Exam

Tuesdy 7-9pm

1404 Siebel

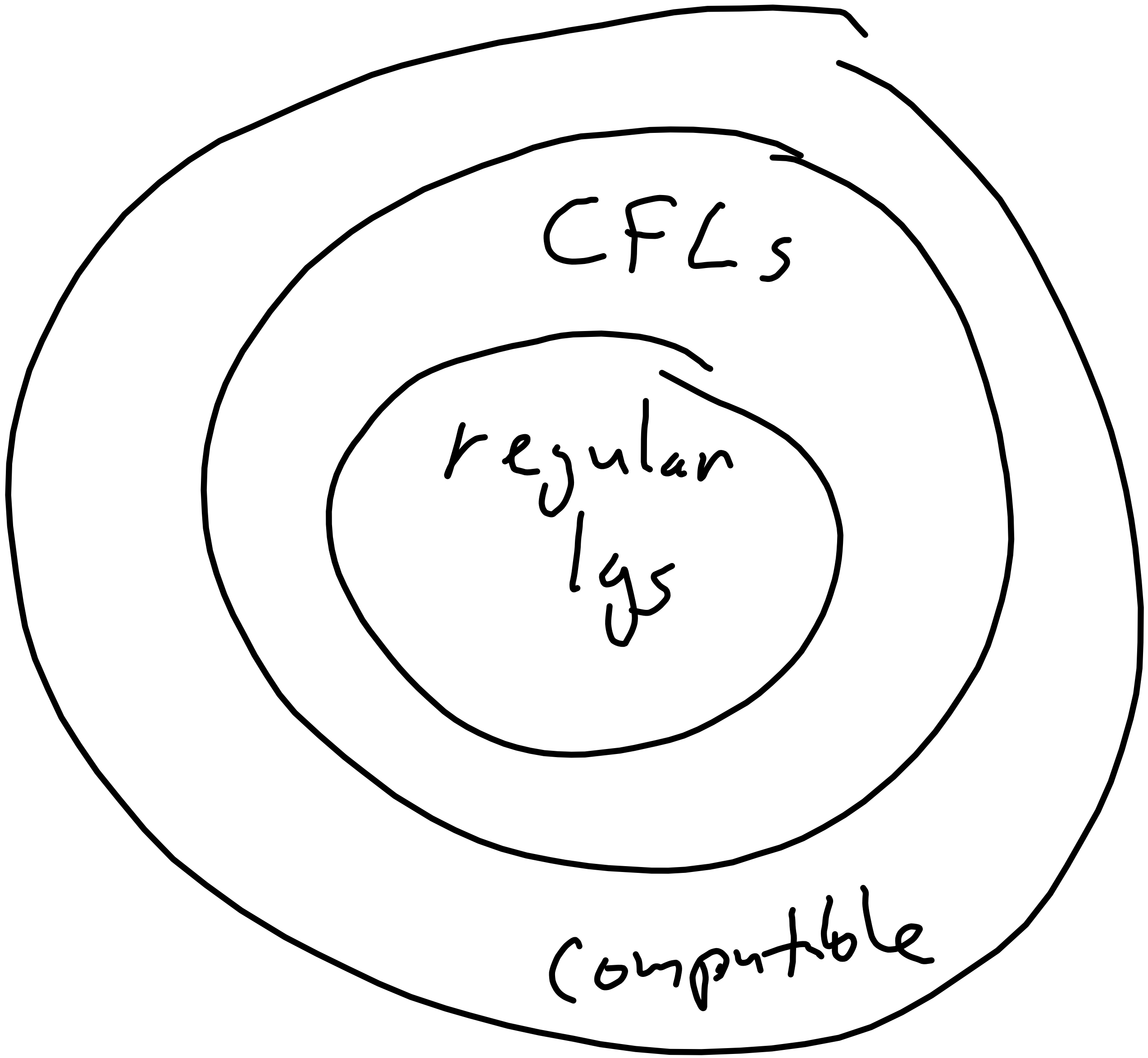
Context-free languages

(CFG's)

Declarative : grammar
CFG

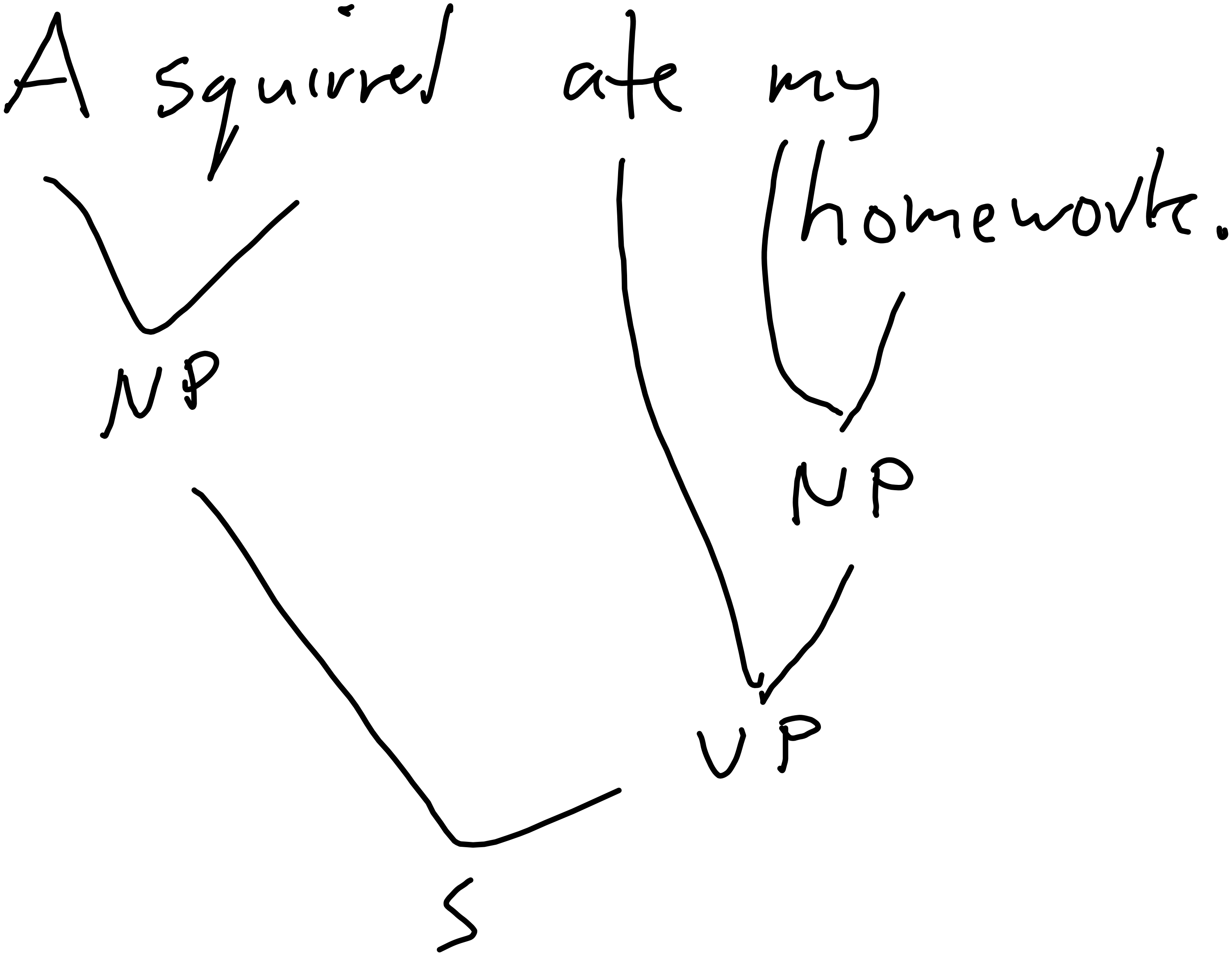
automata : pushdown
automata
PDAs

efficient algorithms : Parsers
Yacc

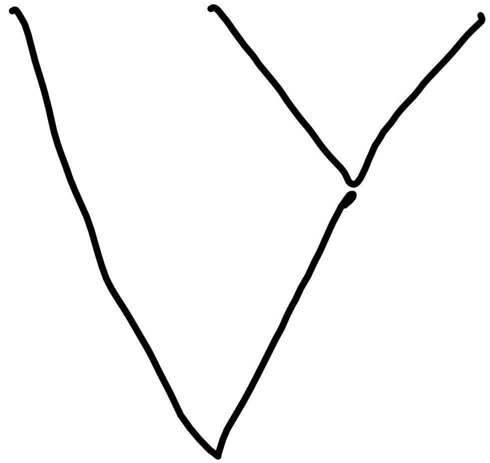

CFLs
regular lgs
computable

Natural language

processing


A squirrel ate my homework.
NP
NP
VP
S

$a + b * c$



$S \to NP \; VP$

$NP \to DET \; NP$

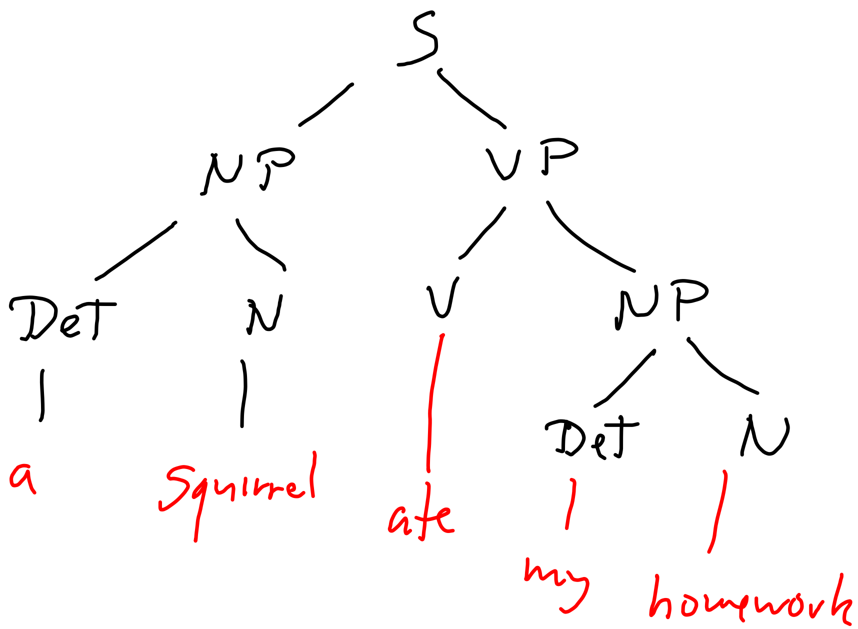
$VP \to V \; NP$

$NP \to N$

$DET \to$ my | a | The | ...

$N \to$ Squirrel | homework | truck ...

$V \to$ ate | slept | fried | ...



Palindromes

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

$$S \to \varepsilon$$

$$S \to aSa \to a\underbrace{bSb}_{S}a$$

$$\to abbSbba$$

$$\to abbabba$$

A Context-free grammar is a 4-tuple

$$\langle V, \Sigma, R, S \rangle$$

V finite set of variables

Σ finite set of terminals

$$V \cap \Sigma = \emptyset$$

S = start symbol $\quad S \in V$

R = set of "production"
rules
substitution rules

each rule is a pair

(A, w) $\quad A \to w$

variable

string of variables and terminals

Let G be a CFG

Define $L(G)$ as follows....

Suppose $A \to w$ is a rule in G

and suppose x and y are strings from $V \cup \Sigma$

Then $xAy \Rightarrow xwy$

"yields"

$$xAAy \Rightarrow xAwy$$
$$\Rightarrow xwwy$$

or

$$xAAy \Rightarrow xwAy$$
$$\Rightarrow xwwy$$

$L(G) \subseteq \Sigma^*$

$w \overset{*}{\Longrightarrow} y$ $\qquad x, y \in (V \cup \Sigma)^*$

iff There is a sequence of strings $x_1, \ldots, x_n$ s.t.

$w = x_1 \Rightarrow x_2 \Rightarrow x_3 \cdots$

$\Rightarrow x_n = y$

$n \geq 0$ $\quad \left(x \overset{*}{\Rightarrow} x \right)$

$$L(G) = \left\{ w \in \Sigma^* \text{ s.t. } S \overset{*}{\Longrightarrow} w \text{ using grammar } G \right\}$$

all terminals

$NP \rightarrow N$

$NP \rightarrow NP$ and NP

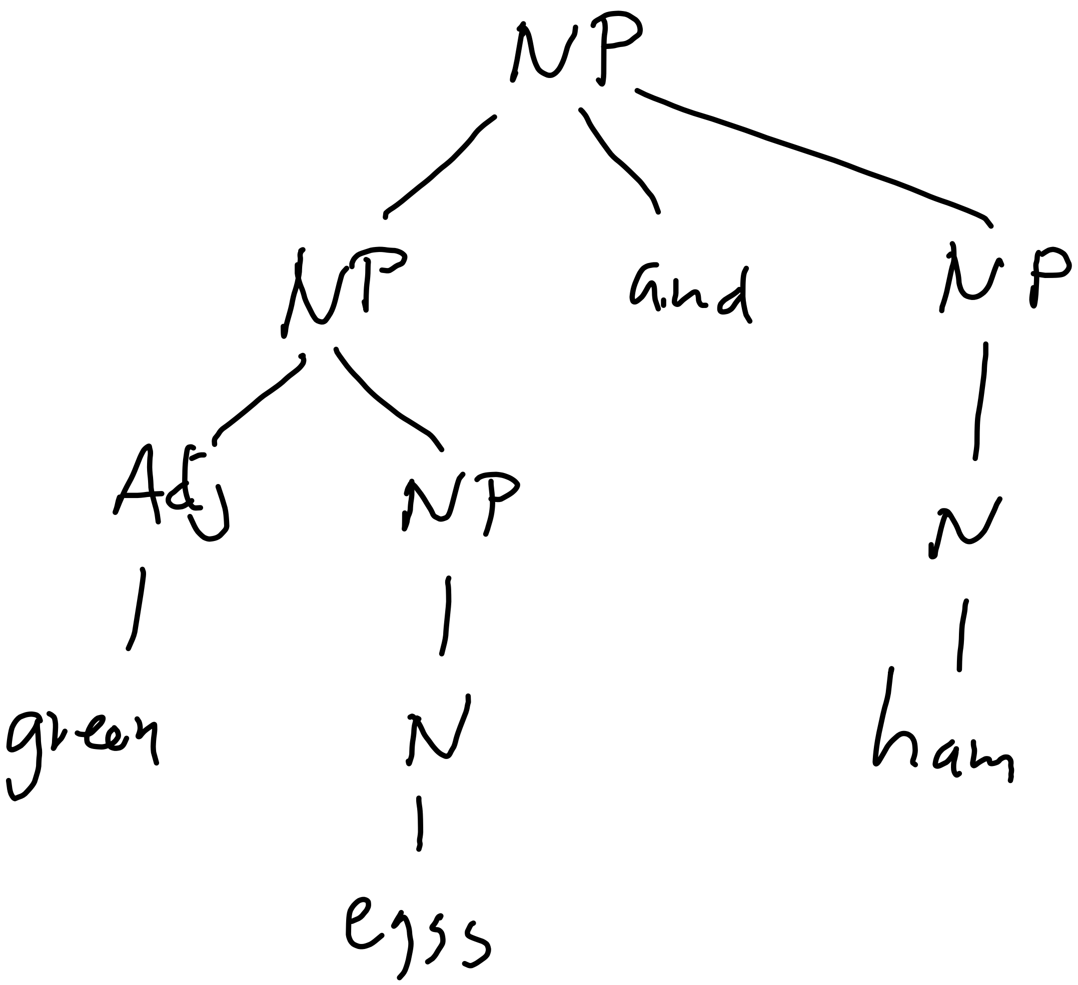
$NP \rightarrow Adj\ NP$

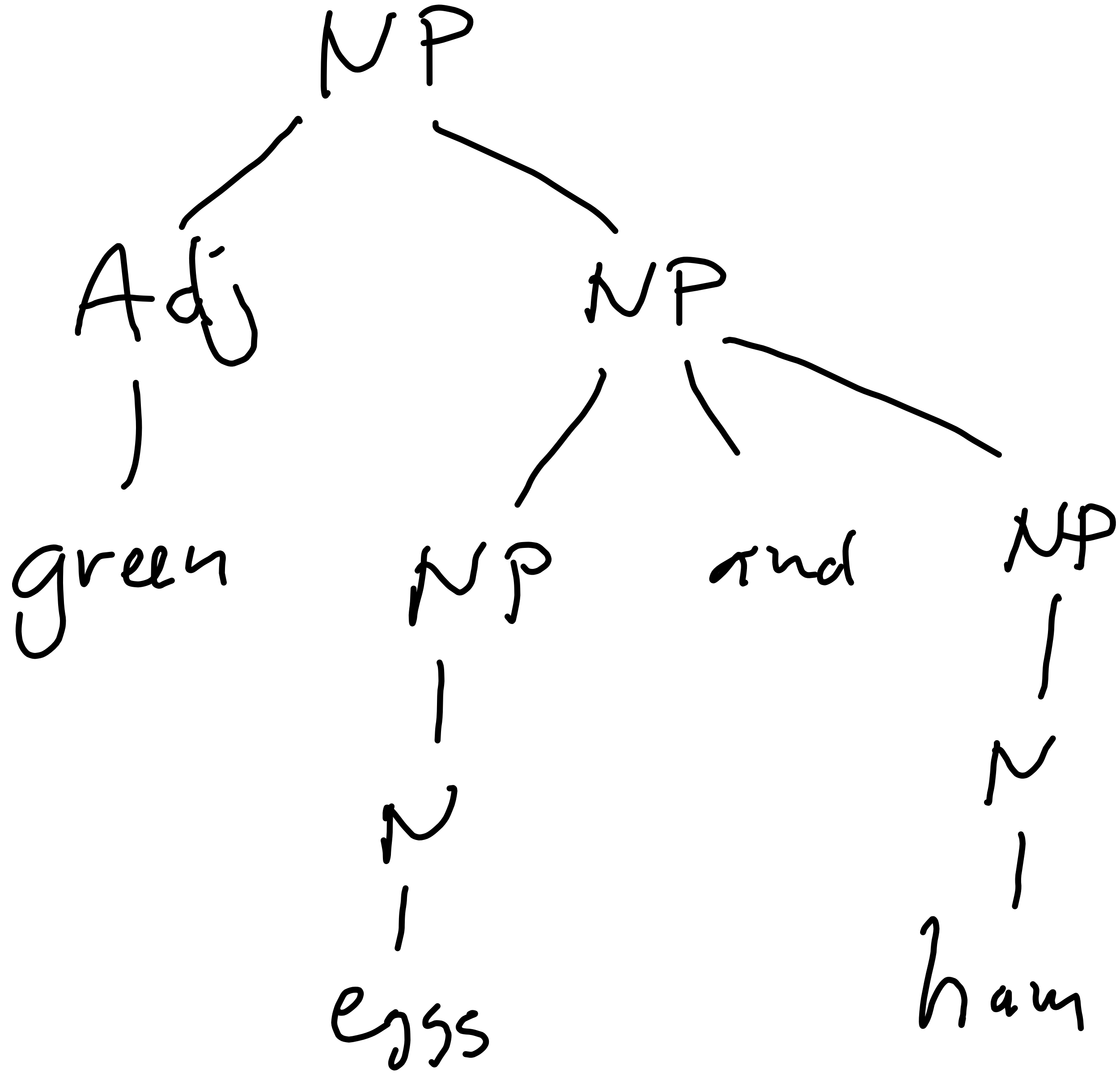
$N \rightarrow$ eggs | ham | ...

$Adj \rightarrow$ green | large | ....

NP is start symbol

(green (eggs) and ham)




NP
Adj
NP
green
NP
and
NP
N
N
eggs
ham

Chomsky – Normal Form CNF

If G is a CFG
G is in CNF if
all rules in G look like

$$A \to BC$$ B, C variables

or $$A \to a$$ a is one terminal

What if $L(G)$ contains ε?
Can G contain "useless" symbol?

If G is CFG

Then $\exists$ CFG G' in CNF

s.t.

$$L(G') = L(G) - \{\varepsilon\}$$

Outline Procedure

[• add new start symbol]

• remove ε productions

$$A \to \varepsilon$$

• remove "unit productions"

$$A \to B$$

[• remove useless symbols]

• rewrite rules with long RHS

$$A \rightarrow B\,C\,D\,E$$

$$A \rightarrow B\,A'$$

$$A' \rightarrow C\,A''$$

$$A'' \rightarrow D\,E$$