Addicter: What’s Wrong With My Translations?

Dan Zeman, Mark Fishel
Jan Berka, Ondřej Bojar

Charles University in Prague
University of Zurich
Visualizer and Error Labeler

• ADDICTER = Automatic Detection and Display of Common Translation Errors

• Error labeling part (Mark)

• Visualizing part (Dan):
  ▪ View word-aligned corpora
  ▪ Look up corpus examples of a word
  ▪ Look up word occurrences in phrase table
  ▪ Alignment summary of a word
  ▪ Browse test data
    • In addition to the above, also shows auto-detected errors
HTML Visualization

• Cheap interface (from the developers point of view)
• Displayed by your favorite browser
• Words are clickable
  ▪ Links to their own examples
• Alignments shown using tables
  ▪ Simple sentence pairs possibly better using graphics
  ▪ Complex reordering? Graphics not that good.
  ▪ Besides, it would be difficult to show in HTML.
**source**

deník aktuálně. cz členy nové rady z ČSSD " vyzkoušel ", jak znají notoricky zavedený slan

**target**

Aktuálně. cz " tested " the Social Democrat members of the new Council in terms of the well
current coalition partners.

**system hypothesis**

newspaper currently. cz new Council members from ČSSD " " tried , as know notoriously e
current coalition partners.

<table>
<thead>
<tr>
<th>deník</th>
<th>aktuálně</th>
<th>čz</th>
<th>členy</th>
<th>nové</th>
<th>rady</th>
<th>z</th>
<th>ČSSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aktuálně</td>
<td>0-0 1-0</td>
<td>2-1</td>
<td>3-2 3-8</td>
<td>4-9 4-10</td>
<td>5-12</td>
<td>6-13</td>
<td>7-14 7-15</td>
</tr>
<tr>
<td>Aktuálně</td>
<td>2-1</td>
<td>cz ČSSD vyzkoušel</td>
<td>9-3</td>
<td>10-4</td>
<td>11-5</td>
<td>14-6</td>
<td>cz 3-8</td>
</tr>
<tr>
<td>newspaper</td>
<td>currently</td>
<td>čz</td>
<td>new</td>
<td>Council</td>
<td>members</td>
<td>from</td>
<td>ČSSD vyzkoušel</td>
</tr>
<tr>
<td>deník</td>
<td>aktuálně</td>
<td>1-1</td>
<td>2-2</td>
<td>3-3</td>
<td>5-4</td>
<td>6-5</td>
<td>členy</td>
</tr>
</tbody>
</table>
You May Be Used to This...

In the first round, half of the amount is planned to be spent.

V prvním kole bude použita polovina částky.
In the first round, half of the amount is planned to be spent.
Alignment Summary

The word 'Obama' occurred 1423 times and got aligned to 627 distinct

1. Obama (249)
2. (242)
3. , (56)
4. Obamovi (17)
5. se (16)
6. prezidenta (15)
7. Obamy (14)
8. Obamu (14)
9. prezident (12)
10. že (11)
11. . (11)
12. Obamou (11)
13. a (10)
14. v (9)
15. (7)
16. Barack (6)
17. na (6)
18. k (6)
19. Obamova (6)
20. s (5)
How to Use

• Word occurrences are first indexed
• Then a Perl script generates the HTML
• Test data browsing: static HTML
• Training data / word examples: dynamic only
  ▪ Do not pre-generate zillions of pages
  ▪ Drawback: web server + CGI needed
Translation Error Analysis

• Any Single-Number Metric may be good for…
  ▪ comparing two systems on given dataset
  ▪ tuning model weights (if easily computable)

• Rarely, if at all…
  ▪ does the absolute value tell anything

• BUT NEVER…
  ▪ points directly to the particular weaknesses of the system
Error detection and labelling

- src: per favore una pizza “ quattro stagioni “ .

- ref: a “ four seasons “ pizza please .

- hyp-1: one “ four seasons “ pie as a favor .

- hyp-2: please , a pizza “ stage four “ .
Error detection and labelling

• Error taxonomy similar to Vilar et al. (2006)
  ▪ Inflection error / untranslated word
  ▪ Lexical choice error
  ▪ Missing (functional/content word)
  ▪ Superfluous
  ▪ Punctuation
  ▪ Misplaced word (locally/globally)
Error detection and labelling

- Works on word-level
- Requires reference and hypothesis
  - Can benefit from source text, lemmas&PoS-tags
- Uses monolingual alignment
  - Addicter's (...) or any other
  - Requires injective (1-to-1) alignments
  - Can find the “optimal injective subset” for non-injective alignments
- Multiple errors per word allowed
Addicter's alignment

- Lightweight (no learning, no external resources)
- Applied to lemmas (can be done with anything else)
  - Only identical lemmas can be aligned
- HMM-based “disambiguation”
  - \( p_{\text{trans}}(a_n \mid a_{n-1}) \sim \exp(-b \times |a_n - a_{n-1} - 1|) \)
  - Stimulates to align similarly to previous alignment
  - Exponential time, solved via beam-search
Lexical errors

- Errors are classified, using the alignments:
  - Unaligned = missing (in ref) / extra (in hyp)
    - Classified into functional/content via pos-tags
  - Aligned: diff. word, same lemma = inflection error
  - Aligned: diff. word and lemma = lex. choice error
  - Any error on punctuation = punctuation error
Order errors

• To find these, alignment is “unscrambled”
  ▪ Find the minimum number of rearrangements to fix the order
• Transposed adjacent elements = local reordering
• Shifted elements = global reordering
Evaluation

• Data: wmt09 en-cz, 200 sentences * 4 systems
  ▪ Tagged manually with translation errors

• Alignments:
  ▪ Addicter
  ▪ METEOR
  ▪ Bilingual (GIZA++, Berkeley)
  ▪ Via source (CzEng)

• Evaluation: precision/recall of all error tags
## Results

<table>
<thead>
<tr>
<th>Alignment Method</th>
<th>Alignment</th>
<th>Translation Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prec.</td>
<td>Rec.</td>
</tr>
<tr>
<td><strong>addicter</strong> &amp; <strong>via_source</strong></td>
<td>86.39</td>
<td>85.89</td>
</tr>
<tr>
<td><strong>addicter</strong></td>
<td>98.89</td>
<td>72.18</td>
</tr>
<tr>
<td><strong>addicter</strong> &amp; <strong>meteor</strong></td>
<td>97.90</td>
<td>71.54</td>
</tr>
<tr>
<td><strong>addicter</strong> &amp; <strong>giza++ intersect</strong></td>
<td>85.99</td>
<td>77.78</td>
</tr>
<tr>
<td><strong>addicter</strong> &amp; <strong>berkeley</strong> &amp; <strong>via_source</strong></td>
<td>73.67</td>
<td>83.50</td>
</tr>
<tr>
<td><strong>addicter</strong> &amp; <strong>berkeley</strong></td>
<td>71.23</td>
<td>78.31</td>
</tr>
<tr>
<td><strong>addicter</strong> &amp; <strong>giza++ grow-diag</strong></td>
<td>65.93</td>
<td>74.58</td>
</tr>
<tr>
<td><strong>via_source</strong></td>
<td>85.00</td>
<td>74.60</td>
</tr>
<tr>
<td><strong>giza++ intersect</strong></td>
<td>81.65</td>
<td>64.09</td>
</tr>
<tr>
<td>berkeley*</td>
<td>68.12</td>
<td>74.38</td>
</tr>
<tr>
<td>meteor</td>
<td>90.37</td>
<td>55.04</td>
</tr>
<tr>
<td>giza++ <strong>grow-diag</strong></td>
<td>61.54</td>
<td>69.95</td>
</tr>
</tbody>
</table>
# Results

<table>
<thead>
<tr>
<th>Wrong hypothesis word</th>
<th>Misplaced word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flag</td>
<td>Prec.</td>
</tr>
<tr>
<td>extra</td>
<td>19.24</td>
</tr>
<tr>
<td>unk</td>
<td>13.39</td>
</tr>
<tr>
<td>form</td>
<td>38.16</td>
</tr>
<tr>
<td>lex/disam</td>
<td>18.48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Missing reference word</th>
<th>Punctuation error</th>
</tr>
</thead>
<tbody>
<tr>
<td>miss_c</td>
<td>2.17</td>
</tr>
<tr>
<td>miss_a</td>
<td>4.78</td>
</tr>
</tbody>
</table>
Experiment Results

• Underaligned translations => miss/extra overkill

• Dependence on a single reference is bad

• Alignment and error detection quality do not correlate
  ▪ 1-to-1 alignment requirement to blame
  ▪ Have to go to phrase-/syntax-/etc.-based alignments
Future (this week?)

- Lots of improvements possible
- Philipp-style corpus occurrences?, aka collocations
- Index of lemmas
  - Find all occurrences of a word regardless form
- Perl-based web server?
- Further integration between visualization and error analysis
- Further testing of error analysis
- Symbiosis with Hjerson