CS 6041
Theory of Computation

Context-free language

Kun Suo

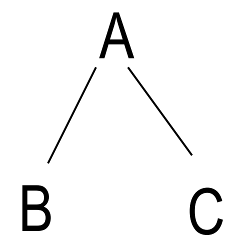
Computer Science, Kennesaw State University

https://kevinsuo.github.io/

Chomsky normal form (CNF)

- CNF: only allow CFG in the following forms
  - $S \rightarrow \varepsilon$
  - $A \rightarrow BC$
  - $A \rightarrow a$

Parse tree for CNF is only binary tree



- Here
  - A,B,C are variables
  - B,C are not start variables

  (start variable does not exist on the right side)
  - a is terminal

Chomsky normal form (CNF)

- CNF: only allow CFG in the following forms
  - $S \rightarrow \varepsilon$
  - $A \rightarrow BC$
  - $A \rightarrow a$

Only start variable S can generate ε

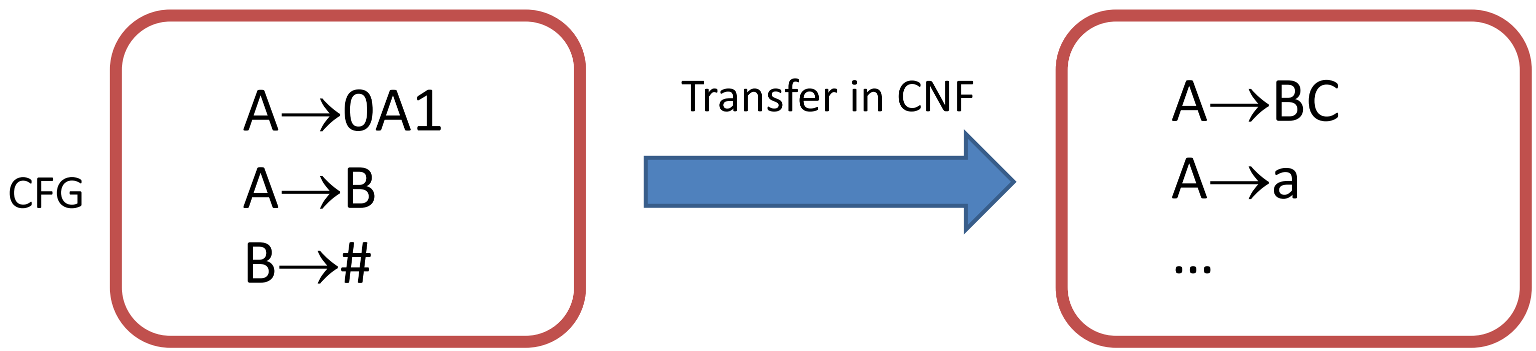
Variables can only generate:
1, two variables
2, single terminal

✓ $S \rightarrow \varepsilon$
✓ $A \rightarrow BC$
✓ $A \rightarrow a$

✗ $A \rightarrow \varepsilon$
✗ $A \rightarrow B$
✗ $A \rightarrow abcd$
✗ $A \rightarrow aB$

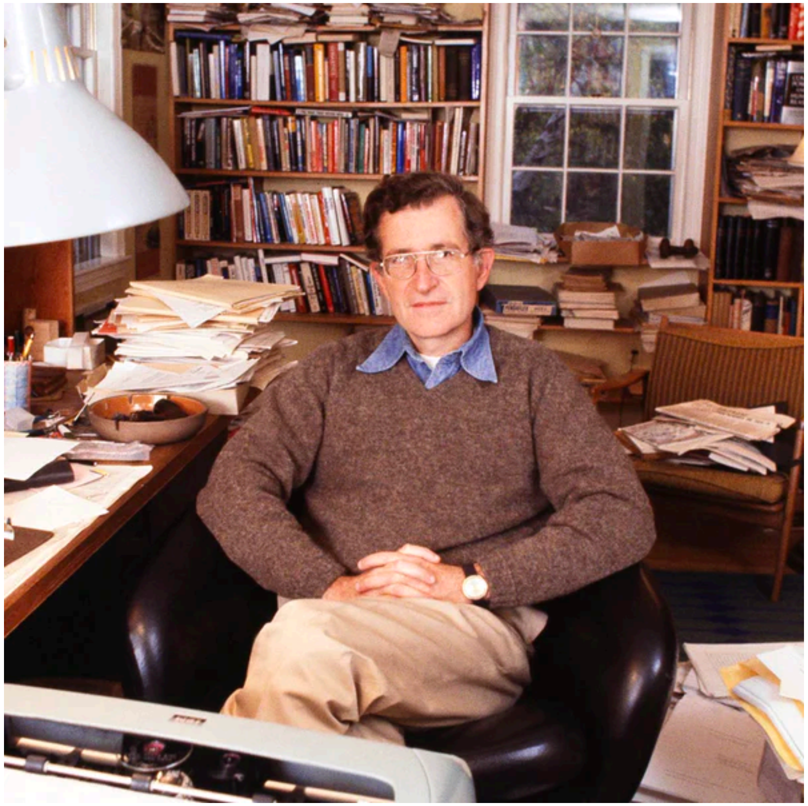
Chomsky normal form (CNF)

- Theorem: Any CFL is generated by a CFG in CNF



Noam Chomsky

- http://linguistics.mit.edu/user/chomsky/



Chomsky normal form example

G_6: $S \rightarrow ASA \mid aB,$

$A \rightarrow B \mid S$

$B \rightarrow b \mid \varepsilon$

Get the CNF for G_6

$S \rightarrow \varepsilon$	$A \rightarrow \varepsilon$
$A \rightarrow BC$	$A \rightarrow B$
$A \rightarrow a$	$A \rightarrow abcd$
	$A \rightarrow aB$

Step 1

G_6: $S \rightarrow ASA \mid aB$,

$A \rightarrow B \mid S$

$B \rightarrow b \mid \varepsilon$

$S \rightarrow \varepsilon$

$A \rightarrow BC$

$A \rightarrow a$

$A \rightarrow \varepsilon$

$A \rightarrow B$

$A \rightarrow abcd$

$A \rightarrow aB$

(1) $S_0 \rightarrow S$

$S \rightarrow ASA \mid aB$

$A \rightarrow B \mid S$

$B \rightarrow b \mid \varepsilon$

To simply the start variable

Step 2

(2a) $S_0 \to S$

$S \to ASA \mid aB$

$A \to B \mid S$

$B \to b \mid \varepsilon$

Only start variable S can generate ε

$S \to \varepsilon$

$A \to BC$

$A \to a$

$A \to \varepsilon$

$A \to B$

$A \to abcd$

$A \to aB$

$S_0 \to S$

$S \to ASA \mid aB \mid a$

$A \to B \mid S \mid \varepsilon$

$B \to b$

Step 2

(2b) $S_0 \rightarrow S$

$S \rightarrow ASA \mid aB \mid a$

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

$B \rightarrow b$

Only start variable S can generate ε

$S \rightarrow \varepsilon$
$A \rightarrow BC$
$A \rightarrow a$
$A \rightarrow \varepsilon$
$A \rightarrow B$
$A \rightarrow abcd$
$A \rightarrow aB$

$S_0 \rightarrow S$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS \mid S$

$A \rightarrow B \mid S$

$B \rightarrow b$

No $A \rightarrow \varepsilon$ anymore

Step 3

(3a) $S_0 \rightarrow S$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS \mid S$

$A \rightarrow B \mid S$

$B \rightarrow b$

$S \rightarrow \varepsilon$

$A \rightarrow BC$

$A \rightarrow a$

$A \rightarrow \varepsilon$

$A \rightarrow B$

$A \rightarrow abcd$

$A \rightarrow aB$

$S_0 \rightarrow S$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$A \rightarrow B \mid S$

$B \rightarrow b$

Step 3

(3b) $S_0 \to S$

$S \to ASA \mid aB \mid a \mid SA \mid AS$

$A \to B \mid S$

$B \to b$

$S \to \varepsilon$

$A \to BC$

$A \to a$

$A \to \varepsilon$

$A \to B$

$A \to abcd$

$A \to aB$

$S_0 \to ASA \mid aB \mid a \mid SA \mid AS$

$S \to ASA \mid aB \mid a \mid SA \mid AS$

$A \to B \mid S$

$B \to b$

Step 3

(3c) $S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$A \rightarrow B \mid S$

$B \rightarrow b$

$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$A \rightarrow b \mid S$

$B \rightarrow b$

$S \rightarrow \varepsilon$

$A \rightarrow BC$

$A \rightarrow a$

$A \rightarrow \varepsilon$

$A \rightarrow B$

$A \rightarrow abcd$

$A \rightarrow aB$

Step 3

(3d) $S_0 \to ASA \mid aB \mid a \mid SA \mid AS$

$S \to ASA \mid aB \mid a \mid SA \mid AS$

$A \to b \mid S$

$B \to b$

$S_0 \to ASA \mid aB \mid a \mid SA \mid AS$

$S \to ASA \mid aB \mid a \mid SA \mid AS$

$A \to b \mid ASA \mid aB \mid a \mid SA \mid AS$

$B \to b$

$S \to \varepsilon$

$A \to BC$

$A \to a$

$A \to \varepsilon$

$A \to B$

$A \to abcd$

$A \to aB$

Step 4

(4) $S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$A \rightarrow b \mid ASA \mid aB \mid a \mid SA \mid AS$

$B \rightarrow b$

$S_0 \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$

$S \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$

$A \rightarrow b \mid AA_1 \mid aB \mid a \mid SA \mid AS$

$B \rightarrow b$

$A_1 \rightarrow SA$

✓ $S \rightarrow \varepsilon$

✓ $A \rightarrow BC$

✓ $A \rightarrow a$

✗ $A \rightarrow \varepsilon$

✗ $A \rightarrow B$

✗ $A \rightarrow abcd$

✗ $A \rightarrow aB$

Step 5

(5) $S_0 \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$

$S \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$

$A \rightarrow b \mid AA_1 \mid aB \mid a \mid SA \mid AS$

$B \rightarrow b$

$A_1 \rightarrow SA$

$S_0 \rightarrow AA_1 \mid UB \mid a \mid SA \mid AS$

$S \rightarrow AA_1 \mid UB \mid a \mid SA \mid AS$

$A \rightarrow b \mid AA_1 \mid UB \mid a \mid SA \mid AS$

$B \rightarrow b$

$A_1 \rightarrow SA$

$U \rightarrow a$

✓ $S \rightarrow \varepsilon$

✓ $A \rightarrow BC$

✓ $A \rightarrow a$

✗ $A \rightarrow \varepsilon$

✗ $A \rightarrow B$

✗ $A \rightarrow abcd$

✗ $A \rightarrow aB$

Step 5

(5) $S_0 \to AA_1 \mid UB \mid a \mid$

$\quad SA \mid AS$

$S \to AA_1 \mid UB \mid a \mid$

$\quad SA \mid AS$

$A \to b \mid AA_1 \mid UB \mid a \mid$

$\quad SA \mid AS$

$B \to b$

$A_1 \to SA$

$U \to a$

G6 in CNF

G_6: $S \to ASA \mid aB$,

$A \to B \mid S$

$B \to b \mid \varepsilon$

✓ $S \to \varepsilon$

✓ $A \to BC$

✓ $A \to a$

✗ $A \to \varepsilon$

✗ $A \to B$

✗ $A \to abcd$

✗ $A \to aB$

Conclusion for CNF

- Add new start variable if needed
- $A \rightarrow \varepsilon$, merge above rules with A
- $A \rightarrow B$, replace B with terminals or other rules
- $A \rightarrow aB$, replace with $U \rightarrow a$, $A \rightarrow UB$
- $A \rightarrow abcd$, replace with $A \rightarrow aU_1$, $U_1 \rightarrow bU_2$, $U_2 \rightarrow cd$
- $A \rightarrow BCD$, similar as the above

✓ $S \rightarrow \varepsilon$
✓ $A \rightarrow BC$
✓ $A \rightarrow a$
✗ $A \rightarrow \varepsilon$
✗ $A \rightarrow B$
✗ $A \rightarrow abcd$
✗ $A \rightarrow aB$

Question

G_7: $S \to AbA$,

$A \to Aa$,

$A \to \varepsilon$

Get the CNF for G_7

Conclusion for CNF

- Add new start variable if needed
- $A \to \varepsilon$, merge above rules with A
- $A \to B$, replace B with terminals or other rules
- $A \to aB$, replace with $U \to a$, $A \to UB$
- $A \to abcd$, replace with $A \to aU_1$, $U_1 \to bU_2$, $U_2 \to cd$
- $A \to BCD$, similar as the above

✓ $S \to \varepsilon$
✓ $A \to BC$
✓ $A \to a$
✗ $A \to \varepsilon$
✗ $A \to B$
✗ $A \to abcd$
✗ $A \to aB$

✓ $S \to \varepsilon$
✓ $A \to BC$
✓ $A \to a$
✗ $A \to \varepsilon$
✗ $A \to B$
✗ $A \to abcd$
✗ $A \to aB$

Question

$S \to AbA,$

$A \to Aa,$

$A \to \varepsilon$

$S \to AbA \mid bA \mid Ab \mid b$

$A \to Aa \mid a$

$S \to \varepsilon$

$A \to BC$

$A \to a$

$A \to \varepsilon$

$A \to B$

$A \to abcd$

$A \to aB$

Question

$S \rightarrow AbA \mid bA \mid Ab \mid b$

$A \rightarrow Aa \mid a$

$S \rightarrow TA \mid bA \mid Ab \mid b$

$A \rightarrow Aa \mid a$

$T \rightarrow Ab$

$S \rightarrow \varepsilon$

$A \rightarrow BC$

$A \rightarrow a$

$A \rightarrow \varepsilon$

$A \rightarrow B$

$A \rightarrow abcd$

$A \rightarrow aB$

Question

$S \rightarrow TA \mid bA \mid Ab \mid b$

$A \rightarrow Aa \mid a$

$T \rightarrow Ab$

$S \rightarrow TA \mid BA \mid AB \mid b$

$A \rightarrow Aa \mid a$

$T \rightarrow AB$

$B \rightarrow b$

$S \rightarrow \varepsilon$

$A \rightarrow BC$

$A \rightarrow a$

$A \rightarrow \varepsilon$

$A \rightarrow B$

$A \rightarrow abcd$

$A \rightarrow aB$

Question

$S \rightarrow TA \mid BA \mid AB \mid b$

$A \rightarrow Aa \mid a$

$T \rightarrow AB$

$B \rightarrow b$

$S \rightarrow TA \mid BA \mid AB \mid b$

$A \rightarrow AC \mid a$

$T \rightarrow AB$

$B \rightarrow b$

$C \rightarrow a$

$S \rightarrow \varepsilon$

$A \rightarrow BC$

$A \rightarrow a$

$A \rightarrow \varepsilon$

$A \rightarrow B$

$A \rightarrow abcd$

$A \rightarrow aB$

Question 2

G_{10}: $S \rightarrow 1A \mid 0B$

$A \rightarrow 1AA \mid 0S \mid 0$

$B \rightarrow 0BB \mid 1$

Get the CNF for G_{10}

Conclusion for CNF

- Add new start variable if needed
- $A \rightarrow \varepsilon$, merge above rules with A
- $A \rightarrow B$, replace B with terminals or other rules
- $A \rightarrow aB$, replace with $U \rightarrow a$, $A \rightarrow UB$
- $A \rightarrow abcd$, replace with $A \rightarrow aU_1$, $U_1 \rightarrow bU_2$, $U_2 \rightarrow cd$
- $A \rightarrow BCD$, similar as the above

$S \rightarrow \varepsilon$
$A \rightarrow BC$
$A \rightarrow a$
$A \rightarrow \varepsilon$
$A \rightarrow B$
$A \rightarrow abcd$
$A \rightarrow aB$

$S \rightarrow \varepsilon$
$A \rightarrow BC$
$A \rightarrow a$

$A \rightarrow \varepsilon$
$A \rightarrow B$
$A \rightarrow abcd$
$A \rightarrow aB$

Question 2

G_{10}: $S \rightarrow 1A \mid 0B$

$A \rightarrow 1AA \mid 0S \mid 0$

$B \rightarrow 0BB \mid 1$

$S \rightarrow \varepsilon$

$A \rightarrow BC$

$A \rightarrow a$

$A \rightarrow \varepsilon$

$A \rightarrow B$

$A \rightarrow abcd$

$A \rightarrow aB$

$S \rightarrow C_1A \mid 0B$

$A \rightarrow C_1AA \mid 0S \mid 0$

$B \rightarrow 0BB \mid 1$

$C_1 \rightarrow 1$

Question 2

$S \rightarrow C_1A \mid 0B$
$A \rightarrow C_1AA \mid 0S \mid 0$
$B \rightarrow 0BB \mid 1$
$C_1 \rightarrow 1$

$S \rightarrow C_1A \mid C_0B$
$A \rightarrow C_1AA \mid 0S \mid 0$
$B \rightarrow 0BB \mid 1$
$C_1 \rightarrow 1$
$C_0 \rightarrow 0$

✓ $S \rightarrow \varepsilon$
✓ $A \rightarrow BC$
✓ $A \rightarrow a$
✗ $A \rightarrow \varepsilon$
✗ $A \rightarrow B$
✗ $A \rightarrow abcd$
✗ $A \rightarrow aB$

Question 2

$S \to C_1A \mid C_0B$
$A \to C_1AA \mid 0S \mid 0$
$B \to 0BB \mid 1$
$C_1 \to 1$
$C_0 \to 0$

$S \to C_1A \mid C_0B$
$A \to C_1C_2 \mid 0S \mid 0$
$B \to 0BB \mid 1$
$C_1 \to 1$
$C_0 \to 0$
$C_2 \to AA$

✓ $S \to \varepsilon$
✓ $A \to BC$
✓ $A \to a$
✗ $A \to \varepsilon$
✗ $A \to B$
✗ $A \to abcd$
✗ $A \to aB$

Question 2

$S \to C_1A \mid C_0B$
$A \to C_1C_2 \mid 0S \mid 0$
$B \to 0BB \mid 1$
$C_1 \to 1$
$C_0 \to 0$
$C_2 \to AA$

$S \to C_1A \mid C_0B$
$A \to C_1C_2 \mid C_0S \mid 0$
$B \to 0BB \mid 1$
$C_1 \to 1$
$C_0 \to 0$
$C_2 \to AA$

✓ $S \to \varepsilon$
✓ $A \to BC$
✓ $A \to a$
✗ $A \to \varepsilon$
✗ $A \to B$
✗ $A \to abcd$
✗ $A \to aB$

Question 2

$S \rightarrow C_1A \mid C_0B$
$A \rightarrow C_1C_2 \mid C_0S \mid 0$
$B \rightarrow 0BB \mid 1$
$C_1 \rightarrow 1$
$C_0 \rightarrow 0$
$C_2 \rightarrow AA$

$S \rightarrow C_1A \mid C_0B$
$A \rightarrow C_1C_2 \mid C_0S \mid 0$
$B \rightarrow C_0C_3 \mid 1$
$C_1 \rightarrow 1$
$C_0 \rightarrow 0$
$C_2 \rightarrow AA$
$C_3 \rightarrow BB$

✓ $S \rightarrow \varepsilon$
✓ $A \rightarrow BC$
✓ $A \rightarrow a$
✗ $A \rightarrow \varepsilon$
✗ $A \rightarrow B$
✗ $A \rightarrow abcd$
✗ $A \rightarrow aB$

Conclusion for CNF

- Add new start variable
- $A \rightarrow \varepsilon$, merge above rules with A
- $A \rightarrow B$, replace B with terminals or other rules
- $A \rightarrow aB$, replace with $U \rightarrow a$, $A \rightarrow UB$
- $A \rightarrow abcd$, replace with $A \rightarrow aU_1$, $U_1 \rightarrow bU_2$, $U_2 \rightarrow cd$
- $A \rightarrow BCD$, similar as the above

✓ $S \rightarrow \varepsilon$
✓ $A \rightarrow BC$
✓ $A \rightarrow a$
✗ $A \rightarrow \varepsilon$
✗ $A \rightarrow B$
✗ $A \rightarrow abcd$
✗ $A \rightarrow aB$

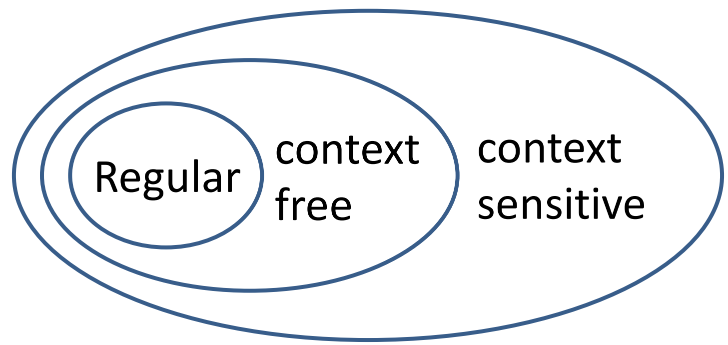
Context-sensitive vs. Context-free

- What is the difference?
  - S->aSb
  - S->ab

  - aSb->aaSbb
  - S->ab

Context-sensitive vs. Context-free

- Context-free
  - S->aSb
  - S->ab



- Context-sensitive
  - aSb->aaSbb
  - S->ab

The rule S->aSb happens only S has a on the left and b on the right