

CS 301

Lecture 10 – Chomsky Normal Form

More CFLs

- $A = \{a^i b^j c^k \mid i \leq j \text{ or } i = k\}$
- $B = \{w \mid w \in \{a, b, c\}^* \text{ contains the same number of as as bs and cs combined}\}$
- $C = \{1^m + 1^n = 1^{m+n} \mid m, n \geq 1\}; \Sigma = \{1, +, =\}$
- $D = \underline{(abb^* \mid bbaa)^*}$
- $E = \{w \mid w \in \{0, 1\}^* \text{ and } w^{\mathcal{R}} \text{ is a binary number not divisible by 5}\}$

Another proof that regular languages are context-free

We can encode the computation of a DFA on a string using a CFG

Give a DFA $M = (Q, \Sigma, \delta, q_0, F)$, we can construct an equivalent CFG $G = (V, \Sigma, R, S)$ where

Another proof that regular languages are context-free

We can encode the computation of a DFA on a string using a CFG

Give a DFA $M = (Q, \Sigma, \delta, q_0, F)$, we can construct an equivalent CFG $G = (V, \Sigma, R, S)$ where

- states of M are variables in G

Another proof that regular languages are context-free

We can encode the computation of a DFA on a string using a CFG

Give a DFA $M = (Q, \Sigma, \delta, q_0, F)$, we can construct an equivalent CFG $G = (V, \Sigma, R, S)$ where

- states of M are variables in G
- q_0 is the start variable, and

Another proof that regular languages are context-free

We can encode the computation of a DFA on a string using a CFG

Give a DFA $M = (Q, \Sigma, \delta, q_0, F)$, we can construct an equivalent CFG $G = (V, \Sigma, R, S)$ where

- states of M are variables in G
- q_0 is the start variable, and
- transitions $\delta(q, t) = r$ become rules $q \rightarrow tr$

Another proof that regular languages are context-free

We can encode the computation of a DFA on a string using a CFG

Give a DFA $M = (Q, \Sigma, \delta, q_0, F)$, we can construct an equivalent CFG $G = (V, \Sigma, R, S)$ where

- states of M are variables in G
- q_0 is the start variable, and
- transitions $\delta(q, t) = r$ become rules $q \rightarrow tr$

If on input $w = w_1w_2\cdots w_n$, M goes through states r_0, r_1, \dots, r_n , then

$$r_0 \Rightarrow w_1r_1 \Rightarrow w_1w_2r_2 \Rightarrow \cdots \Rightarrow w_1w_2\cdots w_nr_n$$

Another proof that regular languages are context-free

We can encode the computation of a DFA on a string using a CFG

Give a DFA $M = (Q, \Sigma, \delta, q_0, F)$, we can construct an equivalent CFG $G = (V, \Sigma, R, S)$ where

- states of M are variables in G
- q_0 is the start variable, and
- transitions $\delta(q, t) = r$ become rules $q \rightarrow tr$

If on input $w = w_1w_2\cdots w_n$, M goes through states r_0, r_1, \dots, r_n , then

$$r_0 \Rightarrow w_1r_1 \Rightarrow w_1w_2r_2 \Rightarrow \cdots \Rightarrow w_1w_2\cdots w_nr_n$$

So G has derived the string wr_n but this still has a variable

What additional rules should we add to end up with a string of terminals?

For each state $q \in F$, add a rule $q \rightarrow \varepsilon$

Formally

Proof.

Given a DFA $M = (Q, \Sigma, \delta, q_0, F)$, we can construct an equivalent CFG $G = (V, \Sigma, R, S)$ where

$$V = Q$$

$$S = q_0$$

$$R = \{q \rightarrow tr : \delta(q, t) = r\} \cup \{q \rightarrow \varepsilon : q \in F\}$$

Formally

Proof.

Given a DFA $M = (Q, \Sigma, \delta, q_0, F)$, we can construct an equivalent CFG $G = (V, \Sigma, R, S)$ where

$$V = Q$$

$$S = q_0$$

$$R = \{q \rightarrow tr : \delta(q, t) = r\} \cup \{q \rightarrow \varepsilon : q \in F\}$$

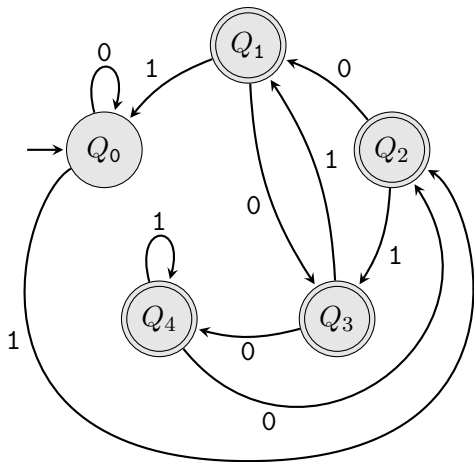
If r_0, r_1, \dots, r_n is the computation of M on input $w = w_1w_2 \cdots w_n$, then $r_0 = q_0$ and $\delta(r_{i-1}, w_i) = r_i$ for $1 \leq i \leq n$

By construction $r_0 \Rightarrow w_1r_1 \Rightarrow w_1w_2r_2 \xRightarrow{*} w_1w_2 \cdots w_nr_n$

Therefore, $w \in L(M)$ iff $r_n \in F$ iff $r_n \Rightarrow \varepsilon$ iff $q_0 \xRightarrow{*} w$ iff $w \in L(G)$ □

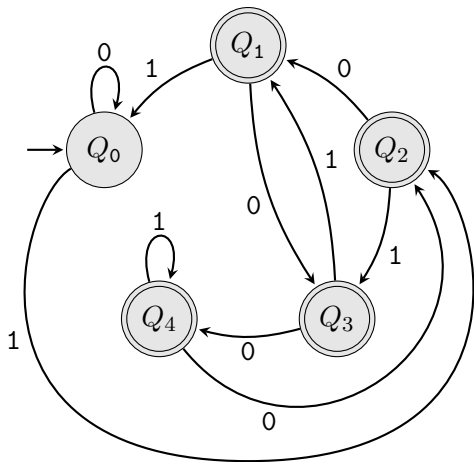
Returning to our language

$E = \{w \mid w \in \{0,1\}^* \text{ and } w^{\mathcal{R}} \text{ is a binary number not divisible by } 5\}$



Returning to our language

$E = \{w \mid w \in \{0,1\}^* \text{ and } w^{\mathcal{R}} \text{ is a binary number not divisible by 5}\}$



$$Q_0 \rightarrow 0Q_0 \mid 1Q_2$$

$$Q_1 \rightarrow 0Q_3 \mid 1Q_0 \mid \varepsilon$$

$$Q_2 \rightarrow 0Q_1 \mid 1Q_3 \mid \varepsilon$$

$$Q_3 \rightarrow 0Q_4 \mid 1Q_1 \mid \varepsilon$$

$$Q_4 \rightarrow 0Q_2 \mid 1Q_4 \mid \varepsilon$$

Chomsky Normal Form (CNF)

A CFG $G = (V, \Sigma, R, S)$ is in **Chomsky Normal Form** if all rules have one of these forms

- $S \rightarrow \varepsilon$ where S is the start variable
- $A \rightarrow BC$ where $A \in V$ and $B, C \in V \setminus \{S\}$
- $A \rightarrow t$ where $A \in V$ and $t \in \Sigma$

Note

- The only rule with ε on the right has the start variable on the left
- The start variable doesn't appear on the right hand side of any rule

CNF example

Let $A = \{w \mid w \in \{a, b\}^* \text{ and } w = w^{\mathcal{R}}\}$.

CFG in CNF

$$S \rightarrow AU \mid BV \mid a \mid b \mid \varepsilon$$

$$T \rightarrow AU \mid BV \mid a \mid b$$

$$U \rightarrow TA$$

$$V \rightarrow TB$$

$$A \rightarrow a$$

$$B \rightarrow b$$

Derivation of baaab

S

CNF example

Let $A = \{w \mid w \in \{a, b\}^* \text{ and } w = w^{\mathcal{R}}\}$.

CFG in CNF

$$S \rightarrow AU \mid BV \mid a \mid b \mid \varepsilon$$

$$T \rightarrow AU \mid BV \mid a \mid b$$

$$U \rightarrow TA$$

$$V \rightarrow TB$$

$$A \rightarrow a$$

$$B \rightarrow b$$

Derivation of baaab

$$S \Rightarrow BV$$

CNF example

Let $A = \{w \mid w \in \{a, b\}^* \text{ and } w = w^{\mathcal{R}}\}$.

CFG in CNF

$$S \rightarrow AU \mid BV \mid a \mid b \mid \varepsilon$$

$$T \rightarrow AU \mid BV \mid a \mid b$$

$$U \rightarrow TA$$

$$V \rightarrow TB$$

$$A \rightarrow a$$

$$B \rightarrow b$$

Derivation of baaab

$$S \Rightarrow BV$$

$$\Rightarrow bV$$

CNF example

Let $A = \{w \mid w \in \{a, b\}^* \text{ and } w = w^R\}$.

CFG in CNF

$$S \rightarrow AU \mid BV \mid a \mid b \mid \varepsilon$$

$$T \rightarrow AU \mid BV \mid a \mid b$$

$$U \rightarrow TA$$

$$V \rightarrow TB$$

$$A \rightarrow a$$

$$B \rightarrow b$$

Derivation of baaab

$$S \Rightarrow BV$$

$$\Rightarrow bV$$

$$\Rightarrow bTB$$

CNF example

Let $A = \{w \mid w \in \{a, b\}^* \text{ and } w = w^R\}$.

CFG in CNF

$$S \rightarrow AU \mid BV \mid a \mid b \mid \varepsilon$$

$$T \rightarrow AU \mid BV \mid a \mid b$$

$$U \rightarrow TA$$

$$V \rightarrow TB$$

$$A \rightarrow a$$

$$B \rightarrow b$$

Derivation of baaab

$$S \Rightarrow BV$$

$$\Rightarrow bV$$

$$\Rightarrow bTB$$

$$\Rightarrow bAUB$$

CNF example

Let $A = \{w \mid w \in \{a, b\}^* \text{ and } w = w^R\}$.

CFG in CNF

$$S \rightarrow AU \mid BV \mid a \mid b \mid \varepsilon$$

$$T \rightarrow AU \mid BV \mid a \mid b$$

$$U \rightarrow TA$$

$$V \rightarrow TB$$

$$A \rightarrow a$$

$$B \rightarrow b$$

Derivation of baaab

$$S \Rightarrow BV$$

$$\Rightarrow bV$$

$$\Rightarrow bTB$$

$$\Rightarrow bAUB$$

$$\Rightarrow baUB$$

CNF example

Let $A = \{w \mid w \in \{a, b\}^* \text{ and } w = w^R\}$.

CFG in CNF

$$S \rightarrow AU \mid BV \mid a \mid b \mid \varepsilon$$

$$T \rightarrow AU \mid BV \mid a \mid b$$

$$U \rightarrow TA$$

$$V \rightarrow TB$$

$$A \rightarrow a$$

$$B \rightarrow b$$

Derivation of baaab

$$S \Rightarrow BV$$

$$\Rightarrow bV$$

$$\Rightarrow bTB$$

$$\Rightarrow bAUB$$

$$\Rightarrow baUB$$

$$\Rightarrow baTAB$$

CNF example

Let $A = \{w \mid w \in \{a, b\}^* \text{ and } w = w^R\}$.

CFG in CNF

$$S \rightarrow AU \mid BV \mid a \mid b \mid \varepsilon$$

$$T \rightarrow AU \mid BV \mid a \mid b$$

$$U \rightarrow TA$$

$$V \rightarrow TB$$

$$A \rightarrow a$$

$$B \rightarrow b$$

Derivation of baaab

$$S \Rightarrow BV$$

$$\Rightarrow bV$$

$$\Rightarrow bTB$$

$$\Rightarrow bAUB$$

$$\Rightarrow baUB$$

$$\Rightarrow baTAB$$

$$\Rightarrow baaAB$$

CNF example

Let $A = \{w \mid w \in \{a, b\}^* \text{ and } w = w^{\mathcal{R}}\}$.

CFG in CNF

$$S \rightarrow AU \mid BV \mid a \mid b \mid \varepsilon$$

$$T \rightarrow AU \mid BV \mid a \mid b$$

$$U \rightarrow TA$$

$$V \rightarrow TB$$

$$A \rightarrow a$$

$$B \rightarrow b$$

Derivation of baaab

$$S \Rightarrow BV$$

$$\Rightarrow bV$$

$$\Rightarrow bTB$$

$$\Rightarrow bAUB$$

$$\Rightarrow baUB$$

$$\Rightarrow baTAB$$

$$\Rightarrow baaAB$$

$$\Rightarrow baaaB$$

CNF example

Let $A = \{w \mid w \in \{a, b\}^* \text{ and } w = w^{\mathcal{R}}\}$.

CFG in CNF

$$S \rightarrow AU \mid BV \mid a \mid b \mid \varepsilon$$

$$T \rightarrow AU \mid BV \mid a \mid b$$

$$U \rightarrow TA$$

$$V \rightarrow TB$$

$$A \rightarrow a$$

$$B \rightarrow b$$

Derivation of baaab

$$S \Rightarrow BV$$

$$\Rightarrow bV$$

$$\Rightarrow bTB$$

$$\Rightarrow bAUB$$

$$\Rightarrow baUB$$

$$\Rightarrow baTAB$$

$$\Rightarrow baaAB$$

$$\Rightarrow baaaB$$

$$\Rightarrow baaab$$

Converting to CNF

Theorem

Every context-free language A is generated by some CFG in CNF.

Proof.

Given a CFG $G = (V, \Sigma, R, S)$ generating A , we construct a new CFG $G' = (V', \Sigma, R', S')$ in CNF generating A .

There are five steps.

START Add a new start variable

BIN Replace rules with RHS longer than two with multiple rules each of which has a RHS of length two

DEL- ϵ Remove all ϵ -rules ($A \rightarrow \epsilon$)

UNIT Remove all unit-rules ($A \rightarrow B$)

TERM Add a variable and rule for each terminal ($T \rightarrow t$) and replace terminals on the RHS of rules

Proof continued

In the following $x \in V \cup \Sigma$ and $u \in (\Sigma \cup V)^+$

START Add a new start variable S' and a rule $S' \rightarrow S$

Proof continued

In the following $x \in V \cup \Sigma$ and $u \in (\Sigma \cup V)^+$

START Add a new start variable S' and a rule $S' \rightarrow S$

BIN Replace each rule $A \rightarrow xu$ with the rules $A \rightarrow xA_1$ and $A_1 \rightarrow u$ and repeat until the RHS of every rule has length at most two

Proof continued

In the following $x \in V \cup \Sigma$ and $u \in (\Sigma \cup V)^+$

START Add a new start variable S' and a rule $S' \rightarrow S$

BIN Replace each rule $A \rightarrow xu$ with the rules $A \rightarrow xA_1$ and $A_1 \rightarrow u$ and repeat until the RHS of every rule has length at most two

DEL- ε For each rule of the form $A \rightarrow \varepsilon$ other than $S' \rightarrow \varepsilon$ remove $A \rightarrow \varepsilon$ and update all rules with A in the RHS

Proof continued

In the following $x \in V \cup \Sigma$ and $u \in (\Sigma \cup V)^+$

START Add a new start variable S' and a rule $S' \rightarrow S$

BIN Replace each rule $A \rightarrow xu$ with the rules $A \rightarrow xA_1$ and $A_1 \rightarrow u$ and repeat until the RHS of every rule has length at most two

DEL- ε For each rule of the form $A \rightarrow \varepsilon$ other than $S' \rightarrow \varepsilon$ remove $A \rightarrow \varepsilon$ and update all rules with A in the RHS

- $B \rightarrow A$. Add rule $B \rightarrow \varepsilon$ unless $B \rightarrow \varepsilon$ has already been removed

Proof continued

In the following $x \in V \cup \Sigma$ and $u \in (\Sigma \cup V)^+$

START Add a new start variable S' and a rule $S' \rightarrow S$

BIN Replace each rule $A \rightarrow xu$ with the rules $A \rightarrow xA_1$ and $A_1 \rightarrow u$ and repeat until the RHS of every rule has length at most two

DEL- ε For each rule of the form $A \rightarrow \varepsilon$ other than $S' \rightarrow \varepsilon$ remove $A \rightarrow \varepsilon$ and update all rules with A in the RHS

- $B \rightarrow A$. Add rule $B \rightarrow \varepsilon$ unless $B \rightarrow \varepsilon$ has already been removed
- $B \rightarrow AA$. Add rule $B \rightarrow A$ and if $B \rightarrow \varepsilon$ has not already been removed, add it

Proof continued

In the following $x \in V \cup \Sigma$ and $u \in (\Sigma \cup V)^+$

START Add a new start variable S' and a rule $S' \rightarrow S$

BIN Replace each rule $A \rightarrow xu$ with the rules $A \rightarrow xA_1$ and $A_1 \rightarrow u$ and repeat until the RHS of every rule has length at most two

DEL- ε For each rule of the form $A \rightarrow \varepsilon$ other than $S' \rightarrow \varepsilon$ remove $A \rightarrow \varepsilon$ and update all rules with A in the RHS

- $B \rightarrow A$. Add rule $B \rightarrow \varepsilon$ unless $B \rightarrow \varepsilon$ has already been removed
- $B \rightarrow AA$. Add rule $B \rightarrow A$ and if $B \rightarrow \varepsilon$ has not already been removed, add it
- $B \rightarrow xA$ or $B \rightarrow Ax$. Add rule $B \rightarrow x$

Proof continued

In the following $x \in V \cup \Sigma$ and $u \in (\Sigma \cup V)^+$

START Add a new start variable S' and a rule $S' \rightarrow S$

BIN Replace each rule $A \rightarrow xu$ with the rules $A \rightarrow xA_1$ and $A_1 \rightarrow u$ and repeat until the RHS of every rule has length at most two

DEL- ε For each rule of the form $A \rightarrow \varepsilon$ other than $S' \rightarrow \varepsilon$ remove $A \rightarrow \varepsilon$ and update all rules with A in the RHS

- $B \rightarrow A$. Add rule $B \rightarrow \varepsilon$ unless $B \rightarrow \varepsilon$ has already been removed
- $B \rightarrow AA$. Add rule $B \rightarrow A$ and if $B \rightarrow \varepsilon$ has not already been removed, add it
- $B \rightarrow xA$ or $B \rightarrow Ax$. Add rule $B \rightarrow x$

UNIT For each rule $A \rightarrow B$, remove it and add rules $A \rightarrow u$ for each $B \rightarrow u$ unless $A \rightarrow u$ is a unit rule already removed

Proof continued

In the following $x \in V \cup \Sigma$ and $u \in (\Sigma \cup V)^+$

START Add a new start variable S' and a rule $S' \rightarrow S$

BIN Replace each rule $A \rightarrow xu$ with the rules $A \rightarrow xA_1$ and $A_1 \rightarrow u$ and repeat until the RHS of every rule has length at most two

DEL- ε For each rule of the form $A \rightarrow \varepsilon$ other than $S' \rightarrow \varepsilon$ remove $A \rightarrow \varepsilon$ and update all rules with A in the RHS

- $B \rightarrow A$. Add rule $B \rightarrow \varepsilon$ unless $B \rightarrow \varepsilon$ has already been removed
- $B \rightarrow AA$. Add rule $B \rightarrow A$ and if $B \rightarrow \varepsilon$ has not already been removed, add it
- $B \rightarrow xA$ or $B \rightarrow Ax$. Add rule $B \rightarrow x$

UNIT For each rule $A \rightarrow B$, remove it and add rules $A \rightarrow u$ for each $B \rightarrow u$ unless $A \rightarrow u$ is a unit rule already removed

TERM For each $t \in \Sigma$, add a new variable T and a rule $T \rightarrow t$; replace each t in the RHS of nonunit rules with T

Proof continued

In the following $x \in V \cup \Sigma$ and $u \in (\Sigma \cup V)^+$

START Add a new start variable S' and a rule $S' \rightarrow S$

BIN Replace each rule $A \rightarrow xu$ with the rules $A \rightarrow xA_1$ and $A_1 \rightarrow u$ and repeat until the RHS of every rule has length at most two

DEL- ε For each rule of the form $A \rightarrow \varepsilon$ other than $S' \rightarrow \varepsilon$ remove $A \rightarrow \varepsilon$ and update all rules with A in the RHS

- $B \rightarrow A$. Add rule $B \rightarrow \varepsilon$ unless $B \rightarrow \varepsilon$ has already been removed
- $B \rightarrow AA$. Add rule $B \rightarrow A$ and if $B \rightarrow \varepsilon$ has not already been removed, add it
- $B \rightarrow xA$ or $B \rightarrow Ax$. Add rule $B \rightarrow x$

UNIT For each rule $A \rightarrow B$, remove it and add rules $A \rightarrow u$ for each $B \rightarrow u$ unless $A \rightarrow u$ is a unit rule already removed

TERM For each $t \in \Sigma$, add a new variable T and a rule $T \rightarrow t$; replace each t in the RHS of nonunit rules with T

Each of the five steps preserves the language generated by the grammar so $L(G') = A$.



Example

Convert to CNF

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

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

START:

Example

Convert to CNF

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

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

START:

$$S \rightarrow A$$

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

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

Example

Convert to CNF

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

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

START:

$$S \rightarrow A$$

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

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

BIN: Replace $A \rightarrow BAB$:

Example

Convert to CNF

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

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

START:

$$S \rightarrow A$$

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

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

BIN: Replace $A \rightarrow BAB$:

$$S \rightarrow A$$

$$A \rightarrow BA_1 \mid B \mid \varepsilon$$

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

$$A_1 \rightarrow AB$$

Example

Convert to CNF

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

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

DEL- ε : Remove $A \rightarrow \varepsilon$:

START:

$$S \rightarrow A$$

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

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

BIN: Replace $A \rightarrow BAB$:

$$S \rightarrow A$$

$$A \rightarrow BA_1 \mid B \mid \varepsilon$$

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

$$A_1 \rightarrow AB$$

Example

Convert to CNF

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

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

START:

$$S \rightarrow A$$

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

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

BIN: Replace $A \rightarrow BAB$:

$$S \rightarrow A$$

$$A \rightarrow BA_1 \mid B \mid \varepsilon$$

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

$$A_1 \rightarrow AB$$

DEL- ε : Remove $A \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B$$

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

$$A_1 \rightarrow AB \mid B$$

Example

Convert to CNF

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

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

START:

$$S \rightarrow A$$

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

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

BIN: Replace $A \rightarrow BAB$:

$$S \rightarrow A$$

$$A \rightarrow BA_1 \mid B \mid \varepsilon$$

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

$$A_1 \rightarrow AB$$

DEL- ε : Remove $A \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B$$

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

$$A_1 \rightarrow AB \mid B$$

Remove $B \rightarrow \varepsilon$:

Example

Convert to CNF

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

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

START:

$$S \rightarrow A$$

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

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

BIN: Replace $A \rightarrow BAB$:

$$S \rightarrow A$$

$$A \rightarrow BA_1 \mid B \mid \varepsilon$$

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

$$A_1 \rightarrow AB$$

DEL- ε : Remove $A \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B$$

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

$$A_1 \rightarrow AB \mid B$$

Remove $B \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A \mid \varepsilon$$

Don't add $A \rightarrow \varepsilon$ because we already removed it

Example

Convert to CNF

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

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

START:

$$S \rightarrow A$$

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

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

BIN: Replace $A \rightarrow BAB$:

$$S \rightarrow A$$

$$A \rightarrow BA_1 \mid B \mid \varepsilon$$

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

$$A_1 \rightarrow AB$$

DEL- ε : Remove $A \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B$$

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

$$A_1 \rightarrow AB \mid B$$

Remove $A_1 \rightarrow \varepsilon$:

Remove $B \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A \mid \varepsilon$$

Don't add $A \rightarrow \varepsilon$ because we already removed it

Example

Convert to CNF

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

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

START:

$$S \rightarrow A$$

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

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

BIN: Replace $A \rightarrow BAB$:

$$S \rightarrow A$$

$$A \rightarrow BA_1 \mid B \mid \varepsilon$$

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

$$A_1 \rightarrow AB$$

DEL- ε : Remove $A \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B$$

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

$$A_1 \rightarrow AB \mid B$$

Remove $B \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A \mid \varepsilon$$

Don't add $A \rightarrow \varepsilon$ because we already removed it

Remove $A_1 \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Don't add $A \rightarrow \varepsilon$ because we already removed it

Example

Convert to CNF

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

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

START:

$$S \rightarrow A$$

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

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

BIN: Replace $A \rightarrow BAB$:

$$S \rightarrow A$$

$$A \rightarrow BA_1 \mid B \mid \varepsilon$$

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

$$A_1 \rightarrow AB$$

DEL- ε : Remove $A \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B$$

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

$$A_1 \rightarrow AB \mid B$$

Remove $B \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A \mid \varepsilon$$

Don't add $A \rightarrow \varepsilon$ because we already removed it

Remove $A_1 \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Don't add $A \rightarrow \varepsilon$ because we already removed it

UNIT: Remove $S \rightarrow A$

Example

Convert to CNF

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

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

START:

$$S \rightarrow A$$

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

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

BIN: Replace $A \rightarrow BAB$:

$$S \rightarrow A$$

$$A \rightarrow BA_1 \mid B \mid \varepsilon$$

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

$$A_1 \rightarrow AB$$

DEL- ε : Remove $A \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B$$

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

$$A_1 \rightarrow AB \mid B$$

Remove $B \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A \mid \varepsilon$$

Don't add $A \rightarrow \varepsilon$ because we already removed it

Remove $A_1 \rightarrow \varepsilon$:

$$S \rightarrow A \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Don't add $A \rightarrow \varepsilon$ because we already removed it

UNIT: Remove $S \rightarrow A$

$$S \rightarrow BA_1 \mid B \mid A_1 \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Example continued

From previous slide

$$S \rightarrow BA_1 \mid B \mid A_1 \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Example continued

From previous slide

$$S \rightarrow BA_1 \mid B \mid A_1 \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow B$

Example continued

From previous slide

$$S \rightarrow BA_1 \mid B \mid A_1 \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow B$

$$S \rightarrow BA_1 \mid A_1 \mid \varepsilon \mid 00$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Example continued

From previous slide

$$S \rightarrow BA_1 \mid B \mid A_1 \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow A_1$

Remove $S \rightarrow B$

$$S \rightarrow BA_1 \mid A_1 \mid \varepsilon \mid 00$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Example continued

From previous slide

$$S \rightarrow BA_1 \mid B \mid A_1 \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow A_1$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow B$

$$S \rightarrow BA_1 \mid A_1 \mid \varepsilon \mid 00$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Don't add $S \rightarrow B$ or $S \rightarrow A$

because we removed them

Example continued

From previous slide

$$S \rightarrow BA_1 \mid B \mid A_1 \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow B$

$$S \rightarrow BA_1 \mid A_1 \mid \varepsilon \mid 00$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow A_1$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Don't add $S \rightarrow B$ or $S \rightarrow A$

because we removed them

Remove $A \rightarrow B$

Example continued

From previous slide

$$S \rightarrow BA_1 \mid B \mid A_1 \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow B$

$$S \rightarrow BA_1 \mid A_1 \mid \varepsilon \mid 00$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow A_1$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Don't add $S \rightarrow B$ or $S \rightarrow A$
because we removed them

Remove $A \rightarrow B$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid A_1 \mid 00$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Example continued

From previous slide

$$S \rightarrow BA_1 \mid B \mid A_1 \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow B$

$$S \rightarrow BA_1 \mid A_1 \mid \varepsilon \mid 00$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow A_1$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $A \rightarrow A_1$

Don't add $S \rightarrow B$ or $S \rightarrow A$
because we removed them

Remove $A \rightarrow B$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid A_1 \mid 00$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Example continued

From previous slide

$$S \rightarrow BA_1 \mid B \mid A_1 \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow A_1$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $A \rightarrow A_1$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid 00 \mid AB$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow B$

$$S \rightarrow BA_1 \mid A_1 \mid \varepsilon \mid 00$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Don't add $S \rightarrow B$ or $S \rightarrow A$ because we removed them

Don't add $A \rightarrow B$ because we removed it

Don't add $A \rightarrow A$ because it's useless

Remove $A \rightarrow B$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid A_1 \mid 00$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Example continued

From previous slide

$$S \rightarrow BA_1 \mid B \mid A_1 \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow A_1$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $A \rightarrow A_1$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid 00 \mid AB$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow B$

$$S \rightarrow BA_1 \mid A_1 \mid \varepsilon \mid 00$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Don't add $S \rightarrow B$ or $S \rightarrow A$ because we removed them

Don't add $A \rightarrow B$ because we removed it

Don't add $A \rightarrow A$ because it's useless

Remove $A \rightarrow B$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid A_1 \mid 00$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $A_1 \rightarrow B$

Example continued

From previous slide

$$S \rightarrow BA_1 \mid B \mid A_1 \mid \varepsilon$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow A_1$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $A \rightarrow A_1$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid 00 \mid AB$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $S \rightarrow B$

$$S \rightarrow BA_1 \mid A_1 \mid \varepsilon \mid 00$$

$$A \rightarrow BA_1 \mid B \mid A_1$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Don't add $S \rightarrow B$ or $S \rightarrow A$ because we removed them

Don't add $A \rightarrow B$ because we removed it

Don't add $A \rightarrow A$ because it's useless

Remove $A \rightarrow B$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid A_1 \mid 00$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid B \mid A$$

Remove $A_1 \rightarrow B$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid 00 \mid AB$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid A \mid 00$$

Example continued

Copied from the previous slide

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid 00 \mid AB$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid A \mid 00$$

Remove $A_1 \rightarrow A$

Example continued

Copied from the previous slide

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid 00 \mid AB$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid A \mid 00$$

Remove $A_1 \rightarrow A$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid 00 \mid AB$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid 00 \mid BA_1$$

Example continued

Copied from the previous slide

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid 00 \mid AB$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid A \mid 00$$

Remove $A_1 \rightarrow A$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid 00 \mid AB$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid 00 \mid BA_1$$

TERM: Add $Z \rightarrow 0$

Example continued

Copied from the previous slide

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid 00 \mid AB$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid A \mid 00$$

Remove $A_1 \rightarrow A$

$$S \rightarrow BA_1 \mid \varepsilon \mid 00 \mid AB$$

$$A \rightarrow BA_1 \mid 00 \mid AB$$

$$B \rightarrow 00$$

$$A_1 \rightarrow AB \mid 00 \mid BA_1$$

TERM: Add $Z \rightarrow 0$

$$S \rightarrow BA_1 \mid \varepsilon \mid ZZ \mid AB$$

$$A \rightarrow BA_1 \mid ZZ \mid AB$$

$$B \rightarrow ZZ$$

$$A_1 \rightarrow AB \mid ZZ \mid BA_1$$

$$Z \rightarrow 0$$

Caution

Sipser gives a different procedure

- 1 START
- 2 DEL- ϵ
- 3 UNIT
- 4 BIN
- 5 TERM

This procedure works but can lead to an exponential blow up in the number of rules!

In general, if DEL- ϵ comes before BIN, then $|G'|$ is $O(2^{2^{|G|}})$;
if BIN comes before DEL- ϵ , then $|G'|$ is $O(|G|^2)$

UNIT is responsible for the quadratic blow up

So use whichever procedure you'd like, but Sipser's can be *very* bad
(Sipser's is bad if you have long rules with lots of variables with ϵ -rules)

Example blow up

$A \rightarrow BCDEEDCB \mid CBEDDEBC$

$B \rightarrow 0 \mid \varepsilon$

$C \rightarrow 1 \mid \varepsilon$

$D \rightarrow 2 \mid \varepsilon$

$E \rightarrow 3 \mid \varepsilon$

has five variables and 10 rules

Converting using START, BIN, DEL- ε , UNIT, TERM gives a CFG with 18 variables and 125 rules

Example blow up

$A \rightarrow BCDEEDCB \mid CBEDDEBC$

$B \rightarrow 0 \mid \varepsilon$

$C \rightarrow 1 \mid \varepsilon$

$D \rightarrow 2 \mid \varepsilon$

$E \rightarrow 3 \mid \varepsilon$

has five variables and 10 rules

Converting using START, BIN, DEL- ε , UNIT, TERM gives a CFG with 18 variables and 125 rules

Converting using START, DEL- ε , UNIT, BIN, TERM gives a CFG with 1394 variables and 1953 rules

Prefix

Recall $\text{PREFIX}(L) = \{w \mid \text{for some } x \in \Sigma^*, wx \in L\}$

Theorem

The class of context-free languages is closed under PREFIX.

Prefix

Recall $\text{PREFIX}(L) = \{w \mid \text{for some } x \in \Sigma^*, wx \in L\}$

Theorem

The class of context-free languages is closed under PREFIX.

Proof idea

Consider the language $\{w\#w^{\mathcal{R}} \mid w \in \{a, b\}^*\}$ generated by

$$T \rightarrow aTa \mid bTb \mid \#$$

Prefix

Recall $\text{PREFIX}(L) = \{w \mid \text{for some } x \in \Sigma^*, wx \in L\}$

Theorem

The class of context-free languages is closed under PREFIX.

Proof idea

Consider the language $\{w\#w^{\mathcal{R}} \mid w \in \{a, b\}^*\}$ generated by

$$T \rightarrow aTa \mid bTb \mid \#$$

Let's convert to CNF

Prefix

Recall $\text{PREFIX}(L) = \{w \mid \text{for some } x \in \Sigma^*, wx \in L\}$

Theorem

The class of context-free languages is closed under PREFIX.

Proof idea

Consider the language $\{w\#w^{\mathcal{R}} \mid w \in \{a, b\}^*\}$ generated by

$$T \rightarrow aTa \mid bTb \mid \#$$

Let's convert to CNF

$$S \rightarrow AU \mid BV \mid \#$$

$$T \rightarrow AU \mid BV \mid \#$$

$$U \rightarrow TA$$

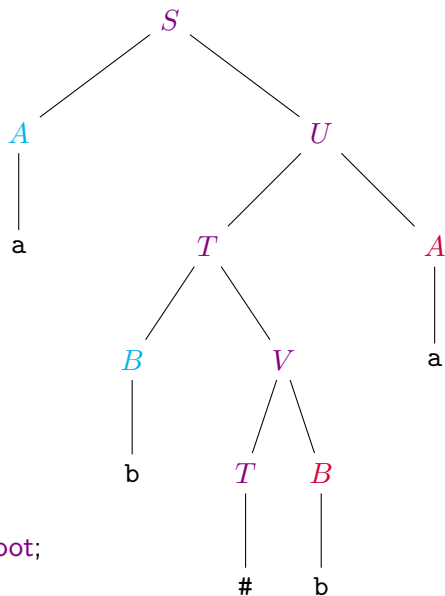
$$V \rightarrow TB$$

$$A \rightarrow a$$

$$B \rightarrow b$$

Derivation of ab#ba

$S \Rightarrow AU$
 $\Rightarrow aU$
 $\Rightarrow aTA$
 $\Rightarrow aBVA$
 $\Rightarrow abVA$
 $\Rightarrow abTBA$
 $\Rightarrow ab\#BA$
 $\Rightarrow ab\#bA$
 $\Rightarrow ab\#ba$



The prefix ab# includes

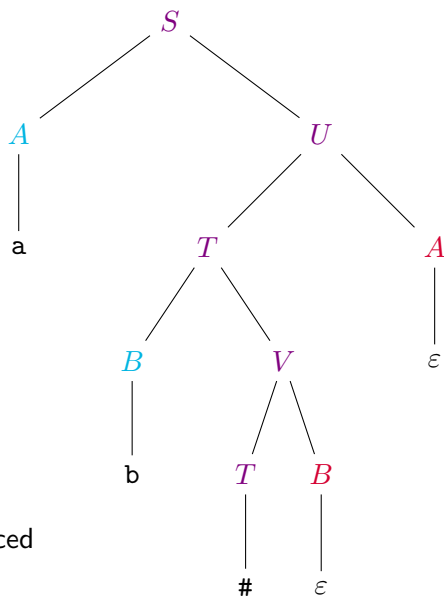
- all terminals from subtrees with a blue root;
- some terminals from subtrees with a violet root;
- no terminals from subtrees with a red root

Desired derivation for the prefix

We would like a derivation like this

$S \Rightarrow AU$
 $\Rightarrow aU$
 $\Rightarrow aTA$
 $\Rightarrow aBVA$
 $\Rightarrow abVA$
 $\Rightarrow abTBA$
 $\Rightarrow ab\#BA$
 $\Rightarrow ab\#\varepsilon A$
 $\Rightarrow ab\#\varepsilon\varepsilon$

Everything left of the **violet path** is produced
Everything right of the **violet path** becomes ε
The leaf connected to the **violet path** is produced



The proof idea

The **violet path** corresponds to the point where we “split” the prefix from the remainder of the string

We want to construct a CFG that keeps track of whether a given variable in the derivation is

L **left** of the split,

S part of the **split**, or

R **right** of the split

We can construct a new CFG whose variables are $\langle A, L \rangle$, $\langle A, S \rangle$, or $\langle A, R \rangle$ where A is a variable in the original CFG

We have to deal with the three types of rules

- $S \rightarrow \varepsilon$
- $A \rightarrow BC$
- $A \rightarrow t$

and produce new rules corresponding to the variable on the LHS being left of, right of, or on the split

Proof

If $L = \emptyset$, then $\text{PREFIX}(L) = \emptyset$ which is CF.

Otherwise, let L be CF and generated by the CFG $G = (V, \Sigma, R, S)$ in CNF.

Construct a new CFG (not in CNF) $G' = (V', \Sigma, R', S')$ where

$$V = \{\langle A, D \rangle \mid A \in V \text{ and } D \in \{L, S, R\}\}$$

$$S' = \langle S, S \rangle$$

Now we just need to specify R' . We'll start with $R' = \emptyset$ and add rules to it

Proof continued

Since L is nonempty, $\varepsilon \in \text{PREFIX}(L)$ so add the rule $\langle S, S \rangle \rightarrow \varepsilon$ to R'

For each rule of the form $A \rightarrow BC$ in R , add the following rules to R'

- | | |
|---|-----------------------------------|
| $\langle A, L \rangle \rightarrow \langle B, L \rangle \langle C, L \rangle$ | left of the split |
| $\langle A, S \rangle \rightarrow \langle B, L \rangle \langle C, S \rangle \mid \langle B, S \rangle \langle C, R \rangle$ | one of B or C is on the split |
| $\langle A, R \rangle \rightarrow \langle B, R \rangle \langle C, R \rangle$ | right of the split |

For each rule of the form $A \rightarrow t$ in R , add the following rules to R'

- $\langle A, L \rangle \rightarrow t$
- $\langle A, S \rangle \rightarrow t$
- $\langle A, R \rangle \rightarrow \varepsilon$

Proof continued

For each $w = w_1w_2\cdots w_n \in L$, $S \stackrel{*}{\Rightarrow} A_1A_2\cdots A_n$ where $A_i \Rightarrow w_i$

By construction,

$$\begin{aligned}\langle S, S \rangle &\stackrel{*}{\Rightarrow} \langle A_1, L \rangle \cdots \langle A_{i-1}, L \rangle \langle A_i, S \rangle \langle A_{i+1}, R \rangle \cdots \langle A_n, R \rangle \\ &\stackrel{*}{\Rightarrow} w_1w_2\cdots w_i\end{aligned}$$

for each $1 \leq i \leq n$

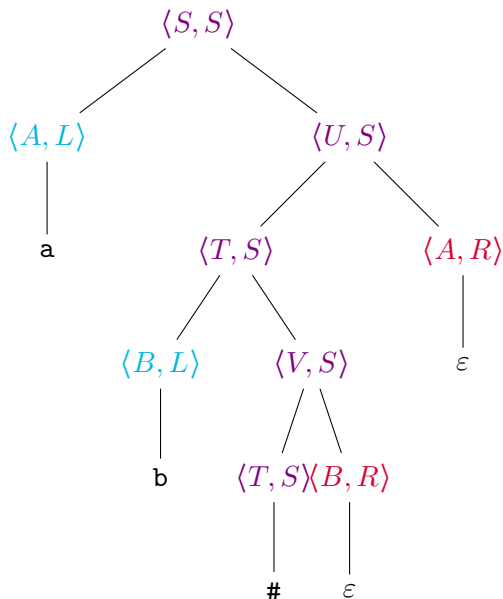
I.e., G' derives the prefix of every string in L

A similar argument works to show that if G' derives a string then it's a prefix of some string in L □

Applying the construction

Deriving ab#

$\langle S, S \rangle \Rightarrow \langle A, L \rangle \langle U, S \rangle$
 $\Rightarrow a \langle U, S \rangle$
 $\Rightarrow a \langle T, S \rangle \langle A, R \rangle$
 $\Rightarrow a \langle B, L \rangle \langle V, S \rangle \langle A, R \rangle$
 $\Rightarrow ab \langle V, S \rangle \langle A, R \rangle$
 $\Rightarrow ab \langle T, S \rangle \langle BA, R \rangle$
 $\Rightarrow ab \# \langle B, R \rangle \langle A, R \rangle$
 $\Rightarrow ab \# \langle A, R \rangle$
 $\Rightarrow ab \#$



Similarities with regular expression

Proving things about

- Regular languages. Assume there exists a regular expression that generates the language and consider the six cases
- Context-free languages. Assume there exists a CFG that generates the language and consider the three types of rules