

Linguagens Formais e Autômatos

Formas Normais sobre GLCs

Andrei Rimsa Álvares andrei@cefetmg.br









Sumário

- Formas normais
 - Chomsky
 - Greibach
- GLCs e autômatos de pilha



FORMAS NORMAIS



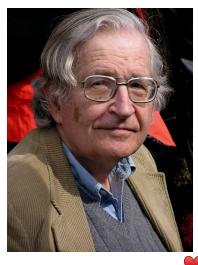
Linguagens Formais e Autômatos





Formas Normais

- Existem duas formas normais especialmente importantes
 - Forma normal de Chomsky
 - Forma normal de Greibach



Noam Chomsky



Sheila Greibach

ttp://en.wikipedia.org/wiki/Noam_Chomsky

ttp://en.wikipedia.org/wiki/Sheila_Greibach





Forma Normal de Chomsky

• Uma GLC $G = (V, \Sigma, R, P)$ está na forma normal de Chomsky (FNC) se cada uma de suas regras tem uma das formas

$$P \rightarrow \lambda$$
 se $\lambda \in L(G)$
 $X \rightarrow YZ$ para $X, Y, Z \in V$
 $X \rightarrow a$ para $a \in \Sigma$





Transformação em FNC

Seja uma GLC G, obter uma GLC equivalente à G tal que

$$P \rightarrow \lambda$$
 se $\lambda \in L(G)$
 $X \rightarrow a$ para $a \in \Sigma$
 $X \rightarrow w$ para $|w| \ge 2$

Basta executar a sequência de eliminações na ordem que garanta a consistência

- Adicionar mais dois passos
 - 1) Modificar cada regra $X \to w$, $|w| \ge 2$ de forma que ela contenha apenas variáveis
 - 2) Substituir cada regra $X \to Y_1 Y_2 ... Y_n$, $n \ge 3$, em que cada Y_i é uma variável, pelo conjunto de regras: $X \to Y_1 Z_1$, $Z_1 \to Y_2 Z_2$, ..., $Z_{n-2} \to Y_{n-1} Y_n$, em que Z_1 , Z_2 , ..., Z_{n-2} são novas variáveis





• Seja a GLC G = ({L, S, E}, {a, (,)}, R, L), em que R consta das regras

$$L \rightarrow (S)$$

$$S \rightarrow SE \mid \lambda$$

$$E \rightarrow a \mid L$$

1) Após eliminação de regras λ

2) Após eliminação de regras unitárias

3) Não há eliminação de variáveis inúteis





• Seja a GLC G = ({L, S, E}, {a, (,)}, R, L), em que R consta das regras

$$L \rightarrow (S)$$

$$S \rightarrow SE \mid \lambda$$

$$E \rightarrow a \mid L$$

4) Regras da forma $X \rightarrow w$, $|w| \ge 2$ somente com variáveis

$$L \rightarrow (S) | ()$$

$$S \rightarrow SE | a | (S) | ()$$

$$E \rightarrow a | (S) | ()$$

$$A \rightarrow ($$

$$B \rightarrow)$$





• Seja a GLC G = ({L, S, E}, {a, (,)}, R, L), em que R consta das regras:

$$L \rightarrow (S)$$

$$S \rightarrow SE \mid \lambda$$

$$E \rightarrow a \mid L$$

5) Substituir regras da forma $X \rightarrow Y_1 Y_2 ... Y_n$ ($n \ge 3$)

$$L \rightarrow ASB \mid AB$$

$$S \rightarrow SE \mid a \mid ASB \mid AB$$

$$E \rightarrow a \mid ASB \mid AB$$

$$A \rightarrow ($$

$$B \rightarrow)$$

$$L \rightarrow AX \mid AB$$

$$S \rightarrow SE \mid a \mid AX \mid AB$$

$$E \rightarrow a \mid AX \mid AB$$

$$A \rightarrow ($$

$$B \rightarrow)$$

$$X \rightarrow SB$$





Mais Manipulação de GLCs

- Antes de apresentar a forma normal de Greibach, serão vistos dois métodos de manipulação que são úteis não apenas como passos intermediários para obtenção de GLC's nesta forma normal, mas também em outros contextos
 - 1) Eliminação de regras recursivas à esquerda
 - 2) Eliminação de variável em regra





Eliminação de Regras Recursivas à Esquerda

- Teorema: para qualquer GLC, existe uma GLC equivalente sem regras recursivas à esquerda
 - Sejam, a seguir, todas as regras X de uma GLC G, onde nenhum $w_{\rm i}$ começa com X

$$X \to Xy_1 | Xy_2 | ... | Xy_n | w_1 | w_2 | ... | w_k$$

— Substituir recursão à esquerda por recursão à direita, em que Z é uma nova variável

$$X \rightarrow w_1 Z \mid w_2 Z \mid ... \mid w_k Z$$

$$Z \rightarrow y_1 Z \mid y_2 Z \mid ... \mid y_n Z \mid \lambda$$

– Eliminando regras λ





 Seja a GLC G = ({E}, {t, +, *, (,)}, R, E), com regras recursivas à esquerda, onde R é dado por

$$E \rightarrow E + E \mid E * E \mid (E) \mid t$$

Remove-se a recursividade à esquerda

$$E \rightarrow (E) \mid t \mid (E)Z \mid tZ$$

 $Z \rightarrow +E \mid *E \mid +EZ \mid *EZ$





Eliminação de Variável em Regra

• **Teorema:** Seja uma GLC $G = (V, \Sigma, R, P)$ tal que $X \rightarrow uYv \in R$, onde $Y \in V$ e $Y \neq X$. Sejam $Y \rightarrow w_1 \mid w_2 \mid ... \mid w_n$ **todas** as regras $Y \in R$. Seja $G' = (V, \Sigma, R', P)$ onde

$$R' = (R - \{X \rightarrow uYv\}) \cup \{X \rightarrow uw_1v \mid uw_2v \mid ... \mid uw_nv\}$$

$$L(G') = L(G)$$





Forma Normal de Greibach

• Uma GLC $G = (V, \Sigma, R, P)$ está na forma normal de Greibach (FNG) se cada uma de suas regras tem uma das formas

$$P \to \lambda$$
 se $\lambda \in L(G)$
 $X \to ay$ para $a \in \Sigma$ e $y \in V^*$

− Uma forma sentencial de uma gramática na FNG, com exceção de λ, é da forma xy, em que $x ∈ Σ^+$ e $y ∈ V^*$

Qual o tamanho de uma derivação de uma palavra?





Transformação em FNG

Seja uma GLC G, obter uma GLC equivalente à G tal que

$$P \rightarrow \lambda$$
 se $\lambda \in L(G)$
 $X \rightarrow a$ para $a \in \Sigma$
 $X \rightarrow w$ para $|w| \ge 2$

Basta executar a sequência de eliminações na ordem que garanta a consistência

– Para cada X \to w para $|w| \ge 2$, substituir por variáveis todos os terminais de w a partir do segundo símbolo. Com isso, obtém-se regras da forma $X \to Yy$, sendo que $Y \in (V \cup \Sigma)$ e $y \in V^+$





Transformação em FNG

- Algoritmo
 - 1) Numerar sequencialmente as variáveis (#P=1)
 - Para cada A ∈ V, começando com P, na ordem dada pela numeração, fazer enquanto possível
 - 2.1) Se existir $A \rightarrow By$, para $|y| \ge 1$, tal que #A > #B, substituir B
 - 2.2) Se existir $A \rightarrow Ay$, para $|y| \ge 1$, eliminar recursão à esquerda
 - 3) Seja B a variável de maior número em que há regra da forma $B \to Cy$, com $C \in V$, substituir C. Repetir até atingir a variável P
 - 4) Para regras da forma $Z \to Ay$, em que Z é variável nova (adicionada na eliminação de recursão à esquerda), substituir a variável A





• Seja a GLC G = ({A, B, C, D}, {b, c, d}, R, A), onde R é dado por

$$A \rightarrow CB$$

 $B \rightarrow BBD \mid b$
 $C \rightarrow BBC \mid Dc$
 $D \rightarrow AD \mid d$

Observação: essa gramática não possui regras λ e unitárias

• Elimina-se a regra C \rightarrow Dc, introduzindo as regras C \rightarrow DE e E \rightarrow c

$$A \rightarrow CB$$

 $B \rightarrow BBD \mid b$
 $C \rightarrow BBC \mid DE$
 $D \rightarrow AD \mid d$
 $E \rightarrow c$

Numeração: #A = 1, #B = 2, #C = 3, #D = 4, #E = 5 (passo 1)





Removendo recursão à esquerda para as regras B (passo 2.2)

#1A
$$\rightarrow$$
 CB
#2B \rightarrow BBD | b
#3C \rightarrow BBC | DE
#4D \rightarrow AD | d
#5E \rightarrow c

A \rightarrow CB
* B \rightarrow b | bZ₁
C \rightarrow BBC | DE
D \rightarrow AD | d
E \rightarrow c
* Z₁ \rightarrow BD | BDZ₁

O asterisco (*) marca as novas regras





• Substitui-se a regra $C \rightarrow BBC$ (passo 2.1: #C > #B)

#1A
$$\rightarrow$$
 CB
#2B \rightarrow b | bZ₁
#3C \rightarrow BBC | DE
#4D \rightarrow AD | d
#5E \rightarrow c
 $Z_1 \rightarrow$ BD | BDZ₁

A \rightarrow CB
B \rightarrow b | bZ₁
* C \rightarrow bBC | bZ₁BC | DE
D \rightarrow AD | d
E \rightarrow c
 $Z_1 \rightarrow$ BD | BDZ₁





• Substitui-se a regra $D \rightarrow AD$ (passo 2.1: #D > #A)

$$\begin{array}{l}
^{\text{#1}}A \rightarrow CB \\
^{\text{#2}}B \rightarrow b \mid bZ_{1} \\
^{\text{#3}}C \rightarrow bBC \mid bZ_{1}BC \mid DE \\
^{\text{#4}}D \rightarrow AD \mid d
\end{array}$$

$$\begin{array}{l}
C \rightarrow bBC \mid bZ_{1}BC \mid DE \\
* D \rightarrow CBD \mid d
\end{aligned}$$

$$\begin{array}{l}
E \rightarrow c \\
Z_{1} \rightarrow BD \mid BDZ_{1}$$

$$\begin{array}{l}
E \rightarrow c \\
Z_{1} \rightarrow BD \mid BDZ_{1}$$





• Substitui-se a regra $D \rightarrow CBD$ (passo 2.1: #D > #C)

#1A
$$\rightarrow$$
 CB
#2B \rightarrow b | bZ₁
#3C \rightarrow bBC | bZ₁BC | DE
#4D \rightarrow CBD | d
#5E \rightarrow c
 $Z_1 \rightarrow$ BD | BDZ₁

A \rightarrow CB
B \rightarrow b | bZ₁
C \rightarrow bBC | bZ₁BC | DE
* D \rightarrow bBCBD | bZ₁BCBD | DEBD | d
E \rightarrow c
Z₁ \rightarrow BD | BDZ₁





Removendo recursão à esquerda para as regras D (passo 2.2)

```
^{\#1}A \rightarrow CB
^{\text{#2}}B \rightarrow b \mid bZ_1
^{\sharp 3}C \rightarrow bBC \mid bZ_1BC \mid DE
^{#4}D \rightarrow bBCBD \mid bZ_1BCBD \mid DEBD \mid d
^{#5}E \rightarrow c
                                                       A \rightarrow CB
   Z_1 \rightarrow BD \mid BDZ_1
                                                       B \rightarrow b \mid bZ_1
                                                        C \rightarrow bBC \mid bZ_1BC \mid DE
                                                   * D \rightarrow bBCBD | bZ<sub>1</sub>BCBD | d | bBCBDZ<sub>2</sub> |
                                                                 bZ_1BCBDZ_2 \mid dZ_2
                                                        E \rightarrow c
                                                       Z_1 \rightarrow BD \mid BDZ_1
                                                   * Z_2 \rightarrow EBD \mid EBDZ_2
```





• Substitui-se a regra $C \rightarrow DE$ (passo 3)

```
^{\#1}A \rightarrow CB
^{\text{#2}}B \rightarrow b \mid bZ_1
^{\#3}C \rightarrow bBC \mid bZ_1BC \mid DE
^{44}D \rightarrow bBCBD \mid bZ_1BCBD \mid d \mid bBCBDZ_2 \mid
            bZ_1BCBDZ_2 \mid dZ_2
                                                            A \rightarrow CB
^{#5}E \rightarrow c
                                                            B \rightarrow b \mid bZ_1
   Z_1 \rightarrow BD \mid BDZ_1
                                                            C \rightarrow bBC \mid bZ_1BC
   Z_2 \rightarrow EBD \mid EBDZ_2
                                                        * C \rightarrow bBCBDE \mid bZ_1BCBDE \mid dE \mid
                                                                      bBCBDZ<sub>2</sub>E | bZ<sub>1</sub>BCBDZ<sub>2</sub>E | dZ<sub>2</sub>E
                                                            D \rightarrow bBCBD \mid bZ_1BCBD \mid d \mid bBCBDZ_2 \mid
                                                                     bZ_1BCBDZ_2 \mid dZ_2
                                                            E \rightarrow c
                                                            Z_1 \rightarrow BD \mid BDZ_1
                                                            Z_2 \rightarrow EBD \mid EBDZ_2
```



 $^{\#1}A \rightarrow CB$



Exemplo

• Substitui-se a regra $A \rightarrow CB$ (passo 3)

```
^{\#2}B \rightarrow b \mid bZ_1
^{\#3}C \rightarrow bBC \mid bZ_1BC \mid bBCBDE \mid bZ_1BCBDE \mid
            dE | bBCBDZ<sub>2</sub>E | bZ<sub>1</sub>BCBDZ<sub>2</sub>E | dZ<sub>2</sub>E
^{\#4}D \rightarrow bBCBD \mid bZ_1BCBD \mid d \mid bBCBDZ_2 \mid
            bZ_1BCBDZ_2 \mid dZ_2
                                                               * A \rightarrow bBCB | bZ<sub>1</sub>BCB | bBCBDEB |
^{\#5}E \rightarrow c
                                                                           bZ<sub>1</sub>BCBDEB| dEB | bBCBDZ<sub>2</sub> EB |
  Z_1 \rightarrow BD \mid BDZ_1
                                                                           bZ_1BCBDZ_2EB \mid dZ_2EB
  Z_2 \rightarrow EBD \mid EBDZ_2
                                                                  B \rightarrow b \mid bZ_1
                                                                  C \rightarrow bBC \mid bZ_1BC \mid bBCBDE \mid bZ_1BCBDE \mid
                                                                           dE | bBCBDZ<sub>2</sub>E | bZ<sub>1</sub>BCBDZ<sub>2</sub>E | dZ<sub>2</sub>E
                                                                  D \rightarrow bBCBD \mid bZ_1BCBD \mid d \mid bBCBDZ_2 \mid
                                                                           bZ_1BCBDZ_2 \mid dZ_2
                                                                  E \rightarrow c
                                                                  Z_1 \rightarrow BD \mid BDZ_1
                                                                  Z_2 \rightarrow EBD \mid EBDZ_2
```





 Finalmente, substitui-se as regras introduzidas pela eliminação de recursão à esquerda (passo 4)

```
#1A \rightarrow bBCB | bZ<sub>1</sub>BCB | bBCBDEB |

bZ<sub>1</sub>BCBDEB | dEB | bBCBDZ<sub>2</sub> EB |

bZ<sub>1</sub>BCBDZ<sub>2</sub>EB | dZ<sub>2</sub>EB

#2B \rightarrow b | bZ<sub>1</sub>

#3C \rightarrow bBC | bZ<sub>1</sub>BC | bBCBDE | bZ<sub>1</sub>BCBDE |

dE | bBCBDZ<sub>2</sub>E | bZ<sub>1</sub>BCBDZ<sub>2</sub>E | dZ<sub>2</sub>E

#4D \rightarrow bBCBD | bZ<sub>1</sub>BCBD | d | bBCBDZ<sub>2</sub> |

bZ<sub>1</sub>BCBDZ<sub>2</sub> | dZ<sub>2</sub>

#5E \rightarrow c

Z<sub>1</sub> \rightarrow BD | BDZ<sub>1</sub>
```

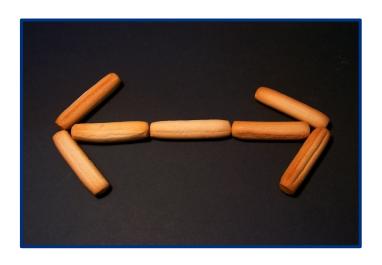
 $Z_2 \rightarrow EBD \mid EBDZ_2$

```
A \rightarrow bBCB \mid bZ_1BCB \mid bBCBDEB \mid
bZ_1BCBDEB \mid dEB \mid bBCBDZ_2EB \mid
bZ_1BCBDZ_2EB \mid dZ_2EB
B \rightarrow b \mid bZ_1
C \rightarrow bBC \mid bZ_1BC \mid bBCBDE \mid bZ_1BCBDE \mid
dE \mid bBCBDZ_2E \mid bZ_1BCBDZ_2E \mid dZ_2E
D \rightarrow bBCBD \mid bZ_1BCBD \mid d \mid bBCBDZ_2 \mid
bZ_1BCBDZ_2 \mid dZ_2
E \rightarrow c
* Z_1 \rightarrow bD \mid bZ_1D \mid bDZ_1 \mid bZ_1DZ_1
```

* $Z_2 \rightarrow cBD \mid cBDZ_2$



GLCS E AUTÔMATOS DE PILHA



Linguagens Formais e Autômatos





Equivalência entre GLCs e APs

- Autômatos de pilha (APs) e gramáticas livres de contexto (GLCs) reconhecem a mesma classe de linguagem: as linguagens livres de contexto
- Será mostrado que
 - 1) Para qualquer GLC G existe um AP que reconhece L(G)
 - 2) Para qualquer APN M existe uma GLC que gera L(M)





$GLC \rightarrow AP$

• **Teorema:** Para qualquer GLC G existe um AP que reconhece L(G)

Seja $G'=(V,\Sigma,R,P)$ uma GLC na FNG equivalente à G. Um APN que aceita L(G') é $M=(\{i,f\},\Sigma,V,\delta,\{i\},\{f\})$, onde δ consta das transições

- $\delta(i, \lambda, \lambda) = \{[f, P]\}$
- $\delta(f, \lambda, P) = \{[f, \lambda]\}, \text{ se } P \rightarrow \lambda \in R$
- $\delta(f, a, X) = \{[f, y] \mid y \in V^* \in X \rightarrow ay \in R\},\$ $\forall a \in \Sigma \in \forall X \in V$





• Seja a GLC $G = (\{P\}, \{0, 1\}, R, P)$, onde R consta das regras

$$P \rightarrow 0P1P \mid 1P0P \mid \lambda$$

• A gramática G' equivalente à G na FNG consta das regras

$$P \rightarrow 0PUP \mid 1PZP \mid \lambda$$
 $Z \rightarrow 0$
 $U \rightarrow 1$

• Pode-se construir um autômato de pilha que reconhece L(G)





 Antes de mostrar que é sempre possível obter uma GLC que gera a linguagem reconhecida por um AP, será mostrada a ideia central por trás do processo

Seja um, APN $M=(E,\Sigma,\Gamma,\delta,I,F)$. Dados $e,e'\in E$ e $A\in\Gamma\cup\{\lambda\}$, considere o conjunto C(e,A,e') de todas as palavras $w\in\Sigma^*$ tais que o APN M, começando em e, com a pilha contendo A, termina no estado e' com a pilha vazia, após consumir w

$$C(e, A, e') = \{w \in \Sigma^* \mid [e, w, A] \stackrel{*}{\vdash} [e', \lambda, \lambda] \}$$

A linguagem reconhecida é dada por

$$L(M) = \bigcup_{(i,f)\in I\times F} C(i,\lambda,f)$$





- Suponha que seja possível gerar o conjunto C(e,A,e') por meio de uma GLC cuja variável de partida seja [e,A,e']
- Então L(M) pode ser gerada por uma GLC constituída de
 - Todas as regras [e, A, e']
 - $P \rightarrow [i, \lambda, f]$, para $\forall i \in I \in \forall f \in F$

Como produzir as regras?





• Teorema: Para qualquer APN M existe uma GLC que gera L(M)

Seja o APN $M = (E, \Sigma, \Gamma, \delta, I, F)$. Será construída uma GLC com variáveis da forma [e, A, e'], para $e, e' \in E$ e $A \in \Gamma \cup \{\lambda\}$, de modo que para todo $w \in \Sigma^*$

$$[e, A, e'] \stackrel{*}{\Rightarrow} w$$
 se e somente se $[e, A, e'] \vdash^* [e', \lambda, \lambda]$





- A gramática G tal que L(G) = L(M) é (V, Σ, R, P) , onde $V = (\{P\} \cup E) \times (\Gamma \cup \{\lambda\}) \times E$ e R contém as regras
 - $\mathbf{P} \rightarrow [i, \lambda, f]$, para cada $i \in I$ e $f \in F$
 - $[e, \lambda, e] \rightarrow \lambda$, para cada $e \in E$

Para cada transição $[e', z] \in \delta(e, a, A)$, onde $a \in \Sigma \cup \{\lambda\}$, $z = B_1B_2...B_n$ ($B_i \in \Gamma$) e $A \in \Gamma \cup \{\lambda\}$, as seguintes regras (tipo 1)

- Se $z = \lambda$: $[e, A, d] \rightarrow a[e', \lambda, d]$, para cada $d \in E$
- Se z $\neq \lambda$: $[e, A, d_n] \rightarrow a[e', B_1, d_1]...[d_{n-1}, B_n, d_n]$, para cada $d_1, d_2, ..., d_n \in E$

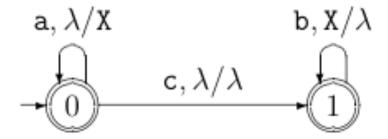
Ainda, se $A = \lambda$, tem-se adicionalmente as regras (tipo 2)

- Se $z = \lambda$: [e, C, d] → a[e', C, d], para cada C ∈ Γ e cada d ∈ E
- Se $z \neq \lambda$: [e, C, d_{n+1}] → a[e', B₁, d_1]...[d_{n-1} , B_n, d_n][d_n , C, d_{n+1}], para C ∈ Γ e cada d_1 , d_2 , ..., d_n , d_{n+1} ∈ E





• Seja o APD que reconhece L = $\{a^ncb^n \mid n \ge 0\} \cup \{\lambda\}$



- Para a gramática, obtém-se as seguintes regras inicialmente
 - $P \rightarrow [0, \lambda, 0] | [0, \lambda, 1]$
 - $[0, \lambda, 0] \rightarrow \lambda$
 - $[1, \lambda, 1] \rightarrow \lambda$

Como são as outras regras?



Tipo 1



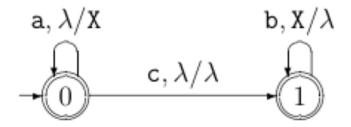
Exemplo

- Para as transições
 - $[0, X] \in \delta(0, a, \lambda)$
 - $\bullet [0, \lambda, 0] \rightarrow a[0, X, 0]$
 - $[0, \lambda, 1] \rightarrow a[0,X,1]$
 - $[0, X, 0] \rightarrow a[0, X, 0][0, X, 0] | a[0, X, 1][1, X, 0]$
 - $[0, X, 1] \rightarrow a[0, X, 0][0, X, 1] | a[0, X, 1][1, X, 1]$
 - $[0, X] \in \delta(0, a, \lambda)$
 - $\underline{\bullet}_{--} [0, \lambda, 0] \rightarrow \mathbf{c}[1, \lambda, 0]$
 - $[0, \lambda, 1] \rightarrow c[1, \lambda, 1]$
 - $[0, X, 0] \rightarrow c[1, X, 0]$
 - $[0, X, 1] \rightarrow c[1, X, 1]$
 - [1, λ] ∈ δ(1, b, X)
 - $[1, \mathbf{X}, \mathbf{0}] \to \mathbf{b}[1, \lambda, \mathbf{0}]$
 - $[1, X, 1] \rightarrow b[1, \lambda, 1]$

Tipo 2

Como variáveis

são inúteis?



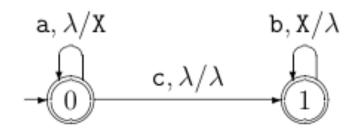




Eliminando as variáveis inúteis obtém-se as seguintes regras

$$P \rightarrow [0, \lambda, 0] \mid [0, \lambda, 1]$$

 $[0, \lambda, 1] \rightarrow a[0, X, 1] \mid c[1, \lambda, 1]$
 $[0, X, 1] \rightarrow a[0, X, 1][1, X, 1] \mid c[1, X, 1]$
 $[1, X, 1] \rightarrow b[1, \lambda, 1]$
 $[0, \lambda, 0] \rightarrow \lambda$
 $[1, \lambda, 1] \rightarrow \lambda$





ISSO É TUDO, PESSOAL!



Linguagens Formais e Autômatos