PORTUÑOL TRANSLATIONS

460 PIERCE STREET
MONTEREY, CA 93940

Team Members: Nathalia Rio Preto, Isabel González-Gutiérrez, and Fiona Maloney-McCrystle

UPDATED PROJECT PROPOSAL

PREPARED ON 3/25/20 FOR:
Professor Adam Wooten
460 Pierce St.
Monterey, CA 93940

INTRODUCTION:

This project was originally intended to estimate the amount of time and costs required to train a statistical machine translation engine designed to translate ocean-related conservation reports. After our client meeting, at which we discussed how we needed to adjust the way we planned to use the data we had found, that intention shifted to estimating the time and costs required to train a statistical machine translation engine designed to translate UN reports about the Sustainable Development Goals.

PROCESS:

The main components of our process were:
-Data research
-Data alignment
-Troubleshooting Microsoft Custom Translator
-Data cleaningMore data research
-More data cleaning
-File conversions
-Data substitution
-Switching data between Tuning and Training
-Glossary creation for official translations of committees, publications, etc

METRICS RESULTS:

Metric	Efficiency	Cost	Quality
Goal	PEMT 35% faster than HT	PEMT 35% cheaper than HT	Translation does not reach 10 error points per 500 words
Result	PEMT 88% faster than HT	PEMT 88% cheaper than HT	Translation failed: for 1000 words, reached 50 points (> 20)

- Note: HT estimates were based on our research suggesting that translators are reasonably expected to translate 250 high quality words/hour (2,000/day)

ESTIMATED COSTS VS. ACTUAL COSTS

Activity	Estimated Hours	Actual Hours	Hourly Rate	Subtotal	Actual Total
Data Identification, File Conversion, and Alignment	15	30 *includes troubleshooting alignment/ platform upload issues	$40.00	$600.00	$1,200.00
Engine Training Rounds and Data Adjustment	30	15	$40.00	$1200.00	$600.00
Post-Editing	6	4	$40.00	$240.00	$160.00
Quality Assessment and Comparison	3	3	$40.00	$120.00	$120.00
Proposal Update	3	6	$40.00	$120.00	$240.00
			Total	$2,280.00	$2,320.00

UPDATED GOALS AND METRICS

Based on the above results, we have set the following updated goals:

-Quality: We wish to keep striving for the same quality metric, as we did not achieve it during the trial, but think with more training it is still a reasonable goal (see QA metrics outlined in original proposal)

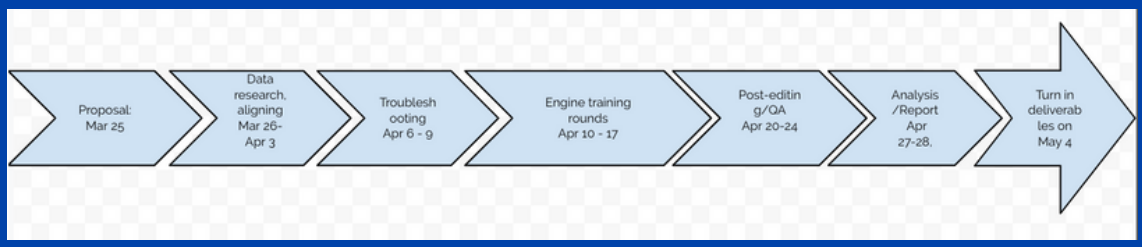
-Efficiency: Our goal would be to gain 5% more efficiency than we did in the pilot phase, making the goal that PEMT be 93% faster than HT

-Cost: Our goal would be to bring costs down 5% more than we did in the pilot project, making the goal that PEMT be 93% cheaper than HT

UPDATED PROJECT DETAILS

We estimate the full completion of the project will take 3x as long as the pilot stage did, coming in at 174 hours, with the same hour allocations as above with one exception: there should be a re-categorization so that troubleshooting has its own category to more accurately reflect where time was spent. At the estimated 174 hours, the full cost would be $6,960.00 to train the engine.

The timeline would look something like this:



UPDATED DELIVERABLES

At the end of the project, will include:

-Report similar to this one updating regarding process and results of the project and analyzing outcome

-Information about data sets used for training, tuning, and testing

-BLEU scores obtained for each training cycle

-Log of changes made to data after each training cycle

-Updated glossary for specific domain, if created

GENERAL TAKEAWAYS/RECOMMENDATIONS

-Seems possible that engine, with more training, could be used by a UN translating division

-Because of high quality needs, would still have to be carefully post-edited

-Cost-benefit analysis:

-Estimated total project timeline and costs: 3x what pilot phase took, totalling at $6,960

-Would pay for itself after 174 hours of post editing (at our rate), or ~22 full days of work

-Seems worthwhile at this cost/timeline, as the system could be used for far longer than one month, all the while making efficiency of translating division at least 88% better

-What we would do next:

-Our first steps to continue working on the training of the system would be:

-Fix capitalization and article issues created by addition of glossary

-Find out exactly where in the data persistent grammar issues were coming from and clean out incorrect segments

FOLLOW UP ON OTHER DELIVERABLES FROM PILOT PROPOSAL

-Information about data sets used for training, tuning, testing:

-Testing: 2016 UN Economic and Social Council Report on “Progress Towards the Sustainable Development Goals”

-Training: UN and UN-organization-specific content: reports on progress about certain SDGs, UN Environmental Program documents, reports about SDG implementation and indicators, etc.

-Tuning: 2017 and 2018 UN Economic and Social Council Reports on “Progress Towards the Sustainable Development Goals”

-BLEU scores for each cycle/changes made to after each cycle

-Please see accompanying “Portuñol Log of Training Rounds” spreadsheet for detailed information on BLEU scores

-Glossary, if used

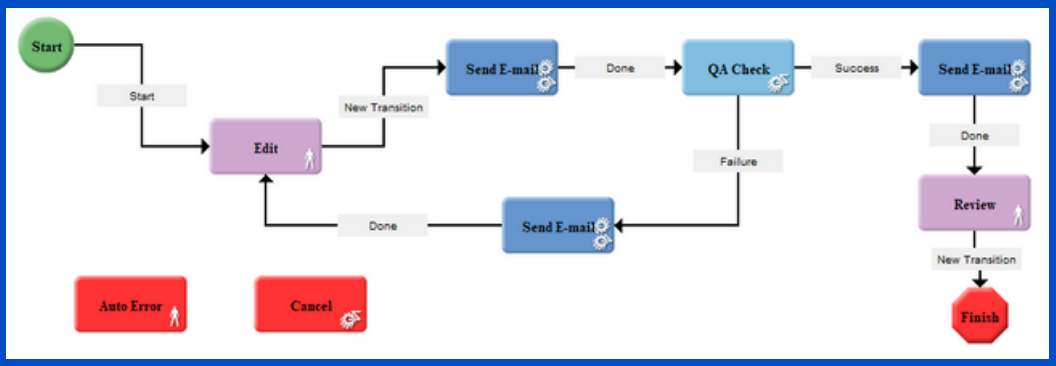
-Please see accompanying “PortuñolGlossary4” document

RECOMMENDED CAT TOOL SETTINGS

We would recommend the following CAT tool settings (list made with Trados in mind) to assist with PEMT or QA of post-edited content:

-Turn on the automatic QA check setting that produces an error when target and source segments are the same, as "untranslated" produces a critical error in our framework, and the post-editor would clearly know translation had failed if they saw that error come up
-Turn on automatic QA check setting that compares length of target and source segments, setting it to produce an error if target is shorter than source by a certain amount. Because our project translates from English into Spanish, the Spanish segments should be up to 30% longer than the English sentences, and it's likely that if they are shorter there would be an omission (one of our error types) that the post-editor would need to mark.
-Turn on the automatic QA check setting specific to numbers for the English-Spanish combination, because if a number is not properly rendered into Spanish, it likely means that it was "untranslated," which, as above, is a critical error in our error severities framework.

RECOMMENDED WORKFLOW



We would recommend using a workflow in WorldServer something like this for the latter stages of our project. The changes we made to this workflow (based on the example) were:

-This workflow does not include translation, as the machine would do that. Instead it starts with the process of post-editing (in which the editor, to whom we assigned that step, would work with the output from the MT engine) and QA checks
-We included "Send Email" steps to automate the process as much as possible.
-We included a final Review step (assigned to a Reviewer) to ensure the utmost quality.