Aspectos Formais da Computação

Prof. Sergio D. Zorzo

Departamento de Computação - UFSCar

1º semestre / 2017

03

Linguagens Formais e Autômatos

Linguagens e Gramáticas

- Alfabeto
- Palavra
- Linguagem Formal
- Gramática
- Hierarquia de Chomsky

Alfabeto

Definição: Símbolo ou Caractere

entidade abstrata básica não definida formalmente

Ex de Símbolo

letras

dígitos

Definição: Alfabeto

conjunto *finito de símbolos*

Ex de Alfabeto

$\Sigma_1 = \{a, b, c\}$

$\Sigma_2 = \{0, 1, ..., 9\}$

$\Sigma_3 = \{\ \}$

Palavra

Definição: Palavra, Cadeia,Sentença sobre um Alfabeto

seqüência finita de símbolos justapostos

Ex: a, abcb são palavras sobre o alfabeto {a, b, c}

ε - palavra vazia - sem símbolos

ε - é palavra sobre qualquer alfabeto

Definição: Tamanho ou Comprimento de uma palavra (representado por |palavra|)

número de símbolos que compõem a palavra

Ex: abcb sobre o alfabeto {a,b,c}

$|abcb| = 4$

$| \varepsilon | = 0$

Conjuntos de palavras sobre o alfabeto Σ

Σ^* - conjunto de todas as palavras sobre Σ

$\Sigma + = \Sigma^* - \{\varepsilon\}$

Ex: para $\Sigma = \{a, b\}$

$\Sigma + = \{a, b, aa, ab, ba, bb, aaa, \ldots\}$

$\Sigma^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, \ldots\}$

Definição: Prefixo, Sufixo, Subpalavra

prefixo (sufixo) - qualquer seqüência de símbolos inicial (final) de uma palavra

Subpalavra - qualquer seqüência de símbolos contígüa de uma palavra

Ex: *para a palavra abcb sobre o alfabeto {a,b,c}*

Prefixos de abcb: ε, a, ab, abc, abcb

Sufixos de abcb: ε, b, cb, bcb, abcb

prefixos e sufixos são subpalavras

Linguagem Formal

Definição: Linguagem Formal

um conjunto de palavras sobre um alfabeto

Exs: Linguagem Formal sobre o alfabeto $\Sigma = \{a, b\}$

conjunto vazio { }

conjunto formado pela palavra vazia { ε }

conjunto das palíndromos (palavras que têm a mesma leitura da esquerda para a direita e vice-versa)

{ ε , a , b , aa , bb , aaa , aba , bbb , bab , ... }

é uma linguagem infinita

Definição: Concatenação

Operação binária, definida sobre uma linguagem palavra formada pela justaposição das palavras

Notação - justaposição dos símbolos que representam as palavras componentes

Satisfaz às seguintes propriedades:

associatividade: v(wt) = (vw)t

elemento neutro (esq/dir): εw = w = wε

Ex: para v = ab e w = cd sobre o alfabeto {a,b,c,d}

vw = abcd vv = abab

Definição: Concatenação Sucessiva

concatenação sucessiva de uma palavra com ela mesma

Exemplo:

w^3 = www w^1 = w a^5 = aaaaa

a^n = aaa...a (a repetido n vezes) $w^0 = \varepsilon$ para $w \neq \varepsilon$

Gramática

Definição: Gramática $G = (V,T,P,S)$

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

V - conjunto finito de símbolos (*variáveis ou não-terminais)*

T - conjunto finito de símbolos (*terminais* - disjunto de V)

P - conjunto finito de pares (α, β) (regra de produção)

α é palavra de $(V \cup T)^+$

β é palavra de $(V \cup T)^*$

S - elemento de V (*variável inicial*)

Notação de (α, β)

$\alpha \rightarrow \beta$

notação abreviada para $\alpha \rightarrow \beta_1, \ldots, \alpha \rightarrow \beta_n$

$\alpha \rightarrow \beta_1 \mid \ldots \mid \beta_n$

Definição: Derivação

Seja $G = (V, T, P, S)$ uma gramática

Derivação é um par da relação denotada por $\Rightarrow$ com domínio em $(V \cup T)^+$ e contra-domínio em $(V \cup T)^*$

$\alpha \Rightarrow \beta$

$\Rightarrow$ é indutivamente definida

a) para qq produção $S \rightarrow \beta$ (S é o símbolo inicial)

$S \Rightarrow \beta$

b) para qq par $\alpha \Rightarrow \beta$ onde $\beta = \beta_u\beta_v\beta_w$

se $\beta_v \rightarrow \beta_t$ é regra de produção de P então

$\beta \Rightarrow \beta_u\beta_t\beta_w$

Derivação é uma substituição de uma subpalavra de acordo com uma regra de produção

Definição: Sucessivos Passos de Derivações

$\Rightarrow^*$

fecho transitivo e reflexivo da relação $\Rightarrow$

zero ou mais passos de derivações sucessivos

$\Rightarrow^+$

fecho transitivo da relação $\Rightarrow$

um ou mais passos de derivações sucessivos

$\Rightarrow^i$

exatos i passos de derivações sucessivos

i é número natural

Gramática é um formalismo Axiomático de Geração

permite derivar ("gerar") todas as palavras da linguagem que representa

Definição: Linguagem Gerada por uma Gramática

$G = (V, T, P, S)$ uma gramática

Linguagem Gerada por G, denotado por L(G) ou GERA(G)

todas as palavras de símbolos terminais deriváveis a partir do símbolo inicial S

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

Ex: números naturais

G = (V, T, P, S)

V = {S, D}

T = {0, 1, 2,..., 9}

P = {$S \to D$, $S \to DS$, $D \to 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$}

uma derivação do número 243 (existe outra?)

$S \Rightarrow DS \Rightarrow 2S \Rightarrow 2DS \Rightarrow 24S \Rightarrow 24D \Rightarrow 243$

portanto

$S \Rightarrow^* 243$

$S \Rightarrow^+ 243$

$S \Rightarrow^6 243$

logo ...

GERA(G) ou L(G) é o conjunto dos números naturais

Definição: Equivalência de Gramáticas

G1 e G2 são equivalentes se e somente se

GERA(G1) = GERA(G2)

Convenções:

A, B, C,..., S, T símbolos variáveis

a, b, c,..., s, t símbolos terminais

u, v, w, x, y, z palavras de símbolos terminais

α, ß,... palavras de símbolos variáveis e/ou terminais

Ex: identificadores em Pascal

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

$V = \{S, C, L, D\}$

$T = \{a, b, ..., z, 0, 1, 2,..., 9\}$

$P = \{\ S \rightarrow LC \mid L,$

$C \rightarrow LC \mid DC \mid L \mid D,$

$L \rightarrow a \mid b \mid ... \mid z,$

$D \rightarrow 0 \mid 1 \mid ... \mid 9\ \}$

Ex: texto com aspas balanceadas

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

$V = \{ S \}$

$T = \{ x , " \}$

$P = \{ S \rightarrow xS \mid \varepsilon,$

$S \rightarrow "S" \}$

Ex: L(G) = {ww | w é palavra de {a, b}*}

G = ({S, X, Y, A, B, F}, {a, b}, P, S)

P = { $S \rightarrow XY$,

$X \rightarrow XaA \mid XbB \mid F$,

$Aa \rightarrow aA$, $Ab \rightarrow bA$, $AY \rightarrow Ya$,

$Ba \rightarrow aB$, $Bb \rightarrow bB$, $BY \rightarrow Yb$,

$Fa \rightarrow aF$, $Fb \rightarrow bF$, $FY \rightarrow \varepsilon$ }

Geração da cadeia baba

$S \Rightarrow XY \Rightarrow XaAY \Rightarrow XaYa \Rightarrow XbBaYa \Rightarrow XbaBYa \Rightarrow$

$XbaYba \Rightarrow FbaYba \Rightarrow bFaYba \Rightarrow baFYba \Rightarrow$

$ba\varepsilon ba = baba$

Definição: Sentença de uma Linguagem L(G)

uma cadeia w $\in$ T* é uma sentença da gramática G = (V, T, P, S) se e somente se S $\Rightarrow^*$ w, ou seja, w é uma cadeia formada apenas de símbolos terminais (pertencentes ao alfabeto da linguagem T) e pode ser obtida a partir do símbolo reservado S da gramática G por meio de sucessivas derivações.

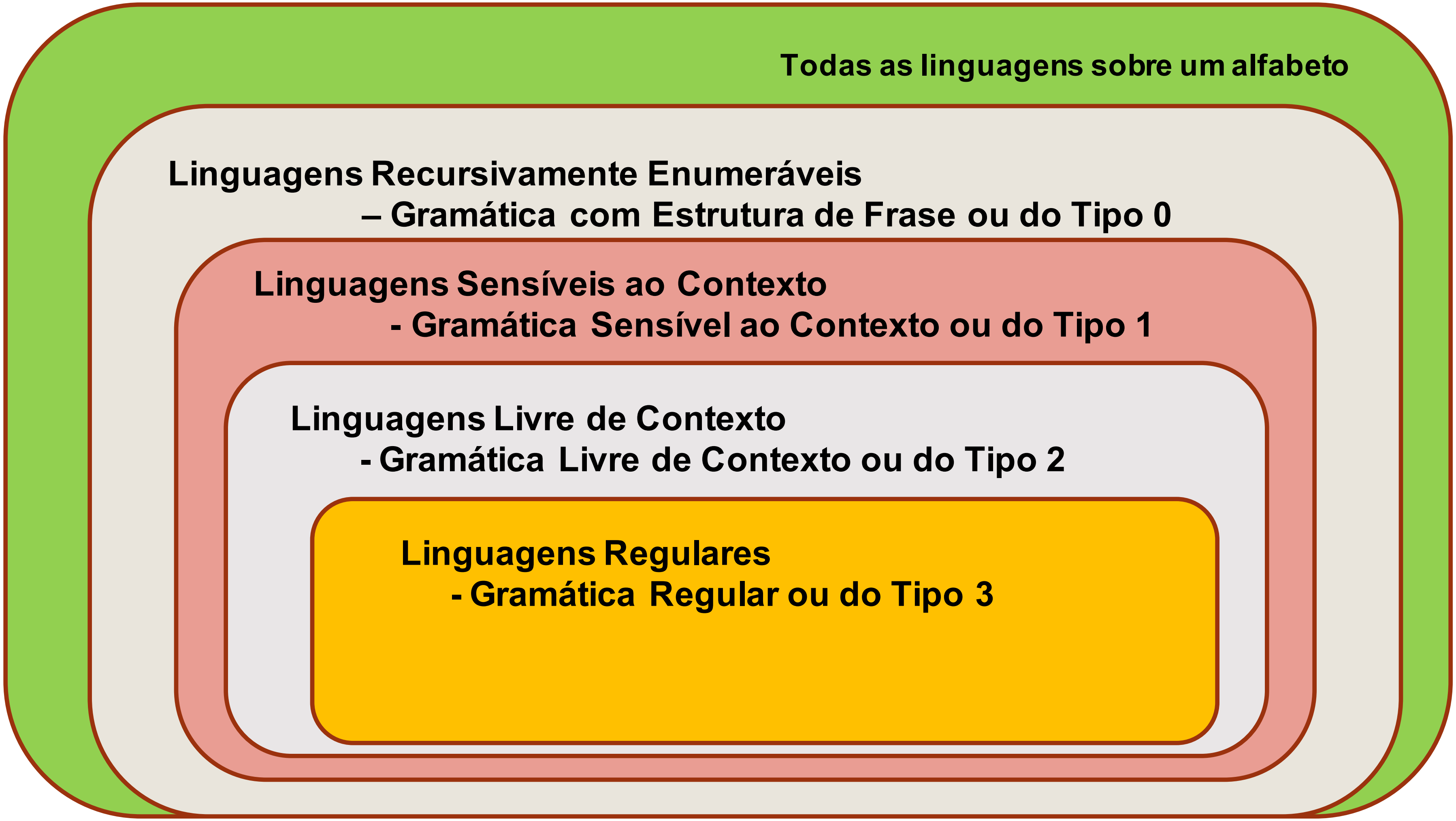
Definição: Forma Sentencial

uma cadeia $\alpha \in$ (V U T)* é uma forma sentencial da gramática G = (V, T, P, S) se e somente se S $\Rightarrow^*$ α, ou seja, α é um “embrião” para alguma sentença gerada pela gramática, ou a própria sentença.

Hierarquia de Chomsky

Hierarquia de Chomsky

A cada classe de linguagem da Hierarquia de Chomsky é associado um tipo de gramática



Definição: Gramática com Estrutura de Frase (GEF) ou do Tipo 0

Uma gramática G = (V, T, P, S) é dita ser do Tipo 0 ou com Estrutura de Frase se todas as regras de produção $\alpha \rightarrow \beta$ são da forma:

$$\alpha \in (V \cup T)^+ \quad e \quad \beta \in (V \cup T)^*$$

ou seja, as cadeias α e β são formadas por símbolos definidos na gramática (terminais ou não terminais) e a cadeia α não pode ser vazia.

Os próximos tipos de gramática são gramáticas com restrições.

Definição: Gramática Sensível ao Contexto (GSC) ou do Tipo 1

Uma gramática $G = (V, T, P, S)$ é dita ser do Tipo 1 ou Sensível ao Contexto se todas as regras de produção

$\alpha \rightarrow \beta$ são da forma:

$$\alpha \in (V \cup T)^+ \ \text{ e } \ \beta \in (V \cup T)^*$$

$$\text{e } |\alpha| \leq |\beta| \ (\text{exceto quando } \beta = \varepsilon)$$

ou seja, as cadeias α e β são formadas por símbolos definidos na gramática (terminais ou não terminais) e a cadeia α tem que ter comprimento menor que a cadeia β, com exceção quando β for vazia.

Há autores que classificam esse tipo de gramática com produções da forma:

$$\alpha A \beta \rightarrow \alpha \mu \beta \text{ com } \alpha, \beta, \mu \in (V \cup T)^* \text{ e } A \in V$$

Exemplo: Gramática Sensível ao Contexto (GSC) ou do Tipo 1

$G_1 = (V, T, P, S)$

$V = \{ S, A, B, C \}$

$T = \{ a, b, c \}$

$$
\begin{aligned}
P = \{\ & S \rightarrow aSBC , \\
& S \rightarrow aBC , \\
& CB \rightarrow BC , \\
& aB \rightarrow ab , \\
& bB \rightarrow bb , \\
& bC \rightarrow bc , \\
& cC \rightarrow cc \ \}
\end{aligned}
$$

$$\mathbf{L(G_1) = \{ a^n\, b^n\, c^n \mid n \geq 1 \}}$$

Definição: Gramática Livre de Contexto (GLC) ou do Tipo 2

Uma gramática $G = (V, T, P, S)$ é dita ser do Tipo 2 ou Livre de Contexto se todas as regras de produção $\alpha \rightarrow \beta$ são da forma:

$$\alpha \in V \text{ e } \quad \beta \in (V \cup T)^*$$

ou seja, as cadeias α e β são formadas por símbolos definidos na gramática (terminais ou não terminais) e a cadeia α tem que ser um símbolo não terminal.

Exemplo: Gramática Livre de Contexto (GLC) ou do Tipo 2

$G_2 = (V, T, P, S)$

$V = \{ S, A , B \}$

$T = \{ 0 , 1 \}$

$P = \{ S \rightarrow AB ,$

$A \rightarrow 0A11 ,$

$A \rightarrow 1 ,$

$B \rightarrow 0B ,$

$B \rightarrow 1 \}$

$$\mathbf{L(G_2) = \{ 0^n\ 1^{2n}\ 1\ 0^m\ 1 \mid n,m \geq 0 \}}$$

Exemplo: Gramática Livre de Contexto ou do Tipo 2

$G_3 = (V, T, P, S)$

$V = \{ S, A, B \}$ $\quad T = \{ a, b \}$

$$P = \{ \quad \begin{aligned} S &\rightarrow aB \\ S &\rightarrow bA \\ A &\rightarrow a \\ A &\rightarrow aS \\ A &\rightarrow bAA \\ B &\rightarrow b \\ B &\rightarrow bS \\ B &\rightarrow aBB \end{aligned} \quad \}$$

$L(G_3) = \{ w \in T^*$ e w contém o mesmo nro de a´s e b´s$\}$ ou

$L(G_3) = \{ w \in T^* \mid |w|_a = |w|_b \}$

Definição: Gramática Regular (GR) ou do Tipo 3

Uma gramática G = (V, T, P, S) é dita ser do Tipo 3 ou Regular se todas as regras de produção $\alpha \rightarrow \beta$ são da forma:

$$\alpha \in V \text{ e } \quad \beta \in T \cup (V \times T)$$

ou seja, as cadeias α e β são formadas por símbolos definidos na gramática (terminais ou não terminais) e a cadeia α tem que ser um símbolo não terminal e a cadeia β tem que ser um símbolo terminal ou um símbolo terminal seguido por um símbolo não terminal.

Definição: Gramática Regular (GR) ou do Tipo 3

Uma gramática G = (V, T, P, S) é dita ser do Tipo 3 ou Regular se todas as regras de produção $\alpha \rightarrow \beta$ são da forma:

$$\alpha \in V \text{ e } \quad \beta \in T \cup (V \times T)$$

Definição: Gramática Regular (GR) ou do Tipo 3

pode ser descrita por uma gramática linear

Tipos de gramática linear:

- **Gramática Linear à Direita – GLD**
- **Gramática Linear à Esquerda – GLE**
- **Gramática Linear Unitária à Direita – GLUD**
- **Gramática Linear Unitária à Esquerda – GLUE**

Definição: Gramática Regular (GR) ou do Tipo 3

Uma gramática G = (V, T, P, S) é dita ser do Tipo 3 ou Regular se for descrita por uma Gramática Linear **(GLD ou GLE ou GLUD ou GLUE)**

Uma gramática é uma GLD se as produções são da forma

$A \rightarrow wB$ ou $A \rightarrow w$

Uma gramática é uma GLE se as produções são da forma

$A \rightarrow Bw$ ou $A \rightarrow w$

Uma gramática é uma GLUD se as produções são da forma

$A \rightarrow wB$ ou $A \rightarrow w$, com $|w| \leq 1$

Uma gramática é uma GLUE se as produções são da forma

$A \rightarrow Bw$ ou $A \rightarrow w$, com $|w| \leq 1$

com $A, B \in V$ e $w \in T^*$

Exemplo: Gramática Regular ou do Tipo 3

$G_4 = (V, T, P, S)$

$V = \{ S, C \}$ $\quad T = \{ a , b , c \}$

$P = \{ \quad S \rightarrow aS ,$

$S \rightarrow bC ,$

$C \rightarrow c \}$

$L(G_4) = \{ a^n \, b \, c \mid n \geq 0 \}$

O padrão GLUD é o mais utilizado e será empregado neste curso.

Classe de Linguagens e Gramáticas

Uma linguagem L é do tipo 3 (ou 2 ou 1 ou 0)

sse

existir uma gramática G = (V, T, P, S) do tipo 3 (ou 2 ou 1 ou 0) que gera L ou seja, L = L(G)

O tipo da Linguagem é determinado pela menor classe da Gramática que a gera.

Fim