

楽曲構造に着目した統計的機械翻訳に基づく 作詞支援

岩下 知美¹ 齋藤 大輔² 峯松 信明¹ 広瀬 啓吉²

概要: 近年では、DTM 用ソフトウェアや Vocaloid 等の歌声合成ソフトの普及により、特別な知識や機材を持たない人でも容易に自作歌唱曲の作成を行うことができるようになった。歌唱曲を制作する上においては作曲・作詞・演奏等の技術が必要であるが、特に作詞は作曲や演奏と比較して学習方法や支援が少なく、歌唱曲を制作する上で難点となっている部分と考えられる。本研究では、作詞のうち特に 1 番から 2 番以降を作詞する過程に着目し、2 番の歌詞を作詞する際に 1 番の歌詞を制約として含めるために、2 番の歌詞を 1 番の歌詞を翻訳元として統計的機械翻訳により生成することによって作詞支援を行う手法を提案する。

1. はじめに

近年、パソコンなどを用いて楽曲の制作を行う DTM (DeskTop Music) 用のソフトや Vocaloid などの歌声合成ソフトの普及を背景に、動画投稿サイト等における自作歌唱曲の発表が増加している。デジタル技術の発展により、先述したソフトが普及し、特別な知識や入手困難あるいは高価な機材を持たない人でもパソコンさえ所有していれば容易に自作曲の作成が行えるようになった。

自作歌唱曲を発表しているのは主に商業的なコンテンツ生成に携わらない一般の消費者である。このように消費者が制作するコンテンツは消費者生成メディア (Consumer Generated Media; CGM) と呼ばれ、今後さらに増加の傾向が顕著になると考えられている。そこで、以前にも増して作曲・作詞・演奏の支援を行う必要が出てきた。作曲・演奏支援の例としては、Orpheus が挙げられる [1]。

しかし、作曲理論や演奏技術と比較すると、作詞は学習方法が少なく、歌唱曲の制作を行う上で難関となっている消費者が多い部分と考えられる。そのため、様々なユーザの意図する歌詞を生成する補助システムが考案されてきた。例えば、作曲補助システム [2] や、形態素リスト上の経路探索による自動作詞 [3] などの研究がすでに行われている。これらの研究は、曲に歌詞を付ける曲先という手法で、作詞を行う際の制約として単語のモーラ数、韻、アクセント等の要素を上げていた。しかし、ともに対象にしている作詞は行単位であり、番などのより大局的な楽曲構造を製作することは想定されていなかった。大局的な楽曲構造を考

慮した研究として渡邊らの研究が挙げられる [4]。この研究では、単語からフレーズ (単語の集合) を生成するモデル、フレーズから行 (フレーズの集合) を生成するモデル、行からブロック (行の集合) を生成するモデル、そしてブロックから番 (ブロックの集合) を生成するモデルで構成されるモデルを考え、これらのモデルを用いることによって 1 番全体の歌詞を自動生成することを行っていた。この研究においては、2 番以降の歌詞の生成は今後の課題として挙げられていたにとどまる。

自動作詞ならびに作詞支援は、基本的にモーラ数などの構造的な制約条件に基づいて言語を生成する技術と考えられる。一方、一般にポップスなどの歌唱曲は同一のメロディーを用いた繰り返し構造を有する。楽曲全体の制作過程を考えた場合、1 番と 2 番の歌詞は構造的に同一になるが、異なる文章となる必要がある。本研究はこのような制約条件のもとで、1 番の歌詞を作成後、2 番以降の歌詞を作詞する過程の支援を検討する。この際、考慮すべき制約には上記の要素の他に 1 番の歌詞そのものが含まれると考えられる。この制約を取り入れるために、2 番の作詞を 1 番の歌詞から 2 番への翻訳として定式化し、統計的機械翻訳に基づいて文章を生成する手法を提案する。この際、翻訳の最小単位として 1) 単語単位での翻訳 および 2) モーラ単位での翻訳 の二つについて検討を行った。

2. 統計的機械翻訳

統計的機械翻訳は、翻訳を翻訳文を S 、原文を T とすると、 T が与えられた時に翻訳文が S となる確率 $P(S|T)$ を最大にする S を求める問題として解く機械翻訳である [5]。ベイズの定理により、 $P(S|T)$ は以下の様に表すことが出

¹ 東京大学大学院工学系研究科電気系工学専攻

² 東京大学大学院情報理工学系研究科電子情報学専攻

来る.

$$P(S|T) = \frac{P(S)P(T|S)}{P(T)} \quad (1)$$

式(1)の分母は S に非依存であるので、前述の問題は $P(S)P(T|S)$ を最大にする S を選ぶ問題と言い換える事ができる.

$P(S)$ は生成された翻訳文がどれほど文らしいかを表すパラメータである. これは前述の N -gram 言語モデルで表す. $P(T|S)$ は翻訳モデルであり, 対訳コーパスからパラメータ推定を行う. パラメータには原文に出現する一単語 e が翻訳文では n 単語になる確率 $P(n|e)$ や, e が翻訳文では f という単語で表される確率 $P(f|e)$, 翻訳文での単語の出現位置 i と原文での単語の出現位置 j の関係を翻訳文での単語数 l を用いて表す $P(i|j, l)$ がある.

3. 統計的機械翻訳を用いた歌詞の生成

先行研究 [2] では, モーラ数だけ与えられているという前提での作詞支援を考えていた. しかし, 楽曲は多くの場合 1 番だけではなく 2 番以降がある. 2 番以降は 1 番の歌詞を踏まえて作成されていることが多い. つまり, モーラ数の制約に加えて 1 番を踏まえるという制約が加わるので, 2 番の作詞が 1 番の作詞よりも困難な場合がある. 本章では, 統計的機械翻訳を用いて 1 番の歌詞から 2 番の歌詞の作成支援を試みた実験に関して述べる. 統計的機械翻訳では対訳コーパスから翻訳モデル, 翻訳先のコーパスから言語モデルを学習する.

歌詞において 2 番以降では, 1 番と同一のメロディーであるため, 結果的にその歌詞についても 1 番と似た傾向を持つと考えられる. そこで, 1 番の歌詞を入力とした際に 2 番の歌詞を生成することを考える. この問題は, 1 番の単語系列を $X = (x_1, x_2, \dots, x_T)$ とした時に同様の構文系列 $Y = (y_1, y_2, \dots, y_T)$ を $P(Y|X)$ の確率に基づいて生成する問題として捉えることができる. この問題を, 翻訳元を 1 番の歌詞, 翻訳先を 2 番の歌詞として統計的機械翻訳ツール Moses を用いて実装し, 解くことによって作詞支援を試みた [6].

4. 翻訳実験

4.1 実験データの作成

RWC 研究用音楽データベース: ポピュラー音楽 [7] の 80 曲のうち, 日本語歌詞を持つ 79 曲を 1 番と 2 番に手作業で分割した. この際, 1 曲については 1 番の歌詞しか存在しなかったため, 最終的に 78 曲を学習データとして用いた.

歌詞においては, 構成が異なるなどの理由で 1 番と 2 番のフレーズが一対一対応しないことがある. そこで今回は次のような規則のもとで対訳コーパスを作成した.

- 原則として 1 番と 2 番の歌詞のみを用い, 3 番以降は用いない (1 番の繰り返し等が多いため)

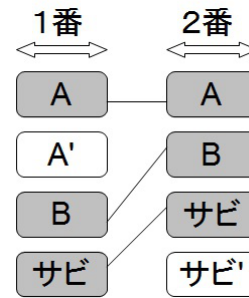


図 1 対訳コーパスに用いるフレーズの選定

- ただし, 3 番と 4 番が存在し, かつどちらも 1 番とも 2 番とも異なる場合はそれぞれ 1 番と 2 番として扱う
- 片方のみで同じフレーズが繰り返される時は, 最初のフレーズのみを用いる (図 1 の 1 番 A' と 2 番サビ' はコーパスに入れない)

例えば図 1 の場合, 線で結ばれている灰色の四角で示されるフレーズを対訳コーパスとした. ただし, 1 番と 2 番の歌詞が完全に一致する場合は除いた. なお, 対訳コーパスは形態素解析エンジン mecab[8] で分かち書き処理を施してある. この対訳コーパスを C とする. この対訳コーパスは, 1 番と 2 番が単語単位で同一の楽曲構造を持つという仮定のもと構築されたコーパスと考える事ができる.

さらに, C を mecab を用いて読みを与え, モーラを一対一対応させた対訳コーパスを手作業で整備した. この際, 「フタリ」と読むべき「二人」を「ニンニン」にする等の mecab で正しく読みが与えられなかった単語は正しい読みにした. また, 1 番と 2 番でわずかにメロディーが異なるなど, モーラを一対一対応させることが不可能な場合があった. その場合には一対一対応しないモーラのみを削除した. これにより単語が一部欠落している場合がある. この対訳コーパスを D とする. この対訳コーパスは, 1 番と 2 番において, メロディーが同一で, モーラ単位での出現に何らかの変動が存在するという仮定のもと構築されたコーパスと考える事ができる. コーパス C に比べると, 日本語的に意味の整合性がとれた文章は生成されにくい, 作詞支援の観点から, 楽曲構造を保存する事を重視しているといえる.

4.2 言語モデルの学習

N -gram 言語モデルを学習する際 N の値を変えて学習を行い, Moses にて翻訳モデルの学習を行った後統計的機械翻訳を実行した結果を比較した. 統計的機械翻訳において, 言語モデル $P(S)$ は, 特定ドメインの言語らしさを反映するものであるため, 用いる学習コーパスの制約は少ない. 本実験では単語単位のモデル化 (コーパス C) については, 2 番のデータを, モーラ単位のモデル化 (コーパス D) については, 1 番のデータを用いた.

```
あなたお願いよ席を立たないで  
息がかかるほどそばにいてほしい  
あなたが好きなんです  
ひとりでいるのがこわくなる  
このまま逢えなくなりそうで  
くちづけさえ知らないけど  
これが愛なのね
```

図 2 「ロマンス」単語単位でのわかち書き歌詞

```
アタイテイヨオキミタタナイデ  
イキカナイワラニカナイテオシイ  
アタヲナイマデ  
タシテイルノサシテクナイ  
コノママハアナクアウデ  
キヅイタワラシラナイタイ  
コシアイナニト
```

図 5 $N = 2$ での出力結果

```
あなたお願いで席を立たないで  
息がかかるほどそばにもてほしい  
あなたが好きなんです  
ひとりでいるのがこわくて  
がの逢えの頃ので  
くちづけさえ知らない時  
そのがになのて
```

図 3 単語単位での翻訳により生成された結果

```
アナタイテイルダレモシラナイデ  
イキスカルワラワカナイオシエテ  
アナタヲナンテク  
イマイデイルノガシテクナル  
コノママナントナクナイワケ  
キヅイタワラシイエナイデ  
コンデアイナノテ
```

図 6 $N = 3$ での出力結果

```
アナタオネガイヨセキヲタタナイデ  
イキガカカルホドソバニイテホシイ  
アナタガスキナンデス  
ヒトリデイルノガコクナル  
コノママアエナクナリソウデ  
クチヅケサエシラナイケド  
コレガアイナノネ
```

図 4 「ロマンス」モーラ単位でのわかち書き歌詞

```
アナタイテイルダレモシラナイデ  
イキルワタシノガワニイテワナイ  
アナタヲナンデマ  
イマイデイルノサソクナル  
コノママナントナクアルイテ  
キヅイタワラシラナイデイ  
コイガアイナノテ
```

図 7 $N = 4$ での出力結果

4.3 実験結果

岩崎弘美「ロマンス」の冒頭七行を入力として学習にコーパス C, D をそれぞれ用いた統計的機械翻訳を行った。

まず、コーパス C を用いて統計的機械翻訳を行った結果を図 3 に示す。「ロマンス」の歌詞は、図 2 の通り入力を行う前に mecab によって単語単位でのわかち書きを行っている。 N を 2 から 5 まで変化させながら N -gram 言語モデルを作成し、統計的機械翻訳を行ったが出力結果はすべて同じであった。

翻訳結果では、翻訳元に出現する単語が高い確率で同じ場所に出現した。「席」など学習コーパスに出現しない単語が翻訳されなかったこと、逆に「あなた」や「好き」など頻出するが 1 番と 2 番で同じ単語が対応する頻度が高い単語がやはり翻訳されなかったことが要因として考えられる。この問題を解決するためには、よりヴァリエーションに富んだ単語及びフレーズを含むコーパスが必要だと考えられる。

次にコーパス D を用いて統計的機械翻訳を行った。「ロマンス」の歌詞は、まず mecab によって読みを出力した後修正しく読みが与えられなかった単語を修正し、図 4 の状態で入力を行った。 $N = 2, 3, 4, 5$ と変えながら N -gram 言語モデルを作成し、統計的機械翻訳を行った出力結果を図 5 から図 8 に示す。

こちらの翻訳結果は、「あなた」のように歌詞コーパスにおいて極端に多く出現する単語は同じ場所に残るものの、

```
アナタイテイルダレモシラナイデ  
イキルワタシノガワニイテワナイ  
アナタヲナンデマ  
イマイデイルノサソクナル  
コノママナントナクアルイテ  
キヅイタワラシラナイデイ  
コイガアイナノテ
```

図 8 $N = 5$ での出力結果

文章全体としては翻訳元の歌詞とは異なるものになっている。モーラ単位で翻訳を行ったことによりモーラの接続が単語にならない可能性もあったが、 N -gram 言語モデルにより単語になる接続が出現しやすくなっていると考えられる。単語とならない接続でも、単語の一部であったり既存の単語に近い接続であったり、人間がある単語を連想しやすい接続となっていた。

以上それぞれ単語単位とモーラ単位による統計的機械翻訳を行った。2 番以降の作詞支援を行うことを考えると、当然 1 番の歌詞に似通った文章では作詞支援とはならない。よって、比較的翻訳元の歌詞とは異なる文章が生成されたモーラ単位での統計的機械翻訳を支援に用いるのがふさわしいと考えられる。

5. 評価実験

5.1 実験方法

4.3 節で行ったモーラ単位での統計的機械翻訳がどの程度作詞支援に有効かを調査するために評価実験を行った。

1.
想うまま叶うなら今すぐ
そっと気持ち届けたい

2.
普通の毎日が何よりも満足なんだったって思えた
そんな日々は今は遠い昔に感じた…

3.
幼い頃描いた
夢をあなたは覚えてる？

図 9 評価実験に用いる歌詞

1.
同じ想いでいられるならまだ
きっと二人でいられた

2.
過去の人は今は居ないけどナイスで普通の毎日と
なによりもなんか変な人も今はなつかしいな

3
確かな今の時を
後であなたは思い出す

図 12 表 10 を見ながら作詞した歌詞

1.
オモイママデナキナライママコ
キットキモチレテイタ

2
フツウノマイニチガナニヨリモ
ナントクナンダッテオモイタ
カンナヒトハイマハイツイヒュカシイナ

3.
ナニカナイコノハイタ
アッテフアナタハオモイテル

図 10 評価実験に用いる歌詞に統計的機械翻訳を施した文章

1
もう二度と会えなくてもまだ
キミの姿追いかける

2.
今輝いていた思い出に別れを告げて走りだそう
水色の空は今は高く秋風がざわめく

3.
ふたりで描いた夢を
今は一人で追いかける

図 11 歌詞以外を見ずに作詞された歌詞

	歌詞のみ	統計的機械翻訳あり
被験者 A	約 50 分	約 20 分
被験者 B	約 20 分	約 15 分

まず、学習データに含まれていない日本語歌詞付きの楽曲を 3 曲選定し、これらの A メロ 8 小節に関して $N = 5$ のモーラ単位での統計的機械翻訳を行った (図 9 と図 10)。次に、用いる楽曲を聞いたことが無く、作詞経験も無い 2 人の被験者に 2 番の A メロ 8 小節の作詞を行ってもらった。

5.2 実験結果

2 名の被験者のうち、被験者 A が作成した歌詞を図 11 と図 12 に示す。一曲作詞を行うのににかかった平均時間は表 1 のようになった。作詞の所要時間に関しては短縮の傾向が見られる。これは作詞に慣れたためとも考えられるが、

被験者から「統計的機械翻訳を行った文章中に参考となる単語があったので作詞が行いやすかった」との意見があったので、統計的機械翻訳を行った文章は作詞の参考になりうることを示唆される。

2. の歌詞は他の二曲に比べてモーラ数が多かったが、この曲に関しては双方の被験者から「他二曲と比較して、統計的機械翻訳を行った文章があまり文章になっていなかったため作詞が困難だった」との意見があった。文章の整形をどのようにして行くかは今後の課題としたい。

6. おわりに

本研究では、統計的機械翻訳を用いた歌詞の生成を試みた。作詞は 1 番だけではなく 2 番以降も作成する場合があることに着目し、1 番から単語単位及びモーラ単位での統計的機械翻訳を用いて 2 番の歌詞を生成し、支援を試みた。提案法による支援によって、作詞時間の短縮を行えることが示されたが、統計的機械翻訳を行うと単語として解釈が難しいモーラ列が生成されてしまうことがあり、その場合支援が難しいということが発覚した。

今後の展望としては品詞を含む N -gram 言語モデルと統計的機械翻訳を統合した作詞支援システムの構築を行いたい。 N -gram 言語モデルから 1 番を、統計的機械翻訳を用いて 2 番以降の作詞を同時に支援することができれば、歌唱曲の作成がより容易になるはずである。

また、モーラ数やアクセントの指定を行うのに、曲先であることを利用してメロディの MIDI (Musical Instrument Digital Interface) データを利用することも検討したい。MIDI データには曲の音符数や音の高低だけでなく音の長さや強弱も含まれているので、比較的長い音符には長音を当てるようにする、強く発音する音符には名詞や動詞、弱く発音する音符には助詞を当てるなど MIDI データから制約を与えることも出来るであろう。MIDI データを使用することによって、自動作詞を行う際のユーザに指定させる項目が少なくなりユーザの負担を減らすことが期待できる。しかし、この場合にはフレーズの区切りをどう判定するかという問題が生じる。ユーザに指定させる、適当な音数で区切るなどの方法が考えられるので、ふさわしい方

法を検討する必要がある.

参考文献

- [1] Orpheus Ver. 3.8 自動作曲システム オルフェウス,
<http://www.orpheus-music.org/v3/>
- [2] 阿部ちひろ, 伊藤彰則, “pâtissier-アマチュア作詞家のための作詞補助システム-”, 研究報告音声言語情報処理, Vol.2012-SLP-90 No.17, pp. 1-6, 2012.
- [3] 堀玄, 齋藤大輔, 嵯峨山茂樹, “形態素リスト上の経路探索による自動作詞”, 計量国語学, Vol.2013-29(3), pp. 104-106, 2013.
- [4] 渡邊研斗, 松林優一郎, 乾健太郎, 後藤真孝, “大局的な構造を考慮した歌詞自動生成システムの提案”, 言語処理学会, 言語処理学会第20回年次大会発表論文集, pp.694-697, 2014.
- [5] Peter F. Brown et al., “A Statistical Approach to Machine Translation”, Computational Linguistics Volume 16, Number 2, pp. 79-85, 1990.
- [6] Moses – Statistical Machine Translation,
<http://www.statmt.org/moses/>
- [7] RWC 研究用音楽 DB,
<https://staff.aist.go.jp/m.goto/RWC-MDB/rwc-mdb-p-j.html>
- [8] 工藤拓, 山本薫, 松本裕治, “Conditional Random Fieldsを用いた日本語形態素解析”, 情報処理学会研究報告, Vol.2004(47), pp. 89-96, 2004.