Computational Linguistics Seminar
LING-696G

Week 5
Administrivia

• Guest lecturer:
  – Bryan Heidorn, School of Information
  – ISTA 455/555: Applied NLP
  – March 2nd
Chapters 3 and 4

• Statistical Machine Translation by Koehn, P.
  – Chapter 3: Probability Theory
  – Chapter 4: Machine Translation by Translating Words
Chapter 3 errata

• p. 70: Text at very top should read *This rule defines a conditional probability distribution* $p(x|y)$ (also called the *posterior*) in terms of its inverse $p(y|x)$ and the two elementary probability distributions $p(x)$ (called the *prior*) and $p(y)$ ... instead of *This rule a conditional probability distribution* $p(x|y)$ expresses in terms of its inverse $p(y|x)$ (called the *posterior*) and the two elementary probability distributions $p(x)$ (called the *prior*) and $p(y)$ ...

• p. 71: Text before Equation 3.17 should read *variance* ... is computed as the *arithmetic* mean of the *squared* difference between ... instead of *variance* ... is computed as the geometric mean of the difference between ...

• p. 71: Final number in Equation 3.18 should read 2.04 instead of 2.4.

• p. 72: Part of paragraph 2 should read *one-in-a-million chance* instead of *one-in-a-million change*.

• p. 73: In caption to Figure 3.4, entropy should be $H(X)$ instead of $E(X)$.

• p. 75: In caption to Figure 3.5, joint entropy should be $H(X,Y)$ and conditional entropy should be $H(Y|X)$. The figure itself is correct.

4.1.4 Alignment

- pg.85: re-ordering, 1 → 2

Translate

<table>
<thead>
<tr>
<th>Chinese</th>
<th>Japanese</th>
<th>English</th>
<th>German - detected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

klein ist das Haus → small is the house

das Haus is klitzeklein → the house is tiny
4.1.4 Alignment

• flavoring particle (pleonastic?)

Translate

<table>
<thead>
<tr>
<th>Chinese</th>
<th>Japanese</th>
<th>English</th>
<th>German - detected</th>
</tr>
</thead>
</table>

das Haus ist ja klein

the house is so small

Ja

• Three main functions:
  1. Expresses surprise (like doch)
  2. Expresses a sense of warning or urgency
  3. Signals something is obvious
4.1.4 Alignment

- **NULL (word 0):**

  Translate

  ![Translation Table]

  **do I (2nd choice)**

  **Subjunctive**
Chapter 4 errata

• p. 84: Text before Equation 4.5 should read \textit{at position }j\textit{ to a German output at position }i\textit{. The indexes are erroneously flipped in the book.}

• p. 90: Text before Equation 4.13 should read \textit{the same simplifications as in Equation (4.10). The book makes an erroneous reference to Equation 4.11.}

• p. 91: The two \textit{end for} on line 12 and 13 should be one indentation to the left.

• p. 91: Left hand side of equation 4.14 should be \textit{instead of} .

• p. 93: The first sum should be a product.

• p. 93: The result of the computation of Equation 4.20 should be \textit{instead of} .

• p. 93: The text should clarify that in Equation 4.21 perplexity is computed over the whole corpus, and not the average per sentence.
4.1.5 IBM Model 1

- \( f = f_1 f_2 \ldots f_{l_f} \)
- \( e = e_1 e_2 \ldots e_{l_e} \)
- \( a : j \rightarrow i \)
- \( \varepsilon = \text{normalization constant} \)

\[
p(e, a | f) = \frac{\varepsilon}{(l_f + 1)^{l_e}} \prod_{j=1}^{l_e} t(e_j | f_{a(j)}) \quad (4.7)
\]

\( l_f + 1 \) input words because word 0 = NULL
corpus not typically word aligned