

An Empirical Study:

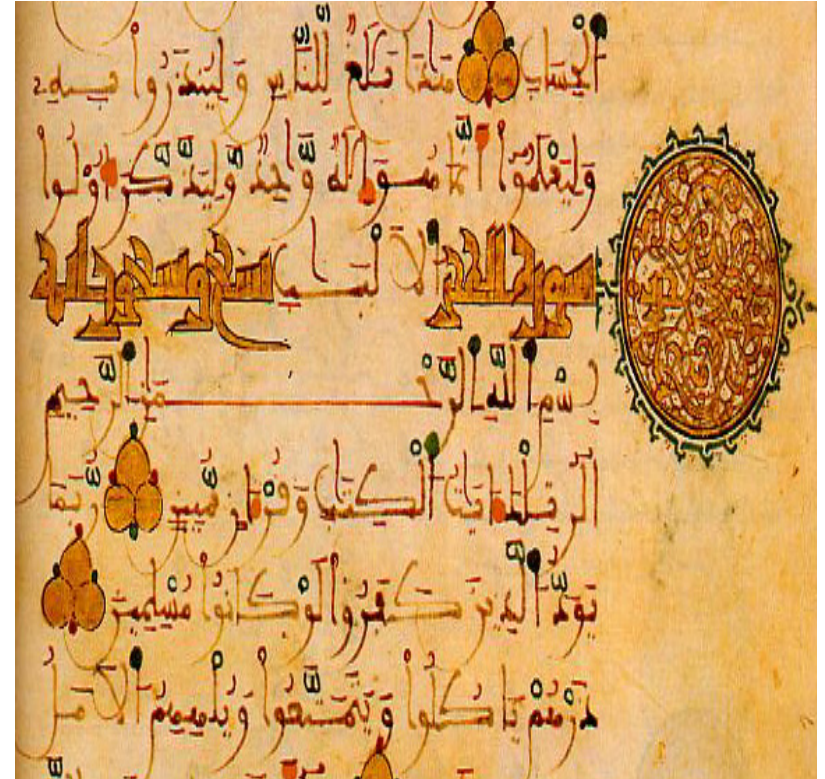
Post-editing Effort for **English to Arabic** Hybrid Machine Translation

**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

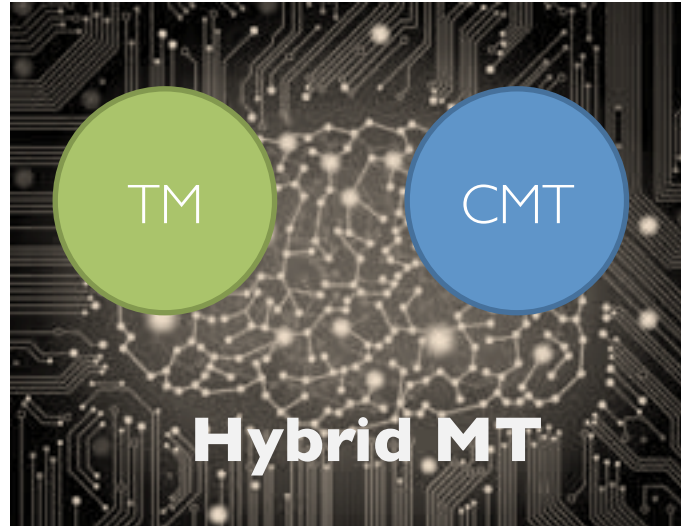
Translators

# Problem

- Various small documents
- Fewer overlap at sentence/segment level
- Few translation memory matches
  - A lot needs to be translated from scratch
- Time and cost inefficient

# Solution: Hybrid Machine Translation

High precision  
translations



100% recall –  
readily available  
translations

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

# Hybrid MT System



## Translation Memory

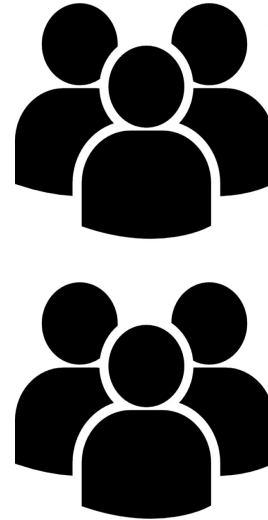
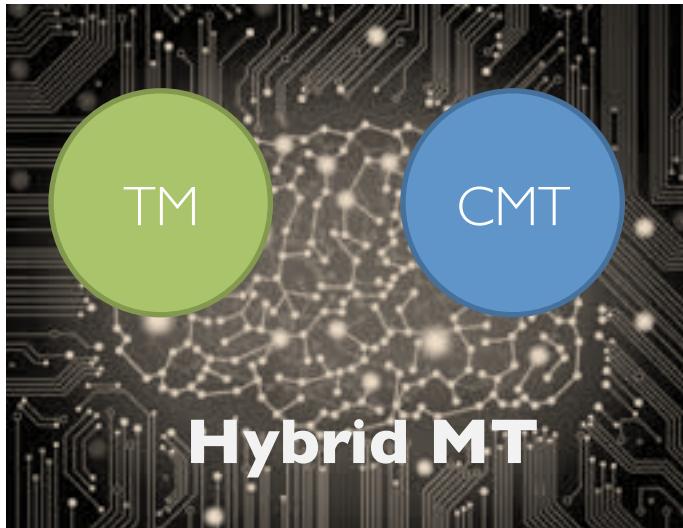
- First pass: use strict matching to translate known words and phrases



## Customized Machine Translation

- Second pass: translate the remaining text using machine translation system

# Aiming higher: Post Editing for Quality



Post Editors

- 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

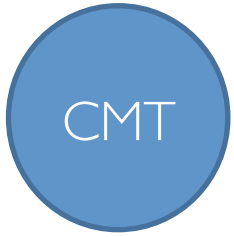
This ensures high quality but **does not guarantee less frustration for post-editors**



# 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)

Looking at:

- Effectiveness
- Quality
- Consistency

Also:

- Translator
- Hybrid MT + Post editing (PE)

# Data

- 1000 documents
  - 90k parallel sentences/segments
  - 953 documents for training
    - 489k tokens
  - Rest for tune and test

# Effectiveness of TM

Exact match

50%  
segments

BUT  
COVERS  
ONLY

7%  
words

Fuzzy match

84%  
segments

BUT  
COVERS  
ONLY

13.5%  
words

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

\* Based on an assessment over X documents

# Effectiveness of CMT

100% AND 99.9%  
segments words

translated!

# Effectiveness of Hybrid MT

- High precision
  - TM exact matches
- High recall
  - CMT to produce high quality translations



# Assessing Quality

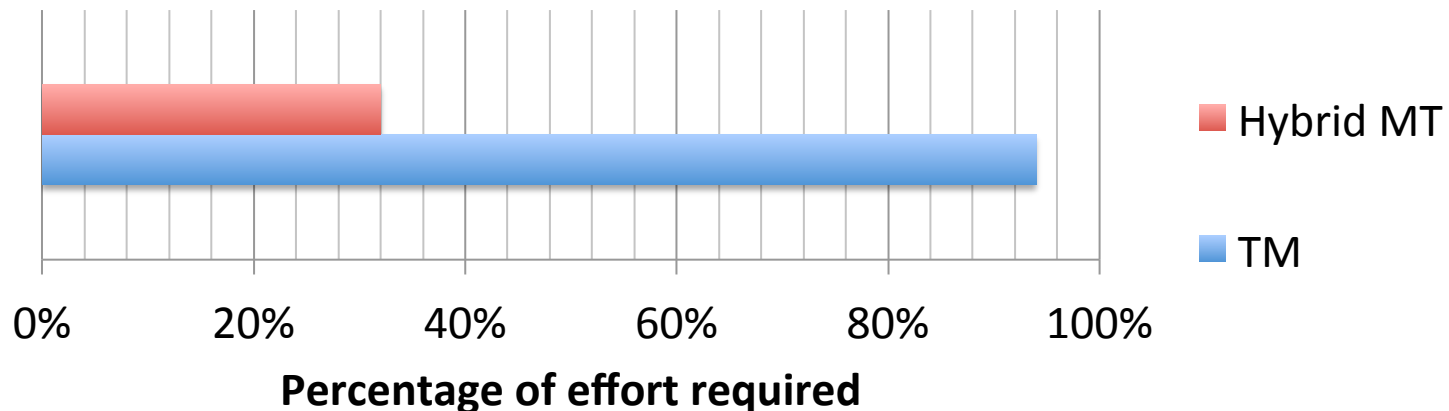
- BLEU
  - Compare output to ‘reference’ translation

	Strict	Partial
TM	7.07	21.01
TM + CMT	54.60	48.54

CMT alone BLEU scores are 53.90

# Assessing Quality

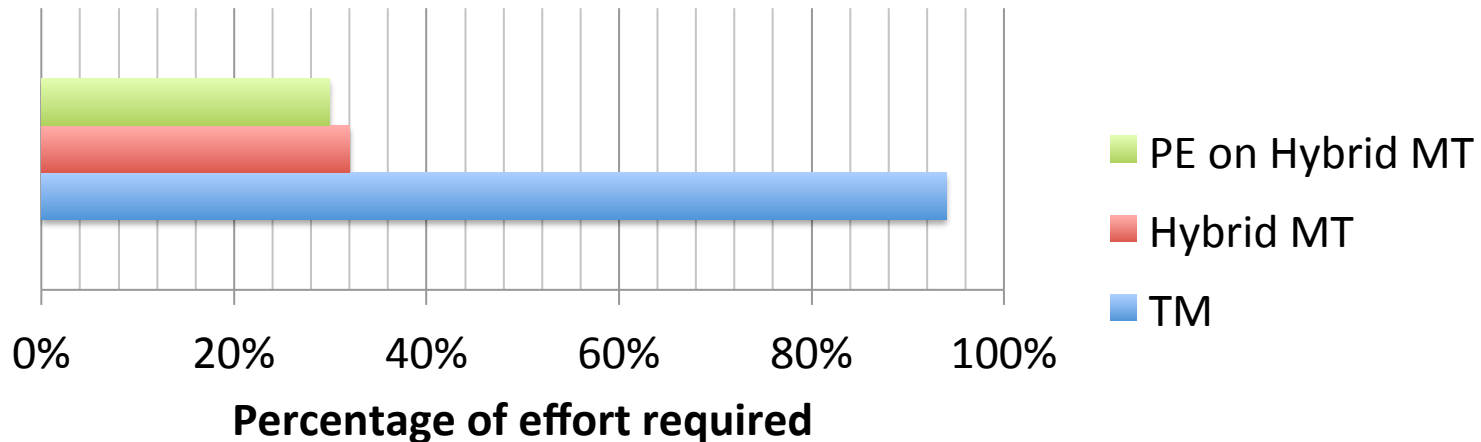
- TER: Translation Error Rate
  - How much effort is needed to get perfect translation?
  - Compare to 'reference' translation



Hybrid MT can improve beyond that!!!

# Assessing Quality

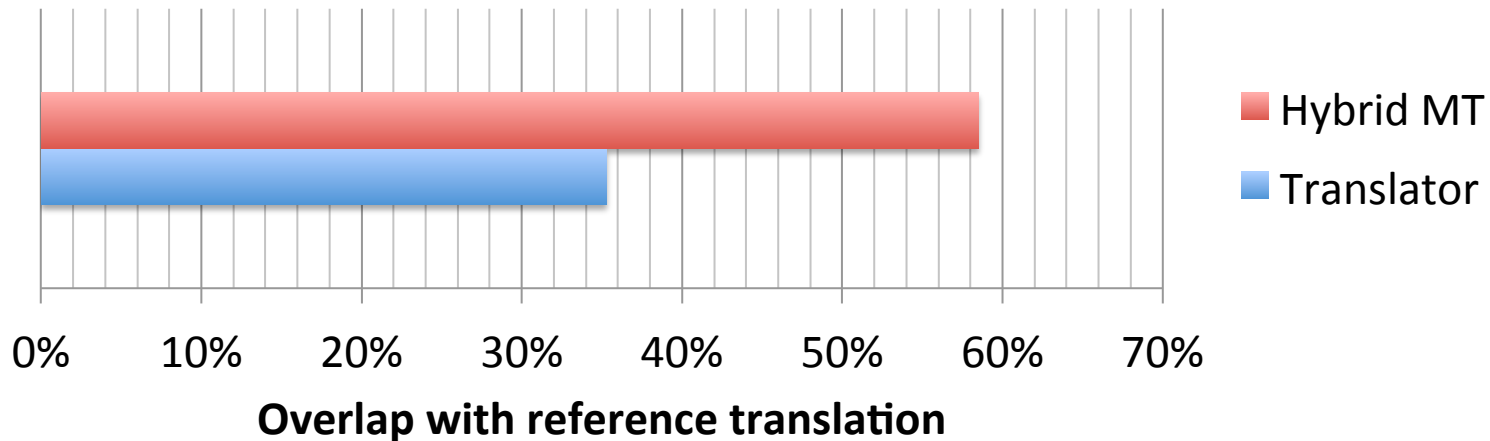
- TER vs. Post editing effort
  - Similar effort estimation using post-editing of Hybrid MT



\* 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

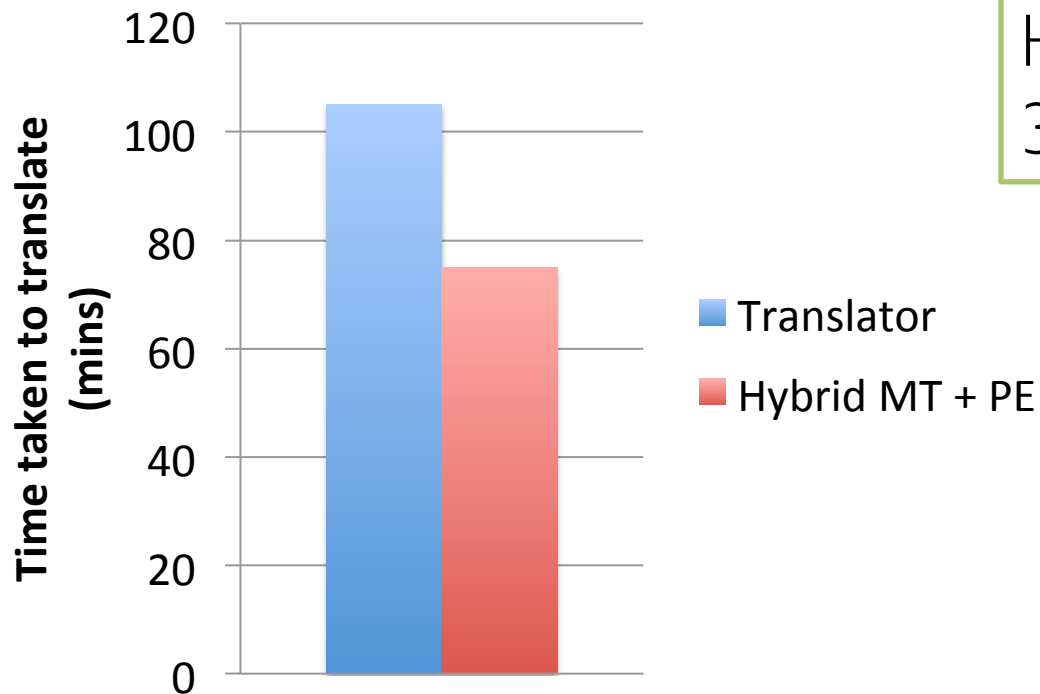


Hybrid MT is more consistent with reference translations

\* Based on an assessment over 4 documents

# Speedup of Hybrid MT

- We compared Hybrid MT versus a junior translator



Hybrid MT+PE is 30% more efficient

\* Based on an assessment over 4 documents

# 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

- Ahmed Abdelali, Kareem Darwish, Nadir Durrani, and Hamdy Mubarak. Farasa: A Fast and Furious Segmenter for Arabic. In NAACL-2016, San Diego, US.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. Moses: Open source toolkit for statistical machine translation. In ACL-2007, Prague, Czech Republic
- Hassan Sajjad, Francisco Guzman, Preslav Nakov, Ahmed Abdelali, Kenton Murray, Fahad Al Obaidli, and Stephan Vogel. QCRI at IWSLT 2013: Experiments in Arabic-English and English-Arabic Spoken Language Translation. In IWSLT-2013, Heidelberg, Germany

Thank you