Seiha Translations
Pilot Project Proposal

Objective

The objective of this project was to train a Custom Translation Engine to translate song lyrics from Japanese to English, ensuring that the English lyrics are written naturally and without grammatical errors, aiming for 30% cost savings, and being 80% more time efficient. The following goals have been reviewed and updated with details to align with the insights gained from the pilot project conducted over a span of six weeks.

	Initial Goal	Findings	Improvements
Quality Goal	Post-edited Machine Translations (PEMT) are written naturally and without grammatical errors.	Human evaluation determined this goal to be achievable.	Although the goal was met, quality assessment could be more thorough. Larger evaluation sample to
Timing Goals	PEMT 80% faster than HT *(1125 Characters/Hour PEMT vs 625 Characters/Hour HT)*	Post-editing rate was calculated based on a sample size of 1,470 words post-edited in 1 hour. This is **3.74 times faster** than the standard human translation rate of 312.5 words/hour.	Very light post-editing was conducted and could be further improved by having a professional bilingual post-editing to have a better assessment.
Pricing Goals	PEMT 30% savings over human translation HT *($0.21/Word for PEMT vs 0.30/Characters for HT)*	Project spending on the machine training process was more than originally estimated.	In order to negotiate better cost savings for post-editing, we would need to find ways to be more efficient with the machine training process.

Pilot Process

Datasets

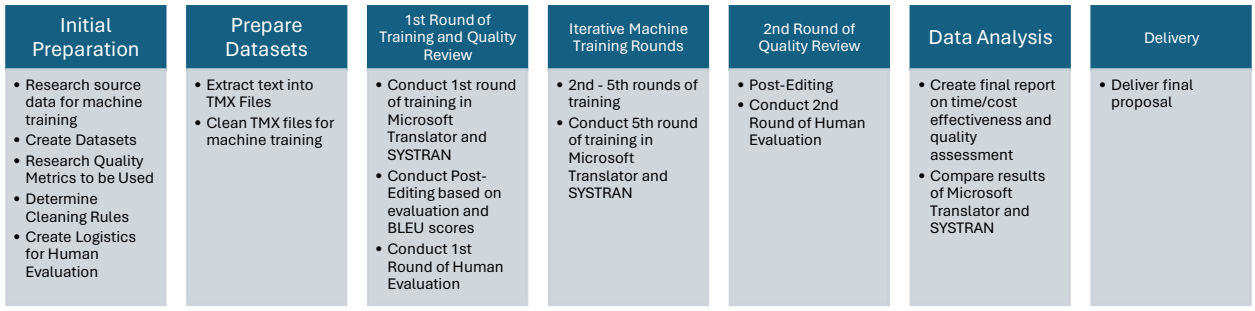
Our process involved collecting more datasets for the training than initially expected. The training set has been updated to include more corpus data that was used in the iterative training process of the pilot project.

Training	Testing	Tuning
•Top 75 pop songs in 2020 in Billboard Japan Charts •Top 75 songs from various genres •10,000 segments JESC Corpus (various fictional TV, movies and books) •5000 segments OpenSubtitles Corpus (various Japanese dramas)	•Top 50-100 pop songs of 2023 in Billboard Japan Charts	•Top 1-50 pop songs of 2023 in Billboard Japan Charts

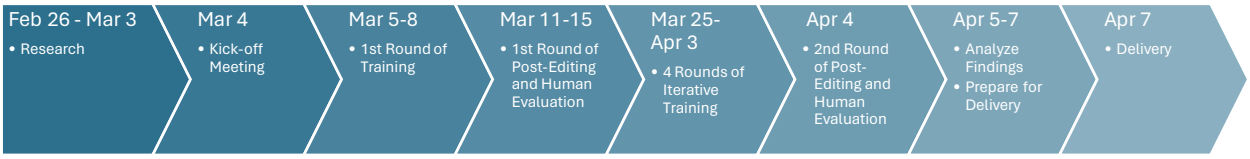
Seiha Translations
Pilot Project Proposal

SONY MUSIC

Workflow



Timeline



Human Evaluation Process

Two human evaluators were selected to participate in the quality assessment of this project. However, to fully assess if this project meets the quality goal, 8 more human evaluators (total 10) would be required.

Evaluators will be shown lyrics to 5 songs that were post-edited and asked the following.

1. **On a scale of 1-4, how grammatically correct do you find the English translated lyrics? (Were there any parts where the grammar seemed unclear?)**
   1 being poor grammar, 4 being excellent grammar

2. **On a scale of 1 – 4, how natural do the translated lyrics feel? (Do they sound forced, awkward, or otherwise flow unnaturally in certain parts?)**
   1 being unnatural, 4 being natural

Updated Pilot Costs

Additional tasks have been added that were not included in the original proposal and times have been updated to reflect the actual pilot project process. Most of the time was spent in the data collection/alignment phase, as well as additional hours spent in team meetings to discuss and review project details. Survey preparation time has also been added.

Seiha Translations
Pilot Project Proposal

SONY MUSIC

Task	Time	Rate	Cost
Research	4 hrs	$40/hr	$160
Data Collection and Alignment	10 hrs	$40/hr	$400
Data Cleaning	1 hrs	$40/hr	$40
MT Training Process 1. **Processing Time** 2. **Quality Assessment of BLEU scores** 3. **Editing Datasets**	6 hrs	$40/hr	$240
Post-Editing (2 Rounds)	2 hrs	$40/hr	$80
Team Meetings	8 hrs	$40/hr	$320
Survey Preparation	2 hrs	$40/hr	$80
Human Evaluation	2 hrs	$40/hr	$80
Quality Review	2 hrs	$40/hr	$80
TOTAL	**37 Hours**		**$1480**

Additional Deliverables

Details of Training Rounds (BLEU Scores)

Training Round	Microsoft	SYSTRAN	Model Details
Round 1	20.12	16.51	Songs (10K)
Round 2	19.00	7.25	JESC (10K)
Round 3	20.00	11.29	Songs + JESC (10K)
Round 4	20.61	14.42	Songs + JESC (20K)
Round 5	20.76	13.93	Songs + JESC + Open Subtitles (25K)
Range of Scores (Max – Min)	1.76	9.26	

Comparison of Microsoft vs Systran

Criteria	Microsoft	Systran
Time	Longer processing time	Shorter processing time
BLEU Scores	Higher BLEU scores	Lower Bleu Score
Dataset	Includes tuning set	Does not include tuning set
Test Results	No time expiration	3-day time expiration
Test Results	Preview option	No preview option
Test Results	Separate files for source and machine translated segments (easier to align in CAT tool)	Merged source and custom machine translated segments (need to manually unmerge for CAT Tool)

Suggestion for improvements

We don't anticipate many more rounds of training, however, there is still a lack of knowledge as to how to improve our BLEU scores. Despite collecting and adding more segments to the training set, the BLEU scores did not show much variance which contributes to the overall ambiguity of how we should edit the datasets in the iterative training process. The slight score improvements so far have led us to note that more volume of training data is perhaps useful and that including song segments shows slightly higher results. Our post-editing process could also be more refined by tracking certain types of errors and changes to reduce those types of errors in the custom translations.