Pilot Project Proposal

FOR:
Wizards of the Coast LLC
P.O. Box 707
Renton, WA 98057

BY:
DnDorks (MIIS)
Dean, Christopher
Guedez, Marianna
Shin, Sook
Soza, Gary
Sublet, Eric

Pilot Project Objectives

Following the news that Wizards of the Coast LLC partnered with Gale Force Nine in localizing the Dungeons & Dragons fifth edition tabletop RPG content into multiple languages, the DnDorks team is proposing the following pilot project. This pilot project aims to determine the effort required in the training of a Neural Machine Translation (NMT) engine intended for use in the translating of content involved in the role-playing game (RPG), **Dungeons & Dragons**, with a particular focus on both official and fan-base campaign materials (English to Spanish). This project's goal is for post-edited machine translations (PEMT) made with this engine to meet the following target criteria to achieve efficiency, cost, and quality.

- Efficiency: PEMT approximately 30% faster than human translation
- Cost: PEMT approximately 40% savings over human translation
- Quality: PEMT with an acceptable score based on the following model:

HUMAN EVALUATION SCORECARD
ADAPTED FROM DQF METRIC

ACCURACY

FLUENCY

VERITY

ERROR TYPES (ACCURACY)
____ Mistranslation
____ Misspelling
____ Omission
____ Punctuation error
____ Grammatical error
____ Formatting(dates,etc)

ERROR TYPES (FLUENCY)
____ Terminology / Polysemy
____ Syntactic Error
____ D&D specific reference

HIT POINTS

Threshold: **25** hit points /500 words

COUNTERSPELL
+1
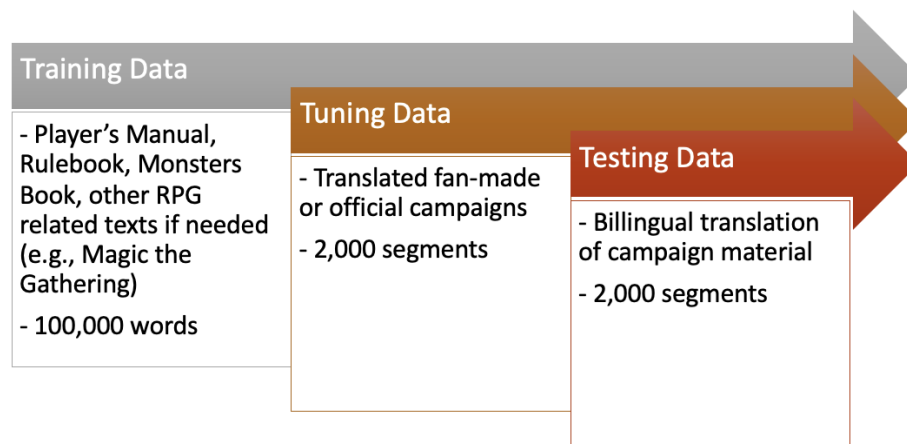
MINOR
-1

MAJOR
-5

CRITICAL
-10

An evaluation of the MT translation will be carried out by a human reviewer using a set of criteria (adapted from DQF model) for error types and severity, performed towards the beginning and end of training of the engine. (The format has been adapted to mimic scorecards typically used in Dungeons and Dragons, in order to motivate and retain human evaluators, who likely have experience with the game.) For each error, the engine will lose *hit-points (HP)*, based on error type and severity. If the HP of the engine falls below 0, the round will be classified as **fail**.

To determine the work involved in bringing the MT to this benchmark, we will train a preliminary engine over the course of two to three weeks. We will estimate the requirements of a completed MT training project based on the costs and work involved in training the machine engine and by assessing the rate of improvement and the final quality.
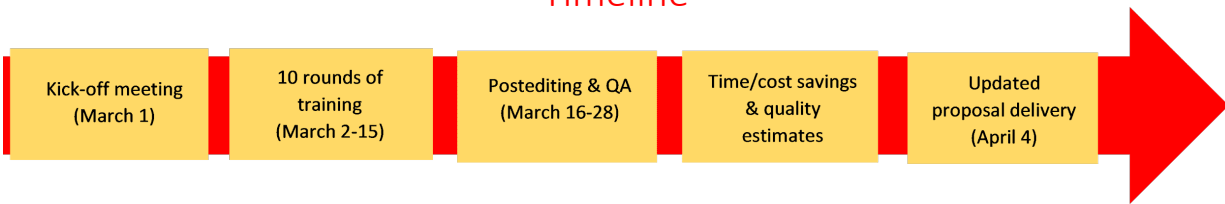
Data Sets

Training Data

- Player's Manual, Rulebook, Monsters Book, other RPG related texts if needed (e.g., Magic the Gathering)

- 100,000 words

Tuning Data

- Translated fan-made or official campaigns

- 2,000 segments

Testing Data

- Billingual translation of campaign material

- 2,000 segments

Process

- **File preparation:** Conversion of data sets to DOC/TEXT format
  - Relevant PDFs will be downloaded online and then converted into Docx using Adobe Acrobat Pro DC
  - Files will be aligned into TMX using SDL Trados, and then cleaned in Olifant to create acceptable segments for training
  - Identify terms to replace with a unique token ID
- **Training rounds:** MT training with Microsoft Translator Hub
  - BLEU score will be evaluated after each round of training
  - Strategies will be implemented and adjusted based on findings of each testing round with the aim of improving BLEU score (if decrease in BLEU score, each subsequent round will be based on the most recent successful model). Including, but not limited to:
    - Addition of new training documents
    - Addition/modification of termbases
    - More detailed cleanup of training/tuning segments
- **Assessment:** Estimation of PEMT time, cost, and quality based on results of training rounds
  - Machine-translated documents will undergo at least two rounds of timed post-editing by human evaluators
    - Post editors aim for grammatically, syntactically, and semantically correct translation. They won't be asked to make stylistic or preferential changes.
  - The time required for post-editing time early and late in the training process will be compared
  - Cost savings will be approximated based on post-editing time, with a standard of $30.00/hr rate
  - A customized form of the DQF quality metrics will be used to determine if the translations are of acceptable quality
- **Proposal update:** Proposal estimates will be reexamined based on the assessment

Timeline

| Kick-off meeting (March 1) | 10 rounds of training (March 2-15) | Postediting & QA (March 16-28) | Time/cost savings & quality estimates | Updated proposal delivery (April 4) |

Costs

The following table details the estimated time and costs for this pilot project.

Task	Estimated Hours	Rate	Subtotal
Document alignment	1	$15.00	$15.00
Document acquisition/cleanup	20	$35.00	$700.00
Glossary Creation	2	$25.00	$50.00
Post-editing (per 2,000 words)	3	$30.00	$90.00
QA	2	$20.00	$40.00
Microsoft Translator Cost			$249.86
Post Mortem	4	15.00	$60.00
PM	N/A	15% of the total	$120.48
		Total	$1,325.34

Deliverables

The following will be delivered to the client upon completion of the pilot project.
- Updated proposal for training an MT engine to translate Dungeons and Dragons campaigns
- Details of the data used for pilot project training (training, tuning, and testing data sets)
- Records of iterations made for training the engine and BLEU score for each round
- Estimates of the feasibility, time, and cost required to achieve the stated goals
- Suggestions for ways to continue training the MT engine

Terms & Agreements

Change of scope will mean more hours of work that will have to be measured after the Client submits a Change Request, this will go into further consideration to see if is possible for the team to make the requested changes and, assessing the timeline impact.