Statistical Machine Translation

- Components: Translation model, language model, decoder

![Diagram showing the components of Statistical Machine Translation]

- Foreign/English parallel text
- Statistical analysis
- Translation Model
- Decoding Algorithm
- Language Model
- English text
- Statistical analysis
Phrase-Based Translation

- Foreign input is segmented in phrases
  - any sequence of words, not necessarily linguistically motivated
- Each phrase is translated into English
- Phrases are reordered
### Phrase Translation Table

- **Phrase Translations for “den Vorschlag”:**

| English           | $\phi(e|f)$ | English           | $\phi(e|f)$ |
|-------------------|------------|-------------------|------------|
| the proposal      | 0.6227     | the suggestions   | 0.0114     |
| ’s proposal       | 0.1068     | the proposed      | 0.0114     |
| a proposal        | 0.0341     | the motion        | 0.0091     |
| the idea          | 0.0250     | the idea of       | 0.0091     |
| this proposal     | 0.0227     | the proposal ,    | 0.0068     |
| proposal          | 0.0205     | its proposal      | 0.0068     |
| of the proposal   | 0.0159     | it                | 0.0068     |
| the proposals     | 0.0159     | ...               | ...        |
Decoding Process

• Build translation left to right
  - select foreign words to be translated
Decoding Process

- Build translation left to right
  - select foreign words to be translated
  - find English phrase translation
  - add English phrase to end of partial translation
Decoding Process

- Build translation left to right
  - select foreign words to be translated
  - find English phrase translation
  - add English phrase to end of partial translation
  - mark foreign words as translated
Decoding Process

- One to many translation
Decoding Process

- Many to one translation
Decoding Process

- Many to one translation
Decoding Process

- Reordering
Decoding Process

- Translation finished
Translation Options

<table>
<thead>
<tr>
<th>Maria</th>
<th>no</th>
<th>dio</th>
<th>una</th>
<th>bofetada</th>
<th>a</th>
<th>la</th>
<th>bruja</th>
<th>verde</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mary</td>
<td>not</td>
<td>give</td>
<td>a</td>
<td>slap</td>
<td>to</td>
<td>the</td>
<td>witch</td>
<td>green</td>
</tr>
<tr>
<td>did not</td>
<td>a slap</td>
<td>by</td>
<td>green witch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>slap</td>
<td>to the</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>did not give</td>
<td></td>
<td>to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>slap</td>
<td>the witch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Look up possible phrase translations
  - many different ways to segment words into phrases
  - many different ways to translate each phrase
Hypothesis Expansion

<table>
<thead>
<tr>
<th>Maria</th>
<th>no</th>
<th>dio</th>
<th>una</th>
<th>bofetada</th>
<th>a</th>
<th>la</th>
<th>bruja</th>
<th>verde</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mary</td>
<td>not</td>
<td>give</td>
<td>a</td>
<td>slap</td>
<td>to</td>
<td>the</td>
<td>witch</td>
<td>green</td>
</tr>
<tr>
<td>did not</td>
<td></td>
<td>a slap</td>
<td>by</td>
<td>green witch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>slap</td>
<td></td>
<td>to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>did not give</td>
<td></td>
<td></td>
<td>to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>slap</td>
<td></td>
<td>the</td>
<td>witch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Start with empty hypothesis
  - e: no English words
  - f: no foreign words covered
  - p: probability 1
Hypothesis Expansion

- Pick translation option
- Create hypothesis
  - e: add English phrase Mary
  - f: first foreign word covered
  - p: probability 0.534
A Quick Word on Probabilities

- Not going into detail here, but...

- Translation Model
  - phrase translation probability $p(\text{Mary}|\text{Maria})$
  - reordering costs
  - phrase/word count costs
  - ...

- Language Model
  - uses trigrams:
  - $p(\text{Mary did not}) = p(\text{Mary}|<s>) \times p(\text{did}|\text{Mary},<s>) \times p(\text{not}|\text{Mary did})$
Hypothesis Expansion

Maria | no | dio | una | bofetada | a | la | bruja | verde

Mary did not give a slap to the witch green
did not a slap by green witch
no slap to
no
did not give

did not slap to the

- Add another hypothesis

e: witch
f: -------*
p: .182

e: Mary
f: *-------
p: .534
Hypothesis Expansion

<table>
<thead>
<tr>
<th>Maria</th>
<th>no</th>
<th>dio una bofetada</th>
<th>a</th>
<th>la</th>
<th>bruja</th>
<th>verde</th>
</tr>
</thead>
</table>

Mary _did not_ give a slap to the witch green

Mary _did not give_ a slap to the green witch

Mary _did not give_ a slap to the witch

● Further hypothesis expansion
Hypothesis Expansion

... until all foreign words covered

- find best hypothesis that covers all foreign words
- backtrack to read off translation
Hypothesis Expansion

- Adding more hypothesis
  ⇒ Explosion of search space
Explosion of Search Space

- Number of hypotheses is exponential with respect to sentence length

⇒ Decoding is NP-complete [Knight, 1999]

⇒ Need to reduce search space
  - risk free: hypothesis recombination
  - risky: histogram/threshold pruning
Hypothesis Recombination

- Different paths to the same partial translation

Mary did not give

\[ p = 0.534 \]
\[ p = 0.164 \]
\[ p = 0.092 \]
\[ p = 0.044 \]
Hypothesis Recombination

- Different paths to the same partial translation
  $\Rightarrow$ Combine paths
  - drop weaker hypothesis
  - keep pointer from worse path
Hypothesis Recombination

- Recombined hypotheses do not have to match completely
- No matter what is added, weaker path can be dropped, if:
  - last two English words match (matters for language model)
  - foreign word coverage vectors match (affects future path)
Hypothesis Recombination

- Recombined hypotheses do not have to match completely
- No matter what is added, weaker path can be dropped, if:
  - last two English words match (matters for language model)
  - foreign word coverage vectors match (effects future path)

⇒ Combine paths
Pruning

- Hypothesis recombination is not sufficient

⇒ Heuristically discard weak hypotheses

- Organize Hypothesis in stacks, e.g. by
  - same foreign words covered
  - same number of foreign words covered
  - same number of English words produced

- Compare hypotheses in stacks, discard bad ones
  - histogram pruning: keep top $n$ hypotheses in each stack (e.g., $n=100$)
  - threshold pruning: keep hypotheses that are at most $\alpha$ times the cost of best hypothesis in stack (e.g., $\alpha = 0.001$)
Hypothesis Stacks

- Organization of hypothesis into stacks
  - here: based on number of foreign words translated
  - during translation all hypotheses from one stack are expanded
  - expanded Hypotheses are placed into stacks
Comparing Hypotheses

- Comparing hypotheses with same number of foreign words covered

Maria no dio una bofetada a la bruja verde

<table>
<thead>
<tr>
<th>e: Mary did not</th>
<th>f: **-------</th>
<th>p: 0.154</th>
</tr>
</thead>
<tbody>
<tr>
<td>better partial translation</td>
<td>covers easier part</td>
<td>--&gt; lower cost</td>
</tr>
</tbody>
</table>

- Hypothesis that covers easy part of sentence is preferred

⇒ Need to consider future cost of uncovered parts