

Gramática Livre de Contexto

Gramática Livre de Contexto

(Merriam-Webster Dictionary of English)

Gramática é o estudo das classes das palavras, de suas inflexões, de suas funções e relacionamentos numa sentença.

Gramática é o estudo do que é preferível e o que é proibido nas inflexões (de caso, gênero, número, tempo, pessoa, modo e voz) e na sintaxe.

(Michaelis, Dicionário Brasileiro da Língua Portuguesa)

Gramática Conjunto de regras e prescrições que estabelecem o conceito de correção da língua escrita e falada.

Gramática, conforme a linguística descritiva, é a exposição sistemática e objetiva dos elementos constitutivos (fonemas, morfemas, palavras, frases etc.), dos processos de construção e dos recursos expressivos próprios do sistema estrutural de uma dada língua.

Gramática Livre de Contexto

Uma gramática $G = (V, T, S, P)$ é dita ser uma Gramática Livre de Contexto (GLC) se todas as suas produções, em P , são da forma:

$$A \rightarrow \alpha$$

onde $A \in V$ e $\alpha \in (V \cup T)^*$.

Uma gramática $\mathcal{G} = (\mathcal{V}, \mathcal{T}, \mathcal{P}, S)$ é expressa por:

- $\mathcal{V}$: conjunto finito de símbolos não terminais, ou variáveis;
- $\mathcal{T}$: conjunto finito de símbolos terminais;
- $\mathcal{P}$: conjunto finito de produções, ou regras de derivação;
- S : símbolo não terminal inicial.

Como $A \in V$, então, do lado esquerdo da produção, aparece somente uma variável. Entretanto, do lado direito, podem aparecer quaisquer combinações, já que $\alpha \in (V \cup T)^*$, ou seja, quaisquer combinações entre as variáveis e os símbolos terminais.

Gramática Livre de Contexto

No âmbito das Linguagens Formais, uma *gramática* $\mathcal{G}$ é um conjunto de regras que, quando aplicado a certos elementos (símbolos *não terminais* e *símbolos terminais*), nos permitem potencialmente gerar todas as cadeias (ou *strings*) que pertencem a uma determinada linguagem $\mathcal{L}$, e somente estas cadeias, nada mais.

Diz-se, por consequência, que a gramática $\mathcal{G}$ **gera** (ou **define**) a linguagem $\mathcal{L}$, denotando-se por $\mathcal{L}(\mathcal{G})$.

As *produções* de $\mathcal{G}$ relacionam as variáveis de $\mathcal{V}$ entre si e/ou os símbolos terminais de $\mathcal{T}$, estabelecendo as regras que guiarão a construção de cadeias da linguagem especificada pela gramática.

Uma produção tem a seguinte forma geral:

$$\alpha \rightarrow \beta$$

Gramática Livre de Contexto

onde $\alpha, \beta \in (\mathcal{V} \cup \mathcal{T})^*$, com $\alpha \neq \lambda$ e contendo pelo menos uma variável, enquanto que β pode inclusive λ (cadeia vazia).

Por convenção:

- Variáveis são denotadas por letras latinas maiúsculas e do início do alfabeto: A, B, C, ...;
- Terminais são representados por letras latinas minúsculas do início do alfabeto, dígitos e símbolos especiais: a, b, c, ..., 0, 1, 2, ..., *, +, etc;
- Formas sentenciais são indicadas por letras gregas minúsculas do início do alfabeto: α, β, γ , etc;
- Cadeias são denotadas por letras latinas do *final* do alfabeto: w, x, y, z, etc.

Exemplo

Seja a gramática $\mathcal{G} = (\mathcal{V}, \mathcal{T}, \mathcal{P}, S)$, com:

- $\mathcal{V} = \{S, A, B\};$
- $\mathcal{T} = \{a, b\};$
- $\mathcal{P} = \{S \rightarrow AB; A \rightarrow a \mid aA; B \rightarrow b \mid bB\}$

algumas cadeias possíveis:

- ab , usando:
 - 1 $S \rightarrow AB$, logo $\alpha = AB$;
 - 2 $A \rightarrow a$, logo $\alpha = aB$;
 - 3 $B \rightarrow b$, logo $\alpha = ab$.

Portanto $w = ab$ é uma cadeia gerada por $\mathcal{G}$.

Exemplo

- **aabb**, usando:

- ① $S \rightarrow AB$, logo $\alpha = AB$;
- ② $A \rightarrow aA$, logo $\alpha = aAB$;
- ③ $B \rightarrow bB$, logo $\alpha = aAbB$;
- ④ $A \rightarrow a$, logo $\alpha = aabB$;
- ⑤ $B \rightarrow b$, logo $\alpha = aabb$.

Portanto $w = aabb$ é uma cadeia gerada por $\mathcal{G}$.

- **aaab**, usando:

- ① $S \rightarrow AB$, logo $\alpha = AB$;
- ② $A \rightarrow aA$, logo $\alpha = aAB$;
- ③ $A \rightarrow aA$, logo $\alpha = aaAB$;
- ④ $A \rightarrow a$, logo $\alpha = aaaB$;
- ⑤ $B \rightarrow b$, logo $\alpha = aaab$.

Portanto $w = aaab$ é uma cadeia gerada por $\mathcal{G}$.

Exemplo

Diversas outras cadeias possíveis:

- abbb;
- abbbbbbbbbb;
- aaabbb;
- aaaaaabb;
- aaaaaabbb;

- não são capazes de gerar a cadeia vazia (ϵ ou λ);
- sempre geram a's à esquerda e b's à direita;
- conseguem gerar qualquer quantidade de a's, sendo UM a quantidade mínima;
- conseguem gerar qualquer quantidade de b's, sendo UM a quantidade mínima;
- as quantidades de a's e b's gerados não são necessariamente relacionadas.

Gramática Livre de Contexto

A partir propriedades anteriormente identificadas, pode-se dizer que:

" $\mathcal{G}$ é capaz de gerar cadeias que contenham uma sequência de a's à esquerda, seguida por uma sequência de b's à direita, sendo que cada uma das sequências (de a's e de b's) deve conter pelo menos um elemento."

Esta *definição* da linguagem gerada, dita $\mathcal{L}(\mathcal{G})$, é formalmente expressa por:

$$\mathcal{L}(\mathcal{G}) = \{w \in \{a, b\}^* \mid w = a^n b^m; n, m \in \mathbb{N}^*\}.$$

$L_1 = \{a^n b^n \mid n \geq 0\}$ é uma LLC
com $G_1 = (\{S\}, \{a, b\}, S, P)$ tal que
 $P = \{ S \rightarrow aSb$
 $S \rightarrow \lambda \}$

Então as cadeias $\{\lambda, ab, aabb, aaabbb \dots\}$ pertencem a L_1

$L_2 = \{ww^R \mid w \in \{a, b\}^*\}$ é uma LLC
com $G_2 = (\{S\}, \{a, b\}, S, P)$ tal que
 $P = \{ S \rightarrow aSa$
 $S \rightarrow bSb$
 $S \rightarrow \lambda \}$

Então as cadeias $\{\lambda, aa, bb, abba, aabbbaa, aabbbbaa \dots\}$ pertencem a L_2

L_3 é composta de expressões aritméticas contendo parênteses balanceados, um operador e dois operandos:
com $G_3 = (\{S\}, \{+, *, (,), a\}, S, P)$ tal que
 $P = \{ S \rightarrow S + S$
 $S \rightarrow S * S$
 $S \rightarrow (S)$
 $S \rightarrow a \}$

Então as cadeias $\{a, a+a, a*a, (a+a), (a*a), (a+a)*a \dots\}$ pertencem a L_3

Gramática Livre de Contexto

- Exercícios

Seja a gramática $\mathcal{G} = (\mathcal{V}, \mathcal{T}, \mathcal{P}, S)$, com:

- $\mathcal{V} = \{S\}$;
- $\mathcal{T} = \{a, b\}$;
- $\mathcal{P} = \{S \rightarrow aSbb \mid aab\}$.

Qual é a linguagem que $\mathcal{G}$ é capaz de gerar?

Seja a linguagem $\mathcal{L} = \{a^n b^{2n} \mid n \in \mathbb{N}^*\}$.

Qual é a gramática $\mathcal{G}$ que é capaz de gerar $\mathcal{L}$?

Gramática Livre de Contexto

- Resolução

A gramática $\mathcal{G} = (\mathcal{V}, \mathcal{T}, \mathcal{P}, S)$, com:

- $\mathcal{V} = \{S\};$
- $\mathcal{T} = \{a, b\};$
- $\mathcal{P} = \{S \rightarrow aSbb \mid abb\}.$

é capaz de gerar a linguagem $\mathcal{L}$ especificada.

Árvore de Derivação

Seja uma Gramática $G = (V, T, S, P)$.

Uma **Derivação à Esquerda** é aquela que, em cada passo, a variável $A \in V$, mais à esquerda, é substituída.

Uma **Derivação à Direita** é aquela que, em cada passo, a variável $A \in V$, mais à direita, é substituída.

Exemplo

Seja $L = \{a^n b^m \mid n \neq m\}$, uma linguagem livre de contexto, com $G = (\{S, T, A, B\}, \{a, b\}, S, P)$, onde as produções são da forma:

$$S \rightarrow AT \mid TB$$

$$T \rightarrow aTb \mid \lambda$$

$$A \rightarrow aA \mid a$$

$$B \rightarrow bB \mid b$$

Verifique as derivações abaixo:

Derivação 01: $S \rightarrow AT \rightarrow aAT \rightarrow aaT \rightarrow aaaTb \rightarrow aaab$

Derivação 02: $S \rightarrow AT \rightarrow AaTb \rightarrow Aab \rightarrow aAab \rightarrow aaab$

Árvore de Derivação

Pela Definição tem-se que a Derivação 01 é uma derivação à esquerda e a Derivação 02 é uma derivação à direita. Percebe-se que as duas derivações são diferentes e resultam na mesma cadeia. Isso quer dizer que a gramática construída é ambígua. Esse conceito será mais bem formalizado adiante.

Para uma determinada Gramática Livre de Contexto, $G = (V, T, S, P)$, a representação da derivação das cadeias na forma de árvore, denominada **Árvore de Derivação**, é como segue:

- a) a raiz é o símbolo inicial da gramática, portanto S ;
- b) os vértices interiores obrigatoriamente são variáveis.

Se A é um vértice interior e $X_1, X_2, \dots, X_n$ são os filhos de A , então $A \rightarrow X_1 X_2 \dots X_n$ é uma produção da gramática e os vértices $X_1, X_2, \dots, X_n$ são ordenados da esquerda para a direita;

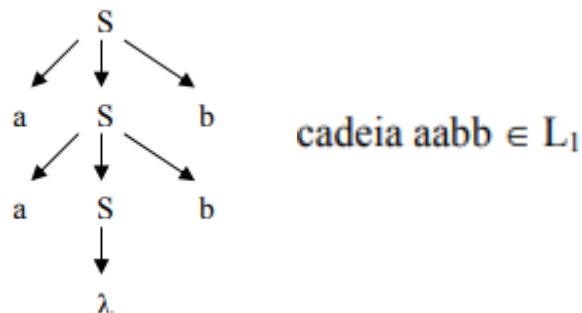
- c) um vértice folha é um símbolo terminal ou o símbolo λ (vazio).

Se a folha é o símbolo vazio, ele é o único filho, ou seja, $A \rightarrow \lambda$.

Exemplo

$L_1 = \{ a^n b^n \mid n \geq 0 \}$ uma LLC, com $G_1 = (\{S\}, \{a,b\}, S, P_1)$ tal que $P_1 = \{ S \rightarrow aSb, S \rightarrow \lambda \}$

Então a cadeia $aabb$ pertence a L_1 . A árvore de derivação é dada pela figura



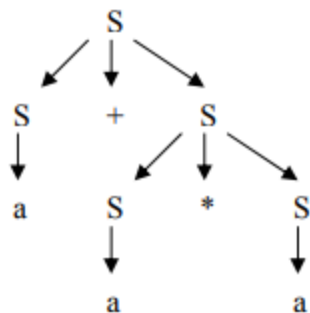
Árvore de Derivação

Exemplo

L_2 é composta de expressões aritméticas contendo parênteses balanceados, um operador e dois operandos,

com $G_2 = (\{S\}, \{+, *, (,), a\}, S, P_2)$ tal que $P_2 = \{S \rightarrow S + S \mid S * S \mid (S) \mid a \}$

Então a cadeia $a+a*a$ pertence a L_2 . A árvore de derivação é dada pela figura



cadeia $a+a*a \in L_2$

Próximos Assuntos

- Ambigüidade
- Linguagens não Ambíguas
- Linguagens Inerentemente Ambíguas
- Simplificação de Gramáticas de Livre Contexto
- Formas Normais
- Autômatos com Pilha