Machine Translation Part 3
Alternatives & Evaluation

CS 585, Fall 2015
Introduction to Natural Language Processing
http://people.cs.umass.edu/~brenocon/inlp2015/

Brendan O’Connor
College of Information and Computer Sciences
University of Massachusetts Amherst

[Some slides borrowed from mt-class.org]
Machine Translation

- MT beyond word-based models
  - Phrase-based
  - Syntax
  - Features
- Evaluation
Phrase-based MT

\[ p(f, a \mid e) = p(f \mid e, a) \ p(a \mid e) \]

**Phrase-to-phrase** translations

- Phrases can memorize local reorderings
- State-of-the-art (currently or very recently) in industry, e.g. Google Translate
Phrase extraction for training:
Preprocess with IBM Models to predict alignments

<table>
<thead>
<tr>
<th>watashi</th>
<th>I</th>
<th>open</th>
<th>the</th>
<th>box</th>
</tr>
</thead>
<tbody>
<tr>
<td>wa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hako</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>akemasu</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

hako wo akemasu / open the box
Maria no dio una bofetada a la bruja verde

Mary did not give a slap to the witch green

did not a slap by hag bawdy

no slap to the green witch

did not give the the witch
Maria no dio una bofetada a la bruja verde

Mary did not give a slap to the witch green

did not a slap by hag bawdy

no slap to the green witch

did not give the the witch
Decoding

Maria no dio una bofetada a la bruja verde

Mary did not give a slap to the witch green

did not a slap by hag bawdy

no slap to the green witch

did not give the the witch
More MT issues

• MT beyond word-based models
  • Phrase-based
  • Syntax
  • Features
    (noisy channel is just two log-linear features)
• Evaluation