COMP-330
Theory of Computation

Fall 2019 -- Prof. Claude Crépeau

Lec. 10 : Context-Free Grammars

Context-Free Grammars

- Let's call the following grammar G_1 :

$$A \rightarrow 0A1$$
$$A \rightarrow B$$
$$B \rightarrow \#$$

- Derivation of a string "000#111" :

$$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 000A111 \Rightarrow 000B111 \Rightarrow 000\#111.$$

Definition of CFG

- Variables $A, B, C, \langle TERM \rangle, \langle EXPR \rangle$
- Alphabet (of terminals) $0, 1, \#$
- Substitution Rules
  $$A \rightarrow 0A1$$
  $$\langle EXPR \rangle \rightarrow \langle TERM \rangle$$
- Start Variable A

  (left-hand side of the first substitution rule)

Definition of CFG

DEFINITION 2.2

A ***context-free grammar*** is a 4-tuple (V, Σ, R, S), where

1. V is a finite set called the ***variables***,
2. Σ is a finite set, disjoint from V, called the ***terminals***,
3. R is a finite set of ***rules***, with each rule being a variable and a string of variables and terminals, and
4. $S \in V$ is the start variable.

Parse Tree

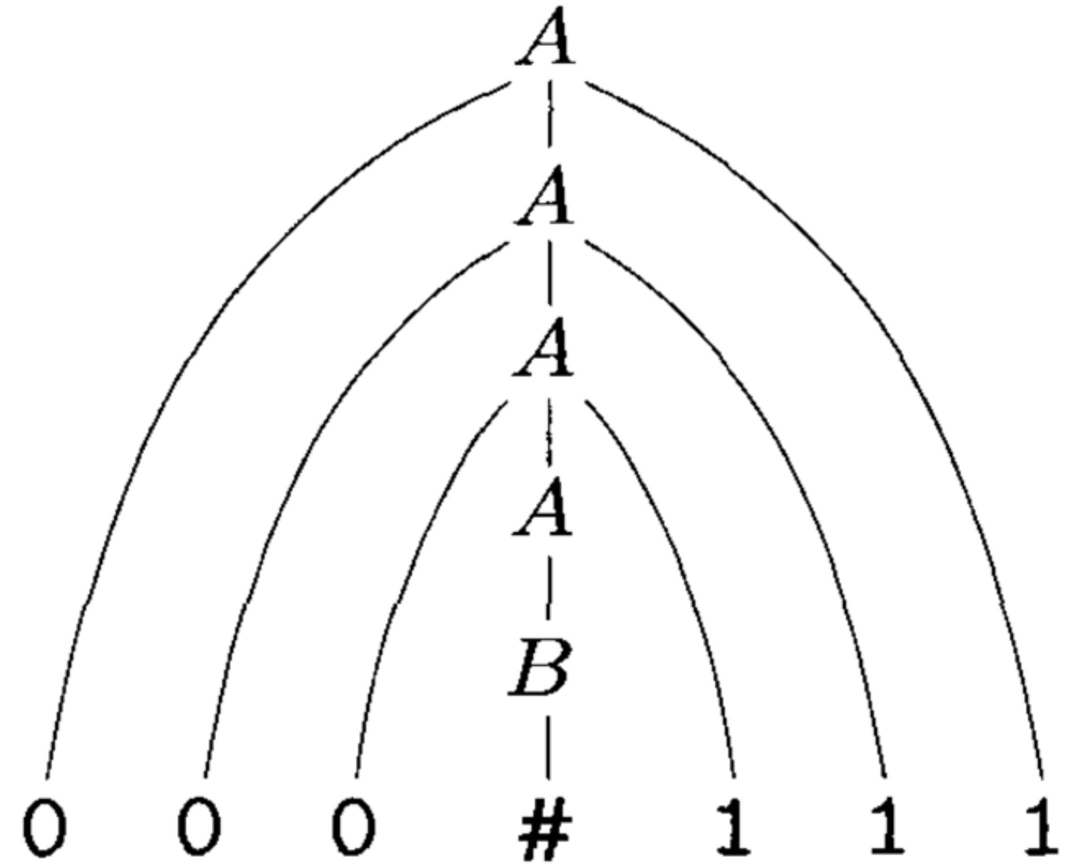


FIGURE 2.1
Parse tree for 000#111 in grammar G_1

Definition of CFL

- If u, v and w are strings of variables and terminals, and $A \rightarrow w$ is a rule of the grammar, we say that uAv yields uwv, written $uAv \Rightarrow uwv$.
- We say that u derives v ($u \overset{*}{\Rightarrow} v$) if u=v or if

$$u \Rightarrow u_1 \Rightarrow u_2 \Rightarrow \ldots \Rightarrow u_k \Rightarrow v,\ k \geq 0.$$

- The language of G is $\{ w \in \Sigma^* \mid S \overset{*}{\Rightarrow} w \}$.

Context-Free Grammars

- Formally, grammar G_1 :

$$V = \{A, B\}$$
$$\Sigma = \{0, 1, \#\}$$
$$R = \{A \rightarrow 0A1 \mid B,$$
$$B \rightarrow \#\}$$
$$S = A$$

- $L(G_1) = \{ 0^n\#1^n \mid n \geq 0 \}$.

Example of CFG

G_2=(
{ ⟨SENTENCE⟩,⟨NOUN-PHRASE⟩,⟨VERB-PHRASE⟩,⟨PREP-PHRASE⟩,
⟨CMPLX-NOUN⟩,⟨CMPLX-VERB⟩,⟨ARTICLE⟩,⟨NOUN⟩,⟨VERB⟩,⟨PREP⟩ },
{a,b,c,…,z," "},
R_2,
⟨SENTENCE⟩
)

R_2:

⟨SENTENCE⟩ → ⟨NOUN-PHRASE⟩⟨VERB-PHRASE⟩
⟨NOUN-PHRASE⟩ → ⟨CMPLX-NOUN⟩ | ⟨CMPLX-NOUN⟩⟨PREP-PHRASE⟩
⟨VERB-PHRASE⟩ → ⟨CMPLX-VERB⟩ | ⟨CMPLX-VERB⟩⟨PREP-PHRASE⟩
⟨PREP-PHRASE⟩ → ⟨PREP⟩⟨CMPLX-NOUN⟩
⟨CMPLX-NOUN⟩ → ⟨ARTICLE⟩⟨NOUN⟩
⟨CMPLX-VERB⟩ → ⟨VERB⟩ | ⟨VERB⟩⟨NOUN-PHRASE⟩
⟨ARTICLE⟩ → a | the
⟨NOUN⟩ → boy | girl | flower
⟨VERB⟩ → touches | likes | sees
⟨PREP⟩ → with

Example of CFG

Rules of grammar G_2 :

⟨SENTENCE⟩	→	⟨NOUN-PHRASE⟩⟨VERB-PHRASE⟩
⟨NOUN-PHRASE⟩	→	⟨CMPLX-NOUN⟩ | ⟨CMPLX-NOUN⟩⟨PREP-PHRASE⟩
⟨VERB-PHRASE⟩	→	⟨CMPLX-VERB⟩ | ⟨CMPLX-VERB⟩⟨PREP-PHRASE⟩
⟨PREP-PHRASE⟩	→	⟨PREP⟩⟨CMPLX-NOUN⟩
⟨CMPLX-NOUN⟩	→	⟨ARTICLE⟩⟨NOUN⟩
⟨CMPLX-VERB⟩	→	⟨VERB⟩ | ⟨VERB⟩⟨NOUN-PHRASE⟩
⟨ARTICLE⟩	→	a | the
⟨NOUN⟩	→	boy | girl | flower
⟨VERB⟩	→	touches | likes | sees
⟨PREP⟩	→	with

Example of CFG

⟨ARTICLE⟩ → a | the

means

⟨ARTICLE⟩ → a
⟨ARTICLE⟩ → the

Rules of grammar G_2 :

⟨SENTENCE⟩	→	⟨NOUN-PHRASE⟩⟨VERB-PHRASE⟩
⟨NOUN-PHRASE⟩	→	⟨CMPLX-NOUN⟩ | ⟨CMPLX-NOUN⟩⟨PREP-PHRASE⟩
⟨VERB-PHRASE⟩	→	⟨CMPLX-VERB⟩ | ⟨CMPLX-VERB⟩⟨PREP-PHRASE⟩
⟨PREP-PHRASE⟩	→	⟨PREP⟩⟨CMPLX-NOUN⟩
⟨CMPLX-NOUN⟩	→	⟨ARTICLE⟩⟨NOUN⟩
⟨CMPLX-VERB⟩	→	⟨VERB⟩ | ⟨VERB⟩⟨NOUN-PHRASE⟩
⟨ARTICLE⟩	→	a | the
⟨NOUN⟩	→	boy | girl | flower
⟨VERB⟩	→	touches | likes | sees
⟨PREP⟩	→	with

$$
\begin{aligned}
\langle \text{SENTENCE} \rangle &\rightarrow \langle \text{NOUN-PHRASE} \rangle \langle \text{VERB-PHRASE} \rangle \\
\langle \text{NOUN-PHRASE} \rangle &\rightarrow \langle \text{CMPLX-NOUN} \rangle \mid \langle \text{CMPLX-NOUN} \rangle \langle \text{PREP-PHRASE} \rangle \\
\langle \text{VERB-PHRASE} \rangle &\rightarrow \langle \text{CMPLX-VERB} \rangle \mid \langle \text{CMPLX-VERB} \rangle \langle \text{PREP-PHRASE} \rangle \\
\langle \text{PREP-PHRASE} \rangle &\rightarrow \langle \text{PREP} \rangle \langle \text{CMPLX-NOUN} \rangle \\
\langle \text{CMPLX-NOUN} \rangle &\rightarrow \langle \text{ARTICLE} \rangle \langle \text{NOUN} \rangle \\
\langle \text{CMPLX-VERB} \rangle &\rightarrow \langle \text{VERB} \rangle \mid \langle \text{VERB} \rangle \langle \text{NOUN-PHRASE} \rangle \\
\langle \text{ARTICLE} \rangle &\rightarrow \texttt{a} \mid \texttt{the} \\
\langle \text{NOUN} \rangle &\rightarrow \texttt{boy} \mid \texttt{girl} \mid \texttt{flower} \\
\langle \text{VERB} \rangle &\rightarrow \texttt{touches} \mid \texttt{likes} \mid \texttt{sees}
\end{aligned}
$$

$$
\langle \text{SENTENCE} \rangle \Rightarrow \langle \text{NOUN-PHRASE} \rangle \langle \text{VERB-PHRASE} \rangle
$$

$$
\begin{aligned}
\langle\text{SENTENCE}\rangle &\to \langle\text{NOUN-PHRASE}\rangle\langle\text{VERB-PHRASE}\rangle \\
\langle\text{NOUN-PHRASE}\rangle &\to \langle\text{CMPLX-NOUN}\rangle \mid \langle\text{CMPLX-NOUN}\rangle\langle\text{PREP-PHRASE}\rangle \\
\langle\text{VERB-PHRASE}\rangle &\to \langle\text{CMPLX-VERB}\rangle \mid \langle\text{CMPLX-VERB}\rangle\langle\text{PREP-PHRASE}\rangle \\
\langle\text{PREP-PHRASE}\rangle &\to \langle\text{PREP}\rangle\langle\text{CMPLX-NOUN}\rangle \\
\langle\text{CMPLX-NOUN}\rangle &\to \langle\text{ARTICLE}\rangle\langle\text{NOUN}\rangle \\
\langle\text{CMPLX-VERB}\rangle &\to \langle\text{VERB}\rangle \mid \langle\text{VERB}\rangle\langle\text{NOUN-PHRASE}\rangle \\
\langle\text{ARTICLE}\rangle &\to \texttt{a} \mid \texttt{the} \\
\langle\text{NOUN}\rangle &\to \texttt{boy} \mid \texttt{girl} \mid \texttt{flower} \\
\langle\text{VERB}\rangle &\to \texttt{touches} \mid \texttt{likes} \mid \texttt{sees}
\end{aligned}
$$

$$
\begin{aligned}
\langle\text{SENTENCE}\rangle &\Rightarrow \langle\text{NOUN-PHRASE}\rangle\langle\text{VERB-PHRASE}\rangle \\
&\Rightarrow \langle\text{CMPLX-NOUN}\rangle\langle\text{VERB-PHRASE}\rangle
\end{aligned}
$$

⟨SENTENCE⟩ → ⟨NOUN-PHRASE⟩⟨VERB-PHRASE⟩
⟨NOUN-PHRASE⟩ → ⟨CMPLX-NOUN⟩ | ⟨CMPLX-NOUN⟩⟨PREP-PHRASE⟩
⟨VERB-PHRASE⟩ → ⟨CMPLX-VERB⟩ | ⟨CMPLX-VERB⟩⟨PREP-PHRASE⟩
⟨PREP-PHRASE⟩ → ⟨PREP⟩⟨CMPLX-NOUN⟩
⟨CMPLX-NOUN⟩ → ⟨ARTICLE⟩⟨NOUN⟩
⟨CMPLX-VERB⟩ → ⟨VERB⟩ | ⟨VERB⟩⟨NOUN-PHRASE⟩
⟨ARTICLE⟩ → a | the
⟨NOUN⟩ → boy | girl | flower
⟨VERB⟩ → touches | likes | sees

⟨SENTENCE⟩ ⇒ ⟨NOUN-PHRASE⟩⟨VERB-PHRASE⟩
⇒ ⟨CMPLX-NOUN⟩⟨VERB-PHRASE⟩
⇒ ⟨ARTICLE⟩⟨NOUN⟩⟨VERB-PHRASE⟩

$$
\begin{aligned}
\langle \text{SENTENCE} \rangle &\rightarrow \langle \text{NOUN-PHRASE} \rangle \langle \text{VERB-PHRASE} \rangle \\
\langle \text{NOUN-PHRASE} \rangle &\rightarrow \langle \text{CMPLX-NOUN} \rangle \mid \langle \text{CMPLX-NOUN} \rangle \langle \text{PREP-PHRASE} \rangle \\
\langle \text{VERB-PHRASE} \rangle &\rightarrow \langle \text{CMPLX-VERB} \rangle \mid \langle \text{CMPLX-VERB} \rangle \langle \text{PREP-PHRASE} \rangle \\
\langle \text{PREP-PHRASE} \rangle &\rightarrow \langle \text{PREP} \rangle \langle \text{CMPLX-NOUN} \rangle \\
\langle \text{CMPLX-NOUN} \rangle &\rightarrow \langle \text{ARTICLE} \rangle \langle \text{NOUN} \rangle \\
\langle \text{CMPLX-VERB} \rangle &\rightarrow \langle \text{VERB} \rangle \mid \langle \text{VERB} \rangle \langle \text{NOUN-PHRASE} \rangle \\
\langle \text{ARTICLE} \rangle &\rightarrow \texttt{a} \mid \texttt{the} \\
\langle \text{NOUN} \rangle &\rightarrow \texttt{boy} \mid \texttt{girl} \mid \texttt{flower} \\
\langle \text{VERB} \rangle &\rightarrow \texttt{touches} \mid \texttt{likes} \mid \texttt{sees}
\end{aligned}
$$

$$
\begin{aligned}
\langle \text{SENTENCE} \rangle &\Rightarrow \langle \text{NOUN-PHRASE} \rangle \langle \text{VERB-PHRASE} \rangle \\
&\Rightarrow \langle \text{CMPLX-NOUN} \rangle \langle \text{VERB-PHRASE} \rangle \\
&\Rightarrow \langle \text{ARTICLE} \rangle \langle \text{NOUN} \rangle \langle \text{VERB-PHRASE} \rangle \\
&\Rightarrow \texttt{a} \ \langle \text{NOUN} \rangle \langle \text{VERB-PHRASE} \rangle
\end{aligned}
$$

$$\langle \text{SENTENCE} \rangle \rightarrow \langle \text{NOUN-PHRASE} \rangle \langle \text{VERB-PHRASE} \rangle$$
$$\langle \text{NOUN-PHRASE} \rangle \rightarrow \langle \text{CMPLX-NOUN} \rangle \mid \langle \text{CMPLX-NOUN} \rangle \langle \text{PREP-PHRASE} \rangle$$
$$\langle \text{VERB-PHRASE} \rangle \rightarrow \langle \text{CMPLX-VERB} \rangle \mid \langle \text{CMPLX-VERB} \rangle \langle \text{PREP-PHRASE} \rangle$$
$$\langle \text{PREP-PHRASE} \rangle \rightarrow \langle \text{PREP} \rangle \langle \text{CMPLX-NOUN} \rangle$$
$$\langle \text{CMPLX-NOUN} \rangle \rightarrow \langle \text{ARTICLE} \rangle \langle \text{NOUN} \rangle$$
$$\langle \text{CMPLX-VERB} \rangle \rightarrow \langle \text{VERB} \rangle \mid \langle \text{VERB} \rangle \langle \text{NOUN-PHRASE} \rangle$$
$$\langle \text{ARTICLE} \rangle \rightarrow \texttt{a} \mid \texttt{the}$$
$$\langle \text{NOUN} \rangle \rightarrow \texttt{boy} \mid \texttt{girl} \mid \texttt{flower}$$
$$\langle \text{VERB} \rangle \rightarrow \texttt{touches} \mid \texttt{likes} \mid \texttt{sees}$$

$$\begin{aligned}
\langle \text{SENTENCE} \rangle &\Rightarrow \langle \text{NOUN-PHRASE} \rangle \langle \text{VERB-PHRASE} \rangle \\
&\Rightarrow \langle \text{CMPLX-NOUN} \rangle \langle \text{VERB-PHRASE} \rangle \\
&\Rightarrow \langle \text{ARTICLE} \rangle \langle \text{NOUN} \rangle \langle \text{VERB-PHRASE} \rangle \\
&\Rightarrow \texttt{a}\ \langle \text{NOUN} \rangle \langle \text{VERB-PHRASE} \rangle \\
&\Rightarrow \texttt{a boy}\ \langle \text{VERB-PHRASE} \rangle
\end{aligned}$$

$\langle$SENTENCE$\rangle \rightarrow \langle$NOUN-PHRASE$\rangle\langle$VERB-PHRASE$\rangle$
$\langle$NOUN-PHRASE$\rangle \rightarrow \langle$CMPLX-NOUN$\rangle \mid \langle$CMPLX-NOUN$\rangle\langle$PREP-PHRASE$\rangle$
$\langle$VERB-PHRASE$\rangle \rightarrow \langle$CMPLX-VERB$\rangle \mid \langle$CMPLX-VERB$\rangle\langle$PREP-PHRASE$\rangle$
$\langle$PREP-PHRASE$\rangle \rightarrow \langle$PREP$\rangle\langle$CMPLX-NOUN$\rangle$
$\langle$CMPLX-NOUN$\rangle \rightarrow \langle$ARTICLE$\rangle\langle$NOUN$\rangle$
$\langle$CMPLX-VERB$\rangle \rightarrow \langle$VERB$\rangle \mid \langle$VERB$\rangle\langle$NOUN-PHRASE$\rangle$
$\langle$ARTICLE$\rangle \rightarrow$ a | the
$\langle$NOUN$\rangle \rightarrow$ boy | girl | flower
$\langle$VERB$\rangle \rightarrow$ touches | likes | sees

$\langle$SENTENCE$\rangle \Rightarrow \langle$NOUN-PHRASE$\rangle\langle$VERB-PHRASE$\rangle$
$\Rightarrow \langle$CMPLX-NOUN$\rangle\langle$VERB-PHRASE$\rangle$
$\Rightarrow \langle$ARTICLE$\rangle\langle$NOUN$\rangle\langle$VERB-PHRASE$\rangle$
$\Rightarrow$ a $\langle$NOUN$\rangle\langle$VERB-PHRASE$\rangle$
$\Rightarrow$ a boy $\langle$VERB-PHRASE$\rangle$
$\Rightarrow$ a boy $\langle$CMPLX-VERB$\rangle$

$$\begin{aligned}
\langle\text{SENTENCE}\rangle &\to \langle\text{NOUN-PHRASE}\rangle\langle\text{VERB-PHRASE}\rangle\\
\langle\text{NOUN-PHRASE}\rangle &\to \langle\text{CMPLX-NOUN}\rangle \mid \langle\text{CMPLX-NOUN}\rangle\langle\text{PREP-PHRASE}\rangle\\
\langle\text{VERB-PHRASE}\rangle &\to \langle\text{CMPLX-VERB}\rangle \mid \langle\text{CMPLX-VERB}\rangle\langle\text{PREP-PHRASE}\rangle\\
\langle\text{PREP-PHRASE}\rangle &\to \langle\text{PREP}\rangle\langle\text{CMPLX-NOUN}\rangle\\
\langle\text{CMPLX-NOUN}\rangle &\to \langle\text{ARTICLE}\rangle\langle\text{NOUN}\rangle\\
\langle\text{CMPLX-VERB}\rangle &\to \langle\text{VERB}\rangle \mid \langle\text{VERB}\rangle\langle\text{NOUN-PHRASE}\rangle\\
\langle\text{ARTICLE}\rangle &\to \texttt{a} \mid \texttt{the}\\
\langle\text{NOUN}\rangle &\to \texttt{boy} \mid \texttt{girl} \mid \texttt{flower}\\
\langle\text{VERB}\rangle &\to \texttt{touches} \mid \texttt{likes} \mid \texttt{sees}
\end{aligned}$$

$$\begin{aligned}
\langle\text{SENTENCE}\rangle &\Rightarrow \langle\text{NOUN-PHRASE}\rangle\langle\text{VERB-PHRASE}\rangle\\
&\Rightarrow \langle\text{CMPLX-NOUN}\rangle\langle\text{VERB-PHRASE}\rangle\\
&\Rightarrow \langle\text{ARTICLE}\rangle\langle\text{NOUN}\rangle\langle\text{VERB-PHRASE}\rangle\\
&\Rightarrow \texttt{a}\ \langle\text{NOUN}\rangle\langle\text{VERB-PHRASE}\rangle\\
&\Rightarrow \texttt{a boy}\ \langle\text{VERB-PHRASE}\rangle\\
&\Rightarrow \texttt{a boy}\ \langle\text{CMPLX-VERB}\rangle\\
&\Rightarrow \texttt{a boy}\ \langle\text{VERB}\rangle
\end{aligned}$$

⟨SENTENCE⟩ → ⟨NOUN-PHRASE⟩⟨VERB-PHRASE⟩
⟨NOUN-PHRASE⟩ → ⟨CMPLX-NOUN⟩ | ⟨CMPLX-NOUN⟩⟨PREP-PHRASE⟩
⟨VERB-PHRASE⟩ → ⟨CMPLX-VERB⟩ | ⟨CMPLX-VERB⟩⟨PREP-PHRASE⟩
⟨PREP-PHRASE⟩ → ⟨PREP⟩⟨CMPLX-NOUN⟩
⟨CMPLX-NOUN⟩ → ⟨ARTICLE⟩⟨NOUN⟩
⟨CMPLX-VERB⟩ → ⟨VERB⟩ | ⟨VERB⟩⟨NOUN-PHRASE⟩
⟨ARTICLE⟩ → a | the
⟨NOUN⟩ → boy | girl | flower
⟨VERB⟩ → touches | likes | sees

⟨SENTENCE⟩ ⇒ ⟨NOUN-PHRASE⟩⟨VERB-PHRASE⟩
⇒ ⟨CMPLX-NOUN⟩⟨VERB-PHRASE⟩
⇒ ⟨ARTICLE⟩⟨NOUN⟩⟨VERB-PHRASE⟩
⇒ a ⟨NOUN⟩⟨VERB-PHRASE⟩
⇒ a boy ⟨VERB-PHRASE⟩
⇒ a boy ⟨CMPLX-VERB⟩
⇒ a boy ⟨VERB⟩
⇒ a boy sees

$$
\begin{aligned}
\langle\text{SENTENCE}\rangle &\rightarrow \langle\text{NOUN-PHRASE}\rangle\langle\text{VERB-PHRASE}\rangle \\
\langle\text{NOUN-PHRASE}\rangle &\rightarrow \langle\text{CMPLX-NOUN}\rangle \mid \langle\text{CMPLX-NOUN}\rangle\langle\text{PREP-PHRASE}\rangle \\
\langle\text{VERB-PHRASE}\rangle &\rightarrow \langle\text{CMPLX-VERB}\rangle \mid \langle\text{CMPLX-VERB}\rangle\langle\text{PREP-PHRASE}\rangle \\
\langle\text{PREP-PHRASE}\rangle &\rightarrow \langle\text{PREP}\rangle\langle\text{CMPLX-NOUN}\rangle \\
\langle\text{CMPLX-NOUN}\rangle &\rightarrow \langle\text{ARTICLE}\rangle\langle\text{NOUN}\rangle \\
\langle\text{CMPLX-VERB}\rangle &\rightarrow \langle\text{VERB}\rangle \mid \langle\text{VERB}\rangle\langle\text{NOUN-PHRASE}\rangle \\
\langle\text{ARTICLE}\rangle &\rightarrow \texttt{a} \mid \texttt{the} \\
\langle\text{NOUN}\rangle &\rightarrow \texttt{boy} \mid \texttt{girl} \mid \texttt{flower} \\
\langle\text{VERB}\rangle &\rightarrow \texttt{touches} \mid \texttt{likes} \mid \texttt{sees} \\
\langle\text{PREP}\rangle &\rightarrow \texttt{with}
\end{aligned}
$$

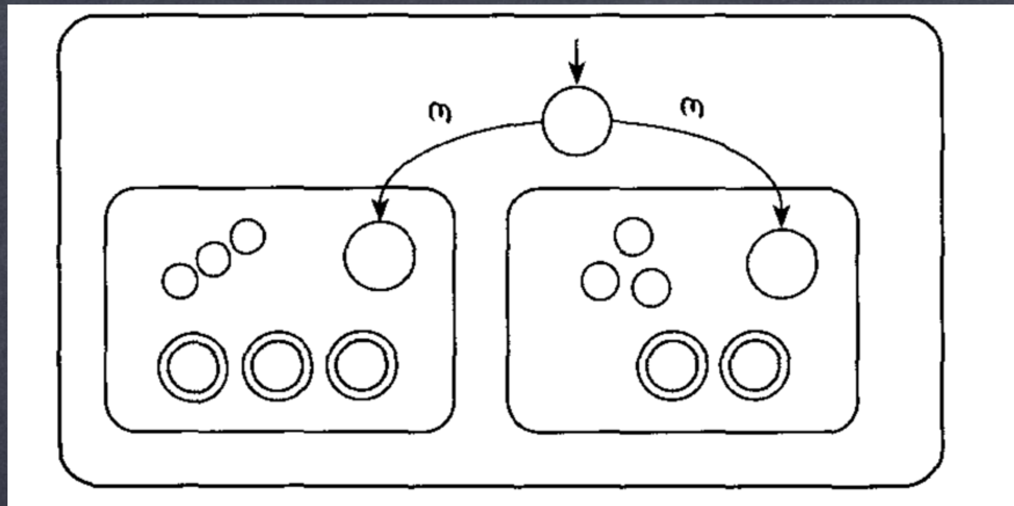
$$\langle\text{SENTENCE}\rangle \overset{*}{\Rightarrow} \texttt{a boy sees}$$

Regular Operations : Kleene's theorem (CFG)

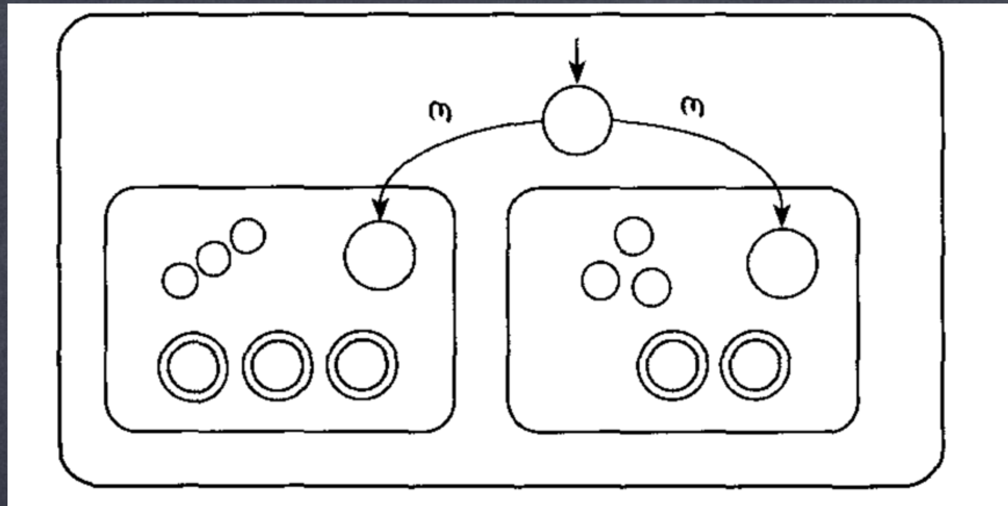
Regular Operations : Kleene's theorem (CFL)

THEOREM

The class of CFLs is closed under the union operation.



Kleene's theorem (CFL)



Kleene's theorem (CFL)

- Let $G_A=(V_A,\Sigma,R_A,S_A)$ be a CFG generating L_A and $G_B=(V_B,\Sigma,R_B,S_B)$ be a CFG generating L_B ($V_A \cap V_B=\varnothing$).
- Consider

– $G_U=(\ \{S_U\} \cup V_A \cup V_B,$

– $\Sigma,$

– $\{S_U \rightarrow S_A \mid S_B\} \cup R_A \cup R_B,$

– $S_U\ ).$

- $L_U = L_A \cup L_B.$

G_1: $V = \{A, B\}$

$\Sigma = \{0, 1, \#\}$

$R_1 = \{A \rightarrow 0A1 \mid B,$

$B \rightarrow \#\}$

$S = A$

R_2:

$$
\begin{aligned}
\langle\text{SENTENCE}\rangle &\rightarrow \langle\text{NOUN-PHRASE}\rangle\langle\text{VERB-PHRASE}\rangle \\
\langle\text{NOUN-PHRASE}\rangle &\rightarrow \langle\text{CMPLX-NOUN}\rangle \mid \langle\text{CMPLX-NOUN}\rangle\langle\text{PREP-PHRASE}\rangle \\
\langle\text{VERB-PHRASE}\rangle &\rightarrow \langle\text{CMPLX-VERB}\rangle \mid \langle\text{CMPLX-VERB}\rangle\langle\text{PREP-PHRASE}\rangle \\
\langle\text{PREP-PHRASE}\rangle &\rightarrow \langle\text{PREP}\rangle\langle\text{CMPLX-NOUN}\rangle \\
\langle\text{CMPLX-NOUN}\rangle &\rightarrow \langle\text{ARTICLE}\rangle\langle\text{NOUN}\rangle \\
\langle\text{CMPLX-VERB}\rangle &\rightarrow \langle\text{VERB}\rangle \mid \langle\text{VERB}\rangle\langle\text{NOUN-PHRASE}\rangle \\
\langle\text{ARTICLE}\rangle &\rightarrow \texttt{a} \mid \texttt{the} \\
\langle\text{NOUN}\rangle &\rightarrow \texttt{boy} \mid \texttt{girl} \mid \texttt{flower} \\
\langle\text{VERB}\rangle &\rightarrow \texttt{touches} \mid \texttt{likes} \mid \texttt{sees} \\
\langle\text{PREP}\rangle &\rightarrow \texttt{with}
\end{aligned}
$$

G_1: $V = \{A,B\}$

$\Sigma = \{0,1,\#\}$

$R_1 = \{A \to 0A1 \mid B,$

$B \to \#\}$

$S = A$

R_2:

$$\begin{aligned}
\langle\text{SENTENCE}\rangle &\to \langle\text{NOUN-PHRASE}\rangle\langle\text{VERB-PHRASE}\rangle \\
\langle\text{NOUN-PHRASE}\rangle &\to \langle\text{CMPLX-NOUN}\rangle \mid \langle\text{CMPLX-NOUN}\rangle\langle\text{PREP-PHRASE}\rangle \\
\langle\text{VERB-PHRASE}\rangle &\to \langle\text{CMPLX-VERB}\rangle \mid \langle\text{CMPLX-VERB}\rangle\langle\text{PREP-PHRASE}\rangle \\
\langle\text{PREP-PHRASE}\rangle &\to \langle\text{PREP}\rangle\langle\text{CMPLX-NOUN}\rangle \\
\langle\text{CMPLX-NOUN}\rangle &\to \langle\text{ARTICLE}\rangle\langle\text{NOUN}\rangle \\
\langle\text{CMPLX-VERB}\rangle &\to \langle\text{VERB}\rangle \mid \langle\text{VERB}\rangle\langle\text{NOUN-PHRASE}\rangle \\
\langle\text{ARTICLE}\rangle &\to \texttt{a} \mid \texttt{the} \\
\langle\text{NOUN}\rangle &\to \texttt{boy} \mid \texttt{girl} \mid \texttt{flower} \\
\langle\text{VERB}\rangle &\to \texttt{touches} \mid \texttt{likes} \mid \texttt{sees} \\
\langle\text{PREP}\rangle &\to \texttt{with}
\end{aligned}$$

- Let **G_1**=({A,B},{0,1,#},R_1,A) be a CFG generating L_1 and **G_2**=({ ⟨SENTENCE⟩,⟨NOUN-PHRASE⟩,⟨VERB-PHRASE⟩,⟨PREP-PHRASE⟩, ⟨CMPLX-NOUN⟩,⟨CMPLX-VERB⟩,⟨ARTICLE⟩,⟨NOUN⟩,⟨VERB⟩,⟨PREP⟩ }, {a,b,c,…,z," "}, R_2 ,⟨SENTENCE⟩) be a CFG generating L_2.

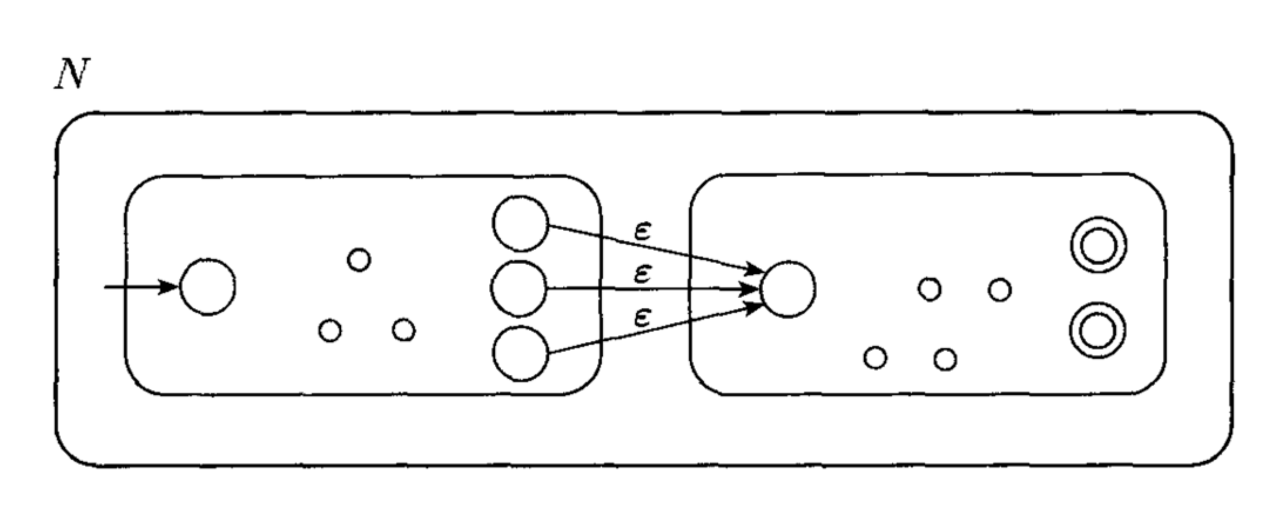
- Let **G_U**=(
  - {S_U,A,B,⟨SENTENCE⟩,⟨NOUN-PHRASE⟩,⟨VERB-PHRASE⟩,⟨PREP-PHRASE⟩, ⟨CMPLX-NOUN⟩,⟨CMPLX-VERB⟩,⟨ARTICLE⟩,⟨NOUN⟩,⟨VERB⟩,⟨PREP⟩},
  - {0,1,#,a,b,c,…,z," "},
  - {$S_U \to A$ | ⟨SENTENCE⟩}$\cup R_1 \cup R_2$,
  - S_U).

- $L_U = L_1 \cup L_2$.

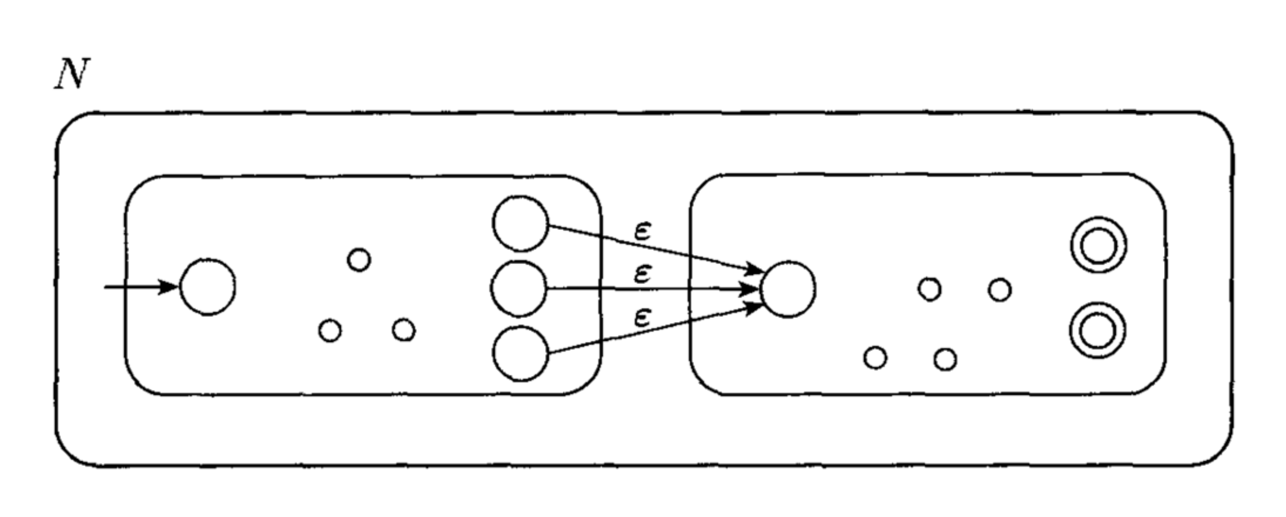
Regular Operations : Kleene's theorem (CFL)

THEOREM

The class of CFLs is closed under the concatenation operation.



Kleene's theorem (CFL)



Kleene's theorem (CFL)

- Let $G_A=(V_A,\Sigma,R_A,S_A)$ be a CFG generating L_A and $G_B=(V_B,\Sigma,R_B,S_B)$ be a CFG generating L_B ($V_A \cap V_B = \emptyset$).
- Consider $G_C=($
  - $\{S_C\} \cup V_A \cup V_B$,
  - Σ,
  - $\{S_C \rightarrow S_A S_B\} \cup R_A \cup R_B$,
  - S_C).
- $L_C = L_A \circ L_B$.

G_1: $V = \{A, B\}$

$\Sigma = \{0, 1, \#\}$

$R_1 = \{A \rightarrow 0A1 \mid B,$

$B \rightarrow \#\}$

$S = A$

R_2:

$$
\begin{aligned}
\langle \text{SENTENCE} \rangle &\rightarrow \langle \text{NOUN-PHRASE} \rangle \langle \text{VERB-PHRASE} \rangle \\
\langle \text{NOUN-PHRASE} \rangle &\rightarrow \langle \text{CMPLX-NOUN} \rangle \mid \langle \text{CMPLX-NOUN} \rangle \langle \text{PREP-PHRASE} \rangle \\
\langle \text{VERB-PHRASE} \rangle &\rightarrow \langle \text{CMPLX-VERB} \rangle \mid \langle \text{CMPLX-VERB} \rangle \langle \text{PREP-PHRASE} \rangle \\
\langle \text{PREP-PHRASE} \rangle &\rightarrow \langle \text{PREP} \rangle \langle \text{CMPLX-NOUN} \rangle \\
\langle \text{CMPLX-NOUN} \rangle &\rightarrow \langle \text{ARTICLE} \rangle \langle \text{NOUN} \rangle \\
\langle \text{CMPLX-VERB} \rangle &\rightarrow \langle \text{VERB} \rangle \mid \langle \text{VERB} \rangle \langle \text{NOUN-PHRASE} \rangle \\
\langle \text{ARTICLE} \rangle &\rightarrow \texttt{a} \mid \texttt{the} \\
\langle \text{NOUN} \rangle &\rightarrow \texttt{boy} \mid \texttt{girl} \mid \texttt{flower} \\
\langle \text{VERB} \rangle &\rightarrow \texttt{touches} \mid \texttt{likes} \mid \texttt{sees} \\
\langle \text{PREP} \rangle &\rightarrow \texttt{with}
\end{aligned}
$$

G_1: $V = \{A, B\}$

$\Sigma = \{0, 1, \#\}$

$R_1 = \{A \rightarrow 0A1 \mid B,$

$B \rightarrow \#\}$

$S = A$

R_2:

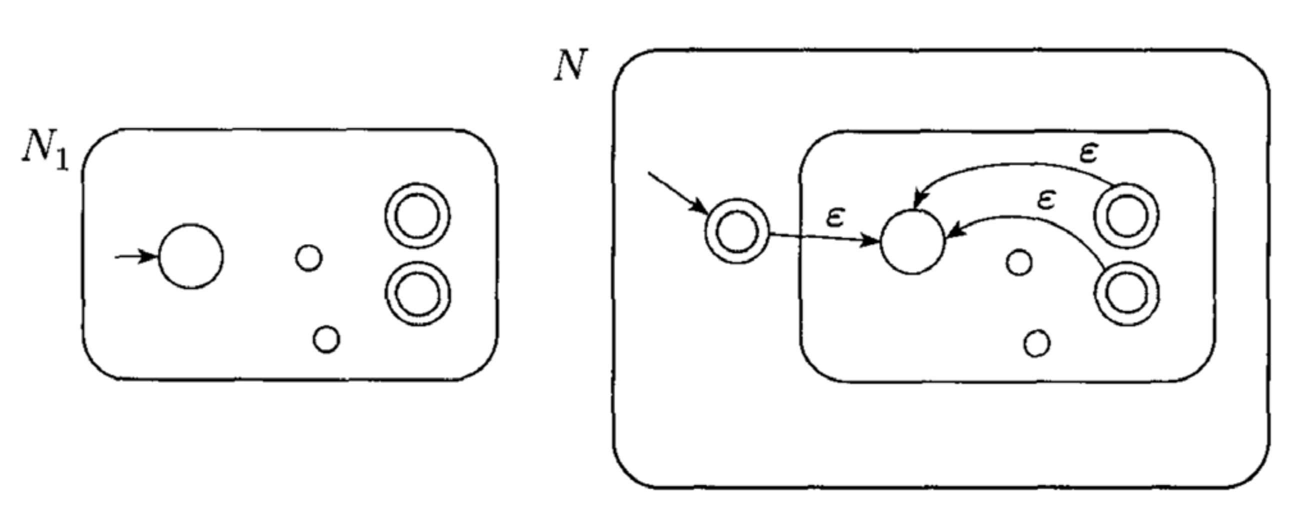
⟨SENTENCE⟩ → ⟨NOUN-PHRASE⟩⟨VERB-PHRASE⟩
⟨NOUN-PHRASE⟩ → ⟨CMPLX-NOUN⟩ | ⟨CMPLX-NOUN⟩⟨PREP-PHRASE⟩
⟨VERB-PHRASE⟩ → ⟨CMPLX-VERB⟩ | ⟨CMPLX-VERB⟩⟨PREP-PHRASE⟩
⟨PREP-PHRASE⟩ → ⟨PREP⟩⟨CMPLX-NOUN⟩
⟨CMPLX-NOUN⟩ → ⟨ARTICLE⟩⟨NOUN⟩
⟨CMPLX-VERB⟩ → ⟨VERB⟩ | ⟨VERB⟩⟨NOUN-PHRASE⟩
⟨ARTICLE⟩ → a | the
⟨NOUN⟩ → boy | girl | flower
⟨VERB⟩ → touches | likes | sees
⟨PREP⟩ → with

- Let G_1=({A,B},{0,1,#},R_1,A) be a CFG generating L_1 and G_2=({ ⟨SENTENCE⟩,⟨NOUN-PHRASE⟩,⟨VERB-PHRASE⟩,⟨PREP-PHRASE⟩, ⟨CMPLX-NOUN⟩,⟨CMPLX-VERB⟩,⟨ARTICLE⟩,⟨NOUN⟩,⟨VERB⟩,⟨PREP⟩ }, {a,b,c,...,z," "}, R_2 ,⟨SENTENCE⟩) be a CFG generating L_2.
- Let G_C=({S_C,A,B,⟨SENTENCE⟩,⟨NOUN-PHRASE⟩,⟨VERB-PHRASE⟩, ⟨PREP-PHRASE⟩,⟨CMPLX-NOUN⟩,⟨CMPLX-VERB⟩,⟨ARTICLE⟩,⟨NOUN⟩, ⟨VERB⟩,⟨PREP⟩}, {0,1,#,a,b,c,...,z," "},
{ $S_C \rightarrow$ A⟨SENTENCE⟩ }$\cup R_1 \cup R_2$,S_C).
- $L_C = L_1 \circ L_2$.

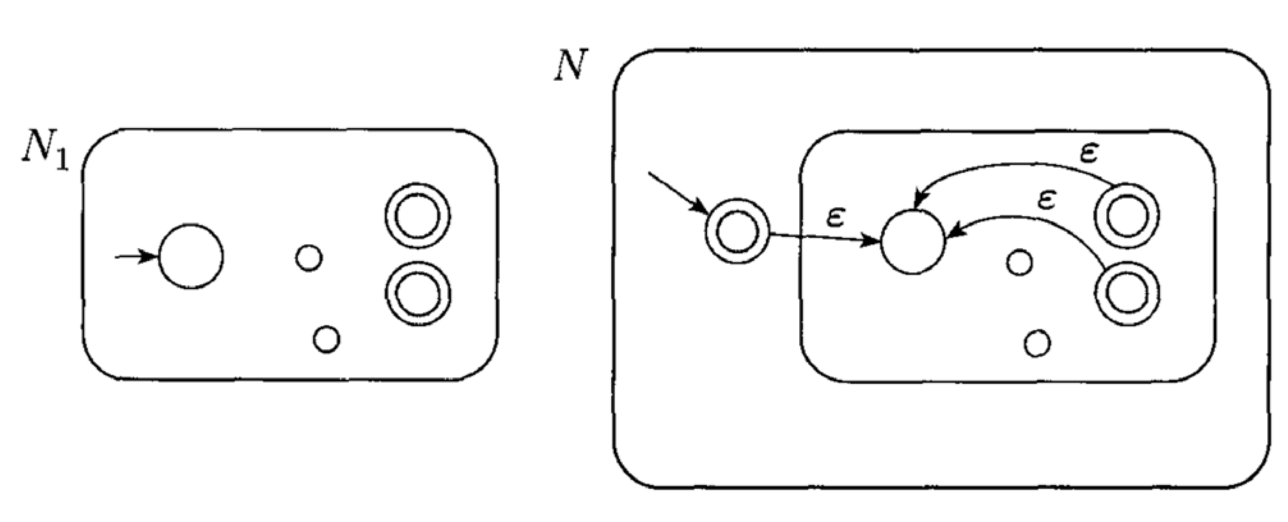
Regular Operations : Kleene's theorem (CFL)

THEOREM

The class of CFLs is closed under the star operation.

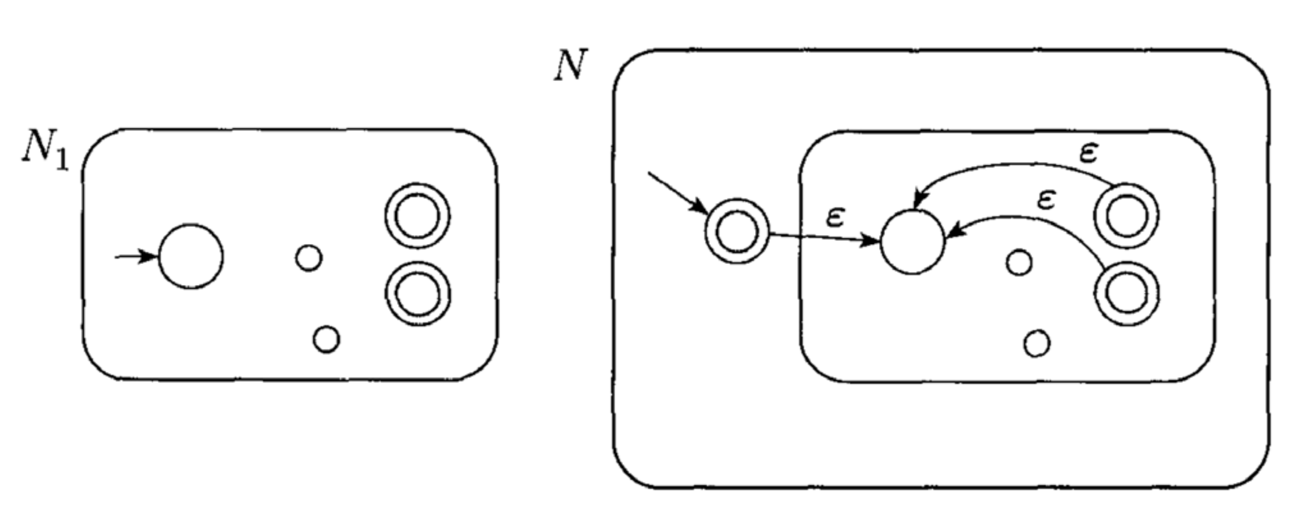


Kleene's theorem (CFL)



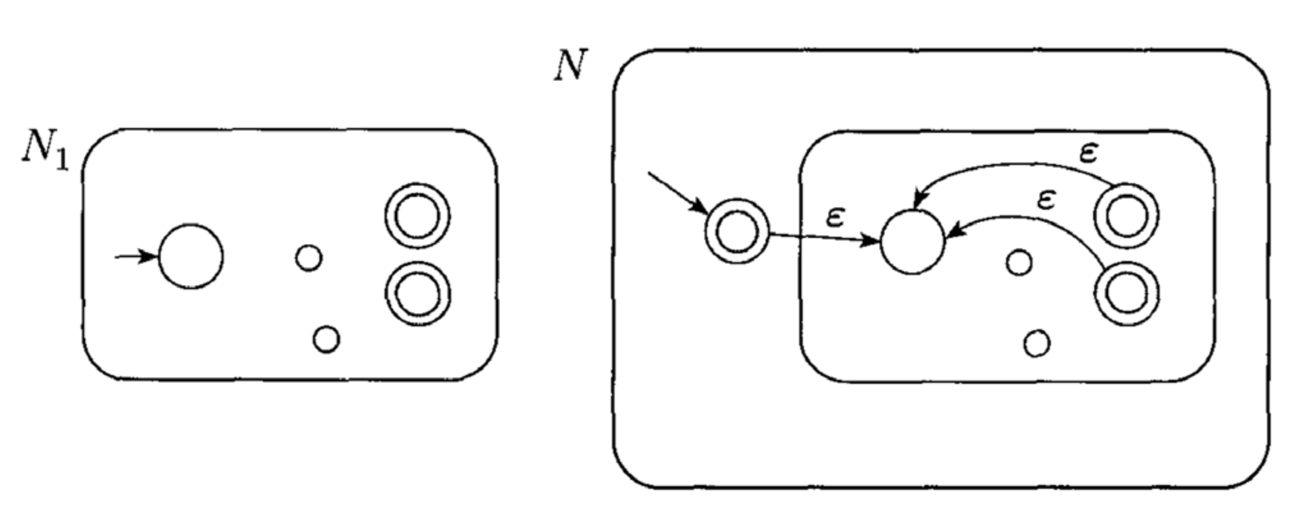
Kleene's theorem (CFL)

- Let $G_A=(V_A,\Sigma,R_A,S_A)$ be a CFG generating L_A.



Kleene's theorem (CFL)

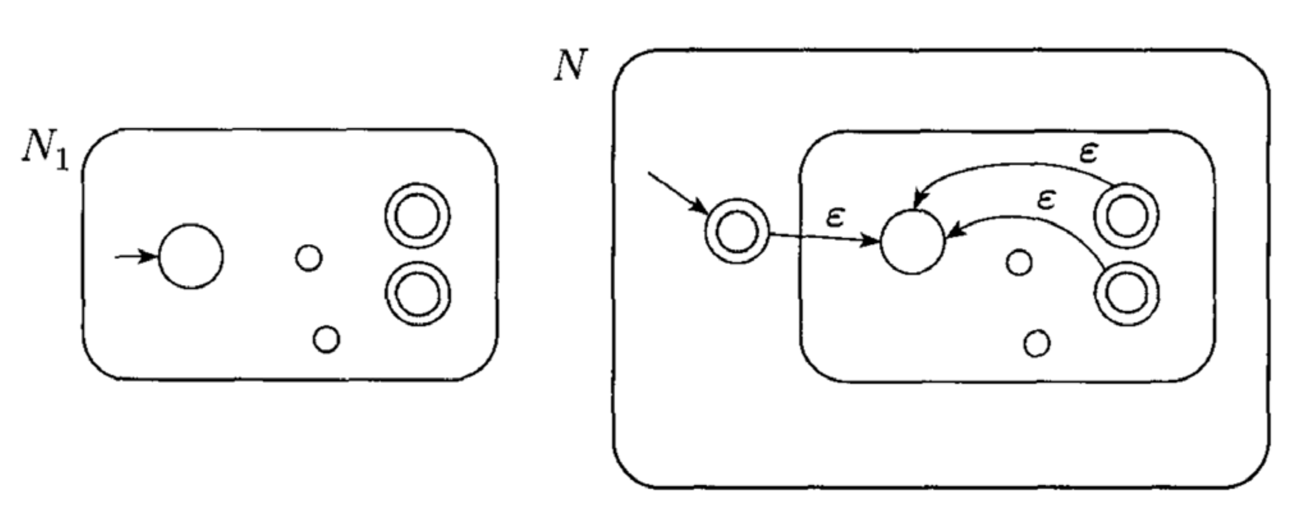
- Let $G_A=(V_A,\Sigma,R_A,S_A)$ be a CFG generating L_A.
- Consider $G_S=($



Kleene's theorem (CFL)

- Let $G_A=(V_A,\Sigma,R_A,S_A)$ be a CFG generating L_A.
- Consider $G_S=($

  $\{S_S\} \cup V_A$,

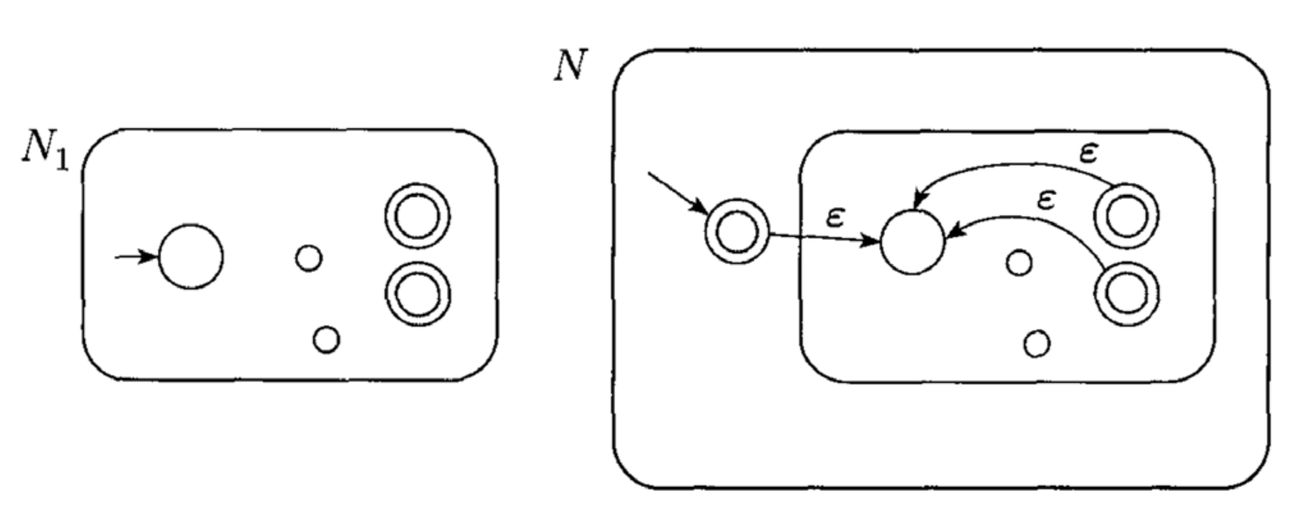


Kleene's theorem (CFL)

- Let $G_A=(V_A,\Sigma,R_A,S_A)$ be a CFG generating L_A.
- Consider $G_S=($

  $\{S_S\} \cup V_A$,

  Σ,



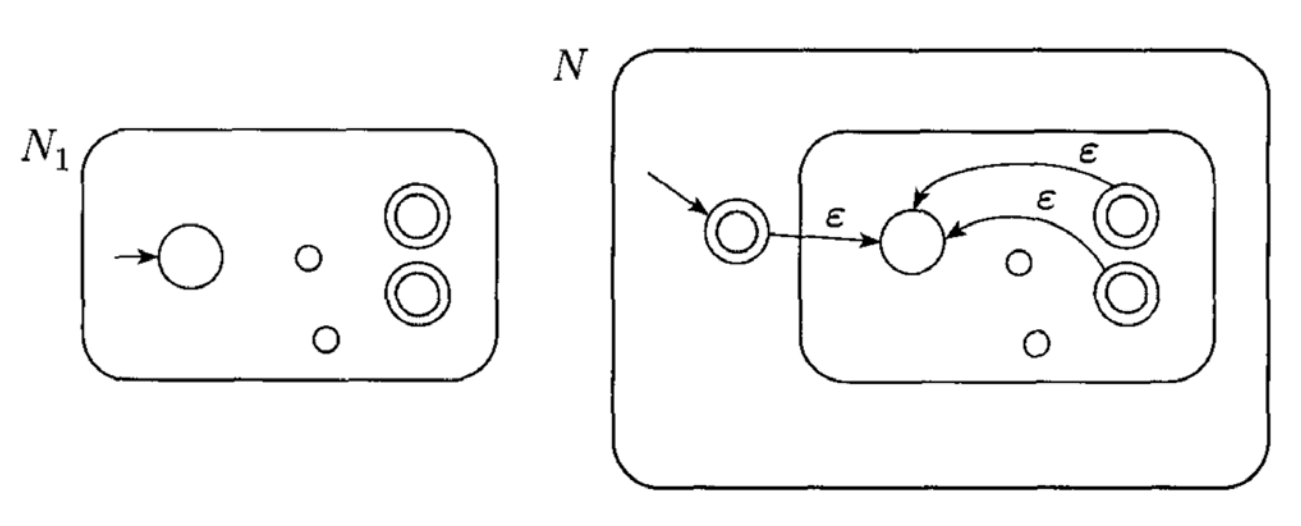
Kleene's theorem (CFL)

- Let $G_A=(V_A,\Sigma,R_A,S_A)$ be a CFG generating L_A.
- Consider $G_S=($

$\{S_S\} \cup V_A$,

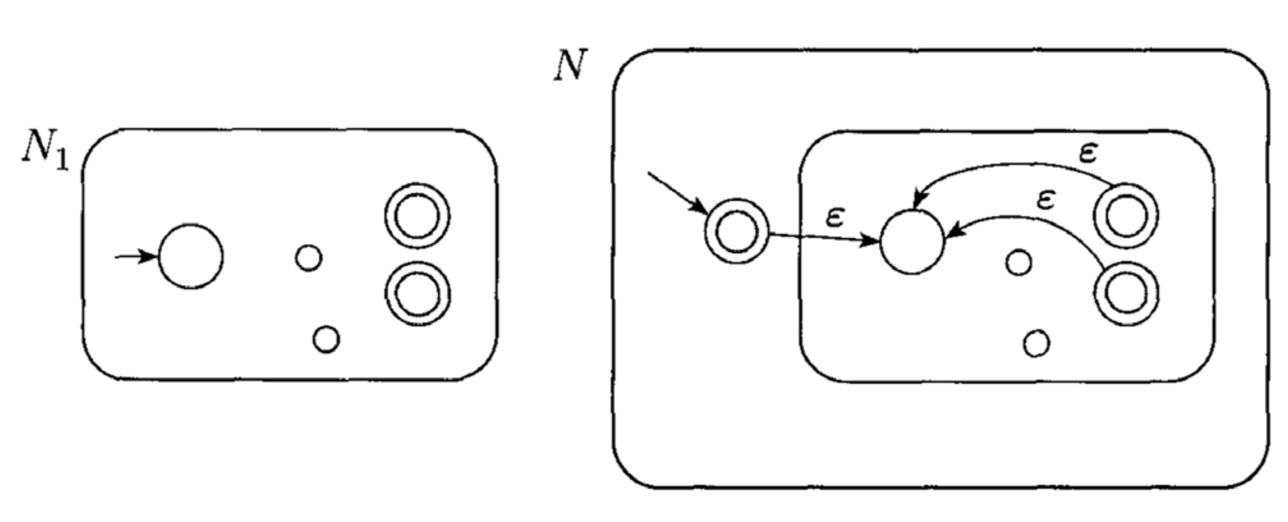
Σ,

$\{S_S \rightarrow \boldsymbol{\varepsilon} \mid S_A S_S\} \cup R_A$,



Kleene's theorem (CFL)

- Let $G_A=(V_A,\Sigma,R_A,S_A)$ be a CFG generating L_A.
- Consider $G_S=($
  - $\{S_S\} \cup V_A$,
  - Σ,
  - $\{S_S \rightarrow \boldsymbol{\varepsilon} \mid S_A S_S\} \cup R_A$,
  - S_S).



Kleene's theorem (CFL)

- Let $G_A=(V_A,\Sigma,R_A,S_A)$ be a CFG generating L_A.
- Consider $G_S=($
  - $\{S_S\} \cup V_A$,
  - Σ,
  - $\{S_S \rightarrow \boldsymbol{\varepsilon} \mid S_A S_S\} \cup R_A$,
  - S_S).
- $L_S = (L_A)^*$.

G_1: $V = \{A, B\}$

$\Sigma = \{0, 1, \#\}$

$R_1 = \{A \rightarrow 0A1 \mid B,$

$B \rightarrow \#\}$

$S = A$

G_1: $V = \{A, B\}$

$\Sigma = \{0, 1, \#\}$

$R_1 = \{A \rightarrow 0A1 \mid B,$

$B \rightarrow \#\}$

$S = A$

- Let $\mathbf{G_1} = (\{A, B\}, \{0, 1, \#\}, R_1, A)$ be a CFG generating L_1.

- Let $\mathbf{G_1}$=({A,B},{0,1,#},R_1,A) be a CFG generating L_1.

G_1: $V = \{A,B\}$

$\Sigma = \{0,1,\#\}$

$R_1 = \{A \rightarrow 0A1 \mid B,$

$B \rightarrow \#\}$

$S = A$

- Let $\mathbf{G_1}=(\{A,B\},\{0,1,\#\},R_1,A)$ be a CFG generating L_1.
- Let

  $\mathbf{G_S}=(\{S_S,A,B\},$

  $\{0,1,\#\},$

  $\{ S_S \rightarrow \boldsymbol{\varepsilon} \mid AS_S,\ A \rightarrow 0A1 \mid B,\ B \rightarrow \#\},$

  $S_S).$
- $L_S = (L_1)^*$.

Construction tools (and Reductions)

CFLs are closed under union, concatenation and star. If there exists a CFL C s. t. either A*=A′, A∪C=A′, A∘C=A′

(**but neither complement nor intersection**)

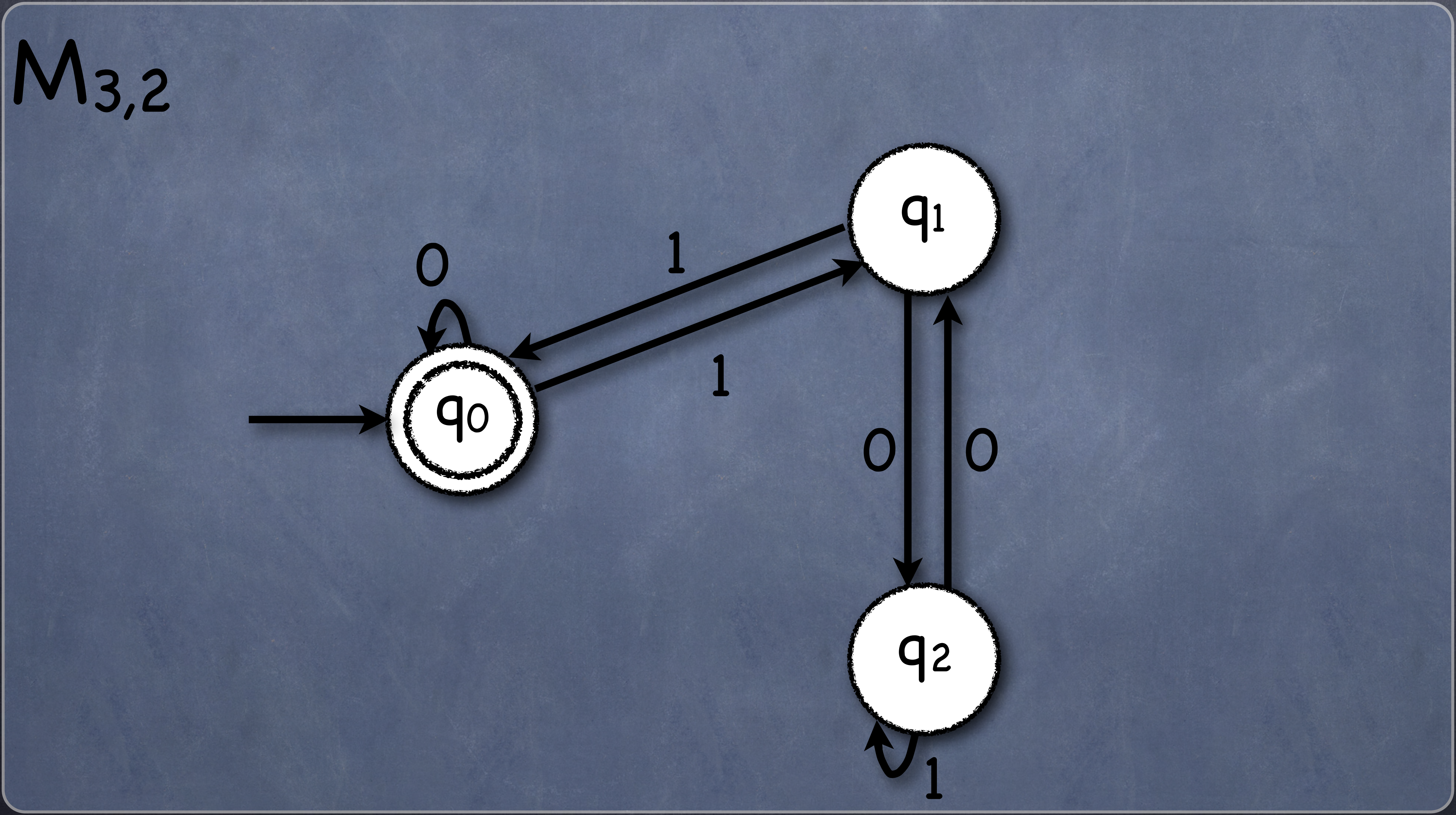
or any combinations of these operations then A′ is a CFL as long as A is.

(If A′ is NON-CFL then so is A.)

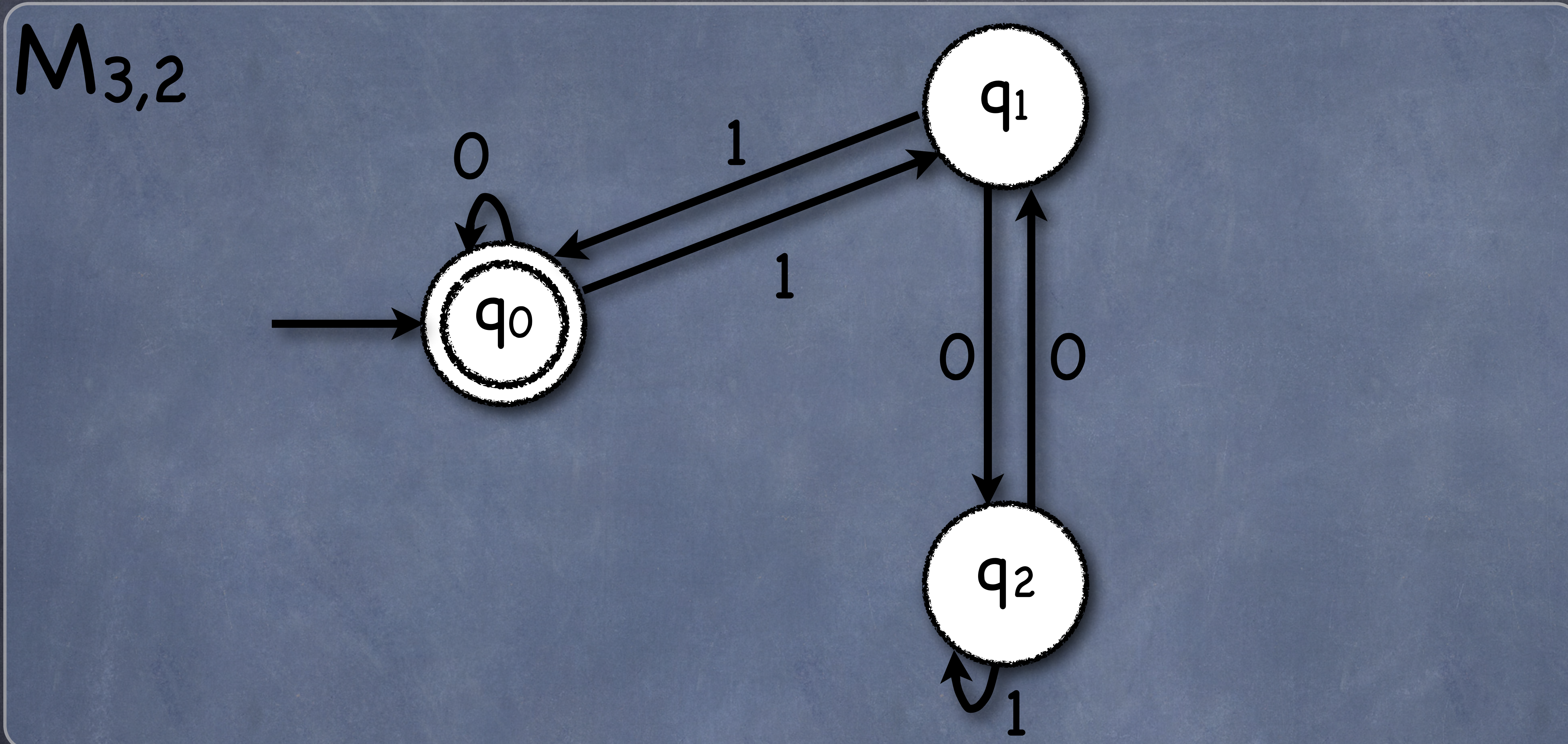
Construction tools

- Constructing a CFG for a regular language L: $M = (Q=\{q_0,q_1,\ldots,q_k\},\Sigma,\delta,q_0,F)$ is converted to $G = (V=\{R_0,R_1,\ldots,R_k\},\Sigma,R,S=R_0)$ where
- R contains rule $R_i \rightarrow aR_j$ for each $\delta(q_i,a) = q_j$ in M, and rule $R_i \rightarrow \varepsilon$ for each accept-state $q_i \in F$.
- R_0 is the start variable.

0 MOD 3 (base 2)



$M_{3,2}$ stops in state $q_r \Leftrightarrow w = r \bmod 3$



- $M_{3,2} = (Q=\{q_0,q_1,q_2\},\{0,1\},\delta,q_0,F)$ is converted to $G_{3,2} = (V=\{R_0,R_1,R_2\},\{0,1\},R,S=R_0)$ where
- R: $R_0 \rightarrow 0R_0 \mid 1R_1 \mid \varepsilon$

  $R_1 \rightarrow 0R_2 \mid 1R_0$

  $R_2 \rightarrow 0R_1 \mid 1R_2$

extra EXAMPLE of CFG

EXAMPLE 2.4

Consider grammar $G_4 = (V, \Sigma, R, \langle \text{EXPR} \rangle)$.

V is $\{\langle \text{EXPR} \rangle, \langle \text{TERM} \rangle, \langle \text{FACTOR} \rangle\}$ and Σ is $\{\texttt{a}, \texttt{+}, \texttt{x}, \texttt{(}, \texttt{)}\}$. The rules are

$$\begin{aligned}
\langle \text{EXPR} \rangle &\rightarrow \langle \text{EXPR} \rangle \texttt{+} \langle \text{TERM} \rangle \mid \langle \text{TERM} \rangle \\
\langle \text{TERM} \rangle &\rightarrow \langle \text{TERM} \rangle \texttt{x} \langle \text{FACTOR} \rangle \mid \langle \text{FACTOR} \rangle \\
\langle \text{FACTOR} \rangle &\rightarrow \texttt{(} \langle \text{EXPR} \rangle \texttt{)} \mid \texttt{a}
\end{aligned}$$

extra EXAMPLE of CFG

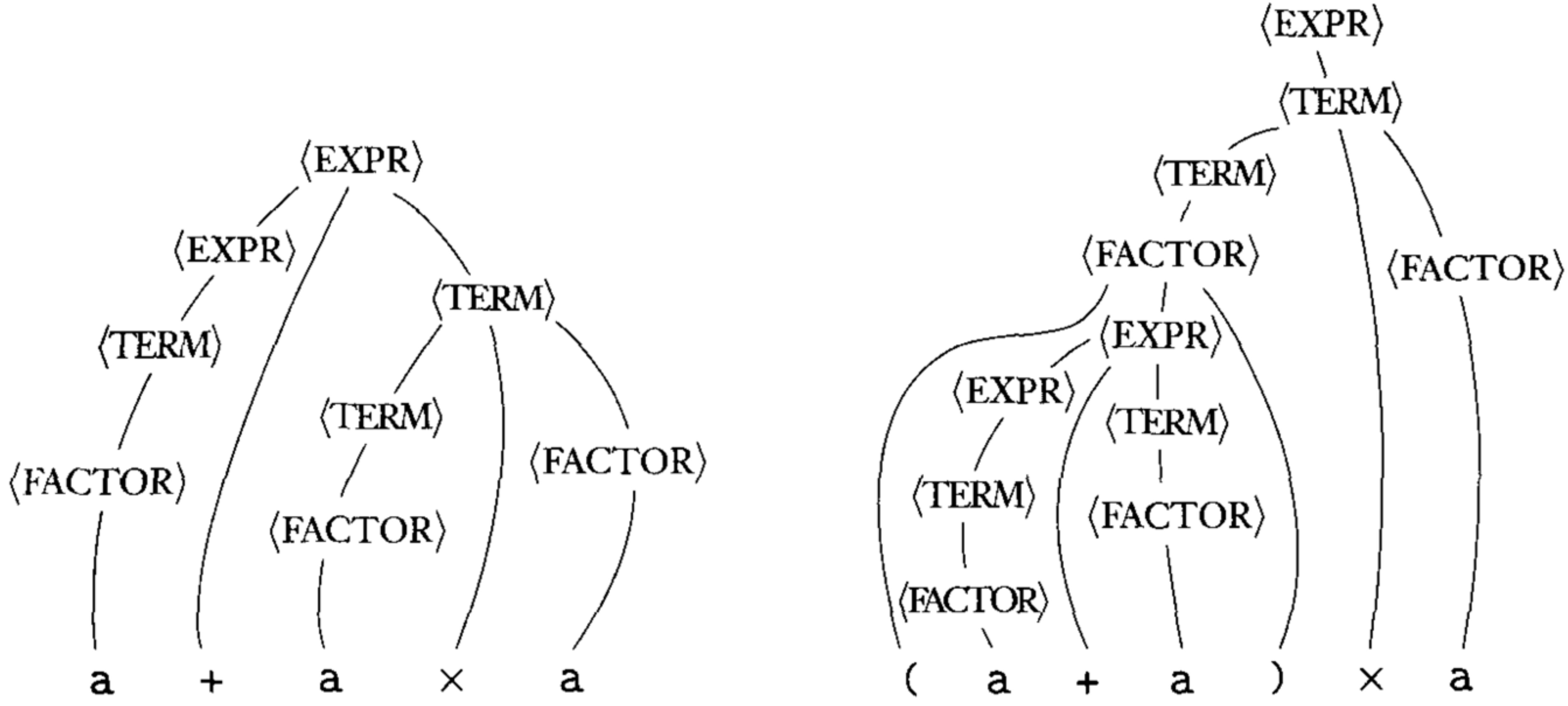


FIGURE 2.5
Parse trees for the strings a+a×a and (a+a)×a

Ambiguity in CFGs

Leftmost Derivation

- A derivation is Leftmost if every time a variable is substituted, it is always the leftmost variable.

EXAMPLE

⟨SENTENCE⟩ ⇒ ⟨NOUN-PHRASE⟩⟨VERB-PHRASE⟩

⇒ ⟨CMPLX-NOUN⟩⟨VERB-PHRASE⟩

⇒ ⟨ARTICLE⟩⟨NOUN⟩⟨VERB-PHRASE⟩

⇒ a ⟨NOUN⟩⟨VERB-PHRASE⟩

⇒ a boy ⟨VERB-PHRASE⟩

⇒ a boy ⟨CMPLX-VERB⟩

⇒ a boy ⟨VERB⟩

⇒ a boy sees

Ambiguity

- A string w is derived ambiguously by a CFG G if it has two or more distinct leftmost derivations. Grammar G is ambigious if it generates some string ambiguously.

Ambiguous version of example 2.4

G_5

$$\langle \text{EXPR} \rangle \rightarrow \langle \text{EXPR} \rangle + \langle \text{EXPR} \rangle \mid \langle \text{EXPR} \rangle \times \langle \text{EXPR} \rangle \mid (\langle \text{EXPR} \rangle) \mid a$$

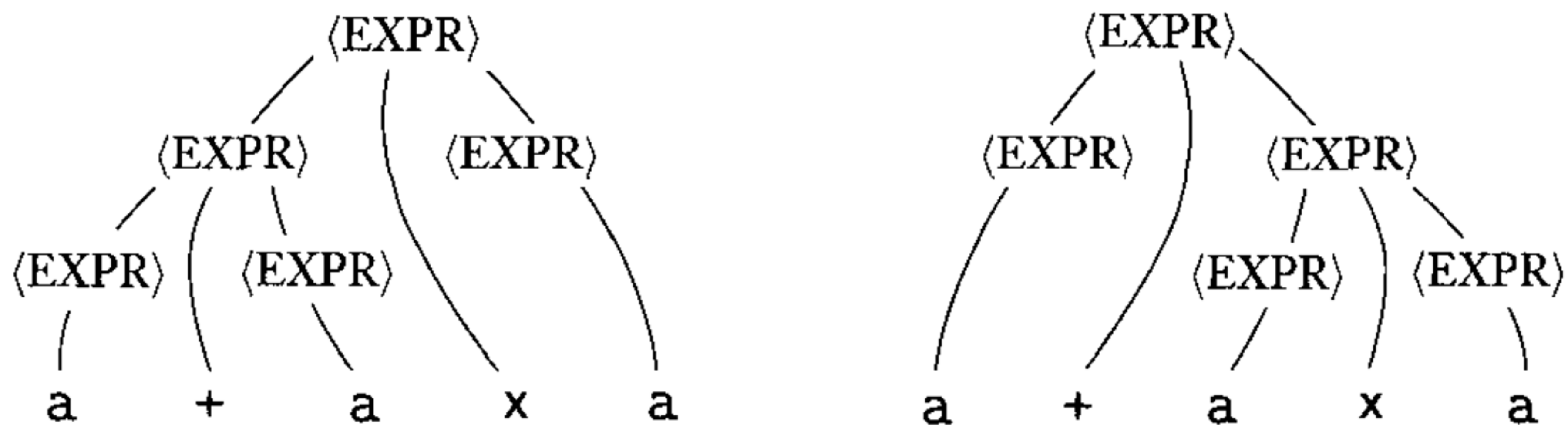


FIGURE 2.6
The two parse trees for the string a+a×a in grammar G_5

Ambiguous CFG

*2.27 Let $G = (V, \Sigma, R, \langle\text{STMT}\rangle)$ be the following grammar.

$$
\begin{aligned}
\langle\text{STMT}\rangle &\to \langle\text{ASSIGN}\rangle \mid \langle\text{IF-THEN}\rangle \mid \langle\text{IF-THEN-ELSE}\rangle \\
\langle\text{IF-THEN}\rangle &\to \texttt{if condition then } \langle\text{STMT}\rangle \\
\langle\text{IF-THEN-ELSE}\rangle &\to \texttt{if condition then } \langle\text{STMT}\rangle \texttt{ else } \langle\text{STMT}\rangle \\
\langle\text{ASSIGN}\rangle &\to \texttt{a:=1}
\end{aligned}
$$

$$
\begin{aligned}
\Sigma &= \{\texttt{if}, \texttt{condition}, \texttt{then}, \texttt{else}, \texttt{a:=1}\}. \\
V &= \{\langle\text{STMT}\rangle, \langle\text{IF-THEN}\rangle, \langle\text{IF-THEN-ELSE}\rangle, \langle\text{ASSIGN}\rangle\}
\end{aligned}
$$

G is a natural-looking grammar for a fragment of a programming language, but G is ambiguous.

a. Show that G is ambiguous.

b. Give a new unambiguous grammar for the same language.

Ambiguity

- Ambiguity is not desirable in CFG because it may lead to unexpected interpretations of a string, for instance in the context of arithmetic expressions or programming languages.
- However, some languages are inherently ambiguous, meaning that all grammars generating this language must be ambiguous.
- example : $\{a^i b^j c^k \mid i=j \text{ or } j=k\}$

Noam Chomsky

Chomsky Normal Form

Chomsky Normal Form

DEFINITION 2.8

A context-free grammar is in ***Chomsky normal form*** if every rule is of the form

$$A \to BC$$
$$A \to a$$

where a is any terminal and A, B, and C are any variables—except that B and C may not be the start variable. In addition we permit the rule $S \to \varepsilon$, where S is the start variable.

Chomsky Normal Form

THEOREM 2.9

Any context-free language is generated by a context-free grammar in Chomsky normal form.

Chomsky Normal Form

THEOREM 2.9

Any context-free language is generated by a context-free grammar in Chomsky normal form.

2.26 Show that, if G is a CFG in Chomsky normal form, then for any string $w \in L(G)$ of length $n \geq 1$, exactly $2n - 1$ steps are required for any derivation of w.

$$A_{\text{CFG}} = \{\langle G, w\rangle \mid G \text{ is a CFG that generates string } w\}.$$

THEOREM 4.7

A_{CFG} is a decidable language.

PROOF IDEA For CFG G and string w we want to determine whether G generates w. One idea is to use G to go through all derivations to determine whether any is a derivation of w. This idea doesn't work, as infinitely many derivations may have to be tried. If G does not generate w, this algorithm would never halt. This idea gives a Turing machine that is a recognizer, but not a decider, for A_{CFG}.

To make this Turing machine into a decider we need to ensure that the algorithm tries only finitely many derivations. In Problem 2.26 (page 157) we showed that, if G were in Chomsky normal form, any derivation of w has $2n - 1$ steps, where n is the length of w. In that case checking only derivations with $2n - 1$ steps to determine whether G generates w would be sufficient. Only finitely many such derivations exist. We can convert G to Chomsky normal form by using the procedure given in Section 2.1.

THEOREM 2.9

Any context-free language is generated by a context-free grammar in Chomsky normal form.

THEOREM 2.9

Any context-free language is generated by a context-free grammar in Chomsky normal form.

- Proof:

THEOREM 2.9

Any context-free language is generated by a context-free grammar in Chomsky normal form.

- Proof:
- First, we add a new start variable S_0 and the rule $S_0 \to S$, where S was the original start variable.

Chomsky Normal Form

EXAMPLE 2.10

Let G_6 be the following CFG and convert it to Chomsky normal form by using the conversion procedure just given. The series of grammars presented illustrates the steps in the conversion. Rules shown in bold have just been added. Rules shown in gray have just been removed.

1. The original CFG G_6 is shown on the left. The result of applying the first step to make a new start variable appears on the right.

$$\begin{aligned} S &\to ASA \mid \mathtt{a}B \\ A &\to B \mid S \\ B &\to \mathtt{b} \mid \varepsilon \end{aligned}$$

$$\begin{aligned} \boldsymbol{S_0} &\boldsymbol{\to S} \\ S &\to ASA \mid \mathtt{a}B \\ A &\to B \mid S \\ B &\to \mathtt{b} \mid \varepsilon \end{aligned}$$

THEOREM 2.9

Any context-free language is generated by a context-free grammar in Chomsky normal form.

THEOREM 2.9

Any context-free language is generated by a context-free grammar in Chomsky normal form.

- Second, we take care of all ε-rules. We remove an ε-rule "$A \to \varepsilon$", where A is not the start variable.
- Then for each occurrence of A on the right-hand side of a rule we add a new rule with that occurrence deleted.
- Accordingly, each rule "$R \to A$" is replaced by "$R \to \varepsilon$" unless it has been already removed.

Chomsky Normal Form

$$\begin{aligned} \boldsymbol{S_0} &\rightarrow \boldsymbol{S} \\ S &\rightarrow ASA \mid \mathtt{a}B \\ A &\rightarrow B \mid S \\ B &\rightarrow \mathtt{b} \mid \varepsilon \end{aligned}$$

2. Remove ε-rules $B \rightarrow \varepsilon$, shown on the left, and $A \rightarrow \varepsilon$, shown on the right.

$$\begin{aligned} S_0 &\rightarrow S \\ S &\rightarrow ASA \mid \mathtt{a}B \mid \mathbf{a} \\ A &\rightarrow B \mid S \mid \boldsymbol{\varepsilon} \\ B &\rightarrow \mathtt{b} \mid \varepsilon \end{aligned}$$

$$\begin{aligned} S_0 &\rightarrow S \\ S &\rightarrow ASA \mid \mathtt{a}B \mid \mathtt{a} \mid \boldsymbol{SA} \mid \boldsymbol{AS} \mid \boldsymbol{S} \\ A &\rightarrow B \mid S \mid \varepsilon \\ B &\rightarrow \mathtt{b} \end{aligned}$$

THEOREM 2.9

Any context-free language is generated by a context-free grammar in Chomsky normal form.

THEOREM 2.9

Any context-free language is generated by a context-free grammar in Chomsky normal form.

- Third, we handle all unit rules by removing each unit rule $A \to B$.
- In consequence whenever $B \to u$ appears, we add the rule $A \to u$ unless this is a unit rule previously removed.

$$
\begin{aligned}
S_0 &\to S \\
S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid \boldsymbol{SA} \mid \boldsymbol{AS} \mid \boldsymbol{S} \\
A &\to B \mid S \\
B &\to \mathtt{b}
\end{aligned}
$$

Chomsky Normal Form

3a. Remove unit rules $S \to S$, shown on the left, and $S_0 \to S$, shown on the right.

$$
\begin{aligned}
S_0 &\to S \\
S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \mid S \\
A &\to B \mid S \\
B &\to \mathtt{b}
\end{aligned}
$$

$$
\begin{aligned}
S_0 &\to S \mid \boldsymbol{ASA} \mid \mathbf{a}\boldsymbol{B} \mid \mathbf{a} \mid \boldsymbol{SA} \mid \boldsymbol{AS} \\
S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\
A &\to B \mid S \\
B &\to \mathtt{b}
\end{aligned}
$$

3b. Remove unit rules $A \to B$ and $A \to S$.

$$
\begin{aligned}
S_0 &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\
S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\
A &\to B \mid S \mid \mathbf{b} \\
B &\to \mathtt{b}
\end{aligned}
$$

$$
\begin{aligned}
S_0 &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\
S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\
A &\to S \mid \mathtt{b} \mid \boldsymbol{ASA} \mid \mathbf{a}\boldsymbol{B} \mid \mathbf{a} \mid \boldsymbol{SA} \mid \boldsymbol{AS} \\
B &\to \mathtt{b}
\end{aligned}
$$

THEOREM 2.9

Any context-free language is generated by a context-free grammar in Chomsky normal form.

THEOREM 2.9

Any context-free language is generated by a context-free grammar in Chomsky normal form.

- Finally, we convert all remaining rules as follows: $A \rightarrow u_1u_2...u_k$ for $k>2$, where each u_i is a variable or terminal with a series of rules $A \rightarrow u_1A_1$, $A_1 \rightarrow u_2A_2,..., A_{k-2} \rightarrow u_{k-1}u_k$ where each A_i is a new variable.
- When $k=2$, and $A \rightarrow u_1u_2$, we may replace any terminal u_i by a variable U_i and the rule $U_i \rightarrow u_i$.

Chomsky Normal Form

$$
\begin{aligned}
S_0 &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\
S &\to ASA \mid \mathtt{a}B \mid \mathtt{a} \mid SA \mid AS \\
A &\to \mathtt{b} \mid \boldsymbol{ASA} \mid \boldsymbol{\mathtt{a}B} \mid \boldsymbol{\mathtt{a}} \mid \boldsymbol{SA} \mid \boldsymbol{AS} \\
B &\to \mathtt{b}
\end{aligned}
$$

4. Convert the remaining rules into the proper form by adding additional variables and rules. The final grammar in Chomsky normal form is equivalent to G_6. (Actually the procedure given in Theorem 2.9 produces several variables U_i and several rules $U_i \to \mathtt{a}$. We simplified the resulting grammar by using a single variable U and rule $U \to \mathtt{a}$.)

$$
\begin{aligned}
S_0 &\to \boxed{AA_1} \mid \boxed{UB} \mid \mathtt{a} \mid SA \mid AS \\
S &\to \boxed{AA_1} \mid \boxed{UB} \mid \mathtt{a} \mid SA \mid AS \\
A &\to \mathtt{b} \mid \boxed{AA_1} \mid \boxed{UB} \mid \mathtt{a} \mid SA \mid AS \\
&\boxed{A_1 \to SA} \\
&\boxed{U \to \mathtt{a}} \\
B &\to \mathtt{b}
\end{aligned}
$$

Chomsky Normal Form

$$
\begin{aligned}
S &\to ASA \mid \mathtt{a}B \\
A &\to B \mid S \\
B &\to \mathtt{b} \mid \varepsilon
\end{aligned}
$$

$$
\begin{aligned}
S_0 &\to AA_1 \mid UB \mid \mathtt{a} \mid SA \mid AS \\
S &\to AA_1 \mid UB \mid \mathtt{a} \mid SA \mid AS \\
A &\to \mathtt{b} \mid AA_1 \mid UB \mid \mathtt{a} \mid SA \mid AS \\
A_1 &\to SA \\
U &\to \mathtt{a} \\
B &\to \mathtt{b}
\end{aligned}
$$

COMP-330
Theory of Computation

Fall 2019 -- Prof. Claude Crépeau

Lec. 10 : Context-Free Grammars