Teorema 4.16 (Regulares $\neq$ livres de contexto) *Existem linguagens livres de contexto que não são regulares.*

<u>Justificativa</u> Considere-se, por exemplo, a linguagem $L = \{0^i 1 0^i \mid i \geq 1\}$. Conforme demonstrado no Exemplo 3.53, L não é regular, uma vez que não satisfaz ao "Pumping Lemma" das linguagens regulares. No entanto, a gramática $G = (\{S, 0, 1\}, \{0, 1\}, \{S \to 0S0, S \to 010\}, S)$ gera exatamente L. Logo, L é livre de contexto, porém não é regular. De maneira semelhante, pode-se demonstrar a existência de inúmeras outras linguagens com essa mesma característica. ■

Conforme definido na Seção 4.1, uma linguagem é dita estritamente livre de contexto se ela for livre de contexto, porém não-regular. A característica desse tipo de linguagens é que elas são geradas apenas por gramáticas que possuam pelo menos um símbolo não-terminal que seja auto-recursivo central e essencial. Assim, é possível gerar, nas formas sentenciais geradas pela gramática, subcadeias da forma:

$$Y \Rightarrow^* \alpha Y \beta, \quad \text{com } \alpha, \beta \in \Sigma^+$$

as quais são responsáveis pelo balanceamento dos termos α e β nas sentenças da linguagem.

De fato, gramáticas lineares à direita não permitem a definição de símbolos não-terminais com essa propriedade. Ao contrário, gramáticas lineares à direita geram apenas formas sentenciais da seguinte forma:

$$Y \Rightarrow^* \alpha Y, \quad \text{com } \alpha \in \Sigma^+$$

em que, obviamente, não há balanceamento de termos nem, portanto, aninhamentos sintáticos. A existência de termos balanceados (ou aninhados) é, por isso, o fator que diferencia uma linguagem estritamente livre de contexto de uma linguagem regular.

4.10 Linguagens que não são Livres de Contexto

Como se sabe, as linguagens regulares podem ser definidas através de expressões regulares, gramáticas regulares ou autômatos finitos. Por outro lado, a existência de linguagens não-regulares foi provada através do uso do "Pumping Lemma" das linguagens regulares.

De maneira análoga, para provar que uma linguagem é livre de contexto, é suficiente apresentar uma gramática livre de contexto que gere esta linguagem, ou ainda um autômato de pilha que a reconheça. A existência de linguagens que não são livres de contexto pode ser demonstrada com o auxílio do "Pumping Lemma" das linguagens livres de contexto, apresentado a seguir.

A importância de um "Pumping Lemma" para as linguagens livres de contexto não se limita à posssibilidade de demonstrar a existência de linguagens que não sejam livres de contexto, e, portanto, de sugerir a existência de outras classes de linguagens. Ele é usado também na demonstração de algumas propriedades importantes das linguagens livres de contexto.

Teorema 4.17 ("Pumping Lemma" para linguagens livres de contexto) *Seja L uma linguagem livre de contexto, com $\epsilon \notin L$. Então, existe uma constante inteira n, dependente apenas de L, que satisfaz às seguintes condições: (i) $\forall\, \gamma \in L, |\gamma| \geq n, \gamma = uvwxy$; ; (ii) $|vwx| \leq n$; (iii) $|vx| \geq 1$; (iv) $\forall\, i \geq 0, uv^i wx^i y \in L$.*

<u>Justificativa</u> Se L é livre de contexto e não contém a cadeia vazia, então $L = L(G)$, sendo $G = (V, \Sigma, P, S)$, G uma gramática livre de contexto na Forma Normal de Chomsky. Portanto, qualquer árvore de derivação que represente uma sentença de L nessa gramática será uma árvore binária.

Por outro lado, árvores binárias de altura i geram sentenças de comprimento máximo 2^{i-1}. A Figura 4.19 ilustra essa relação para os casos i=1, 2 e 3, e facilita a generalização da mesma:

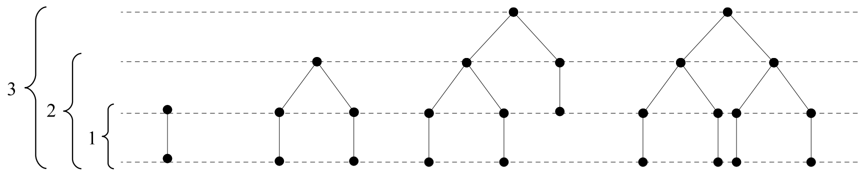

Figura 4.19. Altura e comprimento de sentenças em árvores binárias geradas por gramáticas na Forma Normal de Chomsky

Através desta figura, é fácil perceber que árvores de altura 1 geram sentenças de comprimento 1 (portanto $\leq 2^0$), árvores de altura 2 geram sentenças de comprimento 2 ($\leq 2^1$) e árvores de altura 3 geram sentenças de comprimento 3 ou 4 ($\leq 2^2$). No caso geral, árvores com altura i geram sentenças de comprimento menor ou igual a 2^{i-1}.

Como conseqüência desse resultado, é possível inferir que, se uma sentença tem comprimento mínimo 2^i, então a árvore de derivação correspondente possuirá altura mínima $i + 1$.

Considere-se $k = |N|$, onde N é o conjunto dos símbolos não-terminais de G, estando esta expressa na Forma Normal de Chomsky, e faça-se $n = 2^{k-1} + 1$.[3]

Se $\gamma \in L$ e $|\gamma| \geq n$, isto é, se $|\gamma| \geq 2^{k-1} + 1$, então, face ao resultado anterior, é certo que a altura da árvore de derivação correspondente à sentença γ será maior ou igual a $k + 1$ (pois com árvores de altura k é possível apenas gerar cadeias de comprimento máximo 2^{k-1}). Suponha-se que o valor desta altura seja p (portanto, $p \geq k + 1$) e considere-se um caminho z qualquer na árvore que possua comprimento p (haverá pelo menos um caminho que satisfaça a essa condição).

Se o caminho selecionado possui comprimento p, então este caminho é formado de q símbolos, $q \geq k + 2$. Desses q símbolos, apenas um será terminal (o último símbolo do caminho, aquele que é folha da árvore) e os demais serão necessariamente não-terminais.

Ignorando o símbolo terminal e concentrando a atenção nos $q - 1$ não-terminais, seus antecessores, se $q \geq k + 2$, então existem r símbolos não-terminais neste caminho, e $r = q - 1$, ou seja, $r \geq k + 1$.

Considerem-se agora apenas os primeiros $k + 1$ símbolos não-terminais consecutivos que antecedem imediatamente a folha da árvore, ignorando os $r - (k + 1)$ símbolos não-terminais situados no início do caminho selecionado. O caminho escolhido z pode, portanto, ser considerado como:

$$z = \mu\rho\sigma, \text{ com } \mu \in N^*, |\mu| = r - (k+1), \rho \in N^+, |\rho| = k + 1 \text{ e } \sigma \in \Sigma$$

A Figura 4.20 ilustra esta interpretação do caminho z:

[3] Alguns autores adotam $n = 2^k$. Apesar de correto (pois se L é infinita sempre haverão cadeias com comprimento maior ou igual a esse valor, ou qualquer outro), tal valor de n desconsidera a possibilidade de uso das cadeias $z \in L$ tais que $2^{k-1} + 1 \leq |z| < 2^k$, as quais também satisfazem os critérios do "Pumping Lemma".

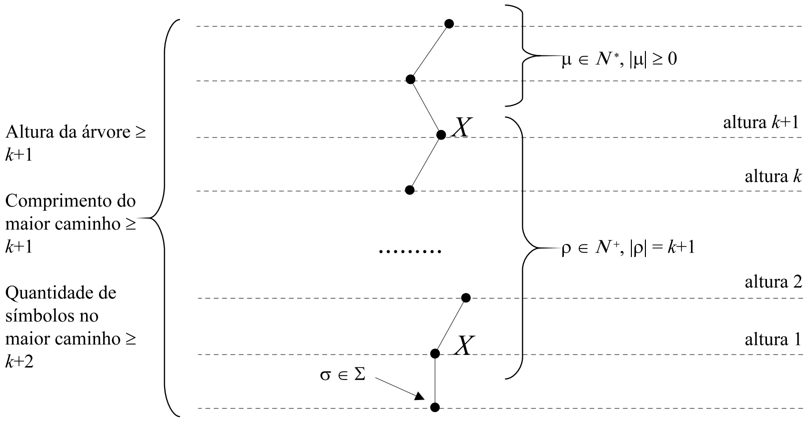

Figura 4.20. Múltiplas ocorrências do símbolo não-terminal X no caminho z

Se ρ contém exatamente $k+1$ símbolos não-terminais e a gramática contém apenas k símbolos não-terminais distintos, é certo que pelo menos um símbolo não-terminal de G aparece mais de uma vez em ρ. Se chamarmos a este símbolo X, as seguintes considerações poderão ser feitas acerca das posições em que os símbolos X podem aparecer na árvore de derivação:

1. Para o *primeiro* X (aquele que está na posição mais alta da árvore, próxima da raiz):

   - Ele pode ser o primeiro símbolo da cadeia ρ (portanto, estar na altura $k+1$);
   - Ele pode ser o penúltimo símbolo da cadeia ρ (portanto, estar na altura 2), uma vez que o segundo X comparece por último nesta mesma cadeia;
   - Ele pode assumir qualquer posição entre essas duas.

2. Para o *segundo* X (aquele que está na posição mais baixa da árvore, próxima da folha):

   - Ele pode ser o segundo símbolo da cadeia ρ (portanto, estar na altura k), uma vez que o primeiro X comparece antes nesta mesma cadeia;
   - Ele pode ser o último símbolo da cadeia ρ (portanto, na altura 1);
   - Ele pode assumir qualquer posição entre essas duas.

Assim, pode-se concluir:

1. Para o *primeiro* X: se $X \Rightarrow^* \alpha_1, \alpha_1 \in \Sigma^*$, então $2 \leq |\alpha_1| \leq 2^k$;
2. Para o *segundo* X: se $X \Rightarrow^* \alpha_2, \alpha_2 \in \Sigma^*$, então $1 \leq |\alpha_2| \leq 2^{k-1}$.

A situação da árvore de derivação da sentença γ pode ser representada como mostra a Figura 4.21 (S representa a raiz de G, e pode, eventualmente, coincidir com o primeiro símbolo da cadeia ρ):

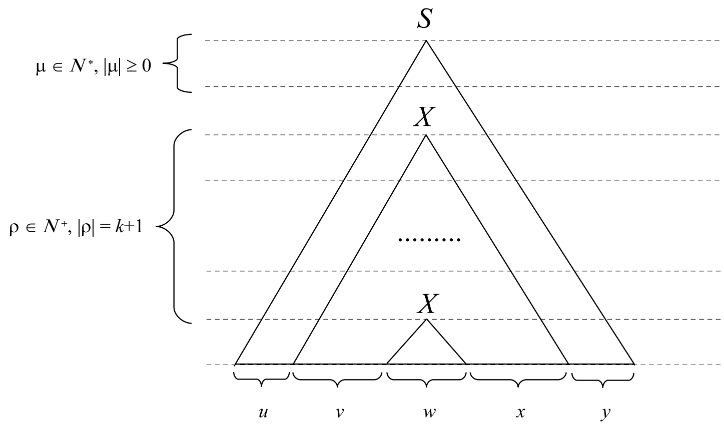

Figura 4.21. Árvore de derivação da cadeia $\gamma = uvwxy$

A inspeção desta árvore nos permite chegar às seguintes conclusões:

- A sentença γ pode ser considerada como sendo composta por cinco partes, $\gamma = uvwxy$;
- Como $w = \alpha_2$, então $1 \leq |w| \leq 2^{k-1}$;
- Como $vwx = \alpha_1$, então $2 \leq |vwx| \leq 2^k$;
- Como $1 \leq |w|$ e $2 \leq |vwx|$, então $|vx| \geq 1$.

Além disso, as regras de G permitem a derivação das seguintes formas sentenciais:

- $S \Rightarrow^* uXy$
- $X \Rightarrow^* vXx$
- $X \Rightarrow^* w$

A inspeção dessas formas sentenciais, assim como a análise da árvore mostrada na Figura 4.21, permite concluir que existem outras possibilidades de derivações (ou seja, de construção da árvore) de sentenças em G, as quais produzem sentenças diversas da original γ, e que, por construção, devem necessariamente pertencer à linguagem L.

Por exemplo, é possível imaginar que, em vez de aplicar regras que fazem o primeiro X derivar vXx, é possível aplicar, a esta ocorrência de X, as regras aplicadas ao segundo X, e dessa maneira gerar a sentença uwy no lugar da sentença $uvwxy$. Da mesma forma, seria possível repetir a derivação aplicada ao primeiro X no lugar da derivação feita para o segundo X, e com isso gerar a sentença $uvvwxxy$. Generalizando, todas as sentenças uv^iwx^iy, com $i \geq 0$, podem ser derivadas em G, e, portanto, pertencem necessariamente a $L(G)$. A Figura 4.22 mostra, graficamente, a aplicação desta idéia:

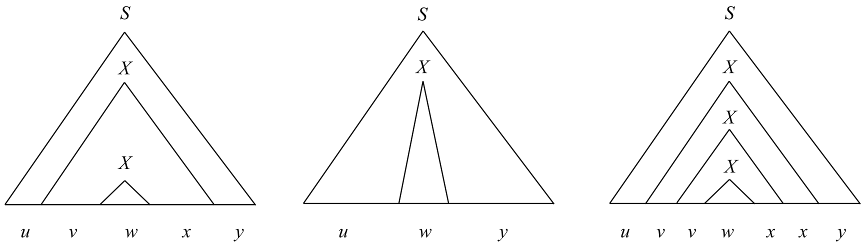

Figura 4.22. Árvores de derivação para as cadeias $uvwxy$, uwy e $uvvwxxy$

Em termos gramaticais, como $S \Rightarrow^* uXy$, $X \Rightarrow^* vXx$ e $X \Rightarrow^* w$, ou seja, $S \Rightarrow^* uvwxy$, então, necessariamente:

- $S \Rightarrow^* uXy \Rightarrow^* uwy$
- $S \Rightarrow^* uXy \Rightarrow^* uvvwxxy$
- $S \Rightarrow^* uXy \Rightarrow^* uvvvwxxxy$
- $\ldots$
- $S \Rightarrow^* uXy \Rightarrow^* uv^iwx^iy, \forall\, i \geq 0$

■

Considerado de maneira informal, este "Pumping Lemma" para linguagens livres de contexto pode ser interpretado da seguinte maneira:

- Sentenças que apresentem um certo comprimento mínimo geram àrvores de derivação com uma certa altura mínima;
- Como a quantidade de símbolos não-terminais da gramática é limitada, necessariamente deverá haver repetição de símbolos em algum caminho da árvore de derivação correspondente;
- Tal repetição permite a geração de uma quantidade infinita de outras sentenças estruturalmente similares, que devem necessariamente pertencer à linguagem.

O símbolo não-terminal X, na demonstração acima, possibilita derivações do tipo $X \Rightarrow^* vXx$ e corresponde ao símbolo que denominamos previamente de auto-recursivo central essencial. A existência de pelo menos um símbolo desse tipo na gramática garante a possibilidade de se gerar uma quantidade infinita de outras sentenças que também pertençam à mesma linguagem.

Se v ou x forem vazios, a linguagem será regular (note-se que o "Pumping Lemma" para linguagens regulares — Teorema 3.16 — é um caso particular do "Pumping Lemma" para linguagens livres de contexto). Se v e x forem simultaneamente diferentes da cadeia vazia, então a linguagem é livre de contexto e não-regular, e os termos v e x ocorrerão sempre de forma balanceada nas sentenças da linguagem.

Exemplo 4.39 A linguagem $\{a^iba^i \mid i \geq 1\}$ é gerada pela gramática livre de contexto $G = (\{S, X, Y, Z, a, b\}, \{a, b\}, \{S \rightarrow XY \mid b, X \rightarrow ZS, Z \rightarrow a, Y \rightarrow a\}, S)$, que está na Forma

Normal de Chomsky e possui $N = \{S, X, Y, Z\}$, $|N| = 4$. De acordo com a demonstração do "Pumping Lemma", a constante n para G é $2^{k-1}+1 = 2^{|N|-1}+1 = 2^{4-1}+1 = 2^3+1 = 8+1 = 9$.

A cadeia $z = a^4ba^4 \in L(G)$ é tal que $|z| \geq 9$. Portanto, $z = uvwxy$, com $|vwx| \leq n$ e $|vx| \geq 1$. A Figura 4.23 apresenta a árvore de derivação correspondente à cadeia z.

A árvore da Figura 4.23 possui altura 9 (portanto maior ou igual a $|N|+1 = 5$) e contém diversos caminhos formados por $|N|+2 = 6$ ou mais símbolos, nos quais é possível isolar os $|N|+1$ últimos símbolos não-terminais.

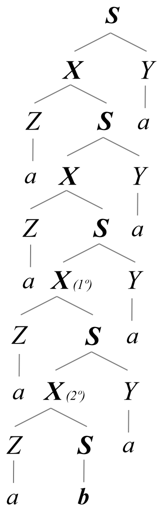

Figura 4.23. Árvore de derivação para a sentença a^4ba^4

Considere-se o caminho $SXSXSXSXSb$, cujo comprimento é 10, e os 5 últimos símbolos não-terminais que o compõe, $SXSXS$. De acordo com o "Pumping Lemma", esse caminho deve conter (e contém) pelo menos um símbolo não-terminal repetido: o símbolo S ocorre três vezes e o símbolo X ocorre duas vezes.

Considere-se o símbolo X. A primeira ocorrência do mesmo gera a cadeia $aaba$, ao passo que a segunda ocorrência gera a cadeia ab. Logo:

$$z = \underbrace{aa}_{u}\underbrace{a}_{v}\underbrace{ab}_{w}\underbrace{a}_{x}\underbrace{aaa}_{y}$$

Conseqüentemente, as cadeias uv^iwx^iy, com $i \geq 0$, devem também pertencer à linguagem. De fato, as cadeias uwy e uv^2wx^2y pertencem a $L(G)$:

$$\underbrace{aa}_{u}\underbrace{ab}_{w}\underbrace{aaa}_{y} = a^3ba^3$$

$$\underbrace{aa}_{u}\underbrace{a}_{v}\underbrace{a}_{v}\underbrace{ab}_{w}\underbrace{a}_{x}\underbrace{a}_{x}\underbrace{aaa}_{y} = a^5ba^5$$

A cadeia uwy pode ser obtida substituindo-se as derivações aplicadas ao primeiro símbolo X pelas aplicadas ao segundo símbolo X. A nova seqüência de derivações torna-se (ver Figura 4.24):

$$S \Rightarrow^* aa\underbrace{X}_{1°}aaa \Rightarrow aa\underbrace{ab}_{w}aaa = a^3ba^3$$

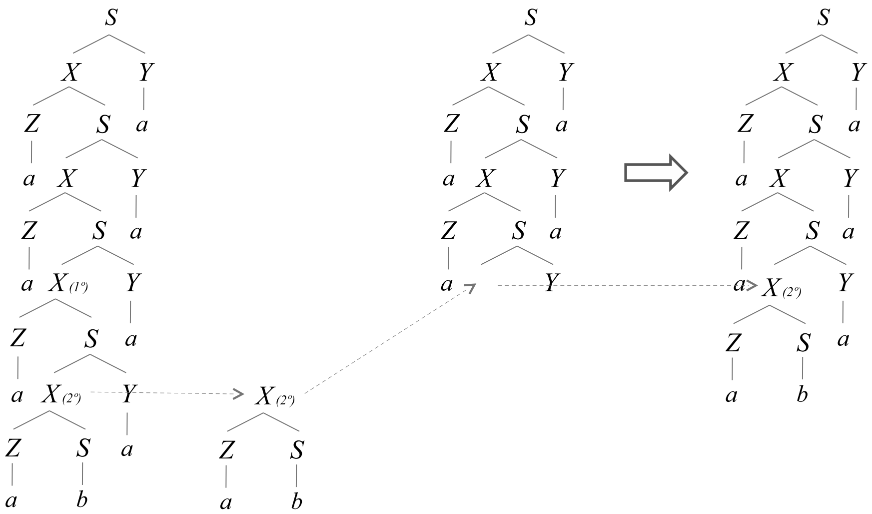

Figura 4.24. Árvore de derivação para a sentença a^3ba^3

A cadeia uv^2wx^2y pode ser obtida substituindo-se as derivações aplicadas ao segundo símbolo X pelas aplicadas ao primeiro símbolo X. A nova seqüência de derivações torna-se (ver Figura 4.25):

$$S \Rightarrow^* aaa \underbrace{X}_{2^o} aaaa \Rightarrow aaa \underbrace{aaba}_{vwx} aaaa = a^5ba^5$$

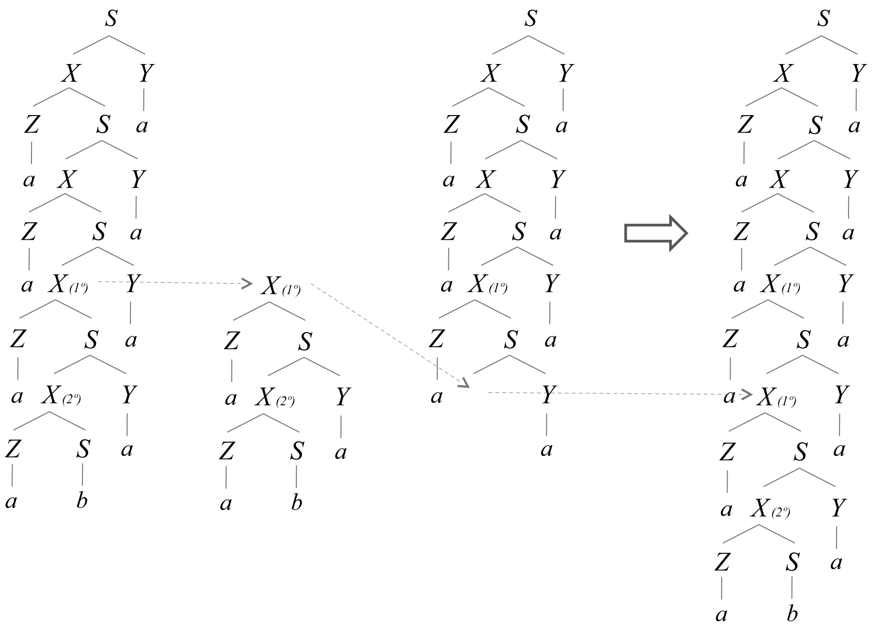

Figura 4.25. Árvore de derivação para a sentença a^5ba^5

A iteração do último passo (Figura 4.25) possibilita a geração das sentenças uv^iwx^iy, com $i \geq 3$. □

Como no caso do "Pumping Lemma" para linguagens regulares, que estabelece uma propriedade inerente a toda e qualquer linguagem regular, o "Pumping Lemma" para linguagens livres de contexto estabelece uma propriedade inerente a toda e qualquer linguagem livre de contexto. De maneira análoga àquele caso, uma de suas principais aplicações é na demonstração da existência de linguagens que não são livres de contexto, conforme mostrarão os exemplos a seguir.

Observe-se que o teorema garante a existência da constante n, mas a aplicação do mesmo não exige que seja determinado o seu valor, como mostram os exemplos seguintes.

Exemplo 4.40 A linguagem $L_1 = \{a^k b^k c^k \mid k \geq 1\}$ não é livre de contexto. Suponha-se, por hipótese, que L_1 seja livre de contexto. De acordo com o Teorema 4.17 ("Pumping Lemma" para linguagens livres de contexto), existe uma constante inteira n tal que, qualquer que seja a sentença $\gamma \in L_1$, $|\gamma| \geq n$, então $\gamma = uvwxy$, $|vwx| \leq n$, $|vx| \geq 1$ e $uv^i wx^i y \in L_1$.

Seja $\gamma = a^n b^n c^n$. Como $|a^n b^n c^n| = 3 * n$, então está satisfeita a condição $|\gamma| \geq n$. Portanto, $\gamma = uvwxy$. Nesta situação, a subcadeia vwx pode assumir um dos seguintes formatos (lembrar que $|vwx| \leq n$):

1. vwx contém apenas símbolos "a" (pelo menos um); portanto, vx também contém apenas símbolos a (pelo menos um);

2. vwx contém apenas símbolos "b" (pelo menos um); portanto, vx também contém apenas símbolos b (pelo menos um);

3. vwx contém apenas símbolos "c" (pelo menos um); portanto, vx também contém apenas símbolos c (pelo menos um);

4. vwx contém símbolos "a" (pelo menos um) seguidos de símbolos "b" (pelo menos um); portanto, vx contém pelo menos um símbolo a ou b, porém nenhum símbolo c;

5. vwx contém símbolos "b" (pelo menos um) seguidos de símbolos "c" (pelo menos um); portanto, vx contém pelo menos um símbolo b ou c, porém nenhum símbolo a.

É impossível que a subcadeia vwx contenha simultaneamente símbolos "a", "b" e "c", uma vez que, para isso acontecer, seria necessário que o comprimento de vwx fosse no mínimo $n+2$, o que contraria o "Pumping Lemma".

Passando-se à análise de cada uma dessas possibilidades, considerando o formato assumido pela sentença uwy, isto é, pela sentença $uvwxy$ após a remoção das subcadeias v e x, o que é previsto pelo "Lemma" quando estabelece que todas as sentenças $uv^i wx^i y$ devem pertencer a L_1 (neste caso, faz-se $i = 0$). Lembrar que $|vx| \geq 1$:

1. uwy conterá n símbolos "b", n símbolos "c" e uma quantidade de símbolos "a" menor que n; logo, uwy não pertence a L_1;

2. uwy conterá n símbolos "a", n símbolos "c" e uma quantidade de símbolos "b" menor que n; logo, uwy não pertence a L_1;

3. uwy conterá n símbolos "a", n símbolos "b" e uma quantidade de símbolos "c" menor que n; logo, uwy não pertence a L_1;

4. uwy conterá n símbolos "c", e quantidades de símbolos "a" e de símbolos "b" respectivamente menores que n; logo, uwy não pertence a L_1;

5. uwy conterá n símbolos "a", e quantidades de símbolos "b" e de símbolos "c" respectivamente menores que n; logo, uwy não pertence a L_1.

Portanto, a sentença $a^n b^n c^n$ contradiz a hipótese inicial, provando que a linguagem L_1 não é livre de contexto. Cumpre observar que, no Exemplo 3.55 da Seção 3.9 ("Pumping Lemma" para as Linguagens Regulares), esta mesma linguagem foi demonstrada como sendo não-regular. □

Exemplo 4.41 A linguagem $L_2 = \{wcw \mid w \in \{a, b\}^+\}$ não é livre de contexto. Esta linguagem sintetiza uma característica bastante comum nas linguagens de programação mais usuais: a necessidade de se usarem identificadores idênticos em partes diferentes de um mesmo programa, por exemplo, na declaração de uma variável e na ocasião de sua utilização posterior.

Suponha-se, por hipótese, que L_2 seja livre de contexto. De acordo com o "Pumping Lemma" para linguagens livres de contexto, existe uma constante inteira n tal que, qualquer que seja a sentença $\gamma \in L_2$, $|\gamma| \geq n$, então $\gamma = uvwxy$, $|vwx| \leq n$, $|vx| \geq 1$ e $uv^iwx^iy \in L_2$.

Considere-se, por exemplo, a sentença $\gamma = a^n b^n c a^n b^n$. Como $|\gamma| = 4 * n + 1 \geq n$, então está satisfeita a condição do "Pumping Lemma" para a escolha da sentença γ. Portanto, $\gamma = uvwxy$ e a subcadeia vwx pode assumir um dos seguintes formatos:

1. vwx contém apenas símbolos "a" (pelo menos um);
2. vwx contém apenas símbolos "b" (pelo menos um);
3. vwx contém símbolos "a" (pelo menos um) seguidos de símbolos "b" (pelo menos um);
4. vwx se inicia com símbolos "b" (zero ou mais) e termina com o símbolo "c";
5. vwx se inicia com o símbolo "c" e termina com símbolos "a" (zero ou mais);
6. vwx se inicia com símbolos "b" (pelo menos um) , continua com um símbolo "c" e termina com símbolos "a" (pelo menos um).

De maneira análoga ao que foi mostrado no Exemplo 4.40, passa-se a examinar o formato das cadeias uwy que são geradas em cada um desses casos:

1. Ao extrair símbolos "a" da subcadeia w que antecede o símbolo "c", esta resulta diferente da subcadeia w posterior ao mesmo "c"; o mesmo acontece se forem extraídos símbolos "a" da subcadeia posterior ao símbolo "c"; logo, uwy não pertence a L_2;
2. Semelhante ao caso anterior;
3. Modifica, exclusivamente, a subcadeia anterior ao símbolo "c" ou a subcadeia posterior ao símbolo "c"; logo, uwy não pertence a L_2;
4. Como $|vx| \geq 1$, vx contém pelo menos um símbolo "b" ou um símbolo "c", que, se removidos da sentença γ, provocam a geração de uma sentença que não pertence a L_2 (seja porque as cadeias w tornam-se diferentes, seja porque o símbolo "c" que as separa desaparece);
5. Semelhante ao caso anterior;
6. Como $|vx| \geq 1$, vx contém pelo menos um símbolo "b" (da primeira subcadeia w) ou um símbolo "c" ou, ainda, um símbolo "a" (da segunda subcadeia w); qualquer que seja o caso, a cadeia resultante uwy não pertence a L_2 (seja porque as cadeias w tornam-se diferentes, seja porque o símbolo "c" que as separa desaparece).

Pelo exposto, fica claro que a hipótese inicial é falsa e que L_2 não é uma linguagem livre de contexto. □

Exemplo 4.42 A linguagem $L_3 = \{a^k \mid k \geq 1 \text{ é um número primo}\}$ não é livre de contexto. Suponha-se que L_3 seja livre de contexto e considere-se a sentença $\gamma = a^p$, cujo comprimento seja maior que o valor da constante n definida pelo "Pumping Lemma" para linguagens livres de contexto, portanto, $p > n$. Se $\gamma = uvwxy$ pertence a L_3, então, de acordo com o "Lemma", a sentença uwy também deve pertencer. Seja $q = |uwy|$.

O comprimento das sentenças uv^iwx^iy, portanto, pode ser calculado da seguinte forma: $|uv^iwx^iy| = |uwy| + i^* |vx|$. Em particular, o comprimento da sentença $|uv^qwx^qy| = |uwy| + q^* |vx| = q + q^* |vx| = q^*(1 + |vx|)$. Logo, o comprimento de $|uv^qwx^qy|$ não é primo e isso prova que L_3 não pode ser livre de contexto.

Vale lembrar que, no Exemplo 3.56 da Seção 3.9 ("Pumping Lemma" para as Linguagens Regulares), esta mesma linguagem foi demonstrada como sendo não-regular. □

4.11 Linguagens Livres de Contexto Determinísticas

Diferentemente da classe das linguagens regulares, que podem ser reconhecidas indistintamente por autômatos finitos, determinísticos ou não-determinísticos, somente autômatos de pilha não-determinísticos são capazes de reconhecer as linguagens livres de contexto, no caso geral.

Esta conclusão, que pode ser provada demonstrando-se a existência de linguagens livres de contexto não-reconhecíveis por autômatos de pilha determinísticos, quaisquer que sejam eles, possui uma importante conseqüência prática, uma vez que os autômatos de pilha não-determinísticos são por natureza pouco eficientes, e por isso constituem modelos de implementação pouco atraentes em diversas situações, entre as quais, na importante área de construção de compiladores.

Exemplo 4.43 A linguagem $L_1 = \{ww^R \mid w \in \{a,b\}^*\}$ é não-determinística, ou seja, é reconhecida apenas por autômatos de pilha não-determinísticos. Um exemplo de tal autômato, com critério de aceitação baseado em pilha vazia, é apresentado na Figura 4.26.

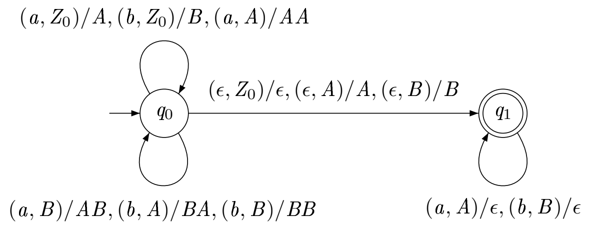

Figura 4.26. Autômato de pilha para a linguagem $\{ww^R \mid w \in \{a,b\}^*\}$

São exemplos de cadeias aceitas por este autômato aa e $abba$, como se pode verificar nos reconhecimentos abaixo:

- $(q_0, aa, Z_0) \vdash (q_0, a, A) \vdash (q_1, a, A) \vdash (q_1, \epsilon, \epsilon)$
- $(q_0, abba, Z_0) \vdash (q_0, bba, A) \vdash (q_0, ba, BA) \vdash (q_1, ba, BA) \vdash (q_1, a, A) \vdash (q_1, \epsilon, \epsilon)$

Como é fácil observar, a característica não-determinística do autômato é fundamental para que ele possa determinar o ponto exato da cadeia de entrada em que termina a subcadeia w e se inicia a subcadeia w^R: apenas uma movimentação em vazio no ponto central da cadeia será capaz de fazer com que o autômato atinja uma configração final. Tal característica é necessária, uma vez que o autômato, por conta de suas próprias limitações, não tem condições de determinar diretamente o ponto exato em que se inicia a cadeia w^R. □

Assim, apesar do elevado interesse prático exibido pelas linguagens livres de contexto, decorrente da relativa facilidade com que elas podem ser representadas e manipuladas, a dificuldade acima mencionada levou os pesquisadores a um estudo mais aprofundado dessa classe de linguagens, visando fornecer aos projetistas e implementadores de linguagens artificiais subsídios que permitissem o aproveitamento prático das características dessa classe de linguagens, sem no entanto sacrificar o desempenho dos respectivos reconhecedores pela presença de eventuais não-determinismos.

Como conseqüência, foi identificada e caracterizada a classe das **linguagens livres de contexto determinísticas**, que se provou ser subconjunto próprio das linguagens livres de contexto genéricas. Por definição, linguagens livres de contexto determinísticas são aquelas que podem ser reconhecidas por autômatos de pilha determinísticos.