

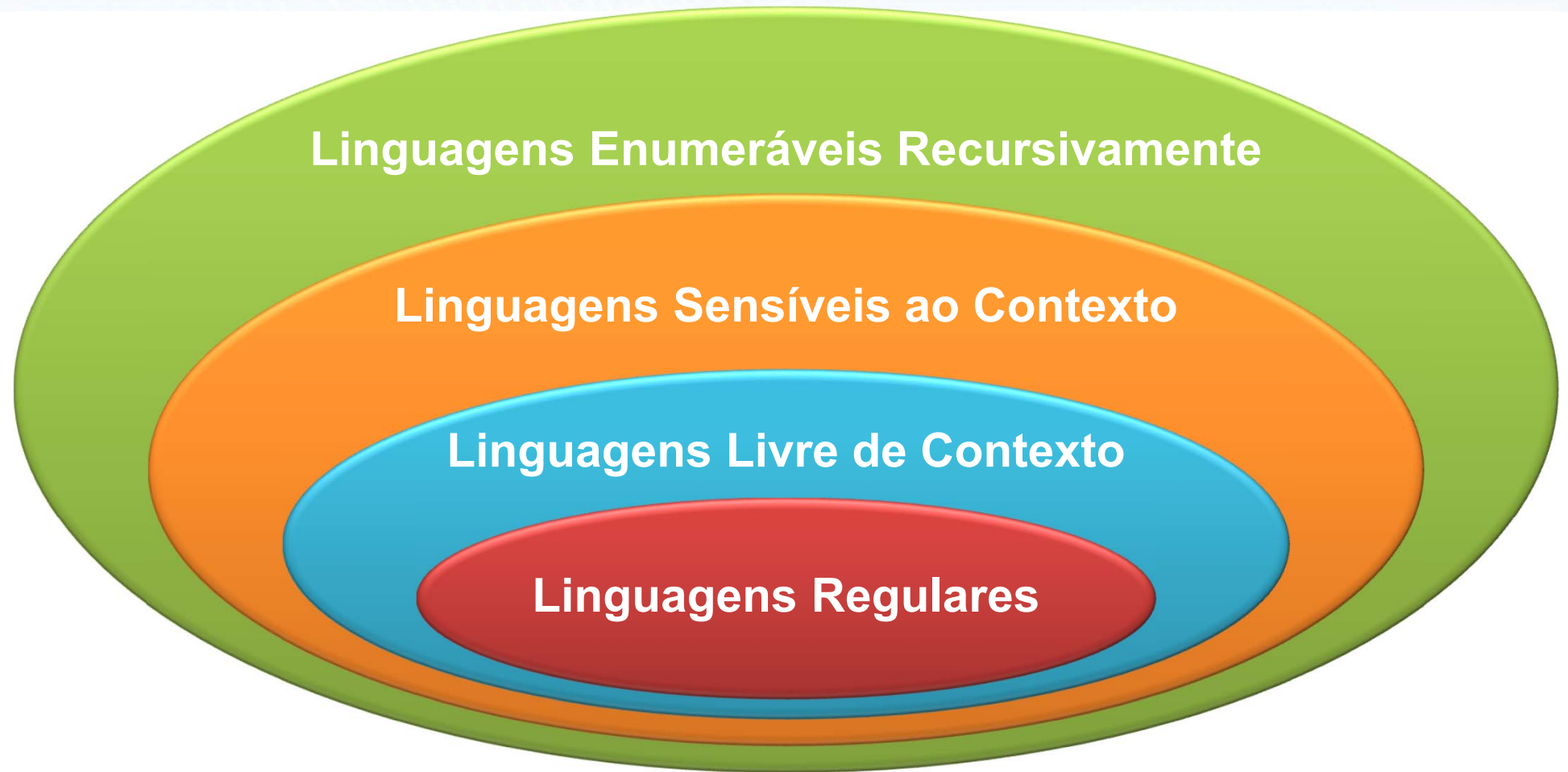
TEORIA DA COMPUTAÇÃO E COMPILADORES

Gramáticas Livres de Contexto - GLC

Prof. Dr. Fernando Kakugawa

fernando.kakugawa@animaeducacao.com.br

Tipos de Linguagens



Tipos de Linguagens

- Linguagens Regulares
 - Autômatos e Expressões Regulares
- Linguagens Livre de Contexto
 - Gramáticas Livre de Contexto e Autômatos de Pilha
- Linguagens Sensíveis ao Contexto
 - Gramáticas Irrestritas e Sensíveis ao Contexto
- Linguagens Enumeráveis Recursivamente
 - Máquinas de Turing

GLC – Introdução

- Gramáticas Livres de Contexto é um método capaz de descrever linguagens que possuam uma estrutura recursiva
 - Ex: $L = \{0^n 1^n, n \geq 1\}$.
- Uma das primeiras aplicações de gramáticas livres de contexto foi sua utilização na descrição de linguagens naturais.

GLC – Aplicações

- Outra utilização importante de gramáticas livres de contexto ocorre na especificação e compilação de linguagens de programação:
 - Para definir uma linguagem de programação é necessário descrever a linguagem a ser criada através de uma gramática;
 - O compilador possui um componente chamado parser (analisador sintático) que extrai o “*significado do programa*” antes de gerar o respectivo código ou executar a interpretação de um código;
 - Existem ferramentas que a partir da gramática podem gerar automaticamente um parser;

Linguagens Livres de Contexto

- A coleção de todas as linguagens associadas as gramáticas livres de contexto são chamadas de linguagens livres de contexto.
 - Esta classe de Linguagens inclui todas as linguagens regulares.

Definição de uma GLC

- É um modelo utilizado para descrever uma linguagem livre de contexto;
 - constituída de um conjunto de regras e relações recursivas entre estas regras.
- Um palíndromo é uma string que é lida da mesma forma da esquerda para direita e vice-versa;
 - Ex.: radar, ovo, 1001, etc.
- Considere uma linguagem formada somente por palíndromos;
 - Esta é dita ser livre de contexto devido a sua característica recursiva e poderia ser descrita da seguinte forma:

Definição de uma GLC

- $L_{\text{pal}} = \{ w \mid w = w^R \}$

➤ onde R indica o reverso de w

- Considerando $\Sigma = \{0, 1\}$ pode-se listar os seguintes exemplos de palíndromos

ε	111
0	0000
1	1111
00	0110
11	1001
010	100001
101	00000
000	111111
	010010

Definição de uma GLC

- Através dos exemplos pode-se perceber que:
 - strings 0, 1, e ε são palíndromos;
 - uma string de tamanho maior que 1 será uma palíndromo, uma vez que inicia e termina com o mesmo símbolo.

Definição de uma GLC

- Assim, pode-se utilizar as seguintes regras para descrever palíndromos:
 1. 0, 1, e ε são palíndromos (BASE);
 2. Se w é um palíndromo então $0w0$ e $1w1$ também são palíndromos (INDUÇÃO);
 3. Nenhuma string é um palíndromo de 0's e 1's a menos que decorra das regras 1 e 2.

Definição de uma GLC

- Uma GLC é uma notação formal para descrever linguagens recursivas como a que definiu o palíndromo
- Basicamente esta consiste:
 - Uma ou mais variáveis que representam as classes de strings, isto é linguagens;
 - Conjunto de regras que informam como são construídas as strings de cada classe;
 - A construção pode utilizar símbolos de um alfabeto, strings que pertencem a uma classe de linguagem ou ambos;

Exemplo

- As regras abaixo pertencem a uma gramática utilizada para descrever a linguagem dos palíndromos:

$$\left. \begin{array}{l} P \rightarrow \varepsilon; \\ P \rightarrow 0; \\ P \rightarrow 1; \end{array} \right\}$$

Definem as regras BASE. Não são definidas recursivamente.

$$\left. \begin{array}{l} P \rightarrow 0P0; \\ P \rightarrow 1P1; \end{array} \right\}$$

Parte recursiva da descrição do palíndromo.

Descrição Formal de uma GLC

- Existem quatro componentes importantes para a descrição gramatical de uma linguagem:
 - 1) Existe um conjunto finito de símbolos que formam as strings da linguagem que está sendo definida.
 - Esse conjunto é o $\{0, 1\}$ para o exemplo do palíndromo.
 - Este conjunto é chamado de alfabeto de símbolos terminais ou simplesmente símbolos terminais;

Descrição Formal de uma GLC

2) Existe um conjunto finito de variáveis também chamado de categorias sintáticas ou não terminais.

- Cada variável representa uma linguagem , isto é, um conjunto de strings
- No exemplo do palíndromo o símbolo P representa uma variável utilizada para representar a classe de palíndromos sobre o alfabeto $\{0,1\}$

Descrição Formal de uma GLC

- 3) Uma das variáveis representa a linguagem que está sendo definida, esta é chamada de símbolo de início.
 - As outras variáveis representam classes de strings que são usadas para ajudar a definir a linguagem do símbolo de início.
 - No exemplo do palíndromo, P é a única variável e portanto o símbolo de início;

Definição Formal de uma GLC

4) Existe um conjunto finito de produções ou regras que representam a definição recursiva de uma linguagem. Cada produção consiste em:

- Uma variável que está sendo (parcialmente) definida pela produção
- O símbolo de produção, \rightarrow
- Uma string de zero ou mais terminais e variáveis

Descrição Formal de uma GLC

- Estes 4 componentes descrevem uma gramática livre de contexto (GLC).
- Uma gramática G pode ser descrita utilizando a seguinte notação
 - $G = (V, T, P, S)$, onde:
 - **V**: conjunto finito de Variáveis;
 - **T**: conjunto finito de Terminais;
 - **P**: conjunto finito de Regras de Produção;
 - **S**: símbolo de Início;

Exemplos

(1) linguagens dos palíndromos

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

Onde :

$$V = \{P\}, T = \{0,1\}, S = P$$

P é dado por:

$$P \rightarrow \varepsilon;$$

$$P \rightarrow 0;$$

$$P \rightarrow 1;$$

$$P \rightarrow 0P0;$$

$$P \rightarrow 1P1;$$

Exemplos

(2) Linguagem formada expressões matemáticas envolvendo somas e multiplicações

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

$$V = \{ E, T, F \}$$

$$T = \{ a, +, *, (,) \}$$

$$S = E$$

P é dado por:

$$E \rightarrow E + T$$

$$E \rightarrow T$$

$$T \rightarrow T * F$$

$$T \rightarrow F$$

$$F \rightarrow (E)$$

$$F \rightarrow a$$

Notação compacta para Regras

- Pode-se escrever as produções para uma gramática listando cada variável uma vez, e depois listando todos os corpos das produções para essa variável, separados por barras verticais. Isto é, as produções:

$$A \rightarrow \alpha_1$$

$$A \rightarrow \alpha_2$$

...

$$A \rightarrow \alpha_n$$

- Podem ser substituídas pela notação
 - $A \rightarrow \alpha_1 \mid \alpha_2 \mid \dots \mid \alpha_n$

Exemplos

(1) $G_p = (V, T, P, S)$ onde :

$$V = \{P\}$$

$$T = \{0, 1\}$$

$$S = P$$

P é dado por

$$P \rightarrow \varepsilon \mid 0 \mid 1 \mid 0P0 \mid 1P1$$

Exemplos

(2) $G = (V, T, P, S)$ onde:

$$V = \{ E, T, F \}$$

$$T = \{ a, +, *, (,) \}$$

$$S = E$$

P é dado por:

$$E \rightarrow E + T \mid T$$

$$T \rightarrow T * F \mid F$$

$$F \rightarrow (E) \mid a$$

Obtenção das strings de uma GLC

- Uma gramática descreve uma linguagem pela geração (derivação) de cada string da linguagem da seguinte maneira:
 - a) Escreva a variável inicial da gramática (símbolo inicial);
 - b) Para a variável que foi escrita no passo (a) encontre uma regra que inicia com esta variável e substitua esta variável pelo lado direito da regra encontrada

Obtenção das strings de uma GLC

c) Repita o passo (b) até que não sobre mais nenhuma variável

- Ao final só restará uma string de símbolos terminais que pertence a linguagem descrita pela gramática;

- Para indicar este processo de derivação das strings da linguagem será utilizado o símbolo relacional \Rightarrow .

Tipos de Derivações

- Pode-se restringir o número de escolhas na derivação de uma string.
 - Existem dois tipos de derivação utilizados para escolher qual variável durante a derivação será substituída por um de seus corpos de produção:
 - Derivações mais à direita
 - Derivações mais à esquerda

Derivação mais à esquerda

- em cada etapa do processo de derivação a variável mais à esquerda é substituída por um de seus corpos de derivação
- Esta derivação é indicada pelos símbolos

$$\begin{array}{c} \Rightarrow \\ lm \end{array} \quad \text{ou} \quad \begin{array}{c} \Rightarrow \\ lm \end{array}^*$$

- Se a gramática para a qual a derivação esta sendo gerada não for subentendida pelo contexto pode-se adicionar o símbolo da gramática abaixo da seta dos símbolos.

Exemplos

$$E \xRightarrow[lm]{*} E + E \xRightarrow[lm]{*} a + E \xRightarrow[lm]{*} a + a$$

$$E \xRightarrow[lm]{*} E * E \xRightarrow[lm]{*} E + E * E \xRightarrow[lm]{*} a + E * E \xRightarrow[lm]{*} a + a * E \xRightarrow[lm]{*} a + a * a$$

Derivação mais à direita

- em cada etapa do processo de derivação a variável mais à direita é substituída por um de seus corpos de derivação
- Esta derivação é indicada pelos símbolos

$$\begin{array}{ccc} & * & \\ \Rightarrow & \text{ou} & \Rightarrow \\ rm & & rm \end{array}$$

- Se a gramática para a qual a derivação esta sendo gerada não for subentendida pelo contexto pode-se adicionar o símbolo da gramática abaixo da seta dos símbolos.

Exemplos

$$\begin{array}{c} * \\ E \Rightarrow E + E \\ rm \end{array} \quad \begin{array}{c} * \\ E \Rightarrow E + a \\ rm \end{array} \quad \begin{array}{c} * \\ a \Rightarrow a + a \\ rm \end{array}$$

$$\begin{array}{c} * \\ E \Rightarrow E * E \\ rm \end{array} \quad \begin{array}{c} * \\ E \Rightarrow E + E * E \\ rm \end{array} \quad \begin{array}{c} * \\ E \Rightarrow E + E * a \\ rm \end{array} \quad \begin{array}{c} * \\ a \Rightarrow E + a * a \\ rm \end{array} \quad \begin{array}{c} * \\ a \Rightarrow a + a * a \\ rm \end{array}$$

Notação para derivações

- As seguintes convenções são adotadas na notação de uma gramática:
 - a) Letras minúsculas próximas ao início do alfabeto (ex: a, b, c ...), parênteses, dígitos, +, -, * , dentre outros símbolos matemáticos representam símbolos terminais;
 - b) Letras maiúsculas próximas ao início do alfabeto (ex: A, B, C ...) representam variáveis;

Notação para derivações

- c) Letras minúsculas próximas ao fim do alfabeto, como x , y , z , são strings de terminais
- d) Letras maiúsculas próximas ao fim do alfabeto, como X e Y , são não terminais ou variáveis;
- e) Letras gregas minúsculas como α e β são strings que representam variáveis e/ou terminais.

Linguagem de um GLC

- Se $G = (V, T, P, S)$ é uma GLC, a linguagem de G , denotada por $L(G)$ é o conjunto de strings de terminais obtidos por derivações a partir do símbolo de início da gramática.

$$L(G) = \{w \text{ em } T^* \mid S \xRightarrow[G]{*} w\}$$

- Se uma linguagem L é a linguagem de uma gramática livre de contexto, então L é dita uma linguagem livre de contexto.

Árvore de Derivação

- Esta estrutura mostra como os símbolos de uma string de terminais são agrupados em substrings,
 - cada um dos quais pertence a linguagem de uma das variáveis da gramática.
- Uma árvore de derivação também é conhecida como árvore de análise sintática quando esta é utilizada em um compilador.
 - Esta é a estrutura preferida para representar o programa fonte pois facilita o processo de geração de código.

Construção de um Árvore

Derivação

- Dada uma gramática $G = (V, T, P, S)$, as árvores de análise sintática para G são árvores com as seguintes condições:
 - 1) O nó raiz da árvore é representado pelo símbolo, ou variável, inicial da gramática;
 - 2) Cada nó interior é rotulado por uma variável em V ;
 - 3) Cada nó folha é rotulado por uma variável, um terminal ou ε .
 - No entanto, se o folha for rotulado por ε , ele deve ser o único filho de seu nó pai;

Construção de um Árvore Derivação

- 4) Se um nó interior é rotulado por A e seu filhos são rotulados por X_1, X_2, \dots, X_k respectivamente, a partir da esquerda, então $A \rightarrow X_1 X_2 \dots X_k$ é uma produção em P .
- Observe que o único momento em que um dos X 's pode ser ε é quando esse é o rótulo de único filho, e se $A \rightarrow \varepsilon$ é uma produção de G

Exemplos

Para a gramática abaixo

$G = (\{ E \}, \{ a, +, *, (,) \}, P, E)$ onde

P é dado por

$E \rightarrow E + E$

$E \rightarrow E * E$

$E \rightarrow (E)$

$E \rightarrow a$

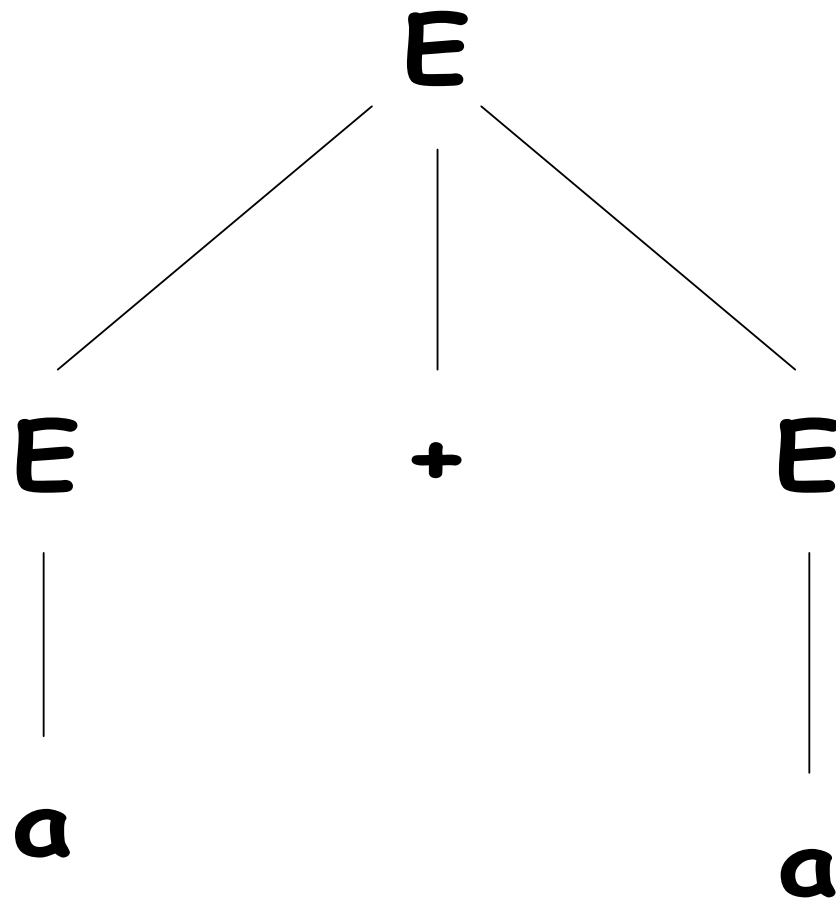
As seguintes derivações

$E \Rightarrow E + E \Rightarrow a + E \Rightarrow a + a$

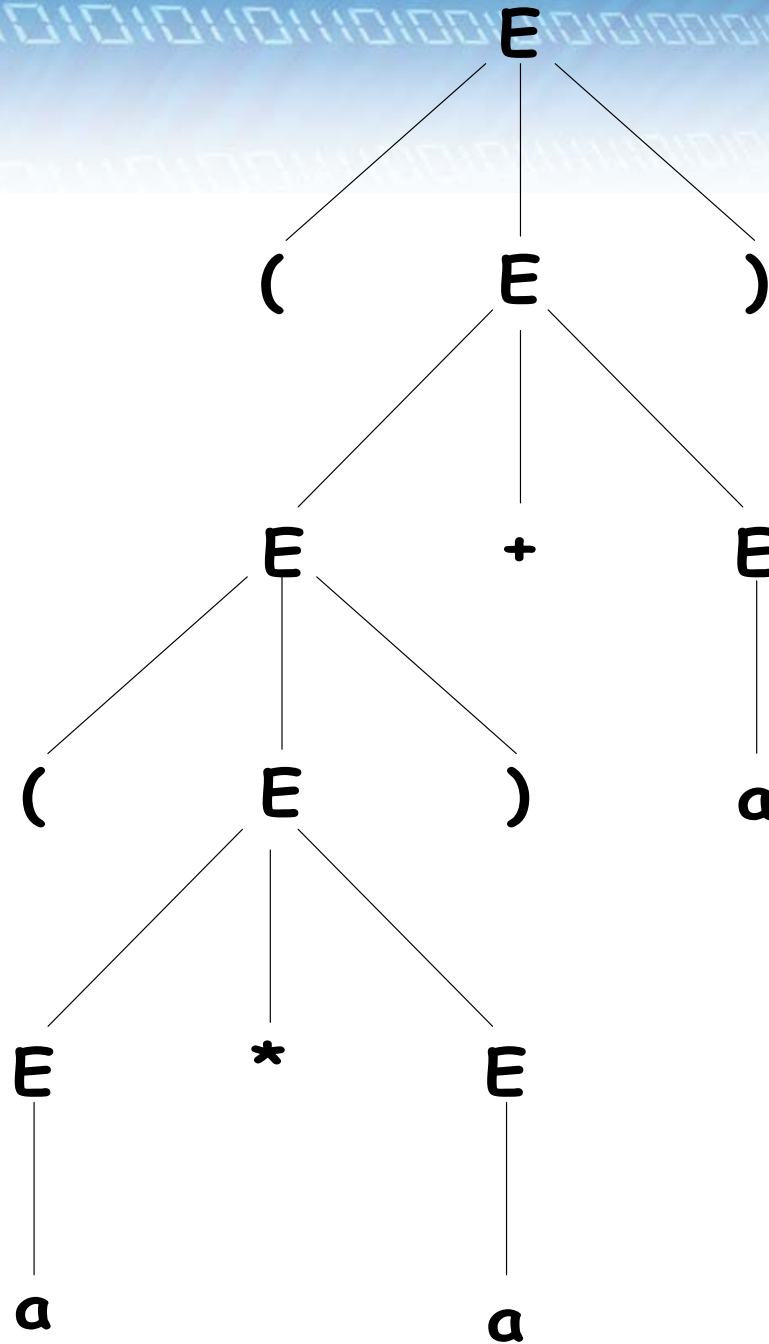
$E \Rightarrow (E) \Rightarrow (E + E) \Rightarrow ((E) + E) \Rightarrow ((E * E) + E) \Rightarrow$
 $((a * E) + E) \Rightarrow ((a * a) + E) \Rightarrow ((a * a) + a)$

Exemplos

- Pode-se desenhar as respectivas árvores de derivação

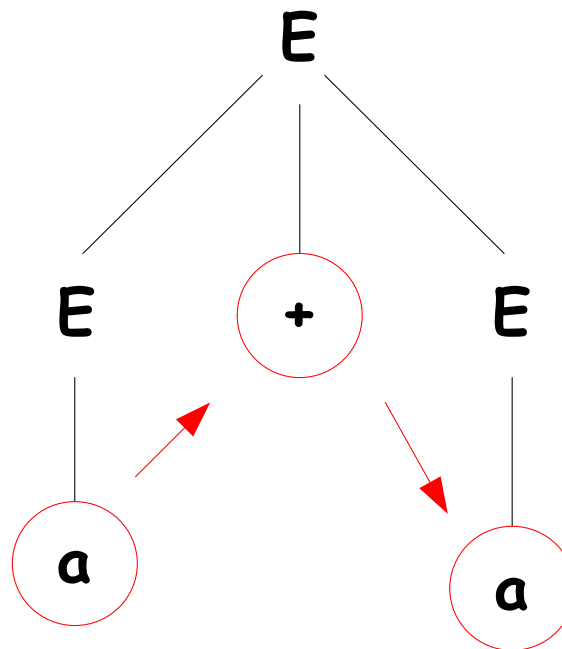


Exemplos



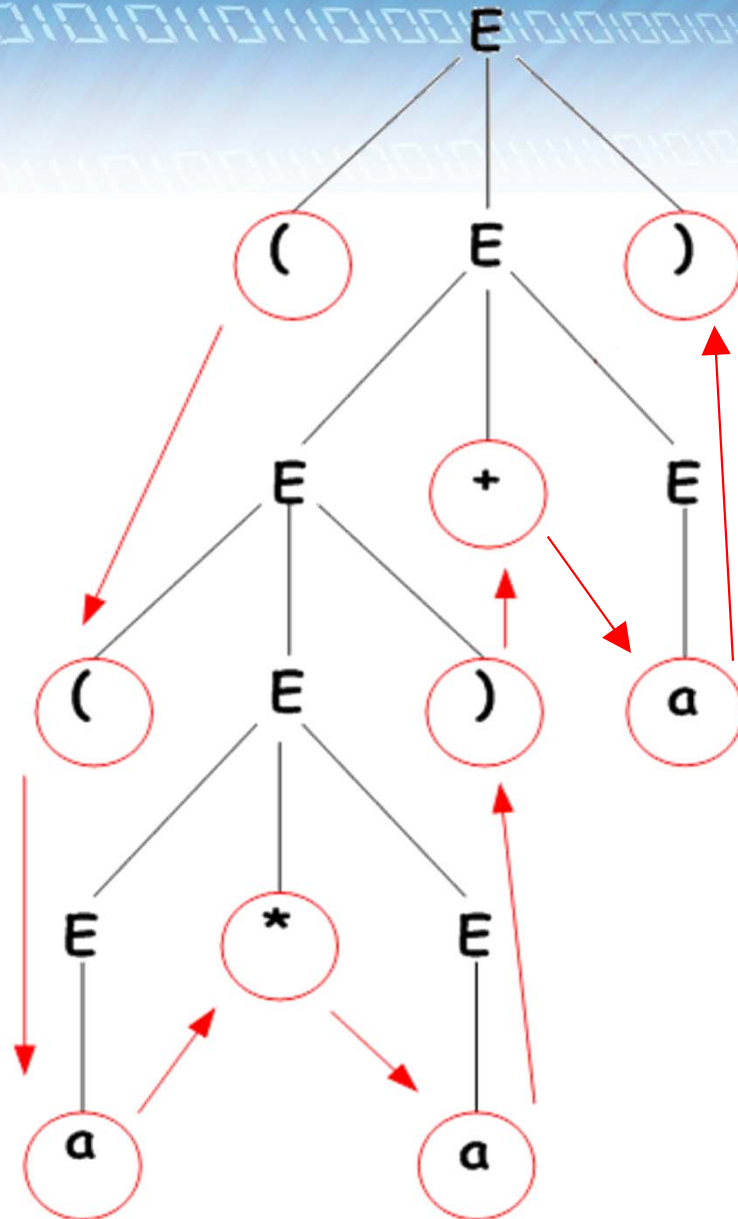
Resultado de uma Árvore de Análise Sintática

- Se os nós folhas de uma árvore forem concatenados a partir da esquerda, obteremos uma string chamada de resultado da árvore
 - que é sempre a string que é derivada da variável raiz



$$E \xRightarrow{*} a + a$$

Exemplo



$$E \Rightarrow ((a * a) + a)$$

Ambigüidade

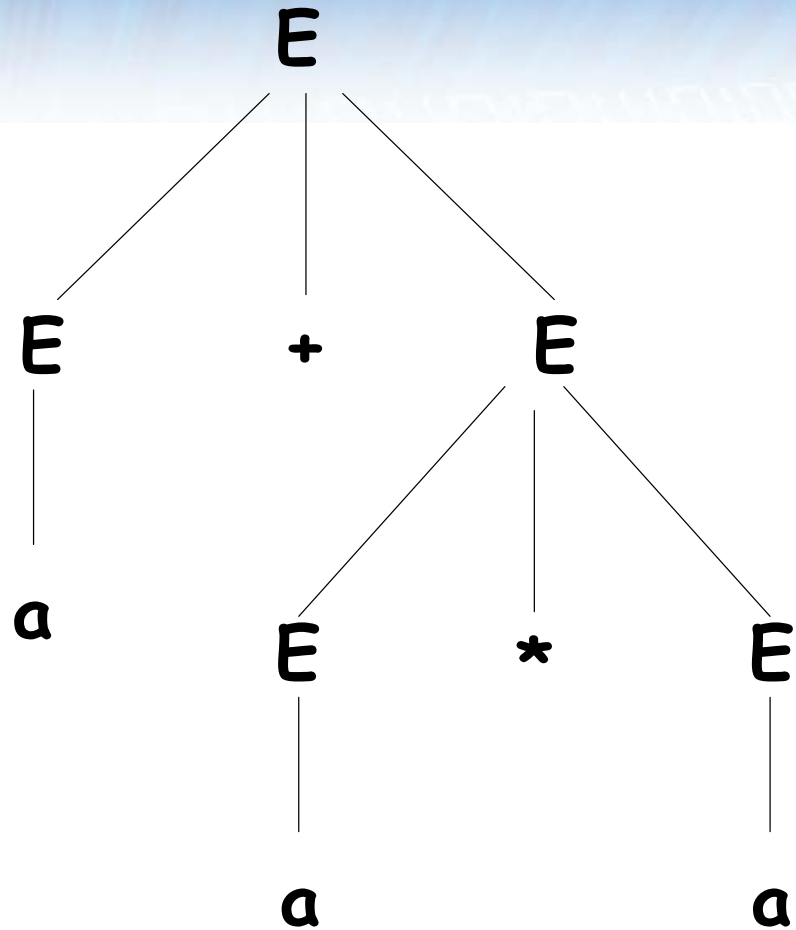
- Uma GLC $G = (V, T, P, S)$ é ambígua se existe pelo menos uma string w em T^* para a qual podemos encontrar duas árvores de derivação, cada qual com uma raiz identificada como S e um resultado w .

- *Exemplos*

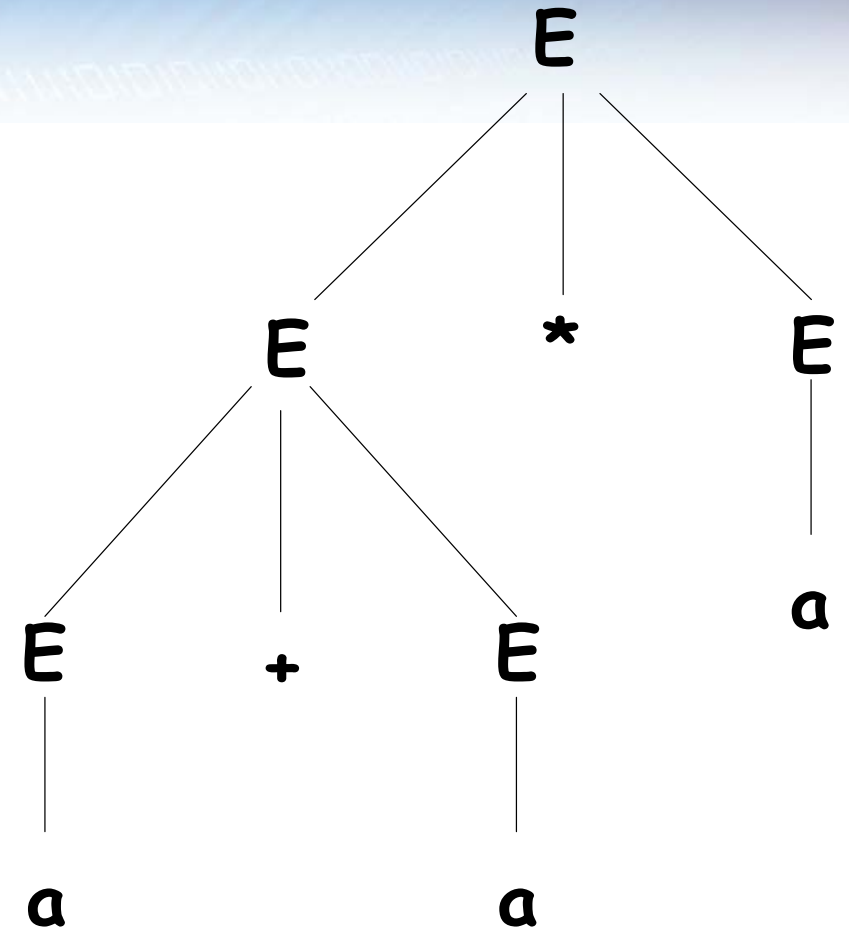
- $E \Rightarrow E + E \Rightarrow E + E * E \overset{*}{\Rightarrow} a + a * a$

- $E \Rightarrow E * E \Rightarrow E + E * E \overset{*}{\Rightarrow} a + a * a$

Exemplo de Ambigüidade



$$E \Rightarrow E + E \Rightarrow E + E * E \Rightarrow a + a * a$$



$$E \Rightarrow E * E \Rightarrow E + E * E \Rightarrow a + a * a$$

Material elaborado por:

Prof. Dr. Augusto Mendes Gomes Jr.

augusto.gomes@animaeducacao.com.br

Prof. Dr. Fernando Kakugawa

fernando.kakugawa@animaeducacao.com.br

