# Compiladores Analise Sintática First Follow.

# Na aula de hoje...

- ☐ Revisão: gramáticas
- □ Relações em uma gramática: Cabeça, Último, Primeiro (*First*) e Seguinte (*Follow*)
  - ☐ Capítulo 4 (seção 4.4.2) do livro *Compiladores : Princípios, técnicas e ferramentas*

- Alfabeto ou vocabulário: Conjunto finito não vazio de símbolos.
  Símbolo é um elemento qualquer de um alfabeto.
- ☐ Cadeia: Concatenação de símbolos de um alfabeto. Define-se como cadeia vazia ou nula uma cadeia que não contém nenhum símbolo.
  - > Ex:
    - > aab
    - ▶ 123094
    - λ cadeia nula

- Comprimento de cadeia: Número de símbolos de uma cadeia. Ex:
  - $\rightarrow$  |aab| = 3
  - ▶ |123094|=6
  - $\rightarrow |\lambda| = 0$

Linguagem é uma coleção de cadeias de símbolos, de comprimento finito. Estas cadeias são denominadas sentenças da linguagem, e são formadas pela justaposição de elementos individuais, que são os símbolos ou átomos da linguagem.

# ☐ Métodos de Representação de Linguagens

- 1) Enumeração (especificação) das cadeias de símbolos que formam as suas sentenças (somente linguagens finitas podem ser representadas por este método)
- 2) Através de um conjunto de **leis de formação** das cadeias (**Gramática**)
- 3) Através de regras de aceitação de cadeias (às regras de aceitação dá-se o nome de **Reconhecedor -** autômatos)

# 2) Leis de Formação

- □ Através de um conjunto de leis de formação das cadeias (ao conjunto de leis de formação dá-se o nome de **Gramática**)
  - ➢ dada uma cadeia de símbolos, só é possível afirmar que tal cadeia pertence à linguagem se for possível, aplicando-se as leis de formação que compõem a gramática da linguagem, derivar (sintetizar) a cadeia em questão.
  - Ao processo de obtenção de uma sentença a partir da gramática dá-se o nome de derivação da sentença.

## □ Gramáticas

Formalmente as gramáticas, são caracterizadas como quádruplas ordenadas

$$G = (Vn, Vt, P, S)$$

- > onde:
  - ✓ Vn representa o vocabulário não terminal (variáveis) da gramática. Este vocabulário corresponde ao conjunto de todos os símbolos dos quais a gramática se vale para definir as leis de formação das sentenças da linguagem.

## □ Gramáticas

Formalmente as gramáticas, são caracterizadas como quádruplas ordenadas

$$G = (Vn, Vt, P, S)$$

> onde:

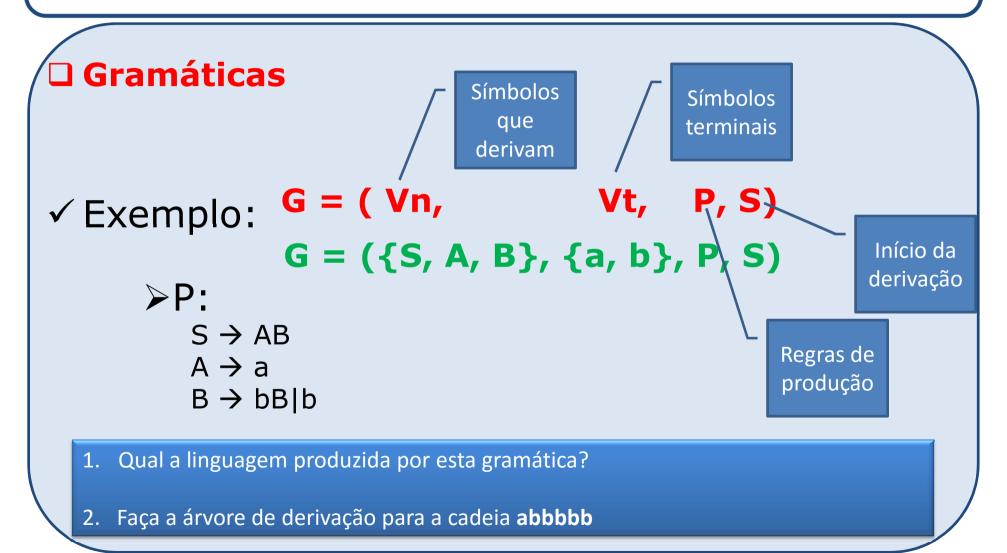
✓ Vt é o vocabulário terminal, contendo os símbolos que constituem as sentenças da linguagem. Dá-se o nome de terminais aos elementos de Vt.

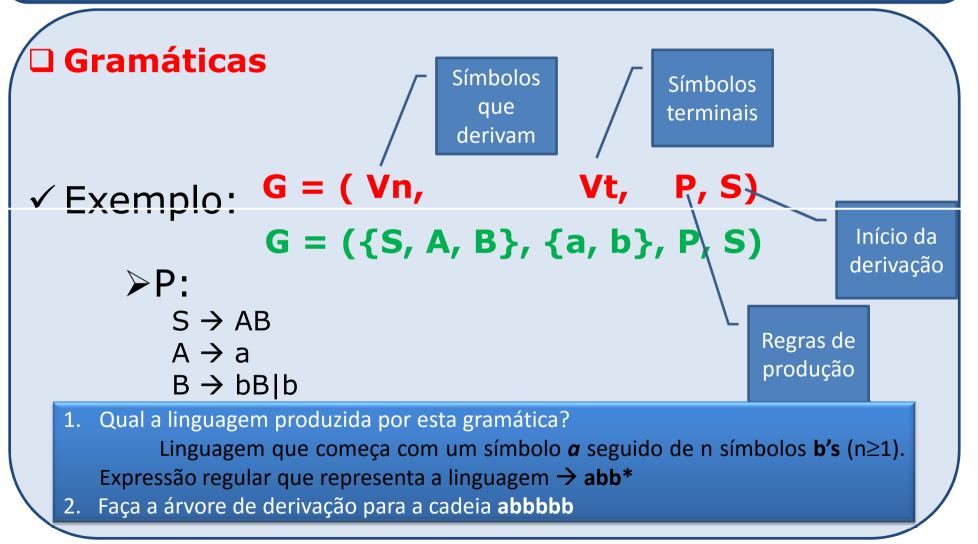
#### ☐ Gramáticas

> Formalmente as gramáticas, são caracterizadas como quádruplas ordenadas

$$G = (Vn, Vt, P, S)$$

- > onde:
  - ✓ P são as regras de produção, que definem o conjunto de todas as leis de formação utilizadas pela gramática para definir a linguagem. Para tanto, cada construção parcial, representada por um não-terminal, é definida como um conjunto de regras de formação relativas à definição do não-terminal a ela referente. A cada uma destas regras de formação que compõem o conjunto P dá-se o nome de produção da gramática





# ☐ Árvore de derivação

- ➤ a raiz tem como rótulo o símbolo inicial S da gramática.
- >a cada nó rotulado por um **não-terminal** A corresponde uma regra de A. Se a regra for A  $\rightarrow$   $X_1X_2$  ...  $X_m$ , os filhos do nó são rotulados, da esquerda para a direita, por  $X_1$ ,  $X_2$ , ...,  $X_m$ . (cada um dos  $X_i$  pode ser um terminal ou um não-terminal.)
- um nó rotulado por um terminal é sempre uma folha da árvore, e não tem filhos.

# Arvore de derivação para a gramática e cadeia

abbbbb

LHS	RHS	
S	$\rightarrow$ AB	S
A	$\rightarrow$ a	
В	$\rightarrow$ bB	В
В	$\rightarrow$ b	b
		b B
		B B
		a b

 Duas gramáticas G1 e G2 são equivalentes se produzem a mesma linguagem

$$L(G1) = L(G2)$$

- 2. Uma sentença é **ambígua** se existem duas ou mais sequências de derivação que a define.
- 3. Uma gramática é **ambígua** se possui alguma sentença ambígua.

```
G: S → AB

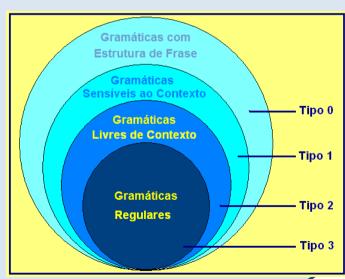
A → AA | B | a

B → Bcd | A

- Verifique se x= aaacd é ambígua.
```

É ambígua > permite derivação mais à direita e mais à esquerda

- Conforme as restrições impostas ao formato das produções de uma gramática, a classe de linguagens que tal gramática gera varia correspondentemente.
- □ A teoria mostra que há quatro classes de gramáticas capazes de gerar quatro classes correspondentes de linguagens, de acordo com a denominada *Hierarquia* de *Chomsky*.
  - a. Gramáticas com Estrutura de Frase ou Tipo 0
  - b. Gramáticas Sensíveis ao Contexto ou Tipo 1
  - c. Gramáticas Livres de Contexto ou Tipo 2
  - d. Gramáticas Regulares ou Tipo 3



#### a. Gramáticas com Estrutura de Frase ou Tipo 0

- ☐ São aquelas às quais nenhuma restrição é imposta.
  - ☐ Exemplo de reconhecedor: Máquina de Turing com fita de entrada infinita

□ Produções da forma

$$\alpha \rightarrow \beta$$

Onde: 
$$\alpha \in (Vn \cup Vt)^+$$
  
 $\beta \in (Vn \cup Vt)^*$ 

$$G = (Vn, Vt, P, S)$$

- □ Lado esquerdo da regra de produção pode conter N símbolos (terminais ou não terminais);
- □ Lado direito da regra de produção pode conter N símbolos (terminais ou não terminais ou vazio);

## a. Gramáticas com Estrutura de Frase ou Tipo 0

☐ Exemplo de GEF:

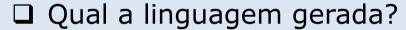
$$G = (\{A, B, C\}, \{a, b\}, P, A)$$

P:  $A \rightarrow BC$ 

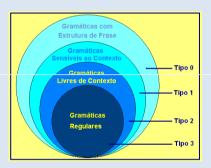
$$BC \rightarrow CB$$

 $B \rightarrow b$ 

 $C \rightarrow a$ 



$$\Box$$
 L(G) = ?



## a. Gramáticas com Estrutura de Frase ou Tipo 0

☐ Exemplo de GEF:

$$G = ({A, B, C}, {a, b}, P, A)$$

P:  $A \rightarrow BC$ 

$$B \rightarrow b$$

$$C \rightarrow a$$

- ☐ Qual a linguagem gerada?
  - $\Box$  L(G) = {ba, ab}

#### b. Gramáticas Sensíveis ao Contexto ou Tipo 1

Restrição: nenhuma substituição pode reduzir o comprimento da forma sentencial à qual a substituição é aplicada.

Compiladores

☐ Produções da forma

```
\alpha \rightarrow \beta
Onde: \alpha \in (Vn \cup Vt)^+
\beta \in (Vn \cup Vt)^+
|\alpha| \leq |\beta|
```

#### b. Gramáticas Sensíveis ao Contexto ou Tipo 1

☐ Exemplo de GSC:

$$G = ({S, B, C}, {a, b, c}, P, S)$$

P:  $S \rightarrow aSBC \mid aBC$ 

 $CB \rightarrow HB$ 

 $HB \rightarrow HC$ 

 $HC \rightarrow BC$ 

 $aB \rightarrow ab$ 

 $bB \rightarrow bb$ 

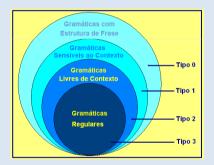
 $bC \rightarrow bc$ 

 $cC \rightarrow cc$ 

□ Qual a linguagem gerada?

```
\Box L(G) = \{a^nb^nc^n|n\geq 1\}
```

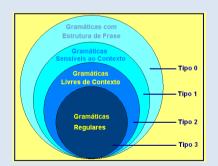
 $\alpha \rightarrow \beta$ Onde:  $\alpha \in (Vn \cup Vt)^+$   $\beta \in (Vn \cup Vt)^+$   $|\alpha| \leq |\beta|$ 



Faça a derivação (mais à esquerda ou mais à direita)

## c. Gramáticas Livres de Contexto ou Tipo 2

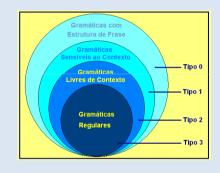
☐ As Gramáticas Livres de Contexto (GLC) ou do Tipo 2 são aquelas que no lado esquerdo da regra há apenas um símbolo não-terminal.



## c. Gramáticas Livres de Contexto ou Tipo 2

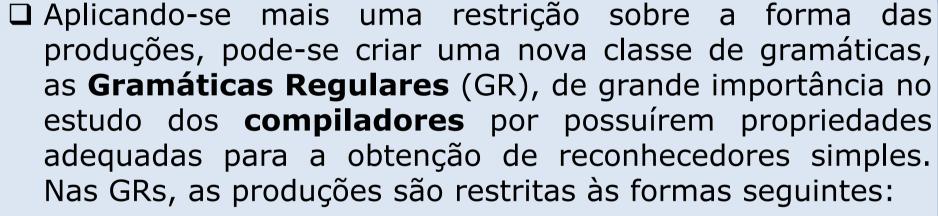
□ Qual a linguagem gerada para:

G = ({S,A,B},{a,b},P,S)  
P: S 
$$\rightarrow$$
 AB  
A  $\rightarrow$  aA | a  
B  $\rightarrow$  bB | b



$$L(G)=\{a^nb^m | n, m \ge 1\}$$

### d. Gramáticas Regulares ou Tipo 3



$$A \rightarrow aB$$
 ou  $A \rightarrow Ba$ 

 $A \rightarrow a$ 

 $A \rightarrow \epsilon$ 

onde  $A,B \in Vn \ e \ a \in Vt$ 

no lado esquerdo da regra há apenas um símbolo nãoterminal

#### d. Gramáticas Regulares ou Tipo 3

☐ Exemplo GR em EBNF:

```
G = ( {<Dig>, <Int>}, {+, -, 0, ..., 9}, P, <Int>)
```

P: <Int> ::= +<Dig> | -<Dig>

<Dig> ::= 0<Dig> | 1<Dig> | ... | 9<Dig> | 0 | 1 | 2 |... | 9

☐ Qual linguagem gerada?

 $\triangleright$  L(G) = conj. números inteiros com sinal ±[0..9]

# Na aula de hoje...

- ☐ Revisão: gramáticas
- □ Relações em uma gramática: Cabeça, Último, Primeiro (*First*) e Seguinte (*Follow*)
  - □ Capítulo 4 (seção 4.4.2) do livro *Compiladores : Princípios, técnicas e ferramentas*

# Relações em uma Gramática

- A construção de analisadores sintáticos é facilitada através de algumas funções associadas a gramática: CABEÇA, ÚLTIMO, PRIMEIRO(First) e Seguinte (Follow).
  - ➤ Ideia: permitem escolher qual produção aplicar com base no próximo símbolo lido.

# Cabeça

- É uma das mais simples de identificar
  - um de seus elementos é a cabeça do lado direito de uma regra
- □ Dada uma produção na forma

$$\alpha \rightarrow \beta \gamma$$

Onde:  $\alpha \in \mathbf{Vn}$ 

$$\beta \in (Vn \cup Vt)$$

$$\gamma \in (Vn \cup Vt)^*$$

 $\Box$  Cabeça ( $\alpha$ ) =  $\beta$ 

$$B \rightarrow bB \mid b$$

# Cabeça

- É uma das mais simples de identificar
  - ☐ um de seus elementos (terminal ou não-terminal) é a cabeça

do lado direito de uma regra

□ Dada uma produção na forma

$$\alpha \rightarrow \beta \gamma$$

Onde:  $\alpha \in \mathbf{Vn}$ 

$$\beta \in (Vn \cup Vt)$$

$$\gamma \in (Vn \cup Vt)^*$$

 $\Box$  Cabeça ( $\alpha$ ) =  $\beta$ 

#### **Exemplo**

P: 
$$S \rightarrow AB$$
  
 $A \rightarrow aA \mid a$   
 $B \rightarrow bB \mid b$ 

29

# Último

Compiladores

- ☐ Relaciona um dado não terminal, existente do ladò esquerdo de uma certa regra, com o último elemento que aparece do lado direito desta regra
- ☐ Dada uma produção na forma

$$\alpha \rightarrow \gamma \beta$$

Onde:  $\alpha \in \mathbf{Vn}$ 

$$\beta \in (Vn \cup Vt)$$

$$\gamma \in (Vn \cup Vt)^*$$

 $\Box$  Último ( $\alpha$ ) =  $\beta$ 

# Último

- □ Relaciona um dado não terminal, existente do lado esquerdo de uma certa regra, com o último elemento (terminal ou nãoterminal) que aparece do lado direito desta regra
- □ Dada uma produção na forma

$$\alpha \rightarrow \gamma \beta$$

Onde:  $\alpha \in \mathbf{Vn}$ 

 $\beta \in (Vn \cup Vt)$ 

 $\gamma \in (Vn \cup Vt)^*$ 

 $\Box$  Último ( $\alpha$ ) =  $\beta$ 

$$P: S \rightarrow AB$$

$$B \rightarrow bB \mid b$$

$$Último(S) = {B}$$

Último (B) = 
$$\{B,b\}$$

- Relação próxima a relação cabeça; entretanto, deve conter somente terminais
- □ Primeiro(A) = x, onde A produz x como seu símbolo mais à esquerda com n derivações, sendo A ∈ Vn e x ∈ Vt\*
  - □ x pode ser a cadeia vazia

- ☐ Para determinar PRIMEIRO(X):
  - 1. Se  $\mathbf{x}$  é um terminal, então PRIMEIRO  $(\mathbf{x}) = \{\mathbf{x}\}$
  - 2. Se  $\mathbf{X}$  é não-terminal e  $X \to a\alpha$  é uma produção, então se acrescenta  $\mathbf{a}$  ao conjunto PRIMEIRO de x
  - 3. Se  $X \rightarrow \epsilon$  é uma produção  $\epsilon$  deve ser adicionado ao conjunto PRIMEIRO de x
  - 4. Se  $X \rightarrow Y_1 Y_2 ... Y_k$  é uma produção, então todo i tal que todos  $Y_1 ... Y_{i-1}$  são não-terminais e PRIMEIRO $(Y_j)$  contém  $\epsilon$ , onde j=1,2...i-1. acrescente todo símbolo diferente de  $\epsilon$  de PRIMEIRO $(Y_j)$  a PRIMEIRO(X). Se  $\epsilon \in PRIMEIRO(X)$ , para todo i=1,2...k. então acrescente  $\epsilon$  a PRIMEIRO(X).

```
E \rightarrow TE'
E' \rightarrow +TE' \mid \epsilon
T \rightarrow FT'
T' \rightarrow *FT' \mid \epsilon
F \rightarrow (E) \mid id
```

```
Primeiro (E) = {?}
Primeiro (E') = {?}
Primeiro (T) = {?}
Primeiro (T') = {?}
Primeiro (F) = {?}
```

## ✓ □ Exemplo

```
T \rightarrow FT'
T' \rightarrow *FT' \mid \varepsilon
F \rightarrow (E) \mid id
```

 $E \rightarrow TE'$ 

```
Primeiro (E) = Primeiro(T) = Primeiro(F) ={(, id}

Primeiro (E') = {+, \varepsilon}

Primeiro (T) = Primeiro(F) = {(, id}

Primeiro (T') = {*, \varepsilon}

Primeiro (F) = {(, id}
```

F deriva em  $\varepsilon$ ?

R: Não, então primeiro(T) = primeiro(F) = {(, id)}

Se F derivasse em **E** era preciso incluir o primeiro(T') em primeiro(T)

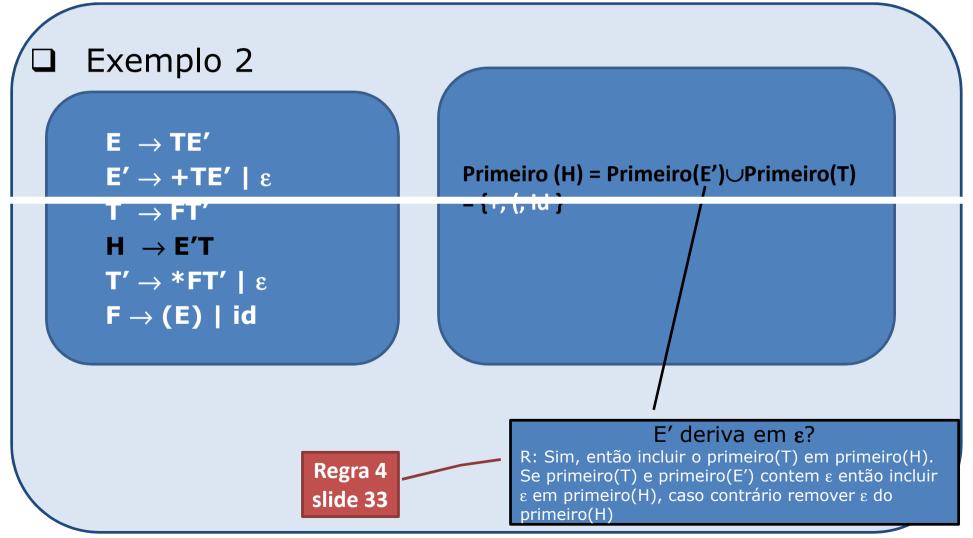
☐ Exemplo 2

```
E \rightarrow TE'
E' \rightarrow +TE' \mid \epsilon
T \rightarrow FT'
H \rightarrow E'T
T' \rightarrow *FT' \mid \epsilon
F \rightarrow (E) \mid ia
```

**Primeiro (H) = {?}** 

Se fosse incluída esta regra na gramática como ficaria o primeiro(H)?

# Primeiro (First)



☐ Se A é um *não-terminal*, Seguinte(A) é o conjunto de terminais que podem figurar imediatamente à direita de A em alguma forma sentencial

□ Seguinte(A)= $\mathbf{x}$  para a regra S $\rightarrow \alpha A\beta$  e primeiro( $\beta$ )= $\mathbf{x}$ 

```
Onde: A \in Vn
X \in Vt^*
\alpha \in \beta \in (Vn \cup Vt)^*
```

- Para determinar Seguinte(A):
  - Colocar \$ em Seguinte(S) se S é o símbolo de partida \$ é o marcador de fim de entrada durante a análise sintática

- 2. Se existe uma produção  $A \rightarrow \alpha B\beta$  e  $\beta \notin \epsilon$  então tudo que estiver em PRIMEIRO( $\beta$ ), exceto  $\epsilon$ , deve ser adicionado em Seguinte(B)
- 3. Se existe uma produção  $A \rightarrow \alpha B$  ou  $A \rightarrow \alpha B\beta$  onde PRIMEIRO( $\beta$ ) contem  $\epsilon$  ( $\beta \rightarrow \epsilon$ ), então tudo que está em Seguinte(A) está em Seguinte(B)

```
E \rightarrow TE'
E' \rightarrow +TE' \mid \epsilon
T \rightarrow FT'
T' \rightarrow *FT' \mid \epsilon
```

```
Primeiro (E) = Primeiro(T) = Primeiro(F) = \{(, id)\}

Primeiro (E') = \{+, \epsilon\}

Primeiro (T) = Primeiro(F) = \{(, id)\}

Primeiro (T') = \{*, \epsilon\}

Primeiro (F) = \{(, id)\}
```

```
E \rightarrow TE'
E' \rightarrow +TE' \mid \epsilon
T \rightarrow FT'
T' \rightarrow *FT' \mid \epsilon
```

```
Primeiro (E) = Primeiro(T) = Primeiro(F) = \{(, id)\}

Primeiro (E') = \{+, \epsilon\}

Primeiro (T) = Primeiro(F) = \{(, id)\}

Primeiro (T') = \{*, \epsilon\}

Primeiro (F) = \{(, id)\}
```

```
E \rightarrow TE'
E' \rightarrow +TE' \mid \epsilon
T \rightarrow FT'
T' \rightarrow *FT' \mid \epsilon
```

```
Primeiro (E) = Primeiro(T) = Primeiro(F) = \{(, id)\}

Primeiro (E') = \{+, \epsilon\}

Primeiro (T) = Primeiro(F) = \{(, id)\}

Primeiro (T') = \{*, \epsilon\}

Primeiro (F) = \{(, id)\}
```

```
E \rightarrow TE'
E' \rightarrow +TE' \mid \epsilon
T \rightarrow FT'
T' \rightarrow *FT' \mid \epsilon
```

```
Primeiro (E) = Primeiro(T) = Primeiro(F) = \{(, id)\}

Primeiro (E') = \{+, \epsilon\}

Primeiro (T) = Primeiro(F) = \{(, id)\}

Primeiro (T') = \{*, \epsilon\}

Primeiro (F) = \{(, id)\}
```

```
E \rightarrow TE'
E' \rightarrow +TE' \mid \epsilon
T \rightarrow FT'
T' \rightarrow *FT' \mid \epsilon
```

```
Primeiro (E) = Primeiro(T) = Primeiro(F) = \{(, id)\}

Primeiro (E') = \{+, \epsilon\}

Primeiro (T) = Primeiro(F) = \{(, id)\}

Primeiro (T') = \{*, \epsilon\}

Primeiro (F) = \{(, id)\}
```

Encontre os conjuntos Primeiro(*First*) para as gramáticas abaixo

Primeiro(A)= {}
Primeiro(B)= {}
Primeiro(C)= {}

Primeiro(S)= {}

b) S  $\rightarrow$  ABd A  $\rightarrow$  aA |  $\epsilon$ B  $\rightarrow$  bB | cA | AC C  $\rightarrow$  cB |  $\epsilon$ 

Encontre os conjuntos Primeiro(*First*) para as gramáticas abaixo

```
b) S \rightarrow ABd
A \rightarrow aA | \epsilon
B \rightarrow bB | cA | AC
C \rightarrow cB | \epsilon
```

```
Primeiro(S)= \{\}

Primeiro(A)= \{\}

Primeiro(B)= \{\}

Primeiro(C)= \{c, \epsilon\}
```

Encontre os conjuntos Primeiro(*First*) para as gramáticas abaixo

```
b) S \rightarrow ABd
A \rightarrow aA | \epsilon
B \rightarrow bB | cA | AC
C \rightarrow cB | \epsilon
```

```
Primeiro(S)= {}

Primeiro(A)= {}

Primeiro(B)= {b, c, \mathbf{U} Primeiro(A)}

Primeiro(C)= {c, \epsilon}
```

Encontre os conjuntos Primeiro(*First*) para as gramáticas abaixo

```
b) S \rightarrow ABd A \rightarrow aA | \epsilon B \rightarrow bB | cA | AC C \rightarrow cB | \epsilon
```

```
Primeiro(S)= {}

Primeiro(A)= {a, \varepsilon}

Primeiro(B)= {b, c, U Primeiro(A)}

Primeiro(C)= {c, \varepsilon}
```

Encontre os conjuntos Primeiro(*First*) para as gramáticas abaixo

```
b) S \rightarrow ABd
A \rightarrow aA | \epsilon
B \rightarrow bB | cA | AC
C \rightarrow cB | \epsilon
```

```
Primeiro(S)= {}

Primeiro(A)= {a, \varepsilon}

Primeiro(B)= {b, c, U Primeiro(A)} = {b, c, a, \varepsilon}

Primeiro(C)= {c, \varepsilon}
```

Encontre os conjuntos Primeiro(*First*) para as gramáticas abaixo

Primeiro(S)= {Primeiro(A)  $\mathbf{U}$  Primeiro(B)  $\mathbf{U}$  d} Primeiro(A)= {a,  $\varepsilon$ }

Primeiro(B)=  $\{b, c, \mathbf{U} \text{ Primeiro}(A)\} = \{b, c, a, \epsilon\}$ 

Primeiro(C)=  $\{c, \epsilon\}$ 

b)  $S \rightarrow ABd$ 

 $A \rightarrow aA \mid \epsilon$ 

 $B \rightarrow bB \mid cA \mid AC$ 

 $C \to cB \mid \epsilon$ 

Quando tem uma regra do tipo A  $\rightarrow$  BCD, o  $\epsilon$  só entra no Primeiro(A) se ele puder ser gerado por B, C e D também.

Primeiro(A)
gera ɛ,
então
Primeiro(B)
entra em
Primeiro(S)

gera ɛ,
então d
também
entra em
Primeiro(S)

Encontre os conjuntos Primeiro(*First*) para as gramáticas abaixo

```
b) S 
ightarrow ABd A 
ightarrow aA | \epsilon B 
ightarrow bB | cA | AC C 
ightarrow cB | \epsilon
```

```
Primeiro(S)= {a, b, c, d, \varepsilon}

Primeiro(A)= {a, \varepsilon}

Primeiro(B)= {b, c, a, \varepsilon }

Primeiro(C)= {c, \varepsilon}
```

```
Primeiro(S)= {}
Primeiro(A)= {}
Primeiro(B)= {}
Primeiro(C)= {}
Primeiro(D)= {}
```

```
Primeiro(S)= {a, g, c, b, f, \epsilon}

Primeiro(A)= {a, g, c, \epsilon}

Primeiro(B)= {b, f, \epsilon}

Primeiro(C)= {c}

Primeiro(D)= {g, c, \epsilon}
```

Encontre os conjuntos Primeiro(*First*) para as gramáticas abaixo

d)  $S \rightarrow aA \mid bB$   $A \rightarrow aAS \mid BD$   $B \rightarrow bB \mid fAC \mid \epsilon$   $C \rightarrow cC \mid Dd$  $D \rightarrow gD \mid C \mid \epsilon$  Primeiro(S)= {}
Primeiro(A)= {}
Primeiro(B)= {}
Primeiro(C)= {}
Primeiro(D)= {}

Encontre os conjuntos Primeiro(*First*) para as gramáticas abaixo

```
d) S \rightarrow aA | bB
A \rightarrow aAS | BD
B \rightarrow bB | fAC | \epsilon
C \rightarrow cC | Dd
D \rightarrow gD | C | \epsilon
```

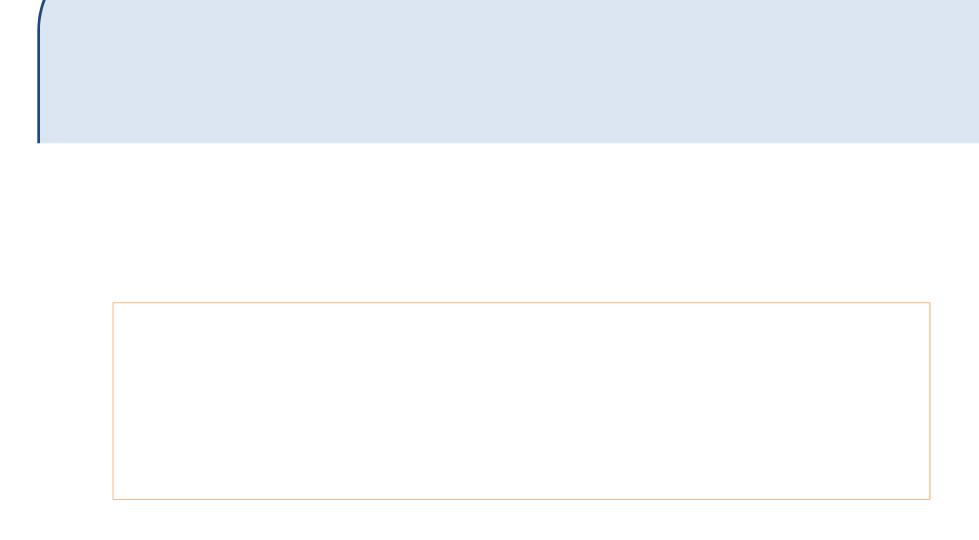
```
Primeiro(S)= \{a,b\}

Primeiro(A)= \{a, b, f, g, c, d, \epsilon\}

Primeiro(B)= \{b, f, \epsilon\}

Primeiro(C)= \{g, c, d, \epsilon\}

Primeiro(D)= \{c, g, d, \epsilon\}
```



Encontre os conjuntos First e Follow para a gramática LALG.