Лекція №9

КВ-граматики та автомати з магазинною (стековою) пам’яттю (МП-автомати)

Визначення МП-автомата.

МП-автомат включає:

1) вхідний рядок, який містить символи вхідного алфавіту;
2) читаючу голівку, яка рухається зліва направо вздовж вхідного рядка;
3) пристрій управління станами;
4) стек (магазин) зі своїм алфавітом.

Формально МП-автомат M описується наступними сімома поняттями:

$M = (Q, T, H, d, q_0, z_0, F)$, де

Q – множина станів,

T – множина вхідних символів,

H – множина символів магазину,

d – функція переходів, яка виконує відображення множини пар $Q\times(T\cup\{\varepsilon\})$ в множину пар $Q\times H$,

q_0 – початковий стан, $q_0 \in Q$,

z_0 – граничний маркер або початковий символ магазина, $z_0 \in H$,

F – множина заключних станів, $F \in Q$.

Функція переходів d описується так:

$d(q, a, z_0) = \{(p_1, h_1), (p_2, h_2)\ldots\}$, де

$q, p_1, p_2, \ldots$ – стани автомата;

$z_0, h_1, h_2, \ldots$ – значення у верхівці магазина;

a – вхідний символ.

Конфігурація і такт роботи МП-автомата

Конфігурація МП-автомата задається трійкою (q, w, z), де

q – поточний стан;

w – непрочитана частина вхідного рядка;

z – ланцюжок символів магазину автомата.

Такт роботи МП-автомата визначається переходом автомата від однієї конфігурації до іншої. Такт записується таким чином:

$(q, aw, hz_1) \vdash (p, w, z_2)$, якщо функція переходу $d(q, a, h)$ містить пару (p, z_2), де

$q, p \in Q$;
$a \in T \cup \{\varepsilon\}$;
$w \in T^*$;
$h \in H$;
$z_1, z_2 \in H^*$.

Позначення:

$\vdash$ – один такт;

$\vdash^{\pm}$ – один або більше тактів;

$\vdash^{*}$ – нуль або більше тактів.

Початкова конфігурація задається так:
(q_0, w, z_0), де
q_0 – початковий стан;
w – непрочитаний рядок;
z_0 – початковий символ магазину.

МП-автомат завершує роботу при порожньому магазині в заключній (кінцевій) конфігурації. При завершенні роботи МП-автомату заключні конфігурації можуть бути коректними (рядок розпізнано) і помилковими (рядок не розпізнано).

Коректні заключні конфігурації:

1. (q, ε, z_0)
2. $(q, \varepsilon, \varepsilon)$

Помилкові заключні конфігурації:

1. (q, w, ε) і немає петлі на кінцевому стані q для продовження розбору.
2. (q, ε, z), тобто у стеку (магазині) ще є хоча б один символ, крім z_0 (інколи така заключна конфігурація є коректною для деяких автоматів).
3. (p, ε, z)
4. (p, ε, z_0)
5. $(p, \varepsilon, \varepsilon)$,

де q – один з кінцевих станів;
p – довільний не кінцевий стан.

МП-автомат M розпізнає мову L, якщо:

$$L(M) = \{w \mid (q_0, w, z_0) \vdash^{*} (q, \varepsilon, \varepsilon)\}$$

або

$$L(M) = \{w \mid (q_0, w, z_0) \vdash^{*} (q, \varepsilon, z_0)\},$$

де q – кінцевий стан.

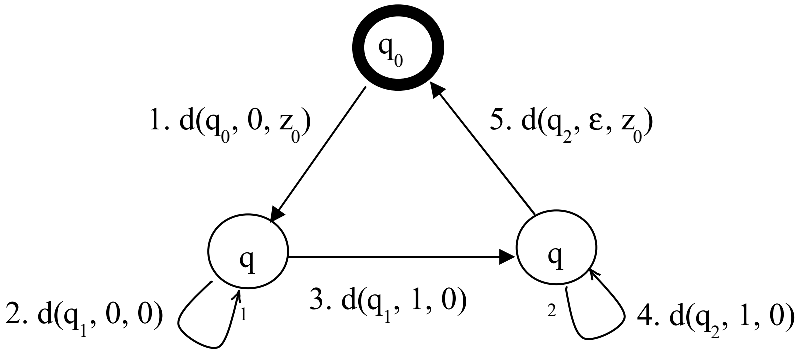
Приклад 1.

Розглянемо МП-автомат, який розпізнає мову $L=\{0^n1^n \mid n \geq 0\}$.

$M=(\{q_0, q_1, q_2\}, \{0, 1\}, \{z_0, 0\}, d, q_0, z_0, \{q_0\})$,

де функція переходів d:

1. $d(q_0, 0, z_0) = \{(q_1, 0z_0)\}$
2. $d(q_1, 0, 0) = \{(q_1, 00)\}$
3. $d(q_1, 1, 0) = \{(q_2, \varepsilon)\}$
4. $d(q_2, 1, 0) = \{(q_2, \varepsilon)\}$
5. $d(q_2, \varepsilon, z_0) = \{(q_0, \varepsilon)\}$



Тоді послідовність тактів для рядка 0011 буде такою:

$(q_0, 0011, z_0) \vdash (q_1, 011, 0z_0)$ по (1)
$(q_1, 011, 0z_0) \vdash (q_1, 11, 00z_0)$ по (2)
$(q_1, 11, 00z_0) \vdash (q_2, 1, 0z_0)$ по (3)
$(q_2, 1, 0z_0) \vdash (q_2, \varepsilon, z_0)$ по (4)
$(q_2, \varepsilon, z_0) \vdash (q_0, \varepsilon, \varepsilon)$ по (5)

Приклад 2.

Розглянемо недетермінований МП-автомат, який розпізнає мову $L = \{ww^{-1} \mid w \in \{a,b\}^+\}$, де:

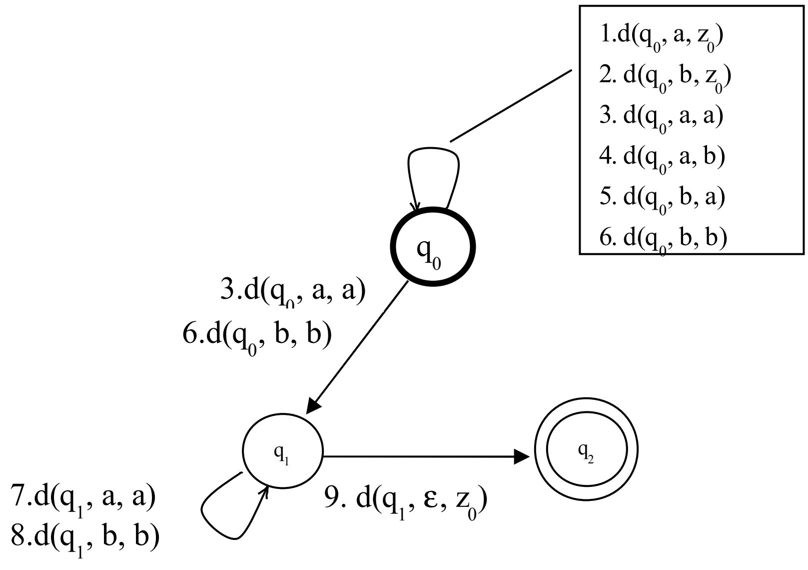
w – непустий ланцюжок, який складається з {a,b}

w^{-1} – «дзеркальне відображення» рядка w.

$M = (\{q_0, q_1, q_2\},\{a, b\},\{z_0, a, b\}, d, q_0, z_0, \{q_2\})$,

де функція переходів d:

1. $d(q_0, a, z_0) = \{(q_0, az_0)\}$
2. $d(q_0, b, z_0) = \{(q_0, bz_0)\}$
3. $\underline{d(q_0, a, a) = \{(q_0, aa), (q_1, \varepsilon)\}}$
4. $d(q_0, a, b) = \{(q_0, ab)\}$
5. $d(q_0, b, a) = \{(q_0, ba)\}$
6. $\underline{d(q_0, b, b) = \{(q_0, bb), (q_1, \varepsilon)\}}$
7. $d(q_1, a, a) = \{(q_1, \varepsilon)\}$
8. $d(q_1, b, b) = \{(q_1, \varepsilon)\}$
9. $d(q_1, \varepsilon, z_0) = \{(q_2, \varepsilon)\}$



Розглянемо послідовність тактів роботи МП-автомата, який:

1) завершує роботу неправильно (остання конфігурація не є заключною);
2) розпізнає вхідний рядок.

1) $(q_0, abba, z_0)$ $\vdash$ (q_0, bba, az_0) по (1)
(q_0, bba, az_0) $\vdash$ (q_0, ba, baz_0) по (5)
(q_0, ba, baz_0) $\vdash$ $(q_0, a, bbaz_0)$ по (6.1)
$(q_0, a, bbaz0)$ $\vdash$ $(q_0, \varepsilon, abbaz_0)$ по (4)

Вхідний рядок є пустим, а коректний заключний стан не досягнуто. => Виконується повернення до найближчої ще не використаної альтернативи в неоднозначній функції переходів.
Якщо всі можливі альтернативи вже були розглянуті, то вхідний рядок не розпізнано, тобто містить помилку.

2) $(q_0, abba, z_0)$ $\vdash$ (q_0, bba, az_0) по (1)
$(q_0, bba, az0)$ $\vdash$ (q_0, ba, baz_0) по (5)
(q_0, ba, baz_0) $\vdash$ (q_1, a, az_0) по (6.2)
(q_1, a, az_0) $\vdash$ (q_1, ε, z_0) по (7)
(q_1, ε, z_0) $\vdash$ $(q_2, \varepsilon, \varepsilon)$ по (9)

Досягнуто коректний заключний стан – вхідний рядок розпізнано.

Відповідність між КВ-граматикою та МП-автоматом

Якщо КВ-граматика G задана як

$G = (T, N, P, S)$,

а МП-автомат M задано як

$M = (Q, T, H, d, q_0, z_0, F)$,

то відповідність між їх елементами буде такою:

КВ-грамматика G	МП-автомат M
T	T
$V=T\cup N$	H
P	d
S	q_0

Множина станів Q і заключних станів F МП-автомата не мають прямої відповідності в КВ-граматиці. Часто практично весь вивід по КВ-граматиці виконується з використанням тільки одного стану МП-автомата.

Приклад.

Розглянемо граматику $G = (N, T, P, S)$, де:

$N = \{E, T, F\}$

$T = \{a, +, *, (,)\}$

$S = E$

$P = \{$ 1. $E \rightarrow T + E$
2. $E \rightarrow T$
3. $T \rightarrow F * T$
4. $T \rightarrow F$
5. $F \rightarrow a$
6. $F \rightarrow (E)$
$\}$.

Граматика G описує простий вираз з двома операціями + та *. Відповідний до неї недетермінований МП-автомат буде таким:

$M = (Q, T, H, d, q_0, z_0, F)$, де

$Q = \{q_0, q_1, q_2\}$;

$z_0 = \#$;

$T = \{a, +, *, (,)\} \cup \{\varepsilon\}$;

$$H = \{\underbrace{E, T, F}_{N}, \underbrace{a, +, *, (,)}_{T}, \underbrace{\#}_{z_0}\}$$

$F = \{ q_2 \}$

Функція переходів d, яка визначена наступними допустимими переходами:

1. $d(q_0, \varepsilon, \#) = \{ (q_1, E\#) \}$
2. $d(q_1, \varepsilon, E) = \{(q_1, T+E), (q_1, T)\}$
3. $d(q_1, \varepsilon, T) = \{ q_1, F*T), (q_1, F) \}$
4. $d(q_1, \varepsilon, F) = \{ (q_1, a), (q_1, (E)) \}$
5. $d(q_1, a, a) = \{ (q_1, \varepsilon) \}$
6. $d(q_1, +, +) = \{ (q_1, \varepsilon) \}$
7. $d(q_1, *, *) = \{ (q_1, \varepsilon) \}$
8. $d(q_1, (, ( ) = \{ (q_1, \varepsilon) \}$
9. $d(q_1,),)) = \{ (q_1, \varepsilon) \}$
10. $d(q_1, \varepsilon, \#) = \{ (q_2, \varepsilon) \}$

Нехай даний недетермінований МП-автомат виконує лівосторонній вивід наступним чином:

1) при кожному переході автомата з стеку виштовхується один символ;
2) якщо виштовхнутий символ виявляється нетермінальним, то замість нього в стек заносяться символи з множини пар, які є допустимі для даної конфігурації (по черзі зліва направо);
3) якщо виштовхнутий символ виявляється термінальним, то:
   a) якщо він співпадає з поточним символом вхідного рядка, то він видаляється зі стеку, а вказівник вхідного рядка переміщується на символ далі;
   b) якщо він не співпадає з поточним символом вхідного рядка, це означає хибний шлях роботи автомата і виконується повернення до найближчої ще нерозглянутої альтернативи;
4) якщо виникає конфігурація, недопустима для визначеної вище функції переходів d, то це також означає хибний шлях роботи автомата і виконується повернення до найближчої ще нерозглянутої альтернативи;
5) якщо виникає недопустима конфігурація, а всі альтернативи вже розглянуті, то даний вхідний рядок не належить мові, визначеній заданою вище граматикою G.

Для реалізації повернень необхідно запам'ятати пройдений шлях станів. При переході з початкового стану автомата q_0 в наступний стан q_1 вміст стеку автомата приймає вигляд E#, а перехід до заключного стану q_2 можливий лише тоді, коли стек пустий (містить тільки символ #) і символи вхідного рядка вже вичерпані. Таким чином, весь процес емуляції лівостороннього виводу виконується автоматом тільки в одному стані q_1.

В наступній таблиці приведена послідовність конфігурацій даного МП-автомата при обробці вхідного рядка a+a.

Пояснення 1. Запис $d(q_1, \varepsilon, E) \vdash (q_1, T+E)$ означає перехід за такими правилами:

1) виконується перехід зі стану q_1 в стан q_1 (тобто в той самий стан);
2) вхідним символом є порожній рядок ε, тобто з вхідного рядка береться ε символів (вказівник вхідного рядка не переміщується), і перехід виконується незалежно від того, який символ стоїть на початку вхідного рядка;
3) перехід виконується, якщо в вершині стеку знаходиться символ E, при переході символ E із стеку виштовхується;
4) після виштовхування символу E в стек заштовхується рядок символів T+E.

Пояснення 2. Запис $d(q_1, a, a) \vdash (q_1, \varepsilon)$ означає перехід за такими правилами:

1) якщо на початку вхідного рядка стоїть символ **a**, і він же знаходиться у вершині стека, то виконується перехід зі стану q_1 в стан q_1 (тобто в той самий стан);
2) виконується перехід на аналіз наступного після **a** символу вхідного рядка;
3) символ **a** виштовхується зі стеку;
4) в стек заштовхується ε символів, тобто нічого.

Таблиця 1

№ конфігурації	Поточна конфігурація			Виконуваний перехід		Примітки
	Стан	Положення покажчика вхідного рядка	Вміст стека	№	Перехід	
1	q_0	$\underset{\uparrow}{a} + a$	#	1	d (q_0, ε, #) $\vdash$ (q_1, E #)	
2	q_1	$\underset{\uparrow}{a} + a$	E #	2.1	d (q_1, ε, E) $\vdash$ (q_1, T + E)	
3	q_1	$\underset{\uparrow}{a} + a$	T + E #	3.1	d (q_1, ε, T) $\vdash$ (q_1, F * T)	
4	q_1	$\underset{\uparrow}{a} + a$	F * T + E #	4.1	d (q_1, ε, F) $\vdash$ (q_1, a)	
5	q_1	$\underset{\uparrow}{a} + a$	a * T + E #	5	d (q_1, a, a) $\vdash$ (q_1, ε)	
6	q_1	$a \underset{\uparrow}{+} a$	* T + E #	Перехід d (q_1, +, *) недопустимий => повертаємося до найближчої альтернативної конфігурації №4 та беремо перехід 4.2		
4	q_1	$\underset{\uparrow}{a} + a$	F * T + E #	4.2	d (q_1, ε, F) $\vdash$ (q_1, (E))	
5	q_1	$\underset{\uparrow}{a} + a$	(E) * T + E #	Перехід d (q_1, a, ( ) недопустимий => повертаємося до найближчої альтернативної нерозглянутої конфігурації №3 і беремо перехід 3.2		
3	q_1	$\underset{\uparrow}{a} + a$	T + E #	3.2	d (q_1, ε, T) $\vdash$ (q_1, F)	
4	q_1	$\underset{\uparrow}{a} + a$	F + E #	4.1	d (q_1, ε, F) $\vdash$ (q_1, a)	
5	q_1	$\underset{\uparrow}{a} + a$	a + E #	5	d (q_1, a, a) $\vdash$ (q_1, ε)	
6	q_1	$a \underset{\uparrow}{+} a$	+ E #	6	d (q_1, +, +) $\vdash$ (q_1, ε)	
7	q_1	$a + \underset{\uparrow}{a}$	E #	2.1	d (q_1, ε, E) $\vdash$ (q_1, T + E)	
8	q_1	$a + \underset{\uparrow}{a}$	T + E #	3.1	d (q_1, ε, T) $\vdash$ (q_1, F * T)	
9	q_1	$a + \underset{\uparrow}{a}$	F * T + E #	4.1	d (q_1, ε, F) $\vdash$ (q_1, a)	
10	q_1	$a + \underset{\uparrow}{a}$	a * T + E #	5	d (q_1, a, a) $\vdash$ (q_1, ε)	
11	q_1	$a + a\underset{\uparrow}{\ }$	* T + E #	Якщо б q_1 було також і завершальним станом, то іноді таку конфігурацію (q, ε, z) вважають коректною для завершення (див. помилкову заключну конфігурацію 2), оскільки, загалом рядок розпізнано як допустимий. Проте дерево виводу було отримано неправильним. Тому, при строгому трактуванні ця конфігурація є недопустимою для даного автомата і виконується повернення до найближчої нерозглянутої альтернативної конфігурації № 9 і беремо перехід 4.2.		

Таблиця 1 (продовження)

<table>
<tr><th rowspan="2">№ конфігурації</th><th colspan="3">Поточна конфігурація</th><th colspan="2">Виконуваний перехід</th><th rowspan="2">Примітки</th></tr>
<tr><th>Стан</th><th>Положення покажчика вхідного рядка</th><th>Вміст стека</th><th>№</th><th>Перехід</th></tr>
<tr><td>9</td><td>q_1</td><td>a + a
↑</td><td>F * T + E #</td><td>4.2</td><td>$d\ (q_1, \varepsilon, F) \vdash (q_1, (\ E\ )\ )$</td><td></td></tr>
<tr><td>10</td><td>q_1</td><td>a + a
↑</td><td>(E) * T + E #</td><td colspan="3">Перехід d (q_1, a, ( ) недопустимий => повертаємося до найближчої альтернативної конфігурації № 8 і беремо перехід 3.2</td></tr>
<tr><td>8</td><td>q_1</td><td>a + a
↑</td><td>T + E #</td><td>3.2</td><td>$d\ (q_1, \varepsilon, T) \vdash (q_1, F)$</td><td></td></tr>
<tr><td>9</td><td>q_1</td><td>a + a
↑</td><td>F + E #</td><td>4.1</td><td>$d\ (q_1, \varepsilon, F) \vdash (q_1, a)$</td><td></td></tr>
<tr><td>10</td><td>q_1</td><td>a + a
↑</td><td>a + E #</td><td>5</td><td>$d\ (q_1, a, a) \vdash (q_1, \varepsilon)$</td><td></td></tr>
<tr><td>11</td><td>q_1</td><td>a + a
↑</td><td>+ E #</td><td colspan="3">Отримали таку ж ситуацію для 11-ої конфігурації, як і для попередньої 11-ї. Див. коментарій вище.
Перехід d (q_1, ε, +) недопустимий => повертаємося до найближчої альтернативної нерозглянутої конфігурації № 9 і беремо перехід 4.2</td></tr>
<tr><td>9</td><td>q_1</td><td>a + a
↑</td><td>F + E #</td><td>4.2</td><td>$d\ (q_1, \varepsilon, F) \vdash (q_1, (\ E\ )\ )$</td><td></td></tr>
<tr><td>10</td><td>q_1</td><td>a + a
↑</td><td>(E) + E #</td><td colspan="3">Перехід d (q_1, a, ( ) недопустимий => повертаємося до найближчої альтернативної нерозглянутої конфігурації № 7 і беремо перехід 2.2</td></tr>
<tr><td>7</td><td>q_1</td><td>a + a
↑</td><td>E #</td><td>2.2</td><td>$d\ (q_1, \varepsilon, E) \vdash (q_1, T)$</td><td></td></tr>
<tr><td>8</td><td>q_1</td><td>a + a
↑</td><td>T #</td><td>3.1</td><td>$d\ (q_1, \varepsilon, T) \vdash (q_1, F * T)$</td><td></td></tr>
<tr><td>9</td><td>q_1</td><td>a + a
↑</td><td>F * T #</td><td>4.1</td><td>$d\ (q_1, \varepsilon, F) \vdash (q_1, a)$</td><td></td></tr>
<tr><td>10</td><td>q_1</td><td>a + a
↑</td><td>a * T #</td><td>5</td><td>$d\ (q_1, a, a) \vdash (q_1, \varepsilon)$</td><td></td></tr>
<tr><td>11</td><td>q_1</td><td>a + a
↑</td><td>* T #</td><td colspan="3">Отримали таку ж ситуацію для 11-ої конфігурації, як і для попередньої 11-ї. Див. коментарій вище.
Перехід d (q_1, ε, *) недопустимий => повертаємося до найближчої альтернативної нерозглянутої конфігурації № 9 і беремо перехід 4.2</td></tr>
<tr><td>9</td><td>q_1</td><td>a + a
↑</td><td>F * T #</td><td>4.2</td><td>$d\ (q_1, \varepsilon, F) \vdash (q_1, (\ E\ )\ )$</td><td></td></tr>
<tr><td>10</td><td>q_1</td><td>a + a
↑</td><td>(E) * T #</td><td colspan="3">Перехід d (q_1, a, ( ) недопустимий => повертаємося до найближчої альтернативної нерозглянутої конфігурації № 8 і беремо перехід 3.2</td></tr>
</table>

Таблиця 1 (продовження)

№ конфігурації	Поточна конфігурація			Виконуваний перехід		Примітки
	Стан	Положення покажчика вхідної рядка	Вміст стека	№	Перехід	
8	q_1	$a + \underset{\uparrow}{a}$	T #	3.2	$d\ (q_1, \varepsilon, T) \vdash (q_1, F)$	
9	q_1	$a + \underset{\uparrow}{a}$	F #	4.1	$d\ (q_1, \varepsilon, F) \vdash (q_1, a)$	
10	q_1	$a + \underset{\uparrow}{a}$	a #	5	$d\ (q_1, a, a) \vdash (q_1, \varepsilon)$	
11	q_1	$a + a\,\underset{\uparrow}{}$	#	10	$d\ (q_1, \varepsilon, \#) \vdash (q_2, \varepsilon)$	
12	q_2	ε	ε	Перейшли в допустиму заключну конфігурацію d (q2, ε, ε) => Вхідний рядок a + a розпізнано з правильним проходом за правилами граматики, при якому є можливість побудови дерева виводу (розбору)		

Як видно з таблиці, в результаті недетермінованості переходів при роботі МП-автомата відбуваються повернення. Тому такий синтаксичний аналізатор виходить неефективним (повільним).

Задача побудови ефективних синтаксичних аналізаторів загалом є непростою. Проте, при накладанні певних обмежень на граматику, яка породжує деяку КВ-мову, побудова ефективного синтаксичного аналізатора на основі детермінованого МП-автомата може стати можливою.

На завершення цього підрозділу сформулюємо деякі теореми (без доведень) і визначення.

Теорема 1. Для довільної контекстно-вільної мови існує приймаючий її недетермінований МП-автомат.

Теорема 2. Якщо деяку мову L приймає недетермінований МП-автомат M, то вона може бути прийнятою також і недетермінованим автоматом M′, який має єдиний кінцевий стан і при переході в який (і тільки в цьому випадку) стек МП-автомата стає пустим.

Теорема 3. Якщо деяку мову L приймає недетермінований МП-автомат M, то ця мова L є контекстно-вільною.

Визначення 1. Детермінованим МП-автоматом називається такий МП-автомат, будь-якій конфігурації якого відповідає тільки один перехід.

Строге визначення детермінованого МП-автомату є таким:
МП-автомат $M = (Q, T, H, d, q_0, z_0, F)$ є детермінованим, якщо для довільних $q \in Q$, $a \in T$ та $A \in H$ виконуються наступні умови:

1) $d(q, a, A)$ має не більше одного елемента в множині переходів;
2) $d(q, \varepsilon, A)$ має не більше одного елемента в множині переходів;
3) якщо $d(q, \varepsilon, A) \neq \varnothing$, то $d(q, a, A) = \varnothing$ для довільного $a \in T$.

Третя умова іншими словами означає, що коли з деякої конфігурації автомат може виконати хоча б один ε-перехід, то цей перехід є єдиним для даної конфігурації.

Визначення 2. Детермінованою контекстно-вільною мовою називають мову, яка приймається детермінованим МП-автоматом.

Теорема 4. Якщо мова $L \subset T^*$ – детермінована, то мова $\overline{L} = T^* \setminus L$ також є детермінованою мовою.

Теорема 5. Довільна регулярна мова є детермінованою мовою, тобто для неї можна побудувати приймаючий її детермінований МП-автомат. Зворотне твердження є невірним, тобто не для всякої детермінованої мови можна побудувати звичайний (не магазинний) детермінований автомат (оскільки детермінована мова може бути також КВ-мовою, а не тільки регулярною).

Теорема 6. Довільна детермінована мова є контекстно-вільною мовою. Зворотне твердження є невірним.