After class, I will

1) Assign HW5 (will be due next Fri)
2) Scan notes for this week.

Modifying grammars to equivalent (nicer) grammars.

Last time: removing useless productions.

Today: λ-productions $(A \to \lambda)$ and unit productions $(A \to B)$

How do we "get rid of" λ-productions?

If we have

$A \to aBycC,$

$B \to \lambda \mid \ldots$

$C \to yxBaA$

we add prods.

$A \to aBycC \mid aycC$

$C \to yxBaA \mid yxaA$

and now we can remove $B \to \lambda$.

We could have a more complicated situation:

$A \rightarrow BC \mid \text{------}$

$B \rightarrow \lambda \mid \text{-----}$

$C \rightarrow \lambda \mid \text{------}$

In this situation we can also produce λ from A (by replacing A w/ BC and then both B + C with λ's)

~~It's~~ We might worry this process gets us in an infinite loop, - let's try to do all our changes at once.

1) Find all vars that can go to λ (<u>nullable</u> vars)

(anything with $X \rightarrow \lambda$ is nullable, anything with $X \rightarrow$ all nullable vars is nullable)

2). In any production that contains a nullable variable, add every possibility of deleting a nullable var as a new production

~~i.e.~~ e.g. If B + C are nullable

$$A \to xByCzB \text{ becomes}$$

$$A \to xByCzB \mid xyCzB \mid xByzB \mid xByCz \mid xyzB \mid yCz \mid xByz \mid xyz$$

So if a prod. has k nullable variables occurences, it is replaced w/ 2^k productions.

3) Then it is safe to delete all the λ-prods - they have all been replaced.

Removing unit productions:

E.g. $A \to B \mid xyC$

$B \to xyABC \mid CxzA \mid xzB$

We can replace $A \to B$ with

$A \to$ all the things B goes to

so we get

$A \to xyC \mid xyABC \mid CxzA \mid xzB$

$B \to xyABC \mid CxzA \mid xzB$.

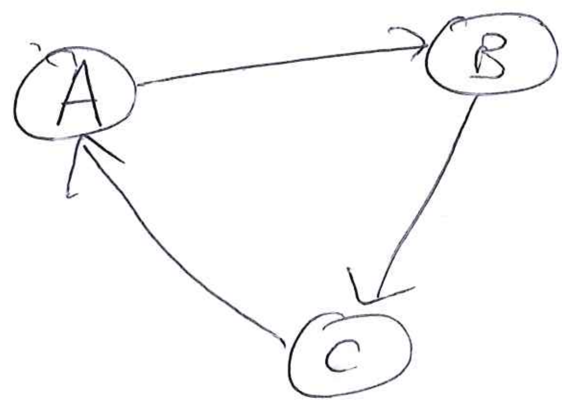
There are complications; we might have

$A \to B \mid \ldots$

$B \to C \mid \ldots$

$C \to A \mid \ldots$

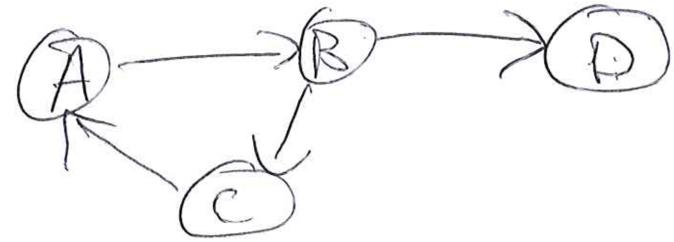
To take care of this we draw a directed graph (nodes = vars
arrows = unit prods)



~~Replace every varia~~

1. Add, as productions for A, all prods for variables one can get to from A in this graph.
2. Now it's safe to delete all unit productions.

In ~~this~~ the above case, one can combine A, B, C into one var, but we can have



where D still needs to be a separate var!

Note: if we act in the right order, we can remove all useless prods, all λ-prods, and all unit prods.

Chomsky Normal Form (CNF)

A grammar is in CNF if all the productions look like

$A \to BC$ (one var $\to$ 2 vars)

($A \to AA$ or $A \to AB$ okay or $A \to BA$)

or $A \to x$ (one var $\to$ one letter)

Turning grammars to CNF:

E.g.: A = start var:

$$A \to xyAB \mid Bx \mid y$$
$$B \to xByB \mid BAx \mid x$$

0) Get rid of useless, λ, + unit prods

① Introduce a var for each letter

$X \to x$

$Y \to y$

$A \to XYAB \mid BX \mid y$

$B \to XBYB \mid B\underline{AX} \mid x$

② Introduce new "glued variables"

$AX \to A\ X$ $\quad AB \to A\ B$

$YAB \to Y\ AB$ $\quad YB \to Y\ B$

$BYB \to B\ YB$

make this readable:

$C = \boxed{AX}$

$D = \boxed{YAB}$

$E = \boxed{BYB}$

$F = \boxed{AB}$

$G = \boxed{YB}$

$C \to AX$	$A \to XD \mid BX \mid y$
$D \to YF$	$B \to XE \mid BC \mid x$
$E \to BG$	$X \to x$
$F \to AB$	$Y \to y$
$G \to YB$	

$\cancel{H \to BX}$

Greibach Normal Form

All productions are of the form

$A \rightarrow x \underline{\text{STUFF}}$

(where STUFF could be λ - you can insist STUFF is all variables)

We won't talk about how to convert to GNF, but it is possible.