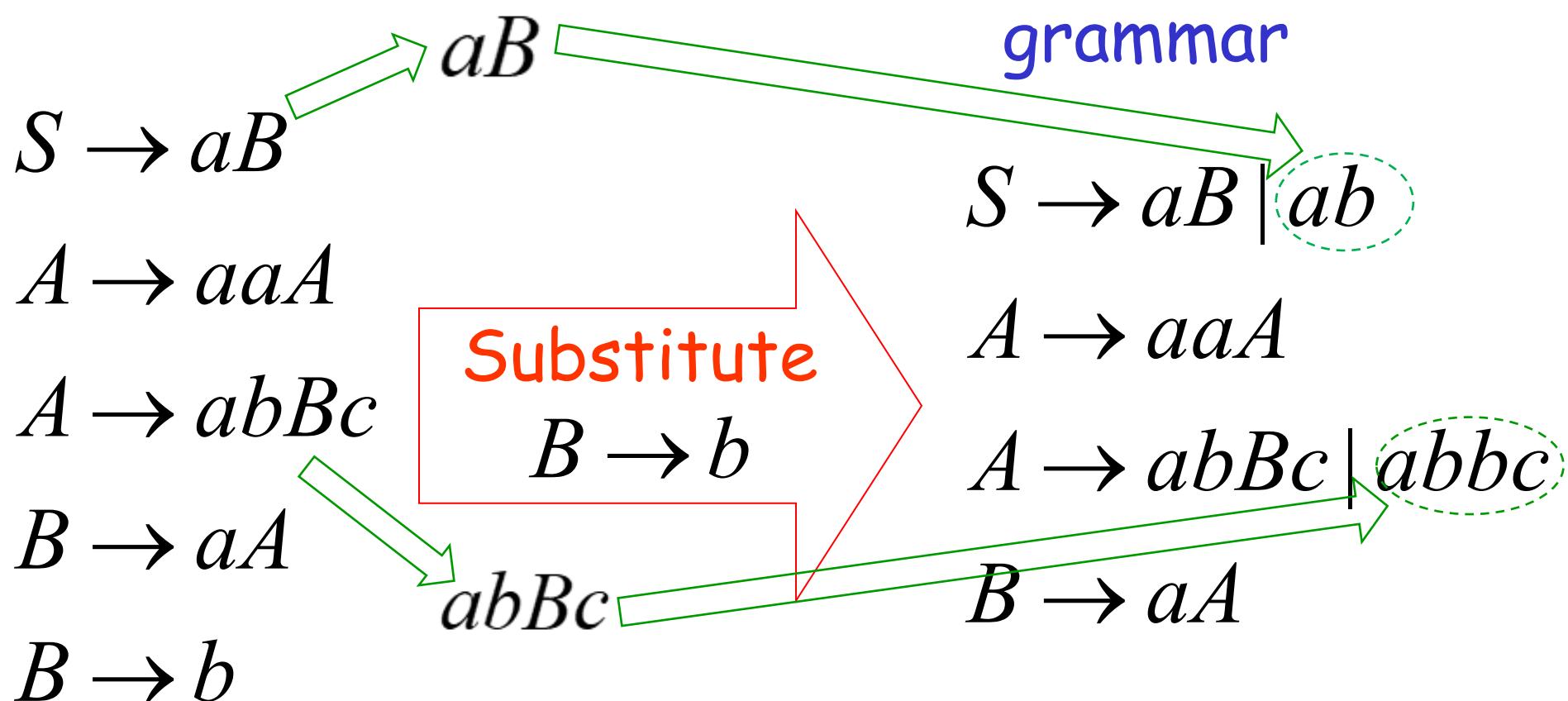


Simplifications of Context-Free Grammars

A Substitution Rule

កំណត់ពាណិជ្ជកម្មសម្រាប់របៀប



ହେଲ୍‌ ବୋଲ୍

A Substitution Rule

$$S \rightarrow aB \mid ab$$

$$A \rightarrow aaA$$

$$A \rightarrow abBc \mid abbc$$

$$B \rightarrow aA$$

Substitute

$$\underline{B \rightarrow aA}$$

$$\cancel{S \rightarrow aB \mid ab \mid aaA}$$

$$A \rightarrow aaA$$

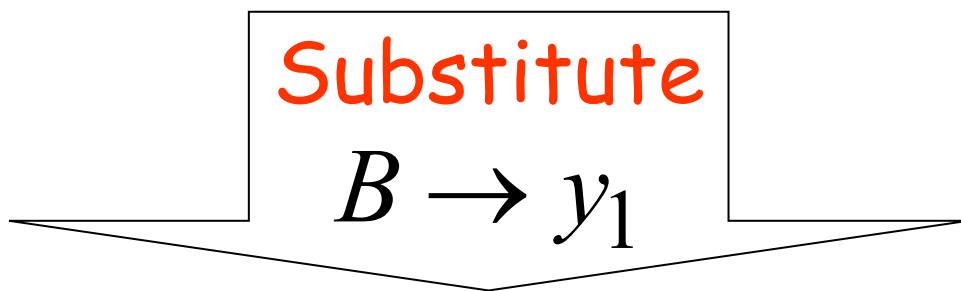
$$\cancel{A \rightarrow abBc \mid abbc \mid abaAc}$$

Equivalent
grammar

In general:

$$A \rightarrow xBz$$

$$B \rightarrow y_1$$



$$A \rightarrow xBz \mid xy_1z$$

equivalent
grammar

Substitution Rule

សំណើនៃ context-free grammar និងការ
បញ្ចូលរូបរាងសមត្ថភាព

Nullable Variables

↳ generate 1 λ

ដើម្បី * នឹង λ និងនៅក្នុង

λ – production:

$$A \rightarrow \lambda$$

Nullable Variable:

$$A \Rightarrow \dots \Rightarrow \lambda$$

Removing Nullable Variables

Example Grammar:

$$S \rightarrow aMb$$

$$M \rightarrow aMb$$

$$M \rightarrow \lambda$$



Nullable variable

Final Grammar

$$S \rightarrow aMb$$

$$M \rightarrow aMb$$

~~$$M \rightarrow \lambda$$~~

Substitute

$$M \rightarrow \lambda$$

$$S \rightarrow aMb$$

$$S \rightarrow ab$$

q_{ini}

$$M \rightarrow aMb$$

$$M \rightarrow ab$$

q_{ini}

វិធីក់ ២

Unit-Productions

ល្អស. V. អាជីវិត ពាណិជ្ជកម្ម និង ការបង្កើត

វិធីក់ និងការបង្កើត

$S \rightarrow aA$
 $A \rightarrow a$
 ~~$A \rightarrow B$~~
 $B \rightarrow A$
 $B \rightarrow bb$

Substitute
 $A \rightarrow B$

ល្អស. V. អាជីវិត និងការបង្កើត
 $S \rightarrow aA | aB$
 $A \rightarrow a$
 $B \rightarrow A | B$
 $B \rightarrow bb$

Unit Production: $A \rightarrow B$

(a single variable in both sides)

Removing Unit Productions

Observation:

X អាជីវកម្មណ៍ និងភាសា /

$$A \rightarrow A$$

Is removed immediately

Example Grammar:

$$S \rightarrow aA$$

$$A \rightarrow a$$

$$A \rightarrow B$$

$$B \rightarrow A$$

$$B \rightarrow bb$$

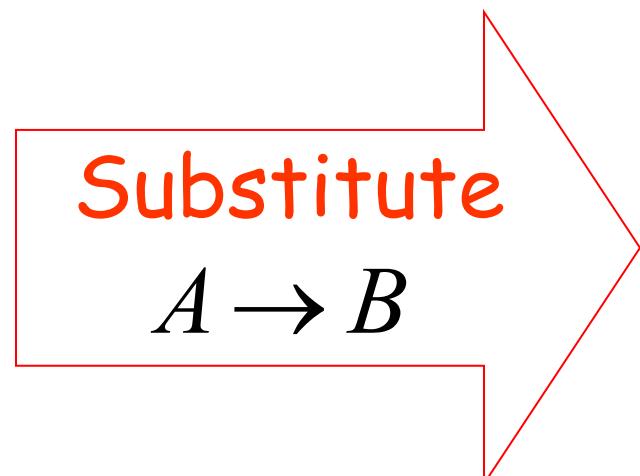
$$S \rightarrow aA$$

$$A \rightarrow a$$

~~$$A \rightarrow B$$~~

$$B \rightarrow A$$

$$B \rightarrow bb$$

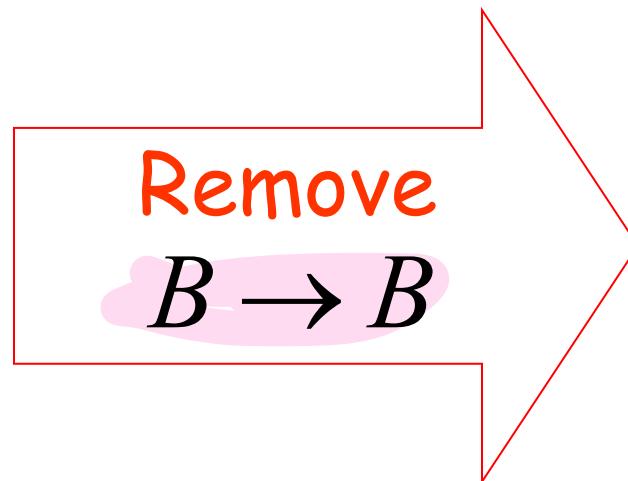


$$S \rightarrow aA \mid aB$$

$$A \rightarrow a$$

$$B \rightarrow A \mid B$$

$$B \rightarrow bb$$

$$S \rightarrow aA \mid aB$$
$$A \rightarrow a$$
$$B \rightarrow A \mid \cancel{B}$$
$$B \rightarrow bb$$

$$S \rightarrow aA \mid aB$$
$$A \rightarrow a$$
$$B \rightarrow A$$
$$B \rightarrow bb$$

$$S \rightarrow aA \mid aB$$

$$A \rightarrow a$$

~~$$B \rightarrow A$$~~

$$B \rightarrow bb$$

Substitute

$$B \rightarrow A$$

$$S \rightarrow aA \mid aB \mid aa$$

$$A \rightarrow a$$

$$B \rightarrow bb$$

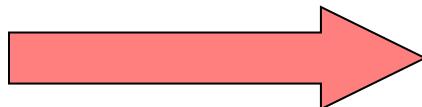
ঘৰ

Remove repeated productions

$S \rightarrow aA \mid aB \mid \cancel{aA}$

ตัด去

$A \rightarrow a$



$B \rightarrow bb$

Final grammar

$S \rightarrow aA \mid aB$

$A \rightarrow a$

$B \rightarrow bb$

ទី៣ ក្នុង ៣ Useless Productions

បានពួកវា មិនមែនជាប្រើប្រាស់

$$S \rightarrow aSb$$

$$S \rightarrow \lambda$$

$$S \rightarrow A$$

មិនមែនជាប្រើប្រាស់ Terminal ក្នុង

$$A \rightarrow aA$$

Useless Production

Some derivations never terminate...

$$S \Rightarrow A \Rightarrow aA \Rightarrow aaA \Rightarrow \dots \Rightarrow aa\dots aA \Rightarrow \dots$$

Another grammar:

$$S \rightarrow A$$

$$A \rightarrow aA$$

$$A \rightarrow \lambda$$

$$B \rightarrow bA$$

S
|
B

X

ລາຍງານ start variable ໄມຕັ້ງ

Useless Production

Not reachable from S

In general:

if

$$S \Rightarrow \dots \Rightarrow xAy \Rightarrow \dots \Rightarrow w$$

$\underbrace{\quad}_{\text{L នឹង នូវការ}} \quad \underbrace{\quad}_{\text{L នឹង នូវការ}} \quad \uparrow$

$w \in L(G)$

then variable

A is useful

មានលទ្ធផល ឬ និង សំណើ

otherwise, variable

A is useless

A production $A \rightarrow x$ is useless
if any of its variables is useless

$$S \rightarrow aSb$$

$$S \rightarrow \lambda$$

Variables

useless

useless

useless

$$S \rightarrow A$$

$$A \rightarrow aA$$

$$B \rightarrow C$$

$$C \rightarrow D$$

Productions

useless

useless

useless

useless

Derivations

Removing Useless Productions

Example Grammar:

$$S \rightarrow aS \mid A \mid C$$

$$A \rightarrow a$$

$$B \rightarrow aa$$

$$C \rightarrow aCb$$

First: find all variables that can produce strings with only terminals

①

Variable γ un Generate γ 's terminals string

1.1

set var variable γ can generate terminal str

$$S \rightarrow aS \mid A \mid C$$

$$A \rightarrow a$$

$$B \rightarrow aa$$

$$C \rightarrow aCb$$

$$S \rightarrow A$$

1.2

อพิจัย γ น Deri oonmiiin Var A, B

$$\text{Round 2: } \{A, B, S\}$$

set var variable γ can generate terminal str

1.3

စာလုပ်နည်း၏ set အမှု

Keep only the variables

that produce terminal symbols: $\{A, B, S\}$

(the rest variables are useless)

$$S \rightarrow aS \mid A \mid C$$

$$A \rightarrow a$$

$$B \rightarrow aa$$

$$\cancel{C \rightarrow aCb}$$



$$S \rightarrow aS \mid A$$

$$A \rightarrow a$$

$$B \rightarrow aa$$

Remove useless productions

Second: Find all variables

②

reachable from S

ທີ່ຖືກຈະນຶກ ດັບຕົວໂດຍ S ຂາດັ່ງຕົວນີ້ໄດ້

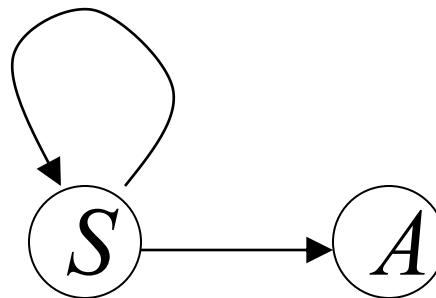
2.1

Use a Dependency Graph

$$S \rightarrow aS \mid A$$

$$A \rightarrow a$$

$$B \rightarrow aa$$



2.2

ກິດຕັ້ງ

B

not
reachable

Keep only the variables
reachable from S
(the rest variables are useless)

Final Grammar

$$S \rightarrow aS \mid A$$

$$A \rightarrow a$$

~~$$B \rightarrow aa$$~~



$$S \rightarrow aS \mid A$$

$$A \rightarrow a$$

Remove useless productions

Removing All

λ

Step 1: Remove Nullable Variables

$A \rightarrow A, A \rightarrow B$

Step 2: Remove Unit-Productions

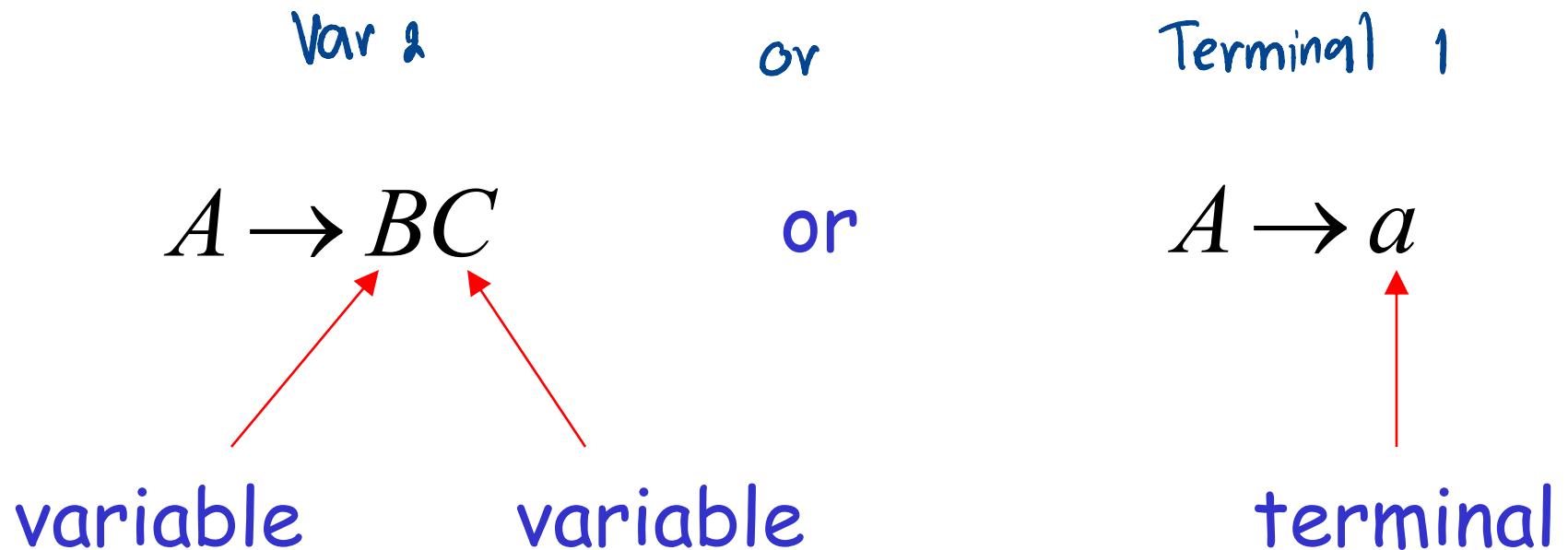
Step 3: Remove Useless Variables

Normal Forms for Context-free Grammars

①

Chomsky Normal Form

Each production has form:



Examples:

ឧស្សាហ៍មិនមែន chomsky
អេ

$$S \rightarrow AS$$

$$S \rightarrow a$$

$$A \rightarrow SA$$

$$A \rightarrow b$$

Chomsky
Normal Form

$$S \rightarrow AS$$

$$S \rightarrow \textcircled{AAS}$$

$$A \rightarrow SA$$

$$A \rightarrow \textcircled{aa}$$

Not Chomsky
Normal Form

Conversion to Chomsky Normal Form

Gramma چุน ក්‍රියාකාර
Chomsky Form = මෙවැදිපිටියා

Example:

$$S \rightarrow ABa$$

$$A \rightarrow aab$$

$$B \rightarrow Ac$$

Not Chomsky
Normal Form

Introduce variables for terminals: T_a, T_b, T_c

$$S \rightarrow ABa$$

$$A \rightarrow aab$$

$$B \rightarrow Ac$$



ఎల్లా

యథిని:

$$\left\{ \begin{array}{l} S \rightarrow ABT_a \\ A \rightarrow T_a T_a T_b \\ B \rightarrow AT_c \\ T_a \rightarrow a \\ T_b \rightarrow b \\ T_c \rightarrow c \end{array} \right.$$

Introduce intermediate variable: V_1

$$S \rightarrow ABT_a$$

$$A \rightarrow T_a T_a T_b$$

$$B \rightarrow AT_c$$

$$T_a \rightarrow a$$

$$T_b \rightarrow b$$

$$T_c \rightarrow c$$



$$S \rightarrow AV_1$$

$$V_1 \rightarrow BT_a$$

$$A \rightarrow T_a T_a T_b$$

$$B \rightarrow AT_c$$

$$T_a \rightarrow a$$

$$T_b \rightarrow b$$

$$T_c \rightarrow c$$

Introduce intermediate variable: V_2

$$S \rightarrow AV_1$$

$$V_1 \rightarrow BT_a$$

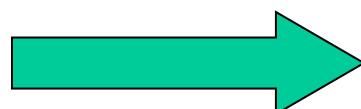
$$A \rightarrow T_a T_a T_b$$

$$B \rightarrow AT_c$$

$$T_a \rightarrow a$$

$$T_b \rightarrow b$$

$$T_c \rightarrow c$$



$$S \rightarrow AV_1$$

$$V_1 \rightarrow BT_a$$

$$A \rightarrow T_a V_2$$

$$V_2 \rightarrow T_a T_b$$

$$B \rightarrow AT_c$$

$$T_a \rightarrow a$$

$$T_b \rightarrow b$$

$$T_c \rightarrow c$$

Final grammar in Chomsky Normal Form:

ຄົມ ດອວຍ ໄກສະ: ດິຈຸນິຍາ
ເຊດ: ນິຕິທີ ລົດນັ້ນຢູ່ມືລົງກົດ

$$S \rightarrow AV_1$$

$$V_1 \rightarrow BT_a$$

$$A \rightarrow T_a V_2$$

$$V_2 \rightarrow T_a T_b$$

$$B \rightarrow AT_c$$

$$T_a \rightarrow a$$

$$T_b \rightarrow b$$

$$T_c \rightarrow c$$

Initial grammar

$$S \rightarrow ABa$$

$$A \rightarrow aab$$

$$B \rightarrow Ac$$

In general: កិច្ចការ
Grammar នៃ នៅលើ Chomsky Normal Form
សមរក នូវការដោះស្រាយ

From any context-free grammar
(which doesn't produce λ)
not in Chomsky Normal Form

we can obtain:

An equivalent grammar
in Chomsky Normal Form

The Procedure

non-ni chomsky NF

① First remove:

Nullable variables

Unit productions

Then, for every symbol a :

1) IH

Add production $T_a \rightarrow a$

In productions: replace a with T_a

New variable: T_a

Replace any production $A \rightarrow C_1C_2\cdots C_n$

with

$$A \rightarrow C_1V_1$$

$$V_1 \rightarrow C_2V_2$$

...

$$V_{n-2} \rightarrow C_{n-1}C_n$$

New intermediate variables: V_1, V_2, \dots, V_{n-2}

ກິບມາມ

ດີ,

Theorem: For any context-free grammar
(which doesn't produce λ)
there is an equivalent grammar
ນີ້ in Chomsky Normal Form ຈະໄດ້ມາເນັດວົດ

Observations

ఏనుక్కాపాణం తీవ్రికుగా గా

- Chomsky normal forms are good for parsing and proving theorems

ఫిస్తుజీగే కాగిచ్చు

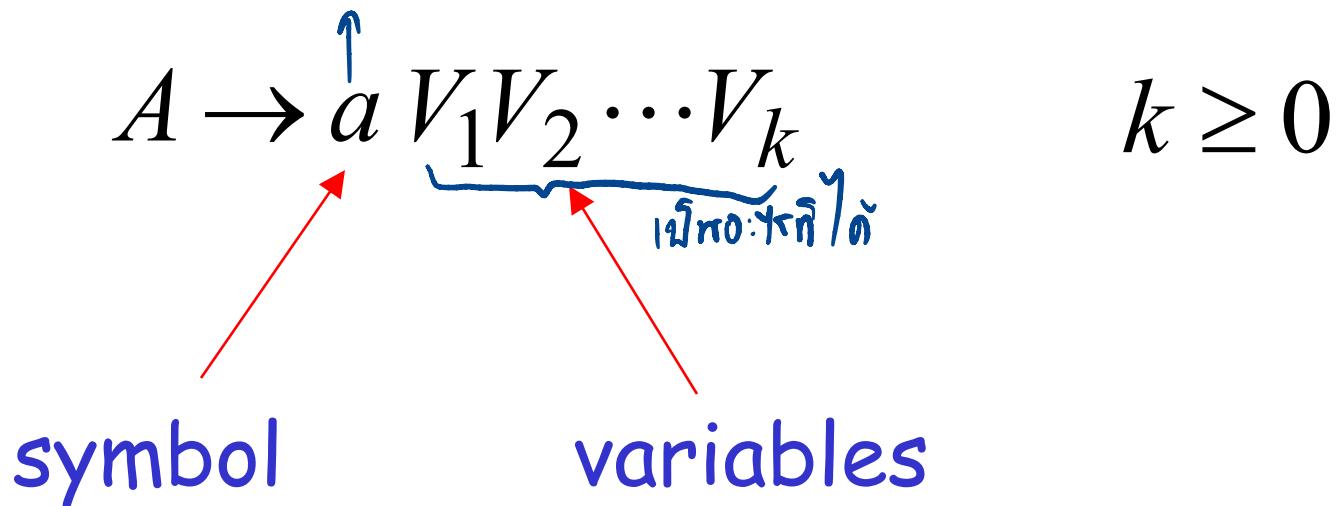
లాహులోయ

- It is very easy to find the Chomsky normal form for any context-free grammar

Greibach Normal Form (กรีบัช)

All productions have form:

fixed = ตัวแรกใน Terminal



Examples:

$$S \rightarrow cAB$$

$$A \rightarrow aA \mid bB \mid b$$

$$B \rightarrow b$$

Greibach

Normal Form

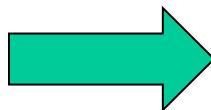
$$S \rightarrow abSb$$

$$S \rightarrow aa$$

Not Greibach

Normal Form

Conversion to Greibach Normal Form:

$$S \rightarrow abSb$$
$$S \rightarrow aa$$

$$S \rightarrow aT_b S T_b$$
$$S \rightarrow aT_a$$
$$T_a \rightarrow a$$
$$T_b \rightarrow b$$

Greibach
Normal Form

Theorem: For any context-free grammar
(which doesn't produce λ)
there is an equivalent grammar
in Greibach Normal Form
గ్రైబాచ్ నోర్మల్ ఫార్మ

Observations

- Greibach normal forms are very good for parsing
- It is possible to find the Greibach normal form of any context-free grammar

Algorithm Parsing

The CYK Parser (Cocke-Younger-Kasami)

The CYK Membership Algorithm

ເກົ່າຈຳເນື້າ

Input:

ຖີ່ອ່ານ w , ແລ້ວ ເສັດວ່າ $w \in L$

- ① • Grammar G in Chomsky Normal Form
- ② • String w

Output:

find if $w \in L(G)$

The Algorithm

Input example:

- Grammar G :
 $S \rightarrow AB$
 $A \rightarrow BB$
 $A \rightarrow a$
 $B \rightarrow AB$
 $B \rightarrow b$
- String w : $aabb$

aabb = w

① สร้าง Table ▷ (3 | 4 จำนวน)

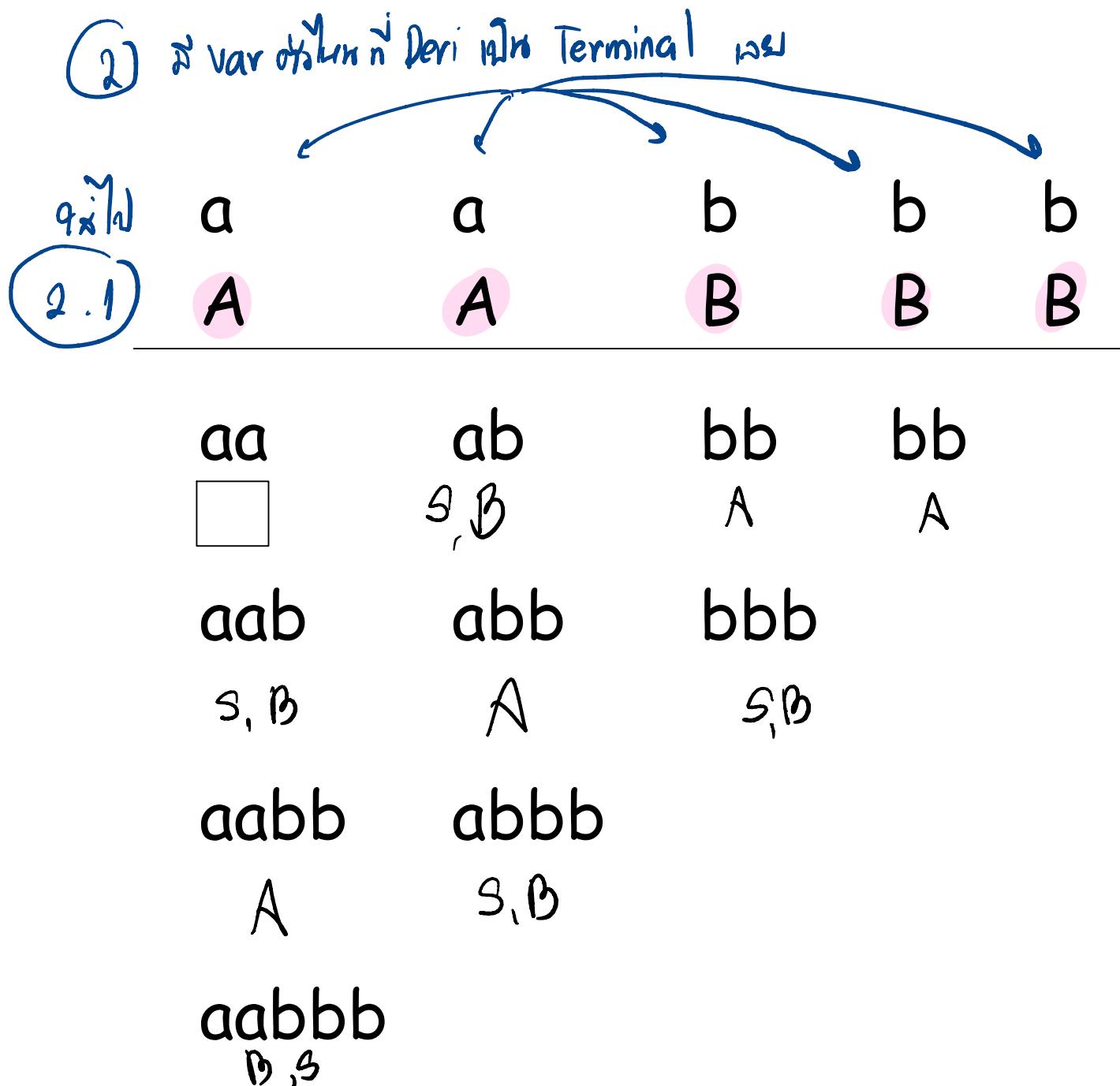
ลำดับ: 1 a a b b b

ลำดับ: 2 aa ab bb bb

ลำดับ: 3 aab abb bbb

ลำดับ: 4 aabb abbb

ลำดับ: 5 aabbb

$S \rightarrow AB$ $A \rightarrow BB$ $A \rightarrow a$ $B \rightarrow AB$ $B \rightarrow b$ 

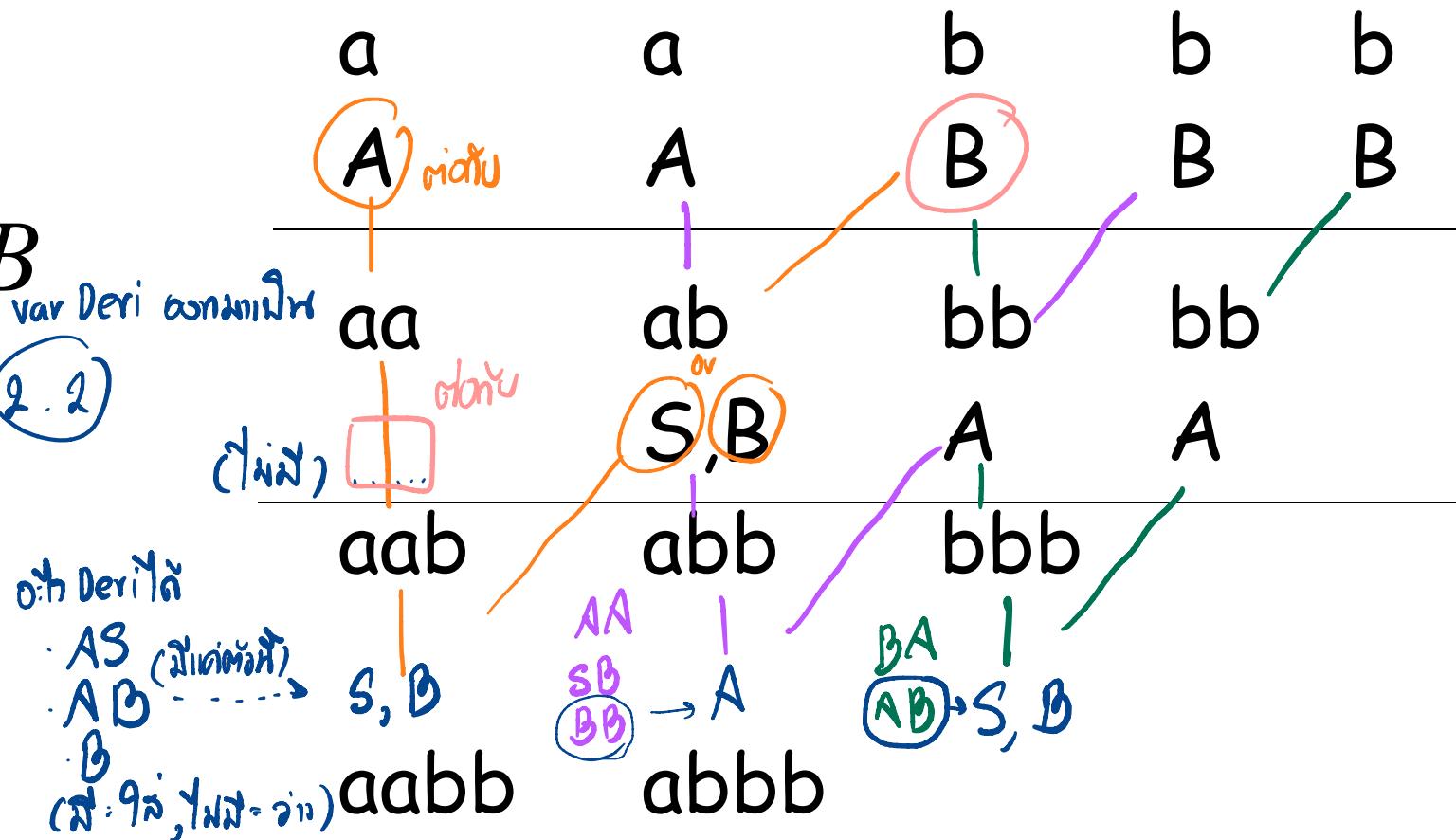
$S \rightarrow AB$

$A \rightarrow BB$

$A \rightarrow a$

$B \rightarrow AB$

$B \rightarrow b$ 2.2



aabb

$S \rightarrow AB$

$A \rightarrow BB$

$A \rightarrow a$

$B \rightarrow AB$

$B \rightarrow b$

(2.4)

AA

A

SB

BB

$\rightarrow A$

aabb

S,B

a
A

a
A

b
B

b
B

b
B

aa
ab
S,B

abb

bb
A

bb
A

bbb

S,B

S,B

aabb

abbb

S,B

AS,AB

SA,BB

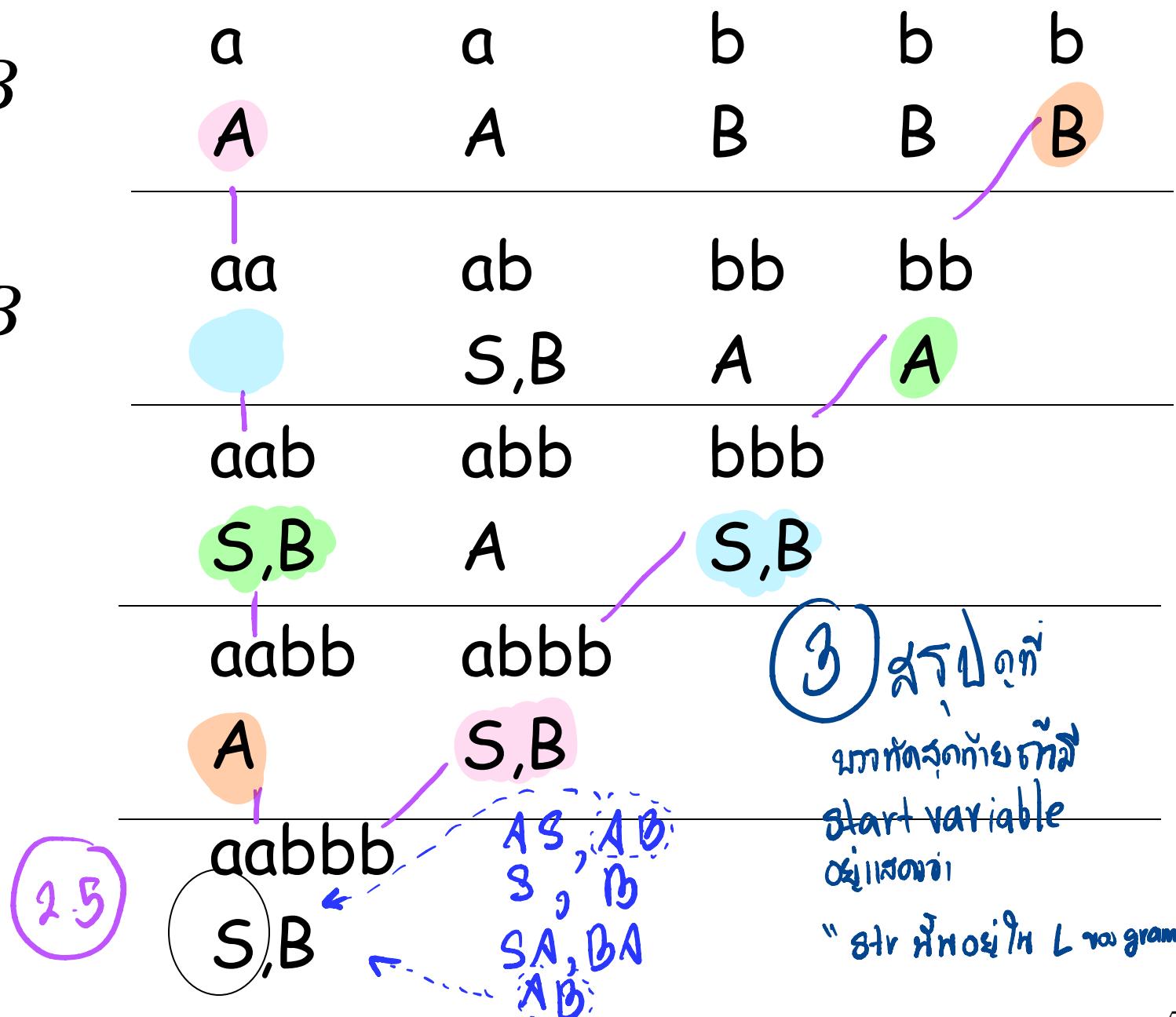
$S \rightarrow AB$

$A \rightarrow BB$

$A \rightarrow a$

$B \rightarrow AB$

$B \rightarrow b$



Therefore: $aabb \in L(G)$

Source code ដោយ loop មិន 3 រូប

លោកវីរ៉ាស់រៀង នៃការ 2 loop (For) នឹងការ

គ្មាន ឬក្បាស់ នៅក្នុងប៉ាក់បីនីត្រូវដោយ vshape ចំណាំ time complexity

Time Complexity:

$$|w|^3$$

Observation: The CYK algorithm can be easily converted to a parser (bottom up parser)