COMPARING SINGABILITY OF TRANSLATED SONGS

Natural Language Processing - Final Project
Umeå University

Author
Barbora Stepankova
30. Oct 2023

Contents

1 Introdiction

The topic of song singability is widely discussed in the linguistic and translation areas. Translation of songs while conserving their singability is a complex task. Even human translators struggle with it, and with machine translation, we are more focused on preserving the meaning than a certain structure.

My goal for this research project is to implement already proposed metrics, as well as come up with my own to quantify the singability of a song in a few numbers.

After implementing the metrics, I will compare the performance of the original human-translated lyrics and machine-translated lyrics.

The goal is to find out how big of a performance gap is between these two methods of lyric translation, and if there is any difference between the experiments when I switch the target and source language.

2 Background

2.1 Singability

Singability is a property of a text, meaning that the text can be easily and comfortably mapped onto a certain melody. This concept of singability has been debated many times, a good definition is provided for example by Peter Low[1]: Singable song has to have

- Singability
- Sense
- Naturalness
- Rhythm
- Rhyme

For this report, we will consider all these criteria together as "Singability".

2.2 Singable translations

The topic of song generation and translation is starting to get more attention recently. Most of the research currently focuses on generating singable lyrics to a melody[2][3] or translating lyrics to fit a given melody[4].

Recently, a new metric for evaluating the singability of translated lyrics was proposed.[5] The metric doesn't rely on the melody and considers just the two texts and their linguistic properties.

3 Research questions

My hypothesis has several parts.

First, **I am expecting the original human translations to be more singable than the machine-translated ones.**

The second part concerns the source and target languages of the lyrics.

As of now, the state-of-the-art language models focus almost purely on English, and using these models for other languages gives worse results.

Because of this, **I am expecting the Czech-English results to be more singable than the English-Czech.**

4 Data

For this task, I am using exclusively lyrics from Disney musicals, originally written in English and translated by a human translator into Czech. All of the song lyrics gathered were freely downloadable from the internet.

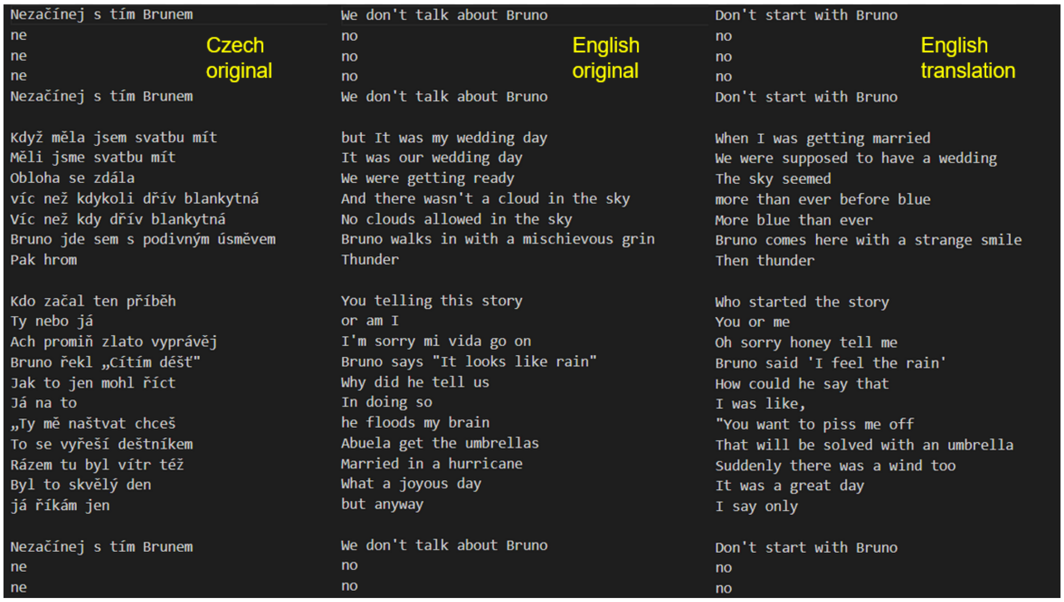
My dataset includes 76 songs from Disney animated musicals, namely:

- Encanto
- Frozen 1
- Frozen 2
- Lion King
- Moana
- Tangled

The limited size of the dataset, as well as the lack of diversity, may have an impact on the results.

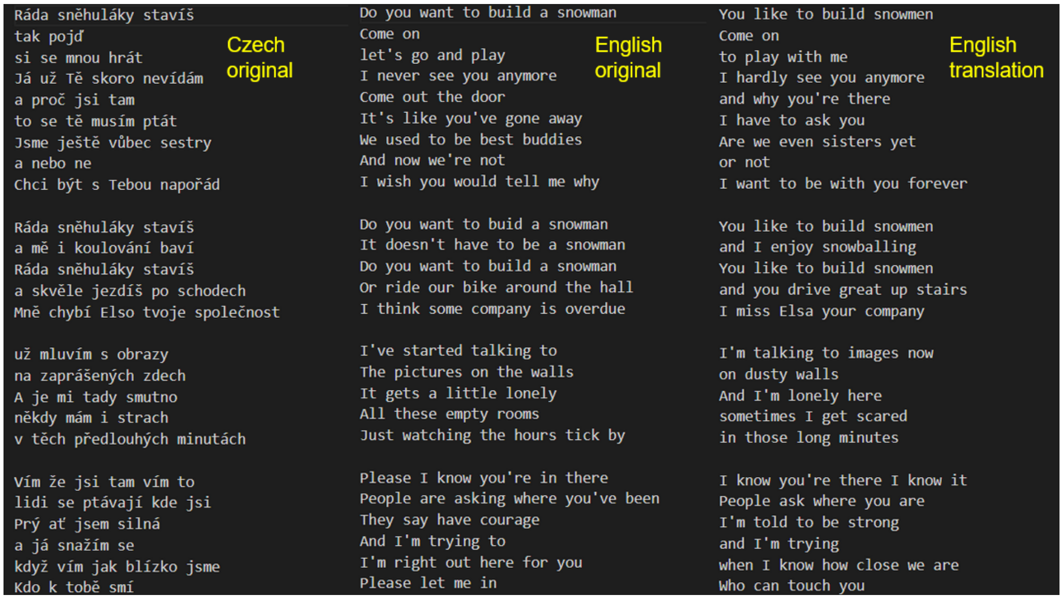
The translated data were obtained by using the Lindat translator API.

The lyrics were aligned using the lyric-aligning function and then hand-corrected to map line by line exactly.



Czech original	English original	English translation
Nezačínej s tím Brunem	We don't talk about Bruno	Don't start with Bruno
ne	no	no
ne	no	no
ne	no	no
Nezačínej s tím Brunem	We don't talk about Bruno	Don't start with Bruno
Když měla jsem svatbu mít	but It was my wedding day	When I was getting married
Měli jsme svatbu mít	It was our wedding day	We were supposed to have a wedding
Obloha se zdála	We were getting ready	The sky seemed
víc než kdykoli dřív blankytná	And there wasn't a cloud in the sky	more than ever before blue
Víc než kdy dřív blankytná	No clouds allowed in the sky	More blue than ever
Bruno jde sem s podivným úsměvem	Bruno walks in with a mischievous grin	Bruno comes here with a strange smile
Pak hrom	Thunder	Then thunder
Kdo začal ten příběh	You telling this story	Who started the story
Ty nebo já	or am I	You or me
Ach promiň zlato vyprávěj	I'm sorry mi vida go on	Oh sorry honey tell me
Bruno řekl „Cítím déšť"	Bruno says "It looks like rain"	Bruno said 'I feel the rain'
Jak to jen mohl říct	Why did he tell us	How could he say that
Já na to	In doing so	I was like,
„Ty mě naštvat chceš	he floods my brain	"You want to piss me off
To se vyřeší deštníkem	Abuela get the umbrellas	That will be solved with an umbrella
Rázem tu byl vítr též	Married in a hurricane	Suddenly there was a wind too
Byl to skvělý den	What a joyous day	It was a great day
já říkám jen	but anyway	I say only
Nezačínej s tím Brunem	We don't talk about Bruno	Don't start with Bruno
ne	no	no
ne	no	no

Figure 1: Aligned Czech and English human translation with Cs-En machine translation of "We don't talk about Bruno" from Encanto.



Czech original	English original	English translation
Ráda sněhuláky stavíš	Do you want to build a snowman	You like to build snowmen
tak pojď	Come on	Come on
si se mnou hrát	let's go and play	to play with me
Já už Tě skoro nevídám	I never see you anymore	I hardly see you anymore
a proč jsi tam	Come out the door	and why you're there
to se tě musím ptát	It's like you've gone away	I have to ask you
Jsme ještě vůbec sestry	We used to be best buddies	Are we even sisters yet
a nebo ne	And now we're not	or not
Chci být s Tebou napořád	I wish you would tell me why	I want to be with you forever
Ráda sněhuláky stavíš	Do you want to buid a snowman	You like to build snowmen
a mě i koulování baví	It doesn't have to be a snowman	and I enjoy snowballing
Ráda sněhuláky stavíš	Do you want to build a snowman	You like to build snowmen
a skvěle jezdíš po schodech	Or ride our bike around the hall	and you drive great up stairs
Mně chybí Elso tvoje společnost	I think some company is overdue	I miss Elsa your company
už mluvím s obrazy	I've started talking to	I'm talking to images now
na zaprášených zdech	The pictures on the walls	on dusty walls
A je mi tady smutno	It gets a little lonely	And I'm lonely here
někdy mám i strach	All these empty rooms	sometimes I get scared
v těch předlouhých minutách	Just watching the hours tick by	in those long minutes
Vím že jsi tam vím to	Please I know you're in there	I know you're there I know it
lidi se ptávají kde jsi	People are asking where you've been	People ask where you are
Prý ať jsem silná	They say have courage	I'm told to be strong
a já snažím se	And I'm trying to	and I'm trying
když vím jak blízko jsme	I'm right out here for you	when I know how close we are
Kdo k tobě smí	Please let me in	Who can touch you

Figure 2: Aligned Czech and English human translation with Cs-En machine translation of "Do you want to build a snowman" from Frozen.

5 Methods

5.1 Proposed evaluation functions

All following metrics are implemented according to "A Computational Evaluation Framework for Singable Lyric Translation" paper[5] and their detailed description can be found there.

5.1.1 Syllable count distance

The distance between the number of syllables per line can be computed as an average of differences between the number of syllables for each line, divided by the length of the line.

5.1.2 Phoneme repetition similarity

Phoneme repetition similarity tells us about the correlation of phoneme repetition across different song parts. If a translation has a high phoneme repetition similarity, it suggests that sections of the song with high and low phoneme repetition correspond to each other.

Usually, choruses are sections with high phoneme repetition and verses have low phoneme repetition similarity.

The phoneme repetition is computed as ”the number of distinct bigrams”/”total number of bigrams” in a section.

5.1.3 Musical structure distance

Musical structure distance shows the distance between inner-song dissimilarities of each of the lyrics.

Inner-song dissimilarity means how different are individual song sections from each other. Meaning that choruses will be more similar to each other than let's say verses, or a chorus and a verse.

This metric tells us how much of the inner song structure the translation kept.

5.1.4 Semantic similarity

This metric shows how much the meaning deviates per section.

I am using a pre-trained BERT model for English. Because of that, I translated all Czech sections into English before getting the section embedding.

I compared these embeddings using cosine similarity and then computed the whole semantic similarity by summing up the cosine similarities of the sections while taking into account the lengths of individual sections.

5.2 My additions to evaluation functions

The original paper focused on the linguistic relationship between English, Japanese and Korean lyrics. I added my metrics to evaluate the more specific Czech-English text relationship.

5.2.1 Rhyme scheme distance

Rhymes are very prominent in both Czech and English texts, compared to more Eastern poetry and songs. Therefore I decided to include a rhyme scheme distance as a metric.

My metric looks into how many lines in each song section have a rhyme in the section and how many don't. It takes the symmetric difference of the rhymes found in both versions and calculates the distance between the two rhyme schemes by dividing this difference by the number of lines.

This way, having a different rhyme scheme in each song gives better results than having a rhyme scheme in one song and nothing in the other.

5.2.2 Stress and phoneme mapping distance

Czech and English both have different definitions of stress. In Czech, the most stress falls on the first syllable, but there are also long and short stresses concerning the vowels.

Each vowel is either long or short, and it is clearly marked by an accent above the vowel.

Take for example *a* (Pronounced as **u** B**u**t) is a short vowel and *á* (Pronounced as **a** in F**a**ther) is the long equivalent. It is pronounced the same, only as a short or a bit longer sound.

Other examples are *e* and *é*, *i* and *í*, *o* and *ó*, *u* and *ú* or *ů* (both long spellings are pronounced the same, it is a grammatical construct) and finally *y* and *ý*.

Compared to that, even though in English the concept of long and short vowels exists too, it isn't clear directly from the text. I tried compensating for this by coming up with a set of overwriting rules and then comparing the modified texts.

I am using a simple average edit distance over a line to determine the phonetic distance between the modified lines.

6 Results

In this section I will present and interpret the results I got from running all of the metric methods on my dataset.

6.1 Syllable count distance

Evaluating the syllable count distance on the dataset gives the average of:

- 0.03559 for human translations
- 0.21583 for Czech to English machine translations
- 0.26194 for English to Czech machine translations

After doing a t-test, I can say that both Czech-to-English and English-to-Czech machine translations have significantly worse values than human translations, with a p-value close to 1.

The results for English as the target language are better than for Czech as the target language, with a p-value of 0.0003.

These results were expected, as the machine translation had no constraint considering the length, while for human translators the syllable length is one of the key features.

6.2 Phoneme repetition similarity

Evaluating the phoneme repetition similarity on the dataset gives the average of:

- 0.67395 for human translations
- 0.67076 for Czech to English machine translations
- 0.72004 for English to Czech machine translations

These results seem similar, and after doing the t-test, we see there is no statistically significant difference in these results. With p-values of 0.95, 0.328 and 0.352, we can say that the human translations and both of the machine translations have comparable phoneme repetition similarity.

This is probably because if the same line is repeated twice in the original lyrics, the machine translator is going to translate the line the same way both of those times.

This metric is not very well suited for this task.

6.3 Musical structure distance

Evaluating the musical structure distance on the dataset gives the average of:

- 0.01633 for human translations
- 0.01531 for Czech to English machine translations
- 0.01620 for English to Czech machine translations

Same as above, there is no statistically significant difference in these results with p-values of 0.57, 0.93 and 0.64, we can say that the human translations and both of the machine translations have comparable musical structure distance.

It makes sense since the machine translator will translate for example all instances of chorus, which is similar or the same, also similarly.

6.4 Semantic similarity

Evaluating the semantic similarity on the dataset gives the average of:

- 0.70659 for human translations
- 0.93112 for Czech to English machine translations
- 0.95113 for English to Czech machine translations

With semantic similarity, we're getting much better results on the machine translation than on human translation, because the machine translation isn't constrained by anything and doesn't have to compromise on meaning.

P-values for testing if human translation is significantly worse than machine translations are both 1.

However, the p-value for testing that the Czech-to-English translation and English-to-Czech translation have the same performance is 0.006, therefore we have to reject this hypothesis.

The English-to-Czech translation has significantly better performance than Czech-to-English.

This goes against our hypothesis, but a possible explanation could be the nature of semantic similarity comparison. To compare the meaning, we are translating the Czech text into English. That means that in the end, we are comparing two English texts, meaning that one time both of the texts are translated, and the other time just one gets translated two ways. That could reduce the error.

6.5 Rhyme scheme distance

Evaluating the rhyme scheme distance on the dataset gives the average of:

- 1.32015 for human translations
- 2.14102 for Czech to English machine translations
- 2.38178 for English to Czech machine translations

After doing the t-test, we can say that human translation have a significantly smaller rhyme scheme distance, with both p-values for both machine translations being equal to 0.99.

For comparison of Czech-to-English and English-to-Czech translations, there is no significant difference between the rhyme scheme distance of these two, with the p-value of 0.3

6.6 Stress and phoneme mapping distance

Evaluating the stress and phoneme mapping distance on the dataset gives the average of:

- 0.65611 for human translations
- 0.67706 for Czech to English machine translations

- 0.67169 for English to Czech machine translations

With stress and phoneme mapping distance, the human translation is also significantly better, with p-values of 0.99.

Comparing the two machine translations also results in a strong similarity of results, with the p-value of 0.387

7 Conclussion

Song translation is a difficult task even for humans, and with the tools we have now, for computers it is nearly impossible. As we saw, many of the important aspects of a singable song translation can't be, or aren't yet, quantified into a function which returns a number.

My first hypothesis was: **I am expecting the original human translations to be more singable than the machine-translated ones.**

When looking at them with the naked eye, they definitely are. According to the metric, 3/6 metrics say that HT is more singable than MT. 2/6 metrics don't see a difference, and 1/6 metrics say that MT is more singable than HT.

By summing up the positive, negative and neutral results, **human translations are more singable than machine translations.**

More important than just answering the hypothesis is thinking about if these metrics actually did a good job of getting the essence of what makes a translation singable, and I think they did not. Especially the phoneme repetition similarity and musical structure difference. Both of them are quantifying a quality of a song that is usually not lost during translation.

Instead of these, I would propose more metrics studying the rhythm and tempo and repetition of the lyrics in general.

Another metric that is necessary to have, but a bit tricky to evaluate is the semantic similarity. The semantic similarity needs to be reasonably high, but it is definitely not the most important aspect of the translation.

As for my second hypothesis: **I am expecting the Czech-English results to be more singable than the English-Czech.**

According to the metrics, 1/6 is slightly favouring the Czech-to-English translations, 1/6 is favouring the English-to-Czech translations, and 4/6 methods are quite strong about them being indistinguishable.

Therefore I conclude that in this case, there is no difference between Czech-to-English and English-to-Czech translations.

References

[1] P. Low, “Singable translations of songs,” *Perspectives*, vol. 11, no. 2, pp. 87–103, 2003. [Online]. Available: https://doi.org/10.1080/0907676X.2003.9961466

[2] F. Guo, C. Zhang, Z. Zhang, Q. He, K. Zhang, J. Xie, and J. Boyd-Graber, “Automatic song translation for tonal languages,” *arXiv preprint arXiv:2203.13420*, 2022. [Online]. Available: https://aclanthology.org/2022.findings-acl.60.pdf

[3] K. Watanabe, Y. Matsubayashi, S. Fukayama, M. Goto, K. Inui, and T. Nakano, “A melody-conditioned lyrics language model,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2018, pp. 163–172. [Online]. Available: https://aclanthology.org/N18-1015.pdf

[4] L. Ou, X. Ma, M.-Y. Kan, and Y. Wang, “Songs across borders: Singable and controllable neural lyric translation,” *arXiv preprint arXiv:2305.16816*, 2023. [Online]. Available: https://arxiv.org/pdf/2305.16816.pdf

[5] H. Kim, K. Watanabe, M. Goto, and J. Nam, “A computational evaluation framework for singable lyric translation,” *arXiv preprint arXiv:2308.13715*, 2023. [Online]. Available: https://arxiv.org/pdf/2308.13715.pdf

A Repository

All the data and all the code used to preprocess data, and evaluate the *Methods functions*, including the functions for syllabification and rhyme-finding can be found at `https://github.com/stepankovab/NLP-Umea/tree/main/FinalProject`