Computation, Information, and Intelligence (ENGRI/CS/INFO/COGST 172), Spring 2007
4/13/07: **Lecture 33 aid — Grammar induction; intro to machine translation**

Topics: Language-model smoothing; issues and approaches to machine translation.

Announcements:

- Regarding question 4(a) on Homework Five: to be clear, a "variant of interpolation smoothing wherein we assume a very simple estimate of the probability of a word occurring" is an expression that looks just like interpolation smoothing, except for a different estimate for the probability of an individual word occurring. For instance, this means that your λ should obey the restriction of lying between 0 and 1. An example, probably incorrect, response would be: "$\lambda \frac{\#(w_i w_j)}{\sum_k \#(w_i w_k)} + (1 - \lambda)\frac{3}{4}$, where $\lambda = \frac{m}{m+1}$".
- Additional self-check for question 1(a) on Homework Five: verify that $acccgggguuugga$ is correctly formatted, but $agcccgggguuugga$ is not.

I. The poverty of the stimulus The classic example, due to Noam Chomsky:

1. Colorless green ideas sleep furiously.
2. Furiously sleep ideas green colorless.

II. Another example of the sparse-data problem A standard dataset used in NLP has *95%* of the instances in the test data not occurring in the data one is allowed to learn from (Collins and Brooks, 1995).

Sometimes the situation is summed up as follows: "lack of evidence is not evidence of lack".

III. Interpolation smoothing For i between 1 and m inclusive, set the probability of a rule $V_i \rightarrow w_i V_j$ (which, in our case, corresponds to the probability that if word w_i occurs then word w_j follows it) to

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

where the interpolation parameter λ (pronounced "lambda") is between 0 and 1 (usually non-inclusive).

IV. Machine-translation paradigms Ordered by the depth of language analysis apparently required.

1. Direct replacement: word-for-word translation.
   But: "I am a fan [of this class]"

2. Syntactic transfer:
   Source-language utterance → source-language parse tree
   → target-language parse tree
   → target-language utterance
   But: "I like $\mathrm{singing}_{\mathrm{gerund}}$" vs. "Ich singe $\mathrm{gern}_{\mathrm{adverb}}$"

3. The interlingual approach:
   Source-language utterance → interlingua representation
   → target-language utterance
   But: What is "blue"?