

Teoria da Computação

Algoritmo CYK

Thiago Alves

Introdução

- ◆ Algoritmo Cocke-Younger-Kasami
- ◆ Algoritmo para decidir se uma string w está em uma linguagem livre de contexto L
- ◆ Linguagem livre de contexto representada por uma GLC G
- ◆ Decidir se G gera w

Testando Pertinência

- ◆ Saber se string w está em $L(G)$
- ◆ Seja $w = a_1 \dots a_n$.
- ◆ Construimos uma matriz triangular $n \times n$ de conjuntos variáveis
- ◆ $X_{ij} = \{A \mid A \Rightarrow^* a_i \dots a_j\}$
- ◆ No final, verificar se S está em X_{1n}
- ◆ $S \Rightarrow^* a_1 \dots a_n$

Testando Pertinência

◆ Se $n = 5$:

X_{15}				
X_{14}	X_{25}			
X_{13}	X_{24}	X_{35}		
X_{12}	X_{23}	X_{34}	X_{45}	
X_{11}	X_{22}	X_{33}	X_{44}	X_{55}
a_1	a_2	a_3	a_4	a_5

Testando Pertinência

- ◆ Vamos preencher a tabela para cada substring de w
- ◆ Preencher as entradas para as substrings de tamanho 1, depois para as de tamanho 2, e assim por diante

Testando Pertinência

- ◆ Preencher as entradas para as substrings de tamanho 1, depois para as de tamanho 2, e assim por diante
- ◆ Podemos usar as entradas de substrings de tamanho menor para determinar as de tamanho maior

Testando Pertinência

- ◆ Para as substrings de tamanho 1 usar as regras da forma $A \rightarrow a$
- ◆ Seja $w = a_1a_2\dots a_n$
- ◆ Colocar na entrada X_{ij} as variáveis que geram a string a_i

Testando Pertinência

- ◆ Suponha que temos todas as entradas para tamanho até 3
- ◆ Para verificar se A gera substring de tamanho 4 podemos quebrá-la de 3 formas diferentes
 - a_1 e $a_2a_3a_4$
 - a_1a_2 e a_3a_4
 - $a_1a_2a_3$ e a_4

Testando Pertinência

- ◆ Para cada quebra, examinar cada regra $A \rightarrow BC$ e verificar se B gera a primeira parte e C a segunda parte
 - Usando as entradas anteriores
- ◆ Se for o caso, A é adicionada a entrada dessa substring

Exemplo

◆ $S \rightarrow AB \mid BC$

◆ $A \rightarrow BA \mid a$

◆ $B \rightarrow CC \mid b$

◆ $C \rightarrow AB \mid a$

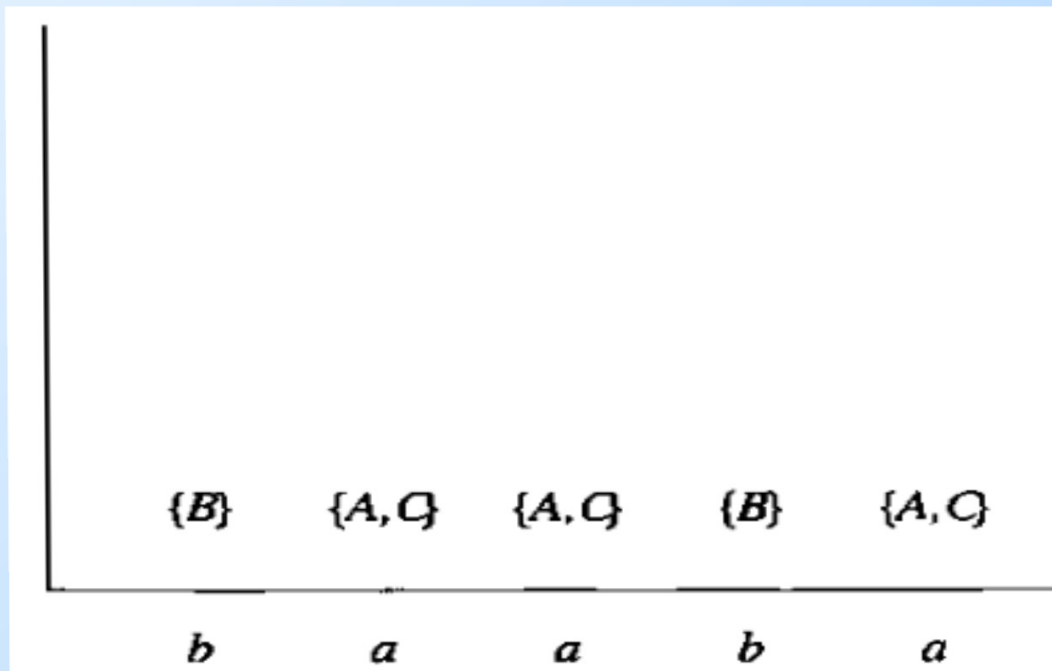
◆ $w = \text{baaba}$

Exemplo

- ◆ Construimos a primeira linha verificando regras com terminais
- ◆ Para substring a: $\{A,C\}$
- ◆ Para substring b: $\{B\}$
- ◆ $X_{11} = \{B\}$, $X_{22} = \{A,C\}$
- ◆ $X_{33} = \{A,C\}$, $X_{44} = \{B\}$
- ◆ $X_{55} = \{A,C\}$

Exemplo

- ◆ Construimos a primeira linha verificando regras com terminais

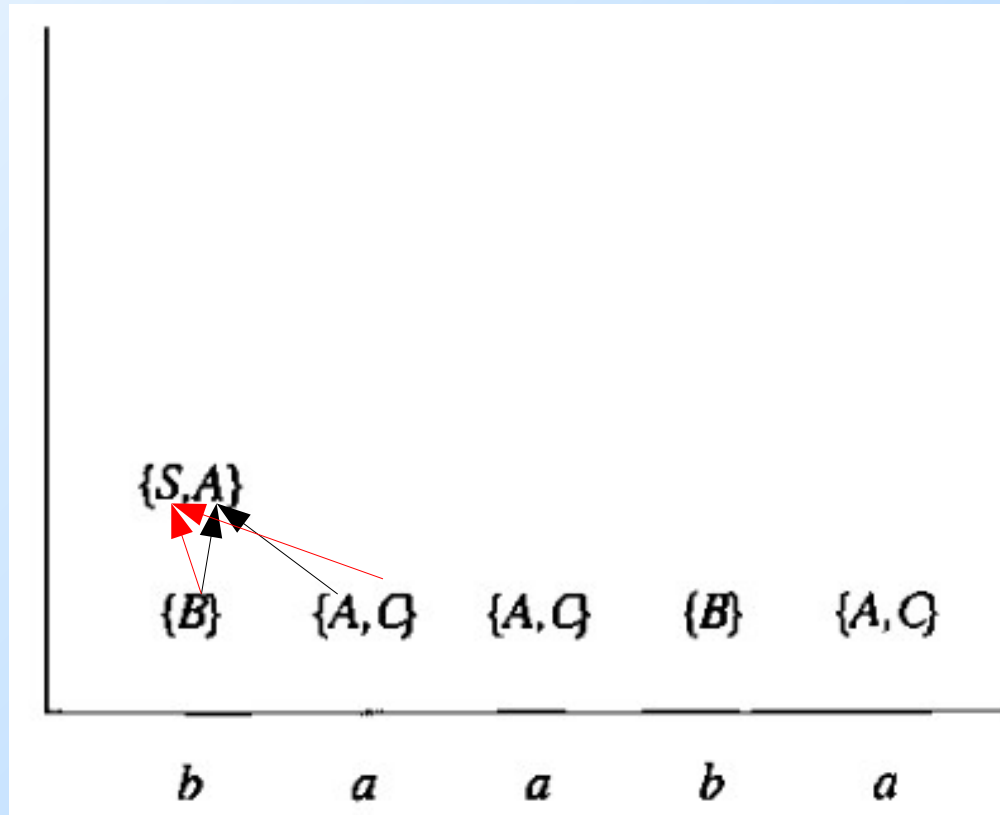


Exemplo

- ◆ Para X_{12} :
- ◆ Substring ba
- ◆ Quebras:
 - b e a
- ◆ Variável deve ter regra com BA ou BC no corpo:
 - $X_{12} = \{A, S\}$

Exemplo

- ◆ Variável deve ter regra com BA ou BC

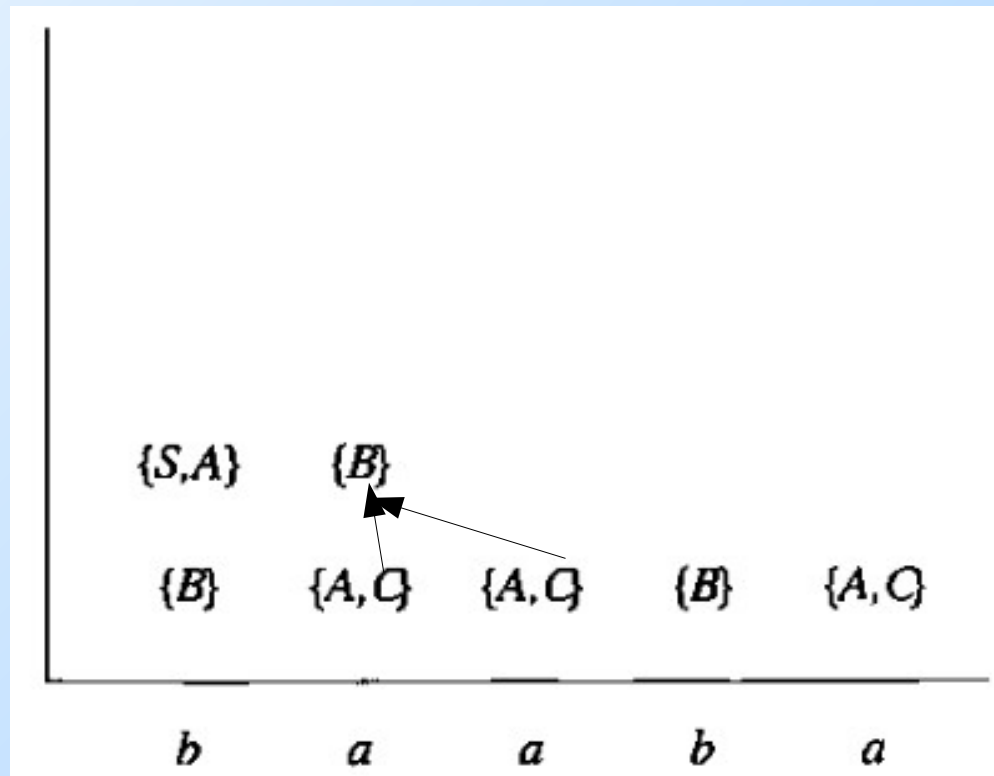


Exemplo

- ◆ Para X_{23} :
- ◆ Substring aa
- ◆ Quebras:
 - a e a
- ◆ Variável deve ter regra com AA ou AC ou CA ou CC no corpo:
 - $X_{23} = \{B\}$

Exemplo

- ◆ Variável deve ter regra com AA ou AC ou CA ou CC no corpo:

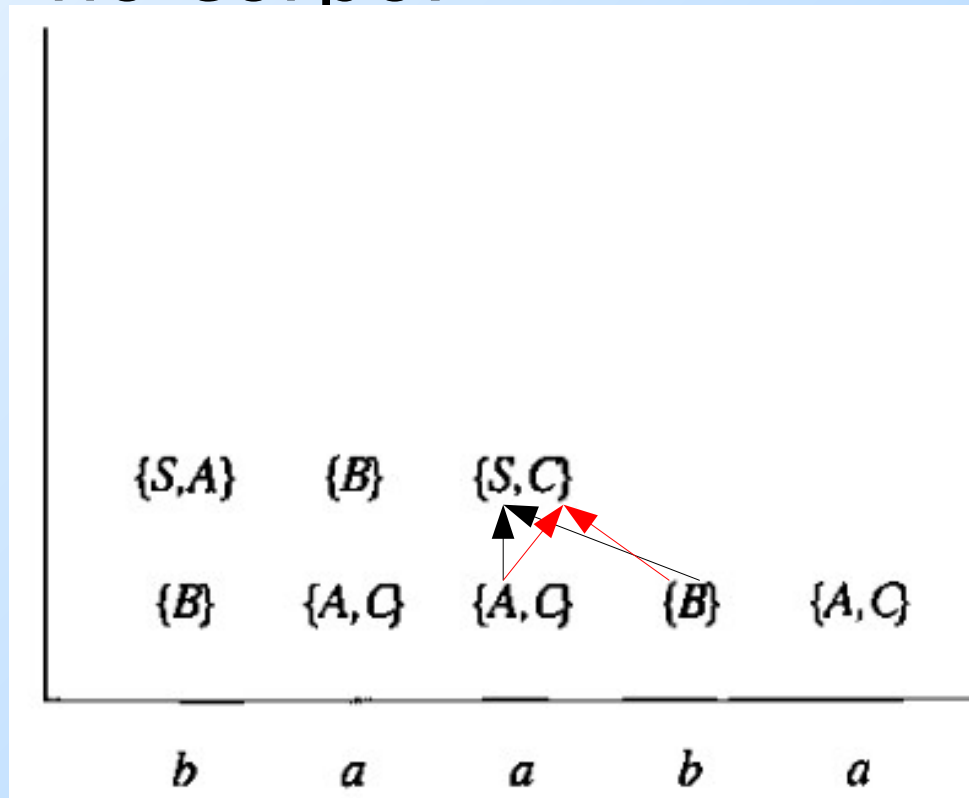


Exemplo

- ◆ Para X_{34} :
- ◆ Substring ab
- ◆ Quebras:
 - a e b
- ◆ Variável deve ter regra com AB ou CB no corpo:
 - $X_{34} = \{S, C\}$

Exemplo

- ◆ Variável deve ter regra com AB ou CB no corpo:

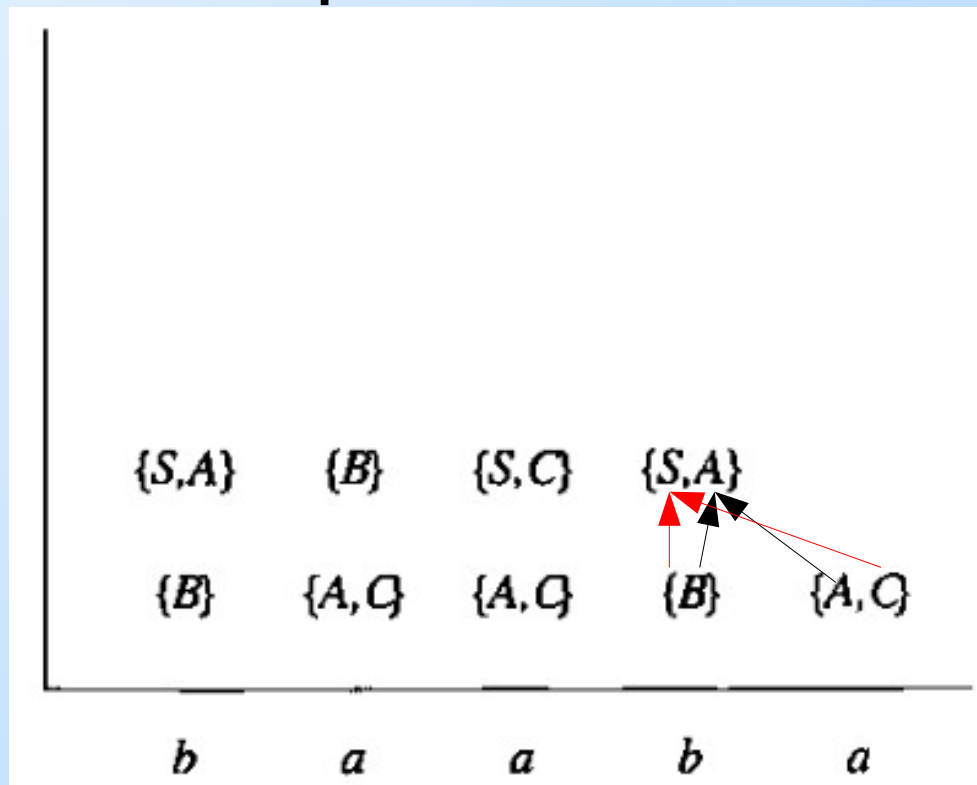


Exemplo

- ◆ Para X_{45} :
- ◆ Substring ba
- ◆ Quebras:
 - b e a
- ◆ Variável deve ter regra com BA ou BC no corpo:
 - $X_{45} = \{S, A\}$

Exemplo

- ◆ Variável deve ter regra com BA ou BC no corpo:

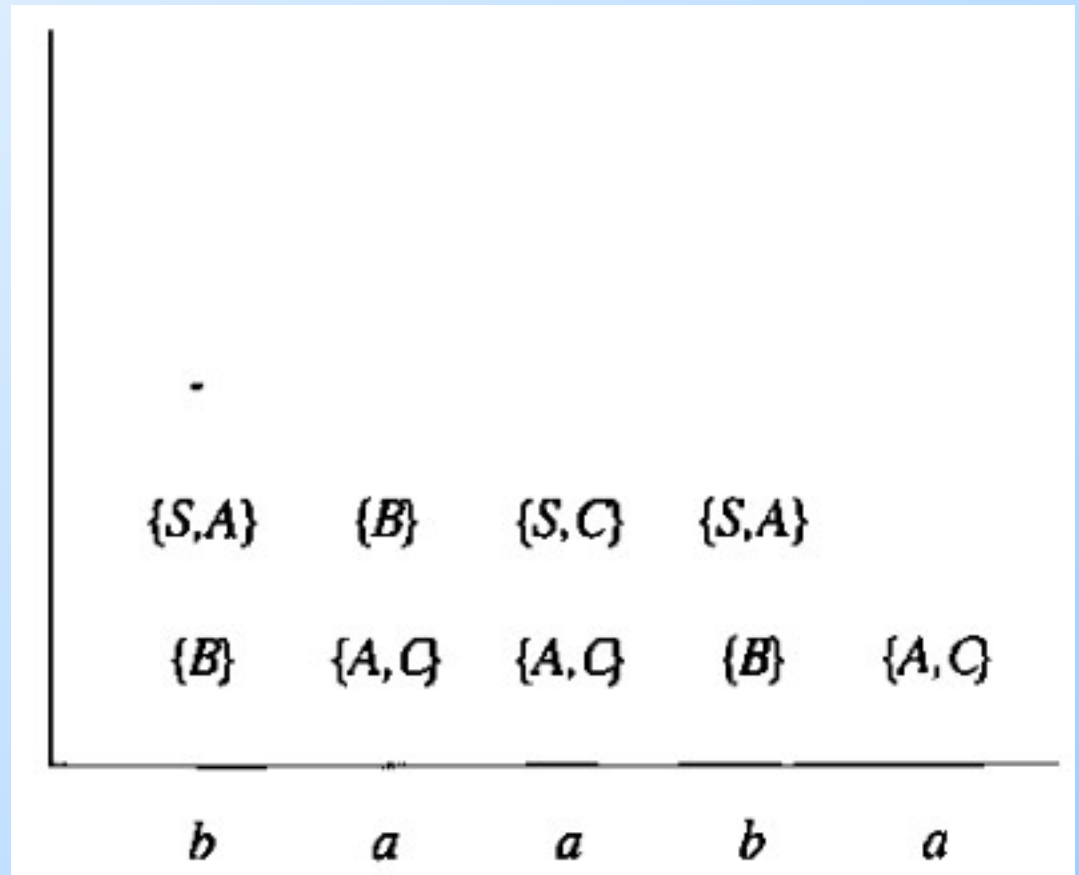


Exemplo

- ◆ Para X_{13} :
- ◆ Substring baa
- ◆ Quebras:
 - b e aa
 - ba e a
- ◆ Variável deve ter regra com BB ou SA ou SC ou AA ou AC no corpo:
 - $X_{13} = \{\}$

Exemplo

$$\blacklozenge X_{13} = \{\}$$

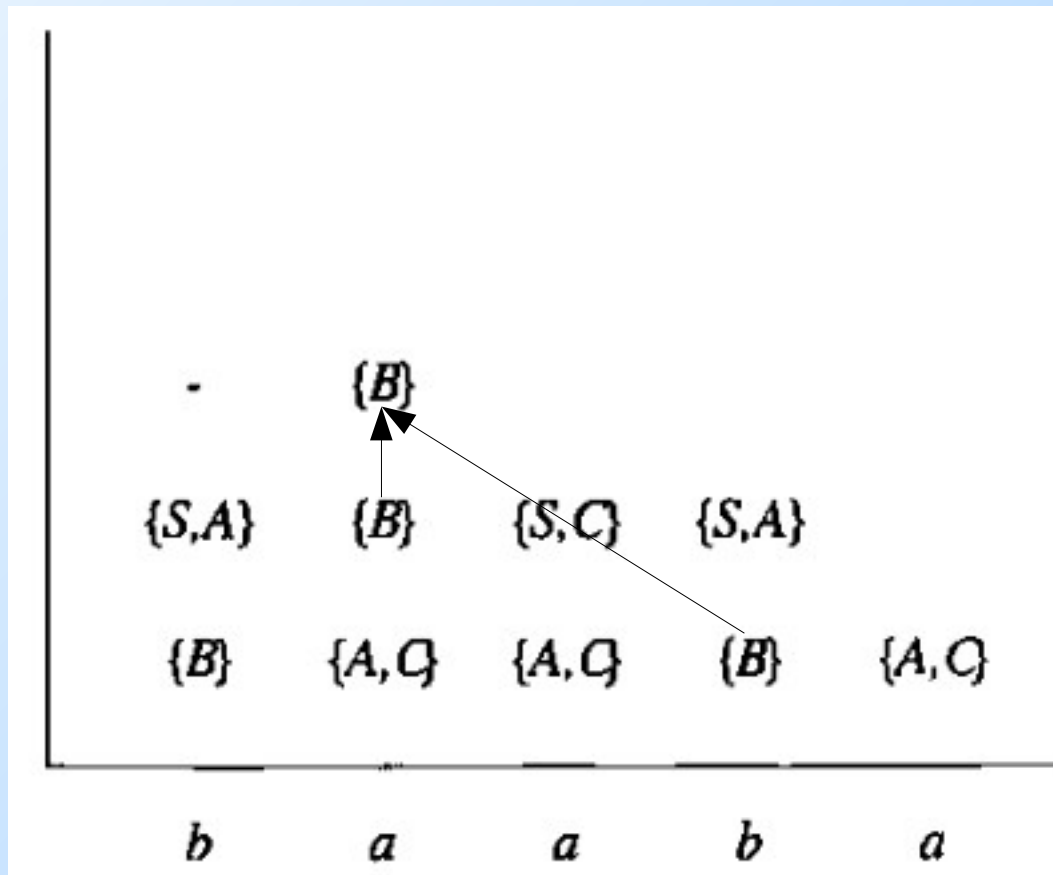


Exemplo

- ◆ Para X_{24} :
- ◆ Substring aab
- ◆ Quebras:
 - a e ab
 - aa e b
- ◆ Variável deve ter regra com AS ou AC ou CS ou CC ou BB no corpo:
 - $X_{24} = \{B\}$

Exemplo

◆ $X_{24} = \{B\}$

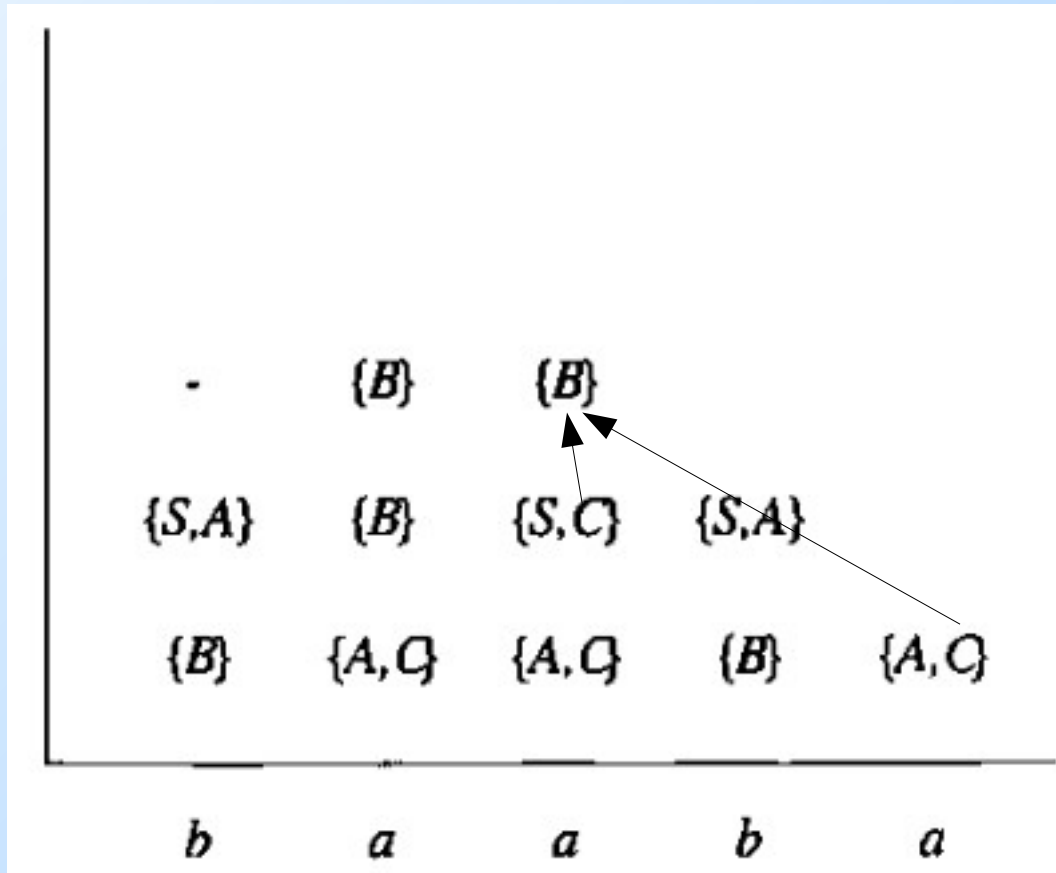


Exemplo

- ◆ Para X_{35} :
- ◆ Substring aba
- ◆ Quebras:
 - a e ba
 - ab e a
- ◆ Variável deve ter algum de $\{AS, AA, CS, CA, SA, SC, CC\}$ no corpo:
 - $X_{35} = \{B\}$

Exemplo

$$\blacklozenge X_{35} = \{B\}$$



Exemplo

- ◆ Para X_{14} :
- ◆ Substring baab
- ◆ Quebras:
 - b e aab
 - ba e ab
 - baa e b

Exemplo

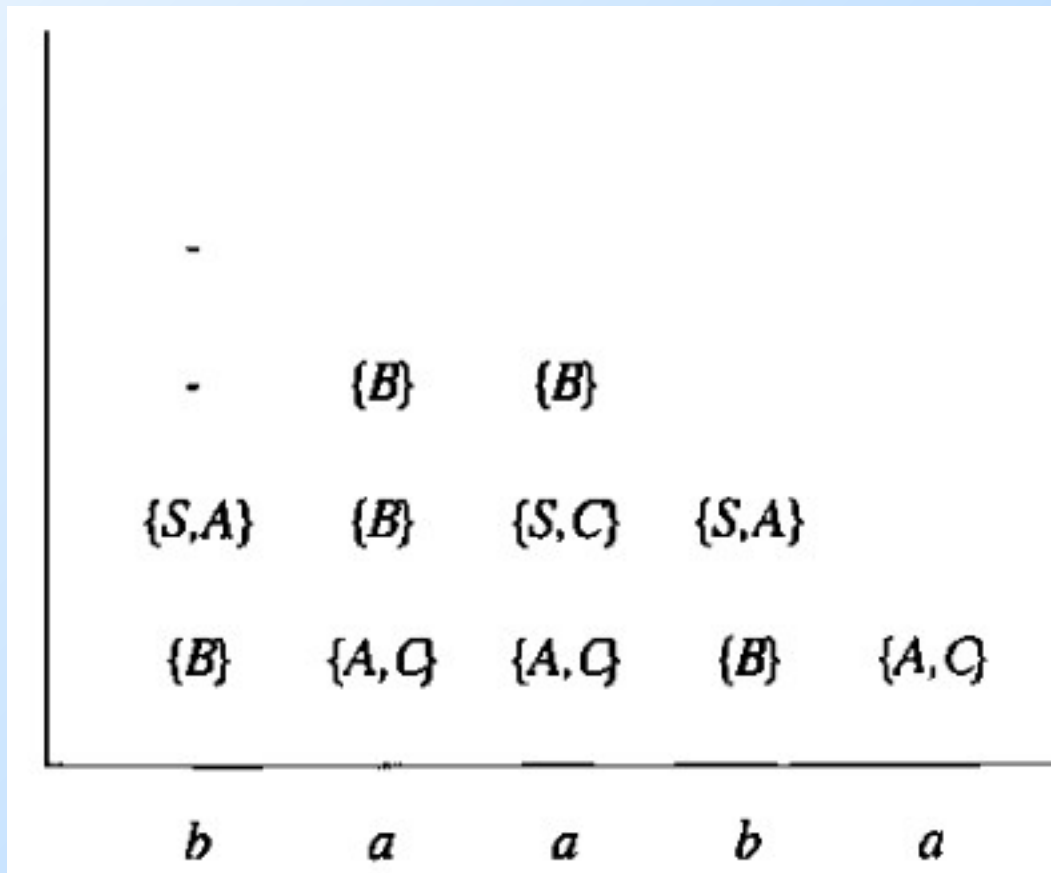
◆ Quebras:

- b e aab
- ba e ab
- baa e b

◆ Variável deve ter algum de {BB,SS,SC,AS,AC} no corpo

Exemplo

$$\blacklozenge X_{14} = \{\}$$



Exemplo

- ◆ Para X_{25} :
- ◆ Substring aaba
- ◆ Quebras:
 - a e aba
 - aa e ba
 - aab e a

Exemplo

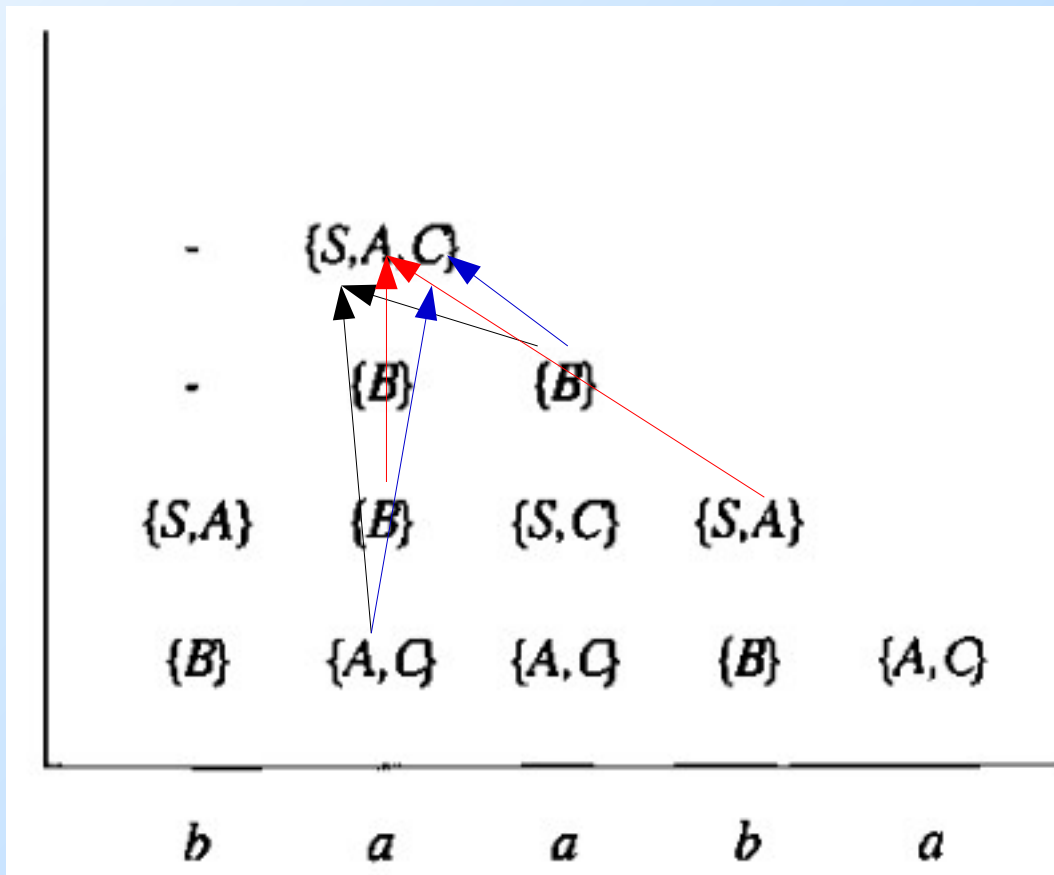
- ◆ Quebras:

- a e aba
- aa e ba
- aab e a

- ◆ Variável deve ter algum de $\{AB, CB, BS, BA, BC\}$ no corpo

Exemplo

◆ $X_{25} = \{S, A, C\}$



Exemplo

- ◆ Para X_{15} :
- ◆ Substring baaba
- ◆ Quebras:
 - b e aaba
 - ba e aba
 - baa e ba
 - baab e a

Exemplo

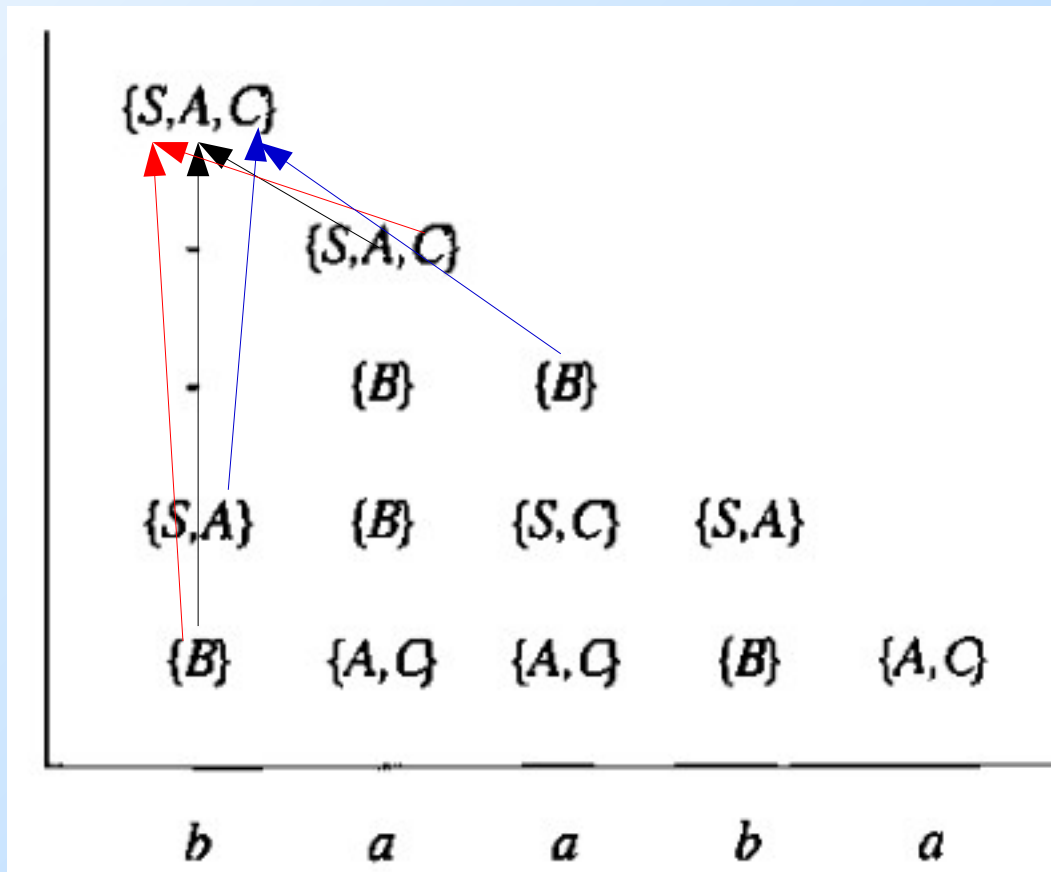
◆ Quebras:

- b e aaba
- ba e aba
- baa e ba
- baab e a

◆ Variável deve ter algum entre {BS,BA,BC,SB,AB} no corpo

Exemplo

◆ $X_{15} = \{S, A, C\}$



Algoritmo CYK

$D =$ “On input $w = w_1 \cdots w_n$:

1. If $w = \epsilon$ and $S \rightarrow \epsilon$ is a rule, *accept*. \llbracket handle $w = \epsilon$ case \rrbracket
2. For $i = 1$ to n : \llbracket examine each substring of length 1 \rrbracket
3. For each variable A :
4. Test whether $A \rightarrow b$ is a rule, where $b = w_i$.
5. If so, place A in $table(i, i)$.
6. For $l = 2$ to n : $\llbracket l$ is the length of the substring \rrbracket
7. For $i = 1$ to $n - l + 1$: $\llbracket i$ is the start position of the substring \rrbracket
8. Let $j = i + l - 1$, $\llbracket j$ is the end position of the substring \rrbracket
9. For $k = i$ to $j - 1$: $\llbracket k$ is the split position \rrbracket
10. For each rule $A \rightarrow BC$:
11. If $table(i, k)$ contains B and $table(k + 1, j)$ contains C , put A in $table(i, j)$.
12. If S is in $table(1, n)$, *accept*. Otherwise, *reject*.”

Exercício

GLC:

$S \rightarrow AB$

$A \rightarrow BC \mid a$

$B \rightarrow AC \mid b$

$C \rightarrow a \mid b$

String $w = ababa$

Exercício

$S \rightarrow AB$

$A \rightarrow BC \mid a$

$B \rightarrow AC \mid b$

$C \rightarrow a \mid b$

String $w = ababa$

$$X_{15} = \{A\}$$

$$X_{14} = \{B, S\} \quad X_{25} = \{A\}$$

$$X_{13} = \{A\} \quad X_{24} = \{B, S\} \quad X_{35} = \{A\}$$

$$X_{12} = \{B, S\} \quad X_{23} = \{A\} \quad X_{34} = \{B, S\} \quad X_{45} = \{A\}$$

$$X_{11} = \{A, C\} \quad X_{22} = \{B, C\} \quad X_{33} = \{A, C\} \quad X_{44} = \{B, C\} \quad X_{55} = \{A, C\}$$