8) THE ROAD SO FAR: LINGUAGENS REGULARES

Linguagens Regulares

→ Por definição: aquelas reconhecidas por AFDs

→ Equivalentemente:

aquelas reconhecidas por AFNs

aquelas descritas por ERs

→ Representam problemas simples:

- $\{w \in \{0,1\}^* : w$ termina em $0\}$ → Decidir se um nº é par
- $\{w \in \{0,1,\dots,9\}^* :$ a soma dos dígitos em w é múltiplo de 3 e o primeiro símbolo de w é $\neq 0\}$ → Decidir se um nº é múltiplo de 3
- $\{w \in \{0\}^* : |w|$ é ímpar$\}$ → Decidir se um nº é ímpar
- {var, function, if, then, else, while, do, let, in, end, printf, getint} → Decidir se uma palavra é reservada
- $\{w \in \{0,1\}^* : w$ começa com 1 e quando interpretado em binário é múltiplo de $5\}$ → Decidir se um nº é múlt.
- $\{w \in \{a,\dots,z\}^* : w$ começa com $lin\}$ → Decidir se a busca foi bem-sucedida

→ É fácil / possível:

- Decidir se um AFD aceita uma dada cadeia
- Decidir se um AFD aceita alguma cadeia
- Decidir se a linguagem de um AFD é finita
- Verificar se dois AFDs são equivalentes

TODAS AS LINGUAGENS

$\{0^n 1^n : n \geq 1\}$

LINGUAGENS REGULARES

$\{0^n 1^m : n \geq 0$ e $m \geq 0\} = 0^* 1^*$

Σ^*

AFD
AFN
ER

Note: $\{0^n 1^n : n \geq 1\} \subseteq 0^* 1^*$

9) GRAMÁTICAS LIVRES DE CONTEXTO

Gramáticas

- → São dispositivos geradores que se aproveitam de recursão
- → Os símbolos do alfabeto das cadeias da linguagem são chamados terminais.
- → Existe um outro alfabeto, cujos símbolos são chamados de variáveis.
- → Cada variável representa uma linguagem.
  - A variável inicial representa a linguagem "principal".
- → Existe um conjunto finito de regras de substituição:
  - ↳ especificam combinações válidas de símbolos
  - ↳ uma regra $\alpha \to \beta$ indica que α pode ser substituído por β sendo que α e β são cadeias de variáveis e terminais.
- → Geramos cadeias de terminais por meio de uma sequência de substituições a partir da variável inicial.

Exemplo

→ Considerando $\Sigma = \{a, 1\}$:

$$A \to aA \qquad\qquad A \to \varepsilon$$

ou

$$A \to aA \mid \varepsilon$$

Gramáticas

→ As restrições sobre como α e β podem ser nas regras de substituição indicam uma gramática diferente:

- Lineares
- Livres de contexto → Dar exemplo de uma linguagem de programação
- Sensíveis ao contexto
- Irrestritas

Gramáticas Livres de Contexto

DEFINIÇÃO: Uma gramática livre de contexto (GLC) é uma 4-upla (V, Σ, R, S), onde

- V é um conjunto finito de variáveis
- Σ é um conjunto finito de terminais, com $V \cap \Sigma = \emptyset$
- R é um conjunto finito de regras de substituição na forma $X \to \alpha$, onde $X \in V$ e $\alpha \in (\Sigma \cup V)^*$
- S é a variável inicial

Exemplo: $(\{A\}, \{0,1\}, \{A \to 0A, A \to \varepsilon\}, A)$

Usando as regras de substituição

DEFINIÇÃO: Se α, β e w são cadeias de variáveis e terminais e $A \to w$ é uma regra de substituição, então dizemos que "$\alpha A \beta$ origina $\alpha w \beta$" e escrevemos $\alpha A \beta \Rightarrow \alpha w \beta$.

Ex: $A \to 0A \mid \varepsilon$

Derivações

DEFINIÇÃO: Sejam α e β cadeias de variáveis e terminais. Dizemos que α deriva β, denotado $\alpha \overset{*}{\Rightarrow} \beta$, se

1) $\alpha = \beta$, ou
2) existem cadeias de var. e term. $\gamma_1, \gamma_2, \ldots, \gamma_k$, $k \geq 0$, tais que
$$\alpha \Rightarrow \gamma_1 \Rightarrow \gamma_2 \Rightarrow \cdots \Rightarrow \gamma_k \Rightarrow \beta$$

Ex: $A \to 0A \mid \varepsilon$

Árvores sintáticas (Parse trees)

→ Uma derivação pode ser representada por uma árvore sintática.
- ↳ Cada nó interno é uma variável.
- ↳ Cada folha é uma variável, um terminal ou ε.
- ↳ Se um nó interno tem rótulo A e seus filhos são, da esquerda para a direita, $X_1, X_2, \dots, X_k$, então $A \to X_1 X_2 \dots X_k$ é uma regra de produção da gramática.

Ex: $A \to 0A \mid \varepsilon$

Leitura: percurso in-ordem

Mais exemplos

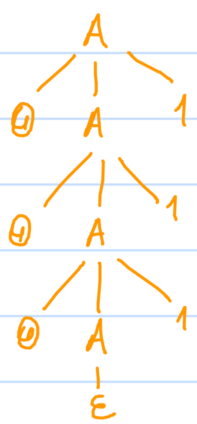
$G_1 = (\{A, B\}, \{0, 1\}, R, A)$, com R dada abaixo

$$A \to 0A1 \mid \varepsilon$$

$A \Rightarrow \varepsilon$

$A \Rightarrow 0A1 \Rightarrow 01$

$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 0011$



$G_2 = (\{P\}, \{0, 1\}, R, P)$, com R dada abaixo

$$P \to \varepsilon \mid 0 \mid 1 \mid 0P0 \mid 1P1$$

$P \Rightarrow \varepsilon$

$P \Rightarrow 0$

$P \Rightarrow 1P1 \Rightarrow 101$

$P \Rightarrow 1P1 \Rightarrow 11P11 \Rightarrow 110P011 \Rightarrow 1101011$

A linguagem de uma gramática

DEFINIÇÃO: Seja G uma gramática cuja variável inicial é S e o conjunto de terminais é Σ.

A linguagem de G, denotada $L(G)$, é o conjunto

$$\{w \in \Sigma^* : S \overset{*}{\Rightarrow} w\}$$

Ex: $P \to \varepsilon \mid 0 \mid 1 \mid 0P0 \mid 1P1$

$L(G_2) = \{w \in \{0,1\}^* : w = w^R\}$
$= \{w \in \{0,1\}^* : w \text{ é palíndromo}\}$

① Se $P \overset{*}{\Rightarrow} w$, argumente que $w \in X$

Se $P \overset{*}{\Rightarrow} w$: $w = \varepsilon, 0, 1$ ou $w = 0\alpha 0$ ou $w = 1\alpha 1$ sendo que $P \overset{*}{\Rightarrow} \alpha$.

Assumindo que $\alpha \in X$, note que $0\alpha 0$, $1\alpha 1$ também pertencem a X.

② Se $w \in X$, argumente que $P \overset{*}{\Rightarrow} w$.

Se $w \in X$: $w = \varepsilon, 0, 1$ ou $w = 0\alpha 0$ ou $w = 1\alpha 1$ com $\alpha \in X$

Assumindo que $P \overset{*}{\Rightarrow} \alpha$, então note que $P \overset{*}{\Rightarrow} w$ também.

TEOREMA: Seja $G = (\{P\}, \{0,1\}, \{P \to \varepsilon, P \to 0, P \to 1, P \to 0P0, P \to 1P1\}, P)$ uma GLC. Seja $X = \{w \in \{0,1\}^* : w = w^R\}$. Então $L(G) = X$.

DEMONSTRAÇÃO: Vamos mostrar que $w \in L(G) \Leftrightarrow w \in X$.

($\Rightarrow$) Suponha que $w \in L(G)$. Por definição, $P \overset{*}{\Rightarrow} w$.

Vamos mostrar por indução no número n de passos da derivação que $w \in X$

BASE: $n = 1$. Então $P \Rightarrow w$, e só pode ser o caso de $w = \varepsilon$, $w = 0$ ou $w = 1$.

nos três casos, w é palíndromo e $\therefore$ $w \in X$.

Seja então o caso de $P \overset{*}{\Rightarrow} w$ em $n > 1$ passos.

Suponha que se $P \overset{*}{\Rightarrow} \alpha$ em k passos, com $1 \leq k < n$, então $\alpha \in X$.

Como $n > 1$, então $P \overset{*}{\Rightarrow} w$ pode ser escrito $P \Rightarrow \gamma \overset{*}{\Rightarrow} w$, sendo que $\gamma = 0P0$ ou $\gamma = 1P1$ apenas.

S.p.g., suponha $\gamma = 0P0$. Então $w = 0\alpha 0$, com $\alpha \in \{0,1\}^*$ tendo sido derivada de P em $n-1$ passos.

Logo, por HI, $\alpha \in X$, ou seja, $\alpha = \alpha^R$.

Mas então $0\alpha 0$ também é palíndromo e, portanto, $w \in X$.

TEOREMA: Seja $G = (\{P\}, \{0,1\}, \{P \to \varepsilon, P \to 0, P \to 1, P \to 0P0, P \to 1P1\}, P)$ uma GLC. Seja $X = \{w \in \{0,1\}^* : w = w^R\}$. Então $L(G) = X$.

(Comentar como seria com mais variáveis)

DEMONSTRAÇÃO (CONT.): Vamos mostrar que $w \in L(G) \Leftrightarrow w \in X$

($\Leftarrow$) Suponha que $w \in X$, ou seja, $w = w^R$.

Vamos provar por indução em $|w|$ que $w \in L(G)$.

BASE: $|w| = 0$ e $|w| = 1$. Nesse caso, $w = \varepsilon$, $w = 0$ ou $w = 1$.

Como $P \Rightarrow \varepsilon$, $P \Rightarrow 1$ e $P \Rightarrow 0$, então $w \in L(G)$.

Seja então $w = w^R$ com $n = |w| \geq 2$.

Suponha que para qualquer $\alpha = \alpha^R$ com $0 \leq |\alpha| < n$ vale que $\alpha \in L(G)$.

Como $w = w^R$ e $|w| \geq 2$, então ou $w = 0\lambda 0$ ou $w = 1\lambda 1$ para algum λ tal que $\lambda = \lambda^R$. Como $|\lambda| = n-2$, então por HI temos $\lambda \in L(G)$.

Logo, $P \overset{*}{\Rightarrow} \lambda$. Como $P \to 0P0$ e $P \to 1P1$ são regras de G, então $P \Rightarrow 0P0 \overset{*}{\Rightarrow} 0\lambda 0$ e $P \Rightarrow 1P1 \overset{*}{\Rightarrow} 1\lambda 1$. Logo, $w \in L(G)$.

CQD

Linguagens Livres de Contexto

DEFINIÇÃO: Uma linguagem é *livre de contexto* se alguma gramática livre de contexto a gera.

TODAS AS LINGUAGENS

LINGUAGENS REGULARES — AFD, AFN, ER

$\Sigma^* a$ $\quad \Sigma^*$ $\quad 0^*1^*$

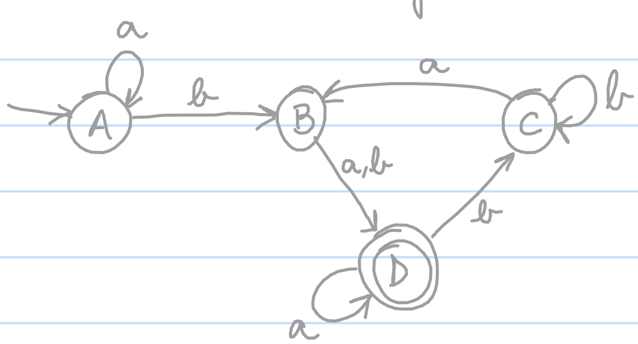
LINGUAGENS LIVRES DE CONTEXTO — GLC

$\{0^n 1^n : n \geq 1\}$

$\{w \in \Sigma^* : w = w^R\}$

TEOREMA: Toda linguagem regular é livre de contexto.

Ideia: Se L é regular, então existe AFD que a reconhece. Vamos transformar esse AFD em uma GLC.



$A \to aA \mid bB$

$B \to aD \mid bD$

$C \to aB \mid bC$

$D \to aD \mid bC \mid \varepsilon$

TEOREMA: Toda linguagem regular é livre de contexto.

DEMONSTRAÇÃO: Seja A uma linguagem regular e $M = (Q, \Sigma, \delta, q_0, F)$ um AFD que a reconhece.

Construa a gramática $G = (Q, \Sigma, R, q_0)$ em que R possui uma regra $X \to aY$ para cada transição $\delta(X, a) = Y$ e uma regra $X \to \varepsilon$ para cada $X \in F$. Claramente G é GLC.

Vamos agora mostrar que $L(G) = L(M) = A$.

Primeiro seja $w \in A$, com $w = x_1 x_2 \dots x_n$ onde $x_i \in \Sigma$.
Como $w \in A$, existe sequência $(r_0, r_1, \dots, r_n)$ de estados tais que $r_0 = q_0$, $r_n \in F$ e $r_{i+1} = \delta(r_i, x_{i+1})$ $\forall$ $0 \leq i < n$.
Mas então $q_0 \Rightarrow x_1 r_1 \Rightarrow x_1 x_2 r_2 \overset{*}{\Rightarrow} x_1 x_2 \dots x_n r_n \Rightarrow x_1 x_2 \cdots x_n$ é uma derivação válida para G, o que significa que $w \in L(G)$.

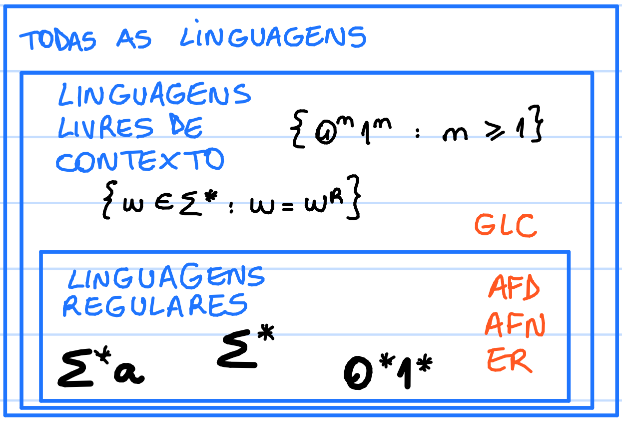
Agora seja $w \in L(G)$, com $w = x_1 x_2 \cdots x_n$ e $x_i \in \Sigma$.
Como $w \in L(G)$, então $q_0 \overset{*}{\Rightarrow} w$, ou seja, existem cadeias de $(\Sigma \cup Q)^*$ $\alpha_0, \dots, \alpha_k$ tais que $\alpha_0 = q_0$ e

$$\alpha_0 \Rightarrow \alpha_1 \Rightarrow \alpha_2 \Rightarrow \cdots \Rightarrow \alpha_k \Rightarrow w$$

Por construção, todas as regras de G são da forma $X \to aY$ ou $Z \to \varepsilon$, com $a \in \Sigma$, $X, Y, Z \in Q$ e $Z \in F$. Então cada α_i só pode ser da forma $\lambda x_i V$, com $\lambda \in \Sigma^*$ e $V \in Q$. Logo, $k = n$.
Vamos denotar por r_i a variável final de α_i $(\alpha_i = \lambda x_i r_i)$.
Assim, $\alpha_i \Rightarrow \alpha_{i+1}$ porque a regra $r_i \to x_{i+1} r_{i+1}$ foi aplicada, para todo $0 \leq i < n$, e $\alpha_n \Rightarrow w$ por aplicação de $r_n \to \varepsilon$.
Isso significa que existem transições $\delta(r_i, x_{i+1}) = r_{i+1}$ em M e que r_n é estado final.
Logo, $(q_0, r_1, \dots, r_n)$ é uma sequência válida de computação em M e portanto $w \in L(M) = A$.

CQD

Linguagens Livres de Contexto



Dada uma GLC, qual linguagem ela gera?

$G_1 = (\{P\}, \{a, b\}, R, P)$ com R dado por

$$P \to aP \mid Pb \mid \varepsilon$$

Lema: Seja $X = \{a^n b^m : n, m \geq 0\}$. Então $L(G_1) = X$.

($\Rightarrow$) Seja $w \in X$. Então $w = a^n b^m$, $n, m \geq 0$. Por indução em $|w|$, vamos provar que $w \in L(G)$.

Se $|w| = 0$, então $w = \varepsilon$ e, de fato, $P \Rightarrow \varepsilon$. $\therefore$ $w \in L(G)$

Seja $|w| = n \geq 1$ e suponha que $\alpha \in X \Rightarrow \alpha \in L(G_1)$ se $|\alpha| < n$.

Como $n \geq 1$, $w = a\lambda$ ou $w = \lambda b$, com $\lambda \in X$. Como $|\lambda| < |w|$, então $P \overset{*}{\Rightarrow} \lambda$, por HI. Como $P \to aP$ e $P \to Pb$ existem, $P \Rightarrow aP \overset{*}{\Rightarrow} a\lambda$ ou $P \Rightarrow Pb \Rightarrow \lambda b$ nos mostram que $w \in L(G_1)$.

($\Leftarrow$) Seja $w \in L(G_1)$. Então $P \overset{*}{\Rightarrow} w$. Por indução no nº de passos k da derivação, vamos provar que $w \in X$.

Se $k = 1$, então $w = \varepsilon$ pois a única regra que gera cadeias com um passo é $P \to \varepsilon$. De fato, $\varepsilon \in X$.

Seja $k > 1$ e suponha que se $\alpha \in L(G_1)$ é gerada com menos de k passos, então $\alpha \in X$.

Como $k > 1$, ocorreu $P \Rightarrow aP \overset{*}{\Rightarrow} a\lambda = w$ ou $P \Rightarrow Pb \overset{*}{\Rightarrow} \lambda b = w$.

Então $P \overset{*}{\Rightarrow} \lambda$ com menos de k passos e, por HI, $\lambda = a^n b^m$, $n, m \geq 0$.

Como $w = a\lambda$ ou $w = \lambda b$, segue que $w \in X$.

RESUMINDO....

① Se $P \overset{*}{\Rightarrow} w$, então $w = \varepsilon$, $w = a\alpha$ ou $w = \alpha b$, com $P \overset{*}{\Rightarrow} \alpha$ por causa dos formatos das regras.
Assumindo que $\alpha \in X$, então $\alpha = a^n b^m$, de forma que $a\alpha$ ou αb continuam pertencendo a X.

② Se $w \in X$, então $w = a^n b^m$ para $n, m \geq 0$, de forma que $w = \varepsilon$, ou começa em a e pode ser escrito $w = a\alpha$ com $\alpha \in X$, ou então termina em b e pode ser escrito $w = \alpha b$ com $\alpha \in X$.
Assumindo que $P \overset{*}{\Rightarrow} \alpha$, então pelas regras temos também $P \overset{*}{\Rightarrow} w$.

$G_2 = (\{P\}, \{a, b\}, R, P)$ com R dado por

$$P \rightarrow aP \mid bP \mid \varepsilon$$

Se $X = \Sigma^*$, então $L(G_2) = X$, pois:

① Se $P \overset{*}{\Rightarrow} w$, então $w = \varepsilon$, $w = a\alpha$ ou $w = b\alpha$, com $P \overset{*}{\Rightarrow} \alpha$.
Diretamente, $\varepsilon \in X$.
Assumindo que $\alpha \in X$, então note que $a\alpha$ e $b\alpha \in X$.

② Se $w \in X$, então $w = \varepsilon$ ou w começa com a ou com b, podendo ser escrita $a\alpha$ ou $b\alpha$ para $\alpha \in X$. Diretamente, $P \Rightarrow \varepsilon$.
Assumindo que $P \overset{*}{\Rightarrow} \alpha$, então note que $P \overset{*}{\Rightarrow} w$.

$G_3 = (\{S\}, \{i, e\}, R, S)$ com R sendo

$$S \rightarrow \varepsilon \mid SS \mid iS \mid iSeS$$

Se $X = \{$sequências válidas de cláusulas if-else em C$\}$, então $L(G_3) = X$, pois:

① Se $S \overset{*}{\Rightarrow} w$, então $w = \varepsilon$, $w = \alpha\beta$, $w = i\alpha$ ou $w = i\alpha e\beta$ com $S \overset{*}{\Rightarrow} \alpha$ e $S \overset{*}{\Rightarrow} \beta$. Diretamente, $\varepsilon \in X$.
Assumindo que $\alpha, \beta \in X$, note que $\alpha\beta$, $i\alpha$, $i\alpha e\beta \in X$.

② Se $w \in X$, então $w = \varepsilon$ ou w é a concat. de duas sequências válidas, cada uma começando com um if, que pode ser sozinho ou casado de um else. ($w = \varepsilon$, $w = \alpha\beta$, $w = i\alpha$, $w = i\alpha e\beta$ com $\alpha, \beta \in X$). Diretamente, $S \Rightarrow \varepsilon$.
Assumindo $S \overset{*}{\Rightarrow} \alpha$ e $S \overset{*}{\Rightarrow} \beta$, então note que $S \overset{*}{\Rightarrow} w$.

Comentar que a árvore não permite saber a ordem das gerações, e que existem várias sequências de derivação e árvores que geram a mesma cadeia.

$G_4 = (\{S,T,X\}, \{0,\#\}, R, S)$ com R dado por

$$S \to TT \mid X$$
$$T \to 0T \mid T0 \mid \#$$
$$X \to 0X00 \mid \#$$

Se $Z = \{w \in \{0,\#\}^* : w = 0^i \# 0^j \# 0^k$ ou $w = 0^n \# 0^{2n}, i,j,k,n \geq 0\}$, $Z_1 = \{w = 0^i \# 0^j : i,j \geq 0\}$ e $Z_2 = \{w = 0^n \# 0^{2n} : n \geq 0\}$, então $L(G_4) = Z$ pois:

(1) Se $X \stackrel{*}{\Rightarrow} w$, então $w = \#$ ou $w = 0\alpha 00$ com $X \stackrel{*}{\Rightarrow} \alpha$. Diretamente, $\# \in Z_2$. Assumindo $\alpha \in Z_2$, $0\alpha 00 \in Z_2$ também.

(2) Se $w \in Z_2$, $w = 0^n \# 0^{2n}$ para $n \geq 0$, logo vale que $w = \#$ ou que $w = 0\alpha 00$ sendo $\alpha \in Z_2$. Diretamente, $X \Rightarrow \#$. Assumindo que $X \stackrel{*}{\Rightarrow} \alpha$, então $X \stackrel{*}{\Rightarrow} 0\alpha 00 = w$.

(1) e (2) mostram que $L(X) = Z_2$.

(3) Se $T \stackrel{*}{\Rightarrow} w$, então $w = 0\alpha$, $w = \alpha 0$ ou $w = \#$ para $T \stackrel{*}{\Rightarrow} \alpha$. Diretamente, $\# \in Z_1$. Assumindo $\alpha \in Z_1$, então $0\alpha, \alpha 0 \in Z_1$.

(4) Se $w \in Z_1$, $w = 0^i \# 0^j$ com $i,j \geq 0$, isto é, $w = \#$ ou $w = 0\alpha$ ou $w = \alpha 0$, com $\alpha \in Z_1$. Diretamente, $T \stackrel{*}{\Rightarrow} w$. Assumindo $T \stackrel{*}{\Rightarrow} \alpha$, então $T \stackrel{*}{\Rightarrow} w$.

(3) e (4) mostram que $L(T) = Z_1$.

(5) Se $S \stackrel{*}{\Rightarrow} w$, então $w = \alpha\beta$ com $T \stackrel{*}{\Rightarrow} \alpha$ e $T \stackrel{*}{\Rightarrow} \beta$ ou então $X \stackrel{*}{\Rightarrow} w$. Se $X \stackrel{*}{\Rightarrow} w$, então diretamente $w \in Z_2$. Se $w = \alpha\beta$, como $\alpha, \beta \in Z_1$, então $w \in Z_1 Z_1$.

Mas então $w \in Z = Z_1 Z_1 \cup Z_2$.

(6) Se $w \in Z$, então $w = 0^i \# 0^j \# 0^k$, $i,j,k \geq 0$, ou $w = 0^n \# 0^{2n}$, $n \geq 0$. No primeiro caso, $w = \alpha\beta$, $\alpha = 0^i \# 0^j$ e $\beta = \# 0^k$. Então $\alpha, \beta \in Z_1$, de forma que $T \stackrel{*}{\Rightarrow} \alpha$ e $T \stackrel{*}{\Rightarrow} \beta$. Então $TT \stackrel{*}{\Rightarrow} w$ e portanto $S \stackrel{*}{\Rightarrow} w$. No segundo, $w \in Z_2$, logo $X \stackrel{*}{\Rightarrow} w$ e $\therefore S \stackrel{*}{\Rightarrow} w$.

$G_5 = (\{S\}, \{a,b\}, R, S)$ com R dada por

$$S \to aSbS \mid bSaS \mid \varepsilon$$

Se $X = \{ w \in \{a,b\}^* : |w|_a = |w|_b \}$, então $L(G_5) = X$ pois:

① Se $S \overset{*}{\Rightarrow} w$, então $w = a\alpha b\beta$, $w = b\alpha a\beta$ ou $w = \varepsilon$, com $S \overset{*}{\Rightarrow} \alpha$ e $S \overset{*}{\Rightarrow} \beta$. Diretamente, $\varepsilon \in X$.

Assumindo que $\alpha, \beta \in X$, e como $|w|_a = |\alpha|_a + |\beta|_a + 1 = |\alpha|_b + |\beta|_b + 1 = |w|_b$ vale que $w \in X$.

② Se $w \in X$, então ou w começa em a ou em b ou é ε.

Se começa em a, existe em w um b tal que $w = a\alpha b\beta$ sendo que $\alpha, \beta \in X$. (ou como por que dá p/ dividir w assim?)

Assumindo que $S \overset{*}{\Rightarrow} \alpha$ e $S \overset{*}{\Rightarrow} \beta$, então $S \overset{*}{\Rightarrow} a\alpha b\beta = w$.

Similar se w começa em b.