

# Linguagens Formais e Autômatos

**P. Blauth Menezes**

blauth@inf.ufrgs.br

**Departamento de Informática Teórica  
Instituto de Informática / UFRGS**



# Linguagens Formais e Autômatos

P. Blauth Menezes

- 1 Introdução e Conceitos Básicos
- 2 Linguagens e Gramáticas
- 3 Linguagens Regulares
- 4 Propriedades das Linguagens Regulares
- 5 Autômato Finito com Saída
- 6 Linguagens Livres do Contexto
- 7 Propriedades e Reconhecimento das Linguagens Livres do Contexto
- 8 Linguagens Recursivamente Enumeráveis e Sensíveis ao Contexto
- 9 Hierarquia de Classes e Linguagens e Conclusões

# 2 – Linguagens e Gramáticas

**2.1 Alfabeto**

**2.2 Palavra**

**2.3 Linguagem Formal**

**2.4 Gramática**

# 2 Linguagens e Gramáticas

## ◆ Linguagem: Dicionário Aurélio

*o uso da palavra articulada ou escrita como meio de expressão e comunicação entre pessoas*

## ◆ Não é suficientemente precisa para

- desenvolvimento matemático de uma teoria baseada em linguagens

## ◆ Linguagem

- conceito fundamental em Computação e Informática

## ◆ Para definir linguagem

- alfabeto
- palavra ou cadeia de caracteres

# 2 – Linguagens e Gramáticas

**2.1 Alfabeto**

**2.2 Palavra**

**2.3 Linguagem Formal**

**2.4 Gramática**

# 2.1 Alfabeto

## ◆ Símbolo ou Caractere

- entidade **abstrata** básica, **não definida** formalmente
- base para definições
- **exemplos**: letras e dígitos

## Def: Alfabeto

Conjunto **finito** de **símbolos** ou **caracteres**

## ◆ Portanto

- conjunto infinito **não** é alfabeto
- $\emptyset$  é um alfabeto

## Exp: Alfabeto

São alfabetos

- $\{ a, b, c \}$
- $\emptyset$  (conjunto vazio)

Não são alfabetos (por quê?)

- $\mathbf{N}$  (conjunto dos números naturais)
- $\{ a, b, aa, ab, ba, bb, aaa, \dots \}$

## Exp: Alfabeto: Linguagem de Programação

Alfabeto de uma linguagem de programação como Pascal

- o conjunto de todos os símbolos usados nos programas
  - \* letras
  - \* dígitos
  - \* caracteres especiais como “>”, “/”, etc
  - \* espaço ou “branco”

### ◆ Alfabeto binário { a, b }

- domínio de valores de um bit
- analogia com a representação interna dos computadores reais
- poucos símbolos: simplifica as diversas abordagens desenvolvidas.

# 2 – Linguagens e Gramáticas

**2.1 Alfabeto**

**2.2 Palavra**

**2.3 Linguagem Formal**

**2.4 Gramática**

## 2.2 Palavra

### Def: Palavra, Cadeia de Caracteres, Sentença

Sobre um alfabeto

- seqüência finita de símbolos justapostos

#### ◆ Cadeia sem símbolos

$\epsilon$  - cadeia vazia ou palavra vazia

## Def: Prefixo, Sufixo, Subpalavra

### Prefixo (Sufixo)

- qualquer seqüência inicial (final) de símbolos da palavra

### Subpalavra

- qualquer seqüência de símbolos contíguos da palavra

## Exp: Palavra, Prefixo, Sufixo, Subpalavra

abcb palavra sobre o alfabeto { a, b, c }

- $\epsilon$ , a, ab, abc, abcb são todos os prefixos
- $\epsilon$ , b, cb, bcb, abcb são todos os sufixos
- qualquer prefixo ou sufixo é uma subpalavra

# Exp: Palavra: Linguagem de Programação

Em uma linguagem de programação como Pascal

uma palavra é um programa

## Def: Concatenação de Palavras

Concatenação de Palavras ou simplesmente Concatenação

- operação binária sobre um conjunto de palavras
- associa a cada par de palavras
  - \* palavra formada pela justaposição da primeira com a segunda

Notação

- justaposição dos símbolos que representam as palavras componentes

Propriedades

- Elemento Neutro:  $\epsilon w = w = w \epsilon$
- Associativa:  $v(wt) = (vw)t$

Associatividade - parênteses podem ser omitidos:  $vw t$

## Exp: Concatenação de Palavras

$\Sigma = \{ a, b \}$  um alfabeto. Para  $v = baaaa$  e  $w = bb$

- $v w = baaaabb$
- $v \varepsilon = v = baaaa$

## Def: Concatenação Sucessiva de uma Palavra

*Concatenação Sucessiva de uma Palavra* (com ela mesma) ou simplesmente *Concatenação Sucessiva*

$w^n$  onde  $n$  é o número de concatenações sucessivas

indutivamente a partir da operação de concatenação

- $w^0 = \varepsilon$
- $w^n = w w^{n-1}$ , para  $n > 0$

## Exp: Concatenação Sucessiva

$w$  palavra e  $a$  símbolo

- $w^3 = w w w$
- $w^1 = w$
- $a^5 = aaaaa$
- $a^n = aaa...a$  (o símbolo  $a$  repetido  $n$  vezes)

## ◆ Se $\Sigma$ é um alfabeto

- $\Sigma^*$  conjunto de **todas** as **palavras** possíveis sobre  $\Sigma$
- $\Sigma^+ = \Sigma^* - \{\epsilon\}$

## Def: Conjunto de Todas as Palavras

$\Sigma$  alfabeto.  $\Sigma^*$  é indutivamente definido

### *Base de Indução*

- $\epsilon \in \Sigma^*$
- para qualquer  $x \in \Sigma$ , vale  $x \in \Sigma^*$

### *Passo de Indução*

- Se  $u$  e  $v$  são palavras de  $\Sigma^*$ ,
- então a concatenação  $uv$  é uma palavra de  $\Sigma^*$

## ◆ Definição alternativa para palavra sobre um alfabeto $\Sigma$

- qualquer elemento  $w$  de  $\Sigma^*$

$$w \in \Sigma^*$$

### Exp: Conjunto de Todas as Palavras

Se  $\Sigma = \{ a, b \}$ , então:

- $\Sigma^+ = \{ a, b, aa, ab, ba, bb, aaa, \dots \}$
- $\Sigma^* = \{ \epsilon, a, b, aa, ab, ba, bb, aaa, \dots \}$

## Def: Comprimento, Tamanho de uma Palavra

De uma palavra  $w$ , representado por  $|w|$

- número de símbolos que compõem a palavra
- função com domínio em  $\Sigma^*$  e codomínio em  $\mathbf{N}$

## Exp: Palavra, Prefixo, Sufixo, Comprimento

$$|abcb| = 4$$

$$|\varepsilon| = 0$$

# 2 – Linguagens e Gramáticas

**2.1 Alfabeto**

**2.2 Palavra**

**2.3 Linguagem Formal**

**2.4 Gramática**

## 2.3 Linguagem Formal

### Def: Linguagem Formal

Linguagem Formal ou simplesmente Linguagem  $L$  sobre um alfabeto  $\Sigma$

$$L \subseteq \Sigma^*$$

### Exp: Linguagem Formal

$\emptyset$  e  $\{\varepsilon\}$  são linguagens sobre qualquer alfabeto

$$\emptyset \neq \{\varepsilon\}$$

$\Sigma^*$  e  $\Sigma^+$  são linguagens sobre um  $\Sigma$  qualquer

$$\Sigma^* \neq \Sigma^+$$

Conjunto de palíndromos sobre  $\Sigma = \{a, b\}$

$\varepsilon, a, b, aa, bb, aaa, aba, bab, bbb, aaaa, \dots$

## Exp: Conjunto de Todas as Linguagens Sobre um Alfabeto

Conjunto das partes de  $\Sigma^*$

$$2^{\Sigma^*}$$

## Exp: Linguagem Formal: Linguagem de Programação

Linguagem de programação como Pascal

conjunto de todos os programas (palavras) da linguagem

# 2 – Linguagens e Gramáticas

**2.1 Alfabeto**

**2.2 Palavra**

**2.3 Linguagem Formal**

**2.4 Gramática**

## 2.4 Gramática

### ◆ Linguagem de programação

- definida pelo conjunto de **todos** os **programas** (palavras)

### ◆ Linguagem de propósitos gerais como Pascal

- **conjunto** de todos os **programas** é *infinito*
- não é definição adequada para implementação em computador

### ◆ Formalismo Gramática

- uma maneira de **especificar** de **forma finita** linguagens (eventualmente) *infinitas*

## ◆ Gramática é, basicamente

- conjunto *finito* de regras
- quando aplicadas sucessivamente, geram palavras
- conjunto de todas as palavras geradas por uma gramática
  - \* define a linguagem

## ◆ Gramáticas para linguagens naturais como Português

- as mesmas que as usadas para linguagens artificiais como Pascal

## ◆ Gramáticas também são usadas para definir semântica

- entretanto, em geral, são usados outros formalismos

## Def: Gramática

Gramática de Chomsky, Gramática Irrestrita ou Gramática

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

- $V$ , conjunto *finito* de símbolos **variáveis** ou **não-terminais**
- $T$ , conjunto *finito* de símbolos **terminais** disjunto de  $V$
- $P: (V \cup T)^+ \rightarrow (V \cup T)^*$ , relação *finita*: **Produções**
  - \* par da relação: **regra de produção** ou **produção**
- $S$ , elemento distinguido de  $V$ : **símbolo inicial** ou **variável inicial**

Representação de uma regra de produção  $(\alpha, \beta)$

$$\alpha \rightarrow \beta$$

Representação abreviada para  $\alpha \rightarrow \beta_1, \alpha \rightarrow \beta_2, \dots, \alpha \rightarrow \beta_n$

$$\alpha \rightarrow \beta_1 \mid \beta_2 \mid \dots \mid \beta_n$$

## ◆ Derivação

- aplicação de uma regra de produção é denominada derivação
- aplicação sucessiva de regras de produção
  - \* fecho transitivo da relação de derivação
  - \* permite derivar palavras da linguagem

## Def: Relação de Derivação

$G = (V, T, P, S)$  gramática

Derivação é um par da Relação de Derivação denotada por  $\Rightarrow$

- domínio em  $(V \cup T)^+$  e codomínio em  $(V \cup T)^*$
- $\langle \alpha, \beta \rangle$  é representado de forma infixada

$$\alpha \Rightarrow \beta$$

$\Rightarrow$  é indutivamente definida como segue:

- para toda produção da forma  $S \rightarrow \beta$  ( $S$  é o símbolo inicial de  $G$ )

$$S \Rightarrow \beta$$

- para todo par  $\eta \Rightarrow \rho \alpha \sigma$  da relação de derivação
  - \* se  $\alpha \rightarrow \beta$  é regra de  $P$ , então

$$\eta \Rightarrow \rho \beta \sigma$$

## ◆ Portanto, derivação

- substituição de uma subpalavra
- de acordo com uma regra de produção

## ◆ Sucessivos passos de derivação

- $\Rightarrow^*$  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 ( $i$  natural)

## ◆ Gramática é um formalismo

- axiomático
- de geração
  - \* permite derivar ("gerar") todas as palavras da linguagem

### Def: Linguagem Gerada

$G = (V, T, P, S)$  gramática

Linguagem Gerada por  $G$ :  $L(G)$  ou  $GERA(G)$

- palavras de símbolos terminais deriváveis a partir de  $S$

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

## Exp: Gramática, Derivação, Linguagem Gerada: Números Naturais

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

- $V = \{ N, D \}$
- $T = \{ 0, 1, 2, \dots, 9 \}$
- $P = \{ N \rightarrow D, N \rightarrow DN, D \rightarrow 0 \mid 1 \mid \dots \mid 9 \}$

Gera, **sintaticamente**, o conjunto dos números naturais

- se **distinguem** os **zeros** à **esquerda**
- exemplo: **123** de **0123**

## Exp: ...Gramática, Derivação, Linguagem Gerada: Números Naturais

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

- $V = \{ N, D \}$
- $T = \{ 0, 1, 2, \dots, 9 \}$
- $P = \{ N \rightarrow D, N \rightarrow DN, D \rightarrow 0 \mid 1 \mid \dots \mid 9 \}$

Uma derivação do número 243

- |                     |                    |
|---------------------|--------------------|
| • $N \Rightarrow$   | $N \rightarrow DN$ |
| • $DN \Rightarrow$  | $D \rightarrow 2$  |
| • $2N \Rightarrow$  | $N \rightarrow DN$ |
| • $2DN \Rightarrow$ | $D \rightarrow 4$  |
| • $24N \Rightarrow$ | $N \rightarrow D$  |
| • $24D \Rightarrow$ | $D \rightarrow 3$  |
| • 243               |                    |

Portanto

- $S \Rightarrow^* 243$
- $S \Rightarrow^+ 243$
- $S \Rightarrow^6 243$

Interpretação indutiva da gramática

- *Base de Indução*: todo dígito é natural
- *Passo de Indução*: se  $n$  é natural, então a concatenação com qualquer dígito também é natural

## Exp: Gramática, Derivação, Linguagem Gerada: Palavra Duplicada

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

na qual:

- $P = \{ S \rightarrow XY,$
- $X \rightarrow XaA \mid XbB \mid F$
- $Aa \rightarrow aA, Ab \rightarrow bA, AY \rightarrow Ya,$
- $Ba \rightarrow aB, Bb \rightarrow bB, BY \rightarrow Yb,$
- $Fa \rightarrow aF, Fb \rightarrow bF, FY \rightarrow \varepsilon \}$

gera a linguagem

$$\{ ww \mid w \text{ é palavra de } \{ a, b \}^* \}$$

## Derivação de **baba**

- $S \Rightarrow$
- $XY \Rightarrow$
- $XaAY \Rightarrow$
- $XaYa \Rightarrow$
- $XbBaYa \Rightarrow$
- $XbaBYa \Rightarrow$
- $XbaYba \Rightarrow$
- $FbaYba \Rightarrow$
- $bFaYba \Rightarrow$
- $baFYba \Rightarrow$
- **baba**

$S \rightarrow XY$   
 $X \rightarrow XaA$   
 $AY \rightarrow Ya$   
 $X \rightarrow XbB$   
 $Ba \rightarrow aB$   
 $BY \rightarrow Yb$   
 $X \rightarrow F$   
 $Fb \rightarrow bF$   
 $Fa \rightarrow aF$   
 $FY \rightarrow \varepsilon$

Existe mais alguma derivação de **baba**?

## Def: Gramáticas Equivalentes

$G_1$  e  $G_2$  são Gramáticas Equivalentes se e somente se

$$\text{GERA}(G_1) = \text{GERA}(G_2)$$

### ◆ Convenções

- $A, B, C, \dots, S, T$  para símbolos variáveis
- $a, b, c, \dots, s, t$  para símbolos terminais
- $u, v, w, x, y, z$  para palavras de símbolos terminais
- $\alpha, \beta, \dots$  para palavras de símbolos variáveis ou terminais

# Linguagens Formais e Autômatos

P. Blauth Menezes

- 1 Introdução e Conceitos Básicos
- 2 Linguagens e Gramáticas
- 3 Linguagens Regulares
- 4 Propriedades das Linguagens Regulares
- 5 Autômato Finito com Saída
- 6 Linguagens Livres do Contexto
- 7 Propriedades e Reconhecimento das Linguagens Livres do Contexto
- 8 Linguagens Recursivamente Enumeráveis e Sensíveis ao Contexto
- 9 Hierarquia de Classes e Linguagens e Conclusões