



PILOT PROJECT UPDATED PROPOSAL

Updated Project Objectives:

This pilot project aims to estimate the work involved in training a neural machine translation (NMT) engine for translating strings on Disney maps at the Walt Disney World location from the source English (EN-US) to Simplified Chinese (ZH-CN). Post-editing Machine Translation (PEMT) should meet the following requirements:

1. Efficiency: PEMT 30% faster than human translation from 2 days to 34 hours (HT: \$0.20USD/word).
2. Cost: PEMT 30% cheaper than human translation \$413.80 USD to \$289.66 USD.
3. Quality: Our goal is to ensure that park guests who arrive go to Disney World have an enjoyable experience. For quality, we must first ensure that the machine translation meets our stringent quality standards for Disney (detailed below). Second, the quality must pass the human evaluation a Chinese speaker using a PEMT map navigate around the park.

Recommended Machine Translation Engine:

We recommend using SYSTRAN Model Studio as the machine translation engine for this project. In the pilot project, SYSTRAN outperformed Microsoft Custom Translator with the needs for the Disney map translation. It proved to be more compatible with the project format at the project dataset because of the quantity of available translation segments and character count.

Recommended Additional NTM Training:

To fully round of the training of this machine translation engine, we recommend an additional 5 rounds of machine translation. Ideally, this will increase the BLEU scores further and allow for a better result in NMT for Disney maps in Mandarin Chinese (Simplified). Additionally, 5-10 more hours for string inputting and corpus pruning will streamline the process further.

Updated Timeline:



*This timeline accounts for the extended deadline due to power outages

Team Members: Nicholas Niculescu, Emilia Primavera, Dayna Brown

Process:

- **Asset Preparation & Clean-Up:** Collect bilingual strings from Disney website, align segments in SDL Trados, and perform text clean up in Okapi Olifant.
- **MT Training (10 Rounds):** Use training, tuning, and testing data in Microsoft Custom Translator. Analysis each round via BLEU scores.
- **Postediting & QA:** Perform PEMT on 2000 word sample group. Use Quality metrics and track time and costs.
- **Evaluation & Estimates:** Implement human and automatic evaluation of PEMT. Compare PEMT sample group quality, efficiency, and cost against those of human translation. Prepare final analysis and breakdown of project.
- **Wrap-Up:** Conduct Pilot Project Post-Mortem, Cost to Savings Ratio, and update proposal.

Dataset:

- Training Dataset - Edited corpus from Wikipedia for en to zh, Hong Kong Disneyland website strings for attractions, food, shops, and entertainment.
- Testing Data - Picture List of Tokyo Disneyland's map.

MT Engine Comparison:

- Within the current limits of collected data the recommendation at the current stage of testing would be to continue with SYSTRAN. When using Microsoft custom translator, we faced several data-based roadblocks that SYSTRAN blew past.
- SYSTRAN also let us easily reevaluate all the rounds we completed with new data once we course corrected.

Recommended Translation Environment Tool:

- Use REGEX to check for untranslated strings, incorrect target language punctuation, and excessively long strings.
- Glossary Converters to handle segments and TMX in Excel Spreadsheets.

BLEU Scores:

Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7	Round 8
6.68(MS)	6.48(MS)	16.82(S)	15.77(S)	16.25(S)	16.31(S)	15.7(S)	17.4(S)
1	5	10					

MS: Microsoft Custom Translator, S: SYSTRAN

Cost-Reduction Recommendations:

- Automate manual collection of segments
- Use existing Disney translation memories
- Use paid, full functionality of NMT training software

QA Metrics:

Type:	Minor	Major	Critical
Abusive	x	x	20
Accuracy	1	5	10
Terminology	1	5	10
Style	1	5	10
Mistranslation	1	5	10
Untranslated	1	5	10
Consistency	1	5	10

Evaluation:

Human Evaluations:

1. Even if the translation meets our quality standards it's vital to have a human attempt to use these maps. Our first human evaluation test would be to give a Chinese speaker our map and have them navigate around the park using google maps.
2. Have a Mandarin linguist rate the quality of the raw MT & PEMT from 1 to 5, where 1 is incomprehensible and 5 is indistinguishable from human translation.

QA & Evaluation Results:

Linguist Evaluation: In the first round, the linguist found a reference to marijuana in Chinese, it confused honey pot thus automatically failing our QA metrics. However, in our best round these issues were fixed. Additionally, post-editing helps compensate for such occurrences.

Costs:

Task	Hour(s)	Price (USD)	Total (USD)
Data Collection & Cleaning	25	\$40.00	\$1000.00
DTP	3	\$60.00	\$180.00
Glossary Creation	3	\$40.00	\$120.00
Management Fee	one-time fee	\$200.00*	\$200.00*
Post-editing	3	\$40.00	\$120.00
QA	2	\$20.00	\$40.00
		Total	\$1660.00 USD

*This is a management fee for the time and labor to wait for the training rounds to finish so customers aren't billed for variable times.

Anticipated Deliverables:

Should this project proceed into a fully trained NMT engine, the following results are expected.

- Record of all glossaries, TMs, corpus, training data, and testing data.
- Details of all the BLEU scores for each round of NMT engine training.
- Expanded results regarding the results of the human evaluation
- Project post-mortem, including cost analysis.