

STATEMENT OF WORK



Project Specification

Client	YouTube
Pilot Project Objective Write-Up	The goal of this pilot project is to ascertain if a custom-trained machine translation (MT) might be the solution for Chinese-language baking Youtubers who want to share their content with the English-speaking audience. Post-edited machine translations (PEMT) of Simplified Chinese subtitles should ideally meet the following criteria.
Pilot Project Goals	Quality Goals: Viewers should be able to follow the recipe, and the terminology used in the subtitles should be palatable. Unpalatable terminology like “embryo” should not be found in the translation. Timing Goals: PEMT approximately 40% faster than human translation. Pricing Goals: PEMT approximately 50% savings compared to human translation.

Human Evaluation

- **Conduct one round of human evaluation on usability after post-editing**
 - Do you think you could successfully bake the treat based on the subtitles? **(1-10)**
 - Were you able to successfully bake the treat? **(1-10)**
 - Could others recognize the treat as the same one in the video? **(0-1)**

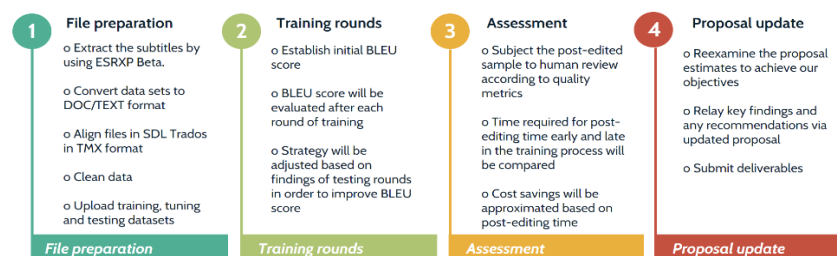
Datasets Explanation

Training Data	<ul style="list-style-type: none">• Subtitles from the YouTube channel 小高姐的 Magic Ingredients (2.39M subscribers)• Subtitles from the YouTube channel 我的爱心食谱 My Lovely Recipes (59.1K subscribers)
Testing Data	<ul style="list-style-type: none">• Subtitles from the YouTube channel 萨姐的南腔北调 Lisa's Kitchen (240K subscribers)
Tuning Data	<ul style="list-style-type: none">• Subtitles from the YouTube channel Ciao! Kitchen (283K subscribers)
Additional Data from the Pilot	<ul style="list-style-type: none">• English subtitles from the YouTube channel Tasty (20.9M subscribers) (decreased the BLEU score by 0.19)• Bilingual glossary of baking terms (increased the BLEU score by 0.09)• Subtitles from the YouTube channel 曼食慢语 Amanda Tastes (1.36M subscribers) were added to the training data (increased the BLEU score by 1)• Subtitles from MASA's Cooking ABC (1.55M subscribers) were added to the tuning data (decreased the BLEU score by 1.21) and later the training data (increased the BLEU score by 0.86)

Timeline

- **Mar 8-10** Phase 1 training: two training rounds with human evaluation and PEMT
- **Mar 11** Review of previous training rounds
- **Mar 12-14** Phase 2 training: two training rounds with human evaluation and PEMT
- **Mar 15** Review of previous training rounds
- **Mar 16-19** Phase 3 training: three training rounds with human evaluation and PEMT
- **Mar 20** Review of previous training rounds
- **Mar 21-22** Calculation of time, cost and savings, and quality evaluation
- **Mar 31** Update proposal

Pilot Process



Pilot Deliverables

- **An updated project proposal pertaining to subtitle translation for baking tutorials and containing the following**
Updated project workflow; Updated project quality expectation;
Updated project duration; Updated project pricing
Updated project outlook and recommendations
- **Data used in training, tuning, and testing of the MT engine**
- **Report on each iteration, including changes made and the ensuing BLEU scores**
- **Post-mortem**

Pilot Costs

<i>TASK</i>	Estimated Hours	Rate	Cost (USD)
<i>Client Meeting and Preparation</i>	4	\$40	\$160
<i>Document Cleaning and Alignment</i>	20	\$40	\$800
<i>MT Training</i>	13	\$40	\$520
<i>Post-editing</i>	3	\$40	\$120
<i>Human Translation</i>	1	\$40	\$40
<i>Project Management Fee</i>		10%	\$164

TOTAL \$1,804

Outlook and Recommendations

Our original budget was \$3,258, but we were \$1,454 under budget. We attribute this extreme cost saving to three factors.

- SYSTRAN was not a viable option for our project and our help request is still pending.
- We overestimated the amount of time we would need for MT training.
- We integrated the human evaluation into post-editing and organized our PEMT by seeing how much the editors could successfully post-edit in 20 minutes, and then evaluating it.

While we still believe in the validity of this idea, our pilot project was unsuccessful since our quality goals were ultimately unmet. If this project were to be initiated, we would suggest completely overhauling the data and increasing the quantity and quality of the bilingual segments used as tuning and training data, as this was the only change we made that positively impacted the quality. Microsoft Custom translator consistently failed to translate multiple words like “ham” and “Celsius” correctly, so we would suggest that rather than the glossary being applied in the form of a dictionary, a term base should pop up near the PEMT segments when the problem words are sighted in the source text, like it does in CAT tools. There should also be a limit on segment length set in the CAT tool.

In the event that the final project is implemented successfully, YouTube’s customers will enjoy calming and palatable subtitles alongside their favorite Chinese baking videos.