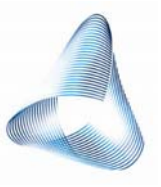
PROMT®
Automated Translation Solutions

Technology for Translators: What Doesn't Kill You, Makes You Stronger

Prepared for the AMTA 2009 Conference

Ottawa, Canada

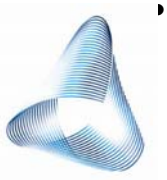
Jordi Carrera and Alex Yanishevsky, proMT



PROMT®
Automated Translation Solutions

Overview

- Background
- Methodology
- Case Studies
- Conclusion



PROMT®
Automated Translation Solutions

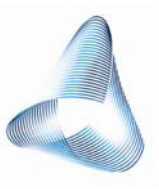
Background

•Initial resistance by translator community to translation memory technology based on fear and distrust

•Now translation memory technology is a valuable part of the translator's toolkit

•Machine translation is following a similar trajectory in the translation community

•Requires a paradigm shift on the part of translators



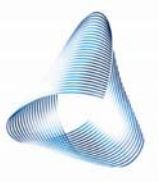
PROMT®
Automated Translation Solutions

Background – Part 2

•Cementing the paradigm shift requires solid proof on the part machine translation technology
- high level discussion of pros/cons of various algorithms for evaluating machine translation

-correlation between human evaluation of machine translated texts and algorithms for evaluating machine translation

-correlation between human evaluation, algorithm evaluation of machine translation and an increase in productivity on specific translation projects

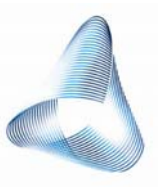


PROMT®
Automated Translation Solutions

Methodology

Tools for evaluating machine translation

- Algorithms for automatic MT evaluation – General features
- BLEU and NIST scores
- METEOR
- Experimental metrics
- The reality



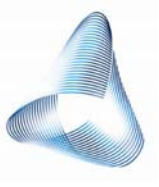
PROMT®
Automated Translation Solutions

Methodology

Algorithms for automatic MT evaluation – General features

Pros

- Perfectible
- Exact, objective and efficient
- Provide broad assessments, but strongly correlate with human judgments
- Human-level reliability when evaluating adequacy and collocational phenomena



Cons

•Calculated on the basis of strict literal similarity between text strings

•Unable to assess fluency, abstract syntactic structure, compositional phenomena or semantic correspondence

•Generally unable to evaluate informativeness, relevance or domain specificity

PROMT®
Automated Translation Solutions

Methodology – Part 2

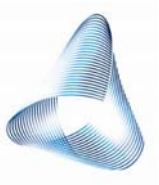
BLEU and NIST scores

Pros

- Computationally inexpensive, algorithmically simple
- Informativeness can be taken into account, but at the cost of vulnerability to noise (NIST)

Cons

- Unable to assess translation recall (only accuracy)
- Virtually unable to evaluate syntactic and most other long-range linguistic relationships
- Unable to correct for exponential decay following from considering increasingly infrequent types of data (NIST)
- Assuming standard error, scores are inversely proportional to the degree of syncretism in a language
- Search space can be expanded, but by introducing noise at the same time. Expansion also presupposes n-fold increases in resource availability



PROMT®
Automated Translation Solutions

Methodology – Part 3

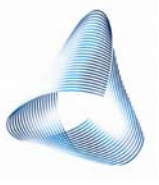
METEOR

Pros

- Able to assess translation recall
- Robust to lexical and semantic variation due to semantic modeling
- Better able to capture syntactic similarity
- Search space can be expanded in a well-founded way (no noise added)

Cons

- Virtually blind to morphological information
- Insensitive to some types of linguistic data due to replacement of high order n-gram models with chunks
- Scores for non-configurational languages are expected to fluctuate



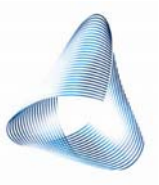
PROMT®
Automated Translation Solutions

Methodology – Part 4

Experimental metrics

Pros

- Non-mutually exclusive metrics (TER, ROUGE, ULC); can complement each other
- Provide valuable insights of how to improve MT evaluation algorithms and experimental background serves as a viability assessment



Cons

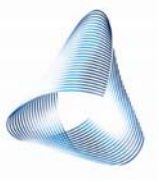
- Increasing sophistication more often than not yields no significant improvement in performance
- Naïve/knowledge-poor (TER, ROUGE); computationally expensive or unclear as far as cognitive modeling is concerned (ULC)

PROMT®
Automated Translation Solutions

Methodology – Part 5

The reality

- Algorithms of automatic MT evaluation constitute a reliable objective measure of performance and the only objective measure of performance
- Correlation with human judgments is strong, but still relates only to a subset of the relevant criteria of evaluation



- Currently, automatic algorithms tend to produce higher scores for statistical MT engines due to erroneously penalizing linguistic variation (lexical, morphological and syntactic) and by occasionally failing to penalize unstructured output

PROMT®
Automated Translation Solutions

And now what?

- How do we track progress?
- How reliable are the algorithms evaluations in relation to post-editing?
- How does this apply to actual localization projects?
- What are the financial and deadline implications?

PROMT®
Automated Translation Solutions

Tracking Progress

- Four different scenarios – ranging from basic to most complicated
  - Out of the box
  - Semi-trained
  - Fully trained without TMs
  - Fully trained with TMs
- Scores based on human (reference) translations

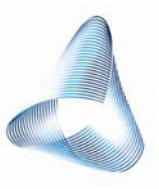
Four different scenarios – ranging from basic to most complicated

  - Out of the box
  - Semi-trained
  - Fully trained without TMs
  - Fully trained with TMs
- Correlation with amount of post-editing effort required

PROMT®
Automated Translation Solutions

Further Suggestions to Improve Quality

- Controlled writing of source
- Advanced leveraging for TMs
- Sub-sentential aligners



PROMT®
Automated Translation Solutions

Conclusion

- Increase in Bleu Plus /Meteor scores is 3-5 fold
- Increase in post-editing throughput is 2-3 fold
- Financial and deadline implications
  - Cost reduction
  - Faster time to market