Back to grammars for a bit...

Chomsky Normal Form

Defn 2.8 A context-free grammar is in **CNF** (Chomsky Normal Form) if every rule is of the form

$A \to BC$ $\quad A, B, C \in V,\ B \neq S,\ C \neq S$

$A \to \sigma$ $\quad \sigma \in \Sigma$

Also allowed is the rule $S \to \varepsilon$ (S is start)

Theorem 2.8 Any CFG can be put into CNF

Proof Sketch:

1. Do we really need $A \to \varepsilon$?

   We can, for each rule like this

   $B \to \sim A \sim$ add the rule

   $B \to \sim \sim$

   ↑ its as if A went to ε within the rule.

   The only "$A \to \varepsilon$" rule we can't simulate by replacing A with ε is all our rules (possibly doubling the number of rules)

is $S \to \varepsilon$, and we can keep that rule.

2. Do we need to have rules like

$A \to B$?

$B \to$  add $A \to$ 

$B \to$  $A \to$ 

3. Do we need to have rules like

$A \to BCD \ldots ZX \Rightarrow A \to BX_1$

X_1 can be terminals or variables.

$X_1 \to CX_2$

$X_2 \to DX_3$

$\vdots$

$X_n \to ZX$.

You can convert any CFG into CNF.

We'll do a small one here:

$S \to AXA \mid Xbb$

$X \to Xc \mid \varepsilon$

$A \to bA \mid \varepsilon$

$S \to \varepsilon \mid AX_A \mid XA \mid AX \mid AA \mid XB_b \mid BB$

$B_b \to BB$

$B \to b$

$X_A \to XA$

$X \to XC \mid c$

$C \to c$

$A \to BA \mid b$

$B \to b$

$B_b \to BB$

Do you have to know how to convert to CNF form for the test?

No, but you should know that any CFG can be converted to CNF.

Lets learn a CFG trick that might be on the test ...

$$\{ a^n \cdot b^m \mid 3n = 2m+1 \}$$

Homework Problems:

Give a CFG for each language below:

1. BalDelim = $\{ w \mid$ w is a string of delimiters $(,), [,], \{, \}$, that are properly balanced $\}$.

2. $\{ a^n b^m \mid 2m = 3n+1 \}$

3. $\{ w \in \{a,b\}^* \mid \#_a(w) = 2 \cdot \#_b(w) \}$

4. $\{ w \in \{a,b\}^* \mid w = w^R \}$

5. $\{ a^i b^j \mid i \neq j \quad i,j \geq 0 \}$

6. $\{ a^i b^j c^k \mid i \neq j \text{ or } j \neq k \}$

---

4. $\{ w \in \{a,b\}^* \mid w = w^R \}$

$$S \to aSa \mid bSb \mid a \mid b \mid \varepsilon$$

1. $(\{\})[]$

Elaine Rich's questions

1. Briefly describe these languages:

a) $(b+ba)(b+a)^*(ab+b)$

b) $((((a^*b^*)^*ab)+((a^*b^*)^*ba))(b+a)^*$

2. r.e.s for: $\Sigma = \{a, b\}$

a) $\forall$ a is immediately followed by b

b) does not end in ba

c) (is nonsensicle)

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

d) binary encodings, no leading 0's, of ints divisible by 4.

e) " " that are powers of 4.

q_2 has 001 as substring.