Capítulo 7

Propriedades das Linguagens Livres do Contexto

As linguagens livres do contexto ocupam uma posição central na hierarquia das linguagens formais. Por um lado, as linguagens livres do contexto incluem importantes, mas restritas, famílias de linguagens como as linguagens regulares e as linguagens de programação usuais pertencem a esta classe de linguagens. Por outro lado, existem famílias de linguagens mais amplas das quais as linguagens livres do contexto são casos especiais. Como anteriormente, para as linguagens regulares, veremos as propriedades de fecho para várias operações, algoritmos para determinar propriedades de membros da família, e resultados estruturais como os lemas do bombeamento.

7.1 O Lema do Bombeamento para Linguagens Livres do contexto

O lema do bombeamento dado para linguagens regulares é uma ferramenta efetiva para mostrar que certas linguagens não são regulares. Existem lemas do bombeamento análogos para outras classes de linguagens, veja por exemplo [Lin90] um lema do bombeamento para linguagens lineares e em [Yu89] um lema do bombeamento para linguagens livres do contexto determinísticas, isto é, reconhecidas por APD's[1]. Porém, aqui, discutiremos um desses lemas de bombeamento para linguagens livres do contexto, em geral.

Teorema 7.1.1 (Lema do bombeamento) *Seja $\mathcal{L}$ uma linguagem livre do contexto infinita. Então existe algum inteiro positivo m tal que qualquer $w \in \mathcal{L}$, com $|w| \geq m$, pode ser decomposto como $w = uvxyz$ com $u, v, x, y, z \in \Sigma^*$ satisfazendo*

$$|vxy| \leq m, \tag{7.1}$$

$$|vy| \geq 1 \tag{7.2}$$

[1] Também existem outras ferramentas para este mesmo propósito, veja por exemplo [HU79, LV95].

e

$$uv^ixy^iz \in \mathcal{L}, \text{ para todo } i = 0, 1, 2, \ldots. \tag{7.3}$$

DEMONSTRAÇÃO: Ver [Lin90]. ■

Exemplo 7.1.2 *Seja a linguagem*

$$\mathcal{L} = \{a^nb^{2n} \;/\; n \geq 1\}.$$

Mostraremos usando o lema do bombeamento que $\mathcal{L}$ é livre do contexto.

Considere, $m = 3$. Então todo $w \in \mathcal{L}$ satisfaz $\mid w \mid \geq m$. Seja, $n \geq 1$.

Claramente, $u = a^{n-1}$, $v = a$, $x = \lambda$, $y = bb$ e $z = b^{2(n-1)}$ é uma decomposição de w satisfazendo as condições (7.1) e (7.2) do lema. Seja $i \in \mathbb{N}$ qualquer, então

$$uv^ixy^iz = a^{n-1}a^i(bb)^ib^{2(n-1)} = a^{n-1+i}b^{2i+2(n-1)} = a^{n-1+i}b^{2(n-1+i)} \in \mathcal{L}$$

Logo, $\mathcal{L}$ é livre do contexto.

Porém, este lema é mais útil para mostrar que uma linguagem não pertence à família das linguagens livre do contexto. Seu uso é análogo aquele visto para linguagens regulares.

Exemplo 7.1.3 *Mostre que a linguagem $\mathcal{L} = \{a^nb^nc^n \;/\; n \geq 0\}$ não é livre do contexto.*

Suponha que $\mathcal{L}$ é livre do contexto. Então, pelo lema do bombeamento, existe m satisfazendo as condições do teorema 7.1.1. A cadeia $a^mb^mc^m$ está em $\mathcal{L}$. Se escolhermos vxy como sendo uma cadeia só de a's, então, pelo lema do bombeamento, a cadeia uv^2xy^2z, por exemplo, deveria estar em $\mathcal{L}$, mas como $|vy| \leq 1$, ela claramente não é da forma $a^nb^nc^n$, pois tem mais a's que b's (e c's). Analogamente, se escolhermos as subcadeias v composta só de a's, digamos j a's, e y composta só de b's, digamos k b's, então a cadeia bombeada uv^2xy^2z, terá a forma $a^{m+j}b^{m+k}c^m$, e portanto, novamente geramos uma cadeia que não está em $\mathcal{L}$. De fato, a única maneira da escolha das subcadeias u, v, x, y e z impedir esse fato é tomar vxy tal que vy tenha o mesmo número de a's, b's e c's. Mas isso não é possível pela restrição (7.2). Portanto, $\mathcal{L}$ não é livre do contexto.

Uma forma direta de se provar isso é dado no seguinte corolário.

Corolário 7.1.4 *Seja $\mathcal{L}$ uma linguagem infinita. $\mathcal{L}$ não é livre do contexto se para cada inteiro positivo m existe $w \in \mathcal{L}$, tal que $|w| \geq m$ e para cada decomposição $uvxyz$ de w, se*

$$|vxy| \leq m \tag{7.4}$$

e

$$|vy| \geq 1 \tag{7.5}$$

então $uv^ixy^iz \notin \mathcal{L}$*, para algum* $i = 0, 1, 2, \ldots$.

DEMONSTRAÇÃO: Direto do lemma do bombeamento. ■

Exemplo 7.1.5 *Considere a linguagem*

$$\mathcal{L} = \{ww \;/\; w \in \{a, b\}^*\}.$$

Apesar desta linguagem parecer muito similar à linguagem livre do contexto do exemplo 5.1.2, ela não é livre do contexto.

Seja $m \in \mathbb{N}$ *qualquer e considere a cadeia* $w = a^m b^m a^m b^m$*. Nos teríamos que analisar todas as possíveis decomposições* $uvxyz$ *de* w*, e se a decomposição satisfazer (7.4) e (7.5) então necessariamente algum bombeamento de* v *e* y *deverá gerar uma cadeia não pertencente à linguagem. Mesmo considerando só as decomposições que satisfazem (7.4) e (7.5) ainda são muitas (além do mais que nós não conhecemos o valor real de* m*). Mas como a restrição (7.4) força que* $|\; vxy \;| \leq m$*, então ela só pode estar em:*

$$\begin{gathered}\underbrace{a\cdots a}_{u}\underbrace{a\cdots a}_{vxy}\underbrace{a\cdots ab^m a^m b^m}_{z} \;\; ou \;\; \underbrace{a^m b\cdots b}_{u}\underbrace{b\cdots b}_{vxy}\underbrace{b\cdots ba^m b^m}_{z} \;\; ou \\ \underbrace{a^m b^m a\cdots a}_{u}\underbrace{a\cdots a}_{vxy}\underbrace{a\cdots ab^m}_{z} \;\; ou \;\; \underbrace{a^m b^m a^m b\cdots b}_{u}\underbrace{b\cdots b}_{vxy}\underbrace{b\cdots b}_{z}\end{gathered} \tag{7.6}$$

ou ainda

$$\begin{gathered}\underbrace{a\cdots a}_{u}\underbrace{a\cdots ab\cdots b}_{vxy}\underbrace{b\cdots ba^m b^m}_{z} \;\; ou \;\; \underbrace{a^m b\cdots b}_{u}\underbrace{b\cdots ba\cdots a}_{vxy}\underbrace{a\cdots ab^m}_{z} \;\; ou \\ \underbrace{a^m b^m a\cdots a}_{u}\underbrace{a\cdots ab\cdots b}_{vxy}\underbrace{b\cdots b}_{z}\end{gathered} \tag{7.7}$$

Note que todas as decomposições em (7.6) e em (7.7) são análogas, e portanto é suficiente considerar um caso de cada.

Assim, no caso de (7.6) consideraremos somente a primeira decomposição. Assim, $u = a^{k_1}$*,* $v = a^{k_2}$*,* $x = a^{k_3}$*,* $y = a^{k_4}$ *e* $z = a^{k_5}b^m a^m b^m$ *para algum* $k_1, k_2, k_3, k_4, k_5 \in \mathbb{N}$ *tal que* $k_1 + k_2 + k_3 + k_4 + k_5 = m$ *e* $k_2 + k_4 \geq 1$*. Assim,*

$$uv^2xy^2z = a^{k_1}a^{2k_2}a^{k_3}a^{2k_4}a^{k_5}b^m a^m b^m = a^{k_1}a^{k_2}a^{k_2}a^{k_3}a^{k_4}a^{k_4}a^{k_5}b^m a^m b^m = a^{m+k_2+k_4}b^m a^m b^m.$$

Como $k_2 + k_4 \geq 1$*, então* $uv^2xy^2z \notin \mathcal{L}$.

No caso de (7.7) consideraremos somente a primeira decomposição. Mas, aqui novamente teremos varias possibilidades, por exemplo v *e* x *podem ser só composto de* $a's$ *e* y *de* $b's$ *ou* y *de* $a's$ *e de* $b's$*, ou ainda,* v *pode ser composto de* $a's$ *e* x *de* $a's$ *e de* $b's$ *e* y *de* $b's$*, etc.*

Mas independente de qual for decomposição ao bombear o v e o y (que satisfazem $\mid vy \mid \geq 1$) será alterada a quantidade de $a's$ e/ou de $b's$ da primeira metade da cadeia w pelo que a cadeia resultante já não terá mais a forma $a^m b^m a^m b^m$ e portanto não pertencerá à linguagem. Logo, $\mathcal{L}$ não é livre do contexto.

Exemplo 7.1.6 *Mostre que a linguagem $\mathcal{L} = \{a^j b^k \ / \ j = k^2\}$ não é livre do contexto.*

Dado um m qualquer, tomamos nossa cadeia como sendo $a^{m^2} b^m$. Claramente, temos várias escolhas para u, v, x, y, z. A única que requer mais cuidado é a seguinte:

$$\overbrace{a\cdots\overbrace{a\cdots a}^{k_1}\cdots a}^{m^2}\ \overbrace{b\cdots\overbrace{b\cdots b}^{k_2}\cdots b}^{m}$$

$$u \qquad v \qquad\quad x \qquad\quad y \qquad z$$

Bombeando as subcadeias v e y, i vezes, geraremos a nova cadeia com $m^2 + (i-1)k_1$ a's e $m + (i-1)k_2$ b's. Se $k_1 = 0$, então, pela restrição (7.5), $k_2 > 1$. Para o caso $i = 0$, teremos m^2 a's e $m - k_2$ b's o qual obviamente não está em $\mathcal{L}$. Usando um raciocínio análogo, obteremos que a escolha $k_2 = 0$ e $k_1 > 1$, também falha. Consideremos, portanto, o caso $k_1 \neq 0$ e $k_2 \neq 0$ e tomemos $i = 0$. Como

$$\begin{aligned} (m-k_2)^2 &\leq (m-1)^2 \\ &= m^2 - 2m + 1 \\ &< m^2 - k_1, \end{aligned}$$

o resultado não está em $\mathcal{L}$. Portanto, $\mathcal{L}$ não é livre do contexto.

7.2 Propriedades de Fecho para Linguagens Livres do Contexto

No capítulo 4, analisamos algumas propriedades de fecho para certas operações e alguns algoritmos de decisão sobre as propriedades da família de linguagens regulares. Ambas questões, eram facilmente provadas. No entanto, as mesmas questões para linguagens livres do contexto são ligeiramente mais difíceis de se responder. De fato, algumas propriedades de fecho que são satisfeitas por linguagens regulares não o são por linguagens livres do contexto. Nesta seção, somente analisaremos as propriedades de fecho das operações mais tradicionais entre linguagens.

Teorema 7.2.1 *A família de linguagens livres do contexto é fechada sobre união, concatenação e fecho estrela $(*)$.*

Demonstração: Sejam L_1 e L_2 duas linguagens livres do contexto qualquer e $G_1 = \langle V_1, T_1, S_1, P_1\rangle$ e $G_2 = \langle V_2, T_2, S_2, P_2\rangle$ duas gramáticas livres do contexto que geram elas. Por simplicidade sempre podemos considerar que $V_1 \cap V_2 = \emptyset$.

As gramática livre do contextos

1. $G_{\cup} = \langle V_1 \cup V_2 \cup \{S\}, T_1 \cup T_2, S, P_1 \cup P_2 \cup \{S \longrightarrow S_1, S \longrightarrow S_2\}\rangle$

2. $G_\circ = \langle V_1 \cup V_2, S_1, (P_1 - P) \cup P_2 \cup \{A \longrightarrow wS_2 \ / \ A \longrightarrow w \in P\}\rangle$ onde $P = \{A \longrightarrow w \in P_1 \ / \ w \in T_1^*\}$ e

3. $G_* = \langle V_1, T_1, S_1, P_1 \cup \{A \longrightarrow wS_1 \ / \ A \longrightarrow w \in P \text{ e } w \in T_1^*\} \cup \{S_1 \longrightarrow \lambda\}\rangle$,

geram as linguagens $L_1 \cup L_2$, $L_2 \circ L_1$ e L_1^*, respectivamente. ■

Teorema 7.2.2 *A família de linguagens livres do contexto não é fechada sobre intersecção e complementação.*

Demonstração: Considere as linguagens

$$\mathcal{L}_1 = \{a^n b^n c^m \ / \ n \geq 0, \ m \geq 0\}$$

e

$$\mathcal{L}_2 = \{a^n b^m c^m \ / \ n \geq 0, \ m \geq 0\}.$$

Claramente $\mathcal{L}_1$ e $\mathcal{L}_2$ são livres do contexto. Mas, como vimos no exemplo 7.1.3, a linguagem

$$\mathcal{L}_1 \cap \mathcal{L}_2 = \{a^n b^n c^n \ / \ n \geq 0\},$$

não é livre do contexto. Logo, as linguagens livres do contexto não são fechadas sobre a intersecção.

Por outro lado, seja a identidade de conjuntos

$$\mathcal{L}_1 \cap \mathcal{L}_2 = \overline{\overline{\mathcal{L}_1} \cup \overline{\mathcal{L}_2}}. \tag{7.8}$$

Se as linguagens livres do contexto fossem fechadas sobre a complementação, então, como pelo teorema anterior elas são fechadas sobre a união, a linguagem em (7.8) também seria livre do contexto, o qual contradiziria o resultado anterior. Conseqüentemente, a família das linguagens livres do contexto não é fechada sobre a complementação. ■

7.3 Propriedades Decidíveis de Linguagens Livres do Contexto

No capítulo 4 já foi visto a existência de um algoritmo de pertinência para linguagens livres do contexto. Isto é uma característica essencial de qualquer família de linguagens útil na prática. Outras propriedades simples de linguagens livres do contexto podem ser determinadas. Mas, algumas questões intuitivamente simples de se responder para linguagens regulares via um algoritmo, não podem ser respondidas algoritmicamente para livres do contexto. Para os propósitos desta discussão, assumiremos que a linguagem é descrita pela sua gramática.

Teorema 7.3.1 *Seja $G = \langle V, T, S, P\rangle$ uma gramática livre do contexto. Então existe um algoritmo para decidir quando $L(G)$, for vazio ou não.*

Demonstração: Por simplicidade assuma que $\lambda \notin L(G)$. Use o algoritmo para remover símbolos e produções inúteis. Se S for inútil, então $L(G)$ é vazio, senão $L(G)$ contém ao menos um elemento. ■

Teorema 7.3.2 *Seja $G = \langle V, T, S, P\rangle$ uma gramática livre do contexto. Então existe um algoritmo para decidir quando $L(G)$, for infinito ou não.*

Demonstração: Assuma que G não contém nenhuma λ-produção, nenhuma produção unitária e nenhum símbolo inútil. Se existe na gramática uma variável se repetindo, $A \in V$, isto é, uma variável tal que

$$A \stackrel{*}{\Longrightarrow} xAy,$$

então, como A não é inútil, deve existir $w \in T^*$ tal que

$$S \stackrel{*}{\Longrightarrow} uAv \Longrightarrow uxAyv \stackrel{*}{\Longrightarrow} w,$$

Logo, $u \stackrel{*}{\Longrightarrow} x_1$, $x \stackrel{*}{\Longrightarrow} x_2$, $A \stackrel{*}{\Longrightarrow} x_3$, $y \stackrel{*}{\Longrightarrow} x_4$ e $v \stackrel{*}{\Longrightarrow} x_5$ com $x_1, x_2, x_3, x_4, x_5 \in T^*$ são tais que $x_1x_2x_3x_4x_5 = w$. Portanto, para todo $n = 1, 2, \ldots$

$$S \stackrel{*}{\Longrightarrow} uAv \stackrel{*}{\Longrightarrow} ux^nAy^nv \stackrel{*}{\Longrightarrow} ux^nzy^nv \stackrel{*}{\Longrightarrow} x_1x_2^nx_3x_4^nx_5$$

e portanto $L(G)$ seria infinita.

Se nenhuma variável pode ser repetida, então o comprimento de qualquer derivação não excede $|V|$. Em cujo caso $L(G)$ é finita.

Assim, para obter um algoritmo para determinar quando $L(G)$ é infinita, somente precisamos determinar quando a gramática tem uma variável repetindo. Isto pode ser conseguido, simplesmente, desenhando o grafo de dependência das variáveis. Se existe um ciclo, então a gramática tem uma variável se repetindo, pois não temos variáveis inúteis, e portanto sempre podemos chegar a ela desde S. ■

7.4 Exercícios

1. Mostre usando o lema do bombeamento que as seguintes linguagens são livres do contexto

   (a) $\mathcal{L} = \{a^n b^n \ / \ n \text{ é um número ímpar}\}$

   (b) $\mathcal{L} = \{a^n b^{2n} \ / \ n \text{ é múltiplo de } 5\}$

   (c) $\mathcal{L} = \{a^n b^p \ / \ n \geq p\}$

   (d) $\mathcal{L} = \{a^n b^p \ / \ n \geq 1 \text{ e } n \leq p \leq 3n\}$

   (e) $\mathcal{L} = \{w \in \{a,b\}^* \ / \ \mathcal{N}_a(w) \text{ é par }\}$

   (f) $\mathcal{L} = \{w \in \{a,b\}^* \ / \ w = w^R\}$

   (g) $\mathcal{L} = \{w \in \{a,b\}^* \ / \ w \neq w^R\}$

2. Mostre que as seguintes linguagens, sobre o alfabeto $\{a, b\}$, não são livres do contexto

   (a) $\mathcal{L} = \{a^m b^n \ / \ n = m^2\}$

   (b) $\mathcal{L} = \{a^m b^n \ / \ m \leq n^2\}$

   (c) $\mathcal{L} = \{a^m b^n \ / \ m \geq (n-1)^3\}$

   (d) $\mathcal{L} = \{uv \in \{a,b\}^* \ / \ \mathcal{N}_a(u) + \mathcal{N}_b(u) = \max\{\mathcal{N}_a(v), \mathcal{N}_b(v)\}\}$

3. Mostre que as seguintes linguagens, sobre o alfabeto $\{a, b, c\}$, não são livres do contexto

   (a) $\mathcal{L} = \{a^k b^m c^n \ / \ n = km\}$

   (b) $\mathcal{L} = \{a^k b^m c^n \ / \ k < m \text{ e } k \leq m \leq n\}$

   (c) $\mathcal{L} = \{a^k b^m c^n \ / \ m > k \text{ e } n > k\}$

   (d) $\mathcal{L} = \{w \ / \ \mathcal{N}_a(w) < \mathcal{N}_b(w) < \mathcal{N}_c(w)\}$

4. Determine quais das seguintes linguagens, sobre o alfabeto $\{a, b\}$, são livres do contexto e quais não

   (a) $\mathcal{L} = \{a^n w w^R a^n \ / \ n \geq 0, \ w \in \{a,b\}^*\}$

   (b) $\mathcal{L} = \{a^m b^n a^m b^n \ / \ m \geq 0, \ n \geq 0\}$

   (c) $\mathcal{L} = \{a^m b^n a^n b^m \ / \ m \geq 0, \ n \geq 0\}$

   (d) $\mathcal{L} = \{a^m b^n a^j b^k \ / \ m + n \leq j + k\}$

   (e) $\mathcal{L} = \{a^m b^n a^j b^k \ / \ m \leq j, n \leq k\}$

5. Mostre que a classe das linguagens livres do contexto é fechada sobre os seguintes operadores:

   (a) Homomorfismos,

   (b) Diferença de conjuntos,

   (c) O operador reverso

(d) Seja $\mathcal{L}$ uma linguagem qualquer. Define $\widehat{\mathcal{L}}_a$ como sendo

$$\widehat{\mathcal{L}}_a = \{awa \ / \ w \in \mathcal{L}\}$$

(e) Seja $\mathcal{L}$ um linguagem sobre o alfabeto $\{0, 1\}$ e $\neg : \Sigma^* \longrightarrow \Sigma^*$ definido por $\neg(\lambda) = \lambda$, $\neg(w0) = 1\neg(w)$ e $\neg(w1) = 0\neg(w)$. Define $\neg\mathcal{L}$ como sendo

$$\neg\mathcal{L} = \{\neg(w) \ / \ w \in \mathcal{L}\}$$

(f) Seja $\mathcal{L}$ uma linguagem qualquer. Define $\widetilde{\mathcal{L}}_a$ como sendo

$$\widetilde{\mathcal{L}}_a = \{wav \ / \ w, v \in \mathcal{L}\}$$

(g) Seja $\mathcal{L}$ uma linguagem qualquer sobre um alfabeto Σ e $\rho : \Sigma^* \longrightarrow \Sigma^*$ definido por $\rho(\lambda) = \lambda$ e $\rho(wa) = \rho(w)aa$ para todo $w \in \Sigma^*$ e $a \in \Sigma$. Define $\mathcal{L}^\rho$ como sendo

$$\mathcal{L}^\rho = \{\rho(w) \ / \ w \in \mathcal{L}\}$$

(h) Seja $\mathcal{L}$ uma linguagem qualquer sobre um alfabeto Σ e $\varrho : \Sigma^* \longrightarrow \Sigma^*$ definido por $\varrho(\lambda) = \lambda$ e $\varrho(wa) = a\varrho(w)aa$ para todo $w \in \Sigma^*$ e $a \in \Sigma$. Define $\mathcal{L}^\varrho$ como sendo

$$\mathcal{L}^\varrho = \{\varrho(w) \ / \ w \in \mathcal{L}\}$$

6. Seja $L(G)$ a linguagem gerada por uma gramática livre do contexto G. Dê um algoritmo para determinar quando $\lambda \in L(G)$ ou não.

7. Mostre que existe um algoritmo que determina quando a linguagem gerada por alguma gramática livre do contexto contém qualquer palavra de comprimento menor que algum número n.