Gramáticas

Prof^a Jerusa Marchi

Departamento de Informática e Estatística

Universidade Federal de Santa Catarina

e-mail: jerusa@inf.ufsc.br

Representação Finita de Linguagens

- Definimos uma **Linguagem** como um conjunto de cadeias sobre um alfabeto Σ, ou seja, qualquer subconjunto de $\Sigma^\star$

$$L \subseteq \Sigma^*$$

- Uma linguagem é *finita* se suas sentenças formam um conjunto finito, caso contrário, a linguagem é *infinita*
- É necessário representar linguagens por especificações finitas. Como?

Representação Finita de Linguagens

- Formas de representar linguagens
  - Enumeração - linguagens finitas são facilmente representadas de maneira finita (pode-se enumerar exaustivamente todas as suas *cadeias*)
  - Sistemas Geradores (gramáticas)
  - Sistemas Reconhecedores (autômatos)

Representação Finita de Linguagens

- É importante observar que o conjunto de todas as possíveis linguagens sobre um dado alfabeto Σ (isto é 2^{Σ}) é incontavelmente infinito
  - o conjunto das partes de qualquer conjunto contavelmente infinito é incontavelmente infinito
- Ou seja, temos um número contável de representações e um número incontável de coisas para representar
  - somente uma quantidade contável de linguagens pode ser representada

Representação Finita de Linguagens

- Exemplo: $L = \{w \mid \in \Sigma^\star = \{a, b\}$ e $\mid w \mid = 2\}$
  - enumeração

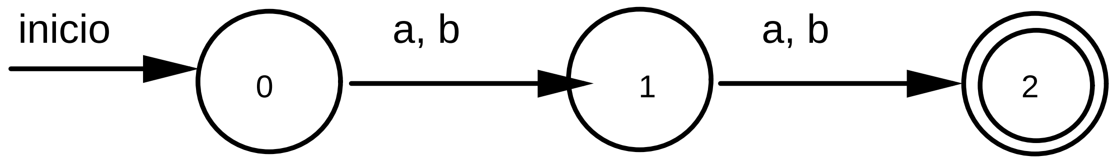
$$L = \{aa, ab, ba, bb\}$$

  - gramática

$$G \therefore \{ \quad \begin{array}{l} S ::= LL \\ L ::= a|b \quad \} \end{array}$$

Representação Finita de Linguagens

- Exemplo: $L = \{w \mid \in \Sigma^\star = \{a, b\}$ e $\mid w \mid = 2\}$
  - autômatos

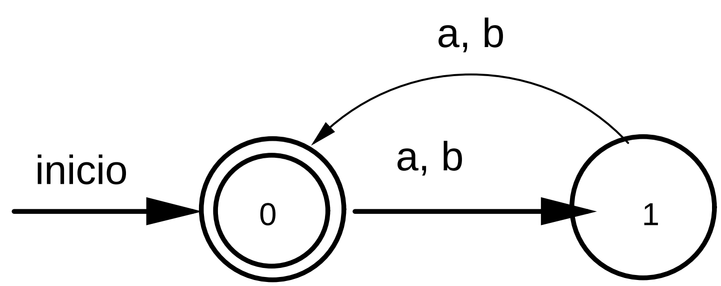


Representação Finita de Linguagens

- Exemplo: $L_1 = \{w \mid \in \Sigma^\star = \{a, b\}$ e $\mid w \mid$ é par $\}$
  - gramática

$$G_1 \therefore \{ \quad S ::= LLS | \varepsilon \\ L ::= a|b \quad \}$$

  - autômatos



Gramáticas

- Uma gramática define uma estrutura sobre um alfabeto de forma a permitir que apenas determinadas combinações sejam válidas, isto é, sejam consideradas sentenças (definindo assim uma linguagem)

Gramáticas

- De maneira informal, uma gramática pode ser definida como sendo:
  - um sistema gerador de linguagens;
  - um sistema de reescrita;
  - uma maneira finita de descrever (representar) uma linguagem;
  - um dispositivo formal usado para *especificar de maneira finita e precisa uma linguagem potencialmente infinita*.

Gramáticas

- Exemplo intuitivo:

$$
\begin{aligned}
<\text{sentença}> \quad & ::= \quad <\text{sujeito}><\text{predicado}> \\
<\text{sujeito}> \quad & ::= \quad <\text{substantivo}> \\
& \qquad \mid <\text{artigo}><\text{substantivo}> \\
& \qquad \mid <\text{artigo}><\text{adjetivo}><\text{substantivo}> \\
<\text{predicado}> \quad & ::= \quad <\text{verbo}><\text{objeto}> \\
<\text{substantivo}> \quad & ::= \quad \text{João} \mid \text{Maria} \mid \text{cachorro} \mid \text{livro} \mid \text{pão} \\
<\text{artigo}> \quad & ::= \quad \text{o} \mid \text{a} \\
<\text{adjetivo}> \quad & ::= \quad \text{pequeno} \mid \text{bom} \mid \text{bela} \\
<\text{verbo}> \quad & ::= \quad \text{morde} \mid \text{le} \mid \text{olha} \\
<\text{objeto}> \quad & ::= \quad <\text{substantivo}> \mid <\text{artigo}><\text{substantivo}> \\
& \qquad \mid <\text{artigo}><\text{adjetivo}><\text{substantivo}>
\end{aligned}
$$

Gramáticas

- Notação utilizada:
  - $< \ldots >$: categoria sintática ou gramatical;
  - $::=$: definido por
  - $|$: ou (alternativa)
  - $\alpha ::= \beta$: regra de sintaxe (ou regra gramatical ou produção)
- "João olha a bela Maria"
- "O cachorro morde o pequeno pão"

Gramáticas: Definição Formal

- **Gramáticas** são constituídas por 4 componentes:
  - **Terminais**: conjunto finito de símbolos básicos que formam as palavras da linguagem (João, le, o, bom ...)
  - **Não-terminais** ou **variáveis sintáticas**: conjunto finito de variáveis que representam um conjunto de palavras (sentença, substantivo, objeto, artigo, adjetivo...).
  - **Símbolo inicial**: Uma variável do conjunto de não-terminais é distinguida como símbolo de partida (sentença).
  - **Produções**

Gramáticas: Definição Formal

- **Produções**: conjunto de regras que especifica como os símbolos terminais e não-terminais podem ser combinados para formar cadeias. Cada produção consiste em:
  - uma símbolo não-teminal chamado de cabeça da produção
  - o símbolo de produção ($::=$)
  - uma palavra de 0 ou mais símbolos terminais ou não-terminais, chamada de corpo da produção

$$< \text{sentença} > \quad ::= \quad < \text{sujeito} >< \text{predicado} >$$

Gramáticas: Definição Formal

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

- N - conjunto de não-terminais (variáveis)
- T - conjunto de terminais ($N \cap T = \emptyset$)
- P - conjunto finito de produções ($P \subseteq (N \cup T)^{+} \times (N \cup T)^{\star}$)
- S - símbolo inicial ($S \in N$)

Gramáticas

- Exemplos:
  - Dado o conjunto de terminais (alfabeto) $T = \{a, b\}$, defina as gramáticas para as seguintes linguagens:
    1. Sentenças que começam e terminam com a.
    2. Sentenças que tenham tamanho par.
    3. Sentenças que tenham tamanho ímpar.
    4. Sentenças na forma $a^n b^n$, com $n > 0$.

Derivação

- Usamos o símbolo $\Rightarrow$ para indicar a derivação de palavras a partir da cabeça para o corpo da produção.
  - Se $\alpha ::= \beta$ é uma produção em P e γ e δ são palavras em $(N \cup T)^\star$, então $\gamma\alpha\delta \Rightarrow \gamma\beta\delta$
    - $\Rightarrow$ significa "deriva em um passo"
    - $\stackrel{*}{\Rightarrow}$ significa "deriva em zero ou mais passos"

Derivação

- Exemplo: Dada a gramática seguinte:

$$S ::= 1A \mid 0S \mid 1$$
$$A ::= 0A \mid 1S$$

A sentença 100101 é gerada pela seguinte sequência de derivações:

$$S \Rightarrow 1A \Rightarrow 10A \Rightarrow 100A \Rightarrow 1001S \Rightarrow 10010S \Rightarrow 100101$$

Redução

- Usamos o símbolo $\Leftarrow$ para indicar a redução de palavras a partir do corpo para a cabeça da produção.
  - Se $\alpha ::= \beta$ é uma produção em P e γ e δ são palavras em $(N \cup T)^\star$, então $\gamma\alpha\delta \Leftarrow \gamma\beta\delta$
    - $\Leftarrow$ significa "reduz em um passo"
    - $\stackrel{*}{\Leftarrow}$ significa "reduz em zero ou mais passos"

Sentença e Forma Sentencial

- Sentença
  - sequência de terminais produzida a partir do símbolo inicial de G

$$S \stackrel{*}{\Rightarrow} w$$

- Forma Sentencial
  - sequência de terminais e/ou não-terminais produzida a partir do símbolo inicial de G
  - Se $S \Rightarrow \alpha \Rightarrow \beta \Rightarrow \cdots \Rightarrow \gamma \cdots$ então α, β, $\cdots$, γ, $\cdots$ são formas sentenciais de G

A Linguagem de uma Gramática

- Se $G = (N, T, P, S)$ é uma Gramática, a linguagem de G, denotada $L(G)$, é o conjunto de **senteças** que tenham derivação a partir do símbolo inicial

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

- Ou seja, uma palavra está em $L(G)$ se:
  - A palavra consiste somente de símbolos terminais
  - A palavra pode ser derivada a partir de S
- Duas gramáticas G_1 e G_2 são equivalentes se $L(G_1) = L(G_2)$

A Linguagem de uma Gramática

- Exemplo: Considere a gramática $G = (\{S\}, \{0, 1\}, P = \{S ::= 0S1, S ::= 01\}, S)$. Aplicando a primeira regra $n-1$ vezes e a segunda regra 1 vez, tem-se:

$$S \Rightarrow 0S1 \Rightarrow 00S11 \Rightarrow 0^3S1^3 \Rightarrow ... \Rightarrow 0^{n-1}S1^{n-1} \Rightarrow 0^n1^n$$

- Portanto, a linguagem $L(G) = \{0^n1^n \mid n \geq 1\}$

Tipos de Gramáticas

- Impondo restrições na forma das produções, pode-se identificar quatro tipos de gramáticas (Hierarquia de Chomsky)
  - Tipo 0 ou *Irrestrita*: Não há restrições nas produções. São o tipo mais geral.
    - $G = (N, T, P, S)$ onde $P = \{\alpha ::= \beta \mid \alpha \in (N \cup T)^{\star} N (N \cup T)^{\star}$ e $\beta \in (N \cup T)^{\star}\}$
    - Exemplo:

$$\begin{array}{ll} S ::= ABS \mid ab & Ba ::= aB \mid a \\ Aa ::= aa \mid a \mid \varepsilon & Bb ::= bb \end{array}$$

Tipos de Gramáticas

- Tipo 1 ou *Sensíveis ao Contexto*:
  - $G = (N, T, P, S)$ onde o símbolo inicial S não aparece no lado direito de nenhuma produção e $P = \{\alpha ::= \beta \mid |\alpha| \leq |\beta|$ e $\alpha \in (N \cup T)^\star N (N \cup T)^\star$ e $\beta \in (N \cup T)^+\}$
  - Exemplo:

$$
\begin{array}{ll}
S ::= aBC \mid aABC & A ::= aABC \mid aBC \\
CB ::= BC & aB ::= ab \\
bB ::= bb & bC ::= bc \\
cC ::= cc &
\end{array}
$$

Tipos de Gramáticas

- Tipo 2 ou *Livre de Contexto*:
  - $G = (N, T, P, S)$ onde $P = \{A ::= \delta,$ com $A \in N$ e $\delta \in (T \cup N)^*\}$
    - Exemplo:

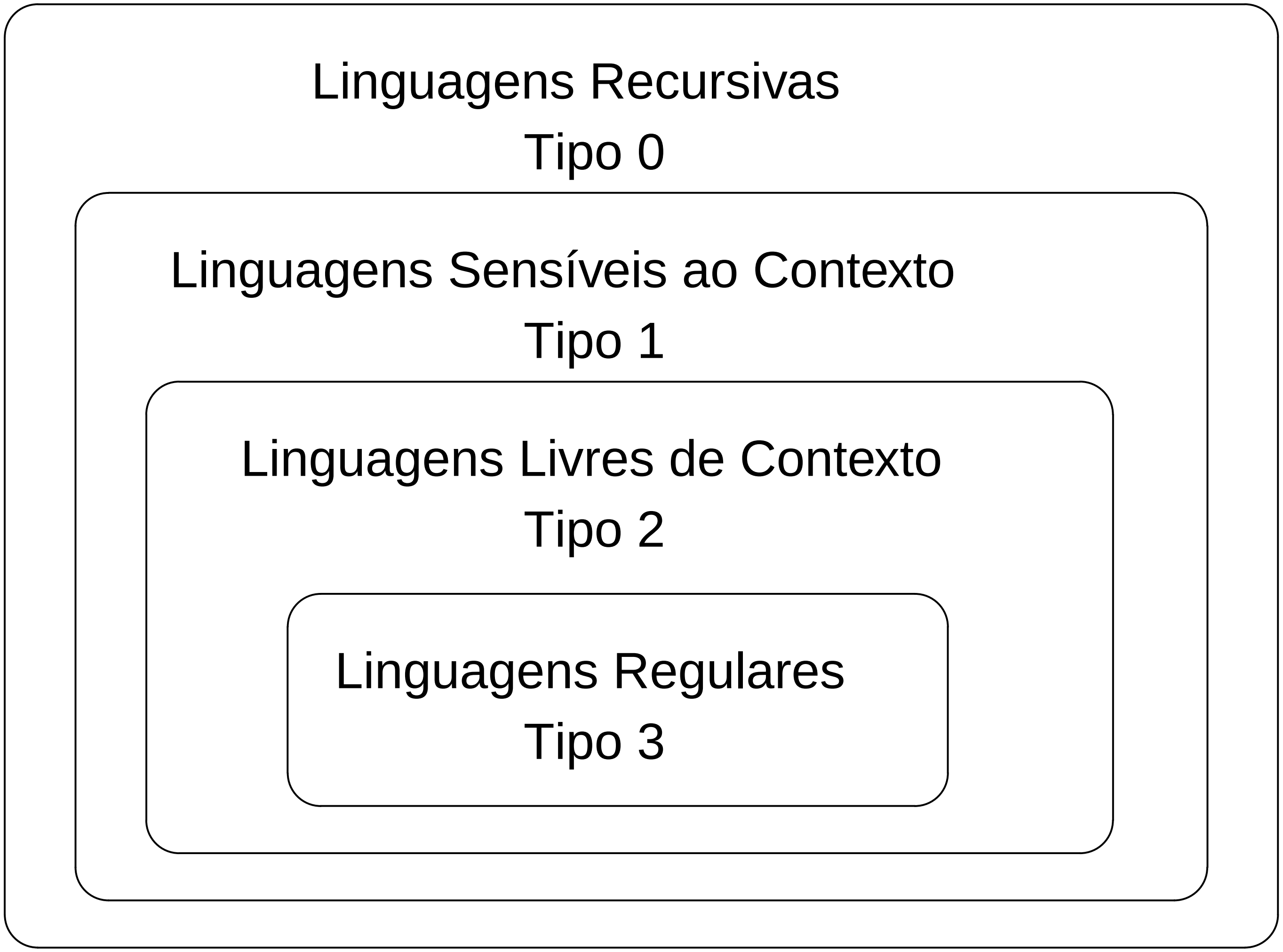
$$S ::= aSb \mid A \quad A ::= cAd \mid e$$

Tipos de Gramáticas

- Tipo 3 ou *Regular*
  - $G = (N, T, P, S)$ onde $P = \{A ::= aX \mid A \in N, a \in T$ e $X \in N \cup \{\varepsilon\}\}$
    - Exemplo:

$$S ::= aS \mid bB \quad B ::= bB \mid bC \quad C ::= cC \mid \varepsilon$$

Tipos de Linguagens



Tipos de Linguagens...

... e sua relação com geradores e reconhecedores

Hierarquia	Linguagem	Gramática	Autômato
Tipo 0	Recursiva	Irrestrita	MT
Tipo 1	Sensível ao Contexto	Sensível ao Contexto	MT
Tipo 2	Livre de Contexto	Livre de Contexto	Pilha
Tipo 3	Regular	Regular	Finito