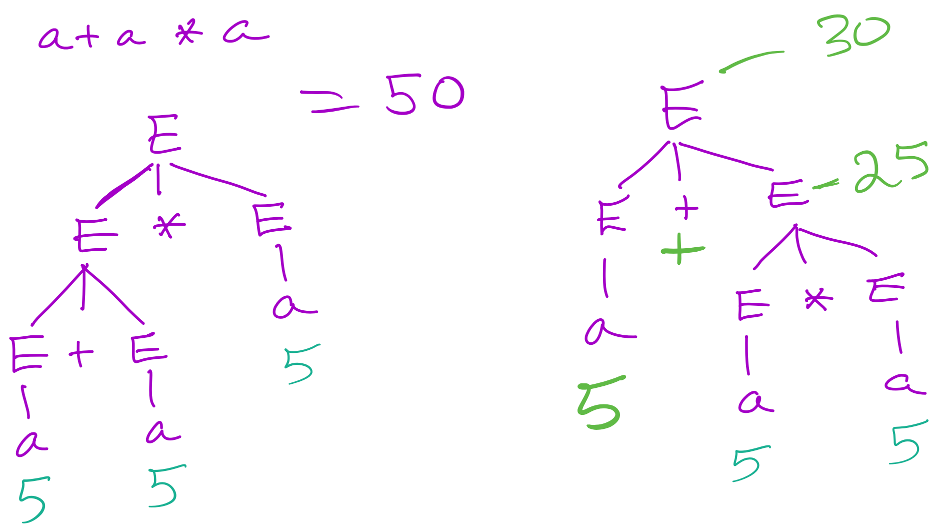
CFGs continued.

We often give grammars just as a set of rules — variables appear on LHS, terminals appear only on RHS. Start variable must be identified.

$\underline{E} \to E+E \mid E*E \mid a$ a+a*a a*a*a+a

a + a * a



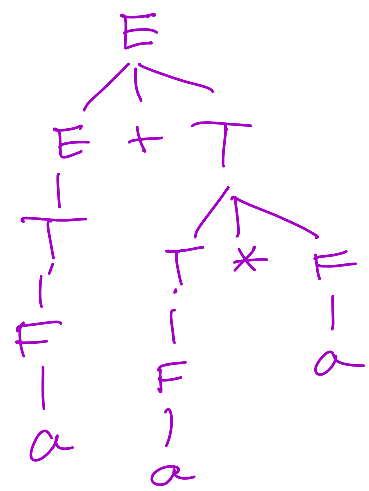
a + a * a is <u>ambiguous</u>.

Def 2.7 A string w is derived <u>ambiguously</u> in CFG G if it has 2 or more different leftmost derivations (derivation tree).
G is ambiguous if it generates some string ambiguously.

Sometimes we can remove the ambiguity — find a new grammar that generates the same language but is not ambiguous.

$$E \to E + T \mid T$$
$$T \to T * F \mid F$$
$$F \to a$$

$a + a * a$



Can every ambiguous CFG be disambiguated?

No. See below. later.

Defn: A language is ***Context-free*** (is a CFL) if $\exists$ a CFG for it.

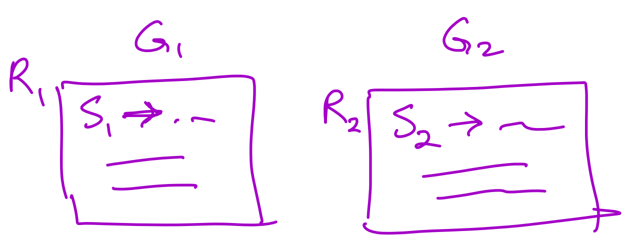
Designing CFGs.

Theorem: The class CFL is closed under $\cup$.

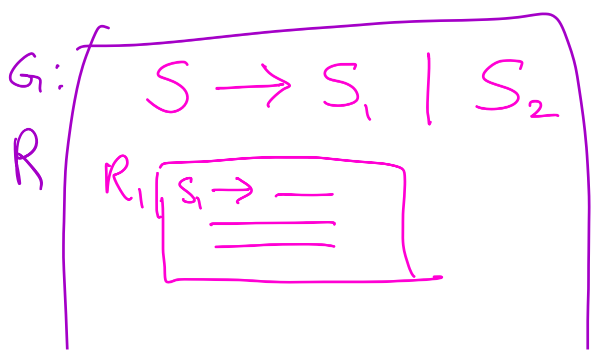
Proof: by construction.

Variables in G_1 Variables in G_2

$$V_1 \cap V_2 = \emptyset.$$



New grammar G:



R_2 $\boxed{S_2 \to \cdots}$

$$L(G) = L(G_1) \cup L(G_2). \quad \blacksquare$$

Eg. G_1: $S_1 \to aS_1b \mid \varepsilon$ $\qquad$ G_2: $S_2 \to bS_2a \mid \varepsilon$

G:

$$S \to S_1$$
$$S \to S_2.$$
$$S_1 \to aS_1b$$
$$S_1 \to \varepsilon$$
$$S_2 \to bS_2b$$
$$S_2 \to \varepsilon$$

Theorem: The Class CFL is closed under concat.

Proof: Exercise.

Exercise Argue that the following language is inherently ambiguous:

$$\{a^i b^j c^k \mid i \geq 0,\ j \geq 0,\ k \geq 0. \text{ either } i = j \text{ or } j = k \text{ or both}\}.$$

$L_1 = \{a^i b^i c^k \mid i \geq 0, k \geq 0\}$

$S_1 \to S_1 c \mid X$

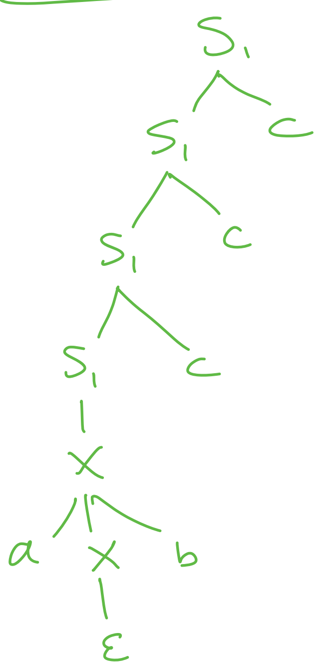
$X \to aXb \mid \varepsilon$

$L_2 = \{a^i b^j c^j : i \geq 0, j \geq 0\}$

$S_2 \to a S_2 \mid Y$

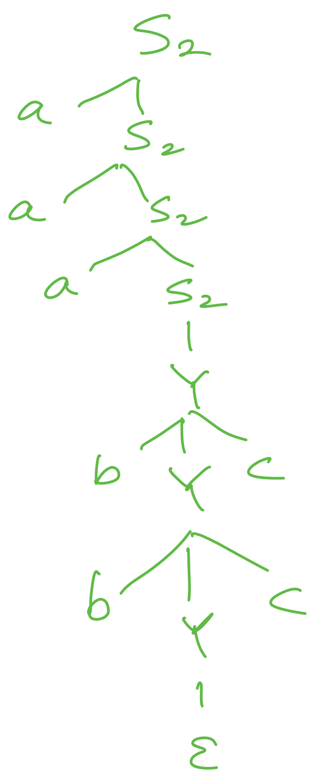
$Y \to bYc \mid \varepsilon$

abccc



$S \to S_1 \mid S_2$

aaabbcc



aabbcc

End of day Feb 2

Start Feb 7.

Defn Chomsky Normal Form (CNF)

A CFG is CNF if every rule is of the form

$A \to BC$ or

$A \to a$ where a is a terminal;

and where A, B, C are any variables, but B, C cannot be the start variable. Also we permit $S \to \varepsilon$, where S is the start variable.

Eg. $S \to \varepsilon \mid E$ $\quad$ $A \to a$

$B \to b$

$E \to aEb$

$E \to AB$

$\Downarrow$

$S \to \varepsilon \mid XB \mid \underline{AB}$

$X \to AE$ $\quad$ // X is "almost" same # a's as b's – needs a b.

$E \to XB$

$E \to AB.$ $\quad$ // E is $a^n b^n$, $n \geq 1$

Theorem: $\forall$ CFL has a CNF grammar that generates it.

$\{ a^i b^i \mid i \geq 0 \}$ $\qquad$ $\{ b^i a^i \mid i \geq 0 \}$

$S_1 \to a S_1 b \mid \varepsilon$ $\qquad$ $S_2 \to b S_2 a \mid \varepsilon$

$\underline{S \to S_1 \mid S_2}$ $\qquad$ $S \to S_1$

$S \to S_2$

$V_1 \cap V_2 = \emptyset$

$L_1 = \{a^i b^i \mid i \geq 0\}$

$S \rightarrow aSb \mid \varepsilon$

$L_2 = \{w \in \{a,b\}^* \mid \#_a(w) = \#_b(w)\}$

$S \rightarrow aSb \mid bSa \mid SS \mid \varepsilon$