

Exercise 3

3.1 正規表現を文脈自由文法に変換せよ

a. $((xy^*x) | (yx^*y))^?$

$$\begin{aligned} S &\rightarrow \epsilon | E \\ E &\rightarrow x Y x | y X y \\ X &\rightarrow x X | \epsilon \\ Y &\rightarrow y Y | \epsilon \end{aligned}$$

b. $((0|1)^+ "." (0|1)^*) | ((0|1)^* "." (0|1)^+)$

$$\begin{aligned} S &\rightarrow E "." H | H "." E \\ E &\rightarrow Z | Z E \\ H &\rightarrow \epsilon | Z H \\ Z &\rightarrow 0 | 1 \end{aligned}$$

3-2 * (* つきなのでパス)

9個の単語 + セミコロンからなる英文の文法を書く。条件が...

3-3 以下の言語の曖昧でない文法を書き下せ。Hint: Yacc に食わせてみればいんじゃないね?

a. a,b からなる回文

$$S \rightarrow a S a | b S b | a | b | \epsilon$$

Yacc に食わせると文句が出るらしい。確かにシフト還元衝突は起きているように見える。

b. a^*b^* で a のほうが多い

$$\begin{aligned} S &\rightarrow A X \\ A &\rightarrow a A | a \\ X &\rightarrow a X b | \epsilon \end{aligned}$$

c. ()[] 対応が取れているもの

$$\begin{aligned} S &\rightarrow T \mid ST \\ T &\rightarrow (S) \mid [S] \mid \epsilon \end{aligned}$$

d. *

e. `public final static synchronized transient` のくりかえしなし順列か部分集合

$$\begin{aligned} S &\rightarrow P F S Y T \mid \dots (5! = 120 \text{ 個書き出すのか } Yo!) \\ P &\rightarrow \epsilon \mid \text{public} \\ F &\rightarrow \epsilon \mid \text{final} \\ S &\rightarrow \epsilon \mid \text{static} \\ Y &\rightarrow \epsilon \mid \text{synchronized} \\ T &\rightarrow \epsilon \mid \text{transient} \end{aligned}$$

実際のコンパイラなら複数回登場を許すね! (or 意味解釈で落とすか、順番を固定するか)

f. セミコロンの分割する文法 (Pascal, ML 的)

$$\begin{aligned} S_0 &\rightarrow (S_1) \\ S_1 &\rightarrow S_2 ; S_2 \mid S_2 \\ S_2 &\rightarrow (S_1) \mid S \\ S &\rightarrow \text{statement} \end{aligned}$$

g. セミコロンの終端以下略 (C)

$$\begin{aligned} S_0 &\rightarrow \{ S_1 \} \\ S_1 &\rightarrow E ; S_1 \mid E ; \mid \{ S_1 \} \\ E &\rightarrow \text{expression} \end{aligned}$$

これだと空ブロックを受理しない。受理させるなら S_1 に ϵ 追加か

3-4. LL(1) を満足するような Grammar 3.1 同等文法を書け

Grammar3.1 は次のとおり。

$$\begin{aligned}
 S &\rightarrow S ; S \\
 S &\rightarrow \text{id} := E \\
 S &\rightarrow \text{print} (L) \\
 E &\rightarrow \text{id} \\
 E &\rightarrow \text{num} \\
 E &\rightarrow E + E \\
 E &\rightarrow (S , E) \\
 L &\rightarrow E \\
 L &\rightarrow L , E
 \end{aligned}$$

現状で何が conflict を起こすのかを認識する必要がある。 ϵ はないので nullable は false であり FOLLOW-set は不要のはず。

$$\begin{aligned}
 \text{FIRST}[S] &= \text{id print} \\
 \text{FIRST}[E] &= \text{id num (} \\
 \text{FIRST}[L] &= \text{FIRST}[E] = \text{id num (}
 \end{aligned}$$

	id	print	num	()	+	,	;	:=
S	0,1	0,2							
E	3,5		4,5	5,6					
L	7,8		7,8	5,6					

こうか? conflict しまくりだ。

明らかにどれも Left-recursion 除去が必要。除去は

$$\begin{aligned}
 X &\rightarrow X \gamma_i \\
 X &\rightarrow \alpha_j
 \end{aligned}$$

とあるとき、

$$\begin{aligned}
 X &\rightarrow \alpha_j X' \quad (\text{each } j) \\
 X' &\rightarrow \gamma_i X' \quad (\text{each } i) \\
 X' &\rightarrow \epsilon
 \end{aligned}$$

と展開できる (p.52) から、以下のようになるはず。

$$\begin{aligned}
 S &\rightarrow \text{id} := E S' \\
 S &\rightarrow \text{print} (L) S' \\
 S' &\rightarrow ; S S' \\
 S' &\rightarrow \epsilon \\
 E &\rightarrow \text{id} E' \\
 E &\rightarrow \text{num} E' \\
 E &\rightarrow (S , E) E' \\
 E' &\rightarrow + E E' \\
 E' &\rightarrow \epsilon \\
 L &\rightarrow E L' \\
 L' &\rightarrow , E L' \\
 L' &\rightarrow \epsilon
 \end{aligned}$$

すると、

	FIRST	nullable	FOLLOW
S	id print	no	, \$ FOLLOW[S']
S'	;	yes	FOLLOW[S]
E	id num (no	FIRST[S'] [E'] [L'] FOLLOW[S] [E'] [L] [L']
E'	+	yes	FOLLOW[E]
L	FIRST[E]	no)
L'	, FIRST[E]	yes	FOLLOW[L]

だから

	FIRST	nullable	FOLLOW
S	id print	no	, \$
S'	;	yes	, \$
E	id num (no	; + , id num) \$
E'	+	yes	; + , id num) \$
L	id num (no)
L'	, id mi, (yes)

	id	print	num	()	+	,	;	:=
S	0	1							
S'							3	2	
E	4		5	6					
E'	8		8		8	7,8	8	8	
L	9		9	9					
L'					11		10		

あらら、 E' で衝突する。天下りの変換ではダメのなようだ。衝突箇所は、

$$E' \rightarrow + E E'$$

$$E' \rightarrow \epsilon$$

これが? 違うな。たとえば

$$E' \rightarrow + E E' \quad \dots(1)$$

$$\sim$$

$$\text{id } E E' \quad \dots(2)$$

だから、(1) の E' と、(1) の E から導出された式 (2) の E' の、どちらを ϵ にするべきかが分からないということか。左再帰除去をきちんと考え直す必要がある。

$$S \rightarrow \text{id} := E S'$$

$$S \rightarrow \text{print} (L) S'$$

$$S' \rightarrow ; S S'$$

$$S' \rightarrow \epsilon$$

$$E \rightarrow \text{id} E'$$

$$E \rightarrow \text{num} E'$$

$$E \rightarrow (S , E) E'$$

$$E' \rightarrow + E$$

$$E' \rightarrow \epsilon$$

$$L \rightarrow E L'$$

$$L' \rightarrow , E L'$$

$$L' \rightarrow \epsilon$$

これでよいのか。文法壊れてないよね?

	id	print	num	()	+	,	;	:=	\$
S	0	1								
S'							3	2		3
E	4		5	6						
E'	8		8		8	7	8	8		8
L	9		9	9						
L'					11		10			

3-5 FIRST, nullable, FOLLOW 集合の計算 とパース表

	FIRST	nullable	FOLLOW
S'	\$ FIRST[S]	no	
S	FIRST[X]	yes	\$ } FIRST[E]
B	\	no	FIRST[S] FIRST[E]
E	\	no	FOLLOW[X]
X	{ \ WORD begin end FIRST[B]	no	FOLLOW[S] FIRST[S]

ゆえに

	FIRST	nullable	FOLLOW
S'	\$ \ WORD begin end	no	
S	{ \ WORD begin end	yes	\$ } \
B	\	no	{ \ WORD begin end
E	\	no	\$ { } \ WORD begin end
X	{ \ WORD begin end	no	\$ { } \ WORD begin end

パース表は

	\	WORD	begin	end	\$	{	}
S'	0	0	0	0	0		
S	2,1	2	2	2			1
B	3						
E	4						
X	5,10	7	8	9		6	

衝突してますな。LL(2) にすると解消できるようだ。

3-6

$$1 S \rightarrow u B D z$$

$$2 B \rightarrow B v$$

$$3 B \rightarrow w$$

$$4 D \rightarrow E F$$

$$5 E \rightarrow y$$

$$6 E \rightarrow$$

$$7 F \rightarrow x$$

$$8 F \rightarrow$$

a. FIRST, nullable, FOLLOW 集合の計算

	FIRST	nullable	FOLLOW
S	u	no	
B	w	no	v z FIRST[D]
D	FIRST[E] FIRST[F]	yes	z
E	y	yes	FOLLOW[D] FIRST[F]
F	x	yes	FOLLOW[D]

より

	FIRST	nullable	FOLLOW
S	u	no	
B	w	no	v z y z
D	x y	yes	z
E	y	yes	x z
F	x	yes	z

b. パース表

	u	v	w	x	y	z
S	1					
B			2,3			
D				4	4	4
E				6	5	6
F				7		8

c. LL(1) でないことの説明

B が左再帰を含むため。

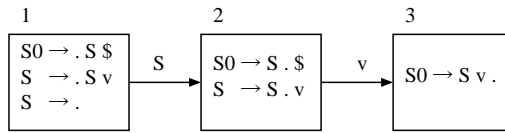


図 1: 3.8 左再帰の場合の状態遷移表

d. LL(1) にするための最小限の変更

$$1 S \rightarrow u B D z$$

$$3 B \rightarrow w B'$$

$$2 B' \rightarrow v B'$$

$$9 B' \rightarrow$$

$$4 D \rightarrow E F$$

$$5 E \rightarrow y$$

$$6 E \rightarrow$$

$$7 F \rightarrow x$$

$$8 F \rightarrow$$

確認してないけど多分これでいいと思う。

3-7 *

(略)

3-8 左再帰を含む tiny 文法を makeup し、LR では問題ないことをデモせよ。また、LR パーサスタックが、右再帰と左再帰でどう成長するか例で示せ。

$$1 S_0 \rightarrow S \$$$

$$2 S \rightarrow S v$$

$$3 S \rightarrow$$

そういえば LR の説明で ϵ が登場しなかったな。

	\$	v	S
1	r3	r3	g2
2	a	s3	
3	r2	r2	

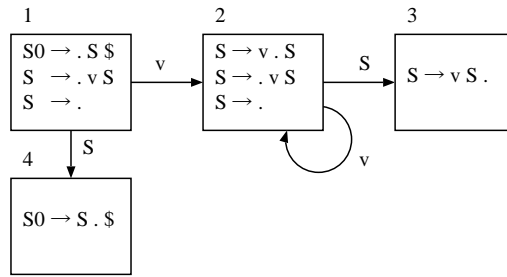


図 2: 3.8 右再帰の場合の状態遷移表

LR(0) ?

ついでに右再帰の場合

1 $S_0 \rightarrow S \$$

2 $S \rightarrow v S$

3 $S \rightarrow$

	\$	v	S
1	r3	s3	g2
2	r3	s2	g3
3	r2	r2	
4	a		

具体的な動作は書かないが、右再帰はスタックが伸びるので LR 法では左再帰で書くべし。

3-9 Grammar3.26 について LR(0) 状態を示し、SLR パース表を構築し衝突を示せ。

Grammar3.26:

0 $S' \rightarrow S \$$

1 $S \rightarrow V = E$

2 $S \rightarrow E$

3 $E \rightarrow V$

4 $V \rightarrow x$

5 $V \rightarrow * E$

,z 因子がないので Fig.3.27 から統合される部分が出てくるわけだ。conflict が起きているのは state 3 のところであり、LR(0) でないのは明らか。FOLLOW(E) を算出して = がなければ SLR になるのだが、

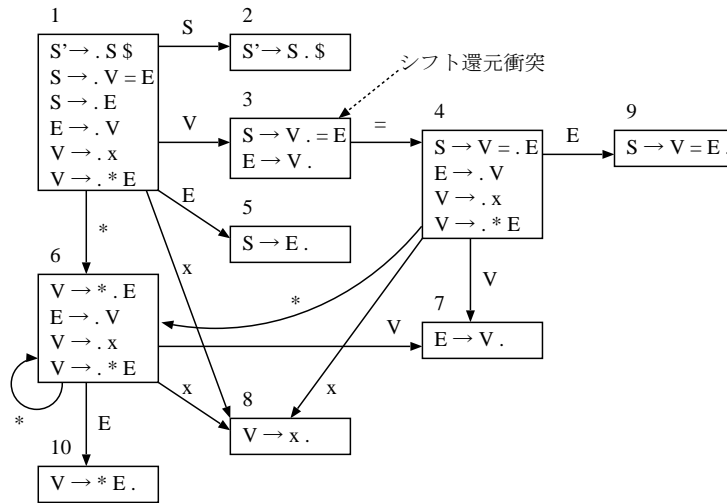


図 3: 3.9 LR(0)/SLR 状態遷移表

	FIRST	nullable	FOLLOW
S'	FIRST[S]	no	
S	FIRST[V][E]	no	\$
E	FIRST[V]	no	FOLLOW[S] FOLLOW[V]
V	x *	no	= FOLLOW[E]

より

	FIRST	nullable	FOLLOW
S'	x *	no	
S	x *	no	\$
E	x *	no	\$ =
V	x *	no	\$ =

と言うわけで FOLLOW(E) 集合に = が現れたので、残念ながら SLR ではありません。

3-10 Exercise 3.7 で LR(1) 状態を示し、パース表を構築し衝突があれば明らかにせよ。

- 0 $S \rightarrow G \$$
- 1 $G \rightarrow P$
- 2 $G \rightarrow P G$
- 3 $P \rightarrow id : R$
- 4 $R \rightarrow$
- 5 $R \rightarrow id R$

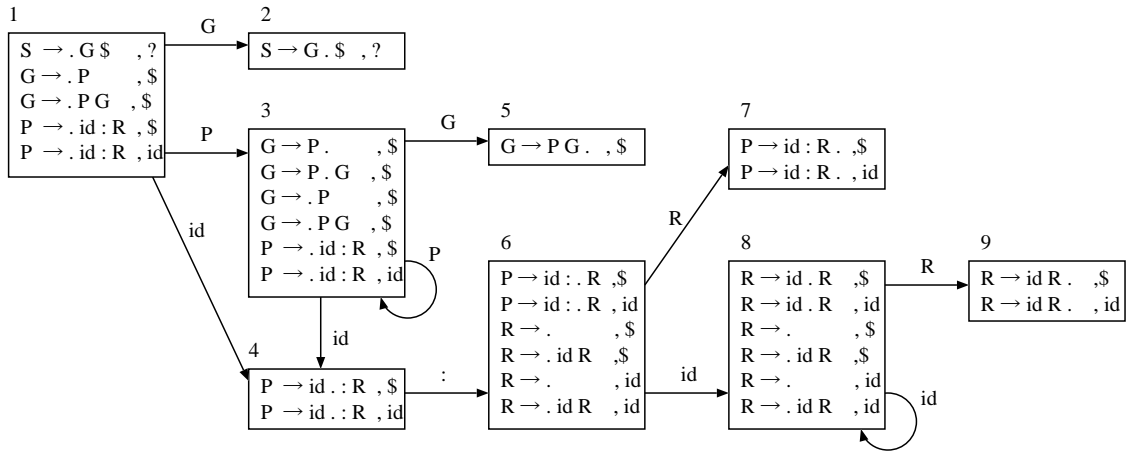


図 4: 3.10 LR(1) 状態遷移表

	id	:	\$	G	P	R
1	s4			g2	g3	
2			a			
3	s4		r1		g3	
4		s6				
5			r2			
6	s7,r4		r4			g7
7	r3		r3			
8	s8,r4		r4			g9
9	r5		r5			

衝突したので LR(1) ではないようだ。Left-Factor すると LL(2) すなわち LR(2) になるようだ。

3-11 以下の文法で LR(0) 状態を作り SLR 文法かどうかを判定せよ。

- 0 $S \rightarrow B \$$
- 1 $B \rightarrow id P$
- 2 $B \rightarrow id (E]$
- 3 $P \rightarrow$
- 4 $P \rightarrow (E)$
- 5 $E \rightarrow B$
- 6 $E \rightarrow B , E$

まず FOLLOW 集合を求めておく。

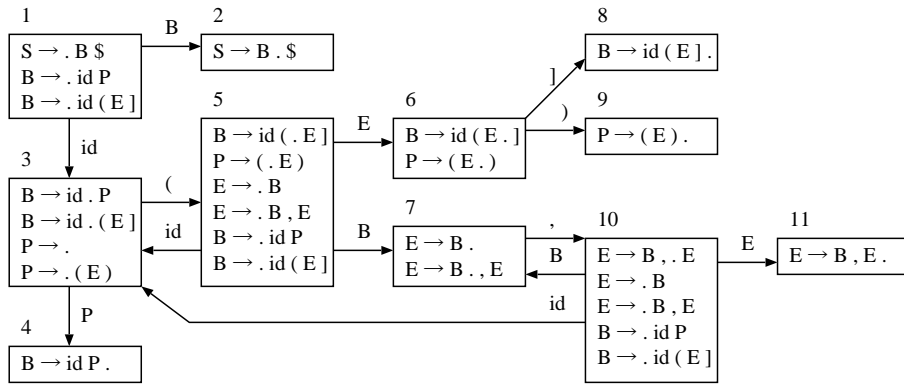


図 5: 3.1 LR(0)/SLR 状態遷移表

	FIRST	nullable	FOLLOW
S	FIRST[B]	no	
B	id	no	, \$ FOLLOW[E]
P	(yes	FOLLOW[B]
E	FIRST[B]	no)] FOLLOW[E]

ゆえに

	FIRST	nullable	FOLLOW
S	id	no	
B	id	no)] , \$
P	(yes)] , \$
E	id	no)]

SLR である。

	id	(])	,	\$	S	B	P	E
1	s3						g2		
2					a				
3		s5	r3	r3	r3			g4	
4			r1	r1	r1				
5	s3						g7		g6
6			s8	s9					
7			r5	r5	s10				
8			r2	r2	r2				
9			r4	r4	r4				
10	s3						g7		g11
11			r6	r6	r6				

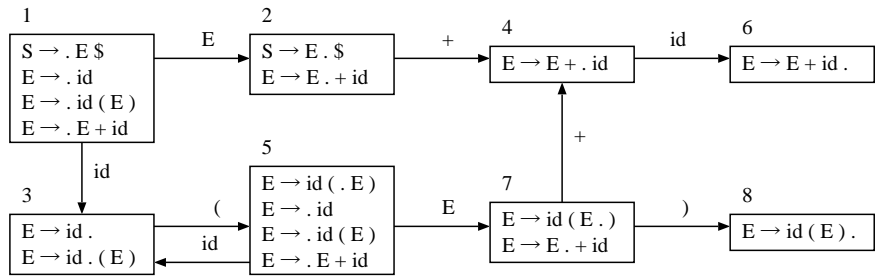


図 6: 3.12 LR(0)/SLR 状態遷移表

3-12

- 以下の文法で LR(0) DFA を構築せよ
- LR(0) かどうか説明せよ
- SLR かどうか説明せよ
- LR(1) かどうか説明せよ

- 0 $S \rightarrow E \$$
- 1 $E \rightarrow id$
- 2 $E \rightarrow id (E)$
- 3 $E \rightarrow E + id$

	FIRST	FOLLOW
E	id) + \$

SLR になってる? s3 は ok (E の FOLLOW 集合に "(" は含まれないから) だが。s2 の状態も ok。つまり SLR。

LR(1) だとどうなるか::

パース表は以下のとおり。

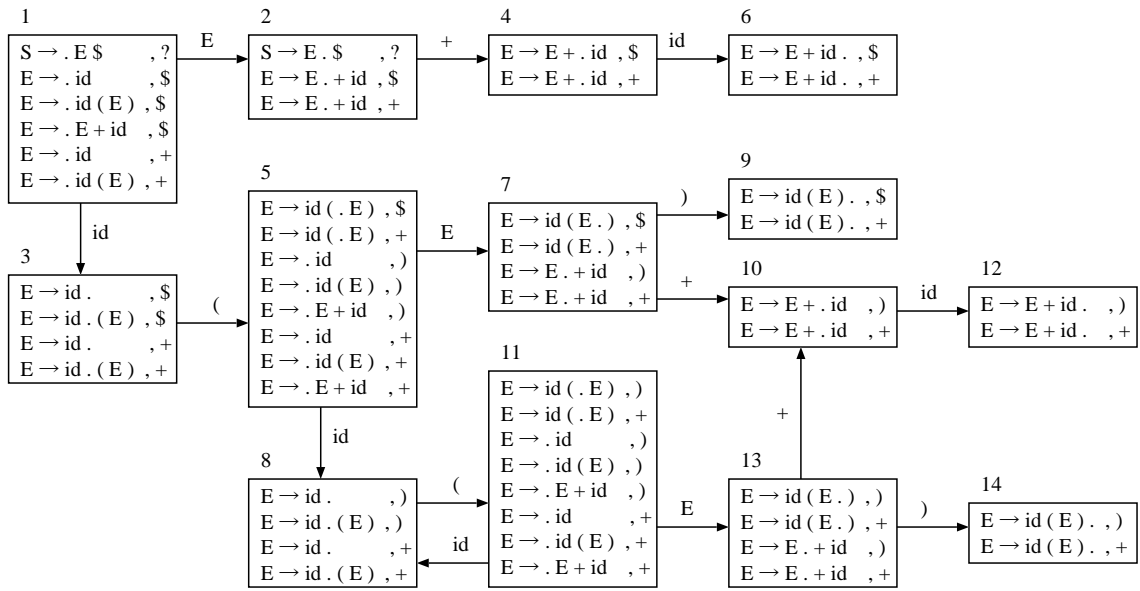


图 7: 3.12 LR(1) 状态转移表

	id	()	+	\$	S	E
1	s3						g2
2				s4	a		
3		s5	r1		r1		
4	s6						
5	s8						g7
6				r3	r3		
7			s9	s10			
8		s11	r1	r1			
9				r2	r2		
10	s12						
11	s8						g13
12			r3	r3			
13			s14	s10			
14			r2	r2			

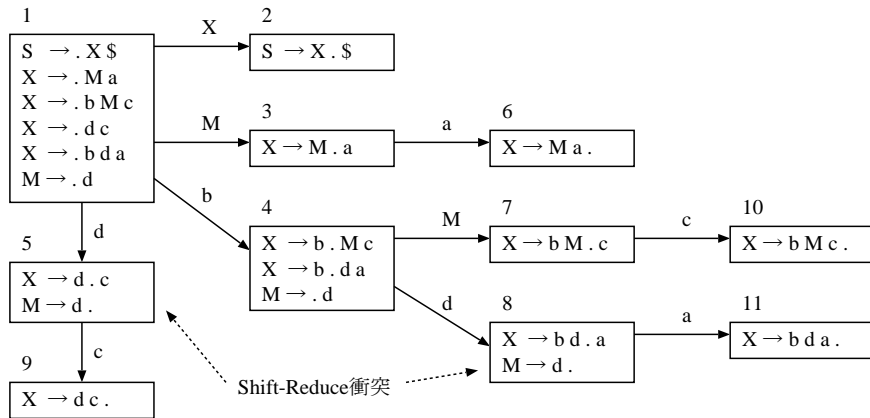


図 8: 3.13 LR(0)/SLR 状態遷移表

3-13 以下の文法が LALR(1) であって SLR でないことを示せ

- 0 $S \rightarrow X \$$
- 1 $X \rightarrow M a$
- 2 $X \rightarrow b M c$
- 3 $X \rightarrow d c$
- 4 $X \rightarrow b d a$
- 5 $M \rightarrow d$

まず SLR でないことの確認

	FIRST	FOLLOW
S	b d	
X	b d	\$
M	d	a c

状態 5 と状態 8 でシフト還元衝突が起きている。

次: LALR(1)

とりあえず LR(1) である。LALR(1) かな? LR(1) パース表を書く

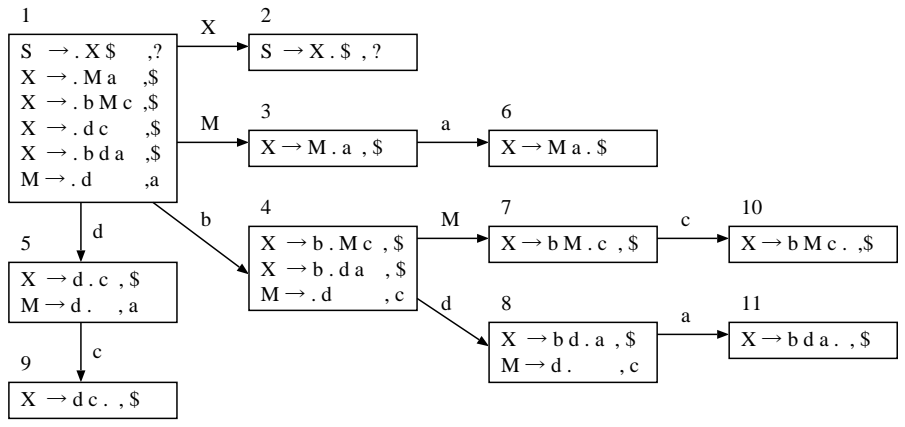


図 9: 3.13 LR(1)/LALR(1) 状態遷移表

	a	b	c	d	\$	S	X	M
1		s4		s5			g2	g3
2					a			
3	s6							
4				s8			g7	
5	r5		s9					
6					r1			
7			s10					
8	s11		r5					
9					r3			
10					r2			
11					r4			

LR(1) から縮退は起きず LALR(1) で同じパース表になる。そもそも状態遷移表の数が SLR と LR(1) で同じだったから当然だが。

3-14 以下の文法が LL(1) であって LALR(1) でないことを示せ

- (0 $S' \rightarrow S \$$)
- 1 $S \rightarrow (X$
- 2 $S \rightarrow E]$
- 3 $S \rightarrow F)$
- 4 $X \rightarrow E)$
- 5 $X \rightarrow F]$
- 6 $E \rightarrow A$
- 7 $F \rightarrow A$
- 8 $A \rightarrow$

	FIRST	nullable	FOLLOW
S'	FIRST[S]	n	
S	(FIRST[E] FIRST[F]	n	\$
X	FIRST[E] FIRST[F]	n	FOLLOW[S]
E	FIRST[A]	y)]
F	FIRST[A]	y)]
A		y	FOLLOW[E] FOLLOW[F]

より

	FIRST	nullable	FOLLOW
S'	(n	
S	()]	n	\$
X)]	n	\$
E		y)]
F		y)]
A		y)]

	()]
S'	0		
S	1	3	2
X		4	5
E		6	6
F		7	7
A		8	8

LL(1) である。

LR(1) ではある。しかし、s6 と s10 がマージされてしまう LALR では還元還元衝突が起きる。

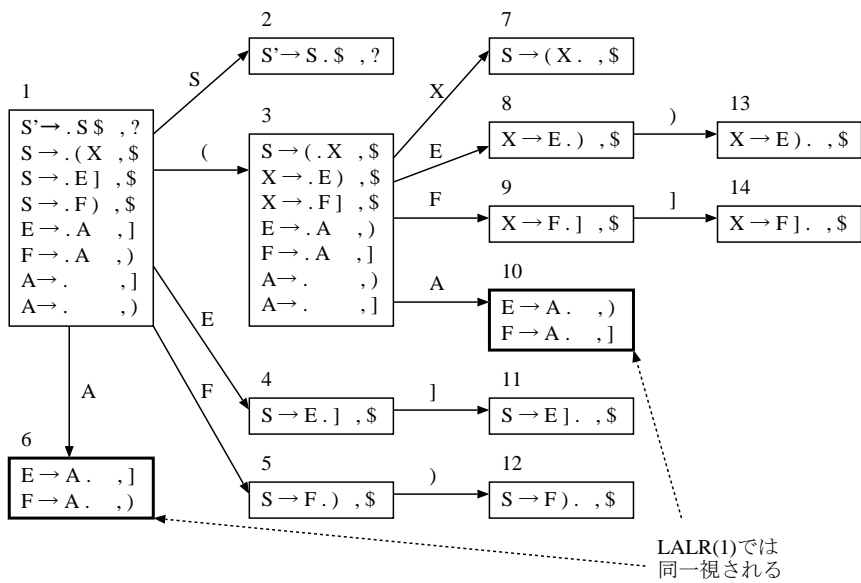


図 10: 3.14 LR(1)/LALR(1) 状態遷移表