BESST Translations
460 Pierce St
Monterey, CA

SMT Training Pilot Project Proposal

ATTN: Adam Wooten, Founder & CEO of Wooten Inc.

OBJECTIVE

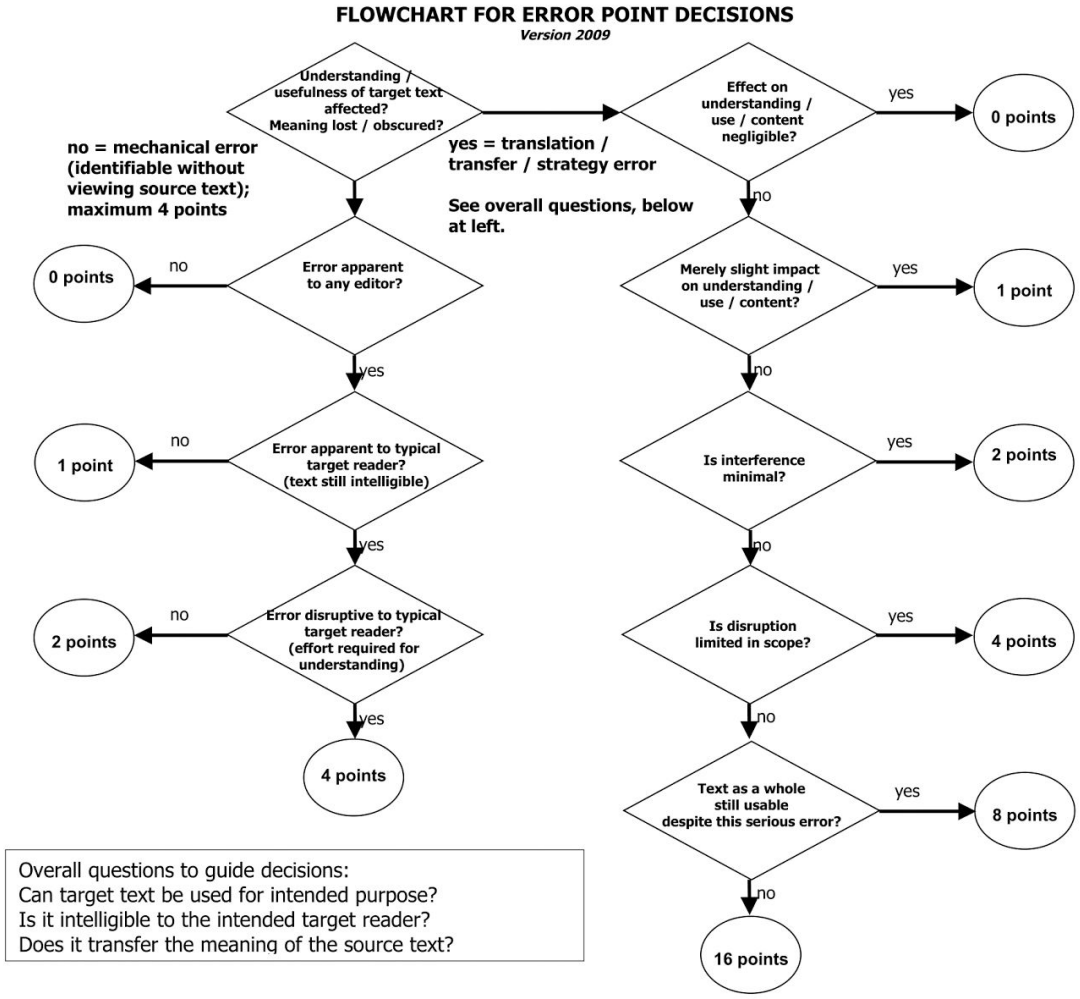
This pilot project will explore the training of an SMT to translate corporate financial statements of KDDI Corporation from Japanese to English. The translation needs to be both automated and of a good enough quality for human post editing.

- Quality goals: A machine translated financial statement that is not only readable but also does not require an excessive amount of human post-editing or QA. We will grade the quality of the MT output using the American Translators Association (ATA) certification error guidelines because of their clearly defined scoring metrics (see chart on next page). The MT output should have no 16-point errors, and overall no more than 20 error points total per 250 words of target text.
- Efficiency goals: Achieve a machine translation that, after human post editing, is significantly faster than an entire human TEP process. Specifically, it should be 70% faster than human translation.
- Pricing goals: Our machine translation should result in a significant price reduction in translation of Japanese financial statements to English. We expect a 70% reduction in cost compared to human translation.

PROCESS

1. Initial machine translation engine training carried out using a large training data set of Japanese financial statements of companies in similar fields, especially telecommunications companies such as Nippon Telegraph and Telephone Corporation and Softbank Group Corp., totaling approximately 100,000 segments, a tuning data set composed of past KDDI financial reports totaling approximately 2,000 segments, and then our testing data consisting of the most recent KDDI financial statements totaling approximately 2,000 segments.
2. Receive an initial BLEU score for the data set described above.

3. Tweak the engine by adding additional related bi-texts to the training data and/or tuning data, or by reevaluating or re-aligning previously used data.
4. Retest the engine and record the resulting BLEU score. If the BLEU score has increased, changes will be kept. If the BLEU score has decreased, changes will be discarded. Submit comments for recommended changes to be applied during the next round of testing.
5. Repeat steps 3-4, making sure to always use the engine that had the highest resulting BLEU score as the base for the next set of tweaks and enhancements.
6. After completing the first round of training for the pilot period, our QA manager will evaluate the quality of the MT output and measure it against our stated goals using the flowchart pictured below.



ATA error guidelines, taken from http://www.atanet.org/certification/aboutexams_flowchart.pdf

TIMELINE AND COSTS

Pilot timeline

We will launch the three-week pilot project following the kickoff meeting on March 30, 2016, pending client approval. On March 31, we will construct an initial translation engine and perform an initial test. During the three-week period terminating on April 19, we will perform regular training and tests as per the above mentioned processes. We will perform a minimum of ten such rounds of training. At the end of our testing period, we will finalize the changes to the translation engine and compare the final BLEU score to the initially recorded BLEU score and perform a QA check to assess our progress.

Pilot costs

Task	Estimated Hours	Hourly Rate	Cost
Document alignment	15	$40.00	$600
MT training	5	$30.00	$150
Post-editing	1	$30.00	$30
QA	1	$30.00	$30
Total estimated pilot cost			$810

DELIVERABLES

Upon completion of this pilot project, BESST Translations will submit a proposal for a full-scale project for training an MT engine to translate KDDI financial statements from Japanese to English based on our findings. This updated proposal will include an overview of the pilot process, with a detailed examination of the training methods and the results of such, as well as a reassessment of our quality, efficiency, and pricing goals, along with our recommendations on how to move forward with this project.