コンパイラ

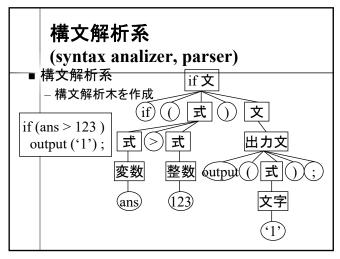
第11回 上昇型構文解析(1)

http://www.info.kindai.ac.jp/compiler E館3階E-331 内線5459 takasi-i@info.kindai.ac.jp コンパイラの構造

- 字句解析系
- 構文解析系
- 制約検査系
- 中間コード生成系
- 最適化系
- 目的コード生成系

1

2



構文解析 文法 G = {N, T, S, P} が与えられたとき。 ω ∈T*に対してS⇒ ω であるか判定, その導出木を得る 下降型解析 上昇型解析

3

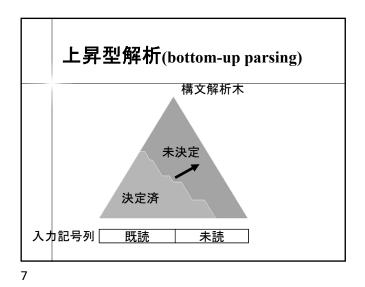
下降型解析(top-down parsing) 構文解析木 決定済 未決定 入力記号列「 既読 未読

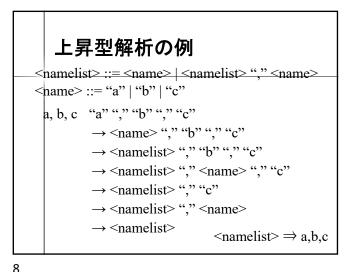
下降型解析の例

<namelist> ::= <name> | <name> "," <namelist> <name> ::= "a" | "b" | "c"

a, b, c < namelist>

- → <name> "," <namelist>
- → "a" "," <namelist>
- → "a" "," <name> "," <namelist>
- → "a" "," "b" "," <namelist>
- → "a" "," "b" "," <name> \rightarrow "a" "," "b" ", "c" \rightarrow "a" "," "b" "," "c" <namelist> \Rightarrow a,b,c





構文解析の種類
「情報システムプロジェクトI の構文解析
「下降型解析
(top-down parsing)

LL解析
(Left to right scan & Left most derivation)
演算子順位構文解析
(bottom-up parsing)

LR解析
(Left to right scan & Right most derivation)

再帰下降構文解析の欠点

■ 受理できる言語の範囲が狭い

- 受理できる原始言語の文法に制限がある
■ 左再帰性を除去すると構文木が変わる

- 演算子の結合性の情報が消失

⇒ 左再帰性の除去時に情報を記録

■ 再帰が必要

- 記述言語が再帰可能な場合のみ使用可能

⇒ スタックを用いる

10

9

LL解析

■ Left to right scan & Left most derivation 解析

- 下降型解析

- スタックと解析表を用いて解析

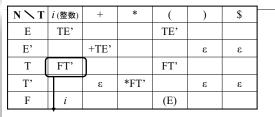
- 再帰下降型解析と本質的に同じ

LL解析 ファイル末 入力記号列 スタック 3 (| 6 | + 4 \$ \$ 開始記号 解析表 Ε N **\ T** | *i* (整数) + (\$ TE' TE' +TE' E, T FT' FT' *FT' F i (E)

11 12

解析表

■ 非終端記号×終端記号の表

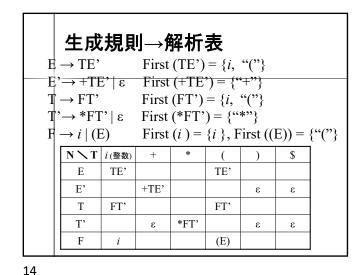


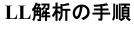
項目M [T, i]: 生成規則 $T \rightarrow FT'$

意味:非終端記号 Tの解析時に記号 i を読めば 次は FT'を解析する

空欄は構文解析エラー

13





X:スタックトップ a∈T:現在の入力記号

- X が終端記号のとき
 - $-X = a \Rightarrow X$ をポップ, 次の文字を読み込む
 - X ≠ a ⇒ 構文解析エラー



X





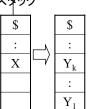
15

LL解析の手順

X: スタックトップ $a \in T:$ 現在の入力記号

- X が非終端記号のとき
 - 解析表M[X,a] = X →Y₁Y₂Y₃...Y_k \Rightarrow X をポップ, Y_1 , Y_2 , Y_3 , ..., Y_k をプッシュ

スタック



_____ 後ろから順にプッシュ

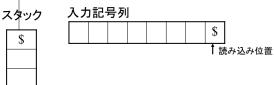
<u></u>								
	a							
	$Y_1Y_2Y_3Y_k$							
	<u>P</u> #	a						

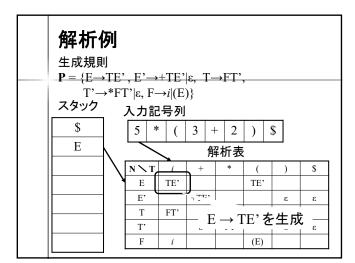
16

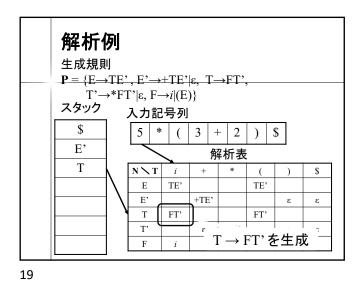
LL解析の手順

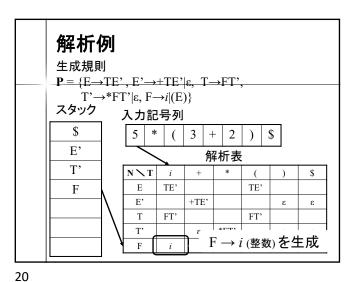
X:スタックトップ a∈T:現在の入力記号

- X = "\$" かつ a = "\$" のとき
 - 解析完了

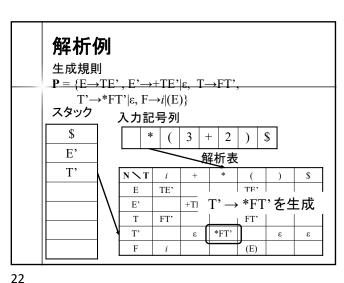




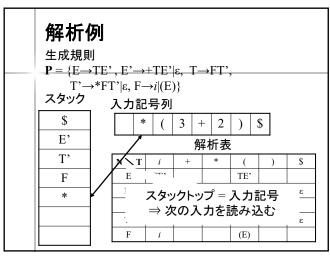


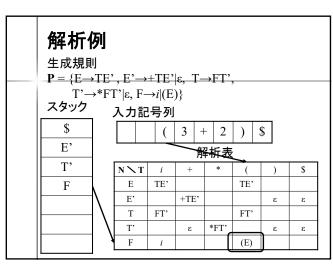


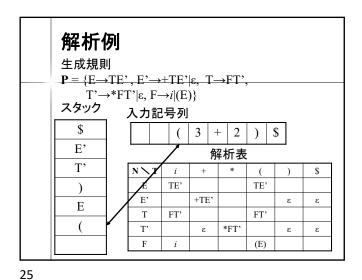
解析例 生成規則 $P = \{E \rightarrow TE', E' \rightarrow +TE' | \epsilon, T \rightarrow FT',$ $T' \rightarrow *FT' | \varepsilon, F \rightarrow i | (E)$ スタック 入力記号列 \$ 5 | * | (| 3 | + | 2 |) | \$ Ε, 解析表 T' N \ T + * \$ Е i(整数) スタックトップ = 入力記号 ⇒ 次の入力を読み込む i (E)

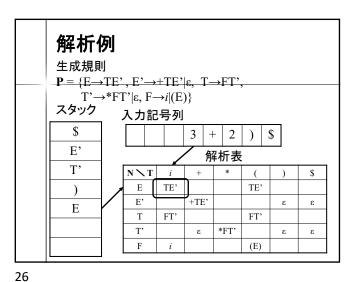


21

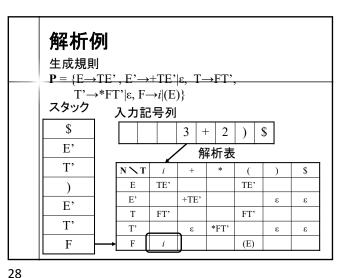




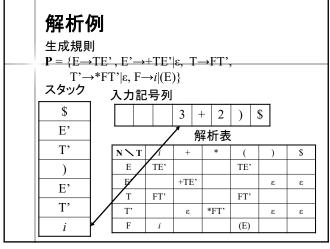


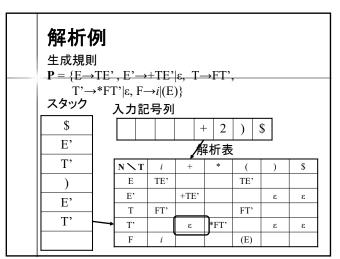


解析例 生成規則 $\mathbf{P} = \{E \rightarrow TE', E' \rightarrow +TE' | \varepsilon, T \rightarrow FT',$ $T' \rightarrow *FT' | \varepsilon, F \rightarrow i | (E)$ スタック 入力記号列 \$ 3 | + | 2 |) | \$ Ε, 解析表 T' $N \setminus T$ + \$ Е TE') E' +TE' 3 Ε, Т FT' T T' *FT' 3 F (E)

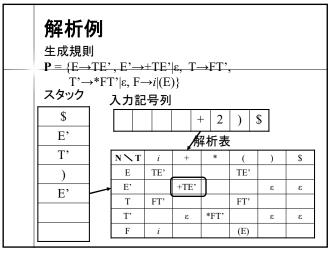


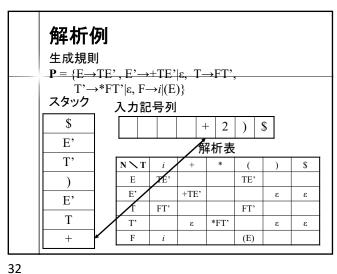
27

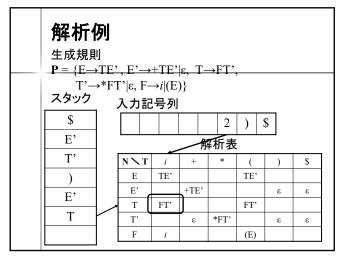


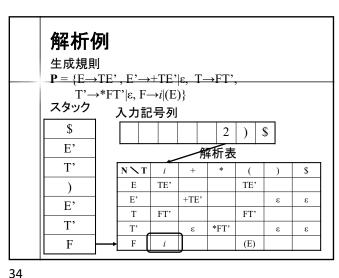


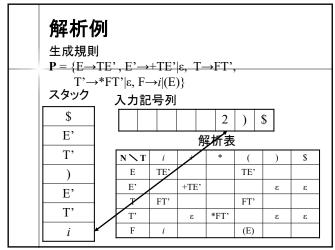
29 30

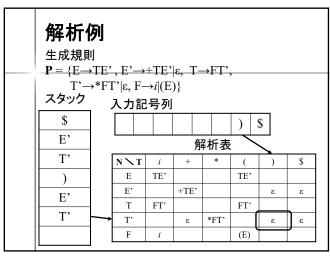




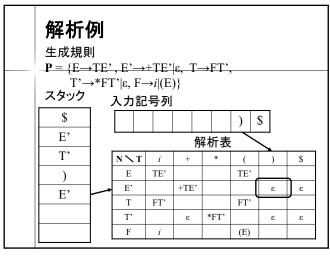


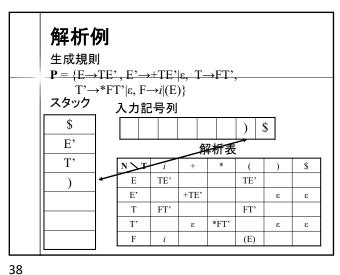


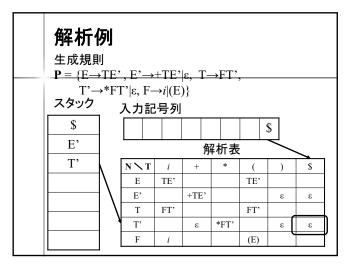


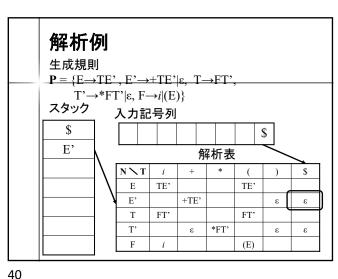


35 36





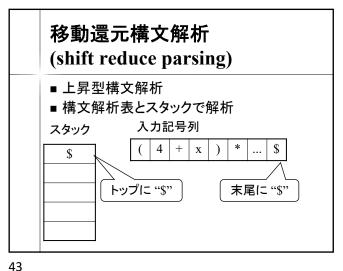




解析例 生成規則 ————————————————————————————————————		+TE'	ε, Τ-	→FT'	,		
T'→	*FΤ' ε, F-	<i>→i</i> (E)}				
スタック	入力記	号列					
\$ +					□	\$	
				- 1			
			角	解析表			
	$N \setminus T$	i	+	*	()	\$
	E	TE'			TE'		
	E'		+TE'			3	3
	T	FT'			FT'		
			3	*FT'		3	ε
	T'		"				

スタック	入力記号列	生成	スタック	入力記号列	生成
E\$	5*(3+2)\$		T'E')T'E'\$	+2)\$	
TE'\$	5*(3+2)\$	T→FT'	E')T'E'\$	+2)\$	E'→+TE'
FT'E'\$	5*(3+2)\$	F→i	+TE')T'E'\$	+2)\$	
iT'E'\$	5*(3+2)\$		TE')T'E'\$	2)\$	T→FT'
T'E'\$	*(3+2)\$	T'→*FT'	FT'E')T'E'\$	2)\$	F→i
*FT'E'\$	*(3+2)\$		iT'E')T'E'\$	2)\$	
FT'E'\$	(3+2)\$	F→(E)	T'E')T'E'\$)\$	T'→ε
(E)T'E'\$	(3+2)\$		E')T'E'\$)\$	Ε'→ε
E)T'E'\$	3+2)\$	Е→ТЕ')T'E'\$)\$	
TE')T'E'\$	3+2)\$	T→FT'	T'E'\$	\$	Τ'→ε
FT'E')T'E'\$	3+2)\$	F→i	E'\$	\$	Ε'→ε
iT'E')T'E'\$	3+2)\$		\$	\$	
			-		

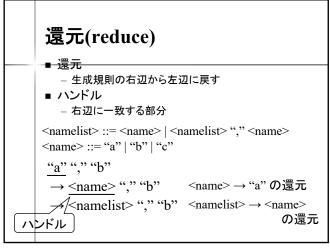
41 42



移動還元構文解析 (shift reduce parsing)

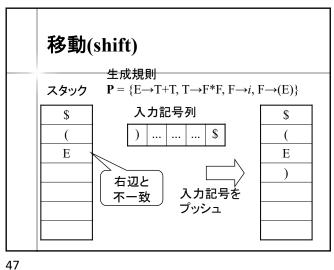
- ■構文解析表とスタックで解析
 - 初期状態
 - 入力記号列末尾に "\$"
 - スタックトップに "\$"
 - if (スタックトップが生成規則の右辺に一致) 生成規則の右辺をポップ, 左辺をプッシュ (還元) else 入力記号をプッシュ (移動)
 - if (スタックトップが開始記号 かつ 入力記号が "\$") 解析終了

44



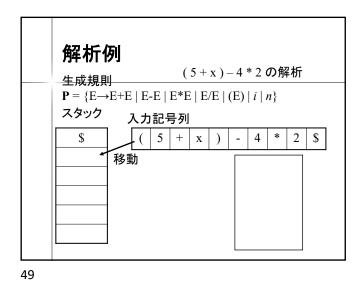
還元 生成規則 $\mathbf{P} = \{E \rightarrow T + T, T \rightarrow F * F, F \rightarrow i, F \rightarrow (E)\}$ スタック \$ 右辺 T+T をポップ ((左辺 E をプッシュ Τ $E \rightarrow T + T \mathcal{D}$ + 右辺と一致 T = ハンドル 対応する ADD コードを出力

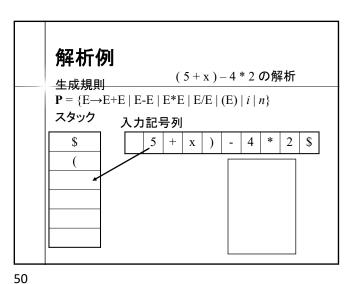
45 46

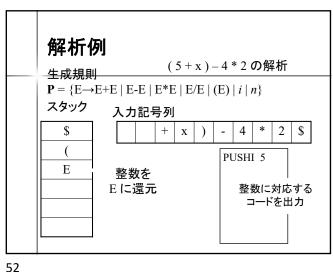


移動と還元 ■ 移動: 右辺を読み込み途中 ■ 還元:右辺を読み込み完了 例: $E \longrightarrow T + T$ 読み込み位置 ⇒ 読み込み位置を移動 $E \rightarrow T_{\blacktriangle} + T$ ⇒ 読み込み位置を移動 ↑読み込み位置 ⇒ 読み込み位置を移動 ⇒ 読み込み完了, 還元する

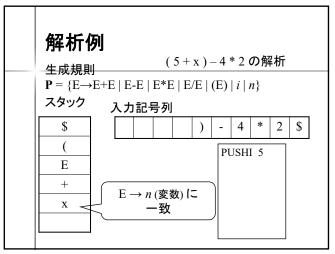
48



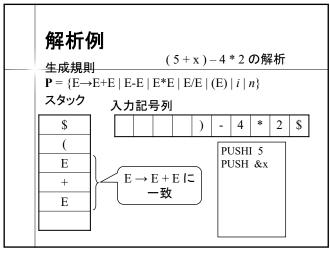


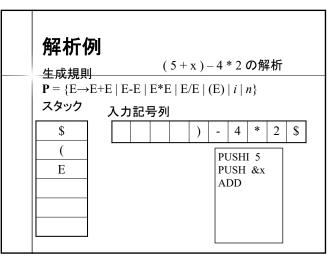






53 54





解析例

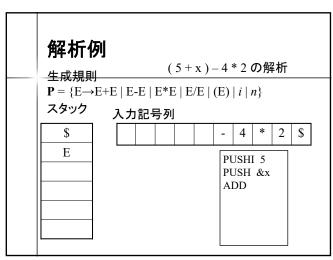
生成規則

P = {E→E+E | E-E | E*E | E/E | (E) | i | n}
スタック
入力記号列

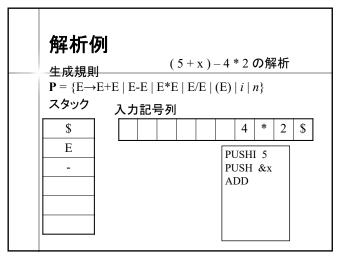
\$
(
E
)

E→(E) に
→致

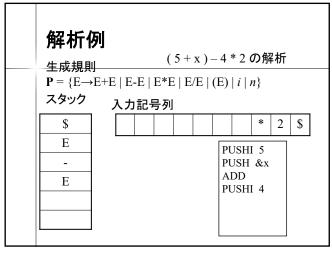
PUSHI 5
PUSH &x
ADD

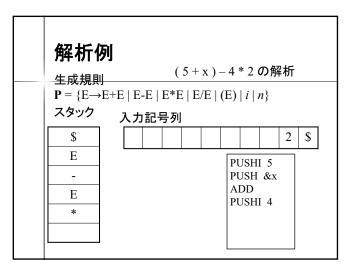


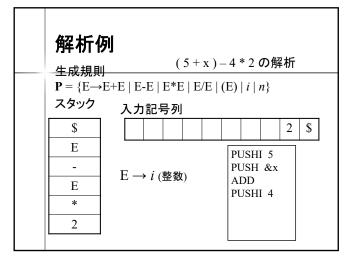
57 58

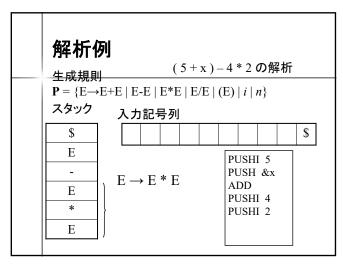




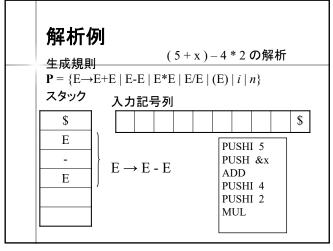


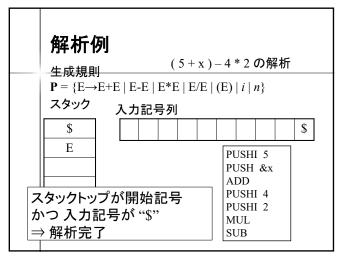






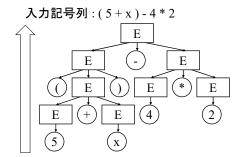
63 64





移動還元構文解析の導出木

■ 最右導出の導出木が生成される



スタック	入力列	操作	出力	
\$	(5+x)-4*2\$	移動		
\$(5 + x) - 4 * 2 \$	移動		
\$ (5	+ x) - 4 * 2 \$	還元 E → i (整数)	PUSHI 5	
\$ (E	+ x) - 4 * 2 \$	移動		
\$ (E +	x)-4*2\$	移動		
\$ (E + x) - 4 * 2 \$	還元 E → n (変数)	PUSH &x	
\$ (E + E) - 4 * 2 \$	還元 E → E + E	ADD	
\$ (E) - 4 * 2 \$	移動		
\$(E)	-4*2\$	還元 E → (E)		
\$ E	-4*2\$	移動		
\$ E -	4 * 2 \$	移動		
\$ E - 4	* 2 \$	還元 E → i (整数)	PUSHI 4	
\$ E - E	* 2 \$	移動 ここは	還元 E → E - E	
\$ E - E *	2 \$	移動で	はなく移動	
\$ E - E * 2	\$	還元 E → i (整数)	PUSHI 2	
\$ E - E * E	\$	還元 E → E * E	MUL	
\$ E - E	\$	還元 E → E - E	SUB	

演算子順位構文解析 (operator precedence parsing)

■ 演算子間の優先順位を定義

67

- A << B : A の優先順位 < B の優先順位

- A >> B: A の優先順位 > B の優先順位

- A == B: A と B は同じハンドル内

例:*>>+ *は+よりも優先順位が高い

例:(==) (と)同じハンドル内

記号の優先順位 (演算子と被演算子)

68

70

■ 記号 A が B よりも優先順位が高い = 逆ポーランド記法にしたときに A が先に来る

(x の優先順位) > (+ の優先順位)

(+の優先順位) < (5の優先順位)

x >> +, + << 5

被演算子と演算子とでは常に被演算子優先

69

記号の優先順位 (+と*)

 $\mathbf{M}: \mathbf{x} + \mathbf{y} * \mathbf{z}$

x, y, z, *, + + << *

例:x*y+z

x, y, *, z, + * >> +

*と+とでは常に*優先

記号の優先順位 (+同士)

例: x +₁ y +₂ z

 $x, y, +_1, z, +_2 +_1 >> +_2$

例:x+y-z

x, y, +, z, - + >> -

例:x-y+z

x, y, -, z, + ->>+

+ 同士, - 同士, + と - とでは 先に来た方が優先 (左結合的)

記号の優先順位 (=同士)

例: $X =_1 y =_2 Z$

 $x, y, z, =_2, =_1 =_1 << =_2$

= 同士では後に来た方優先 (右結合的)

記号の優先順位(\$)

例:\$(数式)\$

全ての処理が終われば\$を処理

⇒ \$ は優先順位最低

\$ << (全て), (全て) >> \$

\$ ** \$: \$同士ならば受理

73

記号の優先順位

例: (5+2)*(7-6)

5, 2, +7, 6, -, *

() がある場合は() 内を優先

(<<(全て)、(全て)<<() >> (全て), (全て) >>)

演算子の優先順位

■ 演算子 f, g

- fがgより優先順位が高い⇒f>>g,g<<f
- fとgの優先順位が同じ
 - ■f, g が左結合的 ⇒ f >> g, g >> f
 - ■f, g が右結合的 ⇒ f << g, g << f

75

76

優先順位表

右側 左側	整数	変数	+, -	*,/	=	()	\$
整数			>>	>>			>>	>>
変数			>>	>>	>>		>>	>>
+, -	<<	<<	>>	<<		<<	>>	>>
*,/	<<	<<	>>	>>		<<	>>	>>
	<<	<<	<<	<<	<<	<<	>>	>>
(<<	<<	<<	<<	<<	<<	==	
)			>>	>>			>>	>>
\$	<<	<<	<<	<<	<<	<<		**
'					- ウ堰	1十樓	ナ解れ	

解析手順

- 1. 入力記号列から非終端記号を取り除く
- 2. 入力記号列に優先順位を挿入する
- 3. 左から見て最初の >> を探す
- 4. 3.の位置から最も手前の << を探す
- 5. << から >> までを還元する

\$ ** \$ になれば受理

77

解析例

■ 入力列 \$ (5+x)-4*2\$

5 を E → *i* で還元

(E + x) - 4 * 2

解析例

優先順位判定では 非終端記号は削除

還元後の記号列: \$(E+x)-4*2\$

\$<<(<<+<<_X>>)>>-<<4>>*<<2>>\$

 $x を E \rightarrow n$ で還元

(E + E) - 4 * 2

\$<<(<<+>>)>>-<<4>>>*<<2>>\$

+ を E → E+E で還元 \$(E)-4*2\$

\$<u><<(==</u>)>>-<<4>>*<<2>>\$

== は同じハンドル内

79

80

解析例

還元後の記号列: \$(E)-4*2\$

\$<<(==)>>-<<4>>*<<2>>\$

()をE→(E)で還元 \$E-4*2\$

\$<<-<4>>>*<-2>>\$

4 を E → *i* で還元 \$ E - E * 2 \$

\$<<-<*<2>>\$

解析例

還元後の記号列: \$E-E*2\$

\$<<-<*><2>>\$

2 を E → *i* で還元 \$ E - E * E \$

\$<<-**<***>>\$

* を E → E*E で還元 \$ E - E \$

\$<<->>\$

81

82

解析例

還元後の記号列: \$ E - E \$

\$<<->>\$

- を E → E-E で還元 \$ E \$

\$**\$

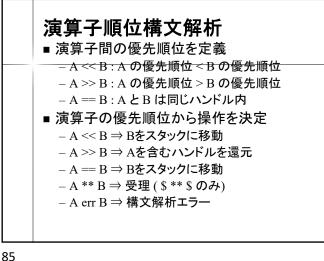
\$ ** \$ になったので受理

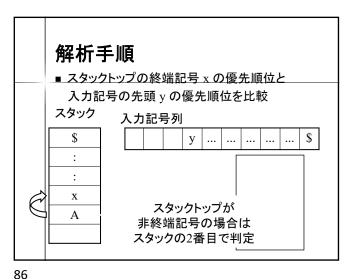
解析例

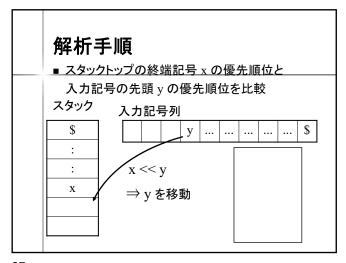
■ 入力列 \$ (5+x)-4*2\$

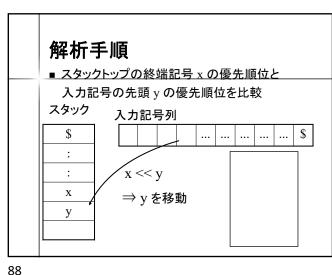
入力列	優先順位付記号列	還元	出力
\$ (<u>5</u> + x) - 4 * 2 \$	\$<<(<u><<5>></u> +< <x>>)>>-<<4>>>*<<2>>\$</x>	$E \rightarrow i$	PUSHI 5
\$ (E + <u>x</u>) - 4 * 2 \$	\$<<(<<+ <u><<x>></x></u>)>>-<<4>>>*<<2>>\$	$E \rightarrow n$	PUSH &x
\$ (<u>E + E</u>) - 4 * 2 \$	\$<<(<u><<+>></u>)>>-<<4>>*<<2>>\$	$E \rightarrow E + E$	ADD
\$ (E) - 4 * 2 \$	\$ <u><<(==)>></u> -<<4>>*<<2>>\$	$E \rightarrow (E)$	
\$ E - <u>4</u> * 2 \$	\$<<- <u><<4>>></u> *<<2>>\$	$E \rightarrow i$	PUSHI 4
\$ E - E * <u>2</u> \$	\$<<-<* <u><<2>></u> \$	$E \rightarrow i$	PUSHI 2
\$ E - <u>E * E</u> \$	\$<<- <u><<*>></u> \$	$E \rightarrow E * E$	MUL
\$ <u>E - E</u> \$	\$ <u><<->></u> \$	$E \rightarrow E - E$	SUB
<u>\$ E \$</u>	<u>\$**\$</u>		

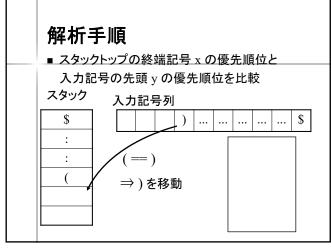
83

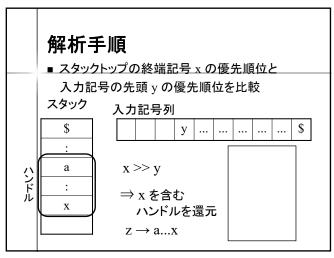


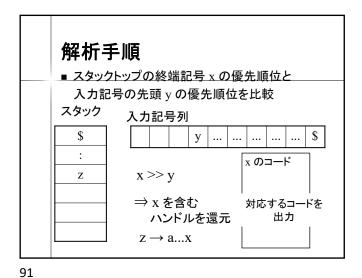


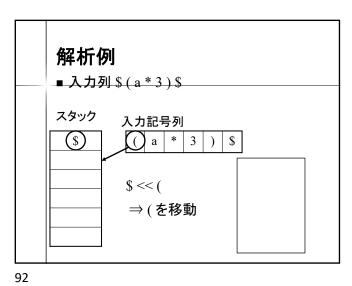




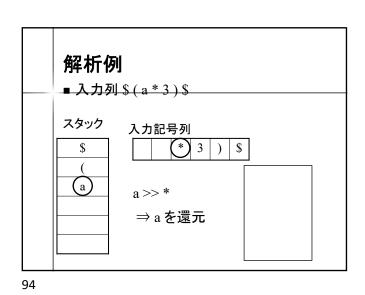




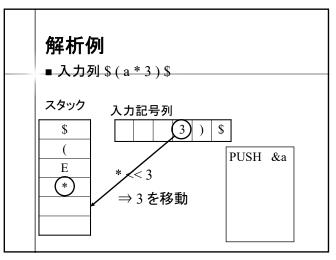




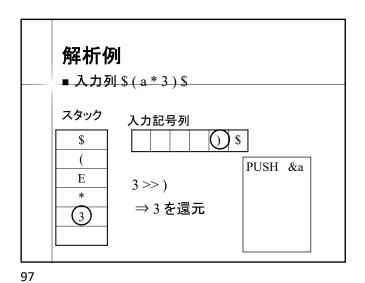
解析例 ■ 入力列 \$ (a*3)\$ スタック 入力記号列 (a) $\frac{|*|}{3|}$ (b) $\frac{|*|}{3|}$ ()(<< a ⇒ a を移動

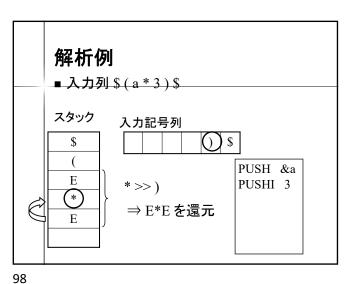


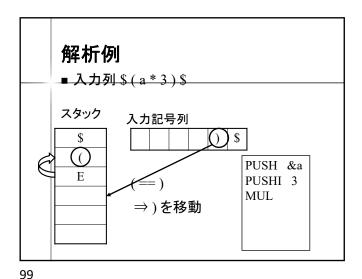
解析例 ■ 入力列 \$ (a*3)\$ スタック 入力記号列 (*) 3 |) | \$ ()PUSH &a (<< * ⇒*を移動 E は非終端記号 95

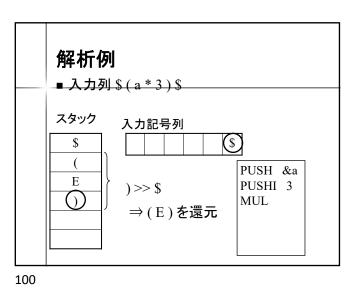


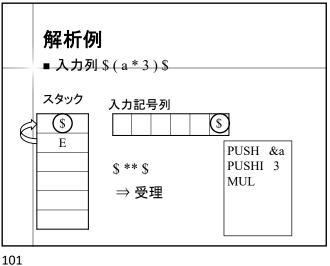
96











解	解析例									
■ ,	入力列 \$(a * 3) \$								
スタック	入力列	優先順位付記号列	判定	操作						
\$	(a*3)\$	\$<< <u>(</u> << a>>*<<3>>)>>\$	\$ << (移動						
\$ (a * 3)\$	\$<<(<< <u>a</u> >>>*<<3>>)>>\$	(<< a	移動						
\$ (a	* 3) \$	\$<<(< <a>><u>*</u><<3>>)>>\$	a >> *	還元						
\$(E *3)\$		\$<<(<< <u>*</u> <<3>>)>>\$	(<< *	移動						
\$(E*	3)\$	\$<<(<< *<< <u>3</u> >>)>>\$	* << 3	移動						
\$(E*3)\$	\$<<(<< *<<3>>)>>\$	x >>)	還元						
\$(E*E)\$	\$<<(<< *>>)>>\$	*>>)	還元						
\$ (E)\$	\$<<(== <u>)</u> >>\$	(==)	移動						
\$(E)	\$	\$<<(==)>> <u>\$</u>)>>\$	還元						
\$ E	\$	\$** <u>\$</u>	\$ ** \$	受理						

優先順位の数値化

- 優先順位には半順序関係が成立
 - 半順序関係
 - $\blacksquare A \ll B \land B \ll C \Rightarrow A \ll C$
 - ⇒各記号に整数値を割り当てる (優先順位が高い方に大きい数値)

例: "+"=3, "*"=5

		優り	<u> </u>	<u>位(</u>	<u>り数</u>	<u>値 (</u>	ے			
:	右側	∥ q(y)	整数	変数	+, -	*,/	=	()	\$
左	則	o(x)	7	_7_	_3	_5_	2	∞	_0_	-1
整	数	7			>>	>>			>>	>>
変	数	7			>>	>>	>>		>>	>>
+,	-	4	<<	~	>>	<<		<<	>>	>>
*,	/	6	<<	<<	>>	>>		<<	>>	>>
=		1	<<	<<	<<	<<	<<	<<	>>	>>
(0	<<	<<	<<	<<	<<	<<	==	
)		∞			>>	>>			>>	>>
\$		-1	<<	<<	<<	<<	<<	<<		**
	±									

解析手順

103

■ スタックトップの終端記号 x の優先順位 p(x) と 入力記号の先頭 y の優先順位 q(y) を比較

例: スタックトップ: + 入力記号の先頭:*

 $p(+) = 4, q(*) = 5 \implies + << *$

演算子順位構文解析の問題点

■ 複数の順位を持つ演算子の存在- 例: 2項演算子の-と単項演算子の-

* >> 2項演算子の - * << 単項演算子の -両者を区別する必要あり

下降型構文解析なら構文解析時に区別可能 上昇型構文解析では構文解析時に区別不可能 ⇒字句解析時に区別

105 106

演算子の区別

- <Exp> ::= <Term> { (+ |) <Term> } - 2項演算子の - は <Term> の後にのみ
 - ⇒ if (一つ前のトークンが <Term>の末尾) 2項演算子の -, else 単項演算子の -

<Term> の末尾:

NAME, INTEGER, "inputint" "]" ")" 等

演算子の区別

104

```
■ 字句解析プログラム(の一部) if (currentChar == "-") {
  if (lookAhead() == "=") {
     currentChar = nextChar();
    token = new Token (ASSIGNSUB);
  } else if (lookAhead == "-") {
                                        直前の
    currentChar = nextChar();
                                      トークンで
    token = new Token (DEC);
                                         判定
  } else if (prevToken == NAME
         | prevToken == INTEGER
         | prevToken == "inputint" ... ) {
    token = new Token (SUB);
                                   // 2項演算子
  } else token = new Token (CSIGN); // 単項演算子
```