1. Overview

**Long-term Goals**
- Bootstrap speech technology for non-written and under-resourced languages
- Given: Audio data
  - Their written translations in another language (e.g., English)
- Collect training data for ASR and MT systems rapidly and at low cost
  - Pronunciation dictionary
  - Parallel corpus, language model
- **Goal of this Paper**
  - Segment phoneme sequences into word units using the written translations
  - Simulate phoneme recognition errors realistically
  - Compare our cross-lingual word segmentation method to monolingual ones, e.g., Adaptor Grammars (Johnson, 2008)

2. Cross-Lingual Alignment

**IBM Model 3**

Problem: Generative story does not fit word-to-phoneme alignment

**Model 3P**

- Extends generative story of IBM Model 3 with additional steps
- Uses GIZA++ alignments to initialize Model 3P parameters
- Then our PISA alignment tool\(^1\) applies EM algorithm

---

3. Experiments and Results

**Compare:**
1. Adaptor Grammars (Monolingual)
2. GIZA++ word-to-phoneme alignments
3. Model 3P

**Experimental Setup**
- English-Spanish BTEC corpus (123k sentence pairs)
- Phoneme recognition errors up to 25.3% were simulated using the confusion matrix of a Spanish phoneme recognizer trained on the Spanish portion of GlobalPhone (Schultz, 2002)

---

**Results**

<table>
<thead>
<tr>
<th>F-Score</th>
<th>0 %</th>
<th>5 %</th>
<th>10 %</th>
<th>15 %</th>
<th>20 %</th>
<th>25 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phoneme Error Rate</td>
<td>0 %</td>
<td>20 %</td>
<td>40 %</td>
<td>60 %</td>
<td>80 %</td>
<td>100 %</td>
</tr>
</tbody>
</table>

---