Seiha Translations
Pilot Project Proposal

SONY MUSIC

Objective

The objective of this pilot project is to train a machine translation engine specializing in translating the lyrics of Japanese pop songs into English, aimed at providing subtitles and closed captioning for YouTube music videos.

1. Quality Goals
- The English translation is written naturally and without grammatical errors

2. Timing Goals
- Post-edited machine translation 80% faster than human translation
  (1125 Characters/Hour PEMT vs 625 Characters/Hour HT)

3. Pricing Goals
- PEMT 30% savings over human translation
  ($0.21/Word for PEMT vs 0.30/Characters for HT)

Pilot Process

Datasets

Training	Testing	Tuning
• Top 50 pop songs in 2020 in Billboard Japan Charts • Top 75 songs from various genres • 25 romantic poems • 5 scripts from television sitcoms and drama episodes	• Top 50-100 pop songs of 2023 in Billboard Japan Charts	• Top 1-50 pop songs of 2023 in Billboard Japan Charts

Workflow

Initial Preparation	Prepare Datasets	1st Round of Training and Quality Review	Iterative Machine Training Rounds	2nd Round of Quality Review	Data Analysis	Delivery
• Research source data for machine training • Create Datasets • Research Quality Metrics to be Used • Determine Cleaning Rules • Create Logistics for Human Evaluation	• Extract text into TMX Files • Clean TMX files for machine training	• Conduct 1st round of training in Microsoft Translator and SYSTRAN • Conduct Post-Editing based on evaluation and BLEU scores • Conduct 1st Round of Human Evaluation	• 2nd - 5th rounds of training • Conduct 5th round of training in Microsoft Translator and SYSTRAN	• Post-Editing • 2 Human Evaluators	• Create final report on time/cost effectiveness and quality assessment • Compare results of Microsoft Translator and SYSTRAN	• Deliver final proposal

Timeline

Feb 26 - Mar 3	Mar 4	Mar 5-8	Mar 11-15	Mar 18-22	Mar 25-29	Apr 1-4	Apr 5
• Research	• Kick-off Meeting	• 1st Round of Training	• 1st Round of Quality Review and Human Evaluation	• 4 Rounds of Iterative Training	• 2nd Round of Quality Review and Human Evaluation	• Analyze Findings • Prepare for Delivery	• Delivery

Seiha Translations
Pilot Project Proposal

SONY MUSIC

Human Evaluation Process

Two rounds of Human Evaluation will be done after the ***first*** and ***last*** rounds of training. We will look to have two evaluators who are native speakers of American English.

Evaluators will be shown lyrics to 5 songs and asked the following.

1. **On a scale of 1 (very hard) – 4 (easy), rate the ease of grammatical comprehension of the lyrics.**
2. **From the same scale of 1 – 4, rate how natural the language seems (for example, strange sentence structures, vocabulary)**

Pilot Costs

Task	Time	Rate	Cost
Research	4 hrs	$40/hr	$160
Source Collection and Alignment	7 hrs	$40/hr	$400
Data Cleaning	2 hrs	$40/hr	$120
Glossary Creation	1 hrs	$40/hr	$40
MT Training Process 1. **Processing Time** 2. **Quality Assessment of BLEU scores** 3. **Editing Datasets**	6 hrs	$40/hr	$480
Post-Editing (2 Rounds) • Different rate for initial training	4 hrs	$40/hr	$160
Meetings	2 hrs	$40/hr	$80
Human Evaluation	2 hrs	$40/hr	$80
Quality Review	2 hrs	$40/hr	$80
TOTAL	**30 Hours**		**$1200**

Deliverables

Deliver final proposal, which will include

- Details of each training round with BLEU scores
- Updated cost/time efficiency of PEMT compared to HT
- Comparison of Microsoft Translator vs SYSTRAN
- Post-Mortem for suggestions on improvement for future projects