AUTÓMATOS E LINGUAGENS FORMAIS

(LCC/LMAT)

3. Gramáticas Independentes de Contexto

Departamento de Matemática

Universidade do Minho

2022/2023

Uma gramática é um modelo matemático, que permite descrever uma linguagem, através de um mecanismo gerador dos elementos da linguagem, semelhante à definição indutiva de um conjunto.

Noam Chomsky introduziu na década de 1950 quatro tipos de gramáticas, que dão origem a uma hierarquia de quatro classes de linguagens, conhecida como Hierarquia de Chomsky:

Hierarquia de Chomsky	Linguagem	Reconhecedor
Tipo 0 (gramática irrestrita)	Recursivamente enumerável	Máquina de Turing
Tipo 1 (gramática dependente de contexto)	Dependente de contexto	Autómato linear limitado
Tipo 2 (gramática independente de contexto)	Independente de contexto	Autómato de pilha
Tipo 3 (gramática regular)	Regular	Autómato finito

Definição

Uma gramática é um quádruplo

$$\mathcal{G} = (V, A, \mathcal{S}, P)$$

onde

1. V é um alfabeto finito, dito não terminal, cujos elementos são chamados variáveis (ou símbolos não terminais);
2. A é um alfabeto finito, dito terminal, cujos elementos são chamados letras (ou símbolos terminais), tal que $V \cap A = \emptyset$;
3. $\mathcal{S}$ é um elemento de V, chamado o símbolo inicial;
4. P é um subconjunto finito de $\big((V \cup A)^* V (V \cup A)^*\big) \times (V \cup A)^*$. Os elementos (α, β) de P são chamados produções (ou regras gramaticais) e, habitualmente, representam-se por $\alpha \to \beta$.

Exemplo

O quádruplo
$$\mathcal{G} = \big(\{\mathcal{S}\}, \{a, b\}, \mathcal{S},\ \{(\mathcal{S}, \epsilon), (\mathcal{S}, a\mathcal{S}b)\}\big)$$
é uma gramática com uma única variável $\mathcal{S}$, duas letras a e b, e duas produções $\mathcal{S} \to \epsilon$ e $\mathcal{S} \to a\mathcal{S}b$, que, simplificadamente, podem ser representadas simplesmente por
$$\mathcal{S} \ \to\ \epsilon \,|\, a\mathcal{S}b\ .$$

As produções de uma gramática permitem transformar palavras de $(V \cup A)^*$ noutras palavras de $(V \cup A)^*$.

Por exemplo, na gramática acima, a palavra $a\mathcal{S}baa \in \{\mathcal{S}, a, b\}^*$ pode ser transformada em:

- $aa\mathcal{S}bbaa$ pela produção $\mathcal{S} \to a\mathcal{S}b$;
- $abaa$ pela produção $\mathcal{S} \to \epsilon$.

Considerando novamente a gramática do exemplo anterior, partindo do seu símbolo inicial $\mathcal{S}$, podemos obter sucessivamente:

1. $a\mathcal{S}b$, usando a produção $\mathcal{S} \to a\mathcal{S}b$;
2. $aa\mathcal{S}bb$, usando novamente a produção $\mathcal{S} \to a\mathcal{S}b$;
3. $aa\epsilon bb$, utilizando $\mathcal{S} \to \epsilon$.

Diz-se, então, que a gramática $\mathcal{G}$ gera a palavra $aabb$ de $\{a, b\}^*$.

A cada gramática $\mathcal{G} = (V, A, \mathcal{S}, P)$ associaremos uma linguagem $L(\mathcal{G})$ sobre o alfabeto A, dita a linguagem gerada por $\mathcal{G}$.

Por exemplo, a linguagem gerada pela gramática do exemplo anterior é:

$$\{a^n b^n : n \in \mathbb{N}_0\}.$$

Para definir rigorosamente o conceito de linguagem gerada por uma gramática, precisamos de formalizar previamente o conceito de derivação.

Definição

Sejam $\mathcal{G} = (V, A, \mathcal{S}, P)$ uma gramática e $\sigma_1, \sigma_2 \in (V \cup A)^*$. Diz-se que:

1. σ_2 **deriva directamente** de σ_1, e escreve-se $\sigma_1 \underset{\mathcal{G}}{\Rightarrow} \sigma_2$, quando

$$\sigma_1 = \gamma\alpha\delta,\ \sigma_2 = \gamma\beta\delta \text{ e } \alpha \to \beta \text{ é uma produção de } \mathcal{G}.$$

2. σ_2 **deriva em $k \in \mathbb{N}$ passos** de σ_1, e escreve-se $\sigma_1 \overset{k}{\underset{\mathcal{G}}{\Rightarrow}} \sigma_2$, quando

$$\sigma_1 = \sigma'_0 \underset{\mathcal{G}}{\Rightarrow} \sigma'_1 \underset{\mathcal{G}}{\Rightarrow} \sigma'_2 \cdots \sigma'_{k-1} \underset{\mathcal{G}}{\Rightarrow} \sigma'_k = \sigma_2$$

   para alguns $\sigma'_0, \sigma'_1, \ldots, \sigma'_k \in (V \cup A)^*$.

3. σ_2 **deriva** de σ_1, e escreve-se $\sigma_1 \overset{*}{\underset{\mathcal{G}}{\Rightarrow}} \sigma_2$, quando

$$\sigma_1 = \sigma_2 \text{ ou } \sigma_1 \overset{k}{\underset{\mathcal{G}}{\Rightarrow}} \sigma_2 \text{ para algum } k \in \mathbb{N}.$$

   A uma sequência de passos elementares (derivações diretas) que permite deduzir $\sigma_1 \overset{*}{\underset{\mathcal{G}}{\Rightarrow}} \sigma_2$ chama-se uma **derivação** de σ_2 a partir de σ_1.

Quando não há dúvida sobre qual a gramática $\mathcal{G}$ que se está a considerar, simplifica-se a notação de $\underset{\mathcal{G}}{\Rightarrow}$, $\overset{k}{\underset{\mathcal{G}}{\Rightarrow}}$ e $\overset{*}{\underset{\mathcal{G}}{\Rightarrow}}$, omitindo a letra $\mathcal{G}$.

Exemplo

Seja $\mathcal{G}_1 = \left(\{\mathcal{S}, \mathcal{B}, \mathcal{C}\}, \{a, b\}, \mathcal{S}, P\right)$ a gramática com produções

$$\begin{aligned} \mathcal{S} &\to a\mathcal{S} \mid b\mathcal{B} \\ \mathcal{B} &\to b\mathcal{B} \mid a\mathcal{C} \\ \mathcal{C} &\to a\mathcal{C} \mid b\mathcal{C} \mid \epsilon. \end{aligned}$$

Em $\mathcal{G}_1$ tem-se a seguinte derivação

$$\mathcal{S} \Rightarrow a\mathcal{S} \Rightarrow aa\mathcal{S} \Rightarrow aab\mathcal{B} \Rightarrow aaba\mathcal{C} \Rightarrow aabab\mathcal{C} \Rightarrow aabab.$$

Podemos então escrever, por exemplo,

$$\mathcal{S} \overset{3}{\Rightarrow} aab\mathcal{B}, \ \ \mathcal{S} \overset{*}{\Rightarrow} aab\mathcal{B}, \ \ a\mathcal{S} \overset{4}{\Rightarrow} aabab\mathcal{C}, \ \ aa\mathcal{S} \overset{*}{\Rightarrow} aaba\mathcal{C}, \ \ \mathcal{S} \overset{*}{\Rightarrow} aabab.$$

Definição

Sejam $\mathcal{G} = (V, A, \mathcal{S}, P)$ uma gramática e $\alpha \in (V \cup A)^*$.

1. O conjunto das palavras de $(V \cup A)^*$ que derivam de α é notado por $D(\alpha)$, i.e.:
$$D(\alpha) = \{\beta \in (V \cup A)^* : \alpha \underset{\mathcal{G}}{\overset{*}{\Rightarrow}} \beta\}.$$
2. O conjunto das palavras de A^* que derivam de α é notado por $L(\alpha)$, i.e.:
$$L(\alpha) = \{u \in A^* : \alpha \underset{\mathcal{G}}{\overset{*}{\Rightarrow}} u\} = D(\alpha) \cap A^*.$$
3. A linguagem gerada pela gramática $\mathcal{G}$ é
$$L(\mathcal{G}) = \{u \in A^* : \mathcal{S} \underset{\mathcal{G}}{\overset{*}{\Rightarrow}} u\} = L(\mathcal{S}).$$

Definição

Duas gramáticas $\mathcal{G}$ e $\mathcal{G}'$ dizem-se equivalentes quando $L(\mathcal{G}) = L(\mathcal{G}')$.

Exemplo

Consideremos a gramática $\mathcal{G}_1 = \big(\{\mathcal{S},\mathcal{B},\mathcal{C}\},\{a,b\},\mathcal{S},P\big)$ do exemplo anterior, de produções

$$\begin{aligned}\mathcal{S} &\to a\mathcal{S} \mid b\mathcal{B}\\ \mathcal{B} &\to b\mathcal{B} \mid a\mathcal{C}\\ \mathcal{C} &\to a\mathcal{C} \mid b\mathcal{C} \mid \epsilon,\end{aligned}$$

e determinemos a linguagem gerada por $\mathcal{G}_1$. Tem-se $L(\mathcal{G}_1) = L(\mathcal{S})$ e

$$\begin{cases} L(\mathcal{S}) = aL(\mathcal{S}) \cup bL(\mathcal{B})\\ L(\mathcal{B}) = bL(\mathcal{B}) \cup aL(\mathcal{C})\\ L(\mathcal{C}) = \{a,b\}L(\mathcal{C}) \cup \{\epsilon\}\end{cases}$$

Estas igualdades podem ser vistas como um sistema de equações lineares à direita nas incógnitas $L(\mathcal{S})$, $L(\mathcal{B})$ e $L(\mathcal{C})$. Resolvendo o sistema obtém-se

$$\begin{cases} L(\mathcal{S}) = a^*bb^*a(a+b)^*\\ L(\mathcal{B}) = b^*a(a+b)^*\\ L(\mathcal{C}) = (a+b)^*.\end{cases}$$

Conclui-se assim que $L(\mathcal{G}_1) = a^*bb^*a(a+b)^* = (a+b)^*ba(a+b)^*$.

Exemplo

Seja $\mathcal{G}_2 = \big(\{\mathcal{S},\mathcal{A},\mathcal{B}\},\{a,b\},\mathcal{S},P\big)$ a gramática com produções

$$\begin{aligned}\mathcal{S} &\to \mathcal{A}\mathcal{B}\\ \mathcal{A} &\to a\mathcal{A}b \mid \epsilon\\ \mathcal{B} &\to b\mathcal{B}a \mid \epsilon.\end{aligned}$$

Em $\mathcal{G}_2$ pode-se realizar, por exemplo, a seguinte derivação

$$\mathcal{S} \Rightarrow \mathcal{A}\mathcal{B} \Rightarrow a\mathcal{A}b\mathcal{B} \Rightarrow aa\mathcal{A}bb\mathcal{B} \Rightarrow aa\epsilon bb\mathcal{B} \Rightarrow aabbb\mathcal{B}a \Rightarrow aabbb\epsilon a$$

donde $a^2b^3a \in L(\mathcal{G}_2)$. Pelo contrário, $abab, ab^2a^2 \notin L(\mathcal{G}_2)$. Tem-se

$$\begin{aligned}L(\mathcal{G}_2) &= L(\mathcal{S}) = L(\mathcal{A}\mathcal{B}) = L(\mathcal{A})L(\mathcal{B})\\ L(\mathcal{A}) &= aL(\mathcal{A})b \cup \{\epsilon\} = a\big(aL(\mathcal{A})b \cup \{\epsilon\}\big)b \cup \{\epsilon\} = a^2L(\mathcal{A})b^2 \cup \{ab,\epsilon\}\\ &= a^{k+1}L(\mathcal{A})b^{k+1} \cup \{a^kb^k,\ldots,ab,\epsilon\} \quad \text{para todo o } k \in \mathbb{N}\\ &= \{a^mb^m : m \in \mathbb{N}_0\}\\ L(\mathcal{B}) &= \cdots = \{b^na^n : n \in \mathbb{N}_0\}\end{aligned}$$

Conclui-se assim que

$$L(\mathcal{G}_2) = \{a^mb^m : m \in \mathbb{N}_0\}\{b^na^n : n \in \mathbb{N}_0\} = \{a^mb^{m+n}a^n : m,n \in \mathbb{N}_0\}.$$

Definição

Uma gramática $\mathcal{G} = (V, A, \mathcal{S}, P)$ diz-se:

1. regular quando é

   linear à direita, ou seja, quando cada produção é da forma
   - $\mathcal{X} \to u\mathcal{Y}$ ou $\mathcal{X} \to u$, onde $\mathcal{X}, \mathcal{Y} \in V$ e $u \in A^*$;

   ou linear à esquerda, isto é, quando cada produção é da forma
   - $\mathcal{X} \to \mathcal{Y}u$ ou $\mathcal{X} \to u$, onde $\mathcal{X}, \mathcal{Y} \in V$ e $u \in A^*$.

2. independente de contexto (GIC) quando cada produção é da forma
   - $\mathcal{X} \to \alpha$, onde $\mathcal{X} \in V$ e $\alpha \in (V \cup A)^*$.

3. dependente de contexto (GDC) quando cada produção é da forma
   - $\alpha\mathcal{X}\beta \to \alpha\sigma\beta$, onde $\mathcal{X} \in V$ e $\alpha, \beta, \sigma \in (V \cup A)^*$ com $\sigma \neq \epsilon$, ou
   - $\mathcal{S} \to \epsilon$, se $\mathcal{S}$ não ocorre no membro direito de outra produção.

Observação

1. Um gramática regular é uma GIC.
2. Estes três tipos de gramática são precisamente os que definem os Tipos 1, 2 e 3 da Hierarquia de Chomsky.

Exemplo

1. A gramática $\mathcal{G}_1 = \big(\{\mathcal{S},\mathcal{B},\mathcal{C}\},\{a,b\},\mathcal{S},P\big)$ com produções

$$\mathcal{S} \to a\mathcal{S} \mid b\mathcal{B},\ \mathcal{B} \to b\mathcal{B} \mid a\mathcal{C},\ \mathcal{C} \to a\mathcal{C} \mid b\mathcal{C} \mid \epsilon,$$

é uma gramática linear à direita e portanto regular, que gera a linguagem regular

$$(a+b)^* ba(a+b)^*.$$

2. A gramática $\mathcal{G}_2 = \big(\{\mathcal{S},\mathcal{A},\mathcal{B}\},\{a,b\},\mathcal{S},P\big)$ com produções

$$\mathcal{S} \to \mathcal{A}\mathcal{B},\ \mathcal{A} \to a\mathcal{A}b \mid \epsilon,\ \mathcal{B} \to b\mathcal{B}a \mid \epsilon,$$

é uma GIC, que gera a linguagem independente de contexto

$$\{a^m b^{m+n} a^n : m,n \in \mathbb{N}_0\}.$$

Note-se que esta linguagem **não** é regular.

Exemplo (continuação)

3. A gramática $\mathcal{G}_3 = \left(\{\mathcal{S}, \mathcal{B}, \mathcal{C}\}, \{a, b, c\}, \mathcal{S}, P\right)$ com produções

$$\begin{aligned}
\mathcal{S} &\to a\mathcal{S}\mathcal{B}\mathcal{C} \mid a\mathcal{B}\mathcal{C} \\
\mathcal{C}\mathcal{B} &\to \mathcal{B}\mathcal{C} \\
a\mathcal{B} &\to ab \\
b\mathcal{B} &\to bb \\
b\mathcal{C} &\to bc \\
c\mathcal{C} &\to cc
\end{aligned}$$

gera a linguagem $\{a^n b^n c^n : n \in \mathbb{N}\}$.

Prova-se (usando um Lema de Iteração para linguagens independentes de contexto, análogo ao Lema de Iteração para linguagens regulares) que esta linguagem **não** é independente de contexto.

A gramática $\mathcal{G}_3$ **não** é uma GDC. (Porquê?)

O próximo objectivo é provar o teorema seguinte:

Teorema

Uma linguagem é regular se e só se é gerada por uma gramática regular.

A demonstração deste teorema utiliza conceitos e resultados auxiliares.

Definição

Uma gramática $\mathcal{G} = (V, A, \mathcal{S}, P)$ diz-se hiper-regular quando é

hiper-linear à direita, ou seja, quando cada produção é da forma

- $\mathcal{X} \to a\mathcal{Y}$, onde $\mathcal{X}, \mathcal{Y} \in V$ e $a \in A$, ou
- $\mathcal{X} \to \epsilon$, onde $\mathcal{X} \in V$;

ou hiper-linear à esquerda, isto é, quando cada produção é da forma

- $\mathcal{X} \to \mathcal{Y}a$, onde $\mathcal{X}, \mathcal{Y} \in V$ e $a \in A$, ou
- $\mathcal{X} \to \epsilon$, onde $\mathcal{X} \in V$.

Como é evidente, toda a gramática hiper-linear à direita é uma gramática linear à direita. Reciprocamente, tem-se o seguinte lema.

Lema

Toda a gramática linear à direita admite uma gramática hiper-linear à direita equivalente.

Demonstração: Seja $\mathcal{G} = (V, A, \mathcal{S}, P)$ uma gramática linear à direita. Definiremos uma gramática hiper-linear à direita $\mathcal{G}' = (V', A, \mathcal{S}, P')$ tal que $L(\mathcal{G}) = L(\mathcal{G}')$. Cada produção de $\mathcal{G}$ é da forma $\mathcal{X} \to u$ ou $\mathcal{X} \to u\mathcal{Y}$, onde $\mathcal{X}, \mathcal{Y} \in V$ e $u \in A^*$.

1. Cada produção $\mathcal{X} \to a_1 a_2 \cdots a_n$ é substituída pelas produções

$$\begin{aligned} \mathcal{X} &\to a_1\mathcal{A}_1 \\ \mathcal{A}_1 &\to a_2\mathcal{A}_2 \\ &\vdots \\ \mathcal{A}_{n-1} &\to a_n\mathcal{A}_n \\ \mathcal{A}_n &\to \varepsilon \end{aligned}$$

   onde $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ são novas variáveis.

2. Cada produção $\mathcal{X} \to a_1 a_2 \cdots a_n \mathcal{Y}$ é substituída pelas produções

$$\mathcal{X} \to a_1 \mathcal{B}_1, \; \mathcal{B}_1 \to a_2 \mathcal{B}_2, \ldots, \; \mathcal{B}_{n-2} \to a_{n-1} \mathcal{B}_{n-1}, \; \mathcal{B}_{n-1} \to a_n \mathcal{Y}$$

   onde $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{n-1}$ são novas variáveis.
3. Elimina-se cada produção $\mathcal{X} \to \mathcal{Y}$ em que a variável $\mathcal{Y}$ não ocorre no membro esquerdo de nenhuma produção de $\mathcal{G}$.
4. Se $\mathcal{X} \to \mathcal{Y}$ e $\mathcal{Y} \to \alpha_1 \mid \cdots \mid \alpha_n$ são produções de $\mathcal{G}$, então consideram-se produções $\mathcal{X} \to \alpha_1 \mid \cdots \mid \alpha_n$ e a cada uma destas aplica-se o processo descrito nos pontos anteriores.

Assim, define-se a gramática $\mathcal{G}' = (V', A, \mathcal{S}, P')$ em que:

- V' é igual à união de V com o conjunto das novas variáveis introduzidas;
- P' é o conjunto de produções que foram obtidas a partir de P pelas substituições descritas acima.

Proposição

Toda a linguagem regular é gerada por uma gramática hiper-linear à direita.

Demonstração: Se $L \subseteq A^*$ é regular, então existe um autómato determinista $\mathcal{A} = (Q, A, \delta, i, F)$ que reconhece L. Considere-se a gramática hiper-linear à direita $\mathcal{G}_\mathcal{A} = (Q, A, i, P)$ com produções $p \to aq$, para cada $q \in \delta(p, a)$,e produções $q \to \varepsilon$,para cada $q \in F$.

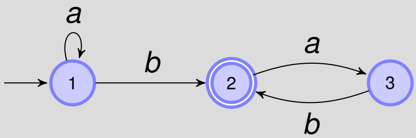
Tem-se, por um lado:

$$
\begin{array}{ll}
 & u = a_1 \cdots a_n \in L \\
\text{sse} & \text{existem } q_1, \ldots, q_{n-1} \in Q,\ q_n \in F \text{ e um caminho em } \mathcal{A} \\
 & i \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \cdots q_{n-1} \xrightarrow{a_n} q_n \\
\text{sse} & \text{existem } q_1, \ldots, q_{n-1} \in Q,\ q_n \in F \text{ e uma derivação em } \mathcal{G}_\mathcal{A} \\
 & i \underset{\mathcal{G}_\mathcal{A}}{\Rightarrow} a_1 q_1 \underset{\mathcal{G}_\mathcal{A}}{\Rightarrow} a_1 a_2 q_2 \underset{\mathcal{G}_\mathcal{A}}{\Rightarrow} \cdots \underset{\mathcal{G}_\mathcal{A}}{\Rightarrow} a_1 a_2 \cdots a_n q_n \underset{\mathcal{G}_\mathcal{A}}{\Rightarrow} a_1 a_2 \cdots a_n \varepsilon \\
\text{sse} & u \in L(\mathcal{G}_\mathcal{A}).
\end{array}
$$

Por outro lado: $\epsilon \in L$ sse $i \in F$ sse $i \underset{\mathcal{G}_\mathcal{A}}{\Rightarrow} \epsilon$ sse $\epsilon \in L(\mathcal{G}_\mathcal{A})$.
Portanto $L = L(\mathcal{G}_\mathcal{A})$.

Exemplo

O seguinte autómato determinista $\mathcal{A}$



reconhece a linguagem regular $L = a^*b(ab)^*$. Então, a gramática hiper-linear à direita $\mathcal{G}_{\mathcal{A}} = (\{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3\}, \{a, b\}, \mathcal{B}_1, P)$ com produções

$$\begin{aligned} \mathcal{B}_1 &\to a\mathcal{B}_1 \mid b\mathcal{B}_2 \\ \mathcal{B}_2 &\to a\mathcal{B}_3 \mid \epsilon \\ \mathcal{B}_3 &\to b\mathcal{B}_2 \end{aligned}$$

gera a linguagem L.

Proposição

Cada gramática linear à direita gera uma linguagem regular.

Demonstração: Seja $\mathcal{G} = (V, A, \mathcal{S}, P)$ uma gramática linear à direita. Pelo último lema podemos supôr que $\mathcal{G}$ é hiper-linear à direita. Então o autómato finito $\mathcal{A}_{\mathcal{G}} = (V, A, \delta, \mathcal{S}, F)$ onde $F = \{\mathcal{X} \in V : \mathcal{X} \to \varepsilon \in P\}$ e, para cada par $(\mathcal{X}, a) \in V \times A$,

$$\delta(\mathcal{X}, a) = \{\mathcal{Y} : \mathcal{X} \to a\mathcal{Y} \in P\},$$

reconhece $L(\mathcal{G})$. De facto,

$$\begin{aligned}
L(\mathcal{G}) &= \{u \in A^* \ : \ \mathcal{S} \overset{*}{\underset{\mathcal{G}}{\Rightarrow}} u\} \\
&= \{u \in A^* \ : (u = \epsilon \wedge \mathcal{S} \underset{\mathcal{G}}{\Rightarrow} \epsilon) \vee \\
&\qquad\qquad \exists n \in \mathbb{N} \ \exists a_1, \ldots, a_n \in A \ \exists \mathcal{A}_1, \ldots, \mathcal{A}_n \in V, \\
&\qquad\qquad \mathcal{S} \underset{\mathcal{G}}{\Rightarrow} a_1 \mathcal{A}_1 \underset{\mathcal{G}}{\Rightarrow} \cdots \underset{\mathcal{G}}{\Rightarrow} a_1 \cdots a_n \mathcal{A}_n \underset{\mathcal{G}}{\Rightarrow} a_1 \cdots a_n = u\} \\
&= \{u \in A^* \ : \ \delta(\mathcal{S}, u) \in F\} \\
&= L(\mathcal{A}_{\mathcal{G}}).
\end{aligned}$$

Deduz-se assim que $L(\mathcal{G})$ é uma linguagem regular.

Como consequência das proposições anteriores, obtém-se:

Proposição

Uma linguagem é regular se e só se é gerada por uma gramática linear à direita.

Dualmente, pode provar-se:

Proposição

Uma linguagem é regular se e só se é gerada por uma gramática linear à esquerda.

Concluiu-se desta forma a demonstração do seguinte teorema (já antes enunciado):

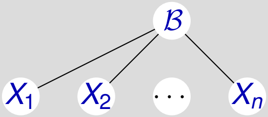
Teorema

Uma linguagem é regular se e só se é gerada por uma gramática regular.

Definição

Seja $\mathcal{G} = (V, A, \mathcal{S}, P)$ uma gramática independente de contexto. Cada derivação de uma palavra de $L(\mathcal{G})$ determina uma árvore de derivação formada do seguinte modo:

- a raiz tem como etiqueta o símbolo inicial $\mathcal{S}$ da gramática;
- por cada passo da derivação em que uma variável $\mathcal{B}$ é reescrita através de uma produção $\mathcal{B} \to X_1 X_2 \cdots X_n$ (com $X_i \in V \cup A \cup \{\epsilon\}$), é criado um nodo etiquetado por $\mathcal{B}$, com filhos etiquetados, da esquerda para a direita, por $X_1, X_2, \ldots, X_n$:



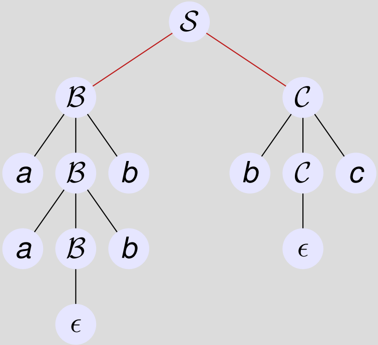
- um nodo com etiqueta em $A \cup \{\epsilon\}$ é sempre uma folha da árvore, e não tem filhos.

Exemplo

Seja $\mathcal{G} = (\{\mathcal{S}, \mathcal{B}, \mathcal{C}\}, \{a,\ b, c\}, \mathcal{S}, P)$ em que P é constituído por:

$$\mathcal{S} \to \mathcal{B}\mathcal{C} \qquad \mathcal{B} \to a\mathcal{B}b \mid \epsilon \qquad \mathcal{C} \to b\mathcal{C}c \mid \epsilon$$

$$\begin{aligned}
\mathcal{S} & \underset{\mathcal{G}}{\Rightarrow} \mathcal{B}\mathcal{C} \\
& \underset{\mathcal{G}}{\Rightarrow} a\mathcal{B}b\mathcal{C} \\
& \underset{\mathcal{G}}{\Rightarrow} aa\mathcal{B}bb\mathcal{C} \\
& \underset{\mathcal{G}}{\Rightarrow} aa\epsilon bb\mathcal{C} \\
& \underset{\mathcal{G}}{\Rightarrow} aabbb\mathcal{C}c \\
& \underset{\mathcal{G}}{\Rightarrow} aabbb\epsilon c
\end{aligned}$$



Uma outra derivação da palavra *aabbbc* a partir de $\mathcal{S}$ é

$\mathcal{S} \Rightarrow \mathcal{B}\mathcal{C} \Rightarrow \mathcal{B}b\mathcal{C}c \Rightarrow \mathcal{B}b\epsilon c \Rightarrow a\mathcal{B}bbc \Rightarrow aa\mathcal{B}bbbc \Rightarrow aa\epsilon bbbc = aabbbc$

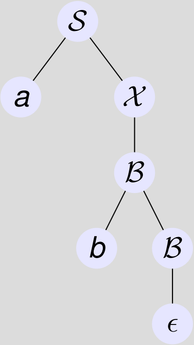
mas, a árvore de derivação determinada é igual à anterior.

Exemplo

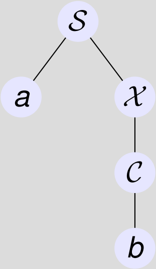
Seja $\mathcal{G}_a = (\{\mathcal{S}, \mathcal{B}, \mathcal{C}, \mathcal{X}\}, \{a, b\}, \mathcal{S}, P)$ em que P é constituído por:

$$\mathcal{S} \to a\mathcal{X} \qquad \mathcal{X} \to \mathcal{B} \mid \mathcal{C} \qquad \mathcal{B} \to b\mathcal{B} \mid \epsilon \qquad \mathcal{C} \to a\mathcal{C} \mid b$$

$$\begin{aligned} \mathcal{S} & \underset{\mathcal{G}_a}{\Rightarrow} a\mathcal{X} \\ & \underset{\mathcal{G}_a}{\Rightarrow} a\mathcal{B} \\ & \underset{\mathcal{G}_a}{\Rightarrow} ab\mathcal{B} \\ & \underset{\mathcal{G}_a}{\Rightarrow} ab\epsilon \end{aligned}$$



$$\begin{aligned} \mathcal{S} & \underset{\mathcal{G}_a}{\Rightarrow} a\mathcal{X} \\ & \underset{\mathcal{G}_a}{\Rightarrow} a\mathcal{C} \\ & \underset{\mathcal{G}_a}{\Rightarrow} ab \end{aligned}$$



Note-se que as duas árvores de derivação são distintas.

Definição

Derivações que determinam a mesma árvore de derivação dizem-se essencialmente iguais.

Exemplo

As derivações do slide anterior **não** são essencialmente iguais.

Já as duas derivações do slide 22 são essencialmente iguais.

Definição

1. Uma GIC $\mathcal{G} = (V, A, \mathcal{S}, P)$ diz-se ambígua quando **alguma** palavra de A^* admite árvores de derivação distintas, por outras palavras, alguma palavra de A^* admite derivações que não são essencialmente iguais.
2. Uma linguagem independente de contexto L diz-se ambígua quando **todas** as GICs que geram L são ambíguas.

Exemplo

A gramática $\mathcal{G}_a = (\{\mathcal{S}, \mathcal{B}, \mathcal{C}, \mathcal{X}\}, \{a, b\}, \mathcal{S}, P)$ do exemplo anterior, de produções

$$\begin{aligned} \mathcal{S} &\to a\mathcal{X} \\ \mathcal{X} &\to \mathcal{B} \mid \mathcal{C} \\ \mathcal{B} &\to b\mathcal{B} \mid \epsilon \\ \mathcal{C} &\to a\mathcal{C} \mid b, \end{aligned}$$

é ambígua. A linguagem gerada por $\mathcal{G}_a$ é

$$\begin{aligned} L(\mathcal{G}_a) &= L(\mathcal{S}) \\ &= aL(\mathcal{X}) \\ &= a\big(L(\mathcal{B}) \cup L(\mathcal{C})\big) \\ &= a(b^* \cup a^* b). \end{aligned}$$

Questão

Será que $a(b^* \cup a^* b)$ é uma linguagem ambígua?

Exemplo

Consideremos a gramática $\mathcal{G}_{na} = (\{\mathcal{S}, \mathcal{B}, \mathcal{C}, \mathcal{X}\}, \{a, b\}, \mathcal{S}, P_{na})$ de produções:

$$
\begin{array}{ll}
\mathcal{S} & \to a\mathcal{X} \\
\mathcal{X} & \to \mathcal{B} \mid a\mathcal{C}b \\
\mathcal{B} & \to b\mathcal{B} \mid \epsilon \\
\mathcal{C} & \to a\mathcal{C} \mid \epsilon
\end{array}
$$

Prova-se que $\mathcal{G}_{na}$ **não** é ambígua, e

$$
\begin{array}{rcl}
L(\mathcal{G}_{na}) & = & L(\mathcal{S}) \\
 & = & aL(\mathcal{X}) \\
 & = & a\big(L(\mathcal{B}) \cup aL(\mathcal{C})b\big) \\
 & = & a(b^* \cup aa^*b) \\
 & = & a(b^* \cup b \cup a^+b) \\
 & = & a(b^* \cup a^*b)
\end{array}
$$

Conclui-se assim que a linguagem $a(b^* \cup a^*b)$ **não** é ambígua.

Proposição

Existem linguagens independentes de contexto ambíguas.

Prova-se que a linguagem

$$\{a^m b^m c^n d^n : m \geq 1, n \geq 1\} \cup \{a^m b^n c^n d^m : m \geq 1, n \geq 1\}$$

é ambígua.