

## **Machine Training Pilot Project Proposal**

Ksenia Sokolova, Amelia Wolford,  
John Nunes, and Hyerim Ko

### **I. Pilot Project Objectives**

This pilot project is intended to train Microsoft Translator Hub's statistical machine translation (SMT) engine to translate medical equipment user manuals from English to Russian. The resulting post-edited machine translations (PEMT) will achieve the following efficiency and cost goals:

- Productivity: no less than 25% faster than human translation
- Cost: no less than 25% cheaper than human translation

Additionally, PEMT quality will be evaluated according to the Translation Automation User Society (TAUS) machine translation quality guidelines, wherein PEMT is “comprehensible (i.e. an end user perfectly understands the content of the message), accurate (i.e. it communicates the same meaning as the source text), [and] stylistically fine, though the style may not be as good as that achieved by a native-speaker human translator. Syntax is normal, grammar and punctuation are correct.” A human QA reviewer will ensure that PEMT meet these guidelines and the following additional requirements:

- Semantically correct translation
- No information accidentally added or omitted
- No offensive, inappropriate or culturally unacceptable content
- As much of the raw MT output as possible used
- Basic rules regarding spelling observed
- No corrections that are of a stylistic nature only
- No restructuring of sentences solely to improve the natural flow of the text
- Key terminology is correctly translated and untranslated terms belong to the client’s list of “Do Not Translate” terms

*Adapted from TAUS MT Post-editing Guidelines:*

<https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines>

The cost and time required to complete this two-week pilot project as well as the productivity gain and achieved quality levels will be used in order to determine the work and cost requirements of fully training an MT engine.

### **II. Process**

To begin building a translation system for medical equipment user manuals, engineers will work with five PDF bilingual pairs of user manuals comprised of approximately 10,000 segments each. Two additional documents of about 1500 segments per document will be used for tuning and testing. Engineers will implement ten rounds of iterated ME testing to obtain the best possible BLEU score.

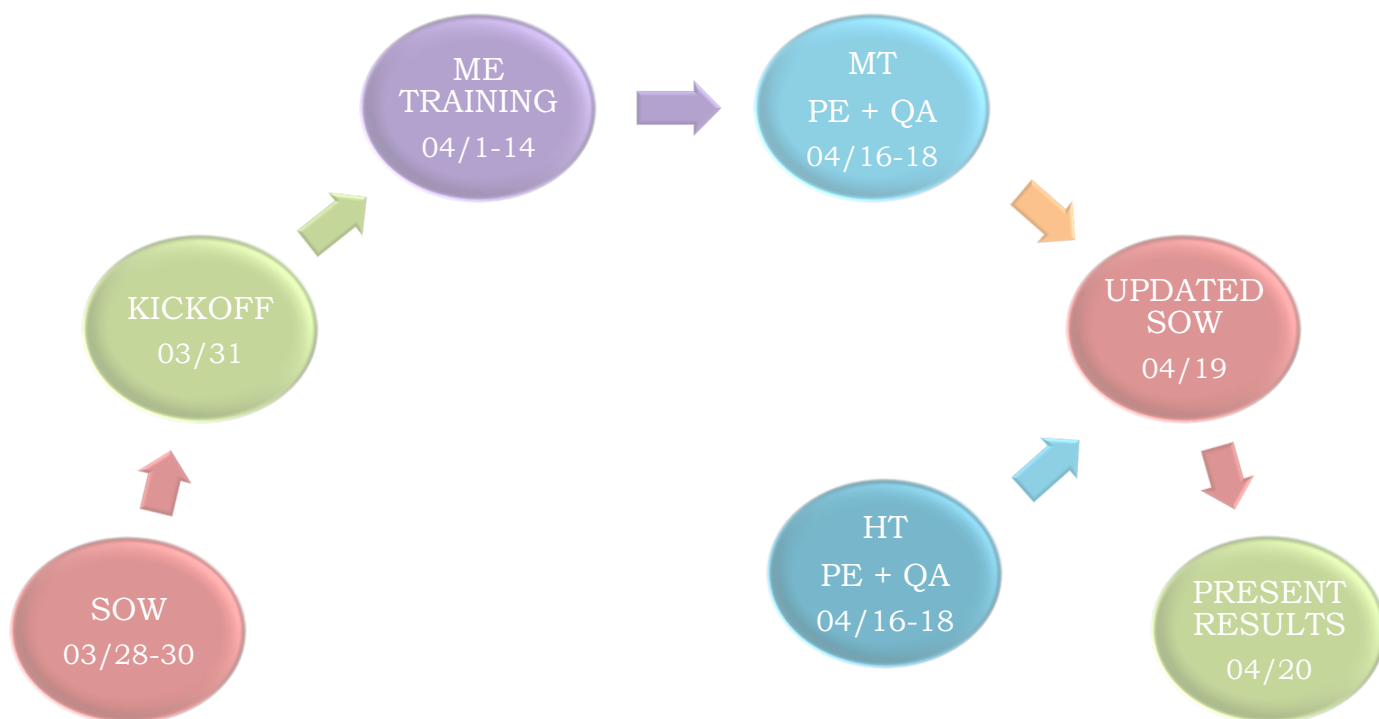
Different strategies will be attempted for each successive round of testing to improve the BLEU score. Such strategies may include converting PDF to DOCX file format, adding a dictionary, switching from auto to manual tuning and switching from auto to manual alignment (extracting text from both languages into a TXT file and uploading with a “.align” extension).

A BLEU score will be recorded for every round; an increase over the previous round's BLEU score will be accepted as a valid engine for further training, while a decrease will be rejected and new strategies will be tried. This process will be repeated until a total of 10 rounds has been reached (including rejected rounds).

Once ten rounds of testing are completed, a trusted human translator (someone who has experience translating similar material and does not have a bias about PEMT) will be hired to first translate a similar test file during 2 hours. Then the human translator will be asked to post-edit the machine-translated text for 2 hours. The productivity gains of MT with human PE versus the HT only will be measured. After the productivity gains are measured, the quality will be assessed according to the TAUS machine translation guidelines. The standard translation rate for HT from English into Russian is \$0.12/word, the PEMT cost being \$.09/word. The assessed comparison will show whether the PEMT has achieved the goals of no less than 25% productivity gain and no less than 25% cost savings.

### III. Timeline and Costs

Following the kickoff meeting on Thursday, March 31, the pilot project will be carried out over a period of two weeks. Starting Friday April 1, one round of training will be completed per weekday, with the final rounds being completed on Thursday, April 14. Human Translation and Post-Editing (PEHT) and Machine Translation and Post-Editing (PEMT), as well as QA will be completed by Monday, April 18. Data collected from the pilot project engine training, HTP/MTPE and QA will be used to calculate time/cost productivity gains of PEMT vs PEHT and to create an updated project proposal to be presented on Wednesday, April 20.



The data collected from multiple sources prior to the start of the pilot project allows us to believe that we can achieve an increase in PE processing speed (that is the processing time in relation to the words processed in that time, which is words divided by time) by approximately 25%.

One of the critical tasks that should take place early in a project involving PEMT is quality assessment, which directly affects the PE costs. The QA can be done with the use of quality metrics, which all as we believe have the same issue - they do not measure the complexity of the correction. Therefore we have found out that the metric that really matters is productivity - or more explicitly - how long does it take a human post editor to correct the MT output to the same level of quality as the human only approach to achieve a “good enough” quality level.

We expect the PE productivity gain (that is the relationship between the number of words per minute done per one translator without any aid and the number of words per minute done by the same translator with the aid of a MT engine) to grow by no less than 25%.

Finally, in relation to the aforementioned, the productivity gain of no less than 25% realized via PEMT allows for establishing the rates and estimating the costs for the pilot project as follows:

TASK	EST. TIME	QUANTITY	EST. RATE	SUBTOTAL
ROUND OF ME TRAINING	0.5	10	\$25	\$125
DATA CONVERSION	0.5	10	\$15	\$75
FILE REFINEMENT	1	5	\$25	\$125
DICTIONARY CREATION	1	1	\$25	\$25
HT		\$0.12/word		
PEHT				
MT				
PEMT		\$0.09/word		
<b>TOTAL</b>				<b>\$350+</b>

#### IV. Deliverables

Upon completion of the pilot project the team will provide deliverables including (but not limited to) an updated proposal for a full-scale SMT training and a compilation of data from the pilot training project such as documents used, techniques used, and BLEU scores for each round of training. This data, along with additional data from post-editing and QA will help determine if the goals listed in section one of the proposal are feasible as well as determine the estimated time and costs required to achieve these goals. The updated proposal will include a list of recommendations for further training of the MT engine, as well as a strategy for using the engine with a translation management system.