第 3 章　　自动机基础

　　自动机 — 是一种语言模型，是语言的一种识别工具，其中的

　　有限自动机(finite automata）FA

被用来处理 正规语言，后者是编译程序设计中词法分析的对象。

【内容提要】

3.1　正规语言及其描述方法

3.2　有限自动机的定义与分类

3.3　有限自动机的等价变换1,2

3.4　有限状态自动机的实现问题

3.5　正规语言描述方法间的相互转换

3.1 正规语言及其描述方法

【定义】 正规语言是指由正规文法定义的语言；程序设计语言中的单词，大都属于此种语言。

▪**正规文法**是指仅有三种形式的产生式：

(1) $A \rightarrow aB$ (2) $A \rightarrow a$ (3) $A \rightarrow \varepsilon$

▪**正规语言**有三种等价的表示方法：

(1) 正规文法 (2) 正规式 (3) 有限自动机

【例3.1】 G(Z)：$A \rightarrow aA \mid \varepsilon$

正规文法

$\because$ $A \Rightarrow \varepsilon$;

$A \Rightarrow aA \Rightarrow aaA \Rightarrow aaaA \Rightarrow \ldots \Rightarrow a^n$, $n \geqslant 0$

$\therefore$ $L(G)=\{ a^n \mid n \geqslant 0 \}$

正规语言

※ 正规语言判定示例:

【例3.2】 L1 $=\{ a^m b^n \mid m \geqslant 0, n \geqslant 1 \}$， 正规语言 ?

∵ 可由正规文法定义:

G1(S): S -> aS|bA ; A -> bA|ε

∴ L1 是正规语言。

【例3.3】 L2 $=\{ (ab)^n \mid n \geqslant 1 \}$， 正规语言 ?

∵ 可由正规文法定义:

G2(S): S -> aA ; A -> bB ; B -> aA|ε

∴ L2 是正规语言。

【例3.4】 L3 $=\{ a^n b^n \mid n \geqslant 0 \}$， 正规语言 ?

∵ 不能由正规文法定义(正规文法无法描述a和b数量上相等！):

∴ L3 不是正规语言！

3.1.1 正规语言的正规式表示法

※ **正规式** 是指由字母表中的符号，通过以下三种运算(也可以使用圆括号)连接起来构成的表达式 e：连接（.） 或（|） 闭包（+，*）

例：

正规式	正规式的 值
ab. cde	abcde
ab|cde	ab ， cde
a^+	a ，aa ，aaa ，… ，a^n ，…
a^*	ε，a，aa，aaa， …，a^n，…

※ 设 val(e)，L(e) 分别表示正规式 e 的值和语言，则：

$$L(e)=\{\ x\ |\ x=val(e)\}$$

即 正规式表示的语言是该正规式所有值的集合(正规集)。

※ 正规式表示正规语言示例:

【例】:(1) $e1 = a^*b^+$

展开:

$L(e1) = \{ a^m b^n \mid m \geq 0, n \geq 1 \}$,

$= \{ b, ab, bb, aaab, aab, abb, aabb, \ldots \}$

(2) $e2 = (ab)^+$

$L(e2) = \{ (ab)^n \mid n \geq 1 \}$,

$= \{ ab, abab, ababab, abababab, \ldots \}$

(3) $e3 = ab^*c \mid b^*$

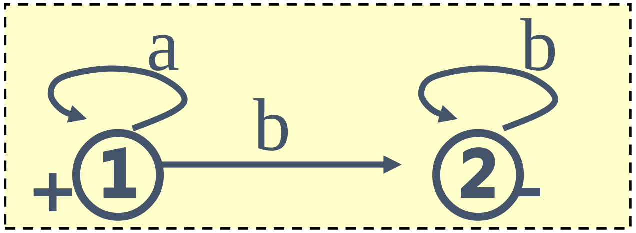
$L(e3) = \{ ab^n c, \; b^n \mid n \geq 0 \}$,

$= \{ ac, abc, abbc, abbbc, \ldots; \; \varepsilon, b, bb, bbb, \ldots \}$

3.1.2 正规语言的有限自动机表示法：

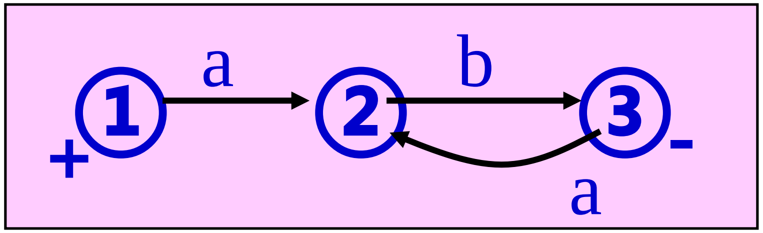
L1= $\{ a^m b^n \mid m \geqslant 0, n \geqslant 1 \}$，

FA1:



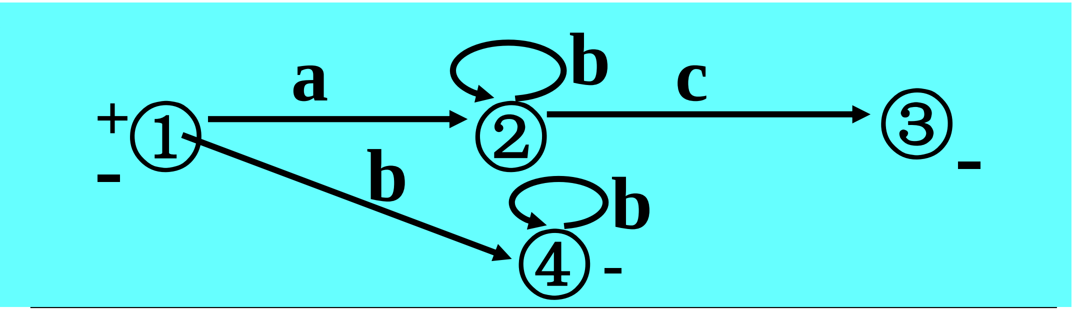
L2= $\{ (ab)^n \mid n \geqslant 1 \}$，

FA2:



L3= $\{ ab^n c, b^n \mid n \geqslant 0 \}$，

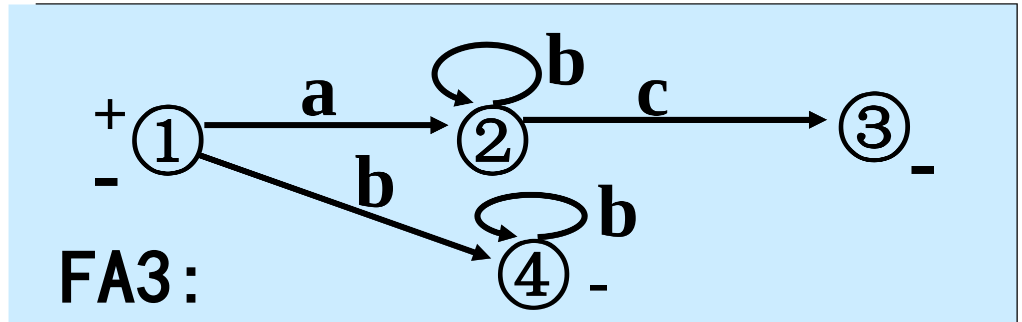
FA3:



※ 初看起来，有限自动机是带标记的有向图！

※有限自动机表示法说明:

L3=
{ ab^nc , b^n | $n \geqslant 0$ }



【图符说明】:

- {1, 2, 3, 4} —— 状态集;

  其中: +(开始状态); -(结束状态)
- {a, b, c} —— 字母表;
- δ (1, a)=2 – 变换 (或 $① \xrightarrow{a} ②$);...

  (表示1状态遇符号a变到2状态, ...);

【运行机制】:

一个 FA, 若存在一条从某开始状态 i 到某结束状态 j 的通路, 且这条通路上所有边的标记连成的符号串为α, 则 α就是一个句子; 所有这样的 α的集合, 就是该 FA 表示的语言。

※ 正规语言的三种表示法综合示例:

【例3.5】 $L = \{ ab^nc,\ b^n \mid n \geqslant 0 \}$, $\Sigma = \{a, b, c\}$;

1. 正规文法描述:

G(S): $S \to aA \mid bB \mid \varepsilon$, $A \to bA \mid c$, $B \to bB \mid \varepsilon$

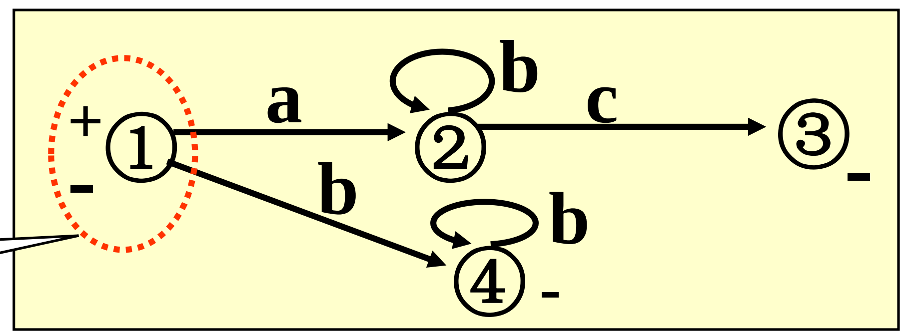
2. 正规式描述:

$e = ab^*c \mid b^*$

3. 有限自动机描述:

表示可接受**空串**!

FA:



【注】 凡是能由上述三种方法表示的语言，一定是正规语言；反之，凡是不能由上述三种方法表示的语言，一定不是正规语言。

3.2 有限自动机的定义和分类

3.2.1 有限自动机的定义

【定义】 有限自动机是一种数学模型，用于描述正规语言，可定义为五元组： $FA=(Q, \Sigma, S, F, \delta)$

其中：

Q(有限状态集)；
Σ(字母表)；
S(开始状态集，$S \subseteq Q$)；
F(结束状态集，$F \subseteq Q$)；
δ：变换（二元函数）：

$i,j \in Q$, $a \in \Sigma+\{\varepsilon\}$;

$\delta(i,a)=j$ **或** $i \xrightarrow{a} j$

【含义】

状态 i 遇符号 a，变换为状态 j

3.2.2 有限自动机怎样描述语言

令 L(FA)为自动机FA所描述的正规语言；则

$$L(FA)=\{\ x \mid i \overset{x}{\Rightarrow} j\ ,\ x\in\Sigma^*\ ,\ i\in S,\ j\in F\ \}$$

含义是:

即:

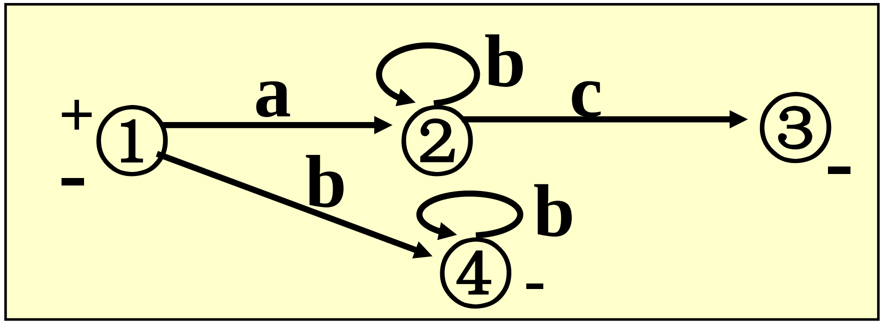
设 $x=a_1\ a_2\ldots a_n$, $a_i\in\Sigma+\{\varepsilon\}$

则有 $i \overset{a_1}{\Rightarrow} i_1 \overset{a_2}{\Rightarrow} i_2 \ \ldots \ \overset{a_n}{\Rightarrow} j$

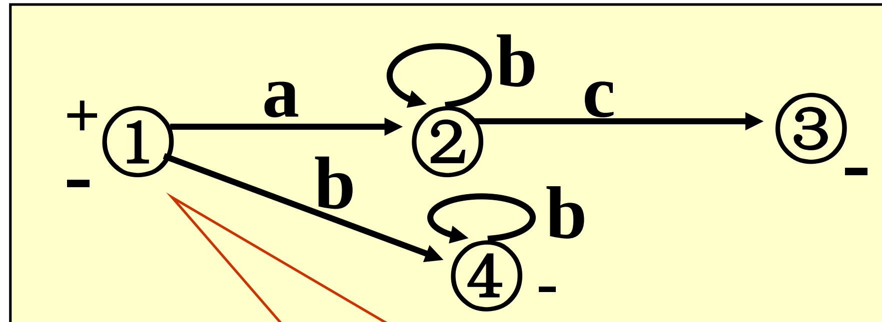
FA 存在一条从某初始状态 i 到某结束状态 j 的连续变换，且每次变换(=>)的边标记连成的符号串恰好为 x ；称x为FA接受，否则拒绝。

如 右图有限自动机:

则 L(FA)的识别过程如下所示:



※ L(FA)的生成(或识别)过程示例:



Ⅰ.第一条通路：FA1

+① $\overset{a}{\Rightarrow}$ ② $\overset{c}{\Rightarrow}$ ③_

+① $\overset{a}{\Rightarrow}$ ② $\overset{b}{\Rightarrow}$ ② $\overset{c}{\Rightarrow}$ ③_

+① $\overset{a}{\Rightarrow}$ ② $\overset{b}{\Rightarrow}$ ② $\overset{b}{\Rightarrow}$ ② $\overset{c}{\Rightarrow}$ ③_

…

Ⅱ.第二条通路：FA2

+① $\overset{\varepsilon}{\Rightarrow}$ ①_

+① $\overset{b}{\Rightarrow}$ ④_

+① $\overset{b}{\Rightarrow}$ ④ $\overset{b}{\Rightarrow}$ ④_

…

接受空串的
FA的典型特征！

$\therefore$ L(FA1)=$\{ ab^nc \mid n \geqslant 0 \}$

$\therefore$ L(FA2)=$\{ b^n \mid n \geqslant 0 \}$

因而

$\therefore$ L(FA)=$\{ab^nc,\ b^n \mid n \geqslant 0\}$

3.2.3 有限自动机的两种表现形式

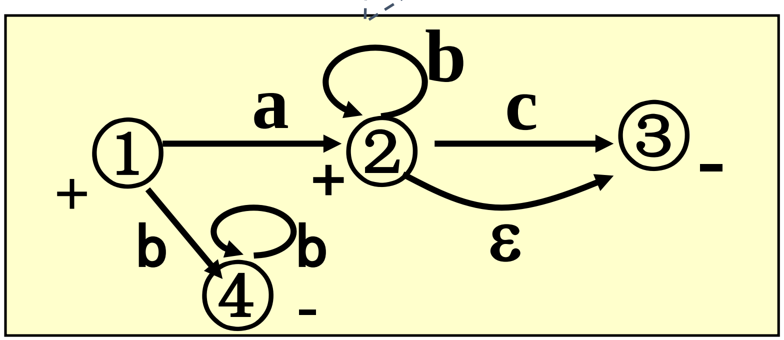
【例3.6】有限自动机 ： FA=(Q, ∑, S, F, δ)

其中：Q={1, 2, 3, 4}，∑={a, b, c}， S={1, 2}， F={3, 4}

δ ： δ (1, a)=2; δ (1, b)=4； δ (2, b)=2；
δ (2, c)=3； δ (2, ε)=3； δ (4, b)=4；

FA 的两种表现形式：

(1) 状态图：



(2) 变换表：

	a	b	c	ε	
1	2	4			+
2		2	3	3	+
3					−
4		4			−

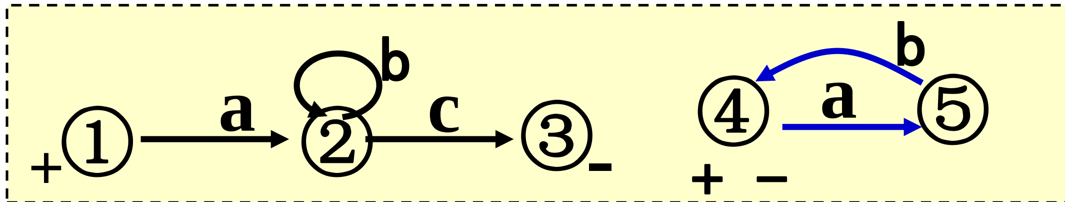
开始状态
结束状态

※ **变换表结构：行(状态)，列(符号)，表项(变换后状态)**

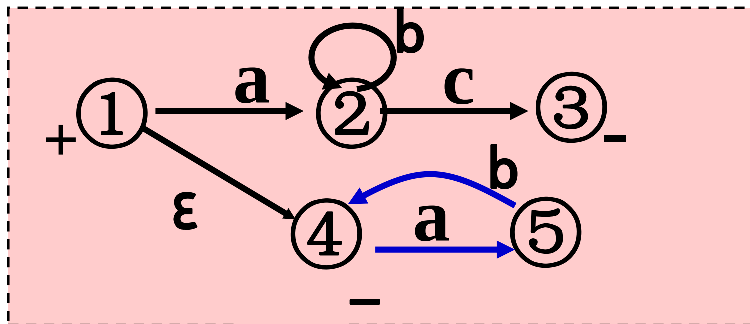
※ 有限自动机的构造示例1

【例3.7】 A={ ab^nc, $(ab)^n$ | $n \geq 0$ } FA的构造：

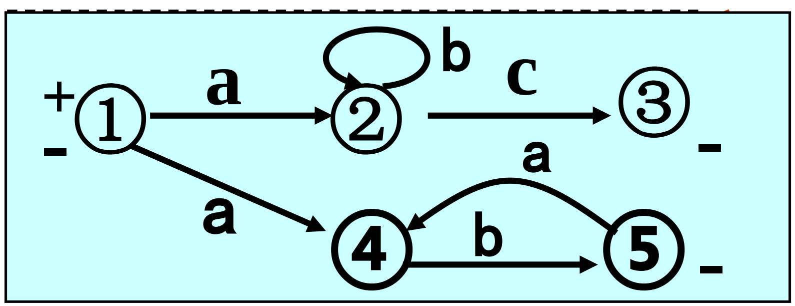
▪方法一：联合式



▪方法二：组合式1



或



有好用的吗?!

比较：

一：开始状态不唯一，不好用！

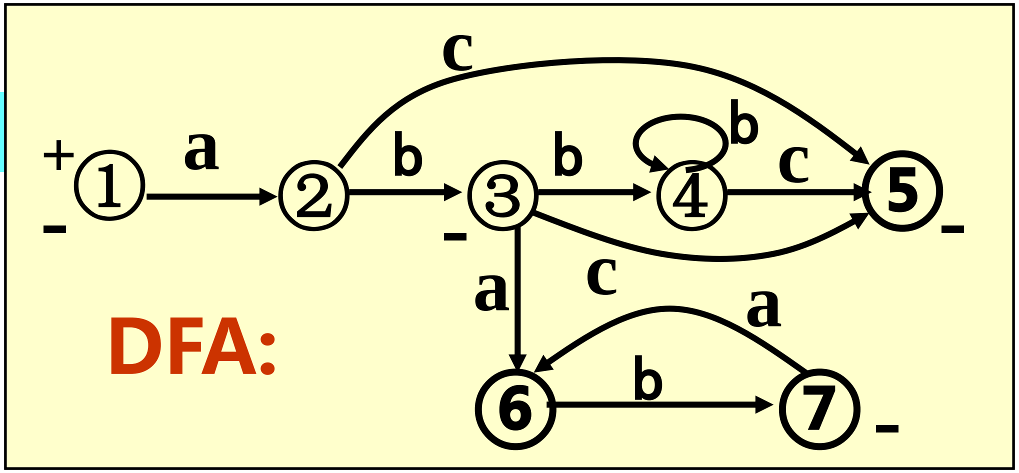
二：带有ε 边，还是不好用！

或：变换函数不单值，如（ $\delta(1, a)=(2|4)$ ），也不好用！

※ 有限自动机的构造示例2

▪方法三：确定式

$A=\{ab^nc,(ab)^n|n\geq 0\}$



※ 确定的有限自动机如右图所示：

3.2.4 有限自动机的分类

1. 确定的有限自动机(DFA)

特征：①开始状态唯一；②变换函数单值；③不带ε边。

2. 非确定的有限自动机(NFA)

-- 不能全部具备上述特征者！

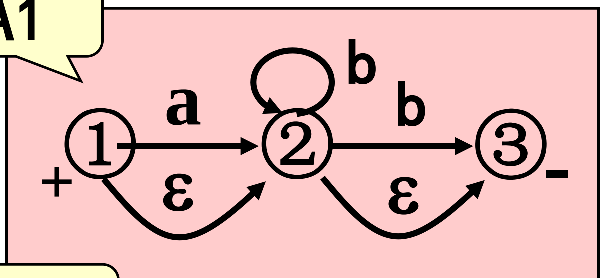
(1) 带有ε边的非确定的有限自动机(εNFA)

(2) 不带有ε边的非确定的有限自动机($\bar{\varepsilon}$NFA)

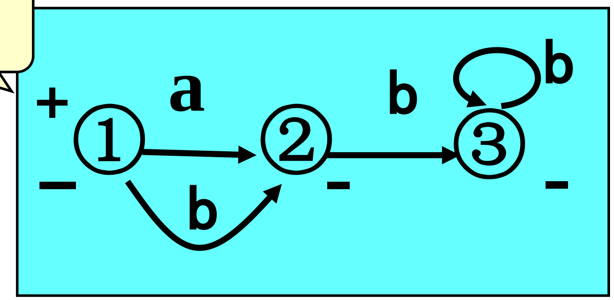
※ 有限自动机的分类示例

【例3.8 】试分别指出下述有限自动机的分类情况：

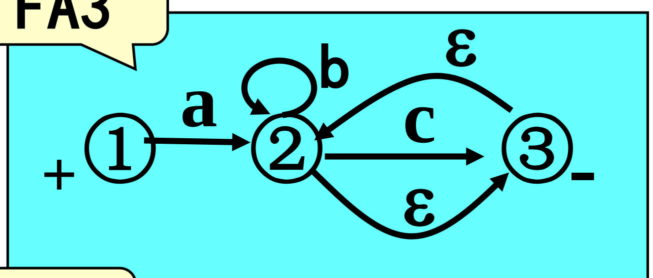
FA1



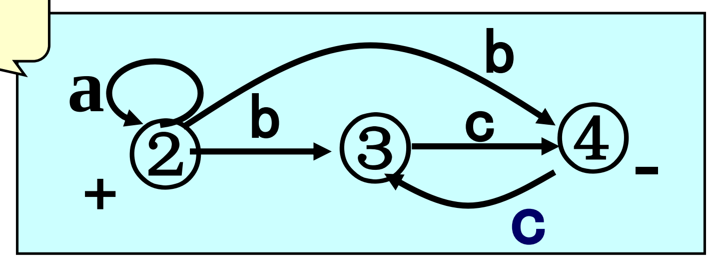
FA2



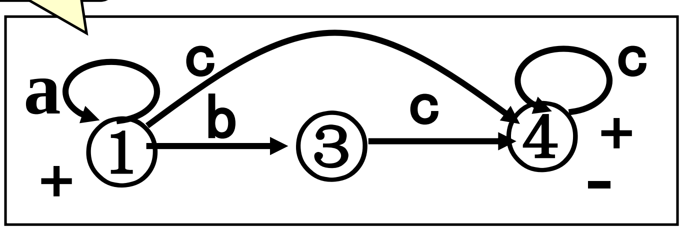
FA3



FA4



FA5



结论：

DFA：FA2

NFA：FA4, FA5（$\bar{\varepsilon}$ NFA）

FA1, FA3 （εNFA）

➢正规式描述语言的简单运算例:

1. $L(a)=\{a\}$
2. $L(a|b)=L(a)+L(b)=\{a, b\}$
3. $L(a.b)=L(a).L(b)=\{a\}.\{b\}=\{ab\}$
4. $L(a(a|b)c)=L(a).L(a|b).L(c)$

   $=\{a\}.\{a, b\}.\{c\}=\{aac, abc\}$
5. $L(b^*)=(L(b))^*=\{b\}^*=\{\varepsilon, b, bb, bbb, \cdots\}$

   $=\{b^n | n\geqslant 0\}$
6. $L(ab^*)=L(a).L(b^*)=\{a, ab, abb, \cdots\}$

   $=\{ab^n | n\geqslant 0\}$
7. $L((a|b)^*)=(L(a|b))^*=\{a, b\}^*$

即:由a, b组成的所有符号串(包括空串)集合。

➢基本图形库

$\overset{+}{=>}$ $\qquad$ $\overset{+}{\bullet=>}$ $\qquad$ $\underset{\bullet\ell}{\overset{+}{=>}}$ $\qquad$ $\underset{\ell}{\overset{+}{=>}}$ $\qquad$ $A \longrightarrow \alpha\beta$

P:
$$E \longrightarrow T \mid E + T \mid E - T$$
$$T \longrightarrow F \mid T * F \mid T / F$$
$$F \longrightarrow i \mid (E)$$

=>*, =>+ , =>.* , =>.+ , =>l* , =>l+ , =>.l+ , =>.l*

$\underset{\ell}{=>}$ $\qquad$ $\underset{\bullet\ell}{=>}$