FBK @ IWSLT-2008

N. Bertoldi, M. Federico, R. Cattoni, †M. Barbaiani

FBK, Trento - Italy
† Rovira i Virgili University, Tarragona - Spain

October 20th, 2008
FBK goal

Pivot translation in real-world condition

• improving translation for low-resourced languages:
  – few parallel data for Italian-centric language pairs: Chinese, Arabic, ...
• improving translation among intra-European languages
• applying pivot-like strategies to adapt SMT systems to different domains

• theoretical foundation of pivot translation task
• mathematically sound definition of approaches
• experimental comparison
Most effort on Pivot Task

- good benchmark:
  - controlled conditions, controlled domain
  - fast development cycle because of small size
  - many competitors

- participation to other IWSLT tasks, but with limited effort:
  - no use of additional data
  - no adaptation to challenge task
  - no optimization for speech input
Task Description

- traveling domain
- Basic Travel Expression Corpus

- BTEC tasks:
  - translation from Chinese into English and from Chinese into Spanish

- Pivot task:
  - translation from Chinese into Spanish without C-S parallel data
  - only independent C-E and E-S parallel data available

- Challenge task:
  - translation from Chinese into English of tourism-related dialogues (no BTEC)

- input condition:
  - automatic and correct transcriptions
  - read (BTEC and Pivot) and spontaneous (Challenge) speech
Task description: data

• training data:
  – monolingual corpora: C1 and C2, E1 and E2, and S1
  – parallel corpora: CE2, ES1, development sets (with multiple refs)
  – CES1 never used as trilingual parallel corpus
  – no additional data (although allowed)

• development data
  – dev set: 506 Chinese sentences with 16 refs in English and Spanish
  – other dev sets for C-E BTEC and Challenge tasks
  – blind devtest set: 1K sentences with 1 reference
  – reduced training corpora (19K sentences) for development

• test set: 507 Chinese sentences

• preprocessing: tokenization, numbers into digits, Chinese word-segmentation
## Pivot Task description: data

<table>
<thead>
<tr>
<th>task</th>
<th>data</th>
<th>sent</th>
<th>source</th>
<th>target</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>words</td>
<td>dict</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Btec</td>
<td>CE1*</td>
<td>18,974</td>
<td>161K</td>
<td>8,017</td>
</tr>
<tr>
<td></td>
<td>CS1*</td>
<td>18,974</td>
<td>161K</td>
<td>8,017</td>
</tr>
<tr>
<td>Pivot</td>
<td>CE2*</td>
<td>18,999</td>
<td>150K</td>
<td>8,114</td>
</tr>
<tr>
<td></td>
<td>ES1*</td>
<td>18,974</td>
<td>172K</td>
<td>8,210</td>
</tr>
<tr>
<td>Btec</td>
<td>CE1+dev</td>
<td>54,021</td>
<td>439K</td>
<td>8,847</td>
</tr>
<tr>
<td></td>
<td>CS1+dev</td>
<td>28,068</td>
<td>229K</td>
<td>8,284</td>
</tr>
<tr>
<td>Pivot</td>
<td>CE2+dev</td>
<td>28,095</td>
<td>217K</td>
<td>8,987</td>
</tr>
<tr>
<td></td>
<td>ES1+dev</td>
<td>19,972</td>
<td>182K</td>
<td>8,385</td>
</tr>
<tr>
<td>Challenge</td>
<td>CE1+dev</td>
<td>55,743</td>
<td>447K</td>
<td>8,864</td>
</tr>
</tbody>
</table>

- training data during development (*)
- training data the final submissions including development sets (+dev)
**Direct baseline system**

- open-source MT toolkit **Moses**
- statistical **log-linear** model with 8 features
- weight optimization by means of a **minimum error training** procedure

- **phrase-based** translation model:
  - direct and inverted frequency-based and lexical-based probabilities
  - phrase pairs extracted from symmetrized word alignments (GIZA++)
- 5-gram word-based LM exploiting Improved Kneser-Ney smoothing (IRSTLM)
- standard negative-exponential distortion model
- word and phrase penalties
Direct system: performance

<table>
<thead>
<tr>
<th>data</th>
<th>BLEU</th>
<th>OOV</th>
<th>applied to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese-English CE1*</td>
<td>26.91</td>
<td>2.00</td>
<td>Btec and Challenge</td>
</tr>
<tr>
<td>Chinese-English CE2*</td>
<td>19.09</td>
<td>3.80</td>
<td>Pivot</td>
</tr>
<tr>
<td>English-Spanish ES1*</td>
<td>49.13</td>
<td>2.01</td>
<td>Pivot</td>
</tr>
<tr>
<td>Chinese-Spanish CS1*</td>
<td>23.67</td>
<td>2.00</td>
<td>Btec</td>
</tr>
</tbody>
</table>

- systems trained on reduced data
- performance on the blind devtest, extracted from CE1 and ES1
- significant mismatch between corpora 1 and 2
- translation from Chinese into English easier than into Spanish
- translation from English into Spanish "easy"
Pivot SMT

- **Goal:**
  - translation from Chinese into Spanish without parallel data

- **Assumption:**
  - two parallel corpora C-E and E-S, with independent English side
  - full-fledged Direct systems trained on C-E and E-S parallel data

- **Approaches:**
  - Coupling C-E and E-S systems at sentence level
  - Coupling C-E and E-S systems at phrase level
  - Synthesizing C-S parallel data and building a full-fledged C-S system
Coupling systems at sentence level

- Corpus
- Train
- Phrase table
- LM
- Moses
- Input
- N = M = 100

- Corpus
- Train
- Phrase table
- LM
- Moses
- 1 best
- N best
- Rescore
- Output

Bertoldi et al. FBK @ IWSLT-2008 October 20th, 2008
Coupling systems at phrase level
Synthesis of parallel data

N = M = 100
Official results of Pivot Task

<table>
<thead>
<tr>
<th>system</th>
<th>run</th>
<th>ASR.1</th>
<th>CRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cascade 1-best</td>
<td>contr6</td>
<td>29.20</td>
<td>33.52</td>
</tr>
<tr>
<td>Cascade Nbest</td>
<td>contr7</td>
<td>32.69</td>
<td>37.41</td>
</tr>
<tr>
<td>PT Composition</td>
<td>contr4</td>
<td>28.52</td>
<td>33.13</td>
</tr>
<tr>
<td>Synthesis</td>
<td><strong>prim</strong></td>
<td><strong>33.11</strong></td>
<td><strong>39.69</strong></td>
</tr>
<tr>
<td></td>
<td>contr1</td>
<td>34.14</td>
<td>39.93</td>
</tr>
</tbody>
</table>

- big gain using 100-best wrt to 1best
- less than 2 BLEU points wrt top performing (39.69 vs 41.57)
- avoiding the CE translation, which poorly performs, is a winning strategy
- ASR (- 13/17% relative) confirms the same results as CRR

- contr1 includes the C-S parallel data of the dev set, **not independent data**
- using correct Spanish translations is better than using synthesized ones
Thank you!
### Official results of all submissions

<table>
<thead>
<tr>
<th>Task</th>
<th>System</th>
<th>Run</th>
<th>BLEU ASR.1</th>
<th>CRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE-btec</td>
<td>Direct</td>
<td>prim</td>
<td>36.91</td>
<td>40.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>contr</td>
<td>36.45</td>
<td>&quot;</td>
</tr>
<tr>
<td>CS-btec</td>
<td>Direct</td>
<td>prim</td>
<td>26.67</td>
<td>30.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>contr</td>
<td>27.05</td>
<td>&quot;</td>
</tr>
<tr>
<td>CE-chal</td>
<td>Direct</td>
<td>prim</td>
<td>23.84</td>
<td>27.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>contr</td>
<td>23.88</td>
<td>&quot;</td>
</tr>
<tr>
<td>CES-pivot</td>
<td>Cascade</td>
<td>contr6</td>
<td>29.20</td>
<td>33.52</td>
</tr>
<tr>
<td></td>
<td>Nbest</td>
<td>contr7</td>
<td>32.69</td>
<td>37.41</td>
</tr>
<tr>
<td></td>
<td>PhraseTable</td>
<td>contr4</td>
<td>28.52</td>
<td>33.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>contr5</td>
<td>30.09</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>Synthesis</td>
<td>prim</td>
<td>33.11</td>
<td>39.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>contr2</td>
<td>35.94</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>contr1</td>
<td>34.14</td>
<td>39.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>contr3</td>
<td>35.98</td>
<td>&quot;</td>
</tr>
</tbody>
</table>