An Empirical Study: Post-editing Effort for English to Arabic Hybrid Machine Translation

Hassan Sajjad, Francisco Guzman, Stephan Vogel
Qatar Computing Research Institute, HBKU
Introduction

• Old Arabic documents

• Translation of metadata from English to Arabic
Traditional Translation Process

British Library → Translation Company → Translators
Problem

• Various small documents
• Fewer overlap at sentence/segment level
  – A lot needs to be translated from scratch
• Time and cost inefficient
Solution: Hybrid Machine Translation

Hybrid MT: Combines the benefits of both! Translation Memory and Customized MT

High precision translations

100% recall – readily available translations
Hybrid MT System

Translation Memory
– First pass: use strict matching to translate known words and phrases

Customized Machine Translation
– Second pass: translate the remaining text using machine translation system
Aiming higher: Post Editing for Quality

- High quality
- High consistency
- Cost and time effective
Customized Machine Translation

• A statistical machine translation system
  – Train specific to the domain of the text that needs to be translated

• General practice
  – Use Moses
  – Train on the data of translation memory
  – Follow recipe of a competition grade system to ensure high quality
English to Arabic CMT

- Best competition grade pipeline involves
  - Arabic (de-) tokenization
    - Splitting morphologically rich words into smaller segments and vice-versa
    - +1.5 BLEU points improvement

- Arabic (de-) normalization
  - Mapping different forms of a letter to one form and vice versa
  - +0.5 BLEU point improvement

This ensures high quality but does not guarantee less frustration for post-editors
Why?

Translation output requires:

• De-tokenization and de-normalization

• De-normalization introduces character-level errors
  – Frustrating for the post-editor to correct
  – Time inefficient
Recommended Practices for CMT of English-Arabic

• Don’t normalize

But

• Always tokenize
  – Improve coverage of words
  – Better translations
Let’s Talk about BL Case Numbers!

We compare:
• Translation Memory (TM) only
• Hybrid MT (TM + CMT)

Also:
• Translator
• Hybrid MT + Post editing (PE)

Looking at:
• Effectiveness
• Quality
• Consistency
Data

• 1000 documents
  – 90k parallel sentences/segments
  – 953 documents for training
    • 489k tokens
  – Rest for tune and test
## Effectiveness of TM

<table>
<thead>
<tr>
<th></th>
<th>Exact match</th>
<th>Fuzzy match</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>50% segments</strong></td>
<td><strong>7% words</strong></td>
<td><strong>84% segments</strong></td>
</tr>
<tr>
<td><strong>BUT COVERS ONLY</strong></td>
<td><strong>BUT COVERS ONLY</strong></td>
<td></td>
</tr>
</tbody>
</table>

More than 85% of words still need to be translated !!!!

* Based on an assessment over X documents
Effectiveness of CMT

100% segments AND 99.9% words translated!
Effectiveness of Hybrid MT

• High precision
  – TM exact matches

• High recall
  – CMT to produce high quality translations
Assessing Quality

- **BLEU**
  - Compare output to ‘reference’ translation

<table>
<thead>
<tr>
<th></th>
<th>Strict</th>
<th>Partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM</td>
<td>7.07</td>
<td>21.01</td>
</tr>
<tr>
<td>TM + CMT</td>
<td>54.60</td>
<td>48.54</td>
</tr>
</tbody>
</table>

CMT alone BLEU scores are 53.90
Assessing Quality

• TER: Translation Error Rate
  – How much effort is needed to get perfect translation?
  – Compare to ‘reference’ translation

Hybrid MT can improve beyond that!!!
Assessing Quality

• TER vs. Post editing effort
  – Similar effort estimation using post-editing of Hybrid MT

* PE is based on an assessment over 4 documents, using a junior translator
Consistency of Hybrid MT

- We compared Hybrid MT versus a junior translator
- We measured consistency with reference translations

* Based on an assessment over 4 documents
Speedup of Hybrid MT

- We compared Hybrid MT versus a junior translator

* Based on an assessment over 4 documents

Hybrid MT+PE is 30% more efficient
Conclusion

• Hybrid MT
  – High precision and high recall

• Hybrid MT plus Post-editing
  – Efficient in terms of both time and cost
  – Improves consistency

• Customized MT for English-Arabic
  – Don’t normalize but always tokenize
References


Thank you