Computation, Information, and Intelligence (ENGRI/CS/INFO/COGST 172), Spring 2007
4/9/07: **Lecture 31 aid — (Lecture 30=prelim) Earley prediction; intro to grammar learning**

Topics: the prediction operation of Earley's algorithm; restricted but probabilistic variants of CFGs.

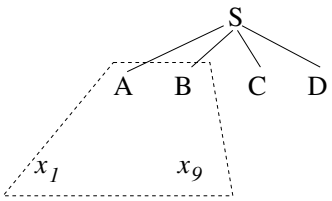
Announcements: The final exam is scheduled for Friday May 18, 2:00-4:30pm, Phillips 219.

I. An alternative interpretation of parse states $(A \rightarrow \alpha \bullet \beta, i, j)$: the branch $A \rightarrow \alpha\beta$ is part of a (partial) parse where

- the part of the sentence that the *entire branch* "covers" starts with x_i, and
- the part of the sentence that the *pre-dot* stuff α covers ends at (and includes) x_j.

II. Example Suppose the CFG has the following rewrite rules (plus others, assumedly), where "duck" is a terminal:

S →A B C D
C →E F C →H I
E →duck



III. Bigram models A bigram CFG would take the following form:

- Terminals: $w_1, w_2, \ldots, w_m$
- Nonterminals: $S, V_1, V_2, \ldots, V_m$
- Start symbol: S
- Rewrite rules: all rewrite rules of the form
  1. $V_i \rightarrow w_i V_j$,
  2. $V_i \rightarrow w_i$, or
  3. $S \rightarrow V_i$

  where $1 \leq i, j \leq m$.

IV. Sentence-ranking example A classic from the speech-recognition literature.

1. It's hard to recognize speech.
2. It's hard to wreck a nice beach.