Computation, Information, and Intelligence (COMS/ENGRI/INFO/COGST 172), Fall 2005
10/28/05: **Lecture 27 aid — Learning Grammars**

Agenda: Finish Earley's algorithm; learning n-gram models

Announcements:

- The Friday November 4th in-class prelim will cover Lecture 14 (9/26/05) through Lecture 22 (10/17/05) inclusive and Homeworks Three and Four. We will assume you have thoroughly understood the lectures and the homework problems and carefully read our solutions sets. You may bring one 8.5" × 11" sheet of paper, on which you have written or printed whatever you please on both sides, for consultation during the exam. No other references, nor calculators or other objects of that ilk, are permitted.

  Due to the prelim, the office hour schedule for *next week only* is (changes italicized):

Monday 10/31	12:45-1:45	Lillian Lee	4152 Upson
	3:30-4:30	Anton Morozov	328C Upson
	8pm-9pm	Ray Doyle	328C Upson
Tuesday 11/1	1-2	Anton Morozov	328C Upson
	2-3	Shannon McGrath	328C Upson
	3-4	Marek Janicki	328C Upson
Wednesday 11/2	*11:15-12*	Lillian Lee	4152 Upson
Thursday 10/3	*12-1*	Marek Janicki	*328A* Upson
	1-2	Shannon McGrath	*328B* Upson
	3-4	Anton Morozov	328C Upson
Friday 11/4	*Office hours cancelled (grading)*		

  You may also make appointments with us.

- The final is 9am-11:30am December 8 (day one), which is indeed Thursday, not Friday.

- To reiterate: the regrade request policy, given on the "Formal Course Descriptions and Policies" handout of 8/26/05, states that you should first read the solutions, and then contact the relevant grader (which will usually not be me) with any remaining questions.

(OVER)

I. Bigram models Let n be a large integer.

- terminals: $w_1, w_2, \ldots, w_n$
- non-terminals: $S, V_1, V_2, \ldots V_n$.
- start non-terminal: S
- rewrite rules:

$$\begin{array}{llll}
V_1 \to w_1 V_1 & V_2 \to w_2 V_1 & \cdots & V_n \to w_n V_1 \quad S \to V_1 \\
V_1 \to w_1 V_2 & V_2 \to w_2 V_2 & \cdots & V_n \to w_n V_2 \quad S \to V_2 \\
\cdots & \cdots & \cdots & \cdots \quad \cdots \\
V_1 \to w_1 V_n & V_2 \to w_2 V_n & \cdots & V_n \to w_n V_n \quad S \to V_n \\
V_1 \to w_1 & V_2 \to w_2 & \cdots & V_n \to w_n
\end{array}$$

Each rule has assigned to it a probability (a number between 0 and 1 inclusive). We require that for any nonterminal, the probabilities of all rules having that nonterminal as the lefthand side sum to one.

II. Sentence-ranking example

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

III. The poverty of the stimulus Chomsky (1957):

> "It is fair to assume that neither sentence
>
> (1) Colorless green ideas sleep furiously
>
> nor
>
> (2) Furiously sleep ideas green colorless
>
> ... has ever occurred .... Hence, in any statistical model ... these sentences will be ruled out ... as equally "remote" from English. Yet (1), though nonsensical, is grammatical, while (2) is not."

IV. Jelinek-Mercer smoothing Set the probability of a rule $V_i \to w_i V_j$ to

$$\lambda \frac{\#(w_i w_j)}{\#(w_i)} + (1 - \lambda) \frac{\#(w_j)}{\sum_k \#(w_k)}$$

where λ is between 0 and 1 (usually non-inclusive).