Computation, Information, and Intelligence (COMS/ENGRI/INFO/COGST 172), Fall 2005
11/7/05: **Lecture 30 aid — Statistical machine translation (cont.)**

Agenda: Finish our very simple approach to learning translation probabilities.

Follow-ups: Our example in last lecture was adapted from Sections 26 ("Chicken and egg") and 27 ("Now for the Magic") of Kevin Knight's (1999) *A Statistical MT Tutorial Workbook* (http://www.isi.edu/natural-language/mt/wkbk.rtf), which you are encouraged to consult if you are seeking an additional reference on this topic. (The tutorial also discusses more advanced models, and is often fairly amusing to boot.)

I. Recall: notation

- For a sentence pair p, let Aligns(p) be the set of all possible alignments of the two sentences in p, and let NumAligns(p) be the size of this set.
- Let Contains($s \leftrightarrow t$) be the set of all alignments A (across all sentence pairs) that contain a position match $i \leftrightarrow j$ where the i^{th} source word was s and the j^{th} target word was t. In the example above, alignment [A1] is in Contains($maison \leftrightarrow house$) but [A2] isn't.
- Let freq($s \leftrightarrow t, A$) be the number of times we have the source word s "matched" to the target word t in alignment A. In our example above, we have freq($bleue \leftrightarrow blue, [A1]$) = 2.

II. Revised presentation: iterative learning algorithm for MT

Upon reflection, it seems better to present the update steps as incorporating the corresponding normalizations, rather than presenting the normalizations as a separate step from the updates. This change should make it more clear what "convergence" means.

1. Init: For every sentence pair p, for every alignment A of p, set $\text{awt}(A) = 1/(\text{NumAligns}(p))$.

2. Repeat the following steps in order until no "significant" change:

3. Update translation weights:
   (a) for every source/target word pair (s, t),
   set $\text{WeightedCount}(s \rightarrow t)$ to $\sum_{A \text{ in } \text{Contains}(s \leftrightarrow t)} \text{freq}(s \leftrightarrow t, A) \times \text{awt}(A)$;
   (b) set $\text{norm}_s = \sum_{t'} \text{WeightedCount}(s \rightarrow t')$;
   (c) set $\text{tr}(s \rightarrow t)$ to $\text{WeightedCount}(s \rightarrow t)/\text{norm}_s$.

4. Update alignment weights:
   (a) for every alignment $A = (1 \leftrightarrow a(1); 2 \leftrightarrow a(2); \cdots; \ell \leftrightarrow a(\ell))$ (ℓ varies for different A),
   set $\text{UnNormedAwt}(A)$ to $\text{tr}(s_1 \rightarrow t_{a(1)}) \times \text{tr}(s_2 \rightarrow t_{a(2)}) \cdots \times \text{tr}(s_\ell \rightarrow t_{a(\ell)})$;
   (b) for each pair p, compute $\text{norm}_p = \sum_{A' \in \text{Aligns}(p)} \text{UnNormedAwt}(A')$;
   (c) for every p, for every A in Aligns(p), set $\text{awt}(A)$ to $\text{UnNormedAwt}(A)/\text{norm}_p$.

III. Example of the word-segmentation problem

社長兼業務部長