Chương 3 PHÂN TÍCH CÚ PHÁP

① Bài toán phân tích cú pháp.

VP PNC : $\boxed{G = (V_T, V_N, P, S)}$

Xâu $w \in V_T^*$.

⊛ Phương pháp phân tích

- Ktra. xâu phân tích từ trái → phải.
- Phg pháp XD cây phân tích → Top-down: $S \overset{*}{\Rightarrow} w$
  → Bottom_up: $w \overset{*}{\Leftarrow} S$
- Phg pháp lựa chọn sx → Quay lui
  → Tất định

⊛ Phân tích trái.

- Là dãy sx được sd trong suy dẫn trái $S \to \alpha$
- Các sx đc đánh từ số từ $1 \ldots p$

VD VP: 1. $E \to T + E$

2. $E \to T$
3. $T \to F * T$
4. $T \to F$
5. $F \to (E)$
6. $F \to a$

Xét xâu $a*(a+a)$.

$E \overset{2}{\Rightarrow} T \overset{3}{\Rightarrow} F*T \overset{6}{\Rightarrow} a*T$
$\overset{4}{\Rightarrow} a*F \overset{5}{\Rightarrow} a*(E)$.
$\overset{1}{\Rightarrow} a*(T+E) \overset{4}{\Rightarrow} a*(F+E)$
$\overset{6}{\Rightarrow} a*(a+E)$.
$\overset{2}{\Rightarrow} a*(a+T)$
$\overset{4}{\Rightarrow} a*(a+F) \overset{6}{\Rightarrow} a*(a+a)$

$\boxed{2 3 6 4 5 1 4 6 2 4 6}$

⊛ Phân tích phải.

⇝ Tương tự phân tích trái.

$E \overset{2}{\Rightarrow} T \overset{3}{\Rightarrow} F*T \overset{4}{\Rightarrow} F*F \overset{5}{\Rightarrow} F*(E)$.
$\overset{1}{\Rightarrow} F*(T+E) \overset{2}{\Rightarrow} F*(T+T) \overset{4}{\Rightarrow} F*(T+F)$
$\overset{6}{\Rightarrow} F*(T+a) \overset{4}{\Rightarrow} F*(F+a) \overset{6}{\Rightarrow} F*(a+a)$
$\overset{6}{\Rightarrow} a*(a+a)$

$\boxed{2 3 4 5 1 2 4 6 4 6 6}$

② Phương pháp phân tích cú pháp quay lui

→ Top-down: Gốc → lá.
→ Bottom-up: PTích gạt nút gọn.

⊛ Phân tích Top-down.

VP PNC: $G = (V_T, V_N, P, S)$

$\forall$ sx $A \to \alpha_1 | \ldots | \alpha_n$ được đánh số 1, 2 ...

Step1: Khởi tạo.

- XD cây chỉ có 1 nút gốc S
  + S: Ký hiệu khởi đầu.
- Gọi kí hiệu cần phân tích là kí hiệu đầu tiên của xâu w.
- Gọi S là nút hoạt động

Step2: Tạo nút con (1 cách đệ quy)

- Chọn nút sx đầu tiên chưa đc áp dụng $A \to X_1 \ldots$
- Tạo k nút con trực tiếp với nhãn $X_1 \ldots X_k$.
  - $k > 0 \Rightarrow$ Lấy X_1 là nút h/đ
  - $k = 0 \Rightarrow$ (sx $A \to \varepsilon$) lấy nút bên phải của A là nút h/động ⇒ Quay lại t/h bước 2

Ghi chú: Nút hoạt động là nút ko k/thúc A.

- So sánh a với kí hiệu cần phân tích h/tại
  - ⊕ Nếu trùng → Nút h/đ là nút bên phải
    → Kh cần phân tích là kí tự t² trên xâu vào.
    → Quay lại b_2.
  - ⊕ K° trùng: Quay lui bước đã sd 1 sx và thử sx tiếp. (Hết khả năng → Quay lại bc trước)

Step3: ĐK dừng.

- AD hết khả năng k° tạo cây ⇒ Xâu k° được đoán nhận.
- Tạo cây suy dẫn cho xâu vào.

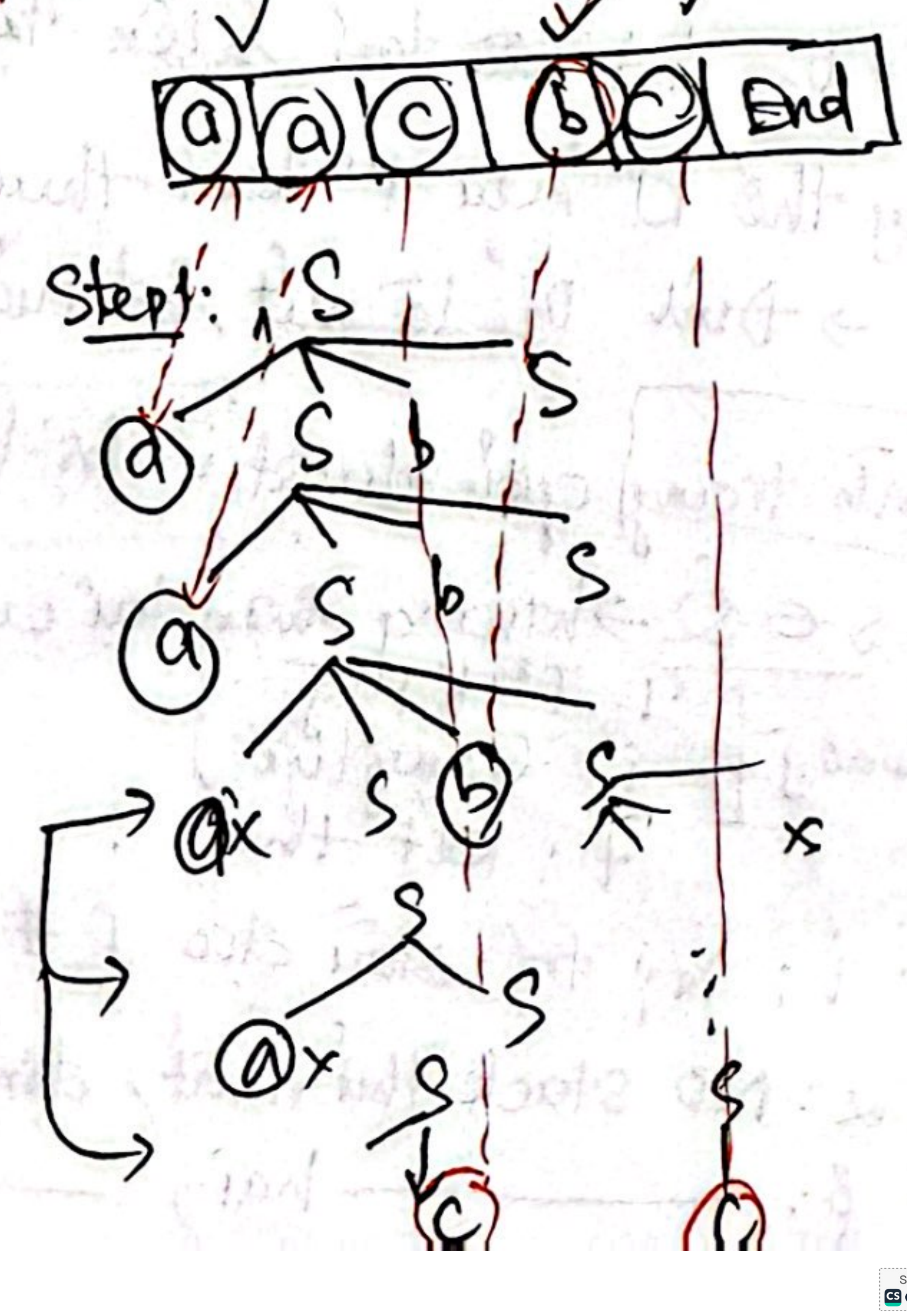
⊛ ĐK dừng. VP k° đệ quy trái

⊛ Cost
- Memory: $C * n$
- Thời gian: K^n

VD $S \to aSbS | aS | c$. Xâu $w = aacbc$

Đánh số sx.

1. $S \to aSbS$
2. $S \to aS$
3. $S \to c$



VD2. Phân tích top-down: ChonP.

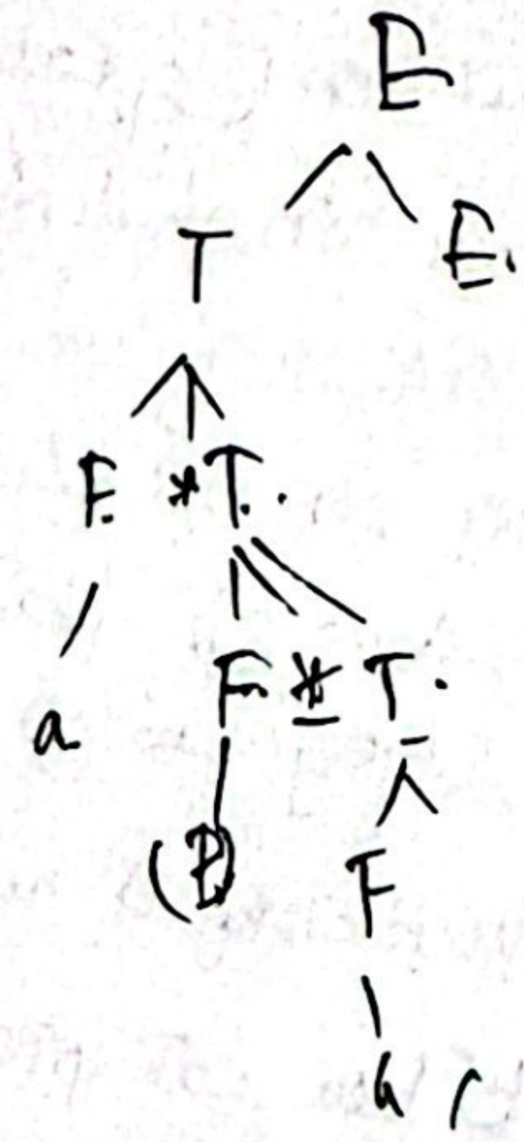
$E \to T + E \mid T$

$T \to F * T \mid F$ XĐ cây suy dẫn $a * a$.

$F \to (E) \mid a$.

a	*	a	End

1. $E \to T + E$
2. $E \to T$
3. $T \to F * T$
4. $T \to F$
5. $F \to (E)$.
6. $F \to a$.



⓪ Giải thuật phân tích Top-down quay lui

- Input:
  - VP PNC k° đệ quy trái
  - Xâu cần ptích $\omega = a_1 \ldots a_n$, $n \geq 0$
  - SX của G đánh số.
- Out:
  - Ptích trái cho ω
  - Thông báo lỗi

Quy ước: $\forall A \in V_N$, $A \to \alpha_1 \mid \alpha_2 \mid \ldots \mid \alpha_n$

Coi sx toàn: $A_1 \to \alpha_1$

$A_2 \to \alpha_2$

...

$A_n \to \alpha_n$

- Phương pháp: Dùng 2 stack D_1 và D_2

+ D_1: ghi lựa chọn và những kí hiệu ~~cần~~ đã phân tích.

+ D_2: bd câu trái hiện tại bằng ptích thay thế kí hiệu k° kết thúc vp tg ứng.

→ Đỉnh D_2 là nút hđ của cây tạo ra.

- Hình trạng giải thuật: Bộ bốn $\boxed{s, i, \alpha, \beta}$

+ $s \in Q$ → htrạng hiện tại của giải thuật.

bao g $\left[\begin{array}{l} q: T^2 \text{ b.thg} \\ b: \text{Quay lui} \\ t: \text{kết thúc} \end{array}\right.$.

+ i: Vị trí đầu đọc. (# kết thúc xâu bằng vào),

+ α: ND stack thứ nhất, đỉnh bên phải.

+ β: ——— hai, ——— trái.

Note: HTrạng ban đầu $(q, 1, \varepsilon, S)$

- Tthện giải thuật
  - + Từ hình trạng đầu → tính htrạng tiếp theo cho đến khi k° tính được nữa
  - + $\vdash$ chỉ 1 thay đổi htrạng.

VD. $(q, i, \alpha, \beta) \vdash (q', i', \alpha', \beta')$

+ Thực hiện theo các dịch chuyển.

- Giải thuật

① Mở rộng cây

$(q, i, \alpha, A\beta) \vdash (q, i, \alpha A_1, \gamma_1 \beta)$

$A \in V_N$, và $A_1 \to \gamma_1$ là lựa chọn đầu tiên của A

② Phù hợp vs kí hiệu vào.

$(q, i, \alpha, \underline{a}\beta) \vdash (q, i+1, \alpha \underline{a}, \beta)$

$a \in V_T$ ở đỉnh D_2 phù hợp vs kí hiệu vào -

⇒ Chuyển a sang D_1 và d/c đầu vào -

③ k° phù hợp vs kí hiệu đầu vào.

$(q, i, \alpha, a\beta) \vdash (b, i, \alpha, a\beta)$ // T^2 quay lui

④ Quay lui trên xâu vào.

$(b, i, \alpha a, \beta) \vdash (b, i-1, \alpha, a\beta)$ $a \in V_T$

Chuyển kí hiệu $D_1 \to D_2$ và d/c đầu đọc sang trái.

⑤ Thử lựa chọn tiếp theo.

$(b, i, \alpha A_J, \gamma_J \beta) \vdash (q, i, \alpha A_{J+1}, \gamma_{J+1} \beta)$ lựa chọn

- γ_{J+1} là sx tiếp theo.
- Nếu là sx cuối của A → quay lui tiếp
  → Thay γ_J bởi A và loại bỏ A_J trên đỉnh của D_1

$(b, i, \alpha A_J, \gamma_J \beta) \vdash (b, i, \alpha, A_J \beta)$

⑥ Hết khả năng quay lui $(i = 1, A \equiv S)$

→ k° đoán nhận.

⑦ KT thành công.

- $(q, n+1, \alpha, \#) \vdash (t, n+1, \alpha, \varepsilon)$

→ KL: Xâu đc đoán nhận.

- Tính hàm ptích trái → h(α)

phích trái $\int$ → $h(\alpha) = \varepsilon$ → kết luận k° KT
$h(A\,J) = \underline{p}$.

p là số hiệu sx liên hệ vs sx $A \to \gamma$
với γ là lựa chọn thứ i của A.

VD. Văn phạm 1. $S_1 \to aSb$
2. $S_2 \to c$.
w: aacbb.

? $w \in L(G)$? Có → Tìm PT trái.

Giải thích.

T^2 đầu:

$(q, 1, \varepsilon, S\#)$ Chưa làm gì, D_2 chứa S.

$\vdash (q, 1, S_1, aSb\#)$ ① $S_1 \to aSb$

$\vdash (q, 2, S_1a, Sb\#)$ ② Đỉnh D_2 là a, đầu đọc tại vị trí 1 là a → khớp.
⇒ Chuyển a sang D1. Dịch đầu đọc

$\vdash (q, 2, S_1aS_1, aSbb\#)$ ① Đỉnh D_2 là S → t^2 mở rộng

$\vdash (q, 3, S_1aS_1a, Sbb\#)$ ② Đỉnh D_2 là a. Đầu đọc tại vtrí 2 là a

$\vdash (q, 3\ S_1aS_1aS_1, \underline{a}Sbbb)$ ①.

$\vdash$ Đầu đọc ở vtrí 3 là c. Đỉnh D_2 là a. } → Không khớp → Quay lui.

$\vdash$ ③ $(b, 3, S_1aS_1aS_1, \underline{aSb}bb\#)$.

Thử tiếp lựa chọn: $S_2 \to c$.

$\vdash (b, 3, S_1aS_1aS_2, cbb\#)$

Đỉnh D_2 là c. Đầu đọc tại 3 là c } ⇒ khớp.

⇒ $\vdash (b, 4, S_1aS_1aS_2c, bb\#)$

Đỉnh D_2 là b. Đầu đọc tại vị trí 4 là b. } ⇒ khớp

$\vdash (b, 5, S_1aS_1aS_2cb, b\#)$.

$\vdash (b, 6, S_1aS_1aS_2cbb, \#)$

D_2 chỉ còn # → chuyển sang t^2 kết thúc

$\vdash (t, 6, S_1aS_1aS_2cbb, \varepsilon)$

⇒ $\boxed{h(S_1aS_2aS_2cbb) = 112}$.

⊕ Phân tích Bottom-up.

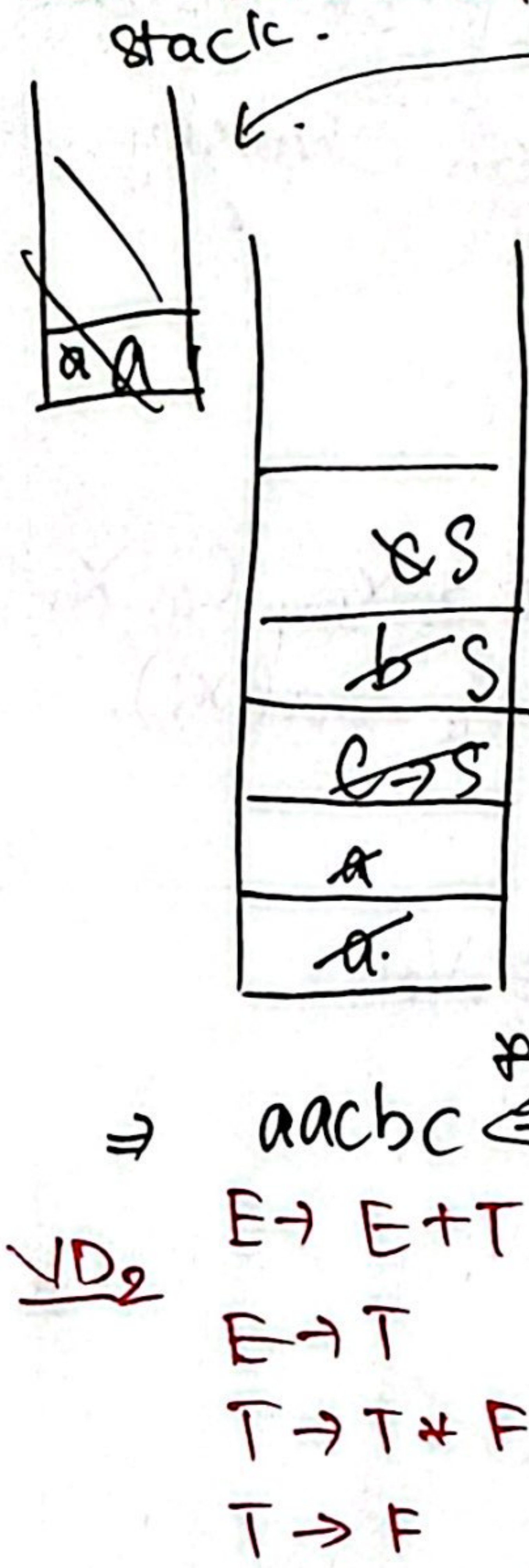
- Sd chu trình phân tích phải.
→ QT gạt - thu gọn.
- Sd stack S ⇒ chứa các ký hiệu của VP đã sinh ra 1 tiền tố nào đó trên xâu vào.

TT.
- Xét các xâu w trên đỉnh stack.
⊕ ∃ sx $A \to w \in P$ ⇒ thu gọn xâu w về A
*(N lựa chọn → đánh số để thử).
⊕ Nếu k° thu gọn → gạt kí hiệu tiếp theo xâu vào stack.
⊕ Đi hết xâu k° thu gọn → quay lui
⊕ Thuật toán dừng.
→ gạt hết kí hiệu và thu gọn về S
→ Thử hết TH mà k° thể thu gọn

VD. $S \to aSbS \mid aS \mid c$, $w = aacbb$.

stack.

a	a	c	b	c	End



$c \to S$.

$c \to S$. } $aS \to S$. } $aSb \to S$

⇒ $aacbc \overset{*}{\Leftarrow} S \in V$ ⇒ xâu được đoán nhận.

VD2
$E \to E+T$
$E \to T$
$T \to T*F$
$T \to F$
$F \to (E)$
$F \to a$.

Xâu $w = a*a$.

a	*	a	E

$a \to F$.

$a \to F \to T$. } $F*F \to T \in V$ ⇒ xâu được đoán nhận.

3. Phương pháp phân tích bảng.

- Cost → n^3 về thời gian (VP đơn ng: n^2 về Tg)
  → n^2 về bộ nhớ
- Phương pháp → CYK
  → Earley.

(*) Giải thuật CYK → xây dựng bảng

- Đk: văn phạm ở dạng chuẩn Chomsky (CNF).
  VP k° chứa ε-sxuất ($A \to \varepsilon$) và sx vô ích.
- Ng tắc: Tạo bảng phân tích ng tắc xuất các pt.
- Dạng chuẩn Chomsky.
  + sx có dạng $A \to BC$ ($A, B, C \in V_N$); $A \to a$ ($a \in V_T$)
  + Mọi VP phi ngữ cảnh k° chứa ε-sx đều chuyển về dạng chuẩn Chomsky.

⇒ Cách chuyển:

$\forall A \to X_1 \ldots X_n$ ($n \geq 2$)

(*) $X_i \in V_T$ ($X_i = a$) → $\begin{cases} A \to X_1 \ldots C_a \ldots X_n \\ C_a \to a\ (x_i). \end{cases}$

(*) $\forall A \to B_1 \ldots B_n$ ($n \geq 3$)

$\begin{cases} \exists D_1, D_2 \ldots D_{n-2} \in V_N. \\ A \to B_1 D_1. \\ D_1 \to B_2 D_2 \\ \vdots \\ D_{n-2} \to B_{n-1} B_n. \end{cases}$

VD: $S \to bA \mid aB$.
$A \to bAA \mid aS \mid a$.
$B \to aBB \mid bS \mid b$.

$S \to bA$: $\begin{cases} S \to C_b A \\ C_b \to b. \end{cases}$

$S \to aB$: $\begin{cases} S \to C_a B. \\ C_a \to a. \end{cases}$

$A \to bAA$: $\begin{cases} A \to C_b AA. \\ A \to C_b D_1 \\ D_1 \to AA. \end{cases}$

$A \to aS$: $A \to C_a S$.

$B \to aBB$: $\begin{cases} B \to C_a BB \\ B \to C_a D_2 \\ D_2 \to BB. \end{cases}$

$B \to bS$: $B \to C_b S$.

Tóm lại:

$S \to C_b A \mid C_a B$
$A \to C_b D_1 \mid C_a S \mid a.$
$B \to C_a D_2 \mid C_b S \mid b.$
$D_1 \to AA$ $D_2 \to BB$
$C_a \to a$ $C_b \to b$

- Giải thuật CYK.

Bảng DLT → Ma trận Δ dưới.

T_{ij} → i cột ; j hàng.

$\forall T_{ij} \supseteq V_N$.

Độ dài bảng T

$A \in T_{ij} \Leftrightarrow A \Rightarrow^+ a_i a_{i+1} \ldots a_{i+j-1}$.
(j là xâu kế từ vị trí i).

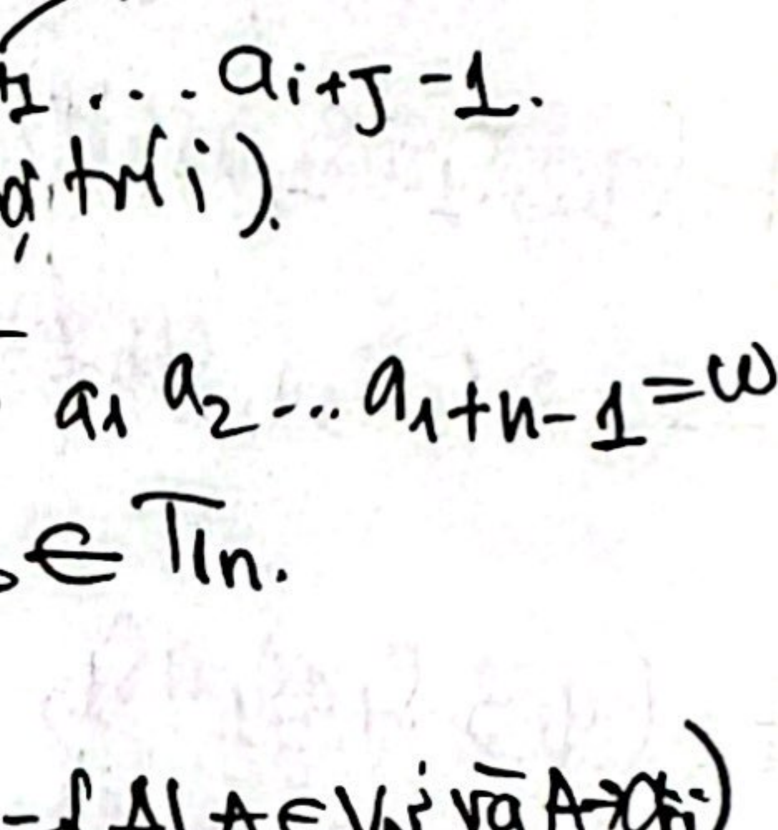
VD. $S \in T_{1n} \Rightarrow S \Rightarrow^+ a_1 a_2 \ldots a_{1+n-1} = \omega$

$\omega \in L(G) \Leftrightarrow S \in T_{1n}$.

$\boxed{\text{Xây dựng}}$

1. XĐ pt cho hàng 1: $T_{i1} = \{A \mid A \in V_N$ và $A \to a_i\}$
2. Đã XĐ xong cho j-1 hàng:
   $T_{ij} = \{A \mid \exists k: 1 \leq k < j$ thỏa mãn $B \in T_{ik}; C \in T_{i+k, j-k} : A \to BC \in P\}$
   $A \in T_{ij}$: $A \to BC \Rightarrow^+ \ldots = a_i \ldots a_{i+j-1}$
3. Lặp lại b2

VD. $S \to AA \mid AS \mid b$
$A \to SA \mid AS \mid a$.
ω: abaab



J					
5	A,S				
4	S,A	A,S			
3	A,S	S	S,A		
2	S,A	A	S	S,A	
1	A	S	A	A	S
i	a	b	a	a	b

T_{12}: $k = 1$ $\{ {22 \atop 41}$

$A \in T_{11}$ $\{ {32 \atop 21}$

$S \in T_{21}$

→ $S \to AA$
→ $A \to SA$ (tìm)
→ $\boxed{\text{Đưa } S, A}$

T_{22}: k = 1. $S \in T_{21}$, $A \in T_{31}$.
Có $A \to SA$ → Điền $\boxed{A}$.

T^2 cho T_{32}, T_{42}

T_{13}: $k=1 \to \begin{cases} T_{11} \ \{A\} \\ T_{22} \ \{A\} \end{cases}$ ∃ $S \to AA$. $\boxed{S}$

~~$k=2 \to \begin{cases} T_{12} \ \{S\} \\ T_{31} \ \{A\} \end{cases}$ ∃ $S, A \to SA$~~

$k=2 \to \begin{cases} T_{12} \ \{S, A\} \\ T_{31} \ \{A\} \end{cases}$ ∃ $A \to SA | AA$ $\boxed{S}\boxed{A}$.

T_{23}: $1 \le k < 3$

$k=1 \to \begin{cases} T_{21} \ \{S\} \\ T_{32} \ \{S\} \end{cases}$ ∄.

$k=2 \to \begin{cases} T_{22} \ \{A\} \\ T_{41} \ \{A\} \end{cases}$ ∃ $S \to AA$. → $\boxed{S}$

T_{33} $1 \le k < 3$.

~~$k=1 \to \begin{cases} T_{31} \\ T_{42} \end{cases}$~~

$\begin{cases} T_{32} \ T_{51} \ \{S\} \\ T_{51} \ T_{32} \ \{S\} \end{cases}$ → ∄ $X \to SS$.

$\begin{cases} T_{31} \ \{A\} \\ T_{42} \ \{S, A\} \end{cases}$ ∃ $X \to SA | AA$ $\boxed{S \text{ và } A}$.

T_{14} $1 \le k < 4$.

→ $k=1 \to \begin{cases} T_{11} \to A \\ T_{23} \to S \end{cases}$ $\boxed{S \to AS}$

$k=2 \to \begin{cases} T_{12} \to S, A \\ T_{32} \to S \end{cases}$ ∄ $X \to SS$. $\boxed{∃ S, A \to AS}$ → $\boxed{S, A}$.

$k=3 \to \begin{cases} T_{13} \to S, A \\ T_{41} \to A \end{cases}$ ∃ $S \to AA$ → $\boxed{A}$.

T_{24}: $k=1 \to \begin{cases} T_{21} \to S \\ T_{33} \to S, A \end{cases}$ ∃ $A \to SA$ $\boxed{A}$

$k=2 \to \begin{cases} T_{22} \to A \\ T_{42} \to S, A \end{cases}$ $\boxed{A, S}$.

$k=3 \to \begin{cases} T_{23} \\ T_{51} \end{cases}$

(*) GT CYK → Tìm cây phân tích.

– Ngược với CYK.

– Gen(i, j, A) sinh ra dãy phân tích trái ứng với suy dẫn $A \Rightarrow^*_L a_i \dots a_{i+j-1}$.

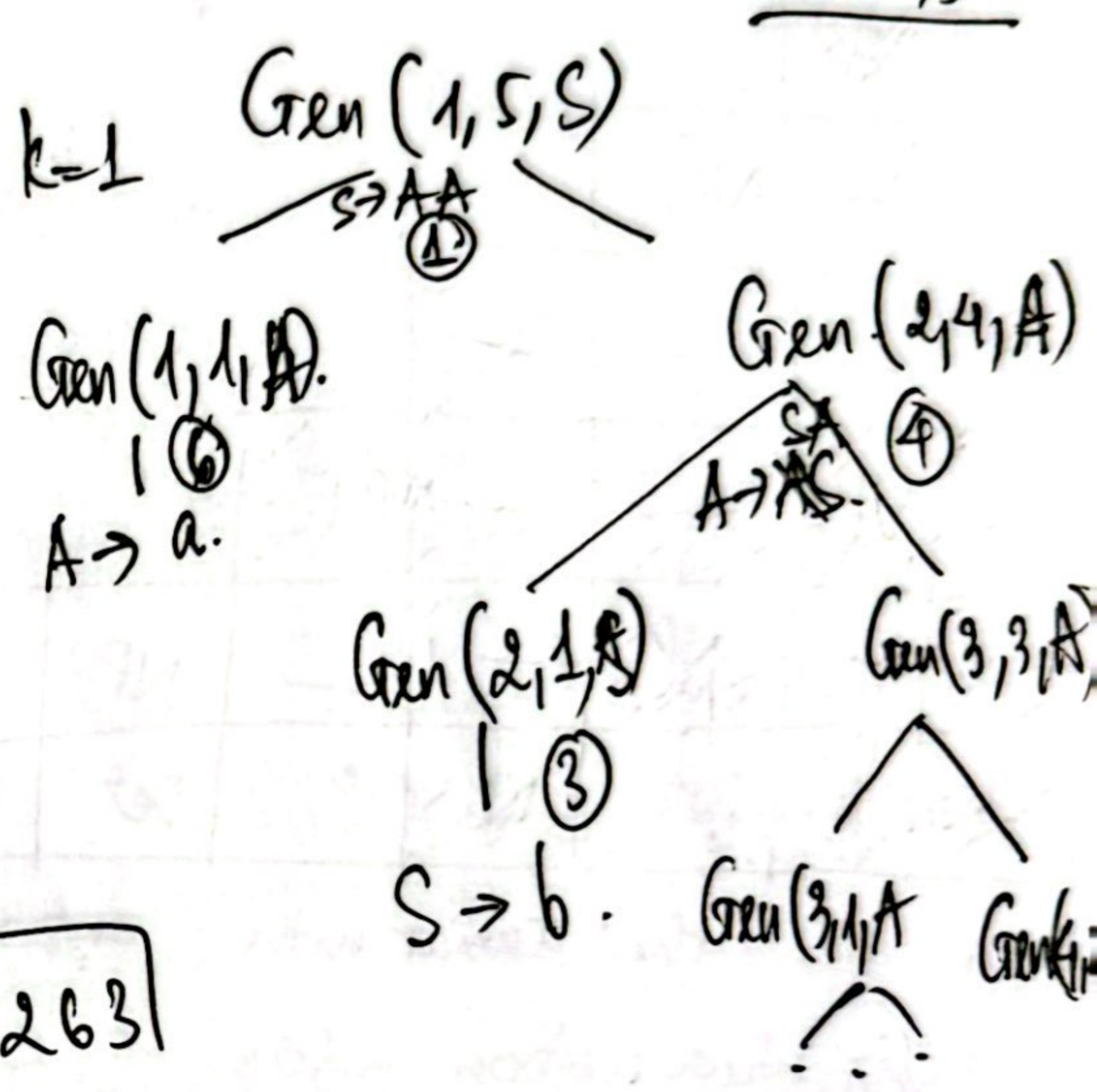
+ $j = 1$ → $A \to a_i$ là 1 sx của m → Viết m

+ $j > 1$: Nếu $k_{min} \in Z$. thì $\begin{cases} B \in T_{ik} \\ C \in T_{i+k, j-k} \\ A \to BC \end{cases}$

⇒ $A \to BC$ là sx m. ⇒ $\begin{cases} \text{Viết m.} \\ \text{gọi } Gen(i, k, B) \\ Gen(i+k, j-k, B) \end{cases}$

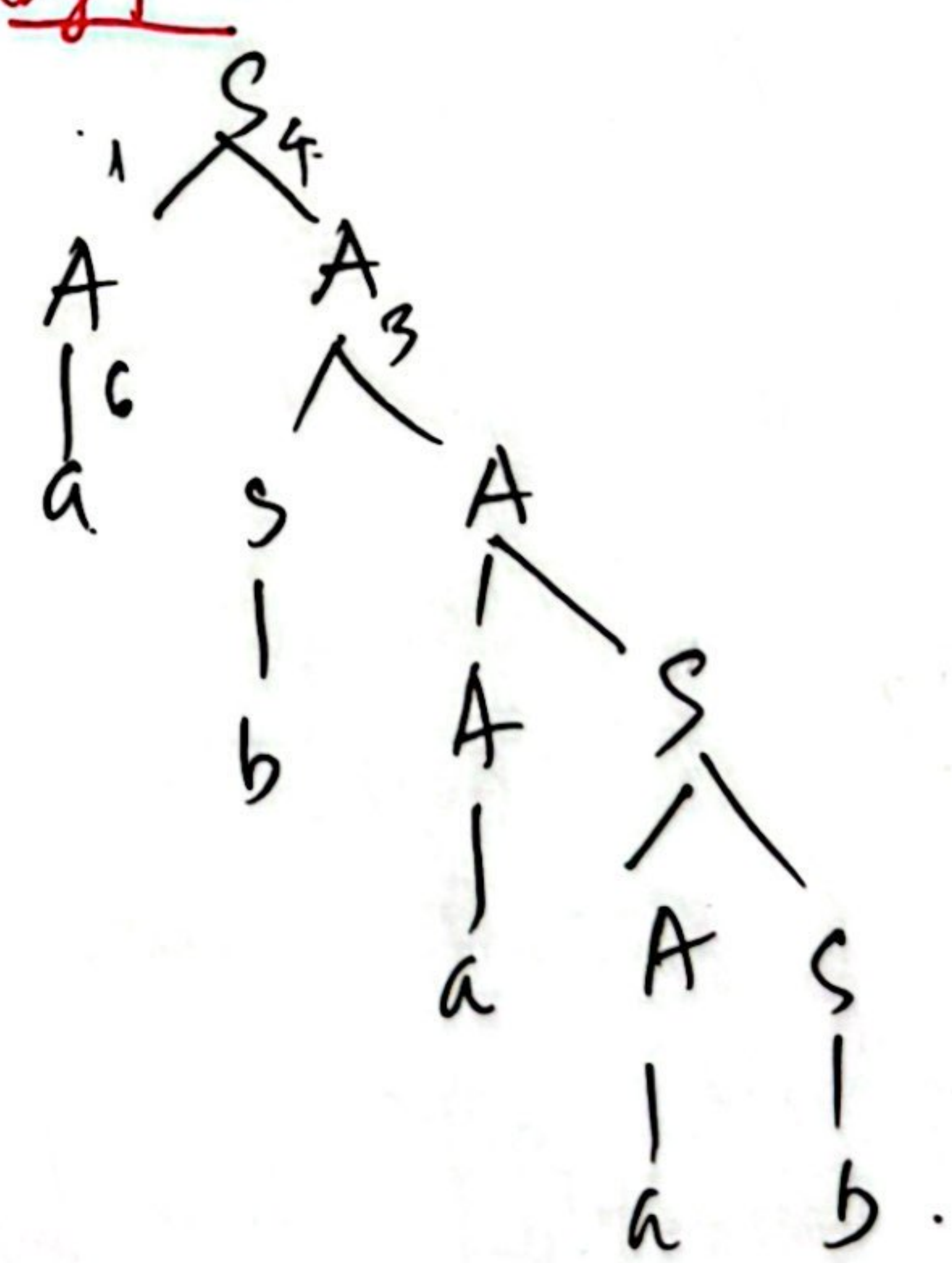
VD.
1. $S \to AA$
2. $S \to AS$
3. $S \to b$
4. $A \to SA$
5. $A \to AS$
6. $A \to a$

w: abaab.



$\boxed{164356263}$

Cây phân tích



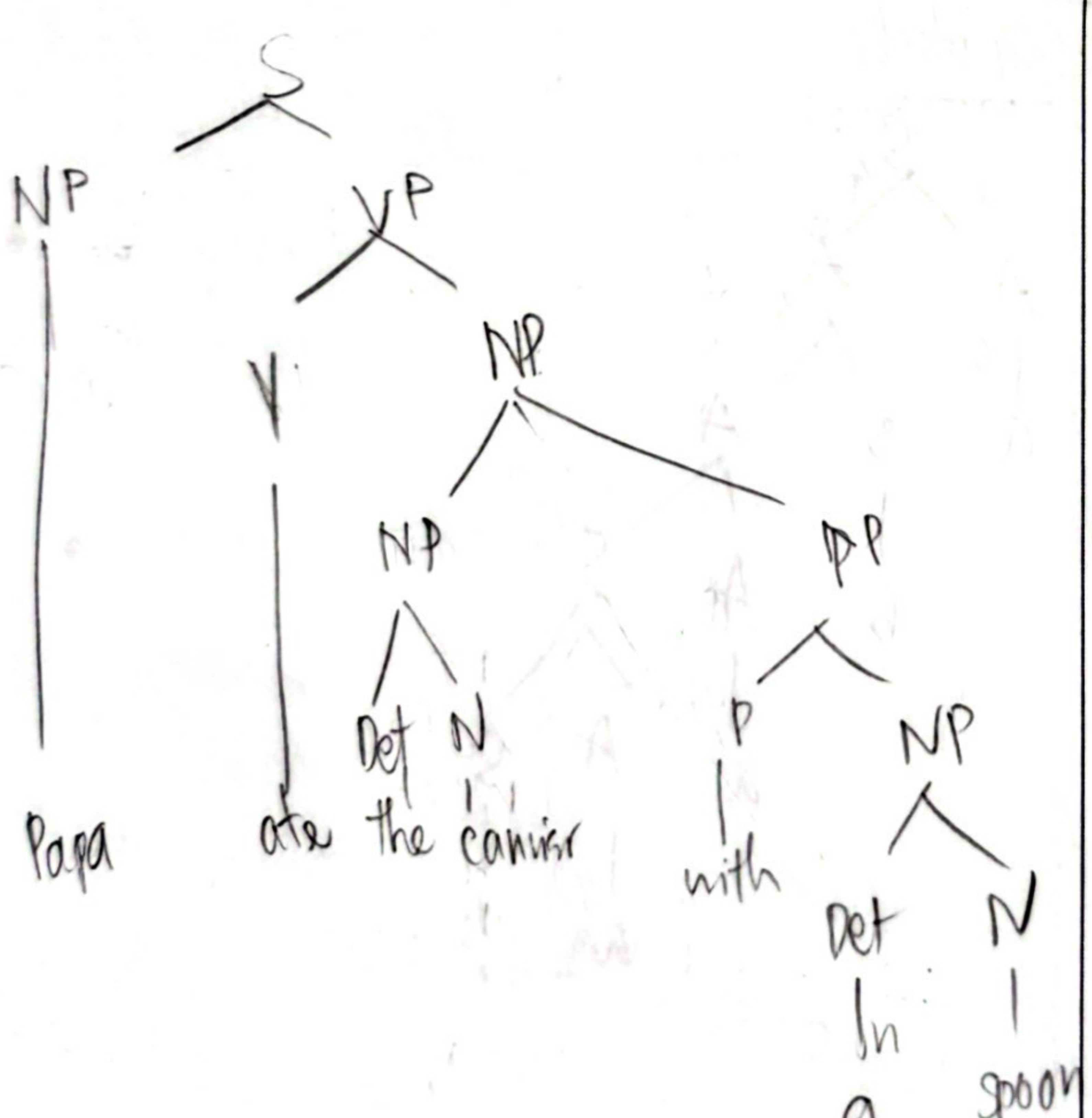
VD1. 1. S → NP VP
2. NP → Det N | NP PP | Papa.
3. VP → VP PP | V NP
4. PP → ~~P NP~~. P NP.
5. N → caviar | spoon
6. V → ate
7. P → with
8. Det → the | a.

w: Papa ate the caviar with a spoon

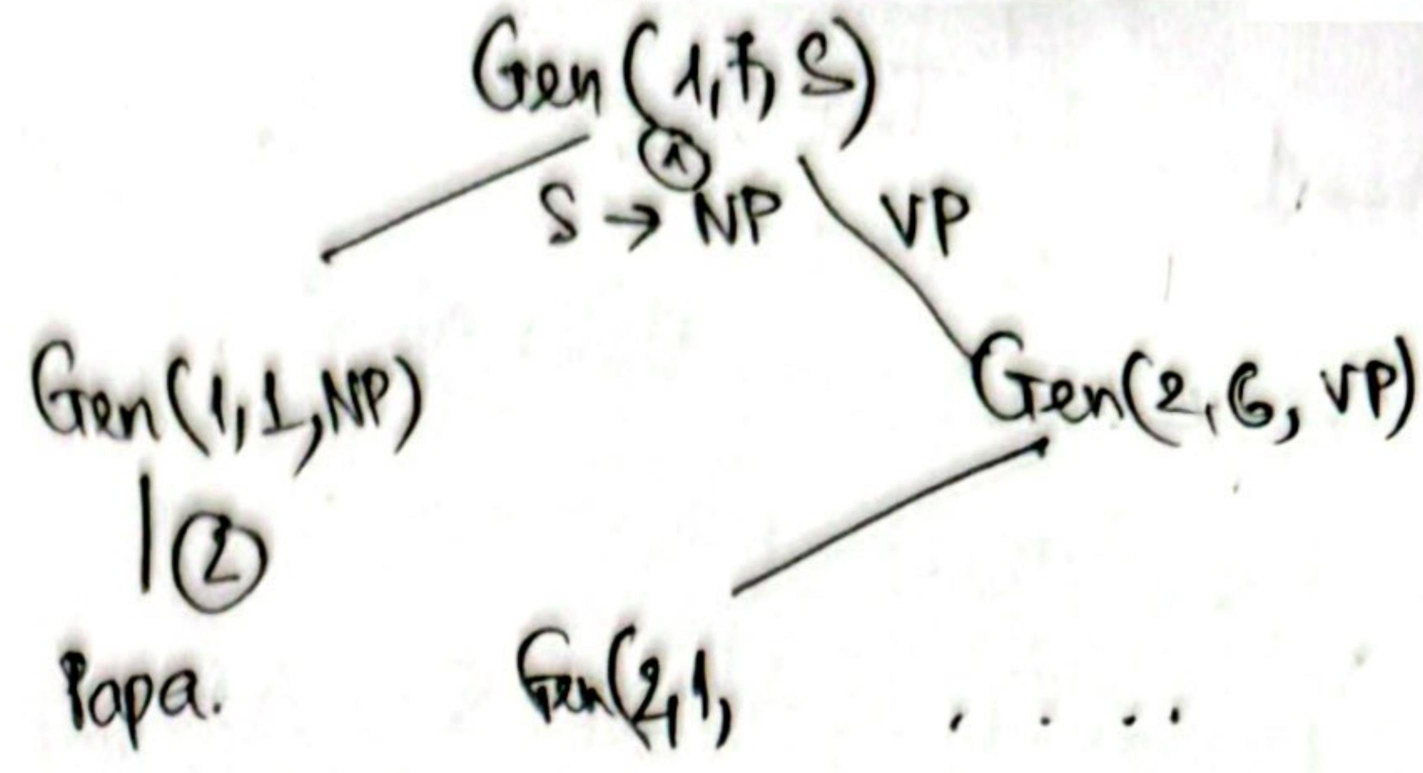
S						
—	VP					
—	—	NP				
S	—	—	—			
—	VP	—	—	PP		
—	—	NP	—	—	NP	
NP	V	Det	N	P	Det	N
Papa	ate	the	caviar	with	a	spoon

→ Xâu được đoán nhận.

C1



C2. Thầy chữa.



④. Phương pháp phân tích cú pháp bất định

- KN: + xem xét 1 lượt xâu từ trái → phải
  Mỗi bước XĐ được duy nhất 1 sx phù hợp vs trạng thái hiện tại.

- VD: S → aA, A → c | bA.
  w: abbc.

Kí hiệu	Xâu cần phân tích	SX
S	abbc	S → aA.
aA	abbc	So sánh đầu đọc
A	bbc	A → bA.
bA	bbc	S²
A	bc	A → bA
bA	bc	A → c.
~~A~~c	c	So sánh.
—	—	

VD2. Có thể bị quay lui (slide T55)
→ Giải pháp: Giới hạn trong 1 số lớp ngôn ngữ: LL(k), LR(k), ...
Left-to-left scanning   Left-to-right scanning.

④ Phân tích topdown.
- Ngôn ngữ LL(k)
  + Khi th phân tích LL(k) → cần nhìn trước k kí hiệu để qđ sản xuất nào được sd.
VP LL(k) cần tìm 1 số đk.

LL(k)
→ Số kí hiệu cần nhìn thấy trước
→ Thực hiện suy diễn trái nhất
→ Cần được phân tích từ trái → phải.

~~FIRST(k)~~ $G(V_T, V_N, P, S) \to$ VPPNC.

- $\boxed{FIRST_k(\alpha)}$ $\{k \in Z^+, \alpha \in V^*, \beta \in V^*$

$= \{x \mid x \in V_T^*, l(x) = k, \alpha \Rightarrow_L^* x\beta\}$.

$= \{x \mid x \in V_T^*, l(x) < k, \alpha \Rightarrow_L^* x\}$.

Đn: $FIRST_k(\alpha)$ là tập hợp các xâu x gồm k kí hiệu kết thúc trái nhất của xâu được suy dẫn ra từ α.

(Case: x k° còn đủ k kí hiệu nhưng $\alpha \Rightarrow x$, k° còn kí hiệu sau x → ~~Được~~ chấp nhận).

$S \Rightarrow_L^* x\delta$

VD

- $\boxed{FOLLOW_k(\alpha)} = \{x \mid S \Rightarrow_L^* \beta\alpha\delta,$
$x \in FIRST_k(\delta)\}$

Đn: $FOLLOW_k(\alpha)$ là tập hợp các xâu chứa k kí hiệu KT, đứng ngay đằng sau α trong bất kì dạng câu nào.

Note: $\alpha \equiv A \in V_N$

$S \Rightarrow_L^* \beta A$, $FOLLOW_1(\alpha) \ni \{\$\}$.

$\to S \to ahSb \mid agb$

VD. Cho văn phạm $S \to \underline{a}Ab \mid \underline{c}$.
$A \to hS \mid g$.

ahaAbb ~~S→ah~~

Tính $First_1(S)$

$k = 1$

$First_1(S) = \{a, c\}$ → $S \to \underline{a}Ab \mid \underline{c}$

$Follow_1(S) = \{b\}$

→ $A \to h\underline{S}$ (k° có kí h nào kết thúc sau S)
$S \to aA\underline{b}$ (sau A là b

$First_1(A) = \{h, g\}$ $First_2(A) = \{ha, g, hc\}$

$Follow_1(A) = \{b\}$. $Follow_2(A) = \{b, bb\}$

$First_2(S) = \{ah, ag, c\}$

$Follow_2(S) = \{b, bb\}$

- Sx uất $\boxed{LL(k)}$

VPPNC G: 1 sx từ A ∈ V_N là sx LL(k) nếu TM

① $S \Rightarrow_L^* xA\alpha \Rightarrow_L x\beta_1\alpha \Rightarrow_L^* xZ_1$ ($x \in V_T^*$

② $S \Rightarrow_L^* xA\alpha \Rightarrow_L x\beta_2\alpha \Rightarrow_L^* xZ_2$.

③ $FIRST_k(Z_1) = FIRST_k(Z_2)$ thì $\beta_1 = \beta_2$

Đn: VPPNC G là LL(k) nếu ∀sx của G là sx của LL(k).

Dc: G → LL(k) | ⇒ k kí hiệu kết thúc suy ra từ $A\alpha$ thì chỉ ∃ 1 sx từ A thoả mãn.
$xA\alpha$ → Dạng câu

VD1. $S \to aAS \mid b$ là VP LL(1)
VP $A \to bSa \mid a$.

Cm:

+) $S \to aAS \mid b$.

Suy dẫn trái: $S \Rightarrow_L^* xS\alpha \Rightarrow x\beta_1\alpha \Rightarrow_L^* xZ_1$

$\Downarrow$

$x\beta_2\alpha \Rightarrow xZ_2$

Các sx cùng 1 non-terminal k° có FIRS giao nhau

Nếu có ε trong FIRST, FIRST ∩ FOLLOW

~~$FIRST_1(S) = \{a, b\}$~~

~~$S \Rightarrow aAS$~~:

$FIRST_1(aAS) = a \neq FIRST_1(b) = b$

+) $A \to bSa \mid a$.

$FIRST_1(bSa) = b$
$FIRST_1(a) = a$
⇒ k° giao nhau.

→ ok.

~~Theo slide~~ Lý thuyết tính tất.

$A \to \alpha \mid \beta$. ⇒ $FIRST_1(\alpha) \cap FIRST_1(\beta) =$

∃ ε: $FIRST_1(\beta) \cap FOLLOW($
$= \emptyset$

VD2 $S \to aAa \mid bAba$
$A \to b \mid \varepsilon$.
Cm k° phải văn phạm LL(1)

Xét sx $S \to b\underline{A}ba$.

Nếu 8 kí hiệu KO KT A, k° biết suy dẫn

$A \to b$ hay $A \to \varepsilon$.

là văn phạm LL(2)

TH1: $S \to \underline{aAa}$, ~~nhìn trc ba~~

_ ~~Phích aAa~~. Nhìn trước ba $\to \boxed{A \to b}$.

_ Nhìn trước a # $\to \boxed{A \to \varepsilon}$

TH2. Ngữ cảnh bAba.

\+ Nhìn thấy bb $\Rightarrow$ AD $A \to b$

\+ Nhìn thấy ba $\to A \to \varepsilon$.

$\Rightarrow$ Điều kiện văn phạm là LL(k)

G_LL(k)

(+) $xA\alpha$ một dạng câu ($S \overset{*}{\underset{L}{\Rightarrow}} xA\alpha$)

$A \to \beta_1 \in P$
$A \to \beta_2 \in P$ $(\beta_1 \neq \beta_2)$.

thì $FIRST_1(\beta_1 a) \cap FIRST_1(\beta_2 a) = \emptyset$

Tóm lại

$\boxed{FIRST(\alpha)} = \{a \mid a \in V_T, \alpha \overset{*}{\underset{L}{\Rightarrow}} a\beta\}$ hoặc

$= \{\varepsilon \mid \alpha \overset{*}{\underset{L}{\Rightarrow}} \varepsilon\}$

Vc: $a \in V_T \Rightarrow FIRST(a) = \{a\}$.
$FIRST(\varepsilon) = \{\varepsilon\}$

$A \to a\gamma \Rightarrow FIRST(A) \supseteq \{a\}$.

$A \to B\gamma \Rightarrow FIRST(A) \supseteq FIRST(B) - \{\varepsilon\}$

~~TH đặc biệt: $B \overset{*}{\underset{L}{\Rightarrow}} b\beta$~~

$\begin{cases} A \to B\gamma \\ B \overset{*}{\Rightarrow} \varepsilon \end{cases}$ $FIRST(A) \supseteq (FIRST(B) - \varepsilon)$
$\cup FIRST(\gamma)$.

$FOLLOW(A) = \{a \mid a \in V_T, S \overset{*}{\underset{L}{\Rightarrow}} \alpha A a \beta\}$

Note: Nếu A xuất hiện bên phải nhất của 1 câu ($S \overset{*}{\underset{L}{\Rightarrow}} \alpha A$) thì $\boxed{FOLLOW(A) \supseteq \{\varepsilon\}}$
$FOLLOW(A) \supseteq \{\#\}$.

Vc:

\+ $A \to \gamma B \alpha$ $\boxed{Follow(B) \supseteq FIRST(\alpha) - \{\varepsilon\}}$

\+ $A \to \gamma B$

$[A \to \gamma B \beta \ (\beta \overset{*}{\underset{L}{\Rightarrow}} \varepsilon)$

$\boxed{FOLLOW(B) \supseteq FOLLOW(A)}$

BT: Cho văn phạm $E \to T \mid E + T$
$T \to F \mid T * F$
$F \to a \mid (E)$

(1)

\+ $FIRST(F) = \{a, (\}$.

~~$FIRST(T) = FIRST(F) = \{$~~

\+ $T \to F \Rightarrow FIRST\{T\} \supseteq FIRST(F) - \{\varepsilon\}$
$= \{a, (\}$.

\+ $E \to T \Rightarrow FIRST(E) \supseteq FIRST(E) - \{\varepsilon\}$
$= \{a, (\}$

(2) Chứng E là start symbol $\Rightarrow \# \in FOLLOW(E)$

\+ $FOLLOW(E) = \{+,), \#\}$.

$E \to E + T \to T + T \to F + F \to (E). + (E)$

\+ ~~$F \to F * \mid T \to T * F$~~

~~$\Rightarrow FOLLOW(T) = FOLLOW$~~

Có: ~~$E \to T$~~

\+ $E \to T \Rightarrow FOLLOW(T) \supseteq FOLLOW(E)$
$= \{+,), \#\}$.

Xét luật $T \to T * F$.

$FOLLOW(T) = FOLLOW(T) \cap \{*\} = \{+,), *, \#\}$

\+ $T \to F \Rightarrow FOLLOW(F) \supseteq FOLLOW(T)$

(8)

* Toán tử $\oplus$

V_1, V_2 là 2 tập kí hiệu.

$\boxed{\text{Định nghĩa}}$ $V_1 \oplus V_2 = V_1$ nếu $\varepsilon \notin V_1$

$= (V_1 - \{\varepsilon\}) \cup V_2$ nếu $\varepsilon \in V_1$

VD.

$\{a,b\} \oplus \{c,d\} = \{a,b\}$.

$\{\varepsilon, a, b\} \oplus \{c,d\} = \{a,b,c,d\}$.

$\{\} \oplus \{a,b\} = \{\}$

$\boxed{\text{Bổ đề.}}$

$$\boxed{FIRST(\alpha\beta) = FIRST(\alpha) \oplus FIRST(\beta).}$$

\+ $\varepsilon \notin$, $a \in FIRST(\alpha)$

$FIRST(\alpha) \oplus FIRST(\beta) \ni a$.

$\alpha \Rightarrow^* ay$

$$\Rightarrow \boxed{FIRST(\alpha\beta) = FIRST(ay\beta) \ni a.}$$

\+ $\varepsilon \in FIRST(\alpha)$

$FIRST(\alpha) \oplus FIRST(\beta) = (FIRST(\alpha) - \{\varepsilon\}) \cup FIRST(\beta)$

$$\Rightarrow \boxed{FIRST(\alpha\beta) = (FIRST(\alpha) - \varepsilon) \cup FIRST(\beta)}$$

$X_i \in V$

$FIRST(X_1 \ldots X_n) = FIRST(X_1) \oplus \ldots \oplus FIRST(X_n)$

* Thuật toán tính $FIRST(A)$

XD các tập $F_i(X)$ với $X \in V$

1. $a \in V_T$ $F_i(a) = \{a\}$ $\forall i$.
2. $F_0(A) = \{a \mid a \in V_T$ và $A \to a\alpha \in P\}$ $\cup \{\varepsilon \mid A \to \varepsilon \in P\}$.
3. Giả thiết tính và $\exists F_0, \ldots, F_{i-1}$

$F(A) = \{a \mid a \in V_T$, if $A \to X_1 \ldots X_n \in P$ thì $a \in F_{i-1}(X_1) \oplus F_{i-1}(X_2) \oplus \ldots \oplus F_{i-1}(X_n)$.

4. Thuật toán dừng khi các F_i k° đổi.

VD. $E \to T \mid E + T$

$T \to F \mid T * F$

$F \to a \mid (E)$.

X	$F_0(X)$	$F_1(X)$	$F_2(X)$	F_3
E.	ϕ	ϕ.	$\{a, (\}$	$\{a, (\}$
T	ϕ	$\{a, (\}$	$\{a, (\}$	k° đổi
F	$\{a, (\}$	$\{a, (\}$	~~k° đổi~~	

Bước 1. Khởi tạo F_0.

$A \to$ Kí hiệu KT $\Rightarrow$ Thêm vào $F_0(A)$

$\left\{\begin{array}{l} F \to a \\ F \to (E) \end{array}\right. \Rightarrow F_0(F) = \{a, (\}$

Bước 2: F_1 ghép

~~$E \to T \Rightarrow FIRST(E) \supseteq FIRST(T) = \phi$.~~

~~$T \to F \Rightarrow FIRST(T) \supseteq FIRST$~~

~~$F_1(E) = F_0(T) = \{\}$ và~~ $E \to T$).

$E \to T \Rightarrow F_1(E) = F_0(T) = \{\phi\}$.

$E \to E + T \Rightarrow F_1(E) = F_0(E) \oplus F_0(+) \oplus F_0(T)$

$= \phi$.

$T \to F \Rightarrow F_1(T) = F_0(F) = \{a, (\}$.

$T \to T * F \Rightarrow F_1(T) = F_0(T) \oplus F_0(*) \oplus F_0(F)$

$= \{\} \oplus \{*\} \oplus \{a, (\}$

$= \phi$.

...

Bước 3. Tính F_2.

$F_2(E) = F_1(T) = (a, ($

$= \underline{F_1(E)} \oplus F_1(+) + F_1(T)$

$= \phi \oplus \{+\} \oplus (a, ($.

Thuật toán tính FOLLOW(A).

Thực hiện cho đến khi k° còn xuất hiện thay đổi trong tập FOLLOW

1. Đặt # vào tập FOLLOW(S)

⊕ S là kí hiệu khởi đầu.

⊕ # ∉ V, dùng để đánh dấu chuỗi kết thúc câu phích.

② ∀ sx $A \to \alpha B \beta$ thêm FIRST(β) − {ε} vào FOLLOW(B)

$\boxed{\text{có kí hiệu k° KT đứng trc 1 kí hiệu \#.}}$

③ ∃ $A \to \alpha B$

[$A \to \alpha B \beta$, {ε} ⊆ FIRST(β) ⇒ Thêm FOLLOW(A) vào FOLLOW(B)

→ $\boxed{\text{kí hiệu k° kết thúc ở cuối}}$

VD. $E \to T \mid E+T$ FIRST(E)

~~FT~~ $T \to F \mid T*F$

$F \to a \mid (E)$.

Quy tắc	sx áp dg	E	T (FOLLOW)	F
~~Khởi tạo~~ 1	Khởi tạo	#	φ	φ
2	E→E+T	+, # Thêm FIRST(+) ~~−{ε}~~ FIRST(+)	φ	φ
	T→T*F		*	
	F→(E)	+,), #		
3	E→T ~~E→E+T~~		*, +,), #	
	T→F ~~F→T*F~~.			*, +,), #
FOLLOW(X)		), +, #	*, +,), #	*, +,), #

VD2. $S \to BA$

$A \to +BA \mid \varepsilon$

$B \to DC$

$C \to *DC \mid \varepsilon$

$D \to (S) \mid a$.

Tính FIRST(X)

FOLLOW(X)

⊕ Tính FIRST(X)

X	F0(X)	F1(X)	F2(X)	F3(X)	FIRST
S	φ	φ.	~~k° đổi~~ (, a.		(, a.
A	ε, +	+, ε	k° đổi		+, ε
B	φ	(, a	(, a	k° đổi	(, a
C	*	*, ε	*, ε	k° đổi	*, ε.
D	(, a	(, a	k° đổi		(, a

⊕ Tính FOLLOW(X)

Quy tắc	SX	S	A	B	C	D
1	Khởi tạo	#	φ	φ	φ	φ
2	S→BA			+		
	A→+BA			+.		
	B→DC					*
	C→*DC					*
	D→(S)	#,)				
3	S→BA		#,)			
	A→+BA ~~A→+BA~~ ~~B→DC~~ A→ε			#,), +	*	
	~~C→*+~~ B→DC				#,), +	
	C→*DC. C→ε					*, #,), +
FOLLOW(X)		#,)	#,)	#, +,)	#, +,)	*, #,), +

⊕ Điều kiện để là văn phạm LL(1)

- $\forall$ sx $A \to \alpha_1 | \dots | \alpha_n$, $n \geq 2$ thì:

$FIRST(\alpha_i) \cap FIRST(\alpha_j) = \phi, \forall i \neq j$

- $\forall A \in V_N$, $A \to \alpha_1 | \dots | \alpha_n$ và $\exists \alpha_j \Rightarrow^* \varepsilon$

$FIRST(\alpha_i) \cap FOLLOW(A) = \phi \ \forall i \neq j$

VPLL(1) đơn giản:

~~$A \to \alpha_1\alpha_2 | \alpha_2\alpha_3 | \dots | \alpha_n$~~

$$\boxed{\begin{array}{l} A \to a_1\alpha_1 | a_2\alpha_2 | \dots | a_n\alpha_n. \\ a_i \in V_T,\ a_i \neq a_j \text{ với } i \neq j. \end{array}}$$

VD. 1. $E \to T | E + T$. Bắt đầu =.

2. $T \to F | T * F$. Kết thúc.

3. $F \to a | (E)$.

→ $F \to a | (E)$ là văn phạm LL(1) đ/giản

→ Xét sản xuất (1).

$FIRST(T) = \{a, (\}$

$FIRST(E+T) = \{a, (\}$

→ $FIRST(T) \cap FIRST(E+T) \neq \phi$.

⇒ Đây ko phải VPLL(1)

VD2. 1. $S \to BA$.

2. $A \to +BA | \varepsilon$.

3. $B \to DC$.

4. $C \to *DC | \varepsilon$

5. $D \to (S) | a$

⊕ (1), (3) là các sx đơn.

⊕ (5) là văn phạm LL(1) đơn giản.

⊕ Xét sx(2).

$FIRST(+BA) = \{+\}$
$FOLLOW(A) = \{\#,)\}$ ⇒ ok

⊕ Xét sx(4).

$FIRST(*DC) = \{*\}$.
$FOLLOW(C) = \{(,), *\}$ ⇒ ok

⊕ Khử đệ quy trái.

- VPLL(1) ko đệ quy trái.

→ Khử

- Đệ quy trái (trực tiếp)

sx: $A \to A\alpha_1 | \dots | A\alpha_n | \beta_1 | \dots | \beta_n \ \forall \beta_i[1] \neq$

⊕ Đệ quy trái đúng khi áp dụng:

$A \to \beta_1 | \beta_2 | \dots | \beta_n$.

⊕ Từ A sinh ra xâu dạng dạng:

$(\beta_1 | \dots | \beta_m)(\alpha_1 | \dots | \alpha_n)^*$.

⊕ Xâu trên sinh ra từ văn phạm:

$A \to \beta_1 A' | \beta_2 A' | \dots | \beta_m A'$.

$A' \to \varepsilon | \alpha_1 A' | \dots | \alpha_n A'$.

VD1: $A \to Aa | b$

⇒ Đây là đệ quy trái trực tiếp

⇒ Văn phạm ko đệ quy trái.

$$\begin{cases} A \to bB. \\ B \to \varepsilon | aB^* \end{cases}$$

⇒ Ngôn ngữ sinh ra từ 2 văn phạm: ba^*

Note: - Loại bỏ đệ quy trái → Cây phân tích thay đổi

- Cần chức lại cây phân tích → cấu trúc gốc.

VD2: $E \to T | E + T$.

$T \to F | T * F$.

$F \to a | (E)$.

~~⊕ $E \to E + T | T$.~~

~~⊕ VP mới: $E \to$~~

⊕ $E \to T | E + T$.

$E \to TE'$.

$E' \to +TE' | \varepsilon$.

⊕ $T \to F | T * F$.

~~$T \to FF'$~~. $T \to FT'$.

$T' \Rightarrow *FT' | \varepsilon$.

(+) Văn phạm LL(1) trên sơ đồ cú pháp.

- Ở mỗi lối rẽ → 1 nhánh bắt đầu = các kí hiệu khác nhau.
- Sơ đồ chứa Lối rỗng → kí hiệu đứng sau k/h đó bỏ bởi sơ đồ phải khác kí hiệu đứng đầu nhánh của sơ đồ.

(+) Văn phạm PL/0 là LL(1) → slide.

(+) Thuật toán phân tích xem trước.

- Chỉ làm việc với VP LL(1).
- 2 gđ → XD ma trận ptích.
  → Thực hiện ptích xem trước.

XD ma trận PT.

PT ma trận PT M trên tập:

$$(V \cup \{\#\}) \times (V_T \cup \{\#\})$$

(1) $A \to \alpha$ là sx thứ k của P

(+) $\forall a \neq \varepsilon$, $a \in FIRST(\alpha)$ → $M[A, a] = (\alpha, k)$

(+) $\varepsilon \in FIRST(\alpha)$ thì $\forall b \in FOLLOW(A)$, $M[A, b] = (\alpha, k)$.

(2) $M[a, a]$ = "Đẩy"

(3) $M[\#, \#]$ = "Nhận".

a	b	b	a	b	#

(4) Mọi ô khác có gtrị sai.

VD

1. $S \to aAS$
2. $S \to b$
3. $A \to a$
4. $A \to bSA$

M	a	b	#
S	aAS, 1	b, 2	sai
A	a, 3	bSA, 4	sai
a	Đẩy	sai	sai
b	sai	Đẩy	sai
#	sai	sai	nhận.

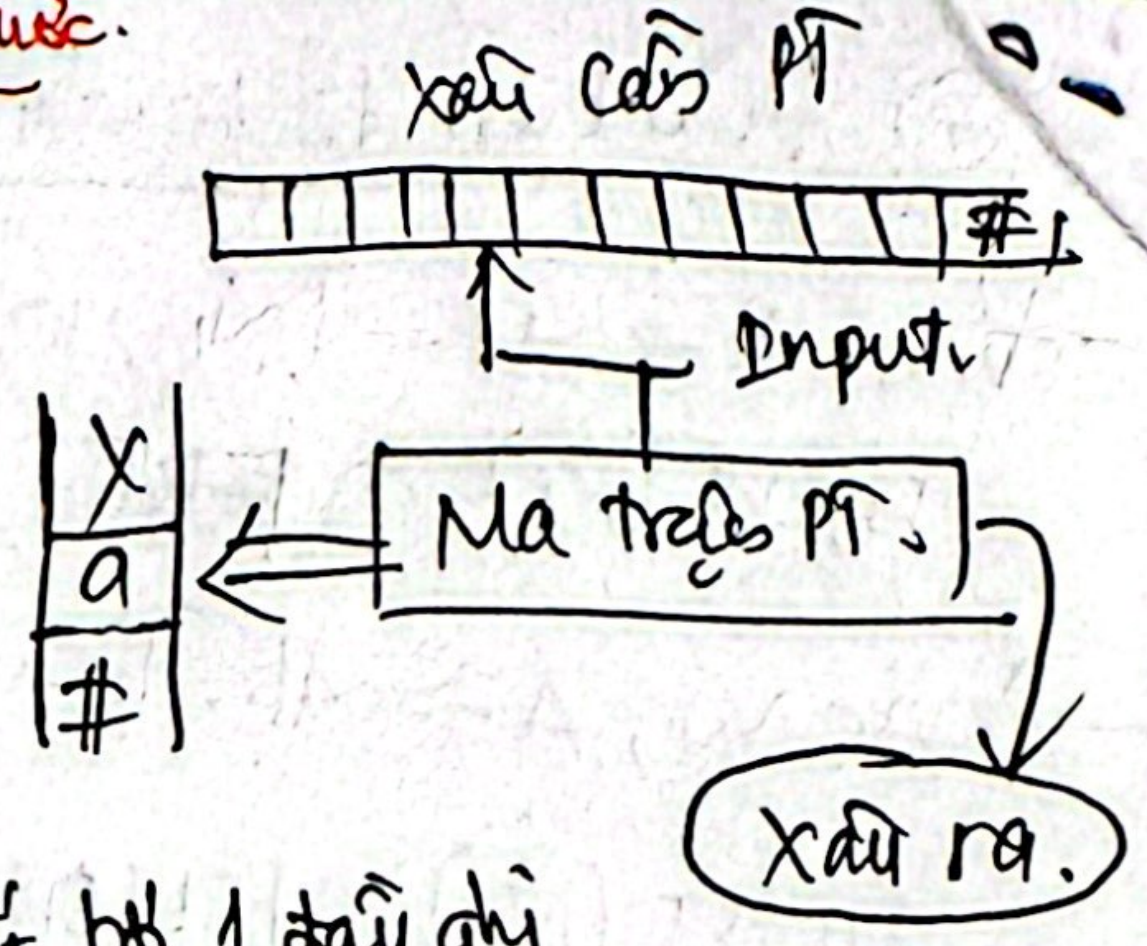
14232

→ Xâu abbab được đoán nhận với sx 14232

Phân tích xem trước.

T² gồm:

- Xâu cần PT
- Ma trận PT
- DS đẩy stack:
  (+) Chứa xâu Xa#
  (+) $Xa \in V^*$
- Xâu ra: hết bởi 1 đầu ghi; chứa thứ tự sx.



Hđộng.

$M[X, a] = (\beta, k)$: X trên đỉnh DS được thay = β. Viết k ra xâu ra

$M[X, a]$ = "Đẩy": đầu đọc dịch phải 1 ô, X bị loại bỏ khỏi DS.

$M[X, a]$ = "Sai": Dừng → Thông báo xâu k0 đoán nhận

$M[X, a]$ = "Nhận": Dừng → xâu đoán nhận.
→ xâu chứa DS các sx đã dùng

Hình trạng T².