**Low-Resource MT**

- **Goal:** Construct an MT system for a new language pair.
- **Problem:** Most language-pairs either have no parallel data or a very limited corpus.

**Pharse Table Triangulation**

- **Triangulation** approximates source-target translations from source-pivot and pivot-target tables:
  \[ \hat{w}(t \mid s) = \sum_p w(t \mid p) \cdot w(p \mid s) \]
  \hfill (1)

  - The lexical weight of an aligned phrase pair \((s, t)\)
  
  \[ \hat{l}_{cx}(t \mid s) = \prod_{s \in a} \sum_{t \in T(s)} \hat{w}(t \mid s) \]

  where \(a\) denotes the word alignment.

**Pharse Table Interpolation**

- Given a source-target table \(w\), we can improve upon triangulation using interpolation:
  \[ \alpha w + (1 - \alpha) \hat{w} \]

**Analysis**

1. Triangulation yields noisy tables
2. \(\hat{w}\) and \(w\) are estimated independently
3. Interpolation can only improve words in the intersection of \(\hat{w}\) and \(w\)

**Research Question:** Can we improve on the triangulated / interpolated lexical weights using very limited source-target bilingual data?

**Monolingual data**

<table>
<thead>
<tr>
<th>Language</th>
<th>Spanish</th>
<th>French</th>
<th>Malagasy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tokens</td>
<td>1.5G</td>
<td>1.5G</td>
<td>58M</td>
</tr>
</tbody>
</table>

Monolingual data: Used to produce word embedding

**Supervised Phrase Table Triangulation for Low-Resource Languages**

Tomer Levinboim, David Chiang  
University of Notre Dame  
{levinboim.1, dchiang}@nd.edu

**Conflated Translations in Triangulation**

- Interpolation may correct conflated translations.
- The independence assumption leads to many conflated translations.
- Further conflations occur when pivoting through English, due to its poor morphology.

**Our approach:**

1. Derive word co-occurrence counts \(c_{\text{emp}}(s, t)\) from limited source-target parallel data / dictionary
2. Train a discriminative model: correct translation (according to \(c_{\text{emp}}\)) should become likely, while incorrect ones should be down-weighted.
3. Generalize beyond the intersection of \(w\) and \(\hat{w}\); by representing words with embedding (word2vec)

**Optimization Problem**

Learn the matrix \(A\) and scalar \(h\):

\[ \max_{A, h} L(A, h) = \max_{A, h} \sum_{s, t} c_{\text{emp}} \log q(t \mid s). \]

- The matrix \(A\) maps between the source/target embedding spaces; \(h\) quantifies how much the triangulation \(\hat{w}\) should be trusted.
- Solved with gradient descent and AdaGrad

**Learning a Discriminative Mapping**

- The matrix \(A\) should map correct Spanish-French translations closer together, and incorrect ones, such as trabajo-empleie, further apart.

**Learning from a 50k Spanish-French Corpus**

<table>
<thead>
<tr>
<th>acceptamos</th>
<th>acceptas</th>
<th>accepter</th>
<th>crayons</th>
</tr>
</thead>
<tbody>
<tr>
<td>norm. (c_{\text{emp}})</td>
<td>1.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(w) (triang.)</td>
<td>0.05</td>
<td>0.39</td>
<td>-</td>
</tr>
<tr>
<td>ours (no (\hat{w}))</td>
<td>0.62</td>
<td>-</td>
<td>0.06</td>
</tr>
<tr>
<td>ours</td>
<td>0.98</td>
<td>0.01</td>
<td>-</td>
</tr>
</tbody>
</table>

Translation probabilities produced by our method better approximate the held out dev set distributions, compared to triangulation (\(\hat{w}\)). The dev set distributions were obtained by deriving \(c_{\text{emp}}\) from an aligned French-Spanish corpus (50k sentences) and normalizing.

**Spanish-English-French**

- Direct: 4k Spanish-French; **Triangulation:** 50k Spanish-English, 50k English-French

**Malagasy-English-French**

- Direct: 1.1k Malagasy-French dictionary entries; **Triangulation:** 100k Malagasy-English, 50k English-French

**Experiment**

1. Produce counts \(c_{\text{emp}}\) from bilingual data
2. Split word-pairs to 70/30% train/dev sets and train the model, selecting the Total Variation minimizing iteration on the dev set
3. Re-score and append lexical weights to the triangulated / interpolated phrase table

**Supervision with few dictionary entries improves upon triangulation by +0.5 BLEU**