,\ A) \vdash^m (q,\ \lambda,\ \lambda), \tag{8.14}$$

причем первый шаг соответствующего вывода имеет вид

$$(q,\ x,\ A) \vdash (q,\ x,\ X_1X_2\ldots X_k). \tag{8.15}$$

Это значит, что в системе команд δ есть команда $q\lambda A \to$ $\to qX_1X_2X_k$ и, следовательно, правило в множестве правил вывода P грамматики G есть правило $A \to X_1X_2\ldots X_k$, по которому указанная команда построена. Из (8.14) и (8.15) следует, что имеет место выводимость

$$(q,\ x,\ X_1X_2\ldots X_k) \vdash^* (q,\ \lambda,\ \lambda). \tag{8.16}$$

Используя индукцию по длине магазинной цепочки $X_1X_2\ldots X_k$, можно доказать, что из (8.16) следует существование таких входных цепочек $x_1, x_2, \ldots, x_k$, что $x = x_1x_2\ldots x_k$ и имеет место выводимость

$$
\begin{aligned}
(q,\, x_1x_2\ldots x_k,\, X_1X_2\ldots X_k) \vdash^{m_1}&\\
\vdash^{m_1} (q,\, x_2\ldots x_k,\, X_2\ldots X_k) \vdash^{m_2}&\\
\vdash^{m_2} (q, x_3\ldots x_k, X_3\ldots X_k) \vdash^{m_3} \ldots \vdash^{m_{k-1}}&\\
\vdash^{m_{k-1}} (q, x_k, X_k) \vdash^{m_k} (q, \lambda, \lambda)&
\end{aligned} \qquad (8.17)
$$

для некоторых m_i, таких, что $1 \leqslant m_i < m-1$, $i = \overline{1,k}$.

Вывод (8.17) можно прокомментировать так: чтобы достичь заключительной конфигурации, M_G должен выбросить все символы $X_1, X_2, \ldots, X_k$ из магазина. Предположим, что к тому моменту, когда X_1 покинет магазин (чтобы уже больше туда не вернуться!), будет прочитано некоторое *начало входной цепочки* x — цепочка x_1, когда X_2 покинет магазин, будет прочитана следующая подцепочка x_2 и так до тех, пока наконец, автомат не прочитает цепочку x_k — *конец цепочки* x.

Из теоремы 8.5 следует, что существуют также выводы

$$
\begin{aligned}
&(q,\, x_1,\, X_1) \vdash^{m_1} q; \lambda; \lambda,\\
&(q,\, x_2,\, X_2) \vdash^{m_2} (q,\, \lambda, \lambda),\\
&\ldots\ldots\ldots\ldots\ldots\\
&(q,\, x_k,\, X_k) \vdash^{m_k} (q,\, \lambda,\, \lambda).
\end{aligned} \qquad (8.18)
$$

Все числа m_i при $i = \overline{1,k}$ не больше $m-1$. Тогда согласно предположению индукции, из (8.18) следует, что для каждого $i = \overline{1,k}$ в грамматике G имеет место $X_i \vdash_G^* x_i$, и в силу того, что в P есть правило вывода $A \to X_1X_2\ldots X_k$ (см. (8.15)),

$$
A \vdash X_1X_2\ldots X_k \vdash^* x_1x_2\ldots x_k = x,
$$

т.е. $A \vdash_G^* x$, что и требовалось доказать.

Отсюда, если $x \in L(M_G)$, то $x \in L(G)$, т.е. $L(M_G) \subseteq L(G)$, и в силу доказанного выше языки $L(G)$ и $L(M_G)$ совпадают. ▶

Алгоритм построения КС-грамматики по МП-автомату. Дадим сначала неформальную мотивировку той конструкции, которая будет приведена ниже. Будем рассматривать МП-автомат как „игрока“, ставящего цели следующего вида: „находясь в состоянии q и имея верхний символ магазина Z, перейти в состояние s“. Условимся записывать такую цель в виде тройки $[qZs]$. Как может наш „игрок“ достичь поставленной цели? Если в множестве команд автомата („правил игры“) есть команда

$$qaZ \to rX_1X_2\ldots X_k,$$

где для каждого i X_i — магазинный символ ($X_i \in \Gamma$), то после выполнения этой команды МП-автомат перейдет в состояние r.

Если $r = s$, то цель достигнута, иначе ставим цель $[rX_1s_1]$, а достигнув ее, ставим цель $[s_1X_2s_2]$ и т.д. Достигнув цели $[s_{k-1}X_ks]$, „игрок“ достигает и цели $[qZs]$. Так как рассматривается допуск с пустым магазином, то символы X_i должны по очереди покинуть магазин, и только в этом случае может быть достигнута „глобальная“ цель МП-автомата: $[q_0Z_0q_f]$, $q_f \in F$ („находясь в начальном состоянии и имея верхним символом магазина начальный магазинный символ, попасть в одно из заключительных состояний, опустошив магазин“). Так как, вообще говоря, „игрок“ не знает наперед последовательности состояний s_1, s_2, ..., s_{k-1}, ведущих к цели, он должен перебрать все такие последовательности.

Эти неформальные соображения лежат в основе следующей конструкции.

Пусть дан МП-автомат

$$M = (Q,\ V,\ \Gamma,\ q_0,\ F,\ Z_0,\ \delta).$$

Определим КС-грамматику $G_M = (V, N, S, P)$, терминальный алфавит которой совпадает со входным алфавитом МП-автомата M, следующим образом.

Нетерминальный алфавит N грамматики есть множество, находящееся во взаимно однозначном соответствии с множеством всех упорядоченных троек вида (q, Z, s), где $q, s \in Q$, $Z \in \Gamma$, пополненное символом S, не принадлежащим ни одному из множеств Q, V, Γ и объявляемым аксиомой грамматики.

Упорядоченные тройки указанного вида записывают обычно как $[qZs]$, интуитивно понимая каждую такую тройку как охарактеризованную выше цель.

Таким образом, $N = \{[qZs]\colon q, s \in Q$ и $Z \in \Gamma\} \cup \{S\}$.

Множество правил вывода P грамматики G_M строится так:

а) если команда $qaZ \to rX_1X_2\ldots X_k$, $k \geqslant 1$, принадлежит системе команд δ, то в P записываются все правила вида

$$[qZs_k] \to a[rX_1s_1][s_1X_2s_2]\ldots[s_{k-1}X_ks_k]$$

для любой последовательности k состояний $s_1, \ldots, s_k$ множества Q (тем самым к P добавляется $|Q|^k$ правил на каждую команду указанного вида);

б) для каждой команды $qaZ \to r\lambda$ в δ в P добавляется правило $[qZr] \to a$;

в) для каждого $q_f \in F$ в P вводится правило $S \to [q_0Z_0q_f]$;

г) никаких других правил в P, кроме определенных пп. а–в, нет.

Пример 8.17. Для МП-автомата

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

с множеством команд δ, имеющих вид

$$q_0aZ \to q_1aZ,$$
$$q_1aa \to q_1aa,$$
$$q_1ba \to q_2\lambda,$$
$$q_2ba \to q_2\lambda,$$
$$q_2\lambda Z \to q_0\lambda,$$
$$q_0\lambda Z \to q_0\lambda,$$

построим эквивалентную ему КС-грамматику. Можно доказать, что этот МП-автомат допускает язык $\{a^nb^n\colon n \geqslant 0\}$. Заметим, что МП-автомат M допускает пустую цепочку, применяя к начальной конфигурации (q_0, λ, Z) последнюю команду. Построенная по данной системе команд грамматика будет иметь следующий вид:

$$
\begin{aligned}
&S \to [q_0 Z q_0],\\
&[q_0 Z s_2] \to a[q_1 a s_1][s_1 Z s_2],\ s_1, s_2 \in \{q_0, q_1, q_2\},\\
&[q_1 a s_2] \to a[q_1 a s_1][s_1 a s_2],\ s_1, s_2 \in \{q_0, q_1, q_2\},\\
&[q_1 a q_2] \to b,\\
&[q_2 a q_2] \to b,\\
&[q_2 Z q_0] \to \lambda,\\
&[q_0 Z q_0] \to \lambda.
\end{aligned}
$$

Подчеркнем, что во второй и третьей строчках имеется не по одному, а по $3^2 = 9$ правил (число всех последовательностей двух состояний из трехэлементного множества состояний). Выведем в этой грамматике цепочку *aabb*:

$$
\begin{aligned}
S \vdash [q_0 Z q_0] \vdash a[q_1 a q_2][q_2 Z q_0] \vdash aa[q_1 a q_2][q_2 a q_2][q_2 Z q_0] \vdash&\\
\vdash aab[q_2 a q_2][q_2 Z q_0] \vdash aabb[q_2 Z q_0] \vdash aabb.&
\end{aligned}
$$

На втором шаге этого вывода мы применяем то правило вывода из девяти правил второй строки, которое получается при подстановке вместо s_1 состояния q_2, а вместо s_2 состояния q_0. Мы „угадываем“ эти состояния, зная (по системе команд исходного МП-автомата), что достичь заключительного состояния по прочтении (непустой) входной цепочки наш МП-автомат может только из состояния q_2. В то же время, если бы мы на втором шаге использовали правило второй строки, получающееся подстановкой $s_1 = q_1$, $s_2 = q_2$, возник бы *бесполезный нетерминал* $[q_1 a q_1]$ и наш вывод зашел бы в тупик.

Аналогично на третьем шаге используется то правило из девяти правил третьей строки, в котором $s_1 = s_2 = q_2$. После этого применяем по очереди правила четвертой, пятой и шестой строк, завершая вывод.

Если мы теперь в построенном выводе „считаем“ по шагам магазинные символы, заключенные в квадратных скобках между состояниями, то получим Z, aZ, aaZ, aZ, Z, т.е. получим изменение содержимого магазина (не считая последнего шага, когда происходит его окончательное опустошение), представленное в следующем выводе на множестве конфигураций МП-автомата M:

$$(q_0,\, aabb,\, Z) \vdash (q_1,\, abb,\, aZ) \vdash (q_1,\, bb,\, aaZ) \vdash \\ \vdash (q_2,\, b,\, aZ) \vdash (q_2,\, \lambda,\, Z) \vdash (q_0,\, \lambda,\, \lambda).$$

Этот вывод есть не что иное, как допускающая последовательность конфигураций для цепочки $aabb$.

Читая же последовательности состояний в квадратных скобках, мы получим в итоге ту последовательность состояний, которую проходит МП-автомат, допуская написанную выше цепочку.

Действительно, после первого шага вывода в грамматике получим последовательность q_0, q_0, что можно интерпретировать так: „из состояния q_0 перейти (вернуться) в это же состояние q_0, прочитав входную цепочку".

После второго шага будем иметь q_0, q_1, q_2, q_0, что означает: „чтобы вернуться в q_0, сначала нужно попасть в q_2 через q_1").

После третьего шага получим q_0, q_1, q_1, q_2, q_2, q_0. Это и есть результирующая последовательность состояний, так как все следующие шаги МП-автомата связаны с „выталкиванием" символов из магазина и не приводят к возникновению новых целей. #

Как правило, грамматика, которая указанным выше способом строится по МП-автомату, оказывается очень громоздкой, содержит много бесполезных и *недостижимых символов*. Это связано с тем, что в ней фигурируют произвольные последовательности состояний МП-автомата фиксированной длины. Что касается разобранного примера 8.17, то в этом случае легко написать грамматику для языка $\{a^n b^n \colon n \geqslant 0\}$ непосредственно:

$$S \to aSb \,|\, ab \,|\, \lambda.$$

По этой грамматике можно построить МП-автомат, используя алгоритм из первой части доказательства основной теоремы, значительно более простой, чем исходный. Его система команд будет иметь следующий вид:

$$
\begin{aligned}
&qaS \to qSb \,|\, qb,\\
&qaa \to q\lambda,\\
&qbb \to q\lambda,\\
&q\lambda S \to q\lambda.
\end{aligned}
$$

В общем же случае проблема распознавания эквивалентности двух МП-автоматов (в отличие от такой же проблемы для конечных автоматов) не разрешима, и не существует общего алгоритма „упрощения“ (в определенном смысле, „минимизации“) МП-автомата, хотя, как мы только что видели, в конкретных случаях это вполне возможно.

Теорема 8.7. МП-автомат M эквивалентен грамматике G_M.

◀ Индукцией по длине n вывода в M докажем, что $Z \in \Gamma$, $(q,\ x,\ Z) \vdash^n (r,\ \lambda,\ \lambda)$ для любых $q,\ r \in Q$, $x \in V^*$ влечет $[qZr] \vdash^* x$ в G_M.

Если $n = 1$, т.е. $(q,\ x,\ Z) \vdash (r,\ \lambda,\ \lambda)$, то $x \in V \cup \{\lambda\}$, и в δ есть команда $qZx \to r\lambda$, откуда в P есть правило $[qZr] \to x$, и $[qZr] \vdash^* x$.

Пусть доказываемое верно для каждого $n \leqslant m-1$, где $m > 1$, и пусть $(q,\ x,\ Z) \vdash^m (r,\ \lambda,\ \lambda)$, причем первый шаг соответствующего вывода имеет вид

$$(q,\ x,\ Z) \vdash (p,\ y,\ X_1X_2\ldots X_k),$$

где $x = ay$ для некоторого $a \in V \cup \{\lambda\}$.

Аналогично доказательству теоремы 8.6 (см. (8.17)) доказывается, что тогда найдутся такие цепочки $x_1x_2\ldots x_k$ и такая последовательность состояний $s_1, \ldots, s_{k-1}$, что $y = x_1x_2\ldots x_k$ и

$$(p,\ x_1x_2\ldots x_k,\ X_1X_2\ldots X_k) \vdash^{m_1} (s_1,\ x_2\ldots x_k,\ X_2\ldots X_k) \vdash^{m_2} \\ \vdash^{m_2} \ldots \vdash^{m_{k-1}} (s_{k-1},\ x_k,\ X_k) \vdash^{m_k} (r,\ \lambda,\ \lambda),$$

где для любого $i = \overline{1, k}$ выполняется $0 \leqslant m_i \leqslant m - 1$. Поэтому в силу теоремы 8.5 для любого $i = \overline{1, k-1}$

$$(s_{i-1},\ x_i,\ X_i) \vdash^{m_i} (s_i,\ \lambda,\ \lambda),$$

где $s_0 = p$, а $s_k = r$, и, согласно предположению индукции,

$$[s_{i-1}X_is_i] \vdash^* x_i.$$

Следовательно, согласно построению грамматики G_M, имеет место выводимость

$$[qZr] \vdash_{G_M} a[pX_is_i][s_iX_2s_2]\ldots[s_{k-1}X_kr] \vdash^*_{G_M} \\ \vdash^*_{G_M} ax_1x_2\ldots x_k = ay = x,$$

что и требовалось доказать.

Пусть цепочка x допускается МП-автоматом M. Тогда

$$(q_0,\ x,\ Z_0) \vdash^*_M (q_f,\ \lambda,\ \lambda),$$

где $q_f \in F$ — одно из заключительных состояний МП-автомата M. Согласно только что доказанному, в этом случае для грамматики G_M выполняется $[q_0Z_0q_f] \vdash^*_{G_M} x$. Но так как в множестве правил вывода грамматики G_M есть правило $S \to$ $\to [q_0Z_0q_f]$, то мы получим $S \vdash_{G_M} [q_0Z_0q_f] \vdash^*_{G_M} x$, т.е. $x \in L(G_M)$. Итак, $L(M) \subseteq L(G_M)$.

Для доказательства обратного включения докажем сначала, что $[qZr] \vdash^*_{G_M} x$ влечет $(q,\ x,\ Z) \vdash^*_M (r,\ \lambda,\ \lambda)$ для любых $q, r \in Q$, $x \in V^*$ и $Z \in \Gamma$. Снова проведем индукцию по длине вывода (в грамматике G_M). При $[qZr] \vdash x$ получаем правило $[qZr] \to x$ в P и, следовательно, команду $qxZ \to r\lambda$ в δ, т.е. $(q,\ x,\ Z) \vdash (r,\ \lambda,\ \lambda)$.

Если же $[qZr] \vdash^m x$ для некоторого $m \geqslant 1$, то $x = ay$ для некоторого $a \in V \cup \{\lambda\}$ и

$$[qZr] \vdash a[pX_1s_1][s_1X_2s_2]\ldots[s_{k-1}X_kr],$$

причем для всех $i = \overline{1,k}$

$$[s_{i-1}X_is_i] \vdash^{m_i} x_i,$$

где $s_0 = p$, $s_k = r$ и $1 \leqslant m_i \leqslant m-1$, так что $y = x_1x_2\ldots x_k$.

Согласно предположению индукции, тогда для каждого такого i

$$(s_{i-1}, x_i, X_i) \vdash^* (s_i, \lambda, \lambda).$$

Но так как указанный выше первый шаг вывода в грамматике возможен только при наличии команды в МП-автомате $qaZ \to \to pX_1X_2\ldots X_k$, то

$$\begin{aligned}(q,\ x,\ Z) \vdash (p,\ y,\ X_1X_2\ldots X_k) \vdash^* \\ \vdash^* (s_1,\ x_2\ldots x_k,\ X_2\ldots X_k) \vdash^* \\ \vdash^* (s_{k-1},\ x_k,\ X_k) \vdash^* (r,\ \lambda,\ \lambda).\end{aligned}$$

Если теперь цепочка x порождается грамматикой G, т.е. $S \vdash^*_{G_M} x$, то первый шаг вывода x из S, согласно определению системы правил грамматики G_M, будет иметь вид $S \vdash [q_0Z_0q_f]$ для некоторого $q_f \in F$, и, следовательно, $[q_0Z_0q_f] \vdash^*_{G_M} x$. Тогда в силу только что доказанного $(q_0,\ x,\ Z_0) \vdash^*_M (q_f,\ \lambda,\ \lambda)$, т.е. $x \in \in L(M)$. Итак, $L(G_M) \subseteq L(M)$, а поскольку обратное включение уже доказано, то $L(M) = L(G_M)$ ▶

Из доказанных теорем 8.6 и 8.7 получаем следующую теорему.

Теорема 8.8. *Язык является контекстно-свободным тогда и только тогда, когда он допускается некоторым МП-автоматом.*

Замечание 8.10. Существует одна полезная модификация построения МП-автомата по КС-грамматике. Вернемся к примеру 8.16. Можно заметить, что в этом примере правая часть каждого правила грамматики начинается некоторым терминалом. Учет этой особенности позволяет найти другой МП-автомат, который, как нетрудно показать, тоже эквивалентен данной грамматике. Система команд этого автомата имеет следующий вид:

$$
\begin{aligned}
&qaS \to qSbS \,|\, qS,\\
&qcS \to q\lambda,\\
&qaa \to q\lambda,\\
&qbb \to q\lambda,\\
&qcc \to q\lambda.
\end{aligned}
$$

Его преимущество в том, что он „видит“ первый непрочитанный символ входной цепочки и, следовательно, имеет меньше альтернатив при выборе команды: например, если очередной символ есть b, то ни одна из команд первых двух строк не может быть применена. Тогда этот автомат имеет меньше возможностей попасть в тупик.

В общем случае, если правая часть любого правила грамматики имеет вид $a\xi$, где $a \in V$, МП-автомат, эквивалентный данной грамматике, определяется командами вида

$$
\begin{aligned}
&qaA \to q\xi,\\
&qbb \to q\lambda, \quad b \in V,
\end{aligned}
$$

(первая — для правила $A \to a\xi$). В такой модификации МП-автомат записывает в магазин „хвост“ правой части правила, следующей за первым терминалом.

Можно доказать, что любая КС-грамматика может быть определена правилами такого вида (так называемая нормальная форма Грейбах*).

*См.: *Ахо А., Ульман Дж.*