

外国語の歌曲を自動訳詞するシステムのための課題と手法の検討

A Study on Task and Methods for Automatic Translation of Foreign Lyrics

西村 綾乃 *1
Ayano Nishimura

伊藤 貴之 *2
Takayuki Itoh

*1 お茶の水女子大学
Ochanomizu University

*2 お茶の水女子大学
Ochanomizu University

Music has been composed all over the world. Opportunities to listen to songs in foreign language are increasing thanks to the evolution of the Internet. However, it is difficult for us to completely understand the lyrics in foreign languages other than our mother tongue. Although sometimes we can see free translation or literal translation of lyrics, it is not a form of lyrics that fits to melody. Therefore, we propose an automatic translation system and discuss the problem to implement this system. We assume that this system can input parameters based user's preference or target, and freely set the abstrusity level of singing or words. In this paper, we report the result of experiment and evaluation on a method of selecting lyrics candidates from sentences generated using 2-gram.

1. はじめに

オペラやミュージカルのような音楽劇が代表するように、音楽と言葉には密接な関係がある。文学を元に音楽劇が創作されるのはもちろん、作家の詩に音楽家が曲をつけることによって、歌唱曲が完成する場合も多い。

現在ではインターネットが普及したことにより、世界中の音楽をいつでも気軽に聞くことが可能になった。しかしながら、外国語の歌曲は音楽そのものを楽しむことは出来ても、馴染みのない言語による歌詞の内容を完全に理解することは難しい。中には、歌詞の意訳や直訳が掲載されていることもあるが、単なる翻訳文はメロディの音数や流れと合致しないため、そのまま歌うと音楽が崩れてしまう。

一方で、海外の多くの歌曲は日本語に訳詞されて馴染み深いものとなっている。しかし、訳詞には語学力や語彙力に加えて音楽の知識も必要であり、非熟達者には難しい作業である。具体的には、訳詞を行う歌曲に対して、原語での意味理解、音楽的なフレーズやリズム構成の理解、歌唱における日本語の発音への理解などが必要になってくる。多数の歌曲で1つの作品となるミュージカルやオペラなどでは、1曲ずつの理解だけではなく、全編を通した話の流れや、メロディフレーズの使い方などの理解も必要となり、その労力は多大なものとなる。

そこで著者は、ユーザのパラメータ設定に応じて音楽に相応しい形で外国語歌詞の日本語訳を出力するインタラクティブな自動訳詞システムの開発を目指している。自動訳詞の手法が確立できることによって、オペラやオペレッタ、ミュージカルなどの大量の歌詞を含む作品の訳詞も容易になり、システムを通じて日本語での上演のための補助を行うことで、日本国内での音楽文化の発展に貢献できるだろう。本稿では、訳詞の際の課題点となるモーラ数に着目し、2-gramを用いて生成した文章群から歌詞候補を選択する手法について、実験・評価を行った結果を報告する。

2. 訳詞の課題

日本語への訳詞の歴史は近藤朔風らが切り拓いた[松田 07]。代表される訳詞の作品には「野ばら」や「ローレライ」などが

あり、これらは現代でも訳詞時と変わらない日本語で歌い継がれている。近藤は原言語の意味と韻律を可能な限り日本語に再現しようとする訳詞を重視した。本提案においても、原言語の意味をなるべく日本語に再現するための自動訳詞を目指す。

日本語への自動訳詞システムの開発に着手するにあたり、まず訳詞特有の課題を列挙する。

原言語の歌詞を日本語に訳すとほとんどの場合、音数（モーラ数）が増えるという問題点にぶつかる。しかしながら、メロディに使われている音符の数は決まっているため、訳詞においてはモーラ数を極端に増やすことはできるだけ避けなければならない。これについて、“訳詞においては、原文の中のどの部分は訳出し、どの部分は省略するかという取捨選択が行われることになる”と、松田は述べている[松田 13]。結果として日本語への訳詞において、原言語の意味を簡略化する必要が生じることが多い。また、モーラ数に関わる課題として、日本語の発音の観点から「撥音」すなわち「ん」と発音される音については、モーラ数を1とする場合と、直前の発音と合わせて歌うことで、モーラ数に含めない場合がある。その他にも、促音、長音、二重母音と呼ばれる発音をモーラとして独立させるかの判断は、その歌唱が与えるリズム感にも影響を与える。

さらに、松田は本文中で、こうも述べている。“曲の終わりに歌いあげる音では、意味のレベルを犠牲にしても「ア」か「オ」の母音にすべき場合がある。”これは、歌唱に適した日本語の発音の課題とも言える。山本は日本語の母音について以下のように述べている[山本 73]。“「あ・お」は会話にも歌唱においてもひびきやすい母音であるが、「え」は歌唱においてとくにひびかせにくい母音である。そこでたえず発声訓練において美しい共鳴をとるよう配慮されている。「い」「う」は歯をくいしばったように口腔中が狭くなり、共鳴のない固い平たい声になりやすいのである。”以上のことから、訳詞を行う際に、選択した単語の母音に対して、言葉の意味を重視するか、歌唱的な面を重視するかといった訳詞の問題が生じることが分かる。

ここで、訳詞をする際に課題となる点をまとめる。

課題 1 日本語への翻訳によってモーラ数が増える

課題 2 原語の意味を取捨選択し、簡略化する必要がある

連絡先: nishimura.ayano@is.ocha.ac.jp, itot@is.ocha.ac.jp

課題 3 撥音などの発音を 1 モーラとして独立させるか、リズムなどを考慮して判断する必要がある

課題 4 音楽的な理由から、特定のモーラに対する母音を限定する必要がある

これらの課題について、人手で訳詞を行った場合でも、結果にばらつきが出ることは自明であり、システムにおいても訳詞の解を 1 つに絞るのは難しいことが予想される。そこで、ユーザが自分の好みに合わせて自由に調節できるパラメータを設ける必要があると考えた。

3. 自動訳詞システムの構想

本研究では、ユーザが訳詞に関するパラメータを設定し、MusicXML などの楽譜情報を与えることで日本語の歌詞候補群を生成する、インタラクティブな Web アプリケーションの開発を想定している。

現段階でユーザが設定すると考えられるパラメータを以下に列挙する。

1. モーラ数の制限
2. 母音の制限
3. 生成された文章の歌詞らしさ

パラメータは現段階での内容であり、今後の実装によっては増減する可能性がある。続いて、各パラメータの詳細について述べる。

3.1 モーラ数の制限

前章で課題 1、課題 2 として挙げた通り、基本的にメロディに使われている音数は決まっており、それを越えたモーラ数の歌詞はなるべく避けるべきである。しかしながら、人手での訳詞においても原言語の意味を反映するために、モーラ数が増えることを許容する場合がある。そこで、歌詞候補を生成する際に、モーラ数の増加についてユーザが任意に選択できるべきであると考えた。

3.2 母音の制限

課題 4 として挙げた通り、歌唱的な理由から、特定の位置のモーラに対して「ア」や「オ」などの母音に限定するべき場合がある。そのため、歌いやすさや韻律を考慮する際に母音の指定が必要になると考えた。

しかしながら、松田が“意味のレベルを犠牲にしても”と述べているように、指定する母音が増えると、歌詞の候補も減り、原言語と全く意味が異なる歌詞が出て来る可能性がある。

3.3 生成された文章の歌詞らしさ

生成された文章が、必ずしも歌詞らしい文章であるとは限らない。そのため、ユーザが歌詞らしさをどれだけ重視するかを設定する必要がある。ここで、piapro[pia 17] から生成したコーパスは、ある程度、歌詞らしくない単語を除外できているという実験結果がある [阿部 12a]。そこで、piapro コーパス中に歌詞候補に含まれる単語が存在した場合は、歌詞らしいと判断できると考えた。

歌詞らしさについて、生成文を形態素解析し、名詞、動詞、形容詞、副詞を抜き出し調査対象の単語とする。そして、生成文の単語が piapro コーパスに含まれる割合を調べる。生成文を s とし、生成文の歌詞らしさを $L(s)$ とした場合、以下のよう

$$L(s) = \frac{\text{piapro コーパスに含まれていた単語の総数}}{s \text{ における調査対象の単語の総数}}$$

ユーザが入力する値は、 $L(s)$ に対する重みを想定している。

4. 歌詞候補群の生成

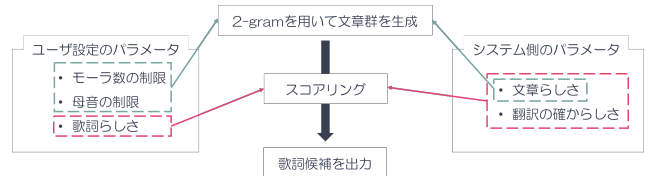


図 1: 歌詞候補群選出の手順

今回用いた歌詞候補群の生成手順を図 1 に示す。最初に、2-gram を用いて文章を生成することで文章群を得る。このとき、モーラ数の制限、母音の制限、文章らしさによって生成する文章を絞り込んだ。得られた文章群に、歌詞らしさ、文章らしさ、翻訳の確からしさを用いてスコアを付けることによって、スコアの高い上位 5 位を歌詞候補群として選出する。ユーザが設定することを想定していない、システム側で利用する文章らしさ、翻訳の確からしさの定義については後述する。本実験において、コーパスは日本語版 Wikipedia コーパスを利用した [wik 17]。

4.1 文章らしさ

2-gram の先頭の単語を w_{i-1} 、後方の単語を $w_i (i > 1)$ とし、生成文を s としたとき、2-gram 確率を $P(w_i|w_{i-1})$ 、生成文の生起確率を $P(s)$ として以下のように定義する。

$$P(w_i|w_{i-1}) = \frac{\text{利用コーパス内において } w_{i-1} \text{ の後に } w_i \text{ が続く頻度}}{\text{利用コーパス内における } w_{i-1} \text{ の出現頻度}}$$

$$s = \{w_1, w_2, w_3, \dots, w_n\}$$

$$P(s) = \prod_{i=2}^n P(w_i|w_{i-1})$$

4.2 翻訳の確からしさ

翻訳の確からしさには、原言語の機械翻訳文を利用する。単語のベクトル化には Word2Vec[Mikolov 13] を使い、次元数を 200 とした。

機械翻訳文と生成文から、それぞれ名詞と動詞を抜き出し、cos 類似度を算出する。このとき、生成文中の単語に対して、機械翻訳文中の単語から cos 類似度が最も高くなる値を求める。

生成文を s とし、生成文中の対象となる単語を w 、生成文中の単語 w に対して最も cos 類似度が高くなる単語を v 、機械翻訳文中の生成文の翻訳の確からしさを $T(s)$ としたとき、以下のよう

$$T(s) = \sum_{k=1}^n \max(\cos(\vec{w}_k, \vec{v}_k))$$

4.3 文章群の生成

文章群を生成する際、制限を設けない場合、大量の文章が生成されてしまう。そのため、モーラ数の制限、母音の制限、文章らしさ、を用いて生成する文章群に条件を課す。

機械翻訳文から重要だと思われる単語を手で 1 つ抜き出し、それを先頭とした、2-gram を取得する。取得した 2-gram の後方の単語を先頭とする 2-gram を更に取得し、指定されたモーラ数と一致するまで探索をする。このとき、探索の途中で形態素解析を行い、助詞が連続すると判断された文章は文章群の候補から外れる。また、特定の位置のモーラに対して母音の制限がある場合、条件を満たさない文章についても文章群の候補から外れる。

文章らしさは、モーラ数が増えるとともに、値が下がっていく。探索の途中において $P(s) > P_1$ を満たさなくなった文章については文章群の候補から外した。また、モーラ数と一致する文章において $P(s) > P_2$ となるものを文章群の候補とする。ここで、別の予備実験より著者らの実装では $P_1 = 0.000114849$, $P_2 = 0.000011099$ とした。

このとき、課題 3 としてあげている問題について、撥音、促音は 0 モーラとして数え、モーラ数が減る場合は考慮せず、モーラ数が指定された数と一致するものだけを考察対象とする。また、形態素解析には MeCab を用いた。

4.4 スコアリング

生成文を s とし、文章らしさを $P(s)$ 、翻訳の確からしさを $T(s)$ 、歌詞らしさを $L(s)$ として、歌詞らしさにユーザが付与できる重みを λ とし、生成文に対して以下のように $Score(s)$ を求める。この $Score(s)$ が文章群中の生成文に対して最も高い 5 つを歌詞候補群とする。

$$Score(s) = P(s) + T(s) + \lambda L(s) \quad (0.0 \leq \lambda \leq 1.0)$$

5. 実験結果と議論

実験には表 1 に示す歌詞および条件を使用した。前回の実験結果より、1 フレーズにおける単語数の平均値は 5.386 であることがわかっている。そのため、音節などにより音数が増えることを想定し 5~8 の範囲で歌詞候補の探索を行った。また、母音については、0~2 音の範囲で指定することとした。この時、先頭の単語および母音は人手で選択した。その実験結果を表 2 に示す。

今回の実験において、スコア上位 5 位に含まれる単語の全てが piapro コーパスに含まれていたため、歌詞らしさの重みによる変動が無かった。そのため、実験条件及び実験結果の歌詞らしさの重みは全て 1.0 として結果を示し、表記は省略する。

実験結果をみると、概ね同じ単語を用いた意味の歌詞候補が選出されている。

A-1 を見ると、「古い時間」という機械翻訳文に含まれる単語が 2 つ使用された文章が歌詞候補として 1 番高いスコアで選出されていることがわかる。しかしながら、これは、原言語の意味とは異なっている。一方で、A-3 の「古い話」はモーラ数は 1 増えているが、原言語の意味を反映していると言える。この 2 つを比較したとき、スコアは「古い時間」の方が高く、翻訳の確からしさについて、スコアのつけかたが不十分であることがわかった。

A-4 については、最後の助詞が変化しているだけで、全ての候補群の意味が同じとなっていた。これは、他のグループにおいても見られる問題である。助詞の選び方については、次のフ

レーズの歌詞によって変わるので、今後、複数のフレーズを考慮した歌詞候補の選出手法が必要になってくる。

A-5 に着目すると「古い話数」とあるが、これは、形態素解析により読み方を「フルイハナシスウ」と判断されたためである。モーラ数は形態素解析の結果に依存してしまい、必ずしも正しい読みにはならない結果となった。

また、A-6 を見ると、他と比較して文章群の生成数が大幅に減り、原言語とは関係の無い意味を持つ歌詞が候補としてあがっている。このことから、母音の指定数を増やすと、全く違った意味を持つ歌詞が候補としてあがるのがわかった。これは、人手の訳詞においても懸念されていた点である。

これらの検証結果より、パラメータを調整し、文章群から歌詞候補を出すことについて以下のような課題点が挙げられる。

1. 歌詞らしさは別の手段によってスコアをつける必要がある
2. 翻訳の確からしさについて、精度を上げる必要がある
3. 歌詞候補の最後の助詞について、次のフレーズを考慮した選出が必要である
4. モーラ数は形態素解析の結果に依存する
5. 原言語の意味を反映するためには、母音の限定は最低限に留める必要がある

課題点がある一方で「古い話」のように原言語の意味を反映した歌詞候補を得られることもわかった。パラメータの調整やスコアの付け方によっては、生成した文章群から歌詞候補を選択する手法は有用であると言えるだろう。

現段階で、使用したパラメータの最適解は不明であるが、クラウドソーシング等を利用し、多くのユーザからデータを集めることで、これらのパラメータの最適解を得られることが出来ると考えている。加えて、同一ユーザが繰り返しシステムを利用することで、遺伝的アルゴリズムを用いてユーザの好みを学習することができよう。

6. 関連研究

歌詞に関連する研究は数多く存在する。阿部らはモーラ数と母音を入力することで、歌詞候補文を生成し提示するシステム「pâtissier」を提案している [阿部 12b]。伊藤らは暗記のための替え歌を自動生成するシステムを実装している [伊藤 15]。このシステムでは、入力文のモーラ数を算出し、その結果から入力文を歌詞として割当可能な楽曲リストを作成し、ユーザに提示する。提示されたリストからユーザが曲を選択すると、モーラ数が完全一致しない部分について音符に変更を加えることで、入力文を歌詞とした歌曲を生成する。

また、パラメータの調整に関する研究では、クラウドソーシングを用いることで、適切なパラメータを推定し視覚デザインの探索を支援するシステムを小山らが提案している [Koyama 14]。

これ以外の関連研究もその大半は「歌詞」か「メロディ」のどちらか一方を新しく作り出すものである。具体的には、歌詞の自動生成、あるいは入力歌詞に対応するメロディの自動生成のいずれかの手法が多数提案されている。それに対して本研究では「歌詞」と「メロディ」がすでに両方存在することを前提として、形式や内容の点で整合性を保ったまま歌詞を翻訳することを目的としている。

表 1: 実験条件

グループ	原言語	機械翻訳文	実音数	指定モーラ数	先頭の単語	母音, モーラ位置
A-1	Tale as old as time	時間として古い物語	5	5	古い	指定なし
A-2				5	古い	(ア, 5)
A-3				6	古い	(ア, 5)
A-4				7	古い	(ア, 5)
A-5				8	古い	(ア, 5)
A-6				5	古い	(ア, 4), (ア, 5)

表 2: 実験結果

グループ	生成文数	結果	スコア	グループ	生成文数	結果	スコア
A-1	982	古い時間	2.00005	A-4	354	古い話を	1.61809
		古い神話	1.57048			古い話が	1.61802
		古い時分	1.44857			古い話で	1.61795
		古い詩を	1.44435			古い話は	1.61795
		古い詩の	1.44433			古い話の	1.61793
A-2	100	古い詩は	1.44426	A-5	216	古い話数	1.99134
		古い詩が	1.44424			古い個体数	1.57231
		古い詩や	1.44422			古い時代ごと	1.55306
		古い歌	1.39155			古い時代のは	1.50141
		古い値が	1.38999			古い時代のを	1.50132
A-3	324	古い話	1.41879	A-6	37	古い球	1.26576
		古い舞台	1.27864			古い穴	1.23765
		古い場合	1.26686			古い花	1.22408
		古い葉先	1.26116			古いマヤ	1.21673
		古い語り	1.20745			古い朝	1.20135

7. 今後の課題

本稿では 2-gram を用いて文章群を作成し、その中から歌詞候補を導き出す手段の検討を行った。適切な訳詞を生成できる可能性はあるものの、適切な歌詞に対して高いスコアを算出するための改良が必要であることがわかった。今後は、得られた結果を元に、さらに訳詞のための実装を行っていく。その一方で、今回の手法では、探索に時間がかかりすぎるため、今後計算量の削減や候補群の決定までの時間を短縮する必要がある。

現段階では、人手で楽曲の解析を行い、単語の抽出やフレーズの仕分けを行っているが、楽曲解析には、GTTM[Tojo 13]を利用することを考えている。

また、西欧語の多くは強弱アクセントを持つ発音が多いが、日本語は高低アクセントを持つ言語である。そのため、音高や拍節構造を考慮した日本語の発音に最適な訳詞を行いたい。

謝辞 歌詞データの収集許可を下さった、クリプトン・フューチャー・メディア株式会社に感謝致します。

参考文献

- [松田 07] 松田 直行：訳詞の誕生：近代日本音楽史における訳詞家近藤朔風の位相, 駒沢短大国文, Vol. 37, pp. 15–43 (2007)
- [松田 13] 松田 直行：「訳詞」とは何を訳すのか：近藤朔風と森外によるオペラ『オルフェウス』訳詞の比較研究のための序章, 駒澤日本文化, No. 7, pp. 165–192 (2013)
- [山本 73] 山本 金雄：歌唱における日本語の発音に関する一考察, 千葉大学教育学部研究紀要. 第 2 部, Vol. 22, pp. 233–249 (1973)

[pia 17] PIAPRO(ピアプロ), <http://piapro.jp> (2017)

[阿部 12a] 阿部 ちひろ, 伊藤 彰則 他：統計的言語モデルを用いた作詞補助システムのための歌詞特徴に関する検討, 研究報告音楽情報科学 (MUS), Vol. 2012, No. 3, pp. 1–6 (2012)

[wik 17] 日本語版 wikipedia コーパス, <http://dumps.wikimedia.org/jawiki/latest/jawiki-latest-pages-articles.xml.bz2> (2017)

[Mikolov 13] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781* (2013)

[阿部 12b] 阿部 ちひろ, 伊藤 彰則 他：patisserie-アマチュア作詞家のための作詞補助システム, 研究報告音楽情報科学 (MUS), Vol. 2012, No. 17, pp. 1–6 (2012)

[伊藤 15] 伊藤 悠真, 寺田 努, 塚本 昌彦 他：Mnemonic DJ: 暗記学習のための替え歌自動生成システム, 情報処理学会論文誌, Vol. 56, No. 11, pp. 2165–2176 (2015)

[Koyama 14] Koyama, Y., Sakamoto, D., and Igarashi, T.: Crowd-powered parameter analysis for visual design exploration, in *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pp. 65–74 ACM (2014)

[Tojo 13] Tojo, S., Hirata, K., and Hamanaka, M.: Computational Reconstruction of Cognitive Music Theory, *New Generation Computing*, Vol. 31, No. 2, pp. 89–113 (2013)