Chương 6 Đơn giản hóa VPPNC và các dạng chuẩn

6.1 Các phương pháp để biến đổi văn phạm

6.2 Hai dạng chuẩn quan trọng

6.3 Giải thuật thành viên cho văn phạm phi ngữ cảnh

Các phương pháp để biến đổi văn phạm

- Chuỗi trống đóng một vai trò khá đặc biệt trong nhiều định lý và chứng minh, và thường cần có một sự chú ý đặc biệt cho nó.
- Nếu $L \ni \lambda$ thì biểu diễn $L = L_1 \cup \lambda$ với $L_1 = L - \lambda$. Nếu

  $$G_1 = (V_1, T, S_1, P_1)$$

  là văn phạm biểu diễn cho L_1 thì

  $$G = (V_1 \cup \{S\}, T, S, P_1 \cup \{S \to S_1 \mid \lambda\})$$

  là văn phạm biểu diễn cho L.
- Trong chương này, chúng ta chỉ xem xét các NNPNC không chứa λ.
- Tuy nhiên những kết luận cho ngôn ngữ không chứa λ vẫn có thể áp dụng cho ngôn ngữ có chứa λ.

Một vài qui tắc thay thế hiệu quả

- Định lý 6.1
  - Cho $G = (V, T, S, P)$ là một VPPNC. Giả sử P có chứa luật sinh

$$A \rightarrow x_1Bx_2$$

trong đó A, B là các biến khác nhau và

$$B \rightarrow y_1 \mid y_2 \mid \ldots \mid y_n$$

là tập tất cả các luật sinh trong P mà có B ở vế trái.
Cho $G_1 = (V, T, S, P_1)$ là VP được xây dựng bằng cách xóa đi

$$A \rightarrow x_1Bx_2$$

từ P, và thêm vào nó

$$A \rightarrow x_1y_1x_2 \mid x_1y_2x_2 \mid \ldots \mid x_1y_nx_2$$

Thì

$$L(G) = L(G_1)$$

Ví dụ

- Xét văn phạm $G = (\{A, B\}, \{a, b\}, A, P)$ với các luật sinh

$$A \to a \mid aA \mid bBc,$$
$$B \to abA \mid b.$$

Sau khi thay thế biến B ta nhận được VP tương đương như sau

$$A \to a \mid aA \mid babAc \mid bbc,$$
$$B \to abA \mid b$$

- Chuỗi *abbc* có các dẫn xuất trong G và G_1 lần lượt như sau:

$$A \Rightarrow aA \Rightarrow abBc \Rightarrow abbc$$
$$A \Rightarrow aA \Rightarrow abbc$$

- Chú ý rằng, biến B và các luật sinh của nó vẫn còn ở trong VP mặc dù chúng không còn đóng vai trò gì trong bất kỳ dẫn xuất nào. Sau này chúng ta sẽ thấy rằng những luật sinh không cần thiết như vậy có thể bị loại bỏ ra khỏi văn phạm.

Loại bỏ đệ qui trái

- Định lý 6.2 (Loại bỏ đệ qui trái)
  - Cho $G = (V, T, S, P)$ là một VPPNC. Chia tập các luật sinh mà vế trái của chúng là một biến đã cho nào đó (chẳng hạn là A), thành hai tập con riêng biệt

$$A \to Ax_1 \mid Ax_2 \mid \ldots \mid Ax_n \qquad (6.2)$$

$$A \to y_1 \mid y_2 \mid \ldots \mid y_m \qquad (6.3)$$

với $x_i, y_i \in (V \cup T)^*$, và A không là prefix của bất kỳ y_i nào.

Xét $G_1 = (V \cup \{Z\}, T, S, P_1)$, trong đó $Z \notin V$ và P_1 nhận được bằng cách thay mọi luật sinh của P có dạng (6.2) và (6.3) bởi

$$A \to y_i \mid y_iZ,\ i = 1, 2, \ldots, m,$$

$$Z \to x_i \mid x_iZ,\ i = 1, 2, \ldots, n,$$

Thì

$$L(G) = L(G_1).$$

Loại bỏ đệ qui trái (tt)

- Chứng minh
  - Các dạng câu mà A sinh ra trong văn phạm G có dạng:

    $$A \stackrel{*}{\Rightarrow} A(x_1 + x_2 + \dots + x_n)^* \Rightarrow y_i(x_1 + x_2 + \dots + x_n)^*$$

    Các dạng câu này cũng có thể được sinh ra trong G_1 bằng cách chú ý Z có thể sinh ra các dạng câu có dạng

    $$Z \stackrel{*}{\Rightarrow} (x_1 + x_2 + \dots + x_n)(x_1 + x_2 + \dots + x_n)^*$$

    mà $A \to y_i \mid y_i Z$ nên

    $$A \stackrel{*}{\Rightarrow} y_i(x_1 + x_2 + \dots + x_n)^*$$

    Vì vậy $L(G) = L(G_1)$.
- Ghi chú
  - Các luật sinh đệ qui-trái chỉ là một trường hợp đặc biệt của đệ qui-trái trong văn phạm như được phát biểu sau.
  - Một văn phạm được gọi là đệ qui-trái nếu có một biến A nào đó mà đối với nó $A \stackrel{*}{\Rightarrow} Ax$ là có thể.

Ví dụ

- Sử dụng Định lý 6.2 để loại bỏ các luật sinh đệ qui-trái khỏi VP

$$A \rightarrow Aa \mid aBc \mid \lambda \qquad\qquad B \rightarrow Bb \mid ba$$

- Áp dụng định lý cho biến A ta được tập luật sinh mới như sau:

$$A \rightarrow aBc \mid \lambda \mid aBcZ \mid Z \qquad B \rightarrow Bb \mid ba$$
$$Z \rightarrow a \mid aZ$$

- Áp dụng định lý một lần nữa lần này cho biến B ta được tập luật sinh kết quả cuối cùng như sau:

$$A \rightarrow aBc \mid aBcZ \mid Z \mid \lambda \qquad B \rightarrow ba \mid baY$$
$$Z \rightarrow a \mid aZ \qquad Y \rightarrow b \mid bY$$

- Nhận xét
  - Việc loại bỏ các luật sinh đệ qui-trái đưa ra các biến mới. VP kết quả có thể là "đơn giản" hơn đáng kể so với VP gốc nhưng một cách tổng quát nó sẽ có nhiều biến và luật sinh hơn.

Luật sinh vô dụng

- Định nghĩa 6.1:
  - Cho $G = (V, T, S, P)$ là một VPPNC. Một biến $A \in V$ được gọi là **khả dụng** nếu và chỉ nếu có ít nhất một chuỗi $w \in L(G)$ sao cho $S \stackrel{*}{\Rightarrow} xAy \stackrel{*}{\Rightarrow} w$,

    với $x, y \in (V \cup T)^*$. Bằng lời, một biến là khả dụng nếu và chỉ nếu nó xuất hiện trong ít nhất một dẫn xuất. Một biến mà không khả dụng thì gọi là **vô dụng.** Một luật sinh được gọi là vô dụng nếu nó có chứa bất kỳ biến vô dụng nào.
- Các dạng vô dụng
  - Vô dụng loại 1: $A \stackrel{*}{\not\Rightarrow} w \in T^*$
  - Vô dụng loại 2: $S \stackrel{*}{\not\Rightarrow} xAy$

Loại bỏ các luật sinh vô dụng

- Định lý 6.3
  - Cho $G = (V, T, S, P)$ là một VPPNC, $\exists$ một VP tương đương $G_0 = (V_0, T, S, P_0)$ mà không chứa bất kỳ biến vô dụng nào.
- Chứng minh
  - Loại bỏ các biến và luật sinh vô dụng loại 1

    Tạo văn phạm $G_1 = (V_1, T, S, P_1)$ với V_1 là tập biến không vô dụng loại 1. Ta tìm V_1 như sau:

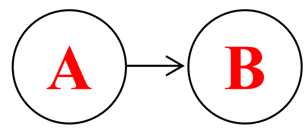
    1. Khởi tạo $V_1 = \varnothing$.
    2. Lặp lại bước sau cho đến khi không còn biến nào được thêm vào V_1.
       - Đối với mỗi $A \in V$ mà có luật sinh $A \to x$, $x \in (V_1 \cup T)^*$, thì thêm A vào V_1.
    3. Loại khỏi P các luật sinh có chứa các biến $\notin V_1$, ta được P_1.

Loại bỏ các luật sinh vô dụng (tt)

- Để loại tiếp các biến và các luật sinh vô dụng loại 2 ta dựa vào G_1 vừa có ở trên và vẽ đồ thị phụ thuộc cho nó, sau đó tìm tập các biến không đạt tới được từ S. Loại các biến này và các luật sinh liên quan đến nó ra khỏi G_1 ta được văn phạm kết quả G_0.

- Đồ thị phụ thuộc (dependency graph)
  - Là một đồ thị có các đỉnh biểu diễn các biến, còn một cạnh nối hai đỉnh A và B khi và chỉ khi có luật sinh dạng

$$A \to xBy$$



- Ví dụ
  - Loại bỏ các biến và các luật sinh vô dụng ra khỏi văn phạm $G = (\{S, A, B, C\}, \{a, b\}, S, P)$, với tập luật sinh P là:

$$S \to aS \mid A \mid C \qquad B \to aa$$
$$A \to a \qquad C \to aCb$$

Ví dụ

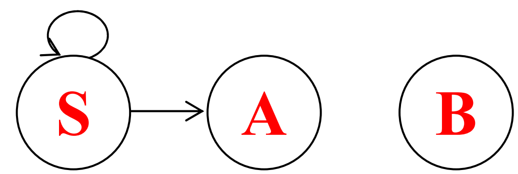
- Loại bỏ các biến vô dụng loại 1 ta được $V_1 = \{S, A, B\}$ và tập luật sinh P_1

$$S \rightarrow aS \mid A$$
$$A \rightarrow a$$
$$B \rightarrow aa$$

- Loại bỏ các biến vô dụng loại 2 ta được văn phạm kết quả

$$S \rightarrow aS \mid A$$
$$A \rightarrow a$$

$$S \rightarrow aS \mid A \mid C$$
$$A \rightarrow a$$
$$B \rightarrow aa$$
$$C \rightarrow aCb$$



- Nhận xét
  - Nếu thay đổi thứ tự loại bỏ (loại bỏ các biến và luật sinh vô dụng loại 2 trước) thì sẽ không loại bỏ được tất cả các biến và luật sinh vô dụng chỉ bằng một lần như ví dụ sau cho thấy.

Ví dụ (tt)

- Xét văn phạm sau

$$S \to aSb \mid ab \mid A$$
$$A \to aAB$$
$$B \to b$$

- Nếu loại bỏ các biến và luật sinh vô dụng loại 2 trước ta thấy văn phạm vẫn không thay đổi vì tất cả các biến đều đạt tới được từ S. Sau đó loại bỏ tiếp các biến và luật sinh vô dụng loại 1 ta sẽ được văn phạm sau:

$$S \to aSb \mid ab \mid A$$
$$B \to b$$

- Rõ ràng văn phạm này còn biến B là vô dụng loại 2.

Loại bỏ luật sinh-λ

- Định nghĩa 6.2
  - Bất kỳ luật sinh nào của VPPNC có dạng

$$A \rightarrow \lambda$$

  được gọi là luật sinh-λ. Bất kỳ biến A nào mà

$$A \stackrel{*}{\Rightarrow} \lambda$$

  - là có thể thì được gọi là khả trống (nullable).
- Định lý 6.4
  - Cho G là một VPPNC bất kỳ mà $L(G)$ không chứa λ, thì tồn tại một văn phạm G_0 tương đương mà không có chứa luật sinh-λ.
- Chứng minh:

  Bước 1
  - Tìm tập V_N tất cả các biến khả trống của G bằng các bước sau.

Loại bỏ luật sinh-λ

1. Đối với mọi luật sinh A $\rightarrow \lambda$, đưa A vào V_N.
2. Lặp lại bước sau cho đến khi không còn biến nào được thêm vào V_N.
   - Đối với mọi luật sinh $B \rightarrow A_1A_2 \ldots A_n$, mà $A_1, A_2, A_n \in V_N$ thì đặt B vào V_N.

Bước 2

- Sau khi có tập V_N ta xây dựng tập luật sinh như sau.
- Ứng với mỗi luật sinh có dạng $A \rightarrow x_1x_2 \ldots x_m$, $m \geq 1$, trong đó mỗi $x_i \in V \cup T$, đặt luật sinh này vào cùng với các luật sinh được sinh ra bằng cách thay thế các biến khả trống bằng λ trong ***mọi tổ hợp có thể***, ngoại trừ nếu tất cả các x_i đều khả trống thì không đặt luật sinh $A \rightarrow \lambda$ vào P_0 của G_0

Ví dụ

- Loại bỏ các luật sinh-λ của văn phạm sau:

$$S \rightarrow ABaC \qquad C \rightarrow D \mid \lambda$$
$$A \rightarrow BC \qquad D \rightarrow d$$
$$B \rightarrow b \mid \lambda$$

- Vì $B \rightarrow \lambda$ và $C \rightarrow \lambda$ suy ra B và C là các biến khả trống.
- Vì $A \rightarrow BC$ nên suy ra A cũng là biến khả trống. Ngoài ra không còn biến nào khác là khả trống.
- Theo Bước 2 ta xây dựng được tập luật sinh mới tương đương như sau:

$$S \rightarrow ABaC \mid BaC \mid AaC \mid Aba \mid aC \mid Aa \mid Ba \mid a$$
$$A \rightarrow BC \mid B \mid C$$
$$B \rightarrow b$$
$$C \rightarrow D$$
$$D \rightarrow d$$

Loại bỏ luật sinh đơn vị

- Định nghĩa 6.3
  - Bất kỳ luật sinh nào của VPPNC có dạng

$$A \rightarrow B$$

  trong đó A, $B \in V$ được gọi là luật sinh-đơn vị.

- Định lý 6.5
  - Cho $G = (V, T, S, P)$ là một VPPNC bất kỳ không có luật sinh-λ, thì tồn tại một VPPNC $G_1 = (V_1, T, S, P_1)$ mà không có bất kỳ luật sinh đơn vị nào và tương đương với G_1.

- Chứng minh
  1. Đặt vào trong P_1 tất cả các luật sinh không đơn vị của P.
  2. Đối với mỗi biến A tìm tất cả các biến B mà $A \overset{*}{\Rightarrow} B$ (*)
     - Điều này thực hiện bằng cách vẽ đồ thị phụ thuộc cho G nhưng một cạnh nối 2 đỉnh A và B khi và chỉ khi có luật sinh-đơn vị $A \rightarrow B$. Hai biến A và B thõa (*) khi và chỉ khi có một con đường trong đồ thị đi từ A đến B.

Ví dụ

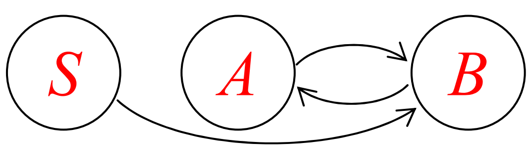
3. Đối với mỗi A, B thõa (*) thêm vào trong P_1 các luật sinh

$$A \rightarrow y_1 \mid y_2 \mid \ldots \mid y_n$$

với $B \rightarrow y_1 \mid y_2 \mid \ldots \mid y_n$ là các luật sinh không đơn vị của B.

- Ví dụ
  - Loại bỏ các luật sinh đơn vị cho VP sau

$$S \rightarrow Aa \mid B$$
$$B \rightarrow A \mid bb$$
$$A \rightarrow a \mid bc \mid B$$



Trước hết, đặt các luật sinh không đơn vị vào trong $P1$

$$S \rightarrow Aa$$
$$A \rightarrow a \mid bc$$
$$B \rightarrow bb$$

Từ ĐTPT ta đưa được thêm các luật sinh sau vào

$$S \rightarrow a \mid bc \mid bb$$
$$A \rightarrow bb$$
$$B \rightarrow a \mid bc$$

Ví dụ (tt)

- Kết quả ta có văn phạm tương đương sau không có luật sinh đơn vị

$$S \to Aa \mid a \mid bc \mid bb$$
$$A \to a \mid bc \mid bb$$
$$B \to bb \mid a \mid bc$$

- Định lý 6.6
  - Cho L là một NNPNC không chứa λ, tồn tại một VPPNC sinh ra L mà không chứa bất kỳ luật sinh vô dụng, luật sinh-λ, hay luật sinh-đơn vị nào.
- Chứng minh:

  B1. Loại bỏ luật sinh-λ

  B2. Loại bỏ luật sinh đơn vị

  B3. Loại bỏ luật sinh vô dụng loại 1, rồi vô dụng loại 2

Một số nhận xét

1. Loại bỏ biến vô dụng loại 1 có thể sinh ra biến vô dụng loại 2.
2. Việc loại bỏ biến vô dụng loại 2 không sinh ra biến vô dụng loại 1.
3. Văn phạm không có luật sinh đơn vị thì việc loại bỏ luật sinh-λ có thể sinh ra luật sinh-đơn vị.
4. Văn phạm không có luật sinh-λ thì việc loại bỏ luật sinh-đơn vị không thể sinh ra luật sinh-λ mới.
5. Loại bỏ luật sinh-λ có thể sinh ra biến vô dụng loại 1.
6. Loại bỏ luật sinh-đơn vị có thể sinh ra biến vô dụng loại 2.
7. Văn phạm không có luật sinh-λ, luật sinh-đơn vị thì việc loại bỏ các luật sinh vô dụng loại 1, loại 2 không sinh ra thêm bất kỳ luật sinh-λ và luật sinh-đơn vị nào mới.

Dạng chuẩn Chomsky

- Định nghĩa 6.4
  - Một VPPNC là thuộc dạng chuẩn Chomsky nếu mọi luật sinh có dạng

$$A \to BC, \text{ hoặc}$$
$$A \to a$$

    trong đó $A, B, C \in V$, còn $a \in T$.

- Định lý 6.7
  - Bất kỳ VPPNC $G = (V, T, S, P)$ nào với $\lambda \notin L(G)$ đều có một văn phạm tương đương $G_1 = (V_1, T, S, P_1)$ có dạng chuẩn Chomsky.

- Chứng minh
  - Không mất tổng quát giả sử G không có luật sinh-vô dụng, luật sinh-đơn vị và luật sinh-λ. Ta xây dựng văn phạm G_1 có dạng chuẩn Chomsky bằng thủ tục sau:

Thủ tục: G-to-G_{Chomsky}

- **Input**: $G = (V, T, S, P)$ với $\lambda \notin L(G)$
- **Output**: $G_1 = (V_1, T, S, P_1)$ có dạng chuẩn Chomsky.

1. Đặt các luật sinh $A \to a$ vào P_1.
2. Đối với các luật sinh $A \to x_1 x_2 \ldots x_n$ với $n \geq 2$, $x_i \in (V \cup T)$ thì thay các kí hiệu kết thúc, chẳng hạn $x_k = a$, bằng các biến đại diện mới B_a, tạo thành các luật sinh trung gian $A \to C_1C_2 \ldots C_n$.
3. Ứng với mỗi biến đại diện B_a đặt vào P_1 các luật sinh $B_a \to a$.
4. Sau khi thực hiện bước 2, ứng với mỗi luật sinh $A \to C_1C_2 \ldots C_n$ mà $n = 2$ đặt nó vào P_1. Ngược lại ứng với $n > 2$ ta giới thiệu các biến mới $D_1, D_2, \ldots$ và đưa vào các luật sinh sau:

$$\begin{aligned}
A &\to C_1D_1 \\
D_1 &\to D_1D_2 \\
&\vdots \\
D_{n-2} &\to C_{n-1}C_n
\end{aligned}$$

Ví dụ

- Hãy biến đổi VP sau thành VP có dạng chuẩn Chomsky.

$$
\begin{aligned}
S &\rightarrow a \mid ABa\\
A &\rightarrow aab\\
B &\rightarrow b \mid Ac
\end{aligned}
$$

Bước 1

$$
\begin{aligned}
S &\rightarrow a\\
B &\rightarrow b
\end{aligned}
$$

Bước 2

$$
\begin{aligned}
S &\rightarrow ABX_a\\
A &\rightarrow X_aX_aX_b\\
B &\rightarrow AX_c
\end{aligned}
$$

Bước 3

$$
\begin{aligned}
X_a &\rightarrow a\\
X_b &\rightarrow b\\
X_c &\rightarrow c
\end{aligned}
$$

Bước 4

$$
\begin{aligned}
S &\rightarrow AD_1\\
D_1 &\rightarrow BX_a\\
A &\rightarrow X_aD_2\\
D_2 &\rightarrow X_aX_b
\end{aligned}
$$

Kết quả

$$
\begin{aligned}
S &\rightarrow a \mid AD_1\\
D_1 &\rightarrow BX_a\\
A &\rightarrow X_aD_2\\
D_2 &\rightarrow X_aX_b\\
B &\rightarrow b \mid AXc\\
X_a &\rightarrow a\\
X_b &\rightarrow b\\
X_c &\rightarrow c
\end{aligned}
$$

Dạng chuẩn Greibach

- Định nghĩa 6.5
  - Một VPPNC là thuộc dạng chuẩn Greibach nếu mọi luật sinh có dạng

$$A \to ax$$

    trong đó $a \in T$ còn $x \in V^*$.
- Định lý 6.8
  - Đối với mọi VPPNC G với $\lambda \notin L(G)$, thì tồn tại một văn phạm tương đương trong dạng chuẩn *Greibach*.
- Chứng minh
  - Không mất tính tổng quát giả sử G không có luật sinh-vô dụng, luật sinh-đơn vị và luật sinh-λ. Ta xây dựng văn phạm có dạng chuẩn Greibach bằng thủ tục sau.

Thủ tục: *G*-to-$G_{Greibach}$

- **Input**: $G = (V, T, S, P)$ với $\lambda \notin L(G)$
- **Output**: $G_1 = (V_1, T, S, P_1)$ có dạng chuẩn Greibach.

1. Đánh số thứ tự cho các biến chẳng hạn là $A_1, A_2, \ldots A_n$.
2. Dùng Định lý 6.1 và 6.2 để viết lại VP sao cho các luật sinh có một trong ba dạng sau

$$A_i \to A_j x_j,\ i < j \qquad a \in T \text{ và } x_i \in (V \cup T)^*$$
$$Z_i \to A_j x_j,\ j \le n$$
$$A_i \to a x_i \qquad Z_i \text{ là các biến mới}$$

- Điều này thực hiện được bằng cách sử dụng Định lý 6.1 và 6.2 cho các biến A_i theo thứ tự i đi từ 1, 2, ... đến n như sau.
- Giả sử xét luật sinh của biến A_i. Nếu có luật sinh $A_i \to A_j x$ mà $i > j$ thì *thay A_j đi đầu* bằng các vế phải của nó, và làm cho đến khi các luật sinh của A_i có dạng $A_i \to A_j x$, $i \le j$. Đến đây loại đệ qui trái cho A_i thì các luật sinh của nó sẽ có dạng như đã nêu.

Thủ tục: *G*-to-$G_{Greibach}$ (tt)

3. Sau khi thực hiện bước 2, tất cả các luật sinh của A_n phải có dạng $A_n \to ax_n$
   - *Thay A_n đi đầu vế phải của các luật sinh* bằng các vế phải của nó. Kết quả các luật sinh của $A_{n\text{-}1}$ có dạng

$$A_{n\text{-}1} \to ax_{n\text{-}1}$$

   - Tương tự thay thế *$A_{n\text{-}1}$ đi đầu vế phải của các luật sinh* bằng các vế phải của nó. Và thực hiện lần lượt cho đến A_1.
4. Thay các kí hiệu kết thúc, chẳng hạn a, không đi đầu vế phải bằng các biến đại diện, chẳng hạn X_a, đồng thời thêm vào các luật sinh mới $X_a \to a$.

- Ví dụ
  - Biến đổi VP sau thành VP có dạng chuẩn Greibach

$$S \to SBb \mid Ab$$
$$A \to Sb \mid Ba$$
$$B \to Sb \mid a$$

Ví dụ

$$S \to SBb \mid Ab$$
$$A \to Sb \mid Ba$$
$$B \to Sb \mid a$$

$S_0 \to S_0B_2b \mid A_1b$ — Loại đệ qui trái →

$S_0 \to A_1b \mid A_1bZ_0$ (1)
$Z_0 \to B_2b \mid B_2bZ_0$ (2)

$A_1 \to S_0b \mid B_2a$ — Thay thế →

$A_1 \to A_1bb \mid A_1bZ_0b \mid B_2a$

Loại đệ qui trái →

$A_1 \to B_2a \mid B_2aZ_1$ (3)
$Z_1 \to bb \mid bZ_0b \mid bbZ_1 \mid bZ_0bZ_1$ (4)

$B_2 \to S_0b \mid a$ — Thay thế →

$B_2 \to A_1ba \mid A_1bZ_0a \mid b$

Thay thế ↓

$B_2 \to B_2aba \mid B_2aZ_1ba \mid B_2abZ_0a \mid B_2aZ_1bZ_0a \mid b$

← Loại đệ qui trái

$B_2 \to b \mid bZ_2$ (5)
$Z_2 \to aba \mid aZ_1ba \mid abZ_0a \mid aZ_1bZ_0a \mid abaZ_2 \mid aZ_1baZ_2 \mid abZ_0aZ_2 \mid aZ_1bZ_0aZ_2$ (6)

Ví dụ (tt)

$$B_2 \rightarrow b \mid bZ_3 \qquad (5)$$

$$A_1 \rightarrow B_2a \mid B_2aZ_1 \quad (3)$$

Thay thế

$$A_1 \rightarrow ba \mid bZ_2a \mid baZ_1 \mid bZ_2aZ_1 \quad (7)$$

$$S_0 \rightarrow A_1b \mid A_1bZ_0 \quad (1)$$

Thay thế

$$\begin{aligned} S_0 \rightarrow\ & bab \mid bZ_2ab \mid baZ_1b \mid \\ & bZ_2aZ_1b \mid babZ_0 \mid bZ_2abZ_0 \mid \\ & baZ_1bZ_0 \mid bZ_2aZ_1bZ_0 \qquad (8) \end{aligned}$$

$$Z_0 \rightarrow B_2b \mid B_2bZ_0 \quad (2)$$

Thay thế

$$Z_0 \rightarrow bb \mid bZ_2b \mid bbZ_0 \mid bZ_2bZ_0 \quad (9)$$

Ví dụ (tt)

Văn phạm gần-Greibach

$$
\begin{array}{ll}
S_0 \rightarrow bab \mid bZ_2ab \mid baZ_1b \mid bZ_2aZ_1b \mid babZ_0 \mid bZ_2abZ_0 \mid baZ_1bZ_0 \mid bZ_2aZ_1bZ_0 & (8) \\
A_1 \rightarrow ba \mid bZ_2a \mid baZ_1 \mid bZ_2aZ_1 & (7) \\
B_2 \rightarrow b \mid bZ_3 & (5) \\
Z_0 \rightarrow bb \mid bZ_2b \mid bbZ_0 \mid bZ_2bZ_0 & (9) \\
Z_1 \rightarrow bb \mid bZ_0b \mid bbZ_1 \mid bZ_0bZ_1 & (4) \\
Z_2 \rightarrow aba \mid aZ_1ba \mid abZ_0a \mid aZ_1bZ_0a \mid abaZ_2 \mid aZ_1baZ_2 \mid abZ_0aZ_2 \mid aZ_1bZ_0aZ_2 & (6)
\end{array}
$$

Thay kí hiệu kết thúc không đi đầu bằng biến đại diện

$$
\begin{array}{l}
S \rightarrow bXY \mid bZ_2XY \mid bXZ_1Y \mid bZ_2XZ_1Y \mid bXYZ_0 \mid bZ_2XYZ_0 \mid bXZ_1YZ_0 \mid bZ_2XZ_1YZ_0 \\
A \rightarrow bX \mid bZ_2X \mid bXZ_1 \mid bZ_2XZ_1 \\
B \rightarrow b \mid bZ_3 \\
Z_0 \rightarrow bY \mid bZ_2Y \mid bYZ_0 \mid bZ_2YZ_0 \\
Z_1 \rightarrow bY \mid bZ_0Y \mid bYZ_1 \mid bZ_0YZ_1 \\
Z_2 \rightarrow aYX \mid aZ_1YX \mid aYZ_0X \mid aZ_1YZ_0X \mid aYXZ_2 \mid aZ_1YXZ_2 \mid aYZ_0XZ_2 \mid aZ_1YZ_0XZ_2 \\
X \rightarrow a \\
Y \rightarrow b
\end{array}
$$

Giải thuật thành viên cho VPPNC

- Giải thuật CYK (J.Cocke, D.H.Younger, T.Kasami)
  - **Input**: Văn phạm Chomsky $G = (V, T, S, P)$

    Chuỗi $w = a_1a_2 \dots a_n$
  - **Output**: "Yes" + DXTN hoặc "No".
  - Chúng ta định nghĩa các chuỗi con

    $$w_{ij} = a_i \dots a_j,$$
  - Và các tập con của V

    $$V_{ij} = \{A \in V : A \overset{*}{\Rightarrow} w_{ij}\},$$
  - Để ý $w = w_{1n}$, vậy $w \in L(G)$ khi và chỉ khi $S \in V_{1n}$.
  - *Vậy để biết w có $\in L(G)$ hay không chúng ta tính V_{1n} và xem S có $\in V_{1n}$ hay không.*

Giải thuật CYK

- A $\in V_{ii}$ nếu và chỉ nếu A có luật sinh $A \rightarrow a_i$.
- Vậy, V_{ii} có thể được tính $\forall i$, $1 \leq i \leq n$.
- Nếu $B \stackrel{*}{\Rightarrow} w_{ik}$ ($\Leftrightarrow B \in V_{ik}$), $C \stackrel{*}{\Rightarrow} w_{(k+1)j}$ ($\Leftrightarrow C \in V_{(k+1)j}$) và đồng thời $A \rightarrow BC$ thì $A \stackrel{*}{\Rightarrow} w_{ij}$ ($\Leftrightarrow A \in V_{ij}$) $\forall\ i \leq k$, $k < j$.
- $V_{ij} = \cup_{k \in \{i,\, i+1,\, \ldots\, ,\, j-1\}} \{A\text{: } A \rightarrow BC$, **với** $B \in V_{ik}$, $C \in V_{(k+1)j}\}$
- Quá trình tính các tập V_{ij}
- $V_{11}, V_{22}, \ldots, V_{nn}$
- $V_{12}, V_{23}, \ldots, V_{(n-1)n}$
- $V_{13}, V_{24}, \ldots, V_{(n-2)n}$
- ...
- V_{1n}

Ví dụ

- Sử dụng giải thuật CYK để PTCP chuỗi $w = aabbb$ trên G sau

$$S \rightarrow AB \qquad (1)$$
$$A \rightarrow BB \mid a \qquad (2, 3)$$
$$B \rightarrow AB \mid b \qquad (4, 5)$$

- Ta có $w = a\ a\ b\ b\ b$

  1 2 3 4 5

- $V_{11} = \{A\}, V_{22} = \{A\}, V_{33} = \{B\}, V_{44} = \{B\}, V_{55} = \{B\},$
- $V_{12} = \varnothing, V_{23} = \{S, B\}, V_{34} = \{A\}, V_{45} = \{A\},$
- $V_{13} = \{S, B\}, V_{24} = \{A\}, V_{35} = \{S, B\},$
- $V_{14} = \{A\}, V_{25} = \{S, B\},$
- $V_{15} = \{S, B\}.$ $\boxed{S \in V_{15} \Rightarrow w \in L(G).}$

Ví dụ (tt)

$S \rightarrow AB$ (1)
$A \rightarrow BB \mid a$ (2, 3)
$B \rightarrow AB \mid b$ (4, 5)

$w = a\ a\ b\ b\ b$
$\quad\ \ 1\ 2\ 3\ 4\ 5$

- Để tìm dẫn xuất cho *w*, chúng ta phải tìm cách "lưu vết"
- $V_{11} = \{A\,(\xrightarrow{3} a)\}$, $V_{22} = \{A(\xrightarrow{3} a)\}$, $V_{33} = \{B(\xrightarrow{5} b)\}$,
  $V_{44} = \{B(\xrightarrow{5} b)\}$, $V_{55} = \{B(\xrightarrow{5} b)\}$,
- $V_{12} = \varnothing$, $V_{23} = \{S(\xrightarrow{1} A_{22}B_{33}), B(\xrightarrow{4} A_{22}B_{33})\}$,
  $V_{34} = \{A(\xrightarrow{2} B_{33}B_{44})\}$, $V_{45} = \{A(\xrightarrow{2} B_{44}B_{55})\}$,
- $V_{13} = \{S(\xrightarrow{1} A_{11}B_{23}), B(\xrightarrow{4} A_{11}B_{23})\}$, $V_{24} = \{A(\xrightarrow{2} B_{23}B_{44})\}$,
  $V_{35} = \{S(\xrightarrow{1} A_{34}B_{55}), B(\xrightarrow{4} A_{34}B_{55})\}$,
- $V_{14} = \{A(\xrightarrow{2} B_{13}B_{44})\}$, $V_{25} = \{S(\xrightarrow{1} A_{22}B_{35}, \xrightarrow{1} A_{24}B_{55}),$
  $B(\xrightarrow{4} A_{22}B_{35}, \xrightarrow{4} A_{24}B_{55})\}$,
- $V_{15} = \{S(\xrightarrow{1} A_{22}B_{35}, \xrightarrow{1} A_{24}B_{55}), B(\xrightarrow{4} A_{22}B_{35}, \xrightarrow{4} A_{24}B_{55})\}$.

Ví dụ (tt)

$$S \rightarrow AB \quad (1)$$
$$A \rightarrow BB \mid a \quad (2, 3)$$
$$B \rightarrow AB \mid b \quad (4, 5)$$

$$w = a\ a\ b\ b\ b$$
$$1\ 2\ 3\ 4\ 5$$

- Kết quả có 3 DXTN như sau:

(1) $S \stackrel{1}{\Rightarrow} A_{11}B_{25} \stackrel{3}{\Rightarrow} aB_{25} \stackrel{4}{\Rightarrow} aA_{22}B_{35} \stackrel{3}{\Rightarrow} aaB_{35} \stackrel{4}{\Rightarrow} aaA_{34}B_{55} \stackrel{2}{\Rightarrow} aaB_{33}B_{44}B_{55} \stackrel{5,5,5}{\Rightarrow} aabbb$ (DXTN: 134342555)

(2) $S \stackrel{1}{\Rightarrow} A_{11}B_{25} \stackrel{3}{\Rightarrow} aB_{25} \stackrel{4}{\Rightarrow} aA_{24}B_{55} \stackrel{2}{\Rightarrow} aB_{23}B_{44}B_{55} \stackrel{4}{\Rightarrow} aA_{22}B_{33}B_{44}B_{55} \stackrel{3,5,5,5}{\Rightarrow} aabbb$ (DXTN: 134243555)

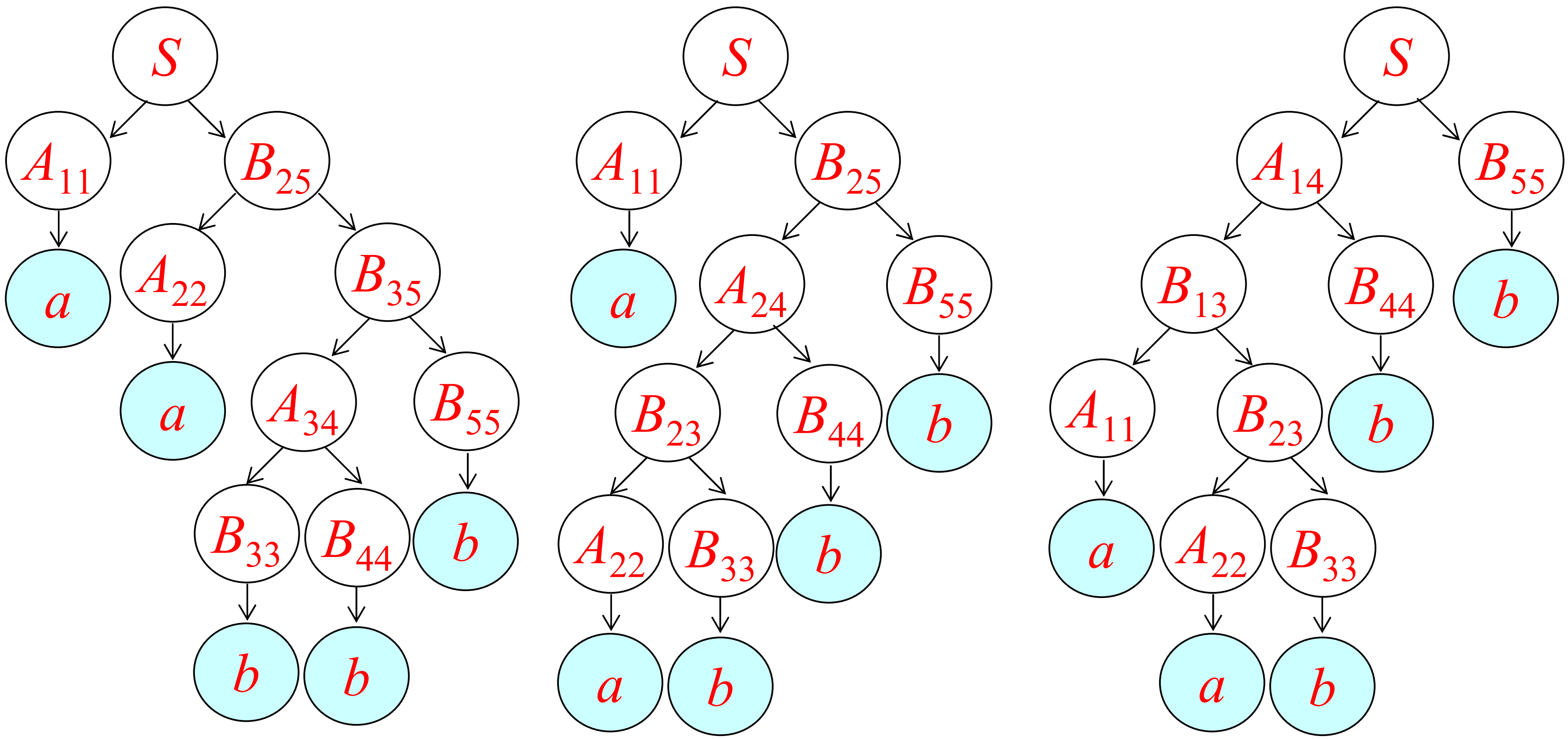
(3) $S \stackrel{1}{\Rightarrow} A_{14}B_{55} \stackrel{2}{\Rightarrow} B_{13}B_{44}B_{55} \stackrel{4}{\Rightarrow} A_{11}B_{23}B_{44}B_{55} \stackrel{3}{\Rightarrow} aB_{23}B_{44}B_{55} \stackrel{4}{\Rightarrow} aA_{22}B_{33}B_{44}B_{55} \stackrel{3,5,5,5}{\Rightarrow} aabbb$ (DXTN: 124343555)

Ví dụ (tt)

$$S \rightarrow AB \qquad (1)$$
$$A \rightarrow BB \mid a \qquad (2, 3)$$
$$B \rightarrow AB \mid b \qquad (4, 5)$$

$$w = a\ a\ b\ b\ b$$
$$\quad\ 1\ 2\ 3\ 4\ 5$$

- 3 CDXtương ứng:



Bài tập

- Dùng giải thuật CYK PTCP các chuỗi sau w_1 = ***abab***, w_2 = ***abaa*** trên các VP G_1, G_2 tương ứng.
- G_1

  $S \rightarrow AB \mid BB \quad (1, 2)$

  $A \rightarrow BA \mid a \quad (3, 4)$

  $B \rightarrow AA \mid a \mid b \quad (5, 6, 7)$

  G_2

  $S \rightarrow AB \quad (1)$

  $A \rightarrow BB \mid a \quad (2, 3)$

  $B \rightarrow BA \mid b \quad (4, 5)$