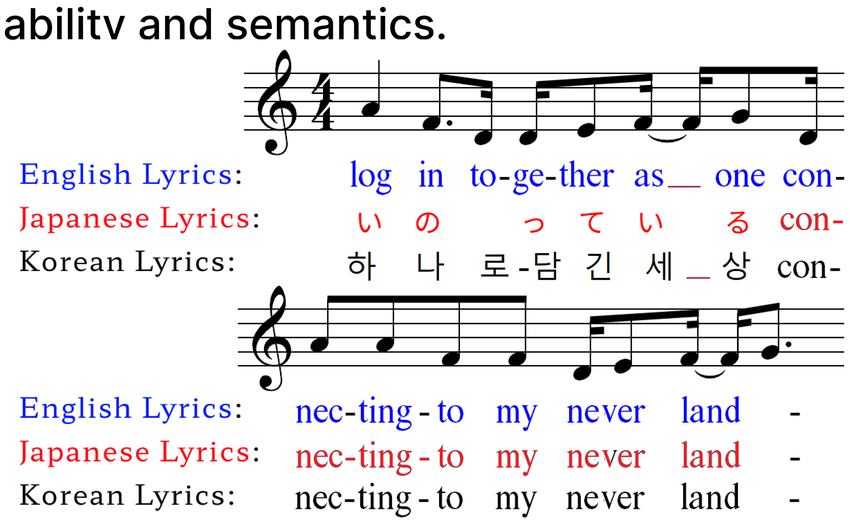
A Computational Evaluation Framework for Singable Lyric Translation

Haven Kim[1], Kento Watanabe[2], Masataka Goto[2], Juhan Nam[1]
[1] Korea Advanced Institute of Science and Technology (KAIST), South Korea
[2] National Institute of Advanced Industrial Science and Technology (AIST), Japan

Singable Lyric Translation

- The goal is to sing in another language and not necessarily deliver the exact meaning.
- Translating lyrics, unlike conventional translation tasks, requires a delicate balance between singability and semantics.



Dataset Construction

- We collected singable lyric translation dataset, aligned line-by-line and section-by-section, across English, Japanese, and Korean.

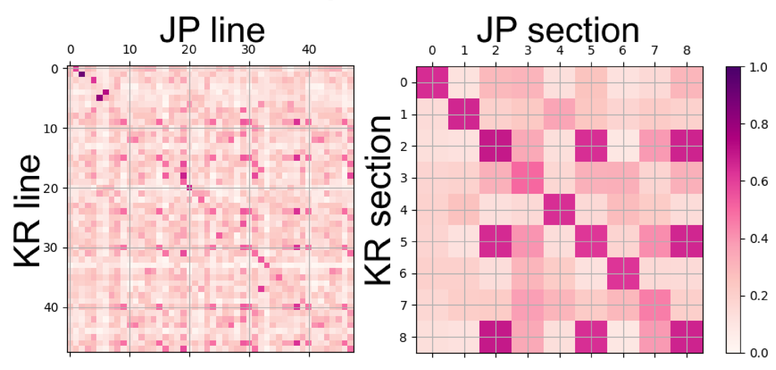
Section #	Line #	English (EN)	Japanese (JP)	Korean (KR)
1	1	Twinkle, twinkle, little star	きらきらひかる	반짝 반짝 작은별
	2	How I wonder what you are	おそらのほしよ	아름답게 비치네
	3	Up above the world so high	まばたきしては	서쪽 하늘에서도
	4	Like a diamond in the sky	みんなをみてる	동쪽 하늘에서도
	5	Twinkle, twinkle, little star	きらきらひかる	반짝 반짝 작은별
	6	How I wonder what you are	おそらのほしよ	아름답게 비치네
2	7	Twinkle, twinkle, little star	きらきらひかる	반짝 반짝 작은별
	8	How I wonder what you are	おそらのほしよ	아름답게 비치네
	9	When the blazing sun is gone	みんなのうたが	서쪽 하늘에서도
	10	When he nothing shines upon	とどくといいな	동쪽 하늘에서도
	11	Then you show your little light	きらきらひかる	반짝 반짝 작은별
	12	Twinkle, twinkle, all the night	おそらのほしよ	아름답게 비치네

- We compared statistical results on singable lyric translation dataset with those on internet-sourced non-singable lyric translation data.

Evaluating Semantics

Semantic Similarity

- Our calculation is based on the cosine similarity between embeddings generated by a pre-trained sentence embedding model.
- We propose evaluating on a section-by-section basis rather than on a line-by-line basis as it is closer to real world translation practices.



Line-wise and section-wise semantic similarity matrix between Japanese and Korean versions of "Wie wird man seinen Schatten los?"

Evaluating Singability

Line Syllable Count Distance

- One of the most important factors that determine the preservation of melodies.
- < 0.15 is an ideal value.

$$Dis_{syl}(\mathbf{X},\tilde{\mathbf{X}}) = \frac{1}{2n}\sum_{i=1}^{n}\left(\frac{|syl(x_i)-syl(\tilde{x_i})|}{syl(x_i)} + \frac{|syl(x_i)-syl(\tilde{x_i})|}{syl(\tilde{x_i})}\right)$$

Phoneme Repetition Similarity

EN	JP	KR
Let it go, let it go (lɛt ɪt goʊ, lɛt ɪt goʊ) Can't hold it back anymore (kænt hoʊld ɪt bæk ˌɛniˈmɔɹ) Let it go, let it go (lɛt ɪt goʊ, lɛt ɪt goʊ) Turn away and slam the door (tɜɹn əˈweɪ ænd slæm ðə dɔɹ)	ありのままの (a-ri-no-ma-ma-no) 姿見せるのよ (su-ga-ta-mi-se-ru-no-yo) ありのままの (a-ri-no-ma-ma-no) 自分になるの (ji-bun-ni-na-ru-no)	다 잊어 다 잊어 (da ijeo da ijeo) 이젠 참지 않을거야 (ijen chamji anheul-geoya) 다 잊어 다 잊어 (da ijeo da ijeo) 문을 열고 나아갈거야 (muneul yeolgo naagal-geoya)

- Some sections call for the high repetition of phonemes (e.g., rhyme) while others do not.
- We decomposed lyrics for each section into phonemes and grouped them as bi-grams. This value, called *pho*, gets lower when the repetition degree is high.

$$pho(X_i) = \frac{\textit{unique bi-gram \# in } X_i}{\textit{total bi-gram \# in } X_i}$$

- Singable lyric translation pairs show high *pho* correlations.

$$Sim_{pho}(\mathbf{X},\tilde{\mathbf{X}}) = corr(\{pho(X_1),...,pho(X_m)\},\{pho(\tilde{X_1}),...,pho(\tilde{X_m})\})$$

Musical Structure Distance

- Musically similar sections share similar linguistic features such as
  a. Choice of phonemes

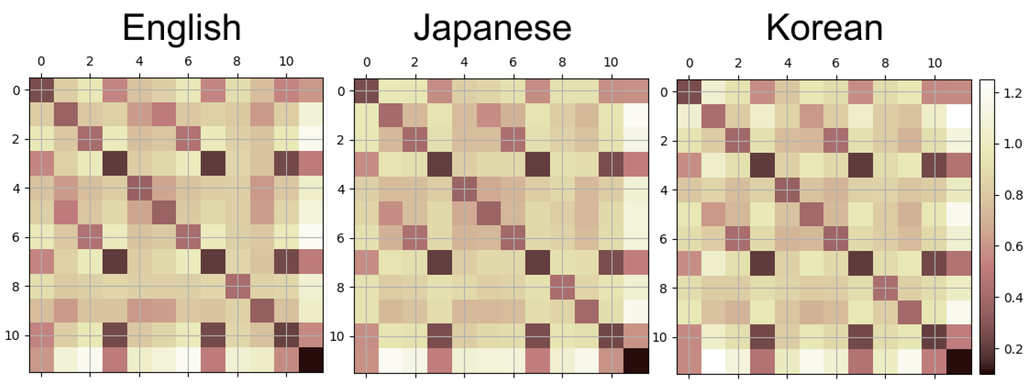
Twinkle, twinkle, little star How I wonder what you are	Then you show your little light Twinkle, twinkle, all the night

  b. Degree of phoneme repetition (e.g., rhyme)

Twinkle, twinkle, little star How I wonder what you are	Then you show your little light Twinkle, twinkle, all the night

- We plotted matrices to represent the self-dissimilarity in these aspects. These matrices look highly similar across different languages when singably translated. Therefore, the distances between matrices are smaller than those between non-singably translated pairs.

$$diss(X_i,X_j) = \underbrace{pho(X_i \mathbin{+\!\!+} X_j)}_{\text{Dissimilarity in the choice of phonemes ('++' denotes the concatenation of text)}} + \underbrace{|pho(X_i) - pho(X_j)|}_{\text{Dissimilarity in the degree of phoneme repetition}}$$



Self-dissimilarity matrix using "Icy" by ITZY