

Kot Translations

460 Pierce Street,
Monterey, CA, 939400

Pilot Project Proposal

For: U.S. Department of State, Embassy of the United States, Moscow
Bolshoy Devyatinsky Lane 8, Moscow, Russia, 121099

Pilot Project Objectives

Train a SMT engine to translate press release documents for U.S. Embassy in Moscow website

Goals

Efficiency	Cost	Quality
20% efficiency	20% savings	High enough quality to not cause an international incident--below 10 points with no critical errors.

QA Metric

	Minor	Major	Critical
Language	1	5	10
Accuracy	1	5	10
Style	1	5	10
Terminology	1	5	10

Adapted from *TAUS Harmonized DQF-MQM Error Typology*:

<https://www.taus.net/evaluate/qt21-project?fbclid=IwAR3rDS82ah01Ba7KT3ysU7BKnpn4PvhAU5Iqhurd4SzsqsEyjYsAFnLwL0Q#harmonized-error-typology>



Kot Translations

Data Sets

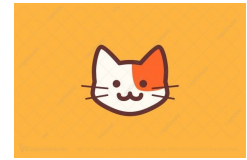
Training Data	Tuning data	Testing data
Bitexts of various kinds of texts filed under “press releases” on the U.S. Embassy’s website	Bitexts of one (at least) of each kind of text filed under “press releases” on the U.S. Embassy in Moscow’s website	A compilation bitext file with all the different kinds of documents filed under “press releases” on the U.S. Embassy in Moscow’s website

Timeline



Process:

File Prep	Translations already exist for all the press releases currently on the embassy’s website. The bitext for the testing file has already been created; bitexts are needed for the training and tuning. We will create documents in Microsoft Word for the texts. Then we will use memoQ to align the tuning files and we will have Microsoft Custom Translator align the docx files for training automatically, at least initially.
Training	BLEU score will be evaluated each round. Tweaks will be made to the training and tuning rounds: additional texts, aligning some of the training data.
Assessment	Assessments will be made by Russian lead linguist, who will also do the post-editing. The metric above will be used to assess quality. Post-editing time will be compared to average speed of translation per page--30 minutes for a high-quality translation. Cost savings will be determined using the metric of \$45/hr.
Proposal Update	Proposal will be updated using the results from these initial training rounds.



Kot Translations

Costs:

Task	Estimated Hours	Rate	Subtotal
Document alignment	15	\$45.00	\$675.00
MT training	15		\$675.00
Glossary creation	10		\$450.00
Post-editing	4		\$180.00
QA	4		\$180.00
			Total: \$2160.00

Deliverables:

- Final updated proposal for SMT engine training to translate press release documents for U.S. Embassy in Moscow.
- Corrected estimates of time and cost required to achieve project goals and execute training. Plans for further improvement of the engine.
- Records of results of engine training and BLEU score evaluation.
- Lessons learned and recommendations on future training for the SMT engine.