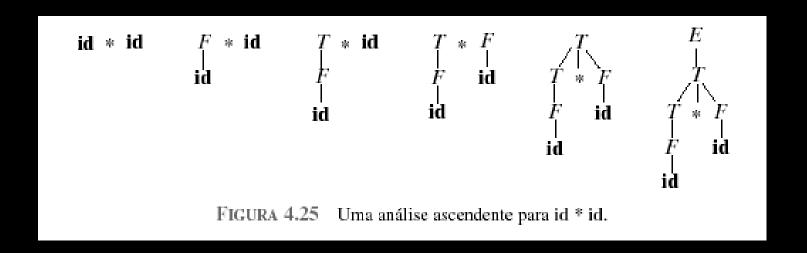
#### Construção da árvore a partir das folhas



# Objetivo - redução da cadeia ao símbolo inicial da gramática

(Empilhar-Reduzir) Shift-Reduce Parsing

Precedência Simples e Precedência de Operadores

Métodos LR

Passo: reduzir uma sub-cadeia a um não-terminal

Ex:

 $S \rightarrow aAcBe$ 

 $A \rightarrow Ab/b$ 

 $B \rightarrow d$ 

sent: abbcde

1. b por A ou d por B?

2. Ab por A ou b por A?

3. d por  $B \rightarrow aAcBe$ 

4. aAcBe por  $S \rightarrow S$ 

# Análise Sintática Ascendente ("Bottom-Up") <sub>Ex:</sub>

Derivação mais à direita:

 $S \rightarrow aAcBe$   $A \rightarrow Ab/b$  $B \rightarrow d$ 

sent: abbcde

$$S \rightarrow aAcBe \rightarrow aAcde \rightarrow aAbcde \rightarrow abbcde$$

#### Handle:

- lado direito de uma regra
- cuja redução representa um passo em direção à construção de uma derivação mais à direita.

Handle: definição - um handle de uma forma sentencial à direita, γ, consiste de:

- uma produção  $A \rightarrow \beta$
- uma posição de  $\gamma$  onde  $\beta$  pode ser encontrada e substituída por A, para produzir a forma sentencial anterior, numa derivação mais à direita de  $\gamma$

Isto é:

se S  $\rightarrow \alpha Aw \rightarrow \alpha \beta w$ , então

 $A \rightarrow \beta$ , na posição que segue  $\alpha$ , é um handle de  $\alpha\beta w$  [ $w \in V_T^*$ ]

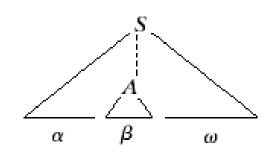


FIGURA 4.27 Um handle  $A \rightarrow \beta$  na árvore de derivação para  $\alpha\beta w$ .

no Ex1: handle

```
abbcde = A \rightarrow b (posição 2)

aAbcde = A \rightarrow Ab (posição 2)

aAcde = B \rightarrow d (posição 4)

aAcBe = S \rightarrow aAcBe (posição 1)
```

Forma Sentencial à Diretta	HANDLE	Produção de redução
$id_1 * id_2$	$id_1$	$F  o \mathrm{id}$
$F*id_2$	F	$T \rightarrow F$
$T*id_2$	$id_2$	$F  o \mathrm{id}$
T * F	T * F	$E \rightarrow T * F$

FIGURA 4.26 Handles durante a análise de id₁ \* id₂.

#### Análise Sintática Ascendente ("Bottom-Up") – Poda do Handle

Uma derivação mais à direita na ordem inversa pode ser obtida através da "poda dos handles".

Para isso, localizamos o handle  $\beta$ n em  $\gamma$ n e substituímos  $\beta$ n pelo lado direito de alguma produção An  $\rightarrow \beta$ n, de modo a obtermos a enésima menos uma forma sentencial à direita  $\gamma$ n-1.

#### Análise Sintática Ascendente ("Bottom-Up") – Poda do Handle

O processo continua até que seja encontrado o símbolo inicial da gramática.

```
Ex: considere a gramática E→E+E / E*E / (E) / id e a cadeia de entrada id1 + id2 * id3. A sequência de reduções é apresentada na tabela abaixo:
```

Analise Sintatica Ascendente				
("Bottom-Up") – Poda do Handle				
Forma sentencial à direita	Handle	Produção Redutora		

id1 + id2 \* id3 id1\_ 
$$E \rightarrow id$$

$$E + id2 * id3 id2 E \rightarrow id$$

$$E + id2 * id3$$
  $id2$   $E \rightarrow id$ 

$$E + E * id3$$
  $id3$   $E \rightarrow id$ 

$$E + E * E$$
 $E * E$ 
 $E \to E * E$ 

#### Análise Sintática Ascendente ("Bottom-Up") – Poda do Handle

Obs: as formas id1, id2, e id3 servem meramente como ilustração para indicar qual elemento está sendo substituído. Todos obedecem à regra  $E \rightarrow id$ .

Dois problemas precisam ser resolvidos se estivermos dispostos a analisar sintaticamente através da poda de handles.

- 1. localizar a subcadeia a ser reduzida numa forma sentencial à direita.
- 2. determinar que produção escolher no caso de existir mais de uma produção com aquela subcadeia no lado direito

#### Antes disso,

Vamos verificar os tipos de estruturas de dados usadas num analisador sintático de empilhar/reduzir

Uma forma conveniente é usar um pilha para guardar os símbolos gramaticais e um *buffer* de entrada para a cadeia w a ser decomposta.

Usamos \$ para marcar o fundo da pilha e também o final à direita da entrada. Inicialmente, a pilha está vazia e a cadeia w está como segue.

PILHA

\$

**ENTRADA** 

w\$

O analisador sintático opera empilhando zero ou mais símbolos até que um handle β surja no topo da pilha. Reduz então β para o lado esquerdo da produção apropriada.

Repete-se o ciclo até que tenha detectado um erro ou que a pilha contenha o símbolo de partida e a entrada esteja vazia

> PILHA ENTRADA \$S \$

Ex: Vamos rastrear as ações que um analisador sintático realizaria para decompor a cadeia **id1 + id2 \* id3**, de acordo com a gramática E→E+E / E\*E / (E) / id, em sua primeira derivação

	Pilha	INPUT	AÇÃO
1	\$	$id_1 + id_2 * id_3 $ \$	shift
2	\$ id <sub>1</sub>	$+ id_2 * id_3 $ \$	reduce $E \rightarrow id$
3	\$ E	$+ id_2 * id_3 $ \$	shift
4	\$ E +	id <sub>2</sub> * id <sub>3</sub> \$	shift
5	\$ E + id <sub>2</sub>	* id <sub>3</sub> \$	reduce $E \rightarrow id$
6	\$ E + E	* id <sub>3</sub> \$	shift
7	\$ E + E *	id <sub>3</sub> \$	shift
8	$$E + E * id_3$	\$	reduce $E \rightarrow id$
9	\$ E + E * E	\$	reduce $E \to E * E$
10	\$ E + E	\$	reduce $E \rightarrow E + E$
11	\$ E	\$	aceita

- 4 operações:
- SHIFT (empilhar) = próximo símbolo entrada é empilhado
- REDUCE (reduzir) = símbolo mais à
- direita de um handle no topo; deve
- localizar extremidade esquerda do handle e
- decidir sobre regra.
- ACEITAÇÃO = término da cadeia e pilha com S (símbolo inicial)
- ERRO = descobre erro sintático e chama rotina de recuperação

- Uso de uma pilha:
- Empilha (SHIFT) símbolos da cadeia até detectar o símbolo mais à direita de um handle;
- Desempilha o handle (REDUCE) e empilha o lado esquerdo da respectiva regra.

Repetir o ciclo até situação de erro ou terminar entrada e pilha = S

Um fato importante que justifica o uso de pilha numa análise sintática de empilhar e reduzir é que o handle irá sempre aparecer no topo da pilha, nunca dentro da mesma.

Como decidir sobre Shift ou Reduce?

Surgem dois tipos de conflitos:

- -conflito shift/reduce: o que fazer com o símbolo do topo da pilha?
- conflito reduce/shift: qual regra aplicar?

Exemplo de conflitos shift/reduce:

Cmd → if Expr then Cmd |
if Expr then Cmd else Cmd |

• • •

Pilha Input ....if Expr then Cmd else.....\$ shift ou reduce?

Obs: não podemos dizer se *if Expr then Cmd* é o handle.

Gramáticas ambiguas

Exemplo de conflito reduce/reduce

L.P. onde A(i,j) \_\_\_\_\_ array chamada de subprograma

Como resolver conflito reduce/reduce: considere hipótese: Análise Léxica retorna "id" para todo identificador

```
1. <cmd> ::= id (<param_list>)
```

- 2. <cmd> ::= <expr> := <expr>
- 3. <param\_list> ::= <param\_list> , <param>
- 4. <param\_list> ::= <param>
- 5. <param> ::= id

cont...

```
6. <expr> ::= id (<expr_list>)
7. <expr> ::= id
8. <expr_list> ::= <expr_list> , <expr>
9. <expr_list> ::= <expr>
```

**Problema:** Reduzir id2 via 5 (procedimento) ou 7 (array)?

- uma solução: mudar token id em 1 para procid — (como? mudando o anal. Léxico)

#### Exercício

- 1) Para a gramática S → 0S1 | 01, indique qual o handle em cada uma das formas sentenciais a direita
  - a) 000111
  - b) 00S11

#### Análise Sintática LR(k)

• O termo LR(k) é usado para representar uma análise cuja varredura de entrada acontece da esquerda para a direita (L = leftto right), com a construção de uma derivação mais à direita ao contrário (R = rightmost derivation), usando o número k de símbolos de entrada de lookahead que serão usados ao se tomar decisões na análise sintática.

### Análise Sintática LR(k) – Pontos que atraem esse tipo de análise

- os analisadores LR podem ser elaborados para reconhecer virtualmente todas as construções de linguagens de programação, que forem escritas em linguagens livres de contexto;
- método de decomposição LR é o mais geral de empilhar/reduzir.
- o analisador LR pode detectar um erro sintático tão cedo quanto possível numa varredura da esquerda para direita.

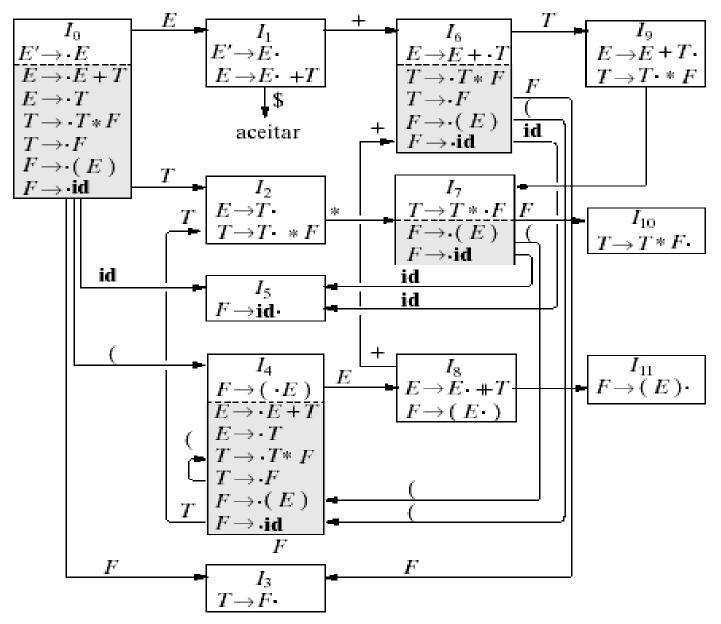


FIGURA 4.31 Autômato LR(0) para a gramática da expressão (4.1).

## Desvantagem Principal

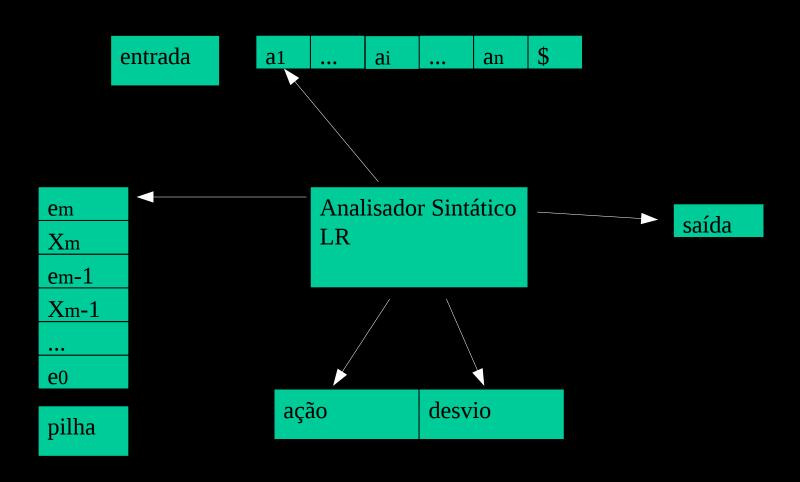
- é um processo muito trabalhoso construir um analisador sintático LR manualmente para uma linguagem de programação
- ESSE PROBLEMA é contornado com o uso de geradores de analisadores LR. ex: o Yacc, que veremos mais tarde!

### Tipos de analisadores LR

- SRL (Simple LR) fáceis de implementar, porém classe restrita de gramáticas
- LR canônicos mais poderosos, podendo ser aplicado a um grande número de linguagens livres de contexto
- LALR (look ahead LR) de nível intermediário e implementação eficiente. Funcionam com a maioria das linguagens de programação. O YACC gera esse tipo de analisador

Agora veremos o analisador SLR(1)

#### Modelo de um analisador sintático LR



• Na pilha, as cadeias são armazenadas na forma e0X1e1X2e2...Xmem, onde Xi é um símbolo gramatical e ei é um estado. Cada estado sumariza a informação contida na pilha abaixo do mesmo e a combinação do símbolo do estado no topo da pilha e o símbolo corrente de entrada é usada para indexar a tabela sintática e determinar ação empilhar.

### Tabela de Análise

• Tabela de transição de estados formada por duas partes:

Ação: **ei** (empilhar estado ei); **ri** (reduzir usando a i-ésima produção) e **a** (aceitar cadeia).

Transição: transição de estados com relação aos não terminais.

## Como construir as tabelas SLR(1)

E := +EE

E ::= \*EE

E ::= a

E := b

#### **Definições iniciais:**

• <u>item</u>: produção na qual foi marcada uma posição do lado direito. A posição é marcada por •

#### Exemplo:

```
E :: = \bullet + EE \mid + \bullet EE \mid + E \bullet E \mid + EE \bullet
```

## Como construir as tabelas SLR(1)

Os itens são usados para construir os estados. A presença, no topo da pilha, de um estado contendo um item da forma  $A := \alpha \bullet \beta$  indica que a parte inicial do handle (redutendo)  $\alpha\beta$ , que é  $\alpha$ , já foi processado e deslocado para a pilha. Num outro exemplo, se tivermos  $A := \alpha \bullet B\beta$ , onde A e B são não terminais e  $\alpha,\beta$ , qualquer símbolo do alfabeto  $(\Sigma^*)$ , sua presença, no topo da pilha indicará que já foi empilhada a cadeia  $\alpha$ , sendo esperado a seguir, a cadeia  $B\beta$ . Isso significa que a entrada corrente poderá ser da forma  $\gamma w$ , com  $\gamma \in \Sigma^*$  ( $\gamma$ , qualquer símbolo do alfabeto),  $w \in T^*$  (w qualquer terminal) e  $B\beta \Rightarrow \gamma$  ( $\gamma$  pode ser obtido a partir de zero ou mais derivações a partir de  $B\beta$ ).

Obs: intuitivamente, um item indica quanto de uma produção já foi examinado num determinado ponto da análise.

### Como construir a tabela de estados

- 1. Definir uma gramática aumentada
- 2. Definir duas funções: fechamento e desvio
  - Se G for uma gramática com símbolo de partida S, então G', a gramática aumentada para G. Ou seja, é G com um novo símbolo de partida S'.
  - A aceitação ocorre quando e somente quando o analisador sintático estiver para reduzir através de  $S' \rightarrow S$ .

#### Operação Fechamento

- Se I = conjunto de itens para G, então fechamento(I) é o conjunto de itens construídos a partir de I por 2 regras:
- cada item de I é adicionado ao fechamento
- se A  $\rightarrow \alpha \bullet B\beta$  estiver em fechamento(I) e B $\rightarrow \gamma$  for uma produção, adicionar o item B $\rightarrow \bullet \gamma$  a I, se já não estiver lá.
- Aplica-se essa regra até que não possam ser adicionados novos itens ao fechamento(I).

#### Operação Fechamento

• Intuitivamente,  $A \rightarrow \alpha \bullet B\beta$  em fechamento(I) indica que, em algum ponto do processo de análise gramatical, esperamos poder ver em seguida na entrada, uma subcadeia derivável a partir de Bβ. Se  $B \rightarrow \gamma$  for uma produção, também esperamos ver uma subcadeia derivável de γ naquele ponto. Por essa razão também incluímos  $B \rightarrow \bullet \gamma$  no fechamento(I).

Exemplo: Dada a gramática de expressões aumentadas:

- E' ::= E
- $E := E + T \mid T$
- T ::= T \* F | F
- F ::= (E) | id

Se I for um conjunto de um item { [E' ::= • E]}, então o fechamento(I) contém os itens:

```
E'::= •E (da primeira regra)
E::= •E + T
E::= • T * F
T::= • F
F::= • (E)
F::= • id
```

(da segunda regra)

- Pela regra 1 E' ::= •E é colocado em fechamento(I). Como existe um E imediatamente à direita do ponto, adicionamos, pela regra 2, as produções E ::= •E + T e E ::= T.
- Existe agora um T imediatamente à direita do ponto e, conseqüentemente, adicionamos F ::= • (E) e F ::= • id. A partir daí nenhum outro item é colocado no fechamento(I).

## Operação Desvio

 Desvio(I,X), onde I é o conjunto de itens e X é um símbolo gramatical é definida como o fechamento do conjunto de todos os itens [A →αX•β] tais que [A →α•Xβ] esteja em I.

• <u>Intuitivamente:</u> se I for o conjunto de itens válidos para algum prefixo viável γ, então desvio(I,X) será o conjunto de itens válidos para o prefixo viável γX.

```
exemplo: se I for o conjunto de dois itens \{[E'::=E \bullet], [E::=E \bullet + T]\}, então o desvio (I,+) consiste em:
```

```
E ::= E + • T
T ::= •T * F
T ::= E + T | T
F ::= • (E)
F ::= • id

E' ::= E

T ::= T * F | F

F ::= (E) | id
```

• Computamos o desvio(I,+) através do exame dos itens com + imediatamente à direita do ponto. E' ::= E • não é um desses itens, mas E ::= E • + T o é. Movemos, então, o ponto por sobre o + para obtermos { E ::= E + • T} e em seguida, realizamos o fechamento do conjunto.

### A construção de um conjunto de itens

• A construção partirá de um estado inicial, e, aplicando as funções fechamento e desvio, calculará novos estados, até que não seja possível prosseguir na construção. Uma vez que o número de itens é finito, o número de estados distintos também o é.

### A construção de um conjunto de itens

- **Convenção 1**: para a raiz S, acrescenta-se a produção S' ::= S#. O símbolo S' passa a ser a nova raiz.
- Convenção 2: toda cadeia de entrada será seguida do símbolo #.

Assim, o primeiro estado **0** deve conter o item S' ::= •S#, e se S' ::= S•# estiver no topo da pilha quando o símbolo seguinte a ser consultado é #, então a cadeia deverá ser aceita.

- Algoritmo para determinar os estados de uma gramática, onde C é uma coleção de estados, ou seja, um conjunto de conjuntos de itens:
- Adota-se o estado 0 = {fechamento {S' ::=
   S#})} como o valor inicial de C
- Se existe um estado e de C, e um símbolo X
   ∈ ∑ (qualquer elemento da gramática), tais
   que e' = desvio(e,X) não é vazio, e e' não
   está em C, então acrescenta-se e' à coleção
   C
- O passo2 é repetido até que não possam mais acrescentar estados à coleção C.
- C é a coleção de estados tipo LR (0) da gramática.

## • Exemplo: Considere a gramática aumentada

Aplicando-se as construções acima, obtém-se os seguintes estados

```
Io: E' ::= •E#
            E ::= •+EE | •*EE | •a | •b
goto(I_0,E) = I_1 = E' ::= E \bullet \#
goto(I_0,+) = I_2 = E := + \bullet EE \mid \bullet + EE \mid \bullet *EE \mid \bullet a \mid \bullet b
goto(I_0, *) = I_3 = E ::= * \bullet EE \mid \bullet + EE \mid \bullet * EE \mid \bullet a \mid \bullet b
goto(I_0,a) = I_4 = E ::= a \bullet
goto(I_0,b) = I_5 = E ::= b \bullet
goto(I_2,E) = I_6 = E := +E \bullet E \mid \bullet + EE \mid \bullet *EE \mid \bullet a \mid \bullet b
goto(I_3,E) = I_7 = E ::= *E \bullet E \mid \bullet + EE \mid \bullet *EE \mid \bullet a \mid \bullet b
goto(I_6,E) = I_8 = E := +EE \bullet
goto(I_7,E) = I_9 = E ::= *EE \bullet
```

- Os valores da função desvio fornecem as entradas ej da tabela. Conforme verificado acima existem estados com itens completos apenas (4,5,8 e 9) ou estados com itens incompletos apenas (0, 1, 2, 3, 6 e 7).
- Esses últimos (os incompletos) indicam que um estado correspondente ao próximo símbolo consultado deve ser deslocado na pilha.

- Os estados constituídos de itens completos indicam que os últimos estados empilhados correspondem a um handle (redutendo) e que, portanto, deve haver redução, independentemente do próximo símbolo.
- Se um estado que está no topo da pilha contém um único item da forma A ::= α •, então deve-se aplicar uma redução utilizando a produção A ::= α. Assim todas as entradas rj (j é o número da regra) são obtidas, e colocadas nas entradas dos Follow(A).

Veja na tabela:

r1 E ::= +EE

r2 E ::= \*EE

r3 E ::= a

$$goto(I_0,E) = I_1 = E' ::= E \bullet \#$$

$$goto(I_0,*) = I_3 = E ::= *\bullet EE \mid \bullet + EE \mid$$
 $\bullet * EE \mid \bullet a \mid \bullet b$ 

$$goto(I_0,a) = I_4 = E := a \bullet$$

$$goto(I_0,b) = I_5 = E ::= b \bullet$$

goto(I<sub>2</sub>,E) = I<sub>6</sub> = E ::= +E
$$\bullet$$
E |  $\bullet$ +EE |  $\bullet$ \*EE |  $\bullet$ a |  $\bullet$ b

$$goto(I3,E) = I7 = E ::= *E \bullet E \mid \bullet + EE \mid \\ \bullet *EE \mid \bullet a \mid \bullet b$$

$$goto(I_6,E) = I_8 = E ::= +EE \bullet$$

$$goto(I_7,E) = I_9 = E ::= *EE \bullet$$

As reduções r = vão nas entradas Follow(X)-  $Follow(E) = \{\#, +, *, a, b\}$ 

acao			~	
	D		ם	$\cap$
ucuc	α	L	a	U

#### transição

						uunsigu
	+	*	a	b	#	Е
0	e2	e3	e4	e5		1
1					a	
2	e2	e3	e4	e5		6
3	e2	e3	e4	e5		7
4	r3	r3	r3	r3	r3	
5	r4	r4	r4	r4	r4	
6	e2	e3	e4	e5		8
7	e2	e3	e4	e5		9
8	r1	r1	r1	r1	r1	
9	r2	r2	r2	r2	r2	

Passo	Pilha	Símbolo Reduzido	cadeia de entrada	ação
0	0		+ * a + b a a #	e2
1	0 + 2		* a + b a a #	e3
2	0 + 2 * 3		a + b a a #	e4
3	0 + 2 * 3 a 4		+ b a a #	r3
4	0 + 2 * 3	E	+ b a a #	e7
5	0 + 2 * 3 E 7		+ b a a #	e2
6	0 + 2 * 3 E 7 + 2		b a a #	e5
7	0 + 2 * 3 E 7 + 2 b 5		a a #	r4
8	0 + 2 * 3 E 7 + 2	E	a a #	e6
9	0 + 2 * 3 E 7 + 2 E 6		a a #	e4
10	0 + 2 * 3 E 7 + 2 E 6 a 4		a#	r3
11	0 + 2 * 3 E 7 + 2 E 6	Е	a#	e8
12	0 + 2 * 3 E 7 + 2 E 6 E 8		a#	r1
13	0 + 2 * 3 E 7	Е	a#	e9
14	0 + 2 * 3 E 7 E 9		a#	r2
15	0 + 2	Е	a#	e6
16	0 + 2 E 6		a #	e4
17	0 + 2 E 6 a 4		#	r3
18	0 + 2 E 6	Е	#	e8
19	0 + 2 E 6 E 8		#	r1
20	0	E	#	e1

#### Exercício

1) Para a gramática abaixo, crie o conjunto de itens e a tabela de analise SRL e os passos do analisador para reconhecer a sentença

id & id v id

 $E \rightarrow E V T$ 

 $E \rightarrow T$ 

 $T \rightarrow T \& F$ 

 $T \rightarrow F$ 

 $F \rightarrow (E)$ 

 $F \rightarrow id$