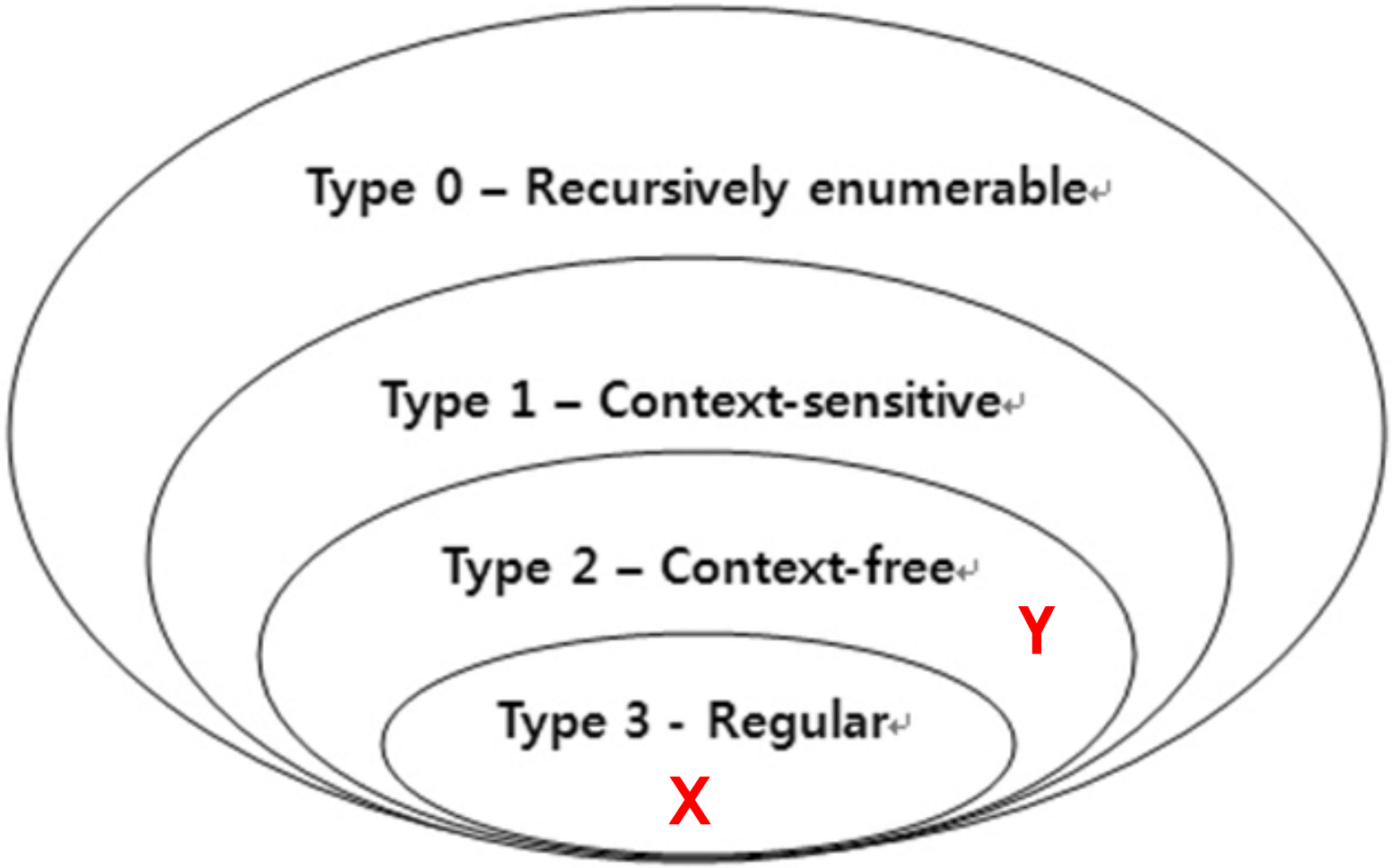
$$S \rightarrow S\,S \mid 1 \mid \varepsilon$$

Theory of Computation

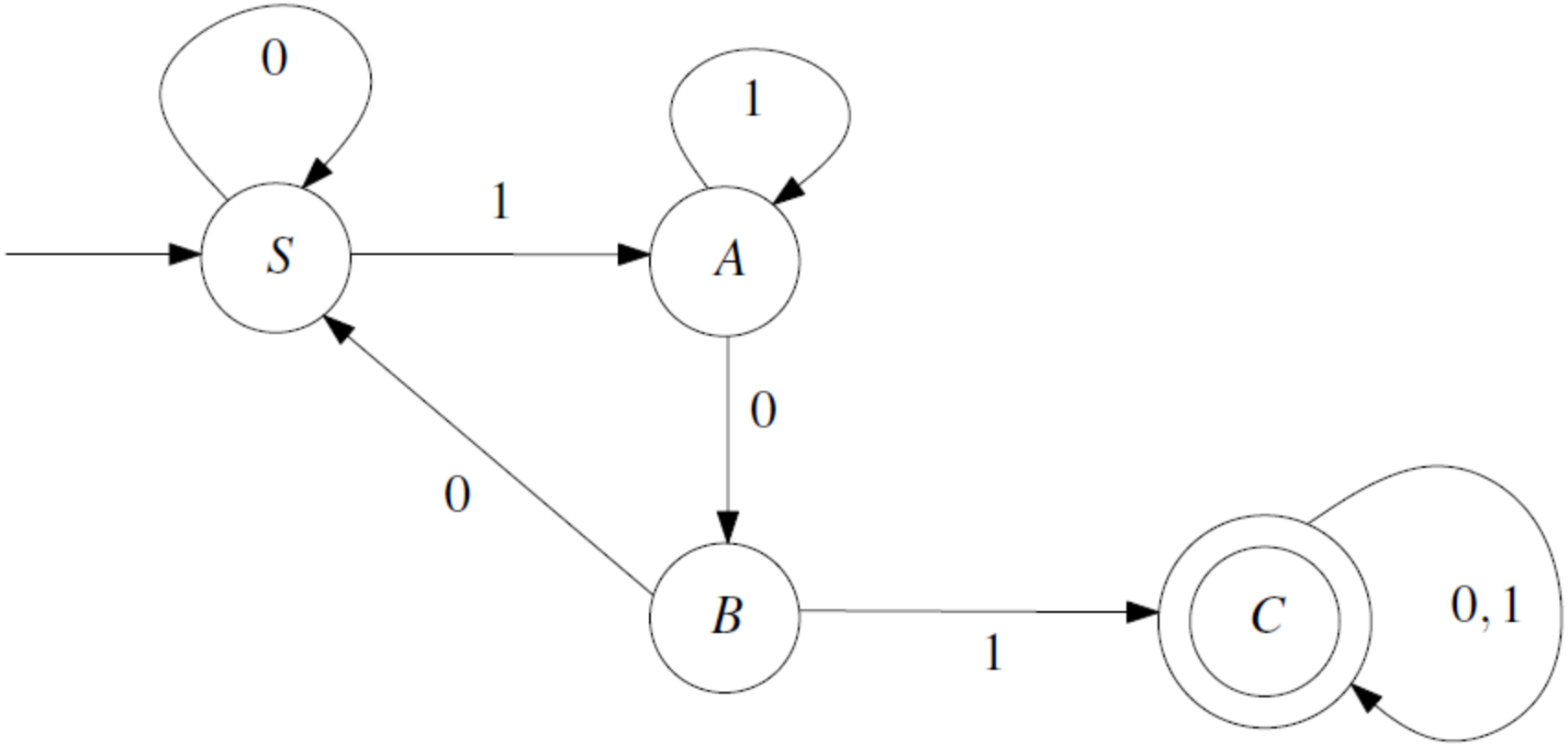
Lesson 8

Chomsky Normal Form

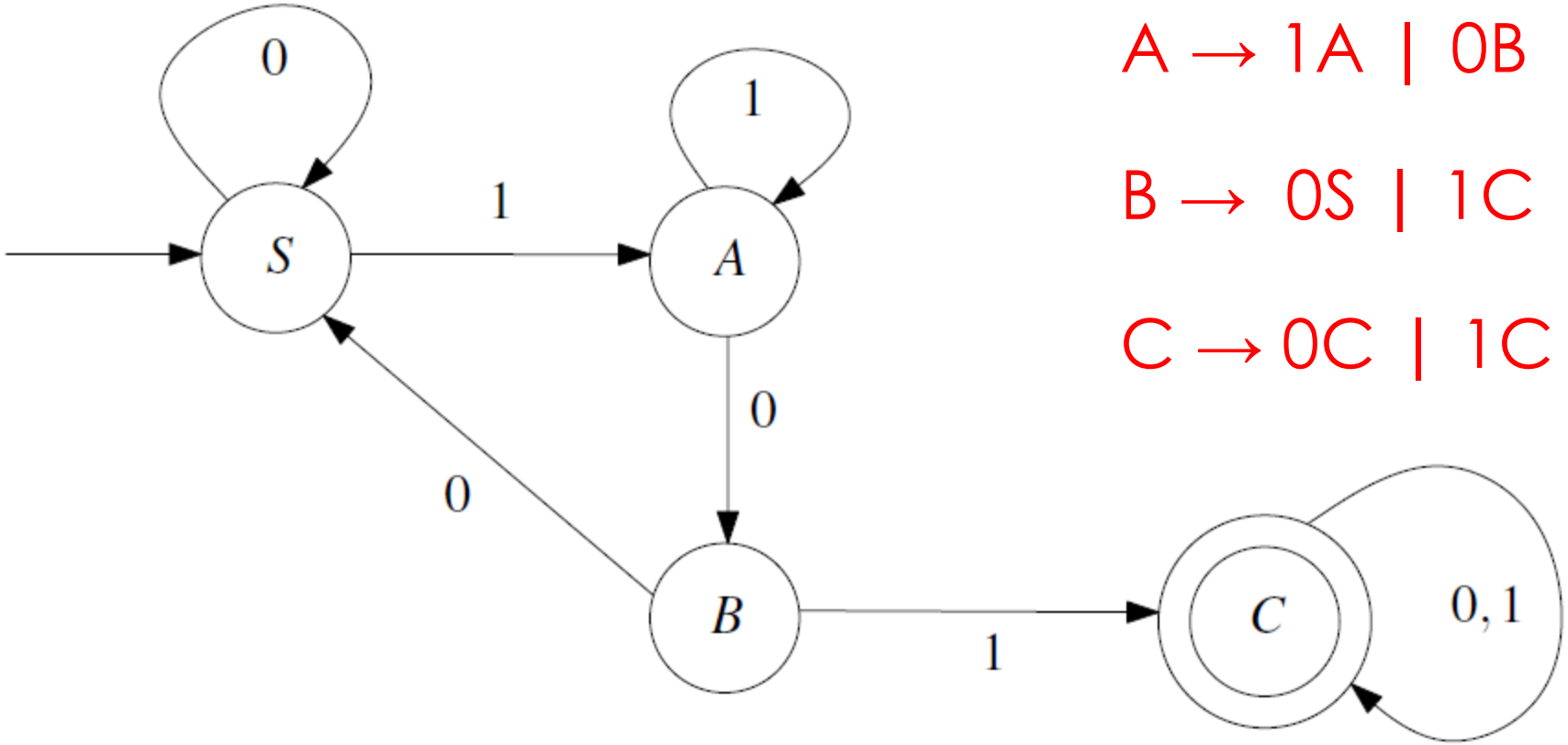
Regular languages are context-free



Regular languages are context-free



Regular languages are context-free



S → 0 S | 1A

A → 1A | 0B

B → 0S | 1C

C → 0C | 1C | ε

Chomsky normal form

Consider the following three grammar rules:

$$S \xrightarrow[1]{} SS \mid \underset{2}{1} \mid \underset{3}{\varepsilon}$$

Using this grammar we can write a '1' symbol in different ways. For example, you can only use the second rule, or 1th, 2nd, 3rd rules respectively.

Chomsky normal form

How can we derivate '1' with this CFG?

$$S \xrightarrow{1} SS \mid \xrightarrow{2} 1 \mid \xrightarrow{3} \varepsilon$$

$S \xrightarrow{2} 1$

$S \xrightarrow{1} SS \xrightarrow{2} 1S \xrightarrow{3} 1$

$S \xrightarrow{1} SS \xrightarrow{1} SSS \xrightarrow{2} 1SS \xrightarrow{3} 1S \xrightarrow{3} 1$

$S \xrightarrow{1} SS \xrightarrow{1} SSS \xrightarrow{1} SSSS \xrightarrow{2} 1SSS \xrightarrow{3} 1SS \xrightarrow{3} 1S \xrightarrow{3} 1$

…

Formal Definition

A context-free grammar $G = (V,\Sigma,R,S)$ is said to be in Chomsky normal form, if every rule in R has one of the following three forms:

1. $A \rightarrow BC$, where A, B, and C are elements of V, $B \neq S$, and $C \neq S$.
2. $A \rightarrow a$, where A is an element of V and a is an element of Σ.
3. $S \rightarrow \varepsilon$, where S is the start variable.

Transforming CFG to CNF

$S \rightarrow SS \mid 1 \mid \varepsilon$

When we revise the above grammar with Chomsky Normal Form, it will turn into the following grammar.

$S \rightarrow AA \mid 1 \mid \varepsilon$

$A \rightarrow AA \mid 1$

Transforming CFG to CNF

Here we will learn a five-step transformation method.

1. Eliminate the start variable on the right.
2. Eliminate all ε-rules.
3. Eliminate all unit-rules.
4. Eliminate all rules having more than two symbols.
5. Eliminate all rules, whose right-hand side contains exactly two symbols, which are not both variables.

An Example on CFG2CNF

Consider the context-free grammar $G = (V, \Sigma, R, A)$, where $V = \{A, B\}$, $\Sigma = \{0, 1\}$, A is the start variable, and R consists of the rules

$$A \rightarrow BAB \mid B \mid \varepsilon$$

$$B \rightarrow 00 \mid \varepsilon$$

An Example on CFG2CNF

A → BAB | B | ε

B → 00 | ε

1. Eliminate the start variable on the right.

S → A

A → BAB | B | ε

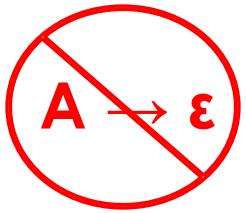
B → 00 | ε

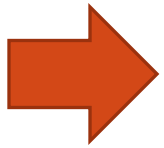
An Example on CFG2CNF

A → BAB | B | ε
B → 00 | ε

2. Eliminate all ε-rules.

S → A
A → BAB | B | ε
B → 00 | ε





S → A | ε
A → BAB | B | BB
B → 00 | ε

S → ε
A → BB

An Example on CFG2CNF

$A \to BAB \mid B \mid \varepsilon$

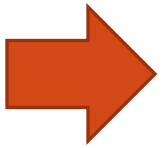
$B \to 00 \mid \varepsilon$

2. Eliminate all ε-rules.

$S \to A \mid \varepsilon$

$A \to BAB \mid B \mid BB$

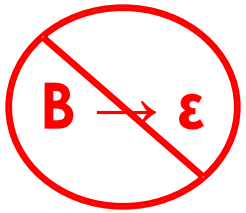
$B \to 00 \mid \varepsilon$



$S \to A \mid \varepsilon$

$A \to BAB \mid B \mid BB \mid AB \mid BA \mid A$

$B \to 00$



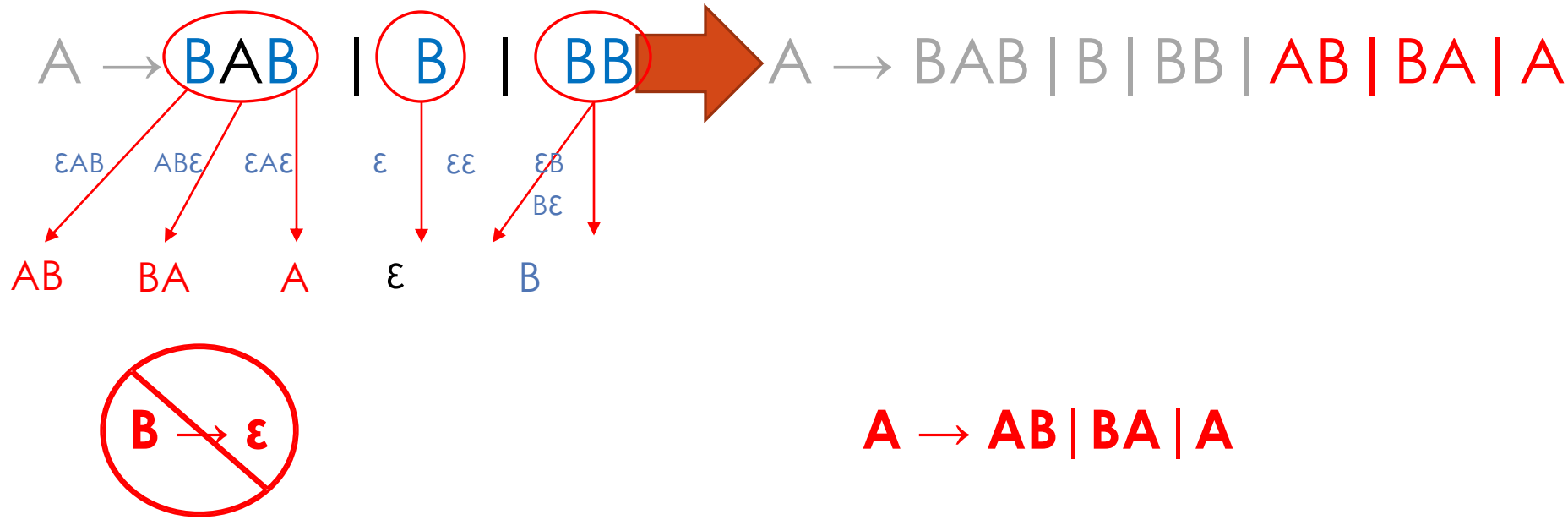
$A \to AB \mid BA \mid A$

An Example on CFG2CNF

$A \rightarrow BAB \mid B \mid \varepsilon$

$B \rightarrow 00 \mid \varepsilon$

2. Eliminate all ε-rules.



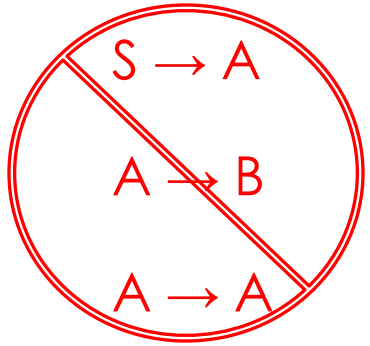
$A \rightarrow AB \mid BA \mid A$

An Example on CFG2CNF

$$A \rightarrow BAB \mid B \mid \varepsilon$$
$$B \rightarrow 00 \mid \varepsilon$$

3. Eliminate all unit-rules.

$$S \rightarrow A \mid \varepsilon$$
$$A \rightarrow BAB \mid B \mid BB \mid AB \mid BA \mid A$$
$$B \rightarrow 00$$



$$S \rightarrow \varepsilon \mid BAB \mid BB \mid AB \mid BA \mid 00$$
$$A \rightarrow BAB \mid BB \mid AB \mid BA \mid 00$$
$$B \rightarrow 00$$

An Example on CFG2CNF

A → BAB | B | ε

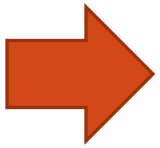
B → 00 | ε

4. Eliminate all rules having more than two symbols.

S → ε | BAB | BB | AB | BA | 00

A → BAB | BB | AB | BA | 00

B → 00



S → ε | BC | BB | AB | BA | 00

A → BC | BB | AB | BA | 00

B → 00

C → AB

An Example on CFG2CNF

A → BAB | B | ε

B → 00 | ε

5. Eliminate all rules, whose right-hand side contains exactly two symbols, which are not both variables.

S → ε | BC | BB | AB | BA | 00

A → BC | BB | AB | BA | 00

B → 00

C → AB

S → ε | BC | BB | AB | BA | DD

A → BC | BB | AB | BA | DD

B → DD

C → AB

D → 0

- That’s all.
- Thanks for listening.