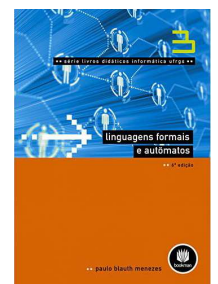


[Aula 11] Linguagem livre de contexto – Gramática livre de contexto

Prof. João F. Mari
joaof.mari@ufv.br

BIBLIOGRAFIA

- MENEZES, P. B. **Linguagens formais e autômatos**, 6. ed., Bookman, 2011.
 - Capítulo 6.
 - + Slides disponibilizados pelo autor do livro.



ROTEIRO

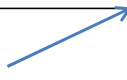
- Linguagens Livres do Contexto
- Gramática Livre de Contexto
- GLC – Linguagem gerada por uma GLC
- EXEMPLO: GLC – Duplo Balanceamento
- Sobre o Duplo Balanceamento
- EXEMPLO: GLC - Expressões Aritméticas
- Forma de Backus Naur (BNF – Backus Naur Form)
- EXEMPLO: BNF – Identificadores (C, Pascal, Java)
- Árvore de Derivação
- Árvore de Derivação × Derivações
- Derivação mais à Esquerda e Derivação mais à Direita
- GLC Ambígua e Gramática Ambígua
- EXEMPLO: GLC ambígua – Expressão Aritmética
- Linguagem Inerentemente Ambígua
- Contraexemplo: Linguagem Inerentemente Ambígua

Linguagens Livres do Contexto

- Classe das Linguagens Livres do Contexto ou Tipo 2
 - Fundamental importância para computação e informática.
 - Universo mais amplo de linguagens quando comparado com as LR
 - Trata, adequadamente, questões típicas de linguagens de programação
 - Parênteses balanceados
 - Construções bloco-estruturadas, etc.
- Algoritmos reconhecedores e geradores
 - São relativamente simples.
 - Possuem eficiência razoável.

Linguagens Livres do Contexto

- Aplicações típicas
 - Centradas em linguagens artificiais
 - Em especial, nas linguagens de programação
 - Analisadores sintáticos
 - Tradutores de linguagens
 - Processadores de texto em geral
 - Analisador sintático do Word (Quando seu Word “marca” seu texto em verde).
- A gente estudamos para a prova.


- Na Hierarquia de Chomsky
 - A Classe das Linguagens Livres do Contexto
 - Contém propriamente a Classe das Linguagens Regulares
 - Porém, ainda é uma classe relativamente restrita
 - É fácil definir linguagens que não pertencem a esta classe.
 - (Linguagens SENSIVEIS ao Contexto e Linguagens Recursivamente Enumeráveis)

Linguagens Livres do Contexto

- **Formalismos:**
 - **Gramática Livre do Contexto** (axiomático ou gerador)
 - Restrições na forma das regras de produção
 - Mais livre do que na gramática regular
 - **Autômato com Pilha** (operacional ou reconhecedor)
 - Análogo ao autômato finito não determinístico
 - Adicionalmente: memória auxiliar tipo pilha
 - Pode ser lida ou gravada

Gramática Livre de Contexto

- Relativamente às GLC
 - Árvore de derivação
 - representa a derivação de uma palavra na forma de árvore
 - arte do símbolo inicial como a raiz
 - termina em símbolos terminais como folhas
 - Gramática Ambígua
 - pelo menos uma palavra com duas ou mais árvores de derivação
 - Simplificação de Gramática (produções)
 - sem reduzir o poder de geração
 - Forma Normal: restrições rígidas na forma das produções
 - sem reduzir o poder de geração da gramática

Gramática Livre de Contexto

- Autômato com pilha construído a partir de uma GLC
 - A construção de um reconhecedor a partir de sua gramática
 - É simples e imediata.
 - Estrutura de pilha é suficiente como única memória
 - Pode ser reconhecida por autômato com pilha com um estado.
 - Os estados não são necessários para "memorizar" o passado.
 - Diferentemente dos **Autômatos Finitos**, em que os estados eram a única forma de armazenar algum tipo de informação.

GLC – Linguagem gerada por uma GLC

- **Gramática Livre do Contexto (GLC)**

$$G = (V, T, P, S)$$

- Qualquer regra de produção é da forma

$$A \rightarrow \alpha$$

- A é variável de V

- lado esquerdo = uma variável

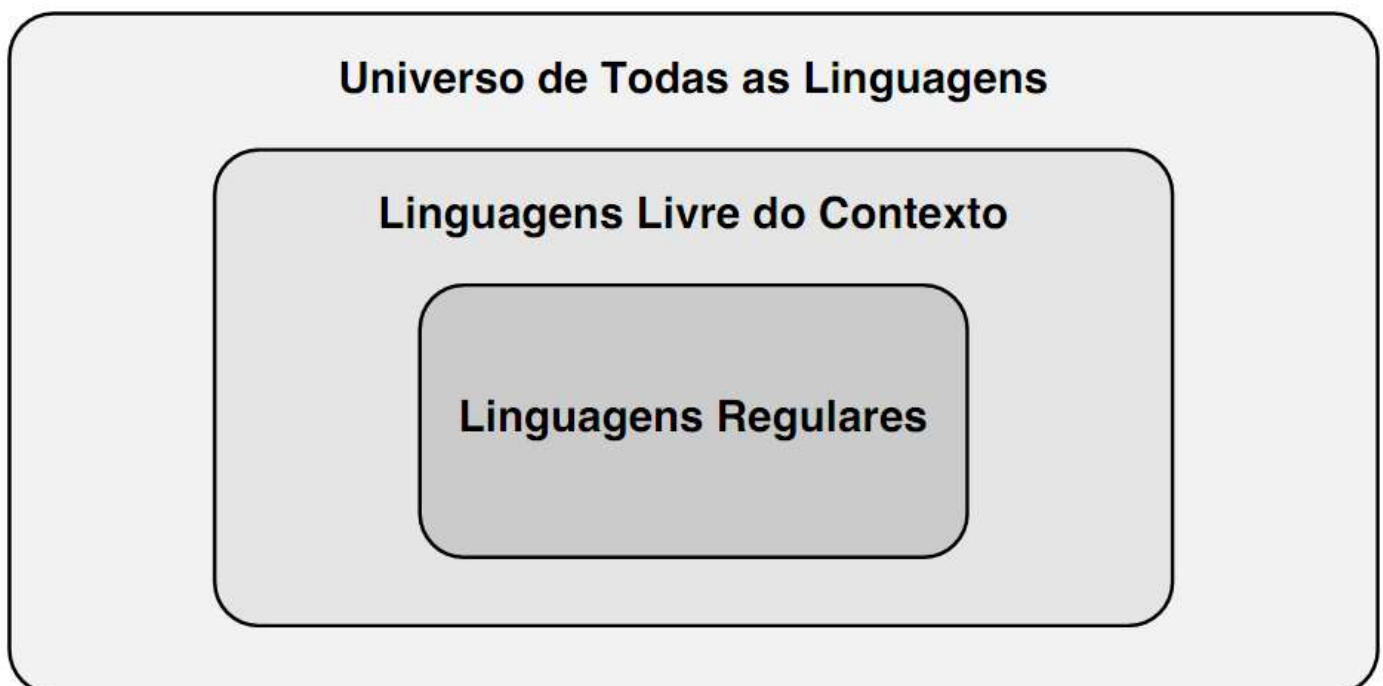
- α é palavra de $(V \cup T)^*$

- **Linguagem gerada pela gramática livre do contexto G**

$$\text{GERA}(G) = \{ w \in T^* \mid S \Rightarrow^+ w \}$$

- Toda Linguagem Regular é Livre de Contexto.

Relação entre as classes de linguagens



Linguagem Livre de Contexto

- Porque o nome “Linguagem Livre do Contexto”?
 - Constitui a mais geral classe de linguagens cuja produção é da forma
 - $A \rightarrow \alpha$
 - Em uma derivação, a variável A deriva α
 - sem depender ("livre") dos símbolos que antecedem ou sucedem A (o "contexto")
 - na palavra que está sendo derivada

EXEMPLO: GLC – Duplo Balanceamento

- Linguagem com Duplo Balanceamento

$$L1 = \{ a^n b^n \mid n \geq 0 \}$$
- $G1 = (\{ S \}, \{ a, b \}, P1, S)$
 - $P1 = \{ S \rightarrow aSb$
 - $S \rightarrow \epsilon \}$
 - $GERA(G1) = L1$
- Derivação da palavra $aabb$
 - $S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aa\epsilon bb = aabb$

Sobre o Duplo Balanceamento

- Importância fundamental em computação
 - Permite estabelecer analogia com estruturas de duplo balanceamento em linguagens de programação
- Linguagens estruturadas em bloco
 - $\text{begin}^n \text{end}^n$, $\{ \}$, e similares
- Linguagens com parênteses balanceados
 - $(^n)^n$

```
int main ( ) {
    printf ("Olá mundo!!!");
}
```

EXEMPLO: GLC - Expressões Aritméticas

- Seja L2 uma linguagem composta por:
 - expressões aritméticas com colchetes balanceados,
 - dois operadores e um operando
- $G2 = (\{ E \}, \{ +, *, [,], x \}, P2, E)$
 - $P2 = \{ E \rightarrow E+E \mid E * E \mid [E] \mid x \}$
- Derivação da expressão $[x+x]*x$
 - $E \Rightarrow E * E \Rightarrow [E] * E \Rightarrow [E+E] * E \Rightarrow [x+E] * E \Rightarrow [x+x] * E \Rightarrow [x+x] * x$

Forma de Backus Naur (BNF – Backus Naur Form)

- Maneira usual de representar uma GLC
 - Variáveis
 - palavras delimitadas pelos símbolos $\langle e \rangle$
 - Terminais
 - palavras não delimitadas

- A regra de produção

$$A \rightarrow \alpha$$

- Na BNF, é representada por

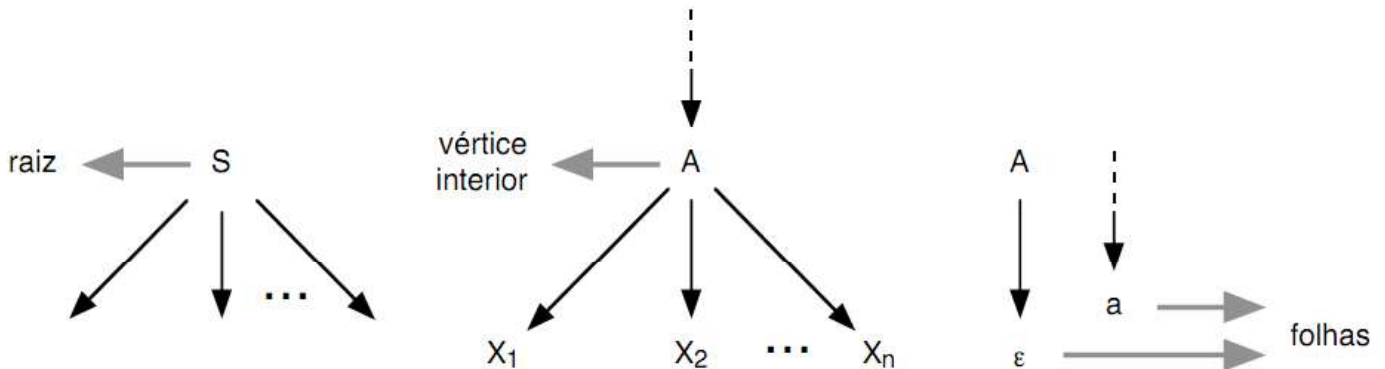
$$\langle A \rangle ::= \alpha$$

EXEMPLO: BNF – Identificadores (C, Pascal, Java)

- A variável $\langle \text{identificador} \rangle$ é o símbolo inicial
 - $\langle \text{identificador} \rangle ::= \langle \text{letra} \rangle | \langle \text{identificador} \rangle \langle \text{letra} \rangle | \langle \text{identificador} \rangle \langle \text{dígito} \rangle$
 - $\langle \text{letra} \rangle ::= a | b | \dots | z$
 - $\langle \text{dígito} \rangle ::= 0 | 1 | \dots | 9$

Árvore de Derivação

- **Raiz:** símbolo inicial
 - se A é um vértice interior e X_1, X_2, \dots, X_n são os "filhos" de A
 - $A \rightarrow X_1X_2\dots X_n$ é uma produção da gramática
 - X_1, X_2, \dots, X_n são ordenados da esquerda para a direita
- **Vértice interior:** variáveis
 - se A é um vértice interior e X_1, X_2, \dots, X_n são os "filhos" de A
 - $A \rightarrow X_1X_2\dots X_n$ é uma produção da gramática
 - X_1, X_2, \dots, X_n são ordenados da esquerda para a direita
- **Vértice folha ou folha:** terminal ou o símbolo vazio
 - Se vazio: único filho de seu pai ($A \rightarrow \epsilon$)



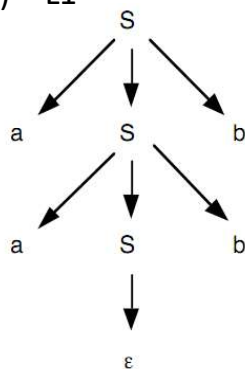
Árvore de Derivação

- Árvores de derivação para as palavras:

aabb

Duplo Balanceamento

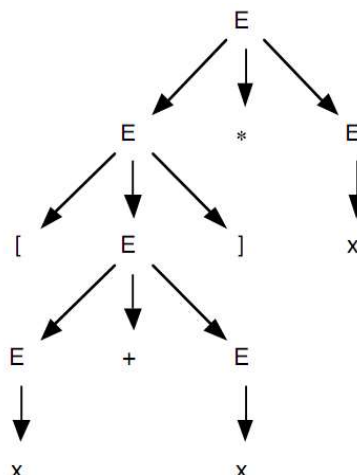
$G1 = (\{ S \}, \{ a, b \}, P1, S)$
 $P1 = \{ S \rightarrow aSb \mid S \rightarrow \epsilon \}$
 $GERA(G1) = L1$



$[x+x]^*x$

Expressão aritmética

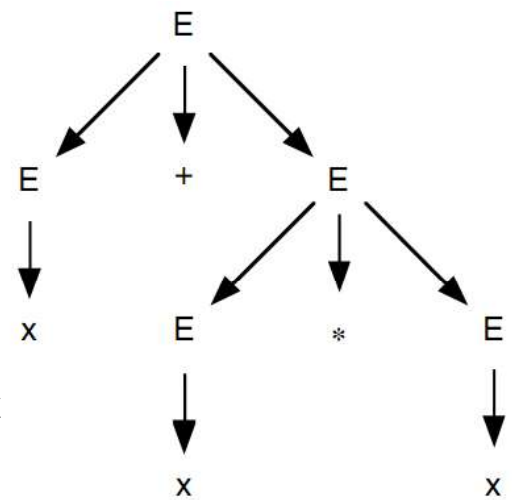
$G2 = (\{ E \}, \{ +, *, [,], x \}, P2, E)$
 $P2 = \{ E \rightarrow E+E \mid E * E \mid [E] \mid x \}$



Árvore de Derivação × Derivações

- Uma árvore de derivação pode representar derivações distintas de uma mesma palavra.

– EXEMPLO: $x+x*x$



– $E \Rightarrow E+E \Rightarrow x+E \Rightarrow x+E*E \Rightarrow x+x*E \Rightarrow x+x*x$

- mais a esquerda

– $E \Rightarrow E+E \Rightarrow E+E*E \Rightarrow E+E*x \Rightarrow E+x*x \Rightarrow x+x*x$

- mais a direita

– $E \Rightarrow E+E \Rightarrow E+E*E \Rightarrow x+E*E \Rightarrow x+x*E \Rightarrow x+x*x$

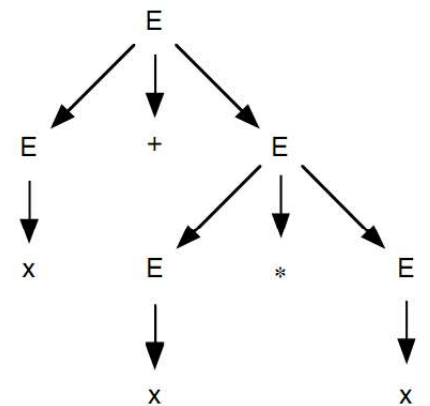
Derivação mais à Esquerda e Derivação mais à Direita

- Derivação mais à Esquerda

– Sequência de produções aplicada sempre à variável mais à esquerda.

– EXEMPLO: $x+x*x$

- $E \Rightarrow E+E \Rightarrow x+E \Rightarrow x+E*E \Rightarrow x+x*E \Rightarrow x+x*x$



- Derivação mais à Direita

– Sequência de produções aplicada sempre à variável mais à direita.

– EXEMPLO: $x+x*x$

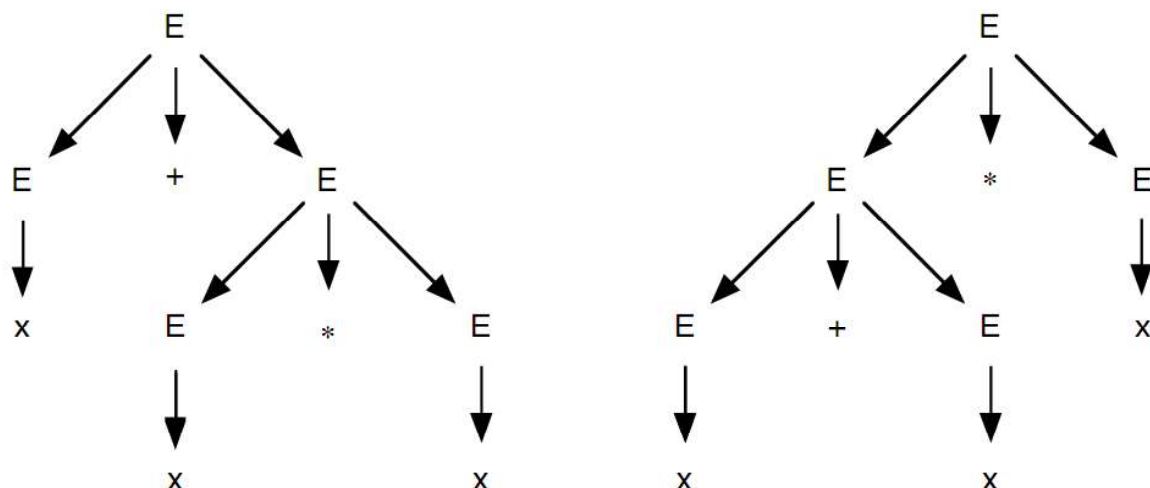
- $E \Rightarrow E+E \Rightarrow E+E*E \Rightarrow E+E*x \Rightarrow E+x*x \Rightarrow x+x*x$

GLC Ambígua e Gramática Ambígua

- Uma GLC é dita ambígua se,
 - Existe pelo menos uma palavra que possua duas ou mais árvores de derivação nessa linguagem.
- Em muitas aplicações é desejável que a gramática usada seja não ambígua.
 - Exemplo: Linguagens de programação
- PORÉM, nem sempre é possível eliminar ambiguidades

EXEMPLO: GLC ambígua – Expressão Aritmética

- Considera a GLC G2:
 - $G2 = (\{ E \}, \{ +, *, [,], x \}, P2, E)$
 - $P2 = \{ E \rightarrow E+E \mid E * E \mid [E] \mid x \}$
 - A palavra $x+x*x$ pode ser gerada por árvores distintas



GLC Ambígua

- Outra forma de definir gramática ambígua é:
 - Verificando se existe pelo menos uma palavra com duas ou mais derivações mais à esquerda OU mais à direita.
- Gramática Ambígua
 - Uma GLC é uma Gramática Ambígua se existe pelo menos uma palavra com:
 - duas ou mais derivações mais à esquerda ou
 - duas ou mais derivações mais à direita

GLC Ambígua

- Considera a GLC G2:
 - $G2 = (\{ E \}, \{ +, *, [,], x \}, P2, E)$
 - $P2 = \{ E \rightarrow E+E \mid E * E \mid [E] \mid x \}$

– A palavra **$x+x*x$**

- Derivação mais à esquerda

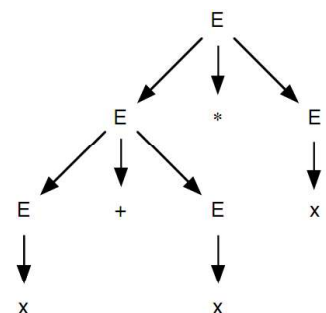
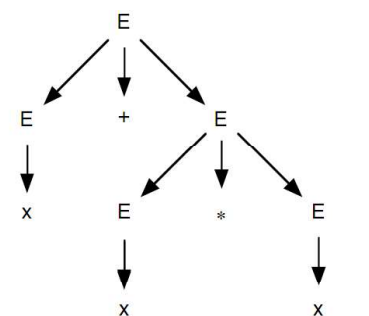
– $E \Rightarrow E+E \Rightarrow x+E \Rightarrow x+E * E \Rightarrow x+x * E \Rightarrow x+x * x$

– $E \Rightarrow E * E \Rightarrow E+E * E \Rightarrow x+E * E \Rightarrow x+x * E \Rightarrow x+x * x$

- Derivação mais à direita

– $E \Rightarrow E+E \Rightarrow E+E * E \Rightarrow E+E * x \Rightarrow E+x * x \Rightarrow x+x * x$

– $E \Rightarrow E * E \Rightarrow E * x \Rightarrow E+E * x \Rightarrow E+x * x \Rightarrow x+x * x$



Linguagem Inerentemente Ambígua

- Uma linguagem L é dita **Linguagem Inerentemente Ambígua** quando:
 - Qualquer GLC que a define é ambígua
 - Ou seja, NÃO existe uma GLC não ambígua que define L .
- EXEMPLO: Linguagem Inerentemente Ambígua
 - $\{ w \mid w = a^n b^n c^m d^m, n \geq 1, m \geq 1 \text{ ou}$
 - $w = a^n b^m c^m d^n, n \geq 1, m \geq 1 \}$

Contraexemplo: Linguagem Inerentemente Ambígua

- A linguagem Expressões Aritméticas não é inerentemente ambígua
 - Dizemos apenas que a Linguagem é não ambígua.
 - Apesar de existirem **várias GLCs ambíguas** que a definem,
 - É **possível** definir **(pelo menos) uma GLC** não ambígua que a define.

[FIM]

- FIM:
 - **[AULA 11]** Linguagem livre de contexto – Gramática livre de contexto.
- Próxima aula:
 - **[AULA 12]** Linguagem livre de contexto – Simplificação de GLC.