# STATEMENT OF WORK

Team Members: Min Chua, Yui Ning Kam, Juang-Lung Lin, Ren Jing, Zhenzhu Zhou

**Prepared for:**
Adam Wooten
Tencent Media
2747 Park Boulevard
Palo Alto, CA 94306

**Quote No.:** 688888-2
**Updated on:** April 10, 2022

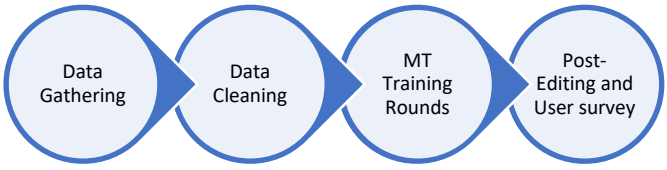
## PROJECT OVERVIEW

From March 7, 2022 to April 2, 2022, our company performed a pilot Chinese-English machine translation (MT) engine customization project on a modern suspense drama *Reset* (“开端”) produced by Tencent. Due to a global streaming boom where demand for translations outstrips supply, Tencent is seeking a partnership with our company to test the feasibility of translating subtitles of some lower-resource productions using machine translation in the future.

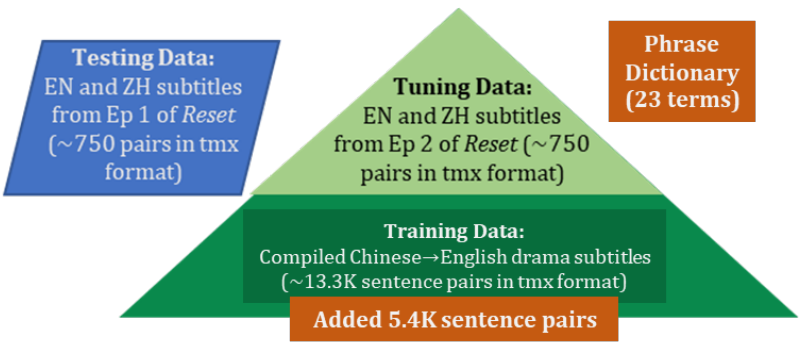
This updated proposal will summarize the results of the pilot project, as well as put forth our recommendation on proceeding with a full-scale MT engine customization project.

## PILOT PROJECT RESULTS

MT Engine Training Process




Chosen Datasets



MT Iterative Training & BLEU Scores

| # | Training Data | Tuning | Testing | Microsoft | SYSTRAN |
|---|---|---|---|---|---|
| 1 | OpenSubtitles Corpus | Reset Ep 2 | Reset Ep 1 | **9.14** ***(chosen for 1st PE + user survey)*** | 12.99 |
| 2 | R1 Training Data **+ data-cleaning** | Reset Ep 2 | Reset Ep 1 | 8.99 | 12.69 |
| 3 | News Commentary Corpus | Reset Ep 2 | Reset Ep 1 | 8.5 | 15.3 |
| 4 | ZH-to-EN Drama Subtitles | Reset Ep 2 | Reset Ep 1 | 13.76 | 21.23 |
| 5 | R4 Training Data **+ data-cleaning** | Reset Ep 2 | Reset Ep 1 | 14.31 | **22.61** ***(highest SYSTRAN – chosen for 2nd PE + user survey)*** |
| 6 | R5 Training Data **+ 5.4K new pairs** | Reset Ep 2 | Reset Ep 1 | **14.41** **(highest Microsoft)** | 21.68 |
| 7 | R6 Training Data **+ Phrase Dictionary** | Reset Ep 2 | Reset Ep 1 | 13.83 | 21.24 |

Survey Results

| | Lowest | Average | Highest |
|---|---|---|---|
| **1st Round PEMT (Round 1 output)** | 6 | 8.1 | 9.5 |
| **2nd Round PEMT (Round 5 output)** | 9 | 9.2 | 10 |

Pilot Project Costing

| Tasks | Estimated Hours | Hourly Rate | Cost |
|---|---|---|---|
| Data Collection | 8.2 | $45 | $367.50 |
| Data Preparation (cleaning, alignment) | 16.5 | $45 | $743.25 |
| MT Training | 6.2 | $45 | $277.50 |
| Human Evaluation | 3.9 | $45 | $173.25 |
| Meeting & Discussion | 30.9 | $45 | $1391.25 |
| Post-Editing | 3.0 | $40 | $120.00 |
| **Total** | **68.6** | | **$3072.75** |

Pilot Project Objectives & Recommendation

- **Quality Goal:** Met quality goal of user survey score higher than 7/10.
- **Pricing Goal:** Since the variable cost of post-editing ($120 per 40-minute episode) is lower than a human translator's rate ($160 per episode), this will allow a full-scale project to be viable as long as the volume of projects can cover the fixed costs. Thus, our pricing goal is also met.
- **Time Goal:** Comparing average post-editing speed of MT output (3 hours per 40-min episode) versus an average human translator's working speed (4 hours per 40-min episode), the fast PEMT speed means that our time goal is also met.
- **As such, with all three goals for the pilot project met, we recommend that Tencent can proceed with the full-scale project.**

---

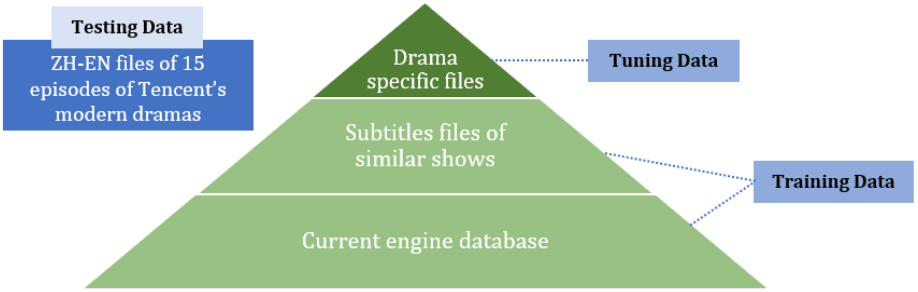
## PROPOSED FULL-SCALE PROJECT

---

### PROJECT OBJECTIVES

- **Quality Goal**
  - In our sets of user surveys, achieve a final average score of 8/10 from all three sections of criteria: Accuracy, Language, and Style.
- **Time Goal**
  - Post-editing of MT (PEMT) output 30% faster than human translator's speed
- **Pricing Goal**
  - Post-editing for MT 25% more affordable than HT, where cost of entire MT project lower than HT when applied to programs with more than 229 episodes (see Break-Even Analysis on page 4).

### DATASETS




### METRICS OF SUCCESS

| | |
|---|---|
| **Quality** | Human Evaluation of Quality through user survey – inviting users to watch a sample of the episode with post-edited MT output subtitles<br>• Did you understand the general storyline of the drama? (Yes/No)<br>• How much of the dialogue did you understand? (Rate 1-10) |
| **Time** | Compare time required for PEMT compared to human translator's working speed (4 hours / 40-minute episode) |
| **Pricing** | Compare total cost required for MT engine + post-editing compared to human translator's rate ($4 / video minute) |

## RECOMMENDED ADDITIONAL TRAINING FOR NMT ENGINE

Based on the pilot project results, the following additional training actions are recommended for the full-scale customization of the NMT engine:

1. Expand the testing dataset to include 5 episodes each of a drama from the top 3 genres of Tencent's modern drama launches: Romance, Mystery/Suspense and Family dramas, for a total of 15 episodes.

2. Add more documents to the training dataset that is proportionate to the increased volume of testing dataset. However, adding more documents does not necessarily mean that the BLEU score would be increased as well, as we didn't get a higher score for R6 in Microsoft even though we added more data to the training dataset. More flexible adjustments are needed in this process.

   At this point, the additional training data should be pegged at the same volume of 10,000 segments of the pilot project, for each of the 3 genres. This equals to a total of 30,000 training segments.

3. More domain-specific data-cleaning would be needed in the data preparation stage to ensure the newly added data would not lower the quality of the dataset.

## COSTING – BREAK-EVEN ANALYSIS

Full-Scale Project Cost Estimate

| Tasks | Estimated Hours | Hourly Rate | Cost |
|---|---|---|---|
| Data Collection | 2 | $45 | $90.00 |
| Data Preparation (cleaning, alignment) | 45 | $45 | $2,025.00 |
| MT Training | 8 | $45 | $360.00 |
| Human Evaluation | 10 | $45 | $450.00 |
| Meeting & Discussion | 25 | $45 | $1,350.00 |
| Post-Editing of MT output | 45 | $40 | $1,800.00 |
| **Total** | **140** | | **$6,075.00** |

- **PEMT speed:**
  - The post-editing speed measured in the pilot project was 15 minutes to post-edit the MT output subtitles of a 3-minute video clip from Episode 1.
  - Therefore, for a full episode (40 minutes), the post-editing time would be around 3 hours.
  - Cost of post-editing for one episode: Post-editor's hourly rate ($40) * 3 = $120
  - Base Rate for PEMT = $120/40mins = **$3/minute video**

| | Base Rate (per minute video) | Rate per episode (as per 40 mins episode) |
|---|---|---|
| **HT** | $4.00 | $160 |
| **PEMT** | $3.00 | $120 |

- The variable cost difference between HT and PEMT per 40 mins episode = $160 - $120 = $40
- The total fixed cost of training the full MT engine = $3,072.75 + $6,075 = $9,147.75

- **Conclusion:**
  - Break Even Point (the volume of projects that can cover the fixed investment of training the MT engine) = $9,147/$40 ≈ 229 Episodes
  - A full-scale project would be viable as long as the engine will be used to translate more than 229 episodes, or approximately 8-12 drama seasons.

**RECOMMENDED CHOICE OF ENGINE**

BLEU Score Comparison:

| | Microsoft | SYSTRAN |
|---|---|---|
| R4 | 14.31 | **22.61** ***(highest SYSTRAN)*** |
| R5 | **14.41 (highest Microsoft)** | 21.68 |

Human Evaluation (Survey Results)

| | Lowest | Average | Highest |
|---|---|---|---|
| **Microsoft** | 6 | 8.1 | 9.5 |
| **SYSTRAN** | 9 | 9.2 | 10 |

Based on the above comparison, we recommend that SYSTRAN be used for the MT engine of the full-scale customization project. With the same sets of data, the BLEU score for SYSTRAN was consistently higher than Microsoft Custom Translator during the iterative training. One possibility is due to the differentiated parent model datasets that SYSTRAN has for different domain specializations, which could fit closer to our project domain.

It should be noted that the human evaluation survey results are imperfect, as the comparison between Microsoft and SYSTRAN originate from different rounds. Thus, a repeat test with constant datasets could be added in the future.

## PROJECT TIMELINE

**APR 8 — Post-Project Meeting**

- On the conclusion of the pilot project, the project results, recommendations for the full-scale project as well as improvements were presented. This is followed by the updated proposal (this document).

**APR 15 — Approval**

- Customer approves the updated proposal, project begins

**APR 18 — Data Compilation**

- Client to send over required dataset volumes of past subtitle translations (.srt format)

**APR 18-29 — Data Preparation**

- Data cleaning using Okapi Olifant, Visual Studio Code
  - File formatting & conversion (.srt to txt)
  - Data cleaning & pre-processing: remove symbols, stitch sentences
  - Alignment and .tmx creation

**MAY 2-4 — MT Training Round 1 + 1st Human Eval**

- Upload training, tuning and testing datasets to SYSTRAN,
- Establish initial baseline BLEU score after 1st training round
- Perform 1st human evaluation of quality with post-edited sample of MT output

**MAY 5-20 — MT Training Rounds 2-10**

- Conduct 9 subsequent rounds of iterative training and testing of MT engine
- Make adjustments each round aiming to improve BLEU score, e.g. adding/removing data, further data cleaning, adding dictionaries, etc.

**MAY 23-25 — 2nd Human Evaluation**

- Conduct post-editing
- Repeat evaluation based on quality metrics to check for improvements in MT quality after training rounds

**MAY 26-27 — Analysis of Pilot Project Findings**

- Analyze evaluation findings and quality improvements (if any)
- Ensure results of MT engine training meet cost, time and quality goals

**MAY 31 — Delivery to Client**

- Deliver MT engine and associated files to client
- Provide support on usage of MT engine as necessary

## FINAL DELIVERABLES

| | |
|---|---|
| **Outcomes of a Full-Scale MT Training Project (in ZIP file)** | A document with link, login access and user guide to the fully-trained NMT Engine workspace |
| | Details of data used for the full-scale machine translation project (training, tuning, and testing data sets) |
| | The translated subtitle files of this full-scale project |
| | A report on the viewer survey |
| | A report on the cost and time for executing the full-scale project |
| | Other related files |