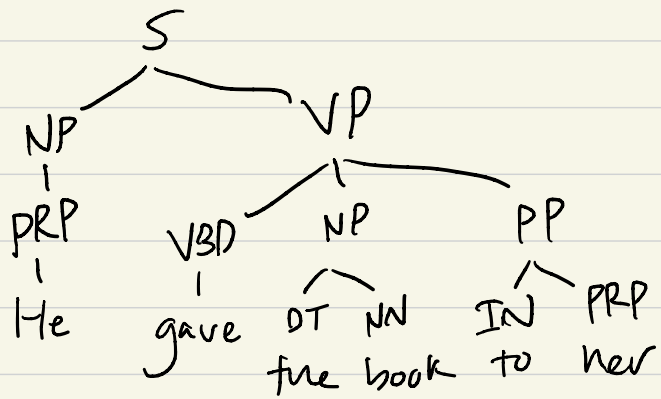
CS 378 Lecture 13

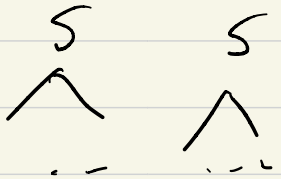
Today

- Finish constituency parsing:
  - Evaluation
  - Better PCFGs
- Dependencies
- (start) Shift-reduce parsing

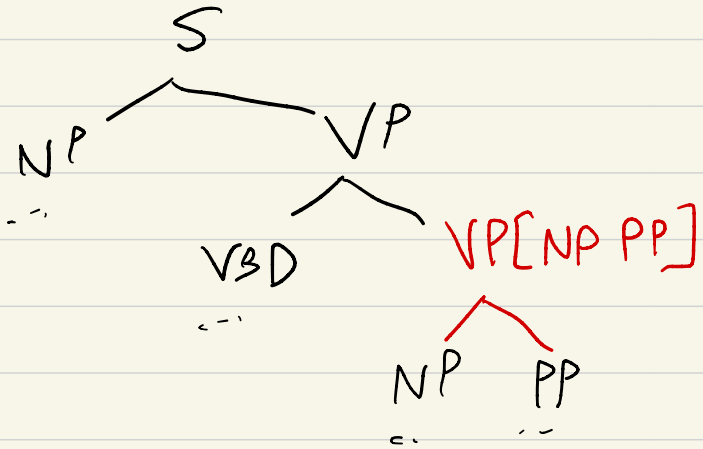
Recap Building a parser:

Input: treebank





① Binarize:

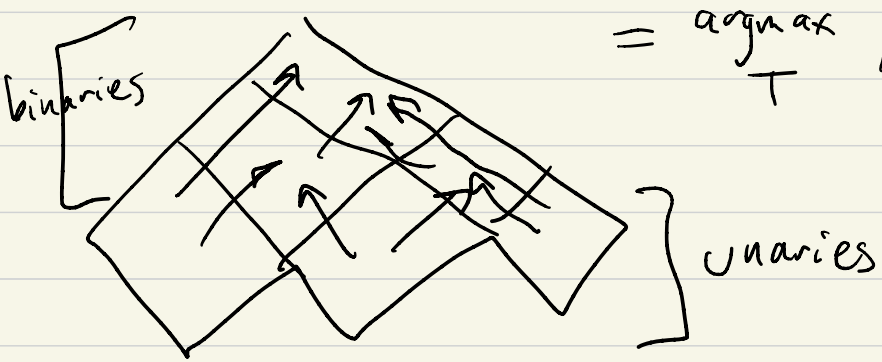


② Count + normalize to get PCFG probs

NP: $P(NP \to PRP \mid NP) = 1/2$
$P(NP \to DT\ NN \mid NP) = 1/2$
} do this over all trees

③ CKY: find $\underset{T}{\text{argmax}}\ P(T \mid \bar{x})$

$= \underset{T}{\text{argmax}}\ P(T, \bar{x})$



Announcements

- A3 updated Q3b
- Final proj: independent proposal deadline pushed back (Oct 27)
- Midterm: list of topics posted
- Next class: review

Finishing CKY

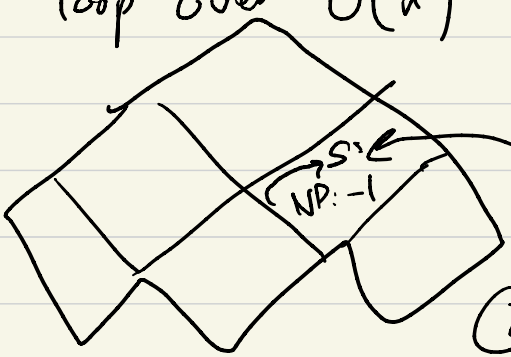
① Runtime: n words in a sentence
$O(n^3)$

Chart has $O(n^2)$ cells

Each cell: loop over $O(n)$ values of k

② Unaries:

$S \rightarrow NP$



① get values in cell w/ binaries

② apply unaries and store values in the same cell

Shift-reduce parsing

Left-to-right
Parse in linear time

State: partially built parse on a stack

Move through the sentence, add words to the partial parse as we go

Greedy, approximate