PROPOSIÇÃO: *Seja $G = (V, I, P, S)$ uma gramática regular. Existe uma expressão regular $\alpha \in R_I$ tal que $L(\alpha) = L_G$.*

Prova(esboço): No que se segue, $\Sigma = V \cup I$.

1. Seja $X \in V$ e

$$\begin{array}{l} X \rightarrow a_1 X \\ \dots \\ X \rightarrow a_k X \\ X \rightarrow b_1 Y_1 \\ \dots \\ X \rightarrow b_m Y_m \\ X \rightarrow \beta_1 \\ \dots \\ X \rightarrow \beta_n \end{array}$$

todas as produções em P que têm X como membro esquerdo onde $k, m, n \in I\!N_0$, $a_1, \dots, a_k, b_1, \dots, b_m \in I$, $Y_1, \dots, Y_m \in V$, $\beta_1, \dots, \beta_n \in I \cup \{\epsilon\}$ e existe no máximo um $1 \leq j \leq n$ tal que $\beta_j = \epsilon$.

Tem-se que a linguagem gerada por X em G é

$$L(X) = \{a_1\}L(X) \cup \dots \cup \{a_k\}L(X) \cup$$
$$\{b_1\}L(Y_1) \cup \dots \cup \{b_k\}L(Ym) \cup \{\beta_1\} \cup \dots \cup \{\beta_n\}$$
i.e
$$L(X) = \{a_1, \dots a_k,\}L(X) \cup \{b_1\}L(Y_1) \cup \dots \cup \{b_m\}L(Ym) \cup \{\beta_1, \dots, \beta_n\}$$
i.e.
$$L(X) = AL(X) \cup B$$

onde $A = \{a_1, \dots, a_k\}$ e $B = \{b_1\}L(Y_1) \cup \dots \cup \{b_m\}L(Ym) \cup \{\beta_1, \dots, \beta_n\}$. Obtém-se assim uma equação envolvendo conjuntos (linguagens) que tem como incógnitas as linguagens geradas por cada um dos símbolos não terminais envolvidos, i.e., $L(X)$, $L(Y_1)$, ..., $L(Y_m)$.

Mostra-se que esta equação tem a solução

$$L(X) = A^*B$$

para a incógnita $L(X)$.

Com efeito, fazendo a substituição obtem-se

$$A^*B = AA^*B \cup B$$
i.e.
$$A^*B = (AA^* \cup \{\epsilon\})B$$
i.e.
$$A^*B = A^*B$$

Assim, dado um símbolo $X \in V$ e considerando todas as produções que têm X no membro esquerdo como acima, tem-se que

$$L(X) = A^*B.$$

Note-se que se $B = \emptyset$ então $L(X) = \emptyset$.

2. <u>Regra-R</u>: Como ficou exposto em 1.,

$$L(X) = A^*B$$

é solução (para a incógnita $L(X)$) da equação

$$L(X) = AL(X) \cup B$$

(sendo $L(X)$, A e B como acima). Nesta situação, diz-se que $L(X) = A^*B$ resulta de $L(X) = AL(X) \cup B$ por aplicação da *regra-R*.

3. Dada a gramática G, L_G é, naturalmente, $L(S)$, isto é, o conjunto das sequências de símbolos terminais para as quais existe uma demonstração em G. Para encontrar $L(S)$ pode proceder-se do seguinte modo:

- para cada símbolo X em V consideram-se todas as produções em P cujo membro esquerdo seja X e constrói-se uma equação entre conjuntos como referido em 1

  (esta equação tem como incógnitas a linguagem gerada por X, ($L(X)$) e as linguagens geradas por todos os símbolos não terminais envolvidos em produções que tenham X como membro esquerdo)

- resolve-se o sistema de equações entre conjuntos assim obtido usando a regra-R, e eventuais substituições e aplicações de propriedades[1] das operações sobre conjuntos envolvidas (união, concatenação (ou justaposição) e fecho de Kleene) com o objectivo de chegar a uma solução para $L(S)$ que explicitamente apenas dependa de subconjuntos de I.

Tendo em conta que as únicas operações sobre conjuntos que estão envolvidas são a união, a concatenação e o fecho de Kleene é possível provar que para cada símbolo não terminal X, $L(X)$ é um conjunto admissível, i.e., um conjunto para o qual existe uma expressão regular α tal que $L(X) = L(\alpha)$.

OBSERVAÇÃO: Usualmente e para simplificar a manipulação das equações, utilizam-se expressões regulares para representar ambos os membros das equações entre conjuntos que resultam das produções em P. Na resolução do sistema utilizam-se então propriedades das expressões regulares que reflectem precisamente as propriedades das operações entre conjuntos envolvidas. Em particular, em termos das expressões regulares, a regra-R é

Regra-R: A equação $X = \alpha X + \beta$ tem solução $X = \alpha^*\beta$

onde α e β são tais que $L(\alpha) = A$ e $L(\beta) = B$ sendo A e B como acima.

Assim, assumindo que $V = \{X_1, \ldots, X_r\}$ e que, para cada $1 \leq j \leq r$,

$$X_j \to a_1^j X_j \mid \ldots \mid a_k^j X_j \mid b_1^j X_1^j \ldots \mid b_m^j X_m^j \mid \beta_1^j \mid \ldots \mid \beta_n^j$$

[1] e.g. associatividade da união e concatenação (ou justaposição), comutatividade da união, distributividade da concatenação face à união

são as produções em P que têm X do lado esquerdo (na notação simplificada usual) e $X_1^j, \ldots, X_m^j \in V$, $a_1^j, \ldots, a_k^j, b_1^j, \ldots, b_m^j \in I$ e $\beta_1^j, \ldots, \beta_n^j \in I \cup \{\epsilon\}$ o sistema é usualmente representado da seguinte forma

$$\begin{cases} X_1 = a_1^1 X_1 + \ldots + a_k^1 X_1 + b_1^1 X_1^1 \ldots + b_m^1 X_m^1 + \beta_1^1 + \ldots + \beta_n^1 \\ \ldots \\ X_r = a_1^r X_r + \ldots + a_k^r X_r + b_1^r X_1^r \ldots + b_m^r X_m^r + \beta_1^r + \ldots + \beta_n^r \end{cases}$$

(onde as incógnitas são os símbolos não terminais) ou, na chamada forma canónica,

$$\begin{cases} X_1 = (a_1^1 + \ldots + a_k^1) X_1 + b_1^1 X_1^1 \ldots + b_m^1 X_m^1 + \beta_1^1 + \ldots + \beta_n^1 \\ \ldots \\ X_r = (a_1^r + \ldots + a_k^r) X_r + b_1^r X_1^r \ldots + b_m^r X_m^r + \beta_1^r + \ldots + \beta_n^r \end{cases}$$

Como se disse acima, o sistema é resolvido usando a regra-R e as propriedades das expressões regulares que reflectem precisamente as propriedades das operações entre conjuntos envolvidas.

EXEMPLO: Dada a gramática regular $G = (V, I, P, S)$ tal que

- $V = \{S, A, C\}$
- $I = \{a, b, c\}$
- $P ::\ \begin{array}{l} S \to aS \mid cS \mid bA \mid bC \\ A \to aA \mid \epsilon \\ C \to cC \mid \epsilon \end{array}$

a expressão regular que denota a linguagem gerada por G obtém-se da seguinte forma:

$$\begin{cases} S = aS + cS + bA + bC \\ A = aA + \epsilon \\ C = cC + \epsilon \end{cases}$$

$$\begin{cases} S = aS + cS + bA + bC \\ A = a^*\epsilon \\ C = c^*\epsilon \end{cases}$$

$$\begin{cases} S = (a + c)S + b(A + C) \\ A = a^* \\ C = c^* \end{cases}$$

$$\begin{cases} S = (a + c)^* b(A + C) \\ A = a^* \\ C = c^* \end{cases}$$

$$\begin{cases} S = (a + c)^* b(a^* + c^*) \\ A = a^* \\ C = c^* \end{cases}$$

A expressão regular é a expressão encontrada para o símbolo S, isto é, $(a + c)^*b(a^* + c^*)$.

EXEMPLO: Dada a gramática regular $G = (V, I, P, S)$ tal que

- $V = \{S, U, D\}$
- $I = \{0, 1, 2\}$
  $P = \{(S, 0S), (S, 1U), (S, \epsilon), (S, 2D), (U, 0S), (U, 1U), (U, \epsilon), (D, 0S), (D, \epsilon)\}$
- $P :: \begin{array}{l} S \to 0S \mid 1U \mid 2D \mid \epsilon \\ U \to 0S \mid 1U \mid \epsilon \\ D \to 0S \mid \epsilon \end{array}$

a expressão regular que denota a linguagem gerada por G obtém-se da seguinte forma:

$$\begin{cases} S = 0S + 1U + 2D + \epsilon \\ U = 1U + 0S + \epsilon \\ D = 0S + \epsilon \end{cases}$$

$$\begin{cases} S = 0S + 1U + 2D + \epsilon \\ U = 1^*(0S + \epsilon) \\ D = 0S + \epsilon \end{cases}$$

$$\begin{cases} S = 0S + 1(1^*0S + 1^*) + 2(0S + \epsilon) + \epsilon \\ U = 1^*(0S + \epsilon) \\ D = 0S + \epsilon \end{cases}$$

$$\begin{cases} S = 0S + 11^*0S + 11^* + 20S + 2 + \epsilon \\ U = 1^*(0S + \epsilon) \\ D = 0S + \epsilon \end{cases}$$

$$\begin{cases} S = (0 + 11^*0 + 20)S + 11^* + 2 + \epsilon \\ U = 1^*(0S + \epsilon) \\ D = 0S + \epsilon \end{cases}$$

$$\begin{cases} S = (0 + 11^*0 + 20)^*(11^* + 2 + \epsilon) \\ U = 1^*(0S + \epsilon) \\ D = 0S + \epsilon \end{cases}$$

A expressão regular é a expressão encontrada para o símbolo S, isto é, $(0 + 11^*0 + 20)^*(11^* + 2 + \epsilon)$.